Identifying Genes Involved in Beta Cell Dysfunction Using High-Throughput Genetic and Genomic Approaches

Antje Kristin Grotz

St Anne's College

Oxford Centre for Diabetes, Endocrinology & Metabolism

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity Term 2020

Declaration

The data and interpretation presented in this thesis is my own work with contributions from others as outlined below. Unless otherwise stated, experiments were performed at the Oxford Centre for Diabetes, Endocrinology and Metabolism (OCDEM) at the University of Oxford under the supervision of Professor Anna Gloyn with input from Daniel Ebner.

Funding for this project was provided by the Wellcome, the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) and the Radcliffe Department of Medicine, University of Oxford.

CRISPR screen optimisation experiments (Chapter 4) and lentiviral production (Chapter 3) were performed in collaboration with Dr Elena Navarro Guerrero at the Target Discovery Institute. EndoC-βH1 culture during the CRISPR screen (Chapter 4) was supported by Sameena Nawaz. FACS staining of the CRISPR screen was performed by Dr Elena Navarro Guerrero and Dr Roberta Baronio. FACS sorting of the CRISPR screen was conducted at the Wellcome Centre for Human Genetics (WCHG) with support from Dr Ruddy Montandon. Lentiviral CRISPR library generation, DNA extraction and PCR-based amplification of screening samples was performed at the Target Discovery Institute by Dr Elena Navarro Guerrero and Dr Roberta Baronio. Sequencing data was generated by the Oxford Genomics Centre at the WCHG and initial data analysis was supported by Dr Agata Wesolowska-Anderson.

Immunofluorescence studies (Chapter 5) were performed at the Department of Developmental Biology and Medicine at Stanford University under the supervision of Prof Seung Kim and Dr Romina Jimena Bevacqua.

This thesis is approximately 46 000 words and has only been submitted to the University of Oxford for the degree of Doctor of Philosophy in Medical Sciences and no part has been submitted for any other degree at this or any other university.

Acknowledgements

The last four years have been the most inspiring and rewarding but also challenging journey for me personally and scientifically. I am immensely grateful for the support, encouragement and help of my friends, family and colleagues. First, I would like to thank my supervisor Anna Gloyn for giving me the opportunity to do this DPhil and even though me joining the lab was initially a surprise to both of us, this group and project were the best choice I could have made. Your support, guidance and your trust and confidence in my abilities to independently conduct my research have made me a better scientist. I would also like to thank Daniel, Elena and Roberta for their support and technical expertise which made the screen possible and the Radcliffe Department of Medicine for their financial support.

Thank you to everyone in the Gloyn and McCarthy group: Agata, Anubha, Benoit, Carla, Emily, Grace, Jason, Mahesh, Mandy, Mark, Marta, Robert, Sameena and Sara for being a great team, creating a stimulating research environment and lots of laughter along the way. In particular, I would like to thank Soren for developing the project with me, Vibe for your calm and clear advice, Fernando for your patience and phenomenal support at all times (while only making me cry once) and Nicole for amazing California road trips and scientific discussions. I am also very grateful to all my colleagues at OCDEM who made it a great place to work even though it never had a normal temperature, thanks to my fellow tissue culture colleagues for bearing with me during my screen and making it fun to look after cells on Saturdays and to Oxlip for adopting me during my final year for coffee and lunch.

My DPhil would not have been such a fun time without my fellow OCDEM students and especially our HPODS committee: Claire, Joely, Katia, Pippa, Shahana and our honorary member Choya, from college formals to 'getting out of control' while setting up OCDEM's most legendary office decoration. Special thanks goes to my partner in crime Shahana, with whom I laughed, cried and ranted in the western room, epically failed to set up minipreps or pH anything and had the best time in the lab based on our motto to just wing it.

Outside of work, I would like to thank my Oxford family and fellow Stockmore Street tenants: Gabi, Lisa, Nicole, Kostas and Hippo. You made Oxford feel like home and I could not have wished for better housemates to enjoy power cuts with, exceed my bedtime with Codenames nights, be deeply committed to Murder Mysteries and complain how hard adulting is. I would also like to thank the great St. Anne's College MCR community who made me feel welcome from day one and where I have made friends from all over the world. I am also very grateful to my friends from back home especially Sabi and Maira who have remained my closest friends since high school despite the distance and who were always happy to come and visit wherever I am.

I am forever thankful to my family, my parents Armin, Catrin, my brother Tim and our dogs Loui and Ella who have always been there for me and from a young age gave me the freedom to explore the world while patiently supporting me and always welcoming me back home. Thank you for always believing in me and raising me to become the person I am today.

My final and biggest thank you goes to my partner Flo, who deserves this success as much as I do. Your unconditional love and support have been my source of strength throughout these years despite living in different countries. Thank you for always putting up with all my western blot talk, being the reasonable of us two and especially for supporting me to chase my dreams while always being by my side. This DPhil would not have been possible without you.

Abstract

Genome-wide association studies have identified >400 signals which are robustly associated with type 2 diabetes (T2D) risk, a large proportion of which are involved in pancreatic beta cell dysfunction. A major barrier to translational insight however is uncertainty over the genes through which they exert their effect. The aim of this thesis was to perform the first genome-wide pooled CRISPR screen in human beta cells and use a high-throughput and unbiased approach to identify candidate causal genes and novel regulators of human beta cell function.

The cell line EndoC-βH1 is the most authentic cellular model of human beta cells but it is a slow growing and technically challenging cell line which has not been established for genome editing yet. Before approaching the genome-wide screen, I therefore developed the first CRISPR/Cas9 editing pipeline in EndoC-βH1 using a dual sgRNA and lentiviral based approach. As demonstrated for six genes, this strategy induced highly efficient genome editing resulting in complete protein depletion and functional gene knockout. Distinct phenotypes upon comparison with an siRNA-based knockdown strategy highlighted the need for complementary validation of individual results to avoid method specific functional interpretations.

After extensive optimisation of all screening parameters, I performed two independent genome-wide pooled loss-of-function CRISPR screens for intracellular insulin in EndoC-βH1 using a FACS based readout. The screen demonstrated an excellent technical performance and was able to identify functionally relevant hits as shown through identification of known regulators of human beta cell function such as *INS, NKX2.2, SLC2A2, G6PC2* or *GCK*, a strong enrichment of T2D, MODY and beta cell pathways and protein-protein networks known to play a crucial role in the beta cell. In total, 580 genes were classified as having a highly reproducible effect on insulin content. Integration with predicted T2D effector genes whose mechanistic role has not been resolved yet, prioritised *C2CD4B, FADS1* and *CALCOCO2* for functional follow-up studies.

Finally, individual functional validation studies confirmed the phenotypic effects from the screen for several genes including for *CALCOCO2,* where silencing of the gene reduced insulin content and secretion in EndoC-βH1, therefore establishing it as a novel regulator of beta cell function and potential causal gene at the *TTLL6* GWAS locus. Cellular localisation studies within primary human islets further confirmed that CALCOCO2 is highly expressed across endocrine and exocrine pancreatic cell types.

Overall, the work presented in this thesis has addressed the outstanding challenge of functionally investigating all potential T2D effector genes in human beta cells and has generated a valuable resource and numerous hypotheses for future studies. The biological insights gained from this work have contributed to a better functional understanding of T2D pathogenesis and predisposition and will inform future therapeutic strategies.

Abbreviations

ANOVA	Analysis of variance
ATAC	Assay for transposase-accessible chromatin
ATP	Adenosine triphosphate
BMI	Body mass index
BSA	Bovine serum albumin
cAMP	Cyclic adenosine monophosphate
ChIP	Chromatin immunoprecipitation
CPM	Counts per million
CRISPR	Clustered regularly interspaced short palindromic repeats
CRISPRa	CRISPR activation
CRISPRi	CRISPR interference
dCas9	Dead Cas9
DSB	Double-strand break
EM	Electron microscopy
eQTL	Expression quantitative trait locus
ER	Endoplasmic reticulum
EV	Empty vector
FACS	Fluorescence activated cell sorting
FDR	False discovery rate
FE	Fold enrichment
GO	Gene ontology
GoF	Gain-of-function
GPCR	G-protein coupled receptor
GSEA	Gene set enrichment analysis
GSIS	Glucose stimulated insulin secretion
GWAS	Genome-wide association study
HbA1c	Glycated haemoglobin
HCL	High confidence gene hit list
HDR	Homology directed repair
HLA	Human leukocyte antigen
hPSC	Human pluripotent stem cells
Indel	Insertions and deletions
iPSC	Induced pluripotent stem cells
K_{ATP}	ATP-sensitive potassium channel
KD	Knockdown
kDa	Kilodalton
KEGG	Kyoto Encyclopedia of Genes and Genomes
KO	Knockout
LD	Linkage disequilibrium
LFC	Log2 fold change
lncRNA	Long non-coding RNA

LoF	Loss-of-function
MAF	Minor allele frequency
MAGeCK	Model-based analysis of genome-wide CRISPR-Cas9 knockout
Mb	Megabase
MODY	Maturity onset diabetes of the young
MR	Mendelian randomization
NDM	Neonatal diabetes mellitus
NES	Normalised enrichment score
NHEJ	Non-homologous end joining
NMD	Nonsense-mediated mRNA decay
NT	Non-targeting
OR	Odds ratio
PAM	Protospacer adjacent motif
PBS(T)	Phosphate buffered saline
PCA	Principal component analysis
PP	Pancreatic polypeptide
P/S	Penicillin-Streptomycin
PSC	Premature stop codon
QC	Quality Control
RNAi	RNA interference
RPKM	Reads per kilobase of transcript, per million mapped reads
SD	Standard deviation
SEM	Standard error of the mean
Seq	Sequencing
sgRNA	Single guide RNA
shRNA	Short hairpin RNA
siRNA	Small interfering RNA
SNP	Single nucleotide polymorphism
TIDE	Tracking of Indels by DEcomposition
TKOv3	Toronto Knockout CRISPR library version 3
tracrRNA	Trans-activating crispr RNA
T1D	Type 1 Diabetes
T2D	Type 2 Diabetes
UPR	Unfolded protein response
WT	Wildtype

Table of Contents

1

Introduction

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

Grotz AK, Gloyn AL and Thomsen SK. Prioritising Causal Genes at Type 2 Diabetes Risk Loci. *Current Diabetes Reports* 2017; 17(9): 76

1.1 Type 2 Diabetes

1.1.1 A global pandemic

Type 2 Diabetes (T2D) affects more than 410 million adults worldwide, having evolved from a disease most prevalent in "Western" countries to a common disease in most parts of the world (International Diabetes Federation, 2019). According to the International Diabetes Federation, 4.2 million people died in 2019 from diabetes and its complications and around 90% of all diabetes cases can be attributed to people living with T2D. Diabetes related annual global expenditures are estimated to rise further from currently 760 billion USD or 10% of the total health care spending and place a significant economic burden on health care systems and individuals (International Diabetes Federation, 2019). T2D can be classified as a global pandemic and is expected to be one of the largest global health challenges of the 21st century. The dramatic increase in T2D prevalence is largely due to a change towards an unhealthy western diet and sedentary lifestyle, however, a significant proportion of disease risk (30-70%) can also be attributed to someone's genetic predisposition (Franks, 2012; Willemsen *et al.*, 2015; Mahajan *et al.*, 2018a).

1.1.2 Disease pathogenesis

T2D is a heterogeneous disorder characterised by prolonged hyperglycaemia from underlying defects in glucose homeostasis. Normal glucose homeostasis to maintain steady-state plasma glucose levels is regulated through a close feedback loop involving glucoregulatory hormones such as the key opposing regulators insulin and glucagon as described in Figure 1.1 (Cerasi & Luft, 1967; Curry, Bennett, & Grodsky, 1968; Blackard & Nelson, 1970; Lins *et al.*, 1983; Freychet *et al.*, 1988). Glucose production in the liver during the fasting state is mediated via glucagon (Unger, 1971; Gerich *et al.*, 1974; Lins *et al.*, 1983). Glucose absorption into skeletal muscle and adipose tissue after feeding and

simultaneous suppression of glucose production is mediated by insulin (Kohn *et al.*, 1996; Zisman *et al.*, 2000; Henquin *et al.*, 2015). Insulin and glucagon are both produced within the pancreatic islets of Langerhans, in beta and alpha cells, respectively. In addition to alpha and beta cells, a small proportion of the islet also consists of delta, gamma cells and epsilon cells which secrete the hormones somatostatin, pancreatic polypeptide and ghrelin, respectively and negatively regulate insulin and glucagon secretion under different conditions to fine-tune glucose homeostasis as shown in Figure 1.1 (Weir *et al.*, 1979; Göpel *et al.*, 2000; Broglio *et al.*, 2001; Egido *et al.*, 2002; Hauge-Evans *et al.*, 2009; Aragón *et al.*, 2015). Other hormones such as the incretin hormones GLP-1 and GIP from the gut also play important roles in maintaining stable glucose levels (Perley & Kipnis, 1967; Nauck *et al.*, 1997).

In T2D, glucose homeostasis is dysfunctional and a downwards spiral involving increased insulin resistance and impaired pancreatic beta cell function has developed (Kahn *et al.*, 1993; Kahn, Cooper, & Del Prato, 2014; Weyer *et al.*, 1999; Stancakova *et al.*, 2009). Insulin resistance develops when excess lipids accumulate in insulin-responsive tissues for instance during obesity, leading to impaired glucose tolerance. Up to a certain extent, this decreased insulin sensitivity can be compensated for by increasing insulin secretion from beta cells to prevent the progression to T2D (Weyer *et al.*, 1999; Stancakova *et al.*, 2009). If beta cells are incapable of increasing their output due to an underlying defect or progressively deteriorate over time due to the increased demand, T2D develops (Butler *et al.*, 2003; Rahier *et al.*, 2008).

Major complications in T2D are due to chronic hyperglycaemia and consequential damage to blood vessels. Complications are very are common in T2D with half of all patients having macrovascular complications affecting the kidney, the eye and the nervous system and a

third having microvascular complications such as cardiovascular disease (Litwak *et al.*, 2013; Zheng, Ley, & Hu, 2018).

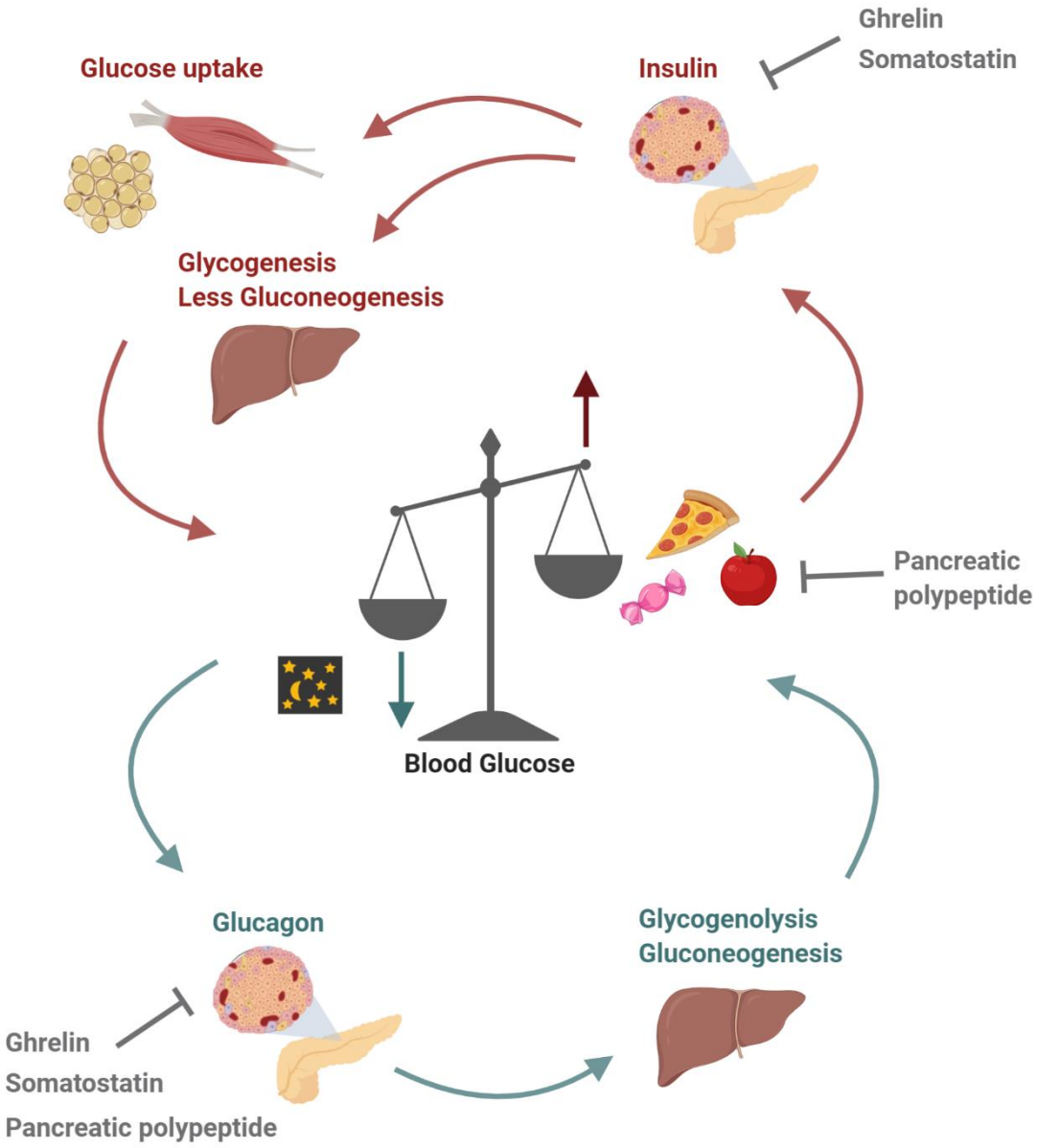


Figure 1.1 - Pancreatic regulation of glucose homeostasis. Simplified schematic depicting how the opposing hormones insulin and glucagon maintain balanced glucose levels in the blood within 4-6 mM. After feeding *(red)*, insulin is released from pancreatic beta cells which promotes glucose uptake in adipose and muscle cells and induces increased glycogenesis and decreased gluconeogenesis in the liver to reduce blood glucose levels. During times of low blood glucose levels *(turquoise)*, glucagon is released from pancreatic alpha cells and promotes glycogenolysis and gluconeogenesis in the liver to increase endogenous insulin blood glucose levels (adapted from Röder *et al.*, 2016).

1.1.3 Current therapeutic approaches are not disease modifying

Initially, lifestyle interventions including diet and exercise changes, primarily aimed at weight reduction are attempted to manage T2D (Marín-Peñalver *et al.*, 2016). Drastic weight-loss interventions have shown to be effective in managing T2D by promoting sustained disease remission in 46% and significant weight loss in 24% of patients within a year (Lean *et al.*, 2019). Even modest but sustained weight-loss can improve glycaemic control and delay the progression towards T2D (Khavandi *et al.*, 2013). Nevertheless, lifestyle interventions alone are usually unsuccessful in managing T2D long term and patients adopt oral and injectable pharmacological regimens to maintain stable blood glucose levels. If the patients have no contraindications, Metformin is the first line treatment (Campbell, White, & Saulie, 1996; Inzucchi *et al.*, 2015). Metformin controls hyperglycaemia through multiple complex modes of actions including but not limited to activation of AMP-activated protein kinase (AMPK) in hepatocytes to lower gluconeogenesis, increase glucose utilisation and alter the microbiome in the gut (Zhou *et al.*, 2001; Rena, Hardie, & Pearson, 2017). The anti-diabetic drug has also prevents the development of T2D in people at high risk by increasing the circulating levels of GDF15 which results in reduced food intake and lower body weight (Coll *et al.*, 2020). Beyond Metformin, a wide range of oral and injectable second and third-line pharmacological agents have been developed to find the best treatment strategy for a patient's personal risk and disease profile. In addition to mediating glycaemic control, newer treatment regimens such as GLP1 receptor agonists (GLP1-RA) and sodium glucose cotransporter 2 (SGLT2) inhibitors, which respectively mimic the action of GLP1 and reduce renal reabsorption of glucose have also shown to be beneficial for T2D associated comorbidities, which is crucial for successful disease management (Scheen, 2015; Wilding, Rajeev, & DeFronzo, 2016; American Diabetes Association, 2020a). Large scale clinical trials demonstrated that

treatment with both classes of drugs was beneficial in T2D patients with high cardiovascular risk and reduced associated hospitalisation and mortality while also better controlling blood glucose levels (Scheen, 2015; Marso *et al.*, 2016; Zinman *et al.*, 2018). Eventually, even after combinatorial drug treatments have not been successful in managing T2D long term, many patients progress to insulin therapy (Centers for Disease Control and Prevention, 2020). Insulin injections are an effective treatment approach, it is however associated with severe risks and side effects such as increased episodes of hypoglycaemia and weight gain.

All currently available therapies to manage T2D are aimed at treating the arising conditions or complications of the disease but are not disease modifying or targeted at the underlying pathophysiology. It is therefore crucial to better understand the disease and the functional mechanisms underlying T2D to inform drug development and be able to monitor disease progression more efficiently through novel biomarkers. Discovering the genetic susceptibility to T2D and identifying the causal genes for the disease can help in understanding these underlying molecular mechanisms and are therefore critical in guiding future T2D therapy and prevention. The resulting personalised or stratified treatments, with patients divided into subclasses are predicted to be more effective with fewer side effects (Fitipaldi *et al.*, 2018).

1.2 The genetic landscape of diabetes

Diabetes is a group of metabolic disorders traditionally grouped into Type 1 Diabetes (T1D), monogenic forms of diabetes and T2D. Due to the heterogeneity of the diseases with varying disease characteristics and progression, a definitive classification is often challenging and misdiagnosis is common. The primary criteria to classify the type of diabetes is age of onset followed by determining distinct disease characteristics such as screening certain genes for mutations common in monogenic diabetes or the presence of specific autoantibodies in T1D (Hattersley *et al.*, 2018; American Diabetes Association, 2020b).

T1D is characterised by an autoimmune reaction against pancreatic beta cells, destruction by immune cells and subsequent insulin deficiency. Genetic predisposition to T1D has its strongest association with the human leukocyte antigen (HLA) region but more than 150 other regions have been identified so far at genome-wide significance ($p < 5x10^{-8}$) (Cudworth & Woodrow, 1974; Todd *et al.*, 2007; Erlich *et al.*, 2008; Barrett *et al.*, 2009; Evangelou *et al.*, 2014; Onengut-Gumuscu *et al.*, 2015; Fortune *et al.*, 2015; Robertson *et al.*, 2020). Interestingly, out of the five associations that colocalised between T1D and T2D, four demonstrated opposing directions of effect, highlighting individual genetic risk factors and the need for distinct therapeutic strategies (Aylward *et al.*, 2018; Inshaw *et al.*, 2020). The genetics of monogenic diabetes and T2D are further described in 1.2.1 and 1.2.2, respectively.

Studying human genetics and naturally occurring variation has enormous potential to better understand disease pathogenesis, guide therapy development, stratify patients based on predicted treatment success and complications, improve disease diagnostics, develop genetic risk scores and identify novel biomarkers for better monitoring of disease progression, which

is essential in a genetically and clinically heterogenous disease such as T2D (McCarthy, 2017; Fitipaldi *et al.*, 2018; Chung *et al.*, 2020).

1.2.1 Monogenic diabetes

Around 2% of all diabetes cases fall into the category of monogenic diabetes, a disorder resulting from mutations in a single gene (International Diabetes Federation, 2019). Most cases of monogenic diabetes can be classified as neonatal diabetes mellitus (NDM) or maturity-onset diabetes of the young (MODY) and usually present an age of onset less than 6 months or 25 years, respectively. The majority of genes involved in monogenic diabetes have been associated with endocrine pancreas development or function with the most common being *GCK, HNF1A, HNF4A* and *HNF1B* in MODY and *ABCC8* and *KCNJ11* in NDM (Froguel *et al.*, 1992; Hattersley *et al.*, 1992; Yamagata *et al.*, 1996a, 1996b; Horikawa *et al.*, 1997; Gloyn *et al.*, 2004; Proks *et al.*, 2006; Babenko *et al.*, 2006). The pathogenic mutations found in MODY and NDM patients are rare, highly penetrant and demonstrate a more severe clinical phenotype than variants found in T2D. Monogenic diabetes therefore offers the unique opportunity to observe the functional and developmental consequences of specific dysfunctional genes in pancreatic beta cells *in vivo* in humans. The functional understanding of genes and pathways leading to monogenic diabetes has also been successful in informing targeted drug treatment as seen for sulphonylureas in patients with mutations in genes encoding subunits of the K_{ATP} channel (Gloyn *et al.*, 2004; Babenko *et al.*, 2006; Pearson *et al.*, 2006). Further, pathophysiological insights from monogenic diabetes can also be exploited to functionally understand and prioritise risk genes in T2D. Around one-third of monogenic genes have so far also been identified as T2D risk genes (*KCNJ11, ABCC8, GCK, SLC2A2, HNF1A, HNF4A, HNF1B, PDX1, PAX4, NEUROD1, WFS1, PPARG*) (Fuchsberger *et al.*, 2016).

1.2.2 Heritability of T2D

T2D is a common multifactorial and complex disease which is influenced by environmental exposures and epigenetic and genetic risk factors. Environmental exposures play a fundamental role in developing T2D and include factors such as a sedentary lifestyle, diet, microbiome and early life influences (Franks & McCarthy, 2016). However, current heritability estimates for T2D range from 30 – 70% based on population, family and twin studies and even though they span a wide range, it clearly highlights the important role of genetics in determining disease risk (Poulsen *et al.*, 1999; Meigs, Cupples, & Wilson, 2000; Almgren *et al.*, 2011). The lifetime risk of developing T2D is projected to be 40% if one parent has T2D and nearly 70% if both parents are affected (Tillil & Kobberling, 1987). Twin studies demonstrated a concordance rate of about 70% in monozygotic twins and 20 – 30% in dizygotic twins (Kaprio *et al.*, 1992). It should be noted that the heritability based on family or twin studies might be overestimated due to shared environmental factors such as intrauterine and pregnancy related influences. Following this initial endeavour to identify the heritability of T2D, studies have now focused on mapping the disease-causing genes and pathways. As opposed to monogenic diabetes, T2D is highly polygenic with many variants that are only associated with small effect sizes, requiring large studies and initiatives to identify genes involved in disease risk.

1.2.3 Discovering T2D risk genes

The familial and twin studies clearly demonstrated a strong heritable component for T2D which was followed by subsequent small-scale linkage studies, candidate gene approaches and large-scale genome-wide association studies (GWAS) to identify genes mediating the disease risk.

Linkage studies identify regions of the genome that are associated with the disease based on coinheritance of certain genes or markers due to their proximity to each other. Successful in detecting rare variants with large effects such as in monogenic diseases, they have not proven to be effective in complex polygenic disease with many variants with low effect sizes such as T2D due to their low resolution (Hanis *et al.*, 1996; Duggirala *et al.*, 1999; Ali, 2013). The studies only utilise a few hundred markers across the genome and the identified linked regions could span up to hundreds of genes without a clear candidate gene. One success story coming out of linkage analysis in T2D was the discovery of the susceptibility gene *TCF7L2*, which has since been replicated in many independent studies and remains the strongest association with T2D with the greatest odds ratio to date (Duggirala *et al.*, 1999; Grant *et al.*, 2006).

Candidate gene studies focus on targeted sequencing of suspected candidate genes that are known to play a role in beta cell function or insulin action. Few genes were identified including the reproducible effects in *KCNJ11* and *PPARG* (Altshuler *et al.*, 2000; Gloyn *et al.*, 2003). *KCNJ11* encodes a subunit of the K_{ATP} channel, is crucial in regulating insulin secretion in the beta cell and activating mutations are associated with NDM (Gloyn *et al.*, 2004). *PPARG* encodes a nuclear receptor and is the target of thiazolidinediones, an insulin sensitising T2D medication (Harris & Kletzien, 1994; Altshuler *et al.*, 2000). The associated variant harbouring a proline for alanine substitution has been linked to lower BMI, increased insulin sensitivity and protection against T2D (Deeb *et al.*, 1998; Altshuler *et al.*, 2000). However, candidate gene studies also failed to deliver many replicated susceptibility genes for T2D due to their limited sample sizes to detect the now known to be small effect sizes of T2D, selection of genes based on limited knowledge and understanding of T2D and low-throughput and limited technologies at the time (McCarthy & Zeggini, 2009).

Technical advances such as novel high-throughput genotyping assays and improved linkage or haplotype maps of the genome in combination with investments into large Biobanks such as the UK Biobank and ever increasing large sample sizes eventually led to the development and successful implementation of large-scale GWAS (Sudlow *et al.*, 2015). In these association studies, many single-nucleotide polymorphisms (SNP) are being genotyped across the genome, relying on the fact that each SNP can be found in linkage disequilibrium (LD) with other SNPs (Gabriel *et al.*, 2002; Slatkin, 2008). It is therefore possible to achieve near genome-wide coverage by assaying only 0.5-1 million SNPs that represent individual haplotypes (The International HapMap Consortium, 2005, 2007; Li, Li, & Guan, 2008). As a result, GWAS allow for an affordable genome-wide interrogation of variants but only by identifying an association with the respective haplotype 'tag SNPs' and not necessarily with the disease-causing variants directly as further described in section 1.3.1 (Johnson *et al.*, 2001).

Early GWAS for T2D in a few thousand samples confirmed the principle study design and identified several loci such as *HHEX* and *SLC30A8,* but it also became clear that larger sample sizes are necessary to identify common variants with weaker effects (Frayling *et al.*, 2007; Sladek *et al.*, 2007; Burton *et al.*, 2007; Scott *et al.*, 2007; Zeggini *et al.*, 2007; Saxena *et al.*, 2007). The second wave of GWAS was based on the formation of large consortia, meta-analysis and the development of custom genotyping arrays to combine efforts and increase sample sizes up to 150 000 individuals (Zeggini *et al.*, 2008; Voight *et al.*, 2010, 2012; Morris *et al.*, 2012). Trans-ethnic GWAS across populations with different ancestry and improved sequencing technologies to perform whole exome or genome studies have further expanded the number of discovered T2D loci (Kooner *et al.*, 2011; Lohmueller *et al.*, 2013; Moltke *et al.*, 2014; Mahajan *et al.*, 2014; Steinthorsdottir *et al.*, 2014; Ng *et al.*, 2014; Fuchsberger *et al.*, 2016; Flannick *et al.*, 2019; Spracklen *et al.*, 2020). Since the first

GWAS for T2D in 2007, the number of distinct association signals for T2D risk has increased to more than 400 (Scott *et al.*, 2017; Mahajan *et al.*, 2018b, 2018a; Spracklen *et al.*, 2020; Vujkovic *et al.*, 2020). The most recent association study investigated T2D susceptibility in more than 200 000 T2D cases and one million controls with a multi-ancestry background resulting in the discovery of 318 loci (Vujkovic *et al.*, 2020). With an ever increasing number of loci for T2D predisposition, it also became apparent that the genetic landscape of T2D is mostly characterised by a long tail of common variants with ever-smaller effect sizes (Yang *et al.*, 2015; Fuchsberger *et al.*, 2016). Approximately 20% of overall T2D risk can now be explained by common variants, equivalent to half of the total estimated heritability (Mahajan *et al.*, 2018a; Barroso & McCarthy, 2019).

The emerging genetic data are in line with the recently described palette model of T2D risk as shown in Figure 1.2 (McCarthy, 2017). Overall T2D susceptibility is dependent on genetic and non-genetic factors with an effect on different processes in pancreas, liver, fat, muscle and brain tissue. All of those processes including islet function and insulin resistance are influenced by some of the identified common variants (Dimas *et al.*, 2014; Mahajan *et al.*, 2018b; Udler *et al.*, 2018). The lifetime T2D risk is eventually an aggregation of all individual risks from each process or in terms of the palette model, a distinct colour which is a mixture of the individual process colours and shades.

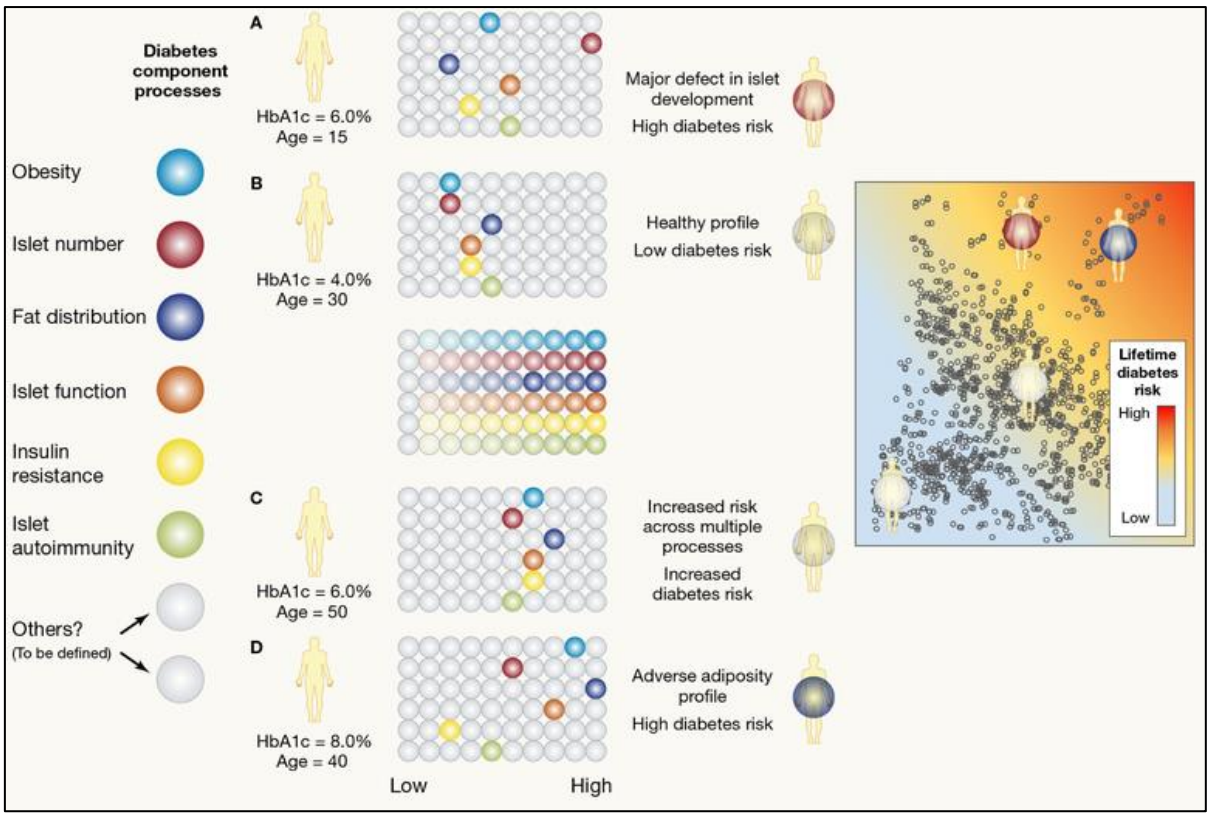


Figure 1.2 – The palette model for T2D. Someone's lifetime risk to develop T2D is influenced by a variety of processes such as islet function and insulin resistance which are influenced by several genetic and non-genetic factors. These processes are indicated by different colours on the left and represent the base colour palette of the T2D risk model. Each individual A, B, C, D has a unique profile for each process ranging from low to high risk. All processes or colour shades in combination make up someone's distinct overall T2D risk. Some individuals might be strongly influenced by one particular process, such as individual A with a major defect in islet development and others might have equally increased risk across all processes such as individual C. Taken from: Barroso and McCarthy (2019). The genetic basis of metabolic disease. *Cell*; 177(1): 146-161

1.2.4 The pancreatic islet is central to T2D

T2D is characterised by a heterogenous multi-organ pathogenesis including the pancreas, liver, muscle, adipose tissue and the brain. Nonetheless, many studies have highlighted the pancreatic islets as central in mediating T2D susceptibility due to defects in insulin, glucagon or somatostatin secretion resulting in abnormal glucose homeostasis (Ingelsson *et al.*, 2010; Voight *et al.*, 2010; Dimas *et al.*, 2014; Pasquali *et al.*, 2014; Mahajan *et al.*, 2018b; Thurner *et al.*, 2018; Udler *et al.*, 2018). When 37 T2D risk variants were initially

associated with quantitative metabolic data in non-diabetic subjects including measures of beta cell function and insulin sensitivity, T2D loci separated into distinct clusters based on their effect on disease risk (Dimas *et al.*, 2014). Most variants fell into one of the clusters associated with beta cell function (insulin processing, reduced beta cell function after glucose stimulation or beta cell function) and fewer variants were assigned into the insulin resistance cluster. Further data from GWAS expanded this analysis by associating 94 loci with physiological measures of a diverse set of metabolic traits (Mahajan *et al.*, 2018b). The largest cluster was formed by loci that acted through a primary effect on insulin secretion and contained 39 out of 71 allocated loci. A similar soft clustering approach integrated 94 variants and 47 diabetes-related traits with the majority of loci mapping to beta cell dysfunction clusters (beta cell or proinsulin) in contrast to insulin response clusters (obesity, lipodystrophy or liver) (Udler *et al.*, 2018). In addition to human physiology, studies refining genomic annotations in islets have also demonstrated an enrichment of T2D and fasting glucose GWAS variants within islet *cis*-regulatory elements such as enhancers (Parker *et al.*, 2013; Pasquali *et al.*, 2014; Varshney *et al.*, 2017; Thurner *et al.*, 2018). Many of these variants affect the downstream gene expression by disrupting islet-specific transcription factor binding motifs. One example is a variant within the *ZFAND3* locus which altered the binding of the key islet transcription factor NEUROD1, resulting in loss of enhancer activity (Pasquali *et al.*, 2014). Due to this comprehensive evidence from physiological clustering and epigenomic annotations, islet cell dysfunction has become the central mechanism to modulate T2D risk and many functional follow-up studies focus on islet cells or more specifically on the insulin secreting beta cell.

1.3 Prioritising causal genes in T2D

Upon identifying genetic loci through GWAS, associated variants and potential target genes should be prioritised in follow-up studies to better understand the underlying functional mechanisms and pathways involved in T2D pathogenesis and to be able to pinpoint the causal gene at each locus. This functional translation from association signals into causal genes and functional mechanisms, however, has often been slow, largely due to the design of GWAS itself. GWAS are so effective in discovering T2D loci across the genome as the assay is designed to only genotype specific SNPs as surrogates of a larger region. However, which one of the coinherited variants within this region actually influence disease risk, is often unclear. In addition, these SNPs are most often (90%) found in non-coding regions (Maurano *et al.,* 2012). For simplicity, GWAS loci are being named after their nearest gene but this proximity does not necessarily entail causality. Associating certain GWAS SNPs with causal genes and disease pathogenesis is therefore challenging in most cases. Causal genes can be prioritised for functional follow-up studies based on traditional approaches such as focussing on low-hanging fruit due to known biology or through integration with coding variants, establishing variant-gene links or unbiased functional screens which are discussed in section 1.3.2 to 1.3.4. To fully understand the underlying functional mechanisms and perform functional follow-up studies in a disease relevant context, determining the causal variant and the relevant tissue, cell type or developmental stage are often a prerequisite for successful identification of causal genes as described in section 1.3.1.

1.3.1 Prerequisites for identifying causal genes

GWAS identify SNPs that serve as surrogates for all variants within a haplotype, but are not necessarily the causal variants that might functionally mediate disease risk through effects on transcriptional, splicing or translational regulation if located within a non-coding region or directly affect protein function as coding variants (Ward & Kellis, 2012). Determining the causal variant can provide a direct functional connection between the genotype, a potential causal gene and the disease, in particular if the causal variant lies within a coding region as discussed in section 1.3.2. To narrow down the causal variant or a set of likely causal variants, many strategies have been deployed including fine-mapping of GWAS loci, experimental investigations or *in-silico* predictions (Fogarty *et al.*, 2013; Morris, 2014; Claussnitzer *et al.*, 2015; Gaulton *et al.*, 2015).

Fine-mapping of a disease locus refines the genetic variation withing a certain haplotype to identify the causal variant using targeted re-sequencing, custom genotyping assays or imputation from previous sequencing projects such as the 1000 Genomes Project and the HapMap project (The International HapMap Consortium, 2007; The 1000 Genomes Project Consortium, 2012; Morris, 2014). These efforts usually identify a list of variants designated as credible set, that contain the smallest set of SNPs that account for a certain posterior probability (usually 99%) to cumulatively drive the association signal. Prioritising variants in fine-mapping studies is often facilitated through trans-ethnic GWAS data and genomic annotations such as chromatin-state maps in disease relevant tissues or transcription factor binding sites (Mahajan *et al.*, 2014, 2018a; Gaulton *et al.*, 2015). Recent improved fine-mapping efforts in T2D created credible sets for 380 of the 403 independent GWAS signals with a median of 42 variants in each set and 18 association signals could be mapped to a single variant (Mahajan *et al.*, 2018a). Exemplary of how a fine-mapped single variant for an association signal can provide direct insights into the underlying biology is the *MTNR1B*

locus. *In silico* analysis predicted that the lone credible variant overlapped a FOXA2 binding site and created a NEUROD1 recognition motif (Gaulton *et al.*, 2015). *In vitro* experimental investigation confirmed that the risk allele specifically binds NEUROD1 in a model of human beta cells, which likely mediates the upregulation of *MTNR1B* expression that is associated with increased T2D risk (Fadista *et al.*, 2014; Gaulton *et al.*, 2015). Therefore, in addition to the causal variant, knowledge of the appropriate tissue and developmental stage is crucial in linking GWAS association signals to functional understanding of the disease mechanism.

1.3.2 Coding variants

Most of the variants identified by GWAS are located within non-coding regions. If the credible set includes a coding variant on the other hand, a direct link between association signals, causal genes and therapeutic insights can be made. Coding variants are enriched within GWAS associations and are more likely to be causal (Gusev *et al.*, 2014; Walter *et al.*, 2015). Nevertheless, this does not imply direct causality for every coding variant, especially if the variants originate from exome-based studies where surrounding non-coding regions were not examined. Recent high-resolution fine-mapping demonstrated that one-third of the detected coding variants in T2D were highly unlikely to be causal, urging caution to draw premature functional conclusions from coding variants (Mahajan *et al.*, 2018b). On the other hand, coding variants can be used to prioritise genes for functional studies and translate association signals into disease insights and potential promising therapeutic targets such as it has been done for *SLC30A8* (Sladek *et al.*, 2007; Flannick *et al.*, 2014; Kleiner *et al.*, 2018; Dwivedi *et al.*, 2019)

The gene *SLC30A8* encodes the zinc transporter ZnT8 and plays an important role in secretory vesicles within the pancreatic beta cell to synthesise, store and stabilise insulin (Chimienti *et al.*, 2004; Pound *et al.*, 2009; Nicolson *et al.*, 2009; Wijesekara *et al.*, 2010). Several coding variants have been identified in *SLC30A8* so far in different studies (Sladek *et al.*, 2007; Flannick *et al.*, 2014, 2019). A common variant (p.Trp325Arg) was identified with its major allele (>70% of the population) demonstrating increased risk for T2D and a protective minor allele (Sladek *et al.*, 2007). Two multi-ancestry studies found more than 30 rare, protective and protein truncating variants in *SLC30A8* (Flannick *et al.*, 2014, 2019). Two variants in particular, p.Arg138* and p.Lys34Serfs*50 were associated with 53% and 83% protection against T2D as demonstrated in populations in western Finland and Iceland, respectively (Flannick *et al.*, 2014). Studies in rodent models have been inconclusive in explaining the protective effects of loss of *Slc30a8* (Lemaire *et al.*, 2009; Wijesekara *et al.*, 2010; Kleiner *et al.*, 2018). A recent comprehensive study combined recruit by genotype metabolic studies, functional investigations in models of human beta cells and a mouse model carrying the human loss of function (LoF) allele to understand the protective effect of coding *SLC30A8* alleles (Dwivedi *et al.*, 2019). Human carriers showed enhanced insulin secretion due to better glucose responsiveness and proinsulin conversion, genome edited human induced pluripotent stem cells (iPSC)-derived beta like cells demonstrated nonsense mediated decay of *SLC30A8* variant transcripts and knockdown of *SLC30A8* in EndoC-βH1 cells resulted in reduced K_{ATP} channel gene expression and function leading to enhanced glucose-stimulated insulin secretion (Dwivedi *et al.*, 2019). These studies demonstrate the protective effect of LoF mutations in *SLC30A8* through reduced transcript expression and its potential as therapeutic target in T2D.

1.3.3 Variant-gene links

In contrast to coding variants, it is more complex and challenging to link non-coding GWAS variants (>90% of all association signals) to their causal effector transcript. Non-coding SNPs are highly likely to influence disease risk through effects on gene regulation such as transcriptional, splicing or translational regulation (Ward & Kellis, 2012). Therefore, intersection with genomic annotations such as expression quantitative trait (eQTL) mapping and epigenomic annotations can link GWAS variants to putative effector genes while also providing a potential functional mechanism how they affect disease risk (Nica & Dermitzakis, 2013). However, genomic annotations are highly context specific and disease relevant cell types and developmental stages have to be considered when performing functional investigations and interpreting emerging data (Thomsen, McCarthy, & Gloyn, 2016b). In T2D, efforts to map genomic annotations have most often focussed on the pancreatic islet due to its central role within disease risk and pathogenesis (Ingelsson *et al.*, 2010; Voight *et al.*, 2010; Dimas *et al.*, 2014; Mahajan *et al.*, 2018b).

eQTL mapping studies identify variants that modify expression levels of either close genes (*cis*-eQTL) or over longer distance (*trans*-eQTL) (Nica & Dermitzakis, 2013). Colocalization of eQTLs and GWAS association signals can be used to associate a variant with a downstream causal gene (Chen *et al.*, 2008; Nica & Dermitzakis, 2008; Nica *et al.*, 2010). Even though access to human islet tissue is limited and small scale studies often have reduced statistical power to identify islet-specific eQTLs, independent investigations have been able to pinpoint novel candidate effector transcripts at several GWAS loci (Fadista *et al.*, 2014; van de Bunt *et al.*, 2015; Varshney *et al.*, 2017; Viñuela *et al.*, 2019). The most recent study using 420 human cadaveric islet preparations highlighted potential effector transcripts at 23 GWAS loci, both confirming previous candidates and identifying novel associations (Viñuela *et al.*, 2019). It is important to keep in mind that 1) these eQTL studies

utilise whole islet samples and are therefore a heterogeneous sample including alpha, beta and delta cells and 2) colocalization does not imply causality as variants can have an effect on disease risk through other horizontal pleiotropic effects such as affecting another phenotypic trait which in turn affects T2D risk. Experimental approaches including perturbations in cellular and animal models are required to validate the effects of candidate causal genes.

Genomic annotations are defined as mapping the regulatory or functional meaning within the genome and are crucial in understanding effects of non-coding variation on gene regulation. Islet chromatin state maps are being generated using next-generation sequencing in combination with enzymes binding to open chromatin (ATAC-Seq), DNA methylation is analysed through genome bisulfite sequencing, genomic interactions such as between enhancer and promoters are defined using chromosome conformation capture (3C, 4C and Hi-C) and protein binding such as transcription factor binding sites can be analysed using chromatin immunoprecipitation followed by sequencing (ChIP-Seq) (Gaulton *et al.*, 2010; Stitzel *et al.*, 2010; Parker *et al.*, 2013; Pasquali *et al.*, 2014; Volkov *et al.*, 2017; Thurner *et al.*, 2018; Manduchi *et al.*, 2018; Greenwald *et al.*, 2019; Miguel-Escalada *et al.*, 2019; Rai *et al.*, 2020). To identify genome-wide interactions between enhancers and promoter, a recent study performed promoter capture Hi-C in human pancreatic islets and mapped 87% of variants overlapping islet enhancers to candidate target genes (Miguel-Escalada *et al.*, 2019). Interestingly, most of these target genes (75%) were not expected based only on their proximity for example *OPTN* at the *CDC123/CAMK1D* locus. Genome editing was used to experimentally validate the regulation of both genes, *OPTN* and *CAMK1D* by the islet enhancer containing the non-coding GWAS SNP. The contribution of each gene and their causal relationship to T2D remains to be determined, as *CAMK1D* has also independently been identified as the target gene of an islet cis-eQTL and an active islet enhancer region,

which both colocalised with a T2D variant (Fogarty *et al.*, 2013; Greenwald *et al.*, 2019). Nonetheless, it highlights how GWAS variants within enhancer can affect multiple genes and can be used to prioritise genes for follow-up studies from GWAS association signals.

1.3.4 Functional screening

Genes at GWAS loci can also be prioritised based on already known involvement in a cellular process relevant to disease susceptibility such as insulin secretion, giving them a higher likelihood to also be involved in T2D pathogenesis. However, many post-GWAS functional follow-up studies have only focussed on validating individual candidate genes at T2D loci. Ideally, all genes at T2D loci would be experimentally investigated for multiple disease-relevant phenotypes in multiple cellular models. This study would be independent of prior assumptions to either exclude any involvement of other genes and confirm the assumed hypothesis or discover genes with an unknown role in disease pathogenesis. High-throughput functional screening can address this need by examining a large number of candidate genes in a highly parallel manner in a relevant cell type and for a phenotype of interest which is involved in disease pathogenesis. Many screening strategies have been developed that can be differentiated based on the setup (arrayed or pooled), direction of effect on gene expression (gain of function (GoF) versus LoF), gene modulation technique (mostly RNA interference (RNAi) versus CRISPR/Cas9) and scale (genome-wide or targeted approach) (Elbashir *et al.*, 2001; Paddison *et al.*, 2004; Qi *et al.*, 2013; Gilbert *et al.*, 2014; Shalem *et al.*, 2014; Shalem, Sanjana, & Zhang, 2015; Doench, 2018). Compared to arrayed screening approaches, pooled genome-wide screens are a cost-effective approach to perform an unbiased investigation of all protein-coding genes for a specific phenotype (section 1.5.3). Several screens have been conducted to study beta cell function or genes involved in T2D pathogenesis and are discussed in sections 1.5.1 and 1.5.4 (Ku *et al.*, 2012; Thomsen *et al.*, 2016a; Akerman *et al.*, 2017; Pappalardo *et al.*, 2017; Fang *et al.*, 2019).

Future functional screens in different cellular models and focussing on other relevant phenotypes will further unravel the genes underlying T2D pathogenesis and playing a role in disease relevant cellular mechanisms. The approach is fast evolving with ever new techniques being developed including imaging-based pooled CRISPR screens, coupling single-cell transcriptomics with CRISPR perturbations, pooled screens targeting non-coding regions and CRISPR base-editing screens (Sanjana *et al.*, 2016; Datlinger *et al.*, 2017; Xie *et al.*, 2018; Wang *et al.*, 2019; Kweon *et al.*, 2020).

1.3.5 Integrating multiple lines of evidence

Each of the discussed strategies offers a complementary approach to prioritise causal genes at T2D GWAS loci, but no approach is sufficient on its own and integrating several lines of evidence is crucial in assigning a causal role. Identification of coding variants require large sample sizes to identify rare variants with large effects and the variants must be naturally occurring. Associating non-coding variants with downstream candidate effector genes can be hindered by potential horizontal pleiotropic effects and most studies are being performed in disease-relevant primary tissue which is often limited. Functional gene-focussed screens interrogate disease relevant phenotypes but are lacking the direct link to GWAS variants and are strongly context specific, meaning they must be performed in a suitable cell type with a relevant phenotypic readout and potentially environmental stimulus. Hence, one line of evidence can give an indication of which gene is causally involved in T2D susceptibility, but several diverse routes of investigations should be pursued to obtain a full picture and support for a certain mechanism (Figure 1.3). One (or several) genes can only be defined with confidence as causal for T2D through combination and convergence of complementary datasets (Gaulton, 2017).

Recent studies have often focussed on integrating genetic, genomic and functional data to associate GWAS variants with genes mediating the effect on disease risk. One such example would be the previously described comprehensive functional studies on *SLC30A8*, combining human genetic data of risk carriers and mouse models with experimental *in vitro* investigations (Dwivedi *et al.*, 2019). A recruit by genotype study established that carriers of the protective allele had increased proinsulin processing and insulin secretion, consistent with a mouse model carrying the mutation that showed increased insulin secretion on a high-fat diet (Dwivedi *et al.*, 2019). Further experimental studies using genome edited iPSC derived beta cells containing the GWAS variants and RNAi mediated knockdown of *SLC30A8* in cellular models of beta cells further confirmed that the variant containing transcript were being degraded through non-sense mediated decay and subsequent loss of ZnT8 reduced K_{ATP} channel expression and increased glucose-stimulated insulin secretion (Dwivedi *et al.*, 2019).

Another computational approach was developed by Fernández-Tajes et al. by calculating positional candidacy scores (PCS) for each gene at a T2D GWAS locus which can be used to prioritise genes within each region (Fernández-Tajes *et al.*, 2019). PCS were calculated based on *cis*-eQTL expression data for GWAS variants in combination with disease-relevant biological annotation such as Gene Ontology (GO) for all protein coding genes within the region at 101 T2D GWAS loci. In total, 1895 genes were scored, located within 1 Mb around the GWAS lead variant and capturing most cis-regulatory effects (GTEx Consortium, 2017).

Future data integration will be supported by ever increasing publicly available datasets and integrated knowledge databases from large scale projects and consortia (Flannick & Florez, 2016). Mapping of functional annotation in many cell types to integrate GWAS variants has been accelerated by ENCODE, Genotype-Tissue Expression (GTEx) and the NIH Epigenomics Roadmap (The ENCODE Project Consortium, 2012; GTEx Consortium, 2015;

Roadmap Epigenomics Consortium *et al.*, 2015). T2D GWAS data from many independent investigations has been accumulated on interactive portals such as the Type 2 Diabetes Knowledge Portal and provides an accessible platform to integrate functional insights with genetic associations (Accelerating Medicines Partnership, 2020).

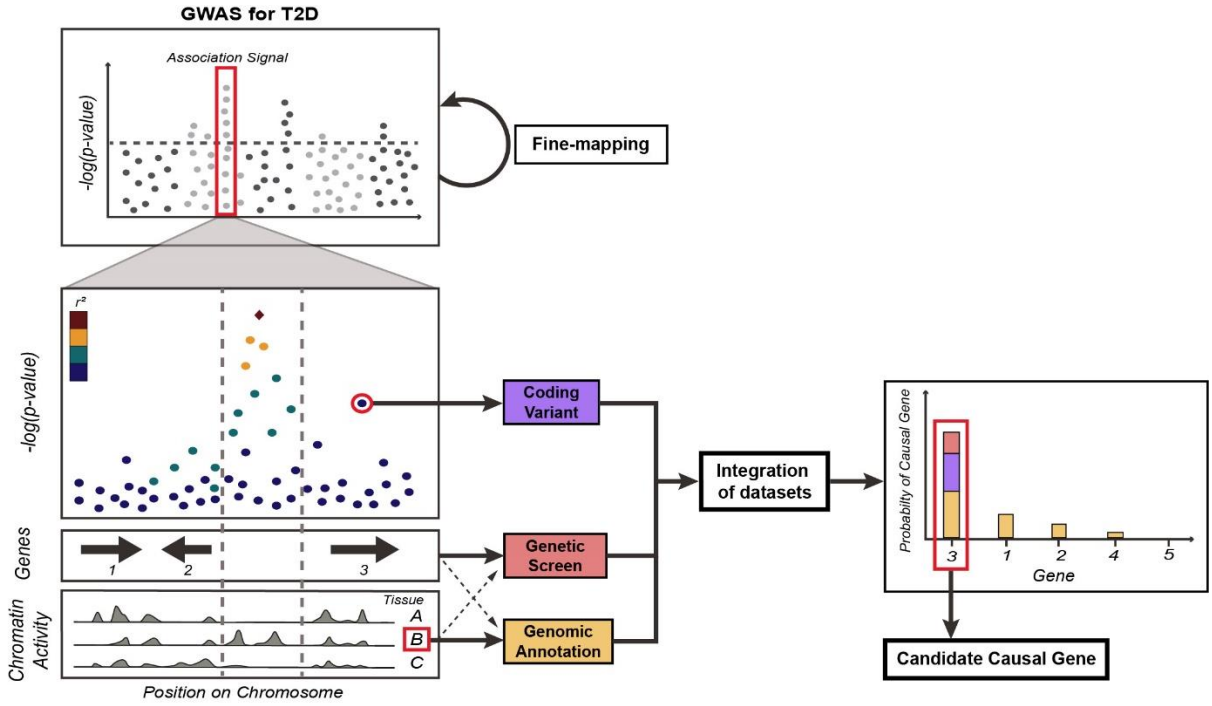


Figure 1.3 – Integrating multiple lines of evidence to prioritise genes at GWAS loci. GWAS have identified more than 400 regions that are associated with T2D risk, but most causal genes are still unknown. Fine-mapping can narrow down the associated variants to identify the causal variant(s) (*top left*, Manhattan plot). Strategies to help prioritise causal genes are demonstrated based on an exemplary risk locus (*bottom left*). An independent coding signal hints at gene 3 as a likely effector transcript (*purple*). Annotations in different tissues (A, B and C) demonstrate that the lead GWAS SNP is located within an active chromatin region of tissue B therefore prioritising tissue B to study context-specific variant-gene links such as *cis*-eQTL (*orange*). Functional screens in a cellular model of tissue B and looking at gene 1, 2, 3 and others within the locus (gene 4 and 5, not shown) can further prioritise a certain gene within the locus (*red*). All prioritisation strategies offer complementary evidence and integration of datasets is crucial (*bottom right*). At this locus, gene 3 demonstrated a functional effect in a genetic screen and was highlighted by genetic data based on a coding variant. Genomic annotations showed some degree of evidence for all genes with the strongest indication for gene 3. Based on integration of all three strategies, gene 3 is most likely to be causal and should be followed-up by in-depth validation studies to understand the functional mechanism and its effect on disease risk (Grotz, Gloyn, & Thomsen, 2017).

1.4 The beta cell within pancreatic islets

The pancreatic islet, especially the beta cell plays a central role in controlling glucose homeostasis and mediating T2D risk as shown through physiological clustering and colocalization with epigenomic annotations (section 1.2.4). Subsequently, islet dysfunction and cellular models of beta cells are the focus of functional follow-up studies to understand T2D pathogenesis and disease risk.

1.4.1 The pancreatic islet

The pancreas is comprised of exocrine and endocrine cells which produce and secrete digestive enzymes into the pancreatic duct and hormones to regulate glucose homeostasis into the bloodstream, respectively. The exocrine tissue consists of acinar and duct cells and compose around 90% of pancreatic mass, whereas endocrine cells form the islets of Langerhans and only make up around 1-2% of the total pancreas (Pandiri, 2014). Human pancreatic islets consist of alpha (35-40%), beta (50%), delta (10-15%), pancreatic polypeptide (<5%) and epsilon cells (<1%) which secrete glucagon, insulin, somatostatin, pancreatic polypeptide and ghrelin, respectively (Shih, Wang, & Sander, 2013; Rorsman & Ashcroft, 2018). The complex structure of pancreatic islets is highly species specific (Rorsman & Ashcroft, 2018). In humans, the different endocrine cell types are dispersed throughout the islet and in direct contact with other cell types (Brissova *et al.*, 2005; Cabrera *et al.*, 2006).

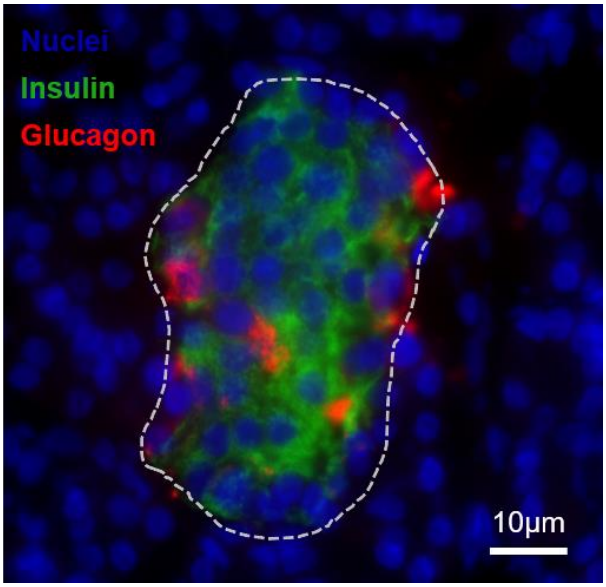


Figure 1.4 - Human pancreatic islet. Immunostaining of a human islet within a pancreas section, the islet region is indicated by the grey dashed line. Cell nuclei are stained by DAPI *(blue)* with endocrine and exocrine cells being shown. Beta cells are indicated by insulin staining *(green)* whereas alpha cells are highlighted by glucagon staining *(red)*. The image was taken at 40x magnification following the immunofluorescence protocol described in Chapter 5 – Methods.

1.4.2 Insulin synthesis and processing

The human *INS* gene is specifically expressed in beta cells which is regulated by cis-regulatory enhancer elements binding key transcription factors such as MafA, NeuroD1 and PDX-1 (Harper, Ullrich, & Saunders, 1981; Artner & Stein, 2008). *INS* transcription, mRNA stability and translation is tightly regulated in response to blood glucose concentrations (Itoh & Okamoto, 1980; Welsh *et al.*, 1985; Andrali *et al.*, 2008). The gene translates a 110-amino acid precursor termed preproinsulin which is folded in the endoplasmic reticulum (ER) and packaged into secretory granules in the Golgi apparatus where it is further processed into mature insulin (Patzelt *et al.*, 1978; Harper *et al.*, 1981; Huang & Arvan, 1995). Preproinsulin is targeted to the rough ER via an N-terminal signal peptide, which is subsequently cleaved to form proinsulin as shown in Figure 1.5 (Patzelt *et al.*, 1978).

The A and B domain of proinsulin are linked through the formation of three disulphide bonds, resulting in the folding and stabilisation of the precursor protein (Huang & Arvan, 1995).

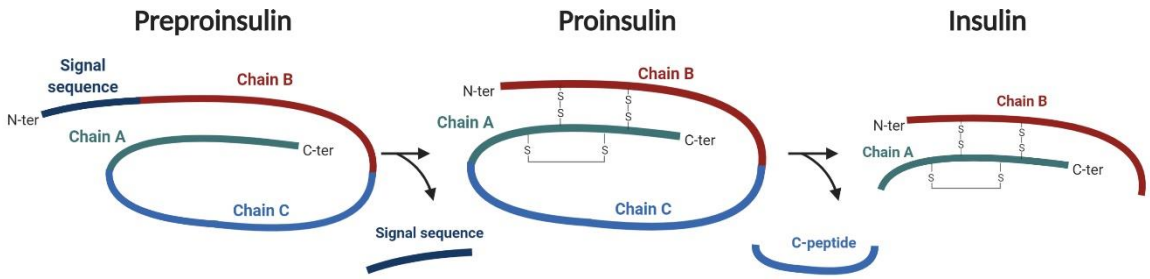


Figure 1.5 – Insulin processing. The insulin precursor preproinsulin (*left*) is being folded and cleaved in the ER to form proinsulin which is stabilised through three disulphide bonds (*middle*). Proinsulin is further cleaved to remove the C peptide in the Golgi apparatus to form mature insulin (*right*). The figure was create based on icons from Biorender.

Correct folding of insulin in the ER is crucial for normal beta cell function and glucose homeostasis as misfolding mutations demonstrate. The dominant negative C96Y mutation in the Akita mouse model leads to misfolded proinsulin and diabetes due to ER stress induced beta cell apoptosis (Wang *et al.*, 1999; Izumi *et al.*, 2003). Identical mutations have been observed in probands with neonatal diabetes (Stoy *et al.*, 2007).

The processing and maturation in the ER are therefore tightly regulated by quality control processes to ensure only properly folded proteins enter the secretory pathway and misfolded proteins are being disposed of (Meusser *et al.*, 2005; Almanza *et al.*, 2019). If ER homeostasis is disrupted due to accumulation of unfolded protein or extrinsic stimuli, the ER is in a state called ER stress. Secretory cells such as pancreatic beta cells are especially susceptible to ER stress due to their high level of insulin production which can easily exceed ER capacity during increased demand (Ghosh, Colon-Negron, & Papa, 2019). To resolve ER stress, the cell activates the cellular stress response pathway called the unfolded protein

response (UPR) (Zhang & Kaufman, 2006). The UPR senses ER stress through the activated trans-membrane signalling proteins activating transcription factor 6 (ATF6), protein kinase R-like ER kinase (PERK) and inositol-requiring enzyme 1 (IRE1), which induce downstream signalling cascades, transcription of UPR target genes and cellular processes to re-establish ER homeostasis (Haze *et al.*, 1999; Harding *et al.*, 2000; Lee *et al.*, 2002; Almanza *et al.*, 2019). These cellular responses include reducing protein translation, disposing of the stress inducing misfolded protein and enhancing capacity of the ER to fold proteins. If the adaptive UPR response is not able to resolve ER stress and is chronically activated, the terminal UPR response induces proapoptotic pathways leading to cell death.

ER stress and the UPR play an important role in normal beta cell function and in diabetes pathogenesis (Cnop, Foufelle, & Velloso, 2012). Mutations in components of the UPR response result in neonatal diabetes such as *WFS1* and *EIF2AK3* (encoding PERK) (Inoue *et al.*, 1998; Delepine *et al.*, 2000). Glucotoxicity, lipotoxicity, amyloid accumulation and the increased demand for insulin due to insulin resistance induce ER stress and contribute to beta cell dysfunction, cell death and loss of beta cell mass in T2D (Haataja *et al.*, 2008; Matsuda *et al.*, 2008; Cnop *et al.*, 2010; Back & Kaufman, 2012).

Correctly folded proinsulin is transported to the Golgi apparatus and packaged into immature secretory granules (Tokarz, MacDonald, & Klip, 2018). Within the granule, proinsulin is cleaved by prohormone convertases PC1/3 and 2 into mature insulin and C-peptide (Hutton, 1994; Ramzy, Asadi, & Kieffer, 2020). Secretory granules process further through the trans-Golgi network and insulin crystallises into its insoluble storage form as hexamer coupled with zinc resulting in mature dense-core granules (Tokarz *et al.*, 2018). Glucose sensing or related stimuli subsequently lead to exocytosis of insulin granules as described in section 1.4.3.

1.4.3 Glucose-stimulated insulin secretion

Pancreatic beta cells secrete insulin according to plasma glucose levels or external stimuli to control glucose homeostasis. Glucose-stimulated insulin secretion (GSIS) can be simplified into glucose uptake and metabolism, membrane depolarization followed by a rise in intracellular Ca^{2+} concentration and insulin granule exocytosis as shown in Figure 1.6 (Tokarz *et al.*, 2018). Glucose enters the cell via glucose transporters, primarily GLUT1 in humans and is phosphorylated by glucokinase (GCK) to generate glucose-6-phosphate (Matschinsky, 1990; McCulloch *et al.*, 2011). GCK acts as the rate limiting step in GSIS and the interplay with GLUT1 mediated glucose uptake result in the concentration-dependant glucose sensing and insulin secretion of the cell (Meglasson *et al.*, 1983; Matschinsky, 1990). Pyruvate is being generated through glycolysis, enters the mitochondria and the Krebs cycle to produce ATP (Jitrapakdee *et al.*, 2010). Subsequently, the resulting increase of the ATP to ADP ratio within the cell mediates closure of the ATP-sensitive potassium channels (K_{ATP}), leading to membrane depolarization and opening of voltage-gated Ca^{2+} channel (Ashcroft, Harrison, & Ashcroft, 1984; Ashcroft & Rorsman, 1989; Rorsman, Braun, & Zhang, 2012). Influx of Ca^{2+} and a rise in intracellular calcium concentration triggers plasma membrane fusion and exocytosis of insulin granules (Rorsman & Ashcroft, 2018). Insulin vesicles docking and fusion with the plasma membrane occurs through complex interplay with the exocytotic SNARE protein complex containing proteins such as SNAP-25, VAMP-2 and Munc18 (Gandasi & Barg, 2014; Gaisano, 2017).

In addition to the K_{ATP} dependant triggering pathway, the K_{ATP} independent amplifying pathway further controls insulin secretion by acting downstream of glucose metabolism and modulating the Ca^{2+} response (Henquin, 2000). Activators of the amplifying pathway include additional amplifying signals from glucose itself and extrinsic inhibitors or potentiators such as the gut-derived hormones glucagon-like peptide-1 (GLP1) and glucose-

dependent insulinotropic polypeptide (GIP) (Sato *et al.*, 1992; Gembal *et al.*, 1993; Henquin, 2011; Drucker, Habener, & Holst, 2017). The incretin hormones GLP-1 and GIP are secreted from the gut upon nutrient ingestion and stimulate insulin exocytosis in the beta cell at near-threshold glucose concentrations (Drucker *et al.*, 2017). They activate their respective G-protein–coupled receptors (GLP1R and GIPR) and increase insulin secretion through cyclic AMP (cAMP) and PKA dependant effects as well as PKA independent effects (Delmeire *et al.*, 2003; Kaihara *et al.*, 2013; Shigeto *et al.*, 2015).

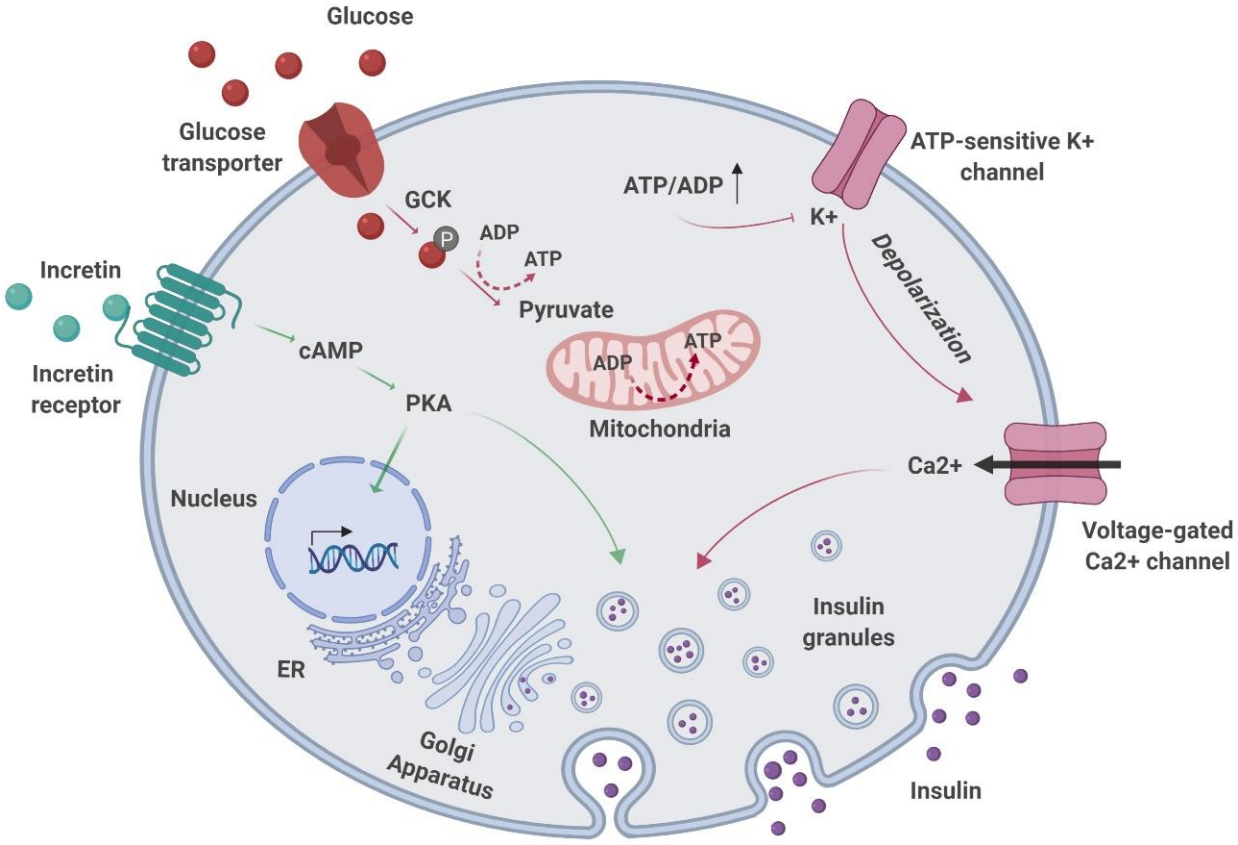


Figure 1.6 - Insulin secretion pathways in the pancreatic beta cell. Glucose-stimulated insulin secretion (pathway highlighted in *red*) is mediated by uptake of glucose via the glucose transporter, increase of the intracellular ATP to ADP ratio through glucose breakdown in glycolysis and the Krebs cycle in the mitochondria. Subsequent closure of the ATP-sensitive potassium channels induces membrane depolarization and Ca^{2+} influx which in turn triggers insulin granule exocytosis. Insulin secretion is additionally controlled by extrinsic stimuli such as the incretin GLP-1 through the amplification pathway (pathway highlighted in *green*).

1.4.4 Rodent vs human pancreatic islets and beta cells

While the availably of suitable *in vitro* models of human beta cells and the restricted access to human cadaveric pancreatic islets have been a long lasting challenge in researching human beta cell function, rodent beta cell models have contributed substantially to the progress in the functional understanding of healthy and dysfunctional beta cells (Efrat *et al.*, 1988; Miyazaki *et al.*, 1990; Asfari *et al.*, 1992). Human and rodent beta cells however have distinct genetic, structural, transcriptional and functional characteristics, therefore hindering the direct translation of functional insights gained in rodent cells into mechanistic understanding in human beta cells. Pancreatic islets in humans have intermingled alpha, beta and delta cells with alpha cells being clustered around blood vessels (Cabrera *et al.*, 2006; Bosco *et al.*, 2010; Cohrs *et al.*, 2017). Rodent islets in contrast are made up of a distinct beta cell core and a non-beta cell mantel resulting in close contact between beta cells and blood vessels (Cabrera *et al.*, 2006; Rorsman & Ashcroft, 2018). In addition, mouse islets have a higher vascular density and more beta cells than human islets (75% compared to 50 %) (Cabrera *et al.*, 2006; Brissova *et al.*, 2015). The interspecies differences also extend to a transcriptomic level. Expression analysis in alpha and beta cells from human and mice revealed a set of common core beta cell genes but also distinct species specific expression signatures especially in receptor genes and long non-coding RNAs (Benner *et al.*, 2014; Baron *et al.*, 2016). The transcriptomic analyses also demonstrated important expression differences in genes involved in disease pathogenesis such as *MAFB*, *IGF2* and *SLC2A2* highlighting the need to study disease relevant genes in the correct species context (Benner *et al.*, 2014; Baron *et al.*, 2016; Cyphert *et al.*, 2019). Moreover, unique characteristics in the glucose stimulated insulin secretion pathway reveal distinct functional signatures (Henquin *et al.*, 2015; Skelin Klemen *et al.*, 2017; Rorsman & Ashcroft, 2018). In mice, glucose uptake is mediated by Glut2 (*Slc2a2*) whereas human beta cells utilise GLUT1 and

GLUT3 (*SLC2A1* and *SLC2A3*) (De Vos *et al.*, 1995; Ferrer, Benito, & Gomis, 1995; McCulloch *et al.*, 2011). These glucose transporters possess different binding affinities for glucose resulting in particular species dependant glucose concentrations to initiate insulin secretion (3 mM in humans and 6 mM in mice) (McCulloch *et al.*, 2011). Rodents and human beta cells also contain different ion channels to regulate membrane potential and depolarization (Braun *et al.*, 2008; Skelin Klemen *et al.*, 2017; Rorsman & Ashcroft, 2018). The species differ in the expression of insulin with rodent beta cells expressing two insulin genes, *Ins1* and *Ins2*, whereas human beta cells only have one gene (*INS*) (Shiao *et al.*, 2008). Finally, rodent and human beta cells have different proliferation capacity *in vivo* and *in vitro* with rodent beta cells being more responsive to expansion stimuli (Parnaud *et al.*, 2008; Butler *et al.*, 2010; Kulkarni *et al.*, 2012; Fiaschi-Taesch *et al.*, 2013).

1.4.5 *In vitro* models of human beta cells

Based on the substantial species differences in beta cells of rodents and humans as described in 1.4.4, studies in mouse or rat beta cells do not adequately replicate human physiology and emphasise the requirement for *in vitro* models of human beta cells. Primary human pancreatic islets are the gold standard cellular models to study beta cell function or diabetes (Hart & Powers, 2019). However, research involving primary islets faces many challenges such as the availability and cost of islet samples, complex isolation procedures, high genetic and environmental variability between donors and preparations, reduced viability and function after *in vitro* culturing and limitations for genetic manipulation (Bottino et al., 2004; Ihm et al., 2006; Kulkarni & Stewart, 2014; Hart & Powers, 2019). Differentiating human pluripotent stem cells (hPSCs) towards beta-like cells to mimic pancreatic development offers another beta cell model to study developmental defects and for easy genetic manipulation, especially when modelling patient-specific mutations (D'Amour *et al.*, 2006; Rezania *et al.*, 2014; Balboa, Saarimäki-Vire, & Otonkoski, 2019b; Velazco-Cruz *et al.*,

2019). Until now, differentiated beta cells are still immature if not implanted into rodents, they require complex, expensive and variable differentiations, in-depth characterisation to detect abnormal karyotypes or *TP53* mutations and often result in heterogeneous populations with several cell types (Merkle *et al.*, 2017; Ihry *et al.*, 2018; Balboa *et al.*, 2019a).

In 2011, an immortalised human beta cell line called EndoC-βH1 cell line was developed by Scharfmann and colleagues (Ravassard *et al.*, 2011). As a stable cell line, EndoC-βH1 cells can be exposed to different experimental conditions and easily expanded *in vitro* with increased reproducibility compared to primary islets and hPSCs due to low genetic and environmental variabilities. The cell line was generated by transducing human fetal pancreatic buds with oncogene simian virus 40 large tumour antigen (SV40LT) under the rat insulin promoter (RIP) followed by expansion in SCID mice (Figure 1.7). The resulting insulinomas were again transduced with human telomerase reverse transcriptase (RIP-hTERT), transplanted into SCID mice, harvested, expanded *in vitro* and termed EndoC-βH1.

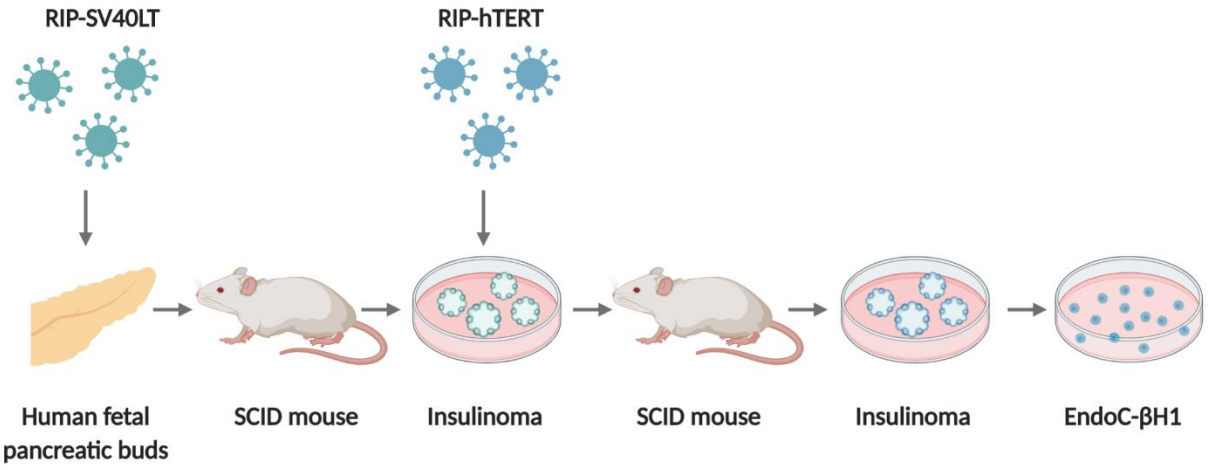


Figure 1.7 – Generation of EndoC-βH1. Human fetal pancreatic buds were transduced with the oncogene SV40LT driven by the rat insulin promoter (RIP) followed by transplantation into SCID mice. The resulting insulinoma were removed, transduced with the human telomerase reverse transcriptase (hTERT) and again transplanted into SCID mice to form insulinoma. The insulinoma cells were then harvested and propagated in tissue culture, resulting in the EndoC-βH1 cell line (Ravassard *et al.*, 2011).

The EndoC-βH1 cell line expressed key beta cell transcription factors at a similar level as primary human islets such as *NEUROD1, NKX6-1, PAX6, PDX1* and *MAFA* and 40% of the expression level of *INS* in islets. Simultaneously, non-beta cell markers such as *GCG* (alpha cells), *SST* (delta cells), *CPA1* (acinar cells) or *CFTR* (ductal cells) were barely detectable (less than 0.05%) or not expressed at all. The cells secreted insulin in a concentration-dependant manner with a 2- to 3-fold stimulation index from low to high glucose. The insulin content of the EndoC-βH1 cells is stable across passages and amounts to 0.46 μg per million cells which is approximately 10% of the content in primary islets (Brandhorst *et al.*, 1998; Gurgul-Convey, Kaminski, & Lenzen, 2015). The stimulation index and secreted insulin as percentage of insulin content on the other hand are similar to primary islets, making the cell line a physiologically relevant model to study insulin secretion and beta cell function (Ravassard *et al.*, 2011; Gurgul-Convey *et al.*, 2015; Henquin *et al.*, 2015). In addition to the initial characterisation by Scharfmann and colleagues, the cell line has now been widely adopted and studied by research groups around the world. These studies independently demonstrated that the EndoC-βH1 cell line has similar electrophysiological and insulin secretion properties as primary beta cells (Andersson *et al.*, 2015; Krishnan *et al.*, 2015; Teraoku & Lenzen, 2017; Hastoy *et al.*, 2018). Further, it has been successfully adopted to study effects of individual T2D risk genes such as *PAM* or *SLC30A8* and for high-throughput screens testing novel drug candidates or candidate effector transcripts (Thomsen *et al.*, 2016a, 2018; Ndiaye *et al.*, 2017; Tsonkova *et al.*, 2018; Dwivedi *et al.*, 2019). In addition, multi-omic profiling including epigenetic, chromosomal, transcriptomic and chromatin interactions characterisations demonstrate a high similarity with primary islets but also distinct characteristics highlighting the fetal and transformed origin of the cell line (Lawlor *et al.*, 2019).

In summary, the human beta cell line EndoC-βH1 has been extensively characterised and demonstrated its suitability as a cellular model to study human beta cell function and T2D risk genes. Due to their slow growth rate, it has to be taken into account that the cells are not able to be clonally expanded for genome editing, have to be closely monitored for their insulin secretion and beta cell characteristics during long term culture and are very sensitive to culture conditions and seeding densities (Weir & Bonner-Weir, 2011; Balboa *et al.*, 2019a).

To further expand the human beta cell line repertoire and generate a cell line with excisable immortalizing transgenes to more closely mimic nearly non-proliferating primary beta cells, EndoC-βH2 and EndoC-βH3 were generated (Scharfmann *et al.*, 2014; Benazra *et al.*, 2015). EndoC-βH2 were created in a similar process as EndoC-βH1 but with *loxP* flanked immortalizing transgenes (Scharfmann *et al.*, 2014). Upon lentiviral transduction of the resulting cell line to express *Cre*, the transgenes were excised, proliferation was stopped, and the cells developed beta cell characteristics that more closely resembled primary islets. EndoC-βH3 incorporated a tamoxifen inducible form of *Cre* to induce proliferation arrest (Benazra *et al.*, 2015).

1.5 Genetic screens in human beta cell models to study diabetes

Genetic screens provide an opportunity to investigate the phenotypic consequences of individually modulating the expression of many genes simultaneously. This high-throughput and unbiased approach is therefore extremely powerful in identifying genes that play a role in beta cell function and can be harnessed to prioritise causal genes at T2D GWAS loci. Compared to studies focussing on individual genes, genetic screens are labour- and cost-efficient and allow to study hundreds or thousands of genes simultaneously for their involvement in regulating a certain phenotype.

1.5.1 Knockdown based screens to study beta cell function

Modulating gene expression to study beta cell function in genetic screens has often been based on knockdown approaches using transient small interfering RNA (siRNA). Due to the required arrayed setup in siRNA-based screens, most functional screens focussed on only a certain subset of gene targets (Thomsen *et al.*, 2016a; Akerman *et al.*, 2017). In 2016, Thomsen et al. systematically screened 300 candidate genes in EndoC-βH1 which are located within 1 Mb of 75 T2D GWAS signals (Thomsen *et al.*, 2016a). As a phenotypic readout of beta cell function, they analysed cell count and insulin secretion under four different stimulation and glucose conditions. To make a large-scale investigation of the expensive insulin secretion measurement possible, Thomsen et al. applied an acoustic liquid handling approach to miniaturise the assay and reduced the costs to around 0.20 $ per data point. With a median residual expression of 43%, the screen was able to identify 67 hits that affected cell count or insulin secretion which mapped to 45 genes at 37 loci while replicating effects of known regulators of beta cell function. Specific screening hits such as *ARL15, ZMIZ1* and *THADA* were confirmed in independent follow-up experiments. This study demonstrated that large-scale functional screening in EndoC-βH1 for insulin secretion can

be used as a powerful tool to identify biologically relevant hits. Pappalardo et al. further expanded these initial strategies by performing a genome-wide siRNA screen for specific regulators of the insulin promoter (Pappalardo *et al.*, 2017). The screen was performed in an engineered mouse beta cell line containing an EGFP reporter regulated by the insulin promoter. In addition, the cell line contained a mCherry reporter controlled by a constitutive RSV promoter to control for gene hits that affect general transcription. The GFP to mCherry ratio upon silencing of approximately 22 000 mouse genes with 3 to 4 individual siRNAs per gene was able to identify regulators of insulin expression. Aside from identifying 26 novel regulators, the screen demonstrated a novel role for *Spry2* in insulin transcription and UPR regulation.

While being able to successfully associate many genes to specific phenotypes of beta cell function such as insulin secretion or expression, siRNA-based screens suffer from potentially undetected effects due to insufficient and variable knockdown (34-88% in Thomsen et al.), incomplete protein depletion, residual protein due to long half-lives and in the case of targeted screens, a biased inclusion of only specific genes (Thomsen *et al.*, 2016a). Recently developed genome-wide pooled CRISPR/Cas9 based genome editing screens can overcome those limitations and allow for an unbiased investigation following complete protein depletion.

1.5.2 CRISPR/Cas9 genome editing

Genome editing is a powerful tool to study the role of certain genes or variants in beta cell function or disease by associating the genetic modifications with cellular and disease relevant phenotypes. Recent advances in gene editing technologies have progressed from initial complex and limited technologies such as zinc finger nucleases (ZFN) and transcription activator like effector nucleases (TALEN) to the simple, flexible and affordable

clustered regularly interspaced short palindromic repeats (CRISPR)-associated proteins (Cas) system (Boch *et al.*, 2009; Moscou & Bogdanove, 2009; Urnov *et al.*, 2010; Gasiunas *et al.*, 2012; Jinek *et al.*, 2012; Cong *et al.*, 2013; Mali *et al.*, 2013).

The CRISPR/Cas9 system is based on the adaptive immune response in bacteria against foreign DNA from invading pathogens (Gasiunas *et al.*, 2012; Jinek *et al.*, 2012). In bacteria, parts of the invading DNA are being incorporated into the CRISPR locus as spacers and together with a repeat element, they are being expressed as pre-crRNA (Mojica *et al.*, 2005; Bolotin *et al.*, 2005; Marraffini & Sontheimer, 2010). Transactivating RNA (tracrRNA) pairs with the pre-crRNA to form the guide RNA (sgRNA) which directs the Cas9 nuclease complex to the specific targeted part of the invading DNA (Brouns *et al.*, 2008). Cas9 becomes activated upon binding to the sgRNA and specifically cleaves the invading DNA by inducing a double strand break (DSB). An important common region within the spacer is the three nucleotide protospacer adjacent motif (PAM), which is responsible for Cas9 target site recognition (Garneau *et al.*, 2010; Gasiunas *et al.*, 2012).

In 2013, the CRISPR/Cas9 system was adapted for genome engineering in mammalian cells as shown in Figure 1.8 (Cong *et al.*, 2013; Mali *et al.*, 2013). Using an adapted system based on the bacterial endonuclease Cas9 and site-specific 20bp single-guide RNAs (sgRNA), the genome editing tool can target and modify specific sites in the genome. Upon targeting, Cas9 cleaves the DNA and introduces DSB. This DSB is subsequently repaired through the cells own repair mechanism using either homology-directed repair (HDR) or non-homologous end-joining (NHEJ) (Garneau *et al.*, 2010). If a DNA repair template with sequence homology is present, HDR can be exploited to introduce precise edits (Capecchi, 1989; Mali *et al.*, 2013). NHEJ on the other hand is an error prone repair mechanism, thus frequently introducing insertions and deletions (Indels) at the target site. These Indels can induce frameshift mutations leading to premature stop codons and nonsense-mediated decay

(NMS), NHEJ is hence the mechanism to introduce gene knockouts (Lino *et al.*, 2018). CRISPR/Cas9 has revolutionised genome editing by making it simple, affordable and flexible leading to widespread possibilities to develop cell lines and animal models, discover genes involved in disease and cellular function and study patient-specific mutations.

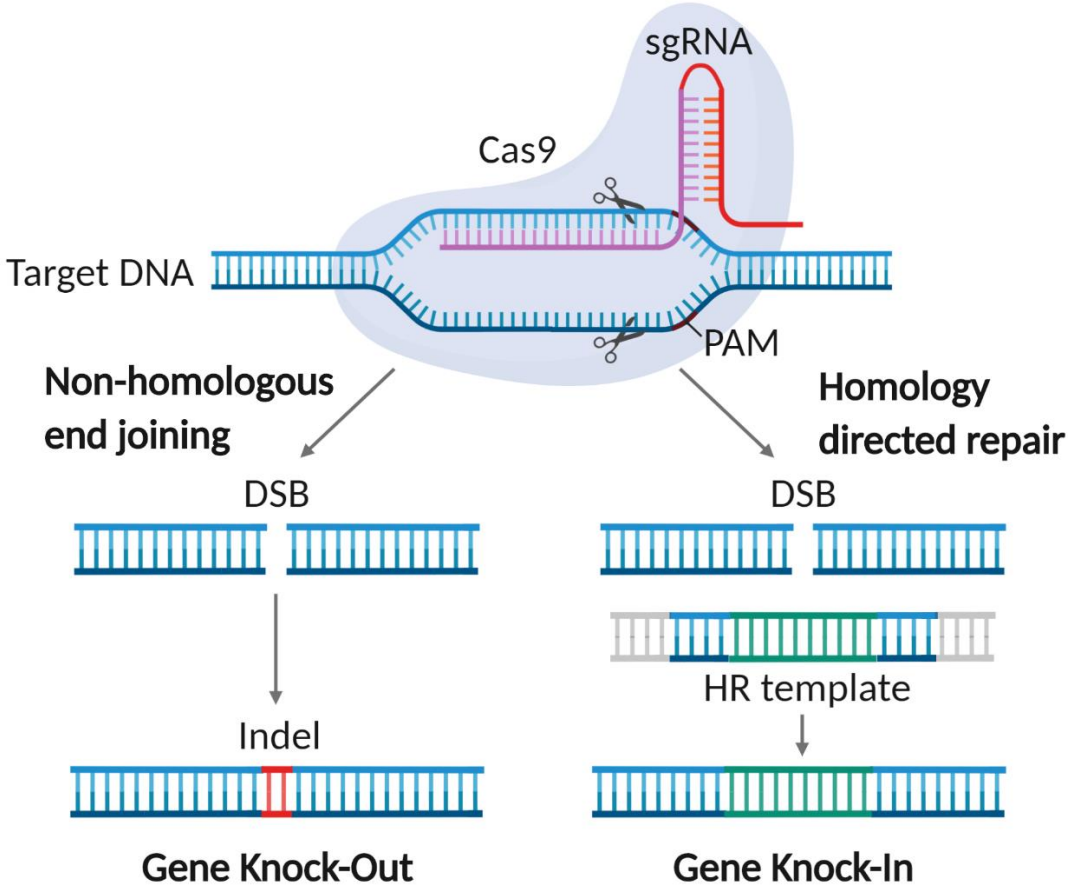


Figure 1.8 – CRISPR/Cas9 genome editing. A sgRNA consisting of a target sequence specific crRNA *(purple)* and a tracrRNA *(red)* that interacts with the endonuclease Cas9 *(grey)*, guides the formed complex to the intended site within the genome. Upon sequence and PAM site recognition, Cas9 initiates sequence-specific double strand cleavage. The PAM region *(brown)*, downstream of the sgRNA sequence is essential for Cas9 induced cleavage. The resulting double strand break will be repaired by DNA repair mechanisms using either non-homologous end joining (NHEJ) or the homology directed repair pathway (HDR). NHEJ is an error-prone repair mechanism that often introduces insertions or deletions (Indel), thus resulting in frameshift mutations and premature stop codons leading to gene KO. If a repair template with homology arms along the DSB has been provided, the cell repairs the DSB based on the template sequence using the precise HDR repair mechanism, therefore allowing to introduce specific mutations or insertions.

The CRISPR system has now been adapted for many other editing approaches using not just Cas9 but also catalytically inactive Cas9 (dCas9), nickase Cas9, other Cas enzymes or Cas fusion proteins (Moon *et al.*, 2019; O'Connell, 2019). In addition to initial Cas9 based genome editing, now the technologies include genome-wide CRISPR screens, base editing without DSB, epigenome editing, RNA targeting, inducible Cas9 systems, gene regulation based on CRISPR interference (CRISPRi) and CRISPR activation (CRISPRa) (Qi *et al.*, 2013; Gilbert *et al.*, 2013; Maeder *et al.*, 2013; Shalem *et al.*, 2014; Wang *et al.*, 2014b; Nihongaki *et al.*, 2015; Kearns *et al.*, 2015; Komor *et al.*, 2016; Adli, 2018; Liu *et al.*, 2020).

1.5.3 Genome-wide pooled CRISPR/Cas9 screens

The development of the CRISPR/Cas9 technology has also facilitated large-scale and affordable genetic screens in mammalian cells in the form of genome-wide pooled CRISPR screens (Koike-Yusa *et al.*, 2014; Shalem *et al.*, 2014; Wang *et al.*, 2014b; Doench, 2018). In arrayed genetic screens, each perturbation is being performed in an individual well and analysed separately, making large-scale or genome-wide applications expensive and labour intensive. In pooled CRISPR screens on the other hand, all cells are located within the same shared cellular vessel but are being individually perturbed with titrated pooled perturbation reagents so each cell will only have one genetic modification (Shalem *et al.*, 2015). This pooled setup provides a simpler and manageable approach for large-scale and cost-effective genetic screens. Perturbation libraries are delivered to the cells via lentivirus and stably integrate into the genome, allowing for a later association of specific phenotypes with individual perturbations. Following a phenotypic selection of interest, the pooled population splits into separate groups. The specific sgRNAs in each population are being identified using next generation sequencing and enriched or depleted sgRNAs give an indication about the importance of a certain gene for the specific phenotype (Doench, 2018). Pooled screens rely on a suitable perturbation system, cellular model and a relevant phenotype.

The most common perturbations in pooled screens are LoF screens using CRISPR/Cas9 to induce gene KO (Shalem *et al.*, 2014, 2015; Wang *et al.*, 2014b). Additional technologies include dCas9 systems to only modulate gene expression levels instead of complete gene KO (CRISPRi), induce GoF expression (CRISPRa) and screens focusing on epigenetic modulations (Gilbert *et al.*, 2014; Konermann *et al.*, 2015; Kearns *et al.*, 2015; Rajagopal *et al.*, 2016). Several predesigned genome-wide and subset sgRNA libraries have been made available for each perturbation approach with differences in the number of sgRNAs per gene, sgRNA design and off-target considerations (Sanjana, Shalem, & Zhang, 2014; Shalem *et al.*, 2014; Hart *et al.*, 2015a, 2017; Doench *et al.*, 2016). The cellular model in a genome-wide screen must be relevant to the cellular phenotype or disease of interest. In addition, the cell system must fulfil practical considerations such as being scalable, therefore often excluding primary cells for genome-scale screens. Extensive cell-type specific optimisation is required for each step of the CRISPR screening protocol to confirm its suitability for screening, often taking longer than the actual screening process (Doench, 2018).

Phenotypic readouts of CRISPR screens include positively or negatively selecting a population of interest (Figure 1.9). Negative selection screens analyse genes that are depleted from the final population, indicating a loss of cellular fitness upon gene perturbation and identifying genes that are critical for cell survival in the specific screening context. These screens have been widely adopted to study essential genes in cancer cell lines to identify cancer dependencies and potential drug targets (Cheung *et al.*, 2011; Hart *et al.*, 2015b; Tsherniak *et al.*, 2017; Behan *et al.*, 2019; Dempster *et al.*, 2019a). Positive selection screens on the other hand enrich genes of interest due to a growth or survival advantage under selective pressure from specific culturing conditions or small molecules (Ma *et al.*, 2015; Jain *et al.*, 2016; Zhang *et al.*, 2016; Wang *et al.*, 2017). The majority of the population dies in a positive selection screen with a small proportion of sgRNAs surviving, whereas

most cells survive under a negative selection while the depleted sgRNAs are of interest. In addition, pooled CRISPR screens can select populations based on a phenotypic cellular marker of interest, which are separated by fluorescence-activated cell sorting (FACS) (Dejesus *et al.*, 2016; Park *et al.*, 2019; Cortez *et al.*, 2020). FACS based CRISPR screens are a powerful tool to identify genes directly influencing a disease relevant phenotype, but the assay requires extensive optimisation beforehand such as for the staining and sorting conditions. FACS based CRISPR screens also restrict the phenotypic readout to cellular proteins where highly specific antibodies are available.

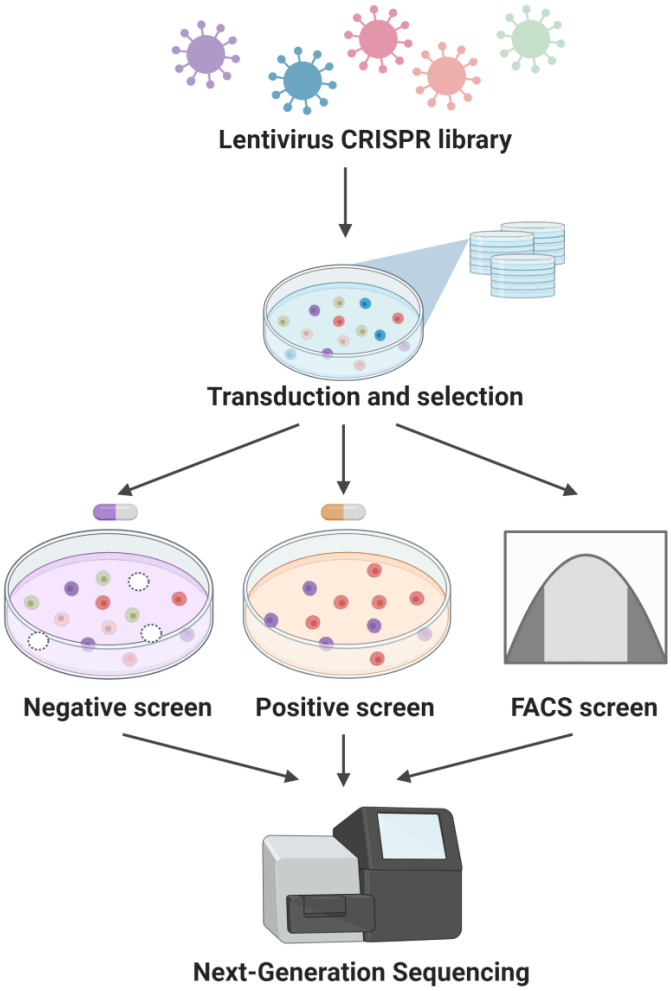


Figure 1.9 – Pooled CRISPR screening. Cells are transduced with a lentiviral pooled CRISPR library at a low MOI, so every cell only has one gene KO and undergo antibiotic selection to remove untransduced cells. In negative screens, a potential condition *(purple)* is applied and only the sgRNAs/cells depleted from the population are of interest. In positive screens, a specific condition *(orange)* induces cell death in most cases and only a small proportion of resistant cells will survive. In addition, FACS screens provide a sophisticated way to separate populations based on the level of a certain protein of interest. All screens are being analysed using Next-Generation Sequencing.

1.5.4 CRISPR screens to study beta cell function

Genetic screens applying CRISPR editing can avoid problems that are associated with the previously described siRNA-based screens as they induce complete protein KO and stable protein depletion. Further, due to the opportunity of performing the screen in a pooled setup as described in section 1.5.3, pooled CRISPR screens can be easily performed on a genome-wide scale, offering an unbiased interrogation of all potential genes of interest. CRISPR editing has been routinely used in iPSC or rodent beta cell lines to model beta cell dysfunction or diabetes, but translating the technique into human beta cell lines or primary islets has proven to be challenging and has not been reported yet (Naylor *et al.*, 2016; Zeng *et al.*, 2016; Atanes *et al.*, 2018; Balboa *et al.*, 2018, 2019a; Xu *et al.*, 2018; Dwivedi *et al.*, 2019). Due to the complex culturing characteristics of EndoC-βH1, as described in section 1.4.5, most studies have focused on siRNA based approaches instead (Chandra *et al.*, 2014; Thomsen *et al.*, 2016a; Ndiaye *et al.*, 2017). Consequently, CRISPR screens in human beta cells have not been approached yet. The first genome-wide CRISPR screen to study genes involved in diabetes and beta cell function was reported in 2019 and was performed in the mouse insulinoma cell line MIN6 (Fang *et al.*, 2019). The authors describe a FACS based genome-wide screen for intracellular insulin using the CRISPR library GeCKOv2. The transduced cells were sorted for the top and bottom 10% of cells in the FACS analysis and two independent screen replicates were performed. The analysis was performed based on clustering the sgRNAs into five tiers according to their significance and reproducibility, from tier-one with p values less than 0.001 for both sgRNAs in the two replicates up to tier-five with both p values between 0.05 and 0.01. Genes were defined as hits if they had at least one tier-one sgRNA, without any contradictory hit sgRNAs, resulting in 373 hit genes. *Ins2*, one of the two mouse orthologous of the insulin gene was among the top hits along with genes known to be involved in insulin secretion such as *Glp1r* and *Xbp1*. The authors

validated specific screening results using individual sgRNAs and the same FACS based insulin readout, claiming to confirm 43 out of 53 hits. No indication is given about the insulin FACS signal that must be achieved to be classified as successfully validated compared to control sgRNAs. KO of many genes seem to only induce a minimal shift in insulin signal or no shift at all, raising doubt about the validation method. Protein interaction networks of the 373 tier-one gene hits identified nine clear protein clusters including GPCR signalling (*Glp1r, Gnas, Nme1, Adcy6*), synaptic vesicle exocytosis (*Vamp2, Snap25, Sytl4, Stxbp1*), mitochondrial ATP production (*Ndufa6, Ndufa9, Ndufb3, Ndufv1*) and protein translation and processing (*Rps25, Eef2, Srp9, Xbp1*). In functional follow-up studies, the authors then focus on genes within the previously unrecognized protein clusters of the cohesin function module and the NuA4/Tip60 histone acetyltransferase (HAT) complex. They demonstrate that the cohesion proteins *Mau2* and *Nipbl* regulate *Ins2* expression and reduce insulin content. This study shows that a genome-wide CRISPR screen to study beta cell function is feasible but due to the highlighted differences between rodents and humans (section 1.4.4), extending the approach into models of human beta cells will be indispensable to better understand genes and pathways regulating human beta cell function.

1.6 Thesis outline

My thesis focuses on identifying genes that are causal for T2D risk or play an important role in beta cell function. Harnessing recent advances in genome editing technologies, I will demonstrate novel pipelines for genome editing in a human beta cell line, both for individual genes and genome-wide. Using a high-throughput genome-wide CRISPR screen, I aim to bridge the gap of connecting GWAS T2D signals to candidate causal genes. Throughout the thesis, I explore various genome editing and silencing approaches to investigate the strengths and weaknesses of different gene and expression modulations, starting with a CRISPR/Cas9 KO pipeline and progressing to a genome-wide CRISPR screen and siRNA-based follow-up studies. The specific outline of each chapter is shown in Figure 1.10 and as follows:

1. In Chapter 3, I describe a novel CRISPR/Cas9 editing pipeline in the human beta cell line EndoC-βH1 and explore the genomic and functional characteristics of the resulting KO cell lines. This genome editing pipeline supports future functional studies of genes and their role in beta cell function and T2D.
2. In Chapter 4, I develop and perform a genome-wide CRISPR LoF screen to identify genes regulating insulin content and human beta cell function. The screen can support prioritisation of candidate causal genes at T2D GWAS loci and identify novel regulators of beta cell function.
3. In Chapter 5, I functionally follow-up on specific screening hits from the genome-wide CRISPR LoF screen. The genes were selected based on their predicted role as potential causal genes in T2D due to genetic evidence but have not yet been linked to any functional evidence. This functional investigation will be able to validate the findings of the screen and demonstrate a novel role of certain genes in beta cell dysfunction and T2D risk.

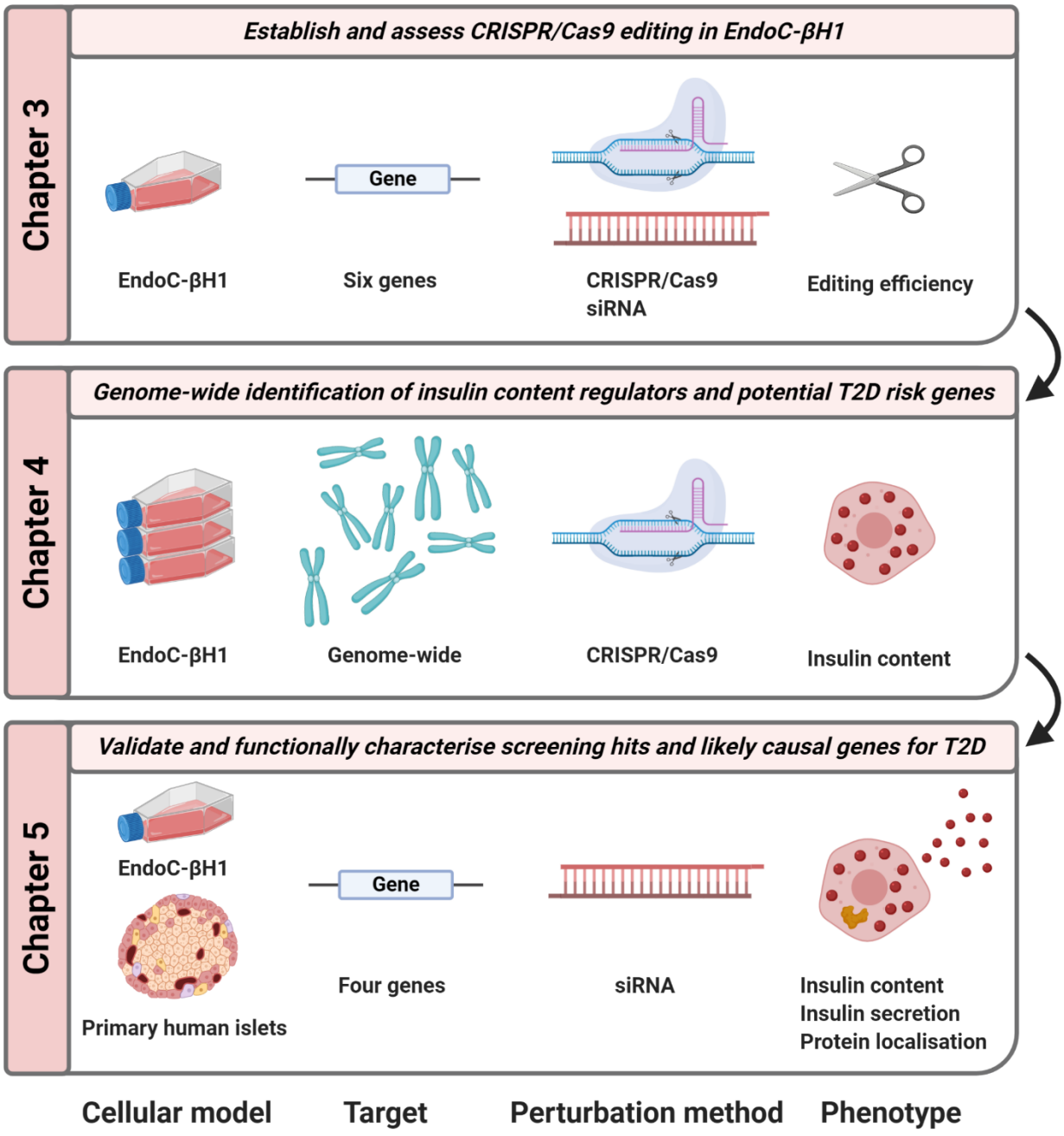


Figure 1.10 – Thesis outline. Chapter 3 *(top)* assessed if genome editing in EndoC-βH1 induces efficient gene KO by targeting six proof of concept genes and compared the permanent protein depletion to a transient siRNA-based strategy. Chapter 4 *(middle)* built on this work by optimising and performing a genome pooled CRISPR screen using a FACS based insulin content readout. Lastly, Chapter 5 *(bottom)* individually validated four screening hits and performed in-depth functional follow-up studies assaying multiple beta cell phenotypes in EndoC-βH1 and primary human islets for a selected gene.

2

Material and Methods

2.1 Cell culture maintenance

EndoC-βH1 were maintained in cell culture flasks coated with 2 μg/ml Fibronectin, 1% extracellular matrix (ECM) and 100 international units (U)/ml penicillin, 100 μg/ml streptomycin (Sigma-Aldrich) in high-glucose (25 mM) Dulbecco's modified Eagle's medium (DMEM, Gibco). They were cultured in freshly made low-glucose (5.5 mM) DMEM (Gibco) containing 2% bovine serum albumin (BSA, Fraction V Fatty acid free, Roche), 2 mM glutamine, 10 mM nicotinamide, 100 U/ml penicillin, 100 μg/ml streptomycin, 50 μM β-2-mercaptoethanol, 5.5 μg/ml transferrin and 6.6 ng/ml sodium selenite (all Sigma-Aldrich). The cells were passaged every 7 days by washing with phosphate buffered saline (PBS, Sigma-Aldrich), incubating with Trypsin-EDTA (Gibco) at 37°C for 4 min followed by neutralisation in 20% foetal bovine serum (FBS, Gibco) in PBS. They were centrifuged at 200 g for 5 min, the cell pellet was washed in PBS and subsequently resuspended in culture media and counted on the automated Cellometer Auto T4 cell counter (Nexcelom Bioscience) before replating at a density of 48,000 cells/cm^2.

Lenti-X HEK293T (Clontech) were grown in high-glucose (25 mM) DMEM (Sigma-Aldrich) containing 10% FBS, 100 U/ml penicillin and 100 μg/ml streptomycin (Gibco). The cells were passaged twice a week and split in a 1:20 ratio.

All cell lines were cultured at 37°C and 5% CO_2 and routinely tested for mycoplasma contamination using the MycoAlert Assay (Lonza).

2.2 Genomic DNA analysis

2.2.1 DNA extraction and quantification

Genomic DNA was extracted from frozen cell pellets using the NucleoSpin Tissue extraction kit (Macherey-Nagel) by following manufacturer's instructions. In brief, cell pellets were lysed at 70°C in lysis buffer containing SDS and Proteinase K, DNA was bound to a silica membrane and subsequently eluted in 20 µl elution buffer. Final DNA concentrations were measured using the NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific). DNA quality was assessed by A260/280 (~1.8) and A260/230 (>2.0) absorbance ratios and only samples within the indicated range were used in downstream applications.

2.2.2 PCR amplification and gel electrophoresis

Polymerase chain reactions (PCR) were prepared to amplify DNA fragments for agarose gel electrophoresis and Sanger Sequencing analysis. Each sample contained 100 ng DNA in a total volume of 10.8 µl and the following reaction components: 2 µl Immobuffer (10x), 0.6 µl MgCl2 (50 mM), 1 µl each of forward and reverse primer (10 µM), 0.4 µl dNTPs (10 mM), 0.2 µl Immolase DNA Polymerase (5 U/µl) (all Bioline) and 4 µl Q-Solution (5x) (Qiagen). The samples were amplified for 10 min at 94°C, 32 cycles of 1 min at 94°C, 1 min at 64°C and 30 sec at 72°C, followed by 10 min at 72°C. PCR samples were run on a 2% agarose gel for 1 h at 120 V and DNA amplification products visualised on a GelDoc transilluminator system (Bio-rad).

2.2.3 Sanger sequencing

To submit samples for Sanger sequencing, PCR products were first purified using exonuclease I (Exo1) and shrimp alkaline phosphatase (SAP) enzymes to remove PCR components that interfere with the sequencing process. Each PCR sample was mixed with 0.05 µl Exo1 (NEB), 0.5 µl SAP, 1 µl 10x SAP buffer (both Agilent) and 0.45 µl nuclease-free H_2O and incubated for 30 min at 37°C followed by 5 min at 95°C. Add 2 µl purified PCR product, 13 µl nuclease-free H2O and 2 µl primer to a Mix2Seq tube (Eurofins Genomics) and sent for sequencing.

2.3 RNA interference

Gene silencing was achieved using a lipid-based small interfering RNA (siRNA) forward transfection strategy. EndoC-βH1 cells were transfected using a pool of four siRNAs at 24h after plating at a final concentration of 15nM (SMARTpool ON-TARGETplus, Dharmacon). siRNA and Lipofectamine RNAiMAX (Invitrogen) transfection mixes were prepared separately in Opti-MEM reduced serum-free medium (Gibco) to account for 0.1% and 0.4% of the same total volume, respectively and incubated for 5 min at room temperature (RT). The dilutions were combined and incubated for further 20 min at RT before dropwise addition to the cells, which were cultured in P/S-free media. Cells were harvested 72 h past transfection for RNA and protein extraction.

2.4 Gene expression analysis

2.4.1 RNA extraction

RNA was extracted using the TRIzol reagent (Invitrogen) following manufacturer's instructions. To obtain enough material for downstream analysis, RNA was extracted from cells cultured in one 6-well or a larger format. Culture media was removed and TRIzol was added directly to the culture plate to detach and lyse cells. Isopropanol precipitation was performed at -20°C overnight and after washing the RNA pellet in diluting concentrations of Ethanol, the pellet was resuspended in RNAse-free H_2O. Final RNA concentrations were measured using the NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific). RNA quality was assessed by A260/280 (~2) and A260/230 (>2.0) absorbance ratios and only samples within the indicated range were used in downstream applications. Extracted RNA was stored at -80°C until cDNA synthesis.

2.4.2 cDNA synthesis

Complementary DNA (cDNA) was synthesised using the GoScript Reverse Transcriptase System (Promega) according to the manufacturer's instructions. 50-500 ng total RNA were used as input material and amplified using oligo(dT) and random primers (Promega). cDNA was stored at -20°C until gene expression analysis.

2.4.3 Real-time TaqMan quantitative PCR

Quantitative qPCR (qPCR) to measure gene expression levels were performed with TaqMan Gene Expression Assays and TaqMan Gene Expression Master Mix on a 7900HT (all Applied Biosystems). cDNA samples were assayed in triplicates per gene and analysed on a 384-well plate. No-template control assays were included for each gene. The following thermocycling conditions were used: 2 min at 50°C, 10 min at 95°C and 40 cycles of 15 sec at 95°C and 1 min at 60°C. Amplification thresholds were verified visually and adjusted if

necessary. Ct-values were analysed using the ΔΔCt method, which uses a two-step normalisation method to obtain expression levels relative to housekeeping genes and control samples. Target genes were normalised to the average of the three housekeeping genes TATA-binding protein (*TBP)*, peptidylprolyl isomerase A *(PPIA)* and glycerinaldehyd-3-phosphat-dehydrogenase (*GAPDH)*.

2.5 Protein expression studies

2.5.1 Cell lysis and protein quantification

To analyse intracellular protein levels, cell pellets were lysed in 30 µl-100 µl RIPA buffer (50 mM Tris pH 7.4, 150 mM NaCl, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS) containing 1x protease inhibitor cocktail (Roche). The samples were incubated for 30 min at 4°C on a tube rotator, followed by centrifugation for 30 min at 13000rpm at 4°C. The supernatant containing the protein lysate was transferred into fresh microcentrifuge tubes and stored at -20°C for short-term storage. Protein concentration was quantified by DC protein assay (Bio-Rad) and absorbance levels, which are representative for protein levels were read on a Versamax microplate reader (Molecular Devices). A standard curve of BSA (Sigma-Aldrich) was used to calculate absolute protein concentrations.

2.5.2 Gel electrophoresis and transfer

Protein samples were size-separated by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). Sample lysates were mixed with loading buffer consisting of 4x Laemmli Buffer (Bio-Rad) and 10% β-mercaptoethanol (Sigma-Aldrich) and denatured at 80°C for 10 min. Per lane, 10 µg of protein were prepared and loaded on a 4-20% Criterion TGX Stain-Free Precast Gel (Bio-Rad). At least one lane of the molecular weight marker Precision Plus All Blue Standard (Bio-Rad) was included. The gel was run for 15 min at 300 V followed by activation on a ChemiDoc MP Imaging System (Bio-Rad). Proteins were

transferred to a 0.2 μm polyvinylidene difluoride (PVDF) membrane using the Trans-Blot Turbo Transfer System (all Bio-Rad) and transfer was confirmed by imaging.

2.5.3 Western Blotting

Unspecific antibody binding was prevented by blocking the membranes in 3% BSA diluted in Tris-buffered saline, supplemented with 0.1% Tween-20 (TBST) for 1 h at RT. The membrane was incubated with primary antibodies overnight at 4°C or for 1 h at RT and with secondary antibodies for 1 h at RT, both with gentle shaking. All antibodies were diluted in 3% BSA in TBST and after each antibody incubation step, the membrane was washed 3x for 5 min with TBST to remove unbound antibodies. Protein bands were detected by incubating the membrane for 4min at RT with Clarity Western enhanced chemiluminescence (ECL) reagent, followed by imaging on the ChemiDoc MP Imaging System (Bio-Rad). To normalise for differences in protein loading, the membrane was further incubated with a loading control antibody of appropriate size (β-tubulin or GAPDH). Band intensities were quantified using the Image Lab 6.0 software (Bio-Rad). Proteins of interest were normalised to their respective loading control on the same blot and displayed relative to a control sample.

2.6 Cellular phenotyping assays

2.6.1 Cell proliferation assay

To assess cell viability and account for differences in cell numbers in insulin secretion experiments, the CyQUANT Direct Cell Proliferation Assay (Thermo Fisher Scientific) was performed by following manufacturer's instructions. Briefly, direct nucleic acid dye (0.2%) and background suppressor dye (1%) were diluted in cell culture medium and the cells were incubated for 1 h at 37°C in the dark. Fluorescence was measured at 480 nm/535 nm (excitation/emission) on an EnSpire Alpha Plate Reader (Perkin Elmer) and fluorescence units were representative of cell numbers.

2.6.2 Insulin secretion assay in EndoC-βH1

Static insulin secretion assays were performed at 37°C and 5% CO_2 and the cells were assayed in 96 well plates in three to six wells per condition and biological replicate. EndoC-βH1 were starved overnight in culture media containing a final concentration of 2.8 mM glucose, followed by 30 min starvation in 0 mM glucose media the next morning. To initiate insulin secretion, the cells were incubated for 1 h in different concentrations of glucose (1 mM to 20 mM Glucose) or secretageous drugs (100 μM Tolbutamide or Diazoxide, Sigma-Aldrich). All conditions were prepared in glucose-free EndoC-βH1 culture medium. Cell count was assayed at this point as described in 2.6.1. Supernatant was collected, spun down at 300g and 4°C for 5 min, half of the total volume was taken off from the top, transferred to a new plate and stored at -20°C until analysis. Following the removal of supernatant, ice-cold acid-ethanol (1.5% HCl, 75% ethanol and 23.5% deionized water) was added to the cells to extract intracellular insulin content. The plates were sealed and stored at -20°C until analysis. All insulin samples were measured using the Insulin (human) AlphaLISA Detection Kit and the EnSpire Alpha Plate Reader in white 96-well 1/2 AreaPlates (all

Perkin Elmer). Supernatant and insulin content samples were diluted 1:10 and 1:200 in assay buffer and ran along a standard curve on each plate. Unknown insulin concentrations were extrapolated using a four-parameter non-linear regression of log-transformed read values in Prism 8 (GraphPad Software). Secreted insulin was displayed as percentage of intracellular insulin and total insulin content was normalised to cell count.

2.7 Lentiviral transduction

2.7.1 Lentivirus production

To produce lentivirus, Lenti-X HEK293T cells were grown in P/S free media to 80% confluency in T175 flasks and co-transfected with the lentiviral packaging vectors pMD2.G (6.85 μg) (Addgene #12259), psPAX2 (10.3 μg) (Addgene#12260), the plasmid containing Cas9 and the sgRNA of interest or the CRISPR library (12.85 μg), 2 ml of JetPrime buffer and 60 μl of JetPrime transfection reagent (Polyplus transfection) per flask. The transfection mix was added to the cells after preincubation at RT for 15 min and the cell culture media was replaced after 16 h into fresh complete culture media. The supernatant containing the viral particles was harvested 48 h after transfection, spun down for 5 min at 2000 rpm to remove any cells or debris and further filtered through a 0.45 μm filter. To concentrate the virus, the supernatant was ultracentrifuged for 2 h at 4°C and 29000 rpm in a swinging-bucket rotor. The ultracentrifugation was performed by Dr Elena Navarro-Guerrero at the Target Discovery Institute (TDI), University of Oxford. The virus pellet was resuspended in 1.5% BSA in PBS, aliquoted to avoid freeze-thaw cycles and stored at -80°C.

2.7.2 Functional viral titer

To determine the amount of infectious viral particles within the virus, the functional titer was determined by measuring the survival after transduction with different viral dilutions and antibiotic selection. EndoC-βH1 were plated at a density of 20000 cells per well in 96 well plates at 48 h before transduction. The cells were transduced for 6 h in 100 µl P/S free media containing different viral dilutions from 1:50 to 1:6400. After 48 h, media was changed on half of the wells per dilution into 4 µg/µl puromycin and the cells were incubated for 7 days. Cell viability was subsequently measured using the CyQUANT Direct Cell Proliferation assay (Invitrogen). Cell counts in puromycin selected wells were normalised to their respective non-selected controls to determine the percentage of survival, which is representative for transduced cells, while also accounting for toxicity of the added viral dilution. The functional titer in transducing units (TU)/µl can then be calculated based on:

(1)
$$\mathrm{TU}/\mu l = \frac{\#\mathrm{Cells} \times m}{\mathrm{Virus\ (\mu l)\ used\ in\ transduction}}$$

The Poisson distribution (PD) (Equation 2) can be used to model the probability *P(k)* that a cell is infected by a certain number *(k)* of viral particles at a given multiplicity of infection (MOI) (*m*) (Figliozzi *et al.*, 2016).

(2)
$$P(k) = \frac{e^{-m}mk}{k!}$$

Simplifying the original PD equation, gives the following:

(3)
$$P(k > 0) = 1 - e^{-m}$$

with *P(k>0)* being the probability that a cell gets infected by at least one viral particle.

Solving equation 3 for a MOI of 0.3 leads to ~26% transduced cells and most of them being infected by a single viral particle (k=1 in equation 2), it is therefore a good constant in determining the functional titer. To apply this constant to the functional titer experiment and determine the MOI relative to the virus (µl) used in transduction, a linear regression for the percentage of alive cells against the amount of infected virus was performed in the linear, unsaturated range of the puromycin selection curve. The linear equation was then solved for a MOI of 0.3 by inserting 26% as the percentage of alive cells, resulting in the amount of virus (µl) needed in the transduction. The TU/µl could then be determined by solving equation 1 along with the number of plated cells.

2.7.3 Transduction of EndoC-βH1

The required amount of virus for a specific MOI was calculated based on the functional titer (2.7.2) and the respective cell numbers. EndoC-βH1 were transduced for 6 h and selected for seven days in 4 µg/ml puromycin. After antibiotic selection, cells were grown and passaged following routine EndoC-βH1 culture conditions.

2.8 Statistical analysis

Statistical analyses were performed in Prism 8.1 (GraphPad Software) and data are shown as mean with standard error of the mean (SEM). For some experiments, fold changes were plotted for better visualisation but statistical tests were performed on log-transformed values. To reduce experimental technical variation such as in western blot data, sample values were normalised to their respective control group and analysed using one-sample Student's t-test. To compare one normally distributed variable between two or more groups, two-sample Student's t-test or one-way analysis of variance (ANOVA) followed by Sidak's multiple comparison test were used, respectively. Results were considered statistically significant if $p<0.05$. The number of biologically independent experiments is indicated as n. Details about statistical tests for individual experiments can be found in the respective method sections or figure legends.

3

A Genome Editing Pipeline in the Human Beta Cell Line EndoC-βH1

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

1. **Grotz AK** et al. A CRISPR/Cas9 genome editing pipeline in the EndoC-βH1 cell line to study genes implicated in beta cell function. *Wellcome Open Research* 2019; 8(4): 150
2. Dwivedi OP et al. Loss of ZnT8 function protects against diabetes by enhanced insulin secretion. *Nature Genetics* 51(11): 1596-1606

3.1 Introduction

3.1.1 Genome editing in human beta cells

The majority of T2D loci from GWAS have been shown to be associated with pancreatic islet function through colocalisation with islet specific regulatory or epigenetic annotations and physiological clustering analysis (Ingelsson *et al.*, 2010; Dimas *et al.*, 2014; Pasquali *et al.*, 2014; Mahajan *et al.*, 2018b; Thurner *et al.*, 2018; Udler *et al.*, 2018). The pancreatic islet and more specifically the beta cell are therefore central in mediating the effects of genetic variants on T2D risk. Studying human beta cell function *in vitro* has long been challenging due to the lack of a suitable human beta cell line along with inadequate or restricted alternative models such as primary human islets. Rodent beta cell lines on the other hand have been fundamental in promoting a better understanding of beta cell function and T2D pathogenesis (Efrat *et al.*, 1988; Miyazaki *et al.*, 1990; Asfari *et al.*, 1992). Direct translation into human physiology and beta cell function is often limited due to the species dependant genetic background along with structural, functional, and transcriptional differences between human and rodent beta cells (Shiao *et al.*, 2008; Bosco *et al.*, 2010; McCulloch *et al.*, 2011; Kulkarni *et al.*, 2012; Benner *et al.*, 2014; Henquin *et al.*, 2015; Baron *et al.*, 2016; Rorsman & Ashcroft, 2018). The development of the EndoC-βH1 cell line in 2011 has eventually provided the opportunity to study human beta cell physiology and pathology *in vitro* (Ravassard *et al.*, 2011). Extensive independent characterisation of the cell line has demonstrated its similarity with primary islets and beta cells regarding their epigenomic, transcriptomic, electrophysiological, and secretory properties and thus its suitability as a model to study human beta cell function (Andersson *et al.*, 2015; Gurgul-Convey *et al.*, 2015; Hastoy *et al.*, 2018; Tsonkova *et al.*, 2018; Lawlor *et al.*, 2019). The cell line however, has not been used for genome editing so far and robust protocols for CRISPR/Cas9 editing have not been described yet, hindering routine genetic manipulation

to study the effects of gene KO of T2D risk genes similar to pipelines that have been successful in rodent beta cell lines (Naylor *et al.*, 2016; Atanes *et al.*, 2018; Xu *et al.*, 2018). The lack of genome editing protocols is due to the cell lines challenging culture characteristics such as a low proliferation rate and high sensitivity to seeding densities which preclude the generation of a clonal cell line combined with batch-to-batch variation and the need for close monitoring across passages.

3.1.2 Experimental outline

In this study, I aim to develop a robust CRISPR/Cas9 editing pipeline in EndoC-βH1 to generate KO cell lines for functional studies of human beta cells and as a proof of concept for other CRISPR based studies such as genome-wide screens. I designed a lentiviral-based dual sgRNA pipeline for genome editing in EndoC-βH1 and characterised the genomic and functional implications of the proof of concept KO cell lines for *INS*, *PAM* and *IDE*. To highlight distinct phenotypes resulting from siRNA based KD or CRISPR/Cas9 KO approaches in EndoC-βH1, I performed a direct comparison for the genes *SLC30A8* and *NEUROD1*. I characterised specific relevant beta cell readouts such as markers of ER stress and apoptosis for *NEUROD1* or insulin secretion and the expression of K_{ATP} channels in *SLC30A8*. In addition to creating a CRISPR pipeline in EndoC-βH1 for KO cells, I demonstrated its suitability for further CRISPR studies by also assessing low MOI transductions similar to protocols used in genome-wide LoF or GoF screens.

3.2 Material and Methods

3.2.1 RNA interference

Transient gene silencing in EndoC-βH1 was performed using siRNAs (*NeuroD1* #L-008667-00, *SLC30A8* #L-007529-01, non-targeted (NT) control pool #D-001810-10, Dharmacon). Experimental conditions are described in more detail in section 2.3. Samples were harvested after 72 h and knockdown was confirmed on both the mRNA level by RT-qPCR and the protein level by western blot.

3.2.2 Genomic DNA analysis

DNA was extracted and PCR amplified as described in section 2.2 using primers depicted in Table 3.1. sgRNA cutting efficiencies and off-target effects in CRISPR KO cells were validated using Tracking of Indels by DEcomposition (TIDE) analysis (Brinkman *et al.*, 2014). Sanger sequencing traces of PCR amplified genomic regions for an EV and CRISPR KO sample were uploaded to https://tide.deskgen.com, resulting in a quantification of all editing events and an insertion and deletion (Indel) profile.

Target	Target Region	Sequence (5' → 3')
LKO1.5R	lentiCRISPRv2	GTTGATAACGGACTAGCCT
Lenti_Cas9	*Cas9*	CAGGCCGATGCTGTACTTCT
PAM_sgRNA1	sgRNA1	ACACCGGAACTAGCAGGCTA
PAM_sgRNA2	sgRNA2	GACGAAACACCGGTTCAGAAC
IDE_sgRNA3	sgRNA3	AAACACCGTACCCACACAGG
IDE_sgRNA4	sgRNA4	CGCATTAATGTGGACTTGACCG
INS_sgRNA5	sgRNA5	CAATGCCACGCTTCTGCAG
INS_sgRNA6	sgRNA6	CATCTGCTCCCTCTACCAGC
NMS_sgRNA7	sgRNA7	AAACACCGTGCAGGATCACA
NMS_sgRNA8	sgRNA8	ACACCGACAATATCCAAGCCA
NEUROD1_sgRNA9	sgRNA9	GACGAAACACCGCTTGCAAA
NEUROD1_sgRNA10	sgRNA10	CTGTAGGCGTGCGGGTTTT
PAM1_F	sgRNA1 target site	GCTGGAGGGAGGAAAGCTTC
PAM1_R	sgRNA1 target site	TTTTTCTGCACGGGGGACTT
PAM2_F	sgRNA2 target site	TTGCTGGCAGATCTAAGGGC
PAM2_R	sgRNA2 target site	TCCCTGGCTGAGATTTTCCTC
IDE3_F	sgRNA3 target site	AGTCGCCGGATTCCTTTACC
IDE3_R	sgRNA3 target site	CTAATGCGGTACCGGCTAGC
IDE4_F	sgRNA4 target site	TCCATGAAACAAAGGCCAAGT
IDE4_R	sgRNA4 target site	CCCCACTTCTGCACCATCTT
INS5_F	sgRNA5 target site	CATCTCTCTCGGTGCAGGAG
INS5_R	sgRNA5 target site	TCCCTCTAACCTGGGTCCAG
INS6_F	sgRNA6 target site	CCTGTAGGTCCACACCCAGT
INS6_R	sgRNA6 target site	AAGACACACAGACGGCACAG
NMS7_F	sgRNA7 target site	ATTCAGATTGCTTGGCCCCA
NMS7_R	sgRNA7 target site	TGGTCTACCTTGTGTCTGTCTG
NMS8_F	sgRNA8 target site	CCCACTGACCTCTTTTGAATGG

NMS8_R	sgRNA8 target site	AGCTTAAATTCTTCAGTCAGCA
sgRNA1_MED15_F	sgRNA1 off-target	GGCCAAACACACAGAGGAGT
sgRNA1_MED15_R	sgRNA1 off-target	TGGACTTGCCCTCTCTTGAC
sgRNA2_GPM6B_F	sgRNA2 off-target	ATCACTGCAGGGAACTGCTT
sgRNA2_GPM6B_R	sgRNA2 off-target	CAGCACCATCCTCAGATCCT

Table 3.1 – Primer specifications

3.2.3 Gene expression analysis

RNA was extracted from EndoC-βH1 cells and converted into cDNA using the GoScript Reverse Transcriptase System (Promega) as described in more detail in section 2.4. mRNA expression levels were assessed using TaqMan Gene Expression Assays (Applied Biosystems, Table 3.2), analysed based on the ΔΔCt method and normalised to the housekeeping genes *TBP*, *PPIA* and *GAPDH*. In CRISPR KO cell lines, TaqMan probes with binding sites outside of the regions targeted by sgRNAs were used.

TaqMan Probe	Assay details	Target region
INS	Hs00355773_m1	Exon 1-2
IDE	Hs00610452_m1	Exon 24-25
PAM	Hs01084034_m1	Exon 22-23
NeuroD1	Hs01922995_s1	Exon 2
NMS	Hs02340859_m1	Exon 5-6
GAPDH	Hs02786624_g1	Exon 8
TBP	Hs00427620_m1	Exon 3-4
PPIA	Hs01634221_s1	Exon 1
SLC30A8	Hs00545182_m1	Exon 2-3
DDIT3	Hs99999172_m1	Exon 1-2
XBP1s	Hs03929085_g1	Exon 5
HSPA5	Hs00607129_gH	Exon 1-2
ATF6	Hs00232586_m1	Exon 6-7
ATF4	Hs00909569_g1	Exon 1-2

Table 3.2 – TaqMan gene expression assays

3.2.4 Protein expression analysis

Protein samples were extracted from EndoC-βH1 cells using RIPA buffer and size-separated by SDS-PAGE as described in 2.5. To detect SLC30A8, protein samples were only mixed with SDS-PAGE loading buffer without heat denaturation. Protein was detected by western blotting using primary and secondary antibodies diluted in 3% BSA (Table 3.3). Protein quantification was performed by normalising sample protein levels to their respective loading control (GAPDH or β-Tubulin) and control sample (non-targeting control in siRNA experiments and EV in KO experiments).

Antibody	Dilution	Species	Manufacturer details
β-Tubulin	1:2000	Mouse monoclonal	Santa Cruz, sc-365791
GAPDH	1:10 000	Rabbit monoclonal	Abcam, ab37168
NEUROD1	1:1000	Mouse monoclonal	Santa Cruz, sc-46684
INS	1:1000	Mouse monoclonal	Santa Cruz, sc-393887
ZnT8	1:1000	Mouse monoclonal	(Merriman *et al.*, 2018)
CHOP	1:1000	Rabbit monoclonal	Abcam, ab179823
pIRE1	1:1000	Rabbit polyclonal	Abcam, ab48187
pPERK	1:1000	Rabbit monoclonal	Cell Signaling, #3179
Cleaved Caspase 3	1:500	Rabbit polyclonal	Cell Signaling, #9661
Cas9	1:1000	Mouse monoclonal	Santa Cruz, sc-517386
PAM	1:1000	Mouse monoclonal	Santa Cruz, sc-514110
IDE	1:1000	Mouse monoclonal	Santa Cruz, sc-393887
Anti-mouse IgG HRP	1:2500	Rabbit polyclonal	Thermo Fisher, 31460
Anti-rabbit IgG HRP	1:2500	Goat polyclonal	Thermo Fisher, 31450

Table 3.3 – Western Blot antibodies and dilutions

3.2.5 Insulin secretion assays

Static insulin secretion assays in EndoC-βH1 KO and control cells were performed in 96 well plates as described in section 2.6.2. The amount of insulin in the supernatant or lysed cells was measured using the AlphaLISA Detection Kit (Perkin Elmer). Secreted insulin was normalised to intracellular insulin and insulin content was normalised to cell count which was assessed using the CyQUANT Direct Cell Proliferation Assay (Thermo Fisher Scientific).

3.2.6 Generation of EndoC-βH1 KO cell lines

3.2.6.1 Cloning of sgRNA into plentiCRISPRv2

sgRNA sequences for individual KO cell lines were retrieved from the Toronto KnockOut CRISPR library version 3 (TKOv3) (Hart et al., 2017) and the plasmid backbone plentiCRISPRv2 was purchased from Addgene (#52961) (Sanjana *et al.*, 2014). Two sgRNAs per gene were chosen based on highest specificity and lowest off-target score which were evaluated on CRISPOR.org (Table 3.4) (Haeussler *et al.*, 2016). To clone the sgRNAs oligonucleotides into the Cas9 and puromycin resistance containing backbone (Figure 3.1), BsmBI compatible tails, 5'CACCGX3' and 5'AAACYC3', with X and Y being complementary sequences of the sgRNA, were added.

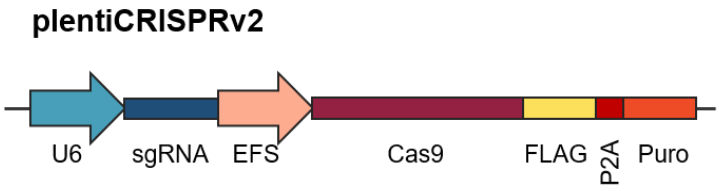


Figure 3.1 – plentiCRISPRv2 plasmid. sgRNAs were cloned into the plentiCRISPRv2 plasmid containing a flag-tagged Cas9 and puromycin resistance cassette (*Puro*) which are separated by the self-cleaving P2A sequence. sgRNA and Cas9 expression are driven by U6 and elongation factor 1α short (EFS) promoter (*arrows*) (Sanjana *et al.*, 2014).

To prepare plentiCRISPRv2 for sgRNA insertion, the vector was digested with FastDigest BsmBI (Fermentas) for 30 min at 37°C and gel-purified using a 0.8% agarose gel. The complementary sgRNA oligonucleotides were annealed (1 µl of each 100 µM stock) and phosphorylated using T4 PNK (NEB) for 30 min at 37°C, 5 min at 95°C, then the heating block was shut off to let the samples could cool down to RT. The BsmBI digested plentiCRISPRv2 (20 ng) and annealed sgRNA oligonucleotides (2 µl of 1:100 dilution) were ligated by incubation with Quick Ligase (NEB) for 1 h at RT. Stbl3 competent cells were transformed with 5 µl of the ligation reaction for plasmid amplification. Bacterial colonies

were further amplified in overnight cultures in 10 ml LB broth (Sigma-Aldrich) containing 100 µg/ml ampicillin. To extract the plasmid DNA for downstream viral production, PureYield Plasmid Miniprep kit (Promega) were used following the manufacturer's instructions. Correct sgRNA integration was confirmed using the LKO1.5R (GTTGATAACGGACTAGCCT) primer in Sanger Sequencing as described in section 2.2.3. sgRNA plasmids were packaged into lentivirus and used to transduce EndoC-βH1 and create KO cell lines.

sgRNA	Target	Target region	Sequence (5' → 3')
sgRNA 1	*PAM*	Exon 1	GAACTAGCAGGCTAGGGACG
sgRNA 2	*PAM*	Exon 12	GTTCAGAACCATACCACCAG
sgRNA 3	*IDE*	Exon 1	TACCCACACAGGCGCTCCGG
sgRNA 4	*IDE*	Exon 8	CATTAATGTGGACTTGACCG
sgRNA 5	*INS*	Exon 2	CACAATGCCACGCTTCTGCA
sgRNA 6	*INS*	Exon 2	CATCTGCTCCCTCTACCAGC
sgRNA 7	*NMS*	Exon 7	TGCAGGATCACACTGCGACC
sgRNA 8	*NMS*	Exon 2	ACAATATCCAAGCCATCTGA
sgRNA 9	*NEUROD1*	Exon 2	CTTGCAAAGCGTCTGAACGA
sgRNA 10	*NEUROD1*	Exon 2	GCTGCGCTGTAGGCGTGCGG
sgRNA 11	*SLC30A8*	Exon 2	GTGTCCCAGAGAGAGACCAG
sgRNA 12	*SLC30A8*	Exon 8	GCACTCACTCACCATTCAGA

Table 3.4 - sgRNA sequences for target genes

3.2.6.2 Transduction of EndoC-βH1 KO cells

Lentivirus was produced in HEK293T cells for each construct, concentrated and titered individually to determine the amount of infectious viral particles as described in section 2.7. Stable cell lines were generated by transducing EndoC-βH1 at a high MOI of 8, followed by selection in 4 μg/ml puromycin for seven days. The cells were then cultured routinely in normal EndoC-βH1 culture medium and passaged weekly. A control cell line referred to as empty vector (EV) control was created in parallel using the plentiCRISPRv2 plasmid without a gene-targeted sgRNA.

3.2.7 Statistical analysis

Statistical analyses were performed in Prism 8.1 (GraphPad Software). Data were normalised to either their respective non-targeting control in siRNA experiments or EV in KO experiments. To account for technical variation, western blot quantifications were normalised within their individual technical replicate. Statistical significance was tested in these samples using a one-sample Student's t-test and the control sample is indicated as dotted line at 100% in the plot. Comparison of two groups such as gene expression differences were analysed using two-sample Student's t-test and more than two groups such as insulin secretion across conditions were analysed using ANOVA followed by Sidak's multiple comparison test. Fold changes were log-transformed for statistical analysis but plotted as percentage of control.

3.3 Results

3.3.1 A CRISPR/Cas9 pipeline to create EndoC-βH1 KO cells

To be able to generate KO cell lines in EndoC-βH1 and create the opportunity to study the role of specific genes in human beta cell function *in vitro*, I developed a robust genome editing pipeline in EndoC-βH1 and characterised three proof of concept genes. The genes peptidyl-glycine alpha-amidating monooxygenase (*PAM*), insulin-degrading enzyme (*IDE*) and *INS* are of importance in the beta cell and therefore represent genes that could be of interest to target in this pipeline (Tager *et al.*, 1979; Sladek *et al.*, 2007; Stoy *et al.*, 2007; Steneberg *et al.*, 2013; Fuchsberger *et al.*, 2016; Thomsen *et al.*, 2018). Briefly, EndoC-βH1 cells were transduced with lentivirus for Cas9 and two sgRNAs per gene, followed by antibiotic selection to remove untransduced cells and characterisation of the resulting stable KO cell lines (Figure 3.2).

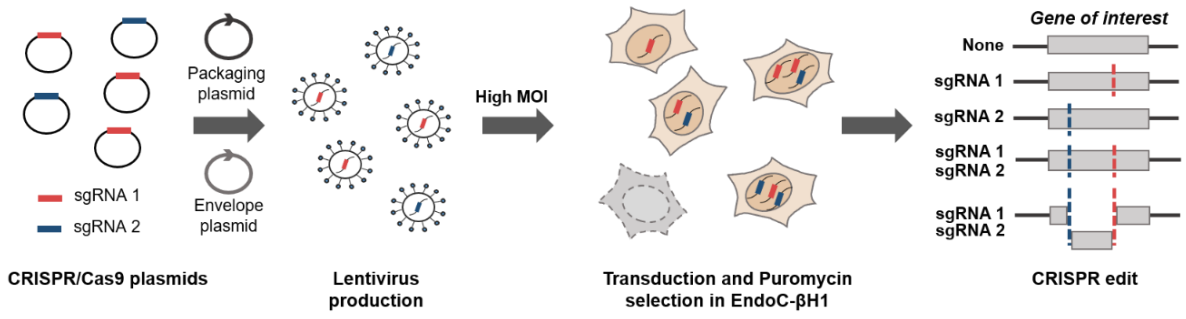


Figure 3.2 – Pipeline to generate EndoC-βH1 KO cells. Two sgRNAs per gene (*red* and *blue*) were cloned individually into plentiCRISPRv2 and separate lentivirus stocks were produced and titered. EndoC-βH1 cells were transduced with the pooled virus at a high MOI of 8 and selected in puromycin. The cells were routinely cultured followed by genomic characterisation to assess if the gene of interest was edited by none, one or both sgRNAs, individually or simultaneously.

The plasmid plentiCRISPRv2 was used in the pipeline as it is a one-vector system expressing Cas9, a puromycin resistance cassette and the cloned sgRNA (Sanjana *et al.*, 2014). Instead of using a single sgRNA per gene, a dual sgRNA strategy using two separate lentiviruses was chosen to increase the KO efficiency. The sgRNAs sequences targeting different parts or exons of each gene were retrieved from the genome-wide CRISPR KO library Toronto KnockOut version 3.0 (TKOv3) which are designed for high on-target and low off-target efficiency (Figure 3.3 A-C). The sgRNAs for *IDE* were chosen to target a specific isoform within the gene to demonstrate the suitability of the pipeline to generate isoform specific KO cells.

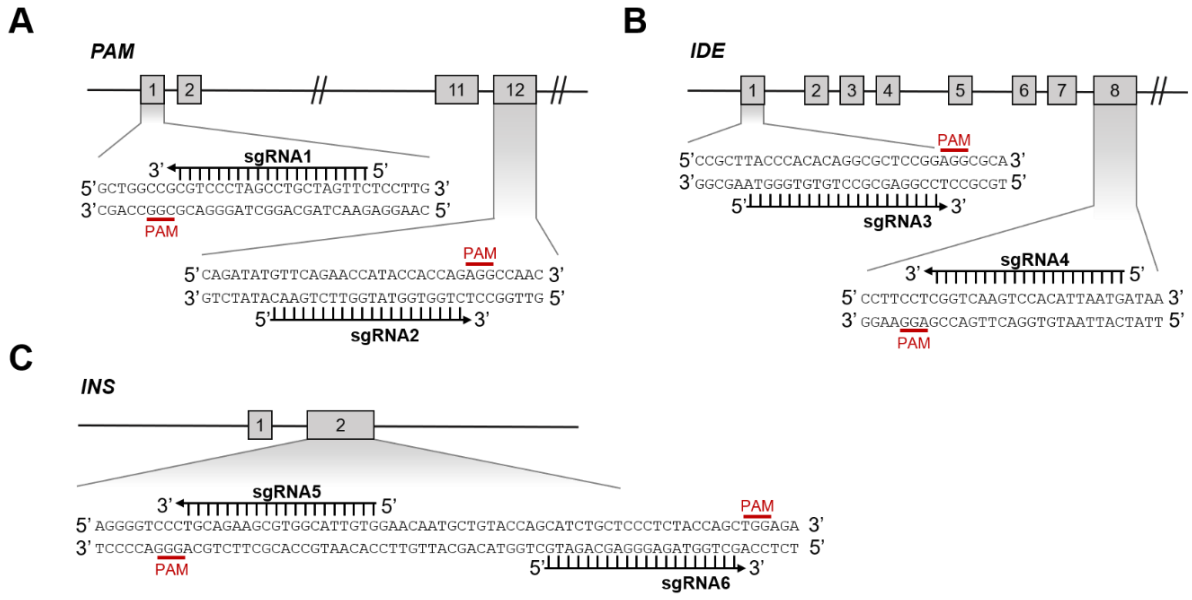


Figure 3.3 – Gene targeting strategy. Two sgRNAs were chosen per gene for *PAM* **(A)**, *IDE* **(B)** and *INS* **(C)**. Exons within each gene are indicated as grey blocks labelled with exon numbers. Specific regions are enlarged to show specific sgRNA target sides and PAM motive *(red)* for Cas9 targeting.

The sgRNAs were cloned into plentiCRISPRv2, packaged into lentivirus and the EndoC-βH1 cells were transduced at a MOI of 8. By using this high MOI, each cell is on average targeted by several lentiviral particles resulting in multiple vector integrations and a higher probability of achieving a gene KO. The viral transduction was performed without the transduction enhancer polybrene, as titration in EndoC-βH1 revealed a high induced toxicity (Figure 3.4 A). To eliminate untransduced cells, transduced EndoC-βH1 were selected in 4 μg/ml puromycin, which was sufficient to induce nearly 100% cell death in untransduced cells within seven days (Figure 3.4 B). For each independent puromycin selection in a different cell passage, the antibiotic dose was determined once again to ensure complete cell death as EndoC-βH1 demonstrated varying susceptibilities depending on their time in culture.

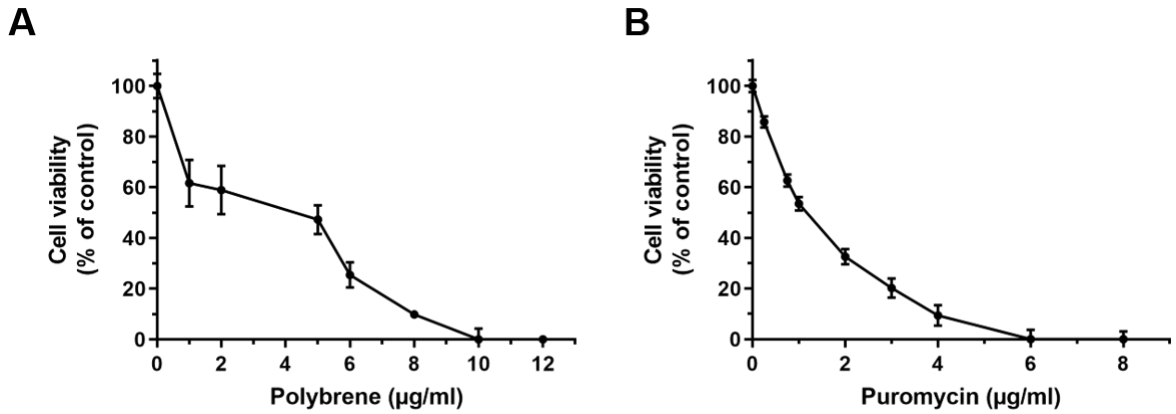


Figure 3.4 – Polybrene and puromycin kill curve. Sensitivity to the transduction enhancer polybrene **(A)** and the antibiotic puromycin **(B)** was determined in EndoC-βH1. **(A)** The cells were incubated with polybrene for 6 h and cell viability was measured after 7 d. **(B)** Puromycin selection was performed for 7 d. Cell viability is normalised relative to vehicle control cells and all data are mean ± SEM from six technical replicates.

The resulting KO cell lines after puromycin selection were cultured like regular EndoC-βH1 cells while their genomic and functional properties were examined. As EndoC-βH1 cells cannot be clonally expanded, these KO cell lines were a heterogenous population with either no edits, Indels based on single sgRNA editing, Indels from dual sgRNA editing or large deletions from simultaneous dual sgRNA cutting. Alongside the KO cell lines, an empty vector (EV) control cell line was created using the same plentiCRISPRv2 backbone but without a sgRNA to assess the effects of stable Cas9 expression and antibiotic selection. In addition, a KO cell line targeting the gene Neuromedin-S (*NMS*) was generated which is not expressed in EndoC-βH1 cells and acts as a control for effects from Cas9 induced DSB.

3.3.2 Genomic modifications of EndoC-βH1 KO cells

To assess the genomic modifications of the KO cell lines for *IDE, PAM* and *INS*, stable integration and expression of each sgRNA and Cas9 were examined followed by characterising the editing efficiency at individual sgRNA target and potential off-target sites. Stably integrated sgRNAs were detected with a PCR-based amplification approach (PCR 1) using a sgRNA specific primer in combination with a generic primer targeting the plasmid backbone (Figure 3.5 A). Both sgRNAs for each gene were detected in their respective KO cell line without amplification in EV control cells, demonstrating successful transduction and integration of both lentiviral constructs for each gene. Consistent with sgRNA integration, stable Cas9 expression was verified in all KO and control cell lines while being absent in untransduced WT cells (Figure 3.5 B).

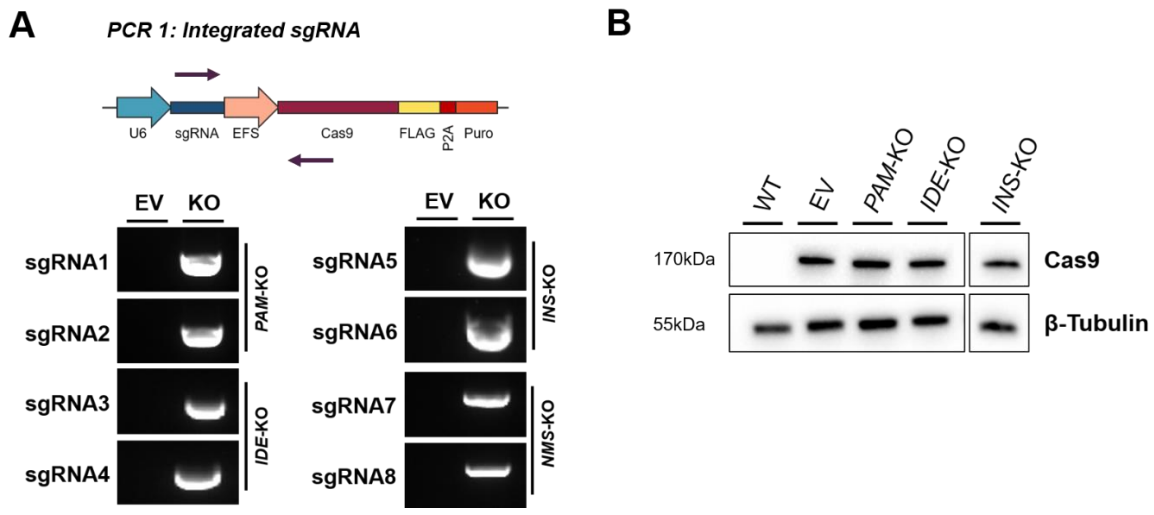


Figure 3.5 – Stable integration of plentiCRISPRv2. (A) Integrated sgRNA were PCR amplified using a sgRNA-specific primer in combination with a primer targeting the common Cas9 region in the plentiCRISPRv2 plasmid (*purple, top panel*), only resulting in a PCR product if the sgRNA is stably integrated *(bottom panel)*. **(B)** Western Blot analysis of Cas9 protein in KO and control cells. WT, wildtype; EV, empty-vector; KO, knockout.

sgRNA integration and Cas9 expression alone do not indicate successful editing, sgRNA target sides were therefore individually PCR amplified (PCR 2), Sanger sequenced and the Indel frequency was assessed by TIDE (Figure 3.6 A). Editing efficiency exceeded 87.5% at all sgRNA target sites in *INS*-KO, *PAM*-KO and *IDE*-KO, resulting in only around 1% of cells being unedited. sgRNA7 and sgRNA8 in the non-expressed control cell line *NMS*-KO had a lower efficiency of 42% and 39.5%, respectively. This lower editing efficiency in a non-expressed gene is consistent with previous studies showing a correlation between the transcriptional or chromatin state and its editing efficiency (Horlbeck *et al.*, 2016; Verkuijl & Rots, 2019). Both sgRNA target sites in *INS*-KO cells (sgRNA5 and 6) are close enough to be amplified within a single PCR reaction and the frequency of large deletions and the simultaneous cutting efficiency of both sgRNAs can therefore be determined using the Inference of CRISPR Edits (ICE) tool (Hsiau *et al.*, 2019). The approximately 50 bp range between the two sgRNAs has been cut out in 33.4% of *INS*-KO cells. To assess the presence

of concurrent sgRNA cutting and large deletions in *IDE*-KO, *PAM*-KO and *NMS*-KO, an adapted PCR approach (PCR 3) was performed using only one primer on either side of the sgRNA target sites in each gene (Figure 3.6 B). This PCR reaction can only occur if the region between the sgRNAs had been deleted as it would otherwise be too large for amplification, which is the case in EV and WT control cells. The presence of a PCR product in KO cells is therefore indicative of the presence of large deletions. This product was detected in all KO cells lines and therefore represents a large deletion of 94 kbp in *PAM*-KO, 66 kbp in *IDE*-KO and 8 kbp in *NMS*-KO. As the sgRNAs are within close proximity in *INS*-KO, two amplification bands are present, indicating a shorter PCR product with the large deletion of around 50 bp and a PCR product with either small Indels or no edits. Image based quantification of this shorter PCR product accounted for 36.7% of the total PCR reaction, consistent with the results from the previous ICE analysis.

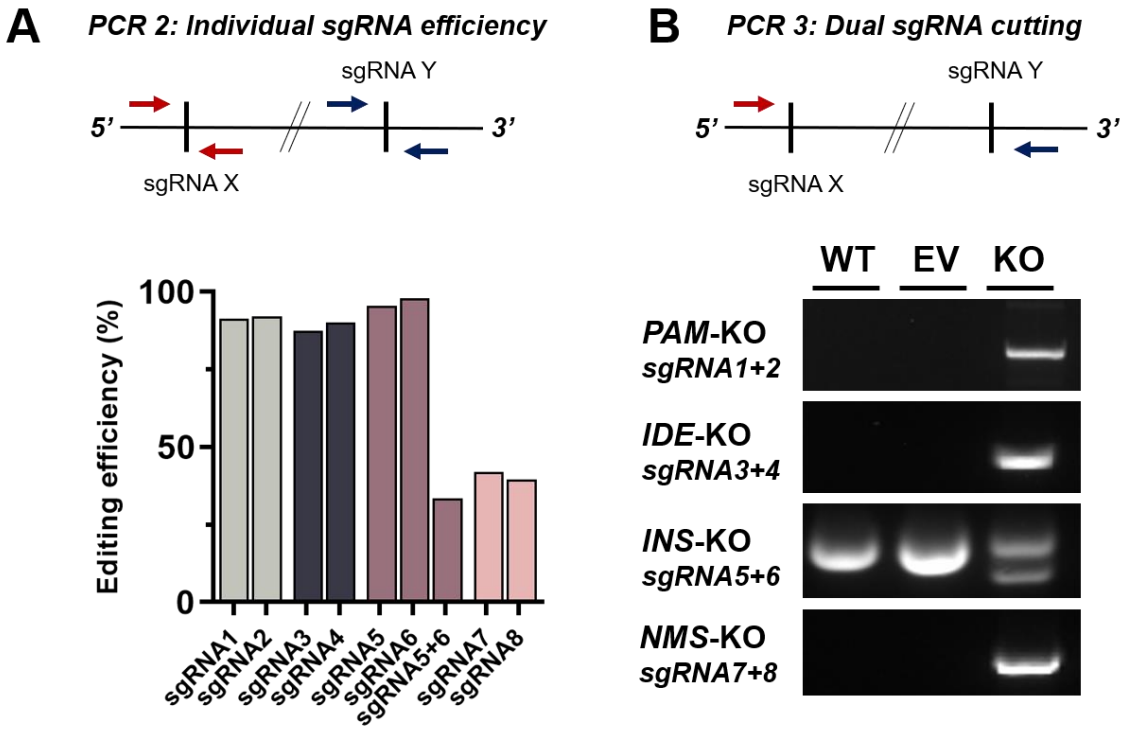


Figure 3.6 – sgRNA editing efficiency. (A) PCR approach (PCR2) amplifying each individual sgRNA targeting side (*top panel*) to determine the editing efficiency of Sanger sequenced PCR products using TIDE and ICE (*bottom panel*). **(B)** PCR approach (PCR3) to assess dual sgRNA cutting using one primer per sgRNA target side (*top panel*), only resulting in a PCR product if a large deletion between both sgRNA is present (*bottom panel*). WT, wildtype; EV, empty-vector; KO, knockout.

The constitutive long-term expression of Cas9 and sgRNAs in this lentiviral based pipeline could lead to an increased chance of editing at off-target sites. The top off-target sites in *PAM*-KO cells were identified based on the cutting frequency determination (CFD) score of sgRNA1 and 2 (Doench *et al.*, 2016). The locations with the highest score were located in intros of *MED15* (0.43) and *GPM6B* (0.54) and TIDE analysis of these regions demonstrated no significant off-target editing (Figure 3.7). Albeit this CRISPR pipeline generates heterogenous KO cell lines, the editing is highly efficient without any detected off-target activity.

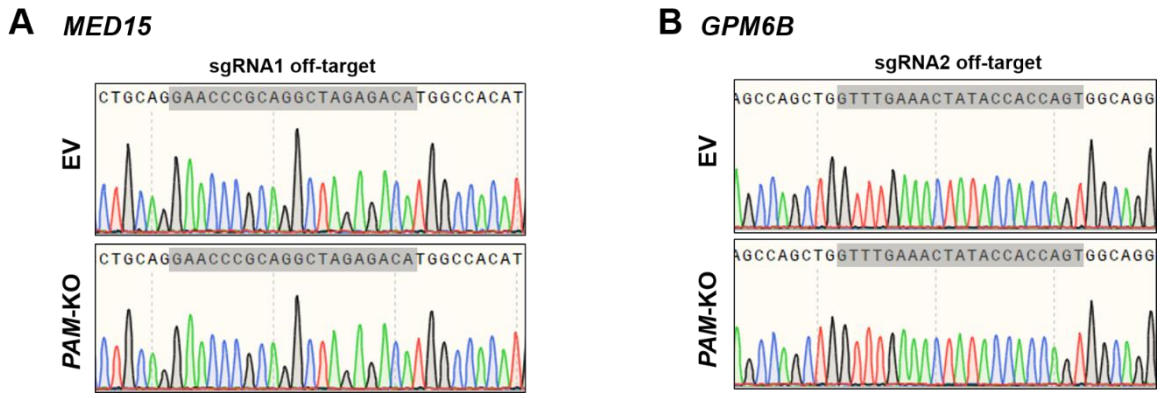


Figure 3.7 – Off-target sequencing in *PAM*-KO. Sanger sequencing traces of predicted top off-target sites in *PAM*-KO cells for sgRNA1 in *MED15* **(A)** and sgRNA2 in *GPM6B* **(B)**. EV, empty-vector; KO, knockout.

3.3.3 Functional characterisation of EndoC-βH1 KO cells

This genome editing strategy was highly efficient in editing EndoC-βH1. However, if this also translates into functional protein KO and if the cells are functionally affected from the transduction and selection process or stable Cas9 expression remains to be determined.

The growth rate was stable across KO and control cell lines throughout the period of culturing and followed common growth trends (Figure 3.8 A). The average growth rate of 2.21 across seven weeks or passages was not significantly different between cell lines ($p=0.96$, Figure 3.8 B). Despite previous reports of Cas9 and DSB induced apoptosis, there was no difference in the growth rate between the control cell lines EV (2.15) and *NMS*-KO (2.30, $p=0.95$), which can be distinguished by inducing no DSB compared to introducing specific DSB, respectively (Ihry *et al.*, 2018; Haapaniemi *et al.*, 2018; van den Berg *et al.*, 2018). The more common EV cell line was therefore used as a control cell line in future characterisations.

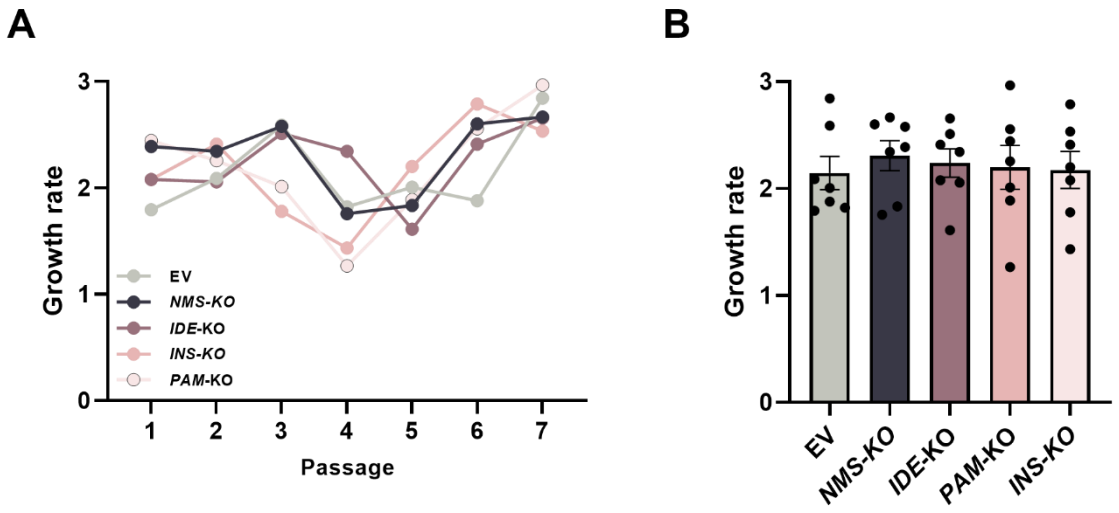


Figure 3.8 – Growth rate in KO and control cells. The growth rate is calculated from the weekly cell counts during splitting of KO and control cell lines and displayed across seven passages for each cell line **(A)** or as average growth rate **(B)**. All data are mean ± SEM from seven passages. Data were analysed using one-way ANOVA with Tukey's multiple comparison test. WT, wildtype; EV, empty-vector; KO, knockout.

To assess if the transduction, antibiotic selection and constitutive expression of Cas9 affects the beta cell phenotype of the cells, I examined insulin secretion and content in EV controls compared to untreated WT cells. The secretory insulin response in WT and EV cells was similar across a range of physiological glucose concentrations (Figure 3.9 A). Insulin secretion could be stimulated in both cell lines using the K_{ATP} channel blocker tolbutamide and reduced using the activator diazoxide. Tolbutamide enhanced insulin secretion in WT and EV cells 3.5- and 2.8-fold (p=0.94) while diazoxide diminished the response by 66.2% and 59.8% (p=0.81), respectively. In response to increased glucose concentrations, EV cells increased their insulin secretion by 2.48 versus 2.64 (p=0.94) in WT cells (Figure 3.9 B). In addition, intracellular insulin was not significantly different between EV and WT cells with around 33.44 ng and 34.83 ng/2×10^4 cells (p=0.83) (Figure 3.9 C).

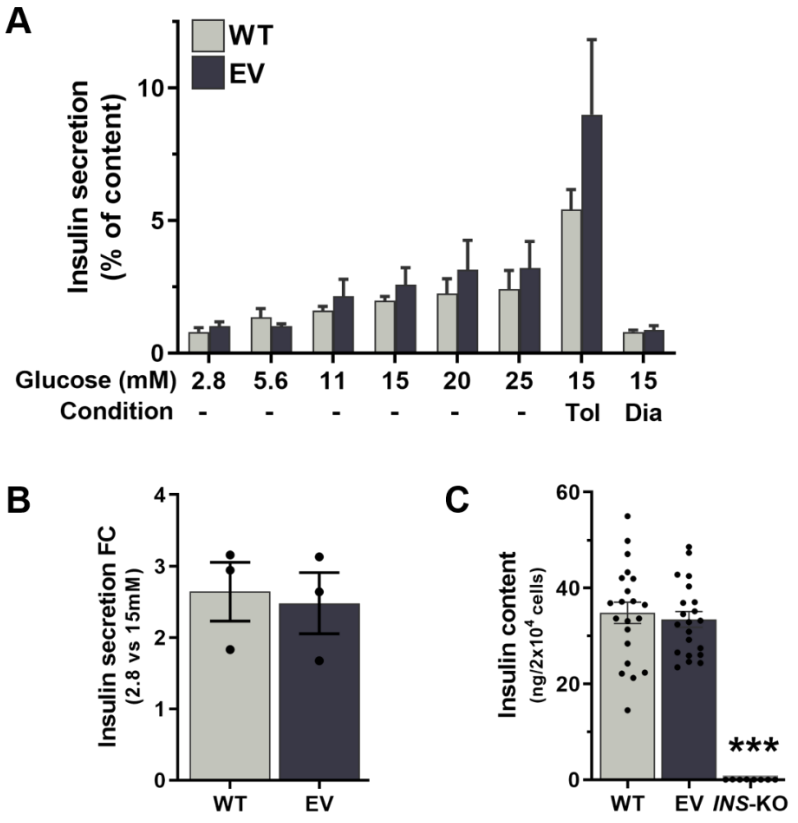


Figure 3.9 – Insulin secretion in WT and EV cells. (A) Insulin secretion as percentage of insulin content in WT and EV cells across different glucose conditions. **(B)** Insulin secretion fold change from 2.8 to 15 mM glucose. **(C)** Insulin content in WT and EV cells. All data are mean ± SEM from three independent experiments and for insulin content as 21 replicates from seven conditions in three independent experiments. Data were analysed using two-way ANOVA Sidak's multiple comparison test **(A)**, two-sample t-test with Holm-Sidak correction **(B)** and one-way ANOVA with Sidak's multiple comparison test **(C)**. WT, wildtype; EV, empty-vector; FC, fold change.

If the EndoC-βH1 cell lines were indeed KO cell lines was assessed by analysing mRNA and protein levels of the target genes. RNA expression of the respective target gene was significantly reduced in *PAM*-KO and *IDE*-KO by 77.8% (p=0.035) and 66% (p=0.034) compared to EV cells, respectively (Figure 3.10 A). *INS*-KO also demonstrated a strong but not significant decrease in *INS* expression by 54.2% (p=0.056). The reduction of the transcript levels demonstrates the introduction of frameshift mutations and premature stop codons (PSC) in the target genes leading to subsequent degradation by the nonsense-mediated mRNA decay (NMD) pathway. The presence of residual transcript is in

accordance with previous studies showing that incomplete NMD can occur in KO cells after introducing frameshift mutations in the coding region of the gene (Reber *et al.*, 2017). As the KO cell lines are a heterogeneous population, it cannot be excluded that a proportion of the detected transcript also originates from unedited cells or transcripts which do not contain a PSC and are not targeted for degradation. In addition, only one isoform was targeted in *IDE*-KO cells while the detected transcript is representative for both isoforms. The isoform specific targeting was confirmed when IDE protein levels were analysed in *IDE*-KO cells with the targeted isoform 1 being undetectable while no difference was observed in protein levels of isoform 2 (Figure 3.10 B). *PAM*-KO and *INS*-KO cells demonstrated complete loss of protein in their respective target genes, all PAM isoforms in *PAM*-KO and insulin precursor and mature insulin were absent in *INS*-KO. The depletion of INS was also confirmed using the sensitive AlphaLISA based detection method ($p<0.0001$ vs EV, Figure 3.9 C). As the KO cell lines demonstrated a complete protein depletion, it can be concluded that these cell lines are indeed functional LoF cell lines. The cell lines were routinely cultured for more than six months without losing any of their KO characteristics. This CRISPR editing pipeline in EndoC-βH1 is therefore able to efficiently create complete KO cell lines without affecting their insulin secretion characteristics in the process.

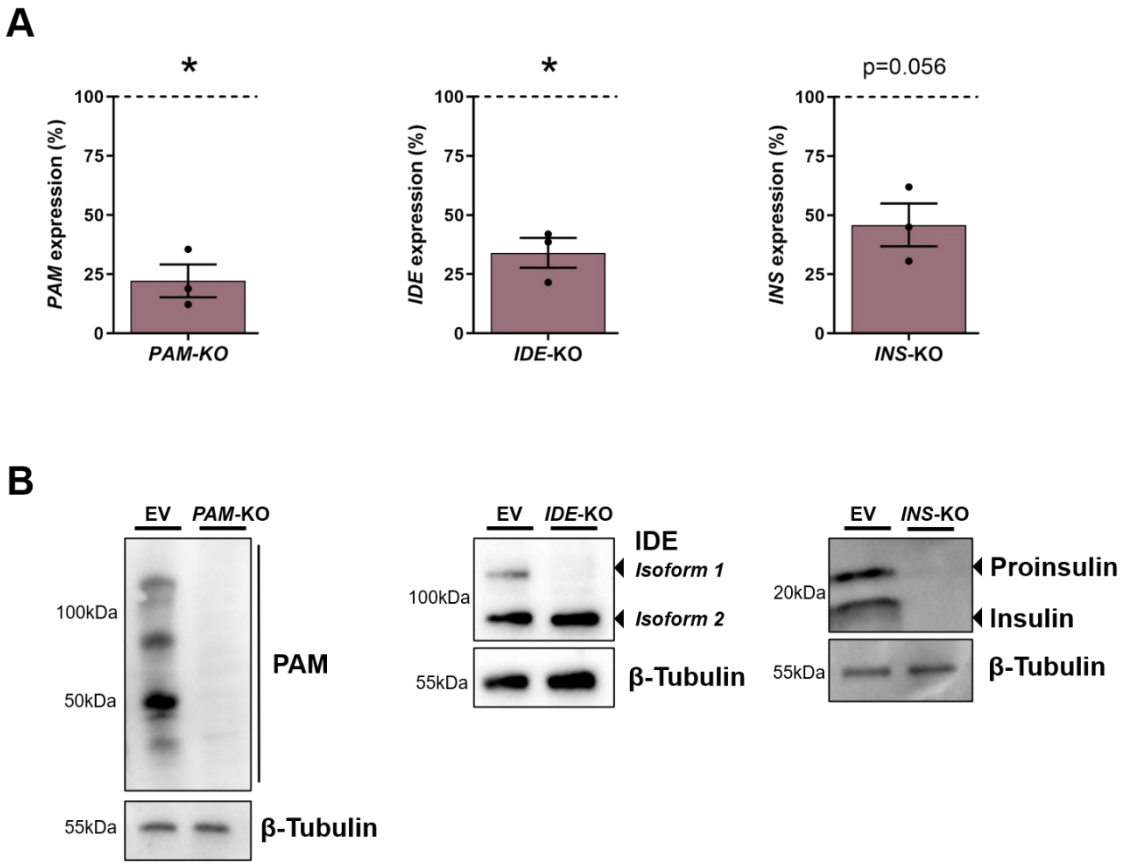


Figure 3.10 – Target gene expression in KO cell lines. (A) mRNA expression of KO target genes, which is normalised to EV control cells within each experiment. The expression level is displayed as percentage of EV which is displayed as dotted line at 100%. **(B)** Protein expression of KO target proteins with β-Tubulin as loading control. All data are mean ± SEM from three independent experiments, analysed using a one-sample t-test. P-values * < 0.05. EV, empty-vector; KO, knockout.

3.3.4 EndoC-βH1 KO cells are distinct loss-of-function models compared to siRNA knockdown

To examine the specific phenotypic effects of EndoC-βH1 KO cells, I compared the LoF model to siRNA based silencing approaches. In contrast to stable KO cells, siRNA mediated silencing is a transient model based on mRNA degradation (Elbashir *et al.*, 2001). Potential differences between the two strategies were assessed in the genes *NEUROD1* and *SLC30A8,* which are known to be important in beta cell function (Malecki *et al.*, 1999; Chimienti *et al.*, 2004; Gu *et al.*, 2010; Wijesekara *et al.*, 2010).

3.3.4.1 NEUROD1-KO cells have increased ER stress and apoptosis

NEUROD1 is a transcription factor involved in pancreas development and mature beta cell function and gene defects are implicated in monogenic diabetes and T2D risk (Malecki *et al.*, 1999; Gu *et al.*, 2010). *NEUROD1*-KO cells were created based on the previously described EndoC-βH1 genome editing pipeline. Even though stable sgRNA integration was demonstrated, the cells only displayed a low editing efficiency (Figure 3.11 A, B). Immediately after antibiotic selection, *NEUROD1*-KO cells showed a reduction in NEUROD1 protein levels, but not a complete depletion as observed in the previous KO cell lines (Figure 3.11 C). After another passage, the level of NEUROD1 protein returned to baseline, similar to the level in EV control cells. It was therefore not possible to create a stable *NEUROD1*-KO cell line as cells harbouring the *NEUROD1* KO alleles were eliminated from the population while cells without edits or LoF alleles were able to survive and expand. To confirm this apoptotic effect upon NEUROD1 deletion, I assessed markers of ER stress and apoptosis in protein samples collected during the short period of NEUROD1 reduction.

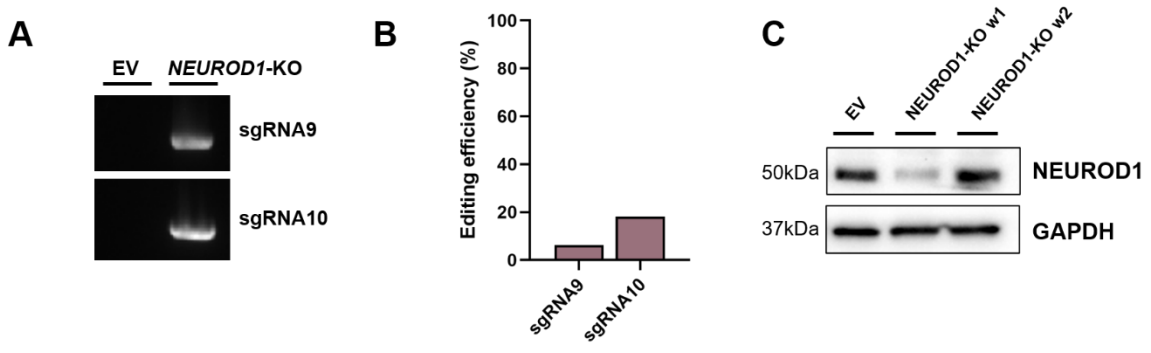


Figure 3.11 – *NEUROD1*-KO cell characteristics. (A) Stable sgRNA integration was assessed using the PCR1 based amplification approach and specific primer for sgRNA 9 and 10. **(B)** sgRNA editing efficiency of Sanger sequenced PCR2 products using TIDE. **(C)** Protein expression of NEUROD1 protein in *NEUROD1*-KO and EV control cells. W1 and w2 indicate the same *NEUROD1*-KO cell line one week apart. GAPDH was indicated as loading control. EV, empty-vector; KO, knockout.

Two of the three key signal activators of the unfolded protein response (UPR) to detect ER stress and activate downstream signalling aimed at restoring ER protein homeostasis, pPERK and pIRE1, were upregulated 10.6- and 8.7-fold in *NEUROD1*-KO cells (Figure 3.12 A, B). Western Blots to detect the third transmembrane protein ATF6 were not successful in detecting the protein due to unspecific antibodies. The pro-apoptotic downstream transcription factor CHOP was increased 2.2-fold, the anti-apoptotic signalling mediator TRAF2 was downregulated by 70% and the apoptosis executioner cleaved caspase 3 was upregulated 1.6-fold, indicating unresolved and severe ER stress leading to apoptosis in *NEUROD1*-KO cells (Figure 3.12 C-E). Levels of ER stress and apoptotic markers were not different between WT and EV cells, as shown by the representative downstream marker CHOP (Figure 3.12 F).

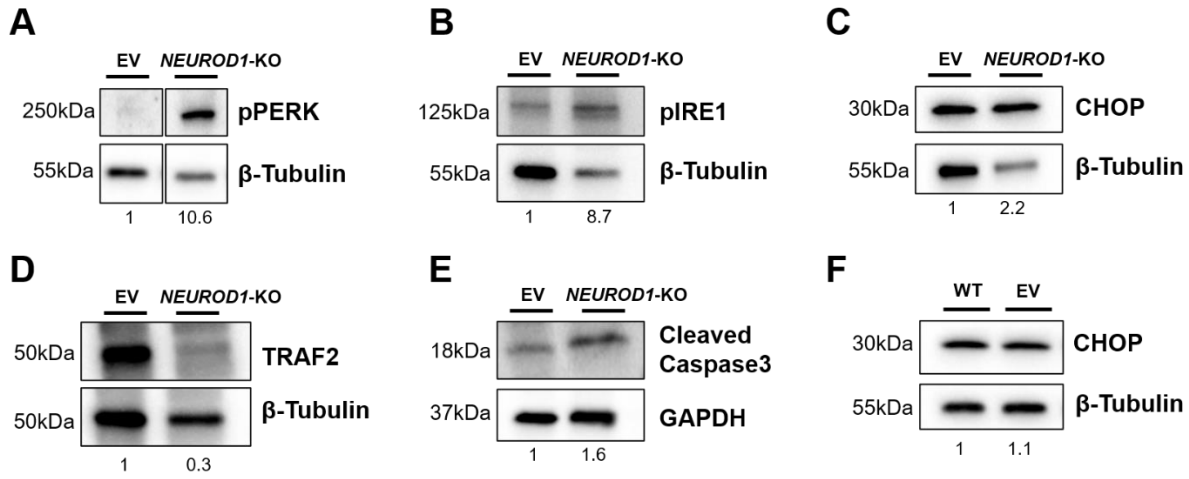


Figure 3.12 – *NEUROD1*-KO cells have increased levels of ER stress. (A-E) Protein expression of ER stress markers pPERK **(A)**, pIRE1 **(B)**, CHOP **(C)**, TRAF2 **(D)** and cleaved caspase 3 **(E)** in *NEUROD1*-KO and EV control cells. **(F)** CHOP expression in WT and EV cells. GAPDH and β-Tubulin were indicated as loading control and normalised protein quantifications are indicated below each lane as fold changes relative to EV protein levels. Data is from one *NEUROD1*-KO cell line. EV, empty-vector; KO, knockout.

To compare these severe implications of NEUROD1 depletion in the *NEUROD1*-KO model to a transient knockdown model, I further characterised markers of ER stress and apoptosis in cells treated with *NEUROD1* siRNA. NEUROD1 was efficiently silenced with a mean protein reduction of 93.7% (p=0.007) (Figure 3.13 A, B). Consistent with the KO model, the pro-apoptotic transcription factor CHOP was significantly increased in *NEUROD1* silenced cells by 34% (p=0.016) compared to cells treated with non-targeting control siRNA (NT) (Figure 3.13 C, D). Other markers of ER stress and apoptosis were unchanged. Total protein levels of the UPR signal activators PERK (90.8%, 0.444) and IRE1 (95.9%, 0.723) were not significantly upregulated and there was no increased phosphorylation resulting in the activated forms pPERK (112.3%, p=0.422) or pIRE1 (227.7%, 0.544, Figure 3.13 E-L). The death protease cleaved caspase 3 (98.0%, p=0.801) and the signalling mediator TRAF2 (96.1%, p=0.541) were not significantly different in *NEUROD1* silenced cells compared to controls (Figure 3.13 M-P).

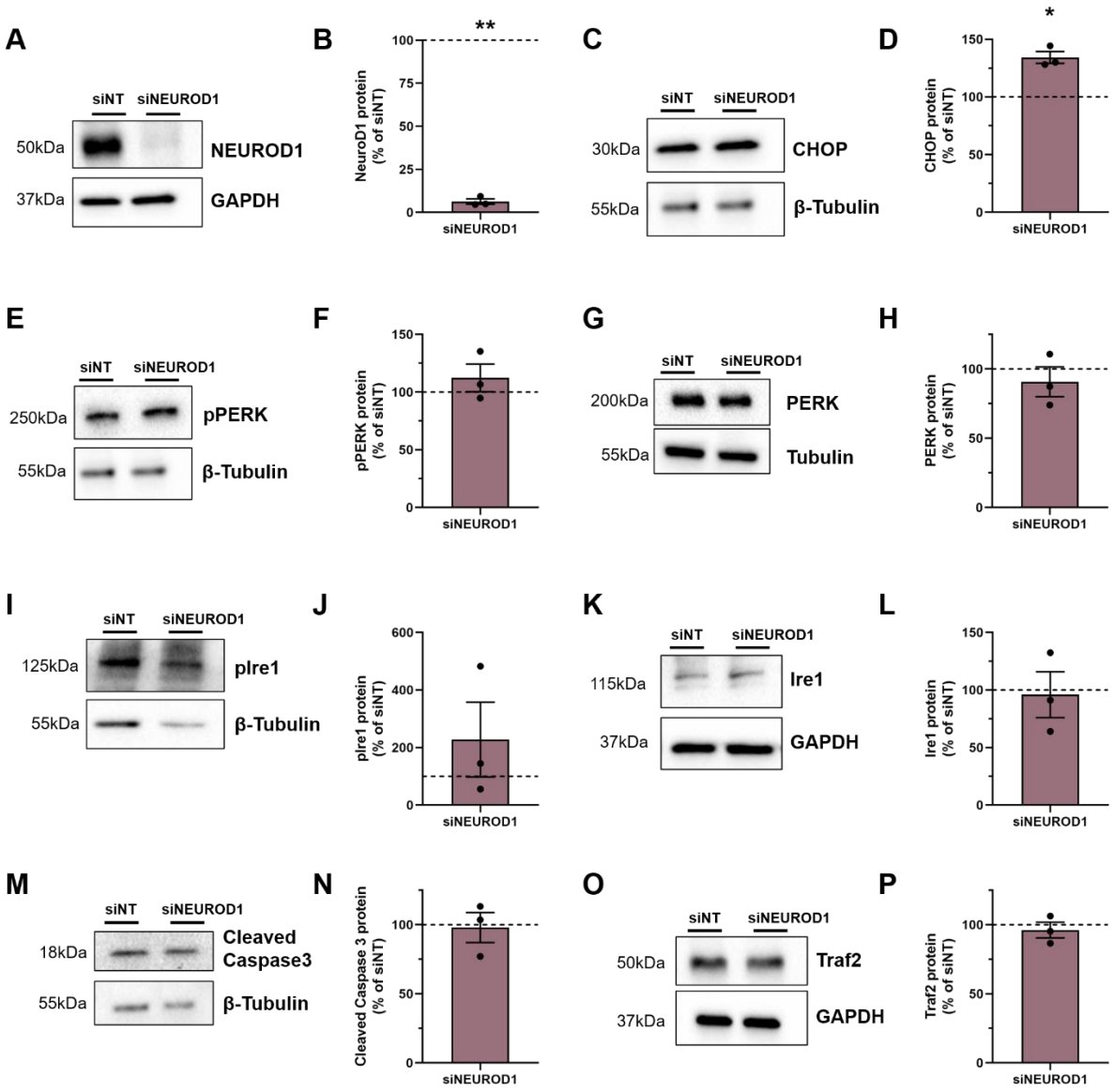


Figure 3.13 – Markers of ER stress and apoptosis in si*NEUROD1* cells. (A-P) Western Blots of NEUROD1 **(A)**, CHOP **(C)**, pPERK **(E)**, PERK **(G)**, pIRE1 **(I)**, IRE1 **(K)**, cleaved caspase 3 **(M)** and TRAF2 **(O)** in *NEUROD1* silenced (si*NEUROD1*) and non-targeting control cells (siNT). **(B, D, F, H, J, L, N, P)** Western Blots were quantified, normalised to their respective loading control GAPDH and β-Tubulin and siNT control. The protein level is displayed as percentage of siNT which is highlighted as a dotted line at 100%. All data are mean ± SEM from three independent experiments, analysed using a one-sample t-test. P-values * < 0.05, ** < 0.01. NT, non-targeting.

Total cell count was not affected in NEUROD1 silenced cells (97.78%, p=0.894), indicating no effect when capturing both, cell death and proliferation (Figure 3.14 A). Based on the proliferation marker Ki-67 (150.2%, p=0.470), it can be concluded that there is no difference in proliferation between si*NEUROD1* and siNT cells and consequently also no apoptotic reduction in cell numbers (Figure 3.14 B,C). To support the proteomic investigations and assess if ER stress occurs in *NEUROD1* silenced cells without resulting in apoptosis, mRNA expression levels were analysed for selected ER stress markers (Figure 3.14 D). Confirming the western blot results, *NEUROD1* expression was drastically reduced in si*NEUROD1* cells by 79.72% (p<0.0001). Unlike the protein level, *DDIT3* expression (encoding for CHOP) was not significantly increased (113.93%, p=0.506). Unexpectedly, other markers of UPR activation were significantly decreased, indicating an adaptive UPR response to chronic ER stress (Gomez & Thomas Rutkowski, 2016). The spliced and active form of the transcription factor *XBP1* (*XBP1*s) induces downstream UPR targets and was reduced by 30.82% (p=0.023). The ER chaperone BiP, encoded by *HSPA5* was also significantly downregulated by 20.27% (p=0.005). *ATF6* (90.80%, p=0.143) and *ATF4* (109.48%, p=0.453) expression was not significantly changed. Even though NEUROD1 is a well-characterised transcription factor for *INS*, there was no difference in *INS* transcription (95.71%, p=0.596).

The *NEUROD1*-KO model in EndoC-βH1 demonstrated a severe phenotype with high ER stress and apoptosis. The siRNA model on the other hand was less pronounced but confirmed a mild effect on ER stress without influencing cell viability as shown by measuring specific protein and RNA markers. Complete and permanent loss of NEUROD1 in *NEUROD1*-KO cells, therefore leads to cell death whereas transient *NEUROD1* silencing does not induce apoptosis.

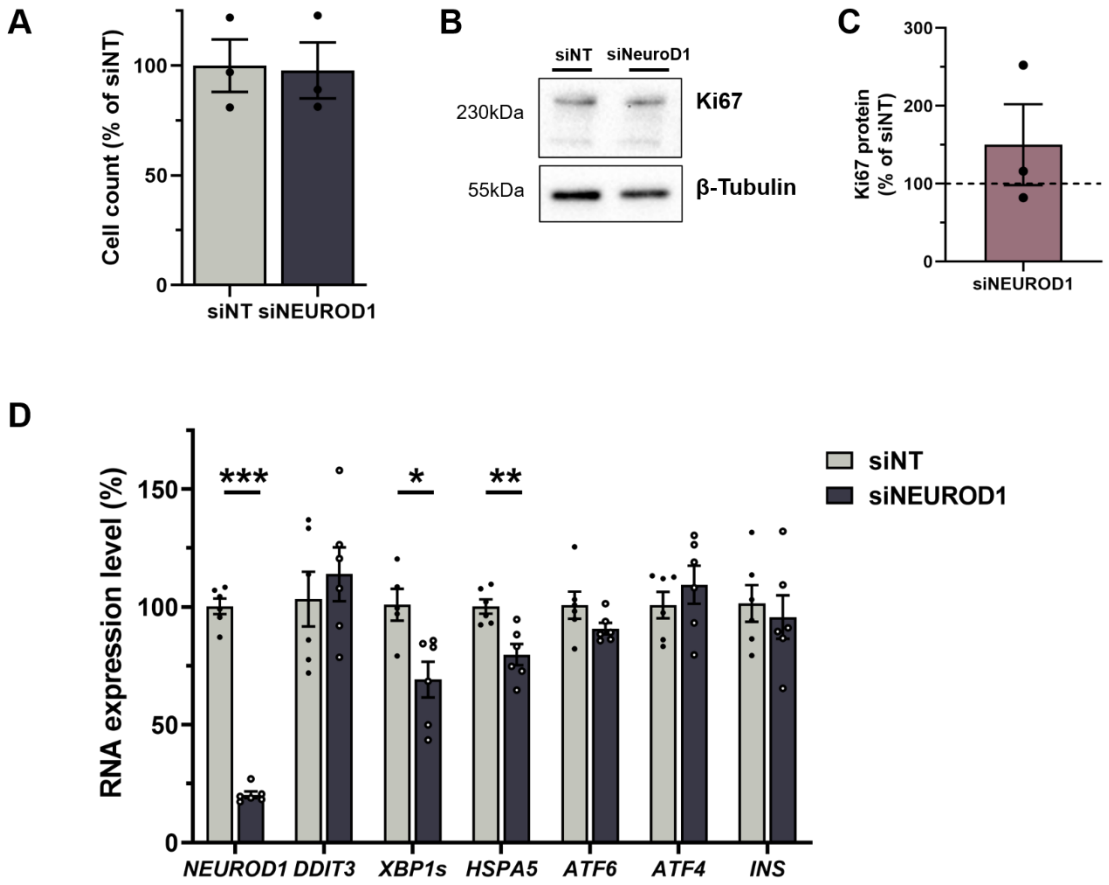


Figure 3.14 – si*NEUROD1* proliferation and RNA expression. (A) Cell count measurements for si*NEUROD1* cells, normalised to siNT controls. **(B)** Protein level of Ki67 and its loading control β-Tubulin. **(C)** Quantification of Ki67 western blot data, normalised to β-Tubulin and siNT control cells. The protein level is displayed as percentage of siNT which is highlighted as dotted line at 100%. **(D)** mRNA expression analysis of selected ER stress markers in si*NEUROD1* cells compared to siNT controls. All data are mean ± SEM from three **(A-C)** or six **(D)** independent experiments, analysed using a one-sample t-test **(C)** and two-sample t-test **(A, D)**. P-values * < 0.05, ** < 0.01, *** < 0.001. NT, non-targeting.

3.3.4.2 SLC30A8-KO cells have unaffected beta cell function

In addition to the first comparison of *NEUROD1* KO and siRNA LoF models, the *SLC30A8* gene which encodes the zinc transporter ZnT8 was chosen as a second example. Specific LoF variants in *SLC30A8* are protective in T2D and previous KD models in EndoC-βH1 have shown to improve glucose sensitivity and reduce expression of K_{ATP} channel subunits (Dwivedi *et al.*, 2019). Insulin secretion and *KCNJ11* and *ABCC8* expression were, therefore selected as suitable phenotypic readouts to compare KD and KO models of *SLC30A8.* Silencing of *SLC30A8* reduced the protein by 76.39% (p=0.001) and *SLC30A8*-KO cells were successfully created based on the previously described EndoC-βH1 CRISPR/Cas9 pipeline and demonstrated complete protein depletion (Figure 3.15 A, B). *SLC30A8* expression was downregulated by 84.54% (p=0.002) in si*SLC30A8* cells and by 57.01% (p=0.041) in *SLC30A8*-KO cells (Figure 3.15 C). Consistent with previous studies, *SLC30A8* silencing reduced *KCNJ11* and *ABCC8* expression by 28.23% (p=0.007) and 28.46% (p=0.011), respectively (Dwivedi *et al.*, 2019). In *SLC30A8*-KO on the other hand, *KCNJ11* (104.20%, p=0.948) and *ABCC8* (94.66%, p=0.529) expression was not significantly different compared to EV control cells.

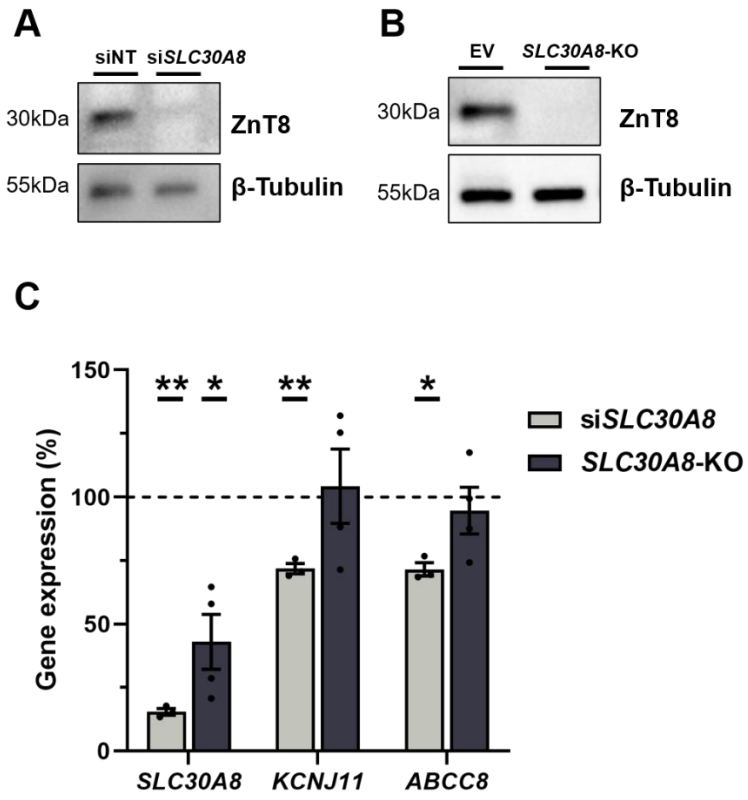


Figure 3.15 – *SLC30A8*-KO and siRNA cells show different effects on KATP channel expression. (A, B) Western Blot analysis of ZnT8 in si*SLC30A8* **(A)** and *SLC30A8*-KO **(B)** cells with their respective loading control β-Tubulin. **(C)** mRNA expression analysis of *SLC30A8, KCNJ11* and *ABCC8* in si*NEUROD1* and *SLC30A8*-KO cells. si*NEUROD1* expression is normalised to siNT and *SLC30A8*-KO expression is normalised to EV cells. The respective controls are indicated as dotted line at 100%. All data are mean ± SEM from three independent experiments, analysed using a one-sample t-test. P-values * < 0.05, ** < 0.01. NT, non-targeting.

In a recent study by Dwivedi et al., *SLC30A8* silencing in EndoC-βH1 increased basal insulin secretion at 1mM glucose without significantly affecting the glucose stimulation index (Dwivedi *et al.*, 2019). In line with the K_{ATP} expression data, *SLC30A8*-KO did not affect insulin secretion under basal or stimulated glucose conditions (p=0.99, Figure 3.16 A). The glucose stimulation index or insulin secretion fold change from low to high glucose was also not significantly different in *SLC30A8*-KO cells compared to EV control cells (2.23 versus 2.27, p=0.915, Figure 3.16 B).

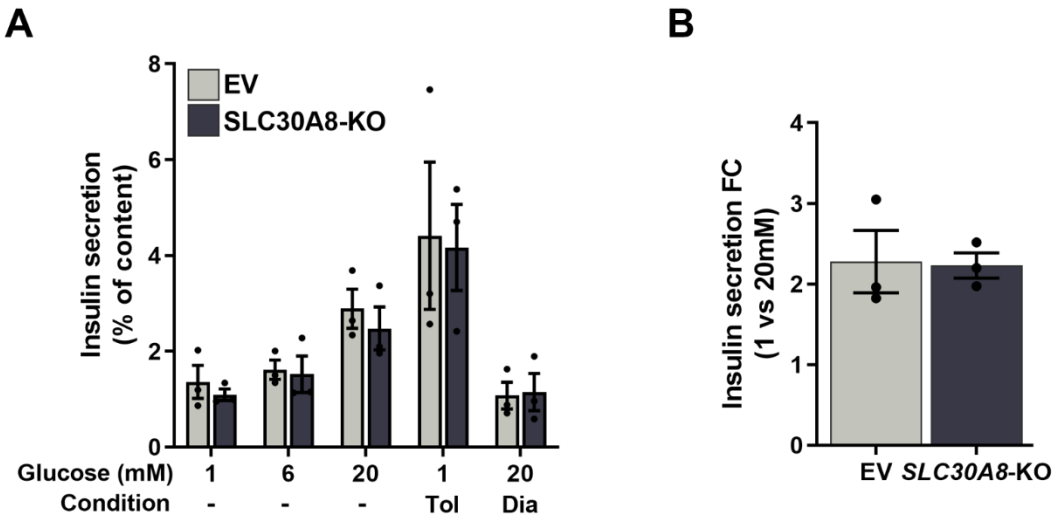


Figure 3.16 – Insulin secretion in *SLC30A8*-KO cells. (A) Insulin secretion as percentage of insulin content in *SLC30A8*-KO and EV cells across different glucose conditions. **(B)** Insulin secretion fold change from 1 to 20 mM glucose. All data are mean ± SEM from three independent experiments. Data were analysed using two-way ANOVA Sidak's multiple comparison test **(A)** and two-sample t-test **(B)**. WT, wildtype; EV, empty-vector; FC, fold change.

As demonstrated based on *NEUROD1* and *SLC30A8*, CRISPR/Cas9 KO and siRNA mediated KD are separate LoF models with distinct phenotypic effects. *NEUROD1*-KO cells with permanent and complete loss of protein showed that gene KO can lead to stronger phenotypes. *SLC30A8*-KO cells on the other hand also demonstrated that phenotypes can be lost or diminished in stable KO cell lines compared to transient models. This comparison highlights the strength of using multiple complementary approaches to understand the role of a specific gene or pathway. This CRISPR/Cas9 pipeline in EndoC-βH1 expands the LoF toolbox by allowing to investigate complete loss of protein in a stable cell line.

3.3.5 EndoC-βH1 KO cells can be generated using a low MOI

In addition to creating stable KO cell lines, CRISPR/Cas9 has been adapted for other techniques such as genome-wide CRISPR screens. Many of these approaches conduct pooled transductions of CRISPR libraries at a low MOI (usually 0.3) to infect each cell with only one viral particle and one sgRNA. To assess if low MOI transductions also efficiently induce gene KOs in EndoC-βH1, the CRISPR pipeline was performed again with a MOI of 0.3 instead of 8. A MOI of 0.3 only transduces around 30% of the cell population and due to the sensitivity of EndoC-βH1 to low seeding densities, it was crucial to assess the confluence of the cells regularly during puromycin selection and perform frequent media changes and additional splitting into smaller cell culture vessels if necessary. The resulting stable cell lines were subsequently characterised for their RNA and protein expression of the target genes *PAM, IDE* and *INS*. Similar to the KO cell lines at a high MOI, mRNA expression was drastically reduced in all low MOI KO cell lines (Figure 3.17 A). Consistent with the highest level of expression, *INS* expression showed the least reduction in high and low MOI *INS*-KO cell lines compared to their respective *PAM*-KO and *IDE*-KO. Analysis of protein abundance by western blot demonstrated that all targeted genes were either not detectable at all (*INS*-KO) or had minimal residual expression (*PAM*-KO and *IDE*-KO, Figure 3.17 B). This low level of remaining protein is in accordance with a transduction at a lower MOI and one sgRNA per cell, as some integrated sgRNAs might only induce synonymous substitutions or mutations that do not change the amino acid reading frame. The integration of multiple sgRNAs at a high MOI reduces the chances of these undesirable edits and the dependency on the efficiency of a single sgRNA. Nevertheless, the strong reduction of both the mRNA and protein level demonstrated that low MOI transductions are also efficient in CRISPR/Cas9 editing in EndoC-βH1 and therefore allow for successful implementation of future low MOI based approaches such as CRISPR screens.

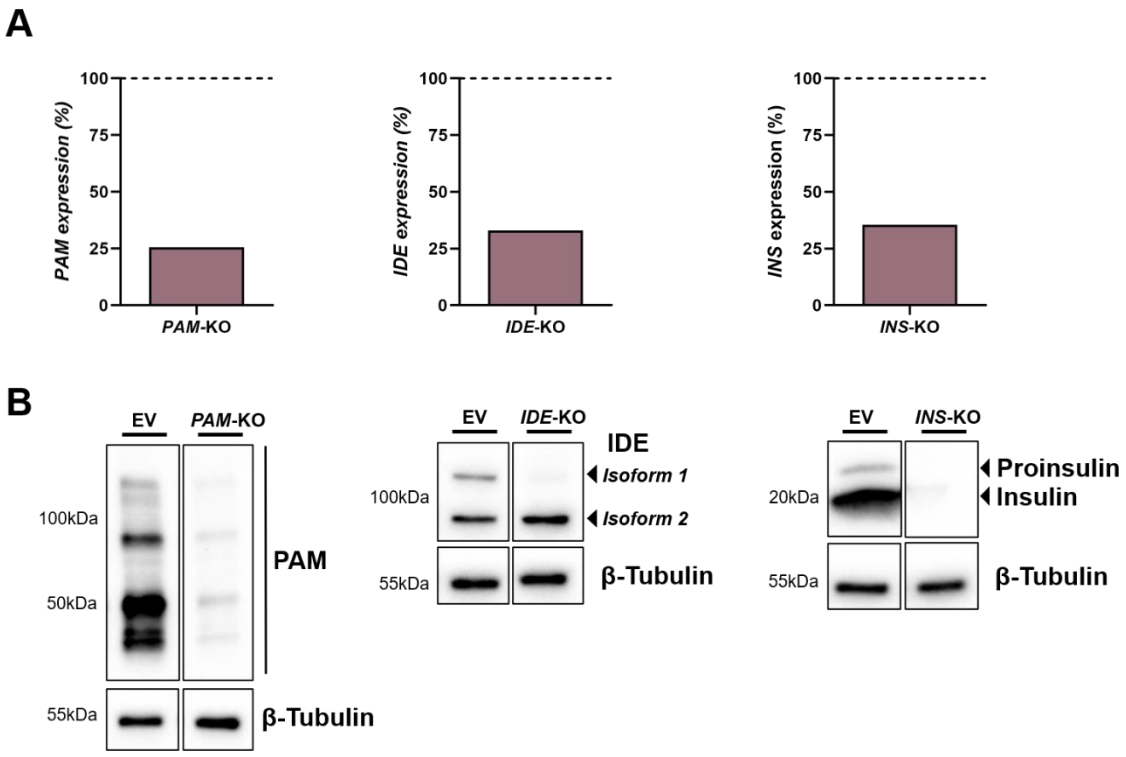


Figure 3.17 – EndoC-βH1 KO cells generated using a low MOI. (A) mRNA expression of KO target genes, which is normalised to EV control cells within each experiment. The expression level is displayed as percentage of EV which is displayed as dotted line at 100%. **(B)** Protein expression of KO target proteins with β-Tubulin as loading control. All data are from one experiment. EV, empty-vector; KO, knockout.

3.4 Discussion

EndoC-βH1 are currently the best available model to study human beta cell function *in vitro* but their challenging growth and culture characteristics have so far prevented the routine use of genome-editing and the establishment of robust CRISPR/Cas9 protocols. In this chapter, I developed a lentiviral based CRISPR/Cas9 editing pipeline in EndoC-βH1 to create stable KO cell lines. I characterised the functional and genomic characteristics in three proof of concept KO cell lines and further assessed the model in comparison to siRNA based silencing strategies in two further independent KO cell lines. The proof of concept cell lines *INS*-KO, *PAM*-KO and *IDE*-KO have distinct gene and isoform structures with different protein localisations and functions, but all cell lines demonstrated highly efficient editing and complete protein depletion highlighting the universal applicability of this pipeline. Adapting the pipeline with low MOI transductions replicated results from the high MOI pipeline and demonstrated the potential application of low MOI based CRISPR approaches such as genome-wide CRISPR screens in EndoC-βH1.

3.4.1 A dual sgRNA strategy efficiently induces gene KO

Instead of a single sgRNA based strategy, this pipeline applied a dual sgRNA based editing approach. This strategy offers the advantage of not having to rely on a single cleavage event and the introduction of a frameshift mutation by only one sgRNA. As only two-thirds of frameshift mutations introduce a PSC and inefficient sgRNA editing or alternative splicing can reduce the chance of achieving a gene KO, a dual sgRNA approach can increase the proportion of functional KO cells (Lappalainen *et al.*, 2013; Canver *et al.*, 2014; Zhou *et al.*, 2014). The probability of achieving a downstream PSC from editing at an individual sgRNA target side is increased by using two separate sgRNAs and in addition, the simultaneous introduction of dual DSB from both sgRNAs might induce large deletions resulting in a non-

functional protein. These large deletions might also excise intronic enhancers and non-coding regulatory elements and affect the expression of other genes resulting in unintended phenotypes. Some of the differences between KO and KD models might be explained by this accidental dysregulation of neighbouring genes in KO cells.

It must also be considered that using several sgRNAs might increase off-target effects. By sequencing selected top off-targets in *PAM*-KO cells, I demonstrated that we did not observe any unintended editing at these sites. It cannot be excluded, however, that off-target effects have occurred at other sites in *PAM*-KO cells that were not assessed, sgRNAs in other KO cells have introduced off-target DSB or large deletions or rearrangements were introduced which could not be captured in this targeted PCR amplification based sequencing approach (Kosicki, Tomberg, & Bradley, 2018). Due to the stable integration and expression of Cas9 and its respective sgRNAs, off-target effects might also increase over time.

The cells were transduced with sgRNAs in separate CRISPR plasmids. Adapting the pipeline with a plasmid containing both sgRNAs would further increase the homogeneity of the population and achieve a higher proportion of dual sgRNA editing. A recent study described a suitable cost-effective and versatile protocol to clone two sgRNAs into a Cas9 containing backbone and applied it to delete transcriptional enhancers in EndoC-βH3 (Beucher & Cebola, 2019; Miguel-Escalada *et al.*, 2019).

3.4.2 This CRISPR/Cas9 is more efficient in generating KO cells compared to similar pipelines

This CRISPR/Cas9 pipeline in EndoC-βH1 was the first robust protocol to describe the generation of KO cells in this cell line and also performed in-depth characterisation if the functionality of the cells was affected by stably expressing Cas9 and going through the transduction and selection process. The only other reported genome editing strategy in EndoC-βH1 by Cardenas-Diaz et al. targeted *HNF1A* and *PDX1* with a similar lentiviral based approach but using only a single sgRNA in a modified plentiCRISPRv2 vector (Cardenas-Diaz *et al.*, 2019, 2020). However, it was not characterised if the cells had phenotypically changed by going through all the steps of the pipeline and at least 10% of the resulting KO cells still contained the targeted protein. The authors used FACS sorting to purify the population and only utilised cells without residual protein for final functional experiments. Based on only those cells, they were able to achieve functional insights into HNF1A's and PDX1's role in the beta cell as a repressor of alpha cell genes and illustrated how EndoC-βH1 KO cells can provide important insights. My previously described KO pipeline demonstrated complete protein depletion and can therefore potentially identify effects that would be masked otherwise due to a low level of residual protein without having to sort for negative cells first. The complete protein depletion in my model compared to the insufficient protein deficiency in the study by Cardenas-Diaz et al. likely results from using the dual sgRNA approach and a higher editing efficiency instead of only relying on a single sgRNA, therefore demonstrating the advantage of this approach.

3.4.3 CRISPR KO can lead to potentiated or diminished phenotypes compared to siRNA models

CRISPR/Cas9 KO and siRNA based KD both offer the possibility to achieve a complete loss or reduction of a transcript and protein of interest to study the phenotypic consequences, they are, however, distinct methods with unique functional outcomes. To compare the CRISPR KO pipeline to a transient model and assess potential different phenotypes, I created *NEUROD1*-KO and *SLC30A8*-KO cells and examined specific beta cell relevant readouts in comparison to siRNA models. Interestingly, *NEUROD1*-KO and *SLC30A8*-KO demonstrated opposite directions of effect compared to transient siRNA KD. *NEUROD1*-KO cells had a more pronounced and severe phenotype on ER stress and apoptosis whereas no effect on K_{ATP} channel expression and insulin secretion was observed in *SLC30A8*-KO cells compared to their respective siRNA model. Previous studies assessing the relationship between KO and KD approaches observed similar discrepancies between the phenotypes arising from the two strategies (Karakas *et al.*, 2007; Rossi *et al.*, 2015; Morgens *et al.*, 2016). Phenotypes might also be diminished in the case of KO cell lines due to the genetic compensation response, an effect describing the nonsense-induced transcriptional compensation through degradation of PSC containing mRNA (El-Brolosy *et al.*, 2019; Ma *et al.*, 2019). This effect could be avoided if a dual sgRNA strategy would be designed to target the complete gene so no degradable mRNA product could be transcribed to initiate a compensatory response. In the case of siRNA-based experiments, it has to be considered if partial sequence complementarity to 3′ UTRs regions of other mRNA transcripts might have induced unwanted seed-based, microRNA-like off-target effects (Doench, Petersen, & Sharp, 2003). To confirm that the observed phenotype is a direct result from the intended siRNA based strategy, multiple individual siRNAs can validate pooled experiments or an alternative confirmation would be shRNA mediated downregulation via lentiviral delivery

or rescue experiments (Jackson & Linsley, 2010; Klinghoffer *et al.*, 2010). Of course it also has to be considered that the KO and KD experiments require substantially different experimental setups and reagents which might have contributed to the observed differences for example by inducing an increased susceptibility to cell death in the case of *NEUROD1*-KO cells.

3.4.4 NEUROD1 is required for beta cell survival

Interestingly, the characterisation of KO cell lines emerging from the CRISPR pipeline also demonstrated that *NEUROD1*-KO cells were not able to survive, consistent with a detected increase in markers of ER stress and apoptosis. Previously, it has been shown that NeuroD1 null mice have 14-fold higher apoptosis in pancreatic cells (Naya *et al.*, 1997). Conditional *NeuroD1*-KO in insulin expressing cells in mice on the other hand did not demonstrate an increased apoptosis as measured by activated caspase 3 (Gu *et al.*, 2010). *Ddit3* (CHOP) has been highlighted previously as a direct downstream target of NeuroD1 when overexpression of *NeuroD1* in rodent beta cells prevented ethanol induced expression of *Ddit3* and reduced apoptosis (Wu *et al.*, 2016). This is consistent with my study showing that both, KO and KD models of NEUROD1 reduction induce an increase in CHOP protein and for the first time to my knowledge, confirm the relationship between NEUROD1 and CHOP in a human beta cell line and without an extrinsic stress stimuli. The siRNA model based on transient silencing caused a phenotype in EndoC-βH1 similar to chronic ER stress but without inducing apoptosis, *NEUROD1*-KO cells in contrast were not able to compensate for the permanent and complete loss of NEUROD1 protein and demonstrated elevated ER stress and apoptosis as shown in a proposed regulatory mechanism in Figure 3.18. It would be an interesting future follow-up investigation beyond this thesis to study the role of *NEUROD*1 in the regulation of ER stress and CHOP and its potential implications for diabetes.

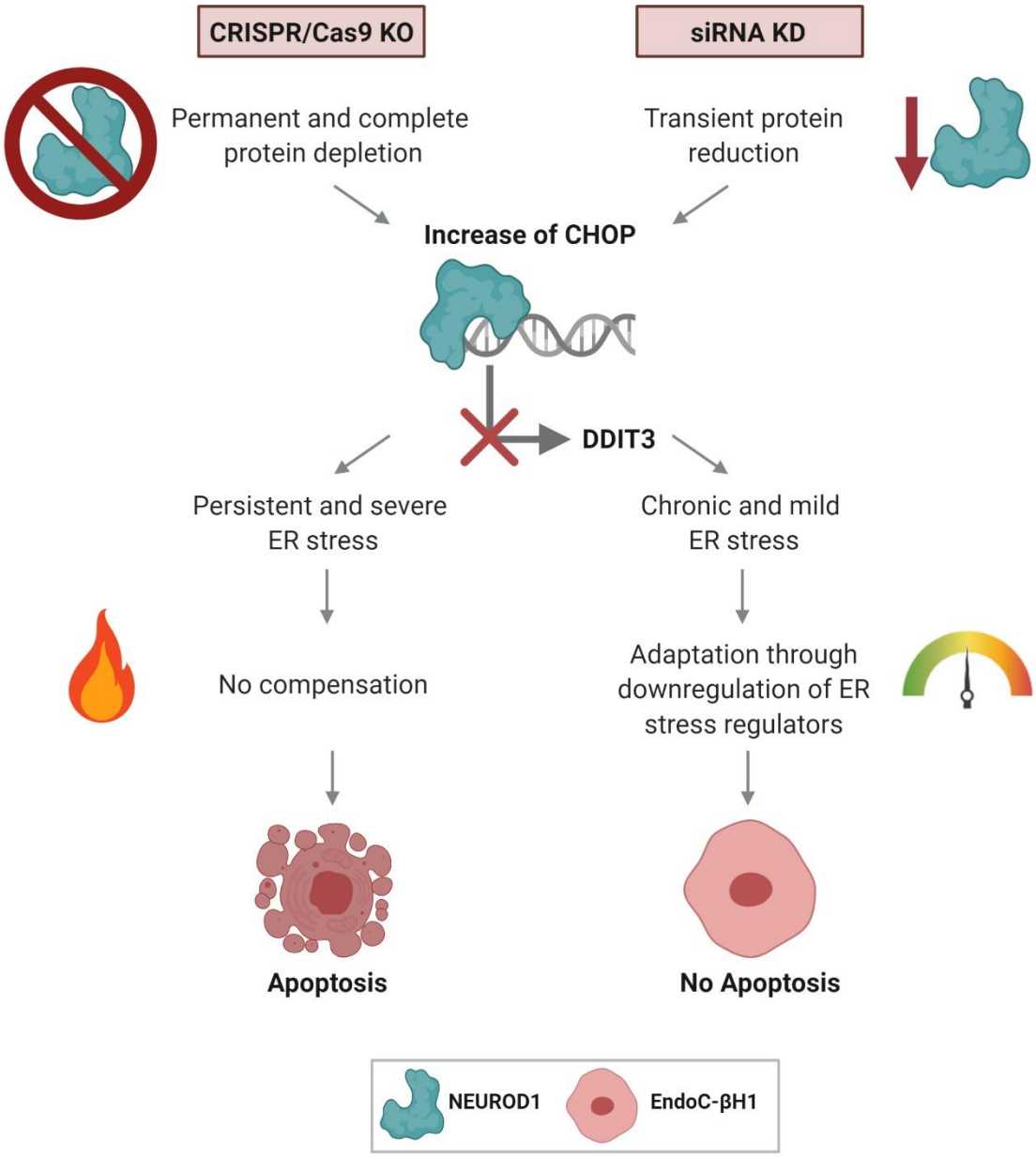


Figure 3.18 - NEUROD1 KO and KD in ER stress. NEUROD1 *(turquoise protein)* is depleted in CRISPR/Cas9 KO cells and reduced in an siRNA model in EndoC-βH1. The protein is proposed to directly regulate *DDIT3* expression, therefore inducing ER stress in both models. *NEUROD1* KO cells cannot adapt or compensate the permanent loss of NEUROD1 and apoptosis is induced, indicated by flames. *NEUROD1* siRNA cells on the other hand can adapt to the transient reduction of NEUROD1 through transcriptional modulation of other ER stress regulators such as *XBP1s* and *HSPA5* and apoptosis is prevented, shown by the equilibrium. The figure was created based on icons from Biorender.

3.4.5 Conclusion

This CRISPR/Cas9 pipeline in EndoC-βH1 is an efficient and robust approach to create LoF cell lines and study the implications of gene KOs in human beta cell function and diabetes pathogenesis. As it is the case with specific LoF models and as I demonstrated when comparing siRNA and CRISPR KO models, functional effects could be specific to this method and should be validated using a complementary approach. This successful creation of KO cell lines and the efficiency of low MOI transductions also support future CRISPR based studies in EndoC-βH1 such as genome-wide CRISPR screening. In total, I created seven KO cell lines using this pipeline, six of which have been described in this chapter. These cell lines also formed the basis of independent functional studies by fellow DPhil students to study the role of these genes in beta cell function (*RREB1*-KO, Dr Katia Mattis) or use the KO cell lines as a protein free cellular system to overexpress and study genetic variants implicated in T2D (*PAM*-KO, Dr Shahana Sengupta and *INS*-KO, Claire Duff).

4

A Genome-Wide Loss-of-Function CRISPR Screen to Identify Genes Involved in Beta Cell Function

4.1 Introduction

4.1.1 Genome-wide CRISPR screen

GWAS have so far identified more than 400 distinct association signals that have been robustly associated with T2D risk (Mahajan *et al.*, 2018b, 2018a; Spracklen *et al.*, 2020; Vujkovic *et al.*, 2020). At the majority of these disease loci, the causal gene and tissue have not been identified yet. Follow-up studies focussing on a single or a few genes at a locus have been successful in linking effector transcripts to their causal role in T2D but the required individual functional studies are highly time and labour-intensive and not a feasible strategy to prioritise potential causal genes on a large scale (Carrat *et al.*, 2017; Thomsen *et al.*, 2018; Dwivedi *et al.*, 2019). High-throughput studies on the other hand can overcome this difficulty by investigating up to thousands of genes and their functional role in disease simultaneously. Genome-wide pooled CRISPR screens are one of these ground-breaking approaches that can prioritise potential causal genes at T2D GWAS loci on a large scale by assigning a functional mechanism and a cell type of action. In pooled LoF CRISPR screens, all cells are located within the same cellular flasks and are being transduced with a genome-wide CRISPR library delivered via titered lentivirus, therefore resulting in only one viral particle or sgRNA integration per cell (Shalem *et al.*, 2015). The cells are then selected for a phenotype of interest, from simple survival readouts to highly complex FACS-based screens and enriched sgRNAs and gene hits are being determined using next generation sequencing. Genome-wide pooled CRISPR screens have not been applied yet to identify novel regulators of human beta cell function or to prioritise T2D risk genes. Previous screens addressing these questions have either applied siRNA based KD approaches, rodent beta cell models or only targeted certain genes, therefore relying on incomplete protein depletion and restricted translation into functional mechanisms in human beta cells (Thomsen *et al.*, 2016a; Pappalardo *et al.*, 2017; Fang *et al.*, 2019).

4.1.2 Experimental outline

In this chapter, I aim to perform the first genome-wide pooled LoF CRISPR screen in a human beta cell line to identify novel regulators of beta cell function and support the prioritisation of causal genes at T2D GWAS loci. The experimental outline of this chapter is shown in Figure 4.1. First, I developed a beta cell relevant CRISPR screening pipeline using a FACS based phenotypic readout for intracellular insulin content in the human beta cell line EndoC-βH1. To be able to have high confidence in the specificity and sensitivity of the antibody-based insulin staining, I extensively optimised and validated all steps and conditions of the FACS protocol. Prior to performing the genome-wide CRISPR screen, I also performed a small-scale CRISPR screen focusing only on a subset of genes to assess the transduction, selection and sorting pipeline. Over several months, the genome-wide CRISPR screen was then performed in EndoC-βH1 followed by extensive computational analysis of the next generation sequencing data to identify genes of interest. Using the extracted sgRNA read counts, I initially performed several quality control steps such as assessing sgRNA representation and replicate consistency to ensure the screen performance was up to high technical standards. To identify genes with a significant effect on insulin content, I initially performed a routine CRISPR screen analysis using the MAGeCK algorithm followed by developing a custom set of criteria to generate a high confidence gene hit list. Based on these hits, I assessed if the screen was able to identify genes relevant for beta cell function by investigating pathway enrichment, protein networks and known regulators of insulin content and beta cell function. I then further integrated the hits with predicted effector genes at T2D GWAS loci to support their potential causal role with functional evidence. I also identified which genes were commonly identified in this CRISPR screen and previous KD and rodent beta cell screens (Thomsen *et al.*, 2016a; Fang *et al.*, 2019).

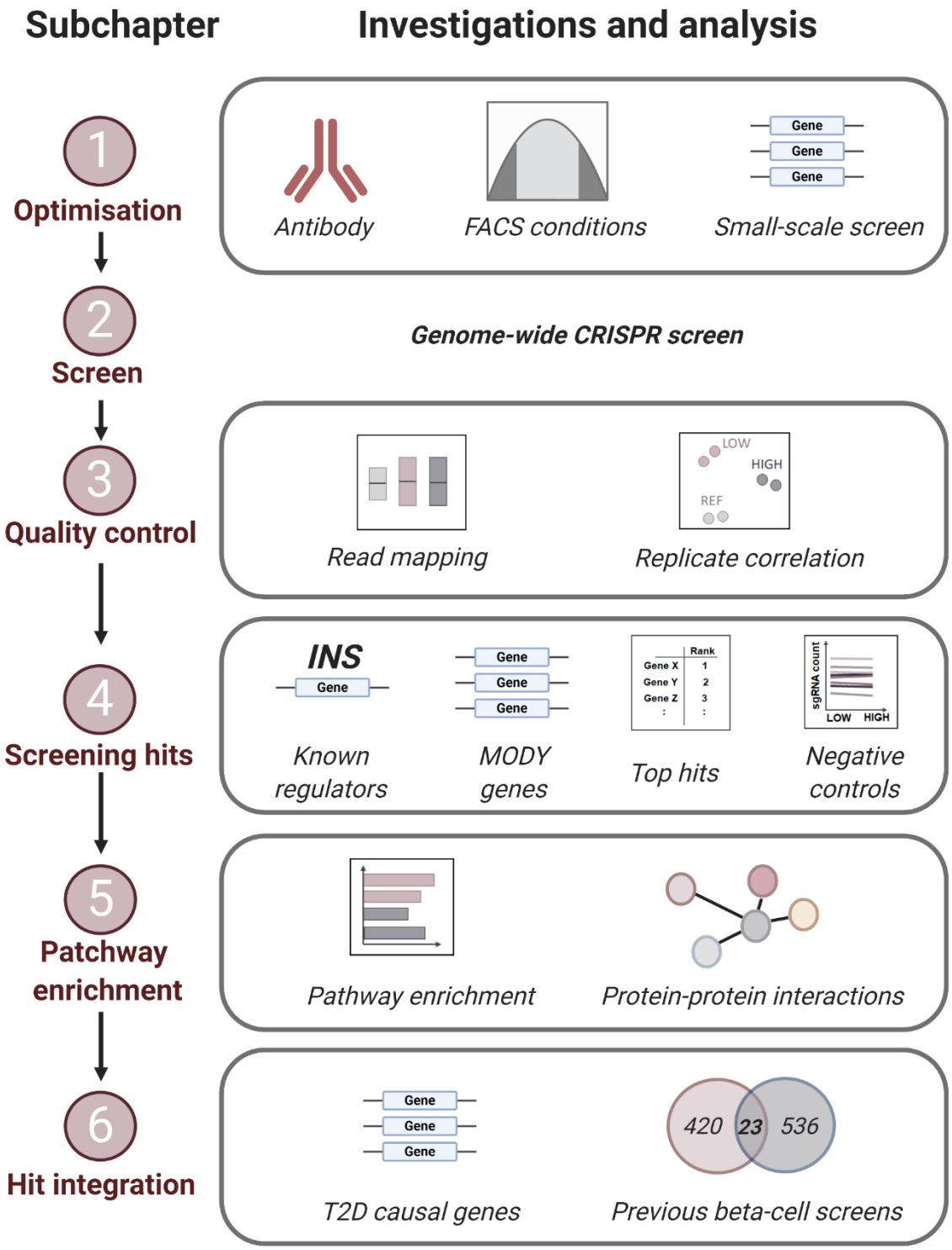


Figure 4.1 – Experimental outline. The subheadings of this chapter are shown on the left with their respective experimental investigations and analysis approaches on the right.

4.2 Material and Methods

4.2.1 CRISPR screen optimisation

4.2.1.1 RNA interference

Transient mRNA silencing was performed in EndoC-βH1 using siRNAs (*INS* #L-011058-00, NT control pool #D-001810-10, Dharmacon) as previously described in section 2.3. In brief, the cells were harvested after 72 h incubation with siRNAs and KD efficiency was determined on the mRNA level by RT-qPCR.

4.2.1.2 TaqMan gene expression analysis

To assess gene expression levels, RNA was extracted and converted into cDNA using the GoScript Reverse Transcriptase System (Promega). The details of the gene expression protocol are described in section 2.4. Gene expression differences between samples were analysed using TaqMan Gene Expression Assays (Applied Biosystems, Table 3.2), analysed based on the ΔΔCt method and normalised to the housekeeping genes *TBP*, *PPIA* and *GAPDH*.

TaqMan Probe	Assay details	Target region
INS	Hs00355773_m1	Exon 1-2
GAPDH	Hs02786624_g1	Exon 8
TBP	Hs00427620_m1	Exon 3-4
PPIA	Hs01634221_s1	Exon 1

Table 4.1 – TaqMan gene expression assays.

4.2.1.3 SYBR green sgRNA analysis

To quantify sgRNA integrations between samples in a small-scale proof of concept screen, I developed and optimised a custom SYBR green analysis pipeline. DNA was extracted from transduced and FACS sorted EndoC-βH1 as described in section 2.2.1. Samples were assayed in triplicates per gene and analysed on a 384-well plate. No-template control assays were included for each gene. Each sample contained 20 ng DNA template in a total volume of 8.75 µl, 10 µl SYBR Green PCR Master mix (Bio-Rad) and 0.625 µl each of forward and reverse primer (10 µM, Table 4.2). The samples were amplified on a 7900HT (Applied Biosystems) for 2 min at 98°C, 35 cycles of 10 sec at 98°C and 30 sec at 60°C, followed by the default melt curve analysis. Melt curves were verified visually to ensure specific amplification and Ct-values were analysed using the ΔΔCt method. sgRNA targets were normalised to the average of the three housekeeping genes *TBP*, *PPIA* and *GAPDH*.

Target	Sequence (5' → 3')
GAPDH_F	AGGCAACTAGGATGGTGTGG
GAPDH_Rev	TTGATTTTGGAGGGATCTCG
TBP_F	TGTTTGTTTCTTGCGAGTGC
TBP_Rev	CCCGGGAGGGTTCTTAGTAG
PPIA_F	AAGGGCTTTCGTGGAAACTT
PPIA_Rev	CCCCTCCATTCTCATCAAGA
LKO1.5R	GTTGATAACGGACTAGCCT
EV	TTTTTGTACGTCTCTGTTTTAGAGC
PAM_sgRNA2	GACGAAACACCGGTTCAGAAC
INS_sgRNA5	CAATGCCACGCTTCTGCAG
INS_sgRNA6	CATCTGCTCCCTCTACCAGC
NMS_sgRNA7	AAACACCGTGCAGGATCACA
NMS_sgRNA8	ACACCGACAATATCCAAGCCA
IDE_sgRNA3	AAACACCGTACCCACACAGG
IDE_sgRNA4	CGCATTAATGTGGACTTGACCG

Table 4.2 – SYBR green primer.

4.2.2 Genome-wide CRISPR screen

To increase the reproducibility and be able to identify true hits, two independent genome-wide CRISPR screens were performed consecutively with separate viral transductions, FACS sorting and sequencing. The following protocols indicate the procedures per replicate if not indicated otherwise.

4.2.3 Amplification and viral production of TKOv3 CRISPR library

The TKOv3 library was purchased on Addgene (#90294) and contains 70948 sgRNAs targeting 18053 genes and 142 control sgRNAs in a plentiCRISPRv2 backbone (Hart *et al.*, 2017). The same amplified plasmid sample was used for both replicates. To maintain library representation, four electroporation reactions were set up using 2 µl of 50 ng/µl TKOv3 library to 25 µl of Endura Competent Cells (Lucigen) each. The electroporation was performed following manufacturer's instructions and 975 µl recovery medium (Lucigen) was added to the electroporated cells. The cells were added to a culture tube containing an additional 1 ml of recovery medium and placed in a shaking incubator at 250 rpm for 1 h at 37°C. All 8 ml of recovered cells were pooled and mixed well. To ensure full library representation, a titer to estimate the transformation efficiency was performed in parallel to the actual library amplification. For the titer, 10 µl of cells were added to 990 µl of recovery medium, mixed and 20 µl were plated onto pre-warmed 10 cm LB agar plates containing 100µg/ml carbenicillin (all Sigma-Aldrich). This dilution plate is equivalent to a 40 000-fold dilution of the full transformation reaction. For the library amplification, 400 µl of the recovered cells were spread on one pre-warmed 15 cm LB agar plate containing 100µg/ml carbenicillin, in total 20 plated were prepared. All plates were incubated for 14-16 h at 30°C. To calculate the transformation efficiency, the number of colonies on the dilution plate were counted and multiplied by 40 000 to get the total number of colonies. To maintain a coverage of 200 colonies per sgRNA and proceed with the library amplification, more than 1.4×10^7 colonies were required. To harvest the library colonies, 7 ml of LB media containing 100µg/ml carbenicillin was added to each 15 cm plate, the colonies were scraped with a cell spreader and pooled into a centrifuge bottle. The bottle was centrifuged for 10 min at 5000 rpm, the media was discarded and the wet pellet was weighed to estimate the amount of plasmid purification columns. The plasmids were extracted using the plasmid mega kit

(Qiagen) and once sgRNA representation was validated, lentivirus was produced following the protocol described in section 2.7.1. sgRNA representation was confirmed by sequencing on a NextSeq500 (Illumina) using 75 base pair (bp) single end reads with a total mapped read depth of 22.7 million reads, equivalent to 319 mapped reads per sgRNA. 70816 sgRNAs or 99.61% could be detected and reads were normally distributed with a low number of over- or underrepresented sgRNAs, indicating even and good library representation (Figure 4.2). The sequencing was performed by the Oxford Genomics Centre and the library amplification was performed by Dr Dylan Jones at the TDI, University of Oxford.

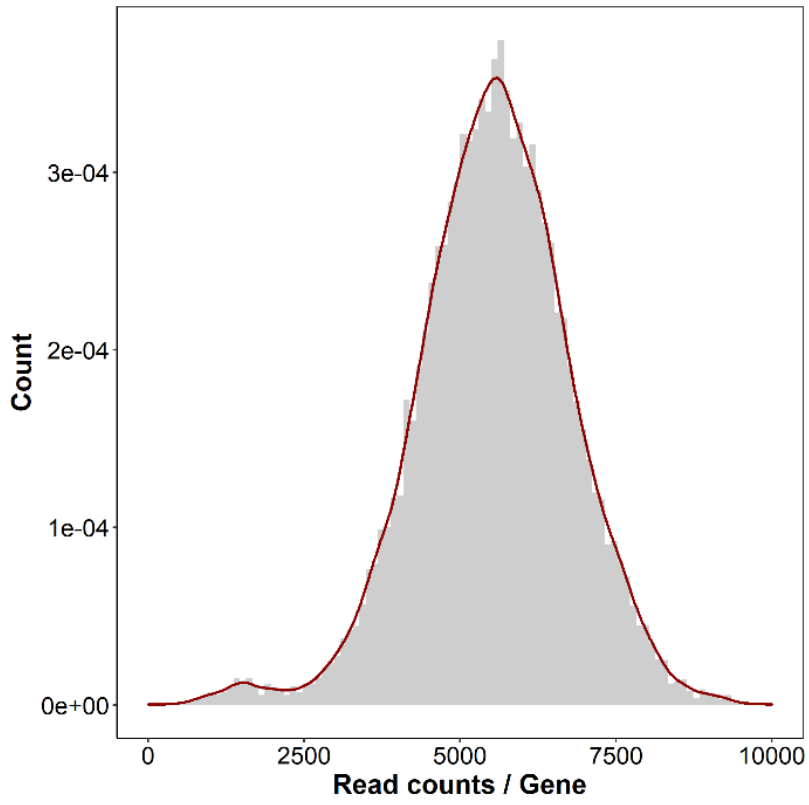


Figure 4.2 – TKOv3 plasmid representation. TKOv3 library sequencing read counts per gene after plasmid amplification are displayed as histogram with 100 bins, the red line indicates the distribution.

4.2.4 Cellular growth and transduction

The genome-wide screen was performed at a coverage of 500 cells per sgRNA and a MOI of 0.3. To have enough cells for the screen and due to the slow growth rate of EndoC-βH1, the cells were grown up over 18 weeks. The first replicate was performed after 10 weeks and the second replicate after another 8 weeks. 670 Million cells were transduced with the CRISPR library in replicate 1 and 744 Million in replicate 2. The average growth rate per week for replicate 1 was 1.56 ± 0.18 and 1.36 ± 0.13 for replicate 2. The cells were grown in T175 flasks at a seeding density of 12 million cells per flask and were split consistently every week on the same day by me. The cells were split in batches of six flasks at once and to keep the population homogenous, different flasks were combined as one batch each week. The cells were transduced for 6 h with the TKOv3 library or Cas9 only vector at a MOI of 0.3. Wildtype, EV cells and antibiotic selection controls were also included for subsequent steps. After 72 h, 5 µg/ml puromycin was added to all transduced cells and the antibiotic selection control and the cells were selected for 7 days. After 3 days in selection, media was changed to remove dead cells and supplement fresh puromycin.

4.2.5 Flow cytometry

Individual FACS analysis for optimisation experiments was performed using the following protocol with specific adaptions as explained in the main results. In the genome-wide screen, cells were harvested for FACS staining and sorting after puromycin selection. Cells were harvested in batches of 6 flasks following routine splitting protocols, as described in section 2.1. The cell pellet was washed in PBS supplemented with 1% BSA and split into separate tubes with a maximum of $5x10^6$ cells per tube. To exclude dead cells in the analysis, a live/dead stain using the LIVE/DEAD Fixable Far Red Dead Cell Stain Kit (Thermo Fisher) was performed. The samples were incubated at RT for 30 min in the dark and washed with

1% BSA in PBS. To fix and permeabilise the cells, the samples were vortexed and incubated with the Fixation/Permeabilization solution from the BD Cytofix/Cytoperm kit (BD Biosciences) for 20 min at 4°C. Subsequently, the cells were washed in Perm/Wash Buffer (BD Biosciences) and incubated with primary antibodies (Table 4.3) overnight at 4°C. The following day, the samples were washed twice with 1% BSA in PBS and incubated with a suitable secondary antibody (Table 4.3) for 1 h at RT in the dark. All antibodies were diluted in Perm/Wash Buffer (BD Biosciences). The samples were then analysed on a BD FACSAria III (BD Biosciences) or SH800 (Sony) flow cytometer using a 100 μm nozzle. Samples were filtered through a 35 μm cell strainer to achieve a single cell suspension. Isotype controls and individual controls stained with each antibody alone were analysed alongside the samples. Initial FACS staining and flow cytometer analysis for optimisation experiments was performed in collaboration with Dr Elena Navarro-Guerrero at the Target Discovery Institute. Flow cytometry data was analysed using Flowjo 10.6 (BD Biosciences).

Antibody	Dilution	Species	Manufacturer details
INS	1:10	Rabbit monoclonal	Cell Signaling, #3014
INS	1:10	Rat monoclonal	DHSB, GN-ID4
INS	1:10	Rat monoclonal	R&D, MAB1417
IgG	1:10	Rabbit monoclonal	Cell Signaling, #3900
IgG	1:10	Rat monoclonal	Invitrogen, 02-9602
Anti-rat IgG-AF488	1:500	Chicken polyclonal	Invitrogen, A-21470
Anti-rabbit IgG-AF488	1:200	Goat polyclonal	Invitrogen, A-11034

Table 4.3 – Flow cytometry antibodies and dilutions

In the genome-wide screen, the harvesting and staining protocol was performed simultaneously in a team of three, with one person harvesting and two colleagues performing different batches of staining. FACS staining was performed by Dr Roberta Baronio and Dr Elena Navarro-Guerrero and flow cytometer sorting in collaboration with Dr Ruddy Montandon at the Wellcome Centre for Human Genetics.

4.2.6 DNA extraction and PCR processing

DNA was extracted from frozen cell pellets using the QIAamp Blood Maxi/Mini Kit (Qiagen) by following manufacturer's instructions. To enrich the integrated sgRNAs within the genome and add barcodes and Illumina adapters for sequencing, two rounds of PCR amplification were performed as shown in Figure 4.4. PCR cycle numbers in PCR1 and PCR2 were optimised for each sample to avoid overamplification and introduce PCR-based biases. PCR1 amplified the integrated sgRNAs and one reaction contained 25 µl of 2x NEBNext Ultra II Q5 Master Mix (NEB), 2.5 µl each of PCR1_F and PCR1_R primer (10 µM regular desalted oligonucleotides, Table 4.4), 2.5 µg genomic DNA and nuclease-free H_2O up to 50 µl. PCR1 reactions were amplified for 30 sec at 98°C, a custom number of cycles per sample for 10 sec at 98°C, 30 sec at 66°C and 15 sec at 72°C followed by 2 min at 72°C and final cooling down to 10°C until each sample was pooled and further processed. To confirm amplification, the 600 bp product of PCR 1 product was visualised on a 1% gel (Figure 4.3). To be able to run all samples on one sequencing lane, each sample was equipped with a unique barcode in PCR2 by using a specific reverse primer for each sample (R01-R04, Table 4.4). To increase the complexity of the low diversity sgRNA samples, forward primers (F01-09, Table 4.4) contained different staggers and were used in a pool with a single reverse primer per sample. For PCR2, 25 µl of 2x NEBNext Ultra II Q5 Master Mix (NEB), 2.5 µl each of one reverse primer (R01-04) per sample and a pool of F01-09 (10 µM HPLC purified, Table 4.4), 5 µl of PCR1 and nuclease-free H_2O up to 50 µl. The

reaction was amplified for 30 sec at 98°C, a custom number of cycles per sample for 10 sec at 98°C, 30 sec at 55°C and 15 sec at 65°C followed by 5 min at 65°C and final cooling down to 10°C. To clean up the PCR samples for sequencing, all PCR2 products were pooled, 50 µl per lane were run on a 2% gel and extracted using the QIAquick Gel extraction kit (Qiagen) followed by isopropanol precipitation if necessary due to low 260/230 ratios (Figure 4.3). The DNA samples were stored at -80°C until they were submitted for sequencing. The DNA extraction and PCR processing were performed by Dr Elena Navarro-Guerrero and Dr Roberta Baronio at the TDI, University of Oxford.

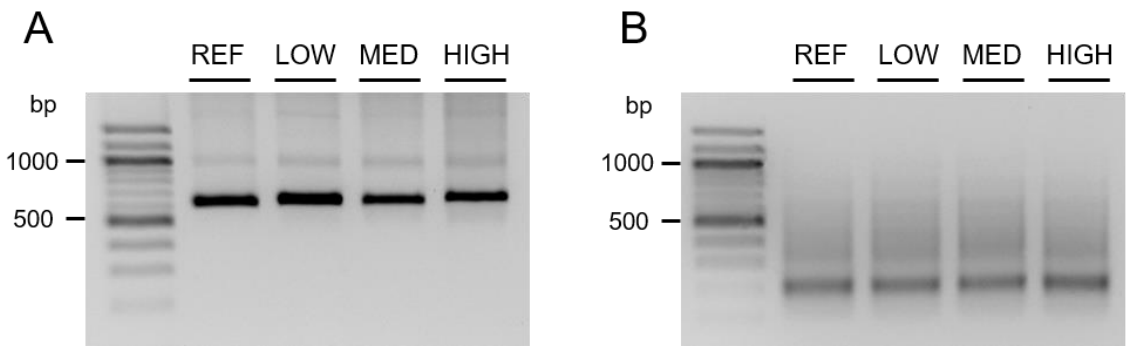


Figure 4.3 – sgRNA PCR amplification. PCR1 (A) and PCR2 (B) products were run on a 1% and 2% gel, respectively to confirm amplification. REF, LOW, MED and HIGH indicate CRISPR screen insulin content samples.

Primer	Sequence (5' → 3')
PCR1_F	AATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCG
PCR1_R	TCTACTATTCTTTCCCCTGCACTGTTGTGGGCGATGTGCGCTCT
R01	*AAGTAGAG#TTCTACTATTCTTTCCCCTGCACTGT
R02	*ACACGATC#ATTCTACTATTCTTTCCCCTGCACTGT
R03	*CGCGCGGT#GATTCTACTATTCTTTCCCCTGCACTGT
R04	*CATGATCG#CGATTCTACTATTCTTTCCCCTGCACTGT
F01	+TAAGTAGAGTCTTGTGGAAAGGACGAAACACCG
F02	+ATACACGATCTCTTGTGGAAAGGACGAAACACCG
F03	+GATCGCGCGGTTCTTGTGGAAAGGACGAAACACCG
F04	+CGATCATGATCGTCTTGTGGAAAGGACGAAACACCG
F05	+TCGATCGTTACCATCTTGTGGAAAGGACGAAACACCG
F06	+ATCGATTCCTTGGTTCTTGTGGAAAGGACGAAACACCG
F07	+GATCGATAACGCATTTCTTGTGGAAAGGACGAAACACCG
F08	+CGATCGATACAGGTATTCTTGTGGAAAGGACGAAACACCG
F09	+ACGATCGATAGGTAAGGTCTTGTGGAAAGGACGAAACACCG

Table 4.4 – CRISPR screen sgRNA PCR primer. The following sequences were abbreviated: CAAGCAGAAGACGGCATACGAGAT (*), GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT (#) and AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT (+).

4.2.7 Illumina sequencing

To assess the differences in sgRNA integrations between samples, the amplified sgRNA DNA samples were sequenced at the Oxford Genomics Centre (Wellcome Centre for Human Genetics). The multiplexed and pooled samples were run on one lane of the NEXTSeq500 (Illumina) as 75 bp paired-end reads. FASTQ were provided by Genomics Centre and used for downstream analysis. All samples in replicate 1 and 2 had a mean read depth of 83.5 ± 8.3 million reads and 82.7 ± 10.3 million reads, respectively.

4.2.8 Data analysis

CRISPR screen analysis was performed using Python 3.8 and R 3.5.

4.2.8.1 Model-based Analysis of Genome-wide CRISPR-Cas9 Knockout

Sequencing reads for each sample were merged using the cat command. To analyse sgRNA enrichment between samples, the Model-based Analysis of Genome-wide CRISPR-Cas9 Knockout v0.5.9.2. (MAGeCK) algorithm was used (Li *et al.*, 2014). In brief, MAGeCK generates read count tables from raw sequencing files followed by ranking sgRNAs based on their level of significance and identifying genes as hits that have consistently higher ranked sgRNAs as shown in Figure 4.4.

The count command was used to extract and count sgRNA sequences from the raw sequencing files. The input for the count command was the TKOv3 library file containing information about sgRNA sequences for each gene and compressed FASTQ files from all samples in both replicates. The output was a read count table for each sample across all genes and mapping information for each sample. All samples had more than 76.5% of all reads mapped to sgRNAs, well over 60% which indicate good sequencing quality. sgRNAs with zero counts in both reference samples were below 1%, demonstrating excellent sgRNA representation in the initial cell populations. Due to the FACS sorting, insulin selection samples should demonstrate sgRNA depletion or enrichment and showed as expected a higher percentage of missing sgRNAs than the reference sample (up to 8% of sgRNAs).

To identify read count changes between samples, MAGeCK first median-normalizes read counts across samples to be able to compare read count values between independent samples and replicates. The algorithm then uses a mean-variance model to estimate the variance of read counts and based on this variance, models the actual read count differences. To assess if these read count changes are significantly different between samples, MAGeCK assumes

that sgRNA read counts follow a negative binomial distribution and generates the probability (p-value) indicating how likely it is that the sample distribution deviation compared to the control population is random. The algorithm generates two one-sided p-values, one p-value for enrichment and one for depletion. MAGeCK subsequently ranks the sgRNA p-values and identifies if the sgRNA ranks for one gene are significantly different using a robust rank aggregation (RRA) algorithm. This algorithm examines if a significant fraction of all sgRNAs targeting one gene are ranked within top ranked sgRNAs. The RRA algorithm performs permutations (100 x number of genes) to test if randomly assigned sgRNAs have a significantly different distribution resulting in the computation of the false discovery rate (FDR) for each gene using the Benjamini-Hochberg procedure. These sgRNA and gene rankings were performed using the test command. This command used the raw count table and information about sample classification (treatment or control) as an input and generated a sgRNA and gene summary table as output. The test command was run with the paired option which performs paired comparison between treatment and control samples within replicates. The MAGeCK RRA analysis was performed to compare low and high insulin samples from both independent replicates. Only the sgRNA ranking was used for the subsequent high confidence hit calling pipeline.

4.2.8.2 High confidence gene hit classification

Due to high variation within the replicates of the insulin low samples, I decided to develop additional stringent criteria to classify genes as high confidence hits instead of using the MAGeCK gene ranking method, similar to a previously published approach (Fang *et al.*, 2019). MAGeCK considers all sgRNAs from both replicates independently, thus only analysing overall enrichment or depletion. Therefore, genes could be wrongly identified as hits even if only a small number of sgRNAs demonstrate the direction of effect which is often the case in data with high variance such as the insulin low sample (Figure 4.4). My

approach on the other hand focussed on the strongest sgRNAs within a gene that are consistent between replicates and therefore only identified genes as hits that have conclusive evidence for reproducibility and a strong effect on insulin content. This modified hit calling used the FDR thresholds for individual sgRNAs from the sgRNA ranking in the MAGeCK analysis (described in 4.2.8.1). To be classified as hits, genes had to have two or more sgRNAs out of their total four with an FDR<0.1 in both replicates. One gene with two significant sgRNAs in either direction and genes that were not expressed in beta cells were excluded. The hits determined based on this strategy were called high confidence hits.

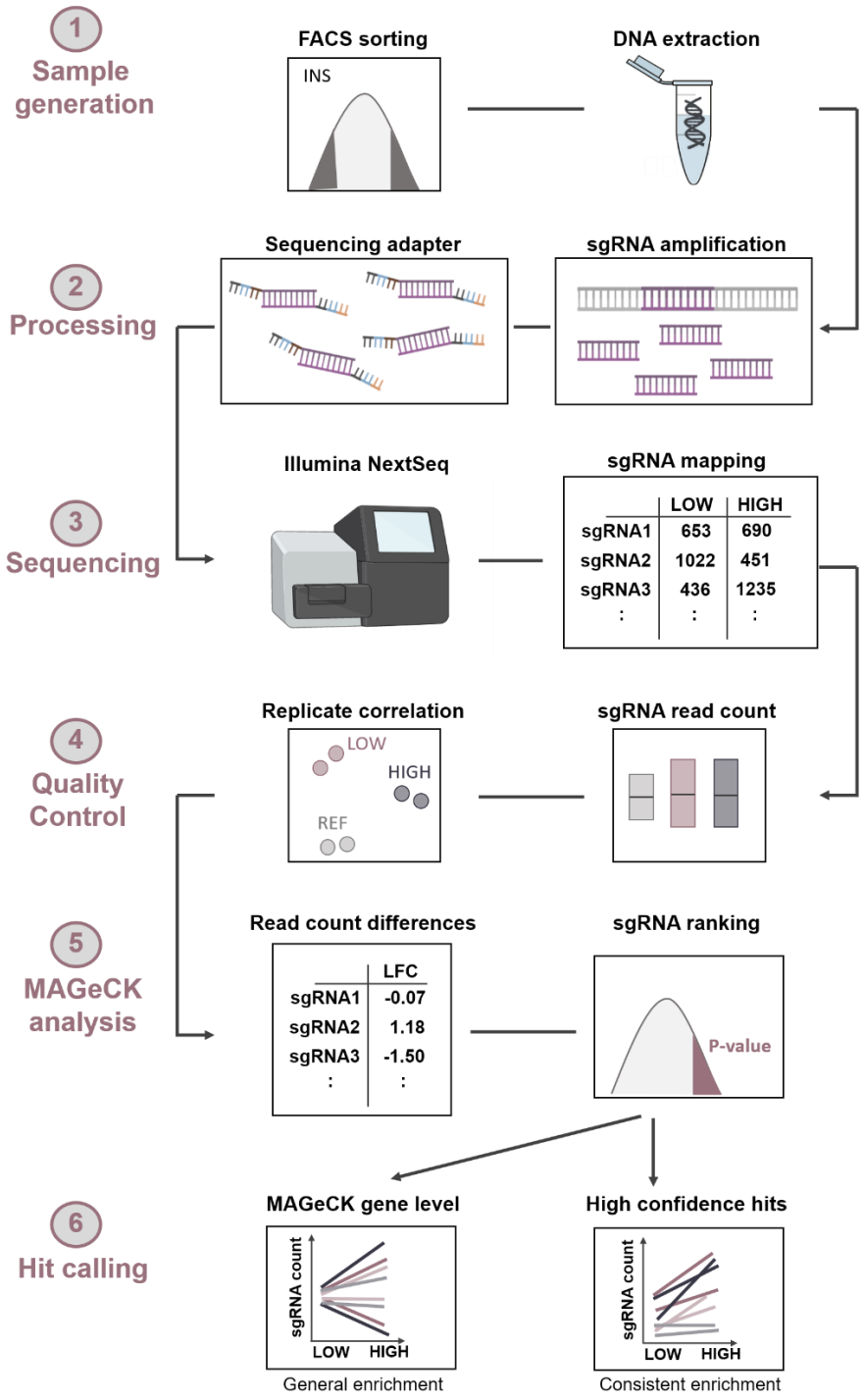


Figure 4.4 - CRISPR screen analysis pipeline. The pooled CRISPR screen was FACS sorted and the DNA extracted from individual samples (1), integrated sgRNAs were amplified and equipped with sequencing adapters in two rounds of PCRs (2), the fragments were sequenced and reads per sgRNA were counted (3), quality control was performed for read count distributions and replicate correlations (4) followed by assessing the sgRNA count differences between samples and assessing if they significantly deviated from the control population distribution (5). In addition to MAGeCK gene level analysis, a custom analysis focusing on high confidence hits was developed (6). More than two sgRNAs (each represented by the same colour per replicate) had to demonstrate consistent effects across replicates to be a high confidence hit whereas MAGeCK only assessed a general effect across sgRNAs without considering individual replicates.

4.2.9 Gene ontology and pathway enrichment

The Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.8 (https://david.ncifcrf.gov/) was used to identify enriched pathways within gene hits (Huang, Sherman, & Lempicki, 2009). The high confidence hits were submitted with homo sapiens as list and background species. The following DAVID categories with significant enrichment were selected for analysis: Online Mendelian Inheritance in Man (OMIM) disease, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway, gene ontology (GO) terms for biological process, cellular component and molecular function. Categories of interest with p-values less than 0.05 were highlighted. Corresponding enrichment and statistical details including the FDR value were indicated for each pathway.

4.2.10 Gene set enrichment analysis

Gene set enrichment analysis was performed on the ranked gene level list using GSEA 4.0.3 (Subramanian *et al.*, 2005). Positive and negative sorted lists of MAGeCK ranked genes were analysed separately using the 'Run GSEA on pre-ranked gene list' feature with 1000 permutations and KEGG pathway and GO terms as outputs. Highest-ranked pathways of interest based on the enrichment score were presented with corresponding p-values, albeit none of them passed the recommended FDR threshold of 0.25.

4.2.11 Protein interaction networks

To identify protein connectivity networks within the identified gene hits, STRING v11 was used to map physical and functional protein-protein interactions (https://string-db.org/) (Szklarczyk *et al.*, 2019). The genes classified as high confidence hits were submitted and correctly identified proteins were selected for mapping. Only interactions with a high confidence score of 0.9 or higher were chosen to reduce the number of false positive interactions. Specific subnetworks were identified manually and replotted individually.

4.3 Results

4.3.1 Establishing a genome-wide CRISPR screen for beta cell function

The successful implementation of a genome-wide CRISPR screen requires a disease relevant and experimentally suitable phenotypic readout and cell line. In T2D, the pancreatic beta cell has shown to be central to disease pathogenesis as described in detail in section 1.2.4 (Dimas *et al.*, 2014; Mahajan *et al.*, 2018b; Thurner *et al.*, 2018). The EndoC-βH1 cell line is the only established *in vitro* human beta cell model and the most similar cell line to primary beta cells (Ravassard *et al.*, 2011; Hastoy *et al.*, 2018; Lawlor *et al.*, 2019). In addition, the cell line is amendable to CRISPR/Cas9 genome editing as investigated in Chapter 3 and therefore an appropriate model system for genome-wide CRISPR screens. Suitable phenotypic readouts for pooled CRISPR screens are either simple sgRNA enrichment readouts based on proliferation or cell death or more complex FACS based sorting for intracellular proteins of interest. The primary function and phenotype of pancreatic beta cells is insulin secretion. However, in the setting of a pooled screen which contains all cells with distinct gene KO in the same flask, insulin secretion is not a suitable phenotypic readout as it cannot be associated with an individual gene KO when all cells secrete into the same shared media. A T2D relevant phenotype in human beta cells which is, in contrast to insulin secretion also suitable for FACS sorting and pooled CRISPR screens is intracellular insulin content. Insulin content is significantly reduced in patients with T2D and has shown to be causally implicated in disease pathogenesis based on studies for the gene *PAM* (Henquin, Ibrahim, & Rahier, 2017; Thomsen *et al.*, 2018). *PAM* has been causally linked to T2D through two coding variants that have been independently associated with disease risk and fine-mapping of one association signal have confirmed its causal role at this locus (Huyghe *et al.*, 2013; Steinthorsdottir *et al.*, 2014; Mahajan *et al.*, 2018b). Further experimental studies have demonstrated that reduction of PAM induces beta cell

dysfunction and reduced insulin secretion through a decrease in insulin content (Thomsen *et al.*, 2018). Insulin content is therefore a phenotypic readout which is not only amendable to a pooled screen as it is an intracellular FACS sortable marker but most importantly, it is also a T2D relevant phenotype. In summary, I created an outline for a genome-wide pooled CRISPR screen for beta cell function based on an insulin content readout in the human beta cell line EndoC-βH1. As shown in Figure 4.5, EndoC-βH1 were transduced with a genome-wide CRISPR library at a low MOI of 0.3 so every cell on average only gets one viral particle, one sgRNA integration and a single gene KO. After puromycin selection to eliminate untransduced cells, FACS staining and sorting were performed for low and high insulin content, which are the populations of interest with altered insulin content. DNA was extracted from the sorted cell populations, integrated sgRNA sequences were amplified using PCR and analysed by next-generation sequencing. sgRNA count was compared between low and high insulin content samples and enriched or depleted sgRNAs within each sample gave an indication about that gene's role in insulin content regulation. To validate the functional role of these genes and confirm the results of the CRISPR screen, subsequent validation studies were performed followed by downstream analysis including pathway enrichment and integration with complementary datasets.

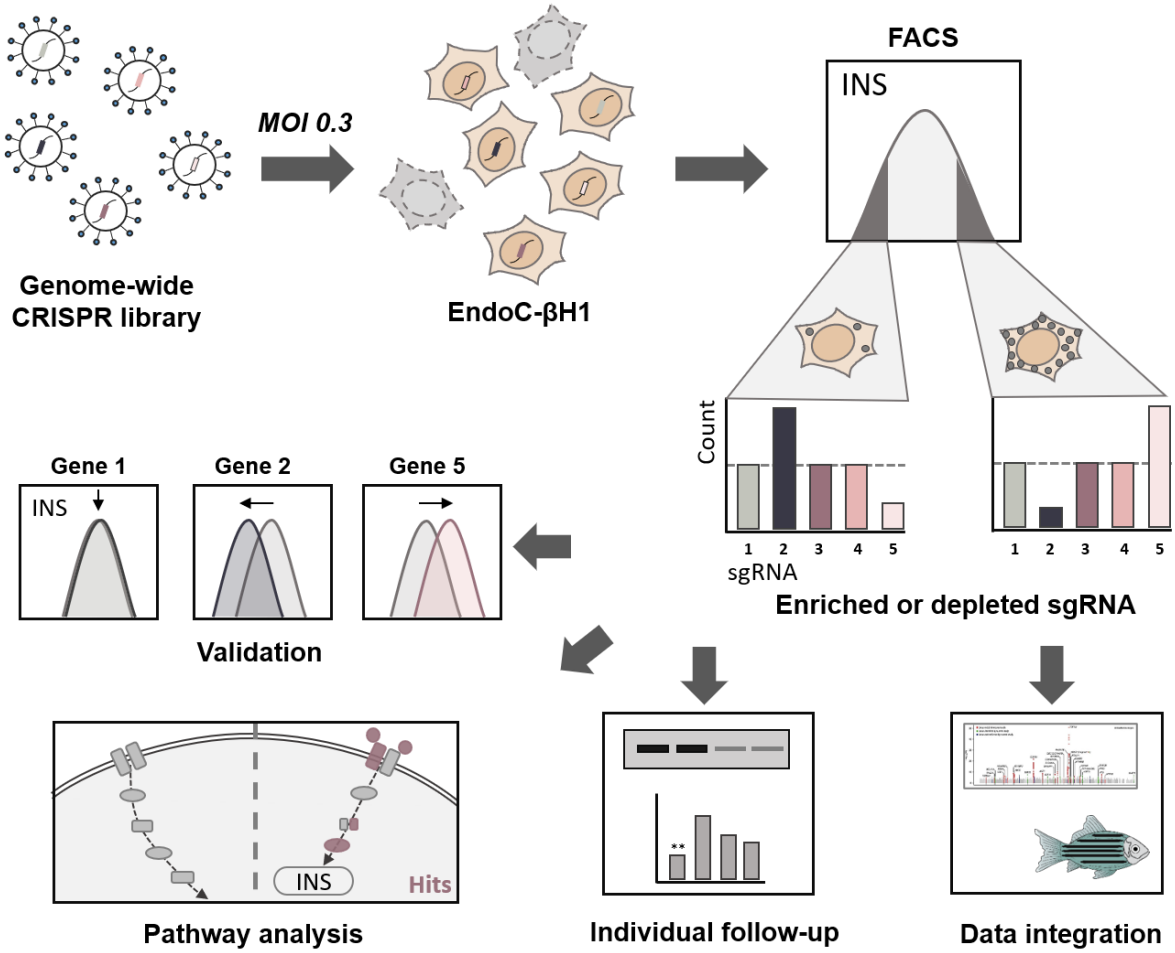


Figure 4.5 - Genome-wide pooled CRISPR screen for human beta cell function. To identify regulators of beta cell function in an unbiased and genome-wide approach, a CRISPR screen pipeline was developed. The human beta cell line EndoC-βH1 was transduced with a genome-wide CRISPR library at a MOI 0.3 to ensure single sgRNA integration. After puromycin selection, populations for different levels of insulin content were sorted using FACS and analysed for sgRNA integration. In this schematic, sgRNA2 is enriched in the low insulin content sample (*dark blue*) and depleted in high insulin content, indicating that KO of gene KO results in decreased insulin content. The sgRNA analysis was followed by individual validation and follow-up studies and downstream analysis such as integrating with pathway analysis and complementary datasets.

4.3.2 CRISPR screen optimisation

Pooled CRISPR screens based on complex FACS staining and sorting require extensive optimisation and validation to ensure a highly specific and sensitive phenotypic readout. Crucial steps in the optimisation pipeline include antibody testing and optimising the staining conditions.

4.3.2.1 Insulin content readout

Initially, antibodies from different manufacturers, species backgrounds and fluorophore conjugations were tested to achieve a specific signal for intracellular insulin in EndoC-βH1 (Figure 4.6). The first tests were performed under standard staining conditions using 4% PFA for cell fixation and 0.2% Triton-X100 for cell permeabilisation. A rabbit antibody from Cell Signaling and a rat antibody from R&D in combination with a species-specific secondary antibody both resulted in specific staining for insulin compared to isotype controls (Figure 4.6 A, B). Maximal separation between the insulin stained cells and isotype controls was not yet achieved, likely due to suboptimal staining conditions. Due to higher consistency between independent replicates and better availability for shipments in larger quantities, the Cell Signaling antibody was superior to the DHSB antibody. Antibodies which are directly conjugated to their fluorophore would allow for a faster staining process without the need for separate secondary antibody incubations. However, the signal for insulin in EndoC-βH1 from directly conjugated antibodies using two independent antibodies (Cell Signaling and R&D) was not different compared to cells stained with isotype controls only (Figure 4.6 C, D). Hence, I decided to proceed with the Cell Signalling antibody in combination with a secondary antibody to perform further optimisation and validation experiments.

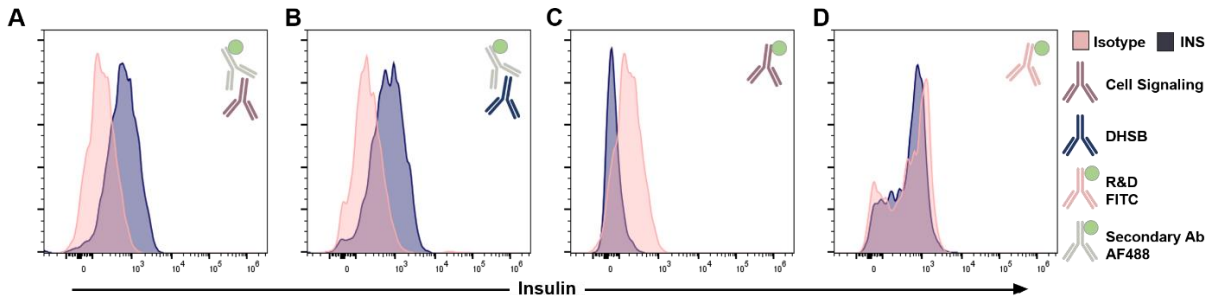


Figure 4.6 – Insulin antibody testing. FACS staining in EndoC-βH1 comparing staining for intracellular insulin *(blue graph)* and respective isotype controls *(pink graph)* using different monoclonal antibodies: **(A)** Cell Signalling rabbit antibody *(purple)* and a secondary Ab *(grey)* conjugated to AF488 *(green)*, **(B)** DHSB rat antibody *(blue)* and a secondary Ab *(grey)* conjugated to AF488 *(green)*, **(C)** Cell Signalling rabbit antibody *(purple)* directly conjugated to FITC *(green)*, **(D)** R&D rat directly conjugated to FITC *(green)*. Three independent experiments were performed and representative FACS plots are shown.

To ensure that the Cell Signaling antibody specifically detects insulin, I performed validation studies in other cell lines. Positive staining in the insulin expressing rat insulinoma cell line INS-1 indicated that the antibody was not specific for human insulin but also cross reacted with rat insulin (Figure 4.7 A). Negative insulin staining in the human embryonic kidney cell line HEK293T compared to isotype controls indicated that the antibody was indeed specific for insulin only and did not demonstrate any unspecific binding to other human proteins (Figure 4.7 B). The absence of insulin expression in HEK293T cells was confirmed on the mRNA level using RT-qPCR.

To further validate the sensitivity of the antibody and its ability to detect different levels of insulin content, I performed siRNA silencing of insulin in EndoC-βH1. An efficient silencing of 97.42% on the mRNA level induced a clear shift in average fluorescence towards less insulin in insulin silenced cells compared to non-targeting control cells (Figure 4.7 C, D). In addition to this shift of average fluorescence intensity, the proportion of insulin negative cells compared to the isotype control increased from 3% in siNT to 25.3% in insulin silenced cells. Less efficient silencing of only 52.79% of the insulin transcript demonstrated

a proportionally smaller reduction in insulin signal (Figure 4.7 E, F). The residual insulin protein and FACS signal even when silencing efficiencies close to 100% were achieved, was in accordance with remaining insulin that was measured by the sensitive and complementary alphaLISA based method (Figure 4.7 G). siRNA induced KD of insulin achieved on average only a mean reduction of 54.23% (p=0.16) in insulin content, likely due to residual insulin granules that usually have a half-life of 3-5 days (Halban & Wollheim, 1980). Conclusively, the Cell Signaling antibody is specific for insulin and able to reflect varying levels of insulin content with a corresponding FACS signal.

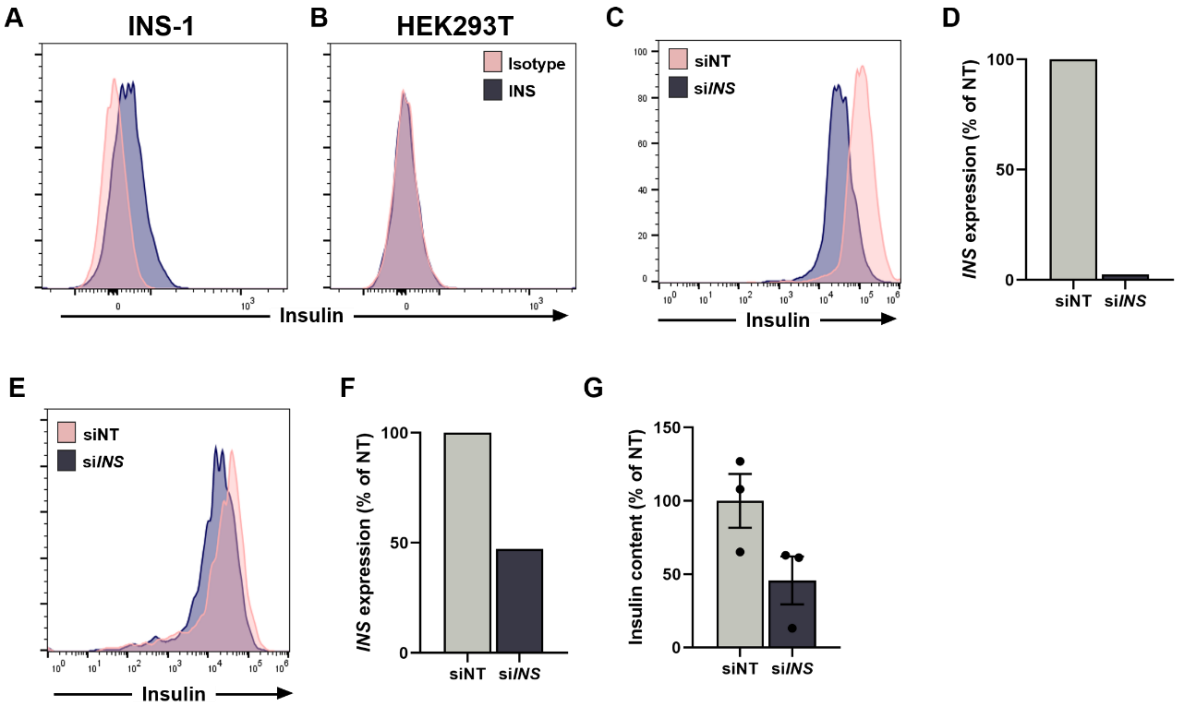


Figure 4.7 – Insulin antibody specificity. (A, B) FACS staining using the insulin targeted Cell Signaling antibody *(blue)* or a respective isotype control *(pink)* in the rat insulinoma cell line INS-1 **(A)** and the human embryonic kidney cell line HEK293 **(B)**. **(C, E)** FACS staining using the insulin targeted Cell Signaling antibody in insulin silenced cells *(blue)* or their respective non-targeting control *(pink)* in EndoC-βH1. **(D, F)** mRNA expression of *INS* in insulin silenced cells and siNT controls. Analysis in D and F correspond to FACS plots in C and E, respectively. **(G)** Insulin content in siNT cells compared to siNT, measured by alphaLISA. All data are mean ± SEM from three **(G)** independent experiments and representative FACS plots are shown **(A, B, C, E)** with their respective silencing efficiency **(D, F).** Data were analysed using two-sample t-test **(G)**. NT, non-targeting.

Although the Cell Signaling insulin antibody resulted in a specific and sensitive FACS signal for insulin in EndoC-βH1, the signal was not fully separated from the isotype control to form a distinct population. I therefore further optimised the fixation and permeabilisation conditions in the FACS protocol to achieve a better separation and improved insulin staining. Changing the fixation reagent in the previously applied staining protocol from 4% PFA into 100% Methanol decreased the proportion of insulin positive cells even further from 51.3% to 39.3% (Figure 4.8 C, D). However, when the commercially available BD Phosflow fixation and permeabilisation reagents were used with their recommended protocol adaptions, the proportion of insulin positive cells increased to 89%, indicating a near complete separation from the isotype control (Figure 4.8 A). Changing the fixation reagent from 4% PFA into BD Phosflow's fixation reagent was thereby the crucial step in improving the insulin signal (87.2%, Figure 4.8 B). Subsequent FACS staining for insulin and the CRISPR screen were performed with the BD Phosflow staining protocol using the Cell Signaling antibody.

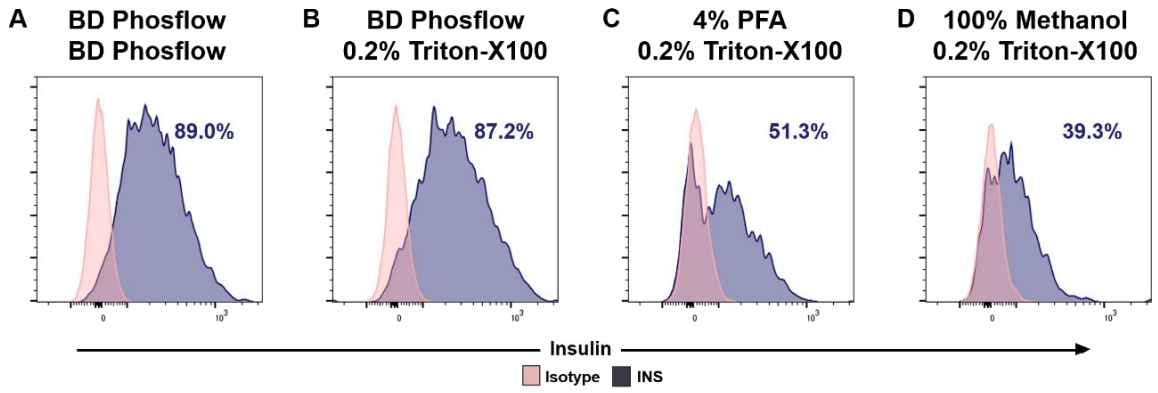


Figure 4.8 - Fixation and permeabilization optimisation. FACS staining in EndoC-βH1 comparing insulin staining for intracellular insulin *(blue graph)* and respective isotype controls *(pink graph)* using different staining protocols: (A) Commercial BD Phosflow fixation and permeabilisation buffer, (B) BD Phosflow fixation buffer and 0.2% Triton-X100 for permeabilisation, (C) 4% PFA as fixation buffer and 0.2% Triton-X100 for permeabilisation, (D) 100% Methanol as fixation buffer and 0.2% Triton-X100 for permeabilisation. Three independent experiments were performed and representative FACS plots are shown.

4.3.2.2 CRISPR library

In addition to optimising the conditions of the phenotypic readout, selecting a suitable CRISPR library is essential for successful pooled screening and depends on a wide range of distinct factors such as cell requirements or sgRNA targets. The choice for this screen was between the TKOv3 and Genome-scale CRISPR Knock-Out v2 (GeCKOv2) human CRISPR LoF libraries as they were well established at our screening collaborators laboratory at the Target Discovery Institute (University of Oxford) (Sanjana *et al.*, 2014; Hart *et al.*, 2017). Both CRISPR libraries use the same sgRNA and Cas9 plasmid backbone (plentiCRISPRv2) but differ in their sgRNA design, the number of sgRNAs per target and the total number of targets. GeCKOv2 targets more genes than TKOv3 (19050 versus 18053) with a higher number of sgRNAs per gene (6 versus 4). More sgRNAs also means a much higher number of cells is required to keep appropriate sgRNA coverage during the screen. Due to the slow growth rate of EndoC-βH1, fewer cells and therefore the TKOv3 library is preferential. In addition, it has been shown that increasing the number of sgRNAs into more than four sgRNAs per gene does not substantially increase the number of hits even further (Hart *et al.*, 2017). Further, it has emerged that a design flaw in the GeCKOv2 library resulted in a significant proportion of duplicate and non-specific sgRNAs. Conclusively, the TKOv3 CRISPR library targeting 18 053 genes with four sgRNAs per gene was the best library for this FACS based insulin content screen.

4.3.2.3 Small-scale proof of concept screen

Before performing the large-scale genome-wide CRISPR screen, I validated the CRISPR screen protocol and pipeline by performing a small-scale proof of concept screen. Instead of a genome-wide CRISPR library, I used the constructs and lentivirus that were previously generated in Chapter 3 for the selected genes of interest *PAM, IDE, INS, NMS* and the EV control. Both, *PAM* and *INS* are known to reduce insulin content upon protein reduction and

function as positive controls in the small-scale screen (Thomsen *et al.*, 2018). *NMS*, which is not expressed in EndoC-βH1 and EV on the other hand are negative controls that should not affect insulin content. Controls for the opposite direction of effect, an increase in insulin content were not as easily identifiable compared to many well characterised genes that induce a reduction in insulin content. Due to *IDE*'s role in the degradation of insulin and its expression in human beta cells, I hypothesised that *IDE* KO might lead to an increase in insulin content due to reduced insulin degradation. However, the consequences of *IDE* deletion on insulin content in EndoC-βH1 have not been tested before and are unknown. EndoC-βH1 cells were transduced with a pooled lentivirus to achieve a combined MOI of 0.3 and including two sgRNAs per gene. In parallel, cells were separately transduced with EV virus as a control for a normal insulin content distribution in FACS. The FACS sorting was performed following puromycin selection using the previously optimised staining protocol which was also adopted for the genome-wide screen. Upon analysis by FACS and after gating for alive and single cells, the screening sample demonstrated a larger proportion of cells with lower insulin content compared to EV control cells (Figure 4.9 A). The sorting gate for low insulin (INS^{low}) was set to the lower limit of EV cells and contained 38.5% of the screening population. All cells with higher insulin content beyond the sorting gate were collected as insulin high sample (INS^{high}, 56.2%). Instead of performing next-generation sequencing to assess sgRNA enrichment in the FACS sorted cell samples, I developed a SYBR green based qPCR-based approach using sgRNA specific primer on the extracted DNA samples. Apart from *PAM*_sgRNA1, all sgRNA amplifications were specific based on melt curve analysis. Based on two independent replicates of this small-scale screen, the control samples NMS and EV showed no change between INS^{low} and INS^{high} samples (EV: log fold change (LFC)=0.042, p=0.935; *NMS*: LFC=0.052, p=0.833, Figure 4.9 B). sgRNAs targeting the positive controls *INS* and *PAM* were, as expected enriched in the INS^{low}

population. More specifically, sgRNAs for *INS* were significantly enriched (p=0.014) with a strong and consistent effect (LFC: 1.476) across sgRNAs and replicates. The single sgRNA for *PAM* demonstrated as anticipated a highly replicable but smaller enrichment than *INS*, consistent with its lower effect on insulin content (LFC=0.373, p=0.061). Deletion of *IDE* did not significantly change insulin content in EndoC-βH1 (LFC=0.251, p=0.259). In contrast to my hypothesis that *IDE* KO might induce an increase in insulin content based on its function to degrade insulin, it did not actually affect the level of insulin in EndoC-βH1. Based on the correct identification of the positive controls *INS* and *PAM*, while the negative controls *NMS* and EV showed no effect, it can be concluded that the screening protocol including the viral transduction, antibiotic selection, FACS sorting and sgRNA enrichment testing is highly specific and sensitive to detect regulators of insulin content and suitable for genome-wide expansion.

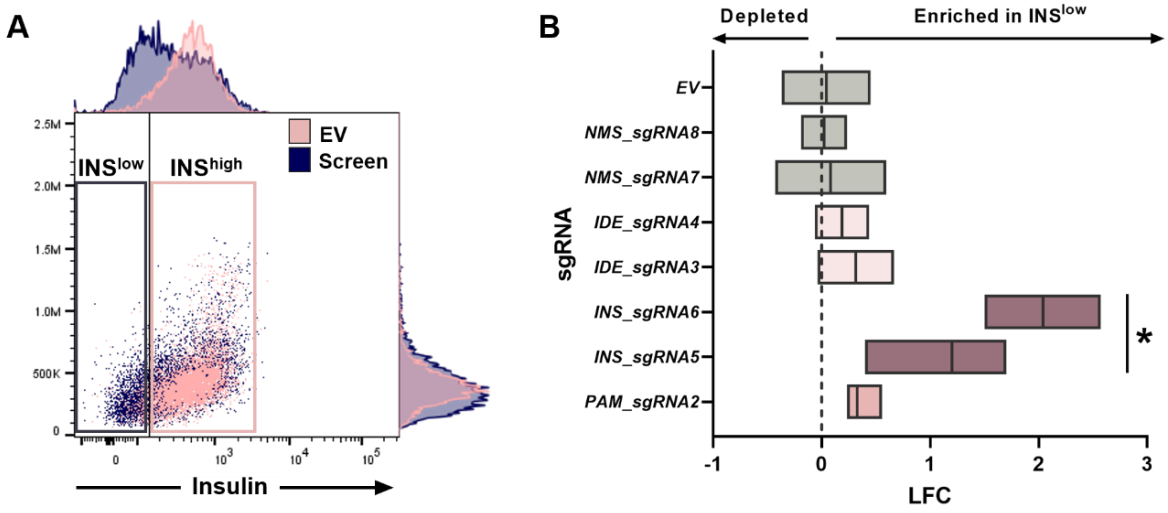


Figure 4.9 – Small-scale CRISPR screen in EndoC-βH1. (A) FACS staining in EndoC-βH1 comparing insulin staining of EV control cells *(pink)* and cells from the small-scale CRISPR screen *(blue)*. Respective boxes indicate sorting gates for low (INS^{low}) and high insulin content samples (INS^{high}). **(B)** sgRNA enrichment for all individual sgRNAs in INS^{low} compared to INS^{high} from control (*grey*) samples and positive controls (INS and PAM, *purple* and *dark pink*). Significance was tested for combined sgRNAs per gene. All data are mean ± SEM from two independent experiments and representative FACS plots are shown **(A)**. Data were analysed using one-sample t-test **(B)**. LFC, log2(fold change); EV, empty vector. P-value * < 0.05.

4.3.3 CRISPR screen performance

4.3.3.1 FACS CRISPR screening

This genome-wide CRISPR screen was independently performed twice in EndoC-βH1 to ensure the effects of certain sgRNAs on insulin content were reproducible and to increase the power to detect associations. As shown in Figure 4.10, each replicate required around 700 million cells to achieve a coverage of 500 cells per sgRNA and factoring in that only around 30% of the cells will survive after puromycin selection. It took around four months of continuous culturing to grow enough EndoC-βH1 in up to 70 large cell culture flasks which took up to 12 h of cell splitting at once to ensure cell synchronization and homogeneity. Within each replicate and after 7 days in puromycin, a reference cell sample was taken in addition to an insulin low and high sample after FACS sorting. The samples were then processed, sequenced and underwent downstream analysis.

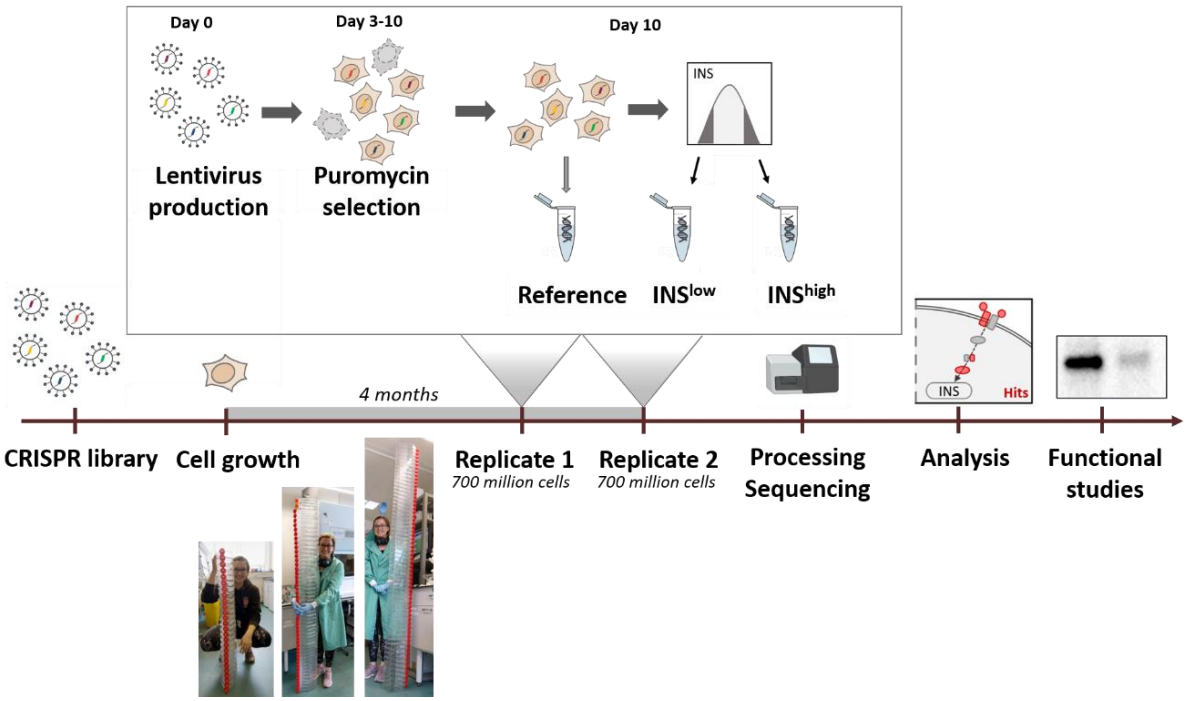


Figure 4.10 – Genome-wide CRISPR screen outline. Key steps and the timeline of the genome-wide CRISPR screen are shown. A common stock of lentiviral CRISPR library was produced and titered followed by bulking up EndoC-βH1 cells over four months to reach the required amount of around 700 million cells per replicate (*bottom images*). Each of the two replicates followed a precise time plan per day for key steps and sample collection as depicted in the enlarged box. After cell collection, the samples were processed and sequenced followed by sgRNA enrichment and downstream analysis and functional follow-up studies.

To assess different levels of insulin content in the screening cells that were transduced with the genome-wide CRISPR library, the extensively optimised FACS sorting for intracellular insulin was performed. Fixed, permeabilised and stained EndoC-βH1 cells were initially analysed based on their granularity and size to exclude debris and doublets (Figure 4.11 A). An additional live/dead stain was implemented to ensure that effects on insulin content were only assessed in viable cells. Only around 67.55% of cells as an average of both replicates passed this viability quality control, highlighting a remaining high percentage of dead cells from the puromycin selection or the staining process that would have compromised the downstream insulin content analysis. The sorting gates for insulin content were determined based on the isotype control and the control WT sample. As the isotype control functioned

as insulin negative control, its upward limit was set as a gate to separate cells with negative or low insulin staining from cells that stained positively for insulin (Figure 4.11 B). The positive control cells clustered as expected beyond that border with around 81.43% of cells that were classified as insulin positive, which is comparable to levels achieved in optimisation experiments (Figure 4.11 C). The screening cells demonstrated a clearly different distribution with a higher percentage of cells having low insulin (INS^{low}) staining compared to control cells, indicating the cells of interest where CRISPR editing induced a reduction in intracellular insulin (Figure 4.11 D-F). The upper limit in insulin content (INS^{high}) was determined as encompassing the top 10% of cells in the control population, resulting in similar proportions in the screening cells. A subset of the final sorted cells underwent a reanalysis to assess the accuracy of the sorting process and demonstrated a high purity of the population (99.1%, Figure 4.11 G).

The cell samples were then processed to extract the DNA, amplify the integrated sgRNA and attach sequencing adapters using a PCR approach with optimised cycles per sample to ensure ideal amplification without introducing bias. The samples were subsequently sent for sequencing at the Oxford Genomics Centre on a NEXTSeq500 followed by quality control and sgRNA mapping of the raw FASTQ sequencing files.

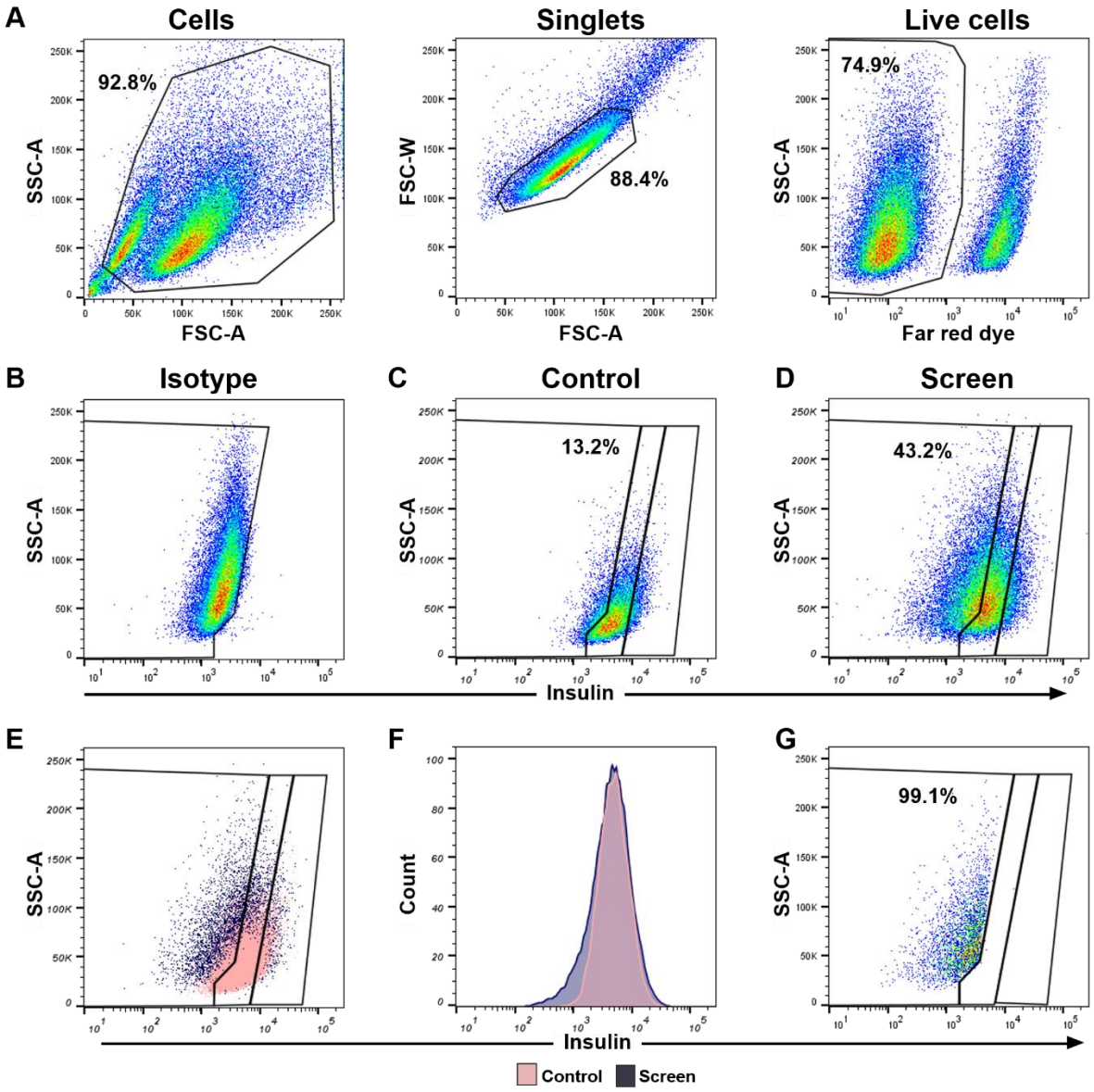


Figure 4.11 - FACS sorting of the genome-wide CRISPR screen. Representative FACS images and sorting gates from replicate one of the two genome-wide CRISPR screens. **(A)** Sorting gates assessing cell size, granularity and viability to exclude debris, define singlets and live cells. **(B-F)** Insulin staining of the isotype control **(B),** control cells **(C)** and screening cells **(D)**. Overlay of control *(pink)* and screening cells *(blue)* as dot plot **(E)** or histogram **(F)**. **(G)** Resorting of low insulin cell population to assess the sample purity. SSC-A, side scatter area; FSC-A, forward scatter area; FSC-W, forward scatter width.

4.3.3.2 Sample quality control

Initially, read counts from the sequencing files were assigned to sgRNAs from the CRISPR library as part of the MAGeCK analysis pipeline (Li *et al.*, 2014). The percentage of mapped reads is a good indication for sequencing and sample quality and should exceed 60% in each sample (Li *et al.*, 2015). With an average mapping proportion of 80.95%, all samples from both replicates passed the sgRNA mapping quality control (Figure 4.12 A). A low percentage of missing sgRNAs with less than 0.5% in both reference samples were indicative of near complete library transduction, stable integration of nearly all sgRNAs and no over selection during puromycin incubation (Figure 4.12 B). A higher percentage in the selected samples (low and high insulin) was expected and desirable due to the positive selection in the FACS sorting leading to sgRNA enrichment and depletion. The normalized read counts were consistent between samples with an average read count of 980.79 per sgRNA, thereby considerably exceeding the minimum requirement of around 300 reads/sgRNA which is necessary to achieve sufficient statistical power (Figure 4.12 C). In addition, the standard deviation of sgRNA read counts was lowest in the reference sample where no sample selection was performed. The larger spread in the low and high insulin samples represented the sgRNAs of interest which are enriched or depleted in either population. Overall, the screen has demonstrated an excellent technical performance and the sequencing and sgRNA read count samples have passed all quality controls.

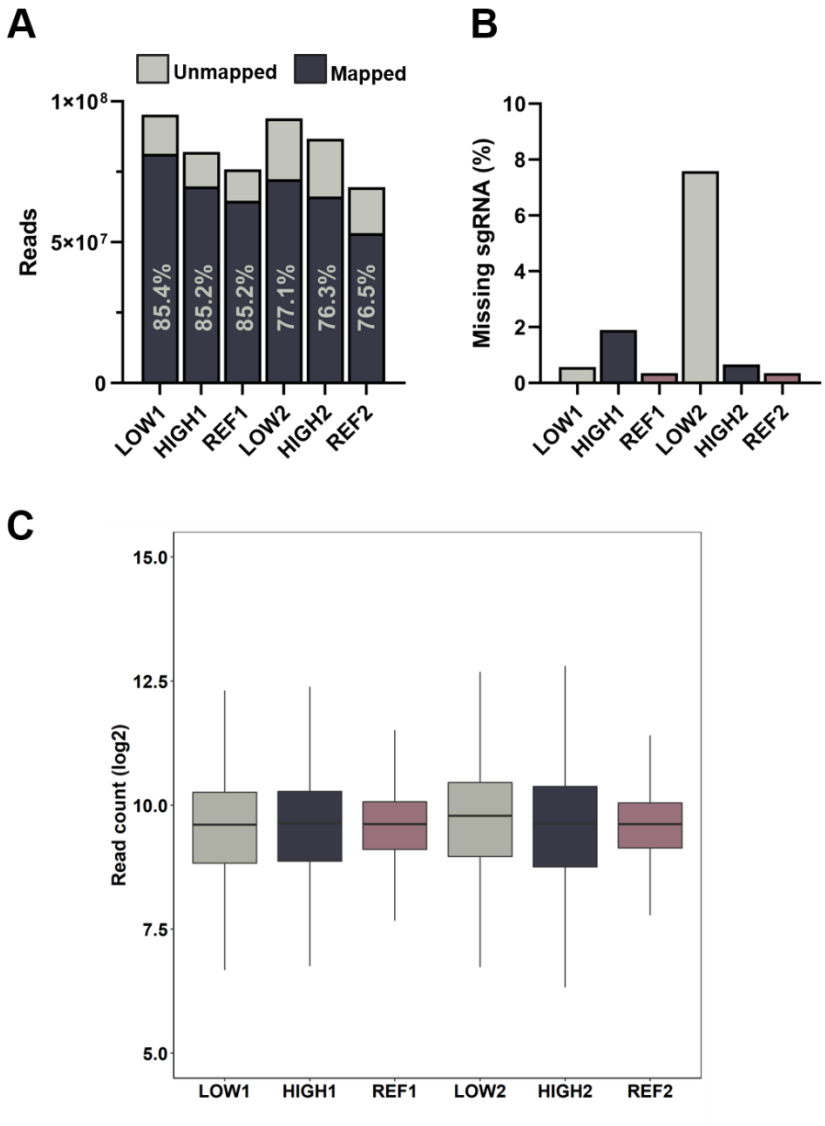


Figure 4.12 – Mapped sequencing read quality control. Read count and mapped sgRNA quality control for low insulin (LOW1, LOW2), high insulin (HIGH1, HIGH2) and reference samples (REF1, REF2) from both replicates including: **(A)** total read counts and the proportion of mapped reads, **(B)** the number of zero-count sgRNAs as proportion of total sgRNAs and **(C)** normalized read count distributions.

Following the sequencing quality control, I assessed the read counts on a sample level to determine the consistency between the samples and replicates. A Principle Component Analysis (PCA) demonstrated clear separation between the individual samples based on their screening group, here shown for the first two PCs (Figure 4.13 A). The reference and high insulin sample showed a close localisation within their replicates while the low insulin

sample, albeit clustering together, demonstrated a larger separation between both replicates. No batch effects between replicate 1 and replicate 2 or sample independent clustering were visible. This correlation between samples was confirmed using a pairwise Pearson correlation analysis on log read counts. Reference and high insulin content samples were highly correlated within their replicates with a correlation coefficient of 0.94 and 0.96 ($p<2.2x10^{-16}$), respectively (Figure 4.13 B, E). The overall correlation between both replicates within the screen was assessed by correlating the fold changes between low and high insulin content samples within each sample, resulting in a correlation coefficient of 0.50 ($p<2.2x10^{-16}$). Low correlation coefficients down to 0.1 are however not a cause of concern in positive selection screens where commonly only a small proportion of sgRNAs are consistently enriched or depleted (Li *et al.*, 2015; Hanna & Doench, 2020). This lower overall correlation compared to the reference and high insulin content sample was driven by variation within the low insulin content samples ($r=0.35$, $p<2.2x10^{-16}$), which was already indicated by a larger separation in the PCA analysis and varying levels of missing sgRNAs (Figure 4.13 D). This higher variability and reduced insulin content were likely a result of cellular stress due to CRISPR screen associated side effects such as the transduction and antibiotic selection and therefore effects on insulin content which are independent of the gene KO introduced by the sgRNA.

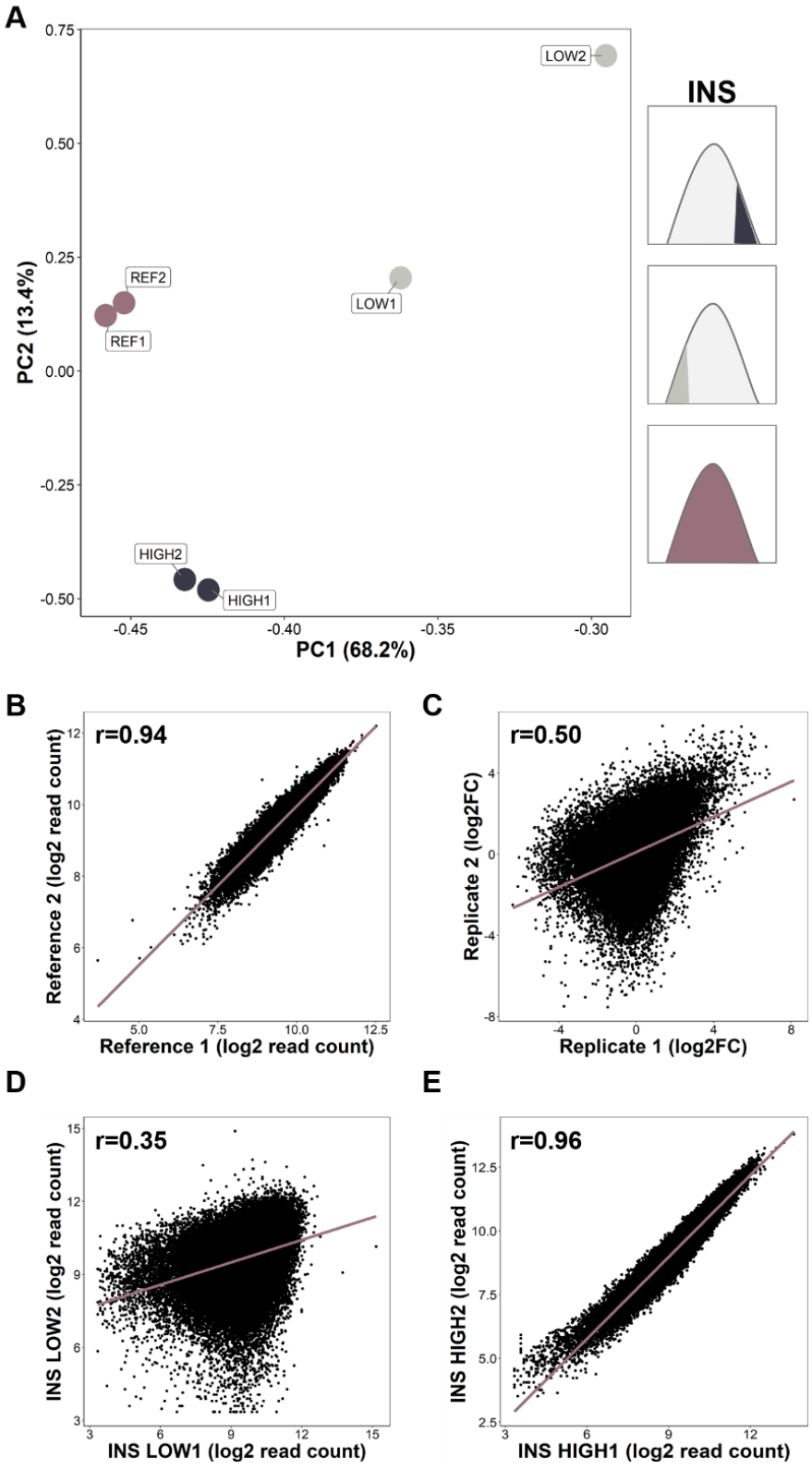


Figure 4.13 - CRISPR screen replicate correlation. (A) Principle Component Analysis for the first two components of low *(grey)* and high *(blue)* insulin content and reference samples *(purple)* from both replicates. **(B-E)** Pairwise Pearson correlation analysis of reference replicates **(B),** fold changes between low and high insulin content for each replicate **(C)**, low insulin content replicates **(D)** and high insulin content replicates **(E)**. The purple line indicates the linear regression line. Log2FC, log2 (fold change); r, Pearson correlation coefficient.

4.3.4 CRISPR screening hits

Significantly enriched or depleted genes in the FACS sorted insulin content samples of this genome-wide CRISPR screen were assessed using the MAGeCK analysis pipeline (Li *et al.*, 2014). This algorithm to analyse CRISPR screens first performs an enrichment analysis for individual sgRNAs between two groups of interest before collapsing the sgRNAs into a gene level-based analysis. Initially, this gene-level analysis was applied comparing the low and high insulin content groups followed by the development of a custom sgRNA-level based identification of high confidence hits.

This initial gene level hit analysis using a less stringent FDR threshold (<0.5), similar to thresholds in previously published studies identified 117 negative regulators and 1382 positive regulators of insulin content (Figure 4.14) (Cortez *et al.*, 2020). Negative regulators demonstrate enriched sgRNAs in the high insulin content population and increased insulin content upon gene KO. sgRNAs which are enriched in the low insulin content population on the other hand reflect positive regulators with a decrease in insulin content upon gene KO. This genome-wide CRISPR screen was strongly biased towards detecting positive regulators, which is consistent with the observed effect in the FACS readout. This is likely a consequence of identifying many factors that are critical for general transcription and translation, in addition to detecting specific regulators of insulin content.

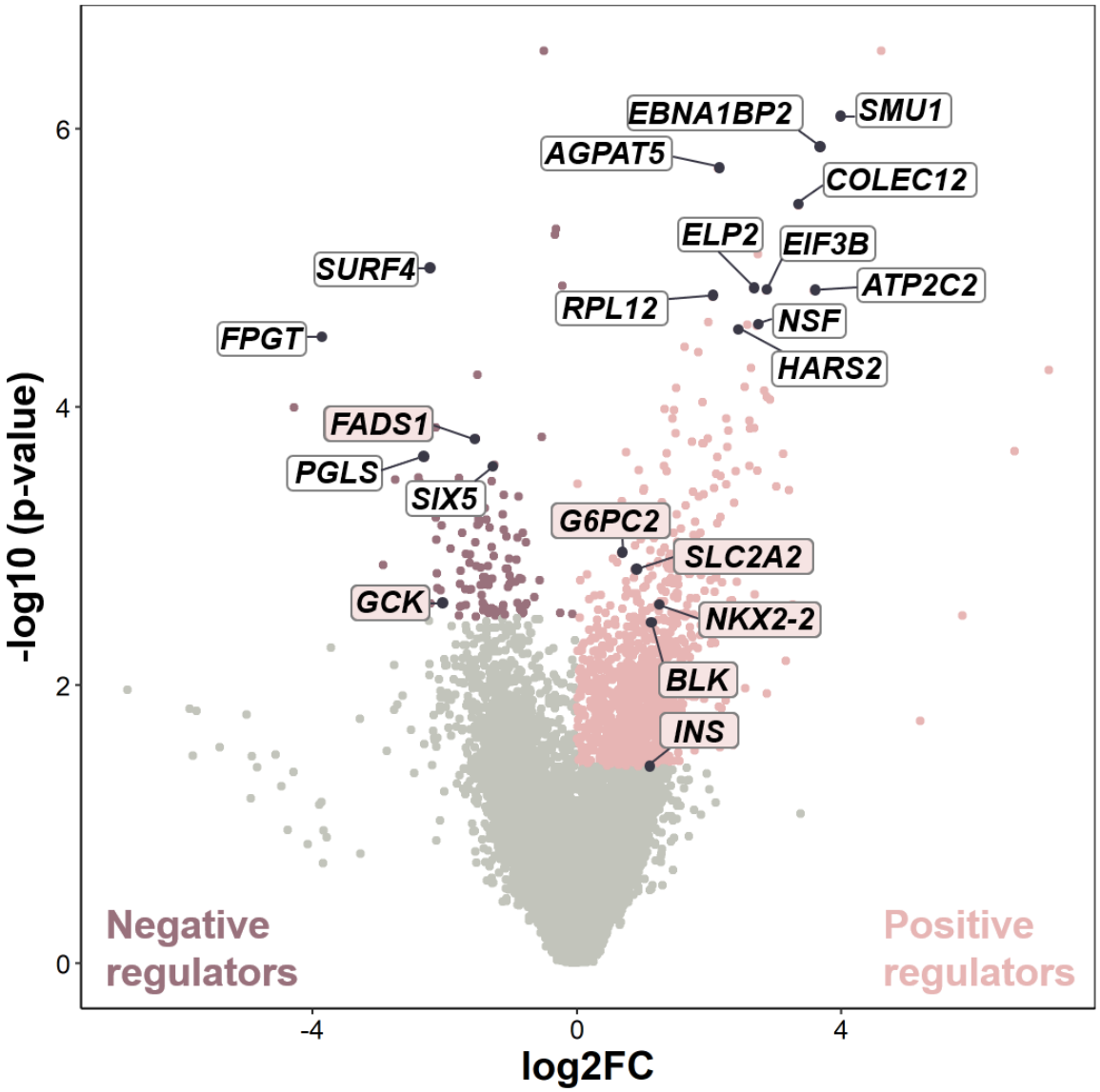


Figure 4.14 - Genome-wide CRISPR screen gene-level volcano plot. Screening hits based on MAGeCK gene-level analysis. Positive regulators *(pink)* and negative regulators *(purple)* leading to reduced and increased insulin content upon gene KO, respectively were indicated based on an FDR<0.5. Specific genes have been highlighted *(pink label)* based on their previous association with T2D risk or known role in beta cell function. Further comprehensive visualisation of these genes is shown in Figure 4.21. Other labelled genes indicate selected top hits. Data are from two independent genome-wide CRISPR screen replicates. FC, fold change; FDR, false discovery rate.

4.3.4.1 Known regulators of beta cell function and diabetes

As expected, sgRNAs targeting *INS* (LFC=1.06, FDR=0.49) were enriched in the low insulin content population. Both sgRNAs that had been used previously to generate *INS*-KO in Chapter 3 were efficient in introducing an effect on insulin content in the CRISPR screen. One of the four sgRNAs however, highlighted in grey in Figure 4.15 A did not affect insulin content in both replicates. It is therefore likely that this sgRNA did not introduce a gene KO in *INS* and no reduction in insulin content, leading to an overall smaller fold change and less significant result compared to other positive regulators with four consistent sgRNAs.

Most importantly, the screen was able to identify known regulators of beta cell function, insulin content and insulin secretion such as the beta cell transcription factor *NKX2.2* (LFC=1.21, FDR=0.26), the glucose transporter *SLC2A2* (LFC=0.87, FDR=0.22), the glucokinase *GCK* (LFC=-2.04, FDR=0.49) and the glucose-6-phosphatase catalytic subunit *G6PC2* (LFC=0.67, FDR=0.21, Figure 4.15). *NKX2.2* is essential in the development of beta cells but has also shown to play a crucial role in mature beta cells where loss of Nkx2.2 in islets of transgenic adult mice led to reduced insulin expression, insulin content and impaired insulin secretion (Sussel *et al.*, 1998; Doyle & Sussel, 2007). Homozygous LoF mutations in *SLC2A2*, which encodes the glucose transporter GLUT2 have been identified as a rare cause of neonatal diabetes, demonstrating its crucial role in the uptake of glucose and human beta cell function (Sansbury *et al.*, 2012). *GCK* is the primary glucose sensor in human beta cells, mediates the rate limiting step to phosphorylate glucose to glucose-6-phosphate and has been associated with MODY and neonatal diabetes (Matschinsky, 1990, 2005; Vionnet *et al.*, 1992; Thomson *et al.*, 2003). *G6PC2* catalyses the hydrolysis of glucose-6-phosphate (G6P) to glucose and has been established as an effector gene for glycaemic traits, a regulator of insulin secretion and loss of G6PC2 has been shown to reduce insulin content, consistent with the observed direction of effect upon KO in this screen (Pound *et al.*, 2013;

Mahajan *et al.*, 2015; Ng, 2016). The opposing effects of *G6PC2* and *GCK* in the beta cell were reflected in the results of this CRISPR screen by showing opposite effects on insulin content.

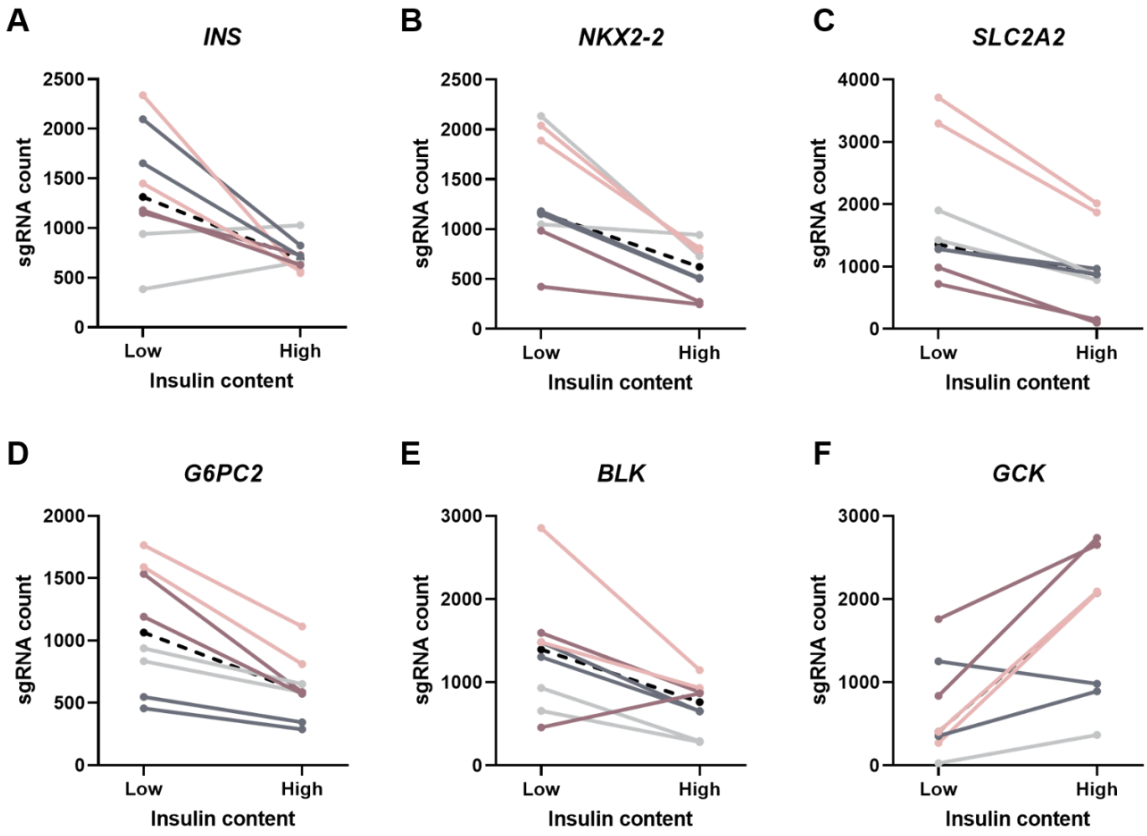


Figure 4.15 – Known regulators of beta cell function. Changes in sgRNA count from low to high insulin content sample with each colour representing the same sgRNA across the two screen replicates for **(A)** *INS*, **(B)** *NKX2-2*, **(C)** *SLC2A2*, **(D)** *G6PC2*, **(E)** *BLK* and **(F)** *GCK*. The black dashed line represents the median sgRNA count for this gene. sgRNA with zero counts were removed.

To provide further confidence in the hits of this genome-wide CRISPR screen, a Fisher's Exact enrichment analysis was performed to assess if genes involved in monogenic diabetes were significantly enriched. A list of genes involved in monogenic diabetes such as MODY or neonatal diabetes was manually curated based on publications reviewing the genetic causes of monogenic diabetes (Greeley *et al.*, 2011; Flanagan *et al.*, 2014; Bansal *et al.*,

2017; Misra & Owen, 2018). It should be noted however that not all genes are widely accepted as a cause of monogenic diabetes (e.g. *BLK*). Nevertheless to be able to capture all potential genes involved in monogenic diabetes, I decided to compile a broad list instead of applying very stringent criteria and only focussing on genes that are well characterised causes of monogenic diabetes such as *KCNJ11*. Out of the 31 tested genes that have been shown to be involved in monogenic diabetes, seven were classified as screening hit (*GCK, INS, GATA6, SLC2A2, SLC19A2, BLK, NKX2.2*), resulting in a significant enrichment (p=0.012).

4.3.4.2 Top screening hits

In addition to assessing well characterised regulators of beta cell function and insulin content, I evaluated the top hits of the CRISPR screen and their potential role in the human beta cell (Figure 4.14). A large proportion of positive regulators were involved in general transcription and translation and thus affected insulin content indirectly through decreasing total protein levels. Within the top hits were the spliceosomal factor *SMU1* (LFC=3.98, FDR=0.007), the 27S pre-rRNA processing factor *EBNA1BP2* (LFC=3.69, FDR=0.008), the transcriptional elongation subunit *ELP2* (LFC=2.67, FDR=0.02), the translation initiation factor *EIF3B* (LFC=2.87, FDR=0.02), the ribosomal protein *RPL12* (LFC=2.05, FDR=0.02) and the mitochondrial tRNA synthetase *HARS2* (LFC=2.43, FDR=0.03). Most of these hits are likely required for survival and the cells would not be viable during longer culturing following gene KO. Consistent with this crucial role in protein synthesis was the classification of all of these genes as common essential genes in the Cancer Dependency Map (DepMap) based on large CRISPR and RNAi screens (Dempster *et al.*, 2019b).

Beside genes that are involved in general transcription and translation, the screen has identified many genes of interest among the top hits who map to important pathways or functions within the beta cell but who have not been characterised in human beta cells yet.

These genes would therefore be compelling targets for in-depth follow up studies and include the vesicle fusing ATPase *NSF* (LFC=2.58, FDR=0.03), the ER cargo receptor *SURF4* (LFC=-2.21, FDR=0.04), the fucose-1-phosphate guanylyltransferase *FPGT* (LFC=-3.87, FDR=0.08), the fatty acid desaturase *FADS1* (LFC=-1.53, FDR=0.23), the transcription factor *SIX5* (LFC=-1.25, FDR=0.30), the calcium-transporting ATPase *ATP2C2* (LFC=3.59, FDR=0.02), the phospholipid biosynthesis acyltransferase *AGPAT5* (LFC=2.13, FDR=0.007) and the scavenger receptor *COLEC12* (LFC=3.35, FDR=0.01). *NSF* has been shown to play a role in beta cell function by catalysing the disassembly of the SNARE complex upon insulin vesicle fusion in exocytosis but the consequences of its loss of function on insulin secretion have not been characterised yet (Littleton *et al.*, 2001; Vikman *et al.*, 2003). *SURF4* is involved in the export of proteins in the ER and has been found in a purification of rough ER from the mouse beta cell line MIN6, highlighting its potential role in beta cell ER homeostasis (Lee *et al.*, 2015). *FPGT* is involved in hexose biosynthesis and has been identified as a target gene of *NKX6.1* in islets of beta cell conditional *Nkx6.1* KO mice, but its role in insulin secretion or beta cell function has not been investigated yet (Taylor, Liu, & Sander, 2013). *FADS1* regulates the synthesis of highly saturated fatty acids and increased islet expression has been associated with raised fasting glucose in GWAS (Dupuis *et al.*, 2010; van de Bunt *et al.*, 2015). Although its lipid associated function might indicate a role linked to insulin sensitivity, the GWAS allele has also been associated with HOMA-B, a measurement of beta cell function and insulinogenic index, hinting at a beta cell mediated effect (Dupuis *et al.*, 2010; Ingelsson *et al.*, 2010). *FADS1* expression might modulate insulin secretion through altered levels of unsaturated fatty acids which has been observed to be associated with the level of insulin secretion and gluco-sensitivity in islets (Pareja *et al.*, 1997). Interestingly, both *AGPAT5* and *COLEC12* have been identified as being differently expressed in human islets from donors with T2D

(HbA1c>6.5) compared to normoglycemic donors (HbA1c<6), but their role in human beta cells remains to be determined (Asplund *et al.*, 2019).

4.3.4.3 High confidence gene hit analysis

Even though the MAGeCK gene level analysis has provided a gene ranking demonstrating an enrichment of known regulators of beta cell function and insulin content, its algorithm to collapse sgRNAs into genes is not ideally suited for all CRISPR screens. The analysis pipeline considers all sgRNAs from both replicates as independent, therefore not putting special focus on sgRNAs which are only enriched or depleted consistently between both replicates. This is especially important in screens with a lower correlation between samples such as in the low insulin sample. Using MAGeCK, a gene could have been classified as hit by only having four or even less sgRNAs from one replicate with an effect. I therefore decided to perform an additional analysis and compiled a high confidence hit list containing only genes that have at least two out of the total four sgRNAs with a highly significant effect (FDR<0.1) in both replicates. This approach is similar to a previously published analysis, albeit using much more stringent criteria (Fang *et al.*, 2019). This high confidence hit list contained 580 genes (Supplementary Table 1), still including the strong and consistent beta cell regulators such as *INS, SLC2A2, NKX2.2* and previously strong hits including *FADS1, NSF, EIF3B* or *AGPAT5*. If not indicated otherwise, the subsequent downstream analyses were performed based on this high confidence hit list.

4.3.4.4 Negative control sgRNAs

In addition to assessing positive or negative hits in the CRISPR screen, it is essential to investigate the distribution of negative control sgRNAs targeting LacZ, luciferase and EGFP to evaluate a potential bias, a skewed distribution and background noise. The mean of the 284 control sgRNAs was not significantly different compared to the mean of all sgRNAs (p=0.298), highlighting that control sgRNAs with no effect on insulin content were not enriched or depleted in either low or high insulin content population (Figure 4.16).

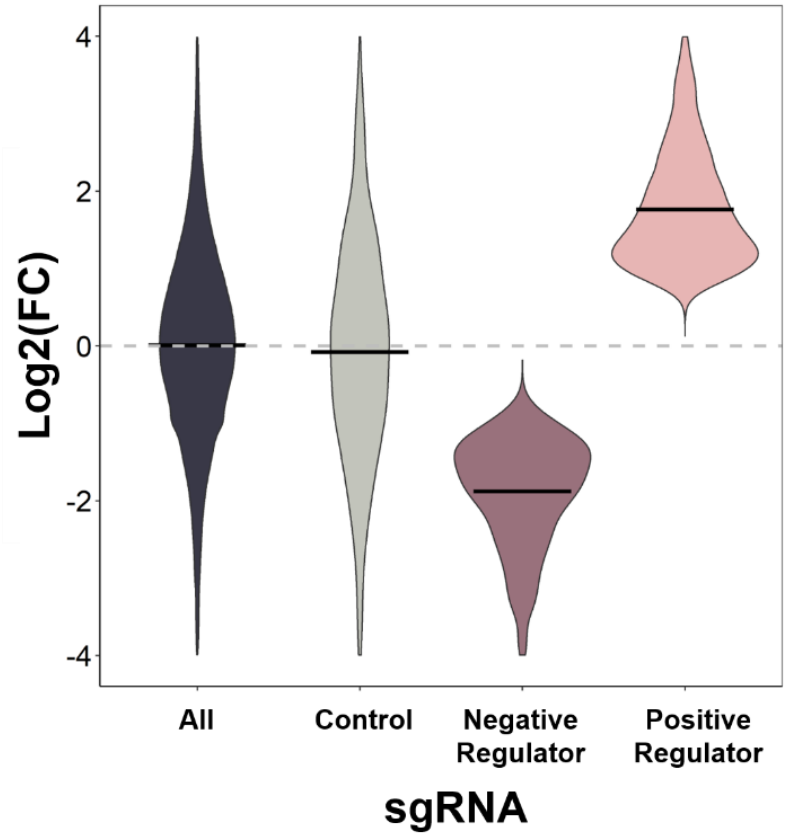


Figure 4.16 – sgRNA distribution in genome-wide CRISPR screen. sgRNA distribution of log2(FC) for the following groups: all sgRNAs independent of their sequence and effect on insulin content *(blue)*, only control sgRNAs targeting LacZ, EGFP and luciferase *(grey)*, sgRNAs from negative *(purple)* or positive regulator *(pink)* hits from the high confidence hit list. The black line indicates the median values for each group. Data are from two independent genome-wide CRISPR screen replicates and differences between all sgRNAs and controls were analysed using a two-sample t-test. Log2(FC), log2(fold change).

4.3.5 The CRISPR screen identified T2D and beta cell pathways

4.3.5.1 Pathway enrichment analysis

Initially, I manually assessed if individual genes known to play a role in beta cell function or T2D could be found within the hits of the CRISPR screen, a pathway analysis of all screening genes however can perform an unbiased overall enrichment analysis. I used two distinct methods to investigate pathway enrichment, Gene Set Enrichment Analysis (GSEA) using the ranked genes of the MAGeCK gene level analysis and DAVID pathway enrichment analysis using only the high confidence hits. GSEA determines enrichment based on the localisation of specific pathway genes across the ranks of the submitted list, therefore not specifically distinguishing between hit or not hit but rather assessing if the genes are higher ranked than others. Pathway enrichment analysis on the other hand evaluates if genes from the submitted hit list are enriched for certain pathways.

GSEA was performed separately for negative and positive regulators to have either at the top of the ranking. Within each category, the highest-ranked pathways based on their enrichment score are presented, albeit none of them passed the recommended FDR threshold of 0.25. Consistent with our bias to detect positive regulators, no KEGG pathway enrichment was detected amongst negative regulators. Enriched KEGG pathways for positive regulators were MODY (normalised enrichment score (NES)=1.32, p=0.019) and T2D (NES=1.30, p=0.003, Figure 4.17 A). Highly enriched GO terms for negative regulators include the beta cell pathways insulin receptor binding (NES=1.45, p=0.003), SNARE complex assembly (NES=1.44, p=0.012) and regulation of vesicle fusion (NES=1.31, p=0.022, Figure 4.17 B). As previously demonstrated for individual top hits, positive regulators are highly enriched for transcription and translation pathways such as regulation of translational elongation (NES=1.38, p=0.018) or mRNA transcription by RNA Polymerase II (NES=1.37, p=0.018, Figure 4.17 C). In addition, pathways important for beta cell function such as ER quality

control (NES=1.37, p=0.013) were also among the most highly enriched pathways. Conclusively, GSEA analysis demonstrated a strong enrichment for genes relevant for beta cell function, insulin content, insulin secretion and total protein levels within the top hits of the MAGeCK ranking, increasing the confidence that this genome-wide CRISPR screen was indeed able to robustly identify genes relevant to the selected phenotype.

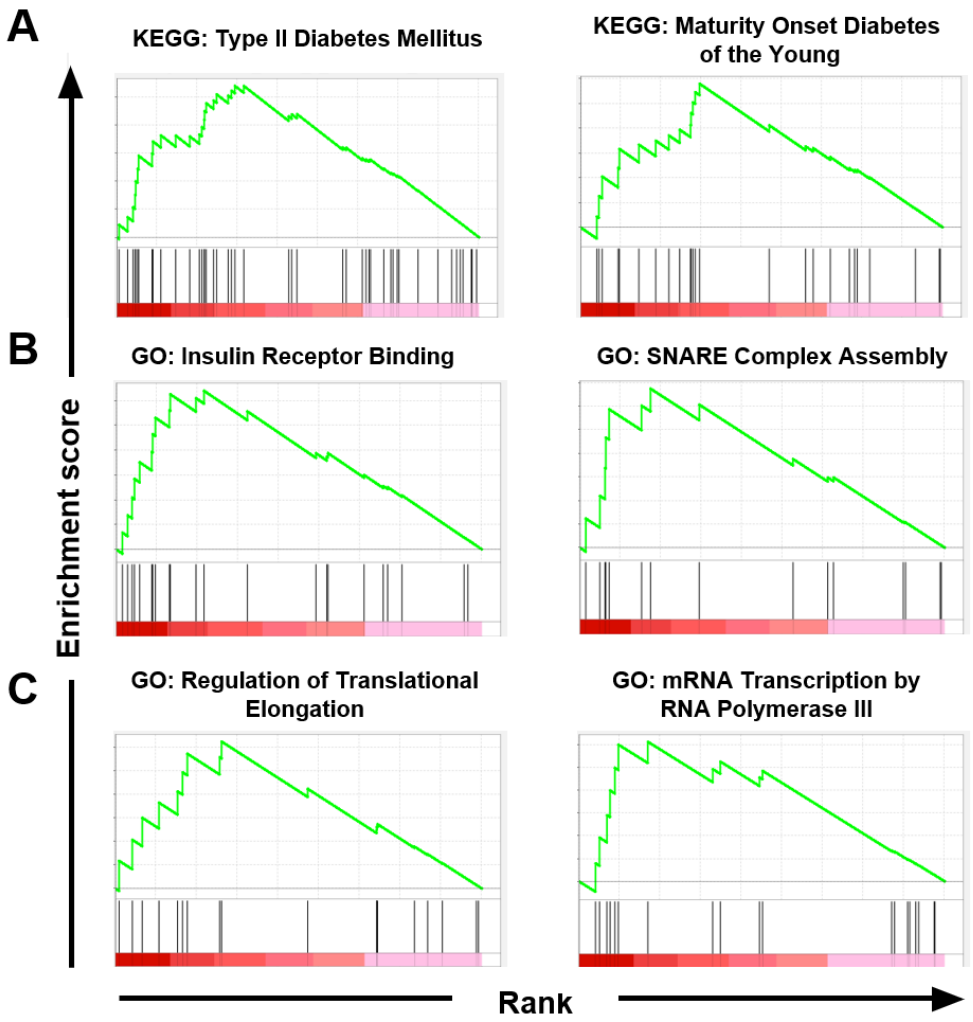


Figure 4.17 – Gene Set Enrichment Analysis. GSEA based on MAGeCK ranked genes for KEGG pathways of positive regulators **(A),** GO-Biological process for negative regulators **(B)** and positive regulators **(C).** Green profile *(top)* reflects the enrichment profile with peaks demonstrating enrichment. Black lines *(middle)* indicates genes from this selected pathway and their position within the sorted ranks of the screening hits going from top-ranked genes *(red)* to bottom-ranked genes *(pink)*.

To assess if beta cell relevant pathways were enriched in the genes classified as hits, I performed pathway analysis using the high confidence hit list as input. The most significantly enriched KEGG pathways among screening hits were 'FoxO signalling pathway' (Fold enrichment (FE)=2.75, p=0.004, FDR=0.05), 'ubiquitin mediated proteolysis' (FE=2.69, p=0.004, FDR=0.06) and 'insulin resistance' (FE=2.84, p=0.008, FDR=0.10). T2D was also among the significantly enriched pathways and demonstrated the strongest enrichment (FE=3.84, p=0.019, FDR=0.21, Figure 4.18 A). GO terms for molecular function showed the most significant enrichment for 'protein binding' (FE=1.22, $p=5.64x10^{-8}$, $FDR=8.47x10^{-7}$) and 'Poly(A) binding' (FE=1.72, $p=5.27x10^{-5}$, $FDR=7.91x10^{-4}$, Figure 4.18 B). Consistent with the previous enrichment of insulin resistance among KEGG pathways, genes mapping to 'insulin receptor binding' (*INS, IGF2, PTPN1, IRS2, DOK1, DOK4*) were highly enriched (FE=6.38, p=0.002, FDR=0.03). The strongest GO terms for biological processes focussed again on transcription and translation including 'regulation of mRNA stability' (FE=3.55, p=0.001, FDR=0.02) and 'translation' (FE=2.37, p=0.001, FDR=0.03, Figure 4.18 C). In line with previous pathways, 'protein ubiquitination involved in ubiquitin-dependent protein catabolic process' (FE=2.83, p=0.002, FDR=0.04) and 'insulin receptor signalling pathway' (FE=3.41, p=0.008, FDR=0.14) were also strongly enriched. Additional beta cell relevant pathways including 'fatty acid homeostasis' (FE=7.67, p=0.056, FDR=0.63) and 'regulation of glucose transport' (FE=4.03, p=0.076, FDR=0.74) could be found just below the p-value threshold.

Collectively, GSEA of ranked genes and pathway analysis of prioritised high confidence hits demonstrated commonly enriched pathways and networks including general protein homeostasis and specific beta cell pathways. Especially insulin receptor signalling and T2D pathways could be found throughout all pathway categories and analysis approaches. The unbiased pathway analysis demonstrated that this CRISPR screen was indeed able to

specifically identify known regulators of insulin content, increasing the confidence to also be able to identify novel regulators.

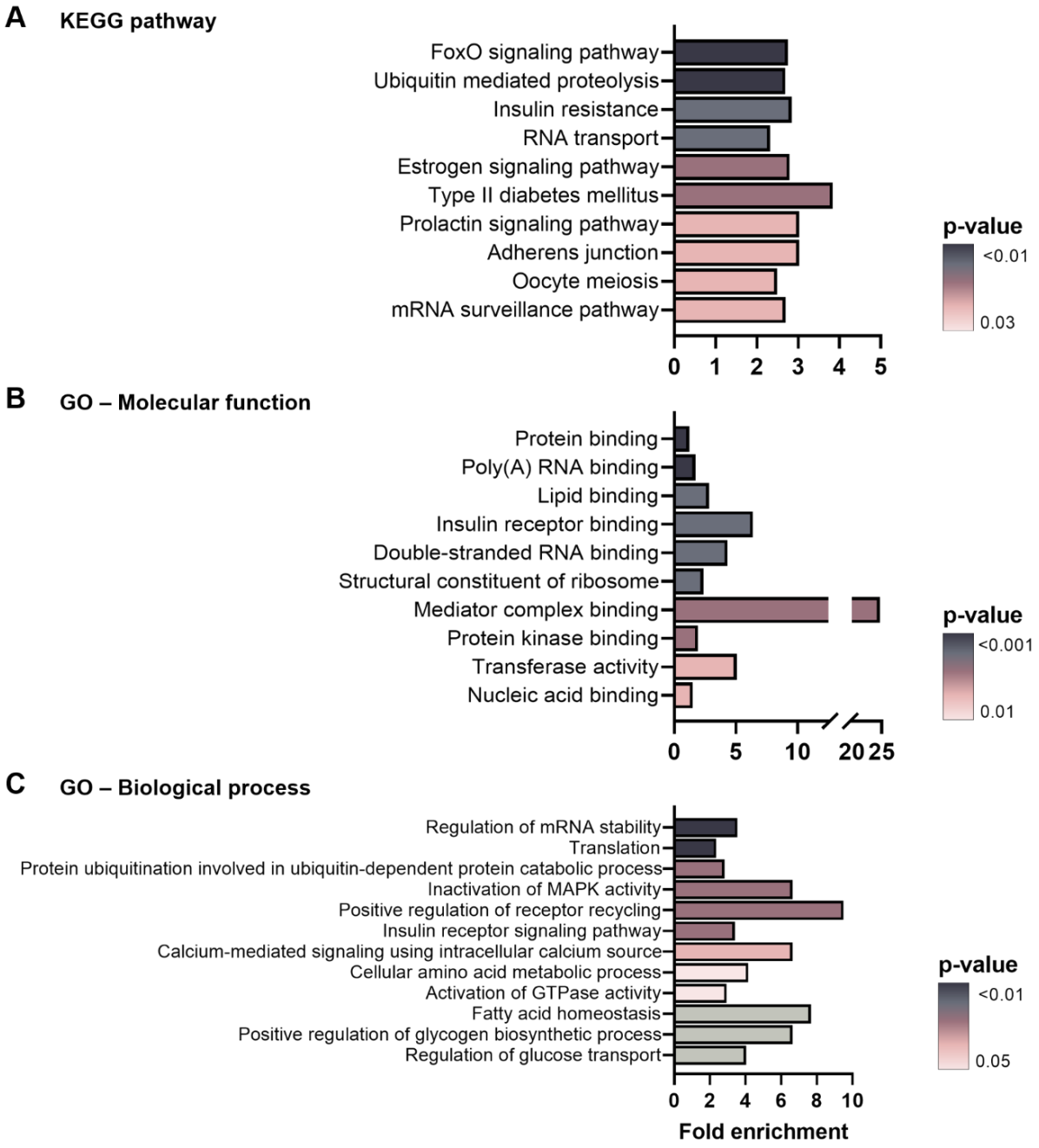


Figure 4.18 – Pathway enrichment analysis. Pathway enrichment analysis using DAVID 6.8 based on the high confidence hit list containing 580 genes and analysing KEGG pathways **(A),** GO terms for molecular function **(B)** and GO terms for biological process **(C)**. Selected pathways are shown, ranked by colour coded p-value.

4.3.5.2 Protein-protein interaction networks

In addition to pathway analysis, protein-protein associations can give an unbiased indication which protein networks are involved in the regulation of insulin content and if the screen is able to identify complete protein clusters. I performed STRING analysis on the high confidence gene hit list to map physical and functional protein association networks (Szklarczyk *et al.*, 2019). The network was significantly enriched for interactions (p=0.0004), demonstrating that the screening hits mapped to protein networks and were not a random compilation of genes. Based on the large network containing all genes, I manually assigned gene functions, identified protein clusters and replotted the clusters individually. As expected, INS took up a central node within the cluster and was widely connected to other proteins (Figure 4.19). Association partners within the network included the transcription factor NKX2.2, cellular receptors such as FFAR1 and components of the insulin signalling pathway such as DOK1 and IRS2.

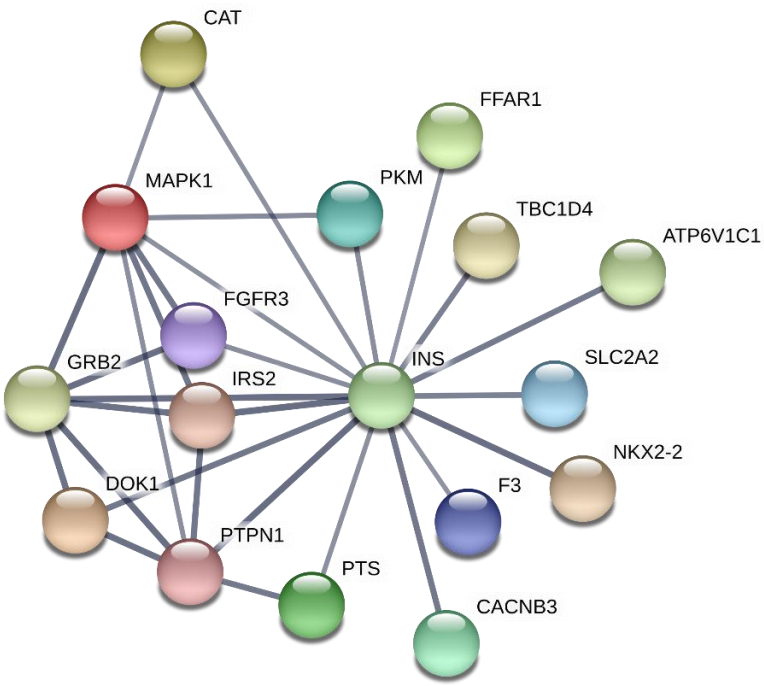


Figure 4.19 – INS protein association network. STRING pathway analysis showing protein-protein associations including physical and functional interactions for INS and other genes within the high confidence hit list. Confidence level to determine interactions was set to high confidence (0.7).

Specific proteins indicated for each of the following networks represent in most cases only a small subset of the full network which is shown in Figure 4.20. Consistent with the pathway analysis, transcription and translation associated networks made up a large proportion of all assigned networks and included transcriptional regulation (MED4, SIN3A, JARID2), nuclear import and export through the nuclear core complex (NUP107, NUP54, NUP188), mRNA splicing (ZPR1, ELAVL1, RNPS1), mRNA processing (PARN, PDCD11, RSRC1) and translational initiation (EIF3D, EIF3J, EIF3B).

The second largest cluster contained proteins related to ubiquitin mediated proteolysis including ubiquitin conjugating enzymes (UBE2J1, UBE2C, UBE2D1), ubiquitin ligases (SMURF1, SMURF2, TRIM39) and deubiquitinases (USP14, OTUB1). Ubiquitin mediated proteolysis is essential in maintaining cellular homeostasis by degrading cellular proteins that are marked for degradation by ubiquitination but has also shown to play an important role in beta cells by modulating insulin synthesis and secretion (Kitiphongspattana *et al.*, 2005; López-Avalos *et al.*, 2006; Hartley, Brumell, & Volchuk, 2009). Previously characterised proteins in beta cells within this cluster include HUWE1, which has been shown to impair insulin exocytosis in pancreas-specific KO mice and RFWD2/COP1, which has been associated with insulin granule docking defects and diabetes in beta cell specific KO mice (Wang *et al.*, 2014a; Suriben *et al.*, 2015). Closely related to ubiquitin mediated degradation was the autophagy network (CALCOCO2, TBK1, SQSTM1).

Mitochondrial function also made up a central component of all protein associations including the clusters mitochondrial translation (MRPL19, MRPS24, MRPL28), ATP synthesis (ATP5H, NDUFA1, NDUFAF4) and membrane proteins (ACO2, TOMM6, SLC25A12). Even though mitochondria are critical for cellular metabolism, many of the associated proteins were not classified as essential common genes within DepMap, hinting at a potential cell type specific role within beta cells in regulating insulin content.

Further networks within the screening hits that are known to play a role in beta cell function or insulin secretion include vesicle trafficking and exocytosis (NSF, VAMP4, SYNJ1), GPCR signalling (SSTR1, NPY, KALRN), GTPase activity (RUFY1, RAP1A, RAB19), lipid metabolism (AGPAT2, AGPAT5, DGAT1) and MAPK pathway (MAPK1, DOK4, DUSP4).

Fine-tuned insulin vesicle trafficking within the cell and subsequent fusion with the plasma membrane are essential for insulin secretion and thus also for the intracellular level of insulin content. The identification of this protein network therefore clearly highlights that the screen was able to specifically identify genes and networks important for the regulation of insulin content.

GPCR are critical to beta cell function by modulating cAMP and calcium levels within the cell. Proteins within the GPCR signalling cluster include the receptor FFAR1 (GPR40) which is well established to enhance insulin secretion upon activation and TIAM1, that has been shown to potentiate insulin secretion upon KD in a rat beta cell line (Veluthakal *et al.*, 2009; Tsuda *et al.*, 2017).

The related small GTPase are a superfamily that bind to GTP and hydrolyse it to GDP to initiate a wide range of downstream signalling processes. Within the GTPase cluster, previously studied proteins in the beta cell include ARAP1, which has initially been associated as a candidate gene mediating T2D risk at a disease locus but more recent studies have pointed to the nearby genes *STARD10* and *FCHSD2* (Kulzer *et al.*, 2014; Carrat *et al.*, 2017; Hu *et al.*, 2020). Another associated gene within the cluster, the Rab GTPase-activating *TBC1D4/AS160* is a T2D associated gene linked to severe insulin resistance through modulating glucose uptake in insulin sensitive tissue but has also been shown to affect insulin secretion upon KD in MIN6 cells and was downregulated in human islets from

diabetic donors (Bouzakri *et al.*, 2008; Sakamoto & Holman, 2008; Dash *et al.*, 2009, 2010; Moltke *et al.*, 2014).

Lipid metabolism is important for beta cell homeostasis with fatty acids supporting and enhancing glucose stimulated insulin secretion while lipotoxicity induces impaired beta cell and apoptosis (Imai *et al.*, 2020). Within this cluster, the sphingolipid phosphatase SGPP2 has been characterised previously in beta cells with KO mice demonstrating increased beta cell ER stress, highlighting its essential role for normal beta cell function (Taguchi *et al.*, 2016).

Another large cluster among screening hits was the MAPK pathway centred around MAPK1/ERK2 which is involved in a wide range of cellular processes including proliferation and transcriptional regulation. *MAPK1* itself has been shown to be regulated by glucose and KD in rodent beta cells has been linked to reduced insulin secretion (Longuet *et al.*, 2005). The MAPK pathway plays an essential downstream role in the insulin signalling pathway and the associated protein cluster contained crucial components of the insulin/IGF1 signal transduction network such as IGF2, IRS2 and PTPN1. Both, overexpression and KO of *IGF2* have been associated with reduced beta cell function and demonstrated inhibited glucose stimulated insulin secretion in beta cell specific overexpression and female KO mice, respectively (Casellas *et al.*, 2015; Modi *et al.*, 2015). KO of the insulin receptor substrate *Irs2* in pancreatic KO mice demonstrated impaired insulin secretion while upregulation of *Irs2* on the other hand increased insulin content and secretion in mice with beta cell specific overexpression (Hennige *et al.*, 2003; Cantley *et al.*, 2007). *PTPN1* acts as a negative regulator of insulin receptor signalling and has been linked to T2D risk and increased insulin secretion in KO mice (Bento *et al.*, 2004; Fernandez-Ruiz *et al.*, 2014).

The majority of genes mapping to the centrosome assembly cluster (AURKA, HAUS2, NPHP4) have not been associated with beta cell function before but could be linked to insulin secretion through regulating the cytoskeleton including microtubules and actin which are critical in mediating insulin granule transport (Henquin, Mourad, & Nenquin, 2012; Zhu *et al.*, 2015). One of the genes within the cluster, *PCNT* has been studied regarding its function in beta cells and demonstrated reduced insulin content mediated through its effect on F-actin upon silencing in MIN6 cells (Zu *et al.*, 2015).

The related protein network containing genes from the cohesion complex (SMC3, SMC1A, PDS5A) is involved in connecting sister chromatids and therefore crucial for chromosome segregation during mitosis (Peters, Tedeschi, & Schmitz, 2008). This protein network has recently been associated with beta cell function for the first time with *Nipbl* regulating *Ins2* expression and insulin secretion in MIN6 cells (Fang *et al.*, 2019).

Collectively, this CRISPR screen for insulin content in EndoC-βH1 was able to identify protein networks that are known to be crucial in beta cell function. The cluster contained previously well characterised genes but in addition also novel genes that had not been associated with insulin homeostasis within the beta cell before. Many of these genes have only been studied in mice or rodent cell lines previously, so the screen has provided a confirmation of the functional mechanisms in human beta cells. Furthermore, the screen has also identified novel and less known protein clusters containing genes that have not been linked to beta cell function previously.

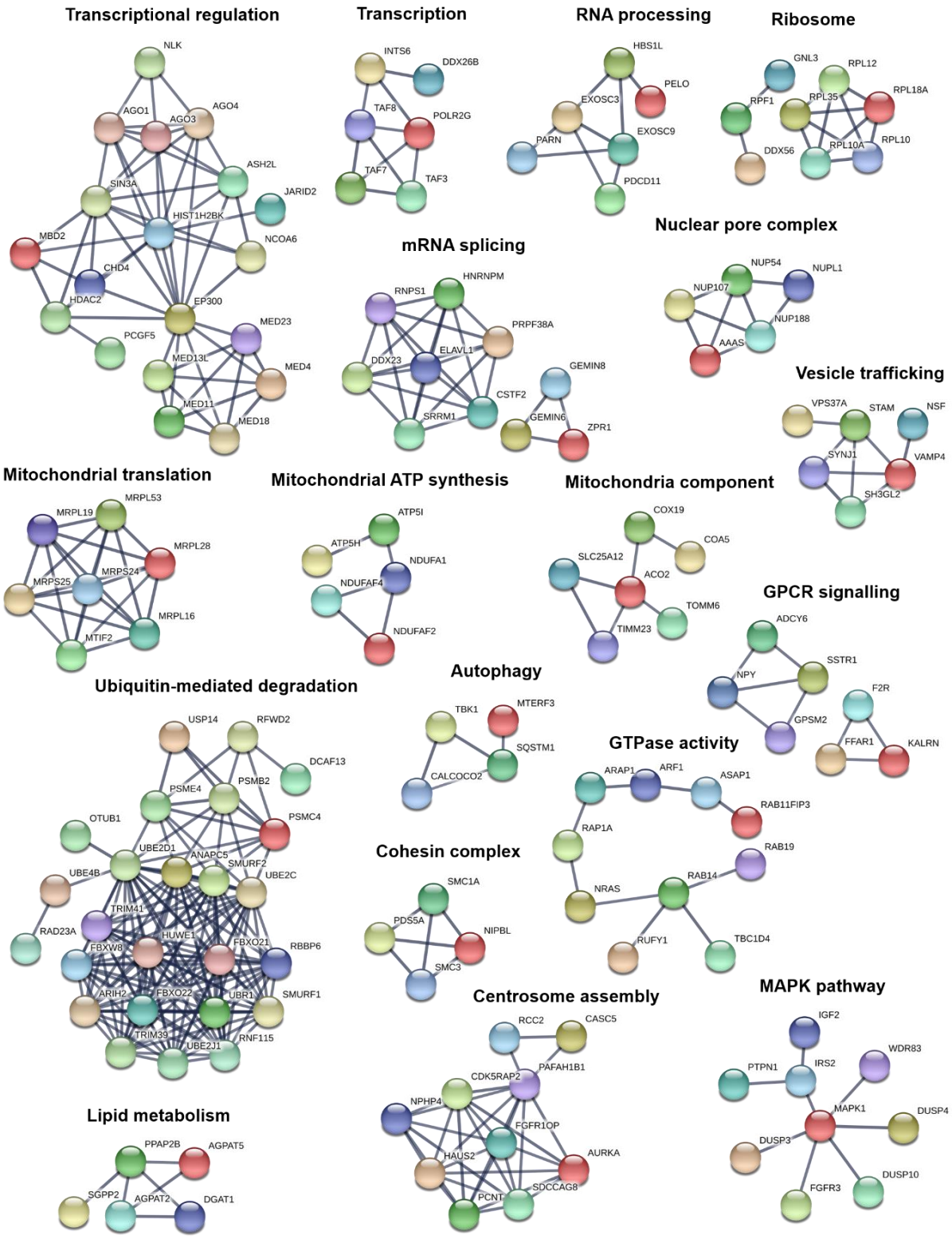


Figure 4.20 – Protein association networks of screening hits. STRING pathway analysis showing protein-protein associations including physical and functional interactions for genes from the high confidence screening hit list. Clusters were manually identified, annotated and individually plotted. Selected clusters are shown. Confidence level to determine interactions was set to highest confidence (0.9).

4.3.6 The CRISPR screen identified T2D effector genes

A primary objective of the genome-wide CRISPR screen in the human beta cell line EndoC-βH1 was to prioritise causal genes at T2D GWAS risk loci. As an initial prioritisation approach, the screening hits were integrated with a list of predicted T2D effector genes that were manually curated by Anubha Mahajan and Mark McCarthy based on combined genetic, regulatory and perturbation evidence for genes associated with T2D (Accelerating Medicines Partnership, 2020). Supporting genetic evidence would for example be a coding variant with a high posterior probability or monogenic associations, regulatory evidence includes tissue-relevant cis-eQTLs while perturbation evidence includes model system mutant phenotypes or siRNA screens. The list classifies the predicted genes into "Causal", "Strong", "Moderate", "Possible" or "Weak" based on their combined evidence and resulting likelihood to be causal for T2D with each class requiring a higher level of supporting evidence. In addition to genes associated with T2D, genes linked to glycaemic traits are also included. In total, 132 T2D GWAS genes met the criteria and are classified as predicted T2D effector genes on this list. Both, MAGeCK gene hits and additional hits from the high confidence hit list were integrated with these predicted T2D effector genes and the resulting 18 common genes were highlighted in Figure 4.21. The additional perturbation evidence from this genome-wide CRISPR screen would support the predicted role of T2D genes that rely on only a single line of evidence, have not been functionally characterised or associated with a predicted tissue of action and not been classified as causal or strong yet. In addition, the screening results further confirm and support the genes that have already been classified as causal.

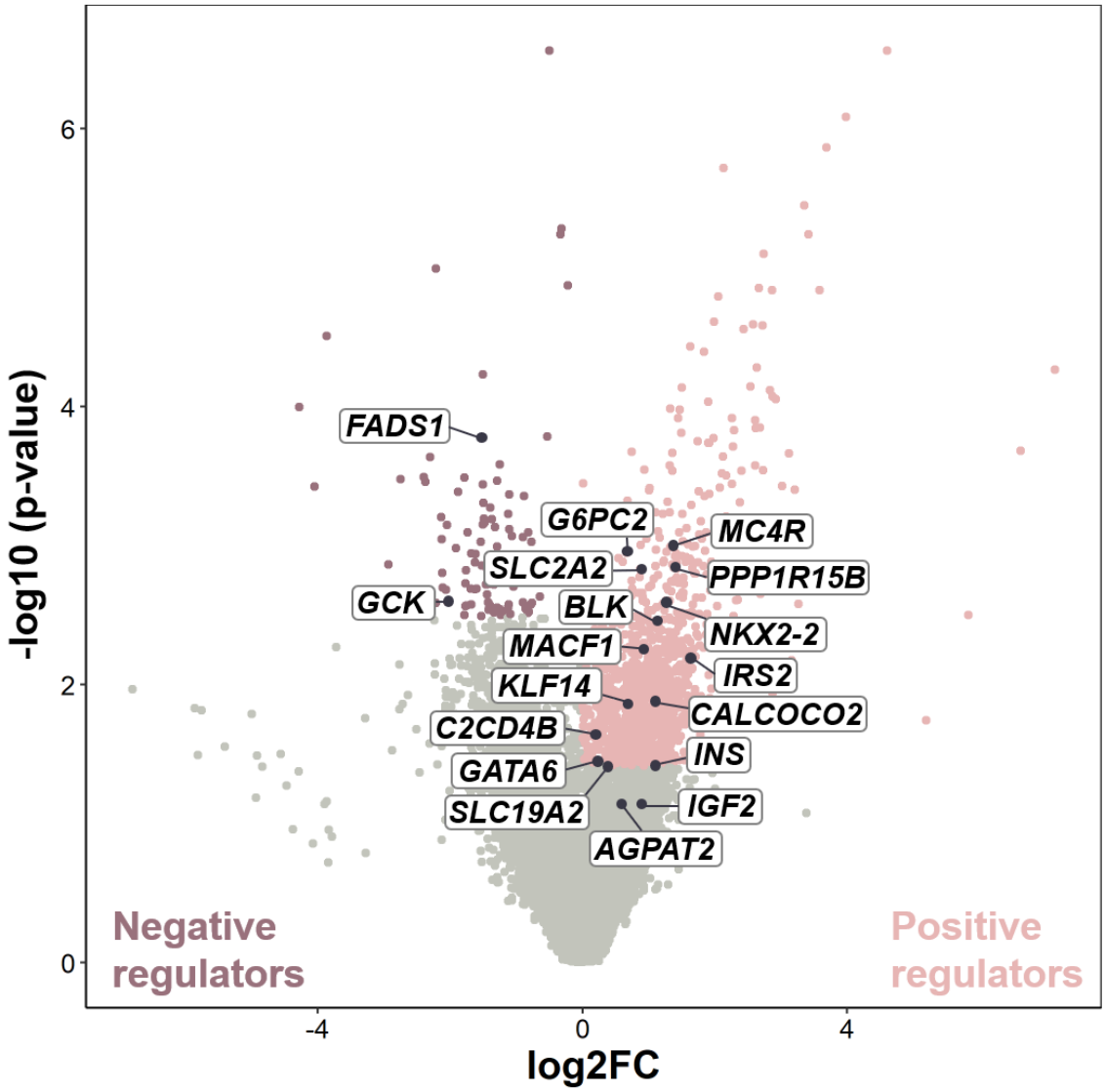


Figure 4.21 – Predicted T2D effector genes in the genome-wide CRISPR screen volcano plot. Screening hits based on MAGeCK gene-level analysis. Positive regulators *(pink)* and negative regulators *(purple)* leading to reduced and increased insulin content upon gene KO, respectively were indicated based on an FDR<0.5. Screening hits that are also part of the predicted T2D effector gene hit list that was generated by Anubha Mahajan and Mark McCarthy were highlighted (Accelerating Medicines Partnership, 2020). Detailed criteria on how the list was curated and genes were classified as predicted casual gene can be found in the associated main text. Data are from two independent genome-wide CRISPR screen replicates. FC, fold change; FDR, false discovery rate

Of the 18 genes that have been predicted to be T2D effector genes and have a phenotypic effect on insulin content in the genome-wide CRISPR screen as shown in more detail in Table 4.5, six genes *(NKX2-2, BLK, GATA6, GCK, INS, SLC19A2)* have a known role in monogenic diabetes and have already been predicted as being causal for T2D or related traits. In addition, their causal role and function in human beta cells has already been characterised previously. Five genes *(G6PC2, SLC2A2, PPP1R15B, IRS2, IGF2)* have been linked to beta cell function and their effect on insulin content has been discussed previously.

The remaining seven genes, however, consist of novel genes of interest that have not been characterised or linked to beta cell function before. *AGPAT2, KLF14, MACF1* and *MC4R* are primarily associated with T2D risk through a different tissue and mechanism of action namely lipodystrophy, adipocyte size and fat distribution, insulin action in muscle and adiposity, respectively (Agarwal *et al.*, 2002; Mahajan *et al.*, 2018b; Small *et al.*, 2018). *KLF14* has recently been shown to mediate its effect on T2D risk through modulating fat distribution and adipocyte cell size while the effects from the risk allele are female-specific and maternally imprinted (Small *et al.*, 2018). The potential effect on insulin content by these genes as shown in the genome-wide CRISPR screen could nevertheless also indicate a potential important role in beta cell function as well, albeit potentially not being the primary functional mechanism through which the associated T2D GWAS loci exert their effect on disease risk. Such a dual role has been previously demonstrated for *TCF7L2* which influences glucose homeostasis and potentially T2D risk through simultaneously modulating insulin production in the beta cell and glucose production in the liver (da Silva Xavier *et al.*, 2009; Boj *et al.*, 2012; Mitchell *et al.*, 2015). Especially the screening hits *AGPAT2* and *MACF1* demonstrated a high expression in EndoC-βH1 and primary pancreatic beta cells, indicating a functional role in human beta cells. These genes would therefore be interesting targets for follow-up studies to better understand their potential role in human beta cells.

Gene	CRISPR screen LFC (FDR)	Classification	Contributing evidence for classification
MC4R	1.32 (0.20)	Causal	Monogenic obesity (Farooqi *et al.*, 2000) Coding variant (Mahajan *et al.*, 2018a; Flannick *et al.*, 2019)
G6PC2	0.67 (0.21)	T2D related	Gene-based signal driven by multiple coding variants and independent non-coding signal for fasting glucose (Mahajan *et al.*, 2015; Ng *et al.*, 2019)
PPP1R15B	1.39 (0.22), HCL	T2D related	Juvenile-onset diabetes (Abdulkarim *et al.*, 2015)
SLC2A2	0.87 (0.22), HCL	Strong	Neonatal diabetes (Sansbury *et al.*, 2012) Glycogen storage disease (Santer *et al.*, 1997) Liver cis-eQTL (Zhou *et al.*, 2016)
FADS1	-1.53 (0.23), HCL	Possible	Islet cis-eQTL for fasting glucose (van de Bunt *et al.*, 2015)
NKX2-2	1.21 (0.26), HCL	Causal	Neonatal diabetes (Flanagan *et al.*, 2014)
BLK	1.09 (0.28)	T2D related	MODY (Borowiec *et al.*, 2009)
MACF1	0.88 (0.30)	Moderate	Coding variant (Mahajan *et al.*, 2018b) Muscle cis-eQTL (Scott *et al.*, 2016)
IRS2	1.57 (0.32)	Causal	Syndrome of severe insulin resistance (Bottomley *et al.*, 2009) Coding variant (Mahajan *et al.*, 2018a)
CALCOCO2	1.07 (0.35), HCL	Possible	Coding variant (Mahajan *et al.*, 2018b)
KLF14	0.64 (0.35)	Strong	Adipose cis- and trans-eQTL (Small *et al.*, 2011, 2018)
C2CD4B	0.16 (0.40)	Strong	Islet cis-eQTL (Varshney *et al.*, 2017) Islet chromatin conformation (Greenwald *et al.*, 2019)

GATA6	0.18 (0.47)	T2D related	Neonatal diabetes (Allen *et al.*, 2011)
GCK	-2.04 (0.49)	Causal	MODY, neonatal diabetes (Hattersley *et al.*, 1992; Njølstad *et al.*, 2001) Coding variant (Flannick *et al.*, 2019)
INS	1.06 (0.49), HCL	Causal	MODY, neonatal diabetes (Stoy *et al.*, 2007; Molven *et al.*, 2008)
SLC19A2	0.33 (0.50)	T2D related	Neonatal diabetes (Shaw-Smith *et al.*, 2012)
AGPAT2	0.54 (0.63), HCL	Moderate	Lipodystrophy (Agarwal *et al.*, 2002)
IGF2	0.85 (0.64), HCL	Causal	Coding variant (Mercader *et al.*, 2017; Flannick *et al.*, 2019)

Table 4.5 – Predicted T2D effector genes in the genome-wide CRISPR screen. Previously associated genes at T2D risk loci that have been predicted to be T2D effector genes by Anubha Mahajan and Mark McCarthy (Accelerating Medicines Partnership, 2020). 'T2D related' genes are not genetically associated with T2D but have strong associations with glycaemic traits instead. Only genetic and regulatory evidence that have been used for this classification are indicated under 'Evidence for role in T2D'. If several lines of evidence are present, the strongest evidence is indicated. HCL, High confidence screening hit list.

The other three genes of interest, *C2CD4B*, *FADS1* and *CALCOCO2* have not been classified as causal yet and little is known about their role in beta cell function.

A T2D risk variant at *C2CD4B* has been associated with islet function through an islet cis-eQTL which also overlaps an islet enhancer (Varshney *et al.*, 2017; Kycia *et al.*, 2018; Greenwald *et al.*, 2019). However, a role in insulin secretion was recently attributed to *C2CD4A* at this V*PS13C*-*C2CD4A*-*C2CD4B* locus using a beta cell specific KO mice model (Kuo *et al.*, 2019). *C2CD4B* was not assessed in parallel in this investigation and a recent study using global KO mice for *C2CD4A* and *C2CD4B* demonstrated defective insulin secretion in female *C2CD4B* KO mice with no effect observed in *C2CD4A* KO mice (Gharavy *et al.*, 2020). The functional role of *C2CD4B* in human beta cells and the mechanism through which it affects beta cell function are still unclear.

The fatty acid desaturase *FADS1* has been linked to insulin secretion through an islet cis-eQTL for fasting glucose while physiological clustering suggested a defect in early insulin secretion, as described in section 4.3.4 (Dupuis *et al.*, 2010; Ingelsson *et al.*, 2010; van de Bunt *et al.*, 2015). The T2D GWAS risk allele is associated with higher fasting glucose and increased *FADS1* expression, which is in line with the CRISPR screen that demonstrated increased insulin content upon loss of *FADS1*. The functional role of *FADS1* in human beta cells and its role in the regulation of insulin content and secretion remain to be characterised.

The autophagy receptor *CALCOCO2* has been associated with T2D risk through a coding variant with a partial role in driving the disease risk and the GWAS locus has been classified as acting through insulin secretion (Mahajan *et al.*, 2018b). However, the gene has not been studied in beta cells so far and its potential role in beta cell dysfunction remains to be explored.

Collectively, these genes that have been associated with T2D risk but their function in human beta cells has not been explored or determined yet, form compelling targets for functional follow-up studies to understand their involvement in human beta cell dysfunction and T2D risk.

4.3.7 Integration with complementary screening approaches

While this has been the first genome-wide CRISPR screen in human beta cells, previous screens as described in section 1.5 have explored insulin secretion in an siRNA screen or used a similar FACS based CRISPR screening approach for insulin content which was performed in mouse MIN6 cells (Thomsen *et al.*, 2016a; Fang *et al.*, 2019).

Thomsen et al. performed an arrayed siRNA based KD screen in EndoC-βH1 analysing 300 candidate genes from 75 T2D GWAS loci for their effect on insulin secretion and cellular proliferation (Thomsen *et al.*, 2016a). A direct comparison with this pooled CRISPR screen in EndoC-βH1 however is not possible as both screens assessed different phenotypic readouts (insulin content vs insulin secretion). Both phenotypes indicate a defect in beta cell function but are not necessarily related or directly associated. A reduction in insulin content could be associated with an increased compensatory mechanism in the actual secretion, leading to no change in total insulin secretion. If the defect affecting insulin content also affects the secretory machinery, it might also lead to impaired insulin secretion. Of the 45 genes that were identified as hits affecting beta cell function by Thomsen et al., 9 hits or 20% (*ADIPOQ, DGKQ, ELAVL4, HEYL, IGF2, INS, PLA2R1, RND3, SOCS7*) also affected insulin content in this genome-wide CRISPR screen (Figure 4.22).

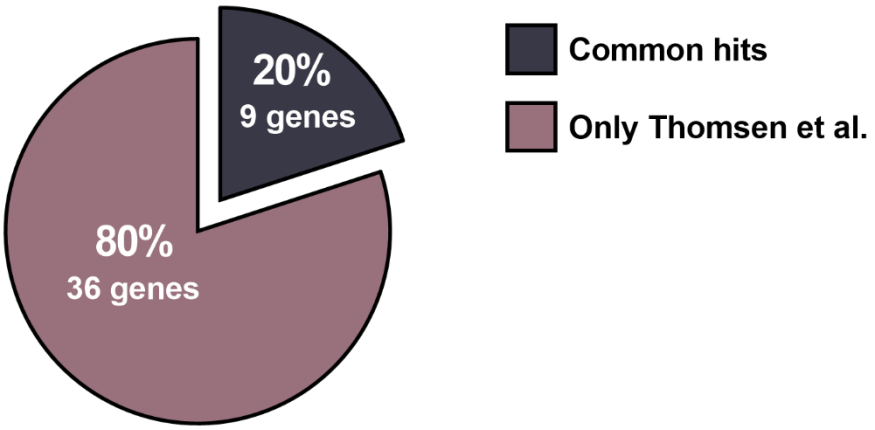


Figure 4.22 – Hit comparison with Thomsen et al. CRISPR screening hits including MAGeCK and high confidence hit genes were compared with hits from the siRNA screen in EndoC-βH1 from Thomsen et al. (Thomsen *et al.*, 2016a). Common genes are indicated as blue, hits specific to Thomsen et al. are indicated as purple.

Fang et al. performed a similar pooled LoF CRISPR screen based on a FACS readout for intracellular insulin (Fang *et al.*, 2019). The main distinction between the two CRISPR screens was that Fang et al. used the mouse beta cell line MIN6 while this screen was performed in the human beta cell line EndoC-βH1. Rodent and human beta cell however have distinct genetic, structural, transcriptional and functional characteristics, which makes it impossible to directly translate functional insights from rodent cells into mechanistic understanding in human beta cells as described in more detail in section 1.4.4. Consistent hits between my CRISPR screen and Fang et al. could therefore indicate which genes are conserved regulators of insulin content in human and mouse beta cells while also demonstrating which hits are highly reproducible between independent experimental investigations. When comparing the 580 high confidence hits from this screen with the 907 tier 1 and 2 hits from Fang et al. that they used for downstream analysis, I identified 43 common genes (Figure 4.23 A). Consistent hits in both screens included as expected *INS* but also important regulators of insulin secretion such as *NSF* and general regulators of transcription and translation such as *EIF3D* and *EIF3B*. The T2D associated gene *FADS1*

which was highlighted as a potential interesting follow-up gene in section 4.3.6, was also detected as a regulator of insulin content in both screens, emphasising its important but yet unresolved role in beta cell function. Another consistent hit between both screens was *NIPBL*, a major subunit of the cohesion complex. *NIPBL* was identified as a novel regulator of insulin content in Fang et al. and one of their target genes for functional follow-up studies that demonstrated its effect on insulin content was mediated through regualting *Ins2* expression.

Even though only around 7.4% of the screening hits from the CRISPR screen in human EndoC-βH1 were also found in the CRISPR screen in mouse MIN6, both screens identified similar protein networks as regulators of insulin content (Figure 4.23 B). More than half (55.6%) of the protein networks from Fang et al. that were identified by STRING were also established as regulatory networks in my CRISPR screen (section 4.3.5). This demonstrated that even though individual genes might not be consistent due to experimental or analytical deviations or distinct roles in human and mouse beta cells, both screens identified consistent results as measured by assessing common protein networks. Beside protein association networks that are known to be essential in beta cell function such as GPCR signalling, vesicle trafficking and exocytosis, mitochondrial ATP production and general protein transcription and translation, both screens also identified a protein cluster containing genes associated with the cohesion loading complex. This network was classified as a novel and previously unknown regulator of beta cell function and insulin content in Fang et al. The results of this CRISPR screen in EndoC-βH1 therefore independently confirm the regulatory effect of this protein network in human beta cell function.

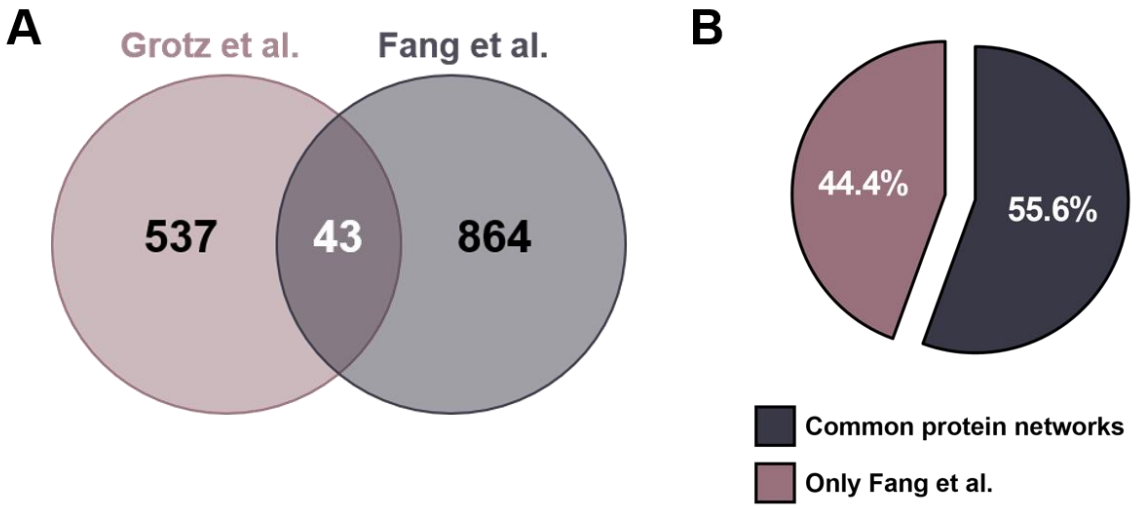


Figure 4.23 - Hit comparison with Fang et al. CRISPR screening hits **(A)** and protein networks **(B)** based on the high confidence hit genes were compared with hits from the CRISPR screen in MIN6 cells from Fang et al. (Fang *et al.*, 2019). **(A)** Venn diagram comparing hits from this CRISPR screen *(Grotz et al, purple)* compared to Fang et al *(blue)*. Numbers indicate the number of genes within each group. **(B)** Common protein networks are indicated as blue, networks specific to Fang et al. are indicated as purple.

In addition to only focussing on individual genes, an interesting line of investigation would have been to compare the variability between individual replicates within both screens to the variability of the final analysis from human and mouse screens to assess if the interspecies variation is substantially different from variation between two replicates of the same screening approach. Unfortunately, this was not possible as the required data to perform this analysis was not provided by Fang et al. (Fang *et al.*, 2019).

Collectively, this CRISPR screen in EndoC-βH1 identified many genes and protein networks that were also discovered in previous screens for beta cell function. The majority of hits were distinct to this screen and suggest a great potential to identify true and specific regulators of human beta cell function and insulin content in functional validation and follow-up studies.

4.4 Discussion

4.4.1 A successful first genome-wide CRISPR screen in human beta cells

Genome-wide pooled CRISPR screens are a powerful and most importantly an unbiased approach to identify regulators of a specific phenotype which can also be exploited to prioritise causal genes at GWAS disease loci. Here, I performed a genome-wide FACS-based pooled LoF CRISPR screen in the human beta cell line EndoC-βH1 to identify regulators of insulin content and beta cell function. This is the first genome-wide CRISPR screen studying beta cell function that has been performed in human beta cells. Previous investigations focused on either rodent beta cell lines which are easier to culture and grow than EndoC-βH1 or only on a subset of genes while using an siRNA silencing based method (Thomsen *et al.*, 2016a; Fang *et al.*, 2019). This screen identified 580 hits which contained both, genes known to play a role in beta cell function and novel regulators of insulin content.

Before the actual genome-wide screen was performed, I spent the first 2.5 years of my PhD developing and optimising the pipeline. The key challenges of this project were the FACS-based readout of the CRISPR screen, which had to be highly optimised and tested for its sensitivity and specificity and the use of the human beta cell line EndoC-βH1, which had not been used for CRISPR editing before and demonstrated extremely challenging growth and culture characteristics. By initially establishing a robust CRISPR editing pipeline in EndoC-βH1 (Chapter 3) followed by in-depth validation and optimisation of antibodies and staining conditions of the FACS readout, all steps and processes crucial to the success of the genome-wide CRISPR screen were previously tested and optimised.

Two independent CRISPR screen replicates were performed separately with EndoC-βH1 at different passages and independent CRISPR library transductions, puromycin selections and FACS sorting. Previous screens have often relied on a single replicate, the two replicate

strategy was essential for this CRISPR screen in determining true regulators of beta cell function due to large variation between the insulin low replicates (Park *et al.*, 2017; Richardson *et al.*, 2018). This is likely a consequence of variability from the EndoC-βH1 cells themselves which can change their characteristics depending on the passage. Their ability to secrete insulin was verified at both time points, the variation is therefore not a result of them losing their beta cell properties. However, the cells could have potentially been more sensitive to the puromycin selection depending on the passage which might have induced a higher cell death, overall worse culturing conditions for the surviving cells due to the pooled setup and therefore affected their level of insulin content. As the EndoC-βH1 are extremely sensitive to different external stimuli or culturing, it is not possible to exactly understand what might have introduced the variation within the low insulin content sample. All possible preventive steps were taken prior to the screen to reduce potential variability between the replicates such as cellular handling by the same person only and reagents originating from the same lot. The two separate FACS staining processes did not introduce significant variation as the high insulin content samples from both replicates were highly correlated. The strategy to perform at least two independent replicates of a genome-wide CRISPR screen is therefore crucial to a successful hit identification in sensitive cell lines such as EndoC-βH1.

The technical performance of this CRISPR screen including the quality of the sequencing reads and the CRISPR library representation was immaculate. In addition to the excellent technical performance, the screen was also able to provide important biological insights into the regulation of insulin content in human beta cells. The specificity of the screen to identify valid regulators of insulin content was verified by identifying many genes, pathways and protein networks known to play a role in insulin homeostasis in the beta cell such as of course *INS* itself, genes involved in monogenic diabetes or well characterised genes that are

involved in insulin transcription and secretion. In addition, this screen has provided a large number of potential novel regulators of insulin which were highlighted through integration with T2D GWAS loci and as part of protein association networks. Especially the ubiquitin-proteasome system and autophagy, transcription and translation and insulin receptor signalling pathways were highly enriched in pathway analyses and formed large protein network clusters among hits.

4.4.2 Limitations of the genome-wide CRISPR screen

This genome-wide pooled CRISPR screen has identified many novel regulators of insulin content and potential causal genes for T2D, but the approach is also associated with a number of limitations.

Among the screening hits were many well characterised regulators of insulin content such as *INS, NKX2.2* or *IRS2*. On the other hand, some genes with an established effect on insulin content such as *PAM* were not classified as hits and indicate that the screen also demonstrated a certain rate of false negatives (Thomsen *et al.*, 2018). Our group has previously shown that KD of *PAM* in EndoC-βH1 induced a reduction in insulin content by around 20%. In addition, this was also confirmed in my small-scale proof of concept CRISPR screen by demonstrating *PAM* sgRNA enrichment in the low insulin content sample. The fact that *PAM* was not discovered as a screening hit in the genome-wide screen can therefore not be attributed to phenotypic differences between siRNA and CRISPR. It also highlights that this pooled CRISPR screen should primarily be regarded as a discovery tool of genes with a certain effect on insulin content. If a gene was not classified as hit, that does not necessarily mean that it does not affect insulin content. Besides false negatives, the screen was also based on a LoF CRISPR approach and tested insulin content only under basal glucose conditions. The effects of gene overexpression and glucose or other external

stimulation were therefore not tested and it cannot be excluded that certain genes only affect insulin content under those conditions. In addition, albeit EndoC-βH1 are the best available beta cell model to date with an established functional profile similar to primary cells, it has to be taken into account that not all gene functions might be directly translatable into human physiology.

As demonstrated by pathway enrichment and protein clusters, the screening hits could be divided into specific regulators of intracellular insulin and genes indirectly affecting total insulin content through regulating general transcription and translation. Even though it is straightforward to distinguish general from specific hits using databases such as DepMap, a potential addition for future similar CRISPR screens would be to include a simultaneous FACS staining of a control gene. If the level of the control or housekeeping gene is affected, the regulator is not a specific regulator of only insulin. These genes could then be excluded through a specific gating strategy. This strategy was initially part of our screening pipeline, however as the first CRISPR screen in EndoC-βH1 which was already associated with many unknown parameters, it was decided not to further pursue and optimise this additional staining pipeline.

Potential limitations of this screen also include drawbacks of the pooled setup itself and the incubation of all cells within the same shared media which could result in the activation of unintended cellular pathways such as insulin receptor signalling in other cells. Insulin receptor signalling is primarily thought of as a pathway within insulin responsive tissue but it has also been shown to affect insulin expression, content and secretion in human beta cells (Xu & Rothenberg, 1998; Hennige *et al.*, 2003; Bento *et al.*, 2004; Cantley *et al.*, 2007; Wang, Gu, & Chen, 2018). It is therefore possible that in addition to being a strongly enriched pathway among hits with a genuine effect on insulin content within those KO cells, the pathway might have caused false positives from changed levels of insulin within the

pooled media. Other KO cells within the same shared flask might have strongly affected insulin secretion, leading to drastically increased or reduced levels of insulin within the flask. This could induce an insulin receptor mediated signalling response in other cells leading to changes in insulin content that were not a result of their own KO. Nonetheless, it can be assumed that many KO cells within a flask do not have an insulin secretion phenotype and due to the presence of both, cells with increased and decreased insulin secretion and the large volume of the media, it is unlikely that drastic changes in the level of insulin in the supernatant have occurred, especially considering it has to be consistently across flasks to affect most sgRNAs of a single gene.

Despite the limitations which are associated with every experimental approach, this CRISPR screen has successfully replicated the effect of many well-known regulators of beta cell function and generated an extremely powerful genome-wide perturbation dataset.

4.4.3 Genome-wide CRISPR screen as basis for future studies

This chapter has focused on establishing that the performed genome-wide CRISPR screen has passed all technical quality controls and the generated screening hits can provide meaningful biological insights as shown by pathway enrichment analyses, protein networks and individual highlighted genes. The generation of the screening data, however, only mark the beginning of future studies focusing on this dataset. I already manually curated interesting genes for follow-up studies based on protein networks and predicted T2D effector genes. This manual curation is going to be expanded by bioinformatic analyses integrating the screening hits with all potential causal genes at T2D GWAS loci instead of only focussing on a certain subset. This analysis exceeded the scope of my thesis and PhD, but it is currently being approached by a collaborator at Stanford University. Compared to only assessing specific genes, this approach will provide an unbiased strategy including all genes

to prioritise causal genes at T2D GWAS loci. Further large-scale datasets can also be integrated to contribute another layer of evidence.

Independent of manual or high-throughput integration with genetic data or by focusing only on the strongest hits, the main priority post CRISPR screening is to prioritise genes for follow-up studies and to independently validate the observed hits. These validation and follow-up studies will be crucial in confirming the observed effects, excluding false positives and connecting genes to their functional mechanism in beta cells and will demonstrate the long-term value of this pooled CRISPR screen by generating a large number of hypothesis for further investigations.

5

Functional Follow-Up of CRISPR Screen Hits

5.1 Introduction

5.1.1 Prioritising causal genes at T2D GWAS loci

GWAS aimed at identifying genes mediating T2D risk have so far identified more than 400 distinct association signals (Scott *et al.*, 2017; Mahajan *et al.*, 2018b, 2018a; Spracklen *et al.*, 2020; Vujkovic *et al.*, 2020). However, translating these association signals at disease loci into causal genes and functional understanding has been slow and often only been focussed on single candidate effector transcripts. Large-scale efforts such as the previously described genome-wide CRISPR screen in Chapter 4 can be integrated with genetic evidence and used as a powerful tool to prioritise genes at T2D loci for functional in-depth studies. This high-throughput pooled CRISPR screen investigated candidate causal genes experimentally in a genome-wide and unbiased manner for their involvement in the regulation of insulin content in the human beta cell line EndoC-βH1. As described in the previous chapter, the genome-wide pooled CRISPR screen identified 580 gene hits with a significant effect on insulin content. An enrichment for genes involved in T2D pathogenesis and key beta cell pathways indicated a strong functional significance of hit genes for beta cell function and a great potential to identify previously unknown causal genes at T2D loci and novel regulators of beta cell function within those hit genes. To be certain that the CRISPR screening hits are genuine regulators of insulin content and not a technical artifact leading to false positives, it is crucial to independently validate the screening hits (Doench, 2018). These validation studies often focus on many hit genes by performing smaller scale secondary screens with focused libraries or alternatively conduct in-depth functional studies on a small number of promising novel genes (Dejesus *et al.*, 2016; Cortez *et al.*, 2020; Li *et al.*, 2020).

5.1.2 Experimental outline

In this chapter, I aim to prioritise candidate causal genes at T2D GWAS loci for functional validation and follow-up studies based on the results from the genome-wide CRISPR screen in EndoC-βH1 in Chapter 4. Successful validation and replication of the observed effects on insulin content would generally confirm the accuracy and correct identification of positive hits in the genome-wide screen while in-depth functional studies of specific genes can generate functional insights into a gene's role in human beta cells. Based on integration with genetic data from T2D GWAS, I prioritised the potential causal genes *CALCOCO2, KCNK17, PLCB3* and *QSER1* and performed functional validation and follow-up studies using an siRNA KD approach. I validated the effects on insulin content and in addition to the readout of the screen, I further characterised their function in human beta cells by assessing cell viability, insulin secretion and expression. As an unknown regulator of beta cell function, I further expanded the studies into *CALCOCO2* by measuring its expression and localisation in primary human islets and exocrine pancreatic cells. I also investigated its potential role as causal gene at the *TTLL6* T2D locus by comparing the sgRNA distribution in the screen for all genes across the locus. To compare my experimental results of *CALCOCO2* KO and KD in human beta cells to other publicly available datasets, I integrated several other lines of evidence such as single-cell transcriptomes in T2D and healthy pancreatic cells (Lawlor *et al.*, 2017).

5.2 Material and Methods

5.2.1 RNA interference

mRNA expression in EndoC-βH1 was transiently silenced using siRNAs (*CALCOCO2* #L-010637-0, *PLCB3* #L-008485-00, *QSER1* #L-014444-00, *KCNK17* #L-006260-00, non-targeted (NT) control pool #D-001810-10, Dharmacon). The specific experimental conditions are described in more details in section 2.3. Briefly, the cells were incubated with siRNAs for 72 h and silencing efficiency was confirmed on both the mRNA level by RT-qPCR and protein level by western blot if specific antibodies were available.

5.2.2 RNA Sequencing

Normalized expression levels of target genes in EndoC-βH1 were assessed based on previously published RNA-Sequencing data (Thomsen *et al.*, 2018). RNA was extracted from EndoC-βH1 using TRIzol as described previously in section 2.4.1. The quality of the RNA sample was assessed using the 2100 Bioanalyzer and only samples with an RNA integrity number higher than 8 were sent for sequencing. Sequencing was performed on a HiSeq2000 at the Oxford Genomics Centre (Wellcome Centre for Human Genetics, University of Oxford). The resulting sequencing reads were aligned to the human genome reference GRCh37.p13 using TopHat2 and gene counts were quantified using RNA-SeQC. This extraction and analysis was performed by Dr Soren Thomsen and the results have been published previously as part of an independent manuscript (Thomsen *et al.*, 2016a). The sequencing data is deposited at the European Nucleotide Archive (ENA; http://www.ebi.ac.uk/ena) under accession number PRJEB15283.

5.2.3 Long read RNA Sequencing

RNA was extracted from two independent samples of primary islets (ISL-021 and ISL-094) by Dr Carla Burrows. The islets were previously isolated at the Islet Isolation Facility in OCDEM. Sample quality was determined as described in section 5.2.2 and sent for sequencing at Pacific Bioscience (PacBio). Sequencing reads were analysed using the PacBio IsoSeq pipeline and resulting transcripts were used as a reference to quantify transcripts from islet RNA-Seq data. The analysis was performed by Dr Anne Ndungu (McCarthy group, Wellcome Centre for Human Genetics, University of Oxford).

5.2.4 Gene expression analysis

To assess differences in gene expression, RNA was extracted and converted into cDNA using the GoScript Reverse Transcriptase System (Promega), which is described in section 2.4. RT-qPCR was performed using TaqMan Gene Expression Assays (Applied Biosystems, Table 3.2) and expression levels were analysed based on the ΔΔCt method with normalisation to the housekeeping genes *TBP*, *PPIA* and *GAPDH*.

TaqMan Probe	Assay details	Target region
INS	Hs00355773_m1	Exon 1-2
GAPDH	Hs02786624_g1	Exon 8
TBP	Hs00427620_m1	Exon 3-4
PPIA	Hs01634221_s1	Exon 1
KCNK17	Hs01572770_m1	Exon 4-5
CALCOCO2	Hs00977443_m1	Exon 9-10
QSER1	Hs00227078_m1	Exon 9-10
PLCB3	Hs01100294_m1	Exon 10-11
KCNK16	Hs01000699_m1	Exon 3-4

Table 5.1 – TaqMan gene expression assays

5.2.5 Protein expression analysis

To assess different protein levels between samples, EndoC-βH1 cells were lysed in RIPA buffer and extracted protein was size-separated and analysed using SDS-PAGE, which is described in more detail in section 2.5. Following size-separation, the protein was detected by western blotting through subsequent incubation with primary and secondary antibodies in 3% BSA (Figure 5.1, Table 3.3). The protein signal was quantified and normalised to its respective loading control (GAPDH or β-Tubulin) and non-targeting control sample.

Antibody	Dilution	Species	Manufacturer details
β-Tubulin	1:2000	Mouse monoclonal	Santa Cruz, sc-365791
GAPDH	1:10 000	Rabbit monoclonal	Abcam, ab37168
INS	1:1000	Mouse monoclonal	Santa Cruz, sc-393887
C-peptide	1:1000	Mouse monoclonal	ExBio, 11-247-C100
CALCOCO2	1:1000	Mouse monoclonal	Santa Cruz, sc-376540
TALK2	1:1000	Mouse monoclonal	Santa Cruz, sc-390435
Anti-mouse IgG HRP	1:2500	Rabbit polyclonal	Thermo Fisher, 31460
Anti-rabbit IgG HRP	1:2500	Goat polyclonal	Thermo Fisher, 31450

Table 5.2 – Western Blot antibodies and dilutions.

5.2.6 Immunofluorescence staining

Cryo-embedded human pancreas and islet sections had been prepared previously by Dr Romina Jimena Bevacqua at Stanford University and were stored at -20°C until immunostaining. The slides were thawed at 42°C for 10 min and dipped into distilled water until no residual embedding compound was left. After washing in PBS (Sigma-Aldrich) three times for 5 min, the slides were boiled for 20 min in a rice cooker in antigen unmasking solution (Dako). After cooling down to room temperature, the samples were washed again

three times for 5 min in PBS. The tissue sections were individually circled using a PAP pen (Sigma-Aldrich) to form a hydrophobic barrier and permeabilised by adding 1% PBS-Triton X-100 (Sigma-Aldrich) to each circle for 10 min at RT. The sections were blocked in universal blocking buffer (UBS) containing 1% BSA (Roche), 0.2% non-fat milk, 0.5% Triton-X 100, 1% DMSO in PBS (all Sigma-Aldrich) for at least 1 h at RT. Primary antibodies (Table 5.3) were diluted together in UBS, each tissue circle was sealed with parafilm and the sections were incubated at 4°C overnight. The sections were washed four times for 10 min in PBS and incubated in secondary antibody diluted in UBS for 1 h at RT in the dark. The sections were washed four times for 10 min in PBS and mounted in VectaShield mounting media (Vector Laboratories), covered with a glass coverslip (VWR) and sealed using nail polish. The slides were stored at 4°C and imaged on a Zeiss Axio imager A1 (Zeiss) microscope with a 20x and 40x objective. Images were collected with different laser settings that were optimized for each sample and channel. Files were exported as Zeiss Vision Images and processed in ImageJ 1.52 (NIH).

Antibody	Dilution	Species	Manufacturer
CALCOCO2	1:100	Mouse monoclonal	Santa Cruz, sc-376540
INS	1:300	Guinea-pig polyclonal	Dako, A0564
Proglucagon	1:100	Rabbit monoclonal	Cell Signalling, #8233
Anti-mouse IgG AF555	1:300	Donkey	Invitrogen, #A-31570
Anti-guinea-pig IgG AF647	1:300	Donkey	Abcam, ab150187
Anti-rabbit IgG AF488	1:300	Donkey	Invitrogen, #A-21206

Table 5.3 – Immunofluorescence antibodies and dilutions

5.2.7 Insulin secretion assays

Insulin secretion and content in EndoC-βH1 were assessed in 96 well plates and measured using the AlphaLISA Detection Kit (Perkin Elmer) as described in section 2.6.2 (Figure 5.1). Insulin measured in the cellular supernatant was displayed as percentage of content or after normalisation to cell count to account for plating and viability differences. Cell count was assessed using the CyQUANT Direct Cell Proliferation Assay (Thermo Fisher Scientific) as described in section 2.6.1.

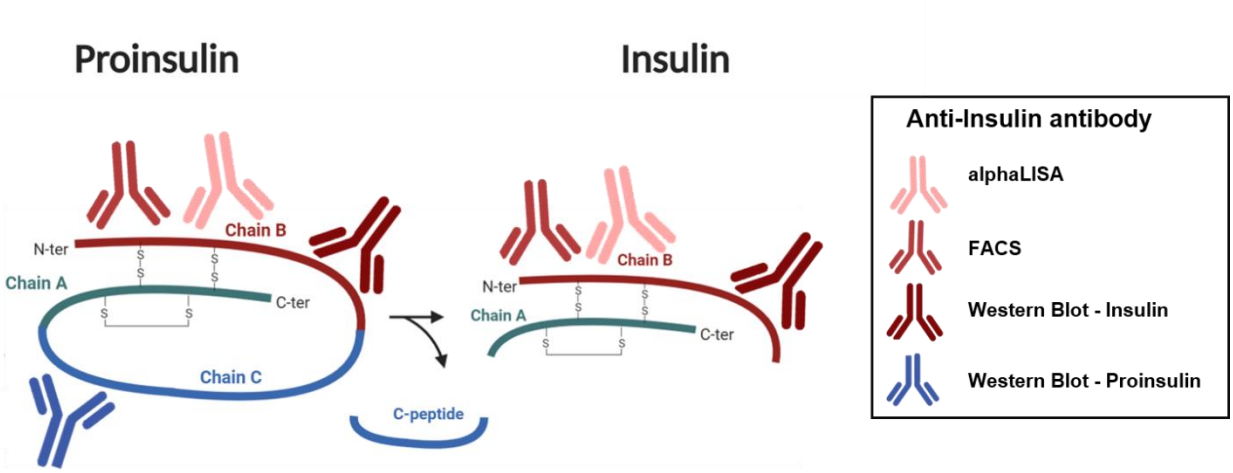


Figure 5.1 – Insulin antibodies used in various detection methods. Antibodies used for alphaLISA, FACS and Western Blot detection shown in red target the insulin B-chain and therefore detect insulin precursors and mature insulin. The antibody shown in blue, targets specifically the C-peptide and does therefore only detect proinsulin in Western Blot as C-peptide alone is too small for Western Blot based detection.

5.2.8 shRNA plasmid amplification

Knockdown in primary human islets is going to be induced through stable expression of SMARTVECTOR short-hairpin RNAs (shRNAs, Dharmacon, Table 5.4). shRNA constructs for genes of interest were selected to target all expressed isoforms. Plasmids were tagged with TurboGFP and shRNAs were expressed from a human CMV promoter. To amplify the shRNA plasmids, the obtained glycerol stocks were scraped with a pipette tip

and used to inoculate 2 ml LB medium containing 100 µg/ml ampicillin. The tubes were placed on a shaker at 220 rpm and 37°C for 16 h. After incubation, DNA was extracted using the Zyppy Plasmid Miniprep kit (Zymo) by following manufacturer's instructions. The plasmid DNA was stored at -20°C until viral production.

Target gene	Sequence (5' → 3')	Dharmacon catalogue number
CALCOCO2	TTCTGATTTACCCATATTC	V3SH11240-226316394
	GTCTAAAACAATCATTTGG	V3SH11240-229421198
	ATTAAAATCCAGACCCATG	V3SH11240-225577390
KCNK17	ACACCCGAAAGTCACATCC	V3SH11240-228518648
	AGTAGAAGCCCTCTGTGTA	V3SH11240-228716515
	AGTGGTTTACTCCCTGCTG	V3SH11240-230128256
RREB1	AAGGGCGGAAGACTGATCT	V3SH11240-228732289
	TTCATCTCAGGCTGGCCCG	V3SH11240-225206372
	GAGAAGTCCGTGAATCCTA	V3SH11240-228985761
FADS1	TAGGCATCTAGCCAGAGCT	V3SH11240-227897456
	TAATCATCCAGGCCAAGTC	V3SH11240-229631309
	TTCACCAACTACCTGCATG	V3SH11240-227466114

Table 5.4 – shRNA specifications. Three shRNA constructs from Dharmacon were chosen per prioritised target gene for future knockdown experiments in primary human islets. Each plasmid construct was amplified separately.

5.2.9 Statistical analysis

Data from siRNA silencing experiments were normalised to their respective non-targeting control siRNA and statistical analyses were performed in Prism 8.1 (GraphPad Software). Protein quantifications from western blot experiments were normalised to their corresponding control to account for technical variation between separate blots. Statistical

significance was subsequently assessed in these samples using a one-sample Student's t-test and the non-targeting control is indicated in the plot as a dotted line at 100%. Two groups such as in gene expression analysis were compared using two-sample Student's t-test and several groups such as in insulin secretion experiments were compared using ANOVA followed by Sidak's multiple comparison test. All fold changes were log-transformed for statistical analysis but are plotted as percentage of control.

5.3 Results

5.3.1 Prioritisation of genes for follow-up studies

To follow-up on the genome-wide CRISPR screen, I focused on several genes in detailed functional studies to primarily validate the observed effects on insulin content from the genome-wide CRISPR screen and once confirmed, to also further understand the underlying functional mechanism and the gene's role in beta cell function. In addition to discovering novel regulators of beta cell function, a key objective of the genome-wide CRISPR screen was to prioritise causal genes at T2D loci and to associate potential candidate causal genes with a cellular mechanism, tissue of action and direction of effect. I therefore focused on screening hits where there is a high degree of confidence that they are effector transcripts at T2D GWAS loci but uncertainty over the biological mechanisms as described in more detail in section 4.3.6. The following genes were chosen for functional studies: calcium-binding and coiled-coil domain-containing protein 2 (*CALCOCO2),* potassium two pore domain channel subfamily K member 17 *(KCNK17), g*lutamine and serine rich 1 *(QSER1)* and 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase beta-3 *(PLCB3).*

In the genome-wide screen, KO of *CALCOCO2* (log2FC (LFC)=1.08, p=0.013, FDR=0.346) induced a reduction in insulin content, corresponding to an enrichment of *CALCOCO2* sgRNA in the low insulin content sample as shown in Figure 5.2. *PLCB3* (LFC= -0.42, p=0.375, FDR=0.998) and *QSER1* (LFC= -0.95, p=0.225, FDR=0.953) on the other hand did not show a consistent sgRNA depletion or enrichment and represent control genes without an effect on insulin content. Due to its strong and consistent effect across sgRNAs in both replicates, *CALCOCO2* was part of the high confidence gene hit list in the genome-wide screen (Figure 5.2 A, D). *KCNK17* (LFC=-0.84, p=0.042, FDR=0.738) caused a small increase in insulin content in the screen but did not meet the stringent significance

threshold to be classified as hit (Figure 5.2 A, E). However, I had already prioritised *KCNK17* and initiated the functional follow-up studies before the final analysis had been completed. Even though not meeting the significance threshold, the gene demonstrated a strong tendency in the final analysis with all sgRNAs showing a consistent effect towards increased insulin and was a T2D candidacy gene, it was therefore an interesting gene for in-depth studies, nonetheless. These genes chosen for follow-up studies thus include a positive and a possible negative regulator of insulin content and two genes without a detected effect to be able to confirm the sensitivity of the screen for all directions of effect.

As highlighted in section 4.3.6, *CALCOCO2* was recently implicated in T2D risk through an exome array based genotyping study in diverse ancestries and demonstrated a high posterior probability of association (PPA) for a single coding variant (rs10278, p.Pro347Ala) at the *TTLL6* locus, consistent with a partial role of this coding variant in driving the association signal (Mahajan *et al.*, 2018b). Albeit the gene's role in beta cell function and T2D pathogenesis has not been explored yet, physiological clustering has strongly indicated the pancreatic beta cell as the tissue of action as the locus exerts its primary effect on T2D risk through insulin secretion (Mahajan *et al.*, 2018b). *CALCOCO2* itself classified as 'possible' causal gene on the list of predicted T2D effector genes, as it has no functional evidence yet connecting the gene to a role in a relevant tissue (Accelerating Medicines Partnership, 2020).

A coding T2D GWAS variant within *KCNK16* (rs1535500) has shown to be associated with the neighbouring gene *KCNK17* through an islet cis-eQTL in two independent studies (Cho *et al.*, 2012; Fadista *et al.*, 2014; Varshney *et al.*, 2017). Based on this regulatory evidence, *KCNK17* has been classified as having a 'moderate' likelihood to be the causal gene at this locus (Accelerating Medicines Partnership, 2020). The causality at both loci

(*CALCOCO2/TTLL6* and *KCNK17/KCNK16*), however, has not been resolved yet and no functional mechanism in the human beta cell has been attributed to these genes.

QSER1 and *PLCB3* both contain coding variants consistent with a causal role but their role in T2D remains to be determined (Mahajan *et al.*, 2018b, 2018a). Both genes were classified as 'causal' on the list of predicted T2D effector genes (Accelerating Medicines Partnership, 2020). Consistent with the lack of an effect in the genome-wide screen, *PLCB3* has been previously shown to play a role in the amplification pathway in the beta cell while it has also been classified as having an effect on insulin action in a multi-trait physiological clustering approach (Thore *et al.*, 2005; Mahajan *et al.*, 2018b). *QSER1*'s general function or role within the beta cell has not been determined yet.

The genes were assessed using an siRNA based silencing approach in EndoC-βH1 as a complementary method to the previous CRISPR KO. If the insulin phenotype can be validated using this orthogonal method, it is certain that the phenotype is caused by the protein reduction itself and is not specific to the LoF approach. In addition, the main advantage of siRNA-based studies in EndoC-βH1 compared to CRISPR is their simplicity and rapidness. If the siRNA KD approach shows a different effect than the CRISPR approach, it could be a consequence of using two separate methods as described in Chapter 3. In the case of this scenario, further experiments using CRISPR should be performed to assess if the effect on insulin content can only be achieved using a permeant and complete protein reduction. In addition to the insulin content screen readout, I also assessed insulin secretion as the primary phenotype in the beta cell for all genes in comparison to non-targeting control siRNAs.

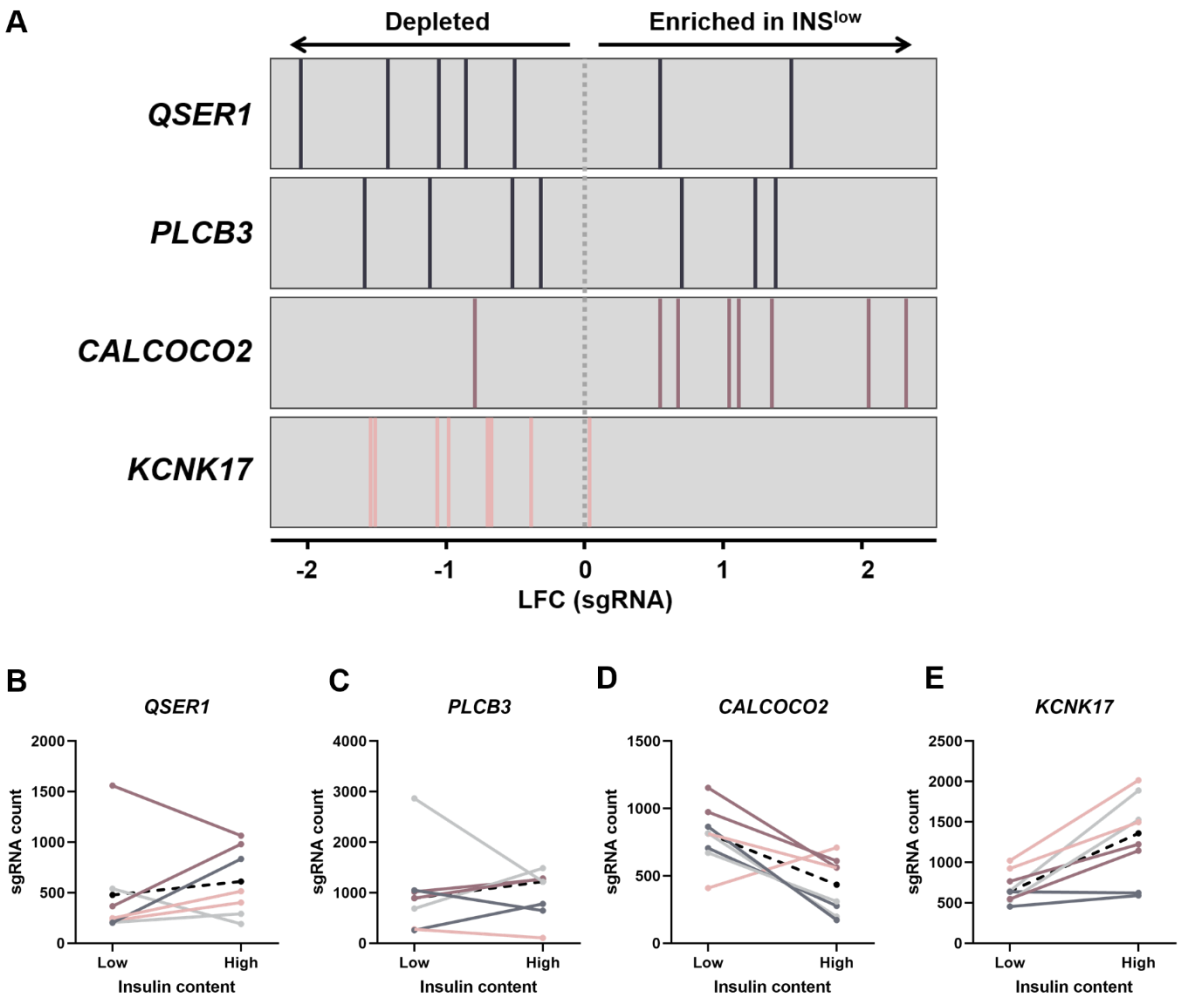


Figure 5.2 - Prioritised genes for follow-up studies. sgRNA fold change **(A)** and count distribution **(B-E)** for all 8 sgRNAs per gene in the genome-wide CRISPR screen. **(A)** Each line indicates a single sgRNA and is either depleted (*left*) or enriched (*right*) in the low insulin content sample for the control genes *QSER1* and *PLCB3* (*black*), the positive regulator *CALCOCO2* (*purple*) and the negative regulator *KCNK17* (*pink*). The grey dashed line at 0 LFC indicates no enrichment or depletion. **(B-E)** Changes in sgRNA count from low to high insulin content sample with each colour representing the same sgRNA across the two screen replicates. The black dashed line represents the median sgRNA count for this gene. sgRNA with zero counts in *PLCB3* and *QSER1* were removed. LFC, log2 (fold change); INS^{low}, low insulin content sample.

5.3.2 Control genes do not affect insulin content and secretion

To confirm the lack of an effect on insulin content as observed in the genome-wide screen and to also determine any potential effects on insulin secretion, the control genes *QSER1* and *PLCB3* were silenced in EndoC-βH1 using siRNA.

The molecular function of *QSER1* in the cell is still unknown and no studies have been published yet to explore its role. In addition to its coding variant association with T2D, *QSER1* has been implicated as an upregulated plasma peptide in ovarian cancer and identified as being close to a GWAS SNP with the strongest association for age of onset in Parkinson disease (Latourelle *et al.*, 2009; Dufresne *et al.*, 2018; Barbitoff *et al.*, 2018; Mahajan *et al.*, 2018a). *QSER1* silencing in EndoC-βH1 was highly efficient as only 21.7% (p=0.001) of residual mRNA expression was detected compared to non-targeting control cells (Figure 5.2 A). It was not possible to determine the level of reduction on the protein level, as all tested *QSER1* antibodies were unspecific (Supplementary Figure 1). Confirming the results of the genome-wide CRISPR screen, there was no difference in insulin content between *QSER1* silenced cells and controls (11 372 pg vs 10 173 pg, p=0.257, Figure 5.3 B). *INS* expression was also not affected by *QSER1* reduction with a mean expression of 97.1% of the non-targeting control (p=0.814, Figure 5.3 C). In addition to not having an effect on insulin content, insulin secretion was not significantly different in *QSER1* silenced cells compared to control cells (p>0.450 across different conditions, Figure 5.3 D). As expected, when normalised to insulin content, insulin secretion was also not significantly changed between the two groups (p>0.474 across different conditions, Figure 5.3 F). In response to raised glucose concentration, *QSER1* silenced cells increased their insulin secretion by 1.55 and 1.64 compared to 1.57 and 1.65 in control cells when assessing raw insulin secretion (p=0.963) or normalised to insulin content (p=0.999), respectively (Figure 5.3 E, G).

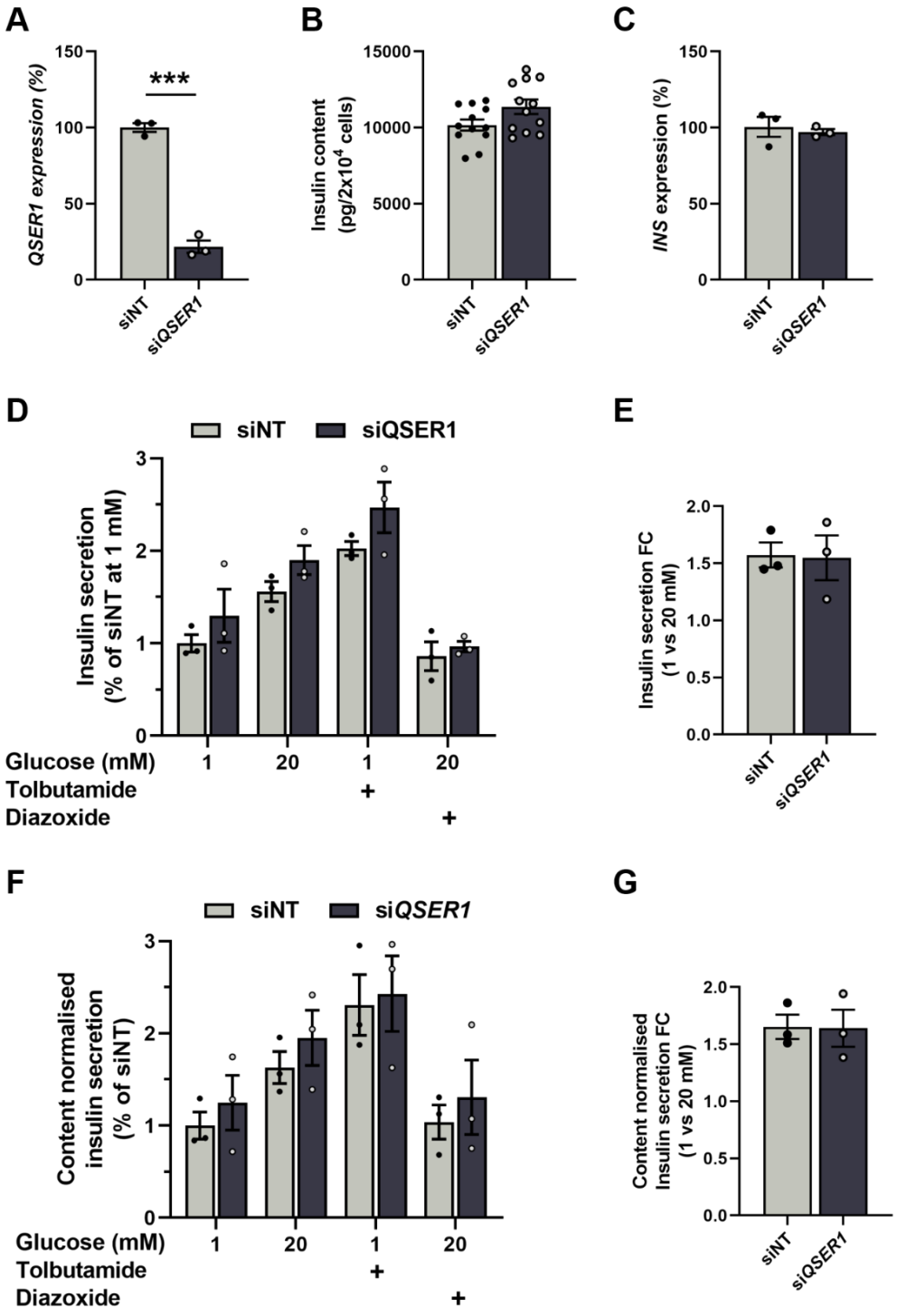


Figure 5.3 - *QSER1* follow-up studies. All data are from si*QSER1* treated cells compared to non-targeting control cells. **(A)** mRNA expression of *QSER1*. **(B)** Insulin content in pg per 20 000 cells. **(C)** mRNA expression of *INS*. **(D, F)** Insulin secretion normalised to siNT **(D)** or to insulin content and siNT **(F)** in 1 mM, 20 mM, 1 mM+100 μM tolbutamide or 20 mM glucose+100 μM diazoxide. **(E, G)** Insulin secretion fold change from 1 to 20 mM glucose. All data are mean ± SEM from three independent experiments and 12 independent replicates from three experiments for insulin content. Data were analysed using two-way ANOVA Sidak's multiple comparison test **(D, F)**, two-sample t-test **(A, E, G)** and one-way ANOVA with Sidak's multiple comparison test **(C, B).** P-values *** < 0.001. FC, fold change; NT, non-targeting.

PLCB3 encodes the phospholipase PLC-β3 and is one of 13 PLC isozymes which play a major role in intracellular signal responses. Upon activation by G protein-coupled receptors (GPCR), PLC-β3 catalyses the hydrolysis of phosphatidylinositol 4,5-biphosphate to generate the protein kinase-activating second messenger diacylglycerol and the Ca^{2+}-mobilising secondary messenger inositol 1,4,5-trisphosphate (IP_3) (Gresset, Sondek, & Harden, 2012). (Zawalich & Zawalich, 1996). In pancreatic beta cells, PLC is involved in the amplifying pathway to mediate the effects of extracellular stimuli on insulin secretion (Thore *et al.*, 2005).

PLCB3 was silenced by 51.2% (p=0.0005) in EndoC-βH1 as measured on the mRNA level (Figure 5.4 A). The silencing on the protein level could not be determined as the antibodies in the western blot-based detection were not specific (Supplementary Figure 1). Consistent with the results observed in the genome-wide CRISPR screen, *PLCB3* silencing did not induce a difference in insulin content with 10 173 pg per $2x10^4$ cells compared to non-targeting control cells at 10 891 pg (p=0.372, Figure 5.4 B). Furthermore, *INS* expression was not affected by *PLCB3* reduction (84.9%, p=0.334, Figure 5.4 C). While PLC plays an important role in mediating the amplification response in insulin secretion downstream of GPCR activation, insulin secretion upon glucose stimulation was not significantly different between *PLCB3* silenced and control cells (p>0.237 across different conditions, Figure 5.4 D, F). The glucose stimulation index between low and high glucose was not affected by silencing of *PLCB3* based on raw insulin secretion (1.48 vs 1.57, p=0.869) and neither when based on values normalised to insulin content (1.47 vs 1.65, p=0.918, Figure 5.4 E, G).

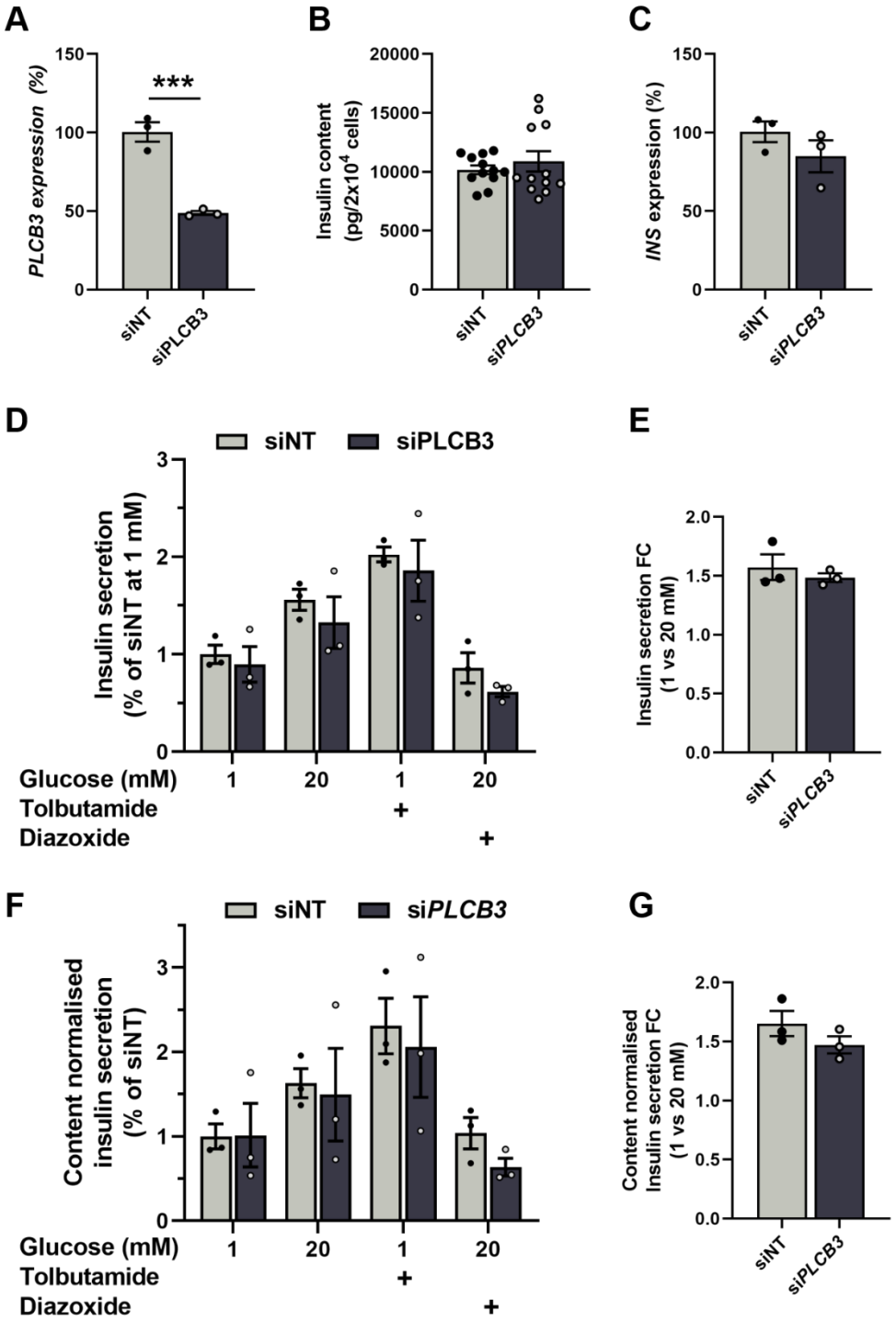


Figure 5.4 – *PLCB3* follow-up studies. All data are from si*PLCB3* treated cells compared to non-targeting control cells. **(A)** mRNA expression of *PLCB3*. **(B)** Insulin content in pg per 20 000 cells. **(C)** mRNA expression of *INS*. **(D, F)** Insulin secretion normalised to siNT **(D)** or to insulin content and siNT **(F)** in 1 mM, 20 mM, 1 mM+100 μM tolbutamide or 20 mM glucose+100 μM diazoxide. **(E, G)** Insulin secretion fold change from 1 to 20 mM glucose. All data are mean ± SEM from three independent experiments and 12 independent replicates from three experiments for insulin content. Data were analysed using two-way ANOVA Sidak's multiple comparison test **(D, F)**, two-sample t-test **(A, E, G)** and one-way ANOVA with Sidak's multiple comparison test **(C, B).** P-values *** < 0.001. FC, fold change; NT, non-targeting.

Taken together, silencing of the T2D GWAS genes *PLCB3* and *QSER1* did not affect the level of intracellular insulin in EndoC-βH1 which confirmed the lack of an effect observed in the genome-wide CRISPR screen. In addition, reduction of both genes did not affect *INS* expression or insulin secretion at different glucose concentrations in human beta cells. *PLCB3* has shown to be involved in the amplification pathway. However, the response to external stimuli such as GLP1 were not specifically tested.

5.3.3 *CALCOCO2* knockdown reduces insulin content and secretion

Macroautophagy (referred to as autophagy) describes a process within the cell to clear protein aggregates, damaged organelles or invading pathogens by packaging them into vesicles called autophagosomes followed by degradation by the lysosome (Glick, Barth, & Macleod, 2010). *CALCOCO2/NDP52* is a ubiquitin-binding autophagy receptor and plays an essential role in selective autophagy (Xie *et al.*, 2015; Boyle, Ravenhill, & Randow, 2019). Autophagy receptors directly bind to prospective autophagosome cargo and act as a bridge between the autophagic machinery and the degradation target. *CALCOCO2* has so far been identified to selectively bind invading pathogens including *Salmonella enterica* and *Mycobacterium tuberculosis* (xenophagy) and damaged mitochondria (mitophagy) (Xie *et al.*, 2015; Furuya *et al.*, 2018; Boyle *et al.*, 2019).

The silencing of *CALCOCO2* in EndoC-βH1 was highly efficient with a mean reduction of 78.1% on the mRNA level ($p=0.0002$, Figure 5.5 A). In line with the RNA expression results, the protein level was also decreased by 80.0% compared to non-targeting control cells ($p=0.0008$, Figure 5.5 B, C). To assess the effects of *CALCOCO2* silencing on the function of human beta cells, insulin content was measured using several independent approaches in addition to investigating its effects on insulin secretion and expression.

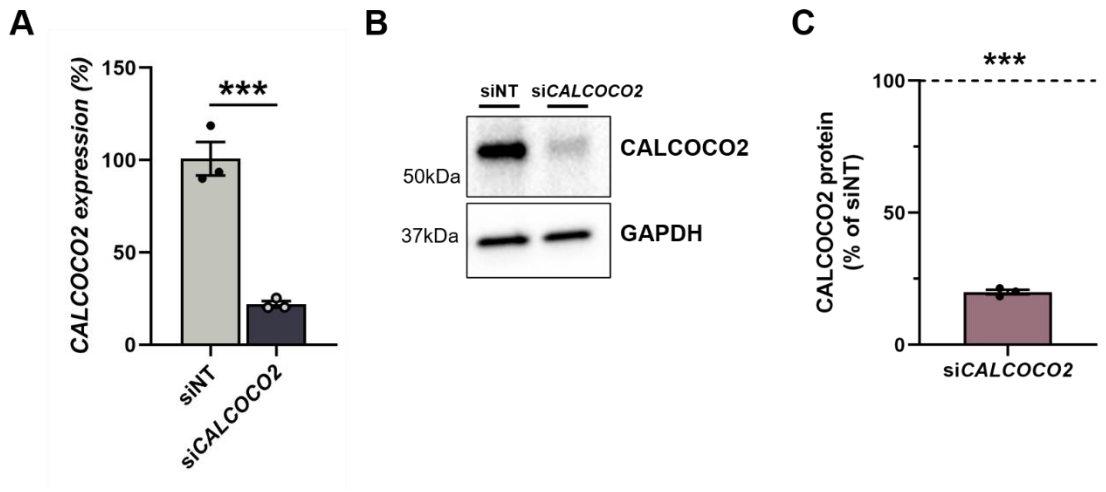


Figure 5.5 – *CALCOCO2* silencing. All data are from si*CALCOCO2* treated cells compared to non-targeting control cells. **(A)** mRNA expression of *CALCOCO2*. **(B)** Protein level of CALCOCO2 and its loading control GAPDH. **(C)** Quantification of CALCOCO2 western blot data, normalised to GAPDH and siNT control cells. The protein level is displayed as percentage of siNT which is highlighted as dotted line at 100%. All data are mean ± SEM from three independent experiments. Data were analysed using two-sample t-test **(A)** and one-sample t-test **(C)**. P-value *** < 0.001. NT, non-targeting.

Insulin content was measured in *CALCOCO2* silenced cells using two complementary antibody-based detection methods (Figure 5.6 A-C). Consistent with the screening results, insulin content was significantly reduced by 24.0% in *CALCOCO2* silenced cells using an alphaLISA based detection method (7735 pg vs 10173 pg, p=0.017, Figure 5.6 A). This reduction in total intracellular insulin was not due to lower cell numbers from decreased cell viability as there was no difference in a combined readout of proliferation and apoptosis between the two groups (94.4%, p=0.5305, Figure 5.6 D). In addition, a western blot-based detection approach also identified a significant reduction of insulin content by 38.0% (p=0.002) upon *CALCOCO2* silencing (Figure 5.6 B, C). Such a reduction in insulin content could potentially be a consequence from defects in various pathways or cellular processes including reduced *INS* transcription or impaired proinsulin processing.

Insulin expression was not significantly different in *CALCOCO2* silenced cells compared to controls (95.9%, p=0.928), indicating no transcriptional effect upon loss of the protein (Figure 5.6 E). To assess a potential processing defect in the maturation from preproinsulin and proinsulin to insulin, an antibody specifically targeting the insulin precursors was used to detect differences in the levels of immature insulin between non-targeting controls and *CALCOCO2* silenced cells. The antibody is C-peptide specific and therefore not able to detect mature insulin while C-peptide alone is too small for SDS-Page based detection and does therefore not produce a signal, as shown in Figure 5.1. Proinsulin was reduced by 46.6% upon *CALCOCO2* KD. However, due to large variation between replicates which is common in western blot based approaches, the difference was not significantly different (p=0.12, Figure 5.6 F, G). Normalisation of the level of proinsulin to insulin contained the same technical variance but confirmed an unaffected ratio with a mean of 85.5% compared to the ratio of control cells (p=0.47, Figure 5.6 H). The equal reduction in both, proinsulin and insulin therefore demonstrated that the processing from insulin precursors to mature insulin is not affected. The decrease in insulin content, both on the level of proinsulin and mature insulin can therefore not be attributed to reduced gene expression or incomplete processing suggesting a mechanism potentially involving increased granule or protein degradation or reduced translation.

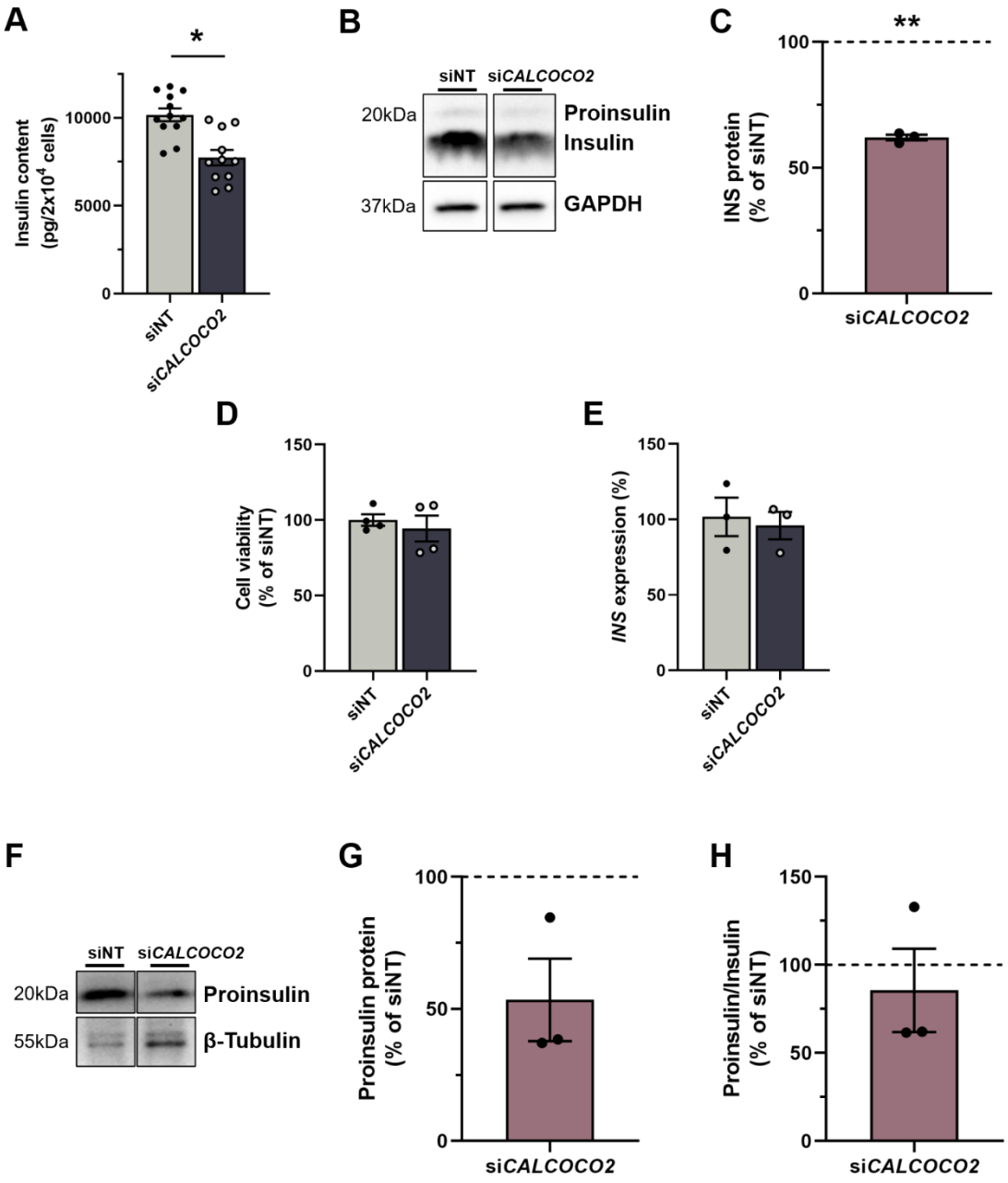


Figure 5.6 – Insulin content in *CALCOCO2* silenced cells. All data are from si*CALCOCO2* treated cells compared to non-targeting control cells. **(A)** Insulin content in pg per 20 000 cells. **(B**) Protein level of Insulin and its loading control GAPDH. The antibody also detects proinsulin but only gives a weak signal compared to mature insulin. **(C, G)** Quantification of Insulin **(C)** and Proinsulin **(G)** western blot data normalised to GAPDH or β-Tubulin and siNT control cells. The protein level is displayed as percentage of siNT which is highlighted as dotted line at 100%. **(D)** Cell count measurements. **(E)** mRNA expression of *INS*. **(F)** Protein level of Proinsulin and its loading control β-Tubulin. **(H)** Proinsulin to insulin ratio based on western blot data from C and G. All data are mean ± SEM from three **(B, C, E-H)** or four **(D)** independent experiments and 12 independent replicates from three experiments **(A)**. Data were analysed using one-way ANOVA with Sidak's multiple comparison test **(A, D, E)** and one-sample t-test **(C, G, H)**. P-values * < 0.05, ** < 0.01. NT, non-targeting.

In addition to validating the insulin content-based readout of the genome-wide CRISPR screen, I also investigated if glucose stimulated insulin secretion was affected by *CALCOCO2* reduction as these effects cannot be concluded on the basis of changes in the level of insulin content alone. Total insulin secretion could be reduced as a result of decreased insulin content with an unchanged proportion of insulin being secreted as seen for *PAM*, the proportion of secreted insulin could be increased to compensate for the decrease in intracellular insulin as seen in unpublished studies for *RREB1* or the reduction in insulin content is a consequence of increased overall secretion (Thomsen *et al.*, 2018; Mattis *et al.*, 2019). When normalised to non-targeting control cells, insulin secretion is significantly decreased ($p<0.034$ in all conditions) in EndoC-βH1 cells treated with *CALCOCO2* siRNA (Figure 5.7 A). Across all glucose and K_{ATP} modulator conditions, insulin release was consistently lower in *CALCOCO2* silenced cells with a mean decrease of 39.30%, ranging from 43.4% in high glucose and diazoxide to 33.9% reduction under high glucose. The glucose stimulation from low to high glucose was not significantly affected ($p=0.166$, Figure 5.7 B). When normalising insulin secretion to insulin content, *CALCOCO2* silenced cells still demonstrated a reduction of around 30.2% across all conditions compared to control cells, but the difference did not meet statistical significance likely due to a smaller sample size ($p>0.412$, Figure 5.7 C). Consistent with the raw secretion, the induction of insulin secretion from low to high glucose was not statistically different between the two groups when normalising to insulin content ($p=0.673$, Figure 5.7 D).

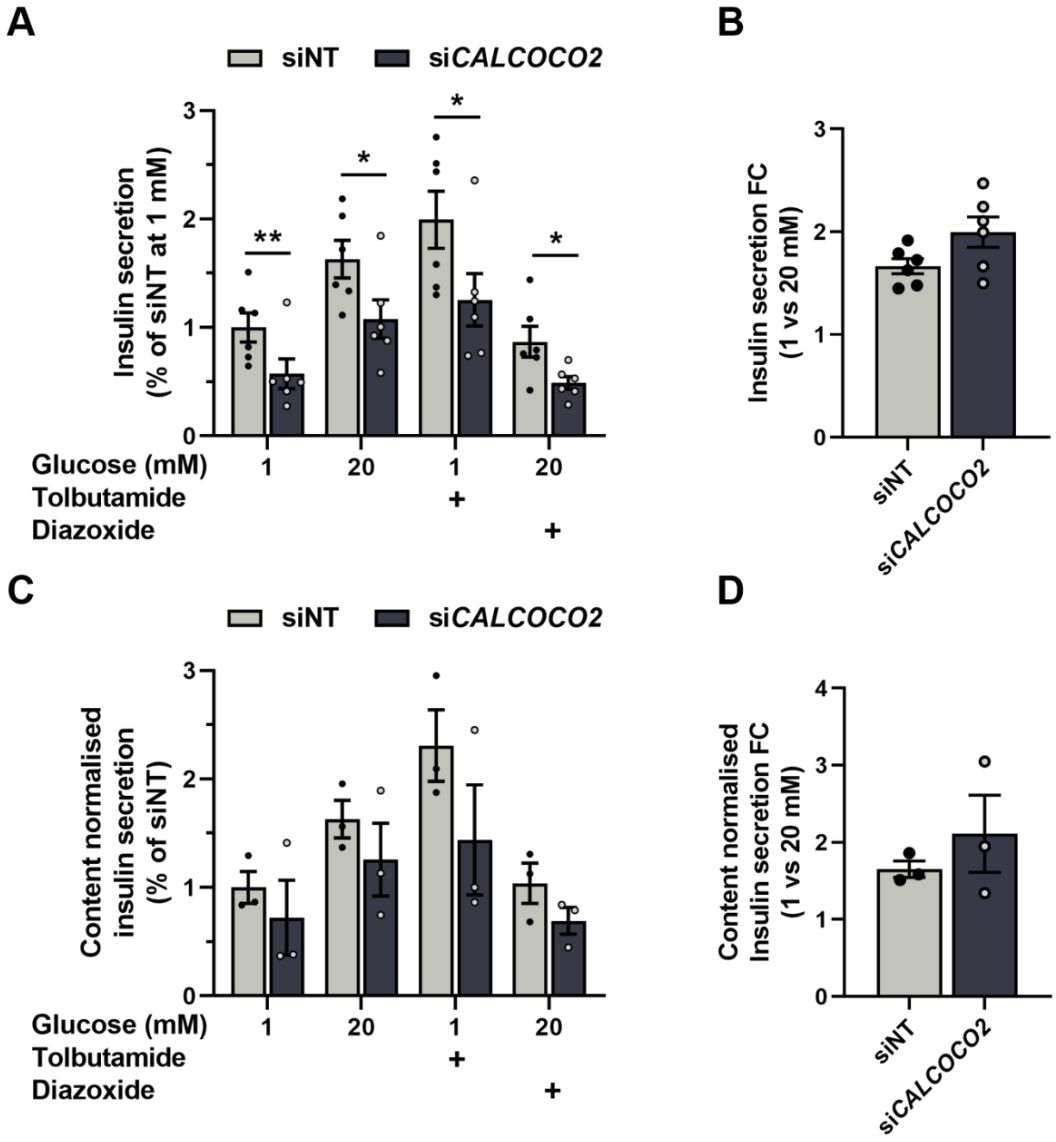


Figure 5.7 – Insulin secretion in *CALCOCO2* silenced cells. All data are from si*CALCOCO2* treated cells compared to non-targeting control cells. **(A, C)** Insulin secretion normalised to siNT **(A)** or to insulin content and siNT **(C)** in 1 mM, 20 mM, 1 mM+100 μM tolbutamide or 20 mM glucose+100 μM diazoxide. **(B, D)** Insulin secretion fold change from 1 to 20 mM glucose. All data are mean ± SEM from three **(C, D)** or six **(A, B)** independent experiments. Data were analysed using two-way ANOVA Sidak's multiple comparison test **(A, C)** and one-way ANOVA with Sidak's multiple comparison test **(B, D)**. P-values * < 0.05. FC, fold change; NT, non-targeting.

5.3.4 *CALCOCO2* is the likely causal gene at the *TTLL6* GWAS locus

CALCOCO2 has initially been associated with T2D at a novel disease locus harbouring three missense variants located within *CALOCOC2, SNF8* and *GIP* (Mahajan *et al.*, 2014; Fuchsberger *et al.*, 2016; Bonàs-Guarch *et al.*, 2018). Further, it has been implicated in an exome array based genotyping study in diverse ancestries through a single coding variant (rs10278, p.Pro347Ala) with a high PPA, indicating a partial causal role for this coding variant in driving the GWAS signal (Mahajan *et al.*, 2018b). This study also attributed a partial role to another coding variant in *SNF8 (*rs57901004, p.Arg155His), albeit to a lesser extent (0.092 vs 0.187 for *CALCOCO2*). Further, an eQTL study in human islets identified an eQTL association which colocalised with the T2D signal and implicated *UBE2Z*, another neighbouring gene at this locus (Khamis *et al.*, 2019). Collectively, the causal roles and involvement in T2D pathogenesis of all genes at the *TTLL6* locus has not been resolved yet and is highly complex.

As I was able to independently reproduce the effect of *CALCOCO2* on insulin content and attribute a potential functional mechanism and causal role to this gene, I further assessed if any of the other genes in the *TTLL6* region demonstrated a significant effect on intracellular insulin in the genome-wide CRISPR screen (Figure 5.8 A). However, none of the other genes at this locus displayed such a strong and consistent effect across sgRNAs as *CALCOCO2* (LFC=1.08, p=0.013, FDR=0.346), which is therefore the only gene to be classified as high-confidence hit (Figure 5.8 B-G). Both of the previously implicated genes *SNF8* (LFC=0.12, p=0.332, FDR=0.992) and *UBE2Z* (LFC=0.11, p=0.641, FDR=0.999) did not show a significant effect on insulin content (Figure 5.8 E, F). In addition, the other genes at this locus including *TTLL6* (LFC=0.22, p=0.552, FDR=0.999), *ATP5G1* (LFC=-0.68, p=0.044, FDR=0.750) and *GIP* (LFC=0.11, p=0.487, FDR=0.999) did not show a significant enrichment of sgRNAs within low or high insulin content (Figure 5.8 C, D, G).

Even though the sgRNAs within *ATP5G1* demonstrated a consistent direction of effect, it was not significant and less pronounced than *CALCOCO2*'s effect on insulin content. Considering the genetic evidence and *CALCOCO2*'s reproducible effect on insulin content and secretion while no significant role in regulation insulin content could be attributed to any other gene at the *TTLL6* locus, *CALCOCO2* is a strong candidate to be the causal gene in this region of chromosome 17q21.32.

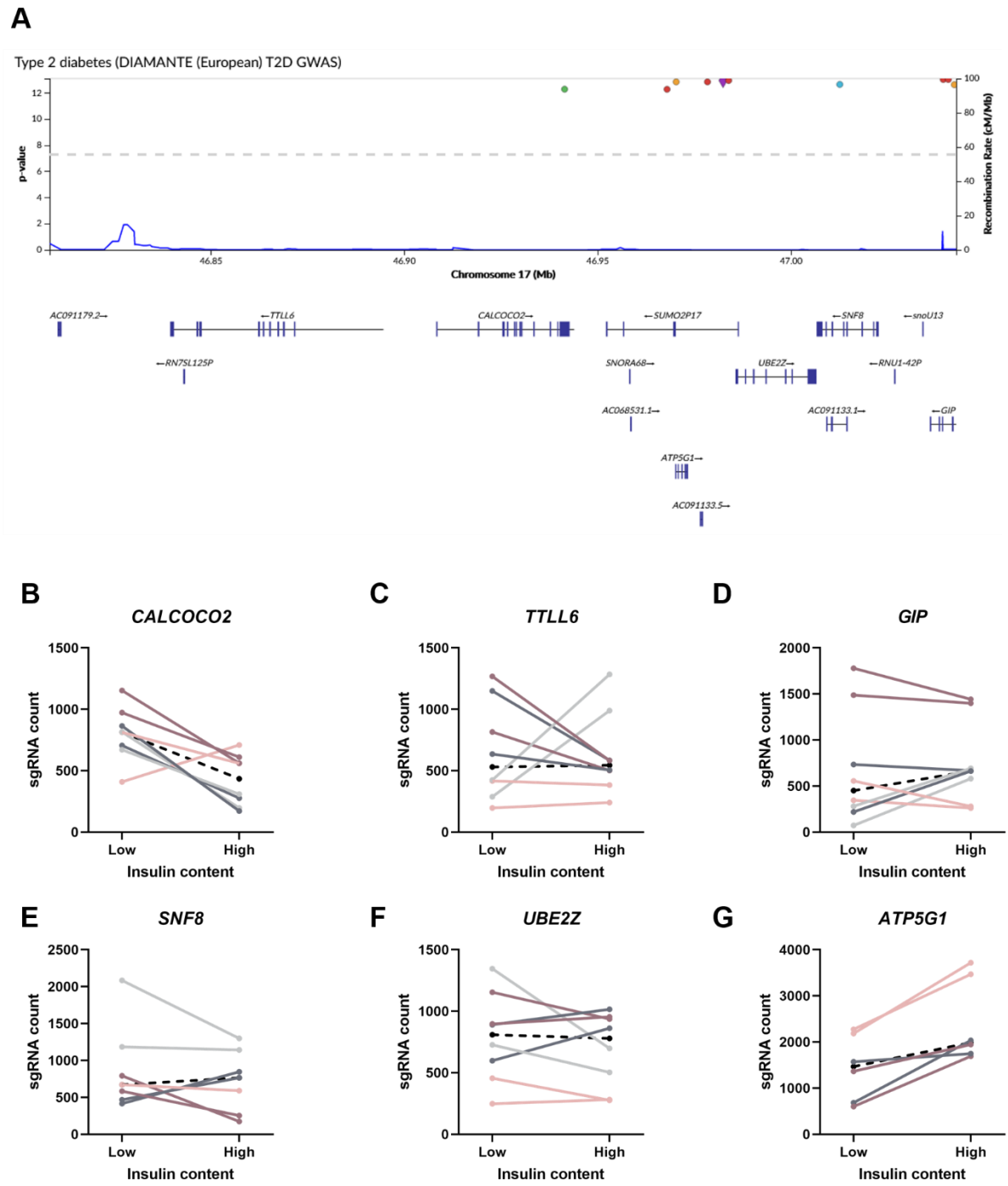


Figure 5.8 – *TTLL6* locus. (A) *TTLL6* locus with credible set GWAS association signals *(top)* and neighbouring genes *(bottom)*. **(B-G)** sgRNA count distribution for all 8 sgRNAs per gene in the genome-wide CRISPR screen for genes at the *TTLL6* locus. Changes in sgRNA count from low to high insulin content sample with each colour representing the same sgRNA across the two screen replicates. The black dashed line represents the median sgRNA count for this gene.

5.3.5 *CALCOCO2* is expressed in primary human islets

CALCOCO2 expression, cellular localisation and function has not been described yet in any model of human or rodent beta cells, I therefore performed further studies to determine its characteristics and role within the human beta cell and the pancreatic islet in more detail.

In line with its elementary and cell type independent role as autophagy receptor, *CALCOCO2* is ubiquitously expressed across tissues and cell types based on the GTEx analysis with medium expression levels ranging from 10 to 110 transcripts per million (TPM) (GTEx Consortium, 2017). In EndoC-βH1, the gene is expressed at a similar level to primary human islets and at a comparable level to other crucial beta cell genes such as *KCNJ11* (Thomsen *et al.*, 2016a). *CALCOCO2* has five described protein coding isoforms, of which two are expressed in islets based on long-read sequencing data (Figure 5.9 A, B). The predominant isoform (ENST00000448105) in islets accounts for 89% of the total *CALCOCO2* expression while the only other expressed isoform (ENST00000258947) contributes 11%. Both RNA isoforms translate into proteins with a similar molecular weight of 55 and 52 kDa, resulting in simultaneous detection of both isoforms by western blot.

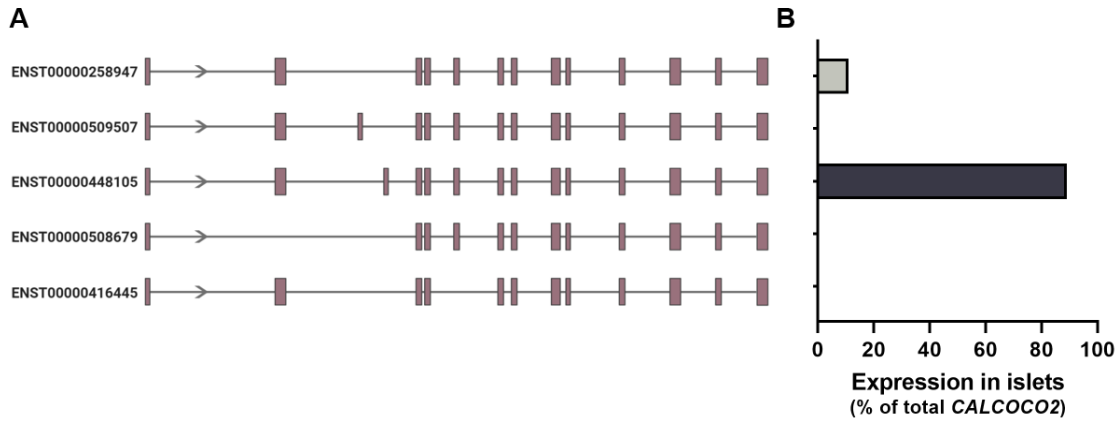


Figure 5.9 – *CALCOCO2* isoform expression in human islets. (A) Schematics of protein coding isoforms of *CALCOCO2* based on the GENCODE basic subset of representative transcripts from the ensemble genome browser. Purple boxes indicate exons, separated by introns represented by a black line. **(B)** Expressed isoforms of *CALCOCO2* in EndoC-βH1 cells relative to the indicated isoforms in **(A)**, based on long read sequencing data by Pacific Bioscience which was performed by Dr. Anne Ndungu.

In addition to demonstrating *CALCOCO2* mRNA expression in the human beta cell line EndoC-βH1 and human islets based on published RNA-Seq datasets, I also assessed the protein localisation of CALCOCO2 within primary human islets. I performed immunofluorescence staining on fixed human pancreas sections to include islet and exocrine cell types and simultaneously stained for INS, RREB1 and CALCOCO2 (Figure 5.10). INS staining indicated beta cells and highlighted islet areas while RREB1 as a known transcription factor served as a positive control for nuclear staining. INS staining colocalised with CALCOCO2, indicating that CALCOCO2 protein is expressed in human beta cells and islets (Figure 5.10 A). Within the cell, CALCOCO2 localises to the cytoplasm as shown by a lack of co-staining with the nuclear counterstain DAPI. The nuclear exclusion is not a result of incomplete permeabilization as shown by nuclear localisation of the positive control RREB1 (Figure 5.10 B).

In line with its ubiquitous expression across cell types and tissues, CALCOCO2 staining is not specific to the beta cells or islets but can also be found in the exocrine tissue of the pancreas (Figure 5.11). More specifically, CALCOCO2 staining also localises to the cytoplasm of ductal and acinar cells.

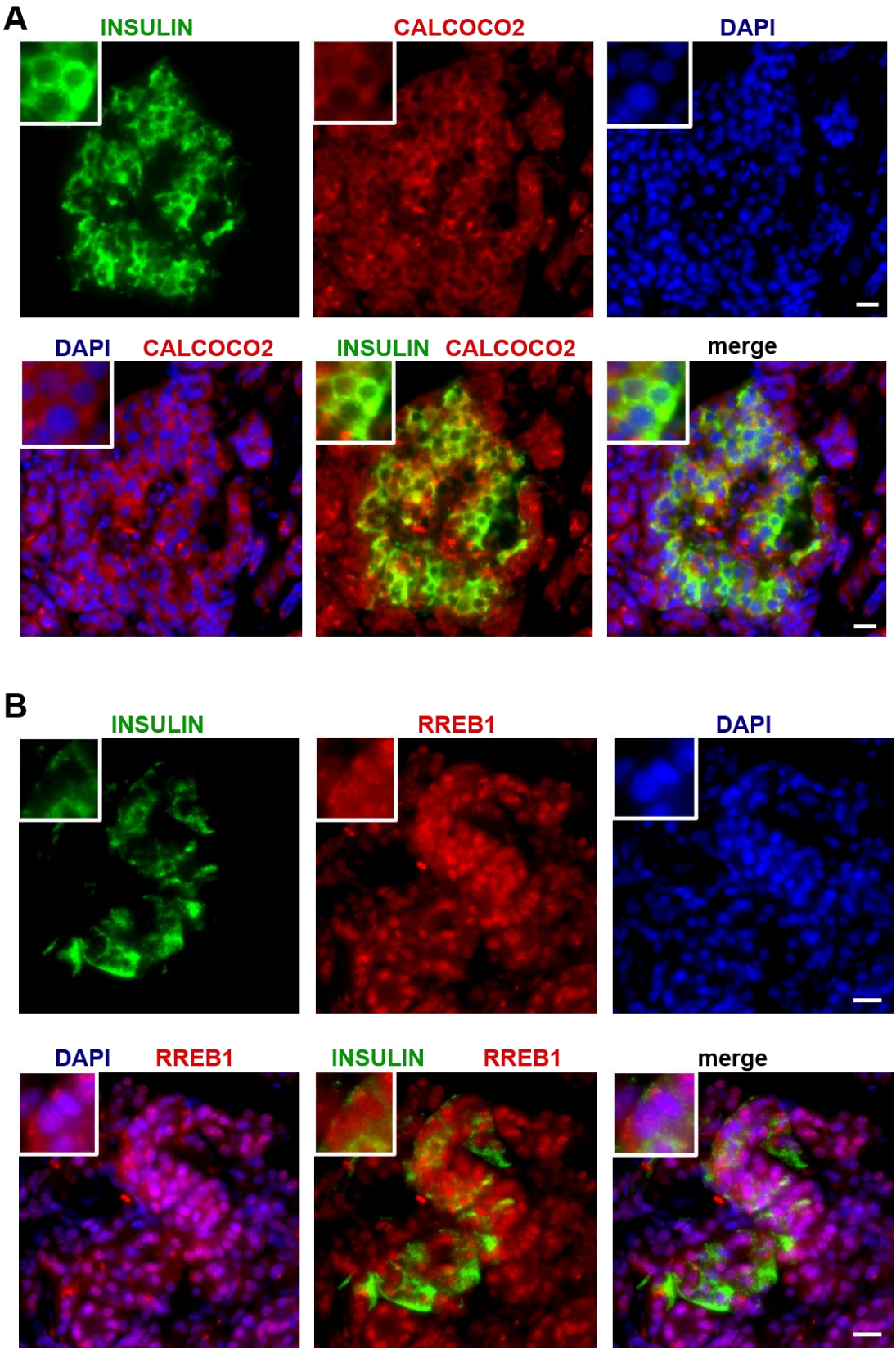


Figure 5.10 – CALCOCO2 in primary human islets. Immunofluorescence staining of primary human islets in pancreas sections. Sections were double immunostained for INS (*green*) and CALCOCO2 **(A)** or RREB1 **(B**, *red*). Cell nuclei were counterstained with DAPI (*blue*). A specific region of interest was enlarged in the top left corner in each panel. Scale bar is 10 μm.

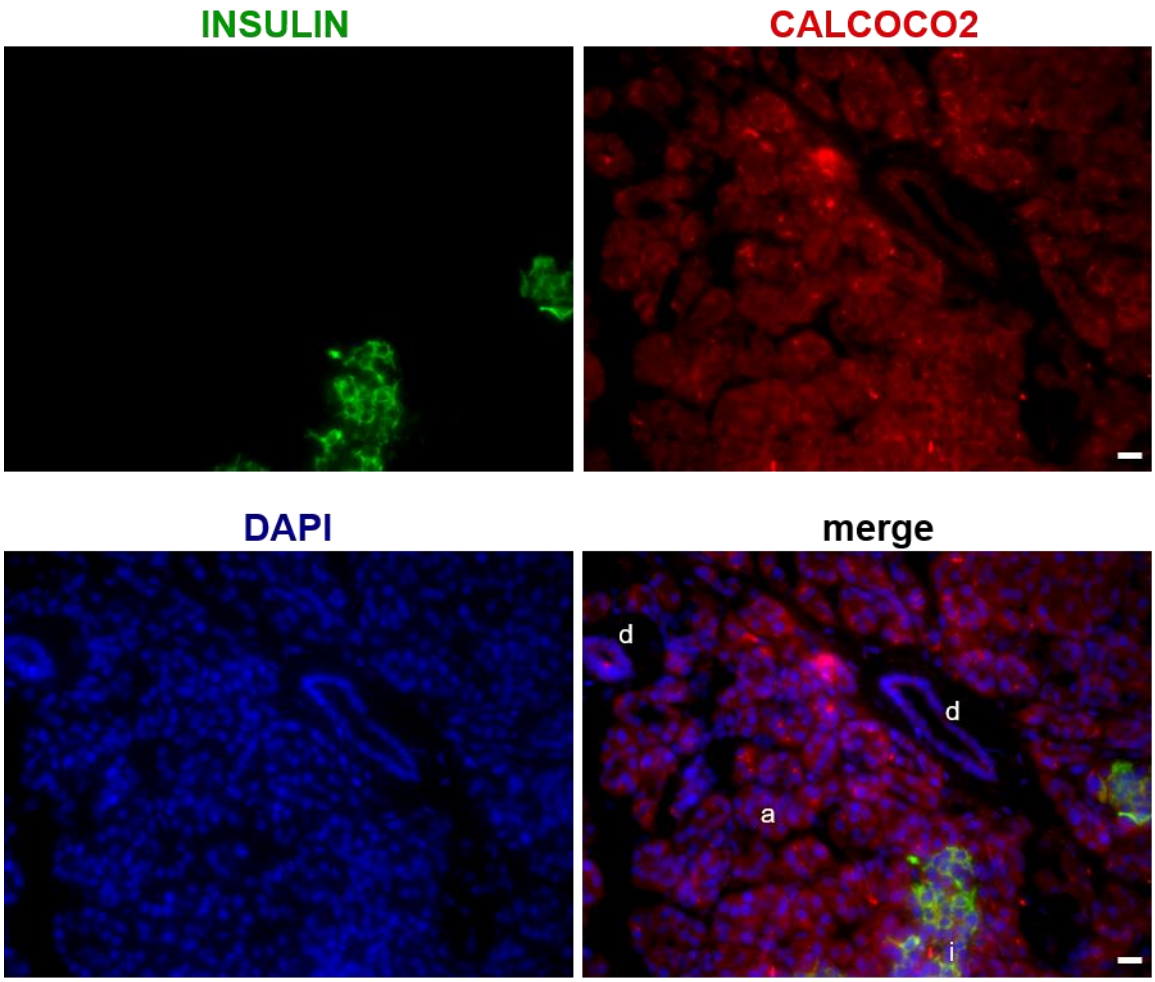


Figure 5.11 – CALCOCO2 in the exocrine pancreas. Immunofluorescence staining of exocrine tissue in pancreas sections. Sections were double immunostained for INS (*green*) and CALCOCO2 (*red*). Cell nuclei were counterstained with DAPI (*blue*). Scale bar is 10 µm. d: ductal cells; a: acinar cells; i: islet.

In addition to generally determining *CALCOCO2*'s expression in human beta cells and due to its genetic association with T2D, I investigated if its expression level changes in patients with T2D which might support a role in T2D pathogenesis or as a therapeutic target. I therefore assessed single-cell transcriptomes for pancreatic cell types in Lawlor et al. from nondiabetic (ND) and T2D donors and extracted *CALCOCO2* expression data (Lawlor *et al.*, 2017). Specific cell types were identified based on expression values in marker genes such as a high expression of *INS* classified a beta cell. Overall and including all cell types, there was no significant difference in *CALCOCO2* expression between ND and T2D samples

(endocrine cells FDR=0.983, exocrine cells FDR=0.737, Figure 5.12 A). When the sample was separated based on its cell types to identify cell type and especially beta cell specific expression signatures, endocrine cells including alpha, beta and delta cells still demonstrated no significant effect on *CALCOCO2* expression in patients with T2D (FDR>0.901, Figure 5.12 B). Even though the study was aimed at measuring single-cell expression levels in islets, a small number of exocrine cells were found in the preparations with were subsequently included in the analysis. Interestingly, *CALCOCO2* expression was significantly reduced in acinar cells of T2D patients (FDR=0.019). *CALCOCO2* expression was not significantly different in ductal cells between ND and T2D donors (FDR=0.655). The cell type specific single cell expression data of *CALCOCO2* in this dataset from Lawlor et al. also confirmed the previously demonstrated ubiquitous expression of *CALCOCO2* and the immunofluorescence staining with protein localisation in all pancreatic cell types.

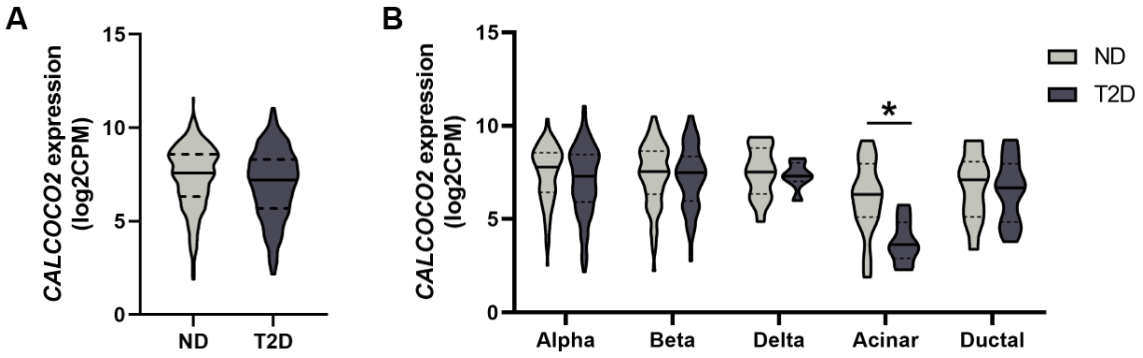


Figure 5.12 – *CALCOCO2* expression in T2D and healthy pancreatic cell types. *CALCOCO2* expression was extracted from the single cell transcriptome dataset from Lawlor et al. (Lawlor *et al.*, 2017). **(A)** *CALCOCO2* expression in ND and T2D cells in all cell types combined, including endocrine and exocrine cell types. **(B)** *CALCOCO2* expression in ND and T2D cells in specific cell types. The following number of single cells was assessed within each sample: **(A)** ND: 387 , T2D: 263; **(B)** *Alpha* ND: 138, T2D: 100; *Beta* ND: 165, T2D: 93; *Delta* ND: 16 , T2D: 7; *Acinar* ND: 15, T2D: 9; *Ductal* ND: 12, T2D: 17. Data were analysed by Lawlor et al. using differential gene expression analyses between different cell types and groups using edgeR. FDR * < 0.05. ND, nondiabetic; T2D, Type 2 Diabetes; log2CPM, log2 counts per million.

5.3.6 *KCNK17* knockdown reduces insulin secretion and *KCNK16* expression

KCNK17 encodes the two-pore-domain potassium (K2P) channel TALK2 and is part of one of the six K2P subfamilies. K2P channels induce background or leak K^+ currents which stabilise the resting membrane potential and during K_{ATP} channel closure, balance the depolarisation plateau to allow for action potential firing (Enyedi & Czirják, 2010). *KCNK17* is highly expressed in human pancreatic beta cells but its functional role is yet to be determined (Nica *et al.*, 2013; van de Bunt *et al.*, 2015; Rorsman & Ashcroft, 2018). Expression of *KCNK17* has shown to be regulated by a T2D GWAS SNP at the *KCNK16* locus, highlighting its potential role as a causal gene at this locus (Fadista *et al.*, 2014; Varshney *et al.*, 2017). Even though *KCNK17* was not classified as a final hit, it had initially been selected for functional follow-up studies as described in section 5.3.1. I silenced *KCNK17* expression in EndoC-βH1 and assessed insulin content, secretion, cell viability and *INS* expression.

KCNK17 expression was efficiently reduced by 76.12% (p=0.0004) in EndoC-βH1 using siRNA while cell viability was unaffected (109.5%, p=0.509, Figure 5.13 A, B). The efficient silencing could not be confirmed on the protein level as the tested western blot antibodies for TALK2 did not demonstrate a specific staining (Supplementary Figure 1). *INS* expression was not affected by *KCNK17* silencing (115.9%, p=0.628), excluding a potential downstream transcriptional effect of TALK2 reduction (Figure 5.13 C). Intracellular insulin was assessed using two independent antibody-based approaches (Figure 5.1). Using the sensitive alphaLISA based assay, *KCNK17* silencing induced a 16.64% increase in insulin content (11 866 pg vs 10 173 pg), albeit just below the significance threshold (p=0.111, Figure 5.13 D). Western blot analysis did not demonstrate a significant difference between non-targeting controls and *KCNK17* silenced cells (99.62%, p=0.911, Figure 5.13 E, F). As in previous studies, I also assessed if insulin processing was affected

in *KCNK17* silenced cells. Large variations in the proinsulin western blot made it challenging to draw any biologically meaningful conclusions, but *KCNK17* silencing did not significantly affect the proinsulin levels based on the mean of all replicates (131.0%, p=0.848), suggesting no defect in insulin maturation which is consistent with the unaffected levels of mature insulin (Figure 5.13 G, H).

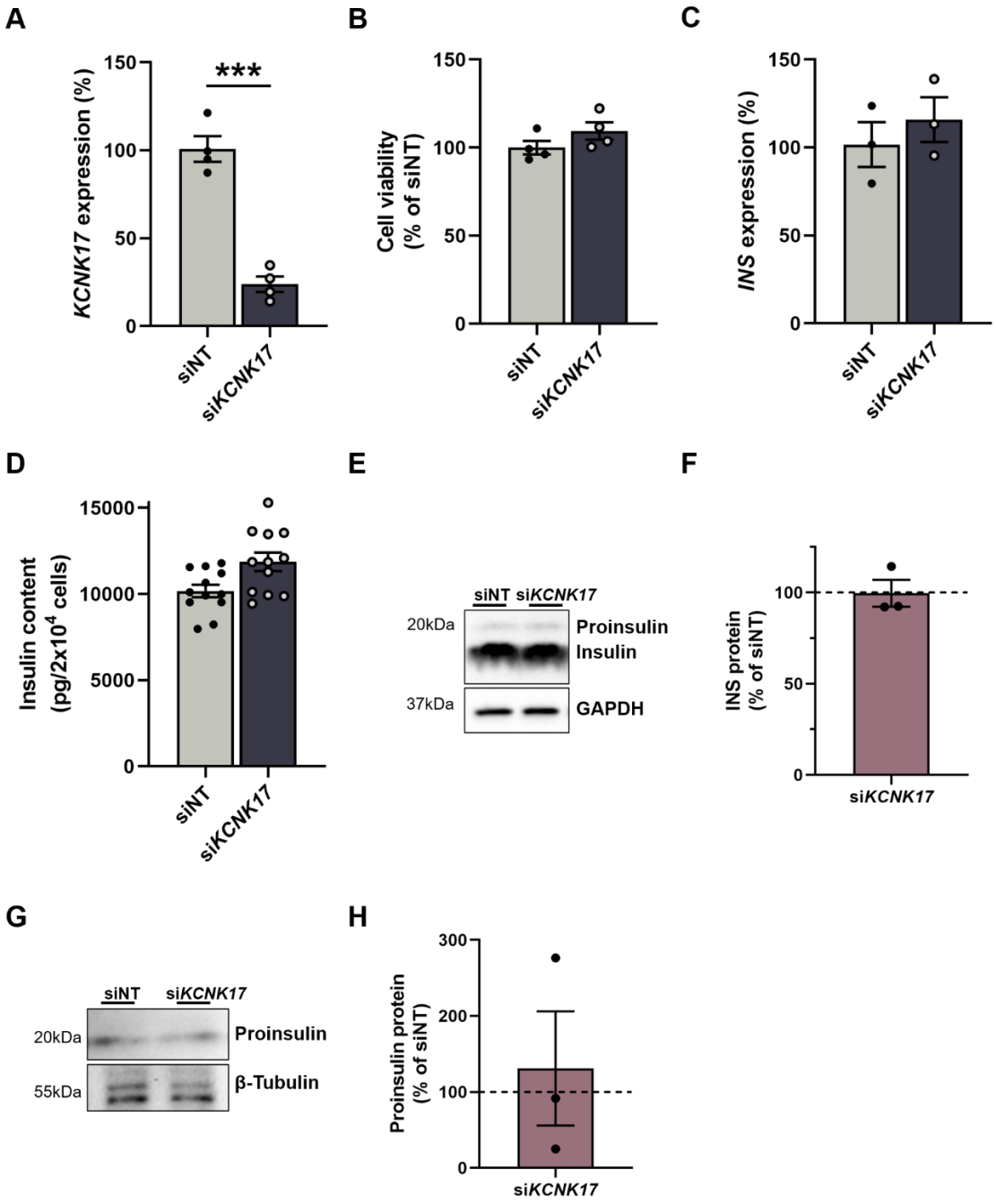


Figure 5.13 – Insulin content in *KCNK17* silenced cells. All data are from si*KCNK17* treated cells compared to non-targeting control cells. **(A)** mRNA expression of *KCNK17*. **(B)** Cell count measurements. **(C)** mRNA expression of *INS*. **(D)** Insulin content in pg per 20 000 cells. **(E)** Protein level of Insulin and its loading control GAPDH. The antibody also detects proinsulin but only gives a weak signal compared to mature insulin. **(G)** Protein level of Proinsulin and its loading control β-Tubulin. **(F, H)** Quantification of Insulin **(E)** and Proinsulin **(G)** western blot data normalised to GAPDH or β-Tubulin and siNT control cells. The protein level is displayed as percentage of siNT which is highlighted as dotted line at 100%. All data are mean ± SEM from three independent experiments **(C, E-H)**, four independent experiments **(A, B)** or 12 independent replicates from three experiments **(D)**. Data were analysed using one-way ANOVA with Sidak's multiple comparison test **(B, C, D)**, two-sample t-test **(A)** and one-sample t-test **(F, H)**. P-value *** < 0.001. NT, non-targeting.

Even though insulin content was not significantly affected by *KCNK17* silencing, I also assessed if the gene played a role in insulin secretion. Interestingly, reduction of *KCNK17* induced a decrease in insulin secretion under high glucose conditions (Figure 5.14 A). Upon incubation with 20 mM glucose or 20 mM glucose and diazoxide, insulin secretion was significantly decreased by 39.44% (p=0.034) and 38.78% (p=0.033), respectively. Subsequently, the induction from low to high glucose was diminished by 23.95% in *KCNK17* silenced cells (1.27 vs 1.67, p=0.026, Figure 5.14 B). The same trend was observed when insulin secretion samples were normalised to insulin content, albeit due to a lower sample size, none of the effects passed the significance threshold (all secretion conditions p>0.460, FC: 1.21 vs 1.65, p=0.288, Figure 5.14 C, D). Collectively, silencing of *KCNK17* did not significantly affect insulin content or *INS* expression but induced a consistent reduction of insulin secretion under high glucose conditions.

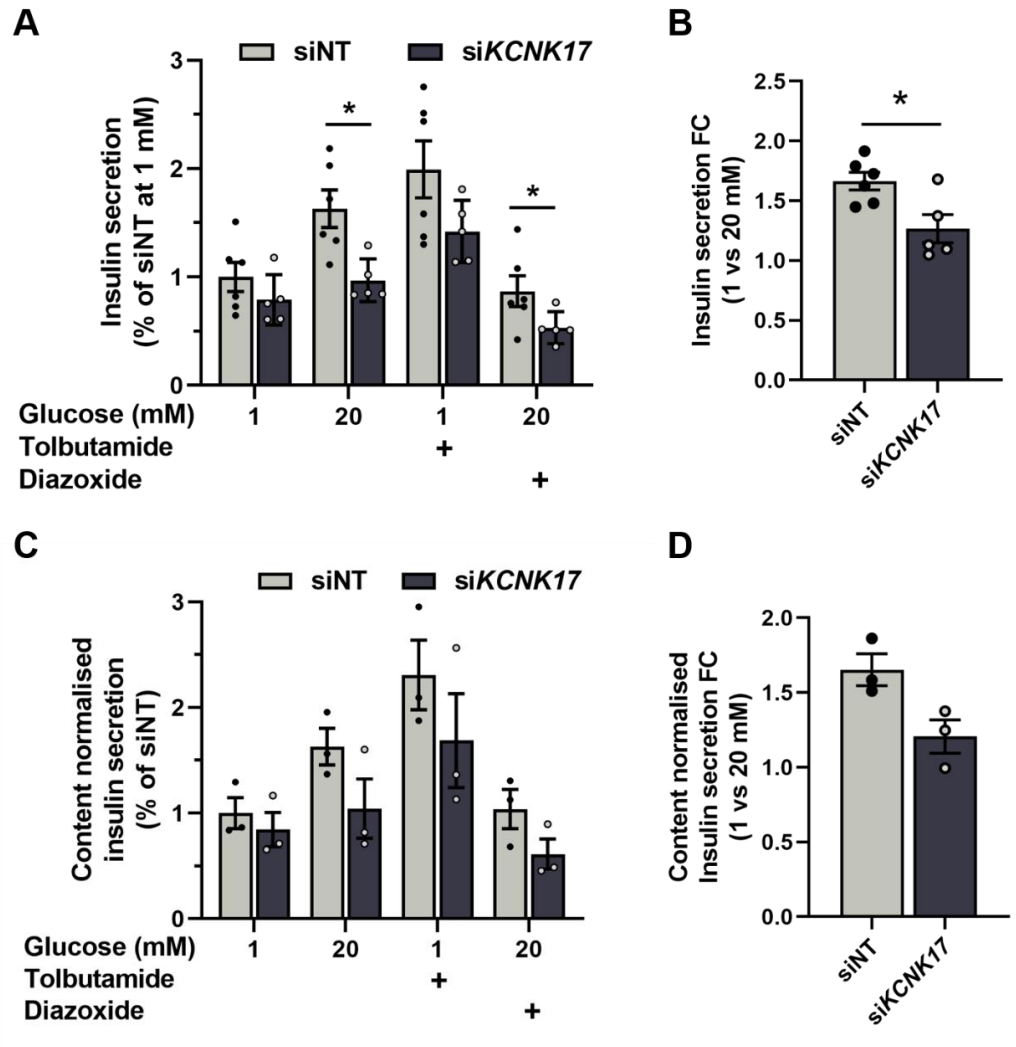


Figure 5.14 – Insulin secretion in *KCNK17* silenced cells. All data are from si*KCNK17* treated cells compared to non-targeting control cells. **(A, C)** Insulin secretion normalised to siNT **(A)** or to insulin content and siNT **(C)** in 1 mM, 20 mM, 1 mM+100 μM tolbutamide or 20 mM glucose+100 μM diazoxide. **(B, D)** Insulin secretion fold change from 1 to 20 mM glucose. All data are mean ± SEM from three **(C, D)** or six **(A, B)** independent experiments. Data were analysed using two-way ANOVA Sidak's multiple comparison test **(A, C)** and one-way ANOVA with Sidak's multiple comparison test **(B, D)**. P-values * < 0.05. FC, fold change; NT, non-targeting.

While I demonstrated a potential functional mechanism how loss of *KCNK17* could influence T2D risk, it is crucial to integrate these functional results with the underlying genetic evidence at this locus to be able to assign a potential causal role. The associated T2D GWAS index SNP rs1535500 is a coding variant within *KCNK16* and has shown to be linked to TALK1 (*KCNK16*) basal channel activity (Vierra *et al.*, 2015). The SNP is in high LD

with four other variants which overlap islet regulatory annotations. Interestingly, the risk haplotype at this locus is also associated with increased expression of the nearby gene *KCNK17* while not affecting *KCNK16* expression, linking both genes to a potential causal role at this locus (Fadista *et al.*, 2014; Varshney *et al.*, 2017). A single haplotype at this T2D GWAS locus is therefore linked to regulatory activity of *KCNK17* and coding variation in *KCNK16*. To investigate a potential direct relationship between these genes in my functional studies, I also assessed *KCNK16* expression in *KCNK17* silenced cells. Interestingly, KCNK16 expression was reduced by 26.01% (p=0.016) compared to non-targeted control cells (Figure 5.15 A). It can therefore not be excluded that the observed effect on insulin secretion upon *KCNK17* silencing was not mediated by *KCNK17* directly but indirectly through the reduction of *KCNK16*. To be certain that the effect is an actual regulatory association between the two genes and not a technical artefact from off-target effects due to the pooled siRNA-based approach, further replication using independent and individual siRNAs or shRNAs should be used.

Even though *KCNK17* did not pass the stringent significance threshold within the genome-wide CRISPR screen, it showed a consistent effect across sgRNAs which might be the consequence of a small effect on insulin content which is challenging to detect robustly compared to other genes with stronger effects within the screen. To investigate if any of the other neighbouring genes at this locus such as *KCNK16, KIF6* or *KCNK5* as shown in Figure 5.15 B, induced a significant effect on insulin content within the genome-wide CRISPR screen and might also be plausible candidates as causal genes, I assessed their sgRNA enrichment within the insulin low and high content populations. None of the other genes at this locus demonstrated a strong and consistent effect for several sgRNAs and were not classified as hits (Figure 5.15 C-F). In addition, based on the MAGeCK gene analysis, *KCNK16*, *KCNK5* and *KIF6* only possessed three or four supporting sgRNAs (sgRNAs

which MAGeCK classifies as having an effect to calculate the gene-level score). *KCNK17* on the other hand demonstrated a much more robust effect with six supporting sgRNAs, highlighting the gene as the only gene at this locus with a consistent effect on insulin content across sgRNAs.

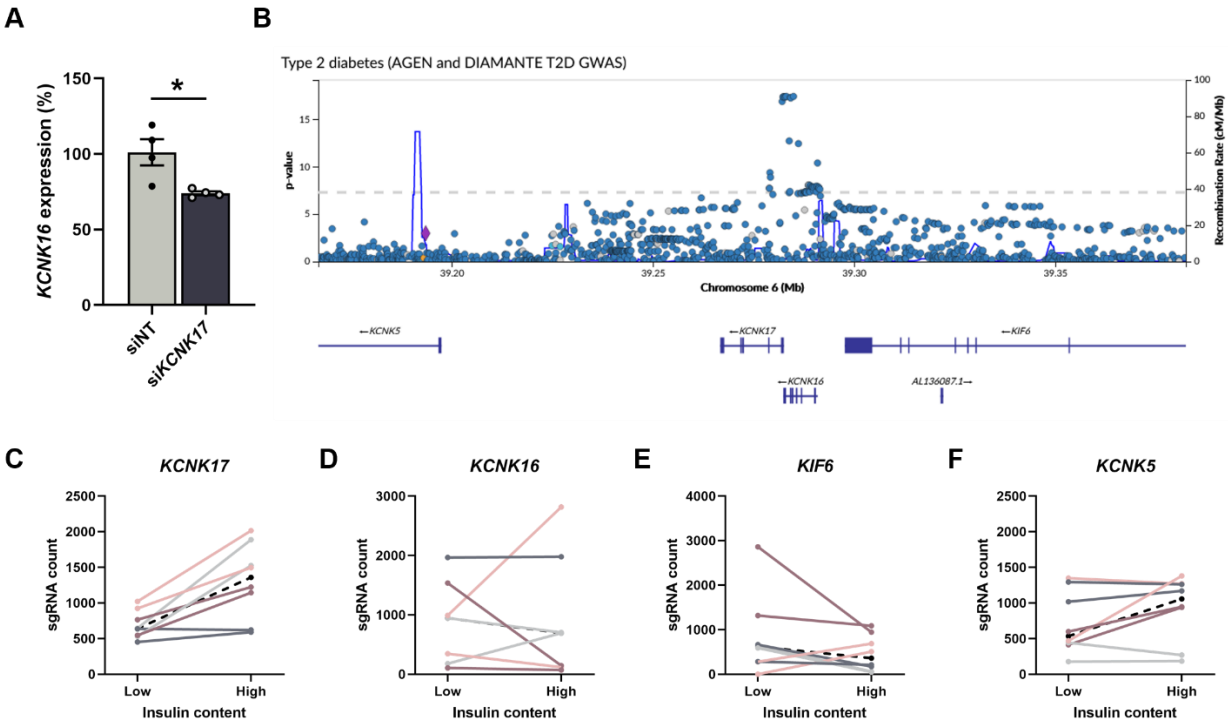


Figure 5.15 – *KCNK16/KCNK17* locus. (A) mRNA expression of *KCNK16* in *KCNK17* silenced cells compared to non-targeting controls. Data are mean ± SEM from four independent experiments and were analysed using two-sample t-test. **(B)** *KCNK16/KCNK17* locus with GWAS association signals *(top)* and neighbouring genes *(bottom)*. **(C-F)** sgRNA count distribution for all 8 sgRNAs per gene in the genome-wide CRISPR screen. Changes in sgRNA count from low to high insulin content sample with each colour representing the same sgRNA across the two screen replicates. The black dashed line represents the median sgRNA count for this gene. P-value * < 0.05. NT, non-targeting.

5.4 Discussion

The previously described genome-wide CRISPR screen is a powerful tool to identify novel regulators of insulin content at scale and to prioritise genes at T2D loci. To be certain about its accuracy and to confirm the observed phenotypic effects, it is crucial to perform functional validation studies for selected hits. In this chapter, I focused on four genes with different directions of effect on insulin content that had been previously implicated in T2D. The genes were chosen due to their association as potential T2D effector gene but whose functional role and association with disease pathogenesis has not been explored or resolved yet. I performed extensive siRNA-based follow-up studies to evaluate beta cell phenotypes such as insulin content and secretion, *INS* expression and cell viability. After having confirmed its effect on insulin content, I further characterised the expression and localisation of *CALCOCO2* in human beta cells and its likely causal role at the *TTLL6* locus. In addition, I demonstrated that *KCNK17* regulates insulin secretion during stimulation with high glucose while its causal role at the *KCNK16/KCNK17* locus remains to be resolved due to its coregulation of *KCNK16*.

5.4.1 Validation of CRISPR screening hits

To validate the effects of the genome-wide CRISPR screen, I chose a complementary siRNA based silencing approach. In contrast to CRISPR KO which leads to stable and complete gene KO, siRNA KD induces a temporary silencing on the mRNA level. As previously described in detail in Chapter 3, both approaches might give rise to technique specific phenotypes. Using siRNA as a validation approach instead of CRISPR has the advantage to be certain that the observed effects within the CRISPR screen are not due to off-target effects or other phenotypes caused by CRISPR such as DSB induced toxicity. If the effect of the screen can be replicated using this complementary and independent method, the phenotype

is most likely a functional consequence of the mRNA and protein reduction of the target gene. If it is not possible to replicate the results from the CRISPR based screen, it does not necessarily mean that the gene is not a true positive hit as residual protein might be masking the functional effects. Even though CRISPR KO have the advantage of complete protein depletion, the generation of KO cell lines in EndoC-βH1 is time-consuming, requires lentiviral production and a high number of cells. Using siRNAs in EndoC-βH1 on the other hand has the essential advantage to be an easy and quick initial validation which can then be followed up by using individual stable models such as CRISPR KO.

The genes *PLCB3, QSER1, CALCOCO2* and *KCNK17* were chosen for follow-up studies due to their association with T2D in genetic studies and because their functional role has not been resolved or studied yet in human beta cells (Fadista *et al.*, 2014; Varshney *et al.*, 2017; Mahajan *et al.*, 2018b, 2018a). The screening hits were prioritised through integration with a list of predicted T2D effector genes and upon filtering for genes without any functional evidence or causal classification, the positive regulator *CALCOCO2* was prioritised (Accelerating Medicines Partnership, 2020). The potential negative regulator *KCNK17* had been prioritised in the same way as *CALCOCO2*. This had been done at an earlier time point before the final screening analysis was performed resulting in *KCNK17* not being one of the final hits. The prioritisation of the negative controls *QSER1* and *PLCB3* was approached in the opposite direction by looking for causal genes within the list of predicted T2D effector genes that did not demonstrate an effect within the screen. Using an siRNA-based validation approach in EndoC-βH1, I was able to confirm all effects on insulin content that were observed in the genome-wide CRISPR screen. *CALCOCO2* silencing induced a significant decrease in insulin content whereas the negative control genes *PLCB3* and *QSER1* did not show any effect on insulin content or secretion. Silencing of *KCNK17* induced a small increase of around 17% in insulin content but did not reach statistical significance, as it was

the case in the CRISPR screen with a modest but robust increase across sgRNAs. By correctly identifying a true positive hit and demonstrating no effects for control genes, the results of the genome-wide CRISPR screen have shown to be accurate and valid.

Both *QSER1* and *PLCB3* did not play a role in insulin content or secretion based on siRNA silencing in EndoC-βH1. Nevertheless, this does not necessarily mean that they are not the causal genes at their respective loci. The genes might act through another T2D associated tissue or cell type such as liver or adipose cells. This is consistent with results from a multi-trait physiological clustering approach which classifies *PLCB3* as having an effect on insulin action and not insulin secretion and *PLCB3*'s association with higher waist-hip ratio and lower body fat percentage, a likely consequence of a storage capacity defect in peripheral adipose tissue (Lotta *et al.*, 2017; Mahajan *et al.*, 2018b). Further, the affected phenotypes within the beta cell might also not have been tested in my specific experimental setup. *PLCB3* plays a role in the beta cell and insulin secretion by acting as a downstream mediator in the amplification pathway (Thore *et al.*, 2005). However, the insulin secretion conditions in my validation approach only included glucose stimulation and K_{ATP} channel modulators and did therefore not target the potentiating pathway. Further experiments should therefore investigate *PLCB3*'s role in insulin secretion and specifically in the amplification pathway by also including specific receptor agonist (Yamada *et al.*, 2016). In line with my results there was a study showing that only deletion of *PLCB1* and not *PLCB3* induced impaired glucose tolerance in KO mice with no change in insulin content (Hwang *et al.*, 2019).

5.4.2 *CALCOCO2* plays a crucial role in beta cell function and T2D

I demonstrated that loss of *CALCOCO2* decreased insulin content in the human beta cell line EndoC-βH1 in a CRISPR screen and in independent complementary siRNA follow-up experiments. This reduction in intracellular insulin also resulted in a consistent inhibition of insulin secretion under both, basal and high glucose conditions. Even though *CALCOCO2* has been associated with T2D through GWAS, its functional role in human beta cells and T2D pathogenesis had not been described previously.

CALCOCO2 has initially been identified in early 2018 as a potential T2D risk gene at a novel locus with common missense variants in *CALCOCO2*, *SNF8* and *GIP* (Bonàs-Guarch *et al.*, 2018; Mahajan *et al.*, 2018b, 2018a). It has then been further implicated as a likely causal gene in an exome array based genotyping study as having a single coding variant with a high PPA which is consistent with a partial role in driving the GWAS signal (Mahajan *et al.*, 2018b). In line with my observed effects of *CALCOCO2* on insulin secretion in the pancreatic beta cell, a physiological clustering approach in this study further identified that the signal at the *TTLL6* region exerts its primary effect on insulin secretion. To further unravel *CALCOCO2*'s role in human beta cells, I established that it is ubiquitously expressed and localises to the cytoplasm in primary human islets. This cellular localisation is in accordance with its role as an autophagy receptor, previous studies in non-beta cells and the Human Protein Atlas (Xie *et al.*, 2015; Thul *et al.*, 2017; Furuya *et al.*, 2018). To further expand the characterisation of *CALCOCO2* and its role in pancreatic islets and beta cells, I set up an shRNA based knockdown approach in primary human islets in collaboration with Prof Seung Kim at Stanford University. This approach offers the advantage of being able to assess the consequences of LoF of *CALCOCO2* in the physiologically more relevant primary cells instead of in a transformed cell line. In addition, the phenotypic readout could be extended from only insulin secretion in beta cells to also include glucagon secretion in alpha

cells and the initially observed effect on insulin content could be validated using a different cellular model and another complementary KD approach. However, due to the worldwide COVID-19 pandemic, the functional experiments have not been performed yet.

The establishment of *CALCOCO2*'s role in insulin secretion and its association in genetic studies for T2D highlight its role as a likely causal gene at this locus and its involvement in disease pathogenesis. Identifying the detailed functional mechanism on how loss of CALCOCO2 affects insulin secretion would exceed the scale of this thesis and remains to be explored. A recent study using protein-protein interaction networks of potential T2D risk genes assigned *CALCOCO2* to a cluster for protein complex disassembly and macroautophagy (Fernández-Tajes *et al.*, 2019). This clustering and CALCOCO2's function as a selective autophagy receptor suggest an autophagy mediated effect on insulin content in human beta cells. Autophagy plays an important role in maintaining beta cell function and ensuring cell survival and is therefore a key pathway in T2D pathogenesis. Loss of the important autophagy enzyme *atg7* in beta cell specific KO mice gave rise to increased beta cell apoptosis leading to reduced insulin secretion and impaired glucose tolerance (Jung *et al.*, 2008). Besides cell survival, autophagy is also crucial for the regulation of insulin homeostasis in the beta cell. siRNA mediated KD of *atg5* and *7* induced increased proinsulin and insulin secretion in INS1 cells while autophagy-hyperactive mice demonstrated decreased insulin secretion due to increased degradation of insulin granules (Riahi *et al.*, 2016; Yamamoto *et al.*, 2018). KD of *CALCOCO2* on the other hand did not affect cell viability, indicating no crucial role in cell survival and demonstrated the opposite direction of effect with reduced insulin content and secretion compared to these studies which assessed modulators of autophagy. It is therefore likely that *CALCOCO2* does not affect beta cell function through regulating general autophagy within the cell.

However, CALCOCO2 has been identified as playing a role in selective autophagy, specifically as being required for efficient Parkin-mediated mitophagy (Lazarou *et al.*, 2015; Furuya *et al.*, 2018). A potential mechanism on how CALCOCO2 could affect insulin content and secretion is therefore potentially by regulating mitophagy.

Mitophagy describes the cellular process to degrade dysfunctional mitochondria and it was shown that CALCOCO2 is an essential receptor to induce mitophagy and localises to damaged mitochondria (Lazarou *et al.*, 2015; Fex *et al.*, 2018; Furuya *et al.*, 2018). In contrast to the previously described autophagy studies, loss of CALCOCO2 is consistent with the functional consequences of other investigations that modulate components of the mitophagy pathway and highlight that functional mitophagy is required for beta cell function. shRNA mediated KD of the mitophagy regulator *Clec16a* in the mouse beta cell MIN6 cells and pancreas specific deletion in mice decreased GSIS and affected mitochondria function (Soleimanpour *et al.*, 2014; Pearson *et al.*, 2018). Overexpression of the downstream mitophagy effector Parkin in MIN6 partly rescued insulin secretion under diabetic conditions (Hoshino *et al.*, 2014). The reduction of insulin secretion upon loss of these autophagy regulators is in line with my results from *CALCOCO2* KO and KD. Defective mitophagy is therefore a likely functional mechanism how loss of *CALCOCO2* affects insulin secretion and beta cell function and should be investigated in future studies (further discussed in Chapter 6).

The hypothesis that mitophagy affects insulin content and is a possible mechanism of action through which *CALCOCO2* affects beta cell function is further supported by additional screening hits mapping to this pathway. The genome-wide CRISPR screen also identified the autophagy receptor *SQSTM1* and the protein kinase *TBK1* as regulators of insulin content. Like CALCOCO2, SQSTM1 is a selective autophagy receptor which links damaged mitochondria to autophagosomes (Geisler *et al.*, 2010; Lazarou *et al.*, 2015). TBK1

phosphorylates these autophagy receptors resulting in a signal amplification enhancing mitophagy of dysfunctional mitochondria (Heo *et al.*, 2015; Richter *et al.*, 2016). Both of these genes also have no attributed role in human beta cell function but further highlight that mitophagy is an interesting pathway to study in the context of insulin regulation and beta cell function.

In addition to its association with T2D risk in GWAS, *CALCOCO2* has also been implicated with T2D in a Mendelian randomization (MR) study of the plasma proteome (Emilsson *et al.*, 2018; Zheng *et al.*, 2019). Briefly, this study assessed protein quantitative trait loci (pQTL) from GWAS for plasma proteins and investigated their potential causal role on human phenotypes through MR. MR based approaches use genetic variants as instruments instead of modifiable exposures or risk factors to reduce confounding factors and effects induced through reverse causation (Smith, 2010). Plasma protein associations are especially informative as these proteins represent easily druggable targets and are often dysregulated in disease (Imming, Sinning, & Meyer, 2006; Santos *et al.*, 2017). A missense variant in *CALCOCO2* (rs550510) with its effect allele A has shown to be associated with reduced blood protein levels of CALCOCO2 (-0.19) (Emilsson *et al.*, 2018). When the MR analysis was applied, Zheng et al. identified that rs550510 was associated with an increased risk of T2D (0.156) (Zheng *et al.*, 2019). It can therefore be concluded that reduced circulating levels of CALCOCO2 are associated with an increased risk of T2D as shown in Figure 5.16, highlighting the protein as a potential biomarker or drug target for T2D treatment. However, it is unclear where the circulating CALCOCO2 originates from and if it also affects beta cell function. As CALCOCO2 can infiltrate mitochondria through a yet unknown mechanism, it might be possible for the protein to also irrupt through the cell membrane from the blood stream (Furuya *et al.*, 2018). To assess the function of circulating CALCOCO protein on human beta cell function, we acquired recombinant full-length CALCOCO2 and will

incubate primary human islets with physiologically relevant levels of the protein followed by investigating potential effects on insulin secretion. This line of investigation had already been initiated during my research stay at Stanford University but had to be stopped prematurely due to the COVID-19 pandemic as the distribution of human tissue was halted followed by a shelter-in-place order and the discontinuation of all non-essential research activities.

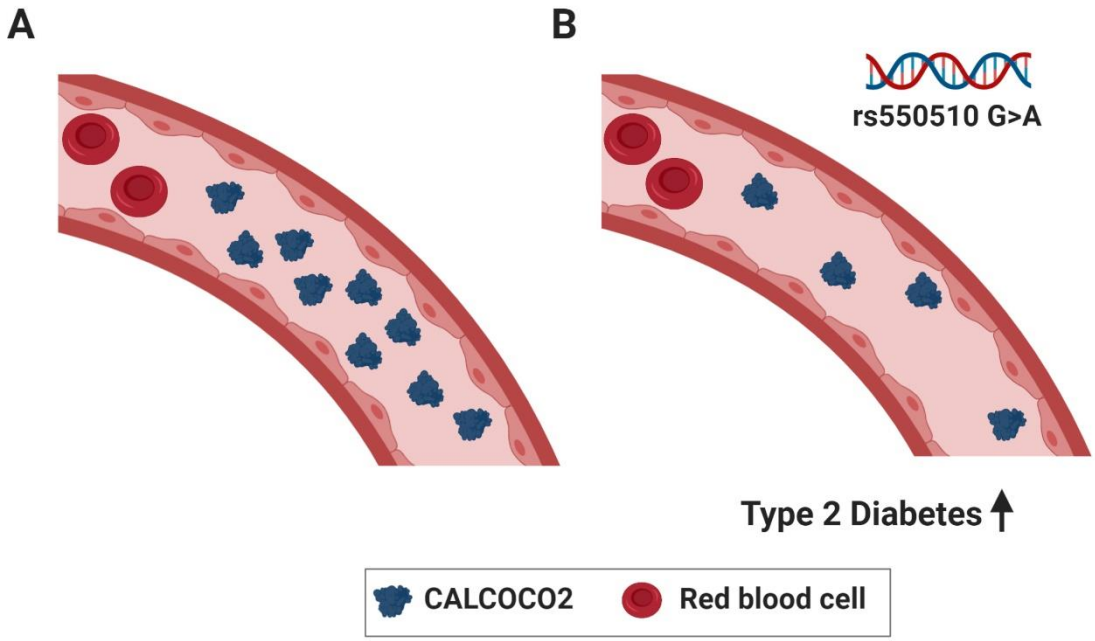


Figure 5.16 – CALCOCO2 blood protein levels in T2D risk. Individuals with the risk genotype rs550510 A within *CALCOCO2* are associated with reduced protein levels of CALCOCO2 (*blue*) in the blood plasma (Emilsson *et al.*, 2018). A decrease in circulating CALCOCO2 has been implicated with an increased risk of T2D in individual B compared to reference individual A using an MR based analysis approach (Zheng *et al.*, 2019).

5.4.3 *KCNK17* may regulate insulin secretion under high glucose conditions

Using siRNA-based silencing of *KCNK17* in the human beta cell line EndoC-βH1, I demonstrated for the first time that a decrease in TALK2 reduced insulin secretion under high glucose conditions and consequently diminished the glucose stimulation index, indicating a glucose-dependant role for TALK2. *KCNK17* has been implicated in T2D risk through being the target gene of an islet cis-eQTL which colocalises with a T2D GWAS SNP (Fadista *et al.*, 2014; Varshney *et al.*, 2017). Interestingly, this GWAS variant maps to a coding region within *KCNK16*, highlighting both genes as potential causal genes at this locus. The SNP is in high LD with four other variants which overlap islet regulatory annotations (promoter chromatin state and ATAC-Seq peak) and have shown to be the mediators of the transcriptional regulation of *KCNK17* (Varshney *et al.*, 2017). The complex haplotype at this locus therefore regulates transcriptional activity of *KCNK17* while also containing a coding variant in *KCNK16*. The risk haplotype has shown to increase *KCNK17* expression, which seems contradictory to my functional experiments that demonstrated a reduction in insulin secretion upon decrease of *KCNK17* expression. However, as I have not directly tested the consequences of *KCNK17* overexpression, it cannot be concluded that both, a decrease and increase in *KCNK17* expression reduce insulin secretion such as previously shown for *ZMIZ1*, which is also implicated in T2D risk (van de Bunt *et al.*, 2015).

K_{ATP} channels are responsible for the majority of K^+ currents while other background K^+ channels such as two-pore domain channels including TALK2 do not significantly affect total beta cell conductance under low glucose conditions (Göpel *et al.*, 1999; Ravier *et al.*, 2009). Under high glucose stimulation, these background channels can significantly affect K^+ currents and total beta cell conductance, in line with the observed effect of *KCNK17* silencing under high and not basal glucose (Seghers *et al.*, 2000). Studies focussing on other two-pore domain channels such as TASK1 (*KCNK3*) or TALK1 (*KCNK16*) also

demonstrated affected insulin secretion under high glucose conditions only (Dadi, Vierra, & Jacobson, 2014; Vierra *et al.*, 2015). Loss of TASK1 or TALK1 induced an increase in insulin secretion in mouse islets which is opposite to the direction of effect upon reduction of TALK2 that was observed in my KD experiment in human EndoC-βH1. In line with my observed effect of *KCNK17* in the beta cell, a previous study from our group investigated potential T2D risk genes and their effect on insulin secretion in EndoC-βH1 and identified as well that silencing of *KCNK17* induced a 17.3% reduction in insulin secretion under high glucose conditions (Thomsen *et al.*, 2016a). Additional experiments to investigate potential functional differences between these channels will be necessary to explain the different directions of effect.

In addition to showing an effect on insulin secretion under high glucose stimulation, I also demonstrated that silencing of *KCNK17* reduced *KCNK16* expression and highlighted why this locus is highly complex while also illustrating a crucial functional aspect that is necessary to uncover the true causal gene(s) at this locus. Before drawing any functional conclusions however, individual follow-up studies using CRISPR KO or individual siRNAs should be performed to be certain that the simultaneous reduction of *KCNK16* expression is not a technical artefact and a consequence of off-target silencing due to high similarity between the two genes. If *KCNK16* and *KCNK17* truly regulate each other's expression while having opposite effects on insulin secretion, that would potentially indicate a compensatory mechanism by each channel to stabilise K+ currents and insulin secretion. The mechanism on how *KCNK17* could regulate *KCNK16* expression is unclear and the gene does not encode regulatory miRNA or lncRNA that might directly affect *KCNK16* expression.

5.4.4 Conclusion

The functional follow-up studies using an siRNA-based silencing approach confirmed the observed effects on insulin content from the genome-wide CRISPR screen, therefore validating the accuracy and demonstrating that the screen can identify true regulators of beta cell function. The prioritised control genes *QSER1* and *PLCB3* with likely causal roles in T2D did not affect insulin content or secretion, suggesting that they influence T2D pathogenesis through beta cell independent mechanisms. *KCNK17* at the complex *KCNK16* GWAS locus decreased insulin secretion under high glucose conditions while also lowering *KCNK16* expression, highlighting that more in-depth experiments are needed to understand the causality at this locus. The autophagy receptor *CALCOCO2* has been identified as a novel regulator of insulin content and insulin secretion in human beta cells. The gene is highly expressed in primary human islets and is the candidate causal gene in the *TTLL6* region while also being a potential biomarker or drug target for T2D.

6

Discussion

6.1 Novel CRISPR/Cas9 approaches in human beta cells to identify T2D effector transcripts

T2D is a common, complex and multifactorial disease and has been associated with environmental, epigenetic and genetic risk factors. The underlying genetic predisposition to disease risk has been the focus of many studies aimed at better understanding disease pathogenesis, identifying novel treatment targets and improving patient stratification (McCarthy, 2017; Fitipaldi *et al.*, 2018). Since the first GWAS for T2D in 2007, more than 400 distinct genetic association signals have been shown to be robustly linked to disease risk (Sladek *et al.*, 2007; Scott *et al.*, 2007, 2017; Saxena *et al.*, 2007; Mahajan *et al.*, 2018b, 2018a; Spracklen *et al.*, 2020; Vujkovic *et al.*, 2020). The majority of these disease loci have not been translated into functional understanding and causal genes yet. The work presented in this thesis addressed this question by performing a genome-wide pooled CRISPR screen in human beta cells aimed at bridging this gap of connecting GWAS T2D signals to causal genes, functional mechanisms and tissue of actions. In addition to providing an improved functional understanding of T2D pathogenesis, this work has also identified novel regulators of beta cell function and explored the strengths and limitations of gene editing and silencing in human beta cells.

The investigation in Chapter 3 described a novel CRISPR/Cas9 pipeline in the human beta cell line EndoC-βH1 followed by in-depth functional and genomic characterisation of several proof of concept cell lines. I developed a lentiviral-based dual sgRNA CRISPR editing pipeline and described the first protocol to robustly genome edit a human beta cell line. Using this approach, I created six KO cell lines whose characterisation formed an independent first-author manuscript, contributed to important experiments on another co-author manuscript and also formed the basis for several independent follow-up studies by fellow DPhil students. This work has provided an important novel CRISPR pipeline for

future investigations into human beta cell function and T2D while also describing essential proof of concept work and setting the foundation for a genome-wide CRISPR screen in EndoC-βH1.

The main focus of my thesis, the genome-wide CRISPR screen in EndoC-βH1 was outlined in Chapter 4. After demonstrating in Chapter 3 that EndoC-βH1 are indeed a suitable cell line for CRISPR based investigations, I developed and performed a genome-wide pooled CRISPR screen for human beta cell function using a FACS based phenotypic readout for insulin content. 580 genes were classified as regulators of insulin content and to assess if potential causal genes for T2D were among the screening hits, I integrated the data with predicted T2D effector genes and prioritised several genes for functional follow-up studies where the supporting evidence was not sufficient to assign a causal role including *CALCOCO2*, *FADS1* and *C2CD4B.*

Chapter 5 continued this work by focusing on the functional follow-up studies of prioritised genes. Based on the previously described integration with predicted T2D effector genes, I chose the autophagy receptor *CALCOCO2* as an interesting gene for in-depth functional validation as it had not been associated with beta cell function before. I also validated negative screening results by choosing genes that were not identified as hits in the genome-wide screen but had been previously linked to a causal role. By confirming their lack of an effect on insulin content in individual follow-up studies, I could evaluate the potential of false negatives within the screen. The genes chosen for this line of investigation were *QSER1* and *PLCB3*. In addition, as I had already initiated some follow-up experiments while the final analysis was still being undertaken, the potential causal gene *KCNK17* was also prioritised for follow-up studies albeit not being a final screening hit. I individually assessed all of these genes using a complementary siRNA approach to confirm that the observed phenotypes were independent of their LoF strategy and using this strategy, I was able to

independently validate all results from the genome-wide CRISPR screen. Based on the initial identification as a screening hit followed by further individual follow-up studies, I was able to assign a functional role to *CALCOCO2* in the human beta cell and highlight its role as a likely causal gene for T2D.

Collectively, this thesis described the process from an initial individual CRISPR proof of concept study to a high-throughput genome-wide pooled CRISPR screen which was then used to prioritise and validate a novel regulator of beta cell function and T2D.

6.2 *CALCOCO2* and its role in human beta cell function and T2D

In Chapter 5, I established for the first time that the screening hit *CALCOCO2* plays an important role in maintaining intracellular insulin and regulating the level of secreted insulin in human beta cells. In combination with its coding genetic association with T2D, it is also the likely causal gene at this GWAS locus, one of 24 genes in this region of chromsome 17q21.32 within 1 Mb of the index/lead variant. Integration with previous studies demonstrated that *CALCOCO2*'s effect on insulin secretion is consistent with the role of other components of the mitophagy pathway in beta cell function, as previously explained in Chapter 5 (Soleimanpour *et al.*, 2014; Pearson *et al.*, 2018; Furuya *et al.*, 2018; Fernández-Tajes *et al.*, 2019). I therefore hypothesised that *CALCOCO2*, in its function as selective autophagy receptor regulates beta cell function, insulin content and secretion through being an essential modulator of the mitophagy pathway in human beta cells (Figure 6.1).

Damaged mitochondria within the cell are being removed in the process of Parkin-mediated mitophagy through interaction of the kinase PINK1 with the ubiquitin ligase Parkin. This process in turn recruits selective autophagy receptors that bind to phagophores and initiate the formation of autophagosomes around the dysfunctional mitochondria (reviewed in Pickles, Vigié, & Youle, 2018). Initially, upon mitochondrial stress or damage, PINK1 stabilizes on the outer mitochondrial membrane and phosphorylates ubiquitin and Parkin, leading to Parkin-mediated formation of ubiquitin chains (Clark *et al.*, 2006; Lazarou *et al.*, 2012; Kondapalli *et al.*, 2012; Koyano *et al.*, 2014; Okatsu *et al.*, 2015). Autophagy receptors subsequently connect ubiquitinated mitochondria and autophagosomes to induce mitophagy (Lazarou *et al.*, 2015). In mammalian cells, five autophagy receptors have been identified as mediating mitophagy: CALCOCO2/NDP52, OPTN, SQSTM1/p62, NBR1 and TAX1BP1.

Only CALCOCO2 and OPTN have shown to be essential with the presence of either of these sufficient to induce mitophagy as shown in single and double KO models in HeLa cells (Lazarou *et al.*, 2015). Distinct expression levels of both receptors within different tissues however hint at tissue specific roles of each receptor. In addition, *SQSTM1* was also a significant hit in the genome-wide CRISPR screen. Based on this crucial role of *CALCOCO2* in mitophagy, I propose that in human beta cells, deletion of *CALCOCO2* is sufficient to impair the recruitment of phagophores to damaged and ubiquitinated mitochondria, which cannot be fully compensated for by OPTN (Figure 6.1). This impaired mitophagy subsequently leads to the accumulation of damaged mitochondria within the cell, inducing mitochondrial respiratory dysfunction and leading to impaired insulin secretion.

To understand the functional role of *CALCOCO2* in human beta cells, future studies should focus on experimentally investigating the proposed hypothesis. Potential experimental strategies to approach the question include assessing the accumulation of damaged mitochondria by electron microscopy and measuring intracellular ATP levels and the level of mitophagy directly. This could be approached by assessing the degradation of mitochondrial DNA or mitochondrial DNA encoded inner membrane protein cytochrome C oxidase subunit II (CoxII) upon extrinsic initiation of mitophagy using the mitochondrial depolarizer protonophore carbonyl cyanidem-chlorophenyl hydrazine (CCCP) or the inhibitors oligomycin and antimycin A (OA).

In addition to these functional investigations, further experiments could focus on specifically connecting the T2D GWAS variant in *CALCOCO2* to its functional mechanism within the beta cell. The coding missense variant with a partial role in driving the GWAS association signal was classified as benign and tolerated on common *in silico* prediction tools, therefore likely not to be damaging whereas my functional studies were performed using a LoF model (Ng & Henikoff, 2003; Adzhubei, Jordan, & Sunyaev, 2013).

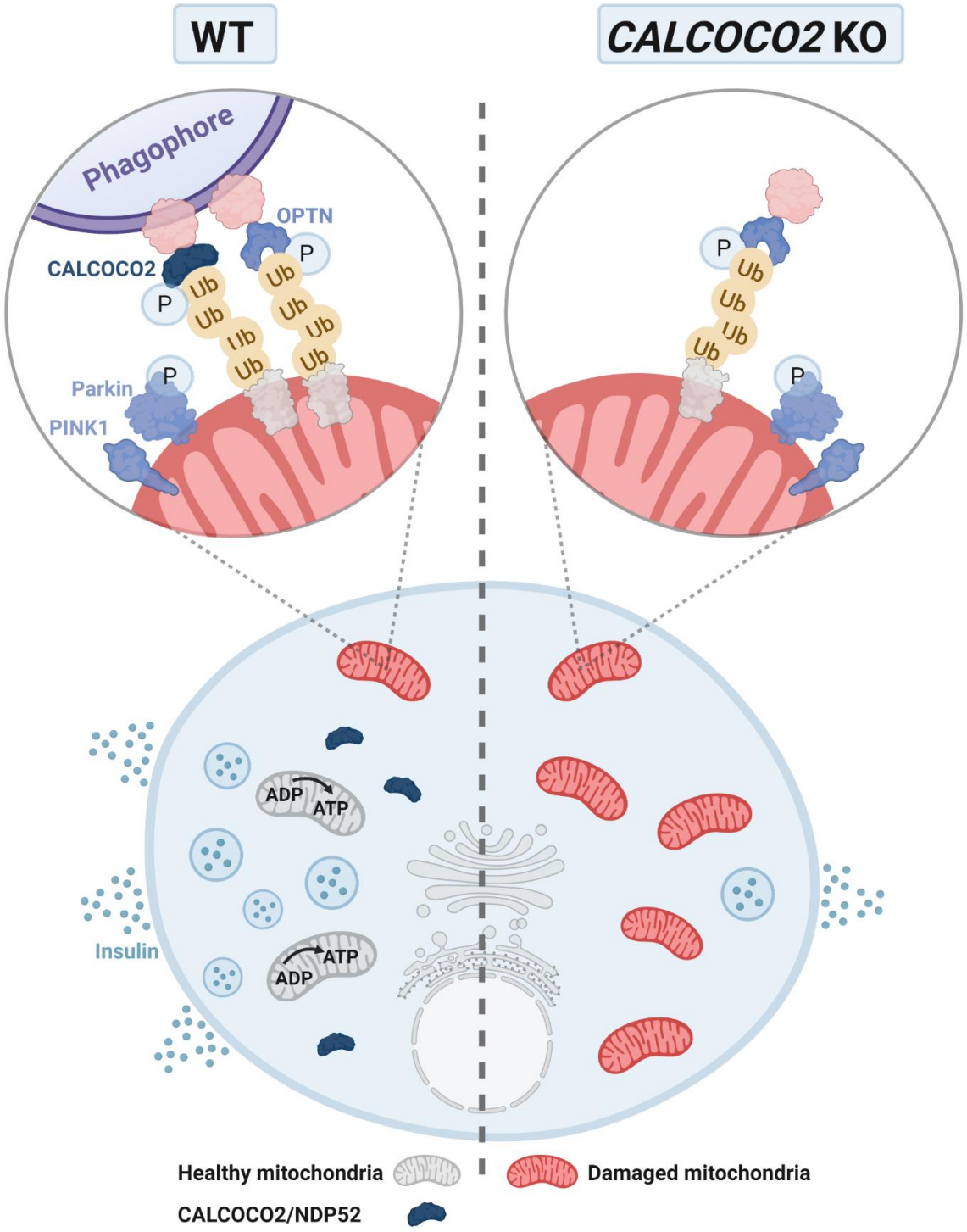


Figure 6.1 – Impaired mitophagy in *CALCOCO2* KO beta cells. WT beta cells *(left)* perform PINK1 and Parkin initiated mitophagy to degrade damaged mitochondria *(red)* using CALCOCO2 and OPTN as autophagy receptors which bind to ubiquitin chains on the mitochondria membrane. WT cells can therefore produce a high level of ATP and secrete insulin accordingly. In *CALCOCO2*-KO beta cells *(right)*, OPTN alone is not sufficient to induce the required level of mitophagy, leading to an accumulation of damaged mitochondria, insufficient ATP production and insulin secretion.

Functionally connecting the variant and the functional investigations would link *CALCOCO2* with certainty to its role as causal gene in T2D. Experiments to make this functional connection include overexpressing the GWAS SNP and a control construct in *CALCOCO2*-KO EndoC-βH1 or assessing beta cell function and insulin secretion in primary islets from risk variant carriers such as previously done for *PAM* (Thomsen *et al.*, 2018).

The identification of *CALCOCO2* as a novel regulator of beta cell function and insulin secretion also justifies the consideration if the gene would be a promising drug target to treat T2D. Based on my functional data, loss of CALCOCO2 reduces beta cell function. Targeted supplementation or gene overexpression would therefore be a potential strategy to restore or improve beta cell function in patients with LoF *CALCOCO2* mutations. General overexpression or activation of CALCOCO2 beyond its normal level has not been experimentally tested yet and it is therefore unclear if it would be beneficial as a generalisable treatment strategy in T2D. Although CALCOCO2 might be a suitable drug target due to its function as cellular receptor, its ubiquitous expression across tissues and its role in innate immunity and bacterial degradation might be associated with severe side effects outside of beta cells upon targeting (Xie *et al.*, 2015; Thul *et al.*, 2017; Boyle *et al.*, 2019). If naturally occurring LoF variants in *CALCOCO2* or "human knockouts" can be found in cohorts with high rates of consanguinity such as the East London Genes & Health, it would support the feasibility of a treatment strategy based on reducing the levels of CALCOCO2 and the specific phenotypic consequences could be further assessed through electronic health record data or genotype-based recall (Finer *et al.*, 2020).

6.3 Challenges associated with translating CRISPR screening hits into causal genes at T2D GWAS loci

One of the main objectives of the genome-wide CRISPR screen and subsequent functional follow-up studies was to prioritise causal genes at T2D GWAS loci. By identifying a functional role in insulin regulation in human beta cells for the potential T2D risk gene *CALCOCO2*, this work has showcased how the screening data can be applied to prioritise and identify a likely causal gene. Based on the screening data alone, it is not possible to distinguish if the hit gene is a causal gene for T2D and mediates disease risk or if it is solely a regulator of beta cell function. It is therefore crucial to line up and integrate the evidence from other complementary studies to be able to pinpoint the causal role and mechanism of the gene within T2D. Other lines of evidence include genetic studies linking the disease-associated variants directly to target genes through fine-mapped coding variants or colocalisation of non-coding variants with eQTLs or epigenomic annotations. In addition, directly investigating the risk alleles in physiological studies or functional experiments can conclusively connect a gene and functional mechanism to its GWAS and disease risk association. The genome-wide screening hits therefore provide a prioritisation approach which presents only one line of evidence connecting a potential causal gene to its direction of effect, tissue of action and functional mechanism. This dataset also offers a validation tool to assess if effector transcripts proposed by other investigations demonstrate a disease-relevant effect in a cell type of interest.

This genome-wide CRISPR screen focussed on a single tissue of interest, the pancreatic beta cell. Even though the pancreatic beta cell and insulin secretion have been highlighted as being central to T2D pathogenesis in a wide range of studies, many causal genes exert their effect on disease risk through other disease relevant tissues such as the liver or muscle (Ingelsson *et al.*, 2010; Mahajan *et al.*, 2018b; Udler *et al.*, 2018). In addition, this screen only assessed insulin content under basal glucose conditions, some genes however might only display an effect upon glucose stimulation or after incubation with other external stimuli. Further, the screen used a CRISPR KO library, thereby only investigation genes and their effect upon LoF. To be able to link all potential causal genes to their tissue of action, direction of effect and functional mechanism, a comprehensive investigation with a range of screens in different tissues and under various conditions would be required. One of the strengths of this screen was its unbiased and genome-wide approach which allows to potentially identify every single regulator of beta cell function. However, if the main goal of the investigation is to primarily identify T2D risk genes, only focussing on a smaller subset of genes would also be suitable. A smaller screen targeting only a few hundred genes at T2D loci would also make the investigation of several cell types and conditions feasible. Depending on the number of target genes, the screen could be performed in an arrayed setting, similar to Thomsen et al. or use a subset CRISPR library in a pooled setting (Thomsen *et al.*, 2016a).

In summary, this screen provides a field-enhancing resource to be able to identify novel regulators of beta cell function in a genome-wide and unbiased manner and as an additional advantage, the data can also be used to prioritise causal genes at T2D GWAS loci and provides a look-up resource to demonstrate that effector transcripts proposed by other complementary approaches are indeed disease relevant. If the main goal is to focus on potential T2D risk genes, a smaller screen which is not genome-wide is also suitable.

6.4 Future directions

The optimisation, execution and follow-up studies of this genome-wide pooled CRISPR screen in EndoC-βH1 have revealed a large number of novel but yet unanswered questions which should be addressed by in-depth future investigations. While developing the CRISPR pipeline in EndoC-βH1 in Chapter 3, I discovered a novel role for NEUROD1 and demonstrated a yet unexplained lack of effect in *SLC30A8*-KO cells, both discoveries would be interesting starting points for future studies to better understand the function of these respective genes within the human beta cell. The CRISPR screen itself is by design a discovery and prioritisation tool to be followed-up by further functionals studies focussing on individual genes. One of these follow-up studies has already been performed for the hit gene *CALCOCO2*. Specific questions arising from the work on *CALCOCO2* have already been discussed in more detail in section 6.2. Throughout Chapter 4, many other interesting hits such as *FADS1* or *C2CD4B* have been highlighted and would provide a rationale for future studies. Due to being one of the strongest hits, its previous association with T2D, its hit classification in the CRISPR screen by Fang et al. and its unresolved role in human beta cells I also prioritised *FADS1* besides *CALCOCO2* for shRNA based functional studies in primary human islets (van de Bunt *et al.*, 2015; Fang *et al.*, 2019). However, due to the COVID-19 associated shortening of my stay at Prof Seung Kim's lab at Stanford University, it was not possible to generate any functional date for *FADS1* as I was only able to set up and amplify the shRNA constructs. Once the normal laboratory-based research will be possible again, our collaborators will continue to perform this functional investigation.

This genome-wide dataset also has the potential to discover novel biological insights and regulators of beta cell function beyond the genes originating from T2D GWAS which were the focus of this follow-up investigation. Unbiased follow-up studies or small-scale screens focussing on a certain number of top hits will reveal genes associated with insulin content regulation that would have not been prioritised based on the integration with other complementary datasets. This work therefore has far-reaching implications beyond this thesis and the initial integration and follow-up approach focussing only on potential causal genes in T2D.

Throughout my PhD, I have also been involved in generating a list of potential T2D target genes which are currently being investigated in *in vivo* screens across disease relevant phenotypes in Drosophila by Prof Seung Kim at Stanford University and in Zebrafish by Prof Marcel den Hoed at Uppsala University. The focused list contains all candidate genes at T2D loci which are most likely to be causal to allow for a simultaneous and high-throughput comparison of several genes at a single locus. The most recent addition of genes also included hits from my CRISPR screen such as *CALCOCO2*. The emerging data from these model system screens will be integrated with my screening hits to combine independent lines of evidence and it will also shed further light on the *in vivo* role of specific genes such as *CALCOCO2*.

To further build on the extensive work I put into optimising the pooled CRISPR screening approach and the FACS readout in EndoC-βH1, we are also going to perform a complementary gain of function (GoF) screen with our collaborators at the Target Discovery Institute (University of Oxford). This screen is going to apply the same screening pipeline based on the insulin content readout but instead of using the TKOv3 CRISPR KO library, this screen makes use of the CRISPR activation library Synergistic Activation Mediator (SAM) which induces a genome-wide transcriptional activation of target genes (Konermann

et al., 2015). The results will provide the opposite direction of effect to the already performed LoF screen resulting in the identification of novel genes which might induce specific effects only upon overexpression and add additional confidence and mechanistic insights for genes already identified by my screen.

Overall, I have developed the first robust CRISPR/Cas9 editing pipeline in EndoC-βH1, successfully performed the first genome-wide CRISPR screen for human beta cell function and identified the autophagy receptor *CALCOCO2* as a novel regulator of beta cell function and likely causal gene mediating T2D risk. This study exemplifies the power of robust individual and high-throughput genome editing approaches to promote an improved functional understanding of human beta cells and T2D pathogenesis.

Bibliography

ABDULKARIM, B. ET AL. (2015) A Missense Mutation in PPP1R15B Causes a Syndrome Including Diabetes, Short Stature, and Microcephaly. *Diabetes* **64**, 3951–3962.

ACCELERATING MEDICINES PARTNERSHIP (2020) Type 2 Diabetes Knowledge Portal. Http://www.type2diabetesgenetics.org/ [accessed 4 April 2020].

ADLI, M. (2018) The CRISPR tool kit for genome editing and beyond. *Nature Communications* **9**, 1911.

ADZHUBEI, I. ET AL. (2013) Predicting functional effect of human missense mutations using PolyPhen-2. *Current protocols in human genetics* **Chapter 7**, Unit7.20-Unit7.20.

AGARWAL, A.K. ET AL. (2002) AGPAT2 is mutated in congenital generalized lipodystrophy linked to chromosome 9q34. *Nature genetics* **31**, 21–23.

AKERMAN, I. ET AL. (2017) Human Pancreatic β Cell lncRNAs Control Cell-Specific Regulatory Networks. *Cell Metabolism* **25**, 400–411.

ALI, O. (2013) Genetics of type 2 diabetes. *World journal of diabetes* **4**, 114–123.

ALLEN, H.L. ET AL. (2011) GATA6 haploinsufficiency causes pancreatic agenesis in humans. *Nature genetics* **44**, 20–22.

ALMANZA, A. ET AL. (2019) Endoplasmic reticulum stress signalling – from basic mechanisms to clinical applications. *The FEBS Journal* **286**, 241–278.

ALMGREN, P. ET AL. (2011) Heritability and familiality of type 2 diabetes and related quantitative traits in the Botnia Study. *Diabetologia* **54**, 2811–2819.

ALTSHULER, D. ET AL. (2000) The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nature genetics* **26**, 76–80.

AMERICAN DIABETES ASSOCIATION (2020a) Standards of Medical Care in Diabetes—2020. *Diabetes Care* **43**.

AMERICAN DIABETES ASSOCIATION (2020b) Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes—2020. *Diabetes Care* **43**, S14 LP-S31.

ANDERSSON, L.E. ET AL. (2015) Characterization of stimulus-secretion coupling in the human pancreatic EndoC-βH1 beta cell line. *PLoS ONE* **10**, 1–18.

ANDRALI, S.S. ET AL. (2008) Glucose regulation of insulin gene expression in pancreatic beta-cells. *The Biochemical journal* **415**, 1–10.

ARAGÓN, F. ET AL. (2015) Pancreatic polypeptide regulates glucagon release through PPYR1 receptors expressed in mouse and human alpha-cells. *Biochimica et biophysica acta* **1850**, 343–351.

ARTNER, I. & STEIN, R. (2008) Transcriptional Regulation of Insulin Gene Expression. In *Pancreatic Beta Cell in Health and Disease* pp. 13–30.

ASFARI, M. ET AL. (1992) Establishment of 2-Mercaptoethaol-Dependent Differentiated Insulin-Secreting Cell Lines. *Endocrinology* **130**, 167–178.

ASHCROFT, F.M. ET AL. (1984) Glucose induces closure of single potassium channels in isolated rat pancreatic beta-cells. *Nature* **312**, 446–448.

ASHCROFT, F.M. & RORSMAN, P. (1989) Electrophysiology of the pancreatic beta-cell. *Progress in biophysics and molecular biology* **54**, 87–143.

ASPLUND, O. ET AL. (2019) Islet Gene View - a tool to facilitate islet research. *bioRxiv*, 435743.

ATANES, P. ET AL. (2018) Identifying Signalling Pathways Regulated by GPRC5B in beta-Cells by CRISPR-Cas9-Mediated Genome Editing. *Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology* **45**, 656–666.

AYLWARD, A. ET AL. (2018) Shared genetic risk contributes to type 1 and type 2 diabetes etiology. *Human Molecular Genetics*, ddy314.

BABENKO, A.P. ET AL. (2006) Activating mutations in the ABCC8 gene in neonatal diabetes mellitus. *The New England journal of medicine* **355**, 456–466.

BACK, S.H. & KAUFMAN, R.J. (2012) Endoplasmic reticulum stress and type 2 diabetes. *Annual review of biochemistry* **81**, 767–793.

BALBOA, D. ET AL. (2019a) Genome editing of human pancreatic beta cell models: problems, possibilities and outlook. *Diabetologia* **62**, 1329–1336.

BALBOA, D. ET AL. (2018) Insulin mutations impair beta-cell development in a patient-derived iPSC model of neonatal diabetes. *eLife* **7**, e38519.

BALBOA, D. ET AL. (2019b) Concise Review: Human Pluripotent Stem Cells for the Modeling of Pancreatic β-Cell Pathology. *STEM CELLS* **37**, 33–41.

BANSAL, V. ET AL. (2017) Spectrum of mutations in monogenic diabetes genes identified from high-throughput DNA sequencing of 6888 individuals. *BMC medicine* **15**, 213.

BARBITOFF, Y.A. ET AL. (2018) Identification of Novel Candidate Markers of Type 2 Diabetes and Obesity in Russia by Exome Sequencing with a Limited Sample Size. *Genes* **9**, 415.

BARON, M. ET AL. (2016) A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. *Cell Systems* **3**, 346-360.e4.

BARRETT, J.C. ET AL. (2009) Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nature genetics* **41**, 703–707.

BARROSO, I. & MCCARTHY, M.I. (2019) The Genetic Basis of Metabolic Disease. *Cell* **177**, 146–161.

BEHAN, F.M. ET AL. (2019) Prioritization of cancer therapeutic targets using CRISPR–Cas9 screens. *Nature* **568**, 511–516.

BENAZRA, M. ET AL. (2015) A human beta cell line with drug inducible excision of immortalizing transgenes. *Molecular Metabolism* **4**, 916–925.

BENNER, C. ET AL. (2014) The transcriptional landscape of mouse beta cells compared to human beta cells reveals notable species differences in long non-coding RNA and

protein-coding gene expression. *BMC Genomics* **15**, 620.

BENTO, J.L. ET AL. (2004) Association of protein tyrosine phosphatase 1B gene polymorphisms with type 2 diabetes. *Diabetes* **53**, 3007–3012.

VAN DEN BERG, J. ET AL. (2018) A limited number of double-strand DNA breaks is sufficient to delay cell cycle progression. *Nucleic acids research* **46**, 10132–10144.

BEUCHER, A. & CEBOLA, I. (2019) One-step dual CRISPR/Cas9 guide RNA cloning protocol. *Protocol Exchange*. Https://doi.org/10.21203/rs.2.1831/v1 [accessed 26 June 2019].

BLACKARD, W.G. & NELSON, N.C. (1970) Portal and peripheral vein immunoreactive insulin concentrations before and after glucose infusion. *Diabetes* **19**, 302–306.

BOCH, J. ET AL. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326**, 1509–1512.

BOJ, S.F. ET AL. (2012) Diabetes risk gene and Wnt effector Tcf7l2/TCF4 controls hepatic response to perinatal and adult metabolic demand. *Cell* **151**, 1595–1607.

BOLOTIN, A. ET AL. (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**, 2551–2561.

BONÀS-GUARCH, S. ET AL. (2018) Re-analysis of public genetic data reveals a rare X-chromosomal variant associated with type 2 diabetes. *Nature Communications* **9**, 321.

BOROWIEC, M. ET AL. (2009) Mutations at the BLK locus linked to maturity onset diabetes of the young and beta-cell dysfunction. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 14460–14465.

BOSCO, D. ET AL. (2010) Unique arrangement of α- and β-cells in human islets of Langerhans. *Diabetes* **59**, 1202–1210.

BOTTINO, R. ET AL. (2004) Response of Human Islets to Isolation Stress and the Effect of Antioxidant Treatment. *Diabetes* **53**, 2559 LP – 2568.

BOTTOMLEY, W.E. ET AL. (2009) IRS2 variants and syndromes of severe insulin resistance. *Diabetologia* **52**, 1208–1211.

BOUZAKRI, K. ET AL. (2008) Rab GTPase-Activating Protein AS160 Is a Major Downstream Effector of Protein Kinase B/Akt Signaling in Pancreatic β-Cells. *Diabetes* **57**, 1195 LP – 1204.

BOYLE, K.B. ET AL. (2019) CALCOCO2/NDP52 initiates selective autophagy through recruitment of ULK and TBK1 kinase complexes. *Autophagy* **15**, 1655–1656.

BRANDHORST, H. ET AL. (1998) Assessment of intracellular insulin content during all steps of human islet isolation procedure. *Cell transplantation* **7**, 489–495.

BRAUN, M. ET AL. (2008) Voltage-Gated Ion Channels in Human Pancreatic Beta-Cells: Electrophysiological Characterization and Role in Insulin Secretion. *Diabetes* **57**, 1618–1628.

BRINKMAN, E.K. ET AL. (2014) Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic acids research* **42**, e168.

BRISSOVA, M. ET AL. (2005) Assessment of Human Pancreatic Islet Architecture and

Composition by Laser Scanning Confocal Microscopy. *Journal of Histochemistry & Cytochemistry* **53**, 1087–1097.

BRISSOVA, M. ET AL. (2015) Human Islets Have Fewer Blood Vessels than Mouse Islets and the Density of Islet Vascular Structures Is Increased in Type 2 Diabetes. *Journal of Histochemistry and Cytochemistry* **63**, 637–645.

BROGLIO, F. ET AL. (2001) Ghrelin, a natural GH secretagogue produced by the stomach, induces hyperglycemia and reduces insulin secretion in humans. *The Journal of clinical endocrinology and metabolism* **86**, 5083–5086.

BROUNS, S.J.J. ET AL. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960–964.

VAN DE BUNT, M. ET AL. (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. *PLoS Genetics* **11**, 1–21.

BURTON, P. ET AL. (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678.

BUTLER, A.E. ET AL. (2010) Adaptive changes in pancreatic beta cell fractional area and beta cell turnover in human pregnancy. *Diabetologia* **53**, 2167–2176.

BUTLER, A.E. ET AL. (2003) ß-Cell Deficit and Increased ß-Cell Apoptosis in Humans With Type 2 Diabetes. *Diabetes* **52**, 102–110.

CABRERA, O. ET AL. (2006) The unique cytoarchitecture of human pancreatic islets has implications for islet cell function. *Proceedings of the National Academy of Sciences* **103**, 2334–2339.

CAMPBELL, R.K. ET AL. (1996) Metformin: A new oral biguanide. *Clinical Therapeutics* **18**, 360–371.

CANTLEY, J. ET AL. (2007) Pancreatic deletion of insulin receptor substrate 2 reduces beta and alpha cell mass and impairs glucose homeostasis in mice. *Diabetologia* **50**, 1248–1256.

CANVER, M.C. ET AL. (2014) Characterization of genomic deletion efficiency mediated by clustered regularly interspaced palindromic repeats (CRISPR)/cas9 nuclease system in mammalian cells. *Journal of Biological Chemistry* **289**, 21312–21324.

CAPECCHI, M.R. (1989) Altering the genome by homologous recombination. *Science* **244**, 1288–1292.

CARDENAS-DIAZ, F.L. ET AL. (2020) A Dual Reporter EndoC-βH1 Human β-Cell Line for Efficient Quantification of Calcium Flux and Insulin Secretion. *Endocrinology* **161**.

CARDENAS-DIAZ, F.L. ET AL. (2019) Modeling Monogenic Diabetes using Human ESCs Reveals Developmental and Metabolic Deficiencies Caused by Mutations in HNF1A. *Cell Stem Cell* **25**, 273–289.

CARRAT, G.R. ET AL. (2017) Decreased STARD10 Expression Is Associated with Defective Insulin Secretion in Humans and Mice. *Am J Hum Genet* **100**, 238–256.

CASELLAS, A. ET AL. (2015) Insulin-like growth factor 2 overexpression induces β-Cell dysfunction and increases beta-cell susceptibility to damage. *Journal of Biological*

Chemistry **290**, 16772–16785.

CENTERS FOR DISEASE CONTROL AND PREVENTION (2020) National Diabetes Statistics Report: Estimates of Diabetes and Its Burden in the United States, 2020.

CERASI, E. & LUFT, R. (1967) The plasma insulin response to glucose infusion in healthy subjects and in diabetes mellitus. *Acta endocrinologica* **55**, 278–304.

CHANDRA, V. ET AL. (2014) RFX6 Regulates Insulin Secretion by Modulating Ca2+ Homeostasis in Human beta Cells. *Cell Reports* **9**, 2206–2218.

CHEN, Y. ET AL. (2008) Variations in DNA elucidate molecular networks that cause disease. *Nature* **452**, 429–435.

CHEUNG, H.W. ET AL. (2011) Systematic investigation of genetic vulnerabilities across cancer cell lines reveals lineage-specific dependencies in ovarian cancer. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 12372–12377.

CHIMIENTI, F. ET AL. (2004) Identification and cloning of a beta-cell-specific zinc transporter, ZnT-8, localized into insulin secretory granules. *Diabetes* **53**, 2330–2337.

CHO, Y.S. ET AL. (2012) Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. *Nat Genet* **44**, 67–72.

CHUNG, W.K. ET AL. (2020) Precision Medicine in Diabetes: A Consensus Report From the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetes care* **43**, 1617–1635.

CLARK, I.E. ET AL. (2006) Drosophila pink1 is required for mitochondrial function and interacts genetically with parkin. *Nature* **441**, 1162–1166.

CLAUSSNITZER, M. ET AL. (2015) FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. *N Engl J Med* **373**, 895–907.

CNOP, M. ET AL. (2012) Endoplasmic reticulum stress, obesity and diabetes. *Trends in Molecular Medicine* **18**, 59–68.

CNOP, M. ET AL. (2010) Causes and cures for endoplasmic reticulum stress in lipotoxic beta-cell dysfunction. *Diabetes, obesity & metabolism* **12**, 76–82.

COHRS, C.M. ET AL. (2017) Vessel network architecture of adult human islets promotes distinct cell-cell interactions in situ and is altered after transplantation. *Endocrinology* **158**, 1373–1385.

COLL, A.P. ET AL. (2020) GDF15 mediates the effects of metformin on body weight and energy balance. *Nature* **578**, 444–448.

CONG, L. ET AL. (2013) Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339**, 819–824.

CORTEZ, J.T. ET AL. (2020) CRISPR screen in regulatory T cells reveals modulators of Foxp3. *Nature*.

CUDWORTH, A.G. & WOODROW, J.C. (1974) Letter: HL-A antigens and diabetes mellitus. *Lancet* **2**, 1153.

CURRY, D.L. ET AL. (1968) Dynamics of insulin secretion by the perfused rat pancreas.

Endocrinology **83**, 572–584.

CYPHERT, H.A. ET AL. (2019) Examining how the MAFB transcription factor affects islet β-Cell function postnatally. *Diabetes* **68**, 337–348.

D'AMOUR, K.A. ET AL. (2006) Production of pancreatic hormone-expressing endocrine cells from human embryonic stem cells. *Nature biotechnology* **24**, 1392–1401.

DADI, P.K. ET AL. (2014) Pancreatic β-cell-specific ablation of TASK-1 channels augments glucose-stimulated calcium entry and insulin secretion, improving glucose tolerance. *Endocrinology* **155**, 3757–3768.

DASH, S. ET AL. (2010) Analysis of TBC1D4 in patients with severe insulin resistance. *Diabetologia* **53**, 1239–1242.

DASH, S. ET AL. (2009) A truncation mutation in TBC1D4 in a family with acanthosis nigricans and postprandial hyperinsulinemia. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 9350–9355.

DATLINGER, P. ET AL. (2017) Pooled CRISPR screening with single-cell transcriptome readout. *Nature Methods* **14**, 297–301.

DEEB, S.S. ET AL. (1998) A Pro12Ala substitution in PPARgamma2 associated with decreased receptor activity, lower body mass index and improved insulin sensitivity. *Nature genetics* **20**, 284–287.

DEJESUS, R. ET AL. (2016) Functional CRISPR screening identifies the ufmylation pathway as a regulator of SQSTM1/p62. *eLife* **5**, 1–16.

DELEPINE, M. ET AL. (2000) EIF2AK3, encoding translation initiation factor 2-alpha kinase 3, is mutated in patients with Wolcott-Rallison syndrome. *Nature genetics* **25**, 406–409.

DELMEIRE, D. ET AL. (2003) Type VIII adenylyl cyclase in rat beta cells: coincidence signal detector/generator for glucose and GLP-1. *Diabetologia* **46**, 1383–1393.

DEMPSTER, J.M. ET AL. (2019a) Agreement between two large pan-cancer CRISPR-Cas9 gene dependency data sets. *Nature Communications* **10**, 5817.

DEMPSTER, J.M. ET AL. (2019b) Extracting Biological Insights from the Project Achilles Genome-Scale CRISPR Screens in Cancer Cell Lines. *bioRxiv*, 720243.

DIMAS, A.S. ET AL. (2014) Impact of Type 2 Diabetes Susceptibility Variants on Quantitative Glycemic Traits Reveals Mechanistic Heterogeneity. *Diabetes* **63**, 2158–2171.

DOENCH, J.G. (2018) Am i ready for CRISPR? A user's guide to genetic screens. *Nature Reviews Genetics* **19**, 67–80.

DOENCH, J.G. ET AL. (2016) Optimized sgRNA design to maximize activity and minimize off- target effects of CRISPR-Cas9. *Nat Biotechnol* **34**, 184–191.

DOENCH, J.G. ET AL. (2003) siRNAs can function as miRNAs. *Genes and Development* **17**, 438–442.

DOYLE, M.J. & SUSSEL, L. (2007) Nkx2.2 regulates beta-cell function in the mature islet. *Diabetes* **56**, 1999–2007.

DRUCKER, D.J. ET AL. (2017) Discovery, characterization, and clinical development of the

glucagon-like peptides. *The Journal of clinical investigation* **127**, 4217–4227.

DUFRESNE, J. ET AL. (2018) The plasma peptides of ovarian cancer. *Clinical proteomics* **15**, 41.

DUGGIRALA, R. ET AL. (1999) Linkage of type 2 diabetes mellitus and of age at onset to a genetic location on chromosome 10q in Mexican Americans. *American journal of human genetics* **64**, 1127–1140.

DUPUIS, J. ET AL. (2010) New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nature Genetics* **42**, 408.

DWIVEDI, O.P. ET AL. (2019) Loss of ZnT8 function protects against diabetes by enhanced insulin secretion. *Nature Genetics* **51**, 1596–1606.

EFRAT, S. ET AL. (1988) Beta-cell lines derived from transgenic mice expressing a hybrid insulin gene-oncogene. *Proceedings of the National Academy of Sciences* **85**, 9037–9041.

EGIDO, E.M. ET AL. (2002) Inhibitory effect of ghrelin on insulin and pancreatic somatostatin secretion. *European journal of endocrinology* **146**, 241–244.

EL-BROLOSY, M. ET AL. (2019) Genetic compensation is triggered by mutant mRNA degradation. *Nature* **568**, 193–197.

ELBASHIR, S.M. ET AL. (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494–498.

EMILSSON, V. ET AL. (2018) Co-regulatory networks of human serum proteins link genetics to disease. *Science* **361**, 1–61.

ENYEDI, P. & CZIRJÁK, G. (2010) Molecular Background of Leak K+ Currents: Two-Pore Domain Potassium Channels. *Physiological Reviews* **90**, 559–605.

ERLICH, H. ET AL. (2008) HLA DR-DQ haplotypes and genotypes and type 1 diabetes risk: analysis of the type 1 diabetes genetics consortium families. *Diabetes* **57**, 1084–1092.

EVANGELOU, M. ET AL. (2014) A method for gene-based pathway analysis using genomewide association study summary statistics reveals nine new type 1 diabetes associations. *Genetic epidemiology* **38**, 661–670.

FADISTA, J. ET AL. (2014) Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism. *Proc Natl Acad Sci U S A* **111**, 13924–13929.

FANG, Z. ET AL. (2019) Single-Cell Heterogeneity Analysis and CRISPR Screen Identify Key β-Cell-Specific Disease Genes. *Cell Reports* **26**, 3132-3144.e7.

FAROOQI, I.S. ET AL. (2000) Dominant and recessive inheritance of morbid obesity associated with melanocortin 4 receptor deficiency. *The Journal of clinical investigation* **106**, 271–279.

FERNANDEZ-RUIZ, R. ET AL. (2014) Protein Tyrosine Phosphatase-1B Modulates Pancreatic β-cell Mass. *PLOS ONE* **9**, e90344.

FERNÁNDEZ-TAJES, J. ET AL. (2019) Developing a network view of type 2 diabetes risk pathways through integration of genetic, genomic and functional data. *Genome*

Medicine **11**, 19.

FERRER, J. ET AL. (1995) Pancreatic islet GLUT2 glucose transporter mRNA and protein expression in humans with and without NIDDM. *Diabetes* **44**, 1369–1374.

FEX, M. ET AL. (2018) The pathogenetic role of β-cell mitochondria in type 2 diabetes. *Journal of Endocrinology* **236**, R145–R159.

FIASCHI-TAESCH, N.M. ET AL. (2013) Human Pancreatic β-Cell G1/S Molecule Cell Cycle Atlas. *Diabetes* **62**, 2450–2459.

FIGLIOZZI, R.W. ET AL. (2016) Using the inverse Poisson distribution to calculate multiplicity of infection and viral replication by a high-throughput fluorescent imaging system. *Virologica Sinica* **31**, 180–183.

FINER, S. ET AL. (2020) Cohort Profile: East London Genes & Health (ELGH), a community-based population genomics and health study in British Bangladeshi and British Pakistani people. *International Journal of Epidemiology* **49**, 20-21i.

FITIPALDI, H. ET AL. (2018) A global overview of precision medicine in type 2 diabetes. *Diabetes* **67**, 1911–1922.

FLANAGAN, S.E. ET AL. (2014) Analysis of transcription factors key for mouse pancreatic development establishes NKX2-2 and MNX1 mutations as causes of neonatal diabetes in man. *Cell metabolism* **19**, 146–154.

FLANNICK, J. & FLOREZ, J.C. (2016) Type 2 diabetes: genetic data sharing to advance complex disease research. *Nature Reviews Genetics* **17**, 535–549.

FLANNICK, J. ET AL. (2019) Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls. *Nature* **570**, 71–76.

FLANNICK, J. ET AL. (2014) Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. *Nat Genet* **46**, 357–363.

FOGARTY, M.P. ET AL. (2013) Allele-Specific Transcriptional Activity at Type 2 Diabetes – Associated Single Nucleotide Polymorphisms in Regions of Pancreatic Islet Open Chromatin at the JAZF1 Locus. *Diabetes* **62**, 1756–1762.

FORTUNE, M.D. ET AL. (2015) Statistical colocalization of genetic risk variants for related autoimmune diseases in the context of common controls. *Nature genetics* **47**, 839–846.

FRANKS, P.W. (2012) The Complex Interplay of Genetic and Lifestyle Risk Factors in Type 2 Diabetes: An Overview. *Scientifica* **2012**, 1–11.

FRANKS, P.W. & MCCARTHY, M.I. (2016) Exposing the exposures responsible for type 2 diabetes and obesity. *Science* **354**, 69–73.

FRAYLING, T.M. ET AL. (2007) A Common Variant in the FTO Gene Is Associated with Body Mass Index and Predisposes to Childhood and Adult Obesity. *Science* **316**, 889–894.

FREYCHET, L. ET AL. (1988) Effect of intranasal glucagon on blood glucose levels in healthy subjects and hypoglycaemic patients with insulin-dependent diabetes. *Lancet* **1**, 1364–1366.

FROGUEL, P. ET AL. (1992) Close linkage of glucokinase locus on chromosome 7p to early-

onset non-insulin-dependent diabetes mellitus. *Nature* **356**, 162–164.

FUCHSBERGER, C. ET AL. (2016) The genetic architecture of type 2 diabetes. *Nature* **536**, 41–47.

FURUYA, N. ET AL. (2018) NDP52 interacts with mitochondrial RNA poly(A) polymerase to promote mitophagy. *EMBO reports* **19**, e46363.

GABRIEL, S.B. ET AL. (2002) The Structure of Haplotype Blocks in the Human Genome. *Science* **296**, 2225 LP – 2229.

GAISANO, H.Y. (2017) Recent new insights into the role of SNARE and associated proteins in insulin granule exocytosis. *Diabetes, obesity & metabolism* **19**, 115–123.

GANDASI, N.R. & BARG, S. (2014) Contact-induced clustering of syntaxin and munc18 docks secretory granules at the exocytosis site. *Nature communications* **5**, 3914.

GARNEAU, J.E. ET AL. (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67–71.

GASIUNAS, G. ET AL. (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences* **109**, E2579–E2586.

GAULTON, K.J. (2017) Mechanisms of Type 2 Diabetes Risk Loci. *Current Diabetes Reports* **17**.

GAULTON, K.J. ET AL. (2015) Genetic fine-mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nature Genetics* **47**, 1415–1425.

GAULTON, K.J. ET AL. (2010) A map of open chromatin in human pancreatic islets. *Nature genetics* **42**, 255–259.

GEISLER, S. ET AL. (2010) PINK1/Parkin-mediated mitophagy is dependent on VDAC1 and p62/SQSTM1. *Nature cell biology* **12**, 119–131.

GEMBAL, M. ET AL. (1993) Mechanisms by which glucose can control insulin release independently from its action on adenosine triphosphate-sensitive K+ channels in mouse B cells. *The Journal of clinical investigation* **91**, 871–880.

GERICH, J.E. ET AL. (1974) Characterization of the Glucagon Response to Hypoglycemia in Man. *The Journal of Clinical Endocrinology & Metabolism* **38**, 77–82.

GHARAVY, S.N.M. ET AL. (2020) Sexually dimorphic roles for the type 2 diabetes-associated C2cd4b gene in murine glucose homeostasis. *bioRxiv*, 2020.05.18.099200.

GHOSH, R. ET AL. (2019) Endoplasmic reticulum stress, degeneration of pancreatic islet β-cells, and therapeutic modulation of the unfolded protein response in diabetes. *Molecular metabolism* **27S**, S60–S68.

GILBERT, L.A. ET AL. (2014) Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* **159**, 647–661.

GILBERT, L.A. ET AL. (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442–451.

GLICK, D. ET AL. (2010) Autophagy: cellular and molecular mechanisms. *The Journal of pathology* **221**, 3–12.

GLOYN, A.L. ET AL. (2004) Activating Mutations in the Gene Encoding the ATP-Sensitive Potassium-Channel Subunit Kir6.2 and Permanent Neonatal Diabetes. *New England Journal of Medicine* **350**, 1838–1849.

GLOYN, A.L. ET AL. (2003) Large-scale association studies of variants in genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes. *Diabetes* **52**, 568–572.

GOMEZ, J.A. & THOMAS RUTKOWSKI, D. (2016) Experimental reconstitution of chronic ER stress in the liver reveals feedback suppression of BiP mRNA expression. *eLife* **5**, e20390.

GÖPEL, S. ET AL. (1999) Voltage-gated and resting membrane currents recorded from B-cells in intact mouse pancreatic islets. *The Journal of physiology* **521**, 717–728.

GÖPEL, S.O. ET AL. (2000) Patch-clamp characterisation of somatostatin-secreting -cells in intact mouse pancreatic islets. *The Journal of physiology* **528**, 497–507.

GRANT, S.F.A. ET AL. (2006) Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nature genetics* **38**, 320–323.

GREELEY, S.A.W. ET AL. (2011) Neonatal diabetes: an expanding list of genes allows for improved diagnosis and treatment. *Current diabetes reports* **11**, 519–532.

GREENWALD, W.W. ET AL. (2019) Pancreatic islet chromatin accessibility and conformation reveals distal enhancer networks of type 2 diabetes risk. *Nature Communications* **10**, 2078.

GRESSET, A. ET AL. (2012) The phospholipase C isozymes and their regulation. *Sub-cellular biochemistry* **58**, 61–94.

GROTZ, A.K. ET AL. (2017) Prioritising Causal Genes at Type 2 Diabetes Risk Loci. *Curr Diab Rep* **17**.

GTEX CONSORTIUM (2015) The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648–660.

GTEX CONSORTIUM (2017) Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213.

GU, C. ET AL. (2010) Pancreatic β cells require NeuroD to achieve and maintain functional maturity. *Cell Metabolism* **11**, 298–310.

GURGUL-CONVEY, E. ET AL. (2015) Physiological characterization of the human EndoC-βH1 β-cell line. *Biochemical and Biophysical Research Communications* **464**, 13–19.

GUSEV, A. ET AL. (2014) Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *American journal of human genetics* **95**, 535–552.

HAAPANIEMI, E. ET AL. (2018) CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nature medicine* **24**, 927–930.

HAATAJA, L. ET AL. (2008) Islet amyloid in type 2 diabetes, and the toxic oligomer hypothesis. *Endocrine reviews* **29**, 303–316.

HAEUSSLER, M. ET AL. (2016) Evaluation of off-target and on-target scoring algorithms and

integration into the guide RNA selection tool CRISPOR. *Genome Biology* **17**, 148.

HALBAN, P.A. & WOLLHEIM, C.B. (1980) Intracellular degradation of insulin stores by rat pancreatic islets in vitro. An alternative pathway for homeostasis of pancreatic insulin content. *The Journal of biological chemistry* **255**, 6003–6006.

HANIS, C.L. ET AL. (1996) A genome-wide search for human non-insulin-dependent (type 2) diabetes genes reveals a major susceptibility locus on chromosome 2. *Nature genetics* **13**, 161–166.

HANNA, R.E. & DOENCH, J.G. (2020) Design and analysis of CRISPR–Cas experiments. *Nature Biotechnology*.

HARDING, H.P. ET AL. (2000) Perk is essential for translational regulation and cell survival during the unfolded protein response. *Molecular cell* **5**, 897–904.

HARPER, M.E. ET AL. (1981) Localization of the human insulin gene to the distal end of the short arm of chromosome 11. *Proceedings of the National Academy of Sciences of the United States of America* **78**, 4458–4460.

HARRIS, P.K. & KLETZIEN, R.F. (1994) Localization of a pioglitazone response element in the adipocyte fatty acid-binding protein gene. *Molecular Pharmacology* **45**, 439 LP – 445.

HART, N.J. & POWERS, A.C. (2019) Use of human islets to understand islet biology and diabetes: progress, challenges and suggestions. *Diabetologia* **62**, 212–222.

HART, T. ET AL. (2015a) High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities Screens Reveal Fitness Genes. *Cell* **163**, 1515–1526.

HART, T. ET AL. (2015b) High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell* **163**, 1515–1526.

HART, T. ET AL. (2017) Evaluation and Design of Genome-Wide CRISPR/SpCas9 Knockout Screens. *Genes Genomes Genetics* **7**, 2719–2727.

HARTLEY, T. ET AL. (2009) Emerging roles for the ubiquitin-proteasome system and autophagy in pancreatic β-cells. *American Journal of Physiology-Endocrinology and Metabolism* **296**, E1–E10.

HASTOY, B. ET AL. (2018) Electrophysiological properties of human beta-cell lines EndoC-βH1 and -βH2 conform with human beta-cells. *Scientific Reports* **8**, 1–16.

HATTERSLEY, A.T. ET AL. (2018) ISPAD Clinical Practice Consensus Guidelines 2018: The diagnosis and management of monogenic diabetes in children and adolescents. *Pediatric Diabetes* **19**, 47–63.

HATTERSLEY, A.T. ET AL. (1992) Linkage of type 2 diabetes to the glucokinase gene. *Lancet* **339**, 1307–1310.

HAUGE-EVANS, A.C. ET AL. (2009) Somatostatin Secreted by Islet δ-Cells Fulfills Multiple Roles as a Paracrine Regulator of Islet Function. *Diabetes* **58**, 403 LP – 411.

HAZE, K. ET AL. (1999) Mammalian transcription factor ATF6 is synthesized as a transmembrane protein and activated by proteolysis in response to endoplasmic reticulum stress. *Molecular biology of the cell* **10**, 3787–3799.

HENNIGE, A.M. ET AL. (2003) Upregulation of insulin receptor substrate-2 in pancreatic β cells prevents diabetes. *The Journal of Clinical Investigation* **112**, 1521–1532.

HENQUIN, J.-C. (2011) The dual control of insulin secretion by glucose involves triggering and amplifying pathways in β-cells. *Diabetes Research and Clinical Practice* **93**, S27–S31.

HENQUIN, J.-C. ET AL. (2015) Dynamics of glucose-induced insulin secretion in normal human islets. *American Journal of Physiology-Endocrinology and Metabolism* **309**, E640–E650.

HENQUIN, J.-C. ET AL. (2017) Insulin, glucagon and somatostatin stores in the pancreas of subjects with type-2 diabetes and their lean and obese non-diabetic controls. *Scientific reports* **7**, 11015.

HENQUIN, J.-C. ET AL. (2012) Disruption and stabilization of β-cell actin microfilaments differently influence insulin secretion triggered by intracellular Ca2+ mobilization or store-operated Ca2+ entry. *FEBS letters* **586**, 89–95.

HENQUIN, J.C. (2000) Triggering and amplifying pathways of regulation of insulin secretion by glucose. *Diabetes* **49**, 1751–1760.

HEO, J.-M. ET AL. (2015) The PINK1-PARKIN Mitochondrial Ubiquitylation Pathway Drives a Program of OPTN/NDP52 Recruitment and TBK1 Activation to Promote Mitophagy. *Molecular Cell* **60**, 7–20.

HORIKAWA, Y. ET AL. (1997) Mutation in hepatocyte nuclear factor-1 beta gene (TCF2) associated with MODY. *Nature genetics* **17**, 384–385.

HORLBECK, M.A. ET AL. (2016) Nucleosomes impede Cas9 access to DNA in vivo and in vitro. *eLife* **5**, e12677.

HOSHINO, A. ET AL. (2014) Inhibition of p53 preserves Parkin-mediated mitophagy and pancreatic β-cell function in diabetes. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 3116–3121.

HSIAU, T. ET AL. (2019) Inference of CRISPR Edits from Sanger Trace Data. *bioRxiv*, 251082.

HU, M. ET AL. (2020) Chromatin 3D interaction analysis of the STARD10 locus unveils FCHSD2 as a new regulator of insulin secretion. *bioRxiv*, 2020.03.31.017707.

HUANG, D.W. ET AL. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**, 44–57.

HUANG, X.F. & ARVAN, P. (1995) Intracellular transport of proinsulin in pancreatic beta-cells. Structural maturation probed by disulfide accessibility. *The Journal of biological chemistry* **270**, 20417–20423.

HUTTON, J.C. (1994) Insulin secretory granule biogenesis and the proinsulin-processing endopeptidases. *Diabetologia* **37 Suppl 2**, S48-56.

HUYGHE, J.R. ET AL. (2013) Exome array analysis identifies novel loci and low-frequency variants for insulin processing and secretion. *Nat Genet* **45**, 197–201.

HWANG, H.-J. ET AL. (2019) Phospholipase C-β1 potentiates glucose-stimulated insulin secretion. *FASEB journal : official publication of the Federation of American Societies*

for Experimental Biology **33**, 10668–10679.

IHM, S.-H. ET AL. (2006) Effect of Donor Age on Function of Isolated Human Islets. *Diabetes* **55**, 1361 LP – 1368.

IHRY, R.J. ET AL. (2018) p53 inhibits CRISPR–Cas9 engineering in human pluripotent stem cells. *Nature Medicine* **24**, 939–946.

IMAI, Y. ET AL. (2020) Connecting pancreatic islet lipid metabolism with insulin secretion and the development of type 2 diabetes. *Annals of the New York Academy of Sciences* **1461**, 53–72.

IMMING, P. ET AL. (2006) Drugs, their targets and the nature and number of drug targets. *Nature reviews. Drug discovery* **5**, 821–834.

INGELSSON, E. ET AL. (2010) Detailed Physiologic Characterization Reveals Diverse Mechanisms for Novel Genetic Loci Regulating Glucose and Insulin Metabolism in Humans. *Diabetes* **59**.

INOUE, H. ET AL. (1998) A gene encoding a transmembrane protein is mutated in patients with diabetes mellitus and optic atrophy (Wolfram syndrome). *Nature genetics* **20**, 143–148.

INSHAW, J.R.J. ET AL. (2020) Divergent genetic effects for type 1 and type 2 diabetes at overlapping association signals. *bioRxiv*, 2020.06.17.156778.

INTERNATIONAL DIABETES FEDERATION (2019) *IDF Diabetes Atlas 2019*. In *International Diabetes Federation* p.

INZUCCHI, S.E. ET AL. (2015) Management of hyperglycemia in type 2 diabetes, 2015: a patient-centered approach: update to a position statement of the American Diabetes Association and the European Association for the Study of Diabetes. *Diabetes care* **38**, 140–149.

ITOH, N. & OKAMOTO, H. (1980) Translational control of proinsulin synthesis by glucose. *Nature* **283**, 100–102.

IZUMI, T. ET AL. (2003) Dominant negative pathogenesis by mutant proinsulin in the Akita diabetic mouse. *Diabetes* **52**, 409–416.

JACKSON, A.L. & LINSLEY, P.S. (2010) Recognizing and avoiding siRNA off-target effects for target identification and therapeutic application. *Nature Reviews Drug Discovery* **9**, 57–67.

JAIN, I.H. ET AL. (2016) Hypoxia as a therapy for mitochondrial disease. *Science* **352**, 54–61.

JINEK, M. ET AL. (2012) A Programmable Dual-RNA – Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816–822.

JITRAPAKDEE, S. ET AL. (2010) Regulation of insulin secretion: role of mitochondrial signalling. *Diabetologia* **53**, 1019–1032.

JOHNSON, G.C. ET AL. (2001) Haplotype tagging for the identification of common disease genes. *Nature genetics* **29**, 233–237.

JUNG, H.S. ET AL. (2008) Loss of Autophagy Diminishes Pancreatic β Cell Mass and

Function with Resultant Hyperglycemia. *Cell Metabolism* **8**, 318–324.

KAHN, S.E. ET AL. (2014) Pathophysiology and treatment of type 2 diabetes: Perspectives on the past, present, and future. *The Lancet* **383**, 1068–1083.

KAHN, S.E. ET AL. (1993) Quantification of the relationship between insulin sensitivity and beta-cell function in human subjects. Evidence for a hyperbolic function. *Diabetes* **42**, 1663–1672.

KAIHARA, K.A. ET AL. (2013) β-Cell–Specific Protein Kinase A Activation Enhances the Efficiency of Glucose Control by Increasing Acute-Phase Insulin Secretion. *Diabetes* **62**, 1527 LP – 1536.

KAPRIO, J. ET AL. (1992) Concordance for type 1 (insulin-dependent) and type 2 (non-insulin-dependent) diabetes mellitus in a population-based cohort of twins in Finland. *Diabetologia* **35**, 1060–1067.

KARAKAS, B. ET AL. (2007) p21 Gene Knock Down Does Not Identify Genetic Effectors Seen with Gene Knock Out. *Cancer Biol Ther* **6**, 1025–1030.

KEARNS, N.A. ET AL. (2015) Functional annotation of native enhancers with a Cas9-histone demethylase fusion. *Nature methods* **12**, 401–403.

KHAMIS, A. ET AL. (2019) Laser capture microdissection of human pancreatic islets reveals novel eQTLs associated with type 2 diabetes. *Molecular metabolism* **24**, 98–107.

KHAVANDI, K. ET AL. (2013) Strategies for preventing type 2 diabetes: an update for clinicians. *Therapeutic advances in chronic disease* **4**, 242–261.

KITIPHONGSPATTANA, K. ET AL. (2005) Proteasome inhibition alters glucose-stimulated (pro)insulin secretion and turnover in pancreatic beta-cells. *The Journal of biological chemistry* **280**, 15727–15734.

KLEINER, S. ET AL. (2018) Mice harboring the human SLC30A8 R138X loss-of-function mutation have increased insulin secretory capacity. *Proceedings of the National Academy of Sciences of the United States of America* **115**, E7642–E7649.

KLINGHOFFER, R.A. ET AL. (2010) Reduced seed region-based off-target activity with lentivirus-mediated RNAi. *Rna* **16**, 879–884.

KOHN, A.D. ET AL. (1996) Expression of a constitutively active Akt Ser/Thr kinase in 3T3-L1 adipocytes stimulates glucose uptake and glucose transporter 4 translocation. *The Journal of biological chemistry* **271**, 31372–31378.

KOIKE-YUSA, H. ET AL. (2014) Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nature Biotechnology* **32**, 267–273.

KOMOR, A.C. ET AL. (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424.

KONDAPALLI, C. ET AL. (2012) PINK1 is activated by mitochondrial membrane potential depolarization and stimulates Parkin E3 ligase activity by phosphorylating Serine 65. *Open Biology* **2**, 120080.

KONERMANN, S. ET AL. (2015) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583–588.

KOONER, J.S. ET AL. (2011) Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. *Nature genetics* **43**, 984–989.

KOSICKI, M. ET AL. (2018) Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nature Biotechnology* **36**, 765–771.

KOYANO, F. ET AL. (2014) Ubiquitin is phosphorylated by PINK1 to activate parkin. *Nature* **510**, 162–166.

KRISHNAN, K. ET AL. (2015) Calcium signaling in a genetically engineered human pancreatic beta-cell line. *Pancreas* **44**, 773–777.

KU, G.M. ET AL. (2012) An sirna screen in pancreatic beta cells reveals a role for Gpr27 in insulin production. *PLoS Genetics* **8**, 1–9.

KULKARNI, R.N. ET AL. (2012) Human β-cell proliferation and intracellular signaling: driving in the dark without a road map. *Diabetes* **61**, 2205–2213.

KULKARNI, R.N. & STEWART, A.F. (2014) Summary of the Keystone islet workshop (April 2014): the increasing demand for human islet availability in diabetes research. *Diabetes* **63**, 3979–3981.

KULZER, J.R. ET AL. (2014) A Common Functional Regulatory Variant at a Type 2 Diabetes Locus Upregulates ARAP1 Expression in the Pancreatic Beta Cell. *Am J Hum Genet* **94**, 186–197.

KUO, T. ET AL. (2019) Identification of C2CD4A as a human diabetes susceptibility gene with a role in β cell insulin secretion. *Proceedings of the National Academy of Sciences* **116**, 20033 LP – 20042.

KWEON, J. ET AL. (2020) A CRISPR-based base-editing screen for the functional assessment of BRCA1 variants. *Oncogene* **39**, 30–35.

KYCIA, I. ET AL. (2018) A Common Type 2 Diabetes Risk Variant Potentiates Activity of an Evolutionarily Conserved Islet Stretch Enhancer and Increases C2CD4A and C2CD4B Expression. *The American Journal of Human Genetics* **102**, 620–635.

LAPPALAINEN, T. ET AL. (2013) Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**, 506–511.

LATOURELLE, J.C. ET AL. (2009) Genomewide association study for onset age in Parkinson disease. *BMC medical genetics* **10**, 98.

LAWLOR, N. ET AL. (2017) Single-cell transcriptomes identify human islet cell signatures and reveal cell-type-specific expression changes in type 2 diabetes. *Genome Research* **27**, 208–222.

LAWLOR, N. ET AL. (2019) Multiomic Profiling Identifies cis-Regulatory Networks Underlying Human Pancreatic β Cell Identity and Function. *Cell Reports* **26**, 788-801.e6.

LAZAROU, M. ET AL. (2012) Role of PINK1 Binding to the TOM Complex and Alternate Intracellular Membranes in Recruitment and Activation of the E3 Ligase Parkin. *Developmental Cell* **22**, 320–333.

LAZAROU, M. ET AL. (2015) The ubiquitin kinase PINK1 recruits autophagy receptors to

induce mitophagy. *Nature* **524**, 309–314.

LEAN, M.E.J. ET AL. (2019) Durability of a primary care-led weight-management intervention for remission of type 2 diabetes: 2-year results of the DiRECT open-label, cluster-randomised trial. *The Lancet Diabetes & Endocrinology* **7**, 344–355.

LEE, J. ET AL. (2015) Proteomics analysis of rough endoplasmic reticulum in pancreatic beta cells. *Proteomics* **15**, 1508–1511.

LEE, K. ET AL. (2002) IRE1-mediated unconventional mRNA splicing and S2P-mediated ATF6 cleavage merge to regulate XBP1 in signaling the unfolded protein response. *Genes & development* **16**, 452–466.

LEMAIRE, K. ET AL. (2009) Insulin crystallization depends on zinc transporter ZnT8 expression, but is not required for normal glucose homeostasis in mice. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 14872–14877.

LI, B. ET AL. (2020) Genome-wide CRISPR screen identifies host dependency factors for influenza A virus infection. *Nature Communications* **11**, 164.

LI, M. ET AL. (2008) Evaluation of coverage variation of SNP chips for genome-wide association studies. *Eur J Hum Gen* **16**, 635–643.

LI, W. ET AL. (2015) Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR. *Genome Biology* **16**, 281.

LI, W. ET AL. (2014) MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome biology* **15**, 554.

LINO, C.A. ET AL. (2018) Delivering CRISPR: a review of the challenges and approaches. *Drug Delivery* **25**, 1234–1257.

LINS, P.E. ET AL. (1983) Minimal increases in glucagon levels enhance glucose production in man with partial hypoinsulinemia. *Diabetes* **32**, 633–636.

LITTLETON, J.T. ET AL. (2001) SNARE-complex disassembly by NSF follows synaptic-vesicle fusion. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 12233–12238.

LITWAK, L. ET AL. (2013) Prevalence of diabetes complications in people with type 2 diabetes mellitus and its association with baseline characteristics in the multinational A1chieve study. *Diabetology & metabolic syndrome* **5**, 57.

LIU, Y. ET AL. (2020) Very fast CRISPR on demand. *Science* **368**, 1265 LP – 1269.

LOHMUELLER, K.E. ET AL. (2013) Whole-exome sequencing of 2,000 Danish individuals and the role of rare coding variants in type 2 diabetes. *American journal of human genetics* **93**, 1072–1086.

LONGUET, C. ET AL. (2005) Extracellularly Regulated Kinases 1/2 (p44/42 Mitogen-Activated Protein Kinases) Phosphorylate Synapsin I and Regulate Insulin Secretion in the MIN6 β-Cell Line and Islets of Langerhans. *Endocrinology* **146**, 643–654.

LÓPEZ-AVALOS, M.D. ET AL. (2006) Evidence for a Role of the Ubiquitin-Proteasome Pathway in Pancreatic Islets. *Diabetes* **55**, 1223 LP – 1231.

LOTTA, L.A. ET AL. (2017) Integrative genomic analysis implicates limited peripheral

adipose storage capacity in the pathogenesis of human insulin resistance. *Nature genetics* **49**, 17–26.

MA, H. ET AL. (2015) A CRISPR-Based Screen Identifies Genes Essential for West-Nile-Virus-Induced Cell Death. *Cell reports* **12**, 673–683.

MA, Z. ET AL. (2019) PTC-bearing mRNA elicits a genetic compensation response via Upf3a and COMPASS components. *Nature* **568**, 259–263.

MAEDER, M.L. ET AL. (2013) CRISPR RNA-guided activation of endogenous human genes. *Nature methods* **10**, 977–979.

MAHAJAN, A. ET AL. (2014) Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nature genetics* **46**, 234–244.

MAHAJAN, A. ET AL. (2015) Identification and Functional Characterization of G6PC2 Coding Variants Influencing Glycemic Traits Define an Effector Transcript at the G6PC2-ABCB11 Locus. *PLoS Genetics* **11**, 1–25.

MAHAJAN, A. ET AL. (2018a) Fine-mapping of an expanded set of type 2 diabetes loci to single-variant resolution using high-density imputation and islet- specific epigenome maps. *Nature Genetics* **50**, 1505–1513.

MAHAJAN, A. ET AL. (2018b) Refining the accuracy of validated target identification through coding variant fine- mapping in type 2 diabetes. *Nature Genetics* **50**, 559–571.

MALECKI, M.T. ET AL. (1999) Mutations in NEUROD1 are associated with the development of type 2 diabetes mellitus. *Nature Genetics* **23**, 323–328.

MALI, P. ET AL. (2013) RNA-Guided Human Genome Engineering via Cas9. *Science* **339**, 823–826.

MANDUCHI, E. ET AL. (2018) A High Resolution Capture-C Promoter 'Interactome' Implicates Causal Genes at Type 2 Diabetes GWAS Loci. *Diabetes* **67**, 1705-P.

MARÍN-PEÑALVER, J.J. ET AL. (2016) Update on the treatment of type 2 diabetes mellitus. *World journal of diabetes* **7**, 354–395.

MARRAFFINI, L.A. & SONTHEIMER, E.J. (2010) Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* **463**, 568–571.

MARSO, S.P. ET AL. (2016) Liraglutide and Cardiovascular Outcomes in Type 2 Diabetes. *The New England journal of medicine* **375**, 311–322.

MATSCHINSKY, F.M. (1990) Glucokinase as glucose sensor and metabolic signal generator in pancreatic beta-cells and hepatocytes. *Diabetes* **39**, 647–652.

MATSCHINSKY, F.M. (2005) Glucokinase, glucose homeostasis, and diabetes mellitus. *Current diabetes reports* **5**, 171–176.

MATSUDA, T. ET AL. (2008) Reduced insulin signaling and endoplasmic reticulum stress act synergistically to deteriorate pancreatic beta cell function. *The Kobe journal of medical sciences* **54**, E114-21.

MATTIS, K.K. ET AL. (2019) 3-LB: The Type 2 Diabetes-Associated Transcription Factor RREB1 Affects Beta-Cell Function and Development. *Diabetes* **68**, 3-LB.

MAURANO, M.T. ET AL. (2012) Systematic Localization of Common Disease-Associated

Variation in Regulatory DNA. *Science* **337**, 1190–1195.

MCCARTHY, M.I. (2017) Painting a new picture of personalised medicine for diabetes. *Diabetologia* **60**, 793–799.

MCCARTHY, M.I. & ZEGGINI, E. (2009) Genome-wide association studies in type 2 diabetes. *Current diabetes reports* **9**, 164–171.

MCCULLOCH, L.J. ET AL. (2011) GLUT2 (SLC2A2) is not the principal glucose transporter in human pancreatic beta cells: Implications for understanding genetic association signals at this locus. *Molecular Genetics and Metabolism* **104**, 648–653.

MEGLASSON, M.D. ET AL. (1983) Chromatographic resolution and kinetic characterization of glucokinase from islets of Langerhans. *Proceedings of the National Academy of Sciences* **80**, 85 LP – 89.

MEIGS, J.B. ET AL. (2000) Parental transmission of type 2 diabetes: the Framingham Offspring Study. *Diabetes* **49**, 2201–2207.

MERCADER, J.M. ET AL. (2017) A Loss-of-Function Splice Acceptor Variant in IGF2 Is Protective for Type 2 Diabetes. *Diabetes* **66**, 2903–2914.

MERKLE, F.T. ET AL. (2017) Human pluripotent stem cells recurrently acquire and expand dominant negative P53 mutations. *Nature* **545**, 229–233.

MERRIMAN, C. ET AL. (2018) Highly specific monoclonal antibodies for allosteric inhibition and immunodetection of the human pancreatic zinc transporter ZnT8. *Journal of Biological Chemistry* **293**, 16206–16216.

MEUSSER, B. ET AL. (2005) ERAD: the long road to destruction. *Nature cell biology* **7**, 766–772.

MIGUEL-ESCALADA, I. ET AL. (2019) Human pancreatic islet three-dimensional chromatin architecture provides insights into the genetics of type 2 diabetes. *Nat Genet* **51**, 1137–1148.

MISRA, S. & OWEN, K.R. (2018) Genetics of Monogenic Diabetes: Present Clinical Challenges. *Current diabetes reports* **18**, 141.

MITCHELL, R.K. ET AL. (2015) Selective disruption of Tcf7l2 in the pancreatic β cell impairs secretory function and lowers β cell mass. *Human molecular genetics* **24**, 1390–1399.

MIYAZAKI, J.-I. ET AL. (1990) Establishment of a Pancreatic β Cell Line That Retains Glucose-Inducible Insulin Secretion: Special Reference to Expression of Glucose Transporter Isoforms. *Endocrinology* **127**, 126–132.

MODI, H. ET AL. (2015) Autocrine Action of IGF2 Regulates Adult β-Cell Mass and Function. *Diabetes* **64**, 4148–4157.

MOJICA, F.J.M. ET AL. (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of molecular evolution* **60**, 174–182.

MOLTKE, I. ET AL. (2014) A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes. *Nature* **512**, 190–193.

MOLVEN, A. ET AL. (2008) Mutations in the insulin gene can cause MODY and autoantibody-negative type 1 diabetes. *Diabetes* **57**, 1131–1135.

MOON, S. BIN ET AL. (2019) Recent advances in the CRISPR genome editing tool set. *Experimental & Molecular Medicine* **51**, 1–11.

MORGENS, D.W. ET AL. (2016) Systematic comparison of CRISPR/Cas9 and RNAi screens for essential genes. *Nat Biotechnol* **34**, 634–636.

MORRIS, A.D.P. ET AL. (2012) Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet* **44**, 981–990.

MORRIS, A.P. (2014) Fine Mapping of Type 2 Diabetes Susceptibility Loci. *Curr Diab Rep* **14**, 549.

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

NAUCK, M.A. ET AL. (1997) Glucagon-like peptide 1 (GLP-1) as a new therapeutic approach for Type 2-diabetes. *Exp Clin Endocrinol Diabetes* **105**, 187–195.

NAYA, F.J. ET AL. (1997) Diabetes, defective pancreatic morphogenesis, and abnormal enteroendocrine differentiation in BETA2/NeuroD-deficient mice. *Genes and Development* **11**, 2323–2334.

NAYLOR, J. ET AL. (2016) Use of CRISPR/Cas9-engineered INS-1 pancreatic beta cells to define the pharmacology of dual GIPR/GLP-1R agonists. *The Biochemical journal* **473**, 2881–2891.

NDIAYE, F.K. ET AL. (2017) Expression and functional assessment of candidate type 2 diabetes susceptibility genes identify four new genes contributing to human insulin secretion. *Molecular metabolism* **6**, 459–470.

NG, M.C.Y. ET AL. (2014) Meta-analysis of genome-wide association studies in African Americans provides insights into the genetic architecture of type 2 diabetes. *PLoS genetics* **10**, e1004517.

NG, N.H.J. (2016) The Role of Glucose-6-Phosphatase Catalytic Domain in Glucose Homeostasis.

NG, N.H.J. ET AL. (2019) Tissue-Specific Alteration of Metabolic Pathways Influences Glycemic Regulation. *bioRxiv*, 790618.

NG, P.C. & HENIKOFF, S. (2003) SIFT: Predicting amino acid changes that affect protein function. *Nucleic acids research* **31**, 3812–3814.

NICA, A.C. & DERMITZAKIS, E.T. (2008) Using gene expression to investigate the genetic basis of complex disorders. *Human molecular genetics* **17**, R129–R134.

NICA, A.C. & DERMITZAKIS, E.T. (2013) Expression quantitative trait loci : present and future. *Phil Trans R Soc B* **368**, 20120362.

NICA, A.C. ET AL. (2010) Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS genetics* **6**, e1000895–e1000895.

NICA, A.C. ET AL. (2013) Cell-type, allelic, and genetic signatures in the human pancreatic beta cell transcriptome. *Genome research* **23**, 1554–1562.

NICOLSON, T.J. ET AL. (2009) Insulin storage and glucose homeostasis in mice null for the granule zinc transporter ZnT8 and studies of the type 2 diabetes-associated variants.

Diabetes **58**, 2070–2083.

NIHONGAKI, Y. ET AL. (2015) CRISPR-Cas9-based photoactivatable transcription system. *Chemistry & biology* **22**, 169–174.

NJØLSTAD, P.R. ET AL. (2001) Neonatal diabetes mellitus due to complete glucokinase deficiency. *The New England journal of medicine* **344**, 1588–1592.

O'CONNELL, M.R. (2019) Molecular Mechanisms of RNA Targeting by Cas13-containing Type VI CRISPR-Cas Systems. *Journal of molecular biology* **431**, 66–87.

OKATSU, K. ET AL. (2015) Phosphorylated ubiquitin chain is the genuine Parkin receptor. *Journal of Cell Biology* **209**, 111–128.

ONENGUT-GUMUSCU, S. ET AL. (2015) Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nature genetics* **47**, 381–386.

PADDISON, P.J. ET AL. (2004) A resource for large-scale screens in mammals. *Nature* **428**, 427–431.

PANDIRI, A.R. (2014) Overview of exocrine pancreatic pathobiology. *Toxicologic pathology* **42**, 207–216.

PAPPALARDO, Z. ET AL. (2017) A whole genome RNA interference screen reveals a role for Spry2 in insulin transcription and the unfolded protein response. *Diabetes* **66**, 1703–1712.

PAREJA, A. ET AL. (1997) Unsaturated fatty acids alter the insulin secretion response of the islets of Langerhans in vitro. *Diabetes Research and Clinical Practice* **38**, 143–149.

PARK, J.S. ET AL. (2019) A FACS-Based Genome-wide CRISPR Screen Reveals a Requirement for COPI in Chlamydia trachomatis Invasion. *iScience* **11**, 71–84.

PARK, R.J. ET AL. (2017) A genome-wide CRISPR screen identifies a restricted set of HIV host dependency factors. *Nature genetics* **49**, 193–203.

PARKER, S.C.J. ET AL. (2013) Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proceedings of the National Academy of Sciences* **110**, 17921 LP – 17926.

PARNAUD, G. ET AL. (2008) Proliferation of sorted human and rat beta cells. *Diabetologia* **51**, 91–100.

PASQUALI, L. ET AL. (2014) Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nature Genetics* **46**, 136–143.

PATZELT, C. ET AL. (1978) Detection and kinetic behavior of preproinsulin in pancreatic islets. *Proceedings of the National Academy of Sciences of the United States of America* **75**, 1260–1264.

PEARSON, E.R. ET AL. (2006) Switching from Insulin to Oral Sulfonylureas in Patients with Diabetes Due to Kir6.2 Mutations. *New England Journal of Medicine* **355**, 467–477.

PEARSON, G. ET AL. (2018) Clec16a, Nrdp1, and USP8 Form a Ubiquitin-Dependent Tripartite Complex That Regulates β-Cell Mitophagy. *Diabetes* **67**, 265 LP – 277.

PERLEY, M.J. & KIPNIS, D.M. (1967) Plasma Insulin Responses to Oral and Intravenous

Glucose: Studies in Normal and Diabetic Subjects. *The Journal of Clinical Investigation* **46**, 1954–1962.

PETERS, J.-M. ET AL. (2008) The cohesin complex and its roles in chromosome biology. *Genes & Development* **22**, 3089–3114.

PICKLES, S. ET AL. (2018) Mitophagy and Quality Control Mechanisms in Mitochondrial Maintenance. *Current Biology* **28**, R170–R185.

POULSEN, P. ET AL. (1999) Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance--a population-based twin study. *Diabetologia* **42**, 139–145.

POUND, L.D. ET AL. (2013) G6PC2: A Negative Regulator of Basal Glucose-Stimulated Insulin Secretion. *Diabetes* **62**, 1547 LP – 1556.

POUND, L.D. ET AL. (2009) Deletion of the mouse Slc30a8 gene encoding zinc transporter-8 results in impaired insulin secretion. *The Biochemical journal* **421**, 371–376.

PROKS, P. ET AL. (2006) A heterozygous activating mutation in the sulphonylurea receptor SUR1 (ABCC8) causes neonatal diabetes. *Human molecular genetics* **15**, 1793–1800.

QI, L.S. ET AL. (2013) Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *CELL* **152**, 1173–1183.

RAHIER, J. ET AL. (2008) Pancreatic beta-cell mass in European subjects with type 2 diabetes. *Diabetes, obesity & metabolism* **10**, 32–42.

RAI, V. ET AL. (2020) Single-cell ATAC-Seq in human pancreatic islets and deep learning upscaling of rare cells reveals cell-specific type 2 diabetes regulatory signatures. *Molecular Metabolism* **32**, 109–121.

RAJAGOPAL, N. ET AL. (2016) High-throughput mapping of regulatory DNA. *Nature biotechnology* **34**, 167–174.

RAMZY, A. ET AL. (2020) Revisiting Proinsulin Processing: Evidence That Human β-Cells Process Proinsulin With Prohormone Convertase (PC) 1/3 But Not PC2. *Diabetes*, db190276.

RAVASSARD, P. ET AL. (2011) A genetically engineered human pancreatic β cell line exhibiting glucose-inducible insulin secretion. *The Journal of clinical investigation* **121**, 3589–3597.

RAVIER, M.A. ET AL. (2009) Glucose controls cytosolic Ca2+ and insulin secretion in mouse islets lacking adenosine triphosphate-sensitive K+ channels owing to a knockout of the pore-forming subunit Kir6.2. *Endocrinology* **150**, 33–45.

REBER, S. ET AL. (2017) CRISPR-Trap: a clean approach for the generation of gene knockouts and gene replacements in human cells. *Molecular Biology of the Cell* **29**, 75–83.

RENA, G. ET AL. (2017) The mechanisms of action of metformin. *Diabetologia* **60**, 1577–1585.

REZANIA, A. ET AL. (2014) Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. *Nature biotechnology* **32**, 1121–1133.

RIAHI, Y. ET AL. (2016) Autophagy is a major regulator of beta cell insulin homeostasis. *Diabetologia* **59**, 1480–1491.

RICHARDSON, R.B. ET AL. (2018) A CRISPR screen identifies IFI6 as an ER-resident interferon effector that blocks flavivirus replication. *Nature microbiology* **3**, 1214–1223.

RICHTER, B. ET AL. (2016) Phosphorylation of OPTN by TBK1 enhances its binding to Ub chains and promotes selective autophagy of damaged mitochondria. *Proceedings of the National Academy of Sciences* **113**, 4039 LP – 4044.

ROADMAP EPIGENOMICS CONSORTIUM ET AL. (2015) Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330.

ROBERTSON, C.C. ET AL. (2020) Fine-mapping, trans-ancestral and genomic analyses identify causal variants, cells, genes and drug targets for type 1 diabetes. *bioRxiv*, 2020.06.19.158071.

RÖDER, P. V. ET AL. (2016) Pancreatic regulation of glucose homeostasis. *Experimental & molecular medicine* **48**, e219.

RORSMAN, P. & ASHCROFT, F.M. (2018) Pancreatic β-Cell Electrical Activity and Insulin Secretion: Of Mice and Men. *Physiological reviews* **98**, 117–214.

RORSMAN, P. ET AL. (2012) Regulation of calcium in pancreatic alpha- and beta-cells in health and disease. *Cell calcium* **51**, 300–308.

ROSSI, A. ET AL. (2015) Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature* **524**, 230–233.

SAKAMOTO, K. & HOLMAN, G.D. (2008) Emerging role for AS160/TBC1D4 and TBC1D1 in the regulation of GLUT4 traffic. *American journal of physiology. Endocrinology and metabolism* **295**, E29-37.

SANJANA, N.E. ET AL. (2014) Improved vectors and genome-wide libraries for CRISPR screening. *Nature Methods* **11**, 783–784.

SANJANA, N.E. ET AL. (2016) High-resolution interrogation of functional elements in the noncoding genome. *Science* **353**, 1545–1549.

SANSBURY, F.H. ET AL. (2012) SLC2A2 mutations can cause neonatal diabetes, suggesting GLUT2 may have a role in human insulin secretion. *Diabetologia* **55**, 2381–2385.

SANTER, R. ET AL. (1997) Mutations in GLUT2, the gene for the liver-type glucose transporter, in patients with Fanconi-Bickel syndrome. *Nature genetics* **17**, 324–326.

SANTOS, R. ET AL. (2017) A comprehensive map of molecular drug targets. *Nature reviews. Drug discovery* **16**, 19–34.

SATO, Y. ET AL. (1992) Dual functional role of membrane depolarization/Ca2+ influx in rat pancreatic B-cell. *Diabetes* **41**, 438–443.

SAXENA, R. ET AL. (2007) Genome-Wide Association Analysis Identifies Loci for Type 2 Diabetes and Triglyceride Levels. *Science* **316**, 1331–1336.

SCHARFMANN, R.R. ET AL. (2014) Development of a conditionally immortalized human pancreatic β cell line. *The Journal of Clinical Investigation* **124**, 1–12.

SCHEEN, A.J. (2015) EMPA-REG OUTCOME: Empagliflozin reduces mortality in patients with type 2 diabetes at high cardiovascular risk. *Revue medicale de Liege* **70**, 583–589.

SCOTT, L.J. ET AL. (2016) The genetic regulatory signature of type 2 diabetes in human skeletal muscle. *Nature communications* **7**, 11764.

SCOTT, L.J. ET AL. (2007) A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316**, 1341–1345.

SCOTT, R.A. ET AL. (2017) An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. *Diabetes* **66**, 2888 LP – 2902.

SEGHERS, V. ET AL. (2000) Sur1 knockout mice. A model for K(ATP) channel-independent regulation of insulin secretion. *The Journal of biological chemistry* **275**, 9270–9277.

SHALEM, O. ET AL. (2014) Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science* **343**, 84–87.

SHALEM, O. ET AL. (2015) High-throughput functional genomics using CRISPR–Cas9. *Nature Reviews Genetics* **16**, 299–311.

SHAW-SMITH, C. ET AL. (2012) Recessive SLC19A2 mutations are a cause of neonatal diabetes mellitus in thiamine-responsive megaloblastic anaemia. *Pediatric Diabetes* **13**, 314–321.

SHIAO, M.S. ET AL. (2008) Adaptive evolution of the insulin two-gene system in mouse. *Genetics* **178**, 1683–1691.

SHIGETO, M. ET AL. (2015) GLP-1 stimulates insulin secretion by PKC-dependent TRPM4 and TRPM5 activation. *The Journal of clinical investigation* **125**, 4714–4728.

SHIH, H.P. ET AL. (2013) Pancreas organogenesis: from lineage determination to morphogenesis. *Annual review of cell and developmental biology* **29**, 81–105.

DA SILVA XAVIER, G. ET AL. (2009) TCF7L2 regulates late events in insulin secretion from pancreatic islet beta-cells. *Diabetes* **58**, 894–905.

SKELIN KLEMEN, M. ET AL. (2017) The triggering pathway to insulin secretion: Functional similarities and differences between the human and the mouse β cells and their translational relevance. *Islets* **9**, 109–139.

SLADEK, R. ET AL. (2007) A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **445**, 881–885.

SLATKIN, M. (2008) Linkage disequilibrium - Understanding the evolutionary past and mapping the medical future. *Nat Rev Genet* **9**, 477–485.

SMALL, K.S. ET AL. (2011) Identification of an imprinted master trans-regulator at the KLF14 locus related to multiple metabolic phenotypes. *Nat Genet* **43**, 561–564.

SMALL, K.S. ET AL. (2018) Regulatory variants at KLF14 influence type 2 diabetes risk via a female-specific effect on adipocyte size and body composition. *Nature genetics* **50**, 572–580.

SMITH, G.D. (2010) Mendelian Randomization for Strengthening Causal Inference in Observational Studies: Application to Gene × Environment Interactions. *Perspectives on Psychological Science* **5**, 527–545.

SOLEIMANPOUR, S.A. ET AL. (2014) The diabetes susceptibility gene Clec16a regulates mitophagy. *Cell* **157**, 1577–1590.

SPRACKLEN, C.N. ET AL. (2020) Identification of type 2 diabetes loci in 433,540 East Asian individuals. *Nature*.

STANCAKOVA, A. ET AL. (2009) Changes in insulin sensitivity and insulin release in relation to glycemia and glucose tolerance in 6,414 Finnish men. *Diabetes* **58**, 1212–1221.

STEINTHORSDOTTIR, V. ET AL. (2014) Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. *Nat Genet* **46**, 294–298.

STENEBERG, P. ET AL. (2013) The type 2 diabetes-associated gene Ide is required for insulin secretion and suppression of α-synuclein levels in β-cells. *Diabetes* **62**, 2004–2014.

STITZEL, M.L. ET AL. (2010) Global epigenomic analysis of primary human pancreatic islets provides insights into type 2 diabetes susceptibility loci. *Cell metabolism* **12**, 443–455.

STOY, J. ET AL. (2007) Insulin gene mutations as a cause of permanent neonatal diabetes. *Proceedings of the National Academy of Sciences* **104**, 15040–15044.

SUBRAMANIAN, A. ET AL. (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* **102**, 15545 LP – 15550.

SUDLOW, C. ET AL. (2015) UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine* **12**, e1001779.

SURIBEN, R. ET AL. (2015) β-Cell Insulin Secretion Requires the Ubiquitin Ligase COP1. *Cell* **163**, 1457–1467.

SUSSEL, L. ET AL. (1998) Mice lacking the homeodomain transcription factor Nkx2.2 have diabetes due to arrested differentiation of pancreatic beta cells. *Development* **125**, 2213–2221.

SZKLARCZYK, D. ET AL. (2019) STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research* **47**, D607–D613.

TAGER, H. ET AL. (1979) A structurally abnormal insulin causing human diabetes. *Nature* **281**, 122–125.

TAGUCHI, Y. ET AL. (2016) Sphingosine-1-phosphate Phosphatase 2 Regulates Pancreatic Islet β-Cell Endoplasmic Reticulum Stress and Proliferation. *The Journal of biological chemistry* **291**, 12029–12038.

TAYLOR, B.L. ET AL. (2013) Nkx6.1 is essential for maintaining the functional state of pancreatic beta cells. *Cell reports* **4**, 1262–1275.

TERAOKU, H. & LENZEN, S. (2017) Dynamics of Insulin Secretion from EndoC-*β*H1 *β*-Cell Pseudoislets in Response to Glucose and Other Nutrient and Nonnutrient Secretagogues. *Journal of Diabetes Research* **2017**, 2309630.

THE 1000 GENOMES PROJECT CONSORTIUM (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65.

THE ENCODE PROJECT CONSORTIUM (2012) An Integrated Encyclopedia of DNA Elements in the Human Genome. *Nature* **489**, 57–74.

THE INTERNATIONAL HAPMAP CONSORTIUM (2005) A haplotype map of the human genome. *Nature* **437**, 1299–1320.

THE INTERNATIONAL HAPMAP CONSORTIUM (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861.

THOMSEN, S.K. ET AL. (2016a) Systematic Functional Characterization of Candidate Causal Genes for Type 2 Diabetes Risk Variants. *Diabetes* **65**, 3805–3811.

THOMSEN, S.K. ET AL. (2016b) The importance of Context : Uncovering Species- and Tissue-Specific Effects of Genetic Risk Variants for Type 2 Diabetes. *Front Endocrinol* **7**.

THOMSEN, S.K. ET AL. (2018) Type 2 diabetes risk alleles in PAM impact insulin release from human pancreatic β-cells. *Nature Genetics* **50**, 1122–1131.

THOMSON, K.L. ET AL. (2003) Identification of 21 novel glucokinase (GCK) mutations in UK and European Caucasians with maturity-onset diabetes of the young (MODY). *Human mutation* **22**, 417.

THORE, S. ET AL. (2005) Feedback activation of phospholipase C via intracellular mobilization and store-operated influx of Ca²⁺ in insulin-secreting β-cells. *Journal of Cell Science* **118**, 4463 LP – 4471.

THUL, P.J. ET AL. (2017) A subcellular map of the human proteome. *Science* **356**, eaal3321.

THURNER, M. ET AL. (2018) Integration of human pancreatic islet genomic data refines regulatory mechanisms at Type 2 Diabetes susceptibility loci. *eLife* **7**, e31977.

TILLIL, H. & KOBBERLING, J. (1987) Age-corrected empirical genetic risk estimates for first-degree relatives of IDDM patients. *Diabetes* **36**, 93–99.

TODD, J.A. ET AL. (2007) Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nature genetics* **39**, 857–864.

TOKARZ, V.L. ET AL. (2018) The cell biology of systemic insulin function. *Journal of Cell Biology* **217**, 2273–2289.

TSHERNIAK, A. ET AL. (2017) Defining a Cancer Dependency Map. *Cell* **170**, 564-576.e16.

TSONKOVA, V.G. ET AL. (2018) The EndoC-bH1 cell line is a valid model of human beta cells and applicable for screenings to identify novel drug target candidates. *Molecular Metabolism* **8**, 144–157.

TSUDA, N. ET AL. (2017) A novel free fatty acid receptor 1 (GPR40/FFAR1) agonist, MR1704, enhances glucose-dependent insulin secretion and improves glucose homeostasis in rats. *Pharmacology research & perspectives* **5**, e00340.

UDLER, M.S. ET AL. (2018) Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: A soft clustering analysis. *PLoS medicine* **15**, e1002654.

UNGER, R.H. (1971) Glucagon Physiology and Pathophysiology. *New England Journal of Medicine* **285**, 443–449.

URNOV, F.D. ET AL. (2010) Genome editing with engineered zinc finger nucleases. *Nature reviews. Genetics* **11**, 636–646.

VARSHNEY, A. ET AL. (2017) Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 2301–2306.

VELAZCO-CRUZ, L. ET AL. (2019) Acquisition of Dynamic Function in Human Stem Cell-Derived beta Cells. *Stem cell reports* **12**, 351–365.

VELUTHAKAL, R. ET AL. (2009) Regulatory roles for Tiam1, a guanine nucleotide exchange factor for Rac1, in glucose-stimulated insulin secretion in pancreatic beta-cells. *Biochemical pharmacology* **77**, 101–113.

VERKUIJL, S.A.N. & ROTS, M.G. (2019) The influence of eukaryotic chromatin state on CRISPR–Cas9 editing efficiencies. *Current Opinion in Biotechnology* **55**, 68–73.

VIERRA, N.C. ET AL. (2015) Type 2 diabetes-associated K+ channel TALK-1 modulates β-cell electrical excitability, second-phase insulin secretion, and glucose homeostasis. *Diabetes* **64**, 3818–3828.

VIKMAN, J. ET AL. (2003) Requirement for N-ethylmaleimide-sensitive factor for exocytosis of insulin-containing secretory granules in pancreatic beta-cells. *Biochemical Society transactions* **31**, 842–847.

VIÑUELA, A. ET AL. (2019) Influence of genetic variants on gene expression in human pancreatic islets – implications for type 2 diabetes. *bioRxiv*, 655670.

VIONNET, N. ET AL. (1992) Nonsense mutation in the glucokinase gene causes early-onset non-insulin-dependent diabetes mellitus. *Nature* **356**, 721–722.

VOIGHT, B.F. ET AL. (2012) The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS genetics* **8**, e1002793.

VOIGHT, B.F. ET AL. (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nature* **42**, 579–589.

VOLKOV, P. ET AL. (2017) Whole-Genome Bisulfite Sequencing of Human Pancreatic Islets Reveals Novel Differentially Methylated Regions in Type 2 Diabetes Pathogenesis. *Diabetes* **66**, 1074 LP – 1085.

DE VOS, A. ET AL. (1995) Human and rat beta cells differ in glucose transporter but not in glucokinase gene expression. *Journal of Clinical Investigation* **96**, 2489–2495.

VUJKOVIC, M. ET AL. (2020) Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. *Nature Genetics* **52**, 680–691.

WALTER, K. ET AL. (2015) The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82–90.

WANG, B. ET AL. (2017) ATXN1L, CIC, and ETS Transcription Factors Modulate Sensitivity to MAPK Pathway Inhibition. *Cell reports* **18**, 1543–1557.

WANG, C. ET AL. (2019) Imaging-based pooled CRISPR screening reveals regulators of lncRNA localization. *Proceedings of the National Academy of Sciences* **116**, 10842 LP – 10851.

WANG, J. ET AL. (2018) Knocking down Insulin Receptor in Pancreatic Beta Cell lines with Lentiviral-Small Hairpin RNA Reduces Glucose-Stimulated Insulin Secretion via Decreasing the Gene Expression of Insulin, GLUT2 and Pdx1. *International journal of molecular sciences* **19**.

WANG, J. ET AL. (1999) A mutation in the insulin 2 gene induces diabetes with severe pancreatic beta-cell dysfunction in the Mody mouse. *The Journal of clinical investigation* **103**, 27–37.

WANG, L. ET AL. (2014a) Dichotomous role of pancreatic HUWE1/MULE/ARF-BP1 in modulating beta cell apoptosis in mice under physiological and genotoxic conditions. *Diabetologia* **57**, 1889–1898.

WANG, T. ET AL. (2014b) Genetic Screens in Human Cells Using the CRISPR-Cas9 System. *Science* **343**, 80 LP – 84.

WARD, L.D. & KELLIS, M. (2012) Interpreting noncoding genetic variation in complex traits and human disease. *Nature biotechnology* **30**, 1095–1106.

WEIR, G.C. & BONNER-WEIR, S. (2011) Finally! A human pancreatic β cell line. *The Journal of clinical investigation* **121**, 3395–3397.

WEIR, G.C. ET AL. (1979) Somatostatin and pancreatic polypeptide secretion: effects of glucagon, insulin, and arginine. *Diabetes* **28**, 35–40.

WELSH, M. ET AL. (1985) Control of insulin gene expression in pancreatic beta-cells and in an insulin-producing cell line, RIN-5F cells. II. Regulation of insulin mRNA stability. *The Journal of biological chemistry* **260**, 13590–13594.

WEYER, C. ET AL. (1999) The natural history of insulin secretory dysfunction and insulin resistance in the pathogenesis of type 2 diabetes mellitus. *Journal of Clinical Investigation* **104**, 787–794.

WIJESEKARA, N. ET AL. (2010) Beta cell-specific Znt8 deletion in mice causes marked defects in insulin processing, crystallisation and secretion. *Diabetologia* **53**, 1656–1668.

WILDING, J.P.H. ET AL. (2016) Positioning SGLT2 Inhibitors/Incretin-Based Therapies in the Treatment Algorithm. *Diabetes Care* **39**, S154 LP-S164.

WILLEMSEN, G. ET AL. (2015) The Concordance and Heritability of Type 2 Diabetes in 34,166 Twin Pairs From International Twin Registers: The Discordant Twin (DISCOTWIN) Consortium. *Twin Research and Human Genetics* **18**, 762–771.

WU, T. ET AL. (2016) Epigallocatechin-3-Gallate Inhibits Ethanol-Induced Apoptosis Through Neurod1 Regulating CHOP Expression in Pancreatic β-Cells. *Anatomical Record* **299**, 573–582.

XIE, N. ET AL. (2018) Novel Epigenetic Techniques Provided by the CRISPR/Cas9 System. *Stem Cells International* **2018**, 7834175.

XIE, X. ET AL. (2015) Molecular basis of ubiquitin recognition by the autophagy receptor CALCOCO2. *Autophagy* **11**, 1775–1789.

XU, G.G. & ROTHENBERG, P.L. (1998) Insulin Receptor Signaling in the Beta-Cell Influences Insulin Gene Expression and Insulin Content: Evidence for Autocrine Beta-

Cell Regulation. *Diabetes* **47**, 1243–1252.

XU, H. ET AL. (2018) Determination of EPAC2 function using EPAC2 null Min6 sublines generated through CRISPR-Cas9 technology. *Molecular and cellular endocrinology* **473**, 114–123.

YAMADA, H. ET AL. (2016) Potentiation of Glucose-stimulated Insulin Secretion by the GPR40-PLC-TRPC Pathway in Pancreatic β-Cells. *Scientific reports* **6**, 25912.

YAMAGATA, K. ET AL. (1996a) Mutations in the hepatocyte nuclear factor-4alpha gene in maturity-onset diabetes of the young (MODY1). *Nature* **384**, 458–460.

YAMAGATA, K. ET AL. (1996b) Mutations in the hepatocyte nuclear factor-1alpha gene in maturity-onset diabetes of the young (MODY3). *Nature* **384**, 455–458.

YAMAMOTO, S. ET AL. (2018) Autophagy Differentially Regulates Insulin Production and Insulin Sensitivity. *Cell reports* **23**, 3286–3299.

YANG, J. ET AL. (2015) Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nature genetics* **47**, 1114–1120.

ZAWALICH, W.S. & ZAWALICH, K.C. (1996) Regulation of insulin secretion by phospholipase C. *The American journal of physiology* **271**, E409-16.

ZEGGINI, E. ET AL. (2008) Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet* **40**, 638–645.

ZEGGINI, E. ET AL. (2007) Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* **316**, 1336–1341.

ZENG, H. ET AL. (2016) An Isogenic Human ESC Platform for Functional Evaluation of Genome-wide-Association-Study-Identified Diabetes Genes and Drug Discovery. *Cell stem cell* **19**, 326–340.

ZHANG, K. & KAUFMAN, R.J. (2006) The unfolded protein response: a stress signaling pathway critical for health and disease. *Neurology* **66**, S102-9.

ZHANG, R. ET AL. (2016) A CRISPR screen defines a signal peptide processing pathway required by flaviviruses. *Nature* **535**, 164–168.

ZHENG, J. ET AL. (2019) Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *bioRxiv*, 627398.

ZHENG, Y. ET AL. (2018) Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. *Nature Reviews Endocrinology* **14**, 88–98.

ZHOU, G. ET AL. (2001) Role of AMP-activated protein kinase in mechanism of metformin action. *The Journal of clinical investigation* **108**, 1167–1174.

ZHOU, J. ET AL. (2014) Dual sgRNAs facilitate CRISPR/Cas9-mediated mouse genome targeting. *FEBS Journal* **281**, 1717–1725.

ZHOU, K. ET AL. (2016) Variation in the glucose transporter gene SLC2A2 is associated with glycemic response to metformin. *Nature genetics* **48**, 1055–1059.

ZHU, X. ET AL. (2015) Microtubules Negatively Regulate Insulin Secretion in Pancreatic β

Cells. *Developmental cell* **34**, 656–668.

ZINMAN, B. ET AL. (2018) Liraglutide and Glycaemic Outcomes in the LEADER Trial. *Diabetes therapy : research, treatment and education of diabetes and related disorders* **9**, 2383–2392.

ZISMAN, A. ET AL. (2000) Targeted disruption of the glucose transporter 4 selectively in muscle causes insulin resistance and glucose intolerance. *Nature medicine* **6**, 924–928.

ZU, Y. ET AL. (2015) Pericentrin Is Related to Abnormal β-Cell Insulin Secretion through F-Actin Regulation in Mice. *PloS one* **10**, e0130458–e0130458.

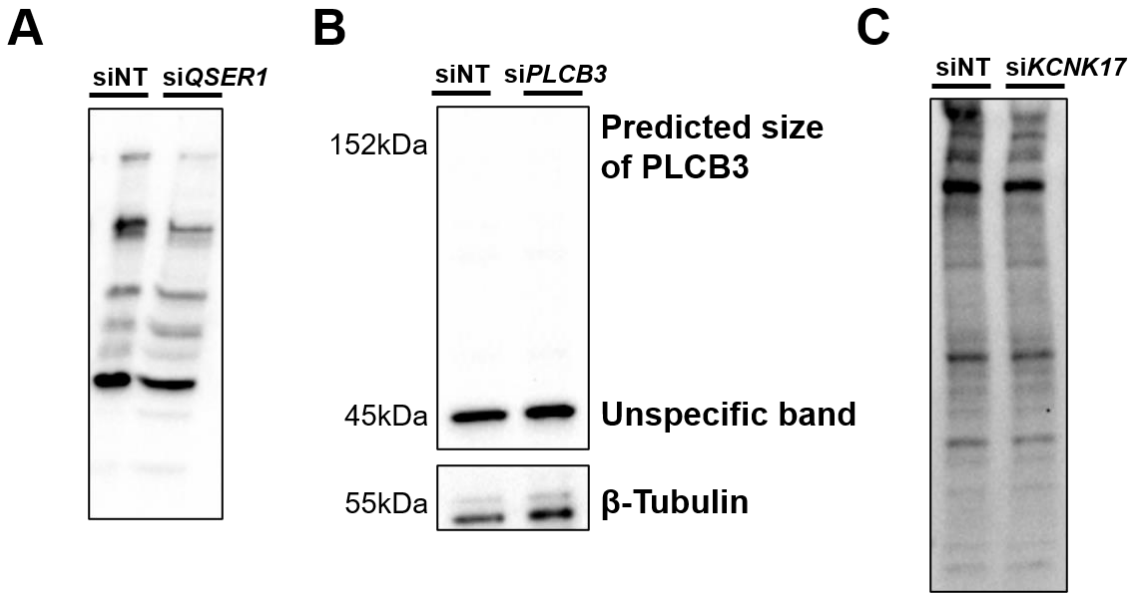
Appendices

Gene					
AAAS	*ARNT2*	*CAPN7*	*DCAF13*	*EP300*	*GALE*
ABCB6	*ASAP1*	*CARHSP1*	*DCUN1D4*	*EPB41L1*	*GALNT7*
ABHD8	*ASCC3*	*CASC5*	*DDAH2*	*EXOSC3*	*GATAD1*
AC018755.1	*ASH2L*	*CAT*	*DDX23*	*EXOSC9*	*GBP3*
ACO2	*ASXL3*	*CBR4*	*DDX26B*	*F2R*	*GEMIN6*
ACTN1	*ATP5H*	*CCDC66*	*DDX3Y*	*F3*	*GEMIN8*
ACTR3B	*ATP5I*	*CDH16*	*DDX56*	*FADS1*	*GINS3*
ADAMTS5	*ATP6V1C1*	*CDK5R2*	*DEDD2*	*FAM109A*	*GJD2*
ADCY6	*ATP8B2*	*CDK5RAP2*	*DERL2*	*FAM172A*	*GJD3*
ADNP	*AURKA*	*CDK5RAP3*	*DGAT1*	*FAM175B*	*GLCE*
AGO1	*B4GALNT4*	*CDYL2*	*DHRS1*	*FAM178A*	*GLI4*
AGO3	*BAZ2B*	*CEP85L*	*DHX36*	*FAM184A*	*GLTSCR1L*
AGO4	*BLCAP*	*CETN3*	*DHX8*	*FAM212B*	*GMEB1*
AGPAT2	*BMPR1A*	*CFLAR*	*DLL3*	*FAM213A*	*GNL3*
AGPAT5	*BTBD10*	*CHD4*	*DMTN*	*FAM46C*	*GPAA1*
AGPS	*BTBD2*	*CHODL*	*DNAJB7*	*FBXL2*	*GPM6A*
AIMP2	*BTF3L4*	*CHRNA5*	*DNAJC1*	*FBXO16*	*GPR119*
ALG2	*BTN3A1*	*CHST8*	*DNAJC4*	*FBXO21*	*GPR137*
AMER3	*C11orf21*	*CHTOP*	*DOCK7*	*FBXO22*	*GPR98*
ANAPC5	*C12orf29*	*CIAO1*	*DOK1*	*FBXO46*	*GPS1*
ANKRD13D	*C12orf52*	*CLK4*	*DOK4*	*FBXO48*	*GPSM2*
ANKRD46	*C14orf39*	*CMTM6*	*DOPEY2*	*FBXW8*	*GRB2*
ANKRD7	*C17orf59*	*CNBP*	*DPCD*	*FDX1*	*GRIP1*
ANO6	*C1orf198*	*CNPY2*	*DPH1*	*FERMT2*	*GSDMB*
ANP32E	*C20orf112*	*COA5*	*DPYSL2*	*FFAR1*	*GSN*
ANTXR2	*C21orf2*	*COL1A1*	*DPYSL5*	*FGFR1OP*	*HARS2*
ANXA10	*C2orf76*	*COL4A1*	*DUSP10*	*FGFR3*	*HAUS2*

ANXA13	*C3orf14*	*COL4A3BP*	*DUSP18*	*FIBP*	*HBS1L*
AP1S1	*C3orf70*	*COX19*	*DUSP3*	*FKBP14*	*HDAC2*
AP5S1	*C7orf41*	*CPM*	*DUSP4*	*FLII*	*HDDC3*
ARAP1	*C7orf49*	*CPNE4*	*DVL2*	*FLNB*	*HDGF*
ARF1	*C7orf60*	*CRB3*	*EID2B*	*FPGS*	*HEXIM2*
ARIH2	*C9orf37*	*CREB3*	*EIF2D*	*FRMD5*	*HIST1H2BK*
ARL10	*CACNB3*	*CSTF2*	*EIF3B*	*FSTL3*	*HM13*
ARL14EP	*CALCOCO2*	*CTNNA2*	*EIF3D*	*FTSJ1*	*HMGXB4*
ARL6IP6	*CALM3*	*CXorf56*	*EIF3J*	*FUT4*	*HNRNPDL*
ARMCX5	*CALY*	*CYB561D2*	*ELAVL1*	*GABRB3*	*HNRNPM*
HOMER2	*MARCH3*	*NISCH*	*PGLS*	*RAB14*	*SERPINA6*
HSPA8	*MARCH9*	*NKX2-2*	*PHACTR2*	*RAB19*	*SERPINH1*
HUWE1	*MB21D2*	*NLK*	*PHF8*	*RAD23A*	*SERTAD2*
IBTK	*MBD2*	*NME3*	*PIBF1*	*RAD52*	*SFR1*
IER2	*MBNL1*	*NOC2L*	*PITPNM2*	*RAI14*	*SGPP2*
IFNGR2	*MCCC1*	*NOL8*	*PKM*	*RAP1A*	*SGTB*
IFT88	*MED11*	*NPHP4*	*PLA2G4C*	*RBBP6*	*SH3GL2*
IGF2	*MED13L*	*NPY*	*PLCL2*	*RBM19*	*SHQ1*
IKBKG	*MED18*	*NRAS*	*PLD3*	*RBM43*	*SIDT2*
IKZF4	*MED23*	*NSF*	*PLK1*	*RBMS1*	*SIN3A*
IL1R2	*MED4*	*NUDT14*	*POLD2*	*RBP1*	*SIPA1L2*
INPP5J	*MESDC2*	*NUDT3*	*POLH*	*RCC2*	*SLC10A3*
INS	*METTL5*	*NUP107*	*POLM*	*RDX*	*SLC16A10*
INTS6	*MGA*	*NUP188*	*POLR2G*	*RELB*	*SLC25A10*
IPO4	*MID1*	*NUP54*	*POP4*	*REXO2*	*SLC25A12*
IQSEC1	*MIF*	*NUPL1*	*PPA2*	*RFWD2*	*SLC25A19*
IREB2	*MIPEP*	*ORAOV1*	*PPAP2B*	*RND3*	*SLC25A39*
IRS2	*MLH3*	*OSBP2*	*PPAPDC2*	*RNF115*	*SLC25A40*
ITM2C	*MMP16*	*OSBPL7*	*PPP1R15B*	*RNPS1*	*SLC25A43*
JAKMIP2	*MOB1B*	*OTUB1*	*PPP2R3A*	*RPF1*	*SLC27A5*
JARID2	*MPDU1*	*PAFAH1B1*	*PPP2R3C*	*RPL10*	*SLC29A4*
KALRN	*MPST*	*PALLD*	*PPP6R3*	*RPL10A*	*SLC2A11*
KANK1	*MRPL16*	*PALM2-*	*PRDM10*	*RPL12*	*SLC2A2*
KIAA0247	*MRPL19*	*PANK3*	*PRKAA1*	*RPL18A*	*SLC35G2*
KIAA1524	*MRPL28*	*PARN*	*PRKAB1*	*RPL35*	*SLC44A3*
KIAA1586	*MRPL53*	*PARP12*	*PRKCD*	*RPS6KC1*	*SLC45A4*
KREMEN1	*MRPS24*	*PARVA*	*PRKD1*	*RPUSD1*	*SLC6A8*
LAMC1	*MRPS25*	*PBK*	*PRPF18*	*RPUSD2*	*SLC7A1*
LAS1L	*MTERFD1*	*PBX3*	*PRPF38A*	*RPUSD3*	*SLC7A2*
LCAT	*MTIF2*	*PCDHA2*	*PRSS23*	*RRAGB*	*SLC7A5*
LETMD1	*MVK*	*PCED1A*	*PSAP*	*RSPH9*	*SLC7A6*

LHFP	*MYO9A*	*PCGF5*	*PSEN1*	*RSPRY1*	*SLC8B1*
LMBR1L	*NAA40*	*PCMT1*	*PSMB2*	*RSRC1*	*SLMAP*
LONRF1	*NAV1*	*PCNT*	*PSMC4*	*RTN3*	*SLX4*
LOXL4	*NCOA4*	*PDCD11*	*PSME4*	*RUFY1*	*SMARCC1*
LRRC61	*NCOA6*	*PDE4D*	*PTP4A1*	*SAE1*	*SMC1A*
LRRN3	*NDUFA1*	*PDS5A*	*PTPN1*	*SAMD4A*	*SMC3*
LURAP1L	*NDUFAF2*	*PELO*	*PTS*	*SCO2*	*SMIM5*
MAFG	*NDUFAF4*	*PEX10*	*QDPR*	*SCRIB*	*SMTN*
MAGOH	*NECAB2*	*PEX13*	*QSOX2*	*SDCCAG8*	*SMURF1*
MAPK1	*NIPBL*	*PGBD4*	*RAB11FIP3*	*SEMA4F*	*SMURF2*
SMYD5	*TAGLN*	*TMEM251*	*TSTD1*	*WDR7*	*ZNF259*
SOCS7	*TAP1*	*TMEM256*	*TTYH3*	*WDR81*	*ZNF329*
SON	*TAPBPL*	*TMEM30B*	*TUBE1*	*WDR83*	*ZNF404*
SOWAHC	*TBC1D19*	*TMEM39B*	*UBE2C*	*WIPF1*	*ZNF426*
SP1	*TBC1D4*	*TMEM68*	*UBE2D1*	*XYLT2*	*ZNF432*
SPATA24	*TBK1*	*TMEM86B*	*UBE2J1*	*YPEL5*	*ZNF503*
SQSTM1	*TCEAL4*	*TMX4*	*UBE4B*	*ZBTB22*	*ZNF510*
SRD5A3	*TCF19*	*TNFAIP8L1*	*UBR1*	*ZBTB24*	*ZNF524*
SRRM1	*TENM2*	*TNFRSF11A*	*USP12*	*ZCCHC2*	*ZNF529*
SRRM3	*TEX10*	*TNPO3*	*USP14*	*ZCCHC4*	*ZNF550*
SSR2	*TEX2*	*TOMM6*	*USP45*	*ZCCHC9*	*ZNF608*
SSTR1	*TIAM1*	*TP53BP2*	*USP6NL*	*ZDHHC4*	*ZNF616*
STAM	*TIGD2*	*TPP1*	*VAMP4*	*ZFP30*	*ZNF654*
STARD4	*TIMM23*	*TRIM39*	*VIL1*	*ZGLP1*	*ZNF672*
SYDE2	*TLDC1*	*TRIM41*	*VPS37A*	*ZNF101*	*ZNF684*
SYNJ1	*TM2D2*	*TRIM45*	*WASF2*	*ZNF141*	*ZNF821*
TAF3	*TMEM132D*	*TRPM2*	*WDFY1*	*ZNF155*	*ZNF846*
TAF7	*TMEM144*	*TSHZ2*	*WDR27*	*ZNF214*	
TAF8	*TMEM2*	*TSPAN2*	*WDR35*	*ZNF24*	

Supplementary Table 1 – High confidence CRISPR screening hits. 580 genes representing the high confidence hits from the genome-wide CRISPR screen as described in Chapter 4. Genes are sorted alphabetically.



Supplementary Figure 1 – Unspecific antibody staining. Western Blot analysis with unspecific protein detection using antibodies for QSER1 **(A)**, PLCB3 **(B)** and TALK2 **(C)** in their respective siRNA silenced samples and no-targeting control (siNT). Western Blots refer to functional follow-up studies in Chapter 5.

Curr Diab Rep (2017) 17: 76
DOI 10.1007/s11892-017-0907-y

GENETICS (AP MORRIS, SECTION EDITOR)

Prioritising Causal Genes at Type 2 Diabetes Risk Loci

Antje K. Grotz[1] · **Anna L. Gloyn**[1,2,3] · **Soren K. Thomsen**[1]

Published online: 31 July 2017

Abstract
Purpose of Review Genome-wide association studies (GWAS) for type 2 diabetes (T2D) risk have identified a large number of genetic loci associated with disease susceptibility. However, progress moving from association signals through causal genes to functional understanding has so far been slow, hindering clinical translation. This review discusses the benefits and limitations of emerging, unbiased approaches for prioritising causal genes at T2D risk loci.
Recent Findings Candidate causal genes can be identified by a number of different strategies that rely on genetic data, genomic annotations, and functional screening of selected genes. To overcome the limitations of each particular method, integration of multiple data sets is proving essential for establishing confidence in the prioritised genes. Previous studies have also highlighted the need to support these efforts through identification of causal variants and disease-relevant tissues.
Summary Prioritisation of causal genes at T2D risk loci by integrating complementary lines of evidence promises to accelerate our understanding of disease pathology and promote translation into new therapeutics.

This article is part of the Topical Collection on *Genetics*

✉ Soren K. Thomsen
soren.thomsen@stx.ox.ac.uk

[1] Oxford Centre for Diabetes, Endocrinology & Metabolism, University of Oxford, Oxford, UK

[2] Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK

[3] National Institute of Health Research Oxford Biomedical Research Centre, Churchill Hospital, Oxford, UK

Keywords Genome-wide association study · Type 2 diabetes · Genetic mechanism · Functional genomics · Causal gene · Effector transcript

Abbreviations

CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CRM	*Cis*-regulatory modules
ENCODE	Encyclopaedia of DNA elements
eQTLs	Expression quantitative trait loci
GOF	Gain of function
GWAS	Genome-wide association studies
LD	Linkage disequilibrium
lncRNAs	Long non-coding RNAs
LOF	Loss of function
MTNR1B	Melatonin receptor 1B
PMCA	Phylogenetic module complexity analysis
RNAi	RNA interference
siRNA	Small interfering RNA
SNP	Single-nucleotide polymorphism
TFBS	Transcription factor binding sites
T2D	Type 2 diabetes

Introduction

In the last decade, genome-wide association studies (GWAS) have evolved as a powerful tool for deciphering the genetic component of type 2 diabetes (T2D) risk. By associating regions of the genome with disease susceptibility, more than 100 loci influencing T2D risk have been identified so far [1–6, 7••]. Moving on from an era of disease locus discovery, post-GWAS methodologies are now advancing to functionally characterise the underlying genes and to interrogate disease

pathways. These comprehensive efforts promise to enable subsequent translation into improved disease diagnostics, treatment, and prevention. However, the progression from association signals at T2D loci to causal genes and a functional understanding of diabetes pathology has been limited. The slow progress is due, in part, to problems arising from the methodology itself and, in part, a consequence of the underlying nature of the association signals.

GWAS exploit the fact that single-nucleotide polymorphisms (SNP) tend to be located in linkage disequilibrium (LD) with other variants [8]. By analysing SNPs that lie in LD with non-genotyped variants, these can serve as representatives for their haplotype ('tag SNPs'), and it is thus possible to achieve reasonable genome-wide coverage of common variation by analysing between 0.5–1 million SNPs [9–12]. Thus, the GWAS paradigm is designed to detect SNPs that act as a proxy for disease-associated regions or loci, and not necessarily the actual causal variants. Additionally, the majority of association signals (~90%) are found in non-coding regions, presumably influencing disease risk through effects on gene regulation [13]. The detected SNPs in non-coding regions are named after the nearest protein-coding gene, but proximity to a gene does not imply causality.

The challenge for functional follow-up studies in elucidating disease mechanisms lies therefore in finding both causal variants and the genes through which they impact on disease risk for the corresponding SNPs. Here, we first discuss the benefits of determining the causal variant(s) and affected tissue(s) as a prerequisite for identifying effector transcripts. We review several approaches for prioritising causal genes at T2D loci and provide recent and prominent examples of likely effector transcripts identified by these strategies. Finally, we highlight the importance of triangulating from multiple datasets and discuss the prospects for future integrative studies.

Prerequisites for Finding Causal Genes

Uncovering the underlying causal mechanisms of T2D risk loci is not exclusively a matter of finding causal genes, since these efforts are complicated by the need to identify both causal variant(s) and the affected tissue(s) in order to obtain a complete picture of disease pathology. Moreover, this additional information is often an inevitable requirement for performing functional follow-up studies in an appropriate model system.

Causal Variants

In GWAS, the variant most strongly associated with disease risk is reported for each locus, though such 'lead SNPs' may only serve as surrogate markers for other genetic perturbations that directly contribute to disease pathology. Identifying the true causal variants can provide a direct functional link between genotype and the observed disease phenotype, especially in cases where the variant is protein altering. To identify a causal variant, or a set of likely causal variants, several strategies have been developed, including fine-mapping of disease-associated regions, experimental prioritisation, and in silico prediction tools.

Fine-mapping of a locus involves analysing SNPs in a defined region of the genome for disease association and is used to refine a GWAS association signal from the surrogate lead SNP to the actual causal variant(s). The SNPs are assayed by deep sequencing, or custom array-genotyping based on GWAS variants and imputation from extensive sequencing efforts such as the 1000 Genomes Project [14, 15]. To achieve sufficient statistical power to detect the association of the true causal variant, large sample sizes are required and the studies often include populations drawn from diverse ancestries to exploit differences in LD patterns [16].

Even so, most fine-mapping efforts uncover a large number of variants that, between them, are likely to be driving a particular association signal—a so-called credible set. In some exceptional cases, however, it is possible to narrow down the credible set to a single variant, as is the case for the melatonin receptor 1B gene (*MTNR1B*) [17•]. The *MTNR1B* locus has previously been implicated in T2D risk and the identification of the single causal variant revealed a likely, direct functional link to the causal gene [18]. The risk allele creates a binding site for the transcription factor NEUROD1 and is associated with preferential binding in human pancreatic beta cells. This additional transcription factor binding event also implicates increased FOXA2-bound enhancer activity and *MTNR1B* expression.

Another way to approach the search for causal variants at GWAS loci is by experimentally testing prioritised SNPs. This strategy was, for example, pursued at the *JAZF1* and *CDC123*/*CAMK1D* loci [19–21]. Variants in high LD ($r^2 > 0.8$) with the lead GWAS SNP were selected for functional analysis based on maps of open chromatin. Effects on gene expression were tested in luciferase reporter assays, and DNA binding capability was analysed through electrophoretic mobility shift assays. The identified potential causal variants at the *JAZF1* and *CDC123*/*CAMK1D* loci appear to act as part of cis-regulatory modules (CRMs). These specific regions harbour combinatorial transcription factor binding sites (TFBS), and the variants affect binding of PDX1 and FOXA1/FOXA2, respectively. However, due to practical limitations, this type of experimental studies mostly analyses a subset of regional variants, opening up the possibility of missing potential true causal variants. Further, the evidence generated is only circumstantial, since establishing functionality is necessary but not sufficient to prove causation. The emergence of new experimental lines of evidence may affect the prioritisation of the true causal variants and should ideally involve integration of different types of analyses (see section on "Integrative approach").

To overcome the practical limitations of functional approaches for identifying causal variants, in silico prediction tools offer an alternative method based on specific assumptions regarding their properties. A recent study, for example, leveraged phylogenetic conservation of TFBS within CRMs to predict causal variants at the *PPARG* and *FTO* T2D risk loci [22, 23•]. This computational approach, termed phylogenetic module complexity analysis (PMCA), identified a clustering of homeobox TFBS at T2D risk loci, and initially proposed a potential causal variant at the *PPARG* locus, which allowed for a subsequent functional interpretation [22]. The risk allele at *PPARG2* leads to enhanced binding of the repressive homeobox transcription factor PRRX1, and thus reduced *PPARG2* expression, defective lipid handling, and insulin sensitivity. PMCA was also successfully applied to identify the causal variant and a potential disease mechanism at the obesity-associated *FTO* locus, a region showing the strongest genetic association in GWAS for obesity and body mass index traits [24, 25]. The proposed causal allele was shown to alter an ARID5B repressor motif, leading to activation of the distant *IRX3* and *IRX5* in adipocyte precursor cells, and pro-obesity consequences for adipocyte thermogenesis regulation [23•]. This work also highlights the additional complexity arising from having multiple causal genes for disease-associated haplotypes. Though post-GWAS efforts have tended to focus on the idea of a single causal gene per locus, causal variant(s) may influence any number of regional genes, and not necessarily in the same manner across different contexts.

Causal Contexts

An important aspect of the prioritisation of causal genes and variants at GWAS loci is to consider the appropriate tissue(s) and developmental stage(s), which allow any functional follow-up studies to be performed in a disease-relevant model. As the majority of T2D association signals are located in non-coding regions and exert regulatory effects, their influence on gene expression may be subject to context-specific activity [26]. Thus, studies analysing the implicated variants and genes need to consider the surrounding genomic context and expression patterns. A notable example is provided by work on the *PTF1A* gene, where a disease-relevant model, human pancreatic progenitor cells, was critical to elucidating a mechanism for isolated pancreatic agenesis [27•]. The identified mutations were found to disrupt an enhancer region that is selectively active in pancreatic progenitor cells and, importantly, show no activity in corresponding adult cell lines.

Strategies for Prioritising Causal Genes

The aim of translating genetic variants into molecular mechanisms will ultimately centre on the identification of causal genes. It is enhanced understanding at this level that holds the key to discovering novel treatments, prevention targets, and diagnostic markers. Several strategies to address this issue are being pursued, including the interrogation of coding variants, establishing variant-gene links for non-coding variants, and using high-throughput screens to prioritise candidate genes.

Coding Variants

Recent GWAS endeavours have shifted attention towards exome-arrays and exome-sequencing to enable identification of rare and low-frequency variants with potentially larger effect sizes—and a more direct biological interpretation—than common variants [7••, 28–30]. Missense variants in coding regions have a protein-altering effect that can directly pinpoint causal genes, offering the possibility of a straightforward and rapid translation into the clinic (Fig. 1).

The importance of coding variants for ascertaining causal mechanisms is illustrated by *SLC30A8*, which encodes a zinc transporter (ZnT8) that is active in the secretory vesicles of beta cells. *SLC30A8* was initially identified as a T2D susceptibility gene harbouring a common missense variant [2]. Contradictory to the supposed negative impact of this risk allele, recent efforts to identify protein-truncating variants leading to loss of function (LOF) in T2D genes discovered several rare protein-truncating variants in *SLC30A8* [31•]. Strikingly, the haploinsufficiency conferred by this class of variants was found to be associated with a 65% reduction in T2D risk. By discovering multiple independent coding variants at this GWAS locus, *SLC30A8* has been validated with high confidence as the causal gene. Furthermore, this study highlights the importance of discovering an extended allelic series to understand functional mechanisms. More broadly, it has established reduced activity of ZnT8 as a protective disease mechanism in T2D and a potential treatment strategy based on antagonism [32].

The power to detect causal genes through coding variants can be further harnessed by performing genetic association studies in isolated populations. These populations, founded by a bottleneck event, show a higher degree of LD, less genetic complexity, and higher allelic frequencies due to genetic drift, which leads to fixation or extinction of specific alleles over time [33]. Furthermore, these studies also benefit from shared non-genetic backgrounds (e.g. common lifestyle and cultural habits), which is a potential confounding factor in larger outbred populations [34]. Exploiting these advantages of studies in isolated populations, a nonsense coding variant in *TBC1D4* was discovered in the Greenlandic population with the largest effect size for a common T2D risk allele (odds ratio = 10.3) [35•]. The variant disrupts the full-length isoform of TBC1D4, which is selectively expressed in skeletal muscle,

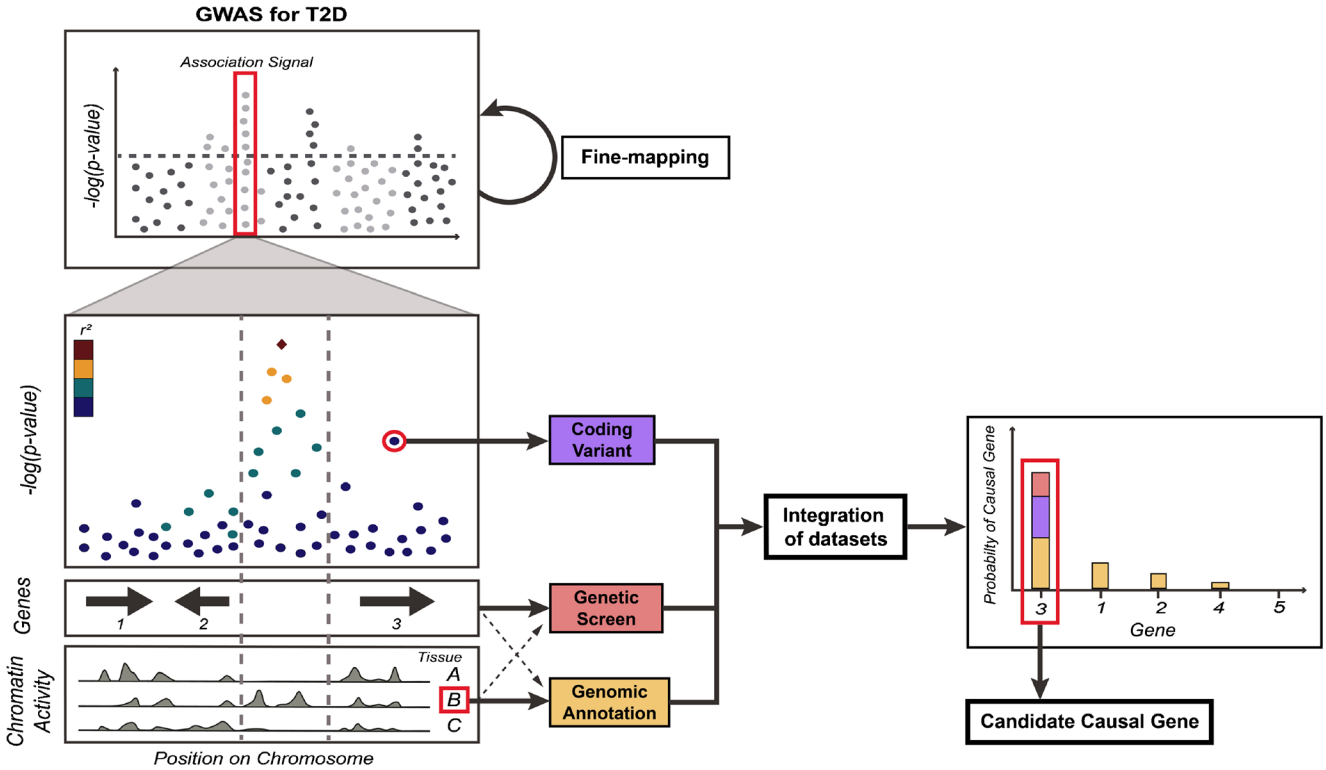


Fig. 1 Using genetic data, genomic annotations, and functional screening for prioritising causal genes at T2D GWAS loci. GWAS for T2D risk have identified more than 100 independent association signals to date (Manhattan plot; *top left*), but the majority of causal genes driving the effects on disease susceptibility remain unknown. Fine-mapping of associated regions can aid the prioritisation efforts by narrowing down the credible sets of causal variants (see main text). Emerging strategies for prioritising causal genes are highlighted for a hypothetical T2D risk locus (*bottom left*); the regional association plot shows a primary, non-coding association signal located upstream of gene 2 and downstream of gene 3 (lead variant; *red diamond*). An independent, coding variant in gene 3 displays moderate (sub-significant) association with T2D risk, providing evidence hinting at this gene as causal at this locus. Further, genomic annotations for different cell types (A, B, and C, for illustration) reveal the primary association signal to be located in a region that displays tissue-specific activity in cell type B. This information provides valuable information for two independent prioritisation strategies. Firstly, functional genetic screening of all regional genes (e.g. genes 1–3 [shown] and 4–5 [not shown]) can be performed in a disease-relevant context, measuring a phenotype specific to cell type B. Further, variant-gene links can be established through experimental studies in tissue B, using, for example, cis-eQTL or chromatin confirmation capture methodologies. Importantly, each of the methods outlined have their own set of limitations (see main text), and integration is thus important for establishing confidence in particular candidates. In this case (graph; *bottom right*), gene 3, which was highlighted by genetic data (*purple bar*), has also been found in a functional screen to cause defects in a disease-relevant tissue, adding further evidence in support of this gene as causal (*red bar*). Finally, variant-gene annotations have shown some degree of evidence for associations between the non-coding signal and genes 1–4 (*yellow bars*), with gene 3 being the most significant target. Taken together, the aggregate burden of priors provides a high degree of confidence in gene 3 as the candidate causal gene at this locus, which can be used to prioritise the gene for follow-up in-depth validation studies

thus exerting its influence on T2D risk through insulin resistance.

Another recent study leveraging the advantages of isolated populations detected a low-frequency coding variant in *AKT2* in the Finnish population [36]. The allele confers T2D risk through increased fasting plasma insulin levels and expands the allelic spectrum from the previously known rare variants in *AKT2* that cause monogenic heterogeneous glycaemic diseases [37, 38]. Collectively, these studies illustrate the importance of identifying coding variants—in isolated and outbred populations—for straightforward translation into molecular mechanisms. While harnessing coding variation can offer powerful insights into causal mechanism, this approach is fundamentally limited by the occurrence of natural variation (in outbred and isolated populations) which necessitates ever-larger association studies to detect rare, coding variation. In addition, identification of a coding signal is not a guarantee for causality, and conditional analysis is often required to estimate the likelihood of a given variant being causal [39]. By design, exome-based studies analyse coding regions only, and thus require additional fine-mapping of non-coding regions to exclude the contribution of non-coding variants as drivers of the association signal.

Establishing Variant-Gene Links

In contrast to missense coding variants, associating GWAS signals in non-coding regions with their downstream causal gene is often a more complex challenge. To identify regulatory effects, non-coding variants can be correlated with genomic

annotations to establish a functional link with their target gene (Fig. 1). Expression quantitative trait loci (eQTLs), for example, describe variants that influence gene expression in close proximity (*cis*-eQTL) or over a long distance (*trans*-eQTL), and provide an approach for directly linking a GWAS variant to its causal gene through effects on expression levels [40]. Crucial for the success of eQTL studies is the interrogation of the correct disease-relevant context(s), since gene expression is often regulated in a cell-type specific manner [41].

For T2D, a large number of disease risk loci have been found through physiological studies to affect insulin processing or secretion in the beta cell, highlighting pancreatic islets as a relevant starting point for annotation studies at these loci [42]. Up to now, islet sample availability has been limiting for large-scale studies, thereby reducing statistical power to detect associations. Nonetheless, recent studies succeeded in mapping islet *cis*-eQTLs and overlapping these with variants driving T2D association signals [43, 44•]. One such coincident locus is *ZMIZ1*, harbouring a gene that had been sparsely characterised for its role in T2D risk [44•]. A recent study confirmed *ZMIZ1* as the likely causal gene at this T2D risk locus, and functional follow-up work has established a role in beta cell function for insulin secretion and exocytosis, thus giving first insights into a potential mechanism [44•, 45].

Tissue availability has so far prevented any progress in finding islet *trans*-eQTLs. T*rans*-eQTLs act over distance and the entire genome is interrogated for any variant-gene associations, thus further limiting power due to more stringent multiple-testing correction [40]. Still, efforts in adipose tissue have demonstrated the power of this approach by elucidating a trans-regulatory network of *KFL14*, a gene linked with both T2D and other metabolic traits [46]. As KLF14 is a transcription factor, the aim of the study was to identify *trans*-genes that are influenced by varied KLF14 levels through *cis*-eQTL variants. Several genes with genome-wide significance were discovered and the study not only connected GWAS, *cis*- and *trans*-associations for the same set of variants, but also defined important disease-related pathways.

The search for causal genes has been pushed ahead by eQTL studies, but the ability to perform large-scale studies containing correlated sets of genotype, phenotype and expression data are still limited by cost obstacles and sample availability. GWAS only measure genetic variation related to a disease phenotype, and expression studies suffer from reduced statistical power due to smaller sample sizes. Predicted expression association studies attempt to circumvent these limitations by integrating existing GWAS and eQTL data [47–50]. This approach aims to identify disease associations based on groups of variants that influence gene expression, directly pinpointing the causal gene instead of tag SNPs. To combine limited available expression sets with large-scale GWAS data, these studies rely on predicted expression modelled from reference panels. The models then impute expression either for publically available summary GWAS data (most large-scale studies) or GWAS data with individual genotypes [47, 49]. This drastically increases power to detect genes that are predicted to show differences in genotype-dependent expression patterns in T2D, and reduces potential confounding factors like reverse causation, where the phenotype and environment influence gene expression [50]. However, similar to *cis*-eQTL studies, predicted expression association studies are unable to detect context-dependent effects that are not captured by the tissues and developmental stages included in the reference panel used for modelling [48, 49]. It is also not possible to exclude the possibility of pleiotropy caused by multiple, correlated effects of groups of variants on gene expression [48]. Despite such limitations, these methods offer a complementary and powerful approach for prioritisation of causal genes and predicted directions of effect.

Genetic Screening of T2D Genes

A third way to identify genes involved in disease risk is prioritisation based on known or observed functions that are perceived to be relevant for disease pathogenesis. T2D risk variants, for instance, would be expected to affect genes involved in cellular processes relevant to disease susceptibility, such as beta cell function and insulin resistance. A gene found to regulate insulin secretion would thus have high prior odds of being the downstream mediator for a nearby T2D association signal known to impact on islet function. Though this is an indirect approach for prioritisation, the strategy benefits from focusing on the relevant processes that ultimately causes effects on disease pathology (Fig. 1). For unbiased generation of priors, all disease-relevant phenotypes should ideally be comprehensively interrogated in a genome-wide fashion. However, most post-GWAS approaches have previously focused on individual candidate genes, with experimental setups that make them poorly suited for systematic assessment of large numbers of genes across multiple tissues.

High-throughput functional genomic screening is an emerging and increasingly powerful approach that allows for highly parallel phenotypic screening to address this gap. Several screening strategies have been established that differ in their direction of modulated gene expression (gain of function (GOF) vs LOF), format (pooled vs arrayed), and gene modulation techniques (RNA interference (RNAi) vs CRISPR/Cas9 modulation) [51–56, 57•]. Screens can either be performed genome-wide, representing an unbiased approach to detect genes that are involved in a specific phenotype, or based on selected genes of interest. A recent study by Thomsen et al. successfully pursued a small interfering RNA (siRNA) arrayed screening approach to systematically interrogate positional candidate genes at T2D GWAS loci in a human beta cell line [45]. Genes located within 1 Mb of 75 GWAS association signals were analysed for insulin secretion

and cell proliferation to reflect beta cell dysfunction. This strategic approach provided 300 genes for screening and identified 45 genes at 37 GWAS loci for having a role in beta cell dysfunction, thus also pinpointing them as potential effector transcripts at these disease loci. Several prioritised genes with poorly characterised connection to beta cell function were separately validated in functional follow-up work including *ARL15*, *THADA*, and *ZMIZ1*. Independently of the previously described *cis*-eQTL study, this work thus attributed a role to *ZMIZ1* in beta cell function, converging multiple lines of evidence to enhance confidence in the candidacy of this gene as causal. Importantly, the study also demonstrated a strong enrichment for known regulators of insulin secretion among significant hits, providing an internal validation that is an essential aspect of any screening strategy.

Taking a more inclusive approach, Pappalardo and colleagues recently pursued the first whole-genome siRNA screen to identify genes involved in glucose homeostasis and T2D [58]. While allowing for a more unbiased approach, performing an arrayed, genome-wide screen restricts the complexity of the phenotype(s) that can be practically measured. This screen focused on a reporter gene readout for insulin promoter activity in a rat beta cell line. The authors were able to identify several novel regulators of insulin promoter activity including *Spry2*, the gene in the closest proximity to a nearby T2D GWAS association signal [59]. The work thus highlights *Spry2* as the likely causal gene at this locus, and follow-up work in cellular and in vivo systems including beta cell specific knockout mice discovered a potential functional mechanism. However, a link between the non-coding association signal and *Spry2* remains to be investigated, ideally through integration with variant-to-gene approaches in human beta cells. This screen also provided robust internal validation by confirming the strongest hits to be known transcription factors targeting the insulin promoter.

Medium-throughput screens and systematic analysis of selected classes of genes represents a related strategy for analysing candidate genes in more depth across a larger spectrum of possible disease phenotypes. This approach was pursued by a recent study that investigated the function of 12 long non-coding RNAs (lncRNAs) in beta cell gene regulation and their potential role in T2D [60]. These lncRNA knockdown targets were selected based on criteria that included expression in a relevant model and an active chromatin profile. The study showed that the beta cell specific lncRNAs jointly regulate enhancer-cluster associated genes with known transcription factors. The lncRNA named as *PLUTO* was established as a regulator of its neighbouring gene *PDX1*, a transcription factor involved in pancreatic development and beta cell function [61]. Based on this overlapping role of lncRNAs and islet transcription factors, and the well-established involvement of the latter in T2D, the work hints at a similarly important role of lncRNAs in T2D pathology.

Future genetic screens hold the potential to play an important role in identifying causal genes for T2D. Pooled approaches are able to extend the scale of arrayed screens in a cost-effective manner and allow for simultaneous perturbation of thousands of genes to promote unbiased interrogation of candidate causal genes. The continuous development and improvement of the differentiation process of induced pluripotent stem cells into beta cells will also allow for investigations of disease-relevant phenotypes at various developmental stages [62, 63]. High-throughput screens thus offer the opportunity to facilitate the transition from T2D GWAS association signals to individual functional follow-up studies by prioritising candidate causal genes based on functional data.

Integrative Approach

All of the above outlined strategies provide complementary approaches for prioritising causal genes for association signals, each with individual advantages and drawbacks. Coding variants are reliant upon large-scale association studies and naturally occurring variation, while variant-gene links are limited by the availability of primary tissue and possible pleiotropy, and gene-centric functional studies establish indirect evidence in a manner that is strongly dependent on context-dependent effects. As a result, one specific line of evidence can only give limited insights into causal mechanisms and is rarely sufficient to provide definitive evidence for a particular mechanism. The true causal gene(s) can only be identified with confidence through integration and convergence of several complementing datasets [64].

The importance of taking an integrative approach is illustrated by the T2D susceptibility locus on chromosome 11q13, which is located near the protein-coding genes *ARAP1*(*CENTD2*) and *STARD10* [3, 65]. Initial studies highlighted *ARAP1* as an effector transcript at the locus, but recent findings contradict this assumption and instead propose *STARD10* as the causal gene [44, 66, 67••]. Fine-mapping, functional annotation data, chromatin accessibility and conformation capture data, promoter-reporter assays in beta cell models, *cis*-eQTL in islet samples, and global and selective mice knockout models were all used to generate complementary data that attribute a role to *STARD10* at this locus. The comprehensive set of data makes it possible to infer causality by triangulation from different results. This point is emphasised by examining the chromatin confirmation capture data in isolation. Physical interactions between both the *STARD10* and *ARAP1* promoters and variants in the credible causal set highlight the possibility of regulatory effects on either gene. Thus, additional information was required to clarify the roles of these genes in disease pathology.

Another recent study outlines how the integration of genomic, expression and functional data can prioritise a potential

causal gene and disease mechanism, and furthermore directly propose a therapeutic hypothesis. The investigated T2D risk allele is common in Mexicans and Latin Americans (~30% allelic frequency) and located near *SLC16A11* and *SLC16A13* [68]. Fine-mapping identified a credible set of causal variants including non-coding variants and missense coding variants in *SLC16A11*. Liver expression data and chromatin modification analysis showed reduced *SLC16A11* expression and less-activating histone modifications in samples from T2D risk allele carriers, thus proposing *SLC16A11* as the candidate causal gene. Further studies into the function of SLC16A11, an H+-coupled monocarboxylate transporter revealed that the coding risk variants exert their effect through decreased chaperone interaction and SLC16A11 plasma membrane localization. Rusu et al. were also able to show how decreased SLC16A11 function might lead to increased T2D risk by having an effect on cellular fatty acid and lipid metabolism, providing a possible therapeutic strategy.

Despite comprehensive integration of datasets, the evidence in these studies still cannot exclude additional pleiotropy (e.g. regulatory effects that remain undetected due to insufficient power, or effects that manifest in cell types not studied). Exhaustively addressing these gaps will require access to data that enable interrogation of variant function in any context (e.g. well-powered cis-eQTL studies across all disease-relevant cell states), and is far from being a feasible aim for current post-GWAS studies. The emergence of ever-greater, publically available datasets of this nature will increasingly facilitate integration with results of individual studies and thereby guide interpretation. Large-scale projects such as ENCODE, Genotype-Tissue Expression (GTEx) and the NIH Epigenomics Roadmap have already generated enormous functional annotation datasets that allow for intersection with potentially causal variants across hundreds of cell types [41, 69, 70]. However, in the case of tissue-specific annotations for inaccessible tissues like islets, these datasets are often lacking or immature. Future studies will expand the possibilities of integrating datasets and improve the prospects for identification of causal genes in T2D.

Conclusion

Connecting GWAS association signals to their corresponding causal genes has proven a major experimental challenge and bottleneck for therapeutic translation. As a consequence of GWAS design and the genetic architecture of T2D, causal variants and genes cannot be easily inferred from genetic association studies, hindering functional interpretation. Thus, prioritising causal genes at T2D loci to aid functional understanding is a central aspect of current studies. These studies must be guided by parallel efforts to identify causal variants and appropriate disease-relevant model systems. A number of strategies have emerged for causal gene prioritisation based on genetic data, genomic annotations, and functional screening, each with limitations that render them insufficient in isolation. Several lines of evidence and different experimental strategies should thus be triangulated to validate the results and increase confidence in a specific causal mechanism. Looking forward, this era of gene prioritisation based on T2D GWAS loci and functional understanding holds the promise to unlock the full potential of genomic medicine and clinical translation.

Acknowledgements ALG is a Wellcome Trust Senior Fellow in Basic Biomedical Science (095101/Z/10/Z and 200837/Z/16/Z), and AKG and SKT are Radcliffe Department of Medicine Scholars. All authors were responsible for drafting the article and revising it critically for important intellectual content. All authors approved the version to be published.

Compliance with Ethical Standards All reported studies/experiments with human or animal subjects performed by the authors have been previously published and complied with all applicable ethical standards (including the Helsinki declaration and its amendments, institutional/national research committee standards, and international/national/institutional guidelines).

Conflict of Interest Antje K. Grotz, Anna L. Gloyn, and Soren K. Thomsen declare that they have no conflict of interest.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Papers of particular interest, published recently, have been highlighted as:
- • Of importance
- •• Of major importance

1. Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PIW, Chen H, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science. 2007;316:1331–6.
2. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature. 2007;445:881–5.
3. Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet. 2010;42:579–89.
4. Cho YS, Chen C-H, Hu C, Long J, Hee Ong RT, Sim X, et al. Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. Nat Genet. 2012;44:67–72.

5. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segre AV, Steinthorsdottir V, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet. 2012;44:981–90.
6. Steinthorsdottir V, Thorleifsson G, Sulem P, Helgason H, Grarup N, Sigurdsson A, et al. Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. Nat Genet. 2014;46:294–8.
7.•• Fuchsberger C, Flannick J, Teslovich TM, Mahajan A, Agarwala V, Gaulton KJ, et al. The genetic architecture of type 2 diabetes. Nature. 2016;536:41–7. **This large-scale association study finds no support for a major role of low-frequency and rare variants in the heritability of T2D risk.**
8. Slatkin M. Linkage disequilibrium- understanding the evolutionary past and mapping the medical future. Nat Rev Genet. 2008;9:477–85.
9. Li M, Li C, Guan W. Evaluation of coverage variation of SNP chips for genome-wide association studies. Eur J Hum Genet. 2008;16:635–43.
10. The International HapMap Consortium. A haplotype map of the human genome. Nature. 2005;437:1299–320.
11. The International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. Nature. 2007;449:851–61.
12. The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature. 2015;526:68–74.
13. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science. 2012;337:1190–5.
14. The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491:56–65.
15. Morris AP. Fine mapping of type 2 diabetes susceptibility loci. Curr Diab Rep. 2014;14:549.
16. Udler MS, Tyrer J, Easton DF. Evaluating the power to discriminate between highly correlated SNPs in genetic association studies. Genet Epidemiol. 2010;34:463–8.
17.• Gaulton KJ, Ferreira T, Lee Y, Raimondo A, Mägi R, Reschen ME, et al. Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. Nat Genet. 2015;47:1415–25. **A fine-mapping effort that identifies a single likely non-coding causal variant for the *MTNR1B* locus with a direct link to the potential causal gene and functional mechanism.**
18. Bouatia-Naji N, Bonnefond A, Cavalcanti-Proenca C, Sparsø T, Holmkvsit J, Marchand M, et al. A variant near MTNR1B is associated with increased fasting plasma glucose levels and type 2 diabetes risk. Nat Genet. 2009;41:89–94.
19. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini L, Hu T, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat Genet. 2008;40:638–45.
20. Fogarty MP, Panhuis TM, Vadlamudi S, Buchkovich ML, Mohlke KL. Allele-specific transcriptional activity at type 2 diabetes—associated single nucleotide polymorphisms in regions of pancreatic islet open chromatin at the JAZF1 locus. Diabetes. 2013;62:1756–62.
21. Fogarty MP, Cannon ME, Vadlamudi S, Gaulton KJ, Mohlke KL. Identification of a regulatory variant that binds FOXA1 and FOXA2 at the CDC123/CAMK1D type 2 diabetes GWAS locus. PLoS Genet. 2014;10:e1004633.
22. Claussnitzer M, Dankel SN, Klocke B, Grallert H, Glunk V, Riess H, et al. Leveraging cross-species transcription factor binding site patterns: from diabetes risk loci to disease mechanisms. Cell. 2014;156:343–58.
23.• Claussnitzer M, Dankel SN, Kim K-H, Quon G, Meuleman W, Haugen C, et al. FTO obesity variant circuitry and adipocyte browning in humans. N Engl J Med. 2015;373:895–907. **Using computational phylogenetic module complexity analysis, this study identifies a likely causal variant and functional mechanism at the strongest obesity locus, *FTO*.**
24. Frayling TM, Timpson NJ, Weedon MN, Freathy RM, Lindgren CM, Perry JRB, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science. 2007;316:889–94.
25. Loos RJF, Yeo GSH. The bigger picture of FTO—the first GWAS-identified obesity gene. Nat Rev Endocrinol. 2014;10:51–61.
26. Thomsen SK, McCarthy MI, Gloyn AL. The importance of context: uncovering species- and tissue-specific effects of genetic risk variants for type 2 diabetes. Front Endocrinol. 2016;7.
27.• Weedon MN, Cebola I, Patch A, Flanagan SE, De Franco E, Caswell R, et al. Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. Nat Genet. 2014;46:61–4. **Important work at the *PTF1A* locus, which shows that studying disease-relevant tissues and developmental stages can be critical to identifying disease mechanisms.**
28. Lohmueller KE, Sparsø T, Li Q, Andersson E, Korneliussen T, Albrechtsen A, et al. Whole-exome sequencing of 2000 Danish individuals and the role of rare coding variants in type 2 diabetes. Am J Hum Genet. 2013;93:1072–86.
29. Huyghe JR, Jackson AU, Fogarty MP, Buchkovich ML, Stringham HM, Sim X, et al. Exome array analysis identifies novel loci and low-frequency variants for insulin processing and secretion. Nat Genet. 2013;45:197–201.
30. Mahajan A, Sim X, Ng HJ, Manning A, Rivas MA, Highland HM, et al. Identification and functional characterization of G6PC2 coding variants influencing glycemic traits define an effector transcript at the G6PC2-ABCB11 locus. PLoS Genet. 2015;11:1–25.
31.• Flannick J, Thorleifsson G, Beer N, Jacobs SBR, Grarup N, Burtt NP, et al. Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. Nat Genet. 2014;46:357–63. **This study discovers several independent loss-of-function coding variants in *SLC30A8* associated with reduced risk of T2D, and thus highlights a possible treatment strategy.**
32. Thomsen SK, Gloyn AL. Human genetics as a model for target validation: finding new therapies for diabetes. Diabetologia. 2017;60:960–70.
33. Andersen MK, Pedersen CT, Moltke I, Hansen T, Albrechtsen A, Grarup N. Genetics of type 2 diabetes: the power of isolated populations. Curr Diab Rep. 2016;16.
34. Hatzikotoulas K, Gilly A, Zeggini E. Using population isolates in genetic association studies. Br Funct Genom. 2014;13:371–7.
35.• Moltke I, Grarup N, Jørgensen ME, Bjerregaard P, Treebak JT, Fumagalli M, et al. A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes. Nature. 2014;512:190–3. **Demonstrating the importance of bottleneck populations, this study identifies a common coding variant in *TBC1D4* with the largest effect size for a common T2D risk variant.**
36. Manning A, Highland HM, Gasser J, Sim X, Tukiainen T, Fontanllas P, et al. A low-frequency inactivating AKT2 variant enriched in the Finnish population is associated with fasting insulin levels and type 2 diabetes risk. Diabetes. 2017;db161329.
37. Tan K, Kimber WA, Luan J, Soos MA, Semple RK, Wareham NJ, et al. Analysis of genetic variation in Akt2/PKB-β in severe insulin resistance, lipodystrophy, type 2 diabetes, and related metabolic phenotypes. Diabetes. 2009;56:714–9.
38. Hussain K, Challis B, Rocha N, Payne F, Minic M, Thompson A, et al. An activating mutation of AKT2 and human hypoglycemia. Science. 2011;334:474.
39. Mahajan A, Wessel J, Willems S, Zhao W, Robertson NR, Chu AY, et al. Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes. bioRxiv. 2017.

40. Nica AC, Dermitzakis ET. Expression quantitative trait loci: present and future. Phil Trans R Soc B. 2013;368.
41. GTEx Consortium. The genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015;348:648–60.
42. Dimas AS, Lagou V, Barker A, Knowles JW, Mägi R, Hivert M, et al. Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. Diabetes. 2014;63:2158–71.
43. Fadista J, Vikman P, Ottosson E, Guerra Mollet I, Lou Esguerra J, Taneera J. Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism. Proc Natl Acad Sci U S A. 2014;111:13924–9.
44.• van de Bunt M, Manning Fox JE, Dai X, Barrett A, Grey C, Li L, et al. 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. PLoS Genet. 2015;11:1–21. **This study makes an important contribution to the intersection of GWAS loci with islet *cis*-eQTLs, and highlights *ZMIZ1* as a likely causal gene through coincidence of signals.**
45. Thomsen SK, Ceroni A, van de Bunt M, Burrows C, Barrett A, Scharfmann R, et al. Systematic functional characterization of candidate causal genes for type 2 diabetes risk variants. Diabetes. 2016;65:3805–11.
46. Small KS, Hedman ÅK, Grundberg E, Nica AC, Kong A, Thorsteindottir U, et al. Identification of an imprinted master trans-regulator at the KLF14 locus related to multiple metabolic phenotypes. Nat Genet. 2011;43:561–4.
47. Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-michaels K, Carroll RJ, et al. A gene-based association method for mapping traits using reference transcriptome data. Nat Genet. 2015;47:1091–8.
48. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet. 2016;48:481–7.
49. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BWJH, et al. Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet. 2016;48:245–52.
50. Mancuso N, Shi H, Goddard P, Kichaev G, Gusev A, Pasaniuc B. Integrating gene expression with summary association statistics to identify genes associated with 30 complex traits. Am J Hum Genet. 2017;100:473–87.
51. Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K, Tuschl T. Duplexes of 21 ± nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature. 2001;411:494–8.
52. Paddison PJ, Silva JM, Conklin DS, Schlabach M, Li M, Aruleba S, et al. A resource for large-scale screens in mammals. Nature. 2004;428:427–31.
53. Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Heckl D, et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. Science. 2014;343:84–7.
54. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell. 2013;152:1173–83.
55. Pritsker M, Ford NR, Jenq HT, Lemischka IR. Genomewide gain-of-function genetic screen identifies functionally active genes in mouse embryonic stem cells. Proc Natl Acad Sci U S A. 2006;103:6946–51.
56. Gilbert LA, Horlbeck MA, Adamson B, Jacqueline E, Chen Y, Whitehead EH, et al. Genome-scale CRISPR-mediated control of gene repression and activation. Cell. 2014;159:647–61.
57.• Shalem O, Sanjana NE, Zhang F. High-throughput functional genomics using CRISPR–Cas9. Nat Rev Genet. 2015;16:299–311. **A review discussing key concepts, comparison with RNAi screening, prospects and future challenges of genome-wide CRISPR/Cas9 knockout and modulation screens.**
58. Pappalardo Z, Chopra DG, Hennings TG, Richards H, Choe J, Yang K, et al. A whole genome RNA interference screen reveals a role for Spry2 in insulin transcription and the unfolded protein response. Diabetes. 2017;66:1703–12.
59. Shu XO, Long J, Cai Q, Qi L, Xiang YB, Cho YS, et al. Identification of new genetic risk variants for type 2 diabetes. PLoS Genet. 2010;6:e1001127.
60. Akerman I, Tu Z, Beucher A, Schadt E, Ravassard P, Ferrer J, et al. Human pancreatic β cell lncRNAs control cell-specific regulatory networks. Cell Metab. 2017;25:400–11.
61. Ahlgren U, Jonsson L, Simu K, Edlund H. β-cell-specific inactivation of the mouse Ipf1/Pdx1 gene results in loss of the β-cell phenotype and maturity onset diabetes. Genes Dev. 1998;12:1763–8.
62. Beer NL, Gloyn AL. Genome-edited human stem cell-derived beta cells: a powerful tool for drilling down on type 2 diabetes GWAS biology. F1000Res. 2016;5:1711.
63. Pagliuca F, Millman J, Guertler M, Segel M. Generation of functional human pancreatic β cells in vitro. Cell. 2014;159:428–39.
64. Gaulton KJ. Mechanisms of type 2 diabetes risk loci. Curr Diab Rep. 2017;in press.
65. Nielsen T, Sparso T, Grarup N, Jorgensen T, Pisinger C, Witte DR, et al. Type 2 diabetes risk allele near CENTD2 is associated with decreased glucose-stimulated insulin release. Diabetologia. 2011;54:1052–6.
66. Kulzer JR, Stitzel ML, Morken MA, Huyghe JR, Fuchsberger C, Kuusisto J, et al. A common functional regulatory variant at a type 2 diabetes locus upregulates ARAP1 expression in the pancreatic beta cell. Am J Hum Genet. 2014;94:186–97.
67.•• Carrat GR, Hu M, Nguyen-Tu M-S, Chabosseau P, Gaulton KJ, De Van BM, et al. Decreased STARD10 expression is associated with defective insulin secretion in humans and mice. Am J Hum Genet. 2017;100:238–56. **This study demonstrates how comprehensive integration of multiple experimental strategies and data sets can be used to infer causal mechanisms, and highlights *STARD10* at the *STARD10/ARAP1* locus as the likely causal gene.**
68. Rusu V, Hoch E, Mercader JM, Tenen DE, Gymrek M, Hartigan CR, et al. Type 2 diabetes variants disrupt function of SLC16A11 through two distinct mechanisms article type 2 diabetes variants disrupt function of SLC16A11 through two distinct mechanisms. Cell. 2017;170:199–212.e20.
69. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489:57–74.
70. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015;518:317–30.

METHOD ARTICLE

REVISED A CRISPR/Cas9 genome editing pipeline in the EndoC-βH1 cell line to study genes implicated in beta cell function [version 2; peer review: 3 approved, 1 approved with reservations]

Antje K. Grotz [1], Fernando Abaitua[2], Elena Navarro-Guerrero [3], Benoit Hastoy [1], Daniel Ebner[3], Anna L. Gloyn [1,2,4]

[1]Oxford Centre for Diabetes, Endocrinology & Metabolism, University of Oxford, Oxford, OX3 7LE, UK
[2]Wellcome Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK
[3]Target Discovery Institute, University of Oxford, Oxford, OX3 7FZ, UK
[4]Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, OX3 7LE, UK

v2 **First published:** 08 Oct 2019, **4**:150
https://doi.org/10.12688/wellcomeopenres.15447.1
Latest published: 29 Apr 2020, **4**:150
https://doi.org/10.12688/wellcomeopenres.15447.2

Abstract

Type 2 diabetes (T2D) is a global pandemic with a strong genetic component, but most causal genes influencing the disease risk remain unknown. It is clear, however, that the pancreatic beta cell is central to T2D pathogenesis. *In vitro* gene-knockout (KO) models to study T2D risk genes have so far focused on rodent beta cells. However, there are important structural and functional differences between rodent and human beta cell lines. With that in mind, we have developed a robust pipeline to create a stable CRISPR/Cas9 KO in an authentic human beta cell line (EndoC-βH1). The KO pipeline consists of a dual lentiviral sgRNA strategy and we targeted three genes (*INS*, *IDE*, *PAM*) as a proof of concept. We achieved a significant reduction in mRNA levels and complete protein depletion of all target genes. Using this dual sgRNA strategy, up to 94 kb DNA were cut out of the target genes and the editing efficiency of each sgRNA exceeded >87.5%. Sequencing of off-targets showed no unspecific editing. Most importantly, the pipeline did not affect the glucose-responsive insulin secretion of the cells. Interestingly, comparison of KO cell lines for *NEUROD1* and *SLC30A8* with siRNA-mediated knockdown (KD) approaches demonstrate phenotypic differences. *NEUROD1*-KO cells were not viable and displayed elevated markers for ER stress and apoptosis. *NEUROD1*-KD, however, only had a modest elevation, by 34%, in the pro-apoptotic transcription factor CHOP and a gene expression profile indicative of chronic ER stress without evidence of elevated cell death. On the other hand, *SLC30A8*-KO cells demonstrated no reduction in K_{ATP} channel gene expression in contrast to siRNA silencing. Overall, this strategy to efficiently create stable KO in the human beta cell line EndoC-βH1 will allow for a better understanding of genes involved in beta cell dysfunction, their underlying functional mechanisms and T2D pathogenesis.

Open Peer Review

Reviewer Status ✓ ? ✓ ✓

	Invited Reviewers			
	1	**2**	**3**	**4**
version 2 (revision) 29 Apr 2020			✓ report	
version 1 08 Oct 2019	✓ report	? report	✓ report	✓ report

1 **Hindrik Mulder**, Skåne University Hospital, Lund University, Lund, Sweden

2 **Shuibing Chen**, Weill Cornell Medical College, New York, USA

3 **Inês Cebola**, Imperial College London, London, UK

4 **Bridget K. Wagner**, Broad Institute of MIT and Harvard, Cambridge, USA

Any reports and responses or comments on the article can be found at the end of the article.

Keywords
EndoC-βH1, human beta cells, CRISPR/Cas9, gene knockout, Type 2 Diabetes, insulin secretion, NEUROD1

Corresponding author: Anna L. Gloyn (anna.gloyn@drl.ox.ac.uk)

Author roles: Grotz AK: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Abaitua F**: Conceptualization, Investigation, Methodology, Supervision, Validation, Writing – Review & Editing; **Navarro-Guerrero E**: Conceptualization, Investigation, Methodology, Resources, Writing – Review & Editing; **Hastoy B**: Conceptualization, Supervision, Writing – Review & Editing; **Ebner D**: Conceptualization, Resources, Supervision, Writing – Review & Editing; **Gloyn AL**: Conceptualization, Funding Acquisition, Project Administration, Supervision, Writing – Review & Editing

Competing interests: ALG has received honoraria from Novo Nordisk and Merck.

Grant information: ALG is a Wellcome Trust Senior Fellow Basic Biomedical Science. ALG is funded by the Wellcome Trust (095101, 200837), Medical Research Council (MR/L020149/1), European Union Horizon 2020 Programme (T2D Systems), and NIH (U01-DK105535; U01-DK085545) and by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. AG is a Radcliffe Department of Medicine Scholar.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Grotz AK, Abaitua F, Navarro-Guerrero E *et al.* **A CRISPR/Cas9 genome editing pipeline in the EndoC-βH1 cell line to study genes implicated in beta cell function [version 2; peer review: 3 approved, 1 approved with reservations]** Wellcome Open Research 2020, **4**:150 https://doi.org/10.12688/wellcomeopenres.15447.2

First published: 08 Oct 2019, **4**:150 https://doi.org/10.12688/wellcomeopenres.15447.1

REVISED **Amendments from Version 1**

This version of the manuscript addresses all comments and suggestions by the reviewers. The conclusions, figures and data of the paper are unchanged. The following changes were made to the text:

- Clarifying the main objective of the manuscript (introduction)

- Expanding the literature review to include Lawlor *et al.* Cell Reports 2019 (introduction)

- Adding 1) that no difference in ER stress was observed between WT and EV cells; and 2) information about the stability of KO cells (main text)

- Discussing 1) seed-based microRNA-like off-target effects from siRNA; 2) the relevance of large deletions in KO cells for regulatory elements; and 3) the contribution of experimental setups to the observed differences for KD and KO cells (discussion)

The changes have been clarified further in responses to the reviewers' specific comments.

Any further responses from the reviewers can be found at the end of the article

Introduction

Type 2 diabetes (T2D) affects around 400 million people worldwide and is a complex disease with genetic and non-genetic risk factors[1]. Genome-wide association studies (GWAS) have so far identified more than 240 loci which are robustly associated with disease risk[2–5]. The vast majority of these exert their impact on T2D-risk through the pancreatic beta cell and therefore authentic human beta cell models are essential for functional follow-up studies[6].

A lack of a stable and functional human beta cell line, restricted access to human cadaveric pancreatic islets and functional limitations of induced pluripotent stem cells (iPSC)-derived beta-like cells have long been a challenge in understanding beta cell biology. Meanwhile, rodent beta cell lines have provided valuable insights into beta cell function and pathophysiology[7–9]. Although they share many similarities with human beta cells, there are also fundamental structural, transcriptional and functional differences, aside from having a distinctive genetic background. Human pancreatic islets have a substantially different architecture than rodent islets as they have fewer beta cells, a mixed cell distribution throughout the whole islet, with alpha, beta and delta cells being adjacent to each other and alpha cells clustering around blood vessels[10,11]. Rodent islets on the other hand have a higher vascular density and are made up of a distinct beta cell core and non-beta cell mantle[10,12]. Transcriptomic analysis in purified beta cells from mice and human demonstrated a set of common core beta cell genes[13,14]. However, the studies also highlighted a substantial number of uniquely expressed genes in either species and significantly differentially expressed genes such as *GAD2, IAPP, MAFB* and *PPARG*, some of which are involved in Type 1 (T1D) and T2D pathology[13–15]. Differences in key components of the glucose-stimulated insulin secretion (GSIS) pathway emphasise unique functional signatures[16]. The principal glucose transporter in rodent beta cells is *SLC2A2*, whereas human beta cells mainly utilise *SLC2A1* and *SLC2A3*, leading to distinct glucose uptake dynamics[17–19]. Furthermore, rodent beta cells express two insulin genes, *INS1* and *INS2*, which is in contrast to human beta cells, which only have one insulin gene[20]. Other distinguishing factors are the proliferative capacity and ion channel composition of rodent cells compared to human beta cells[21–25].

In 2011, Scharfmann and colleagues released the EndoC-βH1 cell line, a human beta cell line which opened up the possibility of studying human beta cell physiology and pathology *in vitro* and provided a valuable alternative to rodent beta cell lines[26]. To generate this cell line, fetal pancreatic buds were transduced with oncogene simian virus 40 large tumour antigen (SV40LT) and human telomerase reverse transcriptase (hTERT). Between each transduction, the cells were transplanted into SCID mice to expand and form insulinomas. The isolated and passaged cells were able to secrete insulin in response to different glucose and secretagogues stimulation, expressed key beta cell markers and were negative for other pancreatic cell markers like glucagon[26]. Insulin content is a magnitude lower than in primary human beta cells, but secreted insulin as percentage of content and the stimulation index are in the same range as for pancreatic islets[16,27]. Multiomic profiling in EndoC-βH1 cells including epigenomic and transcriptomic maps largely recapitulate primary human islets signatures and along with their similar electrophysiological properties, EndoC-βH1 are therefore a representative model of human beta cells and physiological insulin secretion[28–30]. Further independent investigations have also demonstrated their suitability for both high-throughput screening[31,32] and individual gene function studies[33–35].

Robust protocols for generating gene KO using CRISPR/Cas9 in EndoC-βH1 studies have not yet been described. This genome editing tool has revolutionised genetic manipulations by being an easily programmable RNA-guided endonuclease[36–39]. The application in EndoC-βH1, however, is not straightforward, as their proliferation rate is low and they are very sensitive to seeding densities. It is thus not possible to expand a culture from a single cell, which precludes the generation of a modified clonal cell line. Furthermore, the cells have a very low transfection efficiency, batch-to-batch variation and have to be closely monitored across passages to ensure their beta cells characteristics. A recent study created an *HNF1A* KO cell line in EndoC-βH1 using CRISPR/Cas9, this cell line, however, does not demonstrate complete *HNF1A* depletion and has not been fully characterised[40].

Despite the technical challenges of this cell line, we have successfully developed a lentiviral-based pipeline to create stable non-clonal CRISPR/Cas9 KO cell lines in EndoC-βH1 and have performed genomic and functional characterisation for several proof of concept genes. This CRISPR/Cas9 pipeline and its resulting KO cell lines could be a valuable tool in understanding human beta cell function and genes underlying the pathology for both T2D and T1D.

Methods

Cloning of individual sgRNA into plentiCRISPRv2

plentiCRISPRv2 was purchased from Addgene (#52961)[41] and sgRNA sequences were retrieved from the TKO Library v3[42]. Two sgRNAs per gene were chosen based on highest specificity and lowest off-target score which were evaluated on CRISPOR.org[43] (Table 1). BsmBI compatible tails, 5'CACCGX3' and 5'AAACYC3', with X and Y being complementary sequences to the sgRNA, were added to each oligonucleotide. plentiCRISPRv2 vector was digested with FastDigest BsmBI (Fermentas) for 30 min at 37°C and gel-purified using a 0.8% agarose gel. sgRNA oligos were annealed (1 µl of each 100 µM stock) and phosphorylated using T4 PNK (NEB) for 30 min at 37°C, 5 min at 95°C, then the heating block was shut off to let the samples could cool down to room temperature (RT). 20 ng BsmBI digested plentiCRISPRv2 and 2 µl of 1:100 diluted annealed sgRNA oligonucleotides were ligated using Quick Ligase (NEB) for 1 h at RT. Next, 5 µl of the ligation reaction were transformed into Stbl3 competent cells and successful sgRNA insertion was confirmed using Sanger sequencing.

Cell culture

EndoC-βH1 cells were cultured as previously described and passaged every 7 days[26]. They were grown in culture vessels coated with 2 µg/ml Fibronectin and 1% extracellular matrix (ECM) (Sigma-Aldrich) and cultured in DMEM containing 5.5 mM glucose (Gibco), 2% bovine serum albumin (BSA), 2 mM glutamine, 10 mM nicotinamide, 100 international units (U)/ml penicillin, 100 µg/ml streptomycin (P/S), 50 µM β-2-mercaptoethanol, 5.5 µg/ml transferrin and 6.6 ng/ml sodium selenite (all Sigma-Aldrich).

Lenti-X HEK293T cells (Clontech) were cultured in DMEM 6429 (Sigma-Aldrich) containing 10% fetal calf serum, 100 U/ml penicillin and 100 µg/ml streptomycin. All cells were tested negative for mycoplasma and grown at 37°C and 5% CO_2.

Lentiviral production

Lenti-X HEK293T cells were grown to 80% confluency in T175 flasks and co-transfected with lentiviral packaging vectors in P/S free media. The transfection mix consisted of pMD2.G (6.85 µg) (Addgene #12259), psPAX2 (10.3 µg) (Addgene #12260), the respective cloned plentiCRISPRv2 (12.85 µg), 2 ml of JetPrime buffer and 60 µl of JetPrime transfection reagent (Polyplus transfection) per flask. After 15 min incubation at RT, the transfection mix was added to the cells and media was replaced after 16 h into fresh complete culture media. Supernatant containing viral particles was collected 48 h after transfection, spun down for 5 min at 2000 rpm and filtered through a 0.45-µm filter. Supernatant was ultracentrifuged for 2 h at 4°C and 29000 rpm in a swinging-bucket rotor. The virus pellet was resuspended in 1.5% BSA in PBS, aliquoted and stored at -80°C.

Functional lentiviral titer

At 48 h before transduction, EndoC-βH1 were plated at 20,000 cells per well in a 96-well plate. A viral dilution curve ranging from 1:50 to 1:6400 was prepared in 100µl P/S free media and the cells were infected for 6 h. After 48 h, media was changed on half of the wells per dilution into 4 µg/µl puromycin and the cells were incubated for 7 days. Cell viability was analysed using the CyQUANT Direct Cell Proliferation assay (Invitrogen). The puromycin selected cell counts were normalised to their respective non-selected controls to determine the percentage of survival, which represents transduced cells. The functional titer in transducing units (TU)/µl can be calculated based on:

(1)

$$\mathrm{TU}/\mu\mathrm{l} = \frac{\#\mathrm{Cells} \times m}{\mathrm{Virus}\ (\mu\mathrm{l})\ \mathrm{used\ in\ transduction}}$$

The probability that a cell is infected by a certain number of viral particles at a given multiplicity of infection (MOI) (m) can be modelled using the Poisson distribution (PD). Simplifying the original PD equation, gives the following:

(2)

$$P(n>0) = 1 - e^{-m}$$

with $P(n>0)$ being the probability that a cell gets infected by at least one viral particle[44]. A MOI of 0.3 would lead to ~26% transduced cells, most of them being infected by a single viral

Table 1. sgRNA sequences for target genes.

sgRNA	Target gene	Target region	Sequence (5' → 3')
sgRNA 1	PAM	Exon 1	GAACTAGCAGGCTAGGGACG
sgRNA 2	PAM	Exon 12	GTTCAGAACCATACCACCAG
sgRNA 3	IDE	Exon 1	TACCCACACAGGCGCTCCGG
sgRNA 4	IDE	Exon 8	CATTAATGTGGACTTGACCG
sgRNA 5	INS	Exon 2	CACAATGCCACGCTTCTGCA
sgRNA 6	INS	Exon 2	CATCTGCTCCCTCTACCAGC
sgRNA 7	NEUROD1	Exon 2	CTTGCAAAGCGTCTGAACGA
sgRNA 8	NEUROD1	Exon 2	GCTGCGCTGTAGGCGTGCGG
sgRNA 9	SLC30A8	Exon 2	GTGTCCCAGAGAGAGACCAG
sgRNA 10	SLC30A8	Exon 8	GCACTCACTCACCATTCAGA

particle, therefore this MOI is a good constant in determining the functional titer. To determine the MOI relative to the virus (µl) used in transduction in Equation 1, a linear regression for the percentage of alive cells against the amount of infected virus was performed in the linear, unsaturated range of the puromycin selection curve. The amount of virus needed for a MOI of 0.3 was then calculated by inserting 26% as the percentage of alive cells and solving the linear equation for the amount of virus (µl) needed in the transduction. Along with the number of plated cells, the TU/µl could then be determined.

Generation of EndoC-βH1 KO cell lines

To generate stable CRISPR KO lines, cells were transduced at a MOI of 8 which was calculated based on the functional titer. They were selected in 4 µg/µl puromycin for 7 days, with media changes if necessary, to remove dead cells and add fresh puromycin. After selection, cells were grown in normal EndoC-βH1 culture medium and passaged weekly.

Insulin secretion assay

Cells were starved overnight in 2.8 mM glucose followed by 30 min starvation in 0 mM glucose and stimulation for 1 h in one of the following conditions: 2.8 mM, 5.6 mM, 11 mM, 15 mM, 20 mM, 25 mM glucose, 15 mM glucose + 100 µM tolbutamide or diazoxide. All conditions were prepared in glucose-free EndoC-βH1 culture medium. Supernatant was collected and cells were lysed in acid ethanol to collect insulin content. Secreted and intracellular insulin were measured using the Insulin (human) AlphaLISA Detection Kit and the EnSpire Alpha Plate Reader (both Perkin Elmer), diluted 1:10 and 1:200, respectively. Secreted insulin was displayed as percentage of insulin content and insulin content was normalised to cell count, which was measured using the CyQUANT Direct Cell Proliferation assay (Invitrogen).

siRNA silencing

Cells were transfected in 6-well plates at 24 h after plating with siRNAs at a final concentration of 15 nM (SMARTpool ON-TARGETplus human NeuroD1 #L-008667-00 and SLC30A8 #L-007529-01, non-targeted control pool #D-001810-10, Dharmacon). The required amount of siRNA (µM) (Gibco) was prepared in Opti-MEM reduced serum-free medium to constitute 0.1% of the volume. A Lipofectamine RNAiMAX (Invitrogen) mix was prepared to account for 0.4% of the same total volume and both transfection mixes were incubated for 5 min at RT. The RNAiMAX and siRNA mix were pooled and further incubated for 20 min before they were added dropwise to the cultured cells. Cells were harvested 72 h past transfection for RNA and protein extraction.

Gene expression analyses

RNA was extracted using the TRIzol reagent (Invitrogen) following manufacturer's instructions. First-strand cDNA was synthesised using the Super Script III First-Strand Synthesis System (Invitrogen), oligo(dT) primer and 50–500 ng total RNA as input. Quantitative PCR (qPCR) to measure gene expression levels were performed with TaqMan Gene Expression Assays and TaqMan Gene Expression Master Mix on a 7900HT (all Applied Biosystems) using the following thermocycling conditions: 2 min at 50°C, 10 min at 95°C and 40 cycles of 15 sec at 95°C and 1 min at 60°C (Table 2). TaqMan probes with binding sites outside of the regions targeted by sgRNAs were used. Ct values were analysed using the ΔΔCt method and target genes were normalised to three housekeeping genes (*TBP*, *PPIA* and *GAPDH*).

Western blot analyses

Cell pellets for protein analyses were lysed in RIPA buffer (50 mM Tris pH 7.4, 150 mM NaCl, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS) containing 1x protease inhibitor cocktail (Roche). Protein concentration was quantified by DC protein assay (Bio-Rad) and 10 µg of protein per lane were prepared. Lysates were denatured at 80°C for 10 min and run on a Mini-PROTEAN TGX 4–20% precast gel (Bio-Rad) at 300 V for 15 min. The gel was activated on a ChemiDoc MP Imaging System and transferred to a Trans-Blot Turbo polyvinylidene fluoride (PVDF) membrane using the Trans-Blot Turbo Transfer System (all Bio-Rad). Membranes were blocked in 3% BSA for 1 h at RT, incubated with primary antibodies overnight at 4°C followed by a 1 h incubation at RT with secondary antibodies (antibodies are given in Table 3). The membranes were subsequently incubated for 4 min at RT with Clarity Western enhanced chemiluminescence (ECL) reagent and imaged on the ChemiDoc MP Imaging System (Bio-Rad). To normalise for protein loading, the membrane was further incubated with a loading control antibody of appropriate size (tubulin or GAPDH) (Table 3). Western Blot images were quantified using Image Lab 6.0 software (Bio-Rad). Protein bands of interest were normalised to a loading control on the same blot and displayed relative to a control sample.

PCR and sequencing analyses

Genomic DNA for PCR amplification was extracted using the NucleoSpin Tissue extraction kit (Macherey-Nagel). PCR reactions were prepared containing the following components per sample: 2 µl Immobuffer (10x), 0.6 µl $MgCl_2$ (50 mM), 1 µl each of forward and reverse primer (10 µM) (Table 4), 0.4 µl

Table 2. TaqMan gene expression assays.

TaqMan Probe	Assay details	Target region
INS	Hs00355773_m1	Exon 1–2
IDE	Hs00610452_m1	Exon 24–25
PAM	Hs01084034_m1	Exon 22–23
NEUROD1	Hs01922995_s1	Exon 2
GAPDH	Hs02786624_g1	Exon 8
TBP	Hs00427620_m1	Exon 3–4
PPIA	Hs01634221_s1	Exon 1
SLC30A8	Hs00545182_m1	Exon 2–3
DDIT3	Hs99999172_m1	Exon 1–2
XBP1s	Hs03929085_g1	Exon 5
HSPA5	Hs00607129_gH	Exon 1–2
ATF6	Hs00232586_m1	Exon 6–7
ATF4	Hs00909569_g1	Exon 1–2

Table 3. Antibody specifications.

Antibody	Company, catalogue number	Dilution	Species	RRID/Ref
INS	Santa Cruz, sc-377071	1:1000	Mouse monoclonal	AB_2800506
IDE	Santa Cruz, sc-393887	1:1000	Mouse monoclonal	AB_2800507
PAM	Santa Cruz, sc-514110	1:1000	Mouse monoclonal	AB_2800508
NEUROD1	Santa Cruz, sc-46684	1:1000	Mouse monoclonal	AB_671759
Cas9	Santa Cruz, sc-517386	1:1000	Mouse monoclonal	AB_2800509
CHOP	Abcam, ab179823	1:1000	Rabbit monoclonal	AB_10703186
pPERK	Cell Signaling, #3179	1:1000	Rabbit monoclonal	AB_2095853
pIRE1	Abcam, ab48187	1:1000	Rabbit polyclonal	AB_873899
Cleaved Caspase 3	Cell Signaling, #9661	1:500	Rabbit polyclonal	AB_2341188
ZnT8	/	1:1000	Mouse monoclonal	45
β-Tubulin	Santa Cruz, sc-365791	1:2000	Mouse monoclonal	AB_10841919
GAPDH	Abcam, ab181602	1:10 000	Rabbit monoclonal	AB_2630358
α-mouse IgG HRP	Thermo Scientific, 31450	1:2500	Rabbit polyclonal	AB_228427
α-rabbit IgG HRP	Thermo Scientific, 31460	1:2500	Goat polyclonal	AB_228341

Table 4. Primer specifications.

Primer name	Target region	Experiment	Sequence (5' → 3')
LKO1.5R	lentiCRISPRv2	sgRNA sequence	GTTGATAACGGACTAGCCT
lentiCRv2_Cas9	*Cas9*	sgRNA integration	CAGGCCGATGCTGTACTTCT
PAM_sgRNA1	sgRNA1	sgRNA integration	ACACCGGAACTAGCAGGCTA
PAM_sgRNA2	sgRNA2	sgRNA integration	GACGAAACACCGGTTCAGAAC
IDE_sgRNA3	sgRNA3	sgRNA integration	AAACACCGTACCCACACAGG
IDE_sgRNA4	sgRNA4	sgRNA integration	CGCATTAATGTGGACTTGACCG
INS_sgRNA5	sgRNA5	sgRNA integration	CAATGCCACGCTTCTGCAG
INS_sgRNA6	sgRNA6	sgRNA integration	CATCTGCTCCCTCTACCAGC
NEUROD1_sgRNA7	sgRNA 7	sgRNA integration	GACGAAACACCGCTTGCAAA
NEUROD1_sgRNA8	sgRNA 8	sgRNA integration	CTGTAGGCGTGCGGGTTTT
PAM1_F	sgRNA1 target site	Editing efficiency	GCTGGAGGGAGGAAAGCTTC
PAM1_R	sgRNA1 target site	Editing efficiency	TTTTTCTGCACGGGGGACTT
PAM2_F	sgRNA2 target site	Editing efficiency	TTGCTGGCAGATCTAAGGGC
PAM2_R	sgRNA2 target site	Editing efficiency	TCCCTGGCTGAGATTTTCCTC
IDE3_F	sgRNA3 target site	Editing efficiency	AGTCGCCGGATTCCTTTACC
IDE3_R	sgRNA3 target site	Editing efficiency	CTAATGCGGTACCGGCTAGC
IDE4_F	sgRNA4 target site	Editing efficiency	TCCATGAAACAAAGGCCAAGT
IDE4_R	sgRNA4 target site	Editing efficiency	CCCCACTTCTGCACCATCTT
INS5_F	sgRNA5 target site	Editing efficiency	CATCTCTCTCGGTGCAGGAG
INS5_R	sgRNA5 target site	Editing efficiency	TCCCTCTAACCTGGGTCCAG
INS6_F	sgRNA6 target site	Editing efficiency	CCTGTAGGTCCACACCCAGT
INS6_R	sgRNA6 target site	Editing efficiency	AAGACACACAGACGGCACAG
sgRNA1_MED15_F	sgRNA1 off-target	Off-targets	GGCCAAACACACAGAGGAGT
sgRNA1_MED15_R	sgRNA1 off-target	Off-targets	TGGACTTGCCCTCTCTTGAC
sgRNA2_GPM6B_F	sgRNA2 off-target	Off-targets	ATCACTGCAGGGAACTGCTT
sgRNA2_GPM6B_R	sgRNA2 off-target	Off-targets	CAGCACCATCCTCAGATCCT

dNTPs (10 mM), 0.2 µl Immolase DNA Polymerase (5 U/µl) (all Bioline), 4 µl Q-Solution (5x) (Qiagen) and 10.8 µl of the DNA sample (100 ng). The PCR amplification was performed for 10 min at 94°C, 32 cycles of 1 min at 94°C, 1 min at 64°C and 30 sec at 72°C, followed by 10 min at 72°C. DNA samples were run on a 2% agarose gel for 1 h at 120 V and PCR bands were visualised on a GelDoc transilluminator system (Bio-rad). Ahead of sending samples for sequencing, excess primers and nucleotides were removed from PCR reactions using an ExoSAP enzymatic clean up. The reaction was performed with 10 µl PCR product, 0.05 µl ExoI, 0.5 µl SAP, 1 µl SAP buffer (all Affymetrix) and 0.45 µl nuclease-free water and incubated for 30 min at 37°C and 5 min at 95°C. The sequencing reaction was premixed using 13.5 µl nuclease-free water, 1.5 µl ExoSAP treated PCR sample and 2 µl sequencing primer (10 µM) and sent to Eurofins Genomics. Sequence traces were visualised using SnapGene Viewer 4.3 and analysed using TIDE 2.0 and ICE 1.1[46,47].

Statistical analysis

Statistical analyses were performed in Prism 8.0 (GraphPad Software) and data are shown as mean with standard error of the mean (SEM). Values displayed as fold changes were analysed as log-transformed values and statistical tests were performed as indicated in the figure legends. In general, values normalised to a control group such as western blot data were analysed using one-sample Student's t-test and two or more groups were compared using two-sample Student's t-test or one-way analysis of variance (ANOVA) followed by Sidak's multiple comparison test.

Results

A CRISPR/Cas9 pipeline to create EndoC-βH1 KO cells

To demonstrate that this lentiviral CRISPR/Cas9 pipeline can robustly generate KOs in EndoC-βH1, we created KO cell lines for three proof-of-concept genes. We chose genes with known relevance in beta cell function, namely peptidylglycine alpha-amidating monooxygenase (*PAM*)[3,33], insulin-degrading enzyme (*IDE*)[48,49] and insulin (*INS*)[50,51]. In brief, we transduced EndoC-βH1 with lentivirus containing Cas9 and two sgRNAs, selected for successfully transduced cells and characterised the generated heterogeneous KO cell lines (Figure 1A). As a vector system, we chose lentiCRISPRv2 which is a one-plasmid system containing Cas9, a puromycin resistance cassette and the cloned sgRNA (Figure 1B). To increase the KO efficiency in our editing approach, we utilised a dual sgRNA strategy using two sgRNAs in separate lentivirus targeting different parts or exons of each gene (Figure 1C–E)[52]. The sgRNA sequences were retrieved from the genome-wide CRISPR KO library Toronto KnockOut version 3.0 (TKOv3) which are optimized for high on-target efficiency and minimal off-target cutting[42,53]. In the case of *IDE*, only one of the two protein-coding isoforms was targeted. We packaged lentiCRISPRv2 into lentivirus and transduced EndoC-βH1 at a high MOI of 8, ensuring that each cell is infected by several lentivirus and increasing the likelihood to achieve KOs. The cells were selected in 4 µg/ml puromycin to remove untransduced cells. The ideal concentration of puromycin was determined right before selection as EndoC-βH1 cells have different susceptibilities to antibiotics depending on their passage (Figure 1F). After antibiotic selection, the transduced cells are a heterogeneous population having either no edit, an insertion or deletion (indel) from one sgRNA, a large deletion from simultaneous cutting of both sgRNA or two indels from both sgRNA cutting individually (Figure 1A). These stable CRISPR cell lines were routinely cultured like regular EndoC-βH1 cells and their genomic and functional characteristics investigated.

Genomic modifications of EndoC-βH1 KO cells

To characterise the genomic modifications resulting from stable lentiCRISPRv2 integration and CRISPR editing, we analysed sgRNA and Cas9 integration, sgRNA efficiency and potential off-target cutting. A control cell line created with the same lentiviral backbone but without sgRNA was included as empty vector (EV) or Cas9 only control in all experiments. A PCR-based approach (PCR 1) using a sgRNA specific primer and a primer targeting the lentiCRISPRv2 backbone was used to detect sgRNA integration (Figure 2A). Both sgRNAs were detectable in each KO cell line and not in EV control cells indicating successful transduction with both sgRNA lentivirus. Stable Cas9 expression was demonstrated in all cells lines transduced with lentiCRISPRv2 (KO cell lines and EV) but not in untransduced wild-type cells (WT) (Figure 2B). Individual sgRNA editing efficiency was assessed by amplifying sgRNA target sides (PCR 2) and measuring indel frequency using TIDE (Figure 2C)[46]. All sgRNA target sides demonstrate an editing efficiency greater than 87.5%, leaving only around 1% of cells without indels. The sgRNA target sides in *INS*-KO cells (sgRNA 5 and 6) are within range of a single PCR and Sanger sequencing reaction, which makes it possible to assess the frequency of large deletions from simultaneous cutting of both sgRNA using the Inference of CRISPR Edits (ICE) tool[47]. In 33.4% of *INS*-KO cells, the approximately 50 bp region between both sgRNA target sides has been deleted through concurrent sgRNA cutting (Figure 2C). In *PAM*-KO and *IDE*-KO cells, the presence of large deletions between both sgRNAs was demonstrated by performing a PCR with primers on either side of the sgRNA target side (PCR 3) (Figure 2D). If the region between the two sgRNAs is still present, the fragment is too large for PCR amplification but if both sgRNAs cut simultaneously and the region between the sgRNA target sides has been deleted, a PCR product can be amplified. Such PCR product is present in both cell lines, *PAM*-KO and *IDE*-KO, indicating a large deletion of 94 kbp and 66 kbp, respectively, that has not occurred in the WT and EV control cells. In *INS*-KO cells and confirming the results of the ICE analysis, the presence of two bands in contrast to controls indicate the presence of both fragments, a shorter PCR product containing

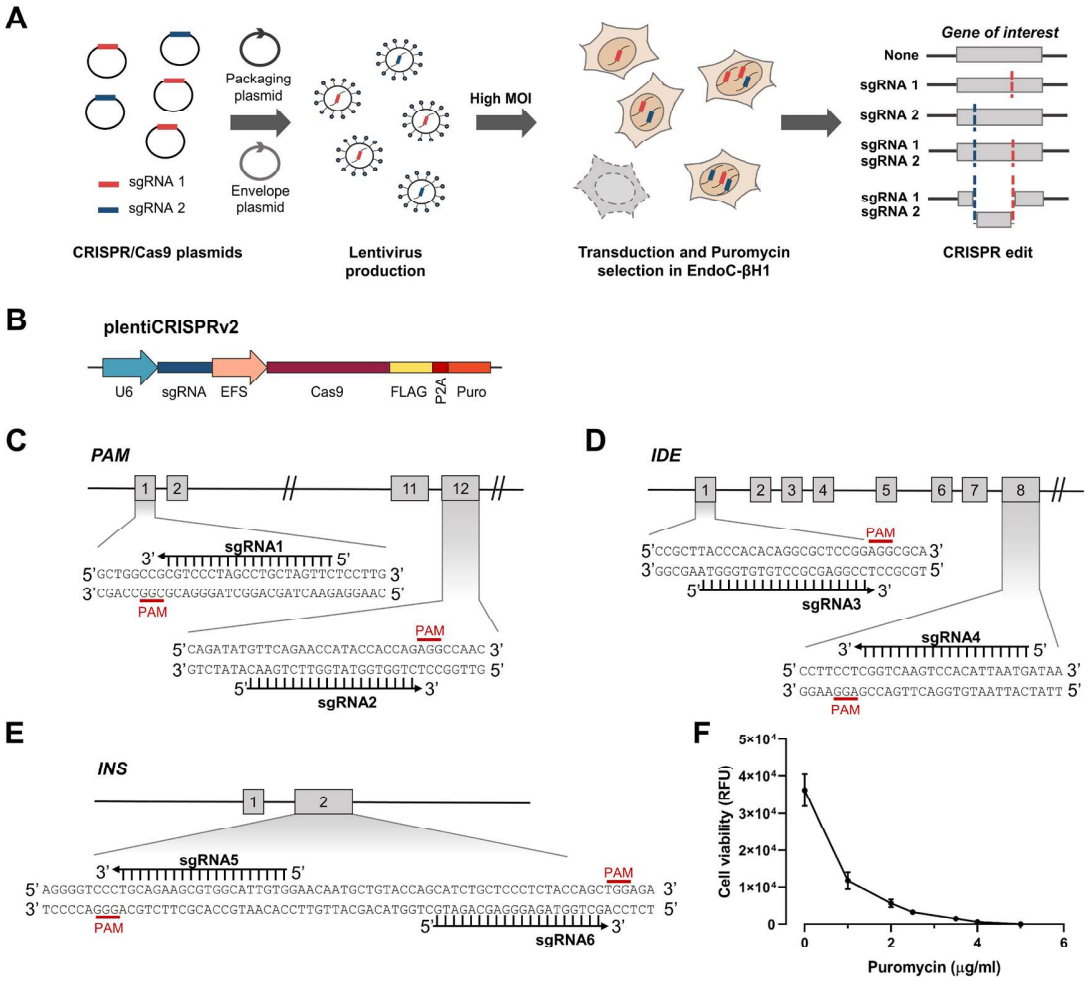


Figure 1. A lentiviral CRISPR/Cas9 pipeline to genome edit EndoC-βH1. (**A**) Strategy to genome edit EndoC-βH1 using a lentiviral dual sgRNA-based approach resulting in a population with cells containing individual or both sgRNA edits. (**B**) CRISPR plasmid, lentiCRISPRv2 containing Cas9 and a single sgRNA. (**C–E**) Gene targeting strategy using two sgRNA per gene for *PAM* (**C**), *IDE* (**D**) and *INS* (**E**). Numbers indicate exons of the respective gene. (**F**) Puromycin kill curve in EndoC-βH1. Relative fluorescence units (RFU) are representative for cell viability. Results are representative of the optimization experiment performed right before selection of the KO cell lines. The graph displays the mean of six technical replicates with standard error of the mean (SEM).

the 50 bp deletion and the normal PCR product which results from cells containing small indels or no edits. Further, we tested if the stable integration of sgRNA and Cas9 in our CRISPR pipeline increased potential off-target activity. We determined the off-target potential by assessing the cutting frequency determination (CFD) score of sgRNA 1 and 2 off-targets in *PAM*-KO cells[54]. The regions with the highest CFD score, introns in *MED15* (0.43) and *GPM6B* (0.54) were sequenced and no off-target activity could be detected (Figure 2E and F). In summary, this CRISPR pipeline in EndoC-βH1 is highly efficient in creating edited populations containing individual indels or large deletions from two double-strand breaks.

Functional characterisation of EndoC-βH1 KO cells

Having established efficient editing of the cells, we next sought to determine if this translated into functional KO cells. We therefore investigated the insulin secretion characteristics of Cas9 expressing cells and determined both mRNA expression and protein levels of the targeted genes in the KO cells. To investigate if the transduction and selection pipeline or general expression of Cas9 affects the functionality of the cells, we compared EV to WT cells and assessed their insulin secretion and content characteristics. The secretory capacities of EV and WT cell lines in response to physiological glucose concentrations stimulation are similar (Figure 3A). Both cell lines were susceptible to the ATP-sensitive potassium (K_{ATP}) channel

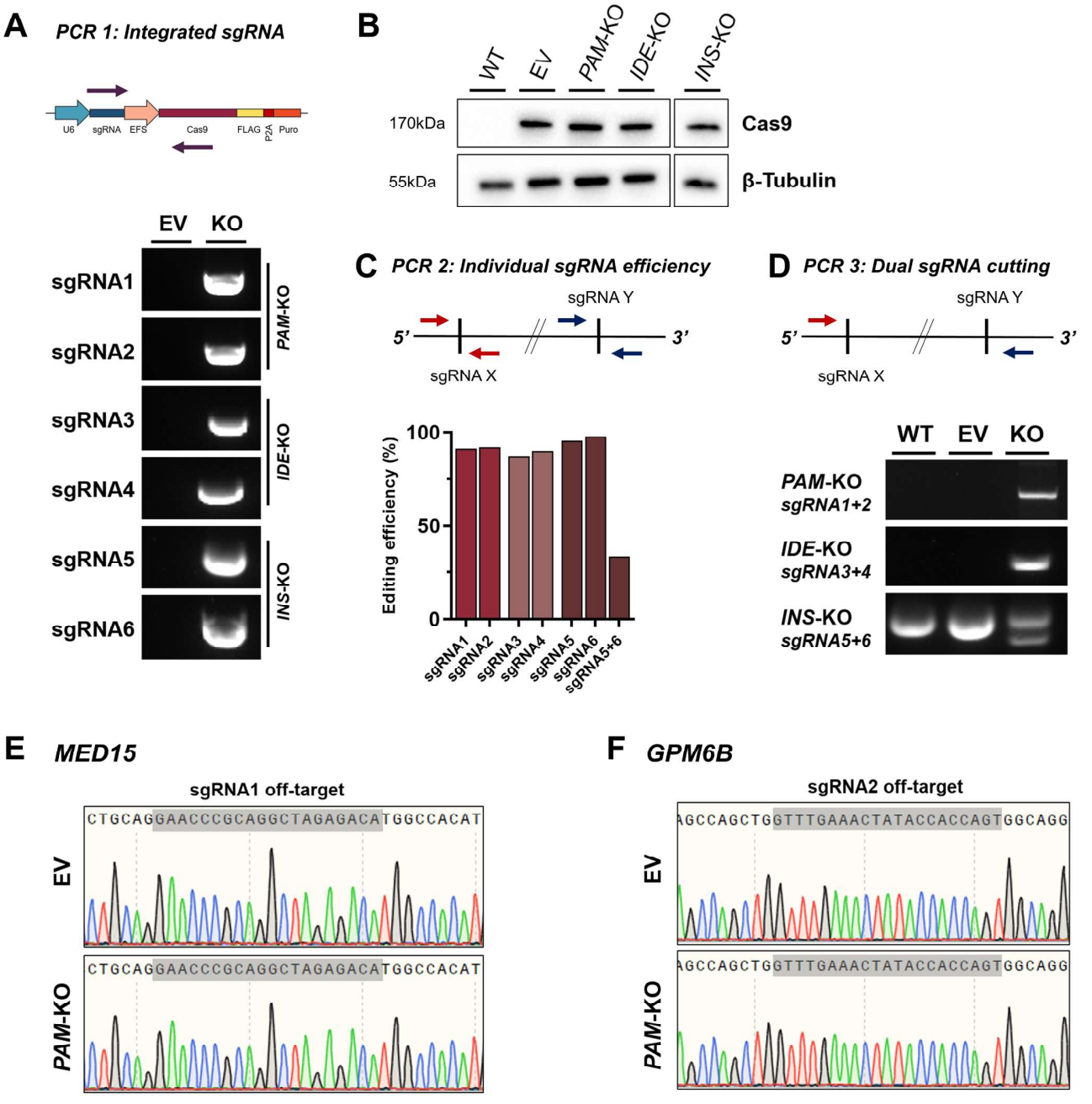


Figure 2. Genomic modifications of EndoC-βH1 KO cell lines. (**A**) Stable genomic integration of lentiCRISPRv2 in transduced EndoC-βH1. *Upper panel:* PCR strategy (PCR 1) using primer specific for each sgRNA together with generic Cas9 primer. PCR product is only present upon successful integration of the specific sgRNA. *Lower panel:* PCR products for individual sgRNAs in control EV cell line or KO cell lines. (**B**) Western blot analysis of Cas9 protein and β-Tubulin in KO cell lines and control cells (WT and EV). (**C**) Editing efficiency of individual sgRNAs. *Upper panel:* PCR strategy (PCR 2) using primer specific for each sgRNA target site. *Lower panel:* PCR products for individual sgRNA target sites were Sanger sequenced and analysed for editing efficiency using TIDE and ICE. (**D**) Editing efficiency of dual sgRNA cutting. *Upper panel:* PCR strategy (PCR 3) using primers flanking the region encompassed by both sgRNAs. PCR product (or shorter fragment) is only present upon deletion of the full region between both sgRNA target sites. *Lower panel:* PCR products indicating large deletion in WT, EV or KO cell lines. (**E**, **F**) Sequencing of top off-target hits in *PAM*-KO cells. Sequencing trace for sgRNA1 off-target *MED15* (**E**) and sgRNA2 off-target *GPM6B* (**F**) are displayed for EV and *PAM*-KO cells.

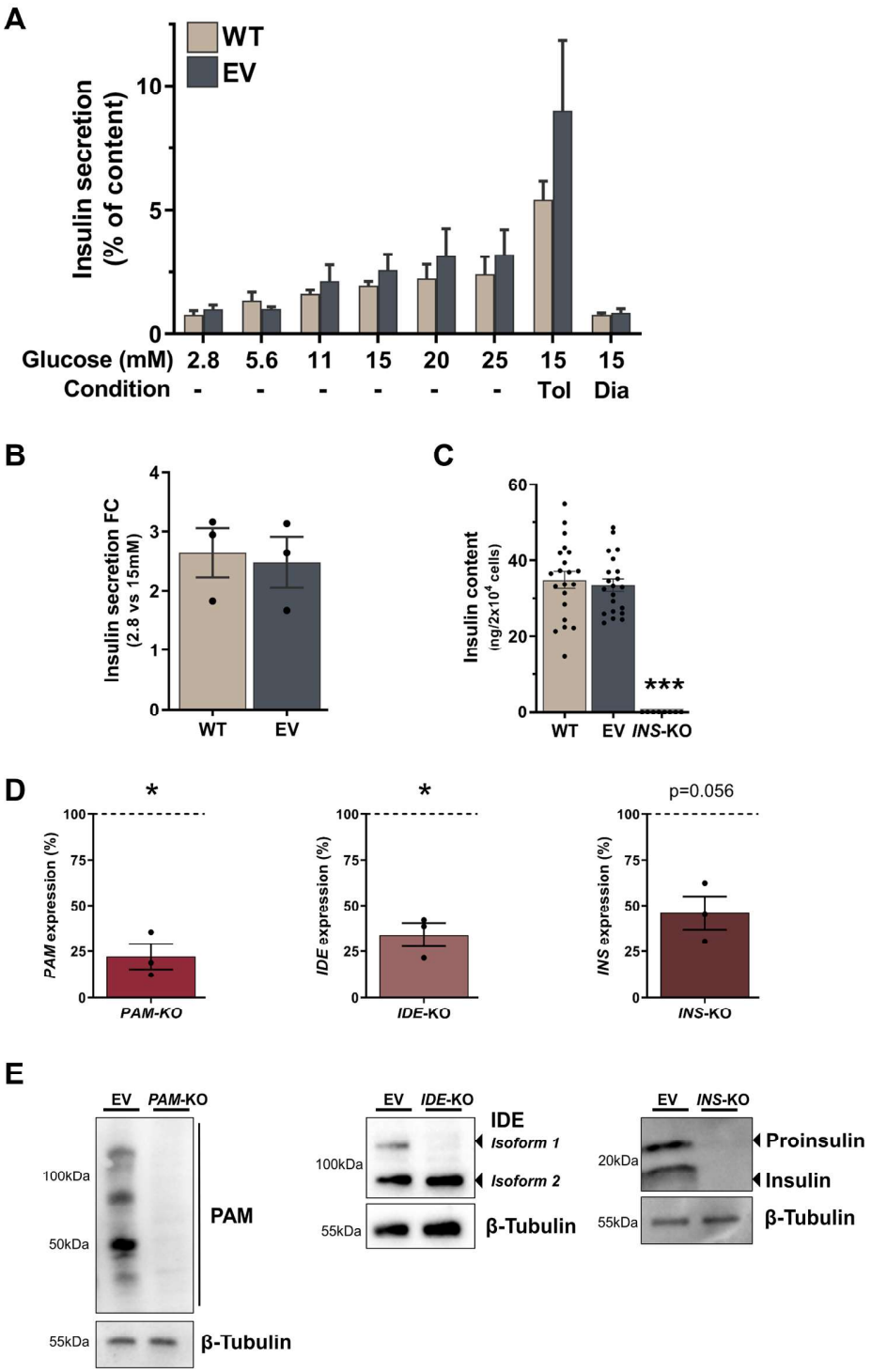


Figure 3. Functional characterisation of EndoC-βH1 KO cell lines. (**A–C**) Insulin secretion as percentage of content in different glucose conditions (**A**), fold change between 2.8 mM and 15 mM glucose stimulation (**B**) and insulin content (**C**) in WT cells compared to EV cells. (**D**) qPCR expression data for *PAM, IDE* and *INS* genes in EV and respective KO cell lines. Values for KO cell lines were normalised to EV within each experiment and fold changes are displayed as percentage of EV, which is indicated as a dotted line at 100%. (**E**) Western blot analysis of PAM, IDE and INS in EV and respective KO cells with β-Tubulin as loading control. All data are mean ± SEM from three independent experiments and for insulin content as 21 replicates from seven conditions in three independent experiments. Fold changes were log-transformed for statistical analysis. P values * < 0.05, ** < 0.01 and *** < 0.001 using two-way ANOVA Sidak's multiple comparison test (**A**), two-sample t-test with Holm-Sidak correction for Diazoxide and Tolbutamide induction and glucose stimulation (**A, B**), one-way ANOVA with Sidak's multiple comparison test for insulin content (**C**) and one-sample t-test for expression data (**D**).

activator and blocker, diazoxide and tolbutamide. Diazoxide reduced insulin secretion by 66.2% and 59.8% (p=0.81), whereas tolbutamide further potentiated insulin secretion 3.5- and 2.8-fold (p=0.94) in WT and EV cells, respectively (Figure 3A). Insulin secretion increased 2.48- vs 2.64-fold (p=0.94) in EV compared to WT cells after stimulation with 15 mM glucose (Figure 3B). Intracellular insulin was equally similar between EV and WT cells, averaging at 33.44 ng and 34.83 ng/2×10^4 cells (p=0.83) (Figure 3C). mRNA levels of targeted genes were measured by RT-qPCR and were significantly decreased in *PAM*-KO and *IDE*-KO by 77.8% (p=0.035) and 66% (p=0.034), respectively. *INS*-KO cells also demonstrate a strong reduction of *INS* transcript by 54.2% (p=0.056) compared to EV control cells (Figure 3D). These reductions of mRNA levels indicate the successful introduction of frameshift mutation and the generation of a premature stop codon (PSC) followed by degradation through nonsense-mediated mRNA decay (NMD). The detection of residual transcript is in line with previous studies demonstrating incomplete NMD in KO genes after induction of frameshift mutations in coding regions of the gene[55]. Due to the heterogeneity of the KO cell population however, it cannot be ruled out that the detected transcript also includes mRNA from unedited cells or cells that do not contain a frameshift mutation or PSC. In *IDE*-KO cells, only one protein-coding isoform was targeted for CRISPR editing. However, the detected *IDE* transcript is representative for both isoforms, suggesting that part of the residual transcript originates from the non-targeted expressed isoform. In fact, when protein levels were assessed by western blot in *IDE*-KO, the targeted isoform (Isoform 1) was not detectable, whereas the non-targeted isoform demonstrates unchanged protein expression. In contrast, we observed complete depletion of PAM in *PAM*-KO and insulin precursor and mature insulin in *INS*-KO cells (Figure 3E). In addition, when insulin content was assessed in *INS*-KO cells using a sensitive AlphaLISA-based method, insulin was not detectable within the dynamic range of the assay (p<0.0001 vs EV) (Figure 3C). The complete absence of protein indicates that the created KO cell lines are all indeed loss-of-function (LoF) cell lines. The KO cell lines were stably cultured for more than six months without losing their KO phenotype. We can conclude that this lentiviral CRISPR pipeline in EndoC-βH1 creates functional KO cells with complete protein depletion and without any adverse effects on insulin secretion due to stable expression of Cas9 or antibiotic selection.

EndoC-βH1 KO cells represent a distinct loss-of-function model

To assess EndoC-βH1 KO cells as a LoF model, we compared KO cells to siRNA silencing strategies. siRNA mediated effects are based on mRNA degradation and unlike stable KO cells, they represent a transient LoF system[56]. To investigate the different strategies, we generated *SLC30A8*-KO and *NEUROD1*-KO cell lines and compared them with their respective siRNA models. *NEUROD1* is a key transcription factor for beta cell function and pancreas development and is implicated in both T2D-risk and monogenic diabetes[57–59]. The *NEUROD1*-KO cell line creation was successful as assessed by sgRNA integration (Figure 4A). However, it was not possible to generate a stable KO cell line as *NEUROD1*-KO cells were not able to survive. Within a few passages and after an initial reduction in NEUROD1, the protein level returned to baseline indicating that *NEUROD1*-KO cells were depleted, and only unedited cells survived and expanded (Figure 4B). Protein samples taken during the brief period of NEUROD1 reduction demonstrated an increased level of markers for endoplasmic reticulum (ER) stress and apoptosis. Unfolded protein response (UPR) signal activators, PERK and IRE1, which are activated through phosphorylation (pPERK and pIRE1) to initiate downstream signalling aimed at restoring ER homeostasis were increased 10.6- and 8.7-fold respectively (Figure 4C and D). However, a 2.2-fold upregulation of the pro-apoptotic transcription factor CHOP and a 1.6-fold increase in the activated death protease cleaved caspase 3 indicated persistent and severe ER stress resulting in apoptosis in *NEUROD1*-KO cells (Figure 4E and F). There was no difference in ER stress or apoptotic markers between WT and EV cells (not shown). To compare these severe consequences of NEUROD1 depletion to a transient model, we performed siRNA KD achieving a mean protein reduction of 93.7% (p=0.007) (Figure 5A and B). CHOP, a downstream transcription factor mediating apoptosis was significantly upregulated by 34% (p=0.016) compared to non-targeting control siRNA (Figure 5C and D). Other markers of UPR activation and apoptosis like pPERK (112.3%, p=0.422) and cleaved caspase 3 (98.0%, p=0.801) were not significantly increased (Figure 5E–H and *Extended data*). A combined readout for cell death and proliferation does not show a difference between control and *NEUROD1* silenced cells (100% vs 97.78%, p=0.894), illustrating no apoptotic effects in cells treated with si*NEUROD1* (Figure 5I). Analysis of expression levels in siRNA treated cells confirm efficient silencing of *NEUROD1* mRNA levels by 79.72% (p<0.0001) (Figure 5J). *DDIT3*, which encodes for CHOP was in contrast to its protein level, not significantly increased (113.93%, p=0.506). The active form of *XBP1*, spliced *XBP1* (*XBP1s*) is a downstream transcription factor inducing the expression of UPR target genes, which was significantly downregulated by 30.82% (p=0.023). *HSPA5*, encoding the ER chaperone BiP was also significantly decreased to 79.73% (p=0.005). There was also a trend towards *ATF6* reduction to 90.80% (p=0.143). *ATF4* and *INS* were not significantly changed (p=0.453 and p=0.596). This observed expression phenotype is consistent with an adaptive UPR response to chronic ER stress[60]. The siRNA silencing approach confirms an effect of *NEUROD1* on ER stress; however, the phenotype is more pronounced and severe in the KO model. Whereas the silencing of *NEUROD1* does not have an apoptotic effect, complete loss of NEUROD1 protein in *NEUROD1*-KO cells leads to cell death.

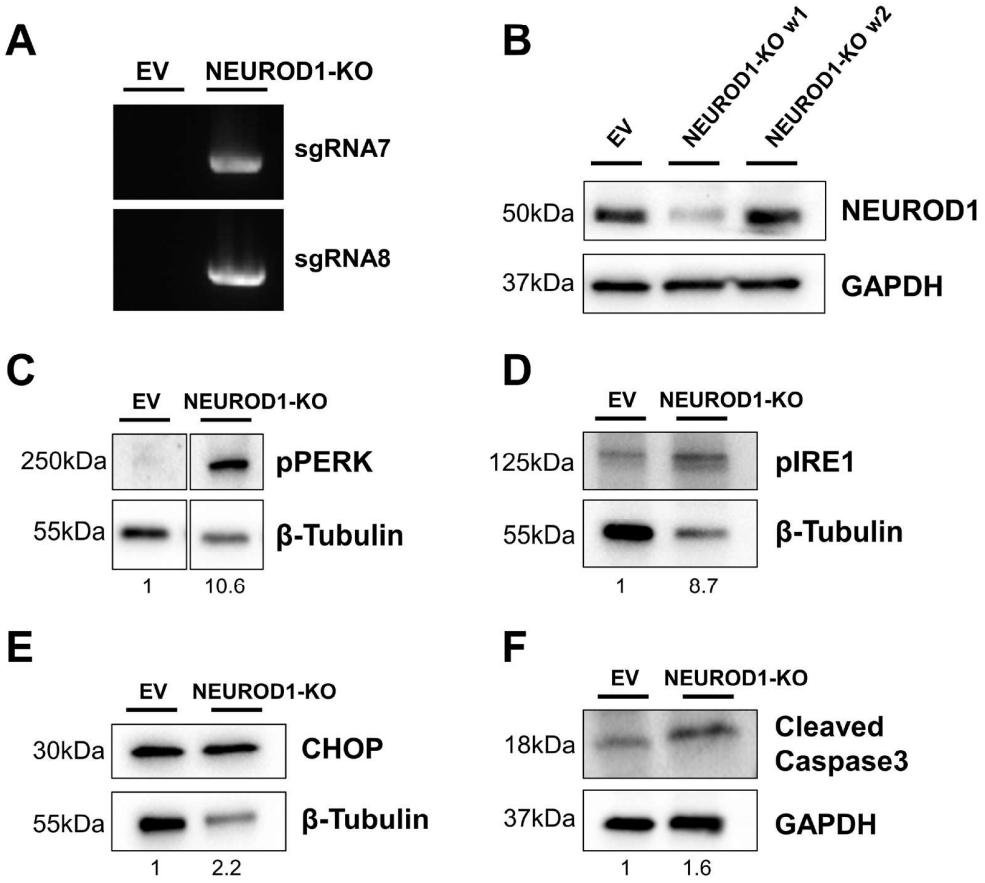


Figure 4. *NEUROD1*-KO cells demonstrate elevated ER stress and apoptosis. (**A**) PCR product for lentiCRISPRv2 and NeuroD1 sgRNA in control EV and *NEUROD1*-KO cells as described in Figure 2A. (**B**) Western Blot analysis for NEUROD1 expression in EV and *NEUROD1*-KO cells one week apart (w1 and w2). (**C–F**) Western blot analysis in EV and *NEUROD1*-KO (w1) for pPERK (**C**), pIRE1 (**D**), CHOP (**E**) and cleaved caspase 3 (**F**). GAPDH or β-Tubulin are indicated as loading controls and data is from one *NEUROD1* KO cell line. Values below each lane represent the fold change of protein levels in *NEUROD1*-KO cells compared to EV cells.

As a second comparison between the KO and siRNA model, we chose the *SLC30A8* gene, which encodes the zinc transporter ZnT8. LoF variants in *SLC30A8* have shown to be protective in T2D and siRNA KD of *SLC30A8* in EndoC-βH1 has been associated with improved glucose sensitivity and reduced expression of K_{ATP} channel subunits, amongst other effects[34]. To assess differences between the siRNA and KO model, we focused on comparing the expression of the K_{ATP} channel subunits, *KCNJ11* and *ABCC8*. siRNA silencing of *SLC30A8* reduced ZnT8 protein by 76.39% (p=0.001) and mRNA expression was decreased by 84.54% (p=0.002) (Figure 6A and C). *SLC30A8*-KO cells demonstrate complete ZnT8 protein depletion and a reduced *SLC30A8* gene expression by 57.01% (p=0.041) (Figure 6B and C). *KCNJ11* and *ABCC8* expression in silenced cells was as previously described, reduced by 28.23% (p=0.007) and 28.46% (p=0.011), respectively (Figure 6C)[34]. In *SLC30A8*-KO cells, however, *KCNJ11* and *ABCC8* mRNA levels were unchanged compared to EV control cells (104.20%, p=0.948 and 94.66%, p=0.529) (Figure 6C).

This comparison between siRNA and KO models in EndoC-βH1 demonstrate that KO LoF models can potentiate the phenotype due to complete and permanent loss of protein, as seen in *NEUROD1*-KO cells. On the other hand, the phenotype in stable cell lines can also be diminished or not detectable at all as described in *SLC30A8*-KO cells. These additional KO cell lines show the importance of using multiple approaches to study the role of genes of interest and highlight how our KO pipeline in EndoC-βH1 cells has added an extra dimension.

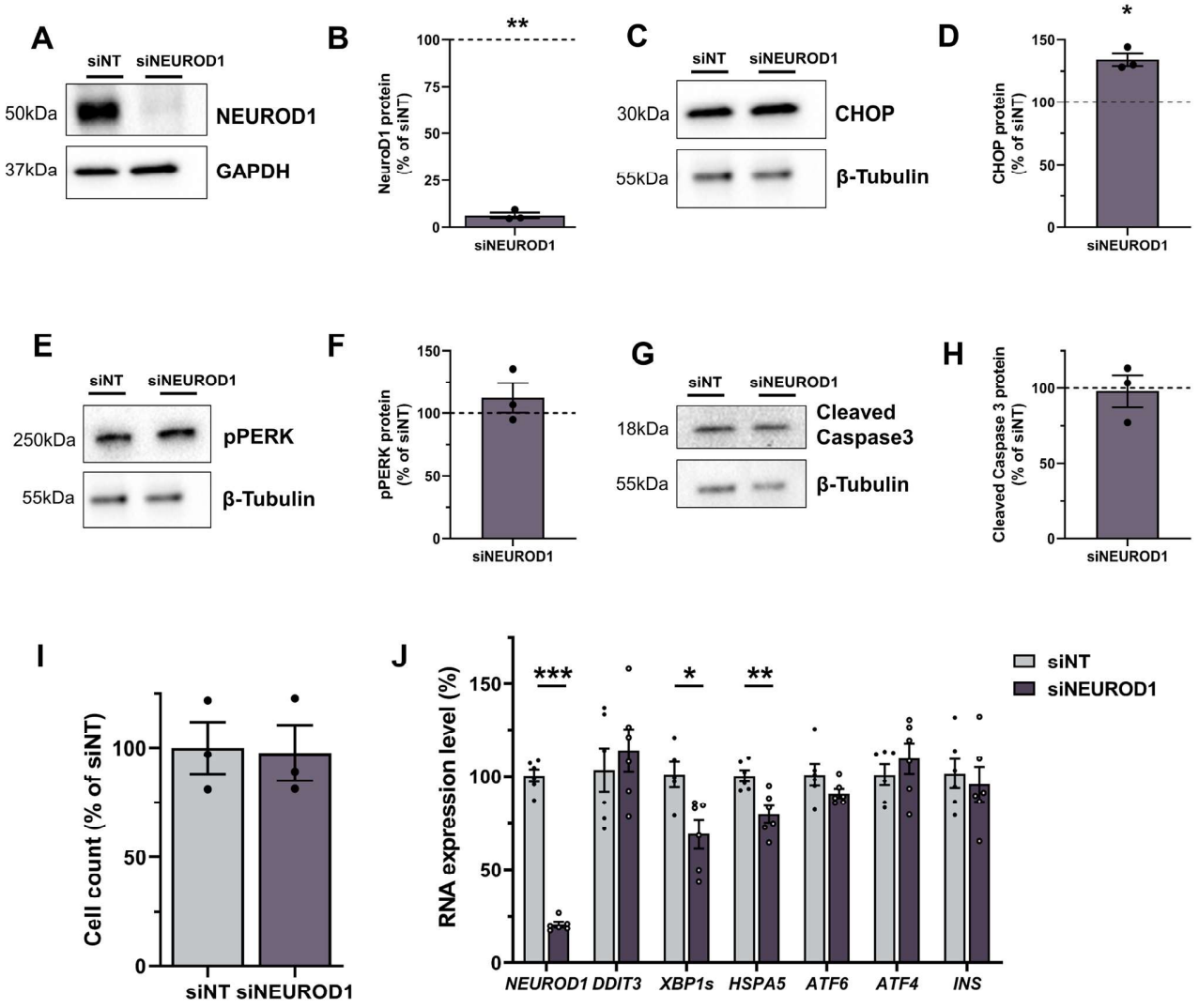


Figure 5. *NEUROD1* siRNA silencing results in increased CHOP and chronic ER stress. (**A–H**) Western Blot analysis in cells treated with siRNA targeting *NEUROD1* (si *NEUROD1*) compared to control siRNA (siNT). Western blots and quantification for NEUROD1 (**A**, **B**), CHOP (**C**, **D**), pPERK (**E**, **F**) and cleaved caspase 3 (**G**, **H**), β-Tubulin and GAPDH are displayed as loading controls. Protein values were normalised to their respective loading controls and siNT within each experiment and fold changes are displayed as percentage of siNT, which is indicated as a dotted line at 100%. (**I**) Cell count data for si *NEUROD1* cells is normalised to siNT. (**J**) Expression data analysis for genes involved in ER stress in siNT and si *NEUROD1* cells. All data are mean ± SEM from three independent experiments for western blots and cell count and six independent experiments for expression data. Fold changes were log-transformed for statistical analysis. P values * < 0.05, ** < 0.01 and *** < 0.001 using one-sample t-test for western blot data (**B**, **D**, **F**, **H**) and two-sample t-test for cell count (**I**) and expression data (**J**).

Discussion

The EndoC-βH1 cell line is an authentic human beta cell line, which is arguably the best current model to study human beta cell function. Due to its challenging growth and culture characteristics, robust protocols for CRISPR/Cas9 disease modelling to study genes implicated in human beta cell function have not been described yet in EndoC-βH1. Here, we have developed a pipeline to create stable EndoC-βH1 KO cell lines and have characterised the modifications in several proof-of-concept KO cell lines. Overall, we have successfully created five independent EndoC-βH1 KO cell lines with different gene structures, isoform expression, protein localisation and function, demonstrating that this pipeline is not restricted to certain subset of genes. The proof-of-concept cell lines,

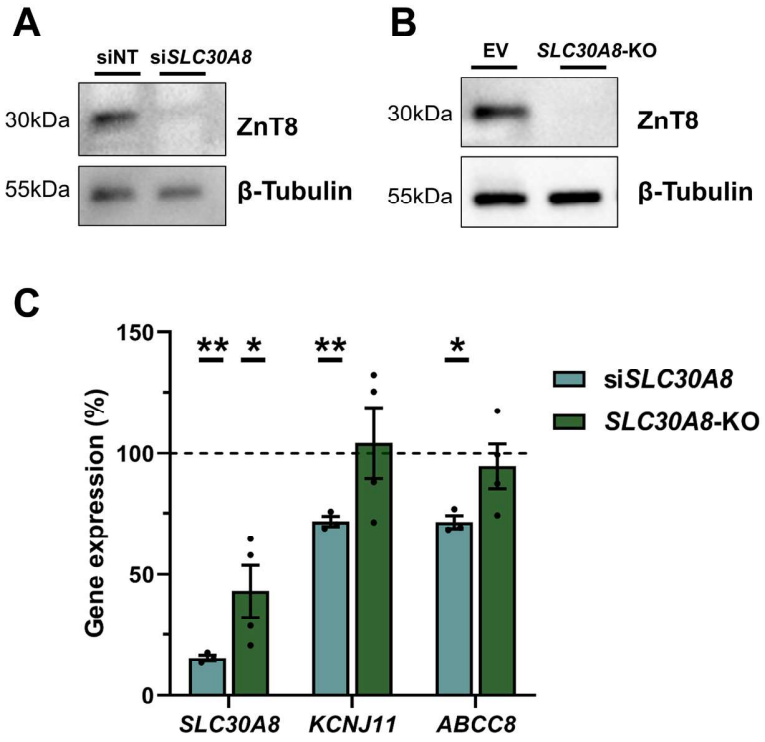


Figure 6. SLC30A8-KO cells do not replicate siRNA mediated effects on K_{ATP} channels. (**A**, **B**) Western blot analysis for Znt8 in cells treated with si *SLC30A8* compared to control siNT (**A**) and *SLC30A8*-KO cells compared to EV cells (**B**). β-Tubulin is indicated as loading control. C, Gene expression analysis of *SLC30A8, KCNJ11* and *ABCC8*. *SLC30A8*-KO cells and si *SLC30A8* were normalised within each experiment to their respective control (EV or siNT) which are displayed as a dotted line at 100%. All data are mean ± SEM from three independent experiments and fold changes were log-transformed for statistical analysis. P values * < 0.05, ** < 0.01 and *** < 0.001 using one-sample t-test comparing treated cells to their control at 100%.

PAM-KO, *IDE*-KO and *INS*-KO cell lines showed an editing efficiency greater than 87.5% for each sgRNA and complete protein depletion, indicating high KO efficiency.

Editing strategies using only one sgRNA rely on a single cleavage event followed by non-homologous end-joining-mediated introduction of a frameshift mutation leading to a PSC. However, only two-thirds of frameshift mutations introduce a PSC and low sgRNA efficiencies, alternative splicing to avoid the introduced PSC and mutations escaping NMD can reduce the KO frequency[61]. To increase the likelihood of achieving a functional KO, we implemented a strategy using two separate sgRNA[52,62]. Performing CRISPR/Cas9 editing with two sgRNA does not only increase the probability of creating a PSC at an individual sgRNA target site, it might also result in a large deletion by creating a pair of double-strand breaks (DSB) and thus rendering the resulting protein non-functional. In addition to the non-functional proteins, large deletions however might also result in the excision of intronic enhancers and non-coding regulatory elements. This might affect the regulation and expression of other genes and potentially introduce unintended phenotypes which could explain some of the differences between KD and KO phenotypes. Using several sgRNAs also increases the chance of off-targetcleavage. Sequencing of selected high chance off-targetsidesin *PAM*-KO has not demonstrated any significant off-target effects. As off-target effects are sgRNA-specific, it cannot be excluded that DSBs have occurred at other sites, for any of the other sgRNAs or large deletions or rearrangements are present which exceed the range of the performed targeted PCR and sequencing reaction[63]. Further, it might be possible that off-target activity increases during long-term culturing due to stable integration of Cas9 and the sgRNAs. We applied this dual sgRNA strategy by transducing EndoC-βH1 with individual lentivirus for each sgRNA. To obtain a more homogenous population, avoid transduction variabilities and achieve efficient dual sgRNA-based deletions, using a single expression vector containing both sgRNAs would further advance this pipeline. This can be achieved through dual-sgRNA cloning into a Cas9 containing backbone. A cost-effective and versatile protocol has recently been described and successfully been used to delete transcriptional enhancers in EndoC-βH3[64,65]. EndoC-βH3 is a drug inducible conditionally immortalized human beta cell line with similar characteristics as EndoC-βH1 albeit not glucose-responsive when left untreated and only demonstrating a stable phenotype for a limited time in culture after transgene excision[66].

A recent study describing an *HNF1A* KO cell line in EndoC-βH1 has utilised a similar approach based on lentiviral transduction with a single sgRNA in a modified lentiCRISPRv2 vector[40]. However, the study has not investigated if the transduction and selection process had any impact on the functionality of the cells and the resulting cell line does not demonstrate a complete protein depletion, as 10% of the cells still contain HNF1A protein. When only HNF1A negative sorted cells were studied, expression analysis could validate their findings from embryonic stem cell models and thus illustrates how KO models in EndoC-βH1 can provide relevant insights into beta cell function. Our KO models in comparison showed a complete protein depletion, which could be due to our dual sgRNA strategy, which results in a higher editing efficiency. Such a cell line with complete protein depletion makes it possible to study the KO consequences without a low level of background expression, which could mask some effects or increase the complexity of the pipeline by having to sort for complete KO cells.

Even though complete protein depletion is present in KO cells, phenotypic consequences can vary greatly compared to temporary protein reduction through siRNA-based approaches as demonstrated in assessing KO and KD strategies for *SLC30A8* and *NEUROD1*. *NEUROD1*-KO and *SLC30A8*-KO cells exhibit opposite directions of effect compared to siRNA KD. *NEUROD1*-KO show a potentiated impact on ER stress and apoptosis whereas effects on target genes in *SLC30A8*-KO are diminished. These differences are in line with previous studies demonstrating contrasting effects between KO and KD approaches[67–69]. The genetic compensation response, which might be masking KO mediated effects, has recently been attributed to nonsense-induced transcriptional compensation through degradation of PSC containing mRNA and should be taken into account when designing future KO sgRNA strategies[70,71]. Another consideration for comparing KO and KD experiments are potential seed-based, microRNA-like off-target effects from siRNA due to partial sequence complementarity to 3′ UTRs regions of other mRNA transcripts[72]. To increase the confidence that the observed phenotype results from silencing of the intended target gene and not due to off-target effects, it is crucial to confirm the pooled siRNA approach using multiple individual siRNAs, lentiviral delivered shRNA or rescue experiments[73,74]. In addition, it cannot be excluded that the experimental setups including distinct reagents have contributed to the observed differences between KO and KD cells for example by inducing different levels of susceptibility to cell death.

Interestingly, EndoC-βH1 *NEUROD1*-KO cells were not viable, which is consistent with a detected increase of ER stress and apoptosis markers. In line with this, pancreases from *NeuroD1* null mice have 14-fold more apoptotic cells[75]. Mice with conditional *NeuroD1*-KO in insulin expressing cells on the other hand do not demonstrate increased apoptosis as measured by activated caspase 3[58]. Overexpression of *NeuroD1* in rodent beta cell lines prevents ethanol induced expression of *Ddit3* (CHOP), reduces apoptosis and highlights *Ddit3* as a downstream target of NeuroD1[76]. This is in accordance with the results of both, our *NEUROD1*-KO and siRNA model which show an increase in CHOP protein, confirming the observed relation between *NEUROD1* and *DDIT3* in a human beta cell model and without extrinsic stress stimuli. Whereas temporary silencing of NEUROD1 induces a phenotype similar to chronic ER stress but without any effects on cell viability, *NEUROD1*-KO cells are not able to compensate for the permanent and complete loss of NEUROD1 protein and demonstrate elevated ER stress and apoptosis. However, further studies are needed to investigate the potential regulatory role of *NEUROD1* in ER stress and apoptosis, and its implications for diabetes.

This genome editing pipeline in EndoC-βH1 is an efficient strategy to robustly create KO cell lines in a human beta cell line. The generated KO cell lines could be used to study the function of genes in human beta cells, investigate their role in diabetes pathology and as a protein free cellular system to overexpress and study genetic variants implicated in disease. As every other LoF model, observed phenotypes might be specific to this strategy and should be validated with complementary approaches such as transient siRNA transfection. The successful generation of these KO cell lines demonstrate the feasibility of CRISPR/Cas9 genome editing in EndoC-βH1 and open up further possibilities for CRISPR/Cas9 based strategies such as CRISPR interference (CRISPRi), CRISPR activation (CRISPRa), genome-wide CRISPR screening, epigenome and base editing.

Data availability

Underlying data

European Nucleotide Archive: A CRISPR/Cas9 genome editing pipeline in the EndoC-βH1 cell line to study genes implicated in beta cell function. Accession number PRJEB34547; http://identifiers.org/ena.embl:PRJEB34547.

Open Science Framework: A CRISPR/Cas9 genome editing pipeline in the EndoC-βH1 cell line to study genes implicated in beta cell function. https://doi.org/10.17605/OSF.IO/2KYAN[77].

This project contains the following underlying data:

- Figure 1: 1F Puromycin kill curve (Raw data in an Excel file)
- Figure 2:
  - 2A IDE sgRNA, 2A INS sgRNA_1D IDE KO, 2A INS sgRNA_1D INS KO, 2A PAM sgRNA, 2B Cas9, 2B Cas9_INSKO, 2B Tubulin, 2B Tubulin_INSKO and 2D PAM KO (Uncropped images in Image Lab and TIFF files)
  - 2E_MED15 EV (Raw sequencing file)
  - 2E_MED15 PAMKO (Raw sequencing file)
  - 2F_GPM6B EV (Raw sequencing file)
  - 2F_GPM6B PAMKO (Raw sequencing file)
  - Figure 2C (Raw sequencing files)

- Figure 3:
  - 3A Secretion data, 3B Fold change, 3C Insulin content and 3D mRNA expression (Raw data in Excel files)
  - 3E IDE, 3E INS, 3E PAM, 3E Tubulin_IDE, 3E Tubulin_INS and (Uncropped images in an Image Lab and TIFF file)
- Figure 4 (All uncropped images in Image Lab and TIFF files)
- Figure 5:
  - 5A GAPDH, 5A NEUROD1, 5C CHOP, 5CE TUBULIN, 5E pERK, 5G Cleaved Caspase3 and 5G TUBULIN (Uncropped images in Image Lab and TIFF files)
  - 5BDFH Quantification (Raw data in an Excel file)
  - 5I Cell Count (Raw data in an Excel file)
  - 5J Gene expression (Raw data in an Excel file)
- Figure 6:
  - 6A SLC30A8, 6A TUBULIN, 6B SLC30A8 and 6B TUBULIN (Uncropped images in Image Lab and TIFF files)
  - 6C Gene expression (Raw data in an Excel file)
- Extended data:
  - A GAPDH_Traf2, A Traf2, C pIre1, E Ki67 and CE Tubulin_pIre1 and Ki67 (Uncropped images in Image Lab and TIFF files)
  - BDF Quantification (Raw data in an Excel file)

Extended data

Open Science Framework: A CRISPR/Cas9 genome editing pipeline in the EndoC-βH1 cell line to study genes implicated in beta cell function. https://doi.org/10.17605/OSF.IO/2KYAN[77].

This project contains the following extended data:

- Extended data.tif: (A–F), Western Blot analysis in cells treated with siRNA targeting *NEUROD1* (si*NEUROD1*) compared to control siRNA (siNT). Western blots and quantification for TRAF2 (A, B), pIRE1 (C, D) and Ki67 (E, F), β-Tubulin and GAPDH are displayed as loading controls. Protein values were normalised to their respective loading controls and siNT within each experiment and fold changes are displayed as percentage of siNT, which is indicated as a dotted line at 100%. Fold changes were log-transformed for statistical analysis. P values * < 0.05, ** < 0.01 and *** < 0.001 using one-sample t-test.

Acknowledgments

We would like to thank members of the Gloyn lab for their support and advice.

References

1. International Diabetes Federation (IDF): **IDF Diabetes Atlas.** IDF Diabetes Atlas, 8th edition. 2017.
Reference Source
2. Morris AP, Voight BF, Teslovich TM, *et al.*: **Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes.** *Nat Genet.* 2012; **44**(9): 981–90.
PubMed Abstract | Publisher Full Text | Free Full Text
3. Fuchsberger C, Flannick J, Teslovich TM, *et al.*: **The genetic architecture of type 2 diabetes.** *Nature.* 2016; **536**(7614): 41–7.
PubMed Abstract | Publisher Full Text | Free Full Text
4. Mahajan A, Wessel J, Willems SM, *et al.*: **Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes.** *Nat Genet.* 2018; **50**(4): 559–71.
PubMed Abstract | Publisher Full Text | Free Full Text
5. Mahajan A, Taliun D, Thurner M, *et al.*: **Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps.** *Nat Genet.* 2018; **50**(11): 1505–13.
PubMed Abstract | Publisher Full Text | Free Full Text
6. Dimas AS, Lagou V, Barker A, *et al.*: **Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity.** *Diabetes.* 2014; **63**(6): 2158–71.
PubMed Abstract | Publisher Full Text | Free Full Text
7. Asfari M, Janjic D, Meda P, *et al.*: **Establishment of 2-mercaptoethanol-dependent differentiated insulin-secreting cell lines.** *Endocrinology.* 1992; **130**(1): 167–78.
PubMed Abstract | Publisher Full Text
8. Miyazaki JI, Araki K, Yamato E, *et al.*: **Establishment of a pancreatic beta cell line that retains glucose-inducible insulin secretion: special reference to expression of glucose transporter isoforms.** *Endocrinology.* 1990; **127**(1): 126–32.
PubMed Abstract | Publisher Full Text
9. Efrat S, Linde S, Kofod H, *et al.*: **Beta-cell lines derived from transgenic mice expressing a hybrid insulin gene-oncogene.** *Proc Natl Acad Sci U S A.* 1988; **85**(23): 9037–41.
PubMed Abstract | Publisher Full Text | Free Full Text
10. Cabrera O, Berman DM, Kenyon NS, *et al.*: **The unique cytoarchitecture of human pancreatic islets has implications for islet cell function.** *Proc Natl Acad Sci U S A.* 2006; **103**(7): 2334–9.
PubMed Abstract | Publisher Full Text | Free Full Text
11. Cohrs CM, Chen C, Jahn SR, *et al.*: **Vessel Network Architecture of Adult Human Islets Promotes Distinct Cell-Cell Interactions *In Situ* and Is Altered After Transplantation.** *Endocrinology.* 2017; **158**(5): 1373–85.
PubMed Abstract | Publisher Full Text
12. Brissova M, Shostak A, Fligner CL, *et al.*: **Human Islets Have Fewer Blood Vessels than Mouse Islets and the Density of Islet Vascular Structures Is Increased in Type 2 Diabetes.** *J Histochem Cytochem.* 2015; **63**(8): 637–45.
PubMed Abstract | Publisher Full Text | Free Full Text
13. Benner C, van der Meulen T, Cacéres E, *et al.*: **The transcriptional landscape of mouse beta cells compared to human beta cells reveals notable species differences in long non-coding RNA and protein-coding gene expression.** *BMC Genomics.* 2014; **15**(1): 620.
PubMed Abstract | Publisher Full Text | Free Full Text
14. Baron M, Veres A, Wolock SL, *et al.*: **A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure.** *Cell Syst.* 2016; **3**(4): 346–360.e4.
PubMed Abstract | Publisher Full Text | Free Full Text

15. Cyphert HA, Walker EM, Hang Y, *et al.*: **Examining How the MAFB Transcription Factor Affects Islet β-Cell Function Postnatally.** *Diabetes.* 2019; **68**(2): 337–48.
PubMed Abstract | Publisher Full Text | Free Full Text

16. Henquin JC, Dufrane D, Kerr-Conte J, *et al.*: **Dynamics of glucose-induced insulin secretion in normal human islets.** *Am J Physiol Endocrinol Metab.* 2015; **309**(7): E640–50.
PubMed Abstract | Publisher Full Text

17. Ferrer J, Benito C, Gomis R: **Pancreatic islet GLUT2 glucose transporter mRNA and protein expression in humans with and without NIDDM.** *Diabetes.* 1995; **44**(12): 1369–74.
PubMed Abstract | Publisher Full Text

18. McCulloch LJ, van de Bunt M, Braun M, *et al.*: **GLUT2 (*SLC2A2*) is not the principal glucose transporter in human pancreatic beta cells: implications for understanding genetic association signals at this locus.** *Mol Genet Metab.* 2011; **104**(4): 648–53.
PubMed Abstract | Publisher Full Text

19. De Vos A, Heimberg H, Quartier E, *et al.*: **Human and rat beta cells differ in glucose transporter but not in glucokinase gene expression.** *J Clin Invest.* 1995; **96**(5): 2489–95.
PubMed Abstract | Publisher Full Text | Free Full Text

20. Shiao MS, Liao BY, Long M, *et al.*: **Adaptive evolution of the insulin two-gene system in mouse.** *Genetics.* 2008; **178**(3): 1683–91.
PubMed Abstract | Publisher Full Text | Free Full Text

21. Fiaschi-Taesch NM, Kleinberger JW, Salim FG, *et al.*: **Human pancreatic β-cell G1/S molecule cell cycle atlas.** *Diabetes.* 2013; **62**(7): 2450–9.
PubMed Abstract | Publisher Full Text | Free Full Text

22. Butler AE, Cao-Minh L, Galasso R, *et al.*: **Adaptive changes in pancreatic beta cell fractional area and beta cell turnover in human pregnancy.** *Diabetologia.* 2010; **53**(10): 2167–76.
PubMed Abstract | Publisher Full Text | Free Full Text

23. Parnaud G, Bosco D, Berney T, *et al.*: **Proliferation of sorted human and rat beta cells.** *Diabetologia.* 2008; **51**(1): 91–100.
PubMed Abstract | Publisher Full Text

24. Braun M, Ramracheya R, Bengtsson M, *et al.*: **Voltage-gated ion channels in human pancreatic beta-cells: electrophysiological characterization and role in insulin secretion.** *Diabetes.* 2008; **57**(6): 1618–28.
PubMed Abstract | Publisher Full Text

25. Rorsman P, Ashcroft FM: **Pancreatic β-Cell Electrical Activity and Insulin Secretion: Of Mice and Men.** *Physiol Rev.* 2018; **98**(1): 117–214.
PubMed Abstract | Publisher Full Text | Free Full Text

26. Ravassard P, Hazhouz Y, Pechberty S, *et al.*: **A genetically engineered human pancreatic β cell line exhibiting glucose-inducible insulin secretion.** *J Clin Invest.* 2011; **121**(9): 3589–97.
PubMed Abstract | Publisher Full Text | Free Full Text

27. Gurgul-Convey E, Kaminski MT, Lenzen S: **Physiological characterization of the human EndoC-βH1 β-cell line.** *Biochem Biophys Res Commun.* 2015; **464**(1): 13–9.
PubMed Abstract | Publisher Full Text

28. Hastoy B, Godazgar M, Clark A, *et al.*: **Electrophysiological properties of human beta-cell lines EndoC-βH1 and -βH2 conform with human beta-cells.** *Sci Rep.* 2018; **8**(1): 16994.
PubMed Abstract | Publisher Full Text | Free Full Text

29. Andersson LE, Valtat B, Bagge A, *et al.*: **Characterization of stimulus-secretion coupling in the human pancreatic EndoC-βH1 beta cell line.** *PLoS One.* 2015; **10**(3): e0120879.
PubMed Abstract | Publisher Full Text | Free Full Text

30. Lawlor N, Márquez EJ, Orchard P, *et al.*: **Multiomic Profiling Identifies cis-Regulatory Networks Underlying Human Pancreatic β Cell Identity and Function.** *Cell Rep.* 2019; **26**(3): 788–801.e6.
PubMed Abstract | Publisher Full Text | Free Full Text

31. Thomsen SK, Ceroni A, van de Bunt M, *et al.*: **Systematic Functional Characterization of Candidate Causal Genes for Type 2 Diabetes Risk Variants.** *Diabetes.* 2016; **65**(12): 3805–11.
PubMed Abstract | Publisher Full Text | Free Full Text

32. Tsonkova VG, Sand FW, Wolf XA, *et al.*: **The EndoC-βH1 cell line is a valid model of human beta cells and applicable for screenings to identify novel drug target candidates.** *Mol Metab.* 2018; **8**: 144–57.
PubMed Abstract | Publisher Full Text | Free Full Text

33. Thomsen SK, Raimondo A, Hastoy B, *et al.*: **Type 2 diabetes risk alleles in *PAM* impact insulin release from human pancreatic β-cells.** *Nat Genet.* 2018; **50**(8): 1122–31.
PubMed Abstract | Publisher Full Text | Free Full Text

34. Dwivedi OP, Lehtovirta M, Hastoy B, *et al.*: **Loss of ZnT8 function protects against diabetes by enhanced insulin secretion.** *bioRxiv.* 2018; 436030.
Publisher Full Text

35. Carrat GR, Hu M, Nguyen-Tu MS, *et al.*: **Decreased *STARD10* Expression Is Associated with Defective Insulin Secretion in Humans and Mice.** *Am J Hum Genet.* 2017; **100**(2): 238–56.
PubMed Abstract | Publisher Full Text | Free Full Text

36. Mali P, Yang L, Esvelt KM, *et al.*: **RNA-guided human genome engineering via Cas9.** *Science.* 2013; **339**(6121): 823–6.
PubMed Abstract | Publisher Full Text | Free Full Text

37. Cong L, Ran FA, Cox D, *et al.*: **Multiplex genome engineering using CRISPR/Cas systems.** *Science.* 2013; **339**(6121): 819–24.
PubMed Abstract | Publisher Full Text | Free Full Text

38. Jinek M, Chylinski K, Fonfara I, *et al.*: **A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.** *Science.* 2012; **337**(6096): 816–21.
PubMed Abstract | Publisher Full Text | Free Full Text

39. Gasiunas G, Barrangou R, Horvath P, *et al.*: **Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria.** *Proc Natl Acad Sci U S A.* 2012; **109**(39): E2579–86.
PubMed Abstract | Publisher Full Text | Free Full Text

40. Cardenas-Diaz FL, Osorio-Quintero C, Diaz-Miranda MA, *et al.*: **Modeling Monogenic Diabetes using Human ESCs Reveals Developmental and Metabolic Deficiencies Caused by Mutations in HNF1A.** *Cell Stem Cell.* 2019; **25**(2): 273–289.e5.
PubMed Abstract | Publisher Full Text

41. Sanjana NE, Shalem O, Zhang F: **Improved vectors and genome-wide libraries for CRISPR screening.** *Nat Methods.* 2014; **11**(8): 783–4.
PubMed Abstract | Publisher Full Text | Free Full Text

42. Hart T, Tong AHY, Chan K, *et al.*: **Evaluation and Design of Genome-Wide CRISPR/SpCas9 Knockout Screens.** *G3 (Bethesda).* 2017; **7**(8): 2719–27.
PubMed Abstract | Publisher Full Text | Free Full Text

43. Haeussler M, Schönig K, Eckert H, *et al.*: **Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR.** *Genome Biol.* 2016; **17**(1): 148.
PubMed Abstract | Publisher Full Text | Free Full Text

44. Figliozzi RW, Chen F, Chi A, *et al.*: **Using the inverse Poisson distribution to calculate multiplicity of infection and viral replication by a high-throughput fluorescent imaging system.** *Virol Sin.* 2016; **31**(2): 180–3.
PubMed Abstract | Publisher Full Text | Free Full Text

45. Merriman C, Li H, Li H, *et al.*: **Highly specific monoclonal antibodies for allosteric inhibition and immunodetection of the human pancreatic zinc transporter ZnT8.** *J Biol Chem.* 2018; **293**(42): 16206–16.
PubMed Abstract | Publisher Full Text | Free Full Text

46. Brinkman EK, Chen T, Amendola M, *et al.*: **Easy quantitative assessment of genome editing by sequence trace decomposition.** *Nucleic Acids Res.* 2014; **42**(22): e168.
PubMed Abstract | Publisher Full Text | Free Full Text

47. Hsiau T, Conant D, Rossi N, *et al.*: **Inference of CRISPR Edits from Sanger Trace Data.** *bioRxiv.* 2019; 251082.
Publisher Full Text

48. Steneberg P, Bernardo L, Edfalk S, *et al.*: **The type 2 diabetes-associated gene *ide* is required for insulin secretion and suppression of α-synuclein levels in β-cells.** *Diabetes.* 2013; **62**(6): 2004–14.
PubMed Abstract | Publisher Full Text | Free Full Text

49. Sladek R, Rocheleau G, Rung J, *et al.*: **A genome-wide association study identifies novel risk loci for type 2 diabetes.** *Nature.* 2007; **445**(7130): 881–5.
PubMed Abstract | Publisher Full Text

50. Tager H, Given B, Baldwin D, *et al.*: **A structurally abnormal insulin causing human diabetes.** *Nature.* 1979; **281**(5727): 122–5.
PubMed Abstract | Publisher Full Text

51. Stoy J, Edghill EL, Flanagan SE, *et al.*: **Insulin gene mutations as a cause of permanent neonatal diabetes.** *Proc Natl Acad Sci U S A.* 2007; **104**(38): 15040–4.
PubMed Abstract | Publisher Full Text | Free Full Text

52. Zhou J, Wang J, Shen B, *et al.*: **Dual sgRNAs facilitate CRISPR/Cas9-mediated mouse genome targeting.** *FEBS J.* 2014; **281**(7): 1717–25.
PubMed Abstract | Publisher Full Text

53. Moffat lab: **TKOv3 sgRNA viewer [Internet].** 2017; [cited 2017 Jun 12].
Reference Source

54. Doench JG, Fusi N, Sullender M, *et al.*: **Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9.** *Nat Biotechnol.* 2016; **34**(2): 184–91.
PubMed Abstract | Publisher Full Text | Free Full Text

55. Reber S, Mechtersheimer J, Nasif S, *et al.*: **CRISPR-Trap: a clean approach for the generation of gene knockouts and gene replacements in human cells.** *Mol Biol Cell.* 2018; **29**(2): 75–83.
PubMed Abstract | Publisher Full Text | Free Full Text

56. Elbashir SM, Harborth J, Lendeckel W, *et al.*: **Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells.** *Nature.* 2001; **411**(6836): 494–8.
PubMed Abstract | Publisher Full Text

57. Malecki MT, Jhala US, Antonellis A, *et al.*: **Mutations in *NEUROD1* are associated with the development of type 2 diabetes mellitus.** *Nat Genet.* 1999; **23**(3): 323–8.
PubMed Abstract | Publisher Full Text

58. Gu C, Stein GH, Pan N, *et al.*: **Pancreatic beta cells require NeuroD to achieve and maintain functional maturity.** *Cell Metab.* 2010; **11**(4): 298–310.
PubMed Abstract | Publisher Full Text | Free Full Text

59. Fajans S, Bell GI, Polonsky KS: **Molecular mechanisms and clinical pathophysiology of maturity-onset diabetes of the young.** *N Engl J Med.* 2001; **345**(13): 971–80.
PubMed Abstract | Publisher Full Text

60. Gomez JA, Rutkowski DT: **Experimental reconstitution of chronic ER stress in the liver reveals feedback suppression of BiP mRNA expression.** *eLife.* 2016; **5**:

pii: e20390.
PubMed Abstract | Publisher Full Text | Free Full Text

61. Lappalainen T, Sammeth M, Friedländer MR, *et al.*: **Transcriptome and genome sequencing uncovers functional variation in humans.** *Nature.* 2013; **501**(7468): 506–11.
PubMed Abstract | Publisher Full Text | Free Full Text

62. Canver MC, Bauer DE, Dass A, *et al.*: **Characterization of genomic deletion efficiency mediated by clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells.** *J Biol Chem.* 2014; **289**(31): 21312–24.
PubMed Abstract | Publisher Full Text | Free Full Text

63. Kosicki M, Tomberg K, Bradley A: **Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements.** *Nat Biotechnol.* 2018; **36**(8): 765–71.
PubMed Abstract | Publisher Full Text | Free Full Text

64. Beucher A, Cebola I: **One-step dual CRISPR/Cas9 guide RNA cloning protocol [Internet].** Protocol Exchange. 2019; [cited 2019 Jun 26].
Publisher Full Text

65. Miguel-Escalada I, Bonàs-Guarch S, Cebola I, *et al.*: **Human pancreatic islet three-dimensional chromatin architecture provides insights into the genetics of type 2 diabetes.** *Nat Genet.* 2019; **51**(7): 1137–48.
PubMed Abstract | Publisher Full Text | Free Full Text

66. Benazra M, Lecomte MJ, Colace C, *et al.*: **A human beta cell line with drug inducible excision of immortalizing transgenes.** *Mol Metab.* 2015; **4**(12): 916–25.
PubMed Abstract | Publisher Full Text | Free Full Text

67. Karakas B, Weeraratna AT, Abukhdeir AM, *et al.*: **P21 gene knock down does not identify genetic effectors seen with gene knock out.** *Cancer Biol Ther.* 2007; **6**(7): 1025–30.
PubMed Abstract | Publisher Full Text | Free Full Text

68. Morgens DW, Deans RM, Li A, *et al.*: **Systematic comparison of CRISPR/Cas9 and RNAi screens for essential genes.** *Nat Biotechnol.* 2016; **34**(6): 634–6.
PubMed Abstract | Publisher Full Text | Free Full Text

69. Rossi A, Kontarakis Z, Gerri C, *et al.*: **Genetic compensation induced by deleterious mutations but not gene knockdowns.** *Nature.* 2015; **524**(7564): 230–3.
PubMed Abstract | Publisher Full Text

70. Ma Z, Zhu P, Shi H, *et al.*: **PTC-bearing mRNA elicits a genetic compensation response via Upf3a and COMPASS components.** *Nature.* 2019; **568**(7751): 259–63.
PubMed Abstract | Publisher Full Text

71. El-Brolosy MA, Kontarakis Z, Rossi A, *et al.*: **Genetic compensation triggered by mutant mRNA degradation.** *Nature.* 2019; **568**(7751): 193–7.
PubMed Abstract | Publisher Full Text | Free Full Text

72. Doench JG, Petersen CP, Sharp PA: **siRNAs can function as miRNAs.** *Genes Dev.* 2003; **17**(4): 438–442.
PubMed Abstract | Publisher Full Text | Free Full Text

73. Jackson AL, Linsley PS: **Recognizing and avoiding siRNA off-target effects for target identification and therapeutic application.** *Nat Rev Drug Discov.* 2010; **9**(1): 57–67.
PubMed Abstract | Publisher Full Text

74. Klinghoffer RA, Magnus J, Schelter J, *et al.*: **Reduced seed region-based off-target activity with lentivirus-mediated RNAi.** *RNA.* 2010**16**(5): 879–884.
PubMed Abstract | Publisher Full Text | Free Full Text

75. Naya FJ, Huang HP, Qiu Y, *et al.*: **Diabetes, defective pancreatic morphogenesis, and abnormal enteroendocrine differentiation in BETA2/neuroD-deficient mice.** *Genes Dev.* 1997; **11**(18): 2323–34.
PubMed Abstract | Publisher Full Text | Free Full Text

76. Wu T, Xiang J, Shan W, *et al.*: **Epigallocatechin-3-Gallate Inhibits Ethanol-Induced Apoptosis Through Neurod1 Regulating CHOP Expression in Pancreatic β-Cells.** *Anat Rec (Hoboken).* 2016; **299**(5): 573–82.
PubMed Abstract | Publisher Full Text

77. Grotz A: **A CRISPR/Cas9 genome editing pipeline in the EndoC-βH1 cell line to study genes implicated in beta cell function**. 2019.
http://www.doi.org/10.17605/OSF.IO/2KYAN

ARTICLES
https://doi.org/10.1038/s41588-019-0513-9
nature genetics

Loss of ZnT8 function protects against diabetes by enhanced insulin secretion

Om Prakash Dwivedi[1,17], Mikko Lehtovirta[1,17], Benoit Hastoy[2,17], Vikash Chandra[3], Nicole A. J. Krentz[4], Sandra Kleiner[5], Deepak Jain[6], Ann-Marie Richard[7], Fernando Abaitua[4], Nicola L. Beer[2], Antje Grotz[2], Rashmi B. Prasad[8], Ola Hansson[1,8], Emma Ahlqvist[8], Ulrika Krus[8], Isabella Artner[8], Anu Suoranta[1], Daniel Gomez[5], Aris Baras[5], Benoite Champon[4], Anthony J. Payne[4], Daniela Moralli[4], Soren K. Thomsen[2], Philipp Kramer[4], Ioannis Spiliotis[2], Reshma Ramracheya[2], Pauline Chabosseau[9], Andria Theodoulou[9], Rebecca Cheung[9], Martijn van de Bunt[2,4], Jason Flannick[10,11], Maddalena Trombetta[12], Enzo Bonora[12], Claes B. Wolheim[8], Leena Sarelin[13], Riccardo C. Bonadonna[14], Patrik Rorsman[2], Benjamin Davies[4], Julia Brosnan[7], Mark I. McCarthy[2,4,15], Timo Otonkoski[3], Jens O. Lagerstedt[6], Guy A. Rutter[9], Jesper Gromada[5], Anna L. Gloyn[2,4,15,18], Tiinamaija Tuomi[1,13,16,18] and Leif Groop[1,8,18*]

A rare loss-of-function allele p.Arg138* in *SLC30A8* encoding the zinc transporter 8 (ZnT8), which is enriched in Western Finland, protects against type 2 diabetes (T2D). We recruited relatives of the identified carriers and showed that protection was associated with better insulin secretion due to enhanced glucose responsiveness and proinsulin conversion, particularly when compared with individuals matched for the genotype of a common T2D-risk allele in *SLC30A8*, p.Arg325. In genome-edited human induced pluripotent stem cell (iPSC)-derived β-like cells, we establish that the p.Arg138* allele results in reduced *SLC30A8* expression due to haploinsufficiency. In human β cells, loss of *SLC30A8* leads to increased glucose responsiveness and reduced K_{ATP} channel function similar to isolated islets from carriers of the T2D-protective allele p.Trp325. These data position ZnT8 as an appealing target for treatment aimed at maintaining insulin secretion capacity in T2D.

Zinc transporters (ZnT) regulate the passage of zinc across biological membranes out of the cytosol, while Zrt/Irt-like proteins transport zinc into the cytosol[1]. ZnT8, encoded by *SLC30A8*, is highly expressed in the membrane of insulin granules within pancreatic β cells, where it transports zinc ions for crystallization and storage of insulin[2]. We previously described a loss-of-function (LoF) allele p.Arg138* (rs200185429, c.412C>T) in the *SLC30A8* gene that conferred 53% protection against T2D (ref. [3]). This allele was extremely rare (0.02%) in most European countries, but more common (>0.2%) in Western Finland[3]. We also reported a protective frameshift allele, p.Lys34Serfs*50, conferring 83% protection against T2D in Iceland[3]. Further, the *SLC30A8* gene harbors a common variant (rs13266634, c.973C>T) p.Trp325Arg in the carboxy (C)-terminal domain[4]. Whereas the major p.Arg325 allele (>70% of the population) confers increased risk for T2D, the minor p.Trp325 allele is protective[5].

The mechanisms by which modulation of ZnT8 activity protects against T2D are largely unknown. Several attempts have been made to study loss of *Slc30a8* function in rodent models, but the results have been inconclusive; global knockout of *Slc30a8* led to either glucose intolerance or had no effect in mice[6–8], whereas overexpression improved glucose tolerance without any effect on insulin secretion[9]. A mouse model harboring the equivalent of the human p.Arg138* allele lacked any detectable ZnT8 protein but showed no effect on glucose tolerance[10]. These rodent studies present a complex picture that may not recapitulate the T2D-protective effects of *SLC30A8* LoF alleles in humans. We therefore performed detailed metabolic studies in human carriers heterozygous for the LoF allele (p.Arg138*) recruited on the basis of their genotype, performed comprehensive functional studies in human β cell models and compared these with the mouse model carrying the human p.Arg138*-*Slc30a8* allele.

[1]Institute for Molecular Medicine Finland, Helsinki University, Helsinki, Finland. [2]Oxford Centre for Diabetes Endocrinology and Metabolism, University of Oxford, Oxford, UK. [3]Stem Cells and Metabolism Research Program and Biomedicum Stem Cell Centre, Faculty of Medicine, University of Helsinki, Helsinki, Finland. [4]Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK. [5]Regeneron Pharmaceuticals, Tarrytown, NY, USA. [6]Department of Experimental Medical Science, Lund University, Lund, Sweden. [7]Pfizer Inc., Cambridge, MA, USA. [8]Lund University Diabetes Centre, Department of Clinical Sciences, Lund University, Skåne University Hospital, Malmö, Sweden. [9]Section of Cell Biology, Department of Medicine, Imperial College London, Imperial Centre for Translational and Experimental Medicine, Hammersmith, Hospital, London, UK. [10]Department of Molecular Biology, Massachusetts General Hospital, Boston, MA, USA. [11]Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA. [12]Department of Medicine, University of Verona and Azienda Ospedaliera Universitaria Integrata of Verona, Verona, Italy. [13]Folkhälsan Research Center, Helsinki, Finland. [14]Department of Medicine and Surgery, University of Parma School of Medicine and Azienda Ospedaliera Universitaria of Parma, Parma, Italy. [15]Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, UK. [16]Abdominal Center, Endocrinology, Helsinki University Central Hospital, Research Program for Clinical and Molecular Metabolism, University of Helsinki, Helsinki, Finland. [17]These authors contributed equally: Om Prakash Dwivedi, Mikko Lehtovirta, Benoit Hastoy. [18]These authors jointly supervised this work: Anna L. Gloyn, Tiinamaija Tuomi, Leif Groop. *e-mail: Leif.Groop@helsinki.fi

Results

Recruitment by genotype. Given the enrichment of the p.Arg138* *SLC30A8* allele in Western Finland, we genotyped >14,000 individuals from the Botnia Study[11] for the *SLC30A8* p.Arg138* and the common p.Trp325Arg variants, and identified 71 p.Arg138* *SLC30A8* carriers (all heterozygotes; 55 nondiabetic individuals, Fig. 1). We then recruited family members of known p.Arg138* carriers to identify additional p.Arg138* carriers and to perform a detailed metabolic study (190 min test meal) in carriers and noncarrier relatives. Of the 79 p.Arg138* carriers (65 new, 14 previously identified) and 103 noncarrier relatives from >21 families (Extended Data Fig. 1), 54 and 47, respectively, participated in a test meal, and 31 and 13 participated in an oral glucose tolerance test (OGTT) during a separate second visit (Fig. 1 and Supplementary Tables 1 and 2). We also had data from previously performed OGTTs within the Botnia Study for 8,436 nondiabetic individuals (55 p.Arg138* carriers; Fig. 1 and Supplementary Tables 2 and 3). Of the 136 p.Arg138* allele carriers, none were homozygous for the protective common variant, p.Trp325, and p.Arg138* segregated with p.Arg325 in all families (Extended Data Fig. 1). Thus, we present the data in three different ways: (1) p.Arg138* versus all p.Arg138Arg, (2) p.Arg138* versus p.Arg138Arg having at least one p.Arg325 allele (p.Trp325Arg or p.Arg325Arg) and (3) p.Arg325 (p.Trp325Arg or p.Arg325Arg) versus p.Trp325Trp on a background of p.Arg138Arg.

Replicating our previous findings[3], carriers of p.Arg138* had a reduced risk of T2D (odds ratio (OR) = 0.40, $P = 0.003$) when analyzing 4,564 T2D (13 p.Arg138* carriers) and 8,183 nondiabetic (55 p.Arg138* carriers) individuals. Additionally, nondiabetic p.Arg138* carriers had lower fasting glucose concentrations than p.Arg138Arg individuals (Supplementary Tables 4 and 5). There were no significant differences in plasma zinc concentrations measured during test meal or OGTT between the groups (data not shown).

Comparison of p.Arg138 versus p.Arg138Arg.* The p.Arg138* carriers had lower blood glucose levels during the test meal, especially during the first 40 min (area under curve, $P = 0.02$), and showed a higher corrected insulin response (CIR) (at 20 min, $P = 0.046$) than noncarriers (Supplementary Table 6 and Fig. 2a–c). Similarly, in the larger population-based OGTT cohort, carriers had higher insulin response during OGTT (Fig. 3a–c, left panel), especially with respect to the early incremental insulin response ($P = 0.008$) and the insulin/glucose ratio (at 30 min, $P = 0.002$; Supplementary Table 4). The higher insulin secretory response during OGTT was consistent across different subsets with OGTT data (meta-analysis: CIR, $P = 0.002$; incremental insulin, $P = 2.4 \times 10^{-4}$; 30 min insulin, $P = 3.8 \times 10^{-4}$) (Supplementary Table 5). Of note, the p.Arg138* carriers had significantly lower proinsulin/C-peptide (20 min: $P = 0.041$; 40 min: $P = 0.043$) and proinsulin/insulin (20 min: $P = 0.006$) ratios during the test meal, suggesting effects on proinsulin conversion (Fig. 2d,e). No differences were seen in glucagon, GLP-1 or free fatty acid (FFA) concentrations during the test meal (Extended Data Fig. 2c–e). Neither the model-based insulin clearance index nor the ratio of insulin and C-peptide areas under the curve during the test meal differed between p.Arg138* and p.Arg138Arg (Extended Data Fig. 2f,g).

Comparison of p.Arg138 versus p.Arg138Arg–p.Arg325.* The above differences were magnified when we restricted the p.Arg138Arg group to carriers of the common risk variant p.Arg325 (Fig. 2, middle panels). The early phase (0–40 min) insulin ($P = 0.026$), insulin/glucose ratio ($P = 0.004$) and CIR ($P = 0.004$; 20 min) were all greater in p.Arg138* carriers versus p.Arg138Arg on a background of p.Arg325 (Supplementary Table 6). Both the proinsulin/C-peptide (20 min: $P = 0.027$, 40 min: $P = 0.044$) and proinsulin/insulin ratios (20 min: $P = 0.003$) were reduced in p.Arg138* carriers (Fig. 2d,e, middle panels).

Comparison of p.Trp325Trp versus p.Arg325. The effect of p.Trp325Trp genotype on glucose and insulin response mimicked the effects of p.Arg138* with pronounced early (20 min) insulin ($P = 0.032$) and C-peptide ($P = 0.030$) responses during the test meal (Fig. 2b,c, right panels, and Extended Data Fig. 2a), as well as increased insulin secretion (30 min insulin: $P = 0.003$, 30 min insulin/glucose: $P = 0.002$, incremental insulin: $P = 0.005$), lower fasting and 120 min proinsulin ($P = 0.006$ and $P = 0.039$, respectively) concentration during OGTT in p.Trp325 carriers (Supplementary Table 4 and Fig. 3b,c, right panels). Moreover, p.Trp325Trp carriers showed a pronounced increased ($P = 0.003$) early incremental insulin secretion during the intravenous glucose tolerance test (IVGTT) (Extended Data Fig. 3a,b and Supplementary Table 4). In patients with newly diagnosed T2D (Supplementary Table 7), the p.Trp325Trp carriers showed a trend ($P = 0.12$) towards enhanced β cell sensitivity to glucose during the OGTT (Extended Data Fig. 3c).

Taken together, all human in vivo data on p.Arg138* show enhanced glucose-stimulated insulin secretion combined with enhanced proinsulin conversion as a potential explanation for T2D protection. The common allele p.Trp325 is also associated with protection from T2D and the metabolic effects are similar but more modest than those of p.Arg138*, suggesting that it might also reduce ZnT8 function.

***SLC30A8* p.Arg138* and p.Lys34Serfs*50 in human iPSCs.** The majority of nonsense *SLC30A8* alleles (including p.Arg138* and p.Lys34Serfs*50) that protect against T2D are located in the first four exons of the eight-exon canonical islet *SLC30A8* transcript (ENST00000456015) and are predicted to undergo nonsense-mediated decay (NMD), a cell surveillance pathway that reduces errors in gene expression by eliminating messenger RNA transcripts with premature stop codons. To confirm that nonsense *SLC30A8* alleles lead to haploinsufficiency through NMD, we used CRISPR–Cas9 to introduce two protective alleles, p.Arg138* and p.Lys34Serfs*50, into the *SLC30A8* locus of the SB Ad3.1 human iPSC line (Extended Data Fig. 4a). Two heterozygous hiPSC lines for the *SLC30A8* p.Arg138* allele (clones B1 and A3) and two homozygous hiPSC lines for *SLC30A8* p.Lys34Serfs*50 (clones B3 and D3) were generated and compared with an unedited wild type and a CRISPR-Sham hiPSC line. All hiPSC lines passed quality control checks, including karyotyping and pluripotency (Extended Data Fig. 4b,c).

We subjected our *SLC30A8*-edited iPSCs to a previously published in vitro endocrine pancreas differentiation protocol[12] (Extended Data Fig. 4). At the end of the seven-stage protocol, *SLC30A8* expression was significantly reduced in cells heterozygous for the p.Arg138* allele (clone A3 0.06 ± 0.03, $P < 0.0001$; clone B1 0.04 ± 0.01, $P < 0.0001$) or homozygous for the p.Lys34Serfs*50 (clone B3 0.04 ± 0.01, $P < 0.0001$; clone D3 0.04 ± 0.01, $P < 0.0001$) compared with unedited control cells (1.02 ± 0.13) (Fig. 4a). In addition, ZnT8 protein was absent in homozygous p.Lys34Serfs*50 hiPSC-derived β-like cells compared with wild-type controls, but was also undetectable in differentiated heterozygous p.Arg138* clones (Fig. 4b,c), suggesting an impact on differentiation. Indeed, the number of *INS*- and *SLC30A8*-transcript expressing cells was reduced in clones with premature stop codons in *SLC30A8* (Extended Data Fig. 4d–g), indicating a reduced formation of β-like cells. To disentangle effects on differentiation from those on expression, allele-specific *SLC30A8* expression was quantified by digital droplet PCR[13] in the heterozygous p.Arg138* lines. Of note, p.Arg138* allele-specific *SLC30A8* expression was reduced compared to the WT allele (clone A3: $24.3 \pm 3.1\%$, $P < 0.0001$; clone B1: $22.2 \pm 1.7\%$, $P < 0.0001$) (Fig. 4d). The reduced expression of the mutant *SLC30A8* allele (clone A3: $18.2 \pm 1.8\%$; clone B1: 18.6%) compared to the wild-type

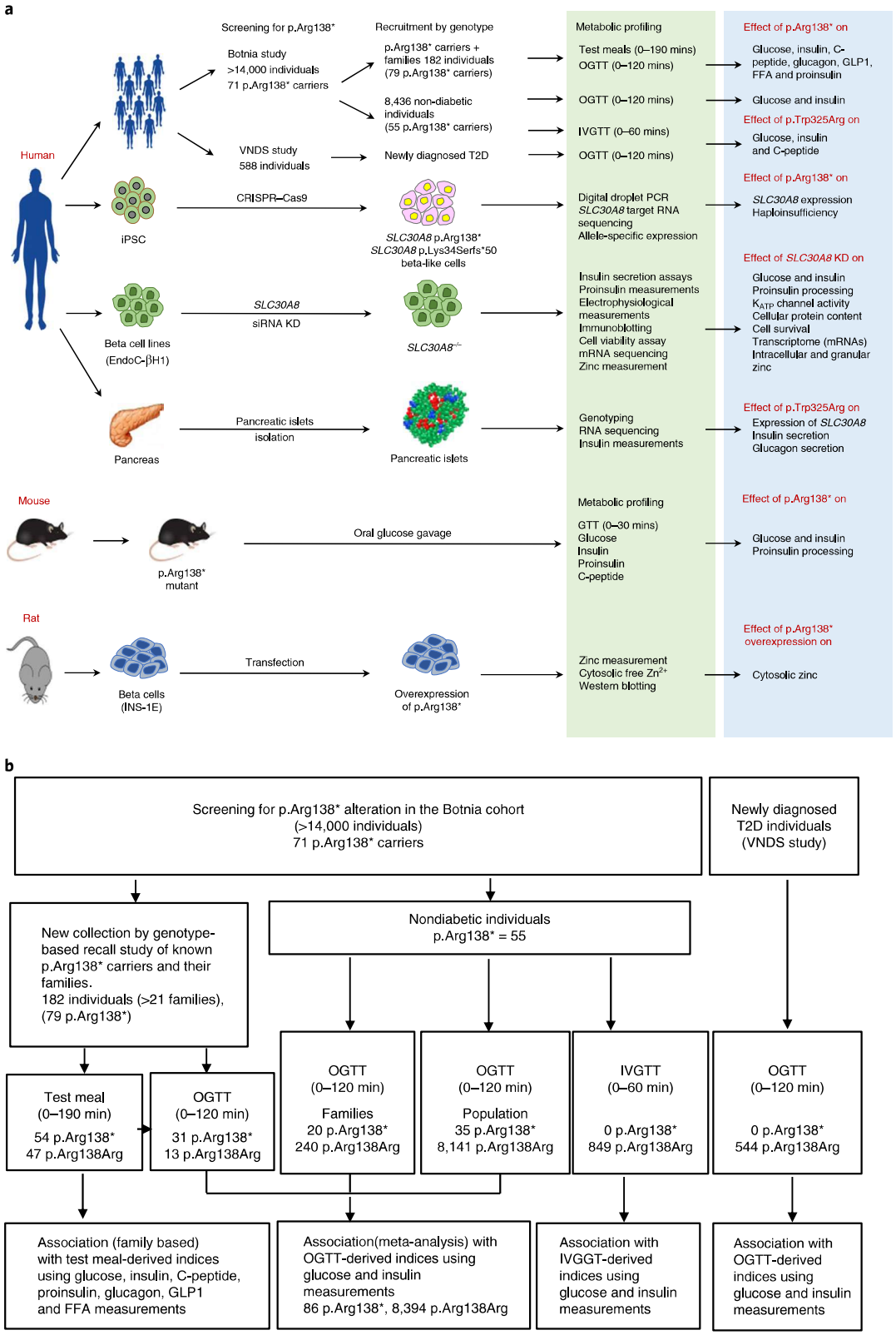


Fig. 1 | A flowchart describing the study design. a, The study design including various model systems (left panels), methods (middle panels) and the purpose of these experiments (right panels). **b**, Detailed description of the human in vivo studies, including a genotype-based recall study for p.Arg138* carriers and their relatives for metabolic studies. OGTT, oral glucose tolerance test; IVGTT, intravenous glucose tolerance test; GTT; glucose tolerance test.

allele (81.7 ± 1.3%) was confirmed ($P < 0.0001$) by targeted *SLC30A8* mRNA sequencing (Fig. 4e and Extended Data Fig. 5c). Although not statistically significant, inhibition of NMD by cycloheximide showed a trend towards greater rescue of the p.Arg138* transcript compared to the p.Arg138Arg transcript (clone A3: 230 ± 61% versus 150 ± 33%, $P = 0.28$ and clone B1:198 ± 45% versus 152 ± 26%,

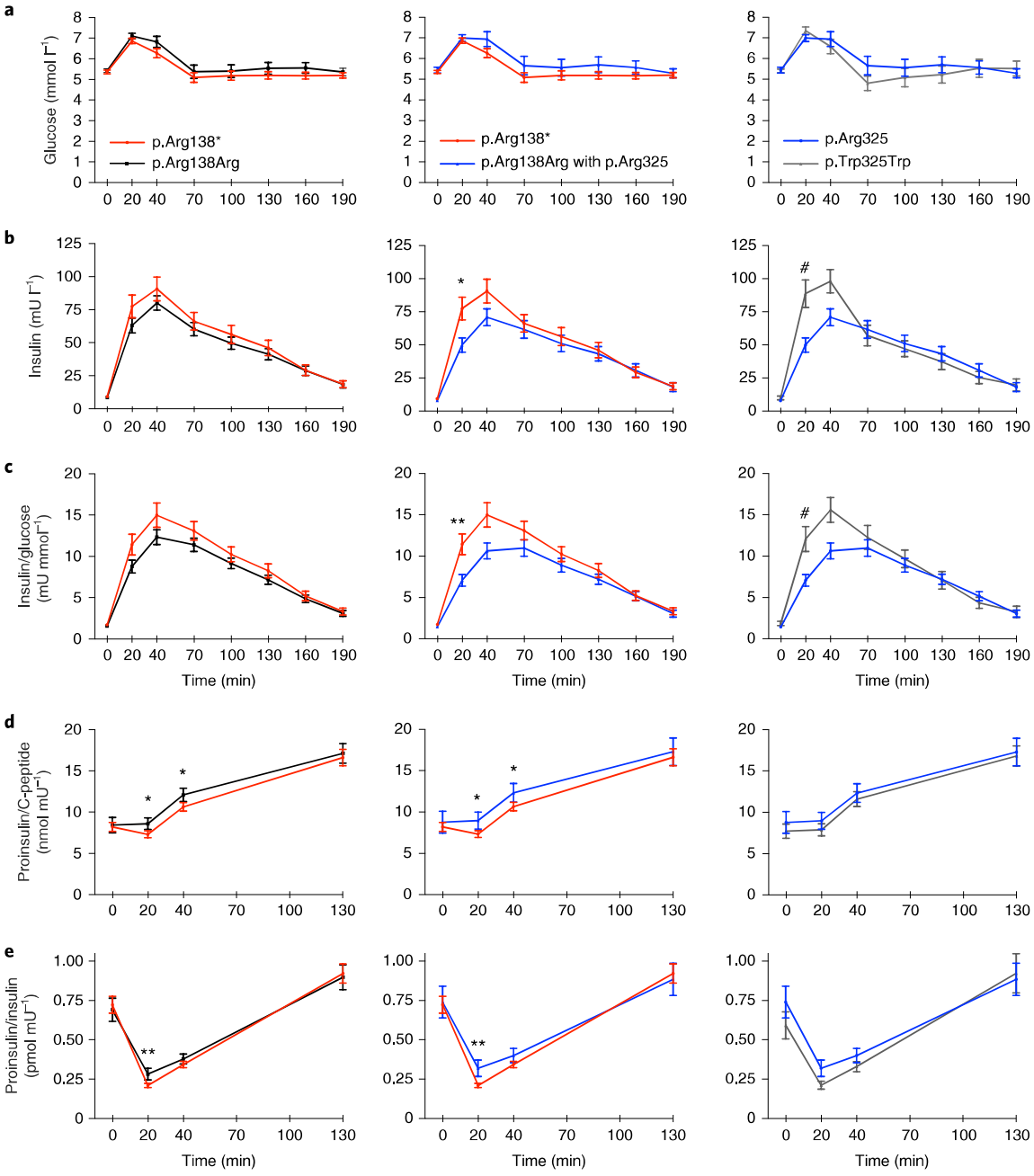


Fig. 2 | *SLC30A8* p.Arg138* enhances insulin secretion and proinsulin processing during test meal. a–e, Association of *SLC30A8* p.Arg138* and p.Trp325Arg variants with plasma glucose (**a**), serum insulin (**b**), insulin/glucose ratio (**c**), proinsulin/C-peptide ratio (**d**) and proinsulin/insulin ratio (**e**) during test meal. Carriers (red, $n = 53$–54) versus noncarriers (black, $n = 37$–47) of p.Arg138* (left panels). Carriers of p.Arg138* (red, $n = 53$–54) versus p.Arg138Arg having the common risk variant p.Arg325 (blue, $n = 25$–31) (middle panels). Carriers of p.Trp325Trp (gray, $n = 12$–16) versus p.Arg325 (blue, $n = 31$) (right panels). Exact numbers used for genetic association analysis are available in the Source Data files. Data are mean ± s.e.m. $^{*}P < 0.05$, $^{**}P < 0.01$ in family-based association (using QTDT[35]) after 100,000 permutations, adjusted for age, sex and BMI (left panels), and age, sex, BMI and genotype of p.Trp325Arg (middle panels). $^{\#}P < 0.05$ in family-based QFAM (as implemented in PLINK[36]) test using 100,000 permutations (see Methods).

$P = 0.39$; Fig. 4f,g). Taken together, these data show that the protective p.Arg138*-*SLC30A8* allele undergoes NMD, resulting in haploinsufficiency for *SLC30A8*.

Impact of *SLC30A8* loss in a human β-cell line. Since human in vivo studies provided strong evidence for a role for the p.Arg138* on insulin secretion and proinsulin processing, we studied the impact of *SLC30A8* loss using short interfering RNA (siRNA)-mediated knockdown (KD) on both phenotypes in a well-characterized human β cell model, EndoC-βH1 (ref. [14]). By siRNA, we achieved a 30–65% decrease in *SLC30A8* mRNA levels (Fig. 5a; $P = 0.003$, and Extended Data Fig. 7a; $P = 5.93 \times 10^{-22}$), which resulted in reduced ZnT8 protein levels (Fig. 5b and Extended Data Fig. 6a) associated with a reduction in intracellular zinc content (34%, $P = 0.002$; Fig. 5c). Assessment of dense core granule zinc content in *SLC30A8*-silenced EndoC-βH1 cells (Extended Data Fig. 6a), through the use of the cell surface-attached fluorescent zinc probe ZIMIR[15] (Extended Data Fig. 6b,c) showed a tendency ($P = 0.10$; Extended

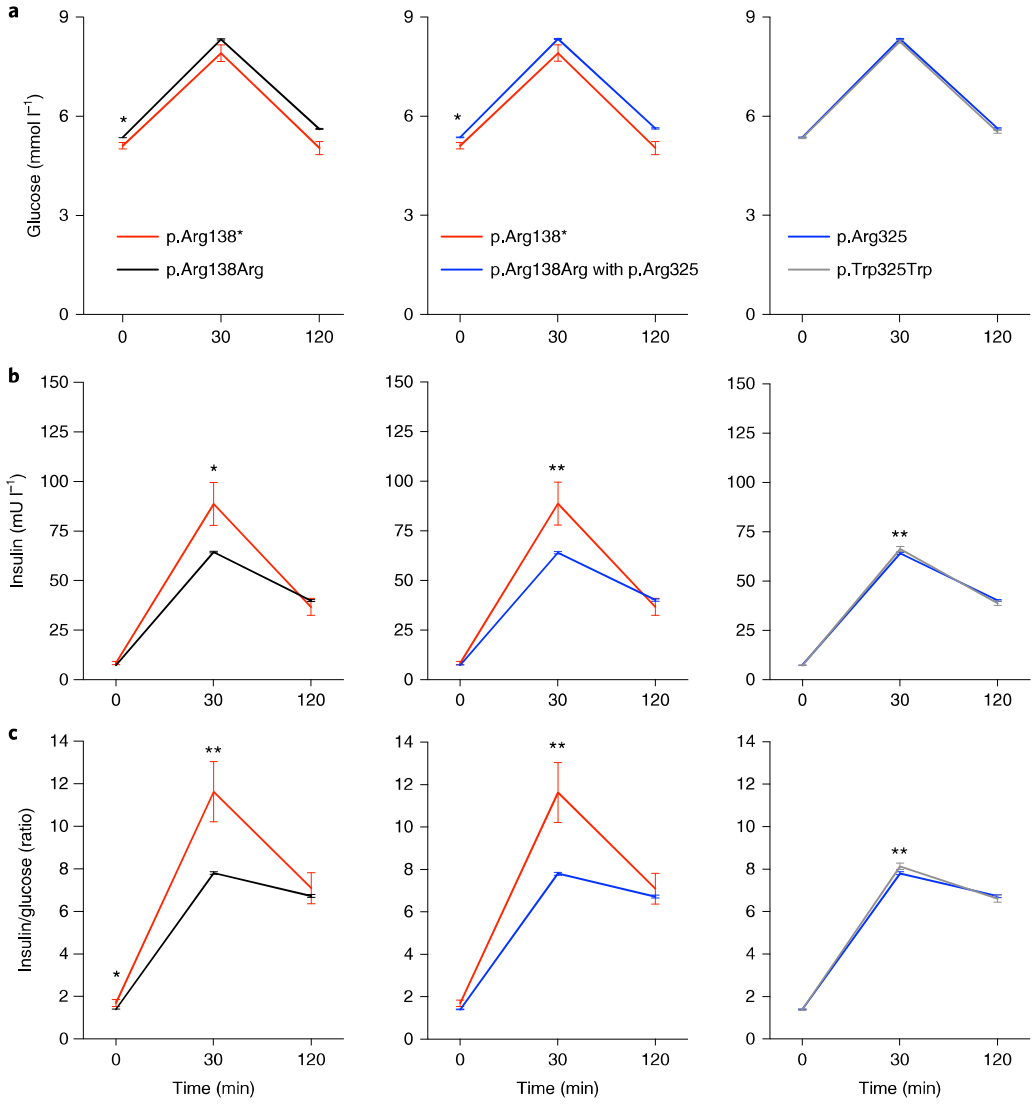


Fig. 3 | *SLC30A8* p.Arg138* and p.Trp325 enhance insulin secretion during OGTT. a–c, Association of *SLC30A8* p.Arg138* and p.Trp325Arg with plasma glucose (**a**), serum insulin (**b**) and insulin/glucose ratio (**c**) during an OGTT. Carriers (red, $n = 34$–35) versus noncarriers (black, $n = 7{,}954$–8,141) of p.Arg138* (left panels). Carriers of p.Arg138* (red, $n = 34$–35) versus p.Arg138Arg having the common risk variant p.Arg325 (blue, $n = 6{,}728$–6,893) (middle panels). Carriers of p.Trp325Trp (gray, $n = 1{,}226$–1,248) versus p.Arg325 (blue, $n = 6{,}728$–6,893) (right panels). Exact numbers used for genetic association analysis are available in Supplementary Table 5. Data are shown as mean ± s.e.m. $^{*}P < 0.05$, $^{**}P < 0.01$ for additive effects, calculated using a mixed model adjusting for genetic relationship, age, sex, BMI and, additionally, the genotype of p.Trp325Arg only for the middle panel, as implemented in GCTA[37] (see Methods).

Data Fig. 6d) for reduced Zn^{2+} release, which is consistent with the loss of zinc from this compartment and also with previous findings in islets from *Slc30a8* null mice[15].

KD of *SLC30A8* had no significant effect on stimulated insulin secretion either in response to 20 mM glucose or to the sulfonylurea tolbutamide (which closes K_{ATP} channels; Fig. 5d). However, basal insulin secretion was significantly increased in si*SLC30A8*-transfected cells ($P = 0.048$), which reduced marginally the glucose stimulation index ($P = 0.26$; Fig. 5e). In addition, the inhibitory effect of diazoxide, a K_{ATP} channel opener, on glucose-stimulated insulin secretion, was significantly reduced ($P = 8 \times 10^{-4}$; Fig. 5d). There was no effect of *SLC30A8* knockdown on insulin content (Fig. 5f). We then measured the resting membrane conductance (G_m) in si*SLC30A8*-transfected cells incubated with 100 µM diazoxide (to activate K_{ATP} channels). In control cells, G_m was in agreement with previous reports[16], while *SLC30A8* knockdown reduced G_m by 65% ($P = 0.002$; Fig. 5g) without altering cell size (Fig. 5h). Indeed, reduced K_{ATP} channel activity is consistent with the reduced expression observed of the genes encoding the K_{ATP} channel subunits, SUR1 (*ABCC8*) and Kir6.2 (*KCNJ11*) ($P = 0.04$ and 0.06 respectively; Fig. 5i). In addition, insulin secretion elicited by the combination of elevated 50 mM extracellular K^+ ($[K^+]_o$) to depolarize the cells and open voltage-gated Ca^{2+} channels and 16.7 mM glucose was significantly higher after *SLC30A8* knockdown ($P = 0.002$; Fig. 5j). The proinsulin–insulin ratios (both total content; $P = 0.007$ and secreted hormones; $P = 0.004$) and proinsulin concentrations (total content; $P = 0.011$, secreted in media; $P = 0.017$) were decreased in si*SLC30A8*-transfected cells, but no differences in either protein or mRNA levels of genes (*PCSK1*, *CPE* and *PCSK2*) involved in proinsulin processing in *SLC30A8* KD cells was observed (Fig. 5m–q and Extended Data Fig. 7a). However, we did observe increased AKT phosphorylation (pAKT-473) and improved cell survival under ER stress ($P = 0.016$ and 0.032, respectively; Fig. 5r–t) in *SLC30A8* KD cells.

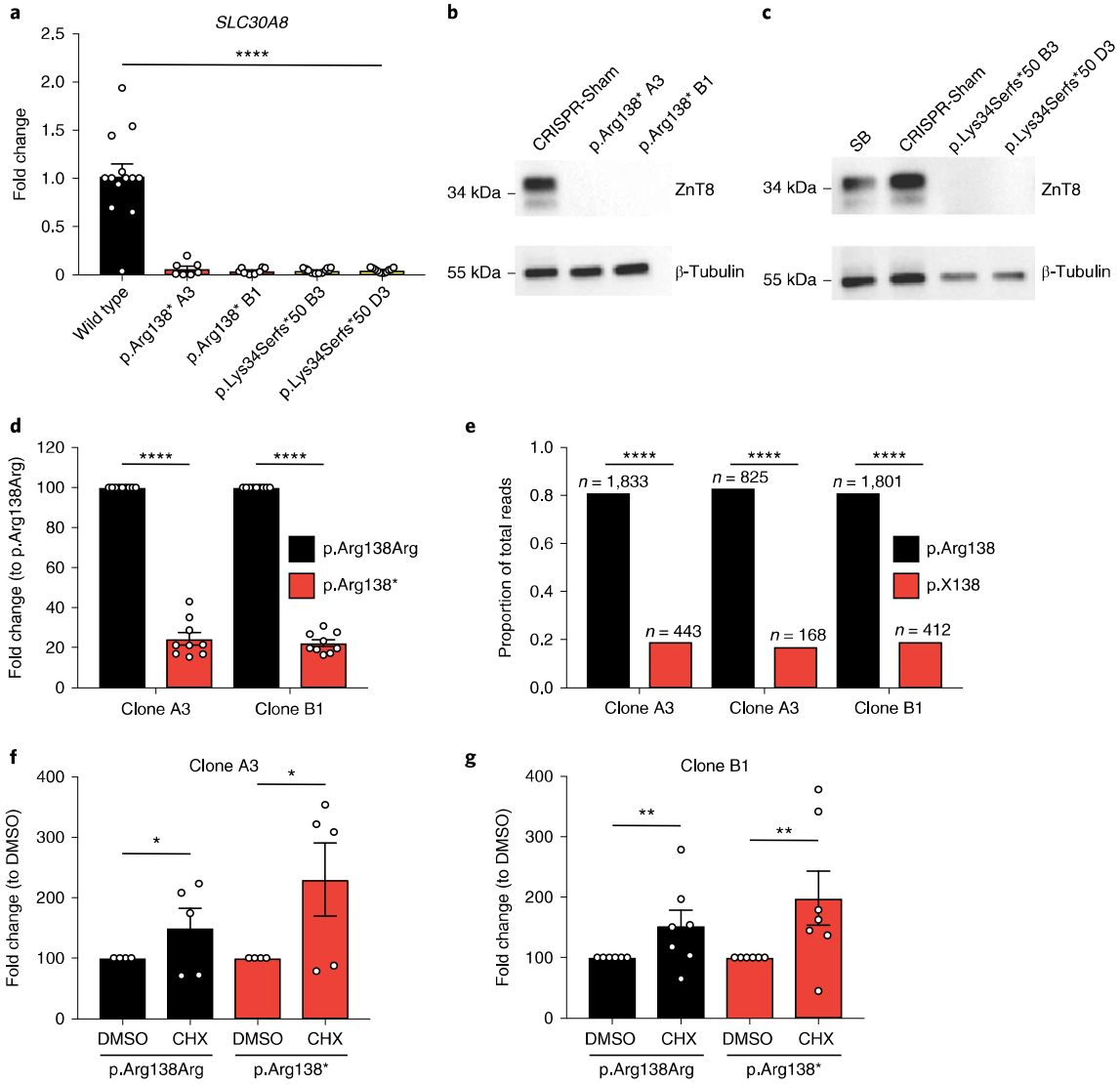


Fig. 4 | β-like cells derived from SLC30A8 p.Arg138* iPSCs display haploinsufficiency of *SLC30A8*. a–c, Characterization of *SLC30A8* expression at the mRNA (**a**) and protein level (**b**,**c**) in cells heterozygous for *SLC30A8* p.Arg138* (**b**) or homozygous for *SLC30A8* p.Lys34Serfs*50 (**c**). Gene expression data normalized to *TBP* and expressed as fold change relative to p.Arg138Arg control ($n = 7$–13 wells from three differentiations). **d**,**e**, Allele-specific expression of p.Arg138Arg (black bar) and p.Arg138* (red bar) in clone A3- or clone B1-derived cells using digital droplet PCR-based probes (**d**), also validated by target *SLC30A8* mRNA sequencing of p.Arg138* clones (*n*, number of unique sequencing reads for each allele) (**e**) . **f**,**g**, Allele-specific expression of p.Arg138Arg (black bar) and p.Arg138* (red bar) in clone A3- (**f**) and clone B1-derived (**g**) cells treated with DMSO or cycloheximide (CHX) for 4 h. Allele-specific expression data (mean ± s.e.m.) determined by digital droplet PCR are presented as fold change relative to p.Arg138Arg transcript (**d**, $n = 9$ wells from three differentiations) or to DMSO control (**f**,**g**, $n = 4$–7 wells from two differentiations). Blots have been cropped and the corresponding full blots are available in the Source Data files. $^*P < 0.05$, $^{**}P < 0.01$, $^{****}P < 0.0001$ one-way ANOVA Holm–Sidak multiple comparison test (**a**) or one-sample *t*-test (**d**,**f**,**g**) or binominal test (one-tailed) considering 0.5 as the expected probability (**e**).

We next examined the effects of *SLC30A8* knockdown on gene expression in EndoC-βH1 cells by mRNA sequencing (RNA-seq) of si*SLC30A8* treated and siScramble cells ($n = 8$ for both, see Supplementary Note). We observed a total of 674 significantly differently expressed genes among 12,956 protein-coding genes that passed the quality control filters (Extended Data Fig. 7 and Supplementary Dataset 1). RNA-seq confirmed the reduction in *KCNJ11* and *ABCC8* expression in *SLC30A8* KD cells ($P = 2.6 \times 10^{-18}$ and $P = 2.04 \times 10^{-9}$, respectively; Extended Data Fig. 7c) as seen earlier by quantitative PCR (qPCR) (Fig. 5i). In addition, further genes involved in the regulation of β cell excitability/exocytosis, including *CACNA1C*, were upregulated in KD cells (Extended Data Fig. 7c). Moreover, expression of genes associated with β cell maturation and development were also influenced by *SLC30A8* knockdown, with decreased expression of *NKX6.1* and *PDX1* and increased expression of *SOX4*, *SOX6* and *SOX11* (Extended Data Fig. 7b). A pathway-enrichment analysis of differentially expressed genes showed enrichment of genes involved in the WNT signaling and insulin secretion pathways (Extended Data Fig. 7 and Supplementary Dataset 1). A global gene-set enrichment analysis of all expressed genes ($N = 12{,}956$) using a gene ontology database showed enrichment of genes involved in positive regulation of TOR signaling (Extended Data Fig. 7) in knockdown cells. Collectively, these data demonstrate a link between *SLC30A8* expression and transcriptional networks involved in cell development, cell fate and plasma membrane polarization.

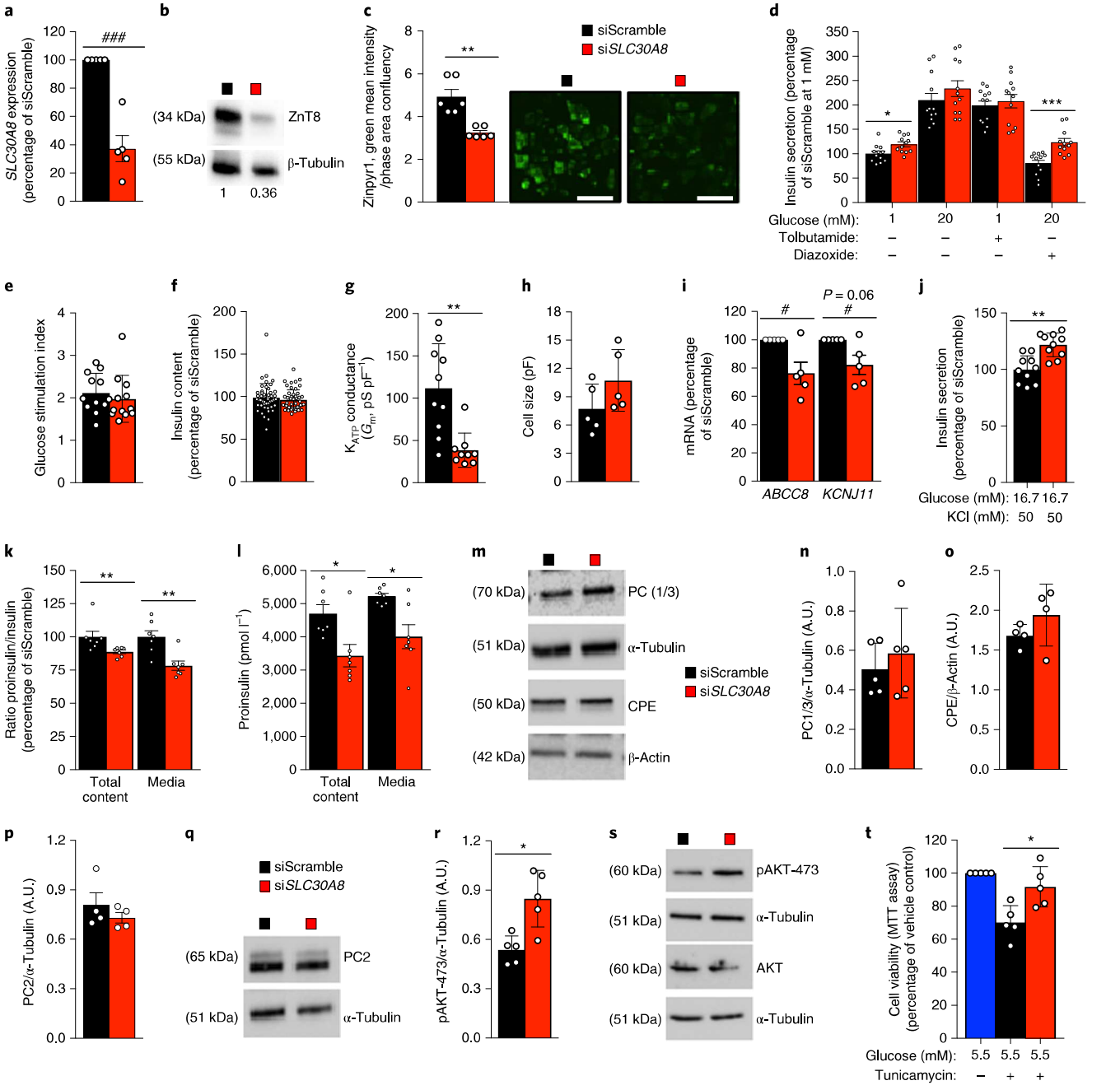


Fig. 5 | *SLC30A8* knockdown leads to enhanced insulin secretion, proinsulin processing and cell viability in the human pancreatic EndoC-βH1 cells. **a**,**b**, Effect of siRNA-mediated KD on *SLC30A8* mRNAs ($n=5$) and protein. Intensities of the signal normalized to control condition are reported below the blot. **c**, Measurement of intracellular zinc using zinc-specific fluorescent dye Zinpyr-1 ($n=6$, scale bars: 50 µm). **d**–**f**, Effect of KD on insulin secretion (normalized to total insulin content and average siScramble basal secretion) stimulated by glucose and K_{ATP} channel regulators (as labeled, and $n=12$) (**d**), glucose stimulation index induced by 20 mmol l^{-1} glucose stimulation ($n=12$) (**e**) and insulin content ($n=48$) (**f**). **g**–**j**, Effect of siSLC30A8 KD on K_{ATP} channel conductance (G_m, siScramble $n=10$ cells, si*SLC30A8* $n=9$ cells) (**g**), cell size ($n=5$) (**h**), expression of K_{ATP} channel subunits ($n=5$) (**i**) and on insulin secretion stimulated by KCl and high glucose ($n=10$) (**j**). **k**,**l**, Effect of KD on proinsulin processing estimated by proinsulin/insulin ratio ($n=7$) (**k**) and proinsulin concentration ($n=7$) (**l**). **m**–**q**, Effect of KD on protein expression of proinsulin processing enzymes PC1/3, CPE (immunoblot (**m**) and densitometry (**n**, $n=5$; **o**, $n=4$)) and PC2 (densitometry (**p**) ($n=4$) and immunoblot (**q**)). **r**–**t**, Effect of KD on basal (5.5 mM glucose) AKT phosphorylation (densitometry (**r**) ($n=5$) and immunoblot (**s**); phospho-AKT-Ser473, total AKT) and cell viability under ER stress (**t**, MTT assay, $n=5$). Black bar graphs: control cells (siScramble); red bar graphs: *SLC30A8* knockdown cells (si*SLC30A8*). Blots have been cropped and the corresponding full blots are available in the Source Data files. Data are mean ± s.e.m. $^{*}P<0.05$, $^{**}P<0.01$, $^{***}P<0.001$ using Mann–Whitney *U*-test; $^{\#}P<0.05$; $^{\#\#\#}P<0.01$ using one-sample *t*-test; Bonferroni multiple correction for four conditions (**d**).

Metabolic phenotype of *Slc30a8* p.Arg138* mice. Since neither global nor tissue-specific *Slc30a8* knockout mouse models have recapitulated the human phenotype in carriers of the *SLC30A8* p.Arg138* allele, we attempted to overcome this problem by using a mouse model carrying the human *Slc30a8* p.Arg138* allele[10]. On a standard chow diet there was no evidence of enhanced insulin

secretion[10] but when subjected to a high-fat diet (HFD) the mice showed a significant increase in insulin secretion ($P=0.0016$; Fig. 6i), and similar changes in proinsulin/insulin ($P=0.0004$) and proinsulin/C-peptide ratios ($P=0.0001$) as seen in human carriers (Fig. 6f,g). No significant changes were seen in insulin clearance, glucose or insulin tolerance (Fig. 6h,j,k).

Impact of p.Arg138* on cytosolic zinc in INS-1 cells. Although we found no evidence for the presence of a truncated protein in the human p.Arg138* β cell model (Fig. 4), as also reported for the p.Arg138* mice[10], we still explored the possibility that a truncated protein could result from mRNA evading NMD. Transient overexpression of tagged ZnT8 p.Arg138* fusion protein in a rat insulinoma cell line, INS-1E, showed distinct punctate distribution patterns, consistent with localization of the truncated ZnT8 protein to secretory granules, as previously observed with the full-length protein[17] (Extended Data Fig. 8a–c). Additionally, western blotting showed stable expression of truncated ZnT8 in native INS-1E cells (Extended Data Fig. 8d).

To investigate the effects of a truncated ZnT8 protein on cytosolic free Zn^{2+}, we used a genetically encoded Zn^{2+} sensor, eCALWY-4 (ref. [18]). Overexpression of the truncated protein (p.Arg138*; Extended Data Fig. 8e,f) had no impact on cytosolic free Zn^{2+} when expressed in INS-1 (832/13) cells (Extended Data Fig. 8g).

Influence of *SLC30A8* p.Trp325Arg in primary human islets. Although we were unable to study the effects of the rare T2D-protective alleles in primary human islets, we were able to assess the impact of the p.Trp325Arg genotype on in vitro insulin and glucagon secretion. Islets obtained from cadaveric p.Trp325 carriers secreted more insulin ($P=0.0153$) than p.Arg325Arg carriers when stimulated with high glucose (16.7 mM) and depolarizing $[K^+]_o$ (70 mM), in line with findings for *SLC30A8* knockdown in EndoC-βH1 (Fig. 7a,b). Interestingly, a trend towards increased glucose responsiveness was already observed at submaximal glucose stimulation (6 mM) (Fig. 7c). Increasing glucose from 1 mM to 6 mM stimulated insulin secretion 2.2- ($P=0.031$) and 2.7 ($P=0.012$)-fold in p.Arg325 and p.Trp325 carriers, respectively, with no effect on insulin content (Fig. 7c,d).

As *SLC30A8* is highly expressed in human alpha cells[1], we also measured glucagon secretion from the same islets (Fig. 7e,f). In islets from p.Arg325Arg donors, 6 mM glucose inhibited glucagon secretion by ~50% compared to 1 mM glucose. In islets from p.Trp325Arg donors, glucagon secretion at 1 mM glucose was reduced by 50% ($P=0.033$) compared to p.Arg325Arg donors, with no effect on glucagon content (Fig. 7e,f).

Finally we explored coexpression of *SLC30A8* with relevant candidate genes (*INS*, *GCG*, encoding proinsulin processing and K_{ATP} channel subunits) as well as the impact of p.Trp325Arg on their expression (Fig. 7g–j). *SLC30A8* transcript levels showed strong positive correlation with expression of all candidate genes (*GCG*, $P\leq1.3\times10^{-7}$, *PCSK1*, $P\leq1.5\times10^{-7}$, *PCSK2*, $P\leq4.6\times10^{-10}$, *CPE*, $P\leq3.2\times10^{-6}$, *KCNJ11*, $P\leq7.1\times10^{-7}$ and *ABCC8*, $P\leq1.6\times10^{-11}$) except *INS*. The protective p.Trp325 allele showed a trend (nonsignificant) of decreased *SLC30A8* expression ($P=0.053$), as well as genes involved in proinsulin processing such as *PCSK1* ($P=0.041$), *PCSK2* ($P=0.045$) and *ABCC8* ($P=0.049$).

Taken together, these data suggest that the common T2D-protective allele (p.Trp325) may also improve the response to a glucose challenge (Figs. 2 and 3) by enhancing insulin secretion and possibly by reducing glucagon secretion in primary human islets.

Discussion

The current study demonstrates the strengths of using human models for studying the consequences of variants associated with human diseases. Although over 30 T2D-protective LoF or missense alleles in *SLC30A8* have been reported[19], previous studies in rodents[6,17,20–22] have failed to provide a mechanistic explanation for this protection. This human study robustly shows that enhanced insulin responsiveness to glucose, combined with enhanced proinsulin processing, contributes to the protection from T2D. As all LoF carriers had the common risk p.Arg325 allele on the same haplotype, the effect of the LoF allele was most evident when compared against p.Arg325 carriers, emphasizing the importance of considering the genetic background of human LoF carriers.

In our human iPSC-derived β-like cells, two different LoF alleles (p.Arg138* and p.Lys34Serfs*50) show a clear reduction in *SLC30A8* expression, suggesting that NMD-induced haploinsufficiency is likely a common mechanism for rare LoF alleles in *SLC30A8*. Although a recent study showed increased zinc transporter activity for the common risk variant p.Arg325 (ref. [23]), it should be kept in mind that the common p.Trp325Arg locus is complex, as another 3′ untranslated region variant (rs3802177) is in strong linkage disequilibrium with p.Trp325Arg. This variant (rs3802177) may potentially affect *SLC30A8* transcription/translation, further modulating the effect of p.Trp325Arg on ZnT8 function.

Interestingly, in our human iPSC-based model of *SLC30A8* LoF mutations (p.Arg138* and p.Lys34Serfs*50), loss of ZnT8 protein decreased the formation of *INS*$^+$ cells. In support of this, we observed decreased expression of multiple genes associated with β cell development (including *PDX1*) and changes in expression of genes involved in WNT signaling (including *TCF7L2*) after partial ZnT8 loss (knockdown) in the human EndoC-βH1 cell line. However, insulin content was not affected, suggesting that partial ZnT8 loss is unlikely to affect β cell development. In support of this, complete ZnT8 null mice were viable, with preserved β cell function[6,9,10]. We have not been able to examine this in humans, as we did not find any human homozygous for *SLC30A8* LoF alleles.

Whereas data from all our substudies are consistent with increased insulin secretion, the precise molecular mechanisms for the involvement of zinc and zinc transporters remain elusive.

Data derived from the human β cell line show a clear decrease in intracellular zinc content and a trend towards a reduction in cosecreted granular zinc after ZnT8 loss. These observations are consistent with previous rodent studies demonstrating that loss of ZnT8 function reduces total islet zinc content (including R138* mice[10]) as well as free Zn^{2+} in the cytosol and granules[6,9,17,24]. The LoF p.Arg138* and p.Lys34Serfs*50 alleles in humans are likely to exert similar effects on intracellular and granular zinc concentrations. In the present study, overexpression of the truncated p.Arg138* protein in INS-1 cells did not result in changes in cytosolic zinc concentrations, which is consistent with haploinsufficiency. In contrast, a recent human study showed that the T2D-risk p.Arg325 allele was associated with higher islet zinc concentrations[25]. In support of a potential role for zinc in the development of diabetes, Zn^{2+} plays an important role as a regulator of cellular excitability[26] in the central nervous system. In β cells, Zn^{2+}-induced inhibition of L-type voltage-gated Ca^{2+} channels could result in inhibition of insulin secretion[27], whereas, Zn^{2+} has been reported to directly activate K_{ATP} channel currents[28]. The downregulation of the K_{ATP} channel subunit genes following *SLC30A8* knockdown suggests that there is coordinated expression of these genes. It is therefore tempting to hypothesize that *SLC30A8* expression may contribute to the normal K_{ATP} channel density in both β and α cells, but this remains to be demonstrated. The combined consequence of a reduction in K_{ATP} channel gene expression and a reduction in para/autocrine electrophysiological regulation (on calcium and K_{ATP} channel activities) by Zn^{2+} could contribute to enhanced insulin secretion, as well as reducing glucagon secretion at low glucose levels. Indeed, a similar effect on the secretory capacity of alpha cells has been observed on inhibition of K_{ATP} channel activity using tolbutamide[29].

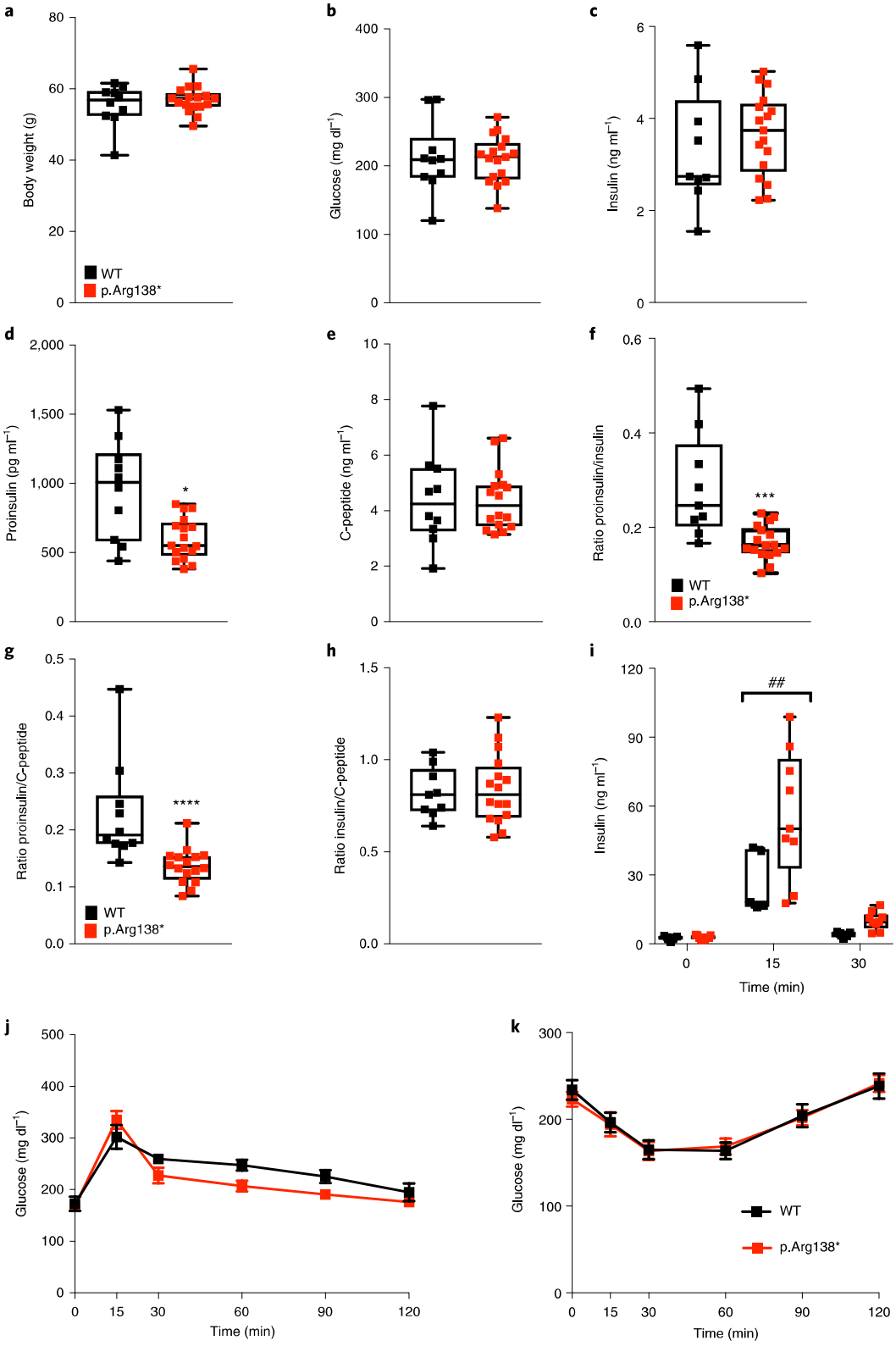


Fig. 6 | Male *Slc30a8* p.Arg138* mice on HFD show enhanced insulin secretion and proinsulin processing. **a**–**h**, Body weight ($n = 10/17$ WT/p.Arg138*) (**a**), circulating glucose ($n = 10/17$ WT/p.Arg138*) (**b**), insulin ($n = 9/17$ WT/p.Arg138*) (**c**), proinsulin ($^{*}P = 0.011$, $n = 10/17$ WT/p.Arg138*) (**d**), C-peptide ($n = 10/16$ WT/p.Arg138*) (**e**), proinsulin/insulin ratio ($n = 9/17$ WT/p.Arg138*) (**f**), proinsulin/C-peptide ratio ($n = 10/16$ WT/p.Arg138*) (**g**) and insulin/C-peptide ratio ($n = 9/16$ WT/p.Arg138*) (**h**) in fasted WT and p.Arg138* mice after 20 weeks on HFD. **i**, Insulin response to oral glucose (2 g kg^{-1}) exposure ($n = 5/9$ WT/p.Arg138*) after 30 weeks on HFD. **j**,**k**, Blood glucose levels over time after oral glucose (2 g kg^{-1}) exposure ($n = 5/11$ WT/p.Arg138*) after 29 weeks on HFD (**j**) and intraperitoneal injection of insulin (1.75 U kg^{-1}) after 28 weeks on HFD ($n = 11/13$ WT/p.Arg138*) (**k**). Black boxes, wild-type mice (WT); red boxes, *Slc30a8* p.Arg138* mice. Box plots depict the interquartile range, median and minimum/maximum values. $^{*}P < 0.05$, $^{***}P < 0.001$, $^{****}P < 0.0001$ using Mann–Whitney *U*-test; $^{\#\#}P < 0.01$ using two-way ANOVA (repeated measurements) and Sidak's multiple comparison test.

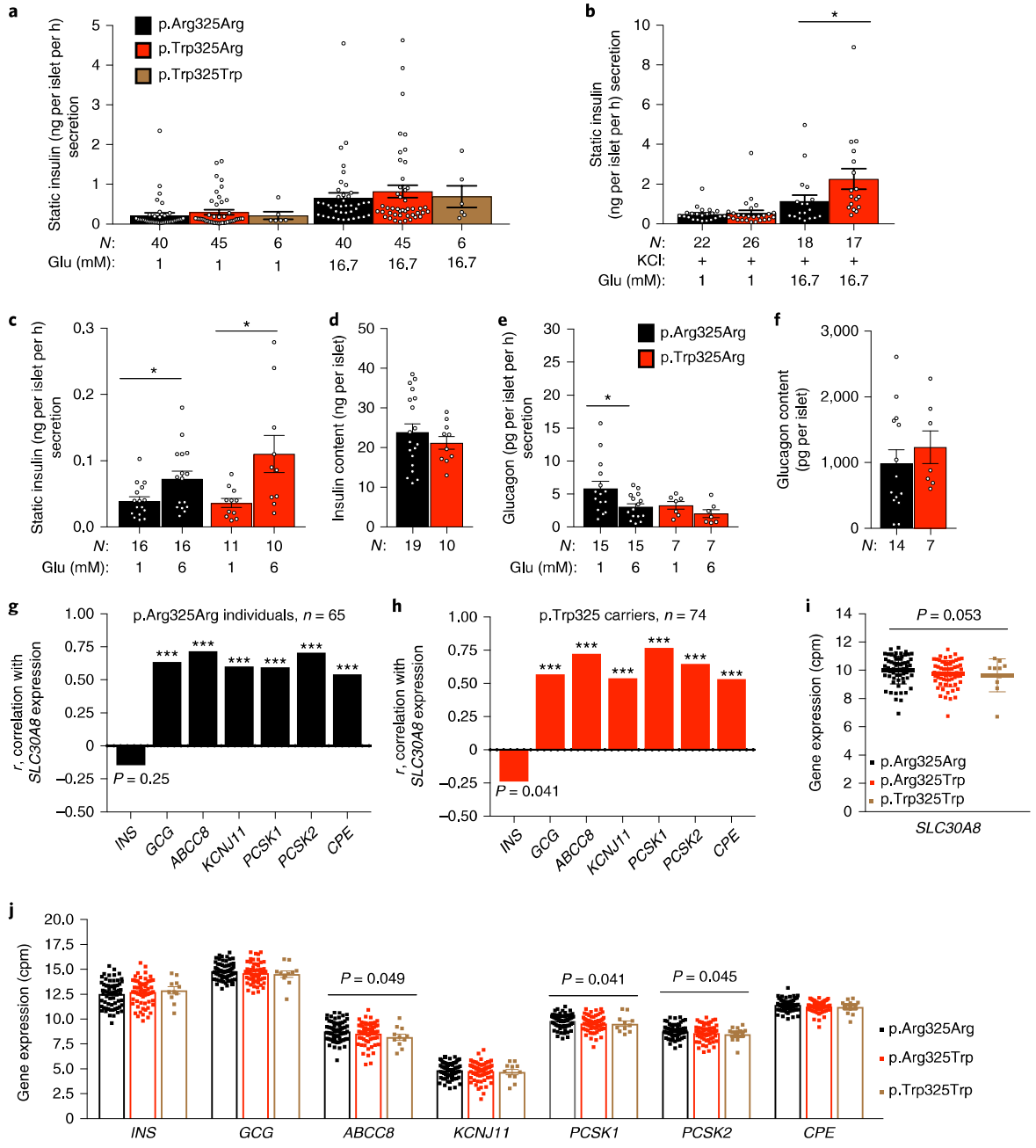


Fig. 7 | *SLC30A8*- p.Trp325 leads to enhanced insulin secretion in human islets. **a**, Effect of p.Trp325Arg genotype on static insulin secretion in the presence of low and high glucose stimulatory conditions. **b**, Effect of p.Trp325Arg genotype on static insulin secretion in the presence of low or high glucose and KCl. **c**,**d**, Effect of p.Trp325Arg genotype on static insulin secretion in submaximal stimulatory conditions (6 mM glucose) (**c**) and their insulin contents (**d**). **e**,**f**, Static glucagon response to glucose (**e**) and glucagon (**f**) content at basal glucose. **g**–**j**, Correlation of *SLC30A8* expression with candidate genes of *INS*, *GCG*, proinsulin processing genes and K_{ATP} channel subunits genes among p.Arg325Arg individuals (**g**) and p.Trp325 carriers (**h**) and effect of p.Trp325Arg genotype (p.Arg325Arg, $n = 65$; p.Trp325Arg, $n = 63$; p.Trp325Trp, $n = 11$ individuals) on expression (cpm = $\log_2$(counts per million)) of *SLC30A8* (**i**) and other candidate genes (**j**). Experiments have been performed in two different centers: LUDC (**a**,**b**,**g**,**h**,**i**,**j**) and Oxford (**c**–**f**). Data are mean ± s.e.m., Glu; glucose, *N*; number of experiments. $^*P < 0.05$, $^{***}P < 0.0001$ using Mann–Whitney *U*-test (**b**–**f**) or Spearman correlation coefficient (*r*) with two-tailed *P* values (**g**,**h**). Three genotype comparison (**a**,**i**,**j**) by linear regression using additive effect adjusting for age, sex and islet purity as implemented in PLINK[36] (see Methods).

Zn^{2+} is essential for proper insulin crystallization, and loss of ZnT8 results in reduced dense core granules and increased granule diameter possibly due to more noncrystalline insulin, which is expected to be released rapidly during exocytosis[6,17,30]. The increased basal insulin secretion seen in the human β cell line following *SLC30A8* knockdown and also reported in islets from *Slc30a8* null mice[17], who also display impaired insulin crystallization, suggest that there is preferential release of noncrystalline insulin. It has also been suggested that increased insulin clearance could explain the decrease in circulating insulin concentrations[20]. However, we did not find any support for changes in insulin clearance in human p.Arg138* carriers or p.Arg138* mice on HFD, nor did we see any effect on the incretin hormone GLP-1 in humans.

In contrast to the consistent effect of human p.Arg138* and p.Trp325Arg on proinsulin processing, studies in mice following the loss of ZnT8 have reported inconsistent effects on proinsulin processing, ranging from inhibitory effects[20,21] to beneficial effects in transgenic Trp325Arg mice on HFD[22]. The reasons for this remain unclear, but since blocking the insulin receptor prevents the positive effect on insulin processing in p.Arg138* mice[10], it is possible that insulin receptor signaling modulates the effect of loss of ZnT8 on proinsulin processing.

The most reproducible finding in all the substudies of *SLC30A8* loss was enhanced glucose-stimulated insulin secretion associated with increased conversion of proinsulin to C-peptide and insulin. Carriers of the p.Trp325 T2D-protective allele also showed a similar phenotype, consistent with previous published studies reporting impaired proinsulin conversion in carriers of the risk p.Arg325 allele[31,32]. It has been suggested that it takes some time for insulin to mature and become biologically active[33,34]. It is therefore possible that the pronounced effect of the LoF T2D-protective allele at 20 and 40 min of the test meal could reflect an effect on the time course of insulin maturation.

Our study has both strengths and limitations. Among its strengths are the comprehensive studies in humans recalled by their genotype in a region of the world where the p.Arg138* allele is enriched. The complementary studies in both human cell lines, as well as the humanized mice, can also be considered as a strength. A limitation is the inconclusive measurement of zinc in different cellular compartments and characterization of the insulin granule cargo. This will require the development of better intracellular sensors for zinc, which will enable such studies in the future.

In conclusion, our data consistently demonstrate that heterozygosity for a LoF allele p.Arg138* and homozygosity for a common allele p.Trp325Trp of *SLC30A8* are associated with increased insulin secretion capacity and a lower risk of developing T2D and in the absence of any on-target adverse events. Therefore, ZnT8 remains an appealing safe target for antidiabetic therapies aimed at preserving β cell function.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-019-0513-9.

Received: 13 October 2018; Accepted: 13 September 2019;
Published online: 1 November 2019

References

1. Chabosseau, P. & Rutter, G. A. Zinc and diabetes. *Arch. Biochem. Biophys.* **611**, 79–85 (2016).
2. Chimienti, F., Devergnas, S., Favier, A. & Seve, M. Identification and cloning of a β-cell-specific zinc transporter, ZnT-8, localized into insulin secretory granules. *Diabetes* **53**, 2330–2337 (2004).
3. Flannick, J. et al. Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. *Nat. Genet.* **46**, 357–363 (2014).
4. Parsons, D. S., Hogstrand, C. & Maret, W. The C-terminal cytosolic domain of the human zinc transporter ZnT8 and its diabetes risk variant. *FEBS J.* **285**, 1237–1250 (2018).
5. Sladek, R. et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **445**, 881–885 (2007).
6. Lemaire, K. et al. Insulin crystallization depends on zinc transporter ZnT8 expression, but is not required for normal glucose homeostasis in mice. *Proc. Natl Acad. Sci. USA* **106**, 14872–14877 (2009).
7. Pound, L. D. et al. Deletion of the mouse Slc30a8 gene encoding zinc transporter-8 results in impaired insulin secretion. *Biochem. J.* **421**, 371–376 (2009).
8. Wijesekara, N. et al. Beta cell-specific Znt8 deletion in mice causes marked defects in insulin processing, crystallisation and secretion. *Diabetologia* **53**, 1656–1668 (2010).
9. Mitchell, R. K. et al. Molecular genetic regulation of Slc30a8/ZnT8 reveals a positive association with glucose tolerance. *Mol. Endocrinol.* **30**, 77–91 (2016).
10. Kleiner, S. et al. Mice harboring the human SLC30A8 R138X loss-of-function mutation have increased insulin secretory capacity. *Proc. Natl Acad. Sci. USA* **115**, E7642–E7649 (2018).
11. Groop, L. et al. Metabolic consequences of a family history of NIDDM (the Botnia study): evidence for sex-specific parental effects. *Diabetes* **45**, 1585–1593 (1996).
12. Rezania, A. et al. Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. *Nat. Biotechnol.* **32**, 1121–1133 (2014).
13. Miyaoka, Y., Chan, A. H. & Conklin, B. R. Using digital polymerase chain reaction to detect single-nucleotide substitutions induced by genome editing. *Cold Spring Harb. Protoc.* https://doi.org/10.1101/pdb.prot086801 (2016).
14. Scharfmann, R. et al. Development of a conditionally immortalized human pancreatic β cell line. *J. Clin. Invest.* **124**, 2087–2098 (2014).
15. Li, D. et al. Imaging dynamic insulin release using a fluorescent zinc indicator for monitoring induced exocytotic release (ZIMIR). *Proc. Natl Acad. Sci. USA* **108**, 21063–21068 (2011).
16. Hastoy, B. et al. Electrophysiological properties of human beta-cell lines EndoC-βH1 and -βH2 conform with human beta-cells. *Sci Rep* **8**, 16994 (2018).
17. Nicolson, T. J. et al. Insulin storage and glucose homeostasis in mice null for the granule zinc transporter ZnT8 and studies of the type 2 diabetes-associated variants. *Diabetes* **58**, 2070–2083 (2009).
18. Vinkenborg, J. L. et al. Genetically encoded FRET sensors to monitor intracellular Zn^{2+} homeostasis. *Nat. Methods* **6**, 737–740 (2009).
19. Flannick, J. et al. Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls. *Nature* **570**, 71–76 (2019).
20. Tamaki, M. et al. The diabetes-susceptible gene SLC30A8/ZnT8 regulates hepatic insulin clearance. *J. Clin. Invest.* **123**, 4513–4524 (2013).
21. Hardy, A. B. et al. Effects of high-fat diet feeding on Znt8-null mice: differences between β-cell and global knockout of Znt8. *Am. J. Physiol. Endocrinol. Metab.* **302**, E1084–E1096 (2012).
22. Li, L., Bai, S. & Sheline, C. T. *hZnT8* (Slc30a8) transgenic mice that overexpress the R325W polymorph have reduced islet Zn^{2+} and proinsulin levels, increased glucose tolerance after a high-fat diet, and altered levels of pancreatic zinc binding proteins. *Diabetes* **66**, 551–559 (2017).
23. Merriman, C., Huang, Q., Rutter, G. A. & Fu, D. Lipid-tuned zinc transport activity of human ZnT8 protein correlates with risk for type-2 diabetes. *J. Biol. Chem.* **291**, 26950–26957 (2016).
24. Gerber, P. A. et al. Hypoxia lowers SLC30A8/ZnT8 expression and free cytosolic Zn^{2+} in pancreatic beta cells. *Diabetologia* **57**, 1635–1644 (2014).
25. Wong, W. P. et al. Exploring the association between demographics, SLC30A8 genotype, and human islet content of zinc, cadmium, copper, iron, manganese and nickel. *Sci. Rep.* **7**, 473 (2017).
26. Vergnano, A. M. et al. Zinc dynamics and action at excitatory synapses. *Neuron* **82**, 1101–1114 (2014).
27. Ferrer, R., Soria, B., Dawson, C. M., Atwater, I. & Rojas, E. Effects of Zn^{2+} on glucose-induced electrical activity and insulin release from mouse pancreatic islets. *Am. J. Physiol.* **246**, C520–C527 (1984).
28. Bloc, A., Cens, T., Cruz, H. & Dunant, Y. Zinc-induced changes in ionic currents of clonal rat pancreatic β-cells: activation of ATP-sensitive K^+ channels. *J. Physiol* **529**, 723–734 (2000).
29. Zhang, Q. et al. Role of KATP channels in glucose-regulated glucagon secretion and impaired counterregulation in type 2 diabetes. *Cell Metab.* **18**, 871–882 (2013).
30. Michael, D. J. et al. Pancreatic beta-cells secrete insulin in fast- and slow-release forms. *Diabetes* **55**, 600–607 (2006).
31. Kirchhoff, K. et al. Polymorphisms in the TCF7L2, CDKAL1 and SLC30A8 genes are associated with impaired proinsulin conversion. *Diabetologia* **51**, 597–601 (2008).
32. Majithia, A. R. et al. Association of the SLC30A8 missense polymorphism R325W with proinsulin levels at baseline and after lifestyle, metformin or troglitazone intervention in the diabetes prevention program. *Diabetologia* **54**, 2570–2574 (2011).
33. Jainandunsing, S. et al. A stable isotope method for in vivo assessment of human insulin synthesis and secretion. *Acta. Diabetol.* **53**, 935–944 (2016).
34. Ivanova, A. et al. Age-dependent labeling and imaging of insulin secretory granules. *Diabetes* **62**, 3687–3696 (2013).
35. Abecasis, G. R., Cardon, L. R. & Cookson, W. O. A general test of association for quantitative traits in nuclear families. *Am. J. Hum. Genet.* **66**, 279–292 (2000).
36. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
37. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).

Methods

Human study population. The Botnia Study has been recruiting patients with T2D and their family members in the area of five primary health care centers in Western Finland since 1990. Individuals without diabetes at baseline (relatives or spouses of patients with T2D) have been invited for follow-up examinations every 3–5 years (ref. [11]). The Prevalence, Prediction and Prevention of diabetes (PPP)–Botnia Study is a population-based study in the same region, including a random sample of 5,208 individuals aged 18 to 75 years from the population registry[38]. Diabetes Registry Vaasa (DIREVA) is a regional diabetes registry of >5,000 diabetic patients from Western Finland (Botnia region)[39]. In the current study, we included >14,000 individuals (Botnia family study = 5,678, PPP = 4,862 and DIREVA = 3,835). All participants gave their written informed consent and the study protocol was approved by the Ethics Committee of Helsinki University Hospital, Finland (the Botnia studies) and the Ethics Committee of Turku University Hospital (DIREVA). Study participants and their clinical measurements in the Verona Newly Diagnosed Diabetes Study (VNDS; Extended Data Fig. 3c) are described in the Supplementary Note.

OGTT, test meal and IVGTT. Subjects maintained a weight-maintaining diet and avoided vigorous exercise for 3 d before the OGTT (Fig. 3) or test meal (Fig. 2 and Extended Data Fig. 2), which were performed after an overnight fast. Height, weight, hip and waist circumferences, fat percentage (%, bioimpedance analyzer) and blood pressure (sitting, three measurements after 5 min rest) were measured. The participants ingested 75 g dextrose (in a couple of minutes, OGTT) or a 526-kcal mixed meal (in 10 min, test meal: 76 g carbohydrate, 17 g protein and 15 g fat). Blood samples were drawn from an antecubital vein for plasma (P-) glucose and serum (S-) insulin and C-peptide at 0, 30, 120 min during the OGTT; for P-glucose, P-glucagon, S- insulin, S-C-peptide, S-zinc and total S-GLP-1 at 0, 20, 40, 70, 100, 130, 160 and 190 min during the test meal. Test-meal samples for S-FFA were collected at 0, 40 and 120 min and for S-proinsulin at 0, 20, 40 and 130 min, respectively. Urine was collected between 0 and 70 and between 70 and 190 min for the determination of glucose and zinc excretion during the test meal. Details about IVGTT (Extended Data Fig. 3) are provided in the Supplementary Note.

Biochemical measurements. See the Supplementary Note for details (Figs. 2 and 3 and Extended Data Fig. 3a,b). CIR was calculated for test meal (at 20 min) and OGTT (at 30 min) using the formula $CIR(t) = Ins(t) \times 100 / [Gluc(t) \times (Gluc(t) - 3.89)]$, where $Ins(t)$ and $Gluc(t)$ are insulin ($mU\,l^{-1}$) and glucose concentrations ($mmol\,l^{-1}$) at the sample time point (min)[40]. Estimation of the insulin clearance index was done on the model-based estimation of glucose-, insulin- and C-peptide curves during the test meal, using the equation $AUC_{ISR} / [AUC_{ins} + (I_{basal} - I_{final}) \times MRT_{ins}]$, where AUC_{ISR} is the area under the curve of insulin secretion rate, AUC_{ins} is the area under the curve of insulin concentration, I_{final} is insulin concentration at the end and I_{basal} is insulin concentration at the beginning of the study[41]. MRT_{ins} is the mean residence time of insulin, and was assumed to be 27 min, as reported previously[42].

Genetic association analysis. The genotype data were obtained as described in the Supplementary Note. All the quantitative traits were inversely normally transformed before the analyses. The family-based recall study included only nondiabetic subjects during the test meal. Association analysis of rare p.Arg138* with glycemic indices obtained during the family-based recall test meal study (Fig. 2 (left and middle panels), Extended Data Fig. 2 and Supplementary Table 6) was performed using family-based association analyses (orthogonal model together with 100,000 Monte Carlo permutations) adjusting for age, sex, body mass index (BMI) and, additionally, other covariates (genotype of p.Trp325Arg for middle panel only) as implemented in QTDT (v.2.6.1, https://csg.sph.umich.edu/abecasis/QTDT/)[35]. Association analysis of common p.Trp325Arg with glycemic indices obtained during the family-based recall test meal study (Fig. 2 (right panels), Extended Data Fig. 2 and Supplementary Table 6) was performed using QFAM family-based association test with 100,000 permutations to correct for any family structure as implemented in PLINK[36] (--qfam-total, --mperm 100,000, http://zzz.bwh.harvard.edu/plink/contact.shtml#cite). The association analysis of p.Arg138* and p.Trp325Arg with glycemic traits during OGTT studies including only nondiabetic individuals (Fig. 2 and Supplementary Tables 4 and 5) and IVGTT (Extended Data Fig. 3) was performed using mixed linear model considering genetic relatedness among samples as implemented in GCTA (v.1.91, http://cnsgenomics.com/software/gcta)[37]. The fixed-effects meta-analysis of all OGTT studies (Supplementary Table 5) was performed using the METAL software package (http://www.sph.umich.edu/csg/abecasis/metal/)[43]. The linear mixed model (adjusting for genetic relatedness) was also used for T2D association analysis.

iPSC generation, differentiation and genome editing. *iPSC generation and maintenance.* The hiPSC SB Ad3.1 was previously generated and obtained through the IMI/EU-sponsored StemBANCC consortium via the Human Biomaterials Resource Centre, University of Birmingham (http://www.birmingham.ac.uk/facilities/hbrc) (Fig. 4 and Extended Data Figs. 4 and 5). Human skin fibroblasts were obtained from a commercial source (Lonza CC-2511, tissue acquisition number 23447). They had been collected from a Caucasian donor with no reported diabetes with fully informed consent and with ethical approval from the National Research Ethics Service South Central Hampshire research ethics committee (REC 13/SC/0179). The fibroblasts were reprogrammed to pluripotency as previously described[44] and were subjected to the following quality control checks: SNP-array testing via Human CytoSNP-12 v.2.1 beadchip (Illumina, catalog no. WG-320-2101), DAPI-stained metaphase counting and mFISH, flow cytometry for pluripotency markers (BD Biosciences, catalog nos. 560589 and 560126) and mycoplasma testing (Lonza, catalog no. LT07-118).

CRISPR–Cas9-mediated generation of p.Arg138 and p.Lys34Serfs*50 hiPSC line.* Several guide RNAs (gRNAs) were designed using the MIT CRISPR tool (http://crispr.mit.edu/) to target near exon 2 and exon 3 of *SLC30A8* (ENST00000456015). The gRNAs were also subjected to an additional BlastN search (www.ensembl.org) to confirm specificity and no additional off-target sites were identified.
To generate *SLC30A8* p.Arg138*, the target site for CRISPR–Cas9 mutagenesis (Supplementary Table 8) was subcloned into the *Bsb*I restriction sites within the gRNA structure in the pX330 (ref. [45]) plasmid that was previously modified to contain a puromycin selection cassette. A single-strand oligonucleotide repair template for homology-directed repair (HDR) was synthesized by Eurogentec, stabilized by addition of a phosphorothioate linkage at the 5′ end, and contained two nucleotide changes: (1) the T2D-protective nonsense mutation at codon-138 (c.412C>T, p.Arg138*), which also mutated the protospacer adjacent motif sequence and (2) a silent mutation at codon-139 (c.417 A>T, p.Ala139Ala) to introduce an *Alu*I restriction site for genotyping. Human iPSCs were cotransfected with the *SLC30A8*-px330-puromycin resistant vectors and the HDR repair template using Fugene6 according to the manufacturer's guidelines (Promega, catalog no. E2691). Following transient puromycin selection, single clones were picked and expanded as described previously[46]. Genotyping PCR was performed using primers (primer pair 1, Supplementary Table 8) to amplify the targeted region followed by *Alu*I restriction digest. Successfully targeted clones were confirmed via Sanger sequence and monoallelic sequencing was performed by TA-cloning (pGEM-T Easy Vector System; Promega, catalog no. A1360) of the PCR amplicons. From 96 clones, 11 clones were heterozygous for p.Arg138*, four of which contained indels in the nontargeted allele. The CRISPR-Sham hiPSC control line (p.Arg138Arg) was generated from hiPSC cells that went through the CRISPR pipeline without being edited at the *SLC30A8* locus. The two p.Arg138* clones (A3 and B1) and the unedited control line (p.Arg138Arg) passed quality control checks that included repeat chromosome counting and pluripotency testing. Both p.Arg138* clones were heterozygous for the c.412C>T, p.Arg138*, while the silent variant at codon-139 (c.417A>T, p.Ala139Ala) was present in both alleles. Description of p.Lys34Serfs*50 hiPSC line generation is provided in the Supplementary Note.

In vitro differentiation of hiPSCs towards β-like cells. Directed differentiation of hiPSCs towards β-like cells was performed using a previously published protocol[12,47]. hiPSCs were seeded on Growth Factor Reduced Matrigel-coated CellBind 12-well tissue culture plates (Corning, catalog nos 356230 and 3336) at a cell density of 1.3×10^6 in mTesR1 (Stem Cell Technologies, catalog no. 05850) with 10 μM Y-27632 dihydrochloride (Abcam, catalog no. ab120129). The following morning, hiPSCs were fed mTesR1 media for >4 h before starting the seven-stage differentiation protocol described in the Supplementary Note.

Quantification of SLC30A8 gene expression in β-like cells derived from CRISPR-edited hiPSCs. Expression of *SLC30A8* was measured at the end of stage 7 using qPCR. Briefly, RNA was extracted using TRIzol Reagent (Life Technologies, catalog no. 15596026) according to the manufacturer's instructions. Complementary DNA was amplified using the GoScript Reverse Transcription Kit (Promega, catalog no. A5000). qPCR was performed using 40 ng of cDNA, TaqMan Gene Expression Master Mix (Applied Biosystems, catalog no. 4369017) and primer/probes for *SLC30A8* (Hs00545182_m1) or the housekeeping gene *TBP* (Hs00427620_m1). Gene expression was determined using the ΔΔCT method by first normalizing to *TBP* and then to the control p.Arg138Arg ($n = 7$–13 wells from three differentiations).

Allele-specific SLC30A8 expression in β-like cells derived from CRISPR-edited hiPSCs. Stage 7 cells were treated with 100 μg ml^{-1} cycloheximide (Sigma, catalog no. C4859) or DMSO (Sigma, catalog no. D2650) for 4 h at 37 °C (ref. [48]) before RNA and cDNA synthesis as above. Allele-specific expression was measured using the QX10 Droplet Digital PCR System and C1000 Touch Thermal Cycler according to manufacturer's guidelines (Bio-Rad). Custom primers (primer pair 2, Supplementary Table 8) and probes for the detection of p.Arg138* variant were designed using Primer3Plus (Applied Biosystems): FAM probe (R138; CT, Supplementary Table 8)), VIC probe (X138; TT, Supplementary Table 8). The specificity of the probes was confirmed by droplet digital PCR using R138 or X138 templates (Extended Data Fig. 5). Results were analyzed using Quanta Soft software (Bio-Rad) and presented as a ratio of wild-type to HDR-edited allele expression ($n > 4$ wells from two differentiations).

Western blot of ZnT8. We used highly specific antibody for ZnT8 as developed by Merriman et al.[49] (Supplementary Note). The signal detected corresponded to a protein of 30–34 KDa consistent with the previous publication[49].

Gene expression measurements in β-like cells derived from CRISPR-edited hiPSCs. Allele-specific *SLC30A8* expression by targeted RNA sequencing (Fig. 4e and Extended Data Fig. 5c) and other transcript expression in hiPSC-derived β-like cells by RNAscope (Extended Data Fig. 4e–g) are detailed in the Supplementary Note.

EndoC-βH1 culture. The results obtained in EndoC-βH1 are from two distinct teams (Helsinki and Oxford) with different batches of EndoC-βH1 cultures (Fig. 5). Here, we report both methods and specify for each experiment the origin of the culture (Helsinki or Oxford). EndoC-βH1 cells were cultured in medium and grown on a matrix as described previously[50] and tested negative for mycoplasma.

SLC30A8 knockdown in EndoC-βH1 cells. In Oxford, EndoC-βH1 cells were transfected with 10 nM siRNA (either SMARTpool ON-TARGETplus SLC30A8 or scramble (Dharmacon, catalog no. L-007529-01)) and Lipofectamine RNAiMAX (Life Technologies, catalog no. 13778-075) according to the manufacturer's instructions for a total of 72 h. In Helsinki, EndoC-βH1 cells were transfected using Lipofectamine RNAiMAX (Life Technologies). 20 nM siRNA ON-TARGETplus siRNA SMARTpool for human *SLC30A8* gene (Dharmacon, catalog no. L-007529-01) and ON-TARGETplus Non-targeting pool (siNT or Scramble) (Dharmacon, catalog no. D-001810-10-05) were used following the protocol described previously[51]. Cells were collected 96 h post-transfection for further studies.

Insulin secretion measurements in EndoC-βH1 cells. In Oxford, cells were subjected to static insulin secretion assays 72 h after siRNA transfection as described previously[51], apart from the following modifications: cells were stimulated for 1 h with 1 mM glucose, 20 mM glucose, 1 mM glucose + 200 µM tolbutamide or 20 mM glucose + 500 µM diazoxide. Insulin levels were measured in both supernatants and cells using the Insulin (human) AlphaLISA Detection Kit and EnSpire Alpha Plate Reader (PerkinElmer, catalog nos AL204C and 2390-0000, respectively). Cell count per well was measured via CyQUANT Direct Cell Proliferation Assay (ThermoFisher, catalog no. C35011). Data are presented as insulin secretion normalized to percentage of insulin content from the control condition. RNA extraction, cDNA synthesis and quantitative PCR with reverse transcription (qRT–PCR) were performed as above (section on 'Quantification of *SLC30A8* gene expression in β-like cells derived from CRISPR-edited hiPSCs') to determine *SLC30A8* knockdown and expression of the K_{ATP} channel genes (*ABCC8* Hs01093752_m1 and *KCNJ11* Hs00265026_s1; ThermoFisher Scientific). In Helsinki, EndoC-βH1 cells were transfected with 20 nM siRNA and Scramble control[52]. Following 96 h of siRNA transfection, cells were incubated overnight in 1 mM glucose containing EndoC-βH1 culture medium. At 1 h before the glucose stimulation assay, the media was replaced by βKREBS (Univercell Biosolution) without glucose. Cells were stimulated with 16.7 mM glucose and 50 mM KCl (Sigma-Aldrich) in βKREBS for 30 min at 37°C in a CO_2 incubator. The cells were then washed and lysed with TETG (Tris pH 8, Triton X-100, glycerol, NaCl and EGTA) solution (Univercell Biosolution) for measurement of total insulin content. Secreted and intracellular insulin were measured using a commercial human insulin ELISA kit (Mercodia) as per manufacturer's instructions (Helsinki).

Electrophysiological measurements in EndoC-βH1 cells (Oxford). SLC30A8 was knocked down in EndoC-βH1 as above. K^+_{ATP} channel conductance was measured in a perforated patch whole-cell configuration, and patch-clamped using an EPC 10 amplifier and HEKA pulse software. KREBS extracellular solution was perfused in at 32 °C and contained: 138 mM NaCl, 3.6 mM KCl, 0.5 mM $MgSO_4$, 10 mM HEPES, 0.5 mM NaH_2PO_4, 5 mM $NaHCO_3$, 1.5 mM $CaCl_2$, 1 mM glucose and 100 µM diazoxide (Sigma-Aldrich, catalog no. D9035). The perforation of the membrane was achieved using an intrapipette solution containing: 0.24 mg ml^{-1} amphotericin B, 128 mM K-gluconate (Sigma, catalog nos Y0000005 and G4500, respectively), 10 mM KCl, 10 mM NaCl, 1 mM $MgCl_2$, 10 mM HEPES, pH 7.35 (KOH). Conductance data are normalized to cell size and presented as pS pF^{-1}. Expression of *ABCC8, KCNJ11, B2M* and *TBP* were measured via qPCR as above (section on 'Quantification of *SLC30A8* gene expression in β-like cells derived from CRISPR-edited hiPSCs').

Other assays in EndoC-βH1 cells. Other assays, including insulin and proinsulin secretion and content, immunoblotting (ZnT8, CPE, PC2, phospho-AKT-Ser473 and AKT), cell viability assay (MTT), Zinpyr-1-based zinc staining, monitoring of stimulated zinc secretion using ZIMIR in EndoC-βH1 cells and RNA (mRNAs) sequencing of EndoC-βH1 cells, are described in the Supplementary Note.

Data analyses. Data are reported as mean (s.e.m.). Statistical analyses were performed using Prism 6.0 (GraphPad Software). All parameters were analyzed using Mann–Whitney *U*-test or one-sample *t*-test as indicated (two-tailed; Fig. 5). See the Supplementary Note for a description of RNA sequencing analysis.

Mouse model. *Animals.* All procedures were conducted in compliance with protocols approved by the Regeneron Pharmaceuticals Institutional Animal Care and Use Committee. The *Slc30a8*$^{p.Arg138*}$ mouse line (Fig. 6) is made in pure C57BL/6 background by changing nucleotide 409 from T into C in exon 3, which changes the arginine into a stop codon[10]. The mutated allele has a self-deleting neomycin selection cassette flanked by *loxP* sites inserted at intron 3, deleting 29 base pairs (bp) of endogenous intron 3 sequence. Mice were housed (up to five mice per cage) in a controlled environment (12-h light/dark cycle, 22 °C, 60–70% humidity) and fed ad libitum with either chow (Purina Laboratory 23 Rodent Diet 5001, LabDiet) or HFD (Research Diets, D12492; 60% fat by calories) starting at age 20 weeks. All data shown are compared to their respective wild-type littermates.

Glucose tolerance test. Mice were fasted overnight (16 h) followed by oral gavage of glucose (Sigma) at 2 g kg^{-1} body weight. Blood samples were obtained from the tail vein at the indicated times and glucose levels were measured using the AlphaTrak2 glucometer (Abbott). Submandibular bleeds for insulin were done at 0, 15 and 30 min postinjection in a separate experiment so as not to interfere with glucose levels.

Hormone measurements. Submandibular bleeds of overnight fasted (16 h) animals were done in the morning. Plasma insulin or proinsulin were analyzed with the mouse insulin/proinsulin ELISA (Mercodia), and C-peptide was analyzed with the mouse C-peptide ELISA (ALPCO). All ELISAs were performed according to the manufacturer's instructions.

Data analyses for mouse studies. Data are reported as mean (s.e.m.). Statistical analyses were performed using Prism 8.0 (GraphPad Software). All parameters were analyzed by two-way analysis of variance (ANOVA) (repeated measurements) combined with Sidak's multiple comparison test or Mann–Whitney *U*-test (two-tailed), as indicated (Fig. 6).

Expression of p.Arg138* alteration in INS-1E. Details are provided in the Supplementary Note (Extended Data Fig. 8).

Human pancreatic islets. Experiments on primary human pancreatic islets (Fig. 7) were independently performed in two places: Oxford and Lund University Diabetes Center (LUDC).

Human pancreatic islets from Oxford. Human pancreatic islets were isolated from deceased donors under ethical approval obtained from the human research ethics committees in Oxford (REC: 09/H0605/2, NRES committee South Central-Oxford B). All donors had given informed research consent as part of the national organ donation program. Islets were obtained from the Diabetes Research & Wellness Foundation Human Islet Isolation Facility, OCDEM, University of Oxford. All methods and protocols using human pancreatic islets were performed in accordance with the relevant guidelines and regulations in the UK (Human Tissue Authority). For in vitro insulin secretion, islets were pre-incubated in Krebs–Ringer buffer (KRB) containing 2 mg ml^{-1} BSA and 1 mM glucose for 1 h at 37 °C, followed by 1-h stimulation in KRB supplemented with 6 mM glucose. Insulin content of the supernatant was determined by radioimmunoassay (Millipore UK) as described previously[53].

Human pancreatic islets from LUDC. Human pancreatic islets were obtained from the Human Tissue Laboratory (Lund University, www.exodiab.se/home) in collaboration with The Nordic Network for Clinical Islet Transplantation Program (www.nordicislets.org)[54,55]. All the islet donors provided their consent for donation of organs for medical research and the procedures were approved by the ethics committee at Lund University (permit no.2011263). Islet preparation for cadaver donors, their purity check and counting procedure have been described previously[56]. Static in vitro insulin secretion assay from 91 islets (nondiabetic individuals) was performed as described previously[56,57]. Briefly, six batches of 12 islets per donor were incubated for 1 h at 37 °C in KRB buffer in the presence of 1 mM or 16.7 mM glucose. Independently, KCl-based insulin secretion assay was performed by incubating in 70 mM KCl together with 1 mM or 16.7 mM glucose in a subset of islets in different batches. Insulin concentrations in the extracts were measured using a radioimmunoassay kit (Euro-Diagnostica). The association of p.Trp325Arg genotype with expression of *SLC30A8* and other candidate genes was performed using RNA sequencing from islets of 139 nondiabetic individuals, as described previously[54,55] (see the Supplementary Note).

Statistics. Detailed information regarding statistical tests (two-tailed; Mann–Whitney *U*-test or one-sample *t*-test, genetic association; mixed model or linear regression or quantitative transmission disequilibrium tests) used for each substudy have been provided in their respective method section or with the figure legends. Sample size details can be found in the figure legends and Source Data. For electrophysiology, *n* are representative of the number of siRNA-transfected cells measured from several passages of the same EndoC-βH1 cell batch. For hormone secretion data, *n* are representative of independent biological replicates from independent passages of cells.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Individual level data for the human study can only be obtained via the Biobank of The Institute of Health and Welfare in Finland (https://thl.fi/en/web/thl-biobank).

Next-generation sequencing data have been deposited in the SRA database (PRJNA563975) and the processed counts data can be found in the Supplementary Dataset 1. The individual processed data from cell lines (Figs. 4 and 5), mice studies (Fig. 6) and human islet work (Fig. 7) are available in the Source Data files. Additional data supporting the findings of this study are available on request from the corresponding author. Source data for Figs. 2 and 4–7 and Extended Data Figs. 2, 3, 6 and 8 are available online.

References

38. Isomaa, B. et al. A family history of diabetes is associated with reduced physical fitness in the prevalence, prediction and prevention of diabetes (PPP)–Botnia study. *Diabetologia* **53**, 1709–1713 (2010).
39. Ahlqvist, E. et al. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *Lancet Diabetes Endocrinol.* **6**, 361–369 (2018).
40. Sluiter, W. J., Erkelens, D. W., Reitsma, W. D. & Doorenbos, H. Glucose tolerance and insulin release, a mathematical approach I. Assay of the beta-cell response after oral glucose loading. *Diabetes* **25**, 241–244 (1976).
41. Mohandas, C. et al. Ethnic differences in insulin secretory function between black African and white European men with early type 2 diabetes. *Diabetes Obes. Metab.* **20**, 1678–1687 (2018).
42. Navalesi, R., Pilo, A. & Ferrannini, E. Kinetic analysis of plasma insulin disappearance in nonketotic diabetic patients and in normal subjects. A tracer study with ^{125}I-insulin. *J. Clin. Invest.* **61**, 197–208 (1978).
43. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
44. van de Bunt, M. et al. Insights into islet development and biology through characterization of a human iPSC-derived endocrine pancreas model. *Islets* **8**, 83–95 (2016).
45. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
46. Krentz, N. A. J. et al. Phosphorylation of NEUROG3 links endocrine differentiation to the cell cycle in pancreatic progenitors. *Dev. Cell* **41**, 129–142.e6 (2017).
47. Perez-Alcantara, M. et al. Patterns of differential gene expression in a cellular model of human islet development, and relationship to type 2 diabetes predisposition. *Diabetologia* **61**, 1614–1622 (2018).
48. Harries, L. W., Hattersley, A. T. & Ellard, S. Messenger RNA transcripts of the hepatocyte nuclear factor-1α gene containing premature termination codons are subject to nonsense-mediated decay. *Diabetes* **53**, 500–504 (2004).
49. Merriman, C., Li, H., Li, H. & Fu, D. Highly specific monoclonal antibodies for allosteric inhibition and immunodetection of the human pancreatic zinc transporter ZnT8. *J. Biol. Chem.* **293**, 16206–16216 (2018).
50. Ravassard, P. et al. A genetically engineered human pancreatic β cell line exhibiting glucose-inducible insulin secretion. *J. Clin. Invest.* **121**, 3589–3597 (2011).
51. Thomsen, S. K. et al. Systematic functional characterization of candidate causal genes for type 2 diabetes risk variants. *Diabetes* **65**, 3805–3811 (2016).
52. Chandra, V. et al. RFX6 regulates insulin secretion by modulating Ca^{2+} homeostasis in human β cells. *Cell Rep.* **9**, 2206–2218 (2014).
53. Ramracheya, R. et al. Membrane potential-dependent inactivation of voltage-gated ion channels in α-cells inhibits glucagon secretion from human islets. *Diabetes* **59**, 2198–2208 (2010).
54. Ottosson-Laakso, E. et al. Glucose-induced changes in gene expression in human pancreatic islets: causes or consequences of chronic hyperglycemia. *Diabetes* **66**, 3013–3028 (2017).
55. Fadista, J. et al. Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism. *Proc. Natl Acad. Sci. USA* **111**, 13924–13929 (2014).
56. Rosengren, A. H. et al. Overexpression of alpha2A-adrenergic receptors contributes to type 2 diabetes. *Science* **327**, 217–220 (2010).
57. Taneera, J. et al. Identification of novel genes for glucose metabolism based upon expression pattern in human islets and effect on insulin secretion and glycemia. *Hum. Mol. Genet.* **24**, 1945–1955 (2015).
58. Bonetti, S. et al. Variants of GCKR affect both β-cell and kidney function in patients with newly diagnosed type 2 diabetes: the Verona newly diagnosed type 2 diabetes study 2. *Diabetes Care* **34**, 1205–1210 (2011).
59. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

Acknowledgements

We thank the Botnia Study Group for recruiting and studying the participants, J. J. Holst for measuring GLP-1 concentrations and L. Boselli for carrying out mathematical modeling of the OGTT studies. We thank D. Fu (Department of Physiology, The Johns Hopkins School of Medicine) for providing monoclonal anti-ZnT8 antibody. We thank W.-h. Li (Departments of Cell Biology and of Biochemistry, University of Texas Southwestern Medical Center) for providing zinc probe ZIMIR. We thank E. Na for her help with the mouse immunohistochemistry and histology, and C. Green and the Chromosome Dynamics & Genome Engineering Cores at the Wellcome Centre for Human Genetics for support with karotyping and genome editing (funded by the Welcome Trust grant no. 203141). We thank the Sequencing Unit core facility at FIMM Technology Centre supported by University of Helsinki and Biocenter Finland. The Botnia and the PPP-Botnia studies (L.G., T.T.) have been financially supported by grants from Folkhälsan Research Foundation, the Sigrid Juselius Foundation, the Academy of Finland (grant nos 263401, 267882, 312063, to L.G., 312072, to T.T. and 317599, to O.P.D.), Nordic Center of Excellence in Disease Genetics, EU (no. EXGENESIS, EUFP7-MOSAIC FP7-600914), Ollqvist Foundation, Swedish Cultural Foundation in Finland, Finnish Diabetes Research Foundation, Foundation for Life and Health in Finland, Signe and Ane Gyllenberg Foundation, Finnish Medical Society, Paavo Nurmi Foundation, Helsinki University Central Hospital Research Foundation, Perklén Foundation, Närpes Health Care Foundation and Ahokas Foundation, as well as by the Ministry of Education in Finland, Municipal Heath Care Center and Hospital in Jakobstad and Health Care Centers in Vasa, Närpes and Korsholm. The work described in this paper has been supported with funding from collaborative agreements with Pfizer Inc., as well as with Regeneron Genetics Center LLC. The work was also supported by Hjelt Foundation (L.G. and C.B.W.) and Rhapsody (L.G.). J.O.L. was supported by Vinnova (Sweden's Innovation Agency) (grant no. 2015-01549), Swedish Diabetes Foundation, Albert Påhlsson Foundation, Hjelt Foundations, Crafoord Foundation, Royal Physiographic Society in Lund, Swedish Foundation for Strategic Research (grant no. IRC15-0067), Swedish Research Council (grant no. 2009-1039, Strategic research area Exodiab). E.A. was supported by Crafoord Foundation, Påhlsson Foundation, Swedish Research Council (grant no. Dnr: 2017-02688). O.H. was supported by Diabetes Research Foundation. R.C.B. was supported by Italian Ministry of University and Research (grant no. PRIN 2015373Z39_004) and University of Parma Research Funds. G.R. was supported by a Wellcome Trust Senior Investigator Award (no. WT098424AIA), MRC Programme grants (nos. MR/R022259/1, MR/J0003042/1, MR/L020149/1) and Experimental Challenge Grant (no. DIVA, MR/L02036X/1), MRC (grant no. MR/N00275X/1), Diabetes UK (grant nos. BDA/11/0004210, BDA/15/0005275, BDA 16/0005485) and Imperial Confidence in Concept grants and a Royal Society Wolfson Research Merit Award. A.L.G. is a Wellcome Trust Senior Fellow in Basic Biomedical Science. M.I.M. and P.R. are Wellcome Senior Investigators. This work was funded in Oxford by the Wellcome Trust (grant nos. 095101 and 200837 to A.L.G., 098381 to M.I.M., 106130 to A.L.G. and M.I.M., 203141 to A.L.G., B.D. and M.I.M., 203141 to. M.I.M. and 090531 to P.R.), Medical Research Council (grant no. MR/L020149/1 to M.I.M., A.L.G. and P.R.), European Union Horizon 2020 Programme (T2D Systems) (A.L.G.) and NIH (grant nos. U01-DK105535 and U01-DK085545 to M.I.M. and A.L.G.). The research was funded by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (A.L.G., M.I.M., P.R.). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Author contributions

M.L., L.S., T.T. and L.G. conducted the human study. E.A., O.H., A.B., O.P.D. and J.F. analyzed the genotype data. M.L., O.P.D., M.T., E.B., R.C.B., T.T. and L.G. analyzed the human data. B.H., A.G., N.L.B., S.K.T., M.v.d.B., V.C., O.P.D., T.O. and A.L.G. characterized the human β cell model. N.A.J.K., F.A., N.L.B., B.C., D.M., P.K., B.D., O.P.D., A.S., M.I.M. and A.L.G. characterized the human iPSC-derived model. U.K., R.B.P., O.P.D., B.H., A.J.P., I.S., R.R., I.A., P.R., M.I.M. and A.L.G. characterized the human islets. S.K., D.G. and J.G. characterized the *Slc30a8* p.Arg138* mice. D.J., J.O.L., P.C., A.T., R.C., A.-M.R., J.B. and G.A.R. characterized the rat insulinoma cell line. M.I.M., A.L.G., T.T. and L.G. supervised the project. O.P.D., M.L., B.H., N.A.J.K., S.K., P.R., C.B.W., A.L.G., T.T. and L.G. wrote the manuscript. All authors revised the manuscript.

Competing interests

L.G. has received research funding from Pfizer Inc., Regeneron Pharmaceuticals, Eli Lilly and Astra Zeneca. N.L.B. and M.v.d.B are now employees of Novo Nordisk, although all experimental work was carried out under employment at the University of Oxford. A.L.G. has received honoraria from Novo Nordisk and Merck. M.I.M. serves on advisory panels for Pfizer, Novo Nordisk, Zoe Global; has received honoraria from Pfizer, Novo Nordisk and Eli Lilly; has stock options in Zoe Global; has received research funding from Abbvie, Astra Zeneca, Boehringer Ingelheim, Eli Lilly, Janssen, Merck, Novo Nordisk, Pfizer, Roche, Sanofi Aventis, Servier, Takeda. G.A.R. is a consultant for Sun Pharma and has received grant funding from Servier. J.O.L. has received research funding from Pfizer Inc. and Novo Nordisk A/S.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41588-019-0513-9.

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-019-0513-9.

Correspondence and requests for materials should be addressed to L.G.

Reprints and permissions information is available at www.nature.com/reprints.