Identifying Causal Mechanisms for Beta Cell Dysfunction at Type 2 Diabetes Risk Loci



Soren Krogsgaard Thomsen

St Cross College, and

Oxford Centre for Diabetes, Endocrinology & Metabolism,

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity Term 2017

For Olivia

Declaration

The work presented herein is the original work of the author, with contributions from others as outlined below. Unless otherwise stated, experiments were conducted at the Oxford Centre for Diabetes, Endocrinology and Metabolism (OCDEM, University of Oxford), under the supervision of Professors Anna Gloyn and Mark McCarthy. Funding was provided by the Radcliffe Department of Medicine, University of Oxford, and the Medical Research Council, UK.

The oral and intravenous glucose tolerance tests that generated the *in vivo* data analysed in Chapter 3 were performed by Dr. Aparna Pal (OCDEM) and Dr. Thomas Potjer (Leiden University Medical Centre, Leiden University, Netherlands). Immunofluorescence studies of p16 were performed together with Dr. Natasha Hui Jin Ng (OCDEM). The results presented in Chapter 3 have previously been published in *Diabetes*, and the resulting manuscript is provided in the Appendix.

The characterisation of variant PAM proteins discussed in Chapter 4 was performed by Dr. Anne Raimondo (OCDEM). Exocytosis data in EndoC-βH1 and primary human beta cells were produced by Dr. Benoit Hastoy (OCDEM) and by Dr. Jocelyn Manning Fox (University of Alberta), respectively. Quantitative immunoelectron microscopy was performed together with Dr. Anne Clark (OCDEM). Part of the results presented in Chapter 4 (*Sections 4.3.1-4.3.5*) have contributed to a BioRxiv preprint, which is provided in the Appendix.

The development and implementation of automated screening procedures, as described in Chapter 5, was carried out at the High Throughput Screening Facility (Target Discovery Institute, University of Oxford) together with Dr. Alessandro Ceroni under the supervision of Daniel Ebner. Isolations and transfection of primary human islets were performed by Dr Manning Fox. Part of the results presented in Chapter 5 (*Sections 5.3.2-5.3.5*) form a manuscript that has been published in *Diabetes,* and a copy is provided in the Appendix.

This thesis is approximately 40,000 words (excluding bibliography, tables, front matter, and appendices), and no part has been submitted for any other degree at this University or elsewhere.

Acknowledgements

When I started out my DPhil studies four years ago, I could not have predicted where this journey of scientific and personal discovery would take me. I am grateful to so many people who have provided support and encouragement along the way, both through times of excitement and through periods of stress. First and foremost, I want to express my most sincere gratitude to my supervisors, Anna Gloyn and Mark McCarthy, without whom none of this would have been possible. I am grateful for your enthusiastic and insightful scientific advice, the confidence you have placed in me to do independent research, and the opportunities you have made available for me to develop as a scientist. I could not have wished for better mentors to navigate me through my DPhil.

I would also like to thank everyone in the Gloyn and McCarthy groups for their invaluable support. Without the many members, past and present, these years would not have been as fun or productive as they were. Special thanks go to my fellow OCDEM students, Agata, Antje, Mahesh, Martijn, Matthias, Natasha, Neelam, Sara, and Shahana for your friendship and companionship, and to other team members, Amy, Anne, Benoit, Carla, Chris, Mandy, and Nicola, for all the helpful and patient advice. I would also like to thank others at the Old Road Campus, Clarissa, Costas, Jason, and Patrick, for keeping spirits high through endlessly fascinating discussions, and my co-founders at BioMe Oxford, Aga, Natasha, Matthias, Patrick, and Rasmus, for joining me on an exciting adventure in a parallel world, providing so many rewarding experiences outside of our studies.

Finally, I would like to thank my family - Far, Mor, Birgitta, Sebastian and Signe - for always being there for me. A special mention to my dad, who is my role model in life, for instilling in me a passion for science, and for raising me to always chase my dreams, wherever they might lead. If not for a memorable trip to Oxford together many years ago, the aspiration to one day come back might never have been ignited. Last, but not least, I would like to thank someone to whom I owe everything - my wife, Inna. If not for your unwavering support since before the start of our graduate studies, the last four years would not have been the same. Having our daughter, Olivia, has been the most exciting challenge of all, and one that continues to be an endless source of inspiration.

Thank you.

Identifying Causal Mechanisms for Beta Cell Dysfunction at Type 2 Diabetes Risk Loci

Soren Krogsgaard Thomsen, St Cross College, University of Oxford
Submitted in candidature for the degree of Doctor of Philosophy, Trinity Term 2017

Genome-wide association studies (GWAS) have identified more than 150 loci associated with type 2 diabetes (T2D) risk. Most association signals are located in noncoding regions of the genome, and uncertainty remains over the causal gene in the vast majority of cases. Motivated by the overall aim of accelerating the translation of genetic discoveries into molecular mechanisms for disease pathogenesis, the work presented in this thesis sought to identify candidate causal genes, and to study their roles in human beta cell function.

The causal mechanism for non-coding variants at the *CDKN2A/B* locus has remained elusive, though animal models have implicated *Cdkn2a* in beta cell function. To determine the effect of *CDKN2A* haploinsufficiency on glucose homeostasis in humans, I analysed data from oral and intravenous glucose tolerance tests in individuals carrying rare *CDKN2A* loss-of-function mutations. Compared with controls, carriers displayed increased insulin secretion, impaired insulin sensitivity, and reduced hepatic insulin clearance. Follow-up studies in the human beta cell line, EndoC-βH1, demonstrated cell-cycle independent effects of *CDKN2A* on insulin secretion.

At the *PAM/PPIP5K2* locus, two non-synonymous variants in *PAM* are associated with both T2D risk and insulinogenic index. To elucidate a causal mechanism, I investigated possible local and extra-pancreatic roles of PAM in beta cell function. I demonstrated that PAM deficiency results in reduced insulin content and altered dynamics of insulin secretion in EndoC-βH1 and primary beta cells. I also identified the amidated granular packaging protein Chromogranin A as an endogenous PAM substrate, establishing a direct link to granulogenesis in human beta cells.

Lastly, to enable systematic, large-scale prioritization of causal genes for T2D risk, I performed a high-throughput RNAi screen for beta cell dysfunction. Among 300 genes selected from 75 risk loci, I identified significant hits at half of these regions. The hits were highly enriched for known monogenic diabetes genes, but also revealed the poorly characterised transcription factor ZMIZ1 to be one of the strongest regulators of insulin secretion. Silencing of *ZMIZ1* in primary human islets subsequently revealed a core beta cell network to be negatively affected, highlighting a plausible causal mechanism.

Overall, this thesis has addressed a significant outstanding challenge in the translation of genetic signals into disease mechanisms. Using a range of different strategies, including genetic data, candidate gene biology, and functional genetic screening, the work has successfully identified several candidate causal mechanisms for beta cell dysfunction at T2D risk loci. These results provide a deeper understanding of the fundamental mechanisms underpinning disease predisposition, and thus highlight potential new therapeutic applications.

Abbreviations

2hGlu	2hr glucose
3C	chromatin conformation capture
AIR	Acute insulin response
AUC	Area under curve
BMI	Body mass index
cAMP	Cyclic adenosine monophosphate
CDK	Cyclin-dependent kinase
CDKI	Cyclin-dependent kinase inhibitor
ChIP	Chromatin-immunoprecipitation
CPM	Counts per million
CV	Coefficient of variation
eQTL	Expression quantitative trait locus
ER	Endoplasmic reticulum
FAMMM	Familial atypical multiple mole melanoma syndrome
FDR	False-discovery rate
FG	Fasting glucose
GPCR	G-protein coupled receptor
GWAS	Genome-wide association study
IBMX	3-isobutyl-1-methylxanthine
iPSC	Induced pluripotent stem cell
IVGTT	Intravenous glucose tolerance test
KO	Knock-out
LD	Linkage disequilibrium
LOF	Loss-of-function
MAF	Minor allele frequency
Mb	Megabase
MODY	Maturity onset diabetes of the young
NDM	Neonatal diabetes mellitus
OGTT	Oral glucose tolerance test
OR	Odds ratio
PheWAS	Phenome-wide association study
PNDM	Permanent neonatal diabetes mellitus
Ppv	Particles per vesicle
RBG	Recruit-by-genotype
RBP	Recruit-by-phenotype
RFU	Relative fluorescence units
RNAi	RNA interference
RPKM	Reads per kilobase per million mapped reads
siRNA	small interfering RNA
SNP	Single-nucleotide polymorphism
T2D	Type 2 diabetes
TGN	Trans-Golgi network
WES	Whole-exome sequencing
WGS	Whole-genome sequencing
WT	Wild type

Table of Contents

Chapter 1: General Introduction……...1

1.1 Pathophysiology of type 2 diabetes 2
1.1.1 Diabetes and global health 2
1.1.2 Key physiological mechanisms of glucose control 2
1.1.3 Aetiology of T2D 4
1.2 Genetics of diabetes 7
1.2.1 Insights from monogenic diabetes 7
1.2.2 Genetic architecture of T2D 8
1.2.3 Recent insights from rare variant associations 9
1.2.4 Monogenic vs polygenic diabetes: a new taxonomy for disease 11
1.3 The genomic context 15
1.3.1 Prerequisites for uncovering GWAS causal mechanisms 15
1.3.2 Causal variants 16
1.3.3 Annotating the genome 17
1.3.4 Causal context 18
1.3.5 Using annotations in post-GWAS follow-up studies 21
1.4 Identifying causal genes 24
1.4.1 Traditional approaches for causal gene prioritization 24
1.4.2 Coding variants 27
1.4.3 Variant-gene associations 28
1.4.4 Functional genetic screening 28
1.4.5 Integration and recent insights 29
1.4.6 Human genetics as model for target validation 30
1.5 The pancreatic beta cell 34
1.5.1 Mechanisms of insulin secretion 34
1.5.2 Cell cycle regulation in the beta cell 37
1.5.3 *In vitro* cellular models of human beta cells 38
1.6 Aims 41

Chapter 2: General Methods……42

2.1 Study approval 42
2.2 Maintenance of cells in culture 42
2.2.1 EndoC-βH1 42
2.2.2 Primary islets 43
2.3 Gene silencing 43
2.4 Gene expression analysis 44
2.4.1 RNA extraction 44
2.4.2 Reverse transcription 45

2.4.3 Real time TaqMan PCR 45
2.5 RNA sequencing 46
2.5.1 Sample QC 46
2.5.2 Illumina sequencing 46
2.6 Cellular phenotyping 47
2.6.1 Cell proliferation assay 47
2.6.2 Static insulin secretion assay 47
2.6.3 Insulin AlphaLISA assay 48
2.7 Immunofluorescence and light microscopy 49
2.8 Protein expression studies 50
2.8.1 Cell lysis and protein quantification 50
2.8.2 Gel electrophoresis and transfer 50
2.8.3 Western blotting 51

Chapter 3: Loss-of-function Mutations in *CDKN2A* Impact on Glucose Homeostasis in Humans 52

3.1 Introduction 53
3.1.1 Disease genetics of the *CDKN2A/B* locus 53
3.1.2 Biological functions of the proteins encoded by *CDKN2A/B* 55
3.1.3 *CDKN2A* mutations in humans 56
3.2 Materials and methods 58
3.2.1 Study participants 58
3.2.2 OGTTs 59
3.2.3 IVGTTs 61
3.2.4 Biochemical assays 61
3.2.5 Cellular assays in EndoC-βH1 62
3.2.6 Protein Kinase A activity assay 62
3.2.7 Statistical analysis 63
3.3 Results 64
3.3.1 Characteristics of the study cohort 64
3.3.2 *CDKN2A* mutation carriers display hyperinsulinemia in OGTTs 66
3.3.3 IVGTT reveals underlying defects in metabolic processes 73
3.3.4 Cell-cycle independent effects of *CDKN2A* deficiency in EndoC-βH1 77
3.4 Discussion 82

Chapter 4: Type 2 Diabetes Risk Alleles in *PAM* Cause Human Beta Cell Dysfunction 88

4.1 Introduction 89
4.1.1 Exome-based association studies identify coding variants for T2D risk 89
4.1.2 T2D risk variants in *PAM* 90

4.1.3 T2D-associated *PAM* alleles cause decreased PAM 91
4.2 Materials and methods 95
4.2.1 Gene expression analyses 95
4.2.2 Immunofluorescent labelling and microscopy 95
4.2.3 Gene silencing and insulin secretion in EndoC-βH1 96
4.2.4 Membrane capacitance assays in EndoC-βH1 and primary islets 96
4.2.5 Electron microscopy 97
4.2.6 Inhibitor studies and western blot analysis of CgA amidation 97
4.2.7 Analysis of exocytosis in primary human beta cells 98
4.2.8 Statistics 98
4.3 Results 99
4.3.1 PAM expression and subcellular localisation 99
4.3.2 PAM regulates glucose-responsive insulin secretion 105
4.3.3 PAM deficiency causes defects in the dynamics of exocytosis 109
4.3.4 Effects of PAM variants on primary beta cell function 111
4.3.5 Chromogranin A is a PAM substrate and possible downstream effector 113
4.3.6 Effects of PAM on extra-pancreatic hormones 116
4.4 Discussion 119

Chapter 5: Systematic Functional Characterisation of Candidate Causal Genes for Type 2 Diabetes Risk Variants…….. 122

5.1 Introduction 123
5.1.1 Unbiased strategies for prioritizing candidate causal genes 123
5.1.2 Functional genetic screens for regulators of beta cell function 125
5.2 Materials & Methods 127
5.2.1 RNA-seq of EndoC-βH1 127
5.2.2 RNA-seq of primary islets 127
5.2.3 Transfection optimization 128
5.2.4 Cellular phenotyping 128
5.2.5 Sample analysis 129
5.2.6 Statistical analysis 129
5.3 Results 130
5.3.1 Establishing a high-throughput screening pipeline 130
5.3.2 Validation of screening pipeline 132
5.3.3 RNA-seq of EndoC-βH1 and target selection 138
5.3.4 Primary screen 141
5.3.5 Follow-up validation 148
5.3.6 *ZMIZ1* regulates key beta cell pathways 151
5.4 Discussion 155

Chapter 6: General Discussion…….. 159

6.1 The candidate gene approach and recruit-by-genotype studies..... 159
6.2 Coding risk variants for causal gene prioritization 161
6.3 Functional genetic screening and data integration 163
6.4 Future directions..... 165

Appendices……... 168

Bibliography…….178

Published Manuscripts……197

CHAPTER 1

General Introduction

Some ideas and concepts presented in the following chapter are based on previously published review articles, with references stated below. Complete manuscripts are provided in the Appendices.

1. **S.K. Thomsen**, and A.L. Gloyn, The pancreatic beta cell: recent insights from human genetics. Trends in Endocrinology & Metabolism 25 (2014) 425-34.

2. **S.K. Thomsen**, M.I. McCarthy, and A.L. Gloyn, The Importance of Context: Uncovering Species- and Tissue-Specific Effects of Genetic Risk Variants for Type 2 Diabetes. Frontiers in Endocrinology 7 (2016).

3. **S.K. Thomsen**, and A.L. Gloyn, Human genetics as a model for target validation: finding new therapies for diabetes. Diabetologia 60 (2017) 960-970.

1.1 Pathophysiology of type 2 diabetes

1.1.1 Diabetes and global health

Over the past decades, diabetes mellitus has been increasing dramatically in prevalence and currently affects 415 million people worldwide (1). Auto-immune and gestational diabetes make up 5-10% of all cases, while monogenic and syndromic forms constitute ~1%, and the remainder is comprised of T2D. Close to half of all individuals with diabetes remain undiagnosed (47%), and a further 318 million people have impaired glucose tolerance, making them at high risk of developing the disease. Diet and lifestyle are key drivers behind the T2D pandemic, but heritable factors are known to contribute to a significant proportion of the phenotypic variance of T2D, with estimates ranging from 20-80% (2-5).

Diabetes and its related complications is widely recognised to be one of the leading health emergencies of this century (1, 6). It is estimated to be the cause of more than 5 million premature deaths annually, and 12% of global health expenditures were spent on diabetes in 2015 ($673 billion) (1). The most common complications of diabetes include retinopathy, cardiovascular disease, kidney failure, and nerve damage due to poor circulation. Importantly, such complications can be prevented or delayed by maintaining normal blood glucose levels, but current treatments fail to provide adequate control for most patients.

1.1.2 Key physiological mechanisms of glucose control

Blood glucose levels are tightly regulated by a multi-hormonal neuroendocrine network. It is a failure of this homeostatic system that underlies the WHO diagnostic definitions of diabetes mellitus (fasting glucose [FG] $\geq$ 7.0 mM; 2hr glucose [2hGlu] $\geq$ 11.1 mM after a glucose tolerance test) and impaired glucose tolerance (FG > 6.0 mM; 2hGlu $\geq$ 7.8 mM) (6). These physiological mechanisms have evolved to balance the availability of nutrients with the requirements for long-term energy storage and prevention of glucose toxicity.

In normoglycemic individuals, this balance is achieved by the opposing actions of two hormones, insulin and glucagon. Insulin, which is produced and secreted by beta cells in the pancreatic islets of Langerhans (*Section 1.5*), is released into the bloodstream in response to elevated glucose levels. Several other signals triggered by food intake can potentiate the secretion of insulin, including gut-derived glucagon like peptide-1 (GLP-1) and gastric inhibitory polypeptide (GIP), as well as high levels of certain amino acids (7, 8). Insulin has a large number of effects on the metabolism of carbohydrates, lipids, and proteins (9). The two major actions in terms of carbohydrate metabolism are: triggering glucose uptake into peripheral tissues (predominantly liver, muscle and adipose tissue); and stimulating the build-up of glycogen stores through glycogenesis. Insulin also stimulates the release of fatty acids and the uptake of amino acids, as well as anabolic processes, such as protein synthesis.

The effects of insulin are mediated through the insulin receptor, a transmembrane tyrosine kinase present at the cell surface of insulin-sensitive tissues (10). Binding of its substrate initiates insulin signalling through insulin receptor substrates 1 and 2 (IRS1/2) (11). Negative feedback at the level of IRS activity is one of the key mechanisms by which post-receptor insulin sensitivity is regulated (12). Further downstream, the insulin signalling cascade impacts on a large number of cellular processes through the phosphoinositide 3-kinase (PI3K) and protein kinase B (PKB/Akt) pathway. One canonical example is the translocation of the glucose transporter type 4 (GLUT4) to the cell surface, which leads to increased uptake of glucose in the fed state (13). The insulin receptor itself is internalised and can be recycled, providing another mechanism for regulating insulin sensitivity through total receptor levels available for insulin to bind (14).

The other major glucoregulatory hormone, glucagon, is secreted by alpha cells of the pancreatic islets and released in response to hypoglycaemia. It has opposite effects to those of insulin, resulting in the increased availability of glucose in the bloodstream. The two primary

actions of glucagon on carbohydrate metabolism are: promoting the breakdown of glycogen stores through hepatic glycogenolysis; and stimulating the production of glucose from gluconeogenesis in the liver and kidney (15). Glucagon also increases breakdown of lipids, and depresses fatty acid synthesis. The effects of the hormone are triggered by a G-protein coupled receptor (GPCR), the glucagon receptor (GCGR), which elevates cyclic adenosine monophosphate (cAMP) levels to bring about downstream signalling events (16).

Aside from the antagonistic actions of insulin and glucagon, metabolic homeostasis is further regulated by the central nervous system and a complex network of hormones, including leptin, the incretins (GLP-1 and GIP) and growth hormone (17-19). The hypothalamus, in particular, plays a key role in integrating nutrient sensing and hormonal signals to control eating behaviour and hepatic glucose production (20). All of these mechanisms work together in a fine-tuned system to maintain steady-state levels of glucose through both fasted and fed conditions.

1.1.3 Aetiology of T2D

Diabetes mellitus is the clinical manifestation of abnormalities in the mechanisms underlying glucose homeostasis. Non-autoimmune forms of the disease, such as T2D and monogenic diabetes, are characterised by an inability of islet beta cells to produce and secrete sufficient insulin. As a result, affected individuals display elevated glucose levels, with long-term complications resulting from prolonged hyperglycaemia. Monogenic forms of diabetes arise from defects in a single gene, and can give insights into the pathways underlying beta cell function (*Section 1.2.1*). In contrast, T2D is a complex disease with a multifactorial aetiology that remains relatively poorly understood (21).

The principal environmental risk factor for T2D is characterised by a Western lifestyle, used as a catch-all term for reduced energy expenditure, a high-fat diet, and excessive calorie

intake (21). Such a lifestyle promotes insulin resistance through the gradual accumulation of lipids in peripheral tissues. Notably, this includes skeletal muscle, which accounts for 75 % of insulin-stimulated glucose uptake, and is among the first tissues to display reduced insulin sensitivity on the path towards T2D (22, 23). Initially, emerging insulin resistance is compensated for by the beta cells, which respond with increased insulin secretion to ensure adequate glucose uptake (24) (**Figure 1A**). Over time, the developing insulin resistance may begin to exceed the compensatory capacity of beta cells, thus causing hyperglycaemia. Prospective studies have shown that the acute insulin response (AIR; a measure of beta cell function) is a better predictor of disease progression than insulin sensitivity, regardless of the baseline level of insulin resistance (25). In other words, the ability of beta cells to progressively compensate in the face of insulin resistance is a key determinant of whether or not an individual develops T2D (**Figure 1B**).

In terms of the genetic risk for T2D, GWAS have identified more than 150 loci that collectively explain about 10% of the estimated heritability (*Section 1.2.2*) (26-28). Physiological characterisations of risk allele carriers have revealed a large number of these associations to be driven through beta cell dysfunction (26, 27). Though physiological tests are limited in the resolution of the phenotypes that can be interrogated in large cohort studies, such findings are broadly consistent with the model for T2D pathogenesis discussed above. Recent genetic studies have thus placed the pancreatic beta cell centre stage in terms of understanding disease risk. Nevertheless, a growing body of evidence has also revealed defects across a range of other pathways, dismissing an overly simplistic view of disease predisposition (29-31). Important contributing factors include dysfunctions in the incretin axis, glucagon regulation, and insulin resistance. T2D is thus a heterogeneous disorder, and the relative contribution and timing of defects in causal pathways may vary across individuals (*Section 1.2.4*).

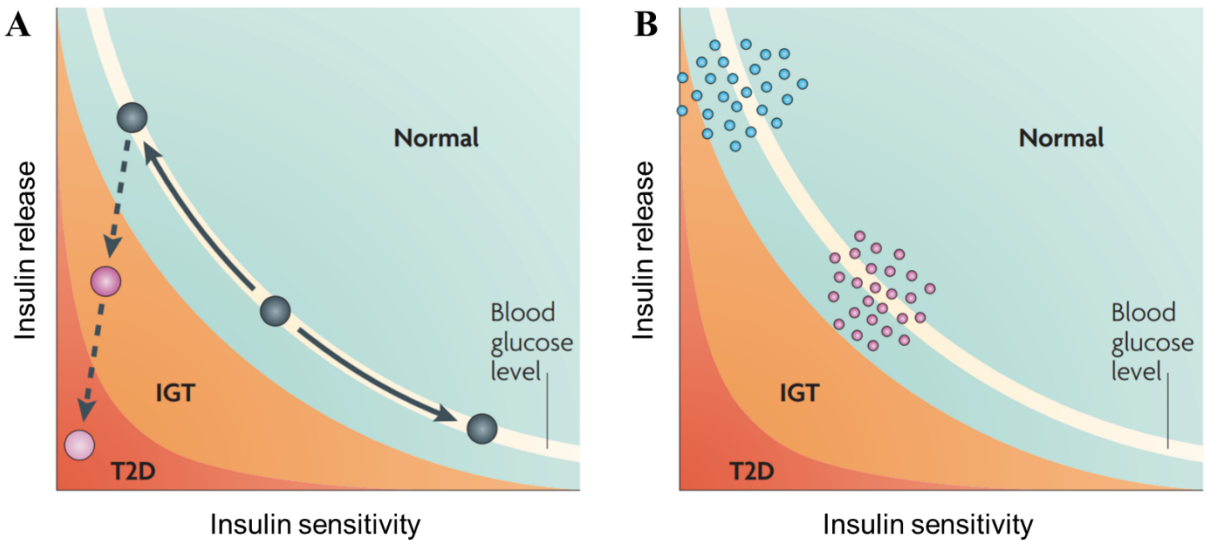


Figure 1 *A simplified model for the aetiology of impaired glucose tolerance and T2D.* The diagrams show overall glucose disposition (contour lines) as a hyperbolic function of insulin sensitivity and insulin release. (A) Perturbations of the normal glucose-tolerance set-point due to changes in either insulin sensitivity or insulin secretion are compensated for to maintain normoglycemia (*solid arrows*). During gradually developing insulin resistance, for example, beta cells are able to increase insulin secretion, resulting in the same level of glucose disposition. At some point, beta cells become unable to compensate further, and glucose tolerance deteriorates, eventually causing T2D (*dashed arrows*). (B) At a population-level, cultural changes have resulted in a shift in the average insulin sensitivity due to a Western lifestyle (*blue vs. pink dots*). *Taken from: Gibson G (2009) Decanalization and the origin of complex disease, Nature Reviews Genetics (32).*

1.2 Genetics of diabetes

Studies into the genetic risk for T2D are motivated by several aims. Most importantly, it provides a powerful tool to understand the fundamental mechanisms underlying disease pathogenesis in humans. These insights can inspire the development of new therapeutic strategies, or identify subsets of individuals that are more likely to benefit from specific treatments. Risk variants might also prove useful in predicting disease progression, even without a deep understanding the affected pathways.

1.2.1 Insights from monogenic diabetes

For neonatal diabetes mellitus (NDM) and maturity onset diabetes of the young (MODY), referred to collectively as monogenic diabetes, the pathogenic variants are rare and highly penetrant. Because of the clear genotype-phenotype correlations, monogenic forms of diabetes constitute an invaluable tool for understanding how the disruption of particular genes can impact on molecular and physiological pathways in humans. Studies into the genetics of monogenic diabetes have been remarkably successful in identifying genes implicated in beta cell function, including both components of the insulin secretory pathways (e.g. *ABCC8, GCK, INS, KCNJ11*) and developmentally important transcription factors (e.g. *HNF1A, HNF4A, PDX1, PTF1A*) (33). In some instances, the insights arising from these studies have enabled substantial therapeutic advances, most notably in the case of sulphonylurea treatment for patients with mutations in genes encoding subunits of the K_{ATP} channel (*ABCC8* and *KCNJ11*) (34).

Causal variants for monogenic diseases have traditionally been identified by linkage analysis and positional cloning. By mapping disease-causing variants relative to co-segregating genetic markers the approximate position in the genome could be established. Typically, this approach would identify a candidate gene region covering millions of base pairs. A number of

overlapping clones covering the region could then be generated to identify the causal gene. Increasingly, however, technological advances have made it possible to use whole exome or whole genome sequencing for variant discovery. Whole-exome sequencing (WES) offers the possibility of discovering coding, pathogenic mutations in an untargeted manner, while whole-genome sequencing (WGS) can be used to identify both coding and regulatory variants. In non-coding regions, however, predicted impact on protein function cannot be used as a predictor of pathogenicity. A challenge for future WGS-based studies is therefore related to the filtering of incidental variation from disease-causing mutations, particularly in cases where linkage information is limited.

1.2.2 Genetic architecture of T2D

In contrast to monogenic forms of diabetes, the genetic susceptibility to T2D is obscured by environmental influences and the relatively low penetrance of risk variants. As a consequence, genetic risk variants are identified through association studies by establishing statistical evidence for overrepresentation in cases compared with controls. The ratio between the frequency of an allele in cases compared with controls is referred to as the odds ratio (OR), a measure of the effect size of a variant. An OR significantly different from 1 indicates either a protective (OR < 1) or harmful effect (OR > 1) of the minor allele.

The majority of risk variants uncovered to date have been identified using microarray technology targeting common single-nucleotide polymorphisms (SNPs) with a minor allele frequency (MAF) > 5 % (35). Despite the relatively large number of loci uncovered in this way (>150), the combined population-attributable risk accounts for only a small fraction (~10%) of the total heritability of T2D (26-28). To account for this "missing heritability" a number of potential sources were initially proposed, including; gene-environment interactions; structural variants not captured by existing GWAS; rare variants of moderately high effect size; incorrect estimates of heritability; and inappropriate study design (28).

Rare variants, in particular, attracted considerable attention, and a number of recent GWAS have focused on discovery of low frequency (0.5 % > MAF > 5 %) and rare variants (MAF < 0.5 %) using WES/WGS or exome arrays (36-38). Findings emerging from these approaches have rejected the notion that rare variants contribute significantly to disease susceptibility. In one early study, Lohmuller and colleagues combined simulations with sequencing of a limited number of individuals, and concluded that their observations were inconsistent with a model wherein rare variants in a modest number of genes explain the majority of the missing heritability (39).

More definitively, a recent study by Fuchsberger et al combined the largest available sets of WGS/WES and exome-array data to test different simulations of T2D architecture (36). The model that best explained empirical observations was based on a polygenic disease architecture with limited selective pressure, shifting the distribution of risk variants towards common alleles. This is in line with previous suggestions that a long "tail" of common variants with progressively smaller effect sizes contributes to T2D risk (26). These variants may explain a majority of the residual heritability, but will require much larger cohorts to establish genome-wide significance. The emerging architecture of T2D genetics is thus consistent with the infinitesimal model also proposed for other complex traits, such as height (40).

1.2.3 Recent insights from rare variant associations

Despite the relatively minor contribution of lower frequency variants to T2D predisposition, efforts to discover rare alleles offer advantages in terms of biological interpretability. Even in the absence of a strong negative selection pressure, the penetrance of detectable disease-associated alleles tends to be inversely related to their population frequency. Thus, most common variants have relatively small effect sizes, while more deleterious mutations are rare (36, 41, 42). Rare variants with higher ORs are more tractable targets for experimental follow-up, and may be more likely to reside in coding regions because of the

potential for greater impact on protein function (*Section 1.4.2*) (35). A number of recent studies have successfully identified such rare and low-frequency risk variants, either by unbiased WES/WGS or through resequencing of loci already implicated by common variants (37, 38, 43, 44).

An initial study of this kind focused on *MTNR1B*, a gene previously associated with T2D and fasting glucose levels (45, 46). The gene encodes the melatonin receptor 1B (MT_2), an inhibitory GPCR that modulates insulin secretion. Since a non-coding risk allele in the region had been found to increase *MTNR1B* expression in primary islets (*Section 1.3.2*), it was expected that any coding T2D risk alleles would operate through increased receptor function (47). Resequencing of the region, however, identified a number of rare variants causing impaired melatonin signalling through MT_2 (43). The variants were too rare to allow for association tests at the level of individual variants but were aggregated to increase statistical power (35). Grouping based on functional effect established that alleles harbouring rare loss-of-function (LOF) variants significantly increase susceptibility to T2D (OR = 5.7). In contrast, a meta-allele consisting of all rare variants shows a diluted effect (OR = 3.3), demonstrating the importance of correctly assigning pathogenicity to increase power. The reason for the opposite directions of effect observed for coding and non-coding risk variants is unclear but may reflect differences between global and islet-specific roles of *MTNR1B* (*Section 1.3.4*).

Targeted resequencing has also been successful in shedding light on the direction of effect at the *SLC30A8* locus (44). *SLC30A8* encodes a Zn^{2+} transporter, ZnT-8, localising to the insulin-containing secretory granules. It was first proposed that a common, non-synonymous risk variant would decrease function of the transporter, leading to an insulin crystallization defect. Nevertheless, follow-up functional studies and animal models have only reported small, incongruent effects on beta cell function (48-52). To resolve this, Flannick et al. genotyped 150,000 individuals looking for T2D-association of rare variants with large functional effects

(44). By aggregating all protein-truncating mutations, they were able to show at an exome-wide level of significance that LOF variants are protective against T2D (OR = 0.34). Demonstrating the power of rare variants in establishing directionality, this surprising finding showed that a 50% reduction of ZnT-8 levels decreases the risk of developing T2D by almost three fold.

Other recent studies have attempted to uncover rare variants in novel regions using untargeted approaches (37, 38). Huyghe et al provided proof-of-principle using an exome-array approach to investigate a number of T2D-relevant traits in healthy adults (37). The study identified associations for the insulinogenic index, a measure of insulin release in response to glucose, and proinsulin levels, a proxy for beta cell stress. The three genes found to be implicated are all strong biological candidates; *PAM* (see Chapter 4) encodes an alpha amidase localising to the membrane of secretory granules, and both *KANK1* and *TBC1D30* encode proteins involved in G-protein signalling.

More recently, the deCODE consortium studied a cohort of Icelanders to test for associations of rare and low frequency variants with T2D (38). By combining WGS of a limited number of individuals with imputation based on Icelandic genealogy, they were able to obtain an effective sample size of close to 300,000 individuals. This provided sufficient power to establish three previously unreported T2D risk loci: *PAM,* in which the same rare allele had been associated with reduced insulinogenic index, *CCND2*, and *PDX1*.

1.2.4 Monogenic vs polygenic diabetes: a new taxonomy for disease

The discovery of *PDX1* as a T2D locus expanded the overlap between T2D risk genes and monogenic diabetes genes (including *KCNJ11, HNF1A, HNF1B, WFS1*). Interestingly, this observation was addressed more systematically by Fuchsberger et al, who established that a gene-set consisting of all known monogenic diabetes genes shows robust association with

T2D risk. The effect does not reflect accidental inclusion of individuals with monogenic diabetes, and is not driven exclusively by previously reported pathogenic alleles (36). In other words, less penetrant alleles predisposing to T2D are enriched in monogenic diabetes genes, suggesting partially overlapping aetiologies between the two groups of diseases. Blurring the boundary between monogenic and complex disease, it has recently been shown that the apparent deterministic effect of certain MODY mutations may reflect inflated estimates of effect sizes due to ascertainment bias (53). As more rare variants with large effect sizes are associated with T2D, the operational cut-off point for classification of a variant as an incompletely penetrant Mendelian mutation or a rare T2D allele with a large effect size will become increasingly arbitrary (54).

Because of an emerging appreciation for the heterogeneous nature of T2D, it is often implied that specific subtypes could be used to inform more precise diagnostic categories. If clinically meaningful subtypes did exist, such diagnostic labels could likely improve treatment efficacy. Certainly, there are individuals suffering from diabetes carrying mutations with high, if not complete penetrance, in specific genes. These individuals may either suffer from a monogenic form of disease or exist somewhere on the spectrum between complex T2D and a Mendelian disorder (53-55). Since disease progression in such individuals is determined by perturbations in a very limited number of pathways, genetic testing could in theory enable precision medicine.

Proof of concept has been provided by a life-changing treatment for individuals with permanent neonatal diabetes mellitus (PNDM). Genetic studies on PNDM has led to the realisation that a subset of individuals harbour pathogenic variants in the genes encoding the subunits of the K_{ATP} channel (56). Similar to the T2D risk haplotype at this locus, the mutations were found to promote opening of the channel, suggesting that sulphonylureas could provide a disease-modifying therapy. This was confirmed in follow-up studies that

demonstrated sustained efficacy in individuals with PNDM (57, 58). Remarkably, most participants were able to discontinue insulin treatment, switching to oral therapy with improved metabolic control. Thus, in the case of a disease with a well-defined genetic aetiology, there is compelling evidence that taking a pharmacogenetics approach can improve quality of life.

An interesting question pertains to whether such pharmacogenomic principles can be generalised to more complex forms of diabetes. Specifically, can genetic testing identify subgroups of individuals with T2D that are more likely to benefit from particular treatments than others? This would likely be the case if the underlying reality of diabetes was a collection of distinct subtypes, each dominated by defects in different pathways. It is clear that some individuals with complex diabetes do carry genetic variants with intermediate-to-high effect sizes that may be suggestive of increased sensitivity to drugs targeting the particular pathways affected. However, as discussed (*Section 1.2.2*), emerging evidence has shown that such individuals are in the minority and now favours a view of disease architecture dominated by a very large number of common variants, each with small effect sizes. Equally, non-genetic factors, though less well understood, appear to be characterised by pervasive environmental perturbations (21). Individual risk profiles are thus mostly determined by exposures that are common and widely shared, arguing against a model for diabetes based on a set of distinct pathologies.

Emerging from the notion that existing disease models may be poorly suited for our current understanding of diabetes, an alternative taxonomy has recently been proposed (**Figure 2**) (59). Referred to as the 'palette' model (as opposed to a subtype-oriented model), it posits that diabetes is caused by a large number of small perturbations (environmental and genetic) across the component pathways of disease (e.g. beta cell function, insulin sensitivity, autoimmunity). Individually, the phenotypic impact of each perturbation is limited, but in

aggregate will push a person on a path away from metabolic homeostasis. By analogy to colours combined in different hues and saturation, the palette taxonomy proposes an unlimited spectrum of disease manifestations. Monogenic and autoimmune forms of diabetes are represented in the extremes of this continuum (60). It is thus an implication of this model that the majority of individuals are not dominated by defects in a single or few processes (59). These individuals cannot meaningfully be categorised into subtypes and attempts at delivering precision medicine will be challenging. It may be that biomarkers for specific processes can be used to glean insights into the pathways that are driving disease progression at any given time (21, 59). In the near future, however, targeting people with high-impact mutations (those at the extremes of the diabetes spectrum) are likely to be a more tractable aim for precision medicine

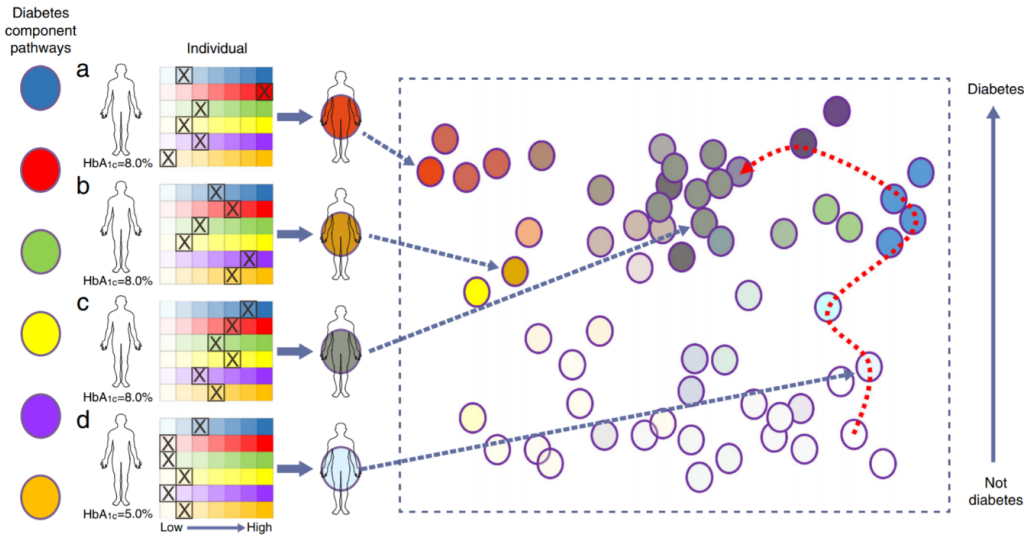


Figure 2 *"Palette model" for the taxonomy of T2D.* The phenotypic spectrum on the right is shown as a two-dimensional diagram, but the underlying reality is a multi-dimensional space made up of all the component pathways of disease. Individuals are coloured and positioned in the phenotype space according to defects in the causal T2D pathways. Four representative individuals (A-D) are highlighted on the left, of which three suffer from diabetes (A-C) and one does not (D). The pathophysiology of A is dominated by a defect in a single process (*red*) and is thus positioned in the extreme of the diabetes spectrum, whereas individuals B and C show lesser defects across several component pathways. Individual D has a mild predisposition to T2D through the blue pathway, and a possible future path toward T2D is illustrated by the red line. *Taken from: McCarthy MI (2017) Painting a new picture of personalised medicine for diabetes, Diabetologia (59).*

1.3 The genomic context

1.3.1 Prerequisites for uncovering GWAS causal mechanisms

Despite advances in identifying genetic variants that modulate diabetes susceptibility, the translation of association signals into molecular mechanisms has mostly been slow. The limited progress is due, in part, to the underlying properties of the association signals and, in part, to the design of GWAS used to identify risk variants. The vast majority of association signals are located in noncoding regions of the genome, presumably influencing nearby genes through regulation of transcriptional, translational or splicing activity (61). Though non-coding variants are named after the nearest protein-coding gene, proximity does not imply causality. Identifying the true causal gene(s) can be a major challenge due to the context-dependent nature of gene regulation.

Complicating things further, traditional microarray-based GWAS are designed to achieve genome-wide coverage of SNPs by sampling only a subset of variants. Exploiting the propensity of alleles in close proximity to be co-inherited (said to be in linkage disequilibrium [LD]) it is possible to achieve reasonable coverage of common variation by sampling just 0.5-1 million "tag" SNPs (62). Thus, the lead SNP (the variant with the lowest p-value of disease association) may simply be a surrogate for a causal variant that is directly genotyped.

Uncovering the causal mechanisms of disease risk loci are thus complicated by the need to identify both causal variant(s) and the disease-relevant context(s) in order to understand disease pathology and inspire potential new therapies. Equally important, narrowing down potential target tissues and causal variants is often a pre-requisite for identifying causal genes (*Section 1.4*).

1.3.2 Causal variants

A causal GWAS risk variant is defined as the genomic position at which the presence of alternative alleles directly modulates disease susceptibility. There may be multiple risk alleles on the same haplotype, in which case the association signal cannot be narrowed down to one specific causal variant. Identifying causal variants can point to regulatory networks and directly link association signals to nearby genes and phenotypes (*Section 1.4.3*). Several types of methods have been used to identify causal variants, including genetic fine-mapping, experimental testing, and *in silico* prediction.

Fine-mapping is the process of refining a disease association signal from tag SNPs to the likely causal variants. The analysis is based on deep sequencing or custom array genotyping to obtain better coverage of genetic variation in a particular region (63). Fine-mapping studies often include diverse ethnic populations to exploit differences in LD patterns across ancestries (64). The result of fine-mapping analyses is a narrowed down list of variants, referred to as a credible set, which has a defined posterior probability of containing the causal variant(s) (often 99%).

In the largest fine-mapping analysis for T2D risk to date, only a handful of the 36 loci included could be refined to credible sets containing fewer than ten variants based on fine-mapping data alone (65). In a few exceptional cases it has been possible to narrow down the credible set to a single causal variant, as is the case for the common risk variant at the *MTNR1B* locus (47). As previously alluded to (*Section 1.2.3*), the identification of a single causal variant was used to demonstrate a direct functional link to the causal gene based on follow-up experimental validation (47). Specifically, the risk allele was found to be associated with preferential binding of NeuroD1 in human pancreatic beta cells and increased enhancer activity towards *MTNR1B*.

1.3.3 Annotating the genome

An equally important prerequisite for establishing causal mechanisms is the identification of the tissue(s) in which the effects of the risk allele manifest. Recent studies have shown that the intersection of genetic data and genomic annotations can be used to understand the disease-relevant context. Broadly speaking, genomic annotation is the process of assigning structural and functional meaning, such as gene structure and chromatin features, to specific sequences in the genome. Cataloguing transcription factor binding sites and epigenetic modifications can shed light on the function of non-coding risk loci. Evidence for this has been provided by a large number of studies, which support a model whereby non-coding SNPs, at least partly, exert their cis-regulatory effects through disruption of transcription factor binding (66-68). The changes in occupancy in turn alter the chromatin landscape both locally and at the promoter region of nearby genes, thereby influencing transcriptional efficiency.

To functionally annotate genomic features, microarrays or sequencing is applied to samples of DNA that are enriched or pre-treated in specific ways. One of the most commonly used methods is chromatin-immunoprecipitation (ChIP) followed by microarray analysis (ChIP-chip) or sequencing (ChIP-seq). ChIP makes use of antibody-based purification to select genomic elements that are associated with either specific transcription factors or modified histones. Alternatively, isolation and identification of DNA fragments not associated with proteins can identify regions of open chromatin (e.g. through DNase 1 hypersensitivity). To map out DNA methylation patterns, approaches based on bisulphite conversion can be used to distinguish unmethylated from methylated cytosines.

A higher level of functional information is present in the relationship between two distinct genomic elements. An important example is the statistical correlation between a variant and the expression of a particular gene. Such variants, so-called expression quantitative

trait loci (eQTLs), can be found by combining transcriptomic and genomic datasets, often requiring large sample sizes to obtain adequate statistical power. Similarly, chromatin conformation capture (3C)-based methods allow identification of genomic regions that are in close spatial proximity due to folding and looping of chromatin.

1.3.4 Causal context

Aside from genetic sequence, transcription factor binding is influenced by the presence of co-factors and the local chromatin environment. As a result, many genomic annotations are dynamically regulated, varying according to developmental stage, metabolic state, or tissue type (69). To establish and maintain cellular identity, cell-type-specific transcription factors tend to bind in clusters, often referred to as cis-regulatory modules (70-72). Intriguingly, association signals for T2D have been found to show a significant overlap with islet-selective enhancer clusters, consistent with physiological studies establishing beta cell dysfunction as a central mechanism of disease-associated variants (27, 72-76).

The overlap between T2D signals and enhancer clusters suggests that the effect of risk variants could be subject to the same tissue specificity observed at the level of regulatory activity (74) (**Figure 3A**). In support of this notion, T2D risk variants are enriched for nearby binding sites of the pioneer transcription factor FOXA2 in islets and liver, but not in other tissues (47). T2D-association signals also show a significant overlap with islet cis-eQTLs, many of which are not identified in other cell types (77). Recent studies have provided more specific evidence to support the notion that context specificity can be a key aspect of causal mechanisms (47, 71, 73, 78).

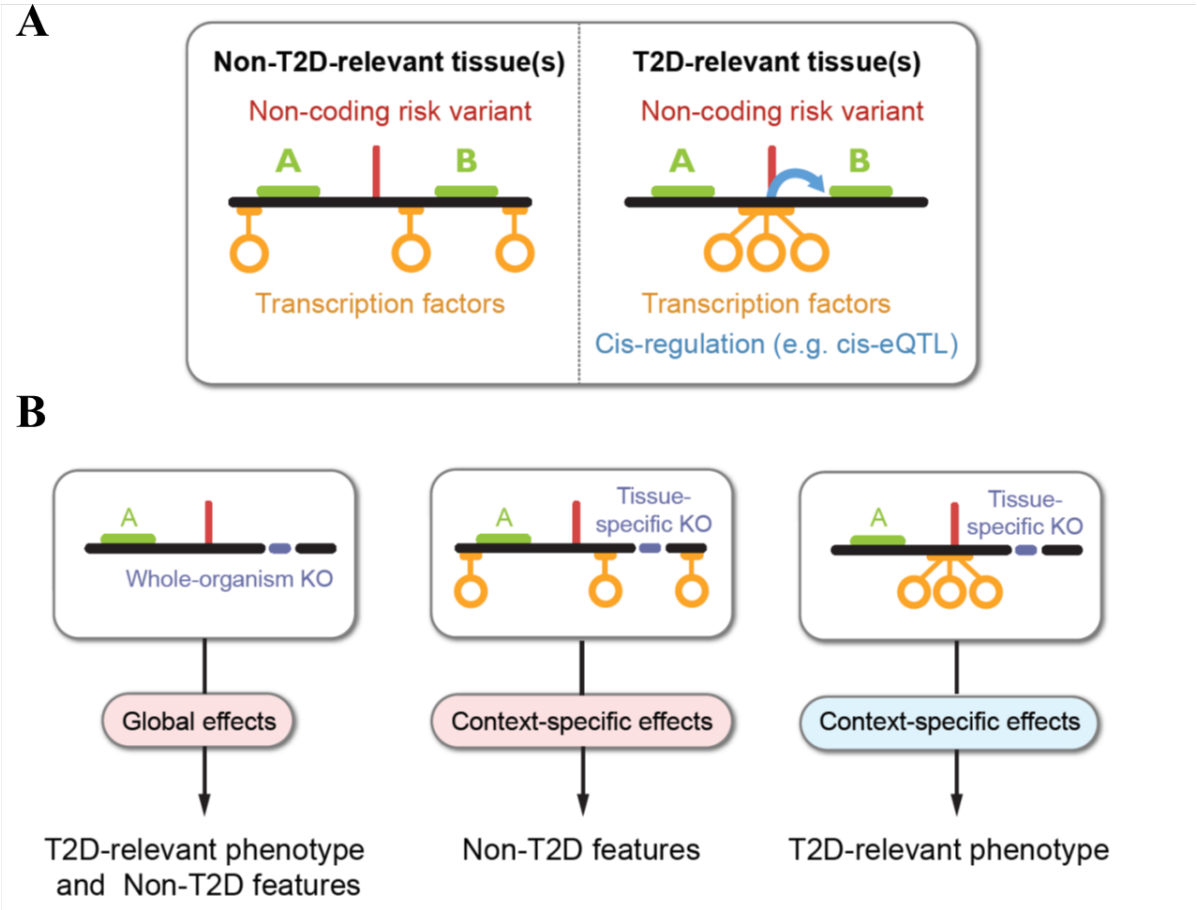


Figure 3 *Context specificity of causal mechanisms for T2D GWAS signals.* (A) Schematic representation of a GWAS locus with a non-coding association signal for T2D risk located near two genes, *A* and *B* (*green*). The left panel shows the regulatory landscape in a tissue type not relevant to understanding the causal mechanism for T2D; the region containing the disease risk variant does not exert any regulatory effects on nearby genes and is not bound by any transcription factors (*yellow*). In contrast, the right-hand side shows the regulatory landscape in a tissue type relevant to understanding the causal mechanism; the association signal overlaps a cluster enhancer defined by binding of multiple different transcription factors. This enhancer influences expression of gene B via a cis-regulatory mechanism that is affected by the risk allele (*blue*). (B) Overview of the observed phenotypes that emerge from studying the role of gene B using different approaches. In the case of manipulating the gene in a whole-body model (left panel), the resulting phenotype is a composite of effects from both disease-relevant and non-relevant tissues. (B-C) If a tissue-specific knockout model is generated, then the relevance of the phenotype depends on the confidence with which the affected tissue(s) can be established. The observed phenotypes may either be irrelevant to the effects contributing to disease risk (central panel) or capture exclusively the T2D-relevant effects without being masked by any confounding features (right panel). In all panels, gene knock-out (KO) is shown as an example of a typical post-GWAS experimental follow-up, but the same limitations are true for over-expression models. *Adapted from: S.K. Thomsen, M.I. McCarthy, and A.L. Gloyn (2016) The Importance of Context: Uncovering Species- and Tissue-Specific Effects of Genetic Risk Variants for Type 2 Diabetes; Frontiers in Endocrinology (79).*

In the case of *PTF1A,* for example, studying both the right tissue and development stage proved necessary to elucidating the underlying mechanism for noncoding mutations in the region. Previous work had identified a group of patients suffering from unexplained isolated pancreatic agenesis, which includes neonatal diabetes as a clinical feature (80). To filter causal mutations from incidental variation, a study by Weedon et al used pancreatic endoderm to define regulatory regions that are active during pancreatic development (78). Their strategy identified a distal enhancer that harbours mutations abolishing enhancer activity towards *PTF1A,* a gene encoding pancreas specific transcription factor 1a (PTF1A) (81). Coding LOF variants in *PTF1A* had previously implicated the gene in syndromic pancreatic agenesis, characterized by severe neurological features in affected individuals. The observation that the identified enhancer is not active in any other cell type tested provides a plausible explanation for the absence of any extra-pancreatic defects (78). Remarkably, even adult pancreatic tissue did not show active chromatin marks in the region, highlighting that studying the right developmental stage was critical.

Coding GWAS variants can also produce context-dependent mechanisms through the restricted expression of gene isoforms. A striking example of this is provided by a coding T2D-associated variant identified in the *TBC1D4* gene in a small founder population of Greenlandic Inuit (82). The risk allele, which produces a truncated transcript that results in nonsense-mediated decay, is positioned in an exon excluded from the widely expressed short isoform. The long form is predominantly expressed in skeletal muscle, and the effect of the risk allele is therefore selectively imposed in this tissue (83). As a result, the risk variant has a different effect on fasting glucose from that observed in individuals carrying LOF mutations affecting both isoforms (82).

The examples above demonstrate that the molecular phenotypes of GWAS signals can be modulated by a multitude of context-dependent factors. In the case of non-coding variants,

tissue- or developmentally restricted activity of the surrounding chromatin can limit effects on gene expression. For coding variants, alternative splicing or expression of homologues can mask a broader phenotype to produce context-dependent effects. These insights have important implications for how we design studies to translate genetic signals into molecular mechanisms.

1.3.5 Using annotations in post-GWAS follow-up studies

To successfully study disease-relevant phenotypes, experimental designs can be guided by the integration of genetic association data and genomic annotations (**Figure 3B**). At the core of this framework is the overlaying of a static dataset – a list of variants linked to disease susceptibility and/or physiological traits – with layers of highly dynamic functional information that provide spatial and temporal dimensions. These layers encompass diverse datasets, and include information centred on single variants, such as histone marks and TF binding sites, and higher-order information that signifies relationships between distinct elements, such as chromatin interactions and cis-eQTL data.

Genomic annotations provide a cellular atlas that can be used to navigate and interpret genetic data in the context of specific cell types and developmental stages. For instance, if a credible set of causal variants has been identified from fine-mapping studies, the tissue of action may be inferred from comparing chromatin states across cell types (71-74, 84-87). Conversely, if the target tissue has been inferred from physiological associations, the causal mechanism can be pinpointed by overlaying with relevant functional annotations in that tissue (*Section 1.4.3*) (47, 78, 88-92). This process generates a plethora of hypotheses that can be followed up with specific functional experiments.

One example is provided by an intronic T2D signal at the *TCF7L2* locus, which has been the focus of conflicting experimental observations. Tissue-specific knockout studies have demonstrated primary roles of *TCF7L2* in a number of different tissues, including liver and

islets, but the non-coding GWAS signal has been consistently associated with a relatively narrow insulin secretion defect (93-97). Genomic annotations provide a clue as to the underlying reason, with the risk variants being located in an islet-specific region of open chromatin (71, 73, 98). Further, the region has chromatin marks indicative of regulatory activity in islets, but not in a wide range of other tissues (73). The annotations can thus be applied as a filter to exclude non-disease-relevant tissues, and guide efforts to study the role of the causal gene in the most appropriate context.

The broader applicability of this approach is, in part, determined by the tractability of individual loci. For regions with extensive LD, the resulting complexity can hinder experimental follow-up. Starting from a limited set of credible variants is thus desirable, though experimental approaches, such as massively parallel reporter assays (MPRA), have emerged to address this gap (99). More generally, the value in taking this approach is dictated by the extent to which genomic annotations for disease-relevant tissues have been made available (or can be obtained). The construction of a truly integrated framework is an incremental and monumental effort, facilitated by the Encyclopaedia of DNA Elements (ENCODE) and the NIH Roadmap Epigenomics project, which together cover hundreds of tissues and epigenetic annotations (85, 100). However, for many hard-to-obtain organs, such as pancreatic islets, power to detect relevant genomic features is still limiting.

Even so, chromatin landscapes for tissues relevant to T2D have begun to emerge in recent years, enabling biological inferences to be made. In most of these studies, pancreatic islets have been used as a proxy for beta cells, which make up 65-80 % of the human islets but are hard to isolate as a pure fraction. Some studies have focused on identifying transcription factor binding sites or regions with specific chromatin states (71-73, 98, 101-103). Other studies have performed transcriptomic analyses to provide information on the expression of both protein-coding genes and non-coding RNAs (104-107). One of the key lessons emerging

from islet annotations has been an improved understanding of how islet-specific genes are regulated. Several studies have independently demonstrated that genes involved in beta cell identity and function tend to be regulated by clusters of transcription factors, rather than orphan enhancers (71-73). The clusters form stretch or super-enhancers that may act as molecular runways for tissue-specific transcription factors (70, 72, 108). Further, as previously discussed (*Section 1.3.4*), these islet cluster enhancers are enriched for variants associated with susceptibility to diabetes and related metabolic traits (72, 73).

1.4 Identifying causal genes

1.4.1 Traditional approaches for causal gene prioritization

A key aspect of translating GWAS signals into disease biology naturally centres on the identification of the causal genes (or "effector transcripts"), mediating the effects of risk alleles on disease pathogenesis (**Figure 4A**). Despite advances in broadly understanding molecular and regulatory mechanisms involved in T2D pathogenesis, progress on individual loci has been slow. A minority of the >150 independent association signals for disease risk have produced a single high-confidence candidate causal gene. As a result, the value in GWAS for understanding fundamental disease pathways and proposing new therapeutic targets is still limited by this experimental bottleneck. Especially since follow-up studies have tended to focus on the 'low-hanging fruits' supported by existing lines of evidence.

Particular candidate causal genes have traditionally been selected for follow-up studies based on a combination of known biology and proximity to an association signal. Such post-GWAS experimental work would typically involve engineering of global animal knock-out (KO) models or the use of genetic testing to identify individuals carrying LOF variants. As previously discussed (*Section 1.3.4*), these studies can shine light on the role of genes in whole-organism physiology, but may have limited relevance for understanding disease-relevant mechanisms. For animal models, context-dependent knockouts can provide improved spatial and temporal resolution for targeting candidate causal genes. Even so, the disease relevance of the observed phenotypes is determined by the confidence with which the target tissue of the risk allele is known. Human physiological associations can narrow down the list of likely relevant tissues, but these measures are often too crude to pinpoint specific cell types.

In principle, using animal models to target non-coding regions could produce more precise disease models, but this strategy is constrained by the low conservation at the level of

regulatory architecture. Transcription factor binding sites have been shown to diverge even faster than the underlying non-coding sequence itself (109). For two key liver transcription factors, the majority of binding events were shown to be species-specific, while only 10-30 % of hepatic enhancer clusters have corresponding rodent orthologues (110, 111). Although subsets of conserved clusters may aid in the prioritization of causal variants, these observations suggest that a different approach is required to delineate GWAS mechanisms (111, 112).

More recently, progress in identifying causal variants and disease-relevant tissues (*Section 1.3*), has inspired a number of different approaches that enable scalable and unbiased prioritization of candidate causal genes. These complementary methods broadly fall into three categories based on coding risk variants, variant-gene interactions, and functional genetic screening (**Figure 4B**).

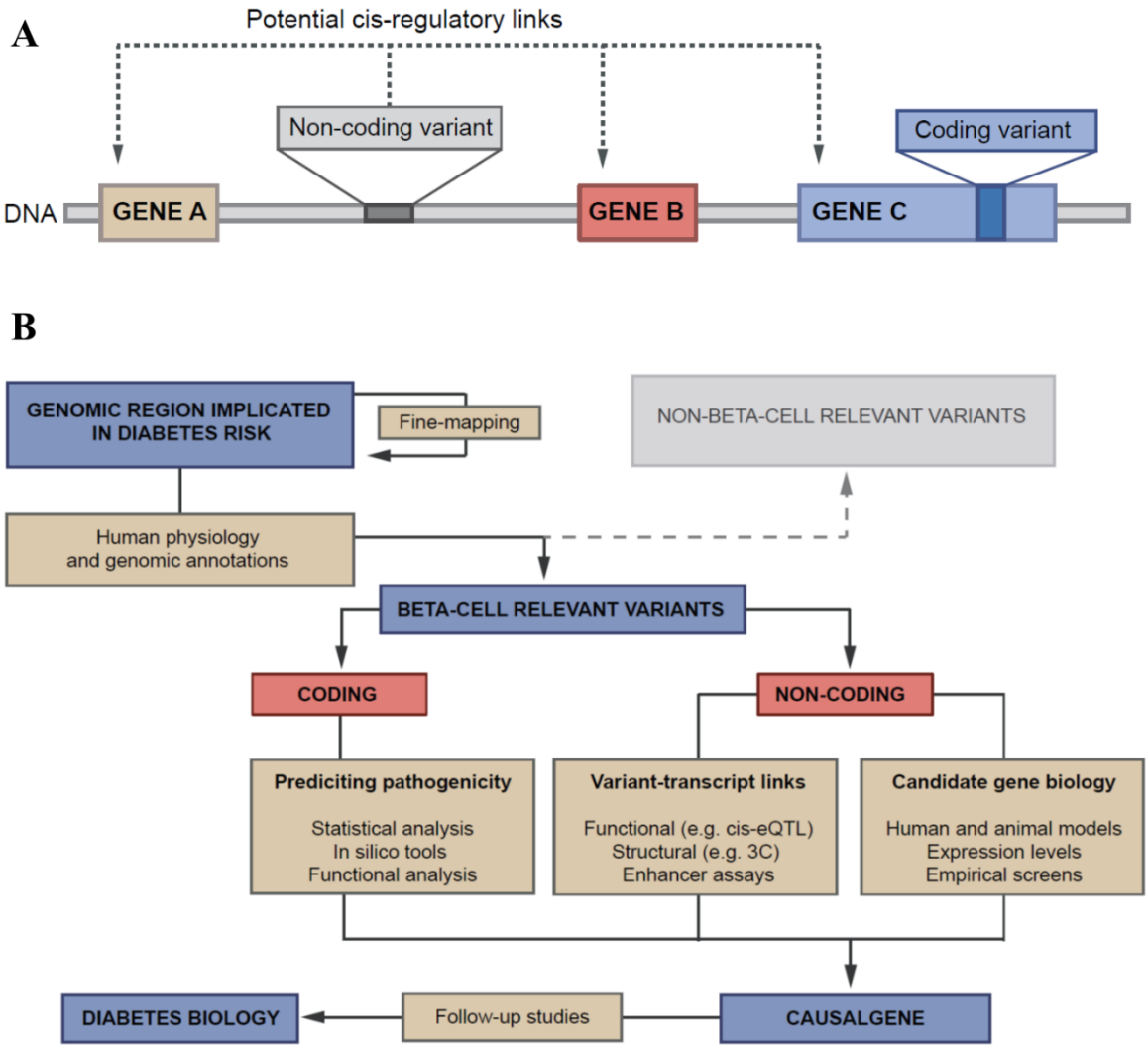


Figure 4 *Translating associations signals into causal genes and mechanisms.* Panel A depicts a hypothetical locus with non-coding and coding disease risk alleles. The coding variant is likely to exert its influence on disease pathogenesis through an effect on the function of gene C. A number of different mechanisms, including amino-acid substitution and protein truncation, could be driving the association. The non-coding variant, in contrast, could be acting through any of the three nearby genes, as shown by the dotted arrows indicating potential cis-regulatory links. Panel B outlines some of the post-GWAS approaches that have emerged for the translation of association signals into disease biology. Human genetics can identify genomic regions implicated in diabetes risk. Subsequently, fine-mapping approaches can refine the genetic associations and identify the causal variant (or credible set). Genomic annotations or physiological characterisations can then provide hints into the tissue(s) affected by the risk alleles, which, in turn, informs efforts to identify the underlying causal genes (see main text). Once a causal gene has been identified, a number of different approaches, including human models, animals models and functional studies, can be used to interrogate the function of the protein and ascertain its role in disease biology. *Adapted from: S.K. Thomsen, and A.L. Gloyn (2014) The pancreatic beta cell: recent insights from human genetics; Trends in Endocrinology & Metabolism.*

1.4.2 Coding variants

Coding association signals can be used to nominate candidate effector transcripts, especially in cases where a nonsynonymous variant in the credible set is predicted to impact on protein function. In contrast with regulatory signals, missense variants thus offer the possibility of a more straightforward translation into causal mechanisms. Even in regions with independent non-coding variants, coding alleles can be used to direct experimental efforts towards particular candidates. This is illustrated by the *G6PC2/ABCC11* locus, which harbours two strong candidate causal genes near a non-coding association signal identified for fasting glucose (113, 114). A more recent effort to map coding variation associated with glycaemic control found coding variants within the *G6PC2* gene (115, 116). Follow-up experimental studies have since explored the effect of the variants to successfully show a functional impact on the encoded glucose-6-phosphatase subunit (115).

Efforts to identify coding variants have been facilitated by a recent shift in the attention of GWAS efforts towards low frequency and rare variants with higher penetrance (*Section 1.2.2*) (36-38). In parallel, the power of such association studies has been further improved by leveraging the genetics of isolated populations. These populations often show less genetic complexity and higher allelic frequencies due to bottleneck events in recent history (117). As previously discussed, a study in Greenlandic Inuit identified a non-sense coding variant in *TBC1D4* with the largest effect size for any common T2D risk allele (OR = 10.3) (82). However, despite some remarkable successes, this approach is fundamentally limited by the occurrence of natural variation (both in outbred and isolated populations) which necessitates ever-larger association studies to detect rare, coding variation. Moreover, fine-mapping analyses of both coding and non-coding regions are required to establish the causality of missense variants, which in some cases may simply be surrogates for a causal regulatory risk allele (65).

1.4.3 Variant-gene associations

As alluded to previously, integration of genetic and genomic data can be used to establish functional regulatory links between non-coding variants and causal genes. In cases where the association signal coincides with a cis-eQTL in a disease-relevant tissue, for example, this can reveal both the target gene and the direction of effect for the risk variant. While many cis-eQTLs are shared across tissues, others appear to show more restricted effects that are specific to one or more tissues (118). As for other genomic annotations, interrogation of the appropriate disease-relevant context is thus essential to the success of this approach. Though power to detect islet cis-eQTLs has been fairly limited by the availability of human islets, the approach has successfully highlighted candidate effector transcripts with previously unknown roles in disease pathogenesis (77, 119); this is the case for the poorly characterised *ZMIZ1* gene that was identified in a recent study (77).

Similar to inference based on coding variants, the strategy is limited by the ability to refine credible sets of causal variants for both the disease association signal and cis-eQTLs. Moreover, even if a true causal risk variant is shown to exert cis-regulatory influences across different tissues, and/or on multiple different genes, some effects may represent pleiotropy that is not important for understanding T2D pathogenesis. Ultimately, hypotheses proposed based on gene-variant interactions should be complemented with functional studies to validate a role of a specific gene in a disease-relevant endophenotype.

1.4.4 Functional genetic screening

A third strategy aims to prioritise genes implicated in cellular processes that are believed to be important for disease pathogenesis, such as insulin secretion or peripheral insulin sensitivity. Accordingly, a gene identified as a regulator of beta cell function would have a high posterior probability of being causal for an association signal linked to insulin

secretion. Some T2D risk loci, for instance, harbour genes implicated in monogenic forms of diabetes (*Section 1.2.4*), which have a high likelihood of being causal for complex forms of diabetes as well. Though this type of evidence is indirect, it can be a powerful method for limiting the number of genes to be taken forward for experimental validation.

Still, the number of loci for which current evidence favours one candidate over others is limiting and tends to be biased towards previously studied genes. High-throughput functional genetic screening is emerging as a potentially powerful approach to address this gap by surveying a multitude of relevant phenotypes across tissue types and developmental stages. Such efforts should ideally allow unbiased functional screening across all disease-relevant contexts, but throughput and access to relevant model systems are still major limitations.

1.4.5 Integration and recent insights

Each of the strategies for prioritizing causal genes have a number of advantages and limitations. Coding variants are reliant upon the occurrence of natural variation, while gene-variant interactions are subject to context-dependency and possible pleiotropy, and functional screening is both experimentally demanding and circumstantial by design. To overcome these individual drawbacks, the different approaches can be integrated to converge on particular causal mechanisms. Such triangulation from complementary datasets has been demonstrated by a number of recent studies investigating T2D risk loci.

At the *ARAP1/STARD10* T2D risk locus, initial reports had implicated the *ARAP1* (*CENTD2*) as a likely causal gene (77, 120). However, subsequent fine-mapping and genomic annotations were used to refine a credible set of variants in an islet-active cluster enhancer (121). A range of functional analyses and cis-eQTL data revealed this non-coding signal to display regulatory activity towards the nearby gene *STARD10,* but not *ARAP1*. Interestingly, 3C-based data showed physical interactions with both the *ARAP1* and *STARD10* promoters,

but the potential to exert a regulatory effect appears only to be actualized for *STARD10* in beta cells. Thus, multiple lines of evidence were required to clarify the specific roles of *ARAP1* and *STARD10* in T2D predisposition.

Another example is provided by work on the *SLC16A11/13* locus, which has attracted attention because of the large fraction of excess T2D prevalence explained by a common risk haplotype in the Latino population (~20%) (122, 123). Fine-mapping identified a credible set of variants that include missense coding variants in the *SLC16A11* gene (123). Further, expression data and genomic annotations showed an additional regulatory effect of non-coding variants on the same risk haplotype. Functional work then confirmed a role for SLC16A11 in lipid and fatty acid metabolism. These results propose *SLC16A11* as a causal gene, through two parallel mechanisms, and suggest a possible therapeutic strategy.

Importantly, in the case of both these examples, the evidence presented is insufficient to exclude the possibility of additional contributing mechanisms (e.g. regulatory pleiotropy in cell types not studied). Exhaustively addressing this gap would require interrogation of variant function across all disease-relevant cell states, and is not a realistic aim for current post-GWAS studies.

1.4.6 Human genetics as model for target validation

Identifying causal mechanisms can provide insights into the underlying disease pathways, as well as inform therapeutic research and development efforts (**Figure 5**). Most obviously, if LOF variants in a particular gene are linked to protective effects, modulating the encoded protein by antagonists would be an attractive aim. Over the past ten years, genome-wide association studies (GWAS) have made significant progress towards mapping the genetic heritability of many complex diseases (124). In the context of drug development, however,

these advances are, broadly speaking, too recent to assess their impact on prospective target discovery.

Still, attempts can be made to validate the use of GWAS associations for target discovery based on existing drugs. One study looked at the proportion of drug mechanisms (defined as a protein target and its approved indication) supported by genetic evidence at the various stages of the drug development pipeline (125). This proportion was found to increase from 2% at the preclinical stage to 8.2% for approved drugs, with the single largest increase occurring between phase II and phase III of clinical trials. Remarkably, based on such historical data, it was estimated that the rate of success was around two-fold higher for a target–indication pair supported by GWAS or other human genetic data compared with pairs with no support.

For the treatment of T2D, thiazolidinediones (TZDs) provide an instructive example of a drug mechanism that has been corroborated by genetic evidence since first being discovered. TZDs are a class of commonly used drugs that act primarily through activation of the peroxisome proliferator-activated receptor γ (PPARγ) to improve insulin sensitivity (126). Within a few years of obtaining market approval in 1996, the gene encoding PPARγ (*PPARG*) was found to contain a missense variant (Pro12Ala) that associates with T2D susceptibility (127-129). Though the functional impact of this common variant remains uncertain, the subsequent discovery of rare, LOF variants associated with disease risk have established direction-of-effect at this locus (130, 131). Genetic evidence thus points to a therapeutic benefit of PPARγ agonists, fully consistent with the clinical effects observed from TZDs. The Pro12Ala association has also since been replicated by GWAS for T2D risk, despite the relatively small effect size of the risk allele (OR 1.16) (132). As illustrated by this case, the effect size of a genetic variant is not necessarily a useful predictor of therapeutic opportunities.

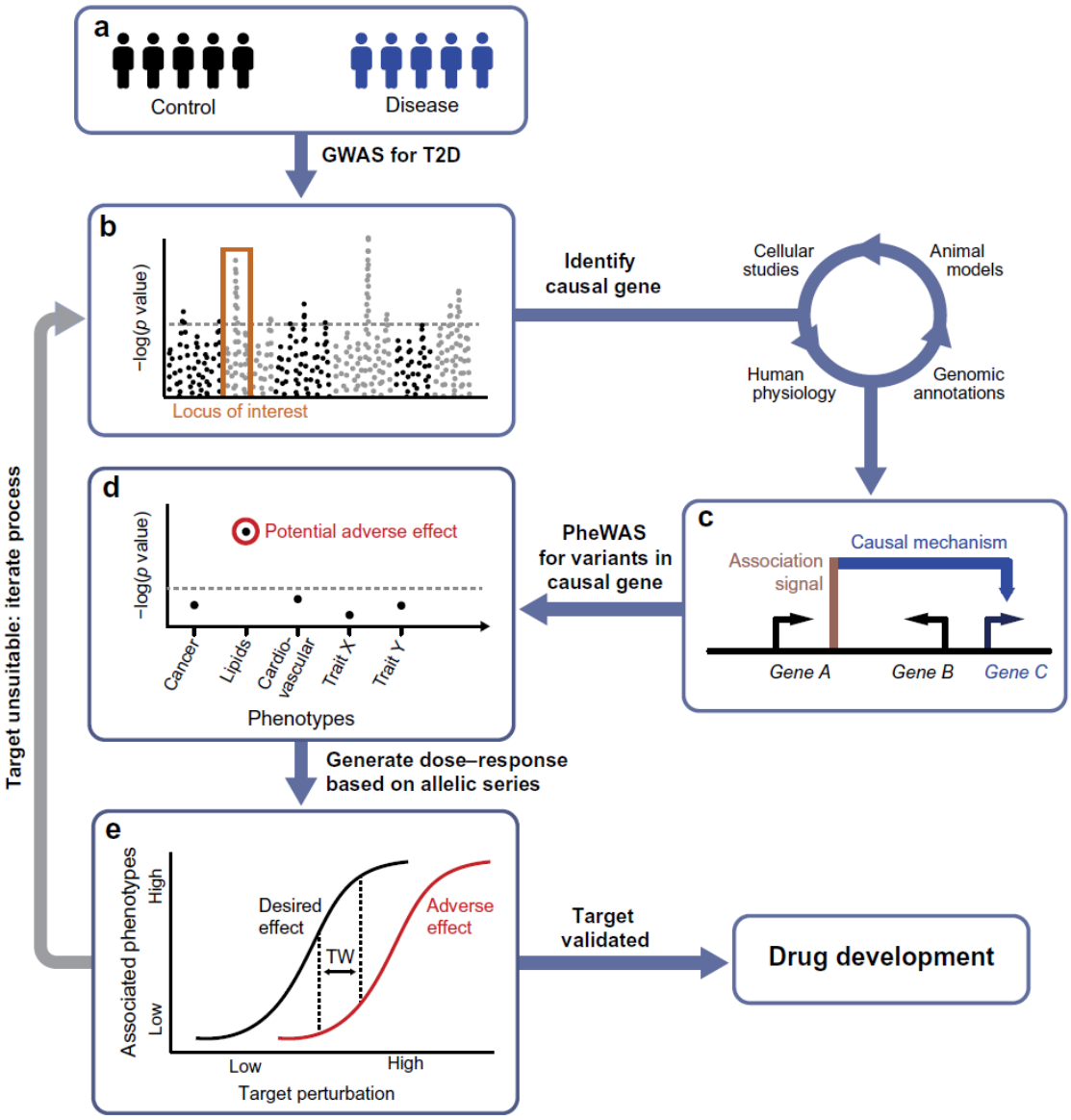


Figure 5 *Using human genetics as a model for drug target validation.* GWAS into the genetic heritability of T2D have identified a large number of variants that are robustly associated with disease risk (A, B). Once a causal gene has been identified (C), the encoded protein may be taken forward for further validation as a potential drug target. Genetic alleles within the causal gene can be interrogated for links to other phenotypes using phenome-wide association studies (PheWAS), which can highlight likely adverse or beneficial effects of long-term treatment (D). For candidate genes harbouring multiple, independent alleles, effects on disease risk can be correlated against their known impact on protein function (E). Some perturbations, such as protein-truncating variants, have predictable effects, while most alleles require extensive experimental follow-up to reliably ascertain their functional impacts. If an allelic series has been established, their phenotypic associations can be used to generate the genetic equivalent of a dose–response curve (E). The therapeutic window (TW) marks the range of perturbations that produce a suitable ratio between desirable effects (i.e. T2D protection) and adverse effects (e.g. raised lipid levels). *Taken from: S.K. Thomsen, and A.L. Gloyn, (2017) Human genetics as a model for target validation: finding new therapies for diabetes; Diabetologia (133).*

Although no validated drug targets have emerged from T2D GWAS to date, recently identified coding variants have highlighted plausible candidates. One candidate that has been the focus of particular interest is the *SLC30A8* gene, as previously discussed (*Section 1.2.3*). An added benefit of finding causal variants in coding regions is the offer of a more straight-forward interpretation for therapeutic strategies. Non-coding variation is subject to the context-dependent activity of cis-regulation and the effects can be restricted to specific tissues or developmental stages (*Section 1.3.4*). As a consequence, drugs that target the affected gene could cause unexpected adverse effects by producing a more global phenotype. Though coding variants can also be subject to such context-dependency (e.g. through tissue-specific isoforms), the effector transcripts are often affected more widely (82).

1.5 The pancreatic beta cell

The pancreatic islets of Langerhans are the endocrine component of the human pancreas, constituting roughly 1-2 % of the total volume (134). There are several million islets distributed non-randomly throughout the pancreas, with clusters of islets surrounding large blood vessels. Each micro-organ is made up of a few thousand cells, and five different endocrine cell types (135). The key function of the islet is in orchestrating glucose homeostasis through the coordinated actions of insulin and glucagon, as previously discussed (*Section 1.1.2*). Beta cells make up ~60% of total islet cells, while alpha cells contribute another ~30%. Less common cells types include somatostatin-producing delta cells (<10%), as well as PP cells (<5%) and epsilon cells (<1%), which produce pancreatic polypeptide (PP) and ghrelin, respectively (134). The cytoarchitecture of islets is known to differ considerably between species (136). Human islets have beta cells scattered throughout and mostly in association with other endocrine cells types. Rich vascularization permit cells to secrete the intracellularly stored hormones, which enter the bloodstream and act in both paracrine and endocrine manners to achieve their function (136).

1.5.1 Mechanisms of insulin secretion

The most important aspect of beta cell function is the coupling of ambient glucose levels to appropriate levels of insulin secretion. This is achieved by a cascade of protein activity that culminates in the exocytosis of insulin-containing vesicles (**Figure 6**). Briefly, rising glucose levels trigger facilitated glucose uptake via dedicated transporters (GLUT1/2/3) (137, 138). Once inside beta cells, glucose is phosphorylated by glucokinase (GCK) to prime the molecule for glycolysis (139). The subsequent breakdown of glucose through the combined action of glycolysis and the Krebs cycle results in an increased ATP:ADP ratio.

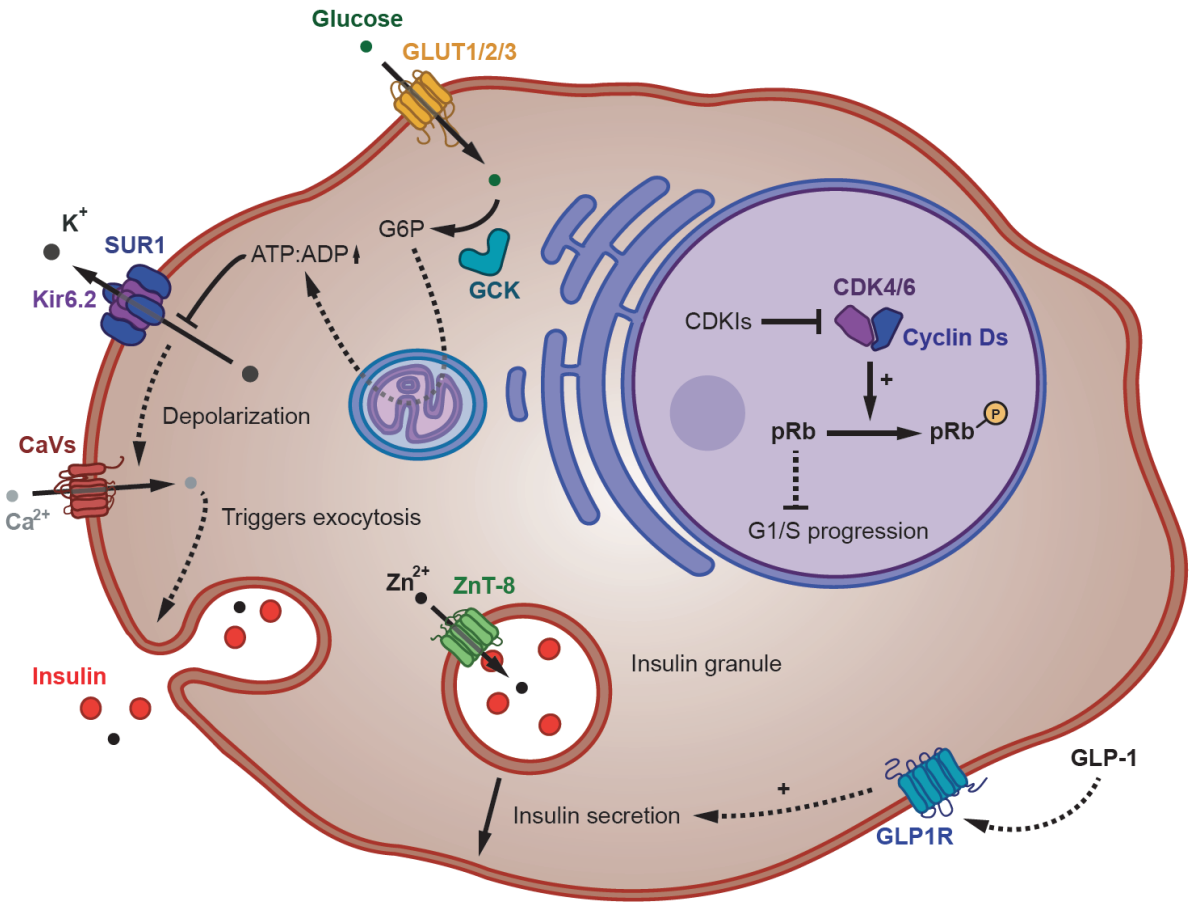


Figure 6 *Key pathways of pancreatic beta cell function.* In the cytosol, components involved in the regulation of insulin secretion are depicted: GCK (encoded by *GCK*), GLUT1/2/3 (*SLC2A1/2/3*), Insulin (*INS*), Kir6.2 (*KCNJ11*), SUR1 (*ABCC8*), voltage-gated Ca^{2+} channels (CaVs), which represent both P/Q type and L-type channels, ZnT-8 (*SLC30A8*), GLP-1-R (*GLP1R*), and GLP-1 (GLP1). In the nucleus, components involved in the regulation of the G1/S progression are depicted: CDK4/6 (*CDK4 and CDK6*), pRb (RB), Cyclins and Cyclin-dependent kinase inhibitors (CDKIs). *Adapted from: S.K. Thomsen, and A.L. Gloyn (2014) The pancreatic beta cell: recent insights from human genetics; Trends in Endocrinology & Metabolism (140).*

This shift in the intracellular energy status of the cell is coupled to a change in electrical activity at the cell membrane through the ATP-sensitive potassium ion channel K_{ATP} (141). Each K_{ATP} channel is made up of four subunits of SUR1 (encoded by *ABCC8*) and four subunits of Kir6.2 (encoded by *KCNJ11*). During low glucose levels, the open channel facilitates the passive diffusion of K^+ out of the cell, thus maintaining a negative resting membrane potential of -70 mV (142). Upon a rise in glucose levels, ATP binds to SUR1 and promotes channel closure. The resulting depolarisation triggers opening of voltage-gated calcium ion channels (CaVs) and, in turn, the fusion of insulin secretory vesicles to the cell membrane (143). During the first phase of insulin secretion, preformed granules are released from the readily releasable pool of vesicles pre-docked at the cell surface (144). In the sustained, second phase of insulin secretion, granules from the reserve pool are mobilized from intracellular stores to become docked and fused at the extracellular membrane. In each case, insulin stored as cargo within the vesicles is released into the bloodstream.

Aside from the primary mechanism of insulin secretion outlined above, several different pathways can contribute to the potentiation of secretion. Most notably, the incretin hormones GIP and GLP-1 can bind their cognate receptors (GIPR and GLP1R, respectively) to stimulate intracellular levels of cAMP (8). Elevated cAMP has a number of downstream effects, including both direct modulation of ion channel activity (both K_{ATP} and CaVs) and vesicle trafficking, as well as upregulation of insulin expression (145). The net effect is a potentiation of insulin secretion that works in synergy with the glucose-dependent pathway.

Several known therapeutic modalities for diabetes impact at various levels of the insulin secretory pathway. As previously described (*Section 1.2.4*), the K_{ATP} channel is the target of sulphonylureas, which are a class of compounds used in the treatment of both monogenic diabetes (caused by *KCNJ11* or *ABCC8* mutations) and T2D. The drugs promote closure of the ATP-sensitive channel, and thus increase insulin secretion in a largely glucose-independent

manner (146). As would be expected from their mechanism of action, hypoglycaemia is a side-effect of sulphonylureas caused by excess insulin secretion. The related meglitinides have weaker SUR1 affinity and a faster half-life, thereby minimizing, but not eliminating, the risk of inducing low blood glucose levels (146). GLP-1 analogues (and DPP-IV inhibitors, which stimulate endogenous GLP-1 levels) are also used as insulin secretagogues in the treatment of T2D (147). Because of their glucose-dependent mechanism of action, the risk of causing hypoglycaemia is lower than that of both sulphonylureas and meglitinides.

In more advanced stages of T2D, when insulin secretagogues are insufficient to provoke the required beta cell response, insulin analogues can be directly administered to bypass endogenous secretion (148). However, to mimic the natural fluctuations in insulin levels that accompany fasted and fed conditions in healthy individuals, injections with both short and fast-acting analogues are required. Even then, intensive glycaemic control is difficult to achieve and typically not sufficient to prevent macrovascular complications (149).

1.5.2 Cell cycle regulation in the beta cell

Total beta cell mass is a key determinant of overall insulin secretion and the ability of beta cells to compensate for developing insulin resistance (150). The number of beta cell can vary by as much as 10-fold between healthy individuals, which may be, at least in part, genetically determined (151, 152). Genetic risk loci for T2D harbour a number of genes encoding cell cycle regulators known to be important in beta cell proliferation, including cyclin-dependent kinase inhibitors (CDKIs; *CDKN2A/B*, *CDKN1C*, *CDKN2C*) and cyclins (*CCNE2*, *CCND2* and *CCNA2*) (26, 27). In humans, beta cell proliferation peaks around a year after birth, but is still low in absolute terms (1-3%) (153). It then declines rapidly, becoming nearly undetectable in adult islets. This represents an important difference to rodent islets, in which beta cell proliferation is considerably higher (~20% postnatally) (152, 154).

The core cellular machinery underlying cell cycle regulation in human beta cells is shared with that of other cell types (**Figure 6**). In non-proliferating cells, the Retinoblastoma protein (pRb) and pocket proteins (p107 and p130) bind E2F transcription factors to inhibit progression past the G1/S restriction point of mitosis (152). CDKIs, such as p16 and p19, inhibit CDK4 and CDK6 to prevent phosphorylation of pRb. In proliferating cells, CDK4/6 form complexes with cyclin partners to phosphorylate pRb. Hyperphosporylated pRb is then unable to inhibit E2Fs, which initiate entry into the cell cycle. In adult human beta cells, expression of the cell cycle regulator p16 (encoded by *CDKN2A*) has been found to correlate with markers of senescence, consistent with its role as a CDKI (see Chapter 3) (155). Thus, even in the presence of pro-mitogenic signals, the relative balance of cyclins, cyclin-dependent kinases (CDK) and CDKIs favours quiescence over proliferation.

1.5.3 *In vitro* cellular models of human beta cells

Due to the limited availability of human islets and the incapacity of primary beta cells to be expanded *in vitro*, several attempts have been made to generate transformed cell culture models. Most early successes were made in rodent beta cells, producing widely-used insulin-secreting cell lines, such as INS-1, MIN6, RIN and derivatives (156). Many of these offer reasonable glucose sensitivity and replicative ability in cell culture, and have inspired research into beta cell physiology over the past decades. Though rodent models continue to be an important tool for studying T2D pathogenesis, it has become increasingly clear that murine pancreatic islets differ in a number of important ways from their human counterparts (33, 136, 157). As a result, certain monogenic forms of diabetes, for example, are not well recapitulated in rodents, and molecular mechanisms elucidated in animal models should be confirmed in humans (80, 137, 158). These results have emphasised the need for a more relevant human beta cell model to study fundamental mechanisms of T2D predisposition.

Eventually, *in vitro* differentiation of induced pluripotent stem cells (iPSCs) may become a source of primary-like human beta cells. Despite recent advances, however, there are a number of technical hurdles to achieving a reliable output of mature iPSC-derived beta cells (159, 160). Meanwhile, significant progress was made recently by Ravassard and colleagues, who produced the first functional human beta cell line, EndoC-βH1 (157, 161). The cell line was derived from foetal pancreatic buds (week 7-11 of gestation), immortalized, and then matured in mice following transplantation. The transformation process was achieved by stable integration of the two transgenes human telomerase reverse transcriptase (hTERT) and SV40 large T antigen (SV40LT) under the control of the rat *Ins II* promoter. While hTERT prevents replicative senescence by extending telomeres, SV40LT is a dominant oncogene that drives proliferation by inhibiting pRb (*Section 1.5.2*).

Importantly, the EndoC-βH1 cell line is able to be expanded for more than 80 passages in cell culture, and displays many key features of primary beta cells (157). This includes expression of beta cell-specific markers (*MAFA*, *PDX1*, *NKX6-1*, *INS*), glucose-responsive insulin secretion, sensitivity to a range of secretagogues (e.g. GLP-1, 3-isobutyl-1-methylxanthine [IBMX], tolbutamide), and the ability to reverse chemically-induced diabetes in mice. Although these features make the EndoC-βH1 the first functional human beta cell line available for research, a number of limitations have constrained possible applications. Firstly, the cells have proven tricky to passage consistently, with a variable doubling time of ~1 week and difficulties with clonal expansion. Further, though the cells show mono-hormonal expression of *INS* (no significant expression of other islet hormones), total insulin protein levels are ~10-fold lower than those observed in primary beta cells. Glucose-responsive insulin secretion is also fairly modest, with a ~2-fold ratio of basal to stimulated secretion in the absence of secretagogues.

Some of these features could suggest that the EndoC-βH1 cells reflect a more immature beta cell state. It is also possible that beta cells, more generally, exist on a trade-off between replication and specialised function, as has previously been suggested (146, 155, 162). Forcing cellular proliferation, even if slow, could therefore result in diminished expression and secretion of insulin. Still, the generation of the first human beta cell line represents an important step forward, and subsequent studies have confirmed the relevance of the EndoC-βH1 in studying stimulus-coupled insulin secretion (163-165).

1.6 Aims

My thesis focuses on the identification of causal genes underlying genetic risk variants for T2D. In the following chapters, I will study the role of candidate causal genes in beta cell function, motivated by the primary aim of shining light on the fundamental mechanisms underpinning T2D predisposition. In so doing, the work will also explore strengths and weaknesses of different strategies for prioritizing causal genes. Starting with a more traditional candidate-gene approach for a locus harbouring non-coding risk variants in *Chapter 3*, I will then demonstrate in *Chapter 4* how coding risk variants can guide experimental follow-up studies, before finally developing a functional genetic screen for the identification of plausible effector transcripts in *Chapter 5*. The overarching aims for each part are as follows:

1. In Chapter 3, I aim to assess the candidacy of *CDKN2A* as a causal gene for T2D risk by characterising the impact of *CDKN2A* LOF mutations on glucose homeostasis in humans.

2. In Chapter 4, I aim to elucidate a causal mechanism for coding T2D risk variants in *PAM* by investigating both local and extra-pancreatic effects on beta cell function.

3. In Chapter 5, I aim to perform unbiased and scalable prioritization of causal genes for T2D risk by developing and implementing a high-throughput genetic screen for beta cell dysfunction.

CHAPTER 2

General Methods

2.1 Study approval

All studies were approved by the University of Oxford's Tropical Research Ethics Committee (OxTREC reference 2–15), or the Oxfordshire Regional Ethics Committee B (REC reference 09/H0605/2). Human pancreatic tissue was obtained from subjects at post-mortem examination according to local and national ethics permissions.

2.2 Maintenance of cells in culture

2.2.1 EndoC-βH1

The EndoC-βH1 cell line was maintained in low-glucose (5.6 mM), pyruvate DMEM culture medium (Thermo Fisher Scientific) supplemented to contain 2% (w/v) bovine serum albumin Fraction V, fatty acid free (BSA; 10775835001, Roche), 2 mM glutamine (Thermo Fisher Scientific), 50 μM 2-mercaptoethanol (Thermo Fisher Scientific), 10 mM nicotinamide (Sigma-Aldrich), 5.5 μg/mL transferrin (Sigma-Aldrich), 6.6 ng/mL sodium selenite (Sigma-Aldrich), 100 U/mL penicillin (Thermo Fisher Scientific), and 100 μg/mL streptomycin (Thermo Fisher Scientific) prepared fresh weekly and filtered through a 0.22 μM filter.

The cell line was passaged once per week by washing adherent cells in phosphate buffered saline (PBS; Sigma-Aldrich), followed by a 5 min incubation in trypsin-EDTA (Thermo Fisher Scientific) at 37°C. The process was then stopped with neutralisation buffer made up of 20% foetal bovine serum (FBS; Thermo Fisher Scientific) in PBS, prior to cells being centrifuged at 200g for 5 min (Beckman GS-15R, Beckman Coulter), washed in PBS,

and centrifuged again. The cell pellet was finally resuspended in fresh culturing media, and the solution of cells counted twice on an automated cell counter (Cellometer Auto T4 cell counter, Nexcelom Bioscience) before re-plating at a density of 48,000 cells/cm^2 on tissue culture flasks pre-coated overnight. The coating medium was based on high-glucose (25 mM) DMEM (Thermo Fisher Scientific) supplemented to contain 2 μg/mL fibronectin (Sigma-Aldrich), 1% ECM gel (Sigma-Aldrich), 100 U/mL penicillin, and 100 μg/mL streptomycin. All tissue culture work was performed in biological safety cabinets at OCDEM, and cells were routinely tested for mycoplasma contamination using the MycoAlert Assay kit (Lonza).

2.2.2 Primary islets

Primary human islets were freshly isolated from deceased donors at the Oxford Centre for Islet Transplantation (OXCIT) in Oxford, UK, or at the Alberta Diabetes Institute IsletCore as previously described (166, 167). Islets were then cultured in CMRL medium (Thermo Fisher Scientific) at 37°C or kept in UW cold storage solution (Viaspan) at 4°C for short-term storage, before being processed as described where relevant.

2.3 Gene silencing

To perform gene silencing by RNA interference (RNAi), a master mix of Opti-MEM reduced serum-free medium (Thermo Fisher Scientific) was pre-mixed with 0.4% RNAiMAX transcription reagent (Thermo Fisher Scientific) and 50 nM small interfering RNA (siRNA; as indicated) by repeated pipetting. The solution was then incubated at room temperature (RT) for 15 min to form siRNA-lipid complexes. During this time, cells were harvested as described above (*Section 2.2.1*). siRNA complexes were then transferred to pre-coated plates, taking care to prevent the coating from drying out. In microplate format, 50 μL of master mix was added to each well on 96-well plates (Corning) using a multichannel pipette. Cells resuspended in standard culturing media were then added in a 1:1 volume to a final density of 75,000 cells/cm^2

and siRNA concentrations of 25 nM. After 24 hours, the transfection media was removed and replaced with complete culturing media. After a further 48 hours, cells could then be processed for gene expression analysis to confirm successful gene silencing.

2.4 Gene expression analysis

2.4.1 RNA extraction

RNA was extracted from cells grown in a monolayer using TRIzol reagent (Thermo Fisher Scientific). Briefly, cells were washed in PBS and lysed in 0.5 mL TRIzol per cm^2, incubated for 5 min at RT, and transferred to 1.5 mL clear tubes following repeated pipetting to ensure complete lysis. Chloroform (Sigma-Aldrich) was added in 1:5 ratio to each sample, which was shaken vigorously for 10 seconds. Samples were then incubated for 1-2 minutes at RT, until an initial phase separation had been achieved, before being centrifuged for 15 min at 12,000g in a benchtop centrifuge (Allegra 64R centrifuge, Beckman Coulter) pre-cooled to 4°C. The transparent, aqueous phase was then carefully transferred to a new tube and mixed with 10 μg of glycogen (Thermo Fisher Scientific) used as an inert precipitation carrier. To initiate the precipitation of RNA, pure isopropanol (Sigma-Aldrich) was added to a final concentration of 50% and the tubes stored at -20°C overnight. Samples were then centrifuged at 12,000g at 4°C for 10 min, and pellets washed in 75% ethanol (Sigma-Aldrich). The washing step was repeated twice, with the supernatant discarded after each wash. Finally, the pellet was air dried for 5-10 min at RT to remove residual ethanol. When the white RNA/glycogen pellet had become gel-like, 20 μL of nuclease-free water (Ambion) was added. To resuspend the pellet fully, the samples were briefly vortexed and left at 4°C for 2-3 h, before the concentration could be quantified for real-time polymerase chain reaction (RT-PCR) (*Section 2.4.2*) or for RNA-sequencing (RNA-seq) (*Section 2.5.1*).

2.4.2 Reverse transcription

In preparation for reverse transcription and RT-PCR, sample quality control and quantification were performed on the NanoDrop 8000 spectrophotometer (Thermo Fisher Scientific). Based on a sample input of 1.5 μL, absorbance measurements were used to estimate RNA concentrations and the presence of impurities. Only samples with an A260/A280 ratio of 1.8-2.0, indicative of low levels of protein contamination, were taken forward for reverse transcription using the SuperScript III First-Strand Synthesis System (Thermo Fisher Scientific). For each reaction, RNA was mixed with 2.5 μM oligo$(dT)_{20}$ primer and 0.5 μM dNTP mix in 96-well PCR plates (4titude), then denatured at 65°C for 5 min in a tetrad thermal cycler (DNA Engine, MJ Research), and placed on ice for 1 min. After annealing, the samples were mixed 1:1 with the following master mix: 2X RT buffer, 10 mM $MgCl_2$, 20 mM DTT, 20 U RNAse OUT, 100 U SuperScript III R in nuclease-free water, and incubated at 50°C for 50 min to perform the cDNA synthesis. Reactions were then terminated by incubation at 85°C for 5 min, briefly centrifuged, and mixed with 1 μL RNase H to degrade RNA. After incubation at 37°C for 20 min, the cDNA was stored at 4°C for short-term storage or -20°C for longer term storage.

2.4.3 Real time TaqMan PCR

For RT-PCR, cDNA was diluted in nuclease-free water as required, and 9.5 μL was added to 0.5 μL of a reaction-specific TaqMan assay (containing target primers and an exon-spanning probe, details provided where relevant; Applied Biosystems) on a 384-well plate (4titude). Each sample was then analysed in technical duplicate for genes of interest (GOIs) and a minimum of two-housekeeping genes (HKGs), *TBP* (TaqMan assay Hs00427620_m1), and *PPIA* (Hs999999904_m1). RT-PCR was performed on the 7900HT Fast Real-Time PCR (Applied Biosystems) for 40-45 cycles, and median C_T values of technical replicates determined using the SDS v2.3 software. Threshold values were visually inspected and

adjusted as necessary. Subsequent analysis was performed in R using the ΔΔCt-method ($2^{\Delta\Delta Ct}$, where ΔΔCt is the average C_T of HKGs subtracted from the C_T of GOIs). Relative expression values ($2^{-\Delta\Delta Ct}$) were normalized to negative controls (e.g. scrambled sequence of siRNA) where available.

2.5 RNA sequencing

2.5.1 Sample QC

In preparation for RNA-seq, sample quality control and quantification was performed on the 2100 Bioanalyzer using the RNA 6000 Nano kit (Agilent) according to manufacturer's instructions. The instrument is based on automated electrophoresis of RNA fragments, and quantification of their intensities relative to a control. The electropherogram also enables assessment of degradation patterns independent of concentrations, producing an RNA integrity number (RIN) from 1 (completely degraded) to 10 (fully intact). Only samples scoring higher than a RIN of 8 were taken forward for RNA-seq.

2.5.2 Illumina sequencing

For sequencing, all samples were processed at the Oxford Genomics Centre (Wellcome Trust Centre for Human Genetics, University of Oxford). Briefly, poly-A selected libraries were prepared using the NEBNext PolyA mRNA Magnetic Isolation Module and NEBNext Ultra Directional RNA Library Prep Kit for Illumina (both New England Biolabs) with 12 cycles of PCR and custom 8bp indexes (168). Libraries were sequenced in multiplex on the Illumina HiSeq2000 as 100-nucleotide paired-end reads, using the TruSeq PE Cluster Generation Kit v3 and TruSeq SBS Kit v3 (both Illumina). TopHat2 on default settings was used to align sequencing reads to the human genome reference GRCh37.p13, and using GENCODE release 18 as the transcriptome reference (169, 170).

2.6 Cellular phenotyping

2.6.1 Cell proliferation assay

EndoC-βH1 cells were counted using the CyQUANT Direct Cell Proliferation kit (Thermo Fisher Scientific). This end-point assay is based on detection of a cell-permeant nucleic acid dye on bottom-read microplate readers. A suppressor dye is added to block direct signal from cells with a compromised cell membrane. To set up for a proliferation assay, cells were plated on clear-bottom black 96-well plates (Corning) in triplicate for each time-point to be measured. At the appropriate times, culture medium was replaced with medium supplemented with 0.2% direct nucleic acid dye and 1% background suppressor dye. Cells were then incubated for 1h at 37°C in the dark, and fluorescence measured on the EnSpire multimode plate reader (PerkinElmer) using FITC filter settings (ex480/em535 nm). Cell counts were expressed as relative fluorescence units (RFU) normalised to the earliest time point.

2.6.2 Static insulin secretion assay

For insulin secretion assays, cells were passaged as described (*Section 2.2.1*) and plated at a density of 75,000 cells/cm2 or pre-treated as required (*Section 2.3*). After 72h, cells were starved overnight in zero-glucose culture medium based on DMEM (Thermo Fisher Scientific) supplemented with glucose to a final concentration of 2.8 mM and additional components as described for the standard culture medium (*Section 2.2.1*). After overnight incubations, cells were further starved for 1h in culture medium with 0 mM glucose. To minimally disturb the monolayer of cells, media changes in 96-well format were performed using multichannel pipettes, adding the fresh media to the sides of the wells after carefully removing existing media. Columns were staggered by 20 seconds to keep incubation times consistent across the plate.

After 1h, starvation medium was replaced with the appropriate conditions to initiate secretion. All media were based on zero-glucose culture medium, supplemented as required, and pre-warmed to 37°C immediately prior to addition. The dimethyl sulfoxide (DMSO)-soluble reagents tolbutamide (Sigma-Aldrich), IBMX (Sigma-Aldrich), diazoxide (Sigma-Aldrich), GLP-1 (Bachem), and TRH (Bachem) were prepared fresh, and an equal volume of DMSO (Sigma-Aldrich) was added to all other conditions. After 1h, supernatants were harvested in a staggered fashion, leaving a residual volume of supernatant (~30 µL in 96-well format) to minimize the number of cells accidentally removed during aspiration. To further prevent insulin contamination from aspirated cells, samples were centrifuged at 300 g in a pre-chilled centrifuge (Heraeus Multifuge 3SR, Thermo Fisher Scientific) for 5 min and a small volume of the supernatant transferred to a new plate. Meanwhile, cells that remained adherent following secretions were processed for measurements of cell counts and/or insulin content, as needed.

If cell count measurements were not required, residual media from the secretion assays were removed and replaced with ice-cold acid-ethanol buffer (1.5% conc. HCl [Sigma-Aldrich], 75% ethanol [Sigma-Aldrich], and 23.5% deionized water) to extract the intracellular insulin content. Plates were then covered in parafilm to prevent evaporation and stored at -20°C for short-term storage or -80°C for long-term storage. If end-point cell counting was required, residual media was removed and the CyQUANT Direction proliferation assay performed as described (*Section 2.6.1*). Cells were then washed once in standard media, before acid-ethanol was added as above. Insulin samples were then taken forward for analysis by AlphaLISA.

2.6.3 Insulin AlphaLISA assay

To analyse insulin samples from EndoC-βH1 cells, the Insulin (human) AlphaLISA Detection Kit (PerkinElmer) was used for measuring concentrations of both supernatants and

cellular contents. As a no-wash extension of conventional enzyme-linked immunosorbent assays (ELISAs), this immunoassay is based on bringing two distinct antibodies in contact with the same analyte molecule. Proximity of the antibodies in the presence of insulin leads to amplified luminescence that can be detected on a plate reader.

Optimal dilution ranges were empirically determined, and found to be 1:10 for supernatants and 1:200 for contents. Samples were diluted in 1X AlphaLISA Immunoassay buffer, and then analysed in white 1/2-AreaPlate-96 (PerkinElmer) according to manufacturer's recommendations. Plates were measured on the EnSpire using the AlphaScreen protocol (PerkinElmer), unless otherwise stated. Each plate included a ten-point standard curve of known insulin concentrations and blanks used for background correction. Raw data were analysed in R, using a 5-parameter logistic regression to fit the standard curve from which insulin concentrations could be calculated for all samples. Normalized secretion values were then calculated on a per-well basis as the ratio of total secreted insulin to total intracellular insulin content (%), or as the ratio of total secreted insulin to cell counts (pg/RFU), which were finally be averaged across replicates.

2.7 Immunofluorescence and light microscopy

Cells cultured in a monolayer on 4-chamber culture slides (Corning) were washed with cold PBS and fixed in 4% paraformaldehyde (Sigma-Aldrich) for 20 min at RT. Samples were then permeabilised for 20 min in PBS/Triton (0.1%; Sigma-Aldrich), blocked in 10 % BSA PBST for 30 min at RT, and incubated at 4°C in 5% BSA PBST with the appropriate antibody. Following overnight incubations, samples were washed five times in PBST and incubated in the dark for 1h at RT in 5 % BSA PBST with Alexa Fluor-conjugated secondary antibody (1:500; Thermo Fisher Scientific). After washing three times in PBST, cell nuclei were stained for 15 min in 20 μM NucRed Dead 647 (Thermo Fisher Scientific), a far-red cell-permeant

dye. Slides were then mounted using a drop of Vectashield Mounting Medium (Vector Labs) and covered with a glass coverslip (VWR). Finally, slides were cured in the dark for 24h, sealed using nail varnish, and stored at 4°C for short-term storage. Images were taken on the Radiance 2100 confocal microscope (BioRad) with a 60X 1.0 N.A. objective. The same laser settings and intensities were applied across all samples.

2.8 Protein expression studies

2.8.1 Cell lysis and protein quantification

Cells to be analysed for whole-cell extract by western blotting were harvested as previously described (*Section 2.2.1*) and pellets obtained after washing in PBS were lysed in ice-cold, freshly prepared RIPA buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, 10 mM NaF, 1X protease inhibitor cocktail [all Sigma-Aldrich]), and rotated for 30 min at 4°C. Lysates were then transferred to 1.5 mL tubes and centrifuged for 20 min at 12,000g at 4°C (Allegra 64R) before transferring supernatants to new tubes. Aliquots of the samples were diluted 1:10 and protein concentrations determined by Bio-Rad Protein Assay according to manufacturer's instructions. Absorbance was read at 595 nm on the Versamax microplate reader using the Softmax Pro v5.4 software (Molecular Devices), and a standard curve of known BSA amounts was used to derive concentrations for all samples based on linear regression.

2.8.2 Gel electrophoresis and transfer

Equal amounts of protein were diluted as required, and denatured in loading buffer (with final concentrations of 1X Laemmli loading buffer [Bio-Rad] and 89 mM 2-mercaptoethanol) at 95°C degrees for 10 min. Samples were briefly centrifuged to collect evaporated liquids, then loaded alongside a molecular weight marker (Precision Plus All Blue Standard; Bio-Rad) on a Criterion TGX stain-free 4-12% pre-cast gel (Bio-Rad) and separated

by sodium dodecyl sulphate polyacrylamide electrophoresis (SDS-PAGE) in 1X Tris/Glycine/SDS (TGS) gel running buffer (Bio-Rad). Gels were run at 200V for 1h, and subsequently activated and imaged using the ChemiDoc MP imaging system (Bio-Rad) to visualise total protein. Proteins were transferred to a membrane using the Trans-Blot Turbo transfer system (Bio-Rad) according to manufacturer's recommendations, and transfer visually confirmed by imaging gels again.

2.8.3 Western blotting

Membranes were blocked in TBST buffer made up of 1 mM Tris-HCl pH 7.4 (Sigma-Aldrich), 5 M NaCl (Sigma-Aldrich), 0.05% Tween-20 (Thermo Fisher Scientific) with 5% non-fat milk powder for 1h at RT under gentle agitation on the Stuart see-saw rocker (Bibby Scientific). The appropriate primary antibodies were then added to membranes in blocking buffer. After overnight incubations with agitation, membranes were washed three times in TBST for 10 min each, and incubated for 1h with appropriate Horseradish peroxidase (HRP)-conjugated secondary antibodies (Thermo Fisher Scientific) in 5% TBST. Finally, membranes were washed three times in TBST for 10 min each, and incubated in western enhanced chemiluminescence (ECL) reagent (Bio-Rad) for 5 min before being visualised on the ChemiDoc system using standard settings optimized for detection of faint bands. ImageLab 5 software was used to perform densitometry analysis using fixed band sizes and subtractive background correction.

CHAPTER 3

Loss-of-function Mutations in *CDKN2A* Impact on Glucose Homeostasis in Humans

---

The results presented in the following chapter contributed to a previously published article, which is referenced below. The complete manuscript is provided in the Appendices.

- Pal*, T.P. Potjer*, **S.K. Thomsen***, et al, Loss-of-function mutations in the cell-cycle control gene CDKN2A impact on glucose homeostasis in humans. Diabetes. 2016 Feb:65(2):527-33

*Joint first authors

3.1 Introduction

3.1.1 Disease genetics of the *CDKN2A/B* locus

Among the first T2D association signals to be discovered by GWAS, non-coding risk variants at the *CDKN2A/B* locus (chromosome 9p21) have been consistently associated with disease risk across multiple ancestries and geographies (26, 27, 171, 172). Two distinct signals (lead SNPs rs10757283 and rs10811661) map to a 12 kb region located upstream of *CDKN2A* and *CDKN2B*, whilst one signal (rs564398) is contained within the long non-coding RNA (*ANRIL*) (**Figure 1**). The *CDKN2A/B* locus also harbours association signals for a number of other conditions, including coronary artery disease, ageing-related disorders, and different forms of cancer, though the T2D risk alleles are physically separate from these regions.

Despite being the focus of intense research, the underlying causal mechanisms driving the genetic associations with T2D risk remain elusive (75, 173). Cis-eQTL analyses have failed to identify any effects on local gene expression in a number of metabolically relevant tissues, including pancreatic islets, liver, and colon (77, 119, 174). In contrast, the nearby cardiovascular risk variants have been shown to influence *ANRIL* expression, which in turn has been correlated with expression of both *CDKN2A* and *CDKN2B (174, 175)*. Deletion of an interval that includes the *Cdkn2a/b* T2D risk variants has established enhancer activity in mouse cardiac tissue, but the relevance to humans has been questioned by subsequent findings (176, 177). The mouse region encodes a long noncoding RNA that is not an orthologue of *ANRIL*, highlighting the possibility of divergent cis-regulatory mechanisms.

It remains possible that human expression analyses have focused on the wrong tissue and/or developmental stage, though several lines of evidence point to a beta cell driven effect. Physiological characterisations in non-diabetic individuals have associated the T2D risk alleles with reduced measures of beta cell function in oral and intravenous glucose tolerance tests, as

well as in a hyperglycaemic clamp (aggregated with seven other T2D risk loci) (173, 178-180). Importantly, none of these studies have uncovered any effects on insulin sensitivity or clearance. In one cohort, the T2D risk alleles were also linked to reduced birth weight, suggestive of decreased foetal secretion in the final trimester of gestation (181). Other studies have not replicated this finding, and the risk variants were not associated with arginine-stimulated insulin secretion (a proxy for beta cell mass) in a hyperglycaemic clamp (179, 182, 183).

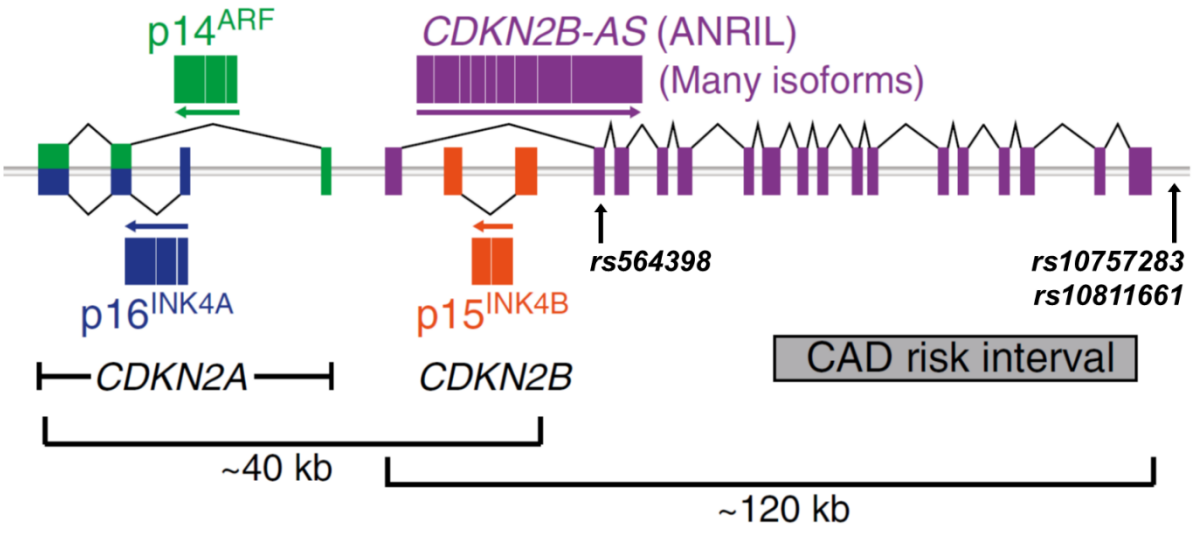


Figure 1. *Schematic of the CDKN2A/B locus showing coding regions and association signals for selected diseases. CDKN2A* encodes two proteins, p14 (green) and p16 (blue), transcribed from two distinct promoters. Though the gene products share two exons (2 and 3), p14 and p16 are in alternative reading frames and thus structurally unrelated. *CDKN2B* encodes p15 (orange), which is structurally homologous to p16. *CDKN2B-AS* (*ANRIL*) encodes a long non-coding RNA (lncRNA) that spans the *CDKN2B* coding region. The three independent T2D risk association signals identified to date are shown by lead RefSNP (rs) numbers (rs564398, rs10757283, rs10811661). CAD, coronary artery disease. Diagram not to scale. Adapted from Kong et al (2016) Diabetologia.

3.1.2 Biological functions of the proteins encoded by *CDKN2A/B*

In the absence of direct evidence nominating a particular causal gene at the *CDKN2A/B* locus, a number of studies have focused instead on characterising regional candidates. *CDKN2B* encodes the cyclin-dependent kinase inhibitor (CDKI) $p15^{INK4A}$ (referred to as p15 below) (**Figure 1**), which has been implicated in adipocyte differentiation and indirectly linked to islet proliferation (184-186). *CDKN2A* encodes the alternatively spliced proteins $p16^{INK4a}$ and $p14^{ARF}$ (p16 and p14 below), which are tumour suppressors acting via distinct signalling pathways (187, 188). Like p15, p16 is a CDKI, involved in the regulation of cell cycle progression through inhibition of CDK4 and CDK6 (*Section 1.5.2*) (189). p14, in contrast, prevents the degradation of the cell-cycle regulator p53 by forming a stable complex with Mdm2 in the nucleus (190). All the proteins have well-characterised roles in tumorigenesis and age-related processes, with p16 also being a key effector of senescence, a cellular state of irreversible growth arrest (174).

In terms of understanding T2D risk, *CDKN2A* has emerged as a plausible causal gene with animal and cellular models directly linking *Cdkn2a* to glucose homeostasis (155, 191-194). In the context of the pancreatic islet, *Cdkn2a* deficiency was found to increase beta cell proliferation and confer resistance to chemically-induced diabetes in a beta cell specific KO mouse (191). Although there was no detectable difference in beta cell proliferation at 10 weeks, a marked increase was observed after 60-70 weeks, highlighting the age-dependent component of *CDKN2A* function. Conversely, overexpression was associated with reduced beta cell proliferation in both young and old mice (191). These observations are also consistent with the effect of *Cdk4*-loss (equivalent to increased p16 activity), which has been shown to result in a reduced number of pancreatic beta cells and insulin-deficient diabetes (192). Complicating this view somewhat, a recent study found that beta cell specific overexpression of p16 in transgenic mice is also associated with an immediate improvement in insulin

secretion and glucose tolerance (155). This effect, however, is later counterbalanced by reduced beta cell mass through diminished proliferation.

Recent mouse studies have also established roles for *Cdkn2a* and *Cdk4* in hepatic glucose production, with cell-cycle independent effects on gluconeogenesis under fasted and fed conditions (193, 194). One of these studies found that 12-week old p16-deficient mice display increased hepatic gluconeogenesis and elevated fasting glucose levels (194). On the basis of their findings, the authors proposed a model whereby p16-dependent sequestration of CDK4 plays a role in regulation of protein kinase A (PKA). Upon loss of p16, CDK4 translocates to the cytoplasm where it increases PKA activity. A similar role in islet beta cells has not been investigated to date.

Taken together, available evidence from experimental models points to a decrease in *Cdkn2a* activity being protective against diabetes. Though p16 deficiency may affect different tissues through both cell-cycle dependent and independent roles, a primary effect on glucose homeostasis appears to be mediated through a long-term increase in beta cell function. To the extent that these findings can be translated into human physiology, elevated p16 levels would thus be expected to confer increased risk of T2D. This, in turn, suggests that any cis-regulatory effect of the non-coding risk alleles at the locus would be predicted to increase *CDKN2A* expression.

3.1.3 *CDKN2A* mutations in humans

While animal studies have provided clues into the possible contribution of *Cdkn2a* to diabetes pathogenesis, less is known about the role of *CDKN2A* in glucose homeostasis in humans. The machinery regulating the G1/S transition in adult human beta cells differs from that of mouse beta cells, which, for example, do not express CDK6 (195-197). *CDKN2A* expression in human beta cells has been correlated with markers of cellular senescence, but

mechanistic studies are difficult to undertake because of the limited availability of primary islets from donors. One study performed silencing of p16 in the human cellular model EndoC-βH2, and confirmed a key role of the protein in triggering senescence and improved insulin secretion (155). Such findings are consistent with the notion that beta cells exist on a trade-off between proliferation and function, but the long-term metabolic effects of altered *CDKN2A* levels in humans are unknown. Individuals heterozygous for germline LOF mutations in the *CDKN2A* gene have a high risk of developing melanoma, a condition known as familial atypical multiple mole melanoma syndrome (FAMMM) (198, 199). FAMMM is characterised by multiple melanocytic naevi, and is caused by *CDKN2A* mutations in approximately 10-40% of cases. The *CDKN2A* melanoma mutations also predispose to increased risk of developing pancreatic neuroendocrine tumours, including insulinomas, in line with the established role of p16 in regulation of beta cell mass in animal models. These FAMMM individuals provide a unique opportunity to study the effects of life-long *CDKN2A* haploinsufficiency on glucose homeostasis in humans.

The aim of the present study was to test the hypothesis that *CDKN2A* LOF mutation carriers would show improved beta cell function compared with matched controls in physiological tests performed by Dr Aparna Pal (OCDEM) and Dr Thomas Potjer (Leiden University Medical Center, Netherlands). I first used data from oral glucose tolerance tests (OGTTs) to demonstrate that *CDKN2A* LOF mutations cause a metabolic phenotype characterised by elevated insulin secretion (*Sections 3.3.1-3.3.2*). Based on more detailed measures from intravenous glucose tolerance tests (IVGTTs), I then derived glycaemic indices that are consistent with improved beta cell function and reduced hepatic insulin clearance in mutation carriers (*Section 3.3.3*). Finally, to shed light on possible cell cycle-independent effects of *CDKN2A* on beta cell function, I performed functional studies following *CDKN2A* silencing in the transformed human model EndoC-βH1 (*Sections 3.3.4*).

3.2 Materials and methods

3.2.1 Study participants

Participants carrying inherited *CDKN2A* LOF mutations (**Table 1**) and first-degree familial controls were recruited from centres in the UK and the Netherlands through collaborations with Professors Timothy Bishop and Julia Newton-Bishop (Leeds Institute of Cancer and Pathology, UK) and Professor Hans Vasen (Leiden University Medical Center, Netherlands). The following inclusion criteria were applied:

- A strong familial history of melanoma (3 or more cases with at least one family member presenting prior to 45 years of age). This criteria was put in place to increase the chance of recruiting mutation carriers, since the ethics protocol in place for the FAMMM cohort in Leeds precluded disclosure of mutation status, due to the incomplete penetrance of *CDKN2A* mutations. Non-mutation carriers were subsequently used as familial controls.
- Participants not suffering from diabetes, or taking any medication that could interfere with glucose tolerance (e.g. corticosteroids).
- Participants aged between 18-80 years.

As far as possible, subjects were matched post-hoc for age, gender and body mass index (BMI) as mutation status became available. Additional controls were recruited, as required, from the Oxford Biobank through a collaboration with Professor Fredrik Karpe (OCDEM). For Leiden cases, a close relative without the *CDKN2A* mutation or, if unavailable, a spouse was recruited as control subject. At the time of recruitment, there was no available estimate for likely effect size, and so power calculations were based on data from a similar study for *PTEN* mutation carriers (200). An effect size of d=0.8 (difference in mean of $\geq$ 0.8 SDs) would require 34 individuals in each group for $\alpha = 0.05$ and $1-\beta$ (power) = 0.9, which was used a target for recruitment.

A total of 33 mutations carriers and 33 controls were successfully recruited, and all participants gave full informed, written consent. Two cases were subsequently excluded due to insufficient evidence of pathogenicity for the identified mutations (OX_013 and OX_014; see also *Section 3.3.1*), and two controls were excluded based on 2-h OGTT glucose levels diagnosing diabetes (serum glucose >11 mmol/L).

3.2.2 OGTTs

Following a 12 hour fast, all participants underwent a 75 g OGTT performed by Dr Aparna Pal (OCDEM) or Dr Thomas Potjer (Leiden University Medical Center). A 23 gauge cannula was placed in an antecubital vein for repeated blood sampling. The oral glucose load was given as a drink over a minute, and blood collected at 0, 15, 30, 60, 90 and 120 min. For UK participants unable to travel (22 in total), the OGTT was performed in their homes using portable centrifuges and dry ice for transportation of samples back to OCDEM. Insulin, glucose, and (for all UK subjects) C-peptide concentrations were measured as described below, and then analysed based on the formulae in **Table 1**. BMI was calculated as weight (in kilograms) divided by height (in metres) squared.

Measure/index	Parameter	Formula or mathematical model used	Reference
iHOMA-B	Beta cell function	Computer model available via (http://www.dtu.ox.ac.uk/homacalculator/).	(201)
iHOMA-S	Insulin sensitivity	Computer model available via (http://www.dtu.ox.ac.uk/homacalculator/).	(201)
BIGTT-AIR$_{0\text{-}30\text{-}120}$	Beta cell function	exp[8.20 + (0.00178 * insulin$_0$) + (0.00168 * insulin$_{30}$) -(0.000383 * insulin$_{120}$) - (0.314 * glucose$_0$) - (0.109 * glucose$_{30}$) + (0.0781 * glucose$_{120}$) + (0.180 * gender (where male=0 and female=1)) - (0.032 * BMI)]	(202)
BIGTT-$S_{I_0\text{-}30\text{-}120}$	Insulin sensitivity	exp[4.90 - (0.00402 * insulin$_0$) - (0.000556 * insulin$_{30}$) -(0.00127 * insulin$_{120}$) - (0.152 * glucose$_0$) - (0.00871 * glucose$_{30}$) - (0.0373 * glucose$_{120}$) - (0.145 * gender (where male=0 and female=1)) - (0.0376 * BMI)]	(202)
Belfiore ISI	Insulin sensitivity	2/[(0.5 * glucose$_0$ + glucose$_{60}$ + 0.5 * glucose$_{120}$)/11.36) * (0.5* insulin$_0$ + insulin$_{60}$ + 0.5* insulin$_{120}$)/638) + 1]	(203)
Matsuda ISI	Insulin Sensitivity	1000 / √(glucose$_0$* insulin$_0$ * glucose$_{mean\text{-}OGTT}$ * insulin$_{mean\text{-}OGTT}$)	(204)
AUC	Multiple	Area under curve estimated using trapezoidal rule	(205)
Insulinogenic index	Beta cell function	(insulin$_{30}$ – insulin$_0$)/(glucose$_{30}$ - glucose$_0$)	(206)
Disposition index	Beta cell function	AUC$_{insulin\text{-}OGTT}$/AUC$_{glucose\text{-}OGTT}$ * Matsuda ISI	(206)
Hepatic insulin clearance	Insulin clearance	AUC$_{C\text{-}Peptide}$/AUC$_{Insulin}$	(207)
Fasting insulin clearance	Insulin clearance	C-peptide$_0$/insulin$_0$	(207)

Table 1 *Definitions of physiological measurements and indices derived from the OGTT*. Subscripts denote time points during the OGTT. Units are pmol/L for insulin, nmol/L for C-peptide and mmol/L for glucose, except in the case of Matsuda ISI and the disposition index, where glucose was inputted in units of mg/dL and insulin as μU/mL.

3.2.3 IVGTTs

IVGTTs were performed on a subset of the UK subjects (eight) who had attended for OGTT and consented to undergo an IVGTT. The tests were performed by Dr Aparna Pal (OCDEM) and a research nurse, Beryl Barrow, at OCDEM. Control subjects, matched for age, gender, BMI and activity were recruited from the Oxford Biobank. Following a 12 hour fast, a 20 gauge cannula was placed in each antecubital vein: one was used for administration of 50% dextrose and the other for blood sampling. Baseline blood samples were taken and a dose of 50 % dextrose (calculated based on weight 0.5mg/kg) was then given from the other cannula over 3 minutes. Subsequent blood samples were then taken at 0, 2, 4, 6, 8, 10, 15, 20, 30, 45, 60, 75, 90, 120, 150 and 180 minutes. Samples were batch-analysed for insulin, glucose and C-peptide as described below. IVGTT data were then analysed in collaboration with Dr Ian Godsland (Imperial College London) using the minimal model approach, according to an algorithm designed to maximise precision and identification success rate, as previously described (208). Quantitative estimates of IVGTT portal delivery rate of newly-secreted insulin, as well fractional hepatic throughput of newly-secreted insulin, and systemic insulin elimination rates were estimated using a combined model of IVGTT insulin and C-peptide concentrations (209).

3.2.4 Biochemical assays

Blood samples were placed immediately on ice and spun at 3000rpm for 10 minutes at 4°C within 30 minutes of being taken. The supernatant from each sample was then pipetted into smaller aliquots and stored in -80°C freezers. Glucose, insulin and C-peptide assays were all performed via a collaboration with Dr Tim James (University of Oxford). Briefly, glucose was analysed with a standard hexokinase assay using Siemens ADVIA 2400 analyser (Siemens, Frimley, UK), and Insulin and C-peptide were analysed by chemiluminescence immunoassays using a Siemens ADVIA Centaur analyser (Siemens).

3.2.5 Cellular assays in EndoC-βH1

The EndoC-βH1 cell line was passaged and transfected as described (*Sections 2.2-2.3*) with commercially available ON-TARGETplus siRNA SMARTpools (Dharmacon) targeting *CDKN2A*. Gene expression analysis by RT-PCR (*Section 2.4*) was performed using the following TaqMan assays: *CDKN2A* (Hs00923894_m1) and *CDKN2B* (Hs00793225_m1), prior to cellular phenotyping (*Section 2.6*). To confirm knockdown at the protein level, immunofluorescence (IF) and western blotting (WB) procedures were performed (*Sections 2.7-2.8*) with the following primary antibody: anti-p16-INK4A (Abcam, ab81278; 1:500 for IF and 1:2000 for WB).

3.2.6 Protein Kinase A activity assay

For PKA activity assays, cells were harvested following gene silencing, as described above, and washed in PBS. Matching input for number of cells, the samples were then processed according to manufacturer's instructions for the PepTag non-radioactive PKA assay (Promega). Briefly, this protein kinase assay is based on phosphorylation of a fluorescent, highly specific PKA substrate peptide (PepTag A1; LRRASLG). Phosphorylation alters the net charge of the PepTag A1 from +1 to -1, allowing unconverted and converted substrate to be separated by electrophoresis on an agarose gel. For each run, partially purified homogenates of samples, matched for cell numbers, were run in technical triplicates, along with a negative and a positive control provided with the kit. Agarose gels (0.8%) were prepared by dissolving agarose (Sigma-Aldrich) in 50 mM Tris-HCl (pH 8.0), and run at 100 V for 20 min. After electrophoresis, gels were visualized using the ChemiDoc MP system and ImageLab 5 software was used to perform densitometry analysis using fixed band sizes and subtractive background correction.

3.2.7 Statistical analysis

Statistical analysis was performed using R version 3.3.1. P-values were determined by unpaired, two-tailed t-tests, except for gender differences where the Chi-squared test was used and for analysis of the IVGTT data where the Mann-Whitney U test was used. Post hoc power calculations were performed assuming an unpaired, two-sample t-test analysis with α=0.05 and β=0.10 (power of 0.90) based on the observed effect sizes of the IVGTT $AUC_{insulin}$ 0-10 min (d = 0.70) and the IVGTT disposition index (d = 0.51).

3.3 Results

3.3.1 Characteristics of the study cohort

Thirty-three participants carrying inherited *CDKN2A* LOF mutations (**Table 2**) were recruited from centres in the UK and the Netherlands. Thirty had been cancer free for at least two years, and the remaining three cases had presented with melanoma between four to twelve months prior to inclusion in the study. For a control group of 33 participants, unaffected first-degree relatives or spouses of carriers were chosen when available and additional controls were recruited from the Oxford Biobank. To confirm the cause of FAMMM in the cohort, targeted sequencing had previously been used to test for variants in probands. This identified *CDKN2A* mutations in all subjects tested (**Table 2**), including the well-characterised Leiden deletion (c.225_243del19) in 18 of the 33 subjects (210).

Participant	Protein affected	Mutation	Melanoma-free > 2 years	Variant Effect Predictor
OX_1	P16-INK4a only	c.88delG	Yes	Frameshift variant
OX_2	P16-INK4a only	c.88delG	Yes	Frameshift variant
OX_3	P16-INK4a only	c.88delG	Yes	Frameshift variant
OX_4	Both	c.458-105A>G	Yes	Intronic variant
OX_5	Both	c.458-105A>G	Yes	Intronic variant
OX_6	Both	c.458-105A>G	Yes	Intronic variant
OX_7	P16-INK4a only	c.52_57dup	Yes	Protein altering variant
OX_8	P16-INK4a only	c.52_57dup	Yes	Protein altering variant
OX_9	Both	c.159G>A	Yes	Missense variant
OX_10	Both	c.159G>A	Yes	Missense variant
OX_11	Both	c.458-105A>G	Yes	Intronic variant

OX_12	*Not tested*	*Not tested*	Yes	*Not tested*
OX_13	Both	c.458-89A>T	Yes	Intronic variant
OX_14	Both	c.540C>T	Yes	3' UTR
LEI_001	Both	c.225_243del19	No	Frameshift variant
LEI_002	Both	c.225_243del19	Yes	Frameshift variant
LEI_003	Both	c.225_243del19	Yes	Frameshift variant
LEI_004	Both	c.225_243del19	Yes	Frameshift variant
LEI_005	Both	c.225_243del19	Yes	Frameshift variant
LEI_006	Both	c.225_243del19	Yes	Frameshift variant
LEI_007	Both	c.225_243del19	Yes	Frameshift variant
LEI_008	Both	c.225_243del19	Yes	Frameshift variant
LEI_009	Both	c.225_243del19	Yes	Frameshift variant
LEI_010	Both	c.225_243del19	Yes	Frameshift variant
LEI_011	Both	c.67G>C	No	Missense variant
LEI_012	Both	c.225_243del19	No	Frameshift variant
LEI_013	Both	c.225_243del19	Yes	Frameshift variant
LEI_014	Both	c.225_243del19	Yes	Frameshift variant
LEI_015	Both	c.225_243del19	Yes	Frameshift variant
LEI_016	Both	c.225_243del19	Yes	Frameshift variant
LEI_017	Both	c.225_243del19	Yes	Frameshift variant
LEI_018	Both	c.225_243del19	Yes	Frameshift variant
LEI_019	Both	c.225_243del19	Yes	Frameshift variant

Table 2 *List of mutations identified in carriers (annotated against CDKN2A-001 [ENST00000304494]).* The second column indicates whether the mutation maps to a region encoding just $p16^{INK4a}$ or both $p16^{INK4a}$ and $p14^{ARF}$. Participant ID is an anonymous number assigned to each case, and the prefix indicates from which cohort the patient was recruited (OX = Oxford, UK; LEI = Leiden, Netherlands). UTR = Untranslated region.

To predict the effect of mutations on protein function I used the Variant Effector Predictor (VEP) tool by ENSEMBL in conjunction with manual curation of the literature. This showed that 21 of the variants identified in subjects were frameshift mutations, causing complete loss of both p14 and p16. The list also highlighted two variants for which the most recent evidence was insufficient to support pathogenicity, and these were removed from all subsequent analysis. This includes the intronic c.458-89C>T mutation and the 3' UTR variant c.540C>T, which have no additional functional support or evidence of co-segregation in affected families (211, 212). All other identified mutations were either predicted protein-altering, missense mutations, or intronic, and have been robustly associated with disease risk (213-218). Though the mutations identified in the cohort may lead to varying degrees of impact on *CDKN2A* function, there is little evidence to suggest any simple genotype-phenotype correlation (199, 219, 220). The analyses below thus considered cases as a single group, except to test for differential effects on p14 and p16 activity.

3.3.2 *CDKN2A* mutation carriers display hyperinsulinemia in OGTTs

After excluding relevant individuals (*Section 3.2.1*), the remaining thirty-one cases and thirty-one controls were well-matched as groups for age ($p = 0.99$), gender ($p = 0.43$) and BMI ($p = 0.97$) (**Table 3**). To test the hypothesis that *CDKN2A*-loss leads to improved beta cell function, Dr Aparna Pal (OCDEM) and Dr Thomas Potjer (Leiden) performed 120-min oral glucose tolerance tests (OGTT) on all subjects. Using these data, I derived standard indices of beta cell function, insulin sensitivity, and hepatic clearance. While no difference in glucose levels was detected, insulin levels were significantly elevated in carriers throughout the test ($p<0.05$ for t_0-t_{100}, $p=0.052$ for t_{120}) (**Figure 2**). As a result, the total output of insulin secretion was 55% higher in carriers compared with controls ($p = 0.01$ for $AUC_{insulin}$; **Table 4**).

	Mutation carriers	Non-carriers	P-value
n	31	31	*NA*
BMI / [cm/kg^2]	27.1 [19; 38]	27.1 [19; 36]	0.97
Age / [yrs]	51.8 [21; 71]	51.8 [25; 84]	0.99
Gender / [% male]	45	32	0.43
Birth weight / [g] (n = 11)	3470 [2090; 4730]	NA	NA
Birth length / [cm] (n = 8)	50 [46; 53]	NA	NA

Table 3 *Baseline characteristics of study participants.* Data are given as mean and range [min; max].

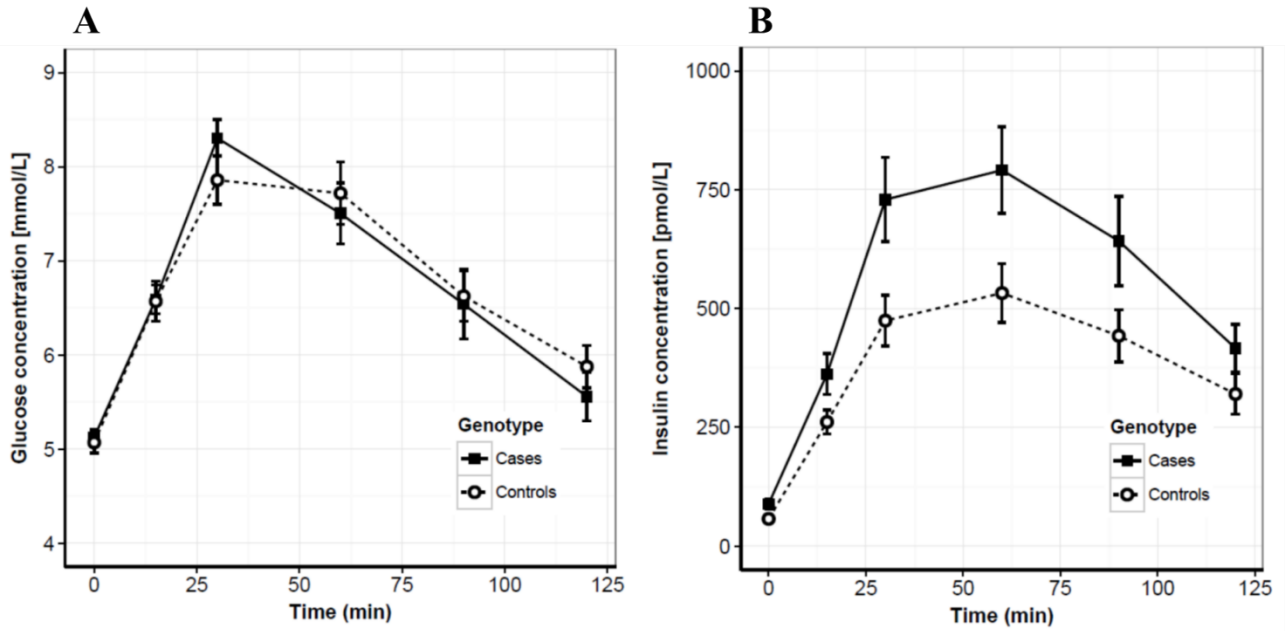


Figure 2 *Results of an OGTT in FAMMM individuals and controls.* Serum glucose (A) and insulin (B) levels during a 120-min OGTT in thirty-one carriers (*black squares, solid line*) and thirty-one controls (*white circles, dashed line*). Data shown as mean +/- SEM.

	Mutation carriers	Non-carriers	P-value
Fasting glucose [mmol/L]	5.2 [4.3; 6.3]	5.1 [3.2; 6.4]	0.65
Fasting insulin [pmol/L]	87 [15; 337]	55 [22; 150]	0.01
Fasting C-peptide [nmol/L]	0.44 [0.24; 0.81]	0.39 [0.20; 0.60]	0.48
iHOMA-B	124 [38.1; 452.4]	96 [45; 236.7]	0.05
iHOMA-S	91 [18; 328.7]	120 [35; 235.6]	0.05
BIGTT-AIR [$*10^3$]	6.4 [0.9; 28]	3.0 [1.2; 12]	0.03
BIGTT-S	5.8 [0.4; 12.8]	7.8 [1.1; 17.8]	0.04
Belfiore ISI	0.78 [0.17; 1.35]	0.97 [0.35; 1.77]	0.02
Matsuda ISI	4.3 [0.8; 11.1]	6.3 [1.5; 20.9]	0.02
$AUC_{glucose}$	839 [563; 1449]	829 [502; 1086]	0.79
$AUC_{insulin}$ [$*10^4$]	7.3 [2.4; 25]	4.7 [1.1; 15]	0.01
$AUC_{C\text{-}Peptide}$	212 [106; 333]	212 [115; 260]	1.00
Insulinogenic index	203 [39; 561]	152 [53; 360]	0.08
C-peptidogenic index	0.45 [0.15; 1.12]	0.45 [0.17; 1.78]	1.00
Disposition index	2.3 [1.1; 3.8]	2.3 [1.0; 3.7]	0.98
Fasting insulin clearance	0.66 [0.12; 1.27]	0.88 [0.65; 1.29]	0.07
Insulin clearance	0.35 [0.16; 0.51]	0.56 [0.32; 1.05]	0.03

Table 4 *OGTT-derived measures of beta cell function, insulin sensitivity and hepatic clearance.* Data are given as mean and range [min; max]. All indices based on C-peptide measurements are based on data from a subset of individuals only (n = 12 carriers and n = 12 controls; OX_001-012). Details on definitions of physiological measures are listed in Table 1.

To explore whether the hyperinsulinemia of mutation carriers was present already during foetal development, birth weight (n = 11) and birth length (n = 8) were obtained for a subset of the Dutch carriers in the cohort. Interestingly, both measures were found to closely match the historic population average for Dutch Caucasian new-borns, suggestive of normal insulin levels in the final trimester of gestation (**Table 3**) (221). As discussed in the general introduction (*Section 1.1*), elevated insulin levels in normoglycemic individuals is an indicator of reduced insulin sensitivity. Accordingly, both dynamic and fasting measures of insulin sensitivity, the BIGTT-S and HOMA-S (homeostatic model assessment of insulin sensitivity), were found to be lower in carriers ($p = 0.04$ and $p = 0.05$, respectively). Other standard measures, the Belfiore and Matsuda ISIs, confirmed the observed reduction in insulin sensitivity of carriers ($p = 0.02$ and $p = 0.02$, respectively).

Deriving equivalent measures for beta cell function revealed these to be increased in carriers ($p = 0.03$ for BIGTT-AIR; $p = 0.05$ for HOMA-B [homeostatic model assessment of beta cell function]). The insulinogenic index, a standard measure of the beta cell response to a glucose challenge, was 31% higher in carriers, though the difference was not significant ($p=0.08$). As a result, the disposition index, which is an aggregate measure of beta cell function relative to insulin sensitivity, remained unaffected compared with controls ($p = 0.98$). Further, these results were not significantly altered by excluding the three carriers that had presented with melanoma within two years prior to the study (**Table 5**).

	Mutation carriers	Non-carriers	P-value
n	28	31	*NA*
BMI [cm/kg^2]	27 [19, 38]	27 [19, 37]	0.66
Age [yrs]	51 [21, 71]	52 [25, 84]	0.96
Gender [% male]	43	33	0.58
Fasting glucose [mmol/L]	5.1 [4.3, 6.3]	5.1 [3.2, 6.4]	0.92
Fasting insulin [pmol/L]	91 [15, 337]	57 [22, 150]	0.01
iHOMA-B	133 [38, 452]	99 [45, 237]	0.03
iHOMA-S	86 [18, 329]	116 [35, 236]	0.04
BIGTT-AIR [*10^3]	6.6 [0.9, 28]	3.2 [1.2, 12]	0.03
BIGTT-S	5.8 [0.4, 12.7]	7.5 [1.1, 17.8]	0.08
Belfiore ISI	0.80 [0.17, 1.35]	0.96 [0.35, 1.77]	0.03
Matsuda ISI	4.3 [0.8; 11.1]	6.1 [1.5; 20.9]	0.02
$AUC_{glucose}$	831 [563, 1449]	832 [502, 1086]	0.98
$AUC_{Insulin}$ [*10^4]	7.2 [2.4, 25]	4.9 [1.1, 15]	0.02
Insulinogenic index	205 [39, 561]	155 [53, 360]	0.10
Disposition index	2.3 [1.1, 3.8]	2.3 [1.0, 3.7]	0.92

Table 5 *Comparison of OGTT-derived measures for carriers and non-carriers after exclusion of three subjects that had presented with cancer within two years prior to the study* (Table 2). Data are given as mean and range [min; max]. Details on definitions of physiological measures are listed in Table 1.

To explore whether the observed hyperinsulinemia phenotype was driven by underlying effects on p16, p14, or both, I re-analysed the data grouping carriers by mutation status (**Figure 3**). Of the mutations identified, 26 affected both p16 and p14, while five were located in regions affecting p16 exclusively. No significant differences were observed between these two groups in either insulin or glucose levels ($p = 1.00$ for $AUC_{insulin}$ and $p = 0.49$ for $AUC_{glucose}$, respectively). Though the analysis is likely underpowered for detecting small effects, the overall trends were qualitatively similarly for the two groups, with both displaying elevated insulin levels compared with controls. The observations therefore suggest that the observed phenotype of mutation carriers is unlikely to be driven primarily through effects on p14.

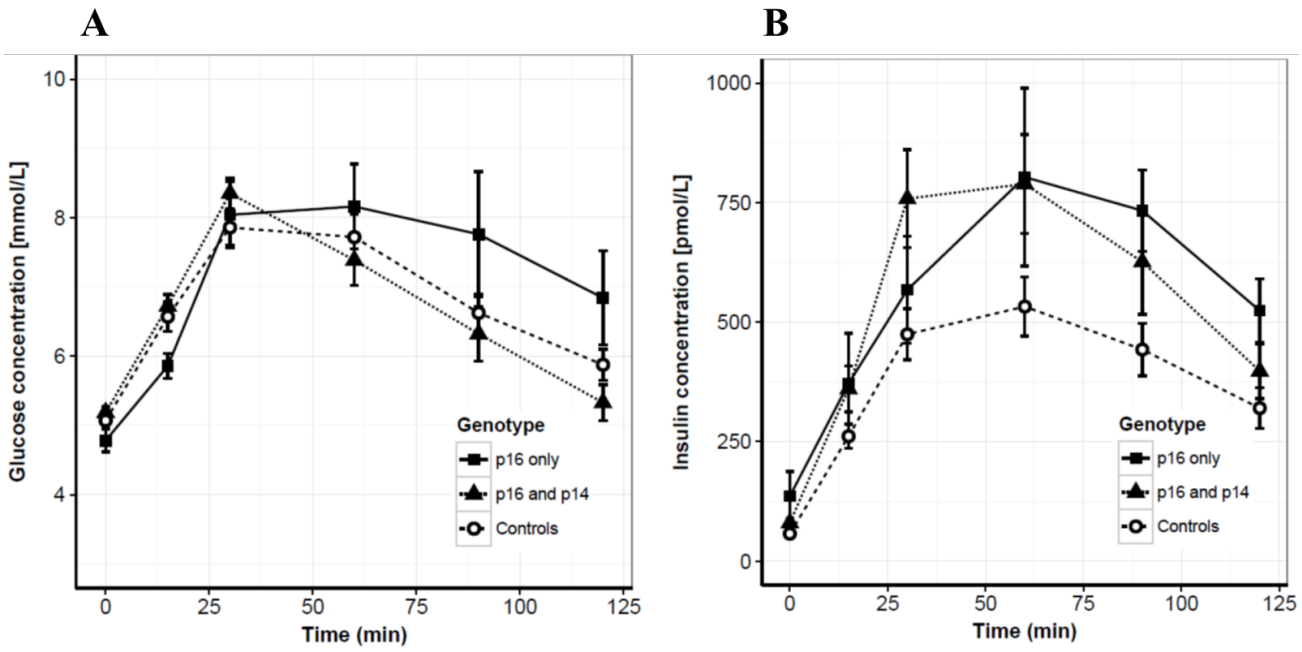


Figure 3 *Differential effects of* CDKN2A *mutations in an OGTT.* Serum glucose (A) and insulin (B) levels during a 120-min OGTT in twenty-six cases with *CDKN2A* LOF mutations affecting both $p16^{INK4a}$ and $p14^{ARF}$ (*black triangles, dotted line*), five cases with *CDKN2A* LOF mutations affecting $p16^{INK4a}$ exclusively (*black squares, solid line*), and thirty-one BMI-, age-, and gender-matched controls (*white circles, dashed line*). Data shown as mean +/- SEM.

OGTT-based indices of beta cell function derived from serum insulin measurements can be confounded by underlying defects in hepatic or peripheral insulin clearance. Since the liver clears a large proportion of insulin during first-pass transit, an apparent increase in first-phase insulin secretion (usually taken to be an indicator of elevated beta cell function) can equally be caused by reduced insulin extraction (222). In contrast, C-peptide (a by-product of proinsulin cleavage) is secreted in equimolar amounts by the pancreatic beta cells, and is not cleared by the liver and peripheral tissues by the same mechanism. Circulating C-peptide levels can thus be used as a more direct indicator of beta cell secretion without the confounding effects of insulin clearance.

For a subset of participants (all Oxford cases and controls) C-peptide measurements were obtained during the OGTT. Despite a tendency towards increased C-peptide levels in the fasted state (+12.8%; $p = 0.48$), the total response was not different for this subset of individuals ($p = 1.00$ for AUC; **Table 4**). Moreover, the C-peptidogenic index (analogous to the insulinogenic index) was unaffected in carriers ($p = 1.00$). Indices of insulin clearance (hepatic and/or peripheral), derived from the ratio between C-peptide and insulin levels, however, showed significantly decreased elimination in mutation carriers (-37.5%, $p = 0.03$ for total clearance) (223). Fasting clearance was also lower, but the observed 25% reduction did not reach significance in the subset of individuals for which C-peptide measurements were available ($p = 0.07$).

3.3.3 IVGTT reveals underlying defects in metabolic processes

The OGTT is a convenient method for generating relatively crude glycaemic measures in a scalable manner. It was selected as the default test for this study due to the benefit it confers in offering the ability to perform a simple glucose tolerance test in the homes of participants. Given the very limited number of *CDKN2A* mutation carriers available in the UK and internationally, this offered an important advantage in terms of facilitating recruitment. However, OGTTs generate less well-validated glycaemic indices than clamp-based techniques. This is partly due to the additional sources of variation introduced by delivering glucose orally (e.g. differences in absorption rate, transit time, incretin effect etc.) and partly a result of the less frequent sampling scheme typically employed, making modelling of the data more challenging.

The IVGTT is a more time-consuming method that can be used to generate glycaemic indices that are better validated against the hyperinsulinemic euglycemic glucose clamp. Two gold-standard measures of beta cell function are the acute insulin response to glucose (IVGTT $AUC_{insulin}$ 0-10 min), which reflects functional beta cell mass and the sensitivity of insulin secretion to glucose,

and the IVGTT disposition index, which quantifies how well beta cells have adapted to variations in insulin sensitivity. To corroborate and extend the findings derived from the OGTT, IVGTTs were performed by Dr Pal on eight cases and eight controls available to come in for follow-up studies at OCDEM (**Figure 4**). I then analysed the data in collaboration with Dr Ian Godsland (Imperial College London) using a minimal model approach, according to an algorithm designed to maximise precision and identification success rate (208).

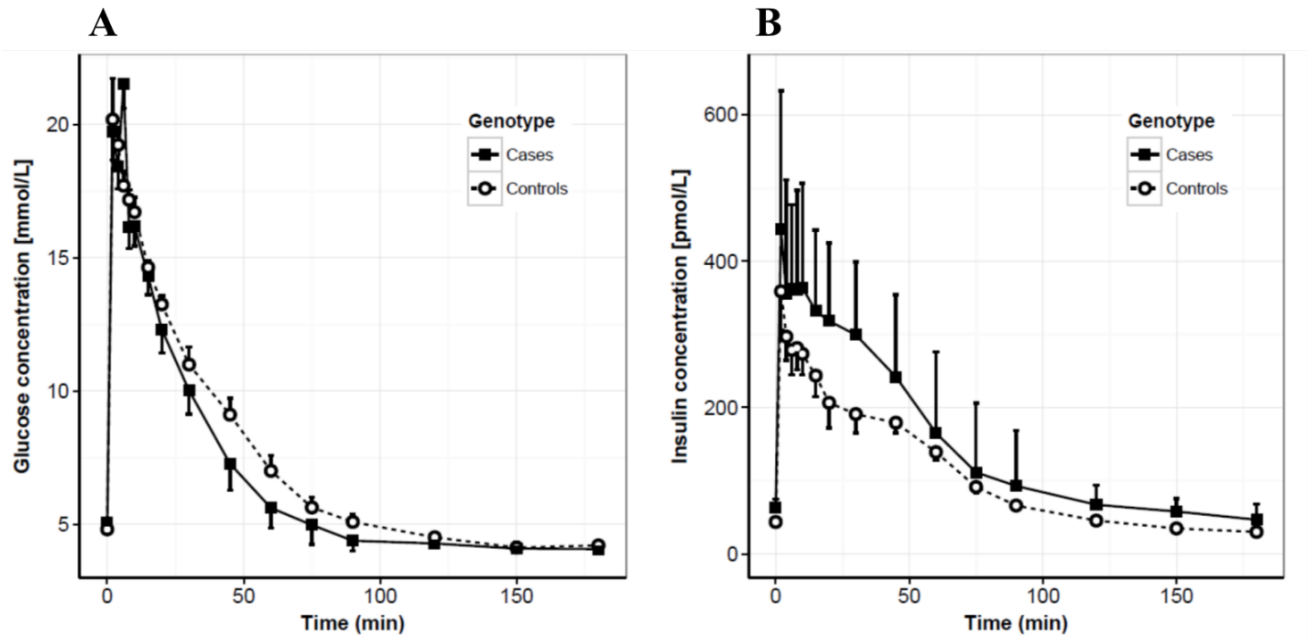


Figure 4 *Results of an IVGTT in FAMMM individuals and controls.* Serum glucose (A) and insulin (B) levels during a 180-min IVGTT in eight carriers (*pink circles*) and eight BMI- (p = 0.72), age- (p = 0.96), and gender-matched (p = 0.50) controls (*blue squares*). Data shown as mean +/- SEM.

None of the indices derived from this test reached statistical significance (**Table 6**). Between 44 and 82 individuals in each group would have been needed to establish significant differences at $p<0.05$ and 90% power based on the observed differences in $AUC_{insulin}$(0-10 min) and disposition index. Considering instead the magnitude and direction of effects in the context of the OGTT results confirms broadly similar trends. Hepatic insulin throughput (inversely proportional to clearance) was increased by 30% ($p = 0.23$), and total insulin output was found to be 66 % and 110 % higher during the first and second phase of secretion, respectively ($p = 0.51$ for $AUC_{insulin}$ 0-10 min; $p = 0.16$ for $AUC_{insulin}$ 10-180 min). The IVGTT-derived insulin sensitivity was slightly higher in mutation carriers (+5.9%, p=0.72), but this index reflects primarily peripheral insulin sensitivity. In contrast, hepatic insulin clearance is a key determinant of OGTT-derived indices of insulin sensitivity.

Interestingly, the C-peptide output was around 30 % higher during both phases of secretion ($p = 0.96$ for $AUC_{C\text{-}peptide}$ 0-10 min; $p = 1.00$ for $AUC_{C\text{-}peptide}$ 10-180 min), indicating a direct contribution of improved beta cell function to the elevated circulating insulin levels of carriers. As a result, both the insulin and C-peptide disposition indices, which reflect the ability of beta cells to adapt to accompanying insulin resistance, were elevated in mutation carriers (+26.3%, $p = 0.51$ for insulin; +14.1%, $p = 0.80$ for C-peptide). Overall, the OGTT and IVGTT results establish that *CDKN2A* mutations cause a metabolic phenotype that manifests predominately as elevated insulin secretion. Both tests are consistent with a contribution of reduced hepatic insulin elimination, while the IVGTT data further suggest that beta cell function is elevated in mutation carriers.

	Mutation carriers	Non-carriers	P value
n	8	8	*NA*
$AUC_{insulin}$ 0-10 min (pmol.L^{-1}.min*10^{-3})	4.8 [1.73, 13.2]	2.9 [1.84, 4.28]	0.51
$AUC_{insulin}$ 10-180 min (pmol.L^{-1}.min*10^{-3})	35.5 [13.7, 94.9]	16.9 [12.5, 22.2]	0.16
$AUC_{C\text{-}peptide}$ 0-10 min (pmol.L^{-1}.min*10^{-3})	15.1 [9.1, 31.9]	11.7 [5.5, 14.5]	0.96
$AUC_{C\text{-}peptide}$ 10-180 min (pmol.L^{-1}.min*10^{-3})	198 [103, 364]	152 [101, 186]	1.00
Net IVGTT insulin secretion 0-10 min (pmol.L^{-1})	0.56 [0.29, 0.84]	0.50 [0.10, 1.04]	0.65
Net IVGTT insulin secretion 10-180 min (pmol.L^{-1})	4.2 [1.20, 11.8]	2.8 [1.10, 4.98]	0.65
Insulin sensitivity, S_I (min^{-1}.$pmol^{-1}$.L)	0.72 [0.19, 1.28]	0.68 [0.30, 0.95]	0.72
Disposition index [*10^{-3}]	2.4 [1.14, 4.44]	1.9 [0.77, 3.40]	0.51
C-peptide disposition index	8.9 [3.45, 12.6]	7.8 [4.00, 13.7]	0.80
Fractional hepatic insulin throughput	0.53 [0.34, 1.12]	0.41 [0.26, 0.74]	0.23
Plasma insulin elimination rate (min^{-1})	0.098 [0.025, 0.207]	0.095 [0.042, 0.235]	0.72

Table 6 *IVGTT-derived measures of beta cell function, insulin sensitivity and hepatic clearance.* Data are given as mean and range [min; max]. All p-values are based on Mann-Whitney U test.

3.3.4 Cell-cycle independent effects of *CDKN2A* deficiency in EndoC-βH1

Finally, I sought to establish the extent to which cell-cycle independent effects of *CDKN2A* on the regulation of insulin secretion could contribute to the phenotype of mutation carriers. Recent work in rodent hepatocyte models has suggested a role of p16 in the regulation of PKA signalling (194). Given the well-characterised effects of PKA on potentiation of insulin secretion (*Section 1.5.1*), I speculated that such signalling events could have a direct impact on beta cell function. To test this hypothesis, I performed knockdown and secretion studies in the human pancreatic beta cell line, EndoC-βH1. As previously discussed, this cell line was transformed by Ravassard et al using the proto-oncogene SV40LT, which acts on the Retinoblastoma protein, thereby masking any effects of p16 on cell-cycle control (157).

I first confirmed expression of p16 by immunofluorescence and found that, consistent with previous reports, the protein localized to both the nucleus and cytoplasm (**Figure 5A**) (224, 225). siRNA-mediated silencing of *CDKN2A* was then performed and efficient knockdown (>80%) could be observed both at the mRNA and protein levels 72 hours after gene silencing **Figure 5**). Transcript levels of the nearby gene *CDKN2B*, which shares a high level of sequence homology, was unaffected by *CDKN2A* knockdown (**Figure 5B**)

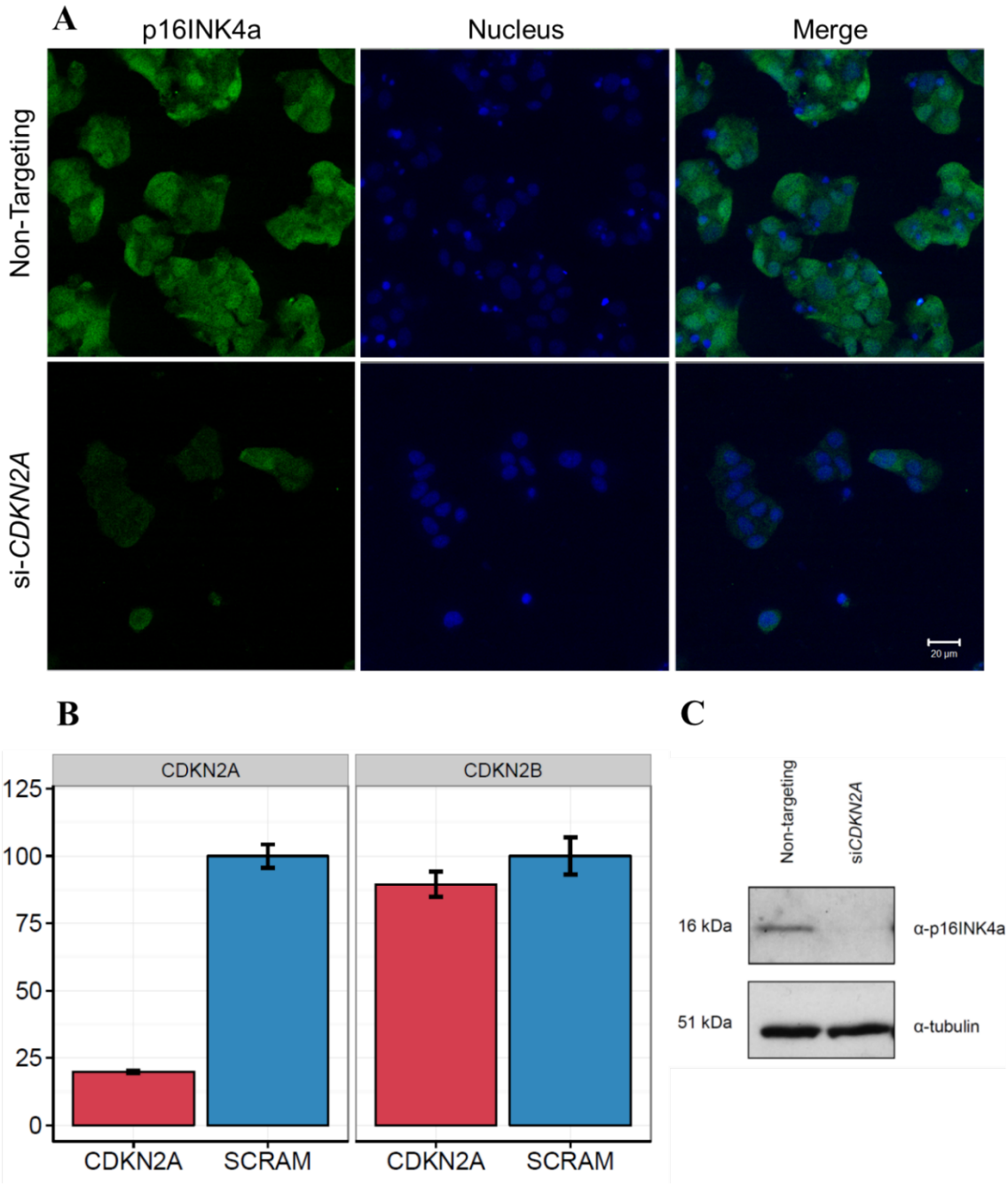


Figure 5 CDKN2A *knockdown in the human beta cell line EndoC-βH1*. Seventy-two hours after silencing of *CDKN2A* (*red bars*) and non-targeting sequence ("Scram"; *blue bars*), knockdown was confirmed both at the mRNA level (B) and at the protein level (A & C; isoform 1 of p16). (A) Scale bar, 20 µm. (B) Data points for transcript expression are means of n = 3, and error bars are SEM. (C) Representative blot of three independent experiments.

As expected, no effect on proliferation was detected when measured over 9 days, consistent with the EndoC-βH1 cells having deregulated cell-cycle control independent of p16 (**Figure 6A**). After 96 h, cells treated with either *CDKN2A* or non-targeting siRNAs were cultured under different conditions to assess the glucose-responsiveness of the cells. In addition to basal and high-glucose media, the effect of exogenous stimulation with the non-selective phosphodiesterase inhibitor (IBMX) was tested. IBMX activates PKA through accumulation of intracellular cAMP (*Section 1.5.1*) and has been shown to potentiate the effects of p16 silencing in a hepatocyte model (194). Across all three conditions, *CDKN2A* knockdown was found to increase insulin secretion as a fraction of total content (basal, $p = 0.02$; high, $p = 0.01$; high glucose with IBMX, $p = 0.04$; **Figure 6C**). A small, but significant reduction in the total insulin content was also observed ($p < 0.01$; **Figure 6B**).

To isolate the effect of *CDKN2A* silencing on glucose responsiveness, I next calculated the fold induction (ratio of stimulated to basal secretion). This identified a trend towards an enhanced response to IBMX stimulation for *CDKN2A* siRNA-treated cells ($p = 0.08$), suggestive of a synergistic effect on PKA activity (**Figure 6D**). To more specifically assess a role for *CDKN2A* in the potentiating pathway of insulin secretion, I lastly performed PKA activity assays following knockdown. Consistent with the trend towards increased glucose-responsiveness, this revealed a corresponding 23 % increase in the activity of PKA following *CDKN2A* knockdown ($p = 0.02$; **Figure 7**).

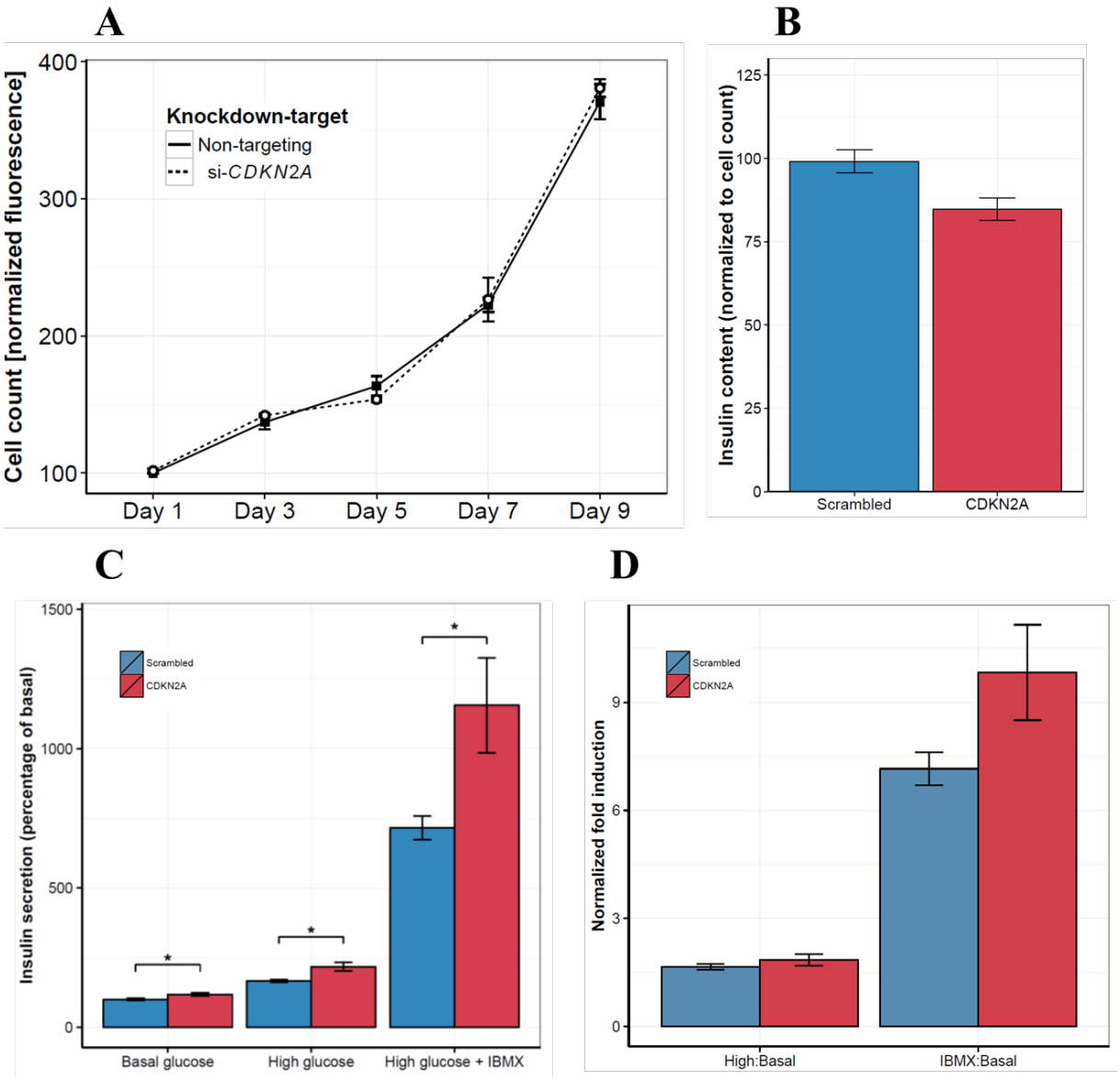


Figure 6 CDKN2A *knockdown in the human beta cell line, EndoC-βH1.* Cells were treated as described in figure 5, followed by measurement of (A) cellular proliferation for si-*CDKN2A* (*white circles, dashed line*) and non-targeting siRNA ("Scrambled"; *black squares, solid)* normalized to day 1. (B-D) After 96 h, static insulin secretion assays were performed under culturing conditions of basal glucose (2.8 mM glucose), high glucose (20 mM) and high glucose with IBMX (100 µM). All secretion results were normalized to total insulin content per well. Bars represent means for n = 11 (si-*CDKN2A*) and 18 (non-targeting) generated in three independent experiments. Data points for cellular proliferation are means for n = 3, and error bars are SEM.

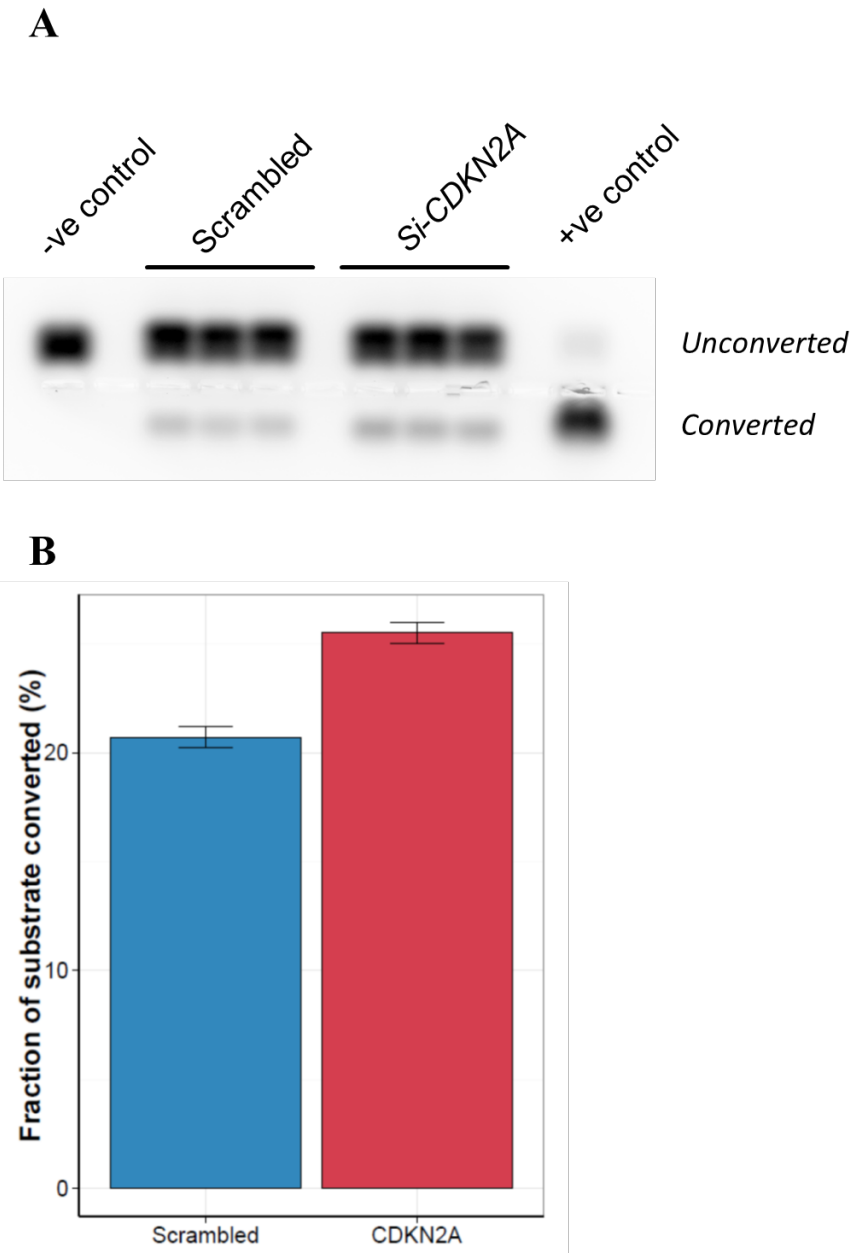


Figure 7 *PKA activity in the EndoC-βH1 cell line following* CDKN2A *knockdown.* Ninety-six hours after treatment with non-targeting ("scrambled"; *blue bar*) or *CDKN2A* (*red bar*) siRNAs, cells were harvested and PKA activity measured on sample input normalized to cell numbers (n = 3). (A) The blot shows replicates generated in the same experiment, and is representative of three independent experiments. (B) Substrate conversion rates were calculated as the ratio of non-phosphorylated to total peptide using RFU quantified with standard software on the ChemiDoc MP system. Data shown as mean +/- SEM.

3.4 Discussion

Individuals carrying heterozygous LOF mutations in the *CDKN2A* gene provide a unique opportunity to study the role of p16 and p14 in glucose homeostasis in humans. Through the analyses of OGTTs and IVGTTs, I have demonstrated that carriers display significantly increased insulin levels compared with matched controls. Further, a contribution of both decreased hepatic insulin clearance and increased beta cell function to the observed hyperinsulinemia could be established. Specifically, indices of beta cell function that reflect the sensitivity of insulin secretion to glucose (iHOMA-B, BIGTT-AIR, insulin and C-peptidogenic indices) or the ability of the beta cells to adapt to accompanying insulin resistance (IVGTT disposition indices), were increased in cases. OGTT and IVGTT-based indices of insulin sensitivity and elimination were all consistent with reduced hepatic insulin extraction in mutation carriers.

Overall, these results propose a combination of two, non-mutually exclusive mechanisms contributing to the phenotype of mutation carriers: *(a)* primary beta cell hyperfunction (increased beta cell mass and/or secretion) driving progressive insulin resistance; and *(b)* primary insulin resistance triggering a compensatory increase in insulin secretion (**Figure 8**). While both explanations are consistent with our data, existing evidence strongly favour a role for *CDKN2A* in beta cell function (191). Further, the observed decrease in hepatic insulin clearance suggests a reduction in the number or affinity of insulin receptors in the liver. Chronic hyperinsulinemia is known to cause a gradual down-regulation both of insulin receptors and post-receptor signalling efficiency, thereby driving general insulin resistance (226). This, in turn, would reduce the compensatory burden placed on the beta cells, such that C-peptide levels would only be moderately elevated.

A similar phenotype was reported recently for patients suffering from long QT syndrome due to *KCNQ1* mutations (227). Although the ion-channel mutations were presumed to cause beta cell dysfunction, carriers showed the largest relative difference in insulin levels and only a modest effect on C-peptide levels. The observed phenotype of *CDKN2A* mutation-carriers in our study is therefore in agreement with the expected physiological adaption to chronic hyperinsulinemia.

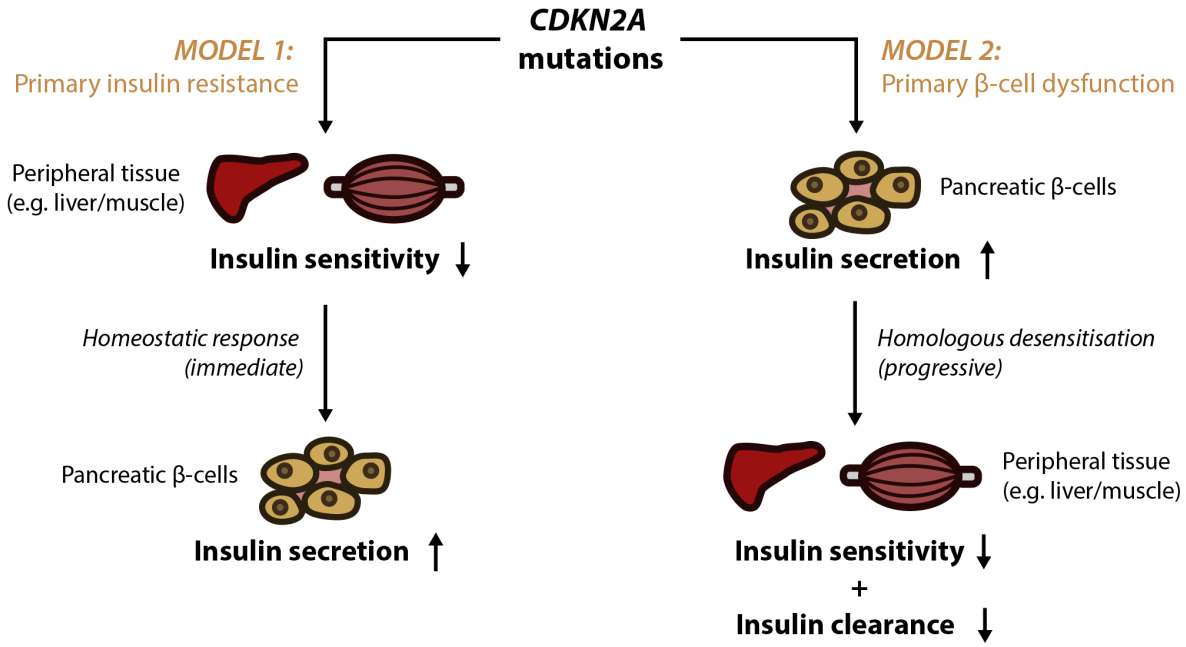


Figure 8 *Outline of two non-mutually exclusive mechanisms compatible with the phenotype of CDKN2A mutations-carriers.* According to model 1 (*left*), a primary reduction in insulin sensitivity leads to a compensatory increase in insulin secretion to maintain glucose homeostasis. Model 2 (*right*) represents another possible contributing mechanism: chronically elevated insulin levels due to beta cell hypersecretion drives a progressive decrease in insulin receptors and insulin signalling through homologous desensitization. In this case, the primary effect of *CDKN2A* loss is on the pancreatic beta cell, but the mechanism ultimately manifests as impaired insulin sensitivity and reduced insulin clearance. As discussed in the main text, it is possible that a combination of the two models contributes to both beta cell hypersecretion and primary insulin resistance.

However, due to limitations on the design of the clinical study, I cannot conclusively address the cause and effect between hyperinsulinemia and insulin resistance in mutation carriers. The underlying changes appear complex, and unequivocal discrimination would require appreciably larger group sizes. Power calculations suggest that an impractically high number of 40-80 individuals would be required to establish significant differences for IVGTT derived measures of beta cell function. It also remains possible that the apparent insulin resistance observed in this study could be an artefact of sampling non-diabetic subjects, leading to preferential exclusion of controls with low insulin sensitivity. Such a truncation effect would be a consequence of the fact that mutation carriers remain normoglycemic even at comparatively high levels of insulin resistance due to improved beta cell function.

To test for a cell-cycle independent role of *CDKN2A* in the regulation of insulin secretion, I performed knockdown studies in the human beta cell line, EndoC-βH1. This identified cell-cycle independent effects on insulin secretion and content. These changes were found to be accompanied by increased PKA activity, in agreement with previous studies establishing such an effect of *CDKN2A* knockdown in liver (193, 194). Notably, Bantubungi et al reported two related effects in a hepatocyte model: firstly, that the effect of *CDKN2A* knockdown on PKA activity is independent of cAMP levels; and secondly, that Forskolin (which, like IBMX, enhances cAMP levels) has a synergistic effect on the cell-cycle independent role of p16 (194). This highlights a possible contribution of the PKA-dependent potentiating pathway to the secretory effects observed in the EndoC-βH1 cell line. On the contrary, a recent study showed that *CDKN2A* deficiency had a strong negative impact on insulin secretion in the related cellular model EndoC-βH2 (without deregulated cell cycle control) (*Section 1.5.2*). While this convincingly showed a key role for p16 in triggering senescence in human beta cells, the study did not explore cell-cycle-independent effects.

The net effect of *CDKN2A* LOF mutations *in vivo* is likely to arise out of a complex interplay between both cell-cycle independent and dependent roles of p16 and p14 (191, 194). Some of these manifest instantly (including direct effects on insulin secretion), while others develop gradually (such as increased proliferation leading to enhanced beta cell mass). Extra-pancreatic effects are also likely to contribute, with emerging roles of *CDKN2A* in other tissues, such as liver, adipose and muscle tissue (174). The absence of an effect on birth weight and birth length in the cohort studied here suggests that the effects of *CDKN2A* deficiency manifest postnatally, consistent with observations in rodent knockout models.

Upstream of *CDKN2A*, several independent association signals for T2D risk have been identified. The underlying effector transcript(s) and disease mechanism have remained elusive, and prior studies have not reported any cis-eQTL effects for these alleles (175). My findings have shown that, similar to known effects of non-coding T2D risk alleles at the locus, coding *CDKN2A* mutations are associated with effects on beta cell function. This provides a plausible link between *CDKN2A* and the common GWAS alleles, and thus points to the gene as a likely effector transcript at this locus.

Interestingly, the T2D-associated variants at the *CDKN2A/B* locus have consistently been linked to a more 'classic' beta cell phenotype than that observed for carriers of coding mutations in this study, with no evidence for an impact on measures of insulin resistance (*Section 3.1.1*) (75, 173). I speculated that any (as yet unidentified) cis-regulatory effect exerted on *CDKN2A* could achieve a more restricted beta cell phenotype through tissue-specific regulation of gene expression. To address this hypothesis, I interrogated existing genomic annotations, and found that the non-coding risk variants map to a cluster of islet enhancer activity and open chromatin. Specifically, the credible sets for the association signals overlap a strong enrichment for islet- and melanocyte-specific FOXA2 binding (**Figure 9**) (73). This highlights a possible mechanism for the more specific beta cell phenotype caused by

common disease-associated variants compared with carriers of *CDKN2A* LOF variants. Based on the direction of effect on measures of beta cell function, the non-coding risk alleles would be predicted to increase expression of *CDKN2A* (75, 173). As previously discussed, no cis-eQTL effects have previously been reported in islets for this region, but larger studies currently underway may be able to shed further light on this hypothesis (119).

Taken together, the data presented here establish *CDKN2A* as an important regulator of glucose homeostasis in humans. I have shown that my results are consistent with *CDKN2A* deficiency affecting a number of tissues, including both islet beta cells and liver. This study thus supports the candidacy of *CDKN2A* as the effector transcript of the T2D-associated alleles in the region, and I have proposed a mechanism to account for the apparent tissue-specificity of the beta cell dysfunction caused by diabetes risk alleles.

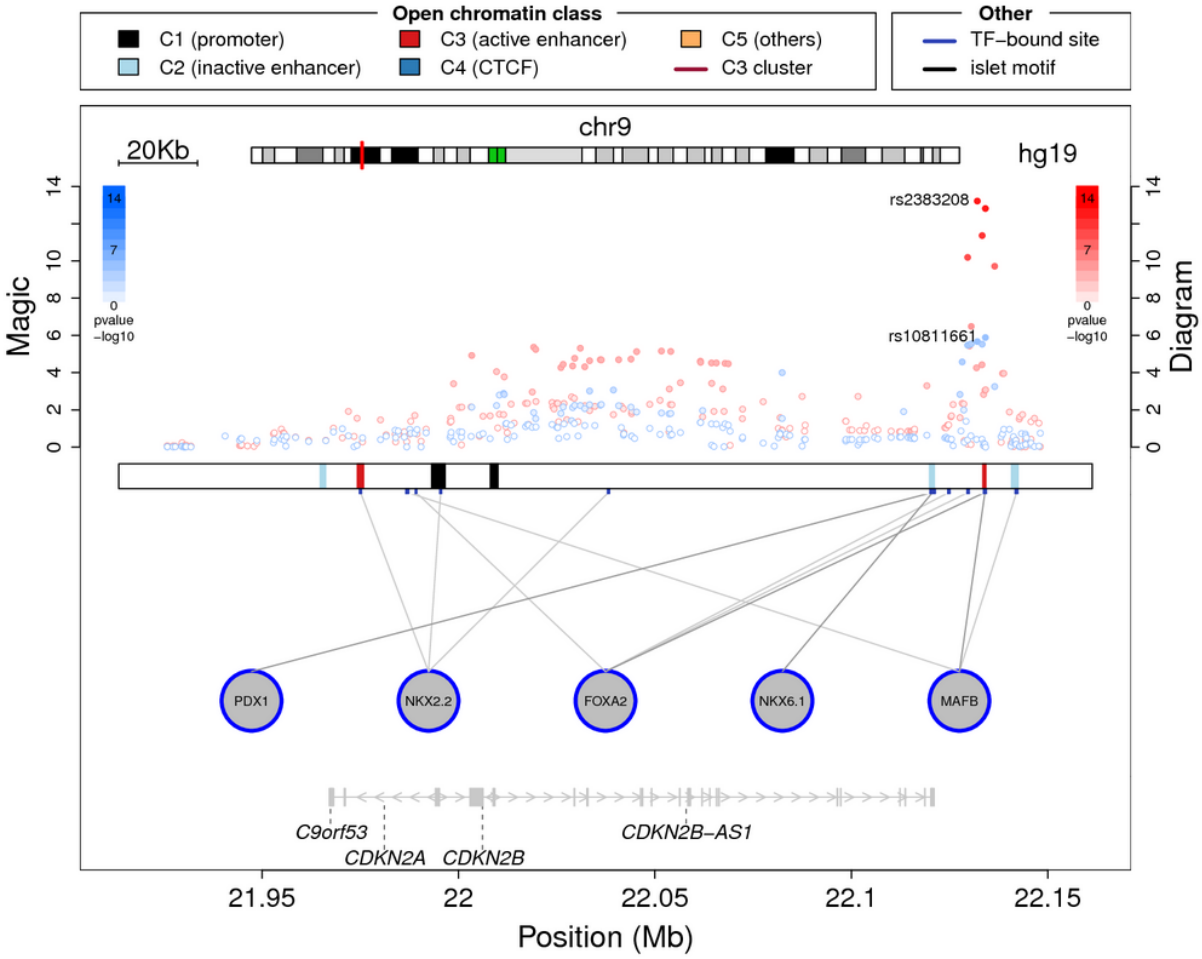


Figure 9 *Regional association plot for the CDKN2A/B locus.* Screenshot from the Human Islet Regulome Browser (73) showing annotations of chromatin state, transcription factor binding and variant association with T2D at the *CDKN2A/B locus*. The genomic binding sites of islet transcription factors (PDX1, NKX2.2, FOXA2, NKX6.1, MAFB) are indicated by lines connected to the respective proteins. A Manhattan plot above shows log(p-values) of association with T2D for variants in the MAGIC (*blue*) and DIAGRAM (*red*) datasets. Lead variants tagging the two genome-wide association signals located downstream of *CDKN2B-AS1* are shown. Fine-mapping efforts have since publication of the islet regulome narrowed down the number of potential causal variants to credible sets of five and six variants (47). Both of these sets contain variants that directly overlap a FOXA2 binding site.

CHAPTER 4

Type 2 Diabetes Risk Alleles in *PAM* Cause Human Beta Cell Dysfunction

---

Some of results presented in the following chapter formed part of a preprint posted on the BioRxiv server, which is currently in review (*Nature Genetics*). The preprint is referenced below, and a copy of the manuscript is provided in the Appendices.

- A. Raimondo*, **S.K. Thomsen***, et al, Type 2 diabetes risk alleles reveal a role for peptidylglycine alpha-amidating monooxygenase in beta cell function. BioRxiv (2017):158642

*Joint first authors

4.1 Introduction

4.1.1 Exome-based association studies identify coding variants for T2D risk

Recent T2D association studies have shifted attention towards discovery of rarer, coding risk variants through exome-centric approaches and WGS (*Section 1.2.2*). An initial motivation was the relatively modest population-attributable risk explained by common variants identified in early GWAS (26, 27). To account for this "missing heritability" a number of possible sources were proposed, including lower frequency and rare variants with higher penetrance (28). Within this category, coding variants attracted particular attention because of their potential to have a relatively large impact on protein function (35).

The "missing heritability" of T2D is now thought to be largely explained by a long tail of common risk variants with increasingly smaller effect sizes (mostly unidentified at current genome-wide thresholds of significance) (36). Despite this, recent exome-based and WGS studies have successfully identified coding association signals, either by considering individual variants or by aggregating the burden of variants in specific genes (36-38). In the largest T2D fine-mapping study to date, 16 genes were identified as harbouring coding risk variant thought to be causal for the association signal (65). Even if their overall contribution to the heritability of T2D is modest, such coding signals offer the possibility of a more straightforward biological interpretation by directly pinpointing effector transcripts.

An added benefit of finding causal variants in coding regions is the offer of a simpler interpretation for therapeutic strategies. Non-coding variation is subject to the context-dependent activity of cis-regulation and the effects can be restricted to specific tissues or developmental stages (79). As a consequence, drugs that target the affected gene could cause unexpected adverse effects by producing a more global phenotype. Though coding variants can

also be subject to such context-dependency (e.g. through tissue-specific isoforms), the effector transcripts are often affected more widely (79, 82).

4.1.2 T2D risk variants in *PAM*

Among the first coding signals to emerge from recent exome-based association studies for T2D risk and intermediate traits, two independent risk variants were recently identified in the *Peptidylglycine Alpha-amidating Monooxygenase* (*PAM*) gene (37, 38). The associations (rs36046591, OR = 1.31; rs78408340, OR = 1.60) were identified and replicated in separate GWAS with case-control datasets from Iceland, Denmark, Finland and Iran. Though one *PAM* variant (rs35658696) was found to be in LD with a coding variant in a nearby gene (*PPIP5K2;* rs36046591), the presence of a second, independent coding variant in *PAM (*rs78408340) provides strong evidence in favour of the gene being causal. More recently, a large-scale fine-mapping effort in individuals of European ancestry also identified a gene-based association for *PAM,* driven primarily by the previously identified coding variant (rs35658696) (65).

Interestingly, both T2D risk alleles in *PAM* have been linked to reduced insulinogenic index in a manner consistent with their effect on T2D risk (rs36046591, β = -1.96; rs78408340, β = -8.42) (37, 38). As a measure of insulin secretion in response to an oral glucose challenge ($\frac{INS_{30}-INS_0}{GLU_{30}-GLU_0}$), the insulinogenic index is frequently used as an indicator of beta cell function (228). Though clamp-based tests can provide more precise indices of beta cell function, the insulinogenic index is a validated measure that can be conveniently assessed across large populations (229). In the case of *PAM*, the association of non-synonymous risk alleles with reduced insulinogenic index hints at a role of PAM in beta cell function, though this could be mediated through both local (islet-intrinsic) and/or extra-pancreatic (islet-extrinsic) mechanisms.

4.1.3 T2D-associated *PAM* alleles cause decreased PAM

PAM encodes peptidylglycine alpha-amidating monooxygenase (PAM), an enzyme in neuroendocrine cells known to modify peptides with a C-terminal glycine to create peptide-amides (230-232). Amidation can dramatically increase the biological potency of a peptide relative to its unmodified glycine-extended conjugate (233). *Pam*$^{-/-}$ mice are embryonic lethal, while *Pam*$^{+/-}$ mice display a mild glucose intolerance with an unknown molecular basis (234). Despite reports of PAM expression in pancreatic islets, a functional role in beta cells has not yet been described (235). Insulin itself is not a PAM substrate, so the effects of *PAM* on beta cell function are therefore mediated by other peptide(s).

The functional PAM enzyme exists in both integral membrane and luminal forms, the latter of which is co-secreted with endocrine peptide(s) (230, 232, 236) (**Figure 1**). Mechanistically, PAM is a bifunctional enzyme, possessing two contiguous catalytic domains: peptidylglycine alpha-hydroxylating monooxygenase (PHM), and peptidyl alpha-hydroxyglycine alpha-amidating lyase (PAL) (232). The p.S539W and p.D563G missense mutations are both located within the PAL domain and are predicted by *in silico* tools (SIFT, PolyPhen2) to be damaging, suggesting that they could affect enzymatic activity (**Table 1**).

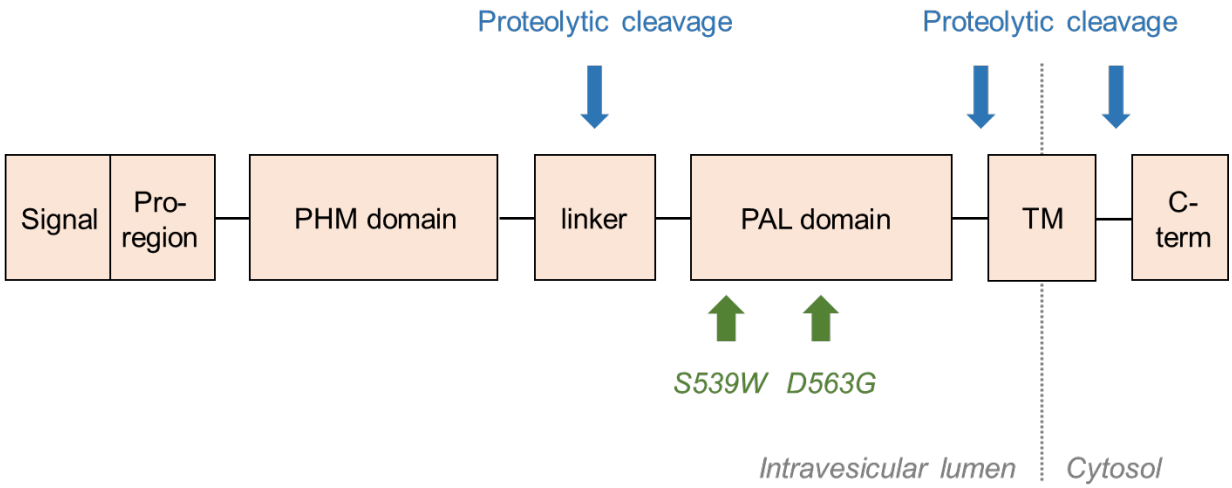


Figure 1 *Schematic of the domain structure of* PAM *protein.* From N-terminus to C-terminus the distinct domains include a Signal peptide (Signal), a proregion, the catalytic peptidylglycine alpha-hydroxylating monooxygenase (PHM) and peptidyl alpha-hydroxyglycine alpha-amidating lyase (PAL) domains separated by a linker, a transmembrane domain (TM) and a C-terminal domain that protrudes into the cytosol. The position of the two T2D risk variants (rs78408340, p.S539W; and rs35658696, p.D563G) are shown in green, and sites of post-translational cleavage are highlighted in blue.

Amino acid substitution	**SIFT Human SNPs**	**PolyPhen-2**	**Conservation h-m-r-e-x**
p.S539W	Damaging	Probably damaging	S-S-S-S-S
p.D563G	Damaging	Probably damaging	D-D-D-D-D

Table 1. *In silico prediction of pathogenicity and amino acid conservation for* PAM *missense variants*. The variants were assessed using the SIFT Human Single Nucleotide Polymorphisms (SNPs) and PolyPhen v.2.2.2 web server algorithms. H, Homo sapiens; M, Mus musculus; R, Rattus norvegicus; E, Equus caballus; X, Xenopus laevis.

To test this, Dr Anne Raimondo (Gloyn lab, OCDEM) previously generated Human Embryonic Kidney (HEK) 293 cell lines expressing recombinant secreted (non-integral membrane) PAM protein for *in vitro* amidation assays. Wild type (WT)-PAM and D563G-PAM were robustly produced, as well as an additional catalytically inactive mutant protein, Y651F-PAM, that was used as a positive control for reduced activity (237). Interestingly, S539W-PAM expression could not be detected across three independently derived cell lines, and this was not due to cellular retention of S539W-PAM (**Figure 2A**).

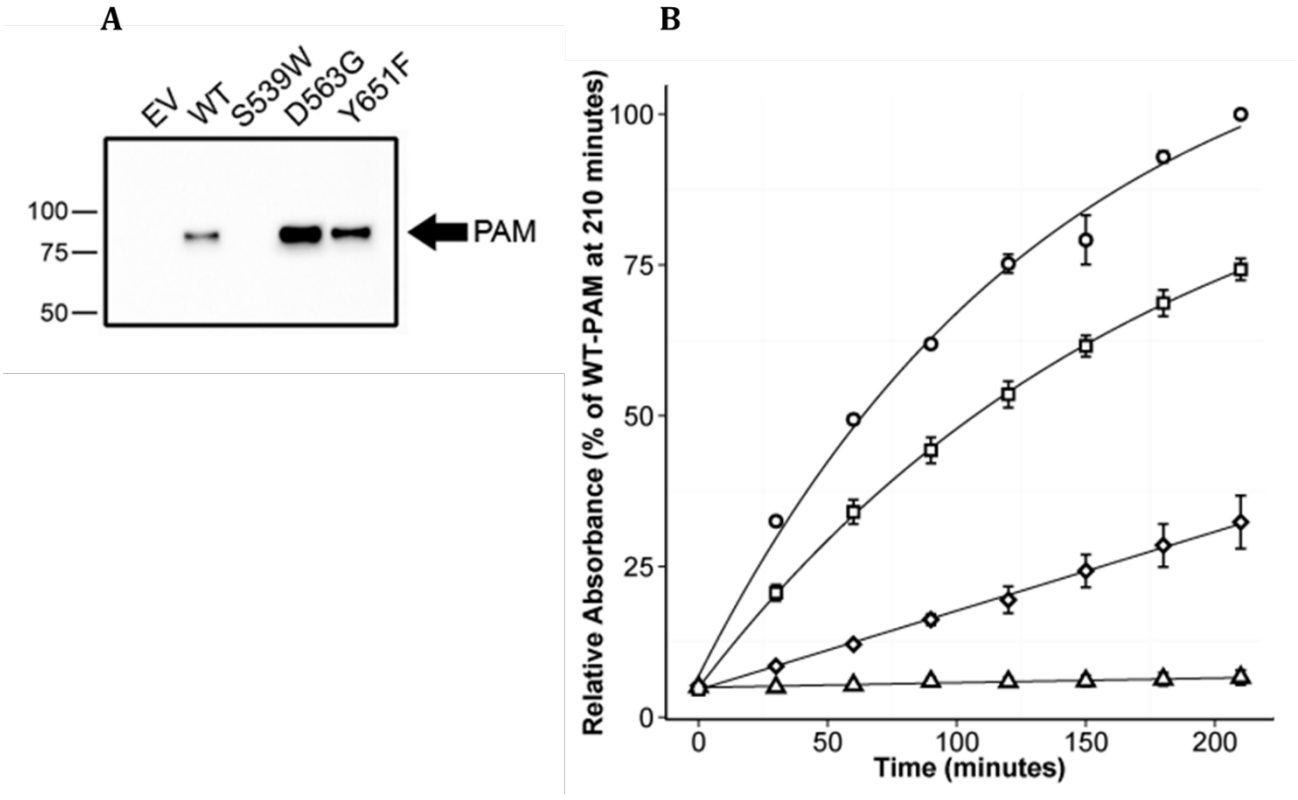


Figure 2. *Analysis of WT and variant PAM function by Dr Raimondo*. (A) Western analysis of recombinant PAM protein production in supernatant from HEK 293 stable cells. Representative of three independent experiments. (B) Amidating activity of WT-PAM (circles), D563G-PAM (squares), and Y651F-PAM (diamonds), and empty vector (triangles) supernatant towards the model substrate hippuric acid measured via spectrophotometric detection of chemically converted glyoxylate. Results shown are the average of n=4 experiments relative to WT-PAM at the final time point (100%). Error bars are mean ±SEM.

Dr Raimondo subsequently developed a cell-free kinetic assay capable of measuring PAM amidating activity via spectrophotometric detection of converted glyoxylate, a by-product of the amidation reaction (238, 239). Matching each reaction for PAM input, reduced amidating activity was observed for D563G-PAM ($p=1.0x10^{-5}$) and Y651F-PAM ($p=4.1x10^{-6}$) (**Figure 2B**). Further analysis showed no significant difference in substrate affinity between WT-PAM and D563G-PAM (K_m 0.95mmol/L vs 1.02mmol/L, p=0.44), suggesting that the p.D563G substitution affects K_{cat}. These results demonstrate that the T2D-associated *PAM* coding alleles decrease PAM function via a combination of defective expression and/or reduced catalysis.

To further elucidate the link between *PAM* variants and T2D risk, this study aimed to investigate a possible role of PAM in islet beta cell function. I first established that PAM has tissue expression and localisation patterns consistent with a role in islet function and insulin secretion (*Section 4.3.1*). I then investigated the impact of *PAM* deficiency on a range of cellular phenotypes in EndoC-βH1, and uncovered defects in insulin content, secretion, and the dynamics of exocytosis (*Sections 4.3.2-4.3.3*). Informed by these results, I confirmed a similar effect of *PAM* genotype status on exocytosis in the largest cohort of primary human islets with functional data available (*Section 4.3.4*). Finally, I explored possible mechanisms by which the effects of *PAM* on beta cell function could be mediated (*Sections 4.3.5-4.3.6*).

4.2 Materials and methods

Materials and methods for experiments discussed in the introduction (*Section 4.1.3*) are provided in Appendix 1, which covers cloning of *PAM* expression plasmids, recombinant PAM protein production, quantitation of recombinant PAM, kinetic analysis of PAM amidating activity, and kinetic analysis of variant PAM substrate affinity.

4.2.1 Gene expression analyses

Gene expression was measured as described in Section 2.4 using the following TaqMan assays: *PAM* (Hs00168596_m1), *CHGA* (Hs00900370_m1), *INS* (Hs00355773_m1), *IAPP* (Hs00169095_m1). RNA was purified by TRIzol extraction (*Section 2.4.1*) or sourced from commercially available RNA tissue panels: Human Total RNA Master Panel II (Clontech), FirstChoice Human Total RNA Survey Panel (Ambion), and mouse Total RNA Master Panel (Clontech). Pooled human adult pancreas RNA from n=3 donors and pooled human adult islet RNA from n=3 donors was sourced from the Oxford Islet Biobank (77). Pooled mouse islet RNA was sourced from four 12–28 week-old NMRI mice. Hypothalamus and pituitary RNA were commercially sourced from AMSBio. C_T values were normalised to the combined average of the housekeeping genes *PPIA* and *TBP*. *PAM* expression in sorted beta cells (n=11) and non-beta cells (n=5) was analysed based on previously published RNA sequencing data (107).

4.2.2 Immunofluorescent labelling and microscopy

Immunofluorescence in EndoC-βH1 was performed as previously described (*Section 2.7*), using the following primary antibodies: 4A6 anti-Myc (EMD Millipore), Belfast anti-insulin (in-house), TGN46 anti-trans Golgi (Sigma-Aldrich), and H-70 anti-calnexin (Santa Cruz Biotechnology). Immunofluorescence in human islets was performed by Dr Benoit Hastoy (Gloyn lab, OCDEM). Briefly, human pancreatic specimens were fixed in 10%

formalin and embedded in wax. Sections of 4μm were de-waxed, rehydrated, and heated to 95°C in 10mmol/L Tris and 1mmol/L EDTA pH 6.0 for 10min for antigen retrieval. Tissue sections were incubated overnight in anti-PAM (Abcam ab109175), anti-insulin (in-house antibody), and mouse anti-glucagon (Sigma-Aldrich G2654) antibodies. PAM immunoreactivity was enhanced using a tyramide amplification step (Thermo Fisher Scientific) and visualised with Alexa Fluor 488 (Thermo Fisher Scientific). Insulin and glucagon were visualised using Alexa Fluor 633 (Thermo Fisher Scientific) and anti-mouse TRITC (Sigma-Aldrich) respectively.

4.2.3 Gene silencing and insulin secretion in EndoC-βH1

For gene silencing studies, EndoC-βH1 cells were transfected (*Section 2.3*) using commercially available SMARTpools (Dharmacon) targeting *PAM*, *CHGA*, and *IAPP*, and analysed for knockdown at the transcript level (*Section 2.4*). For stimulation with TRH and GLP-1 (both Bachem), cells were cultured for 72h prior to overnight starvation in 2.8 mM glucose medium. Insulin secretion and cell proliferation assays were then performed as described (*Section 2.6*). Supernatant and cellular measurements of insulin content were normalised to cell counts on a per-well basis to account for differences in the number of plated cells. Corrected values were then normalised to basal insulin secretion rates for negative control on a per-experiment basis to eliminate variability in baseline secretion rates across replicates.

4.2.4 Membrane capacitance assays in EndoC-βH1 and primary islets

Electrophysiological measurements performed by Dr Hastoy were conducted using an EPC10 Patch Clamp Amplifier (HEKA) at 32°C using the standard whole cell configuration as previously described (240). The pipette solution (intracellular medium) contained 125mM Cs-glutamate, 20mM CsCl, 15mM NaCl, 1mM $MgCl_2$, 5mM HEPES, 0.05mM EGTA, 0.1mM

cAMP, and 3mM Mg-ATP (pH 7.15 with CsOH). The extracellular solution contained 1mM glucose, 118mM NaCl, 5.6mM KCl, 1.2mM $MgCl_2$, 5mM HEPES, and 2.6mM $CaCl_2$ (pH 7.4 with NaOH). Tetraethylammonium chloride (20mM) was added to block residual K^+ currents.

4.2.5 Electron microscopy

Transfected EndoC-βH1 cells (*Section 2.3*) were cultured for 72h, and harvested as described (*Section 2.4.1*). The cell pellet obtained after washing in PBS was then fixed in 2.5% glutaraldehyde, post-fixed in 2% uranyl acetate, dehydrated in graded methanol, and embedded in London Resin Gold (Agar Scientific). Ultrathin sections cut onto nickel grids were immunolabelled for PAM (Abcam ab109175) followed by anti-rabbit biotin (Vector Laboratories) and streptavidin gold 15nm (Biocell). Insulin was immunolabelled using an in-house antibody followed by anti-guinea pig gold 10nm (Biocell). Sections were viewed on a Joel 1010 microscope (accelerating voltage 80kV) with a digital camera (Gatan). To estimate PAM concentration in secretory vesicles, quantification of immunogold labelling was performed in a blinded fashion using Image J software. The following parameters were determined: cross-sectional vesicle area, vesicle density (number of vesicles/cytoplasmic area), and insulin or PAM labelling/vesicle (number of immunogold particles/vesicle).

4.2.6 Inhibitor studies and western blot analysis of CgA amidation

To mimic the effect of gene silencing, cells were cultured for 24 h with the irreversible *PAM* inhibitor 4-phenyl 3-butenoic acid (4P3BA; Sigma-Aldrich) at a concentration of 500 μM (241). Whole-cell extracts (WCE) from siRNA or inhibitor-treated EndoC-βH1 cells were then prepared and quantified as previously described (*Section 2.8.1*). Supernatants were harvested directly from cells and filtered with a 0.22μm filter. 2-10μg WCE and 0.5% total supernatant volume were analysed via denaturing SDS-PAGE (*Sections 2.8.2-2.8.3*). Antibodies specific

for CgA-Gly (Abcam ab52983) and total CgA (Abcam ab15160) were used to calculate the ratio of amidated to non-amidated CgA using Image Lab 5.0 (242).

4.2.7 Analysis of exocytosis in primary human beta cells

Human pancreatic islets were isolated and dispersed at the Alberta Diabetes Institute IsletCore as previously described (167). Single-cell capacitance responses were monitored by Dr Jocelyn Manning Fox (University of Alberta) as described (*Section 4.2.4*) with an additional 1h pre-incubation period at 1 mM glucose. Following recordings, cells were positively identified by insulin immunostaining. All islet samples had previously been genotyped (77).

The kinetics of exocytosis were examined at 1 and 10 mM glucose for seven risk allele carriers and seven controls, matched as groups for age (54.9 years [carriers] vs. 58.4 years [controls]; $p = 0.52$), gender (2/7 females vs. 4/7 females) and BMI (27.7 vs. 26.7; $p = 0.68$). Mean capacitance was calculated for each depolarisation step and ANOVA with Tukey's post-hoc test applied to detect statistically significant differences. In an extended analysis, mean capacitance measurements were adjusted for the top five covariates in terms of variance explained across conditions (date of isolation, cause of death, smoking status, alcohol consumption, and islet purity) by fitting a multiple linear regression model for each condition. Residuals were found to be approximately normally distributed, and the adjusted values were then used to estimate the mean effect of *PAM* genotype status on each of the exocytosis measures.

4.2.8 Statistics

All statistical analyses were performed in R version 3.3.1, except electrophysiological measurements in EndoC-βH1 cells, which were analysed by Dr Hastoy using OriginPro v8.5.1. Graphs show means of the indicated number of replicates, and error bars are SEM. For kinetic

data, rates were found to be dependent on substrate concentration, and exponential models were fitted. Kinetic, electrophysiological, electron microscopy, and cellular data were analysed using two-sided, unpaired t-tests, and Western data using two-sided, paired t-tests.

4.3 Results

4.3.1 PAM expression and subcellular localisation

To identify possible tissues through which PAM could be acting to influence insulin secretion, I first assessed expression patterns across human and mouse tissue panels. Transcript expression profiling by RT-PCR detected *PAM* in multiple tissue types, with highest expression in human islets followed by the heart and salivary glands (**Figure 3A**). In mouse tissues, *Pam* expression was highest in the pituitary (not included in the human panel) followed by the heart and islets (**Figure 3B**). I confirmed broadly similar expression patterns in the publicly available Genotype-Tissue Expression (GTEx) database, though data for pancreatic islets are not available from this project (118).

I next analysed published RNA sequencing data from human pancreatic islets, and found *PAM* to be equally expressed in enriched fractions of beta cells and non-beta cells (**Figure 4A**) (107, 243). This expression pattern was verified at the protein level by immunofluorescent staining of human islets, which showed co-localisation of endogenous PAM with both insulin and glucagon (**Figure 4B**). Consistent with a role in beta cell function, *PAM* is thus highly expressed in human islets and co-localises to the secretory pathway.

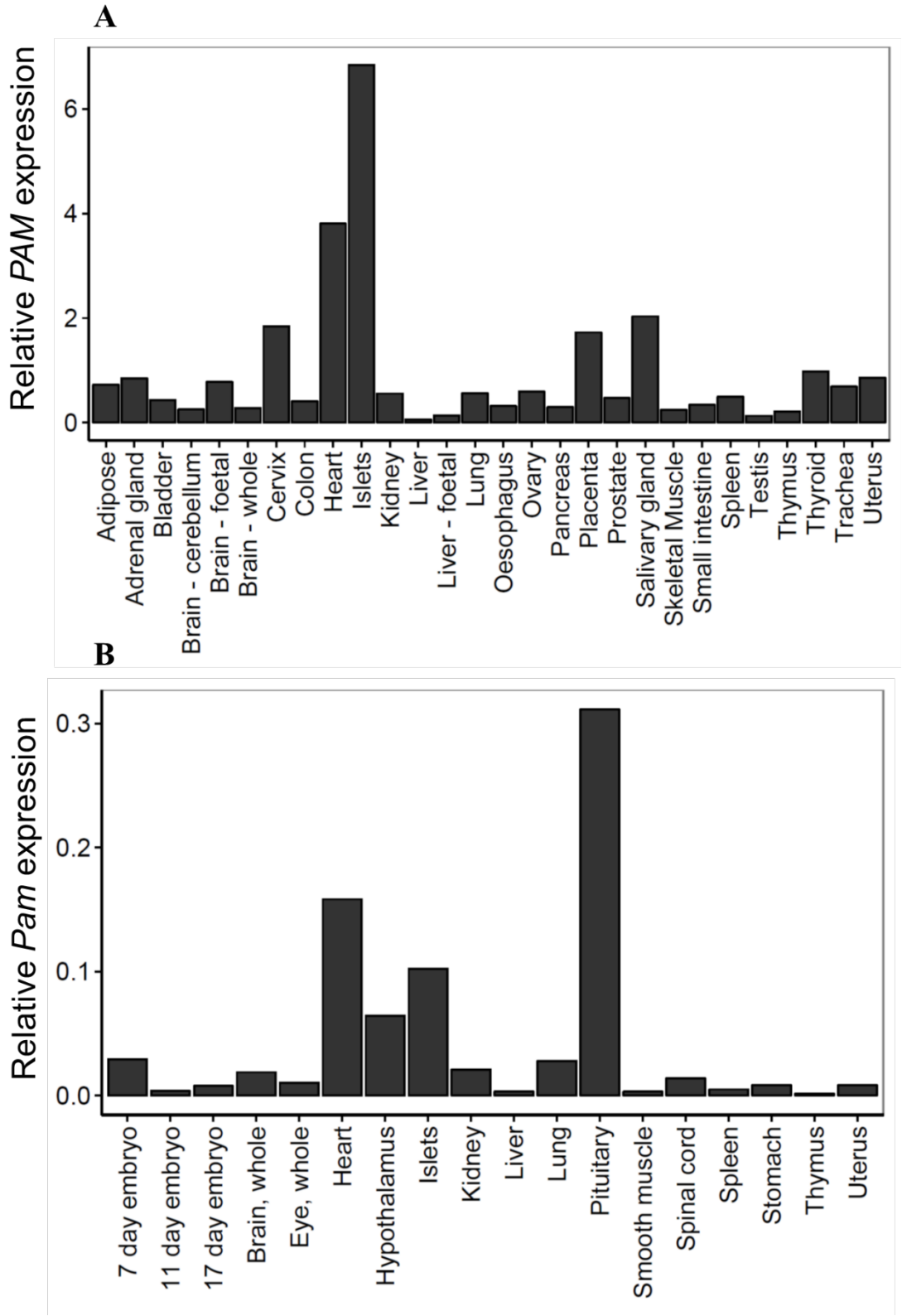


Figure 3. *Analysis of* PAM *transcript expression across selected human and mouse tissues. PAM* expression is shown relative to the average of two housekeeping genes (*PPIA* and *TBP*) in a human tissue panel (A) and a mouse tissue panel (B).

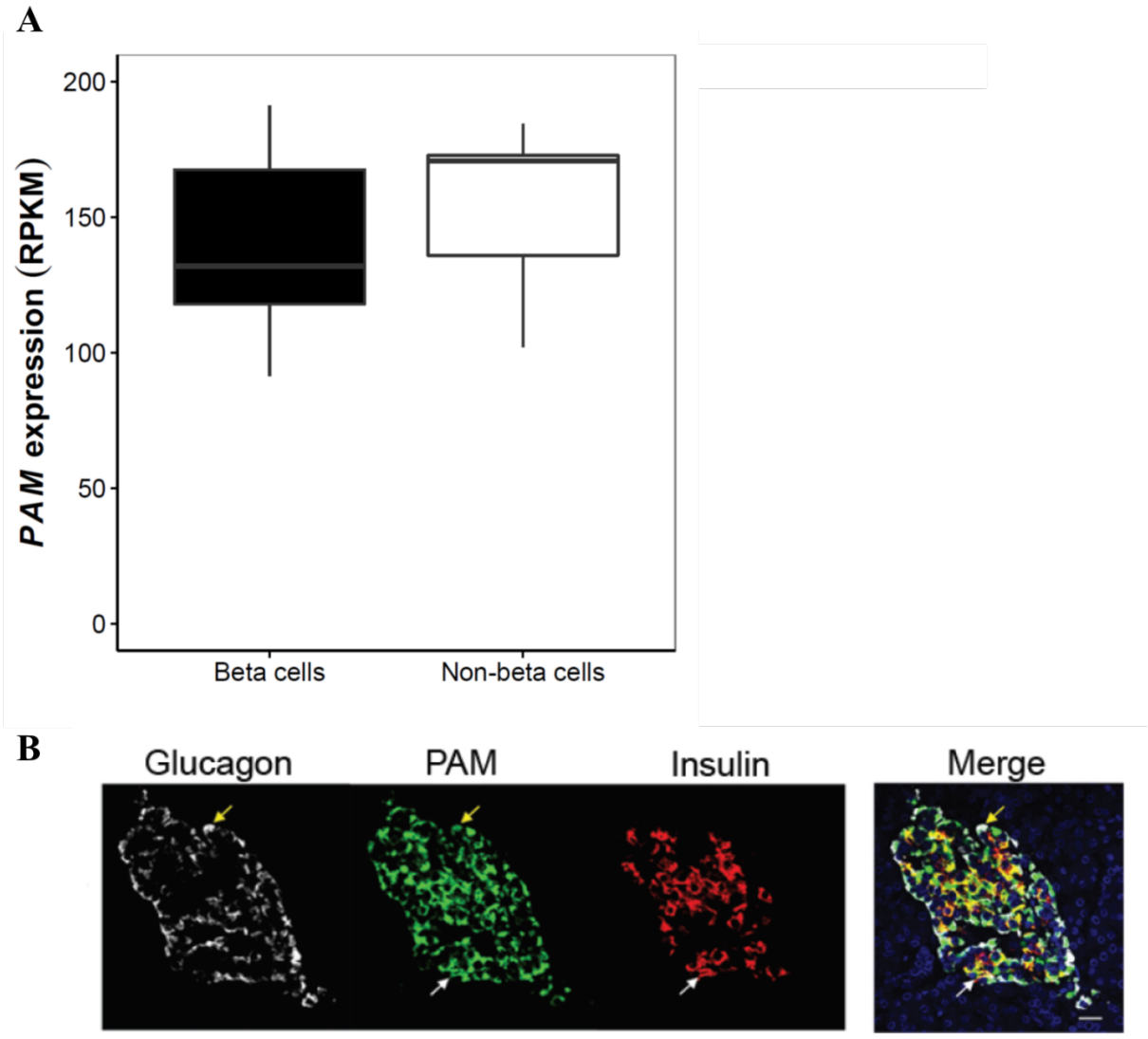


Figure 4. PAM *expression in different cell types from primary human islets*. (A) *PAM* transcript expression in FACS-sorted human islet beta cells and islet non-beta cells. Box plots show the interquartile range for *PAM* expression, and whiskers indicate variability outside the upper and lower quartiles (n=11 for enriched beta cells, and n=5 for non-beta cells). (B) Immunofluorescent labelling of glucagon (white), PAM (green), and insulin (red) in human pancreas by Dr Hastoy. White and yellow arrows indicate examples of co-localisation of PAM with insulin and glucagon respectively. Scale bar, 20µm.

Prompted by these findings, detailed analysis of the subcellular localisation of WT and risk allele PAM proteins was then performed by Dr Anne Raimondo in the human beta cell line EndoC-βH1. This analysis included both integral and secreted forms of PAM, which differ in the presence of a transmembrane and cytosolic C-terminal domain (157). Integral WT-PAM immunofluorescence localised to the trans-Golgi network (TGN), where secretory granules originate, whereas secreted WT-PAM displayed a vesicular staining pattern and co-localised with endogenous insulin (**Figure 5A-B**). Similar results were observed for D563G-PAM and Y651F-PAM (**Figure 5C-D**).

As in HEK cells (*Section 4.1.3*), secreted S539W-PAM expression was undetectable (**Figure 5A-B**). However, integral S539W-PAM displayed an aberrant staining pattern that localised to the endoplasmic reticulum (ER) as well as the TGN. These results confirm that PAM localises to components of the beta cell secretory pathway, in agreement with studies in other neuroendocrine cell types (163, 230, 244-246). The S539W substitution may interfere with PAM protein folding such that its translocation from the ER to the TGN is prevented or delayed.

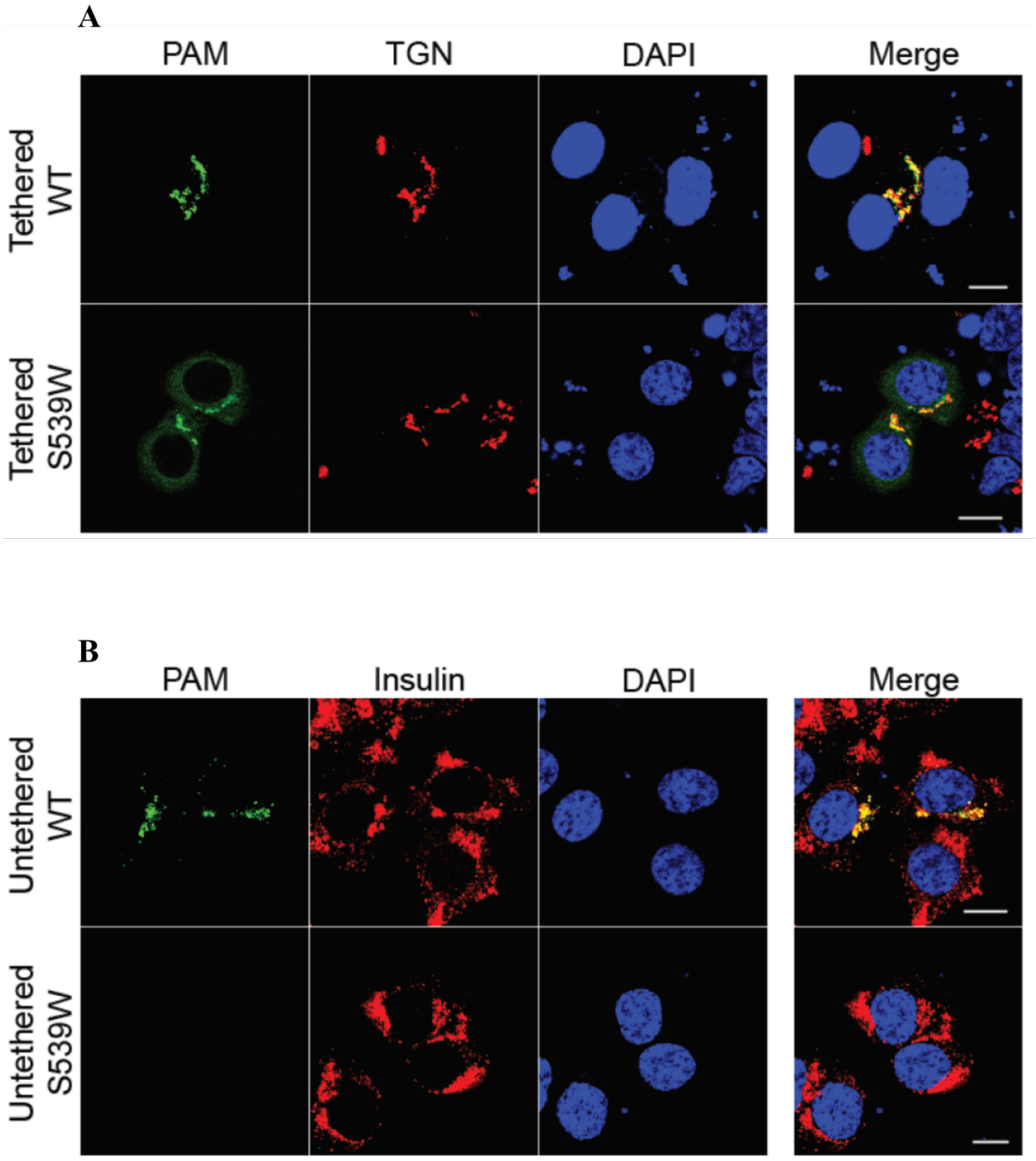


Figure 5. *PAM localisation in human beta cell-derived EndoC-βH1 cells*. Cells were transfected by Dr Raimondo with expression vectors encoding either membrane-tethered or untethered WT or variant PAM, then fixed and stained with fluorescent antibodies against either (A) PAM and TGN, or (B) PAM and insulin. DAPI was used as a nuclear marker. Results shown are representative of at least two independent experiments. Scale bar, 2μm.

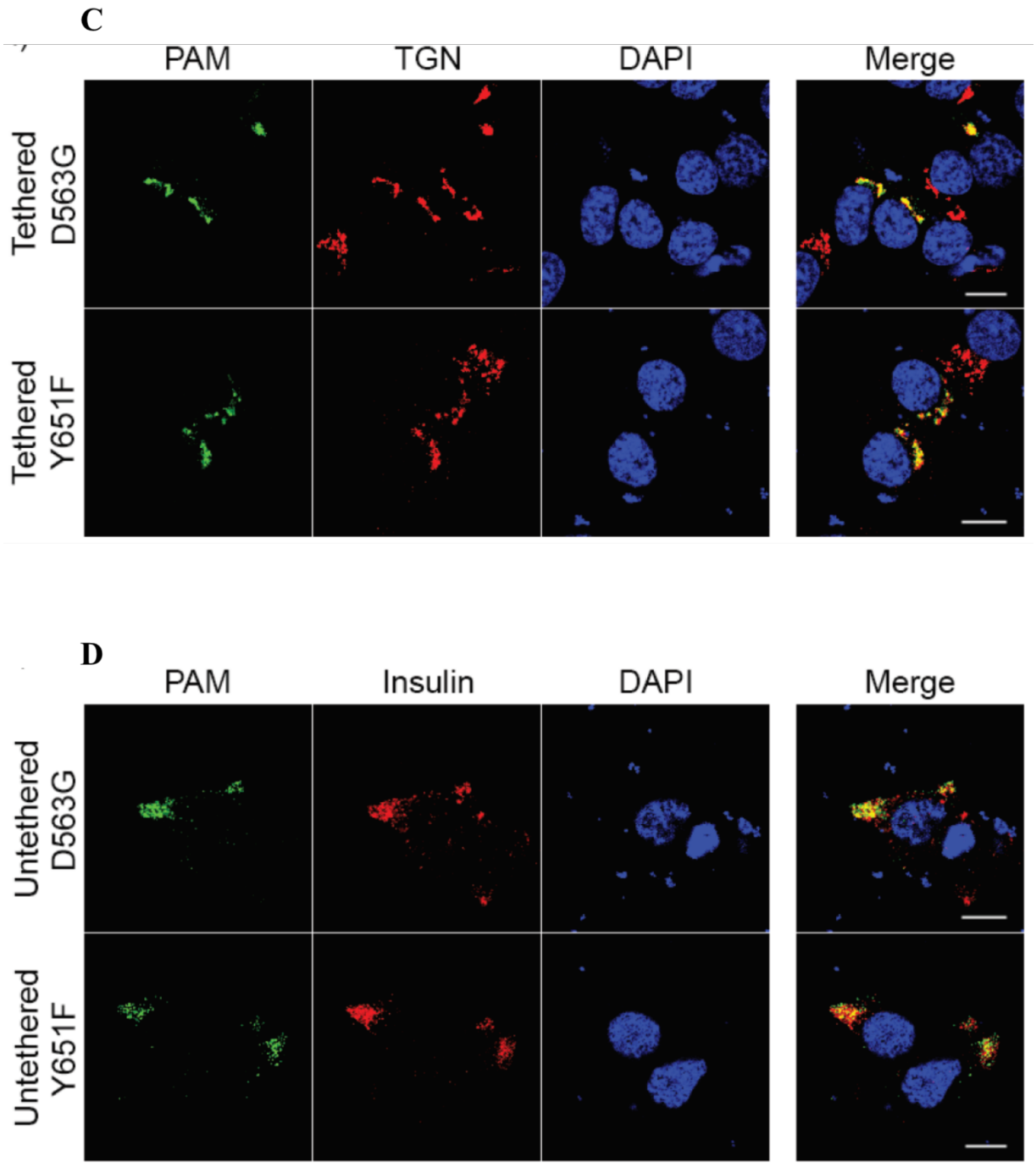


Figure 5 (*continued*). See figure legend on previous page.

4.3.2 PAM regulates glucose-responsive insulin secretion

Taken together, the human genetic and cellular expression data suggest that PAM may play a direct role in the regulation of beta cell function. Given the negative direction of effect for both T2D risk alleles, I hypothesised that reduced PAM function would lead to decreased beta cell function *in vitro*. To address this, I performed siRNA-mediated silencing of *PAM* in EndoC-βH1. Based on a time course, the point of lowest *PAM* expression was determined to be four days post transfection, and I selected this time for all subsequent experiments (**Figure 6A**). To assess the impact of reduced PAM levels on beta cell function, I then performed static insulin secretions combined with end-point measurements of total cell numbers and transcript levels.

I first confirmed satisfactory *PAM* knockdown in this assay, with a 79% reduction in *PAM* transcript levels compared with negative control (**Figure 6B**). This was associated with decreased insulin secretion under both basal (-6.9%, p=0.04) and high glucose conditions (-14.4%, $p=8.2x10^{-4}$) (**Figure 7A-B**). Furthermore, *PAM* knockdown cells displayed a significantly blunted response to glucose (ratio of high to basal secretion; -7.7 %, p=0.03), indicative of a greater impact on secretion under stimulated conditions. Cellular insulin content was also significantly reduced in *PAM* knockdown cells (-17.6%, $p=4.9x10^{-5}$) despite normal viability and *INS* expression (**Figure 7C-D**).

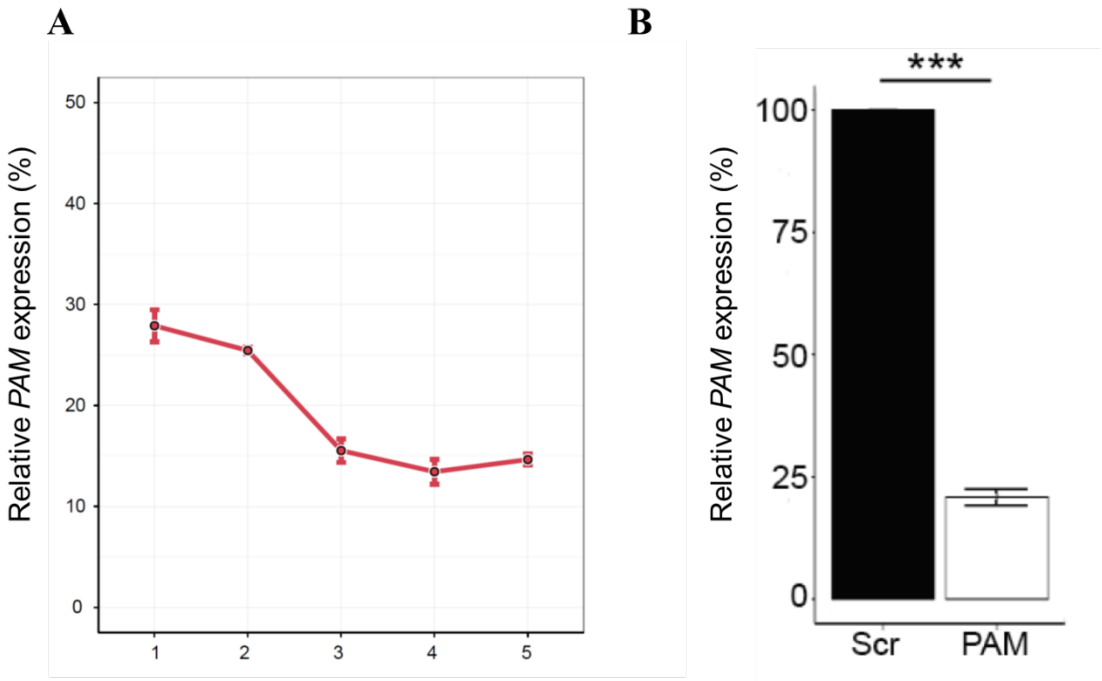


Figure 6. PAM *expression following siRNA-mediated gene silencing*. EndoC-βH1 cells were transfected with pools of siRNAs targeting *PAM* and scrambled (Scr) control. (A) *PAM* expression normalised to control at the indicated times after transfection (n = 3). (B) Relative *PAM* expression immediately following static insulin incubations and end-point measurements of cell counts (see *figure 7*), four days after transfection (n = 4). All expression levels have been normalized to two house-keeping genes (*PPIA* and *TBP*), and indexed to control. Error bars are mean ±SEM. ***p-value < 0.001.

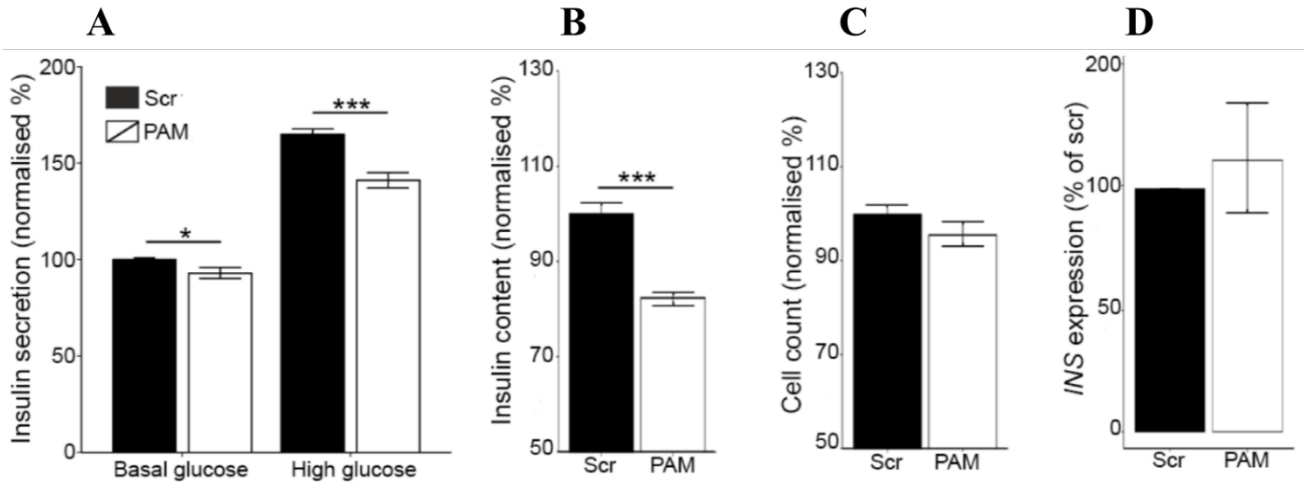


Figure 7. *Effects of PAM gene silencing on beta cell function*. EndoC-βH1 cells were treated as described in figure 6 and then measured for (A) insulin secretion in static incubation assays with basal (1 mM) or high (20 mM) glucose media (n=8), (B) cellular insulin content (n=16), (C) total cell counts (n = 16), and (D) insulin expression (n = 3). In the case of insulin secretion and content, measurements have been normalized to total cell numbers on a per-well basis. All values are expressed as a percentage of scrambled control (and basal levels in A). Error bars are mean ±SEM. P values * <0.05, *** <0.001.

To further probe the impact of *PAM* silencing on beta cell function, I performed quantitative immunoelectron microscopy. Using immunogold-labelling, I co-stained for insulin and PAM, and subsequently scored all electron micrographs in a blinded fashion (**Figure 8A**). The results of this analysis confirmed successful knockdown of PAM at the protein level, with a 47% reduction in the amount of immunogold-labelled particles per vesicle (ppv) (**Figure 8B**). In agreement with the insulin content data presented above, *PAM*-treated cells also showed a 23% decrease in immunogold-labelled insulin per vesicle compared with control (6.26±0.28 vs 8.1±0.41 ppv, $p=1.77x10^{-4}$) (**Figure 8B**). I observed trends towards decreased cross-sectional vesicle area (0.048±0.003μm^2 vs 0.056±0.005μm^2, p=0.18) and increased vesicle density (2.30±0.55 vesicles/μm^2 vs 1.16±0.09 vesicles/μm^2, p=0.056) in *PAM* knockdown cells. Overall, this analysis corroborates the defects observed in my static incubations, and establish that *PAM* silencing impacts negatively on the insulin secretory pathway in EndoC-βH1.

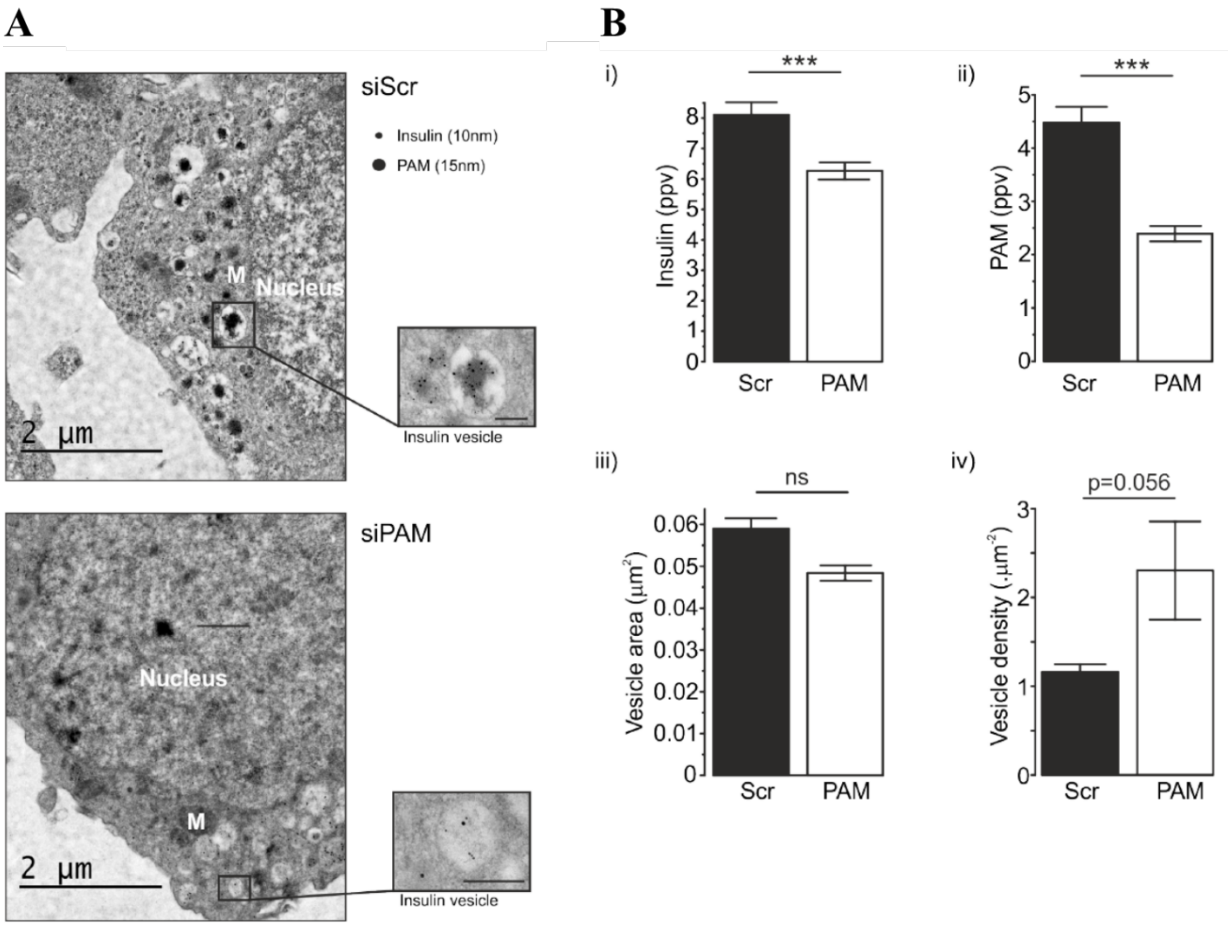


Figure 8. *Effects of* PAM *silencing on ultrastructural features of the human beta cell line, EndoC-βH1.* (A) Representative electron micrographs from siRNA-treated EndoC-βH1 cells. (B) i) number of immunogold particles (insulin) per vesicle, ii) number of immunogold particles (PAM) per vesicle, iii) vesicle area, iv) vesicle density. Results shown are the average of n=290 (*PAM*) or n=235 (scrambled) vesicles except for iv) where n=10 cells per treatment. M, mitochondrion. Error bars are mean ±SEM. P value *** <0.001. ns, not significant.

4.3.3 PAM deficiency causes defects in the dynamics of exocytosis

Net insulin secretion is a function of a number of underlying cellular processes, including insulin production, granulogenesis, glucose metabolism and exocytosis. Thus, an interesting emerging question is the extent to which the observed effects of *PAM* knockdown on insulin secretion are separate from the effects on insulin production. In other words: is the apparent decrease in secreted insulin in static incubations an indirect consequence of the reduced insulin available? To get at this question, Dr Hastoy performed single-cell capacitance measurements on cells I had treated similarly to above.

The protocol provides high-resolution data on the electrophysiological characteristics of individual cells. Following stimulation with a train of 10 depolarisations, an increase in plasma membrane surface area due to vesicle exocytosis can be detected at each step. The size of the steps, normalized to overall cell size, is a function of the number of fused vesicles during the depolarisation phase. The increase in cell capacitance during the first two depolarizations provides an estimate for the size of the readily releasable pool of secretory granules, and subsequent pulses represent mobilization of granules from the more intracellular reserve pool (247). In parallel, the size of different currents can be isolated to assess their contribution to exocytosis.

Comparing *PAM*-treated and control cells, there was no difference in calcium entry (peak current 4.5±0.7pA.pF^{-1} vs 3.6±0.6pA.pF^{-1} for *PAM* knockdown vs control cells, p=0.27), and the cumulative increase in exocytosis following ten pulses was comparable (43.9±8.6fF.pF^{-1} vs 38.0±8.1fF.pF^{-1}, p=0.64) (**Figure 9A-B**). Nevertheless, an apparent delay in exocytosis was observed in *PAM*-treated cells, with a reduced increment during the first depolarisation. In controls cells, nearly 60% of the exocytotic response was elicited by the first pulse, compared with less than 40% in *PAM* knockdown cells (p=0.013) (**Figure 9C**). These

results thus indicate that the dynamics of vesicle mobilization are affected by PAM knockdown, with a reduction in the readily releasable pool and delayed compensation from the reserve pool.

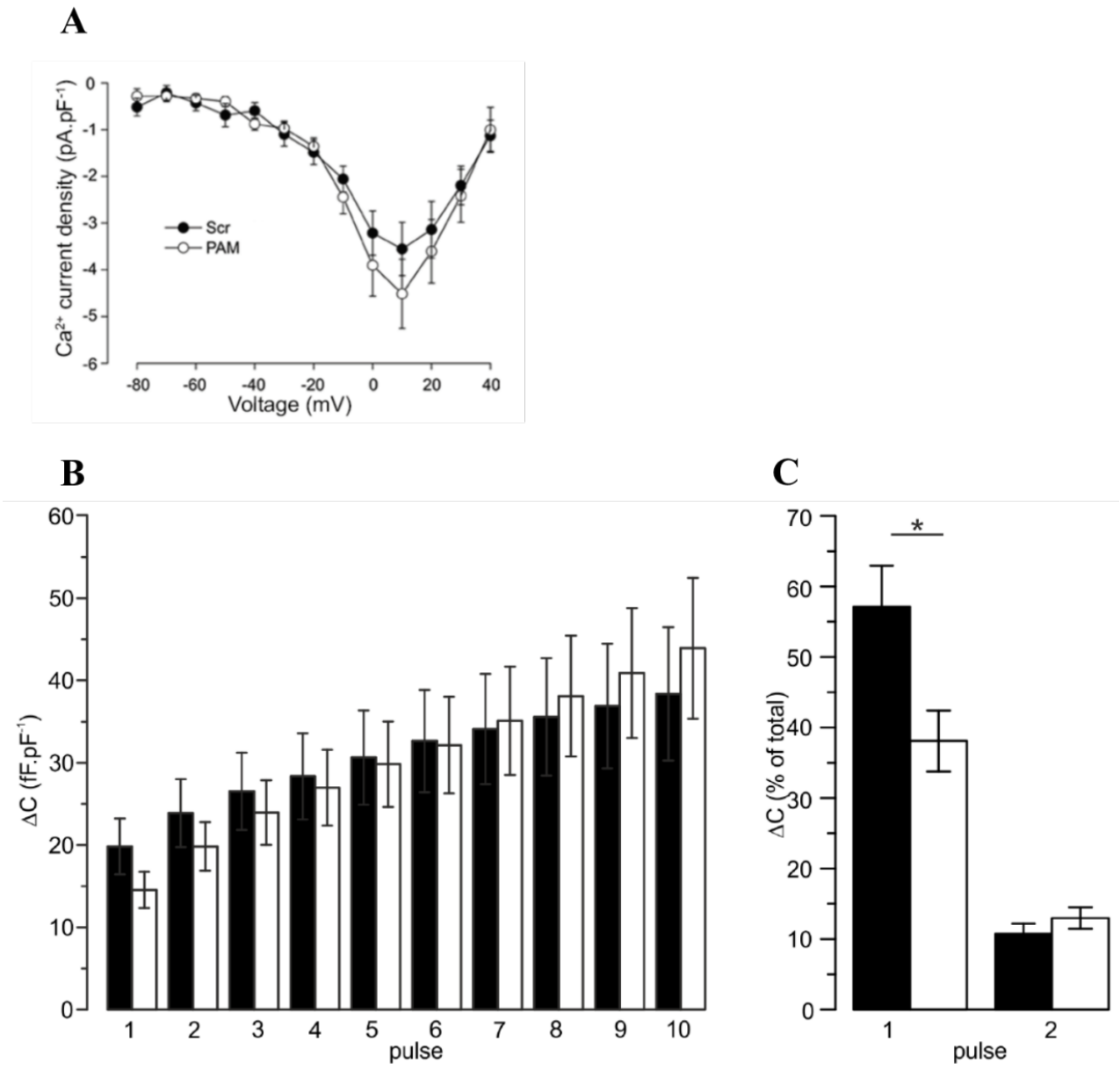


Figure 9. *Effects of* PAM *silencing on exocytosis in EndoC-βH1.* Cells were treated as described in figure 7, and then subjected to electrophysiological measurements by Dr Hastoy: (A) calcium current density (n = 15), and (B) granule exocytosis as a function of depolarisation-evoked increases in membrane capacitance (n = 15). (C) shows the incremental capacitance for the first two pulses as a percentage of the total increase. Error bars are mean ±SEM. * <0.05.

4.3.4 Effects of PAM variants on primary beta cell function

Having established a role for PAM in a human beta cell model, I sought to corroborate these results by examining effects for the common *PAM* variant (rs35658696) on primary beta cell function. In the largest available dataset of human pancreatic islets with functional data, I analysed exocytosis in single beta cells from dispersed islets. Data collection was comparable to that used for exocytosis in EndoC-βH1 above, with capacitance measurements following a train of 10 depolarisations. In this case, cells were starved for 1h in low glucose prior to performing the protocol in either 1 or 10 mM glucose.

Using available genotyping data for the cohort, I found the risk allele to be present in 4.7% of individuals, in line with the reported frequency in European populations (37, 38). Seven individuals heterozygous for the T2D risk allele were compared to seven age-, gender-, and BMI-matched individuals homozygous for the reference (non-risk) allele (**Figure 10A**). To test for differences in the size of the readily releasable pool, I calculated the total increase in capacitance during the first two depolarisations (**Figure 10B**). Although no differences were detected at low glucose, a significant decrease in capacitance was seen in risk allele carriers at 10 mM glucose (-26.5%; p = 0.04).

I next compared exocytosis in the risk allele carriers to a full set of 80 control individuals. In this extended analysis, total exocytosis measurements were available under three different conditions. Using a linear modelling approach, capacitance measurements were analysed after adjusting for the five most significant covariates. No statistically significant effects were observed, but exocytosis under medium and high glucose conditions for risk allele carriers was found to be reduced by 4.63fF.pF^{-1} (-21.3%; p = 0.44) and 4.36fF.pF^{-1} (-17.1%; p = 0.44) compared to non-risk carriers, while exocytosis at basal glucose was increased by 1.21fF.pF^{-1} (+14.7%; p = 0.66) (**Figure 10C**). This is in line with data from EndoC-βH1,

which did not show any significant differences in cumulative exocytosis following ten depolarisation. Overall, these results are consistent with an effect of rs35658696 status on the dynamics of granule release.

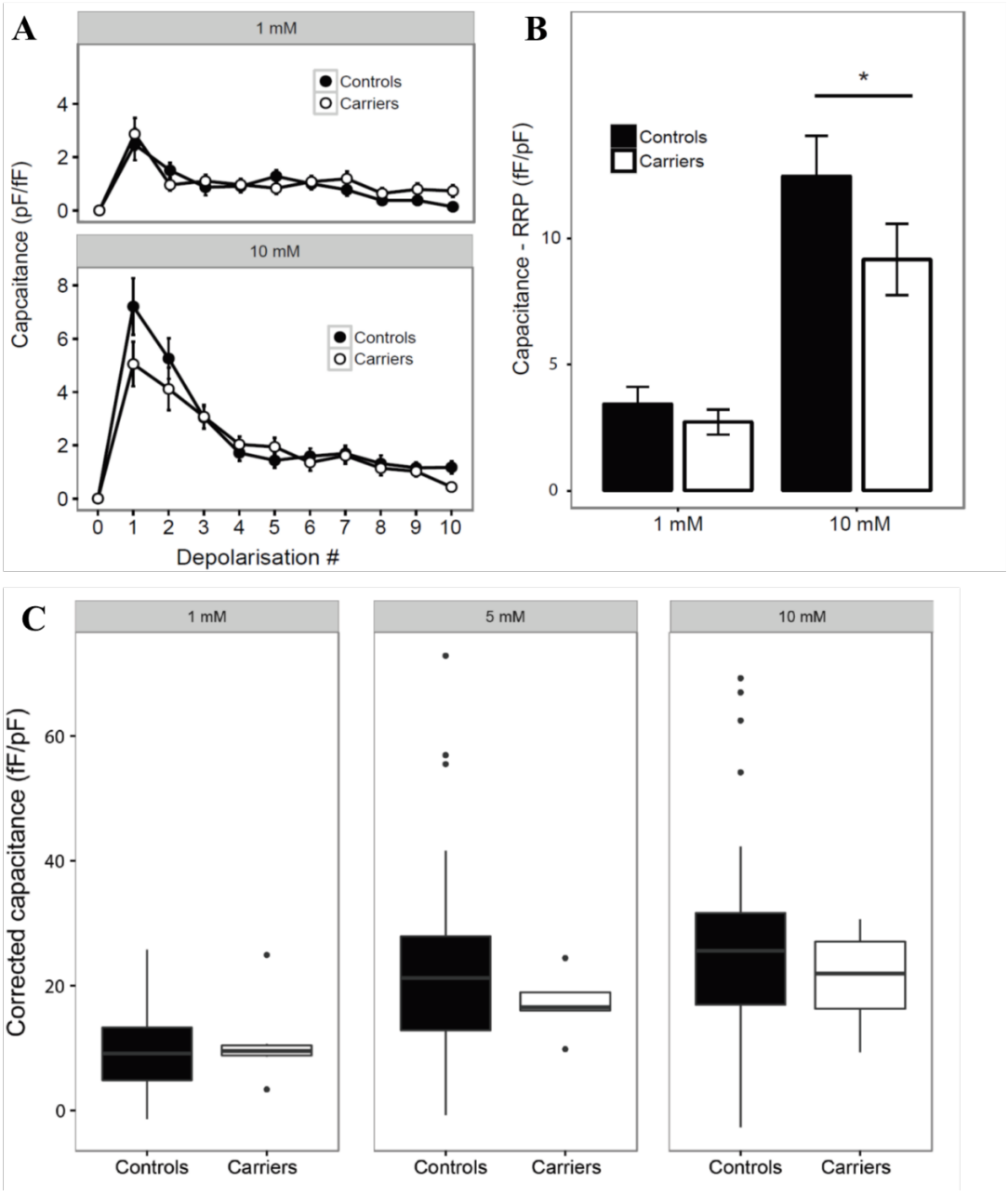


Figure 10. *Effects of rs35658696 genotype status on exocytosis in primary human islets.* Capacitance measurements for individuals heterozygous for the rs35658696 risk allele ("Carriers") and individuals homozygous for the non-risk allele ("Controls") are shown in (A-C). (A-B) display the stepwise capacitance measurements after each of ten depolarisations, and (B) is the cumulative increase in capacitance after the first two depolarisations, indicative of the size of the readily releasable pool. (C) Total increases in capacitance in beta cells from 87 islets. Error bars are mean ±SEM. * P value < 0.05.

4.3.5 Chromogranin A is a PAM substrate and possible downstream effector

To explore possible mechanisms for the PAM-regulated pathway in beta cells, I sought to identify target proteins that could act as downstream effectors of PAM activity. While any C-terminal glycine-extended peptide could in theory be a substrate for PAM, I reasoned that a key downstream mediator of the observed effects in beta cells would need to be co-expressed with PAM along the secretory pathway. I therefore searched a published catalogue of the islet secretome for potential targets of PAM amidation (248). This highlighted two proteins, Chromogranin A (CgA, encoded by *CHGA*) and Islet Amyloid Polypeptide (IAPP, encoded by *IAPP*), both of which are highly expressed in beta cells and have plausible links with PAM function (248). CgA is a granular packaging protein that self-aggregates to condense large numbers of neuroendocrine peptides into nascent secretory granules (249-251). It can also be cleaved to produce smaller peptides with additional biological roles (252). The role of IAPP in beta cells is less well understood, although deposition of elongated fibrils in islets due to IAPP aggregation is a feature of T2D (253).

Interestingly, *CHGA* silencing (**Figure 11A**) revealed a phenotype similar to that observed for *PAM*, with significant reductions in both insulin content and secretion (**Figure 11B-D**). Using an antibody specific to the glycine-extended (non-amidated) form of CgA (CgA-Gly), the accumulation of CgA-Gly was observed in EndoC-βH1 cells treated with *PAM* siRNA by Western blot analysis (**Figure 12A-B**). Treatment of cells with 4P3BA, a PAM antagonist, produced similar results, demonstrating that amidation of CgA in beta cells is dependent on PAM activity (**Figure 12A&C**) (242, 254). An antibody that recognises all forms of CgA (irrespective of amidation status) detected an additional lower molecular weight (~70kDa) form of CgA, presumably due to cleavage near its C-terminus. Production of this cleaved form was altered in response to decreased PAM expression and activity, suggesting that processing of CgA may be regulated by PAM-dependent amidation.

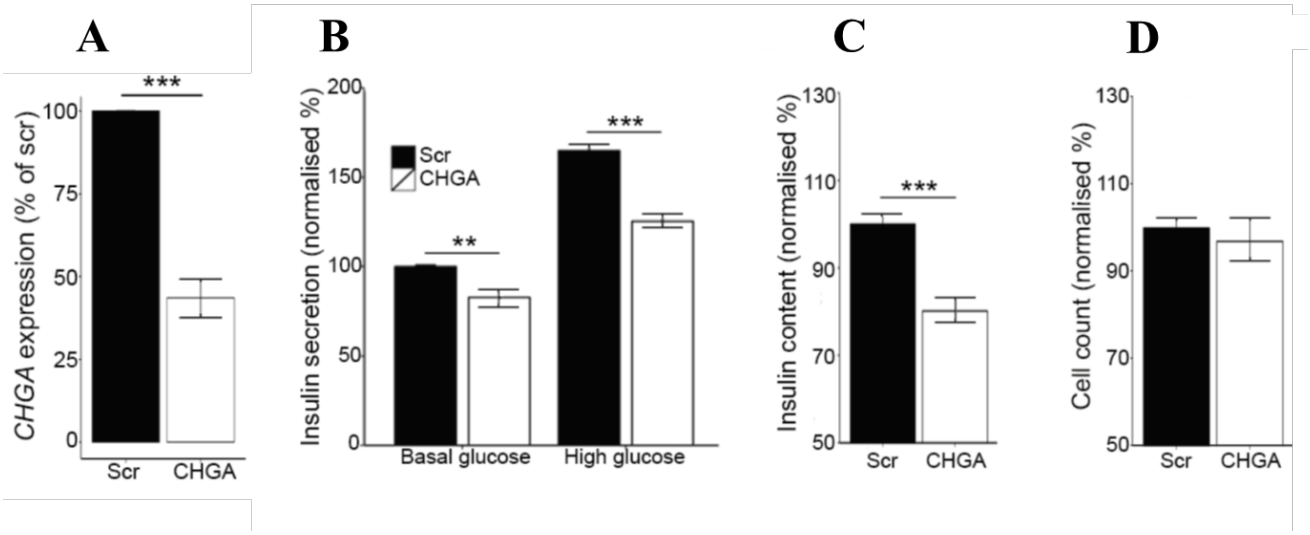


Figure 11. *Effect of* CHGA *gene silencing on beta cell function.* EndoC-βH1 cells were transfected with scrambled (Scr) or *CHGA* siRNA and then measured for (A) *CHGA* expression normalised to scrambled control (n = 4), (B) insulin secretion in static incubation assays with basal (1 mM) or (20 mM) glucose media (n=8), (C) cellular insulin content (n=16), and (D) total cell counts (n = 16). All values have been normalized as described in figure 7. P values ** <0.01, *** <0.001.

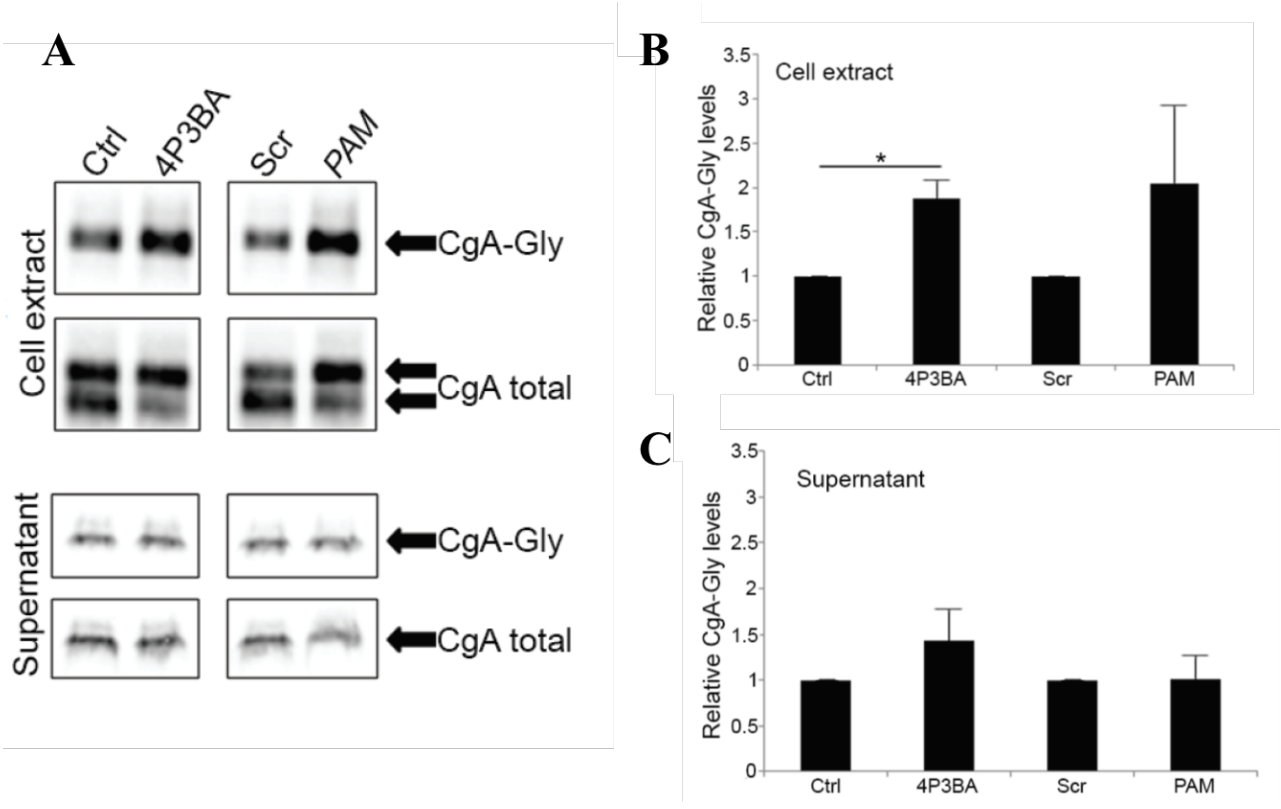


Figure 12 *CgA amidation patterns following PAM silencing and inhibition*. (D) EndoC-βH1 cells were transfected with scrambled or *PAM* siRNA or treated with vehicle (Control) or 4P3BA, then analysed by Western blot using antibodies specific for non-amidated CgA (CgA-Gly) and total CgA. (A) Representative blots for whole cell extracts and supernatants, and (B) quantification by densitometry. Error bars are mean ±SEM.

These results confirm that CgA is a PAM substrate in beta cells, and a strong candidate for mediating the effects of PAM on beta cell function. In contrast, silencing of *IAPP* in EndoC-βH1 cells had no effect on insulin content or secretion (**Figure 13**).

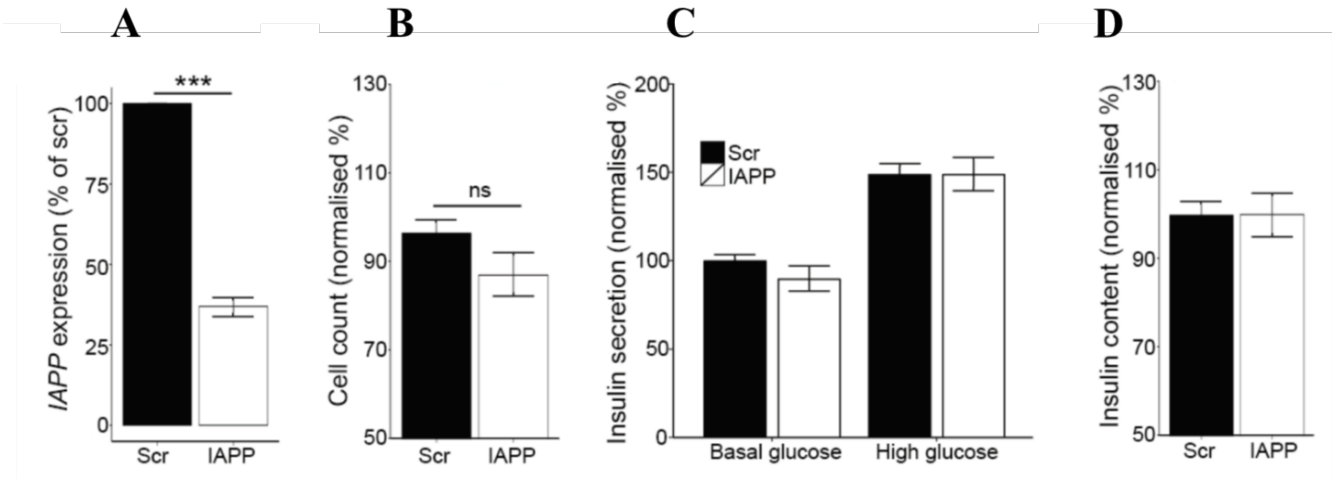


Figure 13. *Effect of* IAPP *silencing on beta cell function.* EndoC-βH1 cells were transfected with scrambled (Scr) or *IAPP* siRNA and then measured for (A) *IAPP* expression normalised to scrambled control (n = 4), (B) total cell counts (n = 16), (C) insulin secretion in static incubation assays with basal (1 mM) or (20 mM) glucose media (n=8), and (D) cellular insulin content (n=16). All values have been normalized as described in figure 7. P values ** <0.01, *** <0.001.

4.3.6 Effects of PAM on extra-pancreatic hormones

Although the tissue panels revealed *PAM* to be most highly expressed in human pancreatic islets, the data also showed *PAM* expression across other tissues known to be important for metabolic homeostasis. It is thus conceivable that *PAM,* in addition to having local (islet-intrinsic) effects, could be affecting beta cell function via amidated signals released from other cell types (islet-extrinsic effects) (243, 255). Glucagon-like peptide 1 (GLP-1) is one such amidated peptide that amplifies insulin secretion following release from intestinal L-cells in response to a meal (8). Though L-cells were not included in the tissue panels used to assess *PAM* expression, others have observed *PAM* to be expressed in both human and mouse L-cells (Prof. Fiona Gribble, personal communication). TRH is another amidated neuroendocrine peptide, released from the hypothalamus, with some previous suggestions of a possible role in modulation of both beta cell and PAM function (256, 257). Hypothalamic expression of *Pam* was fourth highest in the mouse tissue panel, and GTEX showed *PAM* to be expressed in the human hypothalamus (median RPKM = 28.8, n = 96) (118). Both GLP-1 and TRH are therefore plausible candidates for islet-extrinsic modulation of insulin secretion. I focused on exploring the effect of these two specific candidates on beta cell function, since exhaustively testing all possible PAM substrates would be beyond the scope of this study.

I first produced a dose-response curve for high-glucose insulin secretion in EndoC-βH1 using exogenous stimulation with unamidated forms of the two hormones (**Figure 14A**). As expected, full-length GLP-1 produced a strong effect under both basal and high-glucose insulin secretion, with maximum stimulation at 10 nM. In contrast, TRH precursor produced no measurable effect on insulin secretion, even at concentrations orders of magnitude above physiological levels.

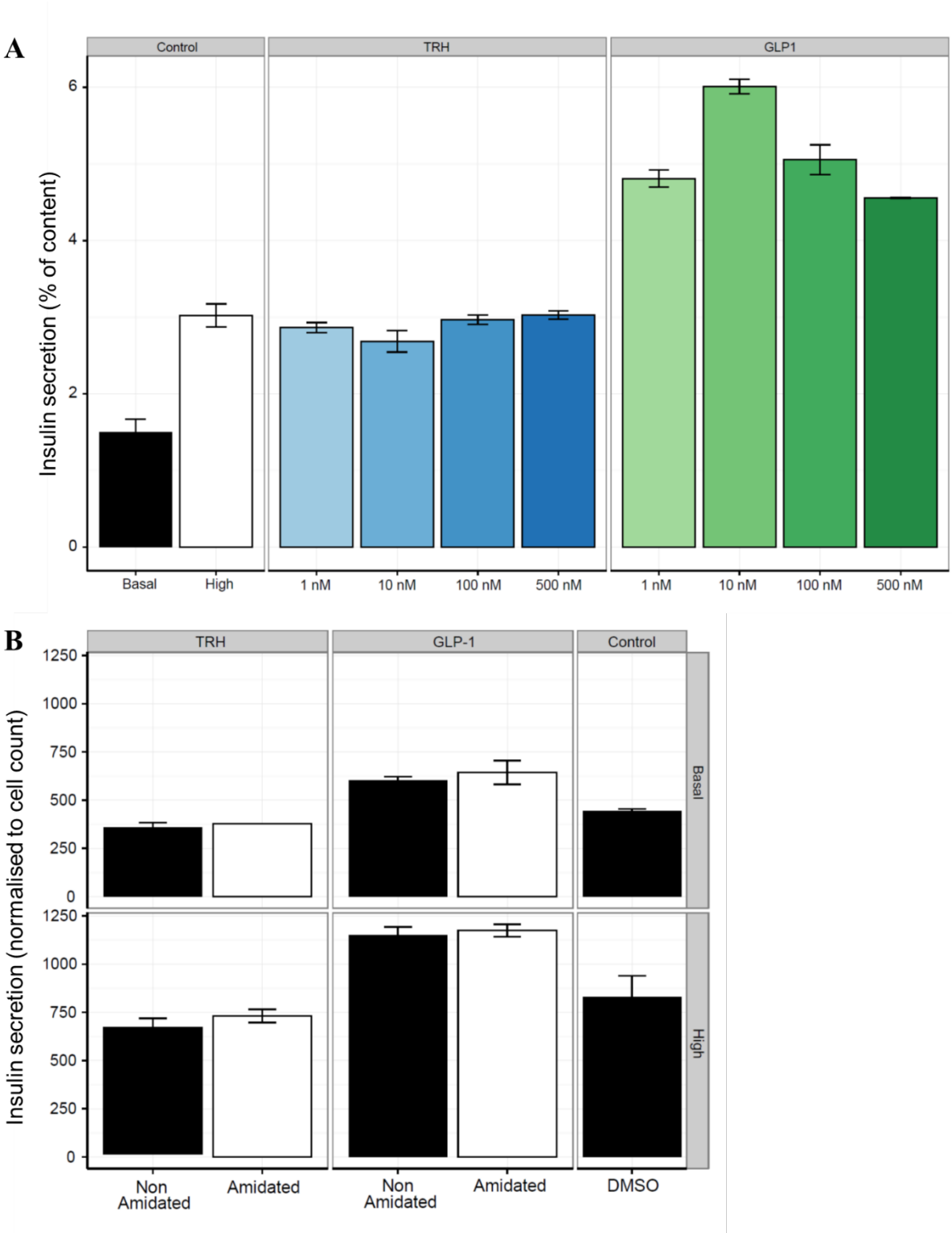


Figure 14 *Exogenous stimulation of insulin secretion with TRH and GLP-1.* (A) Content-normalized insulin secretion in the EndoC-βH1 for DMSO stimulation in basal and high glucose (left panel), unamidated TRH stimulation in high glucose (centre panel), and unamidated GLP-1 stimulation in high glucose (right). (B) Cell-count normalized insulin secretion for exogenous stimulation with unamidated and amidated peptides (all 10 nM) and DMSO (control) in basal and high glucose. Data points are the mean of n = 3, and error bars are SEM.

I next tested the effect of stimulation with amidated versus unamidated forms of the peptides at supraphysiological concentrations (10 nM) (**Figure 14B**). In both cases the amidation status of the hormones was found to be irrelevant to their activity in the assay; amidated GLP-1 displayed the same strong potentiating effect as unamidated GLP-1, while the two forms of TRH were both inactive. These findings are consistent with previous reports, which have shown the potencies of GLP-1-(7-36)-NH2 (amidated form) and GLP-1-(7-37) (unamidated form) in stimulating insulin secretion to be indistinguishable *in vivo* (258, 259). In the case of TRH, studies in rodent models have produced conflicting data. One report has suggested a stimulatory effect of exogenous TRH on basal insulin secretion, while other have failed to replicate this, with one study hinting instead at a possible role in regulation of glucagon secretion (256, 257, 260),

The results presented here show that, in the case of GLP-1 and TRH, directly altering the ratio of amidated to unamidated peptide is unlikely to play a role in influencing human beta cell function. Other possibilities of an islet-extrinsic PAM effect driven through these hormones remain unexplored. Specifically, it cannot be excluded that peptide amidation could affect the total amount of hormone in circulation, e.g. through effects on secretion rates or plasma half-life. On-going clinical studies in the Gloyn group are exploring these possibilities.

4.4 Discussion

Recent efforts to catalogue coding variants associated with T2D risk have identified independent association signals in the *PAM* gene. The coding risk alleles are also associated with decreased insulinogenic index, a physiological measure of insulin secretion in response to a glucose challenge, indicating a likely endogenous role for PAM in beta cell function. Previous work by Dr Anne Raimondo in our group has provided evidence that both T2D-risk alleles reduce overall PAM activity. Here, I show that reduced PAM expression causes beta cell dysfunction through multiple effects on insulin availability, secretion, and the dynamics of granule exocytosis. Importantly, these defects preferentially affect insulin secretion under stimulated conditions, leading to a blunted glucose response. In line with these results, I observed a consistent effect of the rs35658696 T2D risk allele (p.D563G) on measures of exocytosis in primary human beta cells. Taken together, the findings support a molecular mechanism whereby T2D risk alleles in *PAM* decrease PAM function, directly influencing the ability of pancreatic beta cells to mobilize insulin in response to glucose.

This novel mechanism provides a possible explanation for the relatively larger effect of rs78408340 (p.S539W PAM) that has been reported in two independent GWAS for T2D and insulinogenic index (261, 262). Mislocalisation and/or misexpression of p.S539W PAM would severely affect its function, particularly in cell types where the secreted enzyme predominates. Retention of integral membrane p.S539W PAM within the ER, as indicated by the subcellular localisation studies, could exacerbate this effect through ER stress. The precise cellular dynamics that yield secreted PAM in beta cells have not been elucidated; however, it appears generally to be more predominant in neuroendocrine cell types (263). This raises the interesting possibility of tissue-specific haploinsufficiency in p.S539W *PAM* carriers.

To explore the molecular mechanism underlying beta cell dysfunction in *PAM* risk allele carriers, I performed detailed characterisations of the human beta cell line EndoC-βH1 following *PAM* knockdown. The observed defects were broadly similar to those seen in primary human beta cells, with differential effects under stimulated conditions and a significant reduction in the size of the readily releasable pool. Interestingly, the effects of reduced PAM function under low glucose conditions in EndoC-βH1 cells were not apparent in primary beta cells. This may reflect differences in the degree of PAM LOF between siRNA-treated EndoC-βH1 cells and rs35658696 risk allele carriers, as well as differences in pre-incubation time prior to capacitance measurements, which could affect the extent of granule repletion.

The molecular studies in EndoC-βH1 cells confirm that CgA is a substrate for PAM amidation in beta cells (242). CgA is a granular packaging protein known to influence vesicle composition in other neuroendocrine cell types by facilitating the condensation of secreted proteins within nascent secretory vesicles (249-251, 264-266). Previous studies using global *Chga* knockout mouse models have attempted to investigate a role for CgA in pancreatic islets, but have been confounded by compensatory increases in other granin proteins, including Chromogranin B (CgB) (267, 268). Overall, these results suggest a mechanism whereby amidation of CgA by PAM enhances its ability to aggregate and package insulin efficiently in granules. In support of this hypothesis, amidation of numerous biological peptides has been shown to promote protein structure and self-aggregation due to improved electrostatic interactions (269-271). C-terminal cleavage of CgA has also been found to increase protein stability in pituitary cells (272, 273). These findings provide insights into the mechanism by which PAM and CgA may influence granule biogenesis, but future work will be needed to clarify their specific roles. The expression data presented in this study highlight the possibility of a role for PAM in a wide range of tissues. Some of these effects may indirectly modulate the

insulinogenic index *in vivo* through hormones that target the pancreatic beta cell. A large number of peptides are post-translationally amidated by PAM during maturation, but the extent to which this step is required for full activity is target dependent. In this study, I assessed the specific effect of GLP-1 and TRH amidation for exogenous stimulation of insulin secretion in the EndoC-βH1, and found no measurable differences. However, I did not test for possible effects on the total amount of hormone available in circulation. One study has hinted that GLP-1 (7-36-amide) may have an increased half-life compared to GLP-1 *in vitro* (7-37), suggesting that amidation could improve GLP-1 stability under some conditions (274). It thus remains possible that PAM could have islet-extrinsic effects on beta cell function via decreased GLP-1 amidation, which would have implications for genotype-specific responses to incretin-based therapies. Further, it is also possible that amidation of some peptides, such as CgA, could affect secretion rates in tissues other than the pancreatic beta cell. CgA is expressed across a range of neuroendocrine tissues, and has also been detected in intestinal L-cells (275, 276). Effects of PAM and CgA on granulogenesis could therefore prove to be a more general mechanism across other cell types where both proteins are expressed.

In conclusion, I have shown that the amidating enzyme PAM is a critical component of the regulated secretory pathway in beta cells, consistent with the observed effects of *PAM* missense variants on T2D risk and the insulinogenic index. My results propose a molecular mechanism for T2D-associated *PAM* risk alleles, and reveal multiple effects of reduced PAM function on insulin granule packaging and release from beta cells, potentially via amidation of CgA. This study illustrates the importance of coding variants implicated in disease risk for the identification of causal genes for mechanistic studies. Further work is required to explore the possible contribution of other tissues to PAM function, particularly those that reveal implications for incretin-based therapies.

CHAPTER 5

Systematic Functional Characterisation of Candidate Causal Genes for Type 2 Diabetes Risk Variants

Some of the results presented in the following chapter contributed to a previously published article, which is referenced below. The complete manuscript is provided in the Appendices.

- **S.K. Thomsen** et al, Systematic Functional Characterization of Candidate Causal Genes for Type 2 Diabetes Risk Variants. Diabetes. 2016 Dec:65(12):3805-3811

5.1 Introduction

5.1.1 Unbiased strategies for prioritizing candidate causal genes

Recently, a number of strategies have emerged to facilitate the translation of risk variants for complex traits into causal mechanisms using unbiased and scalable approaches (*Section 1.4*). Broadly, these methods rely on genetic data, genomic annotations and functional genetic screening to prioritize candidate causal genes (**Figure 1**). Importantly, each line of evidence carries a number of limitations, and is rarely sufficient in isolation to confidently identify causal genes. Rather, triangulation from multiple types of evidence can be used to build confidence through convergence on a particular mechanism.

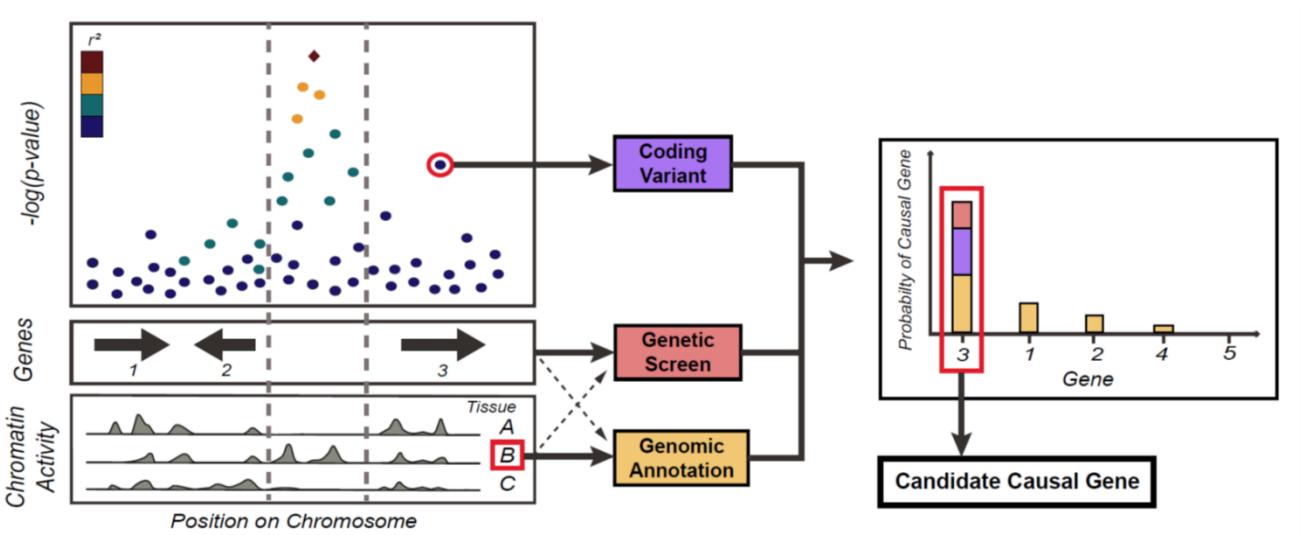


Figure 1 *Prioritizing candidate causal genes for GWAS signals using integration of scalable and unbiased approaches.* (*Left*) A regional association plot for a hypothetical T2D risk locus shows a primary, non-coding association signal with the lead variant (purple diamond) positioned between two genes. An independent association signal within the coding sequence of gene 3 implicates it as a candidate causal gene at the locus. Beyond genetic association data, genomic annotations have revealed the non-coding association signal to overlap a region of chromatin that displays context-dependent activity in tissue "B". Based on this, further studies can be performed in the appropriate disease-relevant context to identify: (1) genomic annotations that can link the non-coding signal to nearby genes; and (2) regulators of disease-relevant processes through genetic screens. In this case, multiple lines of evidence have converged on gene 3 as a likely candidate causal gene that can be taken for more in-depth follow-up studies. *Adapted from Grotz AK, Gloyn AL & Thomsen SK (2017) Prioritizing Causal Genes at Type 2 Diabetes Risk Loci. Current Diabetes Reports (277).*

Firstly, large-scale genetic association studies have the potential to identify coding risk variants that immediately propose candidate effector transcripts (as exemplified in Chapter 4), but this type of evidence is limited by the occurrence of natural variation. Despite an increased focus on exome and whole-genome based genetic association studies, the number of coding variants associated with T2D risk is still relatively low (36, 65). Further, coding signals may be in LD with non-coding variants, and conditional analyses in fine-mapped populations are required to ascertain the likely true causal variant(s) (65).

In contrast, genomic annotations offer the potential to identify direct variant-gene associations that can link non-coding GWAS signals to regional transcripts. These relationships can be either functional (e.g. cis-eQTLs) or structural (e.g. chromatin-chromatin interactions), and tend to be dynamically regulated, varying according to developmental stage, metabolic state or tissue type (79). Large-scale consortia, such as GTEX and the ENCODE project have made significant progress in cataloguing genomic annotations across a large number of cell types, but typically do not include pancreatic islet cells. Since physiological characterisations of carriers of T2D risk variants have implied a central role for islet dysfunction in disease susceptibility, a number of individual studies have attempted to address this gap by identifying islet cis-eQTLs coinciding with non-coding GWAS signals (77, 107, 119, 278, 279). Though a few notable successes have provided proof-of-principle of this strategy, statistical power for wider application has been limited by the scarcity and variable quality of islet preparations (77, 119).

Finally, a complementary strategy to identify candidate causal genes is based on known or observed protein functions that are thought to be relevant in disease pathogenesis. T2D risk variants, for example, 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 a high posterior probability of being the downstream

mediator for a nearby T2D association signal known to impact on islet function. However, the number of loci for which current evidence favours one candidate over others is limiting and tends to be biased towards previously studied genes. High-throughput genetic screening combined with cellular phenotyping holds the potential to overcome this limitation through unbiased identification of novel genes implicated in disease-relevant processes. While any such study will be limited to a particular cell state, focusing on those tissues with high expected relevance to disease, such as islet cells, are likely to be most informative.

5.1.2 Functional genetic screens for regulators of beta cell function

To date, large-scale functional genetic screens in beta cells have been limited by inadequacies of available human cellular models and the high cost of insulin immunoassays (~£1 per data point), the gold standard for measuring insulin. To circumvent these issues, previous efforts have focused on non-human model systems (280-284), reporter-based proxy measurements for insulin (282-284), and/or phenotypes not directly related to insulin production and secretion (281, 285).

The most comprehensive genetic screen in a beta cell model to date was pursued by Pappalardo and colleagues, who performed a whole-genome siRNA screen to identify regulators of insulin transcription (284). Subject to the limitations mentioned above, this study used a reporter-based assay for insulin promoter activity in the mouse beta cell model Min6. Another recent study demonstrated the utility of more comprehensive genetic characterizations in a human beta cell model, EndoC-βH1, but did not achieve sufficient scalability to be applicable for broad prioritization of causal disease genes (a total of 12 lncRNAs were screened by RNAi) (286).

Over the last few years, pooled CRISPR/Cas9 screens have emerged as a robust tool for interrogation of gene function across a number of cell types (287). While this strategy offers

genome-wide scalability, the phenotypes that can be measured are restricted to low-dimensional read-outs, such as cell death/viability or a few protein markers measured by fluorescence-activated cell sorting (FACS). Further, to establish genotype-phenotype correlations in a pooled format, these protein markers are required to be localized either intracellularly or at the cell-surface membrane. This limitation is particularly pertinent in the context of the highly-specialized pancreatic beta cell, where measurement of extracellular insulin secretion is not amenable to pooled screening.

To accelerate the discovery of causal genes for T2D, this study aimed to address the existing limitations in genetic screening for regulators of beta cell function. I first implemented a scalable and robust assay for insulin secretion in the human beta cell line EndoC-βH1 (*Sections 5.3.1-5.3.2*). Based on this work, I applied the pipeline in a primary siRNA screen of 300 positional candidate genes (*Sections 5.3.3-5.3.4*). The results, firstly, allowed internal validation through replication of known mechanisms of beta cell dysfunction; and, secondly, facilitated the prioritization of novel candidate causal genes through integration with complementary lines of evidence. I subsequently validated selected hits in a follow-up study to assess the contribution of off-target effects (*Section 5.3.5*). Finally, I focused on a single gene, *ZMIZ1,* emerging as a strong hit from the primary screen, to explore the pathways affected by ZMIZ1 deficiency in primary human islets (*Section 5.3.6*)

5.2 Materials & Methods

5.2.1 RNA-seq of EndoC-βH1

The EndoC-βH1 cell line was cultured as described (*Section 2.2.1*) and grown to near confluency prior to RNA extraction in TRIzol (*Section 2.4.1*) and RNA-sequencing (*Section 2.5*). Based on raw sequencing data, gene level read counts were quantified as reads per kilobase of transcript per million mapped reads (RPKM) using RNA-SeQC with the "strictMode" flag set (288). The sequencing data for EndoC-βH1 have been deposited at the European Nucleotide Archive (ENA; http://www.ebi.ac.uk/ena) under accession number PRJEB15283.

5.2.2 RNA-seq of primary islets

Human islets were prepared for RNA-seq by Dr Jocelyn Manning Fox at the University of Alberta. Briefly, islets were hand-picked to purity and dispersed using enzyme-free cell dissociation buffer (Thermo Fisher Scientific). Cells were then plated on 35mm dishes and transfected with *ZMIZ1* Silencer Select siRNA (Thermo Fisher Scientific) or a scrambled negative control using DharmaFECT (Thermo Fisher Scientific). After 48hrs, RNA was extracted using TRIzol (*Section 2.4.1*) and prepared for RNA-sequencing (*Section 2.5*). Based on the raw sequencing data, gene level read counts were quantified as counts per million (cpm). The gene set was then filtered to exclude genes on the sex chromosomes and genes not expressed in all samples (cpm > 1), leaving 13,135 genes before normalizing to logCPM values using voom (289). The transformed data were then used to perform principal component analysis and differential expression analysis using Empirical Bayes moderation of a linear model fitted with donor and experiment type as covariates 290).

5.2.3 Transfection optimization

For a test of different transfection conditions, the procedure for gene silencing (*Section 2.3*) was modified to replace RNAiMAX with INTERFERin-HTS (Polyplus transfection) or Lullaby (OZ Biosciences) at the indicated concentrations of transfection reagents and siRNAs (SMARTpools, Dharmacon). Viability was subsequently measured using alamarBlue (Thermo Fisher Scientific) as an indicator of cellular redox levels. Due to the relatively low metabolic activity of the EndoC-βH1, cells were incubated in the recommended concentration of alamarBlue (1:10) for an extended period of 24h before reading fluorescence at 545nm/590nm (ex/em) on a CytoFluor multi-well plate reader (PerSeptive Biosystems). Gene expression was analysed as described (*Section 2.4*) using the following TaqMan assays: *INS* (Hs00355773_m1) and *PAM* (Hs00168596_m1). Optimal transfection conditions were selected to achieve a reasonable trade-off between cell viability (>90%) and silencing efficiency (<10% residual expression).

5.2.4 Cellular phenotyping

Cellular screens were either performed manually or adapted for automated screening on a Perkin Elmer Janus liquid handling workstation based on previously described assays (*Sections 2.3 & 2.6*) (291). 25,000 cells/well were reverse transfected in 96-well format at final siRNA concentrations of 25 nM. Custom libraries of SMARTpool siRNAs were designed for the pilot and primary screens. Follow-up validation was performed using commercially available Silencer Select siRNA pools (Thermo Fisher Scientific) for the following genes: *ABCC8, ARL15, HNF4A, KIF11, MIER3, ZMIZ1, TH,* and *THADA*. Non-targeting sequences from the corresponding suppliers were used as negative controls. Gene expression analyses were performed as described (*Section 2.4*) using the following TaqMan assays: *ABCC8* (Hs00165861_m1), *CAMK1D* (Hs00945341_m1), *CDKAL1* (Hs01554296_m1), *DGKB*

(Hs00391660_m1), *GCK* (Hs00277220_m1), *HHEX* (Hs01074519_m1), *IDE* (Hs00610452_m1), *INS* (Hs00355773_m1), *KCNJ11* (Hs00265026_s1), *KCNK16* (Hs01000699_m1).

5.2.5 Sample analysis

Following manually performed secretion assays, analysis of supernatants, cellular insulin content, and cell counts were performed as described (*Section 2.6*). For the primary screen, supernatants were analysed using the AlphaLISA immunoassay on a PHERAstar FS plate-reader (BMG). Supernatants (50-250 nL) and beads pre-diluted in water (500 nL) were dispensed into ProxiPlate-384 shallow-well microplates (PerkinElmer) with an Echo 550 (Labcyte) acoustic liquid handler, before manual addition of immunoassay buffer to a final volume of 5 µL. Cell counts were measured using the CyQUANT Direct Cell Proliferation assay on an EnVision multimode plate reader with an AlphaScreen module (Perkin Elmer). All responses were normalized as indicated, and expressed as a percentage of non-targeting (NT) control for each phenotype. Effect sizes are given as the percentage difference from NT ($Response_{\text{Gene}}$ - $Response_{\text{NT}}$), and the absolute values hereof ($|Response_{\text{Gene}}$ - $Response_{\text{NT}}|$).

5.2.6 Statistical analysis

Data analyses were performed using R version 3.3.1. To identify significant responses, cell counts and normalized insulin secretion measurements for each gene were compared to non-targeting control using unpaired, two-sample t-tests, The false-discovery rate (FDR) was controlled at 5% by applying the Benjamini-Hochberg procedure to produce adjusted p-values (q-values) for each phenotype. The Z-factor measuring the control response for each phenotype was calculated as $Z' = 1 - \frac{3(\sigma_{INS}+\sigma_{NT})}{|\mu_{INS}-\mu_{NT}|}$. Hierarchical clustering of Z-scores derived from the primary screen was performed using complete linkage clustering of Euclidian distances, as implemented in the gplot R package.

5.3 Results

5.3.1 Establishing a high-throughput screening pipeline

As previously discussed, physiological studies have highlighted beta cell dysfunction as a central mechanism for a majority of T2D risk alleles. I therefore reasoned that a genetic screen for disease-relevant phenotypes in the human beta cell line EndoC-βH1 could capture some of the underlying molecular defects at these loci, allowing the causal genes to be identified. I first designed an experimental strategy for parallel gene silencing and subsequent interrogation of five molecular phenotypes. In overview, selected targets were silenced in arrayed format on microplates using pools of four siRNAs per gene. Three days after knockdown, cells were measured for five end-point phenotypes; total cell number, and insulin secretion under different conditions, including low glucose (1 mM), high glucose (20 mM), and high glucose with tolbutamide (100 μM), or IBMX (100 μM). Low and high glucose conditions were included to simulate fasted and fed states *in vivo*. Tolbutamide and IBMX act on the depolarizing and the potentiating pathways of insulin secretion, respectively, and were included to provide additional mechanistic insights through modulation (e.g. synergy or pharmacologic rescue) of any primary defects observed in low or high glucose.

To achieve optimal sensitivity and minimise the potential influence of technical artefacts, I first optimized transfection conditions for a single, representative gene (*PAM*) based on cell viability and gene expression. While the exact parameters required for optimal gene silencing efficiency are likely to be dependent on the properties of the specific gene being targeted, defining the optimal transfection conditions across all genes would be unpractical. On the basis of a total of more than 60 different conditions, parameters were identified that resulted in greater than 90 % viability and less than 10 % residual *PAM* expression (**Figure 2**).

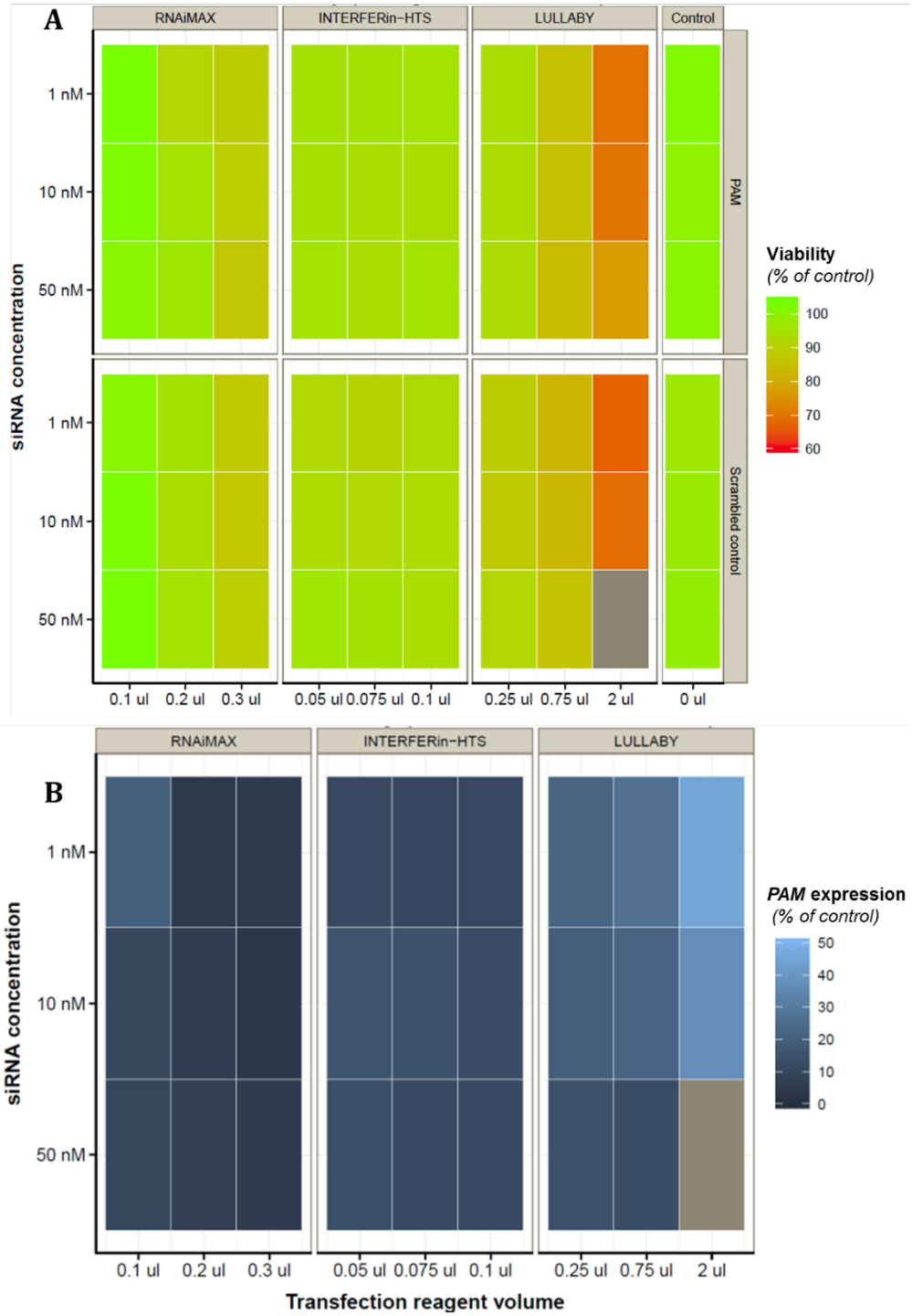


Figure 2. *Optimizing knockdown efficiency and viability of* PAM *knockdown.* Three different transfection reagents (RNAiMAX, INTERFERin-HTS and LULLABY, as well as no-reagent "control") were tested in three concentrations against three siRNA concentrations (1, 10 and 50 nM) of both *PAM* and non-targeting control. (*A*) Cell viability 72 h after transfection. AlamarBlue was used as an indicator of cellular redox levels, and fluorescence measured following 24 h treatment. Cellular viability was then obtained by first background-correcting and then normalizing values to no-transfection controls. (*B*) *PAM* residual expression following 96 h knockdown. RNA was harvested after four days, and expression assessed by qPCR relative to two housekeepers (*TBP* and *PPIA*) and normalized to corresponding non-targeting controls.

5.3.2 Validation of screening pipeline

Using the insulin gene (*INS*) as a positive control, I confirmed that I was able to robustly detect the effects of gene silencing on the phenotypes of the assay (**Figure 3**). As a quantitative measure of robustness, the Z-factor (Z', a standard indicator of statistical effect size in high-throughput screens) was derived for all secretory phenotypes: Z' (low glucose) = 0.75, Z' (low glucose) = 0.68, Z' (IBMX) = 0.51, Z' (Tol) = 0.44. With a Z' > 0.5 generally considered an excellent assay, this highlighted three of four conditions as appropriate for implementation on a high-throughput scale. Using this stringent cut-off (corresponding to 12 standard deviations between positive and negative controls), tolbutamide stimulation was considered to produce a borderline (or marginal) phenotype.

Encouraged by these findings, I next performed a manual (non-automated) pilot screen to establish scalability across multiple plates for a selected subset of phenotypes (basal and high glucose). Further, to assess the sensitivity of detecting 'hits' across a range of expected effect sizes, I designed a target library enriched for genes with prior evidence of a role in beta cell function (**Table 1**). This semi-systematic selection strategy considered all genes located in close proximity to a T2D association signal (<1 Megabase [Mb]), which could conceivably be subject to cis-regulation by non-coding risk variants (see also *Section 5.3.3*). I then prioritized genes previously implicated in monogenic diabetes (Online Mendelian Inheritance in Man, phenotype MIM numbers 606391 [MODY] and 606176 [PNDM]), or genes that had been reported as targets of islet cis-eQTLs overlapping risk variants for T2D (77, 119).

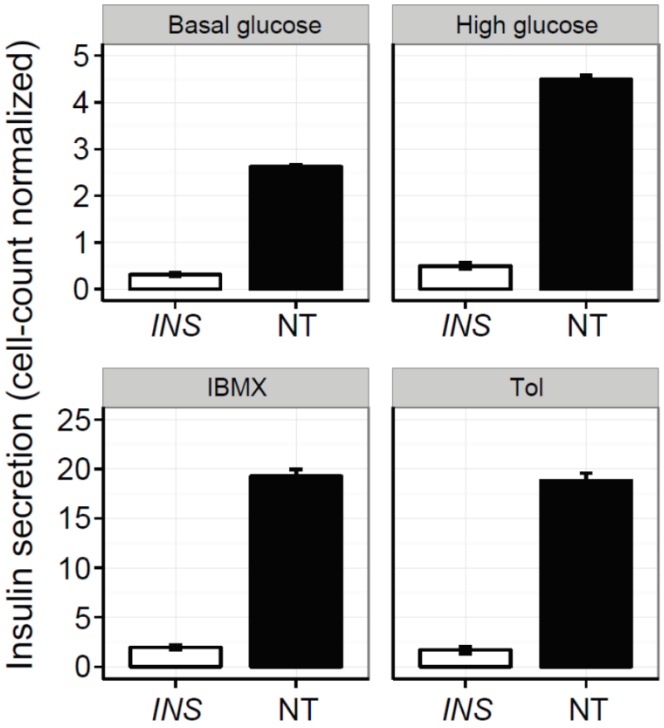


Figure 3 *Validation of screening assay using INS control.* Insulin secretion under the indicated conditions following knockdown of insulin (*INS*) positive control (white bars) and non-targeting (NT) negative control (black bars). All measurements were normalized on a per-well basis to cell-counts, and then averaged across conditions. Data points are the mean of n = 3, and error bars are SEM. Tol = Tolbutamide.

Locus	Gene
ADCY5	*ADCY5*
AP3S2	*AP3S2*
CAMK1D	*CAMK1D*
CDKAL1	*CDKAL1*
DGKB	*DGKB*
GCK	*CAMK2B*
	GCK
	YKT6
HHEX/IDE	*HHEX*
	IDE
	KIF11

Locus	Gene
KCNJ11/ABCC8	*ABCC8*
	KCNJ11
	NUCB2
KCNK16	*KCNK16*
	KCNK17
KCNQ1	*KCNQ1*
NA	*INS*
SLC30A8	*SLC30A8*
TCF7L2	*TCF7L2*
WFS1	*WFS1*
ZMIZ1	*ZMIZ1*

Table 1. *Genes selected as targets for the pilot screen library.* Genes are listed according to their T2D risk locus affiliation, listed in alphabetical order. For some loci, multiple genes were selected for inclusion, either where more than one plausible candidate was located near an association signal (e.g. at the *KCNK16* locus), or to include a negative control where the causal gene for T2D is known with a high degree of certainty (e.g. the *KCNJ11/ABCC8* locus).

Based on content-normalized data, average insulin secretion rates were calculated across all genes (**Figure 4**). To identify those genes for which insulin secretion was significantly affected by knock-down, two-sample t-tests were performed and the Benjamini-Hochberg procedure was used to control the false-discovery rate (FDR). At an FDR of 5 %, nine and ten 'hits' were established for basal and high-glucose insulin secretion, respectively (**Table 2**); this included both negative and positive regulators of these phenotypes. Each of the nine genes shown to affect basal secretion were also found to impact significantly on high-glucose secretion with a consistent direction of effect.

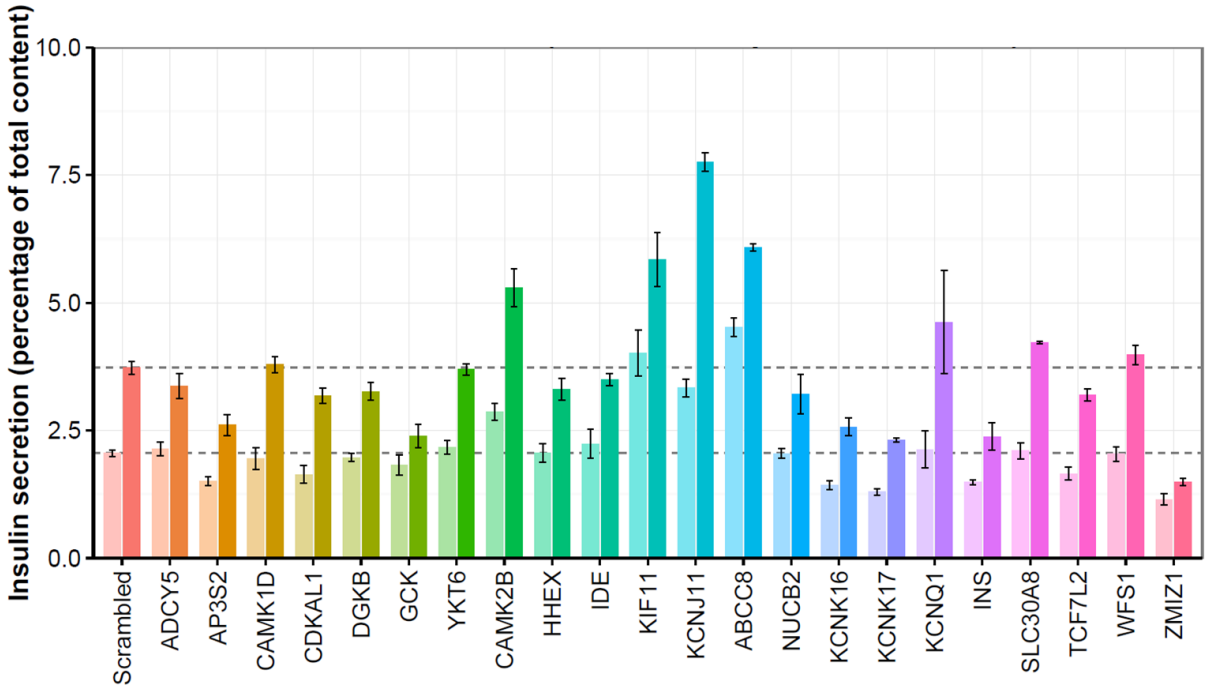


Figure 4. *Pilot screen for identification of genes influencing insulin secretion.* All genes were assayed in triplicate for each condition, except for scrambled, which was assayed for n = 8. Genes are ordered alphabetically according to locus assignment with scrambled control on the left. All bars represent the mean of normalized basal (pale) and high-glucose (dark) insulin secretion, with dotted lines highlighting scrambled levels for ease of reference. Error bars are SEM.

		Basal-glucose insulin secretion		High-glucose insulin secretion	
Locus	**Gene**	Mean	SD	Mean	SD
Control	Scrambled	20.5	3.3	37.3	6.1
ADCY5	*ADCY5*	21.3	2.3	33.7	4.2
AP3S2	*AP3S2*	**15***	1.6	**26.1***	3.6
CAMK1D	*CAMK1D*	19.5	3.8	37.9	2.8
CDKAL1	*CDKAL1*	16.4	3.1	31.8	2.7
DGKB	*DGKB*	19.7	1.4	32.6	3.1
GCK	*GCK*	18.2	3.4	**23.9****	4.1
	YKT6	21.6	2.3	36.9	2
	CAMK2B	**28.6***	2.3	**53****	5.2
HHEX/IDE	*HHEX*	20.6	3.1	33.1	3.7
	IDE	22.4	4.8	35	2.2
	KIF11	**40.2*****	7.8	**58.5*****	9.2
KCNJ11	*KCNJ11*	**33.4*****	3	**77.6*****	3.1
	ABCC8	**45.2*****	3.2	**60.9*****	1.2
	NUCB2	20.5	1.6	32.1	6.7
KCNK16	*KCNK16*	**14.3***	1.5	**25.7****	3.1
	KCNK17	**13****	1.1	**23.1****	0.7
KCNQ1	*KCNQ1*	21.3	5.1	46.2	14.3
NA	*INS*	**14.8***	0.8	**23.8****	4.8
SLC30A8	*SLC30A8*	21	2.8	42.2	0.5
TCF7L2	*TCF7L2*	16.6	2.2	31.9	2.1
WFS1	*WFS1*	20.4	2.5	39.8	3.3
ZMIZ1	*ZMIZ1*	**11.5*****	1.8	**14.9*****	1.3

Table 2. *Results table for pilot screen in EndoC-βH1.* Genes included in the screen are listed according to locus assignment. Results given are means and standard deviations (SD) for basal-glucose insulin secretion and high-glucose insulin secretion. All tests performed are two-sample t-tests for non-targeting control vs. the gene of interest, and p-values were corrected using the BH-procedure. Significant results are bolded and asterisks indicate the level of significance for corrected p-values (q-values); * for $q < 0.05$, ** for $q < 0.01$ and *** for $q < 0.001$.

At the *KCNJ11/ABCC8* locus, for example, both of the well-known monogenic diabetes genes, *KCNJ11* and *ABCC8,* were identified as hits for increased basal and high-glucose insulin secretion, whereas a third unrelated gene, *NUCB2*, was not. *KCNJ11* and *ABCC8* encode subunits of the ATP-sensitive potassium channel K_{ATP} (*Section 1.5.1*), and LOF mutations are known to cause congenital hyperinsulinism, consistent with the direction of effect observed in the pilot screen (292). Similarly, *GCK* emerged as a strong hit for decreased insulin secretion under high glucose conditions. This is in line with GCK being an important glycolytic enzyme that is required for initial priming of glucose (*Section 1.5.1*). In humans, *GCK* haploinsufficiency alters the threshold for glucose-stimulated insulin secretion, and thus reduces insulin secretion leading to mild hyperglycemia (293). In addition to these positive controls, a number of potential candidate genes were found to significantly alter insulin secretion, including *ZMIZ1*, *CAMK2B* and *KCNK16*/*17*.

Having established a robust experimental pipeline, I finally sought to reduce the cost of sample analysis by making use of acoustic liquid handling to miniaturize insulin immunoassays. As discussed previously, a central limitation of previous screens for beta cell dysfunction has been the prohibitively high cost of insulin immunoassays (~£1/data-point). Using automated, acoustic dispensing I performed nanoliter insulin measurements (~50 nL sample input), producing a ten-fold reduction in the cost of sample analysis (£0.10/data-point). The coefficient of variation (CV) was found to increase exponentially at low concentrations, and an optimal dilution range (1,000-10,000 pg/mL) that resulted in a CV < 3 % could be empirically determined [CV] < 3%; **Figure 5**).

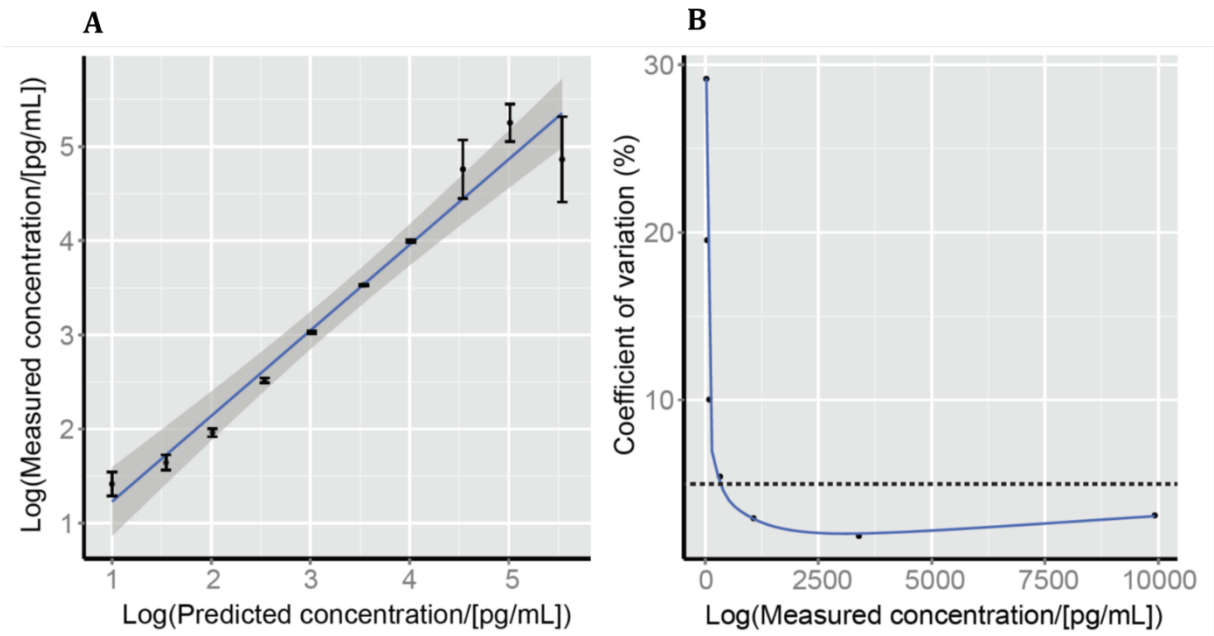


Figure 5 *Optimization of nanoliter insulin immunoassays.* Using a range of known insulin concentrations, the concentration following dilution by acoustic liquid handling was measured by AlphaLISA insulin immunoassays. (A) Measured versus predicted insulin concentration (in pg/mL; both log-transformed). A linear regression (*blue*) and 95% confidence intervals for the linear fit (*grey*) are shown, highlighting an excellent correlation for predicted concentrations between 100 and 10,000 pg/mL. (B) The CV for a range of measured concentrations of insulin. A logarithmic regression has been fitted to the data, and the dashed line indicates a CV of 5 %. At low input concentrations, the CV rises exponentially, but the optimal dilution range (1,000-10,000 pg/mL) for insulin samples could be determined for a CV < 3 %. Data points are the mean of n = 4, and error bars are SEM.

5.3.3 RNA-seq of EndoC-βH1 and target selection

Based on this combined assay and analysis pipeline, I designed a primary screen to assess the role of positional candidate genes for T2D GWAS loci in beta cell function (**Figure 6A**). For target selection, the following strategy was applied (**Figure 6B**); first, all protein-coding genes within 1 Mb of a T2D association signal were considered for inclusion. Since previous studies have shown cis-eQTLs to form a relatively tight, symmetrical distribution around the target-gene transcription start site, this 1 Mb cut-off was defined to capture most effector transcripts subject to cis regulation (118, 294-296).

Next I performed RNA-seq of the EndoC-βH1 to exclude genes not expressed in the cellular model (RPKM = 0), thereby minimizing unproductive silencing. I further removed genes not expressed in a published dataset for enriched primary human beta cells to focus on genes that could theoretically mediate a causal role through beta cell function *in vivo* (107). As expected for a relevant cellular model system, the expression data for EndoC-βH1 correlated strongly with primary beta cells (ρ = 0.78; **Figure 7**) and showed robust expression of key beta cell genes (107, 157) (**Table 3**).

Finally, for loci with fewer than five remaining genes, all candidates were represented in the library for the primary screen. For loci with five or more qualified genes, I implemented a systematic selection strategy developed by Juan Fernandez-Tajes and Martijn van de Bunt (unpublished) to reduce the search space where available evidence strongly favoured a single candidate (based on coding association signals for T2D, islet cis-eQTLs overlapping an association signal, and/or an established role in beta cell function based on automated text mining of PubMed abstracts). In these cases, two genes with no prior evidence of causality were also included per locus. Overall, 300 genes were selected for inclusion (**Appendix 2**).

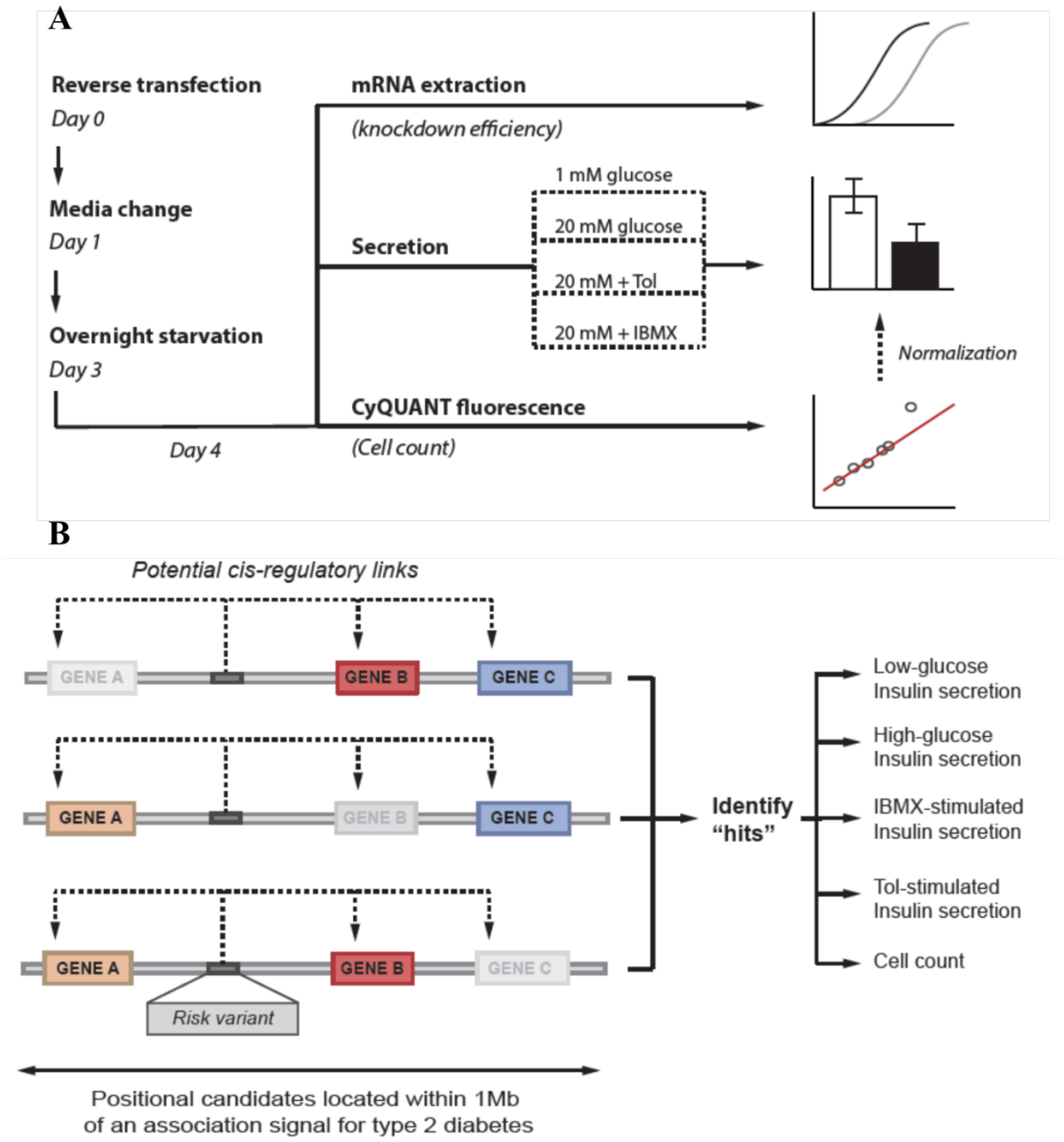


Figure 6 *Overview of assay and screening strategy.* (A) At day 0, cells in 96-well format were transfected with pools of four siRNAs per gene. Seventy-two hours after knockdown, cells were starved overnight before measurements of the indicated phenotypes on day 4. For a subset of genes, mRNA was extracted to assess knockdown efficiency. For all genes, cell counts and insulin secretion were measured under four different conditions. (B) Based on the assay outlined above, we developed a screen of positional candidate genes located within 1Mb of a genome-wide significant signal for T2D risk. For the hypothetical locus shown, the three protein-coding genes *A*, *B* and *C* would be silenced in parallel. Each gene would then be assayed for multiple disease-relevant phenotypes, and significant effects identified according to a stringent FDR-based threshold (see main text for details).

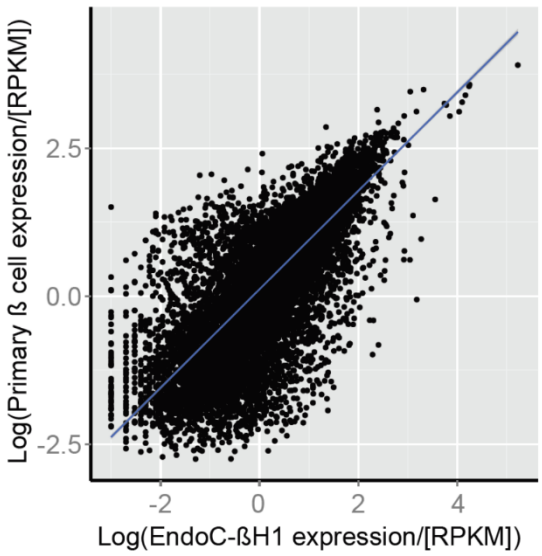


Figure 7 *RNA-seq of sorted primary beta cell versus EndoC-βH1.* Expression levels in a published data-set for sorted primary human beta cells versus expression levels for the EndoC-βH1 cell line as measured by RNA-seq (1). All values are log-transformed RPKM, and a linear regression (*blue*) has been fitted to the data.

Ensembl gene ID	Gene Name	EndoC-βH1 (RPKM)	Primary beta cells (RPKM)
ENSG00000254647.2	*INS*	8012.452	166347.111
ENSG00000087086.9	*FTL*	3090.7454	2066.464
ENSG00000100604.8	*CHGA*	2853.107	1113.162
ENSG00000099194.5	*SCD*	1411.631	238.714
ENSG00000118271.5	*TTR*	1312.8928	1488.996
ENSG00000184009.5	*ACTG1*	1108.0277	839.681
ENSG00000124172.5	*ATP5E*	863.42334	254.779
ENSG00000156508.13	*EEF1A1*	831.17316	637.281
ENSG00000112972.10	*HMGCS1*	719.7504	22.104
ENSG00000167552.9	*TUBA1A*	689.1415	511.699

Table 3 *Top ten most highly expressed protein-coding genes in the EndoC-βH1 cell line.* Expression levels provided for non-mitochondrial genes in EndoC-βH1 and the corresponding expression levels in sorted primary human beta cells (107).

5.3.4 Primary screen

Adapting the cellular assays for automated screening on a robotic liquid handling station, I performed the primary screen in triplicate and assessed performance using standard quality control measures. Since validation of silencing across the entire library would be unpractical, I included the essential gene *PLK1* as a positive control, and visibly confirmed extensive cell death across all conditions. In a subset of 16 genes, I determined residual expression at the transcript level, and found the median knockdown efficiency to be 57% (**Figure 8**), roughly equivalent to mono-allelic LOF.

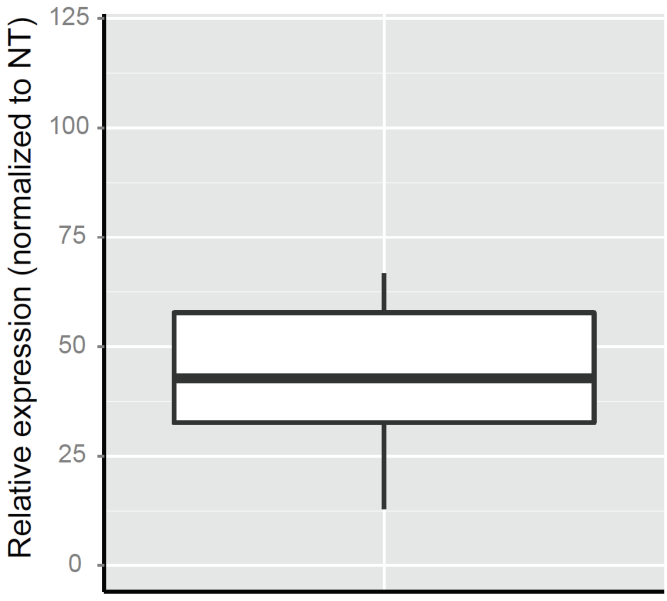


Figure 8 *Knockdown efficiency observed in the primary screen.* Boxplot of residual expression levels for a subset of 16 genes included in the primary screen following 96 h knockdown. Expression levels were measured by quantitative PCR (qPCR) using commercially available TaqMan probes, and normalized first to the two housekeeping genes (*TBP* and *PPIA*) and then to NT control. The genes included for assessment of knockdown efficiency were: *ABCC8*, *CAMK1D*, *CDKAL1*, *DGKB*, *GCK*, *HHEX*, *IDE*, *INS*, *KCNJ11*, *KCNK16*, *KCNK17*, *KIF11*, *PAM*, *TCF7L2*, *YKT6*, *ZMIZ1*.

To account for differences in plating efficiency and proliferation, cell counts were used to normalize insulin secretion data on a per-well basis. Average values were then expressed as the percentage deviation from non-targeting control for each phenotype. Quantile-quantile (Q-Q) plots showed these standardised phenotypes to be approximately normally distributed, with tails for each of the insulin secretion measures suggesting enrichment for non-random effects (**Figure 9**). To identify significant hits, the cell count and insulin secretion responses for each gene were compared to non-targeting control using Student's two-sample t-test, and the Benjamini-Hochberg procedure was used to produce adjusted p-values (q-values). Two criteria were then applied to identify robust effects ("hits"): (1) an FDR-adjusted q-value < 0.05; *and* (2) an absolute effect size among the top 5%. This identified a total of 67 hits (15 for cell count and 52 for insulin secretion phenotypes) among 45 genes at 37 loci (**Table 4**).

For differences in cell numbers, the effect sizes for each gene were estimated based on 12 independently plated replicates (four conditions in triplicate), and therefore likely represent true biological effects rather than systematic plating artefacts. Further, due to the end-point nature of the assay, any difference in cell numbers could represent defects in either proliferation and/or viability. Aside from *KIF11* and *PLK1*, which are genes with a known role in cell division, the largest effect sizes compared with control were observed for *ZMIZ1* (-15.2%; $q = 6.5 \times 10^{-5}$) and *PRDX3* (+16.5%; $q = 9.2 \times 10^{-5}$). The monogenic neonatal diabetes gene *GLIS3* was also found to be associated with significantly decreased cell numbers (-10.3%, $q = 4.6 \times 10^{-3}$), consistent with reports linking reduced expression to beta cell apoptosis (297, 298).

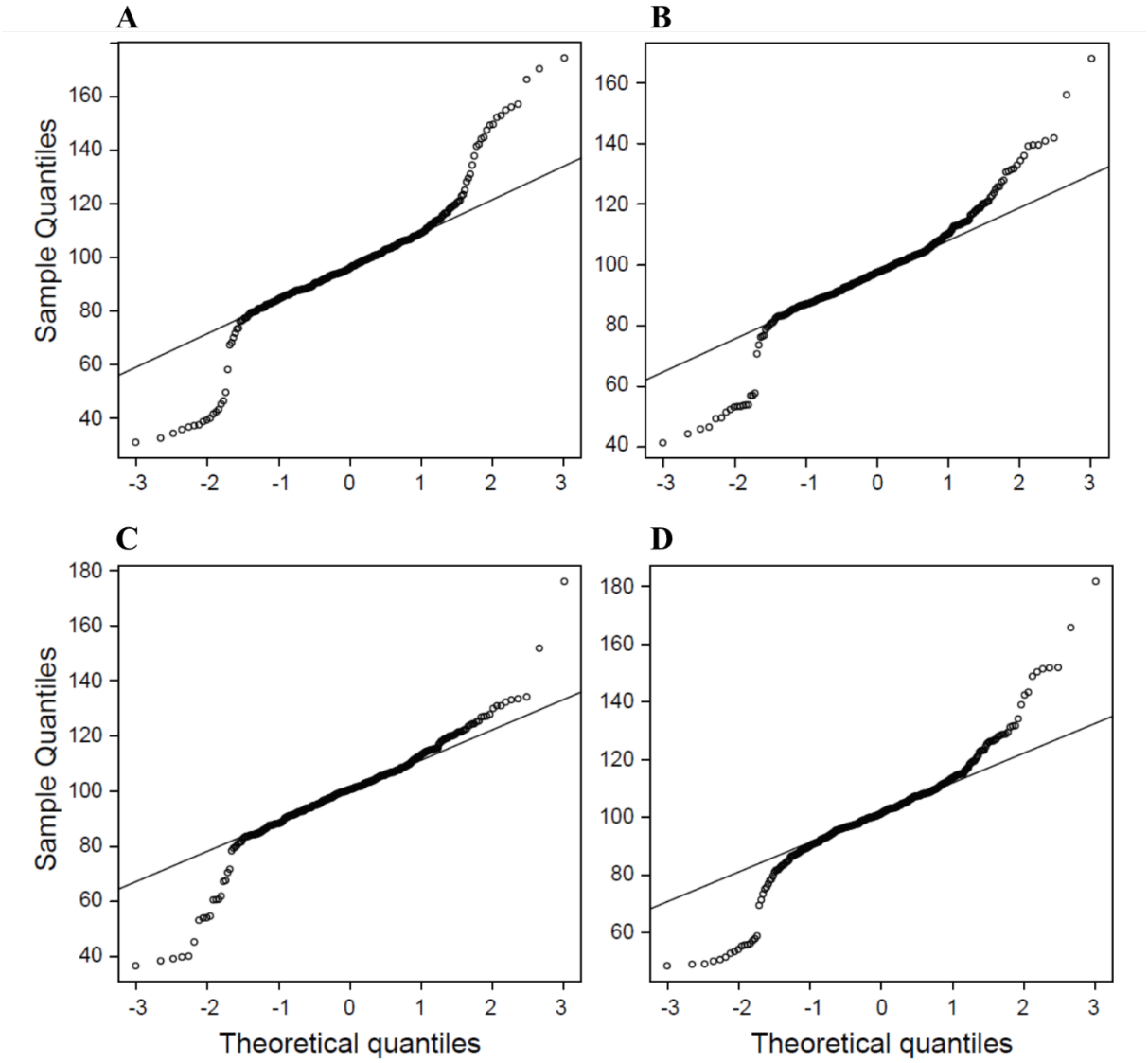


Figure 9 *Q-Q plots for normalized insulin secretion data.* Sample quantiles for insulin secretion under (A) low glucose, (B) high glucose, (C) IBMX, and (D) Tolbutamide (Tol) versus theoretical quantiles for normally distributed variables. A linear fit (*black*) is shown for each of the four plots to indicate regions of approximate normality.

Table 4 *(next page). Effects of significant hits identified in a primary screen for beta cell dysfunction.* The table lists effect sizes (% deviation from control) for each gene with a least one significant effect across the five phenotypes measured. All insulin secretion measurements were normalized on a per-well basis to cell-counts, and the mean percentage deviations from control were then calculated for each condition. For cell counts, values were median-normalized for interplate differences, and the mean percentage deviations from control were calculated across conditions. *q < 0.05 by Student's t-test (FDR-adjusted).

Gene	Locus	Low glucose	High glucose	IBMX	Tolbutamide	Cell count
ABCC8	*KCNJ11*	48.2*	24	26.7*	1.8	-1.4
ADAMTS9	*ADAMTS9*	6.3	-4.8	2.8	-8	12.2*
ADIPOQ	*ST64GAL1*	87*	23.2	23.1	8.8	-7.3
ARL15	*ARL15*	-5.5	-25.9*	-2.1	-15.5	-1.5
BCAR1	*BCAR1*	5.9	25.2	28.5*	9.5	-7
BCL6	*LPP*	-20.7*	-8.9	-1	-12	5.6
BMP8B	*MACF1*	7.8	16.5	9.5	26.9*	-1
CCNT2	*TMEM163*	-32.6*	1.8	4.7	5	-2.8
CDKAL1	*CDKAL1*	2.4	1.7	-10.6	-23.5*	7.5
DGKQ	*MAEA*	3.3	17.5	20.4	33.6*	-9.8
DMRTA2	*FAF1*	32.3*	24.5	13.1	22.7	-0.3
ELAVL4	*FAF1*	-3.6	9.1	21.2	22.8	-11.4*
ETV5	*IGF2BP2*	-12.4	-25.6*	-10.9	-12.6	-2.5
FAH	*ZFAND6*	-23.1*	-20.6*	-16.9	-27.2	-2.3
FBXW7	*TMEM154*	45.2*	9.6	8.7	11.7	9.9*
GINS4	*ANK1*	-16.1	-14.2	-9.5	-17	8.7*
GLIS3	*GLIS3*	-13	-10.6	6.4	-9.3	-10.3*
HEYL	*MACF1*	29.2	29.2*	0.4	20	-7.2
HMGA2	*HMGA2*	16.5	4.1	14.2	24.1*	-1.2
HNF1A	*HNF1A*	21.7	38.3*	23.1	25.5*	5.2
HNF4A	*HNF4A*	36.7*	66.9*	74.9	92.9*	-8.8
IGF2	*DUSP8*	-26.3*	-10.4	1.2	0.1	-1.3
INS	*DUSP8*	-53.5*	-44.8*	-48.7	-38.7*	0.9
KCNK17	*KCNK16*	-9.4	-17.3	-2.5	-1.2	9.8*
KCTD15	*PEPD*	21.7	9.4	12.5	0.8	-10.9*
KIF11	*HHEX/IDE*	45.4*	35*	26.9	55.4*	-40.1*
LINGO1	*HMG20A*	0	19.1	-12.4	-14.9	7.9*
MFGE8	*AP3S2*	30*	3.4	2.8	-1.7	-5.7
MIER3	*ANKRD55*	5.9	36.5*	5.8	18.2	1.9
NDUFS4	*ARL15*	1.6	-4.9	3.9	-1.7	10.8*
PABPC1L	*HNF4A*	-9.7	-12.3	-10.4	-28.8*	-1.2
PHF23	*SLC16A11*	-25.6*	-1.7	-5.9	3.5	-0.9
PLA2R1	*RBMS1*	8.6	1.6	10.8	1.6	9.1*
PRDX3	*GRK5*	24	31.7*	23.4	9.6	16.6*
PTHLH	*KLHDC5*	-2.8	-6.5	-0.5	-25*	-5.9
RND3	*RND3*	-8.7	-6.1	0	-3	-14.9*
SLC2A4	*SLC16A11*	14.5	0	27.2*	18.2	-1.6
SOCS7	*HNF1B*	3.9	-18.5*	11.2	-14.7	-1.6
SPPL3	*HNF1A*	-11.9	-21.9*	-6.1	-10	-10.8*
STK38L	*KLHDC5*	15.2	40.9*	4.7	25.2*	-3
THADA	*THADA*	-1.9	6.5	27.5	24.8*	-10.3
TLE1	*TLE1*	4.3	-5	-23*	16.2	4.6
TM6SF2	*CILP2*	-22.6*	-8.3	-0.8	-12	-8.7
UPF2	*CDC123*	7.5	-12.5	4.7	-24.9*	-3.2
ZMIZ1	*ZMIZ1*	-29.5*	-21.4*	-19.8	-16.8	-15.2*

Table 4 *(continued).* See legend on previous page.

For the insulin secretion data, I first performed an enrichment analysis for genes implicated in MODY. MODY describes a collection of monogenic subtypes of diabetes, characterized by insufficient release or production of insulin. As would be expected for a set of *bona fide* regulators of beta cell function, I observed a strong enrichment of MODY genes among the significant hits ($p = 5.5x10^{-9}$*;* Fisher's exact test). Aggregating absolute effect sizes for MODY and non-MODY genes revealed this enrichment to be driven by altered insulin secretion, and not through effects on cell numbers (**Figure 10**). Specifically, mean absolute effects for MODY genes were increased by more than 60% for each of the secretory phenotypes (q-values < 0.001), but remained unaffected for cell counts (q = 0.67).

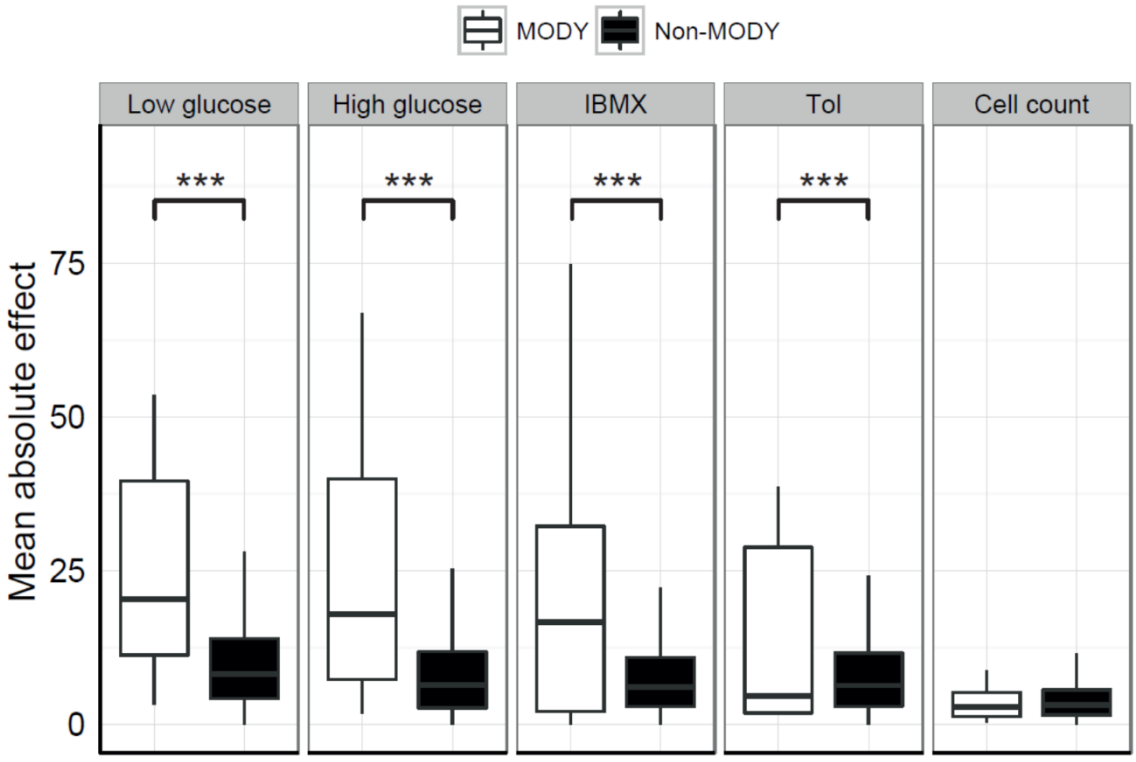


Figure 10 *Comparing absolute effect sizes for MODY and non-MODY genes.* Boxplots of mean absolute effects for MODY genes (*white*) and non-MODY genes (excluding controls, *black*) across the five phenotypes measured. Effect sizes were calculated as described for table 1, and the absolute values were then averaged for the two categories of genes. Among 14 identified MODY genes, eight fulfilled criteria for inclusion in the screen: *HNF4A, GCK, HNF1A, HNF1B, PAX4, INS, ABCC8* and *KCNJ11.* Boxplots show median and interquartile ranges for groups of n = 8 and 292 data-points. ***q-value < 0.001 by Student's t-test (FDR-adjusted). Tol = Tolbutamide.

Further validating the secretion data, I observed strong positive correlations between the normalized responses across conditions (p-values < $2.2x10^{-16}$; **Figure 11A**), and found 10 of 35 genes to cause significant effects under two or more conditions. This included four known MODY genes and *ZMIZ1*, which, even after correcting for the negative effect on cell numbers, remained one of the strongest hits for reduced insulin secretion ($q < 0.01$ for low and high glucose). Such broad effects, independent of specific conditions, hint at cellular defects that are central to the regulation of insulin secretion. In other cases, the pattern of modulation by secretion conditions can be used to pinpoint specific biological pathways affected by gene silencing. For instance, knockdown of the *ABCC8* gene, which encodes a subunit of the ATP-sensitive potassium channel, was found to significantly increase insulin secretion under low glucose and IBMX stimulation. The depolarization caused by silencing of *ABCC8* was masked under high glucose (since the cells would already be partially depolarized), and fully rescued by tolbutamide (due to full, pharmacological depolarization of the cells). To explore the relationship between conditions in greater detail, I performed clustering analysis on Z-scores derived from the normalized secretion values. This revealed high glucose and tolbutamide to be most similar in terms of modulating knockdown effects, with low glucose and tolbutamide being most dissimilar (**Figure 11B**).

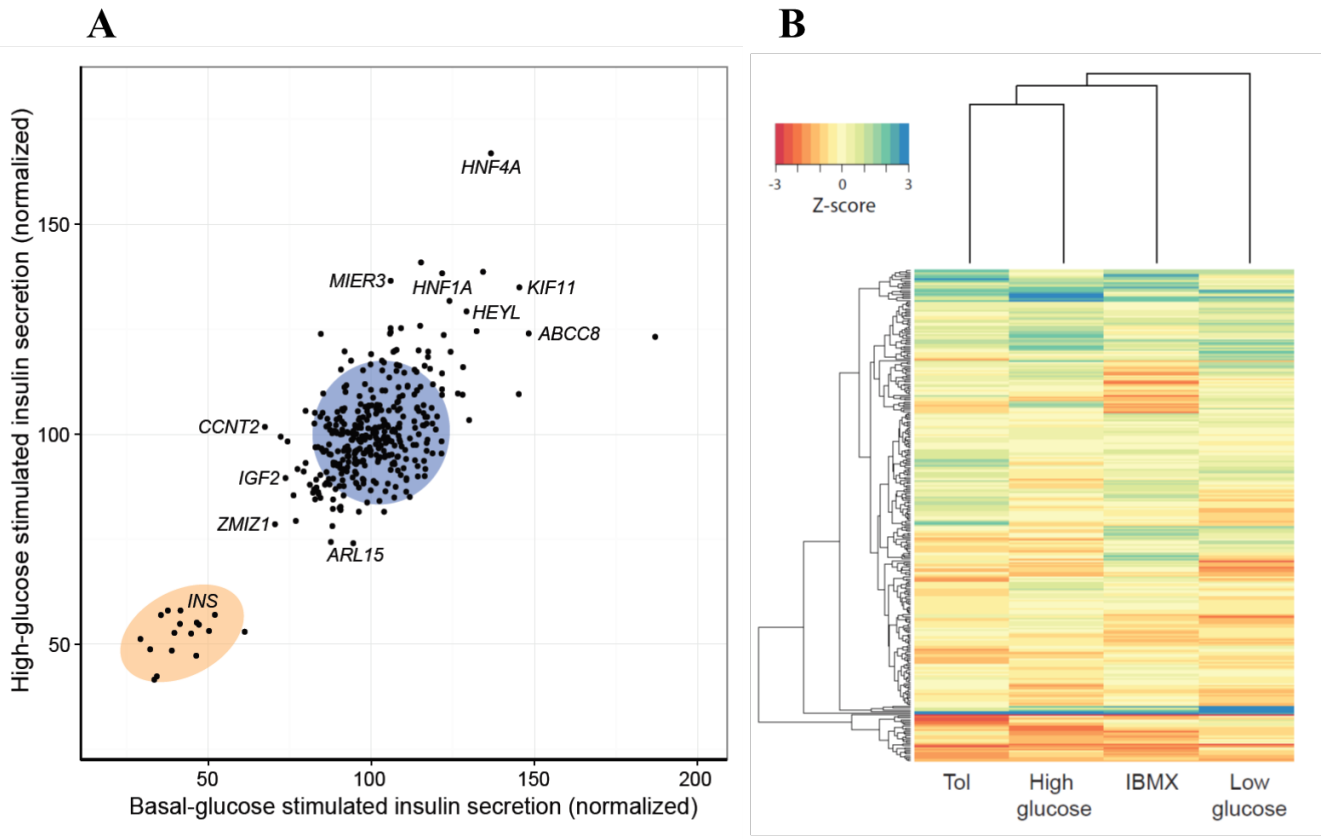


Figure 11 *Comparison of insulin secretion data under different conditions.* (A) Normalized insulin secretion responses under high glucose versus low glucose, with selected hits annotated. The blue circle indicates the 95 % confidence contour for non-targeting control, and the orange circle indicates the 95 % confidence contour for insulin (*INS*) positive controls. All measurements were normalized on a per-well basis to cell-counts, and averages for each condition were then subsequently normalized to the mean of non-targeting control. Data points are mean of n = 3 and shown as percentage of control. (B) Dendrograms and heatmap of Z-scores derived from normalized insulin secretion values with length indicating the Euclidian distance between gene effects (vertical) or conditions (horizontal). Conditions (or genes) clustered together tend to similarly modulate the effect of genes (or behave similarly across conditions). Calculated distances between conditions (low glucose, “A”; high glucose, “B”; IBMX, “C”; Tolbutamide, “D”): A-B = 16.2, A-C = 18.2, A-D = 19.0, B-C = 17.8, B-D = 15.9, C-D = 16.6.

5.3.5 Follow-up validation

I next assessed the contribution of off-target effects by performing a small-scale validation experiment using siRNAs designed with an alternate algorithm. The sequences were confirmed to be different from those of the primary screen, and could thus be used to establish the biological relevance of positive hits. I selected eight target genes, representing hits for both positive and negative defects across the four conditions, and confirmed that knockdown efficiency was satisfactory (median knockdown efficiency = 80.7 %; **Figure 12A**). Compared with insulin secretion results from the primary screen, I observed an excellent linear correlation (ρ = 0.85, p = 6.7×10^{-10}, **Figure 12B)** and 88% directional consistency in normalized responses. The validated hits included several genes with limited prior evidence of a role in the regulation of beta cell function, including; *ARL15* and *ZMIZ1*, which were found to significantly reduce insulin secretion across conditions (q-values < 0.05; **Fig. 13A-B**), and *THADA,* which modestly elevated insulin secretion across three conditions, though the effect under low glucose was not observed in the primary screen (q = 5.6×10^{-3}; **Fig. 13C**). Interestingly, gene silencing of the known MODY gene *HNF4A* was confirmed to cause a paradoxical increase in insulin secretion across all four conditions tested (q-values < 0.001, **Fig. 13D**).

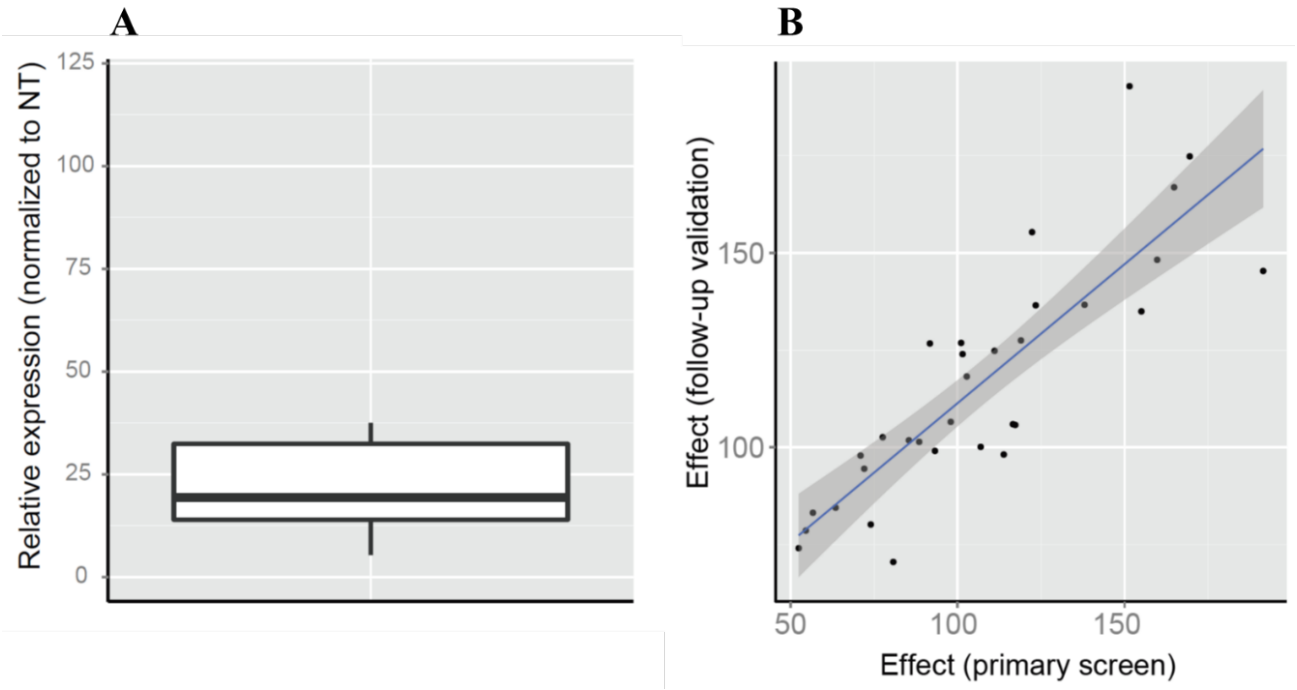


Figure 12 *Quality control measures in follow-up validation*. (A) Boxplot of residual expression levels for 8 genes taken forwarded for validation following 96 h knockdown. (B) Effect sizes for the eight selected hits in a follow-up experiment versus corresponding effect sizes for the same subset of genes measured in the primary screen. The genes taken forward for validation (*THADA, ARL15, MIER3, ZMIZ1, KIF11, ABCC8, HNF4A* and *TH*) were selected to represent targets with varying degrees of prior evidence supporting causality and effect sizes of different magnitudes across conditions. Measurements were processed as described for Figure 11. A linear regression (blue) and 95% confidence intervals for the linear fit (grey) are shown. The follow-up experiment was performed using pools consisting entirely of different sequences of siRNA. The strong correlation observed (r^2 = 0.72) suggests that off-target effects had minimal or no effect on the recorded effects for this subset genes.

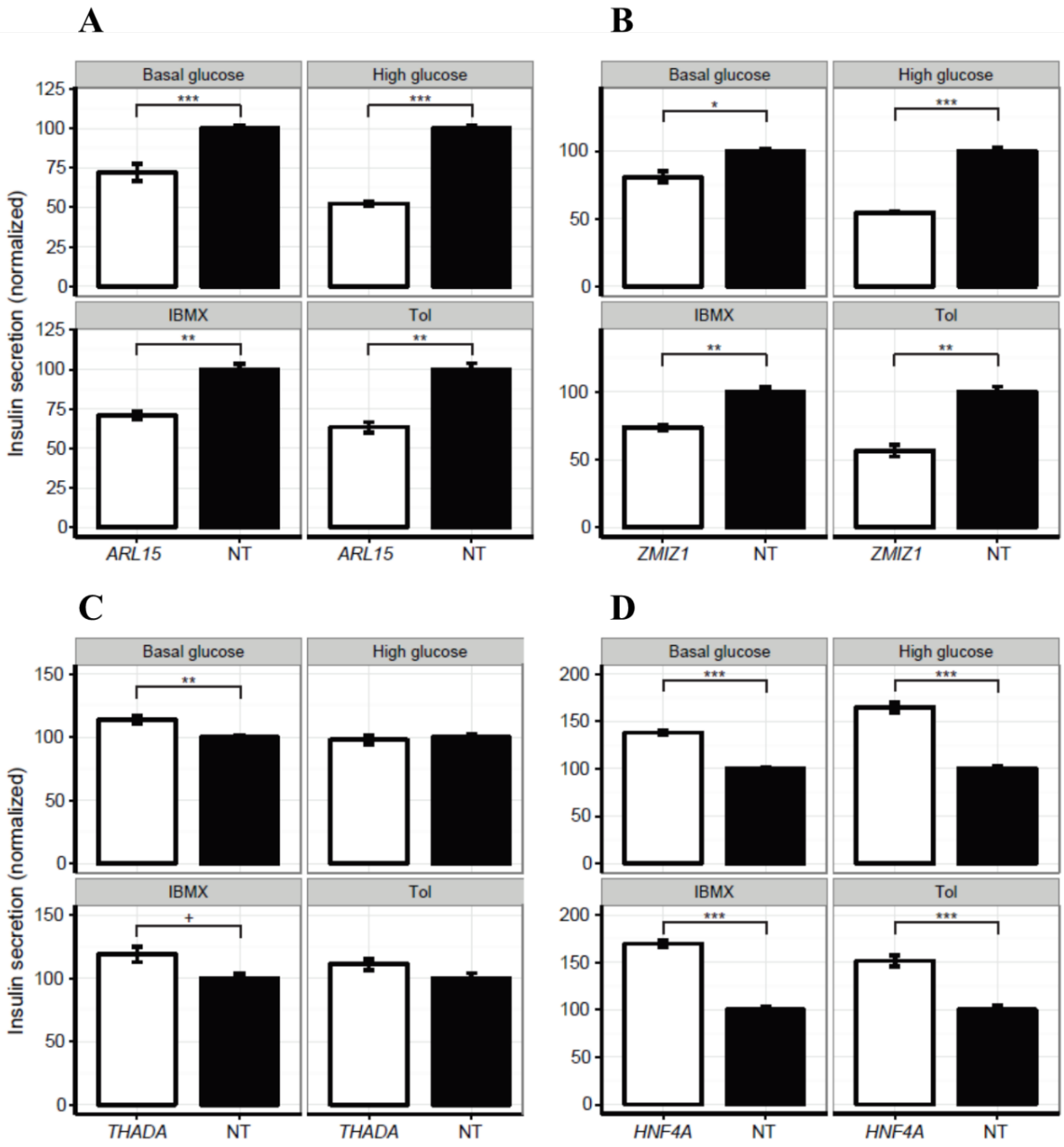


Figure 13 *Insulin secretion data for selected genes in a follow-up validation experiment.* Insulin secretion for (A) ARL15, (B) ZMIZ1, (C) THADA, and (D) HNF4A (white) versus non-targeting (NT) control (black) under the indicated conditions. Measurements were processed as described for Figure 11, and shown as percentage of control. Data points are the mean of n = 6 for NT and n = 3 for other genes, and error bars are SEM. +q-value < 0.1, *q-value < 0.05, **q-value < 0.01, ***q-value < 0.001 by Student's t-test (FDR-adjusted). Tol = Tolbutamide.

5.3.6 *ZMIZ1* regulates key beta cell pathways

Finally, having uncovered several plausible causal genes, I focused on a single candidate to understand the underlying mechanism in greater detail. *ZMIZ1* encodes Zinc Finger, MIZ-Type Containing 1 (ZMIZ1), a PIAS-family transcription factor known to be a co-activator of TGFβ and Notch signalling in non-islet cell types (299, 300). Though dysregulation of the gene has been implicated in leukaemia and autoimmune disorders, a possible role in beta cell function remains poorly understood (301, 302). Importantly, a recent study reported that the intronic T2D association signal coincides with a cis-eQTL for *ZMIZ1* in primary human islets, supporting its candidacy as a causal gene (77). The same study demonstrated that overexpression and silencing of *ZMIZ1* negatively impacts on insulin secretion in islets, consistent with the effects observed in EndoC-βH1. Overall, *ZMIZ1* emerges as a strong candidate causal gene, though the cellular pathways that contribute to beta cell dysfunction are unknown.

To address this, I obtained primary human islets from six donors, each transfected with si*ZMIZ1* and scrambled siRNA by Dr Manning Fox (Patrick MacDonald lab, University of Alberta). I then performed RNA-seq on these islets to identify pathways affected by *ZMIZ1* deficiency. I first confirmed consistent, albeit relatively modest, gene silencing across the six paired samples (-35.5%, p = 0.001; **Figure 14A**). Next, I performed principal component analysis to explore the primary sources of variation in the dataset (**Figure 14B**). Unsurprisingly, between-donor variation was the main driver of separation across the first six principal components. This presumably reflects both biological variation, such as inter-individual differences in genetic and environmental influences, as well as technical sources of variation, such as preparation quality, purity, and handling. Nevertheless, *ZMIZ1* knockdown could be separated from the negative controls in the space defined by principal components seven and eight.

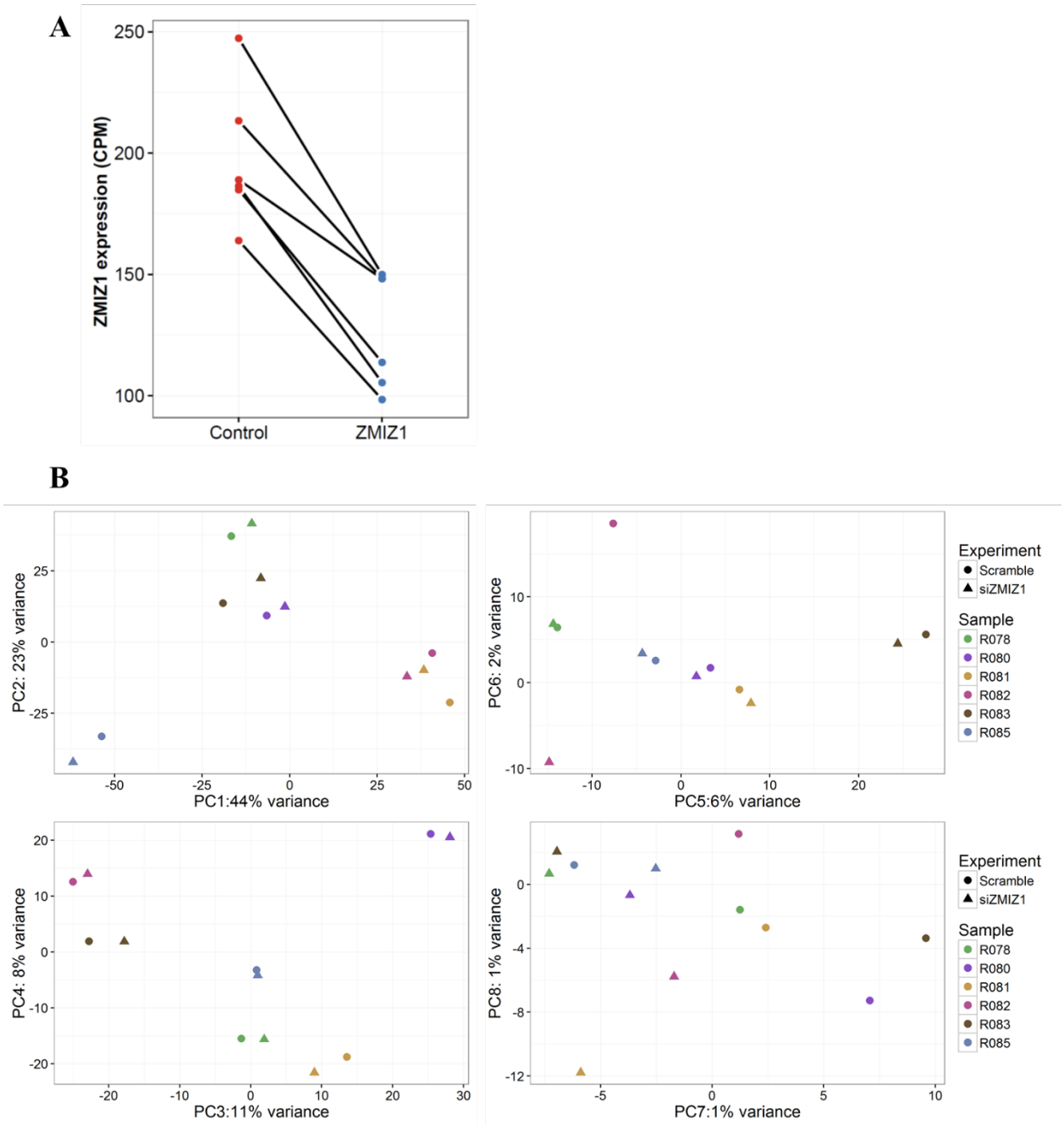


Figure 14 *Gene expression patterns following* ZMIZ1 *silencing in primary human islets*. Gene expression changes were measured 48 hours after dispersed islets were transfected with either si*ZMIZ1* or scrambled negative control. RNA-seq data were analysed (A) for changes in *ZMIZ1* levels, and (B) for overall variation in gene expression patterns using principal component analysis. The first eight principal components (PCs) and the variance in gene expression explained by each is shown.

Despite being dominated by between-donor variation, the effect of *ZMIZ1* silencing on gene expression could be isolated using a paired analysis. I thus performed differential expression analysis to identify specific genes regulated (directly and indirectly) by *ZMIZ1*. This identified a total of 64 genes that each showed both nominally significant differences in expression levels ($p < 0.01$) and directional consistency across all samples. Among these genes, three KEGG terms were identified as significantly enriched: "Maturity onset diabetes of the young" ($q = 0.000282$), "TGFβ signalling pathway" ($q = 0.0134$), and "Type 2 diabetes" ($q = 0.0345$).

To explore the molecular processes contributing to these enrichments, I generated functional protein association networks using STRING analysis. The set of differentially regulated genes showed a higher level of connectivity than expected from a random set of similar size ($p = 1.11*10^{-5}$), and 16 genes coalesced on two distinct networks (**Figure 15A**). The largest network was formed by genes encoding proteins involved in TGFβ signalling (including *TGFBR2*, *SMURF1*, and *BAMBI*), consistent with the previously reported role of ZMIZ1 as a co-regulator of TGFβ effectors. Second, a network of well-established beta cell regulators emerged (including *ABCC8*, *MAFA*, *PDX1*), with all of the component genes robustly downregulated following *ZMIZ1* silencing (**Figure 15B**). Overall, these results establish *ZMIZ1* as an important regulator of core beta cell pathways, and highlight possible mechanisms for the cellular phenotypes observed in EndoC-βH1 and primary islets following *ZMIZ1* silencing.

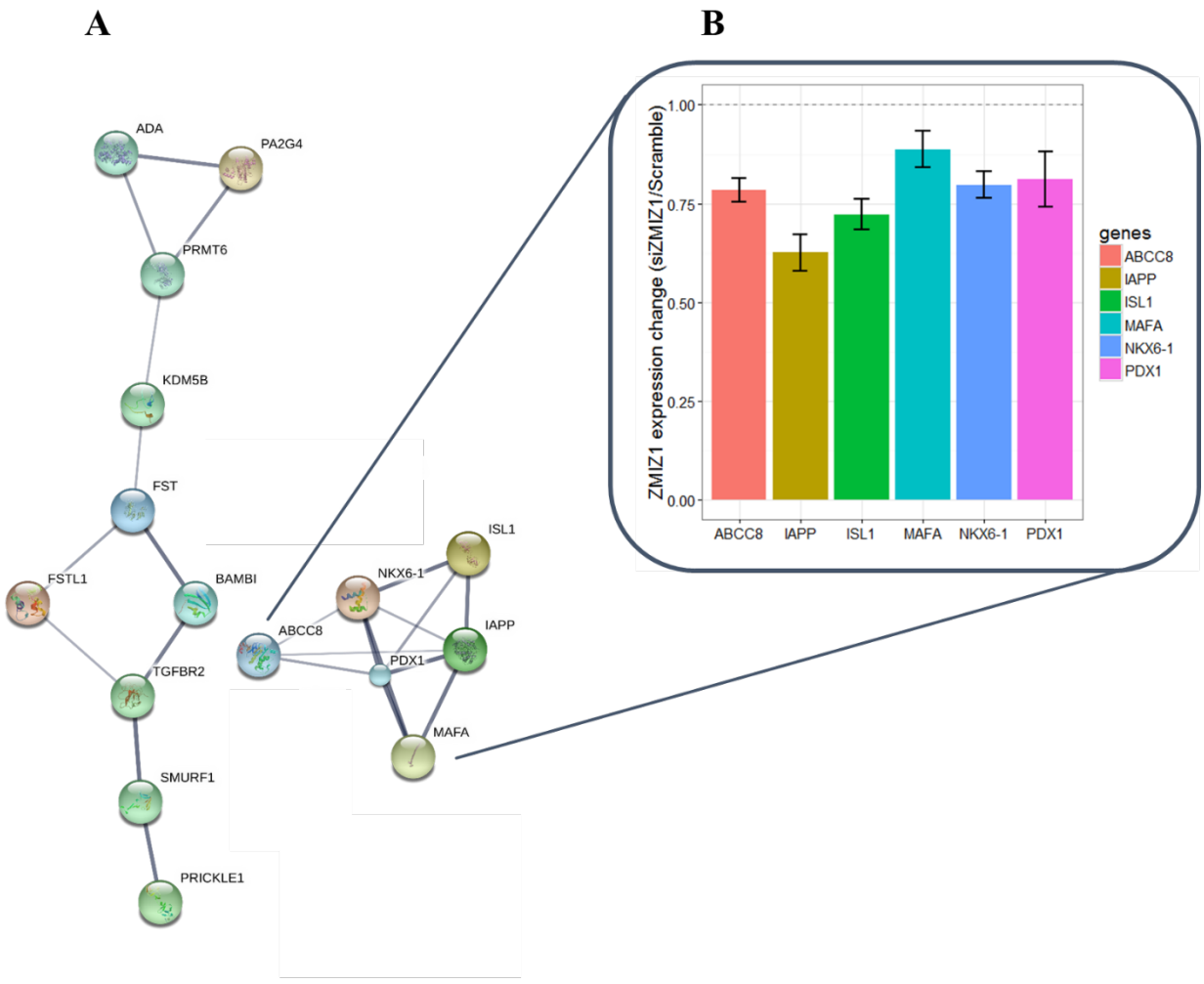


Figure 15 *Functional protein networks for differentially regulated genes following* ZMIZ1 *knockdown in primary islets.* (A) STRING analysis of 64 differentially regulated genes showed 16 of these to be focused on two networks (n = 10 and n = 6). (B) Expression changes following *ZMIZ1* knockdown of six genes identified as members of a core beta cell network.

5.4 Discussion

High-throughput screens for beta cell dysfunction offer the opportunity to systematically characterize the role of genes in a disease-relevant tissue for T2D. Previous studies have relied on rodent beta cell models, and either used indirect proxies for insulin measurements or focused on cellular proliferation (280-285). Here, I report a genetic screening strategy for the interrogation of multiple disease-relevant phenotypes in the human beta cell line EndoC-βH1. In a primary screen of 300 positional candidates, I successfully identified 15 genes regulating cell number (a function of proliferation and viability), and 35 genes regulating insulin secretion. This is the first systematic, large-scale effort to identify genes involved in insulin secretion. Importantly, the identified hits can be used to prioritize novel effector transcripts for T2D GWAS loci, and may shed further light on mechanisms underlying genes previously implicated in beta cell dysfunction.

The known MODY gene *HNF4A* was, perhaps unexpectedly, observed to cause a consistent increase (> 40%) in insulin secretion across all conditions. *HNF4A* encodes the transcription factor Hepatocyte nuclear factor 4 alpha (HNF4α), and is mutated in about 10% of all MODY cases (303). *HNF4A* LOF mutations that cause monogenic diabetes later in life have also been associated with increased birth-weight (indicative of increased foetal insulin secretion) and congenital hyperinsulinism in early infancy (304). The underlying reason for this switch from elevated to reduced insulin secretion is unknown, but it has been speculated that gradual beta cell exhaustion or, alternatively, a shift in the modulating co-factors of HNF4α may underlie this phenomenon (305, 306). To further elucidate the contribution of these two mechanisms, it will be important to clarify the extent to which the EndoC-βH1 cell line may reflect an immature state of beta cells (*Section 1.5.3*).

Among the hits with limited prior evidence of a role in beta cell function, I independently validated *ARL15*, *THADA,* and *ZMIZ1*. *ARL15* encodes ADP-Ribosylation Factor-Like 15, a relatively uncharacterized member of the ARF-family of proteins involved in regulation of vesicle trafficking and biogenesis. The gene is highly expressed in beta cells and located downstream of an islet-active enhancer bound by key beta cell transcription factors (107, 307). *THADA* encodes the protein Thyroid Adenoma-Associated (THADA), and contains a coding disease-association signal that has also been associated with reduced beta cell function (308). Consistent with the directionality of our findings, expression profiling has shown the gene to be more highly expressed in patients with T2D compared with controls (309). Both genes thus emerge as strong candidates for future studies.

To demonstrate the utility in prioritizing genes for post-GWAS follow-up, I focused on one particular gene, *ZMIZ1,* for a number of reasons. Prior to this study, there was limited evidence to directly implicate the gene in beta cell function, but the candidacy of *ZMIZ1* as a causal gene was supported by overlap between the T2D GWAS signal and a cis-eQTL in islets (77). *ZMIZ1* manipulation in primary islets had also been shown to cause decreased exocytosis and insulin secretion, suggesting that the effects observed in my screen are not artefacts of using a transformed cell line (77). Further, the defects arising from *ZMIZ1* silencing were among the largest detected for both insulin secretion and cell numbers, hinting at a key role of *ZMIZ1* in beta cell function.

Based on RNA-seq analysis of primary islets from deceased donors, I identified differentially expressed genes following *ZMIZ1* knockdown. This uncovered two key processes affected by *ZMIZ1* deficiency, implicating both TGFβ signalling and a network of core beta cell regulators. The genes in this beta cell network were significantly and consistently downregulated across all samples, thus offering a possible explanation for the cellular phenotypes detected. These data thus point to a previously unknown role of ZMIZ1 as a key

regulator of beta cell function. Taken together, the integration of multiple lines of evidence now converge on a plausible causal mechanism for the *ZMIZ1* locus, mediated, at least partly, through dysregulation of ZMIZ1 in human islets. Though additional effects on disease susceptibility mediated through other tissues (or genes) cannot be excluded, the lead T2D SNP is not in LD with cis-eQTL variants in other cell types (77). Moreover, the T2D association signal is located in an islet cluster enhancer, raising the possibility of tissue-specific regulatory activity (*Section 1.3.4*). Future experiments will be required to address the extent to which other mechanisms may contribute to defects in metabolic homeostasis.

Despite successfully enabling unbiased prioritization of causal genes for T2D risk, the current screening strategy has a number of limitations. False negatives (i.e. true causal genes not identified as hits) could arise as a result of primary effects of the causal gene on non-beta cell tissues, or through effects on genes expressed at different developmental stages. Likewise, overexpression or greater knockdown efficiency may in some cases be required to expose a disease-relevant phenotype. Among the targets analysed for silencing efficiency, a variable range of knockdown was observed (34% - 88%), and some genes might remain undetected due to insufficient silencing. Conversely, false-positive effects (i.e. non-beta cell regulators identified as hits) also cannot be excluded, and unexpected findings should be further functionally validated. Though the EndoC-βH1 cell line has been found to recapitulate many aspects of beta cell function, it remains a possibility that some findings would not translate directly into human physiology. Finally, a subset of the identified hits may represent true beta cell regulators that are independent of any disease risk variants and, though still of biological importance, not genuine effector transcripts for T2D. In addition to the possibility of more than a single effector transcript per locus, this phenomenon likely also contributes to the relatively high proportion of multi-hit loci observed in the primary screen (8/37).

Despite these limitations, our screening strategy successfully replicated well-established biological mechanisms, and identified genes involved in beta cell function at half of the loci investigated. This demonstrates, for the first time, the feasibility of performing scalable screens for insulin secretory defects in human pancreatic beta cells, and opens up the possibility, not only for large-scale genetic manipulations, but also for compound screens to therapeutically manipulate human beta cells. Insights from this and subsequent functional screens can be integrated with complementary lines of evidence from exome-wide association studies, 3C-based datasets and cis-eQTL studies to prioritize genes for follow-up studies. Ultimately, this could accelerate the translation of genetic association signals into molecular mechanisms for beta cell dysfunction, insulin insufficiency, and T2D.

CHAPTER 6

General Discussion

T2D risk is determined by a complex interplay between environmental and heritable factors (310). Human genetic studies have the potential to identify novel mechanisms for disease, which in turn present opportunities for clinical translation. Over the past decade, GWAS have identified over 150 independent genomic regions robustly associated with T2D predisposition (26, 36). However, the translation of GWAS signals into molecular mechanisms for disease has been slow, primarily due to uncertainty over the genes through which non-coding association signals operate (311). The work presented in this thesis aimed to address this translational bottleneck through the application of multiple complementary approaches for prioritizing causal genes. In doing so, the work has provided new insights into T2D pathogenesis as well as highlighting strengths and limitations of the different strategies.

6.1 The candidate gene approach and recruit-by-genotype studies

The study presented in Chapter 3 exemplifies the traditional approach often taken by experimental studies following up on genetic discoveries for complex disease. Broadly, this strategy uses a heuristic based on proximity to an association signal and prior knowledge about the function of the encoded proteins to select particular candidates for follow-up work. In the case of the *CDKN2A/B* locus, studies in rodent and cellular models had implicated the *CDKN2A* gene in glucose homeostasis, but little was known about a corresponding role in humans (191, 192, 194). Rare *CDKN2A* LOF mutations are a cause of familial melanoma and offer the unique opportunity to determine the impact of p14 and p16 haploinsufficiency on human metabolism *in vivo*. The data presented in Chapter 3 demonstrates that mutation carriers

display impaired insulin sensitivity, increased insulin secretion, and reduced hepatic insulin clearance compared with matched controls. These results propose a model whereby *CDKN2A*-loss affects multiple tissues to impact on glucose homeostasis. This is in line with recent reports that have shown the gene to play a role in both beta cells and several other cell types, including liver and adipose tissue (174).

The results also highlight the advantage in doing recruit-by-genotype (RBG) and recruit-by-phenotype (RBP) studies to increase statistical power. Compared with the large cross-sectional cohorts usually required to study common variants with modest effects, this paradigm enables in-depth physiological studies in carriers with rare variants or phenotypes (312). In this case, using an RBG/RBP hybrid approach, recruitment was facilitated by a striking phenotype (FAMMM) that allowed mutation carriers to be easily identified from the general population. Increasingly, genotyped biobanks with consent to recall are being used to apply the RBG design more systematically for the study of genetic mechanisms (313-317). An important limitation, though, is the trade-off between the effect size of the risk allele and population frequency; larger effects provide increased power for a fixed study size, but necessitate larger biobanks to identify the required number of variant carriers (if subject to negative selection). The study of FAMMM individuals is a case in point, with comprehensive recruitment from centres in both the UK and the Netherlands. Still, the number of available volunteers was insufficient to establish power for IVGTT-derived indices. Future studies will thus be required to derive more specific glycaemic measures in individuals carrying mutations across the allelic spectrum. To specifically address the role of *CDKN2A* in beta cell proliferation in humans, such studies could, for example, seek to estimate total islet mass through arginine stimulation in clamp-based tests (318).

From the perspective of understanding T2D predisposition, it is important to note that the non-coding risk alleles at the *CDKN2A/B* locus have been implicated in reduced beta cell

function, but not indices of insulin clearance or sensitivity (75, 173, 179, 180). It thus remains possible that the causal mechanism for these variants is mediated through genes other than *CDKN2A*. Alternatively, tissue-specific regulatory activity could mask any effect of *CDKN2A* in non-islet tissues, effectively restricting the phenotype produced by the T2D risk variants. In support of this hypothesis, the association signals are located in an islet cluster enhancer that shows FOXA2 binding specific to islets and melanocytes (73). More generally, the challenges posed by the interpretation of these data illustrates a fundamental limitation of using global KO models. Whether based on recombinant animals or human LOF mutation carriers, this approach provides only circumstantial evidence to implicate a gene as causal for non-coding signals.

Future studies will be required to establish a cis-regulatory mechanism for the risk allele at the *CDKN2A/B* locus. Though previous cis-eQTL studies have failed to identify any signal overlapping the T2D risk variants in the tissues studied, larger cohorts may be required to unravel any subtle effects on gene expression (175). It is equally possible that the T2D-related effect could be driven through a developmental phenotype that is undetectable in adult cell types. Advances in genome editing and iPSC differentiation are going to enable studies in isogenic cell lines that allow cis-regulatory effects to be interrogated at different stages of development.

6.2 Coding risk variants for causal gene prioritization

The approach demonstrated in Chapter 4 uses coding variants to eliminate some of the uncertainty in ascertaining causal genes. The *PAM* gene harbours two independent, non-synonymous variants associated with T2D risk (36, 37, 319). One (rs36046591) is in LD with a coding variant in the nearby gene *PPIP5K2* (rs36046591) but has a high posterior probability of causality (88.5%) (319). Taken together with the presence of a second coding variant

(rs78408340), the genetic evidence strongly implicates *PAM* as causal at the *PAM/PPIP5K2* locus. Following on from work showing that both T2D risk alleles impact negatively on overall PAM activity, the results in Chapter 4 explored the effect of *PAM* deficiency on beta cell function. This established a role for PAM in the regulation of insulin availability and exocytosis. The amidated granular packaging protein Chromogranin A was identified as an endogenous substrate, with evidence in support of a downstream role as a PAM effector in beta cells. These findings thus provide molecular and functional explanations for the observed effects of the T2D risk alleles.

The identification of non-synonymous risk variants offers additional advantages aside from explicitly nominating causal genes. Most importantly, the underlying mechanisms are often more experimentally tractable than those of non-coding variants. This is partly due to the direct impact that coding variants have on protein function, and partly due to (generally) less context-dependent mechanisms. Gene expression patterns can therefore be used to identify the cell types likely to be affected by protein dysregulation. For *PAM*, high expression levels in pancreatic islets pointed to a beta cell-driven effect. This was further corroborated by physiological studies, which have associated T2D risk variants with measures of beta cell dysfunction (37, 319). Still, additional extra-pancreatic effects on beta cell function cannot be excluded. Chapter 4 explored one such possible mechanism, mediated through reduced amidation of GLP-1 or TRH precursor. Though no direct impact of amidation status on stimulation of insulin secretion was found, PAM deficiency may either directly or indirectly influence total circulating levels of these (or other) hormones. This hypothesis is the focus of an on-going RBG study at OCDEM, which could have implications for incretin-based therapies for T2D.

More broadly, the therapeutic potential of targeting PAM remains unclear. Given the acute impact of *PAM* silencing on beta cell function, it is conceivable that even short-term

treatment could produce a clinical effect. Activation would presumably confer a protection against T2D, but addressing this requires a more complete understanding of the allelic spectrum in *PAM*, including the effect of gain-of-function variants. It is also possible to predict on-target side-effects using the PheWAS approach, through additional phenotypic associations of non-synonymous variants. The low frequency risk allele in *PAM* (rs78408340) is nominally associated ($p<0.05$) with increased height, waist circumference, and cholesterol, as well as reduced systolic blood pressure (320). Larger genotyped cohorts with comprehensive phenotyping will be required to more conclusively pinpoint any specific concerns regarding possible side-effects of targeting PAM.

Though coding variants provide a powerful approach for prioritizing causal genes, wider applications are constrained by the ability of genetic studies to catalogue such variants. The most comprehensive analysis for T2D risk to date uncovered just 13 loci harbouring non-synonymous coding variants that achieved genome-wide significance (including *PAM/PPIP5K2*) (36). Furthermore, conditional analyses with coverage of non-coding regions are ultimately required to establish causality. A recent fine-mapping study suggested that a third of coding association signals "naively" identified through exome-focused discovery may, in fact, represent false leads (65). For these reasons, the total number of loci for which GWAS offer a signal likely to be driven through coding variation is still limiting for experimental follow-up.

6.3 Functional genetic screening and data integration

Chapter 5 presented a complementary strategy for prioritizing causal genes that is applicable to interrogation of all association signals. The concept is based on systematic and comprehensive functional characterisation of positional candidate genes. By extension of the candidate gene approach illustrated in Chapter 3, it is based on the premise that implicating a

gene in a disease-relevant process raises the posterior probability of that gene being causal for a nearby risk variant. Ideally, it requires unbiased evaluation of all candidate genes, and should be focused on endophenotypes proximal to the component pathways of disease pathogenesis.

The screen developed and implemented in Chapter 5 was based on parallel gene silencing in the human beta cell line EndoC-βH1. In a primary screen of 300 positional candidates selected from 75 T2D regions, each gene was assayed for effects on multiple cellular phenotypes known to be important in disease risk. This successfully pointed to possible causal mechanisms at half of the disease-associated loci investigated, with a strong enrichment for monogenic diabetes genes. The potential for causal gene discovery was illustrated by follow-up work on a single hit, *ZMIZ1*, emerging from the primary screen. Differential expression analysis uncovered two pathways, TGFβ signalling and a core beta cell network, affected by ZMIZ1 deficiency in primary human islets. These findings propose a molecular mechanism for the defects in insulin secretion and proliferation observed at the cellular level.

The value in the data generated by genetic screening is determined both by the confidence with which disease-relevant tissues are known and the relevance of the phenotypes being interrogated. In the case of T2D, defects in insulin secretion has been established as a central mechanism of disease pathogenesis, pointing to the beta cell as a useful starting point for defining causal mechanisms (75). Even so, many association signals are likely to be mediated through other tissues, different development stages, or non-insulin secretion pathways of beta cell function. Future screening efforts should aim to target all such phenotypes, but experimental capacity is currently limited by practical and financial considerations.

Another key limitation of genetic screening for causal gene prioritization is the circumstantial nature of the evidence generated. Being gene-centric by design, the strategy can highlight likely targets of non-coding risk variants but not directly establish cis-regulatory links. Accordingly, screening results should ideally be integrated with other datasets, such as cis-eQTLs or chromatin interactions. These variant-gene annotations can reveal regulatory interactions but not predict which perturbations generate functional effects relevant to disease risk. Triangulation from different datasets that are individually restricted in their ability to predict causal mechanisms can thus build confidence in particular candidates. An example of this is provided by insights into the *ZMIZ1* locus discussed in Chapter 5. The convergence of a tissue-specific cis-eQTL for *ZMIZ1* in islets and evidence of a role for the encoded protein in insulin secretion provides strong support for the candidacy of this gene as causal (77). Further work could seek to understand the regulatory mechanisms involved using chromatin immunoprecipitation, or explore the physiological consequences in human or animal models.

6.4 Future directions

Looking ahead, causal gene prioritization is likely to remain critically important, as biological and therapeutic insights will be needed to justify GWAS in ever larger cohorts. Given the genetic architecture of T2D, the majority of novel association signals will be driven by common, non-coding variants with small ORs. Nonetheless, more comprehensive enumeration of coding variants, in some cases with relatively large effect sizes, will be a powerful tool for translating genetic findings into molecular mechanisms. As illustrated by recent studies on *SLC30A8* and *SLC16A11*, as well as the work on *PAM* discussed in Chapter 4, such coding variants can be particularly useful in evaluating therapeutic strategies (44, 123).

In terms of translating non-coding variation into disease biology, a number of emerging tools are beginning to transform the endeavour through unbiased, genome-wide datasets.

Future efforts are likely to be rooted in the integration of such "platform datasets" that can be applied prospectively to prioritize causal mechanisms. Integration will rely on a combination of data from genomic annotations, functional screens, and fine-mapped genetic data. Genomic annotations for tissues relevant to T2D have begun to emerge, but power is still limiting, especially for hard-to-obtain organs such as islets. Moreover, since each set of annotations is merely a snapshot of a given cell type in a particular metabolic and developmental state, future work will continue to produce an atlas with ever-finer spatiotemporal resolution. Phenotypic screening, as demonstrated in Chapter 5, will benefit from emerging experimental tools, such as CRISPR-Cas9 and iPSCs. Pooled CRISPR screens have simplified genome-wide screening for low-dimensional phenotypes, and recent variations have enabled more complex readouts (287, 321-324). In parallel, progress on differentiation protocols for iPSCs will make surveying a multitude of tissue types and developmental stages an increasingly tractable goal (325).

The ultimate vision will be to unravel causal mechanisms by overlaying fine-mapped loci with available maps of chromatin features, variant-gene interactions, and functional effects across disease-relevant tissues. Physiological associations, if available, may be used to assign higher weights to particular tissues or endophenotypes that are better proxies for whole-body phenotypes. Though not all regions will be amenable to this approach (e.g. due to extensive LD), the strategy could inform hypotheses to direct follow-up experiments. In this way, the process should seek to prioritize mechanisms based on posterior probabilities, but not substitute for detailed mechanistic studies. Traditional approaches will be used to confirm and elaborate on insights derived prospectively from platform datasets. These experiments will continue to involve generation of animal and cellular models, but could also focus increasingly on RBG studies in humans. As larger biobanks become available, detailed physiological phenotypes could be probed in a similar way to that demonstrated in Chapter 3, providing the most relevant "model system" based on *in vivo* human studies.

In conclusion, this thesis has addressed a significant outstanding challenge in the translation of genetic signals into disease mechanisms. The individual chapters have successfully applied different experimental strategies to prioritize causal genes for T2D risk loci. Taking a candidate gene approach in Chapter 3, *CDKN2A* was found to be a plausible causal gene through physiological tests in carriers of rare LOF mutations. Molecular studies in Chapter 4 were then used to establish a causal mechanism for coding risk alleles in the *PAM* gene. Finally, a genetic screen demonstrated in Chapter 5 the feasibility of systematically evaluating positional candidates in an unbiased and scalable manner. Through integration with existing datasets, the *ZMIZ1* gene emerged as a strong candidate causal gene for follow-up studies. Building on the advantages and limitations of the individual approaches, future applications promise to accelerate the translation of genetic discoveries into a deeper understanding of the underlying biology. This, in turn, could pave the way for new disease-modifying treatments to benefit both common and rare forms of diabetes.

Appendices

Appendix 1

A. Cloning of PAM expression plasmids

A pCMV6 expression vector encoding integral membrane H. sapiens *PAM* (NM_138821.2) with an in-frame C-terminal Myc-Flag tag (PAM-Myc-Flag) was purchased from Origene. Soluble functional PAM-Myc-Flag encompassing the first 710 amino acids was generated via PCR using primers 5'GTACCCAGAATGTAGCTCC3' and 5'CGGACGCG-TAACTGATCGATGTTCCAA3'. Puromycin-resistant PAM-Myc-Flag expression vectors were generated via *Bgl*II/*Xma*I digestion, blunt ending with T4 polymerase, and subcloning into *Hpa*I-digested pIRES puro 2 (Clontech). Mutations were introduced via site-directed mutagenesis using the Stratagene QuikChange II kit (Agilent Biotechnologies) according to manufacturer's instructions. Primer sequences were 5'CATGTCTGG-GATGGAAACTGGTTTGACAGCAAGTTTGTTTAC3' and 5'GTAAACAAACTT-GCTGTCAAACCAGTTTCCATCCCAGACATG3' for S539W, 5'GACACTATTCTT-GTCATAGGTCCAAATAATGCTGCAGTAC3' and 5'GTACTGCAGCATTATTTGGAC-CTATGACAAGAATAGTGTC3' for D563G, and 5'CATTTATGTATCAGATGGTTTCTG-CAACAGCAGGATTGTGC3' and 5'GCACAATCCTGCTGTTGCAGAAAC-CATCTGATACATAAATG3' for Y651F. All plasmids were verified by sequencing.

For generation of PAM stable cell lines, HEK 293 cells were transfected using Fugene 6 (Promega) according to manufacturer's instructions. Cells were routinely passaged and maintained in 1μg/mL puromycin. For fluorescence microscopy, EndoC-βH1 cells were transfected in chamber slides (BD Biosciences) using Fugene 6. For siRNA experiments, EndoC-βH1 cells were seeded in Opti-MEM reduced serum medium (Life Technologies)

containing 25nM ON-TARGETplus SMARTpool siRNA complexes (Dharmacon) pre-formed with Lipofectamine RNAiMAX (Life Technologies) according to manufacturer's instructions. Growth medium was replaced 24hr later and incubated for a further 48hr. For PAM inhibitor experiments, seeded cells were treated with 500μM 4-phenyl-3-butenoic acid (4P3BA) or DMSO vehicle (Sigma-Aldrich) for 72hr.

B. Recombinant PAM protein production

HEK 293 stable cell lines were seeded into Triple Flasks (Nunc). Medium was replaced with DMEM 1145 supplemented with 1mM sodium pyruvate and incubated for 3-5 days. Supernatant containing secreted recombinant PAM protein was harvested directly from flasks, filtered into a sterile storage bottle (Corning), and pH corrected to 5.5 with HCl. Filtered supernatant was concentrated using a Centricon Plus-70 30kDa MW cut-off Centrifugal Filter Device (Millipore) according to manufacturer's instructions.

C. Quantitation of recombinant PAM

For relative quantitation of recombinant PAM proteins, 3x1μL aliquots of concentrated supernatant were separated on a 4–20% Criterion TGX Stain-Free Gel (Bio-Rad) and transferred to PVDF using a Trans-blot Turbo Transfer Pack (Bio-Rad) according to manufacturer's instructions. Membranes were incubated overnight in 4A6 anti-Myc tag antibody (Millipore) at 4°C and developed using the Clarity Western ECL Substrate (Bio-Rad) according to manufacturer's instructions. Proteins were visualised on a ChemiDoc MP (Bio-Rad). Lanes and bands were drawn in Image Lab 5.0 and the average band volume (intensity) quantified for each triplicate of samples. The amount of recombinant variant PAM protein/μL could then be calculated relative to WT-PAM.

D. Kinetic analysis of PAM amidating activity

Amidating activity was measured via spectrophotometric detection of chemically converted glyoxylate (239). An assay mix of 100mM MES pH 5.5, 2μM $CuSO_4$, 10mM L-ascorbic acid, 0.1mg/mL bovine liver catalase, and 20mM hippuric acid (Sigma-Aldrich) was split into glass vials. Equimolar amounts of recombinant PAM were added to each vial and incubated in a 37°C water bath. Every 30min, samples were removed to a conical 96-well plate (4titude) containing 40mM EDTA pH 8.0. The glyoxylate produced was converted to glyoxylate phenylhydrazone via addition of 0.1% phenylhydrazine HCl (Sigma-Aldrich), incubated at room temperature for 15min, and saturated with concentrated (37%) HCl. The contents of the entire plate were then aspirated and dispensed into a flat-bottomed 96-well plate (Corning) containing 0.2% $K_3[Fe(CN)_6]$ using an i-Pipette Automated Pipettor (Apricot Designs). Absorbance of 1,5-diphenylformazan was measured at 515nm using a VersaMax microplate reader (Molecular Devices).

E. Kinetic analysis of variant PAM substrate affinity

For K_m calculations, a 12-point two-fold serial dilution of hippuric acid in which the maximum concentration was 20mM was prepared in assay mix (100mM MES pH 5.5, 2μM $CuSO_4$, 10mM L-ascorbic acid, 0.1mg/mL bovine liver catalase). Recombinant PAM or empty vector control supernatant was then added. The amount of D563G-PAM added to the assay was increased such that it displayed equal V_{max} to WT-PAM. The volume of empty vector supernatant matched the volume of WT-PAM used in that experiment. Vials were incubated in a 37°C water bath for 1hr. Samples were then removed into a conical 96-well plate containing 40mM EDTA pH 8.0.

EndoC-βH1 cells transfected with *PAM* expression vectors were fixed in 4% paraformaldehyde, permeabilised in 0.1% (v/v) Triton X-100, and blocked in 10% (w/v) BSA. Primary antibodies (4A6 anti-Myc, Belfast anti-insulin, TGN46 anti-trans Golgi [Sigma-

Aldrich], and H-70 anti-calnexin [Santa Cruz Biotechnology sc-11397]) were incubated overnight at 4°C. Slides were then incubated in Alexa Fluor-conjugated secondary antibodies (Life Technologies) and mounted in Vectashield Mounting Medium with DAPI (Vector Biolaboratories). Cells were visualised using an LSM 10 META confocal laser scanning module arranged on an Axiovert 200 microscope and a Plan-Apochromat 63x/1.4 oil immersion objective (Carl Zeiss). An argon laser was used to excite Alexa Fluor 488 at λ=488nm and a HeNe laser was used to excite Alexa Fluor 543 at λ=543nm. DAPI was excited in two-photon mode using the 740nm line of an infrared light Chameleon laser.

Appendix 2

A. List of target genes included in the primary screen by locus affiliation

Gene Name	T2D Locus	EndoC-βH1 [RPKM]	Primary beta cells [RPKM]
MACF1	*MACF1*	10.727805	15.879
BMP8A	*MACF1*	1.4541069	0.64
HEYL	*MACF1*	1.5134418	0.791
BMP8B	*MACF1*	12.184336	5.887
MFSD2A	*MACF1*	1.2909733	0.846
RHBDL2	*MACF1*	0.6474848	0.224
CDKN2C	*FAF1*	80.097336	5.212
FAF1	*FAF1*	15.062617	6.613
DMRTA2	*FAF1*	0.046803616	0.014
ELAVL4	*FAF1*	10.931084	19.561
TTC39A	*FAF1*	14.173553	10.022
PROX1	*PROX1*	10.218364	12.548
PTPN14	*PROX1*	0.01770482	0.14
SMYD2	*PROX1*	7.387975	2.934
PPM1G	*GCKR*	66.50968	27.658
KHK	*GCKR*	25.9325	1.467
CENPA	*GCKR*	9.154673	0.098
PLB1	*GCKR*	0.04874327	0.456
THADA	*THADA*	3.465101	2.575
LRPPRC	*THADA*	49.46855	16.941
DYNC2LI1	*THADA*	5.1710835	4.216
PLEKHH2	*THADA*	1.5512799	2.889
BCL11A	*BCL11A*	0.42303053	0.44
PAPOLG	*BCL11A*	8.144624	3.191
MGAT5	*TMEM163*	11.647037	6.124
CCNT2	*TMEM163*	14.048298	10.553
ACMSD	*TMEM163*	0.021140289	0.875
MAP3K19	*TMEM163*	0.002687708	0.056
R3HDM1	*TMEM163*	20.273561	7.659
RIF1	*RBM43/RND3*	13.902107	5.398
NMI	*RBM43/RND3*	4.603304	2.955
RBM43	*RBM43/RND3*	0.9375855	3.668
NEB	*RBM43/RND3*	0.28890586	0.038
RND3	*RBM43/RND3*	5.034543	1.76
ITGB6	*RBMS1*	0.052311223	0.845
PLA2R1	*RBMS1*	0.003079499	0.115
LY75	*RBMS1*	11.025683	3.095
RBMS1	*RBMS1*	12.231243	2.013
CD302	*RBMS1*	7.510972	2.322
GRB14	*GRB14*	8.159008	2.218
COBLL1	*GRB14*	5.161108	9.288
SCN2A	*GRB14*	0.29049847	2.111
SCN3A	*GRB14*	0.4411591	4.009
IRS1	*IRS1*	4.0152564	2.761
NKIRAS1	*UBE2E2*	7.2671003	12.391
UBE2E1	*UBE2E2*	11.214179	14.05
PSMD6	*PSMD6*	39.243126	32.951
PRICKLE2	*PSMD6*	0.29158774	4.445
SYNPR	*PSMD6*	0.34662268	0.02
THOC7	*PSMD6*	26.802912	11.24
ATXN7	*PSMD6*	6.243315	8.72

ADAMTS9	*ADAMTS9*	0.006904925	0.497
ADCY5	*ADCY5*	0.10015087	0.78
MYLK	*ADCY5*	0.67700344	0.167
PDIA5	*ADCY5*	3.5262358	2.34
PTPLB	*ADCY5*	41.775444	10.987
SEC22A	*ADCY5*	6.3320513	5.297
DIRC2	*ADCY5*	6.195492	6.982
IGF2BP2	*IGF2BP2*	0.044629	1.983
SENP2	*IGF2BP2*	14.213694	7.035
DGKG	*IGF2BP2*	1.6199253	2.127
MAP3K13	*IGF2BP2*	5.584649	8.417
LIPH	*IGF2BP2*	0.029294837	12.143
ETV5	*IGF2BP2*	6.104882	6.329
TRA2B	*IGF2BP2*	48.647205	27.016
ST6GAL1	*ST64GAL1*	25.443888	6.183
ADIPOQ	*ST64GAL1*	0.006442239	0.001
AHSG	*ST64GAL1*	0.07925758	0.012
TBCCD1	*ST64GAL1*	6.7398887	5.311
CRYGS	*ST64GAL1*	1.1851472	1.43
SST	*LPP*	43.108746	3555.411
BCL6	*LPP*	2.189676	10.494
LPP	*LPP*	3.889396	5.031
CTBP1	*MAEA*	19.709608	18.553
RNF212	*MAEA*	0.007606029	0.677
DGKQ	*MAEA*	6.5478916	4.204
FAM53A	*MAEA*	0.5075767	0.237
TMEM175	*MAEA*	14.789575	15.925
WFS1	*WFS1*	48.648613	64.697
S100P	*WFS1*	0.094338536	0.239
KIAA0232	*WFS1*	13.799467	20.42
ARFIP1	*TMEM154*	15.90032	12.549
TIGD4	*TMEM154*	1.2456352	1.317
FBXW7	*TMEM154*	5.181745	13.471
FHDC1	*TMEM154*	2.2825546	3.897
TMEM154	*TMEM154*	0.09626208	0.536
NDUFS4	*ARL15*	19.788746	11.809
ARL15	*ARL15*	15.944361	23.853
MAP3K1	*ANKRD55*	3.182025	3.707
MIER3	*ANKRD55*	9.665144	10.573
PDE8B	*ZBED3*	1.3867581	9.642
F2R	*ZBED3*	1.0097425	1.087
ZBED3	*ZBED3*	8.214433	10.218
S100Z	*ZBED3*	5.69416	0.415
RREB1	*SSR1/RREB1*	8.097911	9.925
BMP6	*SSR1/RREB1*	0.097318925	0.079
SNRNP48	*SSR1/RREB1*	17.41273	6.136
CAGE1	*SSR1/RREB1*	0.026015453	0.022
E2F3	*CDKAL1*	7.951786	3.897
MBOAT1	*CDKAL1*	0.6750357	0.225
CDKAL1	*CDKAL1*	9.097854	4.014
HLA-C	*POU5F1/TCF19*	20.314272	128.262
HLA-B	*POU5F1/TCF19*	1.0356976	89.07
PRR3	*POU5F1/TCF19*	10.210759	8.001
GNL1	*POU5F1/TCF19*	18.412489	10.526
BTBD9	*ZFAND3*	3.0280964	5.045
GLO1	*ZFAND3*	60.216225	38.375
ZFAND3	*ZFAND3*	32.21161	30.611

SAYSD1	*ZFAND3*	12.916478	13.804
KIF6	*KCNK16*	1.6660756	0.227
GLP1R	*KCNK16*	1.8521343	54.753
KCNK5	*KCNK16*	0.2531769	0.361
DAAM2	*KCNK16*	0.24726911	0.065
KCNK17	*KCNK16*	104.62451	52.748
KCNK16	*KCNK16*	102.129074	169.1
DGKB	*DGKB*	0.056520004	2.545
CREB5	*JAZF1*	0.03463798	2.023
TAX1BP1	*JAZF1*	15.722485	37.226
JAZF1	*JAZF1*	1.3256978	5.059
HIBADH	*JAZF1*	7.88914	21.946
GCK	*GCK*	42.563545	21.444
YKT6	*GCK*	36.720226	26.003
UBE2D4	*GCK*	8.489575	8.93
TMED4	*GCK*	41.99634	44.895
PAX4	*GCC1*	1.3463995	1.025
SND1	*GCC1*	54.31868	38.648
RBM28	*GCC1*	24.012737	7.622
METTL2B	*GCC1*	16.067684	4.139
COPG2	*KLF14*	28.098372	6.619
CPA1	*KLF14*	21.122427	215.01
MKLN1	*KLF14*	13.815692	10.359
KLF14	*KLF14*	0.30431783	0.105
NKX6-3	*ANK1*	46.473755	4.079
SFRP1	*ANK1*	19.806135	0.178
GINS4	*ANK1*	4.1673384	0.309
AP3M2	*ANK1*	14.139829	5.208
TP53INP1	*TP53INP1*	34.056793	21.408
CCNE2	*TP53INP1*	11.706554	0.69
NDUFAF6	*TP53INP1*	3.0254307	1.929
DPY19L4	*TP53INP1*	10.406852	5.737
SLC30A8	*SLC30A8*	50.082935	517.613
RAD21	*SLC30A8*	125.50391	43.451
MED30	*SLC30A8*	5.853778	7.912
EIF3H	*SLC30A8*	30.243889	23.136
GLIS3	*GLIS3*	9.963094	16.547
AK3	*GLIS3*	41.39607	41.579
SLC1A1	*GLIS3*	0.668251	2.219
CDC37L1	*GLIS3*	6.7730727	8.744
CDKN2B	*CDKN2A/B*	14.910821	2.787
DMRTA1	*CDKN2A/B*	2.3605123	4.108
CDKN2A	*CDKN2A/B*	16.370024	3.16
MTAP	*CDKN2A/B*	7.5099783	2.603
TLE4	*TLE4*	6.7753386	1.565
TLE1	*TLE1*	18.546194	10.735
GPSM1	*GPSM1*	6.6489964	9.399
FAM69B	*GPSM1*	12.96178	20.863
C8G	*GPSM1*	0.25058675	0.447
CAMK1D	*CDC123/CAMK1D*	3.047827	1.783
CDC123	*CDC123/CAMK1D*	21.241869	13.717
UPF2	*CDC123/CAMK1D*	5.952305	7.418
SEC61A2	*CDC123/CAMK1D*	4.153639	3.573
NUDT5	*CDC123/CAMK1D*	11.880336	6.168
DHTKD1	*CDC123/CAMK1D*	7.0249395	4.896
HK1	*VPS26A*	0.039167475	0.905
NEUROG3	*VPS26A*	1.7002876	0.003

VPS26A	*VPS26A*	17.768085	10.95
STOX1	*VPS26A*	4.8269176	0.077
C10orf35	*VPS26A*	12.514925	7.626
PPIF	*ZMIZ1*	33.738903	8.455
SFTPA1	*ZMIZ1*	0.25245523	12.246
ZMIZ1	*ZMIZ1*	25.735308	34.923
ZCCHC24	*ZMIZ1*	0.43032268	0.357
HHEX	*HHEX/IDE*	0.06326138	1.421
IDE	*HHEX/IDE*	6.702786	5.403
KIF11	*HHEX/IDE*	19.037043	0.48
CPEB3	*HHEX/IDE*	2.9978027	5.982
TCF7L2	*TCF7L2*	2.292974	5.015
VTI1A	*TCF7L2*	2.0465448	4.366
GRK5	*GRK5*	4.9136653	0.907
NANOS1	*GRK5*	3.1578026	7.825
PRDX3	*GRK5*	49.93399	32.619
SFXN4	*GRK5*	6.9072022	5.75
INPP5F	*GRK5*	2.946096	41.344
PRLHR	*GRK5*	2.050672	1.39
KCNJ11	*KCNJ11*	28.18143	40.775
ABCC8	*KCNJ11*	67.27029	244.683
RPS13	*KCNJ11*	468.69717	83.633
PLEKHA7	*KCNJ11*	6.322894	3.168
NCR3LG1	*KCNJ11*	24.444202	13.894
ARAP1	*ARAP1 (CENTD2)*	13.448667	14.159
INPPL1	*ARAP1 (CENTD2)*	44.3585	9.872
STARD10	*ARAP1 (CENTD2)*	111.17294	36.176
CLPB	*ARAP1 (CENTD2)*	6.619748	4.168
ATG16L2	*ARAP1 (CENTD2)*	7.742456	2.648
MTNR1B	*MTNR1B*	0.015433707	0.642
FAT3	*MTNR1B*	0.14323217	0.037
SLC36A4	*MTNR1B*	19.859741	15.109
PTHLH	*KLHDC5*	0.07151515	0.192
PPFIBP1	*KLHDC5*	4.0497327	2.349
MRPS35	*KLHDC5*	49.13422	21.838
ARNTL2	*KLHDC5*	5.9386883	0.446
STK38L	*KLHDC5*	9.114885	6.608
REP15	*KLHDC5*	0.3843115	0.272
KLHL42	*KLHDC5*	26.8217	30.478
HMGA2	*HMGA2*	0.013419024	0.009
IRAK3	*HMGA2*	0.22330433	0.329
HELB	*HMGA2*	0.51851743	1.773
TMBIM4	*HMGA2*	3.3154817	1.907
LGR5	*TSPAN8/LGR5*	0.09165835	0.173
PTPRR	*TSPAN8/LGR5*	0.03522559	0.137
TPH2	*TSPAN8/LGR5*	0.2344639	0.175
RAB21	*TSPAN8/LGR5*	25.835938	22.635
HNF1A	*HNF1A (TCF1)*	11.920588	4.427
CABP1	*HNF1A (TCF1)*	0.47474712	0.92
CAMKK2	*HNF1A (TCF1)*	5.4568853	2.067
KDM2B	*HNF1A (TCF1)*	4.485531	2.082
SPPL3	*HNF1A (TCF1)*	21.66794	9.827
MLEC	*HNF1A (TCF1)*	52.709465	36.949
CDK2AP1	*MPHOSPH9*	12.791418	28.311
OGFOD2	*MPHOSPH9*	3.5529344	4.656
ABCB9	*MPHOSPH9*	2.5140104	1.566
EIF2B1	*MPHOSPH9*	23.559034	9.742

TCTN2	*MPHOSPH9*	9.149691	5.229
ZCCHC8	*MPHOSPH9*	6.5567923	6.602
RASGRP1	*RASGRP1*	0.6741565	7.274
SPRED1	*RASGRP1*	2.1910295	1.428
FAM98B	*RASGRP1*	12.422814	5.6
VPS13C	*C2CD4A*	23.849556	23.607
HMG20A	*HMG20A*	14.298759	12.474
PSTPIP1	*HMG20A*	0.33292773	0.032
LINGO1	*HMG20A*	7.0082254	0.117
TBC1D2B	*HMG20A*	9.462256	7.787
FAH	*ZFAND6*	0.85486287	1.303
ARNT2	*ZFAND6*	24.756863	11.261
MTHFS	*ZFAND6*	4.262472	3.327
ZFAND6	*ZFAND6*	37.31066	19.454
MESP1	*AP3S2*	3.0201075	0.914
MESP2	*AP3S2*	0.10175843	0.035
ZNF710	*AP3S2*	6.5410094	4.404
MFGE8	*AP3S2*	19.445135	0.704
VPS33B	*PRC1*	7.206141	4.765
BLM	*PRC1*	4.386412	0.217
HDDC3	*PRC1*	7.861545	5.238
RCCD1	*PRC1*	14.031963	1.311
FTO	*FTO*	5.057489	7.094
RBL2	*FTO*	21.088936	17.104
RPGRIP1L	*FTO*	2.059937	0.574
AKTIP	*FTO*	6.018698	6.684
CHD9	*FTO*	8.3386135	13.784
CFDP1	*BCAR1*	23.719404	10.403
BCAR1	*BCAR1*	14.373125	20.272
ZFP1	*BCAR1*	7.834037	6.043
ADAT1	*BCAR1*	13.082058	8.633
SRR	*SRR*	6.8297243	2.776
PAFAH1B1	*SRR*	59.24745	43.372
MNT	*SRR*	5.6960053	9.685
RTN4RL1	*SRR*	2.3101637	10.227
METTL16	*SRR*	9.847925	12.353
SGSM2	*SRR*	23.97471	34.762
SLC16A11	*SLC16A11*	5.575549	1.285
SLC2A4	*SLC16A11*	0.0136352	0.013
PHF23	*SLC16A11*	13.260954	10.724
TNFSF13	*SLC16A11*	2.1491694	6.405
HNF1B	*HNF1B (TCF2)*	0.030431787	2.781
ACACA	*HNF1B (TCF2)*	44.620758	10.826
GPR179	*HNF1B (TCF2)*	1.1049079	0.087
DUSP14	*HNF1B (TCF2)*	7.0064955	5.688
SOCS7	*HNF1B (TCF2)*	6.7027836	5.675
DDX52	*HNF1B (TCF2)*	10.564817	5.468
TBC1D3F	*HNF1B (TCF2)*	0.32611838	0.114
LAMA1	*LAMA1*	0.92843896	0.15
PTPRM	*LAMA1*	1.0576878	7.344
ARHGAP28	*LAMA1*	0.99569255	0.08
MC4R	*MC4R*	0.09446975	0.001
PMAIP1	*MC4R*	35.006386	3.308
CCBE1	*MC4R*	0.057355918	0.151
BCL2	*BCL2A*	0.53563124	0.175
KDSR	*BCL2A*	9.550464	10.958
PHLPP1	*BCL2A*	6.758193	2.535

VPS4B	*BCL2A*	24.030014	14.894
TM6SF2	*CILP2*	0.92930496	1.832
YJEFN3	*CILP2*	1.3774142	1.29
ZNF486	*CILP2*	0.0070848	1.712
TDRD12	*PEPD*	0.051692348	0.076
CEBPA	*PEPD*	14.847472	0.377
CEBPG	*PEPD*	29.904247	20.203
KCTD15	*PEPD*	2.694448	0.526
CEP89	*PEPD*	4.883483	3.482
HNF4A	*HNF4A*	21.057198	6.233
PABPC1L	*HNF4A*	19.464327	3.764
OSER1	*HNF4A*	23.073265	20.905
INS	*DUSP8*	8012.452	166347.111
IGF2	*DUSP8*	95.30156	8.054
MRPL23	*DUSP8*	15.220552	7.366
CTSD	*DUSP8*	150.01717	134.269
TH	*DUSP8*	1.5446588	0.735

Bibliography

[1] IDF, International Diabetes Federation Diabetes Atlas, 7 ed., Belgium: International Diabetes Federation, http://www.diabetesatlas.org, 2015.

[2] P. Almgren, M. Lehtovirta, B. Isomaa, L. Sarelin, M.R. Taskinen, V. Lyssenko, T. Tuomi, and L. Groop, Heritability and familiality of type 2 diabetes and related quantitative traits in the Botnia Study. Diabetologia 54 (2011) 2811-9.

[3] F.B. Hu, Globalization of diabetes: the role of diet, lifestyle, and genes. Diabetes Care 34 (2011) 1249-57.

[4] J.B. Meigs, L.A. Cupples, and P.W. Wilson, Parental transmission of type 2 diabetes: the Framingham Offspring Study. Diabetes 49 (2000) 2201-7.

[5] P. Poulsen, K.O. Kyvik, A. Vaag, and H. Beck-Nielsen, Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance--a population-based twin study. Diabetologia 42 (1999) 139-45.

[6] WHO, Global report on diabetes, World Health Organization, Geneva, Switzerland, http://www.who.int/diabetes/global-report/en/, 2016.

[7] J.C. Floyd, Jr., S.S. Fajans, J.W. Conn, R.F. Knopf, and J. Rull, Stimulation of insulin secretion by amino acids. J Clin Invest 45 (1966) 1487-502.

[8] P.E. MacDonald, W. El-Kholy, M.J. Riedel, A.M. Salapatek, P.E. Light, and M.B. Wheeler, The multiple actions of GLP-1 on the process of glucose-stimulated insulin secretion. Diabetes 51 Suppl 3 (2002) S434-42.

[9] E.A. Newsholme, and G. Dimitriadis, Integration of biochemical and physiologic effects of insulin on glucose metabolism. Exp Clin Endocrinol Diabetes 109 Suppl 2 (2001) S122-34.

[10] J. Lee, and P.F. Pilch, The insulin receptor: structure, function, and signaling. Am J Physiol 266 (1994) C319-34.

[11] J. Boucher, A. Kleinridders, and C.R. Kahn, Insulin receptor signaling in normal and insulin-resistant states. Cold Spring Harb Perspect Biol 6 (2014).

[12] B. Draznin, Molecular mechanisms of insulin resistance: serine phosphorylation of insulin receptor substrate-1 and increased expression of p85alpha: the two sides of a coin. Diabetes 55 (2006) 2392-7.

[13] M.J. Birnbaum, Identification of a novel gene encoding an insulin-responsive glucose transporter protein. Cell 57 (1989) 305-15.

[14] S. Marshall, and J.M. Olefsky, Characterization of insulin-induced receptor loss and evidence for internalization of the insulin receptor. Diabetes 30 (1981) 746-53.

[15] K.M. Habegger, K.M. Heppner, N. Geary, T.J. Bartness, R. DiMarchi, and M.H. Tschop, The metabolic actions of glucagon revisited. Nat Rev Endocrinol 6 (2010) 689-97.

[16] L.J. Jelinek, S. Lok, G.B. Rosenberg, R.A. Smith, F.J. Grant, S. Biggs, P.A. Bensch, J.L. Kuijper, P.O. Sheppard, C.A. Sprecher, and et al., Expression cloning and signaling properties of the rat glucagon receptor. Science 259 (1993) 1614-6.

[17] N. Moller, J.O. Jorgensen, J. Moller, L. Orskov, P. Ovesen, O. Schmitz, J.S. Christiansen, and H. Orskov, Metabolic effects of growth hormone in humans. Metabolism 44 (1995) 33-6.

[18] T.H. Meek, and G.J. Morton, The role of leptin in diabetes: metabolic effects. Diabetologia 59 (2016) 928-32.

[19] J.E. Campbell, and D.J. Drucker, Pharmacology, physiology, and mechanisms of incretin hormone action. Cell Metab 17 (2013) 819-37.

[20] A.P. Coll, and G.S. Yeo, The hypothalamus and metabolism: integrating signals to control energy and glucose homeostasis. Curr Opin Pharmacol 13 (2013) 970-6.

[21] P.W. Franks, and M.I. McCarthy, Exposing the exposures responsible for type 2 diabetes and obesity. Science 354 (2016) 69-73.

[22] M. Bjornholm, and J.R. Zierath, Insulin signal transduction in human skeletal muscle: identifying the defects in Type II diabetes. Biochem Soc Trans 33 (2005) 354-7.

[23] F. Mauvais-Jarvis, and C.R. Kahn, Understanding the pathogenesis and treatment of insulin resistance and type 2 diabetes mellitus: what can we learn from transgenic and knockout mice? Diabetes Metab 26 (2000) 433-48.
[24] M. Stumvoll, B.J. Goldstein, and T.W. van Haeften, Type 2 diabetes: principles of pathogenesis and therapy. Lancet 365 (2005) 1333-46.
[25] A. Festa, K. Williams, R. D'Agostino, Jr., L.E. Wagenknecht, and S.M. Haffner, The natural course of beta-cell function in nondiabetic and diabetic individuals: the Insulin Resistance Atherosclerosis Study. Diabetes 55 (2006) 1114-20.
[26] A.P. Morris, B.F. Voight, T.M. Teslovich, T. Ferreira, A.V. Segre, V. Steinthorsdottir, R.J. Strawbridge, H. Khan, H. Grallert, A. Mahajan, I. Prokopenko, H.M. Kang, C. Dina, T. Esko, R.M. Fraser*, et al.*, Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet 44 (2012) 981-90.
[27] B.F. Voight, L.J. Scott, V. Steinthorsdottir, A.P. Morris, C. Dina, R.P. Welch, E. Zeggini, C. Huth, Y.S. Aulchenko, G. Thorleifsson, L.J. McCulloch, T. Ferreira, H. Grallert, N. Amin, G. Wu*, et al.*, Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet 42 (2010) 579-89.
[28] T.A. Manolio, F.S. Collins, N.J. Cox, D.B. Goldstein, L.A. Hindorff, D.J. Hunter, M.I. McCarthy, E.M. Ramos, L.R. Cardon, A. Chakravarti, J.H. Cho, A.E. Guttmacher, A. Kong, L. Kruglyak, E. Mardis*, et al.*, Finding the missing heritability of complex diseases. Nature 461 (2009) 747-53.
[29] J.J. Holst, T. Vilsboll, and C.F. Deacon, The incretin system and its role in type 2 diabetes mellitus. Mol Cell Endocrinol 297 (2009) 127-36.
[30] A.F. Godoy-Matos, The role of glucagon on type 2 diabetes at a glance. Diabetol Metab Syndr 6 (2014) 91.
[31] R. Taylor, Insulin resistance and type 2 diabetes. Diabetes 61 (2012) 778-9.
[32] G. Gibson, Decanalization and the origin of complex disease. Nat Rev Genet 10 (2009) 134-40.
[33] F.M. Ashcroft, and P. Rorsman, Diabetes mellitus and the beta cell: the last ten years. Cell 148 (2012) 1160-71.
[34] S.E. Flanagan, S. Clauin, C. Bellanne-Chantelot, P. de Lonlay, L.W. Harries, A.L. Gloyn, and S. Ellard, Update of mutations in the genes encoding the pancreatic beta-cell K(ATP) channel subunits Kir6.2 (KCNJ11) and sulfonylurea receptor 1 (ABCC8) in diabetes mellitus and hyperinsulinism. Hum Mutat 30 (2009) 170-80.
[35] O. Zuk, S.F. Schaffner, K. Samocha, R. Do, E. Hechter, S. Kathiresan, M.J. Daly, B.M. Neale, S.R. Sunyaev, and E.S. Lander, Searching for missing heritability: Designing rare variant association studies. Proc Natl Acad Sci U S A 111 (2014) E455-64.
[36] C. Fuchsberger, J. Flannick, T.M. Teslovich, A. Mahajan, V. Agarwala, K.J. Gaulton, C. Ma, P. Fontanillas, L. Moutsianas, D.J. McCarthy, M.A. Rivas, J.R. Perry, X. Sim, T.W. Blackwell, N.R. Robertson*, et al.*, The genetic architecture of type 2 diabetes. Nature 536 (2016) 41-7.
[37] J.R. Huyghe, A.U. Jackson, M.P. Fogarty, M.L. Buchkovich, A. Stancakova, H.M. Stringham, X. Sim, L. Yang, C. Fuchsberger, H. Cederberg, P.S. Chines, T.M. Teslovich, J.M. Romm, H. Ling, I. McMullen*, et al.*, Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. Nat Genet 45 (2013) 197-201.
[38] V. Steinthorsdottir, G. Thorleifsson, and P. Sulem, Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. Nat Genet 46 (2014) 294-8.
[39] K.E. Lohmueller, T. Sparso, Q. Li, E. Andersson, T. Korneliussen, A. Albrechtsen, K. Banasik, N. Grarup, I. Hallgrimsdottir, K. Kiil, T.O. Kilpelainen, N.T. Krarup, T.H. Pers, G. Sanchez, Y. Hu*, et al.*, Whole-exome sequencing of 2,000 Danish individuals and the role of rare coding variants in type 2 diabetes. Am J Hum Genet 93 (2013) 1072-86.
[40] E.A. Boyle, Y.I. Li, and J.K. Pritchard, An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell 169 (2017) 1177-1186.
[41] G.V. Kryukov, L.A. Pennacchio, and S.R. Sunyaev, Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. Am J Hum Genet 80 (2007) 727-39.

[42] D.B. Goldstein, A. Allen, J. Keebler, E.H. Margulies, S. Petrou, S. Petrovski, and S. Sunyaev, Sequencing studies in human genetics: design and interpretation. Nat Rev Genet 14 (2013) 460-70.
[43] A. Bonnefond, N. Clement, K. Fawcett, L. Yengo, E. Vaillant, J.L. Guillaume, A. Dechaume, F. Payne, R. Roussel, S. Czernichow, S. Hercberg, S. Hadjadj, B. Balkau, M. Marre, O. Lantieri, *et al.*, Rare MTNR1B variants impairing melatonin receptor 1B function contribute to type 2 diabetes. Nat Genet 44 (2012) 297-301.
[44] J. Flannick, G. Thorleifsson, N.L. Beer, and S.B. Jacobs, Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. (2014).
[45] I. Prokopenko, C. Langenberg, J.C. Florez, R. Saxena, N. Soranzo, G. Thorleifsson, R.J. Loos, A.K. Manning, A.U. Jackson, Y. Aulchenko, S.C. Potter, M.R. Erdos, S. Sanna, J.J. Hottenga, E. Wheeler, *et al.*, Variants in MTNR1B influence fasting glucose levels. Nat Genet 41 (2009) 77-81.
[46] V. Lyssenko, C.L. Nagorny, M.R. Erdos, N. Wierup, A. Jonsson, P. Spegel, M. Bugliani, R. Saxena, M. Fex, N. Pulizzi, B. Isomaa, T. Tuomi, P. Nilsson, J. Kuusisto, J. Tuomilehto, *et al.*, Common variant in MTNR1B associated with increased risk of type 2 diabetes and impaired early insulin secretion. Nat Genet 41 (2009) 82-8.
[47] K.J. Gaulton, T. Ferreira, Y. Lee, A. Raimondo, R. Magi, M.E. Reschen, A. Mahajan, A. Locke, N. William Rayner, N. Robertson, R.A. Scott, I. Prokopenko, L.J. Scott, T. Green, T. Sparso, *et al.*, Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. Nat Genet 47 (2015) 1415-25.
[48] L.D. Pound, S.A. Sarkar, A. Ustione, P.K. Dadi, M.K. Shadoan, C.E. Lee, J.A. Walters, M. Shiota, O.P. McGuinness, D.A. Jacobson, D.W. Piston, J.C. Hutton, D.R. Powell, and R.M. O'Brien, The physiological effects of deleting the mouse SLC30A8 gene encoding zinc transporter-8 are influenced by gender and genetic background. PLoS One 7 (2012) e40972.
[49] L.D. Pound, S.A. Sarkar, R.K. Benninger, Y. Wang, A. Suwanichkul, M.K. Shadoan, R.L. Printz, J.K. Oeser, C.E. Lee, D.W. Piston, O.P. McGuinness, J.C. Hutton, D.R. Powell, and R.M. O'Brien, Deletion of the mouse Slc30a8 gene encoding zinc transporter-8 results in impaired insulin secretion. Biochem J 421 (2009) 371-6.
[50] K. Lemaire, M.A. Ravier, A. Schraenen, J.W. Creemers, R. Van de Plas, M. Granvik, L. Van Lommel, E. Waelkens, F. Chimienti, G.A. Rutter, P. Gilon, P.A. in't Veld, and F.C. Schuit, Insulin crystallization depends on zinc transporter ZnT8 expression, but is not required for normal glucose homeostasis in mice. Proc Natl Acad Sci U S A 106 (2009) 14872-7.
[51] N. Wijesekara, F.F. Dai, A.B. Hardy, P.R. Giglou, A. Bhattacharjee, V. Koshkin, F. Chimienti, H.Y. Gaisano, G.A. Rutter, and M.B. Wheeler, Beta cell-specific Znt8 deletion in mice causes marked defects in insulin processing, crystallisation and secretion. Diabetologia 53 (2010) 1656-68.
[52] T.J. Nicolson, E.A. Bellomo, N. Wijesekara, M.K. Loder, J.M. Baldwin, A.V. Gyulkhandanyan, V. Koshkin, A.I. Tarasov, R. Carzaniga, K. Kronenberger, T.K. Taneja, G. da Silva Xavier, S. Libert, P. Froguel, R. Scharfmann, *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 (2009) 2070-83.
[53] J. Flannick, N.L. Beer, A.G. Bick, V. Agarwala, J. Molnes, N. Gupta, N.P. Burtt, J.C. Florez, J.B. Meigs, H. Taylor, V. Lyssenko, H. Irgens, E. Fox, F. Burslem, S. Johansson, *et al.*, Assessing the phenotypic effects in the general population of rare variants in genes for a dominant Mendelian form of diabetes. 45 (2013) 1380-5.
[54] S. Althari, and A.L. Gloyn, When is it MODY? Challenges in the Interpretation of Sequence Variants in MODY Genes. Rev Diabet Stud 12 (2015) 330-48.
[55] D.G. MacArthur, T.A. Manolio, D.P. Dimmock, H.L. Rehm, J. Shendure, G.R. Abecasis, D.R. Adams, R.B. Altman, S.E. Antonarakis, E.A. Ashley, J.C. Barrett, L.G. Biesecker, D.F. Conrad, G.M. Cooper, N.J. Cox, *et al.*, Guidelines for investigating causality of sequence variants in human disease. Nature 508 (2014) 469-76.
[56] A.L. Gloyn, E.R. Pearson, J.F. Antcliff, P. Proks, G.J. Bruining, A.S. Slingerland, N. Howard, S. Srinivasan, J.M. Silva, J. Molnes, E.L. Edghill, T.M. Frayling, I.K. Temple, D. Mackay, J.P.

Shield, *et al.*, Activating mutations in the gene encoding the ATP-sensitive potassium-channel subunit Kir6.2 and permanent neonatal diabetes. N Engl J Med 350 (2004) 1838-49.
[57] E.R. Pearson, I. Flechtner, P.R. Njolstad, M.T. Malecki, S.E. Flanagan, B. Larkin, F.M. Ashcroft, I. Klimes, E. Codner, V. Iotova, A.S. Slingerland, J. Shield, J.J. Robert, J.J. Holst, P.M. Clark, *et al.*, Switching from insulin to oral sulfonylureas in patients with diabetes due to Kir6.2 mutations. N Engl J Med 355 (2006) 467-77.
[58] J.V. Sagen, H. Raeder, E. Hathout, N. Shehadeh, K. Gudmundsson, H. Baevre, D. Abuelo, C. Phornphutkul, J. Molnes, G.I. Bell, A.L. Gloyn, A.T. Hattersley, A. Molven, O. Sovik, and P.R. Njolstad, Permanent neonatal diabetes due to mutations in KCNJ11 encoding Kir6.2: patient characteristics and initial response to sulfonylurea therapy. Diabetes 53 (2004) 2713-8.
[59] M.I. McCarthy, Painting a new picture of personalised medicine for diabetes. Diabetologia (In press).
[60] E.A.M. Gale, Declassifying diabetes. Diabetologia 49 (2006) 1989-1995.
[61] L.D. Ward, and M. Kellis, Interpreting noncoding genetic variation in complex traits and human disease. Nat Biotechnol 30 (2012) 1095-106.
[62] M. Li, C. Li, and W. Guan, Evaluation of coverage variation of SNP chips for genome-wide association studies. Eur J Hum Genet 16 (2008) 635-43.
[63] A.P. Morris, Fine mapping of type 2 diabetes susceptibility loci. Curr Diab Rep 14 (2014) 549.
[64] M.S. Udler, J. Tyrer, and D.F. Easton, Evaluating the power to discriminate between highly correlated SNPs in genetic association studies. Genet Epidemiol 34 (2010) 463-8.
[65] A. Mahajan, J. Wessel, S. Willems, W. Zhao, N.R. Robertson, A.Y. Chu, W. Gan, H. Kitajima, R.A. Scott, I. Barroso, T.M. Frayling, M.O. Goodarzi, J.B. Meigs, M. Boehnke, D. Saleheen, *et al.*, Refining The Accuracy Of Validated Target Identification Through Coding Variant Fine-Mapping In Type 2 Diabetes. bioRxiv (2017).
[66] M. Kasowski, S. Kyriazopoulou-Panagiotopoulou, F. Grubert, J.B. Zaugg, A. Kundaje, Y. Liu, A.P. Boyle, Q.C. Zhang, F. Zakharia, D.V. Spacek, J. Li, D. Xie, A. Olarerin-George, L.M. Steinmetz, J.B. Hogenesch, *et al.*, Extensive Variation in Chromatin States Across Humans. Science 342 (2013) 750-752.
[67] H. Kilpinen, S.M. Waszak, A.R. Gschwind, S.K. Raghav, R.M. Witwicki, A. Orioli, E. Migliavacca, M. Wiederkehr, M. Gutierrez-Arcelus, N.I. Panousis, A. Yurovsky, T. Lappalainen, L. Romano-Palumbo, A. Planchon, D. Bielser, *et al.*, Coordinated Effects of Sequence Variation on DNA Binding, Chromatin Structure, and Transcription. Science 342 (2013) 744-747.
[68] G. McVicker, B. van de Geijn, J.F. Degner, C.E. Cain, N.E. Banovich, A. Raj, N. Lewellen, M. Myrthil, Y. Gilad, and J.K. Pritchard, Identification of Genetic Variants That Affect Histone Modifications in Human Cells. Science 342 (2013) 747-749.
[69] T. Handstad, M. Rye, R. Mocnik, F. Drablos, and P. Saetrom, Cell-type specificity of ChIP-predicted transcription factor binding sites. BMC Genomics 13 (2012) 372.
[70] W.A. Whyte, D.A. Orlando, D. Hnisz, B.J. Abraham, C.Y. Lin, M.H. Kagey, P.B. Rahl, T.I. Lee, and R.A. Young, Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell 153 (2013) 307-19.
[71] K.J. Gaulton, T. Nammo, L. Pasquali, J.M. Simon, P.G. Giresi, M.P. Fogarty, T.M. Panhuis, P. Mieczkowski, A. Secchi, D. Bosco, T. Berney, E. Montanya, K.L. Mohlke, J.D. Lieb, and J. Ferrer, A map of open chromatin in human pancreatic islets. Nature Genetics 42 (2010) 255-9.
[72] S.C.J. Parker, M.L. Stitzel, D.L. Taylor, J.M. Orozco, M.R. Erdos, J.A. Akiyama, K.L. van Bueren, P.S. Chines, N. Narisu, B.L. Black, A. Visel, L.A. Pennacchio, F.S. Collins, and N.C. Sequencing, Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. Proc Natl Acad Sci U S A 110 (2013) 17921-17926.
[73] L. Pasquali, K.J. Gaulton, S.A. Rodriguez-Segui, L. Mularoni, I. Miguel-Escalada, I. Akerman, J.J. Tena, I. Moran, C. Gomez-Marin, M. van de Bunt, J. Ponsa-Cobas, N. Castro, T. Nammo, I. Cebola, J. Garcia-Hurtado, *et al.*, Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. Nat Genet 46 (2014) 136-43.

[74] G. Trynka, C. Sandor, B. Han, H. Xu, B.E. Stranger, X.S. Liu, and S. Raychaudhuri, Chromatin marks identify critical cell types for fine mapping complex trait variants. Nat Genet 45 (2013) 124-30.
[75] A.S. Dimas, V. Lagou, A. Barker, J.W. Knowles, R. Magi, M.F. Hivert, A. Benazzo, D. Rybin, A.U. Jackson, H.M. Stringham, C. Song, A. Fischer-Rosinsky, T.W. Boesgaard, N. Grarup, F.A. Abbasi, *et al.*, Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. Diabetes 63 (2014) 2158-71.
[76] J.M. Torres, E.R. Gamazon, E.J. Parra, J.E. Below, A. Valladares-Salgado, N. Wacher, M. Cruz, C.L. Hanis, and N.J. Cox, Cross-tissue and tissue-specific eQTLs: partitioning the heritability of a complex trait. Am J Hum Genet 95 (2014) 521-34.
[77] M. van de Bunt, J.E. Manning Fox, X. Dai, A. Barrett, C. Grey, L. Li, A.J. Bennett, P.R. Johnson, R.V. Rajotte, K.J. Gaulton, E.T. Dermitzakis, P.E. MacDonald, M.I. McCarthy, and A.L. Gloyn, 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 11 (2015) e1005694-e1005694.
[78] M.N. Weedon, I. Cebola, A.M. Patch, S.E. Flanagan, E. De Franco, R. Caswell, S.A. Rodriguez-Segui, C. Shaw-Smith, C.H. Cho, H. Lango Allen, J.A. Houghton, C.L. Roth, R. Chen, K. Hussain, P. Marsh, *et al.*, Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. Nat Genet 46 (2014) 61-4.
[79] S.K. Thomsen, M.I. McCarthy, and A.L. Gloyn, The Importance of Context: Uncovering Species- and Tissue-Specific Effects of Genetic Risk Variants for Type 2 Diabetes. Frontiers in Endocrinology 7 (2016).
[80] H. Lango Allen, S.E. Flanagan, C. Shaw-Smith, E. De Franco, I. Akerman, R. Caswell, J. Ferrer, A.T. Hattersley, and S. Ellard, GATA6 haploinsufficiency causes pancreatic agenesis in humans. Nat Genet 44 (2012) 20-2.
[81] G.S. Sellick, K.T. Barker, I. Stolte-Dijkstra, C. Fleischmann, R.J. Coleman, C. Garrett, A.L. Gloyn, E.L. Edghill, A.T. Hattersley, P.K. Wellauer, G. Goodwin, and R.S. Houlston, Mutations in PTF1A cause pancreatic and cerebellar agenesis. Nat Genet 36 (2004) 1301-5.
[82] I. Moltke, N. Grarup, M.E. Jorgensen, P. Bjerregaard, J.T. Treebak, M. Fumagalli, T.S. Korneliussen, M.A. Andersen, T.S. Nielsen, N.T. Krarup, A.P. Gjesing, J.R. Zierath, A. Linneberg, X. Wu, G. Sun, *et al.*, A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes. Nature 512 (2014) 190-3.
[83] D. Baus, K. Heermeier, M. De Hoop, C. Metz-Weidmann, J. Gassenhuber, W. Dittrich, S. Welte, and N. Tennagels, Identification of a novel AS160 splice variant that regulates GLUT4 translocation and glucose-uptake in rat muscle cells. Cellular Signalling 20 (2008) 2237-2246.
[84] Y. Okada, D. Wu, G. Trynka, T. Raj, C. Terao, K. Ikari, Y. Kochi, K. Ohmura, A. Suzuki, S. Yoshida, R.R. Graham, A. Manoharan, W. Ortmann, T. Bhangale, J.C. Denny, *et al.*, Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature 506 (2014) 376-81.
[85] A. Kundaje, W. Meuleman, J. Ernst, M. Bilenky, A. Yen, A. Heravi-Moussavi, P. Kheradpour, Z. Zhang, J. Wang, M.J. Ziller, V. Amin, J.W. Whitaker, M.D. Schultz, L.D. Ward, A. Sarkar, *et al.*, Integrative analysis of 111 reference human epigenomes. Nature 518 (2015) 317-330.
[86] S. Onengut-Gumuscu, W.M. Chen, O. Burren, N.J. Cooper, A.R. Quinlan, J.C. Mychaleckyj, E. Farber, J.K. Bonnie, M. Szpak, E. Schofield, P. Achuthan, H. Guo, M.D. Fortune, H. Stevens, N.M. Walker, *et al.*, Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. Nat Genet 47 (2015) 381-6.
[87] H.K. Finucane, B. Bulik-Sullivan, A. Gusev, G. Trynka, Y. Reshef, P.R. Loh, V. Anttila, H. Xu, C. Zang, K. Farh, S. Ripke, F.R. Day, S. Purcell, E. Stahl, S. Lindstrom, *et al.*, Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat Genet 47 (2015) 1228-35.
[88] K.S. Small, A.K. Hedman, E. Grundberg, A.C. Nica, G. Thorleifsson, A. Kong, U. Thorsteindottir, S.Y. Shin, H.B. Richards, N. Soranzo, K.R. Ahmadi, C.M. Lindgren, K. Stefansson, E.T. Dermitzakis, P. Deloukas, *et al.*, Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes. Nat Genet 43 (2011) 561-4.

[89] K. Musunuru, A. Strong, M. Frank-Kamenetsky, N.E. Lee, T. Ahfeldt, K.V. Sachs, X. Li, H. Li, N. Kuperwasser, V.M. Ruda, J.P. Pirruccello, B. Muchmore, L. Prokunina-Olsson, J.L. Hall, E.E. Schadt, *et al.*, From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. Nature 466 (2010) 714-9.
[90] I. Adrianto, F. Wen, A. Templeton, G. Wiley, J.B. King, C.J. Lessard, J.S. Bates, Y. Hu, J.A. Kelly, K.M. Kaufman, J.M. Guthridge, M.E. Alarcon-Riquelme, J.M. Anaya, S.C. Bae, S.Y. Bang, *et al.*, Association of a functional variant downstream of TNFAIP3 with systemic lupus erythematosus. Nat Genet 43 (2011) 253-8.
[91] B.P. Fairfax, S. Makino, J. Radhakrishnan, K. Plant, S. Leslie, A. Dilthey, P. Ellis, C. Langford, F.O. Vannberg, and J.C. Knight, Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. Nat Genet 44 (2012) 502-10.
[92] K.K. Farh, A. Marson, J. Zhu, M. Kleinewietfeld, W.J. Housley, S. Beik, N. Shoresh, H. Whitton, R.J. Ryan, A.A. Shishkin, M. Hatan, M.J. Carrasco-Alfonso, D. Mayer, C.J. Luckey, N.A. Patsopoulos, *et al.*, Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature 518 (2015) 337-43.
[93] O. Le Bacquer, J. Kerr-Conte, S. Gargani, N. Delalleau, M. Huyvaert, V. Gmyr, P. Froguel, B. Neve, and F. Pattou, TCF7L2 rs7903146 impairs islet function and morphology in non-diabetic individuals. Diabetologia 55 (2012) 2677-81.
[94] G. da Silva Xavier, A. Mondragon, G. Sun, L. Chen, J.A. McGinty, P.M. French, and G.A. Rutter, Abnormal glucose tolerance and insulin secretion in pancreas-specific Tcf7l2-null mice. Diabetologia 55 (2012) 2667-76.
[95] S.F. Boj, J.H. van Es, M. Huch, V.S. Li, A. Jose, P. Hatzis, M. Mokry, A. Haegebarth, M. van den Born, P. Chambon, P. Voshol, Y. Dor, E. Cuppen, C. Fillat, and H. Clevers, Diabetes risk gene and Wnt effector Tcf7l2/TCF4 controls hepatic response to perinatal and adult metabolic demand. Cell 151 (2012) 1595-607.
[96] A.H. Rosengren, M. Braun, T. Mahdi, S.A. Andersson, M.E. Travers, M. Shigeto, E. Zhang, P. Almgren, C. Ladenvall, A.S. Axelsson, A. Edlund, M.G. Pedersen, A. Jonsson, R. Ramracheya, Y. Tang, *et al.*, Reduced insulin exocytosis in human pancreatic beta-cells with gene variants linked to type 2 diabetes. Diabetes 61 (2012) 1726-33.
[97] G. da Silva Xavier, M.K. Loder, A. McDonald, A.I. Tarasov, R. Carzaniga, K. Kronenberger, S. Barg, and G.A. Rutter, TCF7L2 regulates late events in insulin secretion from pancreatic islet beta-cells. Diabetes 58 (2009) 894-905.
[98] M.L. Stitzel, P. Sethupathy, D.S. Pearson, P.S. Chines, L. Song, M.R. Erdos, R. Welch, S.C.J. Parker, A.P. Boyle, L.J. Scott, E.H. Margulies, M. Boehnke, T.S. Furey, G.E. Crawford, F.S. Collins, *et al.*, Global Epigenomic Analysis of Primary Human Pancreatic Islets Provides Insights into Type 2 Diabetes Susceptibility Loci. Cell Metab 12 (2010) 443-455.
[99] A. Melnikov, A. Murugan, X. Zhang, T. Tesileanu, L. Wang, P. Rogov, S. Feizi, A. Gnirke, C.G. Callan, Jr., J.B. Kinney, M. Kellis, E.S. Lander, and T.S. Mikkelsen, Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. Nat Biotechnol 30 (2012) 271-7.
[100] An integrated encyclopedia of DNA elements in the human genome. Nature 489 (2012) 57-74.
[101] M. Volkmar, S. Dedeurwaerder, D.A. Cunha, M.N. Ndlovu, M. Defrance, R. Deplus, E. Calonne, U. Volkmar, M. Igoillo-Esteve, N. Naamane, S. Del Guerra, M. Masini, M. Bugliani, P. Marchetti, M. Cnop, *et al.*, DNA methylation profiling identifies epigenetic dysregulation in pancreatic islets from type 2 diabetic patients. Embo Journal 31 (2012) 1405-1426.
[102] R. Bhandare, J. Schug, J. Le Lay, A. Fox, O. Smirnova, C. Liu, A. Naji, and K.H. Kaestner, Genome-wide analysis of histone modifications in human pancreatic islets. Genome Research 20 (2010) 428-433.
[103] T. Dayeh, P. Volkov, S. Salo, E. Hall, E. Nilsson, A.H. Olsson, C.L. Kirkpatrick, C.B. Wollheim, L. Eliasson, T. Ronn, K. Bacos, and C. Ling, Genome-wide DNA methylation analysis of human pancreatic islets from type 2 diabetic and non-diabetic donors identifies candidate genes that influence insulin secretion. PLoS Genet 10 (2014) e1004160.

[104] I. Moran, I. Akerman, M. van de Bunt, R. Xie, M. Benazra, T. Nammo, L. Arnes, N. Nakic, J. Garcia-Hurtado, S. Rodriguez-Segui, L. Pasquali, C. Sauty-Colace, A. Beucher, R. Scharfmann, J. van Arensbergen*, et al.*, Human beta cell transcriptome analysis uncovers lncRNAs that are tissue-specific, dynamically regulated, and abnormally expressed in type 2 diabetes. Cell Metab 16 (2012) 435-48.
[105] M. van de Bunt, K.J. Gaulton, L. Parts, I. Moran, P.R. Johnson, C.M. Lindgren, J. Ferrer, A.L. Gloyn, and M.I. McCarthy, The miRNA profile of human pancreatic islets and beta-cells and relationship to type 2 diabetes pathogenesis. PLoS One 8 (2013) e55272.
[106] V. Kameswaran, N.C. Bramswig, L.B. McKenna, M. Penn, J. Schug, N.J. Hand, Y. Chen, I. Choi, A. Vourekas, K.J. Won, C. Liu, K. Vivek, A. Naji, J.R. Friedman, and K.H. Kaestner, Epigenetic Regulation of the DLK1-MEG3 MicroRNA Cluster in Human Type 2 Diabetic Islets. Cell Metab 19 (2014) 135-45.
[107] A.C. Nica, H. Ongen, J.C. Irminger, D. Bosco, T. Berney, S.E. Antonarakis, P.A. Halban, and E.T. Dermitzakis, Cell-type, allelic, and genetic signatures in the human pancreatic beta cell transcriptome. Genome Res 23 (2013) 1554-62.
[108] J. Loven, H.A. Hoke, C.Y. Lin, A. Lau, D.A. Orlando, C.R. Vakoc, J.E. Bradner, T.I. Lee, and R.A. Young, Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell 153 (2013) 320-34.
[109] K. Stefflova, D. Thybert, M.D. Wilson, I. Streeter, J. Aleksic, P. Karagianni, A. Brazma, D.J. Adams, I. Talianidis, J.C. Marioni, P. Flicek, and D.T. Odom, Cooperativity and rapid evolution of cobound transcription factors in closely related mammals. Cell 154 (2013) 530-40.
[110] D. Schmidt, M.D. Wilson, B. Ballester, P.C. Schwalie, G.D. Brown, A. Marshall, C. Kutter, S. Watt, C.P. Martinez-Jimenez, S. Mackay, I. Talianidis, P. Flicek, and D.T. Odom, Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. Science 328 (2010) 1036-40.
[111] B. Ballester, A. Medina-Rivera, D. Schmidt, M. Gonzalez-Porta, M. Carlucci, X. Chen, K. Chessman, A.J. Faure, A.P. Funnell, A. Goncalves, C. Kutter, M. Lukk, S. Menon, W.M. McLaren, K. Stefflova*, et al.*, Multi-species, multi-transcription factor binding highlights conserved control of tissue-specific biological pathways. Elife 3 (2014) e02626.
[112] M. Claussnitzer, S.N. Dankel, B. Klocke, H. Grallert, V. Glunk, T. Berulava, H. Lee, N. Oskolkov, J. Fadista, K. Ehlers, S. Wahl, C. Hoffmann, K. Qian, T. Ronn, H. Riess*, et al.*, Leveraging cross-species transcription factor binding site patterns: from diabetes risk loci to disease mechanisms. Cell 156 (2014) 343-58.
[113] N. Bouatia-Naji, G. Rocheleau, L. Van Lommel, K. Lemaire, F. Schuit, C. Cavalcanti-Proenca, M. Marchand, A.L. Hartikainen, U. Sovio, F. De Graeve, J. Rung, M. Vaxillaire, J. Tichet, M. Marre, B. Balkau*, et al.*, A polymorphism within the G6PC2 gene is associated with fasting plasma glucose levels. Science 320 (2008) 1085-8.
[114] W.M. Chen, M.R. Erdos, A.U. Jackson, R. Saxena, S. Sanna, K.D. Silver, N.J. Timpson, T. Hansen, M. Orru, M. Grazia Piras, L.L. Bonnycastle, C.J. Willer, V. Lyssenko, H. Shen, J. Kuusisto*, et al.*, Variations in the G6PC2/ABCB11 genomic region are associated with fasting glucose levels. J Clin Invest 118 (2008) 2620-8.
[115] A. Mahajan, X. Sim, H.J. Ng, A. Manning, M.A. Rivas, H.M. Highland, A.E. Locke, N. Grarup, H.K. Im, P. Cingolani, J. Flannick, P. Fontanillas, C. Fuchsberger, K.J. Gaulton, T.M. Teslovich*, et al.*, Identification and functional characterization of G6PC2 coding variants influencing glycemic traits define an effector transcript at the G6PC2-ABCB11 locus. PLoS Genet 11 (2015) e1004876.
[116] J. Wessel, A.Y. Chu, S.M. Willems, S. Wang, H. Yaghootkar, J.A. Brody, M. Dauriz, M.F. Hivert, S. Raghavan, L. Lipovich, B. Hidalgo, K. Fox, J.E. Huffman, P. An, Y. Lu*, et al.*, Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility. Nat Commun 6 (2015) 5897.
[117] M.K. Andersen, C.E. Pedersen, I. Moltke, T. Hansen, A. Albrechtsen, and N. Grarup, Genetics of Type 2 Diabetes: the Power of Isolated Populations. Curr Diab Rep 16 (2016) 65.
[118] Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348 (2015) 648-60.

[119] J. Fadista, P. Vikman, E.O. Laakso, I.G. Mollet, J.L. Esguerra, J. Taneera, P. Storm, P. Osmark, C. Ladenvall, R.B. Prasad, K.B. Hansson, F. Finotello, K. Uvebrant, J.K. Ofori, B. Di Camillo, *et al.*, Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism. Proc Natl Acad Sci U S A 111 (2014) 13924-9.

[120] J.R. Kulzer, M.L. Stitzel, M.A. Morken, J.R. Huyghe, C. Fuchsberger, J. Kuusisto, M. Laakso, M. Boehnke, F.S. Collins, and K.L. Mohlke, A common functional regulatory variant at a type 2 diabetes locus upregulates ARAP1 expression in the pancreatic beta cell. Am J Hum Genet 94 (2014) 186-97.

[121] G.R. Carrat, M. Hu, M.S. Nguyen-Tu, P. Chabosseau, K.J. Gaulton, M. van de Bunt, A. Siddiq, M. Falchi, M. Thurner, M. Canouil, F. Pattou, I. Leclerc, T.J. Pullen, M.C. Cane, P. Prabhala, *et al.*, Decreased STARD10 Expression Is Associated with Defective Insulin Secretion in Humans and Mice. Am J Hum Genet 100 (2017) 238-256.

[122] A.L. Williams, S.B. Jacobs, H. Moreno-Macias, A. Huerta-Chagoya, C. Churchhouse, C. Marquez-Luna, H. Garcia-Ortiz, M.J. Gomez-Vazquez, N.P. Burtt, C.A. Aguilar-Salinas, C. Gonzalez-Villalpando, J.C. Florez, L. Orozco, C.A. Haiman, T. Tusie-Luna, *et al.*, Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. Nature 506 (2014) 97-101.

[123] V. Rusu, E. Hoch, J.M. Mercader, D.E. Tenen, M. Gymrek, C.R. Hartigan, M. DeRan, M. von Grotthuss, P. Fontanillas, A. Spooner, G. Guzman, A.A. Deik, K.A. Pierce, C. Dennis, C.B. Clish, *et al.*, Type 2 Diabetes Variants Disrupt Function of SLC16A11 through Two Distinct Mechanisms. Cell 170 (2017) 199-212.e20.

[124] A.L. Price, C.C. Spencer, and P. Donnelly, Progress and promise in understanding the genetic basis of common diseases. Proc Biol Sci 282 (2015) 20151684.

[125] M.R. Nelson, H. Tipney, J.L. Painter, J. Shen, P. Nicoletti, Y. Shen, A. Floratos, P.C. Sham, M.J. Li, J. Wang, L.R. Cardon, J.C. Whittaker, and P. Sanseau, The support of human genetic evidence for approved drug indications. Nat Genet 47 (2015) 856-60.

[126] H. Hauner, The mode of action of thiazolidinediones. Diabetes Metab Res Rev 18 Suppl 2 (2002) S10-5.

[127] D. Altshuler, J.N. Hirschhorn, M. Klannemark, C.M. Lindgren, M.C. Vohl, J. Nemesh, C.R. Lane, S.F. Schaffner, S. Bolk, C. Brewer, T. Tuomi, D. Gaudet, T.J. Hudson, M. Daly, L. Groop, *et al.*, The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. Nat Genet 26 (2000) 76-80.

[128] S.S. Deeb, L. Fajas, M. Nemoto, J. Pihlajamaki, L. Mykkanen, J. Kuusisto, M. Laakso, W. Fujimoto, and J. Auwerx, A Pro12Ala substitution in PPARgamma2 associated with decreased receptor activity, lower body mass index and improved insulin sensitivity. Nat Genet 20 (1998) 284-7.

[129] C.J. Yen, B.A. Beamer, C. Negri, K. Silver, K.A. Brown, D.P. Yarnall, D.K. Burns, J. Roth, and A.R. Shuldiner, Molecular scanning of the human peroxisome proliferator activated receptor gamma (hPPAR gamma) gene in diabetic Caucasians: identification of a Pro12Ala PPAR gamma 2 missense mutation. Biochem Biophys Res Commun 241 (1997) 270-4.

[130] A.R. Majithia, B. Tsuda, M. Agostini, K. Gnanapradeepan, R. Rice, G. Peloso, K.A. Patel, X. Zhang, M.F. Broekema, N. Patterson, M. Duby, T. Sharpe, E. Kalkhoven, E.D. Rosen, I. Barroso, *et al.*, Prospective functional classification of all possible missense variants in PPARG. Nat Genet 48 (2016) 1570-1575.

[131] A.R. Majithia, J. Flannick, P. Shahinian, M. Guo, M.A. Bray, P. Fontanillas, S.B. Gabriel, E.D. Rosen, and D. Altshuler, Rare variants in PPARG with decreased activity in adipocyte differentiation are associated with increased risk of type 2 diabetes. Proc Natl Acad Sci U S A 111 (2014) 13127-32.

[132] A. Mahajan, Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nat Genet 46 (2014) 234-44.

[133] S.K. Thomsen, and A.L. Gloyn, Human genetics as a model for target validation: finding new therapies for diabetes. Diabetologia 60 (2017) 960-970.

[134] C. Ionescu-Tirgoviste, P.A. Gagniuc, E. Gubceac, L. Mardare, I. Popescu, S. Dima, and M. Militaru, A 3D map of the islet routes throughout the healthy human pancreas. Sci Rep 5 (2015) 14634.
[135] A. Pisania, G.C. Weir, J.J. O'Neil, A. Omer, V. Tchipashvili, J. Lei, C.K. Colton, and S. Bonner-Weir, Quantitative analysis of cell composition and purity of human pancreatic islet preparations. Lab Invest 90 (2010) 1661-75.
[136] O. Cabrera, D.M. Berman, N.S. Kenyon, C. Ricordi, P.O. Berggren, and A. Caicedo, The unique cytoarchitecture of human pancreatic islets has implications for islet cell function. Proc Natl Acad Sci U S A 103 (2006) 2334-9.
[137] L.J. McCulloch, M. van de Bunt, M. Braun, K.N. Frayn, A. Clark, and A.L. Gloyn, GLUT2 (SLC2A2) is not the principal glucose transporter in human pancreatic beta cells: implications for understanding genetic association signals at this locus. Mol Genet Metab 104 (2011) 648-53.
[138] A. De Vos, H. Heimberg, E. Quartier, P. Huypens, L. Bouwens, D. Pipeleers, and F. Schuit, Human and rat beta cells differ in glucose transporter but not in glucokinase gene expression. J Clin Invest 96 (1995) 2489-95.
[139] F. Matschinsky, Y. Liang, P. Kesavan, L. Wang, P. Froguel, G. Velho, D. Cohen, M.A. Permutt, Y. Tanizawa, T.L. Jetton, and et al., Glucokinase as pancreatic beta cell glucose sensor and diabetes gene. J Clin Invest 92 (1993) 2092-8.
[140] S.K. Thomsen, and A.L. Gloyn, The pancreatic beta cell: recent insights from human genetics. Trends Endocrinol Metab 25 (2014) 425-34.
[141] C.G. Nichols, KATP channels as molecular sensors of cellular metabolism. Nature 440 (2006) 470-6.
[142] P.E. MacDonald, J.W. Joseph, and P. Rorsman, Glucose-sensing mechanisms in pancreatic beta-cells. Philos Trans R Soc Lond B Biol Sci 360 (2005) 2211-25.
[143] P. Rorsman, M. Braun, and Q. Zhang, Regulation of calcium in pancreatic alpha- and beta-cells in health and disease. Cell Calcium 51 (2012) 300-8.
[144] Z. Wang, and D.C. Thurmond, Mechanisms of biphasic insulin-granule exocytosis - roles of the cytoskeleton, small GTPases and SNARE proteins. J Cell Sci 122 (2009) 893-903.
[145] A.R. Meloni, M.B. DeYoung, C. Lowe, and D.G. Parkes, GLP-1 receptor activated insulin secretion from pancreatic beta-cells: mechanism and glucose dependence. Diabetes Obes Metab 15 (2013) 15-27.
[146] A. Vetere, A. Choudhary, S.M. Burns, and B.K. Wagner, Targeting the pancreatic beta-cell to treat diabetes. Nat Rev Drug Discov 13 (2014) 278-89.
[147] L. Prasad-Reddy, and D. Isaacs, A clinical review of GLP-1 receptor agonists: efficacy and safety in diabetes and beyond. Drugs Context 4 (2015) 212283.
[148] S.G. Swinnen, J.B. Hoekstra, and J.H. DeVries, Insulin therapy for type 2 diabetes. Diabetes Care 32 Suppl 2 (2009) S253-9.
[149] Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). UK Prospective Diabetes Study (UKPDS) Group. Lancet 352 (1998) 837-53.
[150] J.J. Meier, and R.C. Bonadonna, Role of reduced beta-cell mass versus impaired beta-cell function in the pathogenesis of type 2 diabetes. Diabetes Care 36 Suppl 2 (2013) S113-9.
[151] Y. Saisho, A.E. Butler, E. Manesso, D. Elashoff, R.A. Rizza, and P.C. Butler, beta-cell mass and turnover in humans: effects of obesity and aging. Diabetes Care 36 (2013) 111-7.
[152] P. Wang, N.M. Fiaschi-Taesch, R.C. Vasavada, D.K. Scott, A. Garcia-Ocana, and A.F. Stewart, Diabetes mellitus--advances and challenges in human beta-cell proliferation. Nat Rev Endocrinol 11 (2015) 201-12.
[153] J.J. Meier, A.E. Butler, Y. Saisho, T. Monchamp, R. Galasso, A. Bhushan, R.A. Rizza, and P.C. Butler, Beta-cell replication is the primary mechanism subserving the postnatal expansion of beta-cell mass in humans. Diabetes 57 (2008) 1584-94.
[154] D.T. Finegood, L. Scaglia, and S. Bonner-Weir, Dynamics of beta-cell mass in the growing rat pancreas. Estimation with a simple mathematical model. Diabetes 44 (1995) 249-56.

[155] A. Helman, A. Klochendler, N. Azazmeh, Y. Gabai, E. Horwitz, S. Anzi, A. Swisa, R. Condiotti, R.Z. Granit, Y. Nevo, Y. Fixler, D. Shreibman, A. Zamir, S. Tornovsky-Babeay, C. Dai*, et al.*, p16(Ink4a)-induced senescence of pancreatic beta cells enhances insulin secretion. Nat Med 22 (2016) 412-20.
[156] M. Skelin, M. Rupnik, and A. Cencic, Pancreatic beta cell lines and their applications in diabetes mellitus research. Altex 27 (2010) 105-13.
[157] P. Ravassard, Y. Hazhouz, S. Pechberty, E. Bricout-Neveu, M. Armanet, P. Czernichow, and R. Scharfmann, A genetically engineered human pancreatic beta cell line exhibiting glucose-inducible insulin secretion. J Clin Invest 121 (2011) 3589-97.
[158] G. da Silva Xavier, E.A. Bellomo, J.A. McGinty, P.M. French, and G.A. Rutter, Animal models of GWAS-identified type 2 diabetes genes. J Diabetes Res 2013 (2013) 906590.
[159] F.W. Pagliuca, J.R. Millman, M. Gurtler, M. Segel, A. Van Dervort, J.H. Ryu, Q.P. Peterson, D. Greiner, and D.A. Melton, Generation of functional human pancreatic beta cells in vitro. Cell 159 (2014) 428-39.
[160] A. Rezania, J.E. Bruin, P. Arora, A. Rubin, I. Batushansky, A. Asadi, S. O'Dwyer, N. Quiskamp, M. Mojibian, T. Albrecht, Y.H. Yang, and J.D. Johnson, Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. 32 (2014) 1121-33.
[161] G.C. Weir, and S. Bonner-Weir, Finally! A human pancreatic beta cell line. J Clin Invest 121 (2011) 3395-7.
[162] R. Scharfmann, S. Pechberty, Y. Hazhouz, M. von Bulow, E. Bricout-Neveu, M. Grenier-Godard, F. Guez, L. Rachdi, M. Lohmann, P. Czernichow, and P. Ravassard, Development of a conditionally immortalized human pancreatic beta cell line. J Clin Invest 124 (2014) 2087-98.
[163] L.E. Andersson, B. Valtat, A. Bagge, V.V. Sharoyko, D.G. Nicholls, P. Ravassard, R. Scharfmann, P. Spegel, and H. Mulder, Characterization of stimulus-secretion coupling in the human pancreatic EndoC-betaH1 beta cell line. PLoS One 10 (2015) e0120879.
[164] E. Gurgul-Convey, M.T. Kaminski, and S. Lenzen, Physiological characterization of the human EndoC-betaH1 beta-cell line. Biochem Biophys Res Commun 464 (2015) 13-9.
[165] K. Krishnan, Z. Ma, A. Bjorklund, and M.S. Islam, Calcium signaling in a genetically engineered human pancreatic beta-cell line. Pancreas 44 (2015) 773-7.
[166] S.E. Cross, S.J. Hughes, A. Clark, D.W. Gray, and P.R. Johnson, Collagenase does not persist in human islets following isolation. Cell transplantation 21 (2012) 2531-5.
[167] J. Lyon, J.E. Manning Fox, A.F. Spigelman, R. Kim, N. Smith, D. O'Gorman, T. Kin, A.M. Shapiro, R.V. Rajotte, and P.E. MacDonald, Research-Focused Isolation of Human Islets From Donors With and Without Diabetes at the Alberta Diabetes Institute IsletCore. Endocrinology 157 (2016) 560-9.
[168] S. Lamble, E. Batty, M. Attar, D. Buck, R. Bowden, G. Lunter, D. Crook, B. El-Fahmawi, and P. Piazza, Improved workflows for high throughput library preparation using the transposome-based Nextera system. BMC Biotechnol 13 (2013) 104.
[169] D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, and S.L. Salzberg, TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14 (2013) R36.
[170] J. Harrow, A. Frankish, J.M. Gonzalez, E. Tapanari, M. Diekhans, F. Kokocinski, B.L. Aken, D. Barrell, A. Zadissa, S. Searle, I. Barnes, A. Bignell, V. Boychenko, T. Hunt, M. Kay*, et al.*, GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res 22 (2012) 1760-74.
[171] E. Zeggini, M.N. Weedon, C.M. Lindgren, T.M. Frayling, K.S. Elliott, H. Lango, N.J. Timpson, J.R. Perry, N.W. Rayner, R.M. Freathy, J.C. Barrett, B. Shields, A.P. Morris, S. Ellard, C.J. Groves*, et al.*, Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science 316 (2007) 1336-41.
[172] R. Saxena, B.F. Voight, V. Lyssenko, N.P. Burtt, P.I. de Bakker, H. Chen, J.J. Roix, S. Kathiresan, J.N. Hirschhorn, M.J. Daly, T.E. Hughes, L. Groop, D. Altshuler, P. Almgren, J.C. Florez*, et al.*, Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 316 (2007) 1331-6.

[173] N. Grarup, C.S. Rose, E.A. Andersson, G. Andersen, A.L. Nielsen, A. Albrechtsen, J.O. Clausen, S.S. Rasmussen, T. Jorgensen, A. Sandbaek, T. Lauritzen, O. Schmitz, T. Hansen, and O. Pedersen, Studies of association of variants near the HHEX, CDKN2A/B, and IGF2BP2 genes with type 2 diabetes and impaired insulin release in 10,705 Danish subjects: validation and extension of genome-wide association studies. Diabetes 56 (2007) 3105-11.
[174] Y. Kong, R.B. Sharma, B.U. Nwosu, and L.C. Alonso, Islet biology, the CDKN2A/B locus and type 2 diabetes risk. Diabetologia 59 (2016) 1579-93.
[175] S.A. Hannou, K. Wouters, R. Paumelle, and B. Staels, Functional genomics of the CDKN2A/B locus in cardiovascular and metabolic disease: what have we learned from GWASs? Trends Endocrinol Metab 26 (2015) 176-84.
[176] A. Visel, Y. Zhu, D. May, V. Afzal, E. Gong, C. Attanasio, M.J. Blow, J.C. Cohen, E.M. Rubin, and L.A. Pennacchio, Targeted deletion of the 9p21 non-coding coronary artery disease risk interval in mice. Nature 464 (2010) 409-12.
[177] A. Congrains, K. Kamide, M. Ohishi, and H. Rakugi, ANRIL: Molecular Mechanisms and Implications in Human Health. Int J Mol Sci 14 (2013) 1278-1292.
[178] A.S. Dimas, V. Lagou, A. Barker, J.W. Knowles, R. Magi, M.F. Hivert, A. Benazzo, D. Rybin, A.U. Jackson, H.M. Stringham, C. Song, A. Fischer-Rosinsky, T.W. Boesgaard, N. Grarup, F.A. Abbasi*, et al.*, Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. Diabetes (2013) http://dx.doi.org/10.2337/db13-0949.
[179] L.M. t Hart, A.M. Simonis-Bik, G. Nijpels, T.W. van Haeften, S.A. Schafer, J.J. Houwing-Duistermaat, D.I. Boomsma, M.J. Groenewoud, E. Reiling, E.C. van Hove, M. Diamant, M.H. Kramer, R.J. Heine, J.A. Maassen, K. Kirchhoff*, et al.*, Combined risk allele score of eight type 2 diabetes genes is associated with reduced first-phase glucose-stimulated insulin secretion during hyperglycemic clamps. Diabetes 59 (2010) 287-92.
[180] F. Peng, D. Hu, C. Gu, X. Li, Y. Li, N. Jia, S. Chu, J. Lin, and W. Niu, The relationship between five widely-evaluated variants in CDKN2A/B and CDKAL1 genes and the risk of type 2 diabetes: a meta-analysis. Gene 531 (2013) 435-43.
[181] N. Pulizzi, V. Lyssenko, A. Jonsson, C. Osmond, M. Laakso, E. Kajantie, D.J. Barker, L.C. Groop, and J.G. Eriksson, Interaction between prenatal growth and high-risk genotypes in the development of type 2 diabetes. Diabetologia 52 (2009) 825-9.
[182] R.M. Freathy, D.O. Mook-Kanamori, U. Sovio, I. Prokopenko, N.J. Timpson, D.J. Berry, N.M. Warrington, E. Widen, J.J. Hottenga, M. Kaakinen, L.A. Lange, J.P. Bradfield, M. Kerkhof, J.A. Marsh, R. Magi*, et al.*, Variants in ADCY5 and near CCNL1 are associated with fetal growth and birth weight. Nat Genet 42 (2010) 430-5.
[183] M. Horikoshi, R.N. Beaumont, F.R. Day, N.M. Warrington, M.N. Kooijman, J. Fernandez-Tajes, B. Feenstra, N.R. van Zuydam, K.J. Gaulton, N. Grarup, J.P. Bradfield, D.P. Strachan, R. Li-Gao, T.S. Ahluwalia, E. Kreiner*, et al.*, Genome-wide associations for birth weight and correlations with adult disease. Nature 538 (2016) 248-252.
[184] M. Moritani, S. Yamasaki, M. Kagami, T. Suzuki, T. Yamaoka, T. Sano, J. Hata, and M. Itakura, Hypoplasia of endocrine and exocrine pancreas in homozygous transgenic TGF-beta1. Mol Cell Endocrinol 229 (2005) 175-84.
[185] P.A. Svensson, B. Wahlstrand, M. Olsson, P. Froguel, M. Falchi, R.N. Bergman, P.G. McTernan, T. Hedner, L.M. Carlsson, and P. Jacobson, CDKN2B expression and subcutaneous adipose tissue expandability: possible influence of the 9p21 atherosclerosis locus. Biochem Biophys Res Commun 446 (2014) 1126-31.
[186] S.D. Horswell, L.G. Fryer, C.E. Hutchison, D. Zindrou, H.E. Speedy, M.M. Town, E.J. Duncan, R. Sivapackianathan, H.N. Patel, E.L. Jones, A. Braithwaite, M.P. Salm, C.K. Neuwirth, E. Potter, J.R. Anderson*, et al.*, CDKN2B expression in adipose tissue of familial combined hyperlipidemia patients. J Lipid Res 54 (2013) 3491-505.
[187] A. Kamb, N.A. Gruis, J. Weaver-Feldhaus, Q. Liu, K. Harshman, S.V. Tavtigian, E. Stockert, R.S. Day, 3rd, B.E. Johnson, and M.H. Skolnick, A cell cycle regulator potentially involved in genesis of many tumor types. Science 264 (1994) 436-40.

[188] T. Kamijo, F. Zindy, M.F. Roussel, D.E. Quelle, J.R. Downing, R.A. Ashmun, G. Grosveld, and C.J. Sherr, Tumor suppression at the mouse INK4a locus mediated by the alternative reading frame product p19ARF. Cell 91 (1997) 649-59.
[189] M. Serrano, G.J. Hannon, and D. Beach, A new regulatory motif in cell-cycle control causing specific inhibition of cyclin D/CDK4. Nature 366 (1993) 704-7.
[190] H.O. Weber, T. Samuel, P. Rauch, and J.O. Funk, Human p14(ARF)-mediated cell cycle arrest strictly depends on intact p53 signaling pathways. Oncogene 21 (2002) 3207-12.
[191] J. Krishnamurthy, M.R. Ramsey, K.L. Ligon, C. Torrice, A. Koh, S. Bonner-Weir, and N.E. Sharpless, p16INK4a induces an age-dependent decline in islet regenerative potential. Nature 443 (2006) 453-7.
[192] S.G. Rane, P. Dubus, R.V. Mettus, E.J. Galbreath, G. Boden, E.P. Reddy, and M. Barbacid, Loss of Cdk4 expression causes insulin-deficient diabetes and Cdk4 activation results in beta-islet cell hyperplasia. Nat Genet 22 (1999) 44-52.
[193] Y. Lee, J.E. Dominy, Y.J. Choi, M. Jurczak, N. Tolliday, J.P. Camporez, H. Chim, J.H. Lim, H.B. Ruan, X. Yang, F. Vazquez, P. Sicinski, G.I. Shulman, and P. Puigserver, Cyclin D1-Cdk4 controls glucose metabolism independently of cell cycle progression. Nature 510 (2014) 547-51.
[194] K. Bantubungi, S.A. Hannou, S. Caron-Houde, E. Vallez, M. Baron, A. Lucas, E. Bouchaert, R. Paumelle, A. Tailleux, and B. Staels, Cdkn2a/p16Ink4a regulates fasting-induced hepatic gluconeogenesis through the PKA-CREB-PGC1alpha pathway. Diabetes 63 (2014) 3199-209.
[195] J. Martin, S.L. Hunt, P. Dubus, R. Sotillo, F. Nehme-Pelluard, M.A. Magnuson, A.F. Parlow, M. Malumbres, S. Ortega, and M. Barbacid, Genetic rescue of Cdk4 null mice restores pancreatic beta-cell proliferation but not homeostatic cell number. Oncogene 22 (2003) 5261-9.
[196] N. Fiaschi-Taesch, T.A. Bigatel, B. Sicari, K.K. Takane, F. Salim, S. Velazquez-Garcia, G. Harb, K. Selk, I. Cozar-Castellano, and A.F. Stewart, Survey of the human pancreatic beta-cell G1/S proteome reveals a potential therapeutic role for cdk-6 and cyclin D1 in enhancing human beta-cell replication and function in vivo. Diabetes 58 (2009) 882-93.
[197] N.M. Fiaschi-Taesch, F. Salim, J. Kleinberger, R. Troxell, I. Cozar-Castellano, K. Selk, E. Cherok, K.K. Takane, D.K. Scott, and A.F. Stewart, Induction of human beta-cell proliferation and engraftment using a single G1/S regulatory molecule, cdk6. Diabetes 59 (2010) 1926-36.
[198] L.A. Cannon-Albright, D.E. Goldgar, L.J. Meyer, C.M. Lewis, D.E. Anderson, J.W. Fountain, M.E. Hegi, R.W. Wiseman, E.M. Petty, A.E. Bale, and et al., Assignment of a locus for familial melanoma, MLM, to chromosome 9p13-p22. Science 258 (1992) 1148-52.
[199] A.M. Goldstein, Familial melanoma, pancreatic cancer and germline CDKN2A mutations. Hum Mutat 23 (2004) 630.
[200] A. Pal, T.M. Barber, M. Van de Bunt, S.A. Rudge, Q. Zhang, K.L. Lachlan, N.S. Cooper, H. Linden, J.C. Levy, M.J. Wakelam, L. Walker, F. Karpe, and A.L. Gloyn, PTEN mutations as a cause of constitutive insulin sensitivity and obesity. N Engl J Med 367 (2012) 1002-11.
[201] D.R. Matthews, J.P. Hosker, A.S. Rudenski, B.A. Naylor, D.F. Treacher, and R.C. Turner, Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. Diabetologia 28 (1985) 412-9.
[202] T. Hansen, T. Drivsholm, S.A. Urhammer, R.T. Palacios, A. Volund, K. Borch-Johnsen, and O. Pedersen, The BIGTT test: a novel test for simultaneous measurement of pancreatic beta-cell function, insulin sensitivity, and glucose tolerance. Diabetes Care 30 (2007) 257-62.
[203] F. Belfiore, S. Iannello, and G. Volpicelli, Insulin sensitivity indices calculated from basal and OGTT-induced insulin, glucose, and FFA levels. Mol Genet Metab 63 (1998) 134-41.
[204] M. Matsuda, and R.A. DeFronzo, Insulin sensitivity indices obtained from oral glucose tolerance testing: comparison with the euglycemic insulin clamp. Diabetes Care 22 (1999) 1462-70.
[205] G. Pacini, and A. Mari, Methods for clinical assessment of insulin sensitivity and beta-cell function. Best Pract Res Clin Endocrinol Metab 17 (2003) 305-22.
[206] D.I. Phillips, P.M. Clark, C.N. Hales, and C. Osmond, Understanding oral glucose tolerance: comparison of glucose or insulin measurements during the oral glucose tolerance test with specific measurements of insulin resistance and insulin secretion. Diabet Med 11 (1994) 286-92.

[207] S.A. Herzberg-Schafer, H. Staiger, M. Heni, C. Ketterer, M. Guthoff, K. Kantartzis, F. Machicao, N. Stefan, H.U. Haring, and A. Fritsche, Evaluation of fasting state-/oral glucose tolerance test-derived measures of insulin release for the detection of genetically impaired beta-cell function. PLoS One 5 (2010) e14194.
[208] I.F. Godsland, O.F. Agbaje, and R. Hovorka, Evaluation of nonlinear regression approaches to estimation of insulin sensitivity by the minimal model with reference to Bayesian hierarchical analysis. Am J Physiol Endocrinol Metab 291 (2006) E167-74.
[209] R.M. Watanabe, A. Volund, S. Roy, and R.N. Bergman, Prehepatic beta-cell secretion during the intravenous glucose tolerance test in humans: application of a combined model of insulin and C-peptide kinetics. J Clin Endocrinol Metab 69 (1989) 790-7.
[210] N.A. Gruis, P.A. van der Velden, L.A. Sandkuijl, D.E. Prins, J. Weaver-Feldhaus, A. Kamb, W. Bergman, and R.R. Frants, Homozygotes for CDKN2 (p16) germline mutation in Dutch familial melanoma kindreds. Nat Genet 10 (1995) 351-3.
[211] R.R. McWilliams, E.D. Wieben, K.G. Rabe, K.S. Pedersen, Y. Wu, H. Sicotte, and G.M. Petersen, Prevalence of CDKN2A mutations in pancreatic cancer patients: implications for genetic counseling. Eur J Hum Genet 19 (2011) 472-8.
[212] A.M. Goldstein, S.N. Stacey, J.H. Olafsson, G.F. Jonsson, A. Helgason, P. Sulem, B. Sigurgeirsson, K.R. Benediktsdottir, K. Thorisdottir, R. Ragnarsson, J. Kjartansson, J. Kostic, G. Masson, K. Kristjansson, J.R. Gulcher*, et al.*, CDKN2A mutations and melanoma risk in the Icelandic population. J Med Genet 45 (2008) 284-9.
[213] M. Harland, C.F. Taylor, S. Bass, M. Churchman, J.A. Randerson-Moor, E.A. Holland, G.J. Mann, D.T. Bishop, and J.A. Newton Bishop, Intronic sequence variants of the CDKN2A gene in melanoma pedigrees. Genes Chromosomes Cancer 43 (2005) 128-36.
[214] M. Harland, S. Mistry, D.T. Bishop, and J.A. Bishop, A deep intronic mutation in CDKN2A is associated with disease in a subset of melanoma pedigrees. Hum Mol Genet 10 (2001) 2679-86.
[215] A.M. Goldstein, M. Chan, M. Harland, N.K. Hayward, F. Demenais, D.T. Bishop, E. Azizi, W. Bergman, G. Bianchi-Scarra, W. Bruno, D. Calista, L.A. Albright, V. Chaudru, A. Chompret, F. Cuellar*, et al.*, Features associated with germline CDKN2A mutations: a GenoMEL study of melanoma-prone families from three continents. J Med Genet 44 (2007) 99-106.
[216] N. Soufir, M.F. Avril, A. Chompret, F. Demenais, J. Bombled, A. Spatz, D. Stoppa-Lyonnet, J. Benard, and B. Bressac-de Paillerets, Prevalence of p16 and CDK4 germline mutations in 48 melanoma-prone families in France. The French Familial Melanoma Study Group. Hum Mol Genet 7 (1998) 209-16.
[217] I. Orlow, C.B. Begg, J. Cotignola, P. Roy, A.J. Hummer, B.A. Clas, U. Mujumdar, R. Canchola, B.K. Armstrong, A. Kricker, L.D. Marrett, R.C. Millikan, S.B. Gruber, H. Anton-Culver, R. Zanetti*, et al.*, CDKN2A germline mutations in individuals with cutaneous malignant melanoma. J Invest Dermatol 127 (2007) 1234-43.
[218] C.B. Begg, I. Orlow, A.J. Hummer, B.K. Armstrong, A. Kricker, L.D. Marrett, R.C. Millikan, S.B. Gruber, H. Anton-Culver, R. Zanetti, R.P. Gallagher, T. Dwyer, T.R. Rebbeck, N. Mitra, K. Busam*, et al.*, Lifetime risk of melanoma in CDKN2A mutation carriers in a population-based sample. J Natl Cancer Inst 97 (2005) 1507-15.
[219] D.K. Bartsch, M. Sina-Frey, S. Lang, A. Wild, B. Gerdes, P. Barth, R. Kress, R. Grutzmann, M. Colombo-Benkmann, A. Ziegler, S.A. Hahn, M. Rothmund, and H. Rieder, CDKN2A germline mutations in familial pancreatic cancer. Ann Surg 236 (2002) 730-7.
[220] A.M. Goldstein, J.P. Struewing, A. Chidambaram, M.C. Fraser, and M.A. Tucker, Genotype-phenotype relationships in U.S. melanoma-prone families with CDKN2A and CDK4 mutations. J Natl Cancer Inst 92 (2000) 1006-10.
[221] J.A. de Wilde, S. van Buuren, and B.J. Middelkoop, Trends in birth weight and the prevalence of low birth weight and small-for-gestational-age in Surinamese South Asian babies since 1974: cross-sectional study of three birth cohorts. BMC Public Health 13 (2013) 931.
[222] M. Campioni, G. Toffolo, R. Basu, R.A. Rizza, and C. Cobelli, Minimal model assessment of hepatic insulin extraction during an oral test from standard insulin kinetic parameters. Am J Physiol Endocrinol Metab 297 (2009) E941-8.

[223] D.L. Horwitz, J.I. Starr, M.E. Mako, W.G. Blackard, and A.H. Rubenstein, Proinsulin, insulin, and C-peptide concentrations in human portal and peripheral blood. J Clin Invest 55 (1975) 1278-83.
[224] H.A. McKenzie, C. Fung, T.M. Becker, M. Irvine, G.J. Mann, R.F. Kefford, and H. Rizos, Predicting functional significance of cancer-associated p16(INK4a) mutations in CDKN2A. Hum Mutat 31 (2010) 692-701.
[225] R. Keller-Melchior, R. Schmidt, and M. Piepkorn, Expression of the Tumor Suppressor Gene Product p16INK4 in Benign and Malignant Melanocytic Lesions. 110 (1998) 932-938.
[226] M.H. Shanik, Y. Xu, J. Skrha, R. Dankner, Y. Zick, and J. Roth, Insulin resistance and hyperinsulinemia: is hyperinsulinemia the cart or the horse? Diabetes Care 31 Suppl 2 (2008) S262-8.
[227] S.S. Torekov, E. Iepsen, M. Christiansen, A. Linneberg, O. Pedersen, J.J. Holst, J.K. Kanters, and T. Hansen, KCNQ1 Long QT Syndrome Patients Have Hyperinsulinemia and Symptomatic Hypoglycemia. Diabetes 63 (2014) 1315-25.
[228] H.S. Seltzer, E.W. Allen, A.L. Herron, Jr., and M.T. Brennan, Insulin secretion in response to glycemic stimulus: relation of delayed initial release to carbohydrate intolerance in mild diabetes mellitus. J Clin Invest 46 (1967) 323-35.
[229] E. Cersosimo, C. Solis-Herrera, M.E. Trautmann, J. Malloy, and C.L. Triplitt, Assessment of pancreatic beta-cell function: review of methods and clinical applications. Curr Diabetes Rev 10 (2014) 2-42.
[230] S.L. Milgram, R.E. Mains, and B.A. Eipper, Identification of routing determinants in the cytosolic domain of a secretory granule-associated integral membrane protein. The Journal of biological chemistry 271 (1996) 17526-35.
[231] A.M. Oyarce, and B.A. Eipper, Neurosecretory vesicles contain soluble and membrane-associated monofunctional and bifunctional peptidylglycine alpha-amidating monooxygenase proteins. Journal of neurochemistry 60 (1993) 1105-14.
[232] B.A. Eipper, S.L. Milgram, E.J. Husten, H.Y. Yun, and R.E. Mains, Peptidylglycine alpha-amidating monooxygenase: a multifunctional protein with catalytic, processing, and routing domains. Protein Sci 2 (1993) 489-97.
[233] D.J. Merkler, C-terminal amidated peptides: production by the in vitro enzymatic amidation of glycine-extended peptides and the importance of the amide to bioactivity. Enzyme and microbial technology 16 (1994) 450-6.
[234] T.A. Czyzyk, Y. Ning, M.S. Hsu, B. Peng, R.E. Mains, B.A. Eipper, and J.E. Pintar, Deletion of peptide amidation enzymatic activity leads to edema and embryonic lethality in the mouse. Developmental biology 287 (2005) 301-13.
[235] O. Garmendia, M.P. Rodriguez, M.A. Burrell, and A.C. Villaro, Immunocytochemical finding of the amidating enzymes in mouse pancreatic A-, B-, and D-cells: a comparison with human and rat. The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society 50 (2002) 1401-16.
[236] B.A. Eipper, C.B. Green, T.A. Campbell, D.A. Stoffers, H.T. Keutmann, R.E. Mains, and L. Ouafik, Alternative splicing and endoproteolytic processing generate tissue-specific forms of pituitary peptidylglycine alpha-amidating monooxygenase (PAM). The Journal of biological chemistry 267 (1992) 4008-15.
[237] M. De, J. Bell, N.J. Blackburn, R.E. Mains, and B.A. Eipper, Role for an essential tyrosine in peptide amidation. The Journal of biological chemistry 281 (2006) 20873-82.
[238] E. Maeda-Nakai, and A. Ichiyama, A spectrophotometric method for the determination of glycolate in urine and plasma with glycolate oxidase. Journal of biochemistry 127 (2000) 279-87.
[239] S.E. Carpenter, and D.J. Merkler, An enzyme-coupled assay for glyoxylic acid. Analytical biochemistry 323 (2003) 242-6.
[240] V. Chandra, O. Albagli-Curiel, B. Hastoy, J. Piccand, C. Randriamampita, E. Vaillant, H. Cave, K. Busiah, P. Froguel, M. Vaxillaire, P. Rorsman, M. Polak, and R. Scharfmann, RFX6 regulates insulin secretion by modulating Ca2+ homeostasis in human beta cells. Cell reports 9 (2014) 2206-18.

[241] P.D. Simpson, B.A. Eipper, M.J. Katz, L. Gandara, P. Wappner, R. Fischer, E.J. Hodson, P.J. Ratcliffe, and N. Masson, Striking Oxygen Sensitivity of the Peptidylglycine alpha-Amidating Monooxygenase (PAM) in Neuroendocrine Cells. J Biol Chem 290 (2015) 24891-901.
[242] P.D. Simpson, B.A. Eipper, M.J. Katz, L. Gandara, P. Wappner, R. Fischer, E.J. Hodson, P.J. Ratcliffe, and N. Masson, Striking Oxygen Sensitivity of the Peptidylglycine alpha-Amidating Monooxygenase (PAM) in Neuroendocrine Cells. The Journal of biological chemistry (2015).
[243] D.M. Blodgett, A. Nowosielska, S. Afik, S. Pechhold, A.J. Cura, N.J. Kennedy, S. Kim, A. Kucukural, R.J. Davis, S.C. Kent, D.L. Greiner, M.G. Garber, D.M. Harlan, and P. diIorio, Novel Observations From Next-Generation RNA Sequencing of Highly Purified Human Adult and Fetal Islet Cell Subsets. Diabetes 64 (2015) 3172-81.
[244] E. Gurgul-Convey, M.T. Kaminski, and S. Lenzen, Physiological characterization of the human EndoC-betaH1 beta-cell line. Biochemical and biophysical research communications 464 (2015) 13-9.
[245] R. El Meskini, R.E. Mains, and B.A. Eipper, Cell type-specific metabolism of peptidylglycine alpha-amidating monooxygenase in anterior pituitary. Endocrinology 141 (2000) 3020-34.
[246] S.L. Milgram, S.T. Kho, G.V. Martin, R.E. Mains, and B.A. Eipper, Localization of integral membrane peptidylglycine alpha-amidating monooxygenase in neuroendocrine cells. Journal of cell science 110 (Pt 6) (1997) 695-706.
[247] K.D. Gillis, R. Mossner, and E. Neher, Protein kinase C enhances exocytosis from chromaffin cells by increasing the size of the readily releasable pool of secretory granules. Neuron 16 (1996) 1209-20.
[248] Y.H. Yang, M. Szabat, C. Bragagnini, K. Kott, C.D. Helgason, B.G. Hoffman, and J.D. Johnson, Paracrine signalling loops in adult human and mouse pancreatic islets: netrins modulate beta cell apoptosis signalling via dependence receptors. Diabetologia 54 (2011) 828-42.
[249] T. Kim, J.H. Tao-Cheng, L.E. Eiden, and Y.P. Loh, Chromogranin A, an "on/off" switch controlling dense-core secretory granule biogenesis. Cell 106 (2001) 499-509.
[250] V. Colomer, G.A. Kicska, and M.J. Rindler, Secretory granule content proteins and the luminal domains of granule membrane proteins aggregate in vitro at mildly acidic pH. The Journal of biological chemistry 271 (1996) 48-55.
[251] S.H. Yoo, and J.P. Albanesi, Ca2(+)-induced conformational change and aggregation of chromogranin A. The Journal of biological chemistry 265 (1990) 14414-21.
[252] K.B. Helle, A. Corti, M.H. Metz-Boutigue, and B. Tota, The endocrine role for chromogranin A: a prohormone for peptides with regulatory properties. Cellular and molecular life sciences : CMLS 64 (2007) 2863-86.
[253] A. Clark, and M.R. Nilsson, Islet amyloid: a complication of islet dysfunction or an aetiological factor in Type 2 diabetes? Diabetologia 47 (2004) 157-69.
[254] A.G. Katopodis, and S.W. May, Novel substrates and inhibitors of peptidylglycine alpha-amidating monooxygenase. Biochemistry 29 (1990) 4541-8.
[255] C. Benner, T. van der Meulen, E. Caceres, K. Tigyi, C.J. Donaldson, and M.O. Huising, 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 (2014) 620.
[256] M. Najvirtova, Z. Bacova, A. Mateffyova, and V. Strbak, A role of thyrotropin-releasing hormone in insulin secretion by isolated rat pancreatic islets. Pflugers Arch 449 (2005) 547-52.
[257] R.N. Kulkarni, Z.L. Wang, K.O. Akinsanya, W.M. Bennet, R.M. Wang, D.M. Smith, M.A. Ghatei, P.G. Byfield, and S.R. Bloom, Pyroglutamyl-phenylalanyl-proline amide attenuates thyrotropin-releasing hormone-stimulated insulin secretion in perifused rat islets and insulin-secreting clonal beta-cell lines. Endocrinology 136 (1995) 5155-64.
[258] C. Orskov, A. Wettergren, and J.J. Holst, Biological effects and metabolic rates of glucagonlike peptide-1 7-36 amide and glucagonlike peptide-1 7-37 in healthy subjects are indistinguishable. Diabetes 42 (1993) 658-61.
[259] T.P. Vahl, B.W. Paty, B.D. Fuller, R.L. Prigeon, and D.A. D'Alessio, Effects of GLP-1-(7-36)NH2, GLP-1-(7-37), and GLP-1- (9-36)NH2 on intravenous glucose tolerance and glucose-induced insulin secretion in healthy humans. J Clin Endocrinol Metab 88 (2003) 1772-9.

[260] J.C. Ebiou, D. Grouselle, and S. Aratan-Spire, Antithyrotropin-releasing hormone serum inhibits secretion of glucagon from isolated perfused rat pancreas: an experimental model for positive feedback regulation of glucagon secretion. Endocrinology 131 (1992) 765-71.
[261] J.R. Huyghe, A.U. Jackson, M.P. Fogarty, M.L. Buchkovich, A. Stancakova, H.M. Stringham, X. Sim, L. Yang, C. Fuchsberger, H. Cederberg, P.S. Chines, T.M. Teslovich, J.M. Romm, H. Ling, I. McMullen*, et al.*, Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. Nature genetics 45 (2013) 197-201.
[262] V. Steinthorsdottir, G. Thorleifsson, P. Sulem, H. Helgason, N. Grarup, A. Sigurdsson, H.T. Helgadottir, H. Johannsdottir, O.T. Magnusson, S.A. Gudjonsson, J.M. Justesen, M.N. Harder, M.E. Jorgensen, C. Christensen, I. Brandslund*, et al.*, Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. Nature genetics 46 (2014) 294-8.
[263] F.A. Tausk, S.L. Milgram, R.E. Mains, and B.A. Eipper, Expression of a peptide processing enzyme in cultured cells: truncation mutants reveal a routing domain. Molecular endocrinology 6 (1992) 2185-96.
[264] S.H. Yoo, and M.S. Lewis, Dimerization and tetramerization properties of the C-terminal region of chromogranin A: a thermodynamic analysis. Biochemistry 32 (1993) 8816-22.
[265] C.A. Mosley, L. Taupenot, N. Biswas, J.P. Taulane, N.H. Olson, S.M. Vaingankar, G. Wen, N.J. Schork, M.G. Ziegler, S.K. Mahata, and D.T. O'Connor, Biogenesis of the secretory granule: chromogranin A coiled-coil structure results in unusual physical properties and suggests a mechanism for granule core condensation. Biochemistry 46 (2007) 10999-1012.
[266] G.K. Bandyopadhyay, and S.K. Mahata, Chromogranin A Regulation of Obesity and Peripheral Insulin Sensitivity. Front Endocrinol (Lausanne) 8 (2017) 20.
[267] J. Wollam, S. Mahata, M. Riopel, A. Hernandez-Carretero, A. Biswas, G.K. Bandyopadhyay, N.W. Chi, L.E. Eiden, N.R. Mahapatra, A. Corti, N.J.G. Webster, and S.K. Mahata, Chromogranin A regulates vesicle storage and mitochondrial dynamics to influence insulin secretion. Cell Tissue Res 368 (2017) 487-501.
[268] A. Bartolomucci, R. Possenti, S.K. Mahata, R. Fischer-Colbrie, Y.P. Loh, and S.R. Salton, The extended granin family: structure, function, and biomedical implications. Endocr Rev 32 (2011) 755-97.
[269] R.P. Nanga, J.R. Brender, S. Vivekanandan, and A. Ramamoorthy, Structure and membrane orientation of IAPP in its natively amidated form at physiological pH in a membrane environment. Biochimica et biophysica acta 1808 (2011) 2337-42.
[270] D.E. Shalev, A. Mor, and I. Kustanovich, Structural consequences of carboxyamidation of dermaseptin S3. Biochemistry 41 (2002) 7312-7.
[271] M.L. Sforca, S. Oyama, Jr., F. Canduri, C.C. Lorenzi, T.A. Pertinhez, K. Konno, B.M. Souza, M.S. Palma, J. Ruggiero Neto, W.F. Azevedo, Jr., and A. Spisni, How C-terminal carboxyamidation alters the biological activity of peptides from the venom of the eumenine solitary wasp. Biochemistry 43 (2004) 5608-17.
[272] T. Kim, and Y.P. Loh, Protease nexin-1 promotes secretory granule biogenesis by preventing granule protein degradation. Molecular biology of the cell 17 (2006) 789-98.
[273] H. Koshimizu, N.X. Cawley, T. Kim, A.L. Yergey, and Y.P. Loh, Serpinin: a novel chromogranin A-derived, secreted peptide up-regulates protease nexin-1 expression and granule biogenesis in endocrine cells. Molecular endocrinology 25 (2011) 732-44.
[274] A. Wettergren, L. Pridal, M. Wojdemann, and J.J. Holst, Amidated and non-amidated glucagon-like peptide-1 (GLP-1): non-pancreatic effects (cephalic phase acid secretion) and stability in plasma in humans. Regul Pept 77 (1998) 83-7.
[275] K.W. Schmid, R. Weiler, R.W. Xu, R. Hogue-Angeletti, R. Fischer-Colbrie, and H. Winkler, An immunological study on chromogranin A and B in human endocrine and nervous tissues. Histochem J 21 (1989) 365-73.
[276] G.M. Portela-Gomes, M. Stridsberg, H. Johansson, and L. Grimelius, Complex co-localization of chromogranins and neurohormones in the human gastrointestinal tract. J Histochem Cytochem 45 (1997) 815-22.

[277] A.K. Grotz, A.L. Gloyn, and S.K. Thomsen, Prioritising Causal Genes at Type 2 Diabetes Risk Loci. Curr Diab Rep 17 (2017) 76.
[278] A. Varshney, L.J. Scott, R.P. Welch, M.R. Erdos, P.S. Chines, N. Narisu, R.D. Albanus, P. Orchard, B.N. Wolford, R. Kursawe, S. Vadlamudi, M.E. Cannon, J.P. Didion, J. Hensley, A. Kirilusha*, et al.*, Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. Proc Natl Acad Sci U S A 114 (2017) 2301-2306.
[279] J. Taneera, S. Lang, A. Sharma, J. Fadista, Y. Zhou, E. Ahlqvist, A. Jonsson, V. Lyssenko, P. Vikman, O. Hansson, H. Parikh, O. Korsgren, A. Soni, U. Krus, E. Zhang*, et al.*, A systems genetics approach identifies genes and pathways for type 2 diabetes in human islets. Cell Metab 16 (2012) 122-34.
[280] W. Wu, J. Shang, Y. Feng, C.M. Thompson, S. Horwitz, J.R. Thompson, E.D. MacIntyre, N.A. Thornberry, K. Chapman, Y.-P. Zhou, A.D. Howard, and J. Li, Identification of glucose-dependant insulin secretion targets in pancreatic beta cells by combining defined-mechanism compound library screening and siRNA gene silencing. Journal of Biomolecular Screening 13 (2008) 128-134.
[281] J.A. Lee, S. Chu, F.S. Willard, K.L. Cox, R.J.S. Galvin, R.B. Peery, S.E. Oliver, J. Oler, T.D. Meredith, S.A. Heidler, W.H. Gough, S. Husain, A.D. Palkowitz, and C.M. Moxham, Open Innovation for Phenotypic Drug Discovery: The PD2 Assay Panel. Journal of Biomolecular Screening 16 (2011) 588-602.
[282] S.M. Burns, A. Vetere, D. Walpita, V. Dancik, C. Khodier, J. Perez, P.A. Clemons, B.K. Wagner, and D. Altshuler, High-Throughput Luminescent Reporter of Insulin Secretion for Discovering Regulators of Pancreatic Beta-Cell Function. Cell Metabolism 21 (2015) 126-137.
[283] G.M. Ku, Z. Pappalardo, C.C. Luo, M.S. German, and M.T. McManus, An siRNA Screen in Pancreatic Beta Cells Reveals a Role for Gpr27 in Insulin Production. PLoS Genet 8 (2012).
[284] Z. Pappalardo, D. Gambhir Chopra, T.G. Hennings, H. Richards, J. Choe, K. Yang, L. Baeyens, K. Ang, S. Chen, M. Arkin, M.S. German, M.T. McManus, and G.M. Ku, A Whole-Genome RNA Interference Screen Reveals a Role for Spry2 in Insulin Transcription and the Unfolded Protein Response. Diabetes 66 (2017) 1703-1712.
[285] D. Walpita, T. Hasaka, J. Spoonamore, A. Vetere, K.K. Takane, D. Fomina-Yadlin, N. Fiaschi-Taesch, A. Shamji, P.A. Clemons, A.F. Stewart, S.L. Schreiber, and B.K. Wagner, A Human Islet Cell Culture System for High-Throughput Screening. Journal of Biomolecular Screening 17 (2012) 509-518.
[286] I. Akerman, Z. Tu, A. Beucher, D.M. Rolando, C. Sauty-Colace, M. Benazra, N. Nakic, J. Yang, H. Wang, L. Pasquali, I. Moran, J. Garcia-Hurtado, N. Castro, R. Gonzalez-Franco, A.F. Stewart*, et al.*, Human Pancreatic beta Cell lncRNAs Control Cell-Specific Regulatory Networks. Cell Metab 25 (2017) 400-411.
[287] O. Shalem, N.E. Sanjana, and F. Zhang, High-throughput functional genomics using CRISPR-Cas9. Nat Rev Genet 16 (2015) 299-311.
[288] D.S. DeLuca, J.Z. Levin, A. Sivachenko, T. Fennell, M.D. Nazaire, C. Williams, M. Reich, W. Winckler, and G. Getz, RNA-SeQC: RNA-seq metrics for quality control and process optimization. Bioinformatics 28 (2012) 1530-2.
[289] C.W. Law, Y. Chen, W. Shi, and G.K. Smyth, voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol 15 (2014) R29.
[290] M.E. Ritchie, B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, and G.K. Smyth, limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43 (2015) e47.
[291] A. Pal, T.P. Potjer, S.K. Thomsen, H.J. Ng, A. Barrett, R. Scharfmann, T.J. James, D.T. Bishop, F. Karpe, I.F. Godsland, H.F. Vasen, J. Newton-Bishop, H. Pijl, M.I. McCarthy, and A.L. Gloyn, Loss-of-Function Mutations in the Cell-Cycle Control Gene CDKN2A Impact on Glucose Homeostasis in Humans. Diabetes 65 (2016) 527-33.
[292] T. Meissner, B. Beinbrech, and E. Mayatepek, Congenital hyperinsulinism: molecular basis of a heterogeneous disease. Hum Mutat 13 (1999) 351-61.
[293] T.J. McDonald, and S. Ellard, Maturity onset diabetes of the young: identification and diagnosis. Ann Clin Biochem 50 (2013) 403-15.

[294] B.E. Stranger, A.C. Nica, M.S. Forrest, A. Dimas, C.P. Bird, C. Beazley, C.E. Ingle, M. Dunning, P. Flicek, D. Koller, S. Montgomery, S. Tavare, P. Deloukas, and E.T. Dermitzakis, Population genomics of human gene expression. Nat Genet 39 (2007) 1217-24.
[295] A.S. Dimas, S. Deutsch, B.E. Stranger, S.B. Montgomery, C. Borel, H. Attar-Cohen, C. Ingle, C. Beazley, M. Gutierrez Arcelus, M. Sekowska, M. Gagnebin, J. Nisbett, P. Deloukas, E.T. Dermitzakis, and S.E. Antonarakis, Common regulatory variation impacts gene expression in a cell type-dependent manner. Science 325 (2009) 1246-50.
[296] J.B. Veyrieras, S. Kudaravalli, S.Y. Kim, E.T. Dermitzakis, Y. Gilad, M. Stephens, and J.K. Pritchard, High-resolution mapping of expression-QTLs yields insight into human gene regulation. PLoS Genet 4 (2008) e1000214.
[297] T.C. Nogueira, F.M. Paula, O. Villate, M.L. Colli, R.F. Moura, D.A. Cunha, L. Marselli, P. Marchetti, M. Cnop, C. Julier, and D.L. Eizirik, GLIS3, a susceptibility gene for type 1 and type 2 diabetes, modulates pancreatic beta cell apoptosis via regulation of a splice variant of the BH3-only protein Bim. PLoS Genet 9 (2013) e1003532.
[298] J. Dooley, L. Tian, S. Schonefeldt, V. Delghingaro-Augusto, J.E. Garcia-Perez, E. Pasciuto, D. Di Marino, E.J. Carr, N. Oskolkov, V. Lyssenko, D. Franckaert, V. Lagou, L. Overbergh, J. Vandenbussche, J. Allemeersch*, et al.*, Genetic predisposition for beta cell fragility underlies type 1 and type 2 diabetes. Nat Genet 48 (2016) 519-27.
[299] X. Li, G. Thyssen, J. Beliakoff, and Z. Sun, The novel PIAS-like protein hZimp10 enhances Smad transcriptional activity. J Biol Chem 281 (2006) 23748-56.
[300] L.A. Rakowski, D.D. Garagiola, C.M. Li, M. Decker, S. Caruso, M. Jones, R. Kuick, T. Cierpicki, I. Maillard, and M.Y. Chiang, Convergence of the ZMIZ1 and NOTCH1 pathways at C-MYC in acute T lymphoblastic leukemias. Cancer Res 73 (2013) 930-41.
[301] N.L. Fewings, P.N. Gatt, F.C. McKay, G.P. Parnell, S.D. Schibeci, J. Edwards, M.A. Basuki, A. Goldinger, M.J. Fabis-Pedrini, A.G. Kermode, C.P. Manrique, J.L. McCauley, D. Nickles, S.E. Baranzini, T. Burke*, et al.*, The autoimmune risk gene ZMIZ1 is a vitamin D responsive marker of a molecular phenotype of multiple sclerosis. J Autoimmun 78 (2017) 57-69.
[302] G. Soler, I. Radford-Weiss, R. Ben-Abdelali, N. Mahlaoui, J.F. Ponceau, E.A. Macintyre, M. Vekemans, O.A. Bernard, and S.P. Romana, Fusion of ZMIZ1 to ABL1 in a B-cell acute lymphoblastic leukaemia with a t(9;10)(q34;q22.3) translocation. Leukemia 22 (2008) 1278-80.
[303] B.M. Shields, S. Hicks, M.H. Shepherd, K. Colclough, A.T. Hattersley, and S. Ellard, Maturity-onset diabetes of the young (MODY): how many cases are we missing? Diabetologia 53 (2010) 2504-2508.
[304] E.R. Pearson, S.F. Boj, A.M. Steele, T. Barrett, K. Stals, J.P. Shield, S. Ellard, J. Ferrer, and A.T. Hattersley, Macrosomia and hyperinsulinaemic hypoglycaemia in patients with heterozygous mutations in the HNF4A gene. Plos Medicine 4 (2007) 760-769.
[305] B. Glaser, Type 2 diabetes: Hypoinsulinism, hyperinsulinism, or both? Plos Medicine 4 (2007) 619-620.
[306] S.A. Rahman, A. Nessa, and K. Hussain, Molecular mechanisms of congenital hyperinsulinism. Journal of Molecular Endocrinology 54 (2015) R119-R129.
[307] L. Pasquali, K.J. Gaulton, S.A. Rodriguez-Segui, L. Mularoni, I. Miguel-Escalada, I. Akerman, J.J. Tena, I. Moran, C. Gomez-Marin, M. van de Bunt, J. Ponsa-Cobas, N. Castro, T. Nammo, I. Cebola, J. Garcia-Hurtado*, et al.*, Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. Nat Genet 46 (2014) 136-43.
[308] A.M. Simonis-Bik, G. Nijpels, T.W. van Haeften, J.J. Houwing-Duistermaat, D.I. Boomsma, E. Reiling, E.C. van Hove, M. Diamant, M.H.H. Kramer, R.J. Heine, J.A. Maassen, P.E. Slagboom, G. Willemsen, J.M. Dekker, E.M. Eekhoff*, et al.*, Gene Variants in the Novel Type 2 Diabetes Loci CDC123/CAMK1D, THADA, ADAMTS9, BCL11A, and MTNR1B Affect Different Aspects of Pancreatic beta-Cell Function. Diabetes 59 (2010) 293-301.
[309] H. Parikh, V. Lyssenko, and L. Groop, Prioritizing genes for follow-up from genome wide association studies using information on gene expression in tissues relevant for type 2 diabetes mellitus. BMC Medical Genomics 2 (2009) 72.
[310] O. Ali, Genetics of type 2 diabetes. World journal of diabetes 4 (2013) 114-23.

[311] A. Bonnefond, and P. Froguel, Rare and Common Genetic Events in Type 2 Diabetes: What Should Biologists Know? Cell metabolism 21 (2015) 357-368.
[312] L.J. Corbin, V.Y. Tan, D.A. Hughes, K.H. Wade, D.S. Paul, K.E. Tansey, F. Butcher, F. Dudbridge, J.M. Howson, M.W. Jallow, C. John, N. Kingston, C.M. Lindgren, M. O'Donavan, S. O'Rahilly, *et al.*, Causal Analyses, Statistical Efficiency And Phenotypic Precision Through Recall-By-Genotype Study Design. bioRxiv (2017).
[313] G.D. Tan, M.J. Neville, E. Liverani, S.M. Humphreys, J.M. Currie, L. Dennis, B.A. Fielding, and F. Karpe, The in vivo effects of the Pro12Ala PPARgamma2 polymorphism on adipose tissue NEFA metabolism: the first use of the Oxford Biobank. Diabetologia 49 (2006) 158-68.
[314] M. Shah, R.T. Varghese, J.M. Miles, F. Piccinini, C. Dalla Man, C. Cobelli, K.R. Bailey, R.A. Rizza, and A. Vella, TCF7L2 Genotype and alpha-Cell Function in Humans Without Diabetes. Diabetes 65 (2016) 371-80.
[315] R.T. Varghese, I. Viegas, C. Barosa, C. Marques, M. Shah, R.A. Rizza, J.G. Jones, and A. Vella, Diabetes-Associated Variation in TCF7L2 Is Not Associated With Hepatic or Extrahepatic Insulin Resistance. Diabetes 65 (2016) 887-92.
[316] T. Tuomi, C.L. Nagorny, P. Singh, H. Bennet, Q. Yu, I. Alenkvist, B. Isomaa, B. Ostman, J. Soderstrom, A.K. Pesonen, S. Martikainen, K. Raikkonen, T. Forsen, L. Hakaste, P. Almgren, *et al.*, Increased Melatonin Signaling Is a Risk Factor for Type 2 Diabetes. Cell Metab 23 (2016) 1067-77.
[317] Y. Tang, A.S. Axelsson, P. Spegel, L.E. Andersson, H. Mulder, L.C. Groop, E. Renstrom, and A.H. Rosengren, Genotype-based treatment of type 2 diabetes with an alpha2A-adrenergic receptor antagonist. Sci Transl Med 6 (2014) 257ra139.
[318] R.P. Robertson, L.D. Bogachus, E. Oseid, S. Parazzoli, M.E. Patti, M.R. Rickels, C. Schuetz, T. Dunn, T. Pruett, A.N. Balamurugan, D.E. Sutherland, G. Beilman, and M.D. Bellin, Assessment of beta-cell mass and alpha- and beta-cell survival and function by arginine stimulation in human autologous islet recipients. Diabetes 64 (2015) 565-72.
[319] V. Steinthorsdottir, G. Thorleifsson, and P. Sulem, Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. 46 (2014) 294-8.
[320] T2D-GENES Consortium, GoT2D Consortium, DIAGRAM Consortium, http://www.type2diabetesgenetics.org/; Accessed 08/09/17.
[321] P. Datlinger, A.F. Rendeiro, C. Schmidl, T. Krausgruber, P. Traxler, J. Klughammer, L.C. Schuster, A. Kuchler, D. Alpar, and C. Bock, Pooled CRISPR screening with single-cell transcriptome readout. Nat Methods 14 (2017) 297-301.
[322] D.A. Jaitin, A. Weiner, I. Yofe, D. Lara-Astiaso, H. Keren-Shaul, E. David, T.M. Salame, A. Tanay, A. van Oudenaarden, and I. Amit, Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. Cell 167 (2016) 1883-1896.e15.
[323] B. Adamson, T.M. Norman, M. Jost, M.Y. Cho, J.K. Nunez, Y. Chen, J.E. Villalta, L.A. Gilbert, M.A. Horlbeck, M.Y. Hein, R.A. Pak, A.N. Gray, C.A. Gross, A. Dixit, O. Parnas, *et al.*, A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. Cell 167 (2016) 1867-1882.e21.
[324] A. Dixit, O. Parnas, B. Li, J. Chen, C.P. Fulco, L. Jerby-Arnon, N.D. Marjanovic, D. Dionne, T. Burks, R. Raychowdhury, B. Adamson, T.M. Norman, E.S. Lander, J.S. Weissman, N. Friedman, *et al.*, Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. Cell 167 (2016) 1853-1866.e17.
[325] N.L. Beer, and A.L. Gloyn, Genome-edited human stem cell-derived beta cells: a powerful tool for drilling down on type 2 diabetes GWAS biology. F1000Res 5 (2016).

Published Manuscripts

Manuscripts of relevance to work discussed in this thesis are included in the following order:

1. **S.K. Thomsen**, and A.L. Gloyn, The pancreatic beta cell: recent insights from human genetics. Trends in Endocrinology & Metabolism 25 (2014).

2. **S.K. Thomsen**, M.I. McCarthy, and A.L. Gloyn, The Importance of Context: Uncovering Species- and Tissue-Specific Effects of Genetic Risk Variants for Type 2 Diabetes. Frontiers in Endocrinology 7 (2016).

3. **S.K. Thomsen**, and A.L. Gloyn, Human genetics as a model for target validation: finding new therapies for diabetes. Diabetologia 60 (2017).

4. Pal*, T.P. Potjer*, **S.K. Thomsen***, et al, Loss-of-function mutations in the cell-cycle control gene CDKN2A impact on glucose homeostasis in humans. Diabetes (2016)

5. A. Raimondo*, **S.K. Thomsen***, et al, Type 2 diabetes risk alleles reveal a role for peptidylglycine alpha-amidating monooxygenase in beta cell function. BioRxiv (2017)

6. **S.K. Thomsen** et al, Systematic Functional Characterization of Candidate Causal Genes for Type 2 Diabetes Risk Variants. Diabetes (2016)

*Joint first authors

[Author comment added August 2018]: the BioRxiv pre-print listed as reference 5 above was accepted for publication in Nature Genetics after the submission of this thesis. The updated reference is provided below:

S.K. Thomsen* et al, Type 2 diabetes risk alleles in *PAM* impact insulin secretion from human pancreatic beta-cells. Nature Genetics (2018)

Review

CellPress

Special Issue: Recent Insights In The Pancreatic Beta Cell

The pancreatic β cell: recent insights from human genetics

Soren K. Thomsen[1] and Anna L. Gloyn[1,2]

[1] Oxford Centre for Diabetes, Endocrinology, and Metabolism, University of Oxford, Headington, OX3 7LE, UK
[2] Oxford National Institute for Health Research (NIHR) Biomedical Research Centre, Churchill Hospital, Headington, OX3 7LE, UK

Diabetes mellitus is a metabolic disease characterised by relative or absolute pancreatic β cell dysfunction. Genetic variants implicated in disease risk can be identified by studying affected individuals. To understand the mechanisms driving genetic associations, variants must be translated through causative transcripts to biological insights. Studies into the genetic basis of Mendelian forms of diabetes have successfully identified genes involved in both β cell function and pancreatic development. For type 2 diabetes (T2D), genome-wide association studies (GWASs) are uncovering an ever-increasing number of susceptibility variants that exert their effect through β cell dysfunction, but translation to mechanistic understanding has in most cases been slow. Improved annotations of the islet genome and advances in whole-genome and -exome sequencing (WHS and WES) have facilitated recent progress.

Interrogating the genetic basis of diabetes

Over the past decades diabetes mellitus has been increasing dramatically in prevalence and is predicted to affect 592 million worldwide by 2035 [1]. Auto-immune forms currently make up 5–10% of all cases, and the remainder predominantly comprises T2D. Changes in diet and lifestyle are key drivers behind the T2D pandemic, but heritable factors are known to contribute to a significant proportion of the phenotypic variance [2,3]. Studies into the genetic susceptibility of T2D have supported the notion that the disease develops due to an inability of the pancreatic β cell to compensate for increasing peripheral insulin resistance [4,5]. The primary function of the β cell is in metabolic homeostasis, achieved through a coupling of ambient glucose levels with insulin secretion. In T2D and related, rare forms of Mendelian diabetes, relative or absolute β cell dysfunction leads to hyperglycaemia, causing severe complications over time.

The study of human genetics provides a powerful tool to understand the mechanisms underlying β cell dysfunction in humans. Studies of individuals or populations with diabetes can be used to identify genetic variants (see Glossary) that either cause or increase susceptibility to disease. For neonatal diabetes mellitus (NDM) and maturity onset diabetes of the young (MODY), referred to collectively as Mendelian diabetes, the variants are rare and highly penetrant. Regions containing such variation

Glossary

Causative transcript: the transcript through which a genetic variant exerts its effect on disease pathogenesis.
Common variant associations study (CVAS): a type of GWAS targeting variants with a minor allele frequency (MAF) higher than 5%; Box 1 for details.
Expression quantitative trait loci (eQTL): a variant linked to the expression of a specific gene. eQTLs that map near the affected transcript are referred to as *cis*-eQTLs and those that are distal to the transcript are *trans*-eQTLs.
Exocrine pancreatic insufficiency: inadequate levels of digestive enzymes secreted by the exocrine part of the pancreas.
Fine-mapping: the process of refining a disease-associated region by targeted genotyping or sequencing in additional individuals, often from multiple ancestries.
Genetic variant: describes any type of variation at a given genomic position, including single-nucleotide changes and structural variation (deletions and insertions). Genetic variants can be classified based on whether they are coding or non-coding (regulatory) and causative or incidental. Coding mutations can be further divided into non-synonymous and synonymous variations depending on the effect they have on the encoded amino acids. Both non-synonymous mutations and structural variants can be protein-truncating if the translated protein is shorter than the wild type.
Genome-wide association study (GWAS): broad term that encompasses both CVASs and RVASs. Box 1 for details.
Haploinsufficiency: deviation from wild type state due to having only a single functional copy of a given allele.
Mendelian diabetes: covers those forms of diabetes which show a Mendelian pattern of inheritance (also sometimes referred to as monogenic diabetes). The term includes both maturity onset diabetes of the young (MODY) and neonatal diabetes mellitus (NDM), the two most common forms of Mendelian diabetes.
Minor allele frequency (MAF): frequency of the least common allele at a particular locus.
Missing heritability: the gap between the heritability accounted for by currently identified susceptibility variants and the variance in disease susceptibility observed by population genetics.
Odds ratio (OR): used in GWASs as a measure of the effect size for a specific variant. It is defined as the ratio between the proportion of cases and controls carrying a genetic variant.
Penetrance: proportion of mutation carriers that express the expected phenotype. A high penetrance is equivalent to a high OR. A highly penetrant, disease-causing variant will be rare in the general population because of reduced fitness. This trade-off between effect-size and frequency is exploited in the search for low-frequency and rare T2D susceptibility variants which could potentially have a higher OR.
Physiological characterisation: in the context of diabetes susceptibility, physiological characterisation refers to the measurement of several quantitative glycaemic traits. Frequently measured parameters include fasting proinsulin (PI) levels, insulinogenic index (IGI), and various indices of glucose tolerance. Such characterisation can provide insights into whether insulin resistance or β cell dysfunction is driving particular associations.
Rare variant association study (RVAS): a type of GWAS targeting variants with a MAF lower than 5% (low-frequency and rare variants); Box 1 for details.
Stretch enhancer: extended regulatory region bound by multiple transcription factors and exhibiting tissue-specific enhancer activity. Stretch enhancers mirror the related concepts of superenhancers, clusters of open regulatory elements (COREs), and locus control regions.

Corresponding author: Gloyn, A.L. (anna.gloyn@drl.ox.ac.uk).
Keywords: pancreatic β cell; human genetics; diabetes; GWAS.

1043-2760/
© 2014 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/3.0/). http://dx.doi.org/10.1016/j.tem.2014.05.001

CrossMark

Box 1. Identifying genomic regions implicated in diabetes risk

Linkage analysis and positional cloning

For highly penetrant mutations causing Mendelian diabetes, positional cloning has been an invaluable tool to identify genes implicated in disease pathogenesis. The principle is based on the tendency of genetic elements in close proximity to be coinherited. By mapping disease-causing variants relative to cosegregating genetic markers (DNA sequences of known position that show variation in sequence or length between individuals) the approximate position in the genome can be established. Typically, this approach will identify a candidate gene region covering millions of basepairs. Following this, several overlapping clones covering the region can be generated and the causative gene identified.

GWAS, RVAS, and CVAS studies

For complex forms of diabetes, numerous genes with small effect sizes and low penetrance contribute to disease susceptibility. Thus, rather than using linkage-based approaches, GWASs are applied to large samples of individuals with and without diabetes to establish statistical evidence for association of particular variants with the disease. The ratio between the frequency of an allele in cases compared with controls is referred to as the odds ratio (OR), a measure of the effect size of a variant. An OR of 1 indicates no effect, whereas an OR significantly different from 1 shows evidence of either a protective (OR <1) or harmful effect (OR >1). GWASs that target common SNPs [minor allele frequency (MAF) $>5\%$)] using microarray technology are referred to in this review as common variant association studies (CVASs). By contrast, studies that target low-frequency ($0.5\% > \text{MAF} > 5\%$) and rare variants ($\text{MAF} < 0.5\%$) by either sequencing-based approaches or microarrays designed to capture such SNPs are referred to as rare variant association studies (RVASs). Although often simply called exome-sequencing studies, we refrain from using this terminology to avoid mixing up statistical and genotyping methodologies [6].

were traditionally identified by linkage analysis and positional cloning (Box 1), but technological advances have recently resulted in WES and WGS becoming more common. For T2D, which is a multifactorial disease, the susceptibility variants have smaller effect sizes and can become more common due to reduced purifying selection. The majority of variants established to date have been identified through common variant association studies (CVASs), a type of GWAS targeting common variants (Box 1) [6]. The associations, however, have been found insufficient to account for measures of T2D heritability. To address this, rare variant association studies (RVASs) have been increasingly used to capture lower-frequency variants of moderate effect sizes (Box 1). The findings from studies on the genetics of both Mendelian and complex types of diabetes provide biological insights that have the potential to be translated directly into clinically relevant disease mechanisms. This advantage to human genetics is particularly evident in cases where rodent models have been found to inadequately recapitulate human phenotypes [7–9].

Despite advances in identifying genetic variants that increase diabetes susceptibility, the translation of association signals into molecular mechanisms has in many cases been slow. Exceptions include regions where biologically relevant genes harbour coding variants [10]. Many signals, however, are located in intergenic regions, and assigning a causative transcript to a particular variant presents a significant challenge (Figure 1A). Recent progress in structurally and functionally annotating the islet genome has provided a context in which to understand the effects of non-coding variants, and has in some cases facilitated translation into biological mechanisms (Figure 1B) [11–14].

In this review we first describe the current state of genomic annotation in islets, focusing on the biological impact of this work. Second, we discuss new insights into β cell biology from variants causing or increasing susceptibility to T2D and related forms of Mendelian diabetes. Where relevant we highlight the use of islet genome annotation to interrogate variation, a strategy that will become increasingly important in understanding how non-coding variants impact on the molecular machinery of the β cell.

Understanding the impact of regulatory variants on islet gene function

Genomic annotation is the process of assigning structural and functional meaning, such as gene structure and chromatin modifications, to specific sequences in the genome (Box 2). Many of these features are dynamically regulated, and therefore vary according to developmental stage, metabolic state, or tissue type. The ENCODE project has made significant progress in cataloguing genomic features (including various histone modifications and transcription

Box 2. Annotating the genome

Cataloguing transcription factor binding sites and epigenetic features, such as histone modifications, can shed light on the function of regulatory genomic regions. Evidence for this has been provided by recent studies in lymphoblastoid cell lines which have demonstrated that non-coding SNPs, at least to some extent, exert their *cis*-regulatory effects through disruption of transcription factor binding (Figure 1B) [68–70]. The changes in occupancy in turn alter the chromatin landscape both locally and at the promoter region of nearby genes, thereby influencing transcriptional efficiency.

Chromatin states and transcription factor networks

To functionally annotate genomic features, microarray analysis or WGS is applied to samples of DNA that are enriched or pretreated in different ways. One of the most commonly used methods is chromatin immunoprecipitation (ChIP) followed by microarray analysis (ChIP-chip) or sequencing (ChIP-seq). ChIP makes use of antibody-based purification to select genomic elements that are associated with either specific transcription factors or modified histones. Alternatively, isolation and identification of DNA fragments not associated with proteins can identify regions of open chromatin through so-called formaldehyde-assisted isolation of regulatory elements (FAIRE). Further, to map out DNA methylation patterns, approaches based on bisulfite conversion can be used to distinguish unmethylated from methylated cytosines.

Variant-transcript regulatory links

A higher level of functional information is present in the relationship between two distinct genomic elements. An important example is the statistical correlation between a variant and the expression of a particular gene. Such variants, so-called expression quantitative trait loci (eQTLs), can be found by combining transcriptomic and genomic datasets from islets, but require large sample sizes to obtain adequate statistical power. Once identified, overlapping GWAS association signals with *cis*-eQTLs (local eQTLs) can in principle provide strong evidence in support of a specific causative transcript (Figure 1). Large-scale studies are coming out, but early studies have already demonstrated that particular expected *cis*-eQTLs can be identified at a relaxed *P* value threshold [24]. Finally, variant-transcript links can also be identified by exploiting the fact that enhancers tend to interact with the promoter regions of target genes. By chromatin-conformation capture (3C)-based approaches it is possible to identify genomic elements that are in close spatial proximity owing to folding and looping of chromatin.

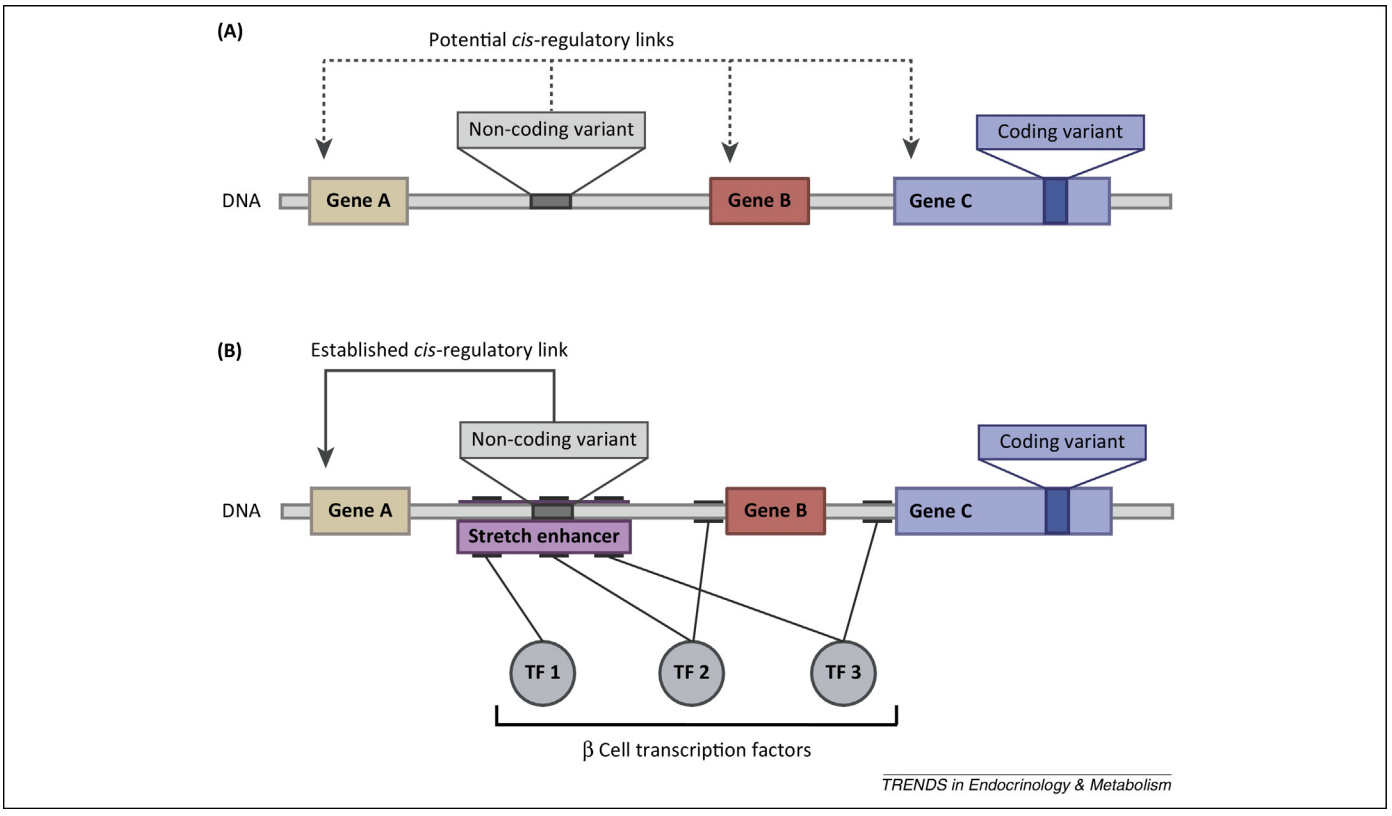


Figure 1. The mechanisms underlying disease-associated genetic variants. **(A)** Location of a non-coding and a coding genetic variant relative to three genes: A, B, and C. The coding variant is likely to exert its influence on disease pathogenesis through an effect on the function of gene C. Several different mechanisms, including amino acid substitution and protein truncation, could be driving the association. The non-coding variant, on the contrary, could be acting through any of the three nearby genes, as shown by the dotted arrows indicating potential *cis*-regulatory links. The variant could, for example, affect either expression levels or splicing patterns, but identifying the causative transcript can be challenging. **(B)** The schematic depicted in panel A with additional functional annotations (Box 2 for details). The binding sites of three β cell specific transcription factors (TF 1, 2 and 3) are shown by lines connected to the respective proteins. Three binding sites cluster to form a stretch enhancer, which is likely to have islet-specific *cis*-regulatory activity (see the section 'Understanding the impact of regulatory variants on islet gene function'). The non-coding variant is seen to overlap with the binding site of TF2 within the stretch enhancer, and could therefore plausibly act through disruption of TF2 binding. The solid arrow shows an established *cis*-regulatory link between the variant and gene 1. This relationship could be supported by different types of functional annotation, including a correlation between the non-coding variant and expression of gene 1 (established by *cis*-eQTL studies) or a physical interaction between the enhancer and the promoter region of gene 1 (demonstrated by chromatin-conformation capture, 3C) (Box 2). With such annotations at hand, identifying the causative transcript at a disease-associated locus is a more tractable problem.

factor binding sites) for a large number of cell types, but does not currently include either pancreatic islets or β cells. To obtain information that is directly relevant to our understanding of those diabetes risk variants that exert their effect through β cell dysfunction, several groups have recently addressed this issue. In most of these studies, islets of Langerhans have been used as a proxy for β cells, which make up 65–80% of the human islets but are hard to isolate as a pure fraction. Some have focused on identifying transcription factor binding sites or regions with specific states of chromatin [13–19].

Other studies have performed transcriptomic analyses to provide a wealth of information on the expression of both protein-coding genes and non-coding RNAs, thereby expanding the classes of structural elements that can be interrogated when considering the impact of genetic variants [20–23]. Such datasets have also been integrated with genomic data to enable genome-wide testing for variants that affect transcript levels, so-called expression quantitative trait loci (eQTL) (Box 2) [23,24]. Early studies have been limited by insufficient statistical power owing to relatively small sample sizes, but large-scale eQTL efforts in islets are currently underway and are anticipated to provide an important layer of regulatory information to the islet genome.

One of the key lessons emerging from the genomic annotations has been an increased understanding of how islet-specific genes are regulated. Several studies have made observations which indicate that genes involved in β cell identity and function tend to be regulated by clusters of transcription factors rather than by orphan enhancers [13,14,17]. The clusters form stretch- or super-enhancers that may act as molecular runways for tissue-specific transcription factors [17,25,26] (Figure 1B). Importantly, the islet stretch enhancers are found to be enriched for variants associated with susceptibility to diabetes and related metabolic traits [13,17].

From association signal to molecular mechanism

Mendelian diabetes

Mendelian forms of diabetes encompass a group of diseases, including both MODY and NDM, which show a Mendelian pattern of inheritance. Because of the clear genotype–phenotype correlation they constitute an invaluable tool for understanding how the disruption of a particular gene can impact on molecular and physiological

pathways in humans. Studies into the genetics of Mendelian diabetes have been remarkably successful in identifying genes implicated in pathogenesis, including both components of the insulin secretory pathways (e.g., *ABCC8, GCK, INS, KCNJ11*) (Figure 2A) and developmentally important transcription factors (e.g., *HNF1A, HNF4A, PDX1, PTF1A*) [27]. In some instances the insights arising from these studies have enabled substantial therapeutic advances, most notably in the case of sulfonylurea treatment for patients with mutations in genes encoding subunits of the K_{ATP} channel (*ABCC8* and *KCNJ11*) [28].

The genetics of Mendelian diabetes is continuing to provide insights into β cell function, facilitated by advances in sequencing-based approaches, such as WES and WGS, and in genomic annotation. WES offers the possibility to discover coding, disease-causing mutations in an untargeted manner, as recently demonstrated by a study investigating the genetic aetiology of pancreatic agenesis [8]. This rare condition, which clinically manifests itself through both diabetes and exocrine pancreatic insufficiency, is caused by complete or partial failure of the pancreas to develop. Lango Allen and colleagues performed WES on a subset of probands from a cohort of individuals with unexplained pancreatic agenesis. The authors used a common strategy of filtering against synonymous and known variation to enrich for likely pathogenic variants, and this identified heterozygous *de novo* mutations in the *GATA6* (GATA-binding factor 6) gene. Targeted sequencing of *GATA6* in further subjects revealed additional non-synonymous mutations in over half the individuals. Mutations in *GATA6* have subsequently been shown in follow-up studies to cause a spectrum of diabetic phenotypes, ranging from pancreatic agenesis to adult-onset diabetes with no overt exocrine insufficiency [29,30]. This incomplete penetrance of *GATA6*-inactivating mutations has also been noted by Bonnefond *et al.*, who showed that cardiac malformation is a more consistent phenotype of mutation carriers [31]. The variability in clinical presentation presumably reflects both mutational burden as well as genetic background. Overall, these studies have established that *GATA6* haploinsuffiency is a major cause of syndromic pancreatic agenesis and point to a crucial role for GATA6 in pancreatic development. Furthermore, this discovery highlights the value of using human genetics to understand gene-dosage relationships for dominant variants because heterozygous *Gata6* knockout mice show no apparent phenotype [8].

Unlike exome-based approaches, WGS can be used to identify both coding and regulatory variants. In non-coding regions, however, predicted disruption of protein function cannot be used as an indicator of pathogenicity. A challenge for WGS-based studies into Mendelian diseases is therefore related to the filtering of incidental variation from disease-causing mutations, and particularly in cases where linkage information is limited. A study into the genetic basis of patients with non-syndromic pancreatic agenesis (no extra-pancreatic features) explored the potential in combining WGS with spatially and temporally relevant genomic annotations to identify a subset of functionally active regulatory variants [11]. The study focused on homozygous mutations, consistent with the observation that close to half of the studied individuals were to born to consanguineous parents [8,11]. Failing to identify any plausible coding variants, the authors extended their search to developmentally active enhancer regions. Such an annotated map was obtained by defining enhancers bound by multiple transcription factors in pancreatic endoderm derived from human embryonic stem cells. In this way, a putative stretch enhancer downstream of *PTF1A* (pancreas transcription factor 1) was found to contain homozygous mutations in 9 of 12 individuals with isolated pancreatic agenesis. Compelling genetic and functional evidence established the pathogenicity of the variants, which all seem to abolish enhancer activity through disruption of transcription factor binding. Recessive coding mutations in *PTF1A* have previously been implicated in Mendelian diabetes, but are in all these cases associated with neurological features [32,33]. Interestingly, the distal enhancer found to harbour mutations causing isolated pancreatic agenesis does not have active marks in any ENCODE-annotated tissue type, suggesting that tissue-specific regulation can account for the absence of extra-pancreatic features [11].

The observation that some mouse models (e.g., *GATA6, HNF1A*) show poor phenotypic correlation with humans questions the degree to which mice can provide relevant insights into human pancreatic development [7,8,34]. To address this issue more systematically, Flanagan *et al.* screened for homozygous mutations in individuals with NDM born to consanguineous parents [9]. The study focused on 29 genes known to be important regulators of mouse pancreatic development. In total, seven of the genes were found to be affected in the studied collection, including *NKX2-2* (NK2 homeobox protein 2) and *MNX-1* (motor neuron and pancreas homeobox 1), which had not previously been associated with NDM. Comparison between the clinical features of the subjects (both pancreatic and extra-pancreatic) and the corresponding phenotypes of homozygous-null mouse models revealed pronounced similarities across all seven genes. Thus, it seems that the phenotypic correlation between humans and mouse models is generally high for biallelic inactivations.

Overall, the genetics of rare Mendelian forms of diabetes have provided a wealth of information into both the function of β cells and more recently also pancreatic development (Figure 2). The insights generated by these and other studies enhance our understanding of the transcriptional networks that regulate human pancreatic development [35,36]. Integration of genetics and functional genomics has been established as a powerful tool to interrogate regulatory variation, but the proportion of non-coding variants causing Mendelian diseases is still uncertain [6].

Type 2 diabetes

In contrast to Mendelian forms of diabetes, the genetic contribution to T2D is obscured by environmental influences and the relatively low penetrance of susceptibility variants. As a consequence, genetic variants are identified through association studies by establishing statistical evidence for overrepresentation in cases compared with controls (Box 1). Once identified at a stringent genome-wide

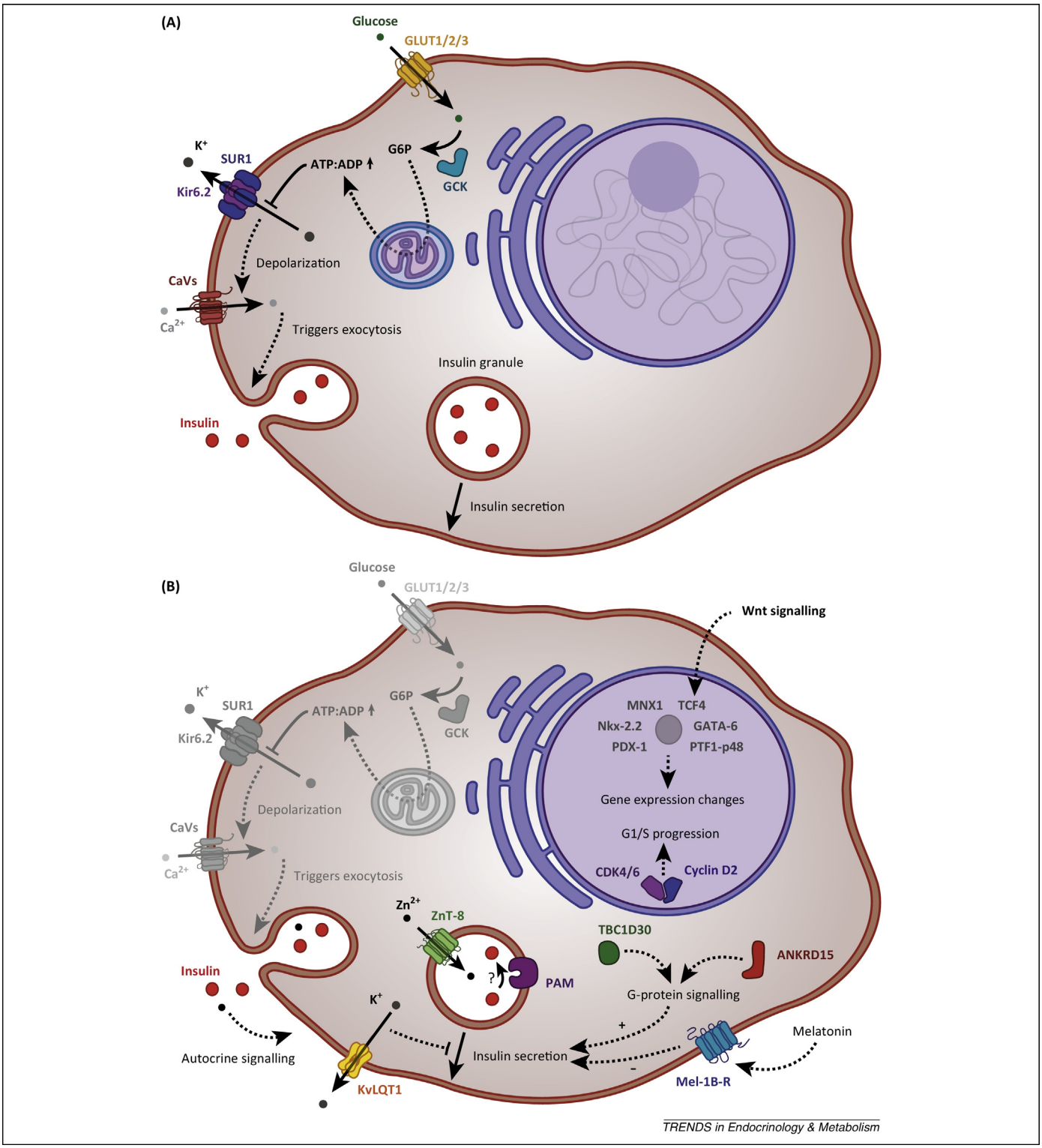


Figure 2. Schematic representation of the pancreatic β cell and the location of key components implicated by human genetics in diabetes pathogenesis. Solid arrows indicate direct mechanisms of action and dotted arrows indicate mechanisms with intermediate events not displayed in the schematic. **(A)** Consensus model of glucose-stimulated insulin secretion (GSIS), including several key proteins involved in Mendelian forms of diabetes. Proteins depicted: GCK (encoded by *GCK*), GLUT1/2/3 (*SLC2A1/2/3*), insulin (*INS*), Kir6.2 (*KCNJ11*), SUR1 (*ABCC8*), and voltage-gated Ca^{2+} channels (CaVs), which represent both P/Q type and L-type channels ([71] for details). Briefly, glucose enters the cell through one of the glucose transporters GLUT1/2/3 and is phosphorylated by GCK to form glucose-6-phosphate (G6P). In the mitochondria, aerobic metabolism of glycolytic products results in the generation of ATP. Changes in the ATP:ADP ratio inactivates the ATP-sensitive potassium channel K_{ATP}, which consists of four SUR1 and four Kir6.2 subunits. Closure of K_{ATP} in turn causes depolarization of the membrane opening voltage-gated Ca^{2+} channels. Finally, the influx of Ca^{2+} triggers exocytosis of insulin granules. **(B)** Location and proposed role in β cell function of proteins encoded by genes discussed in this review. Proteins depicted: ANKRD15 (encoded by *KANK1*), CDK4/6 (*CDK4 and CDK6*), cyclin D2 (*CCND2*), KvLQT1 (KCNQ1), Mel-1B (*MTNR1B*), PAM (*PAM*), TBC1D30 (*TBC1D30*), ZnT-8 (*SLC30A8*), and the transcription factors GATA-6 (*GATA6*), MNX1 (*MNX1*), Nkx-2.2 (*NKX2-2*), PDX-1 (*PDX1*), PTF1-p48 (*PTF1A*), and TCF4 (*TCF7L2*). See main text for details of individual genes.

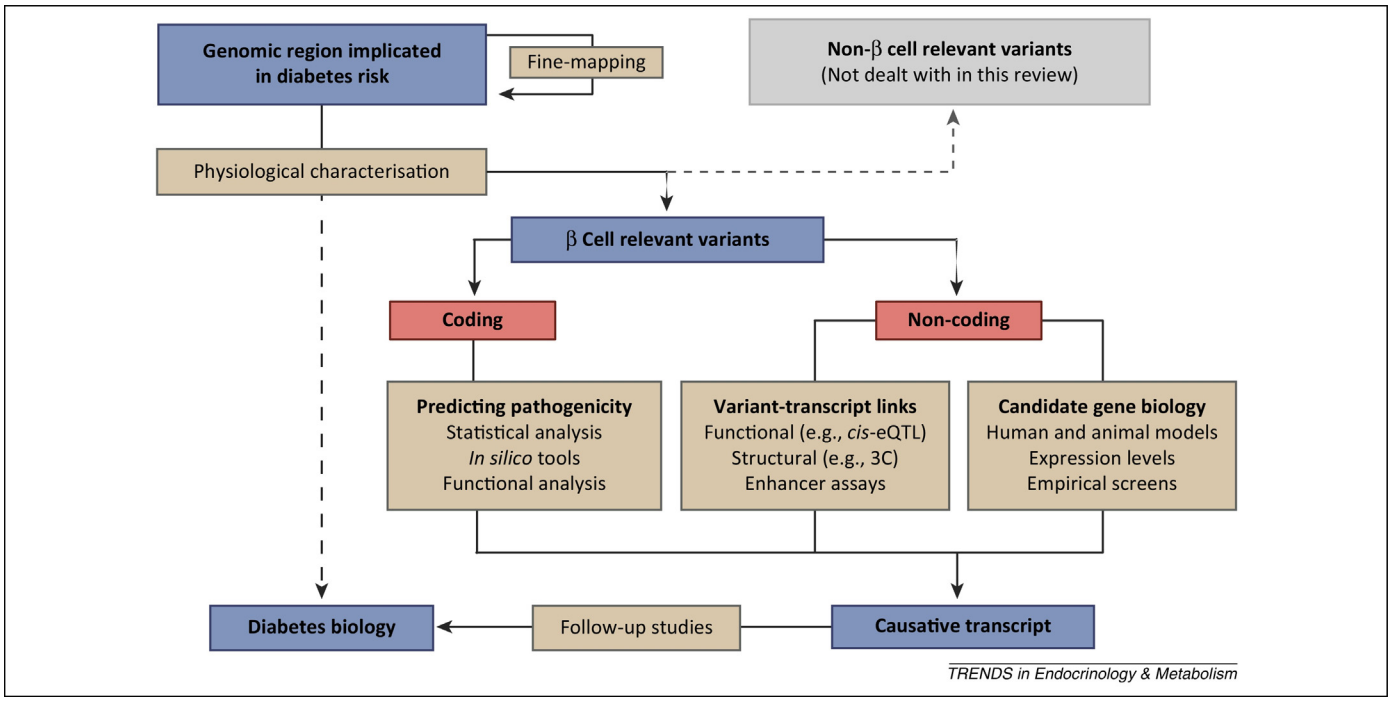


Figure 3. From association signal, through causative transcript, to biology. Beige boxes contain key experimental procedures; blue boxes indicate the entity of focus along the process of translating genetic associations to biology; and red boxes are divisions of such entities. As summarised in Box 1, human genetics can identify genomic regions implicated in diabetes risk. Subsequently, fine-mapping approaches can refine the genetic associations and identify the causative variant. Physiological characterisation of individuals carrying those risk alleles can in turn reveal effects indicative of the tissue(s) affected, but only indirect evidence for the underlying mechanisms (dotted line). Instead, identifying the transcripts driving the association of β cell relevant variants provides more direct insight into disease biology. For coding variants this can be achieved by prediction or validation of pathogenicity, whereas for non-coding signals consideration of prior candidate gene biology for nearby genes or variant transcript links can prioritize regional transcripts. Once a causative transcript has been identified, several different approaches, including human models, animal models, and functional studies, can be used to interrogate the function of the protein and ascertain its role in disease biology. Abbreviations: 3C, chromatin conformation capture; *cis*-eQTL, *cis*-expression quantitative trait loci (Box 2).

significance threshold, variants must be linked to causative transcripts, which can be interrogated for biological function (Figure 3). A few of the 77 susceptibility regions previously found to be associated with T2D contain a non-synonymous variant in the coding region of a gene [4,5,37]. In such cases (e.g., *KCNJ11*, *SLC30A8*, and *GCKR*), identifying the causative transcript is relatively straightforward and can be immediately followed up with functional studies and animal models [10]. For the remaining loci, however, the variants reside in non-coding regions of the genome, making it a challenge to establish a causal link between the signal and a nearby gene (Figure 1). We discuss here progress concerning variants from across the frequency spectrum.

Common variants

Most of the established T2D-associated loci are common variants that have been identified through CVASs [4,5,37]. Physiological characterisations of carriers have demonstrated that, although T2D is etiologically heterogeneous, the majority of these association signals appear to act through β cell dysfunction [4,38,39]. Thus, inferring the causative transcripts from risk variants holds the promise to identify genes that are important for β cell function.

In the case of the *TCF7L2* (transcription factor 7-like 2, T cell specific, HMG-box) locus, improvements in islet genomic annotations have provided new insights into the mechanism of an intronic risk variant [13,14]. The association signal has been the focus of particular interest because of its relatively large effect size [odds ratio (OR) = 1.4 in the CEU (Central European) cohort], the highest of any common variant in Caucasian populations [5]. Despite the high OR, functional studies have shown inconsistent effects of *TCF7L2* levels on β cell function, and the primary tissues through which the effect is expressed has remained controversial [40–44]. Genomic annotations, however, have provided several lines of evidence supporting islet-specific enhancer activity. First, Gaulton *et al.* were able to show that the *TCF7L2* association signal is located in an islet-selective region of open chromatin, indicative of regulatory activity [14]. The risk allele is itself associated with a more accessible state of chromatin and increased enhancer activity in β cell lines [14,18]. Further, it has been demonstrated by a comprehensive analysis of transcription factor binding sites in islets that the region containing the lead single-nucleotide polymorphism (SNP) is a stretch enhancer, bound by three key islet transcription factors [13]. Together, these observations provide strong evidence in favour of a β cell driven effect on T2D susceptibility (Figure 2B).

A different type of regulatory relationship has been described for the intronic variants at the *KCNQ1* (potassium voltage-gated channel, KQT-like subfamily, member 1) locus. The region contains both multiple independent association signals and biologically plausible transcripts, including *KCNQ1* (Figure 2B). Consistent with the imprinted status of the locus, it is one of only two T2D loci to display a strong parent-of-origin effect [45,46].

Physiological characterisation of individuals with loss-of-function mutations in *KCNQ1* has demonstrated increased insulin secretion, highlighting *KCNQ1* as a likely causative transcript [47]. This is reinforced by a study investigating the imprinting status of regional transcripts in foetal and adult islets [12]. Evidence of monoallelic expression was observed for *KCNQ1* in the foetal but not in the adult state, arguing that any potential effect on T2D must be developmental. Thus, *KCNQ1* is emerging as a strong candidate at the locus, although the possibility of effects on other regional transcripts, such as *CDKN1C* (cyclin-dependent kinase inhibitor 1C; encoding $p27^{KIP2}$), cannot be excluded [19].

Low-frequency and rare variants

Despite the large number of common variants having been discovered by CVASs, the combined population-attributable risk accounts for only a small fraction of the total heritability of T2D (Box 3) [4,5]. This has prompted speculations that rare and low-frequency variants of moderate effect sizes could explain a significant proportion of the 'missing heritability' [48]. Early RVASs have addressed this conundrum directly by targeting either specific regions already established by CVAS, or by examining the whole exome or genome in an unbiased fashion [49–52].

An initial study of this kind focused on *MTNR1B* (melatonin receptor 1B), a gene previously implicated in both T2D and fasting glucose levels by CVASs [53,54]. The melatonin receptor 1B is an inhibitory G protein-coupled receptor (GPCR) that modulates insulin secretion (Figure 2B). It has attracted attention both because of the therapeutic value of GPCRs but also because of the potential link between circadian rhythm and metabolic control; release of melatonin from the pineal gland is inhibited by light, and thus the endocrine hormone is thought to reduce insulin secretion during the night. Because *MTNR1B* expression was initially observed to be higher in individuals with T2D than in controls it was expected that the common risk variant would increase expression [54]. This, in turn, would lead to inappropriately high inhibition of insulin secretion by melatonin. Resequencing of the region, however, identified several rare variants that caused pronounced decreases in receptor function [49]. The variants were too rare to allow association tests at the level of individual variants but were aggregated to increase statistical power, a common practice in RVASs [6]. Grouping based on functional effect established that alleles harbouring rare loss-of-function variants significantly increases susceptibility to T2D (OR = 5.7). By contrast, a meta-allele consisting of all rare variants shows a diluted effect (OR = 3.3). This demonstrates the importance of correctly assigning pathogenicity to exclude incidental mutations and thereby increase power. In addition to discovering an unexpected direction of effect, the study also provides proof-of-principle that rare variants with large effect sizes can be found by resequencing loci prioritized by CVAS.

Targeted resequencing has also been remarkably successful in shedding light on the direction of effect at the *SLC30A8* (solute carrier family 30 member 8) locus [50]. *SLC30A8* encodes a Zn^{2+} transporter, ZnT-8, that localises to the insulin-containing secretory granules (Figure 2B). It was first proposed that a common non-synonymous variant identified by CVAS would decrease function of the transporter, leading to an insulin crystallization defect. Nevertheless, follow-up functional studies and animal models have only reported small, incongruent effects on β cell function [55–59] (reviewed in [60]). To resolve this, Flannick *et al.* genotyped 150 000 individuals looking for T2D-association of rare variants with large functional effects [50]. By aggregating all protein-truncating mutations, they were able to show at an exome-wide level of significance that loss-of-function variants are protective against T2D (OR = 0.34). Demonstrating the power of human genetics to determine mechanistic directionality, this surprising finding shows that a 50% reduction of ZnT-8 levels decreases the risk of developing T2D by almost threefold. To understand the underlying protective mechanism and evaluate the potential of ZnT-8 as a therapeutic target, these new insights will need to be reconciled with existing data, as discussed elsewhere [60].

Two RVASs so far have attempted to uncover new loci by non-targeted approaches [51,52]. Huyghe *et al.* provided proof-of-principle using an exome-chip approach to investigate several T2D-relevant traits in healthy adults [51]. The study identified associations for insulinogenic index (IGI), a measure of insulin release in response to glucose, and proinsulin (PI) levels, a proxy for β cell stress. The three genes found to be implicated are all strong biological candidates: *PAM* (peptidylglycine α-amidating monooxygenase; associated with IGI) encodes an α-amidase localising to the membrane of secretory granules, and both *KANK1* (KN motif and ankyrin repeat domains 1) and *TBC1D30* (TBC1 domain family, member 3; associated with PI) encode proteins involved in G-protein signalling (Figure 2B).

Box 3. Missing heritability in T2D

Common variants identified to date together account for less than 10% of the variance in diabetes susceptibility estimated by twin and segregation studies [5,48]. To account for this 'missing heritability' several potential sources have been proposed, including gene–environment interactions, structural variants not captured by existing GWASs, rare variants of moderately high effect size, incorrect estimates of heritability, and inappropriate study design. Rare variants in particular have attracted attention recently and the focuses of many GWASs have shifted from CVAS towards RVAS. The notion of rare variants contributing to disease susceptibility is supported by resequencing of the diabetes susceptibility genes *KCNJ11* and *HHEX*, which revealed an unexpectedly large number of rare variants arising because of recent explosive population growth [72]. Conversely, modelling of the distribution of common variants has shown that a very large number of such variants may be associated with T2D [5]. These would explain a majority of the heritability but require unfeasibly large populations for genome-wide significance owing to progressively smaller effect sizes. Supporting this, Lohmueller and colleagues combined simulations with sequencing of a limited number of individuals, and concluded that their observations were inconsistent with a model where rare variants in a modest number of genes explain the majority of diabetes susceptibility [73]. However, recent studies with larger samples sizes have been able to establish exome-wide levels of statistical significance for rare variants at both new and previously established T2D loci (main text for details). As for common variants, it appears that the limited number of rare variants identified so far primarily affect β cell function. Overall, despite their relatively high ORs, the population-attributable risks remain low because of their low frequencies, and it remains to be seen whether rare variants contribute significantly to T2D heritability.

More recently, the deCODE consortium has studied a cohort of Icelanders to test for associations of rare and low-frequency variants with T2D [52]. By combining WGS of a limited number of individuals with imputation based on Icelandic genealogy they were able to obtain an effective sample size of close to 300 000 individuals. This provided sufficient power to establish four previously unreported association signals in three regions: *PAM*, in which a rare variant had also previously been associated with IGI (see above), *CCND2* (cyclin D2), and *PDX1* (pancreatic and duodenal homeobox 1) (Figure 2B). Interestingly, the intronic variant in *CCND2* was protective against T2D (OR = 0.5) but was associated with higher body mass index (BMI) and height. Cyclin D2 is a widely expressed regulator of the G1/S cell cycle transition and has been shown to influence the proliferative capacity of β cells among many other cell types [61]. *PDX1* encodes a transcription factor important for β cell maturation and is known to be involved in Mendelian forms of diabetes [27]. Accordingly, a rare frameshift mutation was shown to have a moderately high effect (OR = 2.3) [52].

The discovery of *PDX1* as a T2D risk locus expands the overlap between genes implicated in T2D susceptibility and genes known to be involved in Mendelian diabetes (e.g., *KCNJ11, HNF1A, HNF1B, WFS1*). Blurring the boundary between the two diseases, it has recently been shown that the apparent deterministic effect of particular MODY-mutations may reflect an inflated estimate of effect size as a result of ascertainment bias [62]. As more rare variants with large effect sizes become associated with T2D, the operational cut-off point for classification of a mutation as an incompletely penetrant Mendelian variant or a rare T2D mutation with a high effect size will become increasingly arbitrary.

Concluding remarks and future perspectives

Human genetics has proven a powerful instrument to explore the genetics of β cell dysfunction. Substantial progress has been made from studies into Mendelian forms of diabetes, which have improved our understanding of β cell biology and delivered validated drug targets. Translation of initial T2D variants has been slower, but recent studies have shown progress in elucidating the mechanisms underlying several common and low-frequency susceptibility variants (Figure 2) [37,51,52]. Several large-scale sequencing efforts by international consortia are currently underway, and these will undoubtedly further expand the number of robustly associated genetic variants. Although it seems unlikely that the first generation of such studies will uncover a large proportion of the missing heritability, they may shed much needed light on the overall distribution of variant frequencies and effect sizes at T2D loci.

One of the greatest challenges in the field remains to bridge the gap between regulatory variants and causative transcripts, but several new methods are promising to accelerate this process. First, the identification of rare, coding variants in regions associated with common regulatory variants can provide prior evidence in favour of particular transcripts. Proof-of-principle RVASs have already been published, but studies with larger sample sizes will provide greater statistical power [51,52]. A second approach is based on annotation of the islet genome through the integration of large-scale genomics and transcriptomics datasets. Understanding the genomic context of genetic associations enables the identification of a functionally active subset of variants and even prediction or direct determination of the impact that specific alleles have on chromatin states. A large number of studies have already contributed to annotating the islet genome, creating an atlas with a growing number of structural and functional layers. As an extension of this principle, *cis*-eQTL studies promise to deliver an additional level of information by identifying correlations between genotypes and expression levels.

Once a causative gene has been inferred, the discovery must be followed up by studies to interrogate the biological function of the encoded gene product. Only then can we gain new insights into β cell biology and appreciate the full implications for disease pathogenesis. For such follow-up studies, traditional characterisation of animal models is being increasingly supplemented by new human model systems. The generation of the first glucose-responsive human β cell line was recently reported [63]. Although still an imperfect model, the cell line represents a major step forward and will complement current work in rodent cell lines. In addition, advances are being made in both the derivation of induced pluripotent stem (iPS) cells and in their subsequent differentiation to the endodermal lineage [64–66]. Soon patient-derived iPS cells will be used routinely to study β cell function. Novel genome-editing tools such as the CRISPR/cas system are already available for fast and simple genetic manipulation of cell lines and iPS cells [67]. In this way it is possible to create both knockouts and models carrying risk alleles with more subtle effects on splicing or function. Finally, appropriate physiological characterisation of human variant-carriers will provide increasingly detailed insight into pathophysiological mechanisms.

Future studies evaluating genetic associations will be supported by such new methods, and thereby accelerate the process of going from variant through transcript to biology. The insights arising from human genetics will continue to improve our understanding of β cell function

Box 4. Outstanding questions

- How is the balance between secretion, proliferation, and apoptosis ultimately regulated and disturbed by T2D susceptibility variants that exert their effect through β cells? Translating more variants to biological mechanisms will improve our understanding of this complex system and thereby also diabetes pathogenesis (Figure 3).
- Will larger *cis*-eQTL studies and other functional annotations, such as chromatin conformation capture (3C) (Box 2), deliver the anticipated advance in facilitating translation of regulatory variants to molecular mechanisms?
- Will rare variants contribute significantly to the missing heritability of T2D? In addition, will rare variants, similarly to common variants, primarily exert their effect on diabetes susceptibility through β cell dysfunction?
- What proportion of rare variants will be found in regulatory regions of the genome? There is reason to believe that rare variants with moderate and high effects are less likely to be regulatory than common variants because of the more dramatic effect that coding mutations can have on protein function.

(Box 4), inspiring the development of novel drugs and facilitating advances in replacement therapies.

Acknowledgements

S.K.T. is a Radcliffe Department of Medicine Scholar and A.L.G. is a Wellcome Trust Senior Fellow in Basic and Biomedical Research. The authors apologise to those whose work they have been unable to cite in this review article owing to space constraints and reference guidelines.

References

1 Guariguata, L. *et al.* (2013) *IDF Diabetes Atlas*, International Diabetes Federation
2 Almgren, P. *et al.* (2011) Heritability and familiality of type 2 diabetes and related quantitative traits in the Botnia Study. *Diabetologia* 54, 2811–2819
3 Hu, F.B. (2011) Globalization of diabetes: the role of diet, lifestyle, and genes. *Diabetes Care* 34, 1249–1257
4 Voight, B.F. *et al.* (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.* 42, 579–589
5 Morris, A.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
6 Zuk, O. *et al.* (2014) Searching for missing heritability: designing rare variant association studies. *Proc. Natl. Acad. Sci. U.S.A.* 111, E455–E464
7 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. *Mol. Genet. Metab.* 104, 648–653
8 Lango Allen, H. *et al.* (2012) GATA6 haploinsufficiency causes pancreatic agenesis in humans. *Nat. Genet.* 44, 20–22
9 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 Metab.* 19, 146–154
10 Ng, H.J. and Gloyn, A.L. (2013) Bridging the gap between genetic associations and molecular mechanisms for type 2 diabetes. *Curr. Diab. Rep.* 13, 778–785
11 Weedon, M.N. *et al.* (2014) Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. *Nat. Genet.* 46, 61–64
12 Travers, M.E. *et al.* (2013) Insights into the molecular mechanism for type 2 diabetes susceptibility at the KCNQ1 locus from temporal changes in imprinting status in human islets. *Diabetes* 62, 987–992
13 Pasquali, L. *et al.* (2014) Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat. Genet.* 46, 136–143
14 Gaulton, K.J. *et al.* (2010) A map of open chromatin in human pancreatic islets. *Nat. Genet.* 42, 255–259
15 Volkmar, M. *et al.* (2012) DNA methylation profiling identifies epigenetic dysregulation in pancreatic islets from type 2 diabetic patients. *EMBO J.* 31, 1405–1426
16 Bhandare, R. *et al.* (2010) Genome-wide analysis of histone modifications in human pancreatic islets. *Genome Res.* 20, 428–433
17 Parker, S.C.J. *et al.* (2013) Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc. Natl. Acad. Sci. U.S.A.* 110, 17921–17926
18 Stitzel, M.L. *et al.* (2010) Global epigenomic analysis of primary human pancreatic islets provides insights into type 2 diabetes susceptibility loci. *Cell Metab.* 12, 443–455
19 Dayeh, T. *et al.* (2014) Genome-wide DNA methylation analysis of human pancreatic islets from type 2 diabetic and non-diabetic donors identifies candidate genes that influence insulin secretion. *PLoS Genet.* 10, e1004160
20 Moran, I. *et al.* (2012) Human beta cell transcriptome analysis uncovers lncRNAs that are tissue-specific, dynamically regulated, and abnormally expressed in type 2 diabetes. *Cell Metab.* 16, 435–448
21 van de Bunt, M. *et al.* (2013) The miRNA profile of human pancreatic islets and beta-cells and relationship to type 2 diabetes pathogenesis. *PLoS ONE* 8, e55272
22 Kameswaran, V. *et al.* (2014) Epigenetic regulation of the DLK1–MEG3 microRNA cluster in human type 2 diabetic islets. *Cell Metab.* 19, 135–145
23 Nica, A.C. *et al.* (2013) Cell-type, allelic, and genetic signatures in the human pancreatic beta cell transcriptome. *Genome Res.* 23, 1554–1562
24 Taneera, J. *et al.* (2012) A systems genetics approach identifies genes and pathways for type 2 diabetes in human islets. *Cell Metab.* 16, 122–134
25 Loven, J. *et al.* (2013) Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* 153, 320–334
26 Whyte, W.A. *et al.* (2013) Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 153, 307–319
27 Ashcroft, F.M. and Rorsman, P. (2012) Diabetes mellitus and the beta cell: the last ten years. *Cell* 148, 1160–1171
28 Flanagan, S.E. *et al.* (2009) Update of mutations in the genes encoding the pancreatic beta-cell K(ATP) channel subunits Kir6.2 (KCNJ11) and sulfonylurea receptor 1 (ABCC8) in diabetes mellitus and hyperinsulinism. *Hum. Mutat.* 30, 170–180
29 De Franco, E. *et al.* (2013) GATA6 mutations cause a broad phenotypic spectrum of diabetes from pancreatic agenesis to adult-onset diabetes without exocrine insufficiency. *Diabetes* 62, 993–997
30 Catli, G. *et al.* (2013) A novel GATA6 mutation leading to congenital heart defects and permanent neonatal diabetes: a case report. *Diabetes Metab.* 39, 370–374
31 Bonnefond, A. *et al.* (2012) GATA6 inactivating mutations are associated with heart defects and, inconsistently, with pancreatic agenesis and diabetes. *Diabetologia* 55, 2845–2847
32 Tutak, E. *et al.* (2009) A Turkish newborn infant with cerebellar agenesis/neonatal diabetes mellitus and PTF1A mutation. *Genet. Couns.* 20, 147–152
33 Al-Shammari, M. *et al.* (2011) A novel PTF1A mutation in a patient with severe pancreatic and cerebellar involvement. *Clin. Genet.* 80, 196–198
34 da Silva Xavier, G. *et al.* (2013) Animal models of GWAS-identified type 2 diabetes genes. *J. Diabetes Res.* 2013, 906590
35 Smith, S.B. *et al.* (2010) Rfx6 directs islet formation and insulin production in mice and humans. *Nature* 463, 775–780
36 Bonnefond, A. *et al.* (2013) Transcription factor gene MNX1 is a novel cause of permanent neonatal diabetes in a consanguineous family. *Diabetes Metab.* 39, 276–280
37 Mahajan, A. (2014) Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat. Genet.* 46, 234–244
38 Ingelsson, E. *et al.* (2010) Detailed physiologic characterization reveals diverse mechanisms for novel genetic loci regulating glucose and insulin metabolism in humans. *Diabetes* 59, 1266–1275
39 Dimas, A.S. *et al.* (2013) Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. *Diabetes* http://dx.doi.org/10.2337/db2313-0949
40 Le Bacquer, O. *et al.* (2012) TCF7L2 rs7903146 impairs islet function and morphology in non-diabetic individuals. *Diabetologia* 55, 2677–2681
41 da Silva Xavier, G. *et al.* (2012) Abnormal glucose tolerance and insulin secretion in pancreas-specific Tcf7l2-null mice. *Diabetologia* 55, 2667–2676
42 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
43 Rosengren, A.H. *et al.* (2012) Reduced insulin exocytosis in human pancreatic beta-cells with gene variants linked to type 2 diabetes. *Diabetes* 61, 1726–1733
44 da Silva Xavier, G. *et al.* (2009) TCF7L2 regulates late events in insulin secretion from pancreatic islet beta-cells. *Diabetes* 58, 894–905
45 Kong, A. *et al.* (2009) Parental origin of sequence variants associated with complex diseases. *Nature* 462, 868–874
46 Hanson, R.L. *et al.* (2013) Strong parent-of-origin effects in the association of KCNQ1 variants with type 2 diabetes in American Indians. *Diabetes* 62, 2984–2991
47 Torekov, S.S. *et al.* (2013) KCNQ1 Long QT syndrome patients have hyperinsulinemia and symptomatic hypoglycemia. *Diabetes* 63, 1315–1325
48 Manolio, T.A. *et al.* (2009) Finding the missing heritability of complex diseases. *Nature* 461, 747–753
49 Bonnefond, A. *et al.* (2012) Rare MTNR1B variants impairing melatonin receptor 1B function contribute to type 2 diabetes. *Nat. Genet.* 44, 297–301
50 Flannick, J. *et al.* (2014) Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. *Nat. Genet.* 46, 357–363
51 Huyghe, J.R. *et al.* (2013) Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. *Nat. Genet.* 45, 197–201

52 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
53 Prokopenko, I. *et al.* (2009) Variants in MTNR1B influence fasting glucose levels. *Nat. Genet.* 41, 77–81
54 Lyssenko, V. *et al.* (2009) Common variant in MTNR1B associated with increased risk of type 2 diabetes and impaired early insulin secretion. *Nat. Genet.* 41, 82–88
55 Pound, L.D. *et al.* (2012) The physiological effects of deleting the mouse SLC30A8 gene encoding zinc transporter-8 are influenced by gender and genetic background. *PLoS ONE* 7, e40972
56 Pound, L.D. *et al.* (2009) Deletion of the mouse Slc30a8 gene encoding zinc transporter-8 results in impaired insulin secretion. *Biochem. J.* 421, 371–376
57 Lemaire, K. *et al.* (2009) Insulin crystallization depends on zinc transporter ZnT8 expression, but is not required for normal glucose homeostasis in mice. *Proc. Natl. Acad. Sci. U.S.A.* 106, 14872–14877
58 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
59 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
60 Davidson, H.W. *et al.* (2014) Zinc transporter 8 (ZnT8) and β cell function. *Trends Endocrinol. Metab.* http://dx.doi.org/10.1016/j.tem.2014.1003.1008
61 Georgia, S. *et al.* (2010) Cyclin D2 is essential for the compensatory beta-cell hyperplastic response to insulin resistance in rodents. *Diabetes* 59, 987–996
62 Flannick, J. *et al.* (2013) Assessing the phenotypic effects in the general population of rare variants in genes for a dominant Mendelian form of diabetes. *Nat. Genet.* 45, 1380–1385
63 Ravassard, P. *et al.* (2011) A genetically engineered human pancreatic beta cell line exhibiting glucose-inducible insulin secretion. *J. Clin. Invest.* 121, 3589–3597
64 Hua, H. *et al.* (2013) iPSC-derived beta cells model diabetes due to glucokinase deficiency. *J. Clin. Invest.* 123, 3146–3153
65 Maehr, R. *et al.* (2009) Generation of pluripotent stem cells from patients with type 1 diabetes. *Proc. Natl. Acad. Sci. U.S.A.* 106, 15768–15773
66 Shang, L. *et al.* (2014) Beta-cell dysfunction due to increased ER Stress in a stem cell model of Wolfram syndrome. *Diabetes* 63, 923–933
67 Cong, L. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823
68 Kasowski, M. *et al.* (2013) Extensive variation in chromatin states across humans. *Science* 342, 750–752
69 Kilpinen, H. *et al.* (2013) Coordinated effects of sequence variation on DNA Binding, chromatin structure, and transcription. *Science* 342, 744–747
70 McVicker, G. *et al.* (2013) Identification of venetic variants that affect histone modifications in human cells. *Science* 342, 747–749
71 Rorsman, P. and Braun, M. (2013) Regulation of insulin secretion in human pancreatic islets. *Annu. Rev. Physiol.* 75, 155–179
72 Coventry, A. *et al.* (2010) Deep resequencing reveals excess rare recent variants consistent with explosive population growth. *Nat. Commun.* 1, 131
73 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. *Am. J. Hum. Genet.* 93, 1072–1086

frontiers in Endocrinology

OPINION
published: 31 August 2016
doi: 10.3389/fendo.2016.00112

The Importance of Context: Uncovering Species- and Tissue-Specific Effects of Genetic Risk Variants for Type 2 Diabetes

*Soren K. Thomsen[1], Mark I. McCarthy[1,2,3] and Anna L. Gloyn[1,2,3]**

[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] Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, UK

Keywords: type 2 diabetes, genome-wide association study, genetic association, genetic mechanism, genomic annotation, gene regulation, epigenetics

Genome-wide association studies (GWAS) have been highly successful in identifying genetic variation associated with type 2 diabetes (T2D) risk and related quantitative traits (1–3). The vast majority of association signals are located in non-coding regions of the genome, influencing nearby genes through regulation of transcriptional, translational, or splicing activity (4). Due to the highly context-dependent nature of gene expression, the effects of many risk variants are restricted to specific cell types and produce more subtle effects than those observed in organism-wide (or "global") knockouts. In addition, identification of the underlying causal genes and target tissues is often a major challenge, hindering translation into disease mechanisms. Recent studies have shown that the intersection of genetic data and genomic annotations can be used to produce a cellular atlas with which to understand the phenotypes of GWAS signals. Through the generation of directed hypotheses, this integrated framework has the potential to bridge the gap between association signals and disease biology.

OPEN ACCESS

Edited by:
Robert Kenneth Semple, University of Cambridge, UK

Reviewed by:
Luca Andrea Lotta, University of Cambridge, UK

****Correspondence:***
Anna L. Gloyn
anna.gloyn@drl.ox.ac.uk

Specialty section:
This article was submitted to Genomic Endocrinology, a section of the journal Frontiers in Endocrinology

Received: *24 April 2016*
Accepted: *04 August 2016*
Published: *31 August 2016*

Citation:
Thomsen SK, McCarthy MI and Gloyn AL (2016) The Importance of Context: Uncovering Species- and Tissue-Specific Effects of Genetic Risk Variants for Type 2 Diabetes. Front. Endocrinol. 7:112. doi: 10.3389/fendo.2016.00112

THE BASIS FOR TISSUE SPECIFICITY

Across the human population, differences in complex traits, such as height and disease susceptibility, are influenced by the presence of single-nucleotide polymorphisms (SNPs). Some of these genetic variants modulate binding of transcription factors (TFs), which in turn drive differences in gene expression (5, 6). TF binding is also influenced by co-factors and chromatin state, which are highly dependent on cell type and developmental stage. To establish and maintain cellular identity, cell-type-specific TFs tend to bind in clusters, often referred to as cis-regulatory modules (7–9). Intriguingly, association signals for T2D have been found to show a significant overlap with islet-selective enhancer clusters (9, 10). Although other tissues have also been implicated in T2D susceptibility, this is consistent with physiological studies establishing islet dysfunction as a central mechanism of disease-associated variants (1, 11–13).

The overlap between T2D signals and enhancer clusters suggests that the effect of risk variants could be subject to the same tissue specificity observed at the level of regulatory activity (11). In other words, a motif-altering allele would only be expected to produce a molecular phenotype in those contexts (cell type or developmental stage) where the binding site has the potential to be occupied. In support of this notion, T2D risk variants were found to be enriched for nearby binding sites of the pioneer TF FOXA2 in islet and liver (14). T2D-association signals also show a significant overlap with SNPs affecting islet expression of regional transcripts, so-called cis-expression quantitative trait loci (cis-eQTL), most of which are not found to be eQTLs in other tissues (15). Together, these general observations outline some of the context-dependent effects that T2D risk variants can be subject to.

CONTEXT SPECIFICITY OF CAUSAL MECHANISMS FOR T2D SUSCEPTIBILITY

Recent studies have provided more specific evidence to support the notion that context specificity is a key aspect of GWAS causal mechanisms. The following cases collectively provide examples of mechanisms where studying the right tissue, species, and developmental stage proved critical to uncovering the relevant phenotypes.

At the *MTNR1B* locus, a convergence of evidence has pointed to effects of a non-coding T2D-association signal on the pancreatic β cell. Physiological studies have revealed a phenotype indicative of β-cell dysfunction in risk-variant carriers, with some evidence for additional effects on insulin action (12, 16, 17). Fine-mapping efforts identified a single likely causal variant that overlaps active islet and liver enhancers, and a cis-eQTL for *MTNR1B* in islets (14, 15, 18, 19). The risk allele, which increases islet *MTNR1B* expression, was predicted to create a NEUROD1 binding site and shown to selectively bind this key TF in human β cells. These results establish a likely causal mechanism for the non-coding risk allele, and illustrate how motif-altering alleles can generate highly tissue-specific effects. Surprisingly, exon re-sequencing of the *MTNR1B* gene has also shown coding loss-of-function (LOF) mutations to be associated with increased risk of T2D (20). The reason for the opposite directions of effect observed for coding and non-coding risk variants is unclear but may reflect differences between global and islet-specific roles of *MTNR1B.*

In the case of *PTF1A,* studying the right tissue proved necessary but not sufficient to elucidating the underlying mechanism for non-coding mutations in the region. Previous work had identified a group of patients suffering from unexplained isolated pancreatic agenesis, which includes neonatal diabetes as a clinical feature (21). To filter causal mutations from incidental variation, one study used pancreatic endoderm to define regulatory regions that are active during pancreatic development (22). Their strategy identified a distal enhancer that harbors mutations abolishing enhancer activity toward *PTF1A* (23). Coding LOF variants in *PTF1A* had previously implicated the gene in syndromic pancreatic agenesis, characterized by severe neurological features in affected individuals. The observation that the identified enhancer region is not active in any cell type other than pancreatic endoderm provides a plausible explanation for the absence of any cerebral defects (22). Remarkably, even adult pancreatic tissue did not show active chromatin marks in the region, highlighting that studying the right developmental stage was critical to the success of the approach.

The mechanisms underlying GWAS signals are sometimes studied using individuals that carry LOF mutations in positional candidate genes. The observed phenotypes will be a function of global effects across all the tissues where the gene is expressed, which may confound or mask the more context-dependent actions of regulatory risk alleles. At the *CDKN2A* locus, non-coding T2D signals have been robustly associated with measures of islet dysfunction, and a number of studies have established effects of *CDKN2A* on insulin secretion and cellular senescence in β-cells (12, 24, 25). By contrast, coding LOF mutations in *CDKN2A*, which are a cause of familial melanoma, were recently shown to result in a metabolic phenotype consistent with effects on both liver and β-cells (26). This discrepancy was proposed to arise from islet-specific TF binding of the enhancer region containing the T2D signals.

In animal studies, context-dependent knockouts can provide improved spatial and temporal resolution for targeting candidate causal genes. Even so, the disease relevance of the observed phenotypes is determined by the confidence with which the target tissue of the risk allele is known. One example is provided by an intronic T2D signal at the *TCF7L2* locus, which has been the focus of conflicting observations. Tissue-specific knockout studies have demonstrated primary roles of *TCF7L2* in a number of different tissues, including liver and islets, whereas the non-coding GWAS signal has been consistently associated with a relatively narrow insulin secretion defect. Genomic annotations provide a clue as to the underlying reason, with the risk variants being located in an islet-specific region of open chromatin (8, 10, 27). Furthermore, the region has chromatin marks indicative of regulatory activity in islets, but not in a wide range of other tissues (10). The annotations can, thus, be applied as a filter to exclude non-disease-relevant tissues, and guide efforts to study the effect of the risk allele in the most appropriate context.

Coding GWAS variants can also produce context-dependent mechanisms through the restricted expression of gene isoforms. A striking example of this is provided by a coding variant identified in the *TBC1D4* gene in a small founder population of Greenlandic Inuit (28). The risk allele, which produces a truncated transcript that results in nonsense-mediated decay, is positioned in an exon excluded from the short isoform of the transcript. Unlike the widely expressed short isoform, the long form is predominantly expressed in skeletal muscle (29). The decreased insulin sensitivity resulting from reduced expression of *TBC1D4* is, therefore, selectively imposed on muscle tissue. As a result, the risk variant has a different effect on fasting glucose from that observed in individuals carrying LOF mutations affecting both isoforms (28).

Similar to alternative spliceforms, gene homologs can contribute to concealing the primary effect of a GWAS signal in a specific context. At the *ADCY5* locus, T2D risk alleles have been linked to both decreased islet expression of the *ADCY5* gene and β-cell dysfunction, though the underlying molecular mechanism remains unclear (12, 15). Expression studies in rodents have shown *Adcy5* to be nearly undetectable compared with the closely related homolog *Adcy6* (30, 31). By contrast, human islets show roughly equal expression of these orthologous genes, hinting at a non-conserved function of *ADCY5* between the species (30, 32). It also highlights an underlying species-specificity that makes rodents less well suited as models for mechanistic studies.

LIMITATIONS OF TRADITIONAL APPROACHES

The examples above demonstrate that the molecular phenotypes of GWAS signals can be modulated by a multitude of context-dependent factors. In the case of non-coding variants, tissue- or developmentally restricted activity of the surrounding chromatin can limit effects on gene expression. For coding

variants, alternative splicing or expression of homologs can mask a broader phenotype to produce context-dependent effects. These insights have important implications for how we design studies to translate genetic signals into molecular mechanisms.

Traditionally, particular genes have been selected for follow-up studies based on a combination of known candidate-gene biology and proximity to the GWAS signal (**Figure 1**, *left*). For whole-organism gene knockouts, this could involve engineering

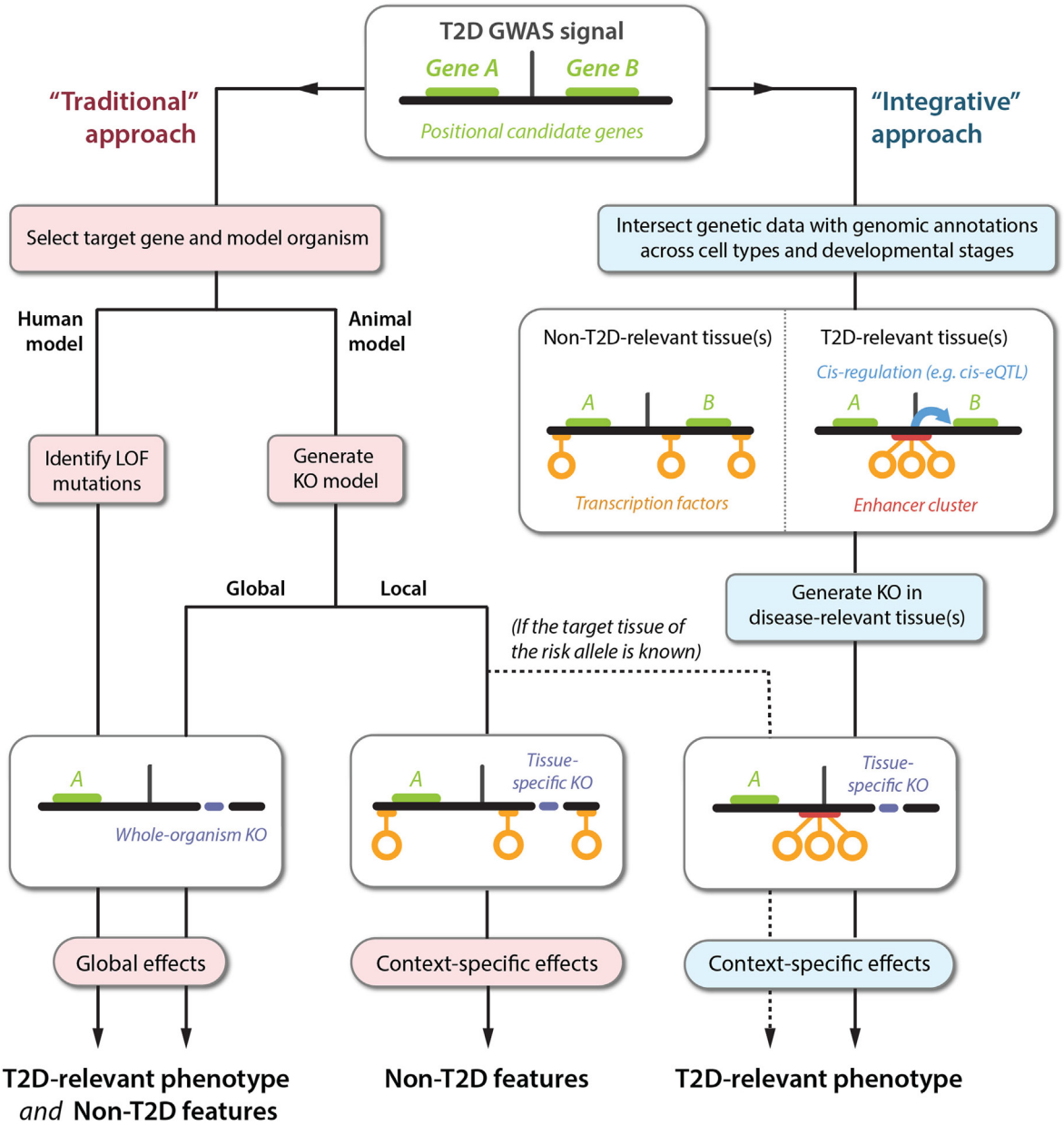


FIGURE 1 | Translating GWAS signals into disease mechanisms using traditional and integrative approaches. The top panel shows a schematic representation of a GWAS locus with a non-coding association signal for T2D risk located near two genes, *A* and *B*. The left-hand side represents the traditional approach followed for elucidating the causal mechanisms leading to a T2D phenotype. This approach can produce a range of phenotypes that are difficult to translate into causal GWAS mechanisms, because the selected genes are often knocked out globally (bottom left panel), or in a tissue-specific context that is irrelevant to understanding the effects of the T2D susceptibility phenotype (bottom center panel). Sequencing approaches can also be used to identify gain-of-function mutations (not shown), which are subject to the same limitations. The right-hand side outlines an alternative approach using integration of emerging datasets to produce directed hypotheses. For optimal resolution, several types of genetic and genomic datasets can be integrated, including TF binding, enhancer regions (defined based on chromatin state), and cis-regulatory relationships (e.g., identified by cis-eQTL or chromatin-conformation capture studies). In this example, the GWAS signal is observed to disrupt an enhancer cluster with cis-regulatory activity toward gene *B*. Importantly, the disruption exclusively affects enhancer activity in disease-relevant tissues. This proposes a follow-up experiment for manipulation of gene *B* in a specific context (bottom right panel), producing a phenotype that is likely to be directly relevant to understanding the molecular basis of T2D susceptibility.

of animal models or using genetic testing to identify individuals carrying LOF variants. As discussed, the relevance of these approaches for delineating disease-relevant mechanisms is limited by the potential for unspecific global phenotypes. Tissue-specific knockouts provide higher spatiotemporal resolution but require the target tissue(s) of the GWAS signal to be known. Human physiological associations can narrow down the list of likely relevant tissues, but these measures are often too crude to pinpoint specific cell types.

In principle, using animal models to target non-coding regions could produce more precise disease models, but this strategy is constrained by the low conservation at the level of regulatory architecture. In a study of the *Cdkn2a* locus in mice, targeted deletion of a 70 kb non-coding interval established enhancer activity toward nearby genes, but the relevance to humans has been questioned by subsequent findings (33, 34). The region encodes a long non-coding RNA that has no clear ortholog in rodents, highlighting the possibility of divergent cis-regulatory mechanisms. More generally, TF binding sites have been shown to diverge even faster than the underlying sequence itself (35). For two key liver TFs, the majority of binding events were shown to be species specific, while only 10–30% of hepatic enhancer clusters have corresponding rodent orthologs (36, 37). Although subsets of conserved clusters may aid in the prioritization of causal variants, these observations suggest that a different approach is required to delineate GWAS mechanisms (37, 38).

Even in those cases where animal models do provide targeted gene manipulation in an appropriate context (whether through tissue-specific knockout or transcriptional dysregulation), the resulting phenotypes may not be directly relevant for understanding human disease. Though rodent models continue to be an important tool for studying type 2 diabetes pathogenesis, it has become increasingly clear that murine pancreatic islets differ in a number of ways from their human counterparts (39–41). Certain monogenic forms of diabetes are, therefore, not well recapitulated in rodents, and molecular mechanisms elucidated in animal models should be interpreted with caution (42).

TOWARD AN INTEGRATED UNDERSTANDING OF GWAS SIGNALS

To successfully study disease-relevant phenotypes, experimental designs can be guided by the integration of genetic association data and genomic annotations (**Figure 1**, *right*). At the core of this framework is the overlaying of a static dataset – a list of variants linked to disease susceptibility and/or physiological traits – with layers of highly dynamic functional information that provide spatial and temporal dimensions. These layers encompass diverse datasets, and include information centered on single variants, such as histone marks and TF binding sites, and higher-order information that signifies relationships between distinct elements, such as chromatin interactions and cis-eQTL data.

Genomic annotations provide a cellular atlas with which to navigate and interpret genetic data in the context of specific cell types and developmental stages. For instance, if a set of likely causal variants has been identified from GWAS or fine-mapping studies, the tissue of action may be inferred from comparing chromatin states across cell types (8–11, 43–46). Conversely, if the target tissue is known from physiological associations, the causal mechanism can be pinpointed by overlaying with relevant functional annotations (14, 22, 47–51). This process generates a plethora of directed hypotheses that can be followed up with specific functional experiments. Increasingly, such studies are likely to be focused on differentiated cells derived from human stem cells, which can provide disease models and chromatin maps that are both functionally and developmentally relevant.

The broader applicability of this approach is, in part, determined by the tractability of individual loci. For regions with extensive linkage disequilibrium, the arising complexity can hinder experimental follow-up. Starting from a limited set of credible variants is, thus, essential. More generally, the value in taking an integrative approach is dictated by the extent to which genomic annotations for disease-relevant tissues have been made available (or can be obtained). The construction of a truly integrated framework is an incremental and monumental effort, facilitated by the Encyclopedia of DNA Elements (ENCODE) and the NIH Roadmap Epigenomics project, which together cover hundreds of tissues and epigenetic annotations. Since each dataset is merely a snapshot of a given cell type in a particular metabolic and developmental state, this on-going process will continue to produce an atlas with ever-finer spatiotemporal resolution. For many hard-to-obtain organs, such as pancreatic islets, power to detect relevant genomic features is still limiting.

Even so, chromatin landscapes for tissues relevant to T2D have begun to emerge in recent years, enabling biological inferences to be made. As we have seen, this has successfully uncovered tissue- and species-specific effects of T2D risk variants. The insights have also provided compelling evidence to demonstrate that context-dependent phenotypes are not the exception but, in fact, a fundamental aspect of GWAS biology. In the coming years, we need to build on this paradigm to accelerate the translation of genetic findings into molecular mechanisms.

AUTHOR CONTRIBUTIONS

ST, MM, and AG wrote, edited, and approved the manuscript.

FUNDING

ST is a Radcliffe Department of Medicine scholar, MM is a Wellcome Trust Senior Investigator, and AG is a Wellcome Trust Senior fellow in Basic Biomedical Science (095101).

REFERENCES

1. 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. doi:10.1038/ng.609
2. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segrè 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. doi:10.1038/ng.2383
3. DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium, Asian Genetic Epidemiology Network Type 2 Diabetes (AGEN-T2D) Consortium, South Asian Type 2 Diabetes (SAT2D) Consortium, Mexican American Type 2 Diabetes (MAT2D) Consortium, Type 2 Diabetes Genetic Exploration by Nex-generation sequencing in muylti-Ethnic Samples (T2D-GENES) Consortium, Mahajan A, et al. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat Genet* (2014) 46:234–44. doi:10.1038/ng.2897
4. Ward LD, Kellis M. Interpreting noncoding genetic variation in complex traits and human disease. *Nat Biotechnol* (2012) 30:1095–106. doi:10.1038/nbt.2422
5. Kasowski M, Kyriazopoulou-Panagiotopoulou S, Grubert F, Zaugg JB, Kundaje A, Liu Y, et al. Extensive variation in chromatin states across humans. *Science* (2013) 342:750–2. doi:10.1126/science.1242510
6. McVicker G, van de Geijn B, Degner JF, Cain CE, Banovich NE, Raj A, et al. Identification of genetic variants that affect histone modifications in human cells. *Science* (2013) 342:747–9. doi:10.1126/science.1242429
7. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* (2013) 153:307–19. doi:10.1016/j.cell.2013.03.035
8. Gaulton KJ, Nammo T, Pasquali L, Simon JM, Giresi PG, Fogarty MP, et al. A map of open chromatin in human pancreatic islets. *Nat Genet* (2010) 42:255–9. doi:10.1038/ng.530
9. Parker SC, Stitzel ML, Taylor DL, Orozco JM, Erdos MR, Akiyama JA, et al. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc Natl Acad Sci U S A* (2013) 110:17921–6. doi:10.1073/pnas.1317023110
10. Pasquali L, Gaulton KJ, Rodríguez-Seguí SA, Mularoni L, Miguel-Escalada I, Akerman I, et al. Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat Genet* (2014) 46:136–43. doi:10.1038/ng.2870
11. Trynka G, Sandor C, Han B, Xu H, Stranger BE, Liu XS, et al. Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat Genet* (2013) 45:124–30. doi:10.1038/ng.2504
12. Dimas AS, Lagou V, Barker A, Knowles JW, Mägi R, Hivert MF, et al. Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. *Diabetes* (2014) 63:2158–71. doi:10.2337/db13-0949
13. Torres JM, Gamazon ER, Parra EJ, Below JE, Valladares-Salgado A, Wacher N, et al. Cross-tissue and tissue-specific eQTLs: partitioning the heritability of a complex trait. *Am J Hum Genet* (2014) 95:521–34. doi:10.1016/j.ajhg.2014.10.001
14. 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. doi:10.1038/ng.3437
15. 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:e1005694. doi:10.1371/journal.pgen.1005694
16. Bouatia-Naji N, Bonnefond A, Cavalcanti-Proença C, Sparsø T, Holmkvist 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. doi:10.1038/ng.277
17. Prokopenko I, Langenberg C, Florez JC, Saxena R, Soranzo N, Thorleifsson G, et al. Variants in MTNR1B influence fasting glucose levels. *Nat Genet* (2009) 41:77–81. doi:10.1038/ng.290
18. Fadista J, Vikman P, Laakso EO, Mollet IG, Esguerra JL, Taneera J, et al. 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. doi:10.1073/pnas.1402665111
19. Tuomi T, Nagorny CL, Singh P, Bennet H, Yu Q, Alenkvist I, et al. Increased melatonin signaling is a risk factor for type 2 diabetes. *Cell Metab* (2016) 23:1067–77. doi:10.1016/j.cmet.2016.04.009
20. Bonnefond A, Clément N, Fawcett K, Yengo L, Vaillant E, Guillaume JL, et al. Rare MTNR1B variants impairing melatonin receptor 1B function contribute to type 2 diabetes. *Nat Genet* (2012) 44:297–301. doi:10.1038/ng.1053
21. Lango Allen H, Flanagan SE, Shaw-Smith C, De Franco E, Akerman I, Caswell R, et al. GATA6 haploinsufficiency causes pancreatic agenesis in humans. *Nat Genet* (2012) 44:20–2. doi:10.1038/ng.1035
22. Weedon MN, Cebola I, Patch AM, 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. doi:10.1038/ng.2826
23. Sellick GS, Barker KT, Stolte-Dijkstra I, Fleischmann C, Coleman RJ, Garrett C, et al. Mutations in PTF1A cause pancreatic and cerebellar agenesis. *Nat Genet* (2004) 36:1301–5. doi:10.1038/ng1475
24. Grarup N, Rose CS, Andersson EA, Andersen G, Nielsen AL, Albrechtsen A, et al. Studies of association of variants near the HHEX, CDKN2A/B, and IGF2BP2 genes with type 2 diabetes and impaired insulin release in 10,705 Danish subjects: validation and extension of genome-wide association studies. *Diabetes* (2007) 56:3105–11. doi:10.2337/db07-0856
25. Krishnamurthy J, Ramsey MR, Ligon KL, Torrice C, Koh A, Bonner-Weir S, et al. p16INK4a induces an age-dependent decline in islet regenerative potential. *Nature* (2006) 443:453–7. doi:10.1038/nature05092
26. Pal A, Potjer TP, Thomsen SK, Ng HJ, Barrett A, Scharfmann R, et al. Loss-of-function mutations in the cell-cycle control gene CDKN2A impact on glucose homeostasis in humans. *Diabetes* (2016) 65:527–33. doi:10.2337/db15-0602
27. Stitzel ML, Sethupathy P, Pearson DS, Chines PS, Song L, Erdos MR, et al. Global epigenomic analysis of primary human pancreatic islets provides insights into type 2 diabetes susceptibility loci. *Cell Metab* (2010) 12:443–55. doi:10.1016/j.cmet.2010.09.012
28. 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. doi:10.1038/nature13425
29. Baus D, Heermeier K, De Hoop M, Metz-Weidmann C, Gassenhuber J, Dittrich W, et al. Identification of a novel AS160 splice variant that regulates GLUT4 translocation and glucose-uptake in rat muscle cells. *Cell Signal* (2008) 20:2237–46. doi:10.1016/j.cellsig.2008.08.010
30. Hodson DJ, Mitchell RK, Marselli L, Pullen TJ, Gimeno Brias S, Semplici F, et al. ADCY5 couples glucose to insulin secretion in human islets. *Diabetes* (2014) 63:3009–21. doi:10.2337/db13-1607
31. Parton LE, McMillen PJ, Shen Y, Docherty E, Sharpe E, Diraison F, et al. Limited role for SREBP-1c in defective glucose-induced insulin secretion from Zucker diabetic fatty rat islets: a functional and gene profiling analysis. *Am J Physiol Endocrinol Metab* (2006) 291:E982–94. doi:10.1152/ajpendo.00067.2006
32. Eizirik DL, Sammeth M, Bouckenooghe T, Bottu G, Sisino G, Igoillo-Esteve M, et al. The human pancreatic islet transcriptome: expression of candidate genes for type 1 diabetes and the impact of pro-inflammatory cytokines. *PLoS Genet* (2012) 8(3):e1002552. doi:10.1371/journal.pgen.1002552
33. Visel A, Zhu Y, May D, Afzal V, Gong E, Attanasio C, et al. Targeted deletion of the 9p21 non-coding coronary artery disease risk interval in mice. *Nature* (2010) 464:409–12. doi:10.1038/nature08801
34. Congrains A, Kamide K, Ohishi M, Rakugi H. ANRIL: molecular mechanisms and implications in human health. *Int J Mol Sci* (2013) 14:1278–92. doi:10.3390/ijms14011278
35. Stefflova K, Thybert D, Wilson MD, Streeter I, Aleksic J, Karagianni P, et al. Cooperativity and rapid evolution of cobound transcription factors in closely related mammals. *Cell* (2013) 154:530–40. doi:10.1016/j.cell.2013.07.007
36. Schmidt D, Wilson MD, Ballester B, Schwalie PC, Brown GD, Marshall A, et al. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* (2010) 328:1036–40. doi:10.1126/science.1186176
37. Ballester B, Medina-Rivera A, Schmidt D, Gonzàlez-Porta M, Carlucci M, Chen X, et al. Multi-species, multi-transcription factor binding highlights conserved control of tissue-specific biological pathways. *Elife* (2014) 3:e02626. doi:10.7554/eLife.02626
38. Claussnitzer M, Dankel SN, Klocke B, Grallert H, Glunk V, Berulava T, et al. Leveraging cross-species transcription factor binding site patterns: from diabetes risk loci to disease mechanisms. *Cell* (2014) 156:343–58. doi:10.1016/j.cell.2013.10.058
39. Ashcroft FM, Rorsman P. Diabetes mellitus and the beta cell: the last ten years. *Cell* (2012) 148:1160–71. doi:10.1016/j.cell.2012.02.010

40. Ravassard P, Hazhouz Y, Pechberty S, Bricout-Neveu E, Armanet M, Czernichow P, et al. A genetically engineered human pancreatic beta cell line exhibiting glucose-inducible insulin secretion. *J Clin Invest* (2011) 121:3589–97. doi:10.1172/JCI58447
41. Cabrera O, Berman DM, Kenyon NS, Ricordi C, Berggren PO, Caicedo A. The unique cytoarchitecture of human pancreatic islets has implications for islet cell function. *Proc Natl Acad Sci U S A* (2006) 103:2334–9. doi:10.1073/pnas.0510790103
42. Thomsen SK, Gloyn AL. The pancreatic beta cell: recent insights from human genetics. *Trends Endocrinol Metab* (2014) 25:425–34. doi:10.1016/j.tem.2014.05.001
43. Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* (2014) 506:376–81. doi:10.1038/nature12873
44. 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. doi:10.1038/nature14248
45. Onengut-Gumuscu S, Chen WM, Burren O, Cooper NJ, Quinlan AR, Mychaleckyj JC, et al. Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat Genet* (2015) 47:381–6. doi:10.1038/ng.3245
46. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh PR, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* (2015) 47:1228–35. doi:10.1038/ng.3404
47. Small KS, Hedman AK, Grundberg E, Nica AC, Thorleifsson G, Kong A, et al. Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes. *Nat Genet* (2011) 43:561–4. doi:10.1038/ng.833
48. Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs KV, et al. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature* (2010) 466:714–9. doi:10.1038/nature09266
49. Adrianto I, Wen F, Templeton A, Wiley G, King JB, Lessard CJ, et al. Association of a functional variant downstream of TNFAIP3 with systemic lupus erythematosus. *Nat Genet* (2011) 43:253–8. doi:10.1038/ng.766
50. Fairfax BP, Makino S, Radhakrishnan J, Plant K, Leslie S, Dilthey A, et al. Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nat Genet* (2012) 44:502–10. doi:10.1038/ng.2205
51. Farh KK, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* (2015) 518:337–43. doi:10.1038/nature13835

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer LL declared a past co-authorship with one of the authors MM to the handling Editor, who ensured that the process met the standards of a fair and objective review.

Copyright © 2016 Thomsen, McCarthy and Gloyn. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Diabetologia (2017) 60:960–970
DOI 10.1007/s00125-017-4270-y

REVIEW

Human genetics as a model for target validation: finding new therapies for diabetes

Soren K. Thomsen [1] · **Anna L. Gloyn** [1,2,3]

Received: 16 December 2016 / Accepted: 14 March 2017 / Published online: 26 April 2017
© The Author(s) 2017. This article is an open access publication

Abstract Type 2 diabetes is a global epidemic with major effects on healthcare expenditure and quality of life. Currently available treatments are inadequate for the prevention of comorbidities, yet progress towards new therapies remains slow. A major barrier is the insufficiency of traditional preclinical models for predicting drug efficacy and safety. Human genetics offers a complementary model to assess causal mechanisms for target validation. Genetic perturbations are 'experiments of nature' that provide a uniquely relevant window into the long-term effects of modulating specific targets. Here, we show that genetic discoveries over the past decades have accurately predicted (now known) therapeutic mechanisms for type 2 diabetes. These findings highlight the potential for use of human genetic variation for prospective target validation, and establish a framework for future applications. Studies into rare, monogenic forms of diabetes have also provided proof-of-principle for precision medicine, and the applicability of this paradigm to complex disease is discussed. Finally, we highlight some of the limitations that are relevant to the use of genome-wide association studies (GWAS) in the search for new therapies for diabetes. A key outstanding challenge is the translation of GWAS signals into disease biology and we outline possible solutions for tackling this experimental bottleneck.

Keywords Adverse effects · Genome-wide association studies · Human genetics · Monogenic diabetes · Precision medicine · Review · Target discovery · Target validation · Therapeutic mechanisms · Type 2 diabetes

Electronic supplementary material The online version of this article (doi:10.1007/s00125-017-4270-y) contains a slideset of the figures for download, which is available to authorised users.

✉ Anna L. Gloyn
anna.gloyn@drl.ox.ac.uk

[1] Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Churchill Hospital, Old Road, Headington, Oxford OX3 7LE, 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

Abbreviations

Cis-eQTL	Cis-expression quantitative trait locus
FOXA2	Forkhead box protein A2
GKRP	Glucokinase regulatory protein
GWAS	Genome-wide association study
K_{ATP}	ATP-sensitive potassium channel
MTRN1B	Melatonin receptor 1B
NEUROD1	Neurogenic differentiation 1
PheWAS	Phenome-wide association study

PNDM	Permanent neonatal diabetes mellitus
PPARγ	Peroxisome proliferator-activated receptor γ
SGLT2	Sodium–glucose cotransporter 2
TZD	Thiazolidinediones
ZnT8	Zinc transporter 8

Introduction

Long-term complications of type 2 diabetes and its related disorders present a major, growing socioeconomic burden to society [1]. Despite incremental advances in the development of therapies for diabetes, current treatments fail to provide adequate glucose control for most patients. Much-needed efforts to develop novel, first-in-class drugs are hampered by the slow rate and escalating costs of research and development programmes in the pharmaceutical industry. The staggering estimated price tag for an average new drug is approaching $3 billion, with the high attrition rate in clinical trials (>80%) imposing a cumulatively higher cost on those drugs that do make it to market [2, 3]. The most common reasons for failure include lack of efficacy and/or unsuitable safety profiles, even in cases where the correct target is engaged [4, 5]. Clearly, these observations attest to the limitations of existing preclinical models in evaluating therapeutic candidates before committing to expensive human studies.

Over the past decades, technological advances have unlocked the possibility of using human genetics as a complimentary strategy for preclinical target validation. Genetic variation offers valuable insights into the effects of manipulating specific proteins or pathways in a system that is directly relevant to human disease. Such 'experiments of nature' can, in principle, inform target identification, predict potential adverse long-term effects and identify suitable indications for treatment. In this review, we first discuss evidence for the benefits provided by human genetics for the treatment of diabetes within each of these three domains. Second, we focus on one of the key challenges facing this paradigm, specifically the identification of causal mechanisms for genetic variants, and provide examples of potential solutions.

Glossary

Cis-eQTL: a genomic region containing variant(s) that influence gene expression in one or more cell types. A cis-eQTL exerts this regulatory function locally, affecting nearby target gene(s) in an allele-specific manner

Druggable: a biological target that binds to (or may bind to) a drug with high affinity

'Experiments of nature': an inherited genetic perturbation; it is constant from birth, preceding the disease state, and may therefore be used to elucidate the long-term impact of modulating specific targets in vivo, the mechanisms underlying these effects and therapeutic strategies for disease

Gene/protein enrichment: a method for identification of biological features that are statistically overrepresented in a set of genes or proteins of interest

GWAS: phenotype-to-genotype approach to elucidate associations between a genome-wide set of variants and specific diseases/traits

Mendelian randomisation: a tool for discovery of causal relationships between environmental exposures and disease status

PheWAS: genotype-to-phenotype approach for identification of diseases/traits associated with a particular genetic variant (or multiple variants within a gene of interest)

Target discovery: identification of drug targets i.e. molecules or molecular interactions that play an important role in disease pathways and may be altered by therapeutic agents

'Experiments of nature' in drug target discovery

Nomination of a therapeutic candidate for clinical testing is based on the expectation that modulating a specific target will result in a net benefit for patients, taking into account both desired and adverse effects. Supporting evidence is usually

derived from extensive preclinical testing, including studies in animal and cellular models, as well as data from observational epidemiology [6]. Importantly, these sources are often unsuitable for establishing definitive proof of causality in humans, with the specific models either lacking direct relevance or being unable to confidently distinguish cause and effect [7]. In both of these areas, human genetics can complement existing lines of evidence with a relevant window into the chain of causality. A genetic perturbation (an inherited mutation in or near an affected protein) is constant from birth and thus precedes the disease state rather than being affected by the disease environment. Unlike other molecular phenotypes, such as metabolite or protein levels, genetic variants are, therefore, not prone to be confounded by reverse causation. Moreover, in well-mixed populations, genotypes are randomly assigned at conception, thus acting, in effect, as natural versions of a randomised controlled trial [8]. Taking advantage of these properties, genetic epidemiology can intersect with other types of preclinical evidence to establish a powerful framework for target discovery and validation (Fig. 1).

Though experiments of nature can be highly relevant tools for prioritising drug targets, a number of challenges limit the applicability of this strategy. First, many disease-causing variants are located in regions of the genome that do not code for proteins, making identification of causal mechanisms a substantial challenge (this is discussed in more detail later) [9]. Further, the impact of individual mutations on protein activity often cannot be determined a priori, which necessitates extensive in vitro work to establish both statistical associations and directions of effect [10]. Once causal mechanisms have been established, these insights can then be used to focus research and development efforts on producing a desirable therapeutic effect. For instance, if loss-of-function variants in a particular gene are linked to protective effects, modulating the encoded protein by antagonists would be an attractive target (and often a more tractable goal for medicinal chemists than protein activation).

Ultimately, genetic epidemiology is reliant upon the occurrence of natural variation (either single-nucleotide and/or structural variants) in genomic regions relevant to disease. Since the penetrance of disease-associated variants tends to be inversely related to their frequency in the population (because of negative selection), most common variants have small effect sizes, while more deleterious mutations are rare [11–13]. As a result, sequencing of large-scale case–control cohorts is required to establish associations across the frequency spectrum [12, 14]. Phenotypic selection and familial studies (e.g. focusing on extreme forms of diabetes) can enrich for high-penetrance mutations, but assigning pathogenicity and estimating effect sizes from such studies can be a challenge [10, 15, 16]. In the ideal scenario, multiple independent alleles with different degrees of effect on a phenotype can be used to calibrate a genetic 'dose–response' curve [7, 17]. Not only does this build confidence in a specific target but the allelic series can also be used to predict the magnitude of effect required to produce a therapeutic response in vivo (Fig. 1). It is worth noting, however, that not all mutations produce effects that can be linearly mapped to a simple dose–response curve. More subtle perturbations could, for example, lead to aspects of both gain- and loss-of-function.

Genetic studies for the validation of drug mechanisms Over the past 10 years, genome-wide association studies (GWAS) have made significant progress towards mapping the genetic heritability of many complex diseases [18]. In the context of drug development, however, these advances are, broadly speaking, too recent to assess their impact on prospective target discovery. Still, attempts can be made to validate the use of GWAS associations for target discovery based on existing drugs. One study looked at the proportion of drug mechanisms (defined as a target paired with its approved indication) supported by genetic evidence at the various stages of the drug development pipeline [19]. This proportion was found to increase from 2% at the preclinical stage to 8.2% for approved drugs, with the single largest increase occurring between phase II and phase III of clinical trials. Remarkably, based on such historical data, it was estimated that the rate of success was around twofold higher for a target–indication pair supported by GWAS or other human genetic data compared with pairs with no support. While any retrospective study will certainly have limitations (e.g. successful drug mechanisms might spur genetic research into particular targets), two observations support the overall conclusion of this study: first, the same correlation was found for GWAS data alone (success ratio: 1.8 for supported vs unsupported mechanisms), which is unlikely to be influenced by known biological mechanisms [19]; second, most potential confounding factors (e.g. unknown causal mechanisms for GWAS signals and incomplete mapping of genetic heritability) actually tend to bias observations towards the null hypothesis.

Thiazolidinediones For the treatment of type 2 diabetes, thiazolidinediones (TZDs) provide an instructive example of

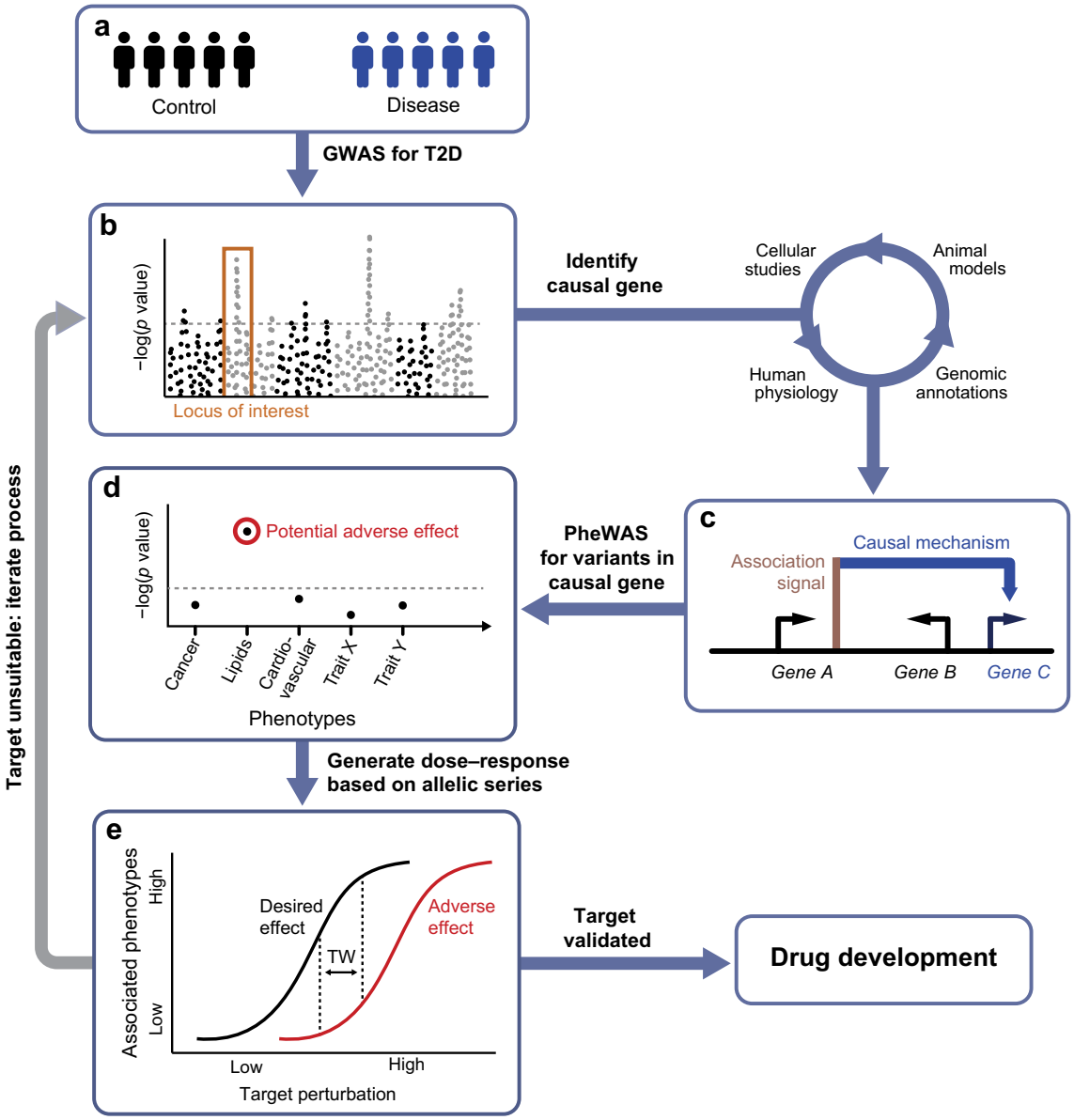


Fig. 1 Using human genetics as a model for drug target validation. GWAS into the heritability of type 2 diabetes (T2D) have identified a large number of variants that are robustly associated with disease risk (**a**, **b**). Nevertheless, establishing the underlying causal mechanisms has proven to be a major experimental bottleneck. The process usually involves an array of approaches, including in vitro and in vivo studies in animal and cellular models, as well as genetic and physiological follow-up studies of risk-allele carriers. Once a causal gene has been identified (**c**), the encoded protein may be taken forward for further validation as a potential drug target. Genetic alleles within the causal gene can be interrogated for links to other phenotypes using PheWAS, which can highlight likely adverse or beneficial effects of long-term treatment (**d**). For candidate genes harbouring multiple, independent alleles, effects on disease risk can be correlated against their known impact on protein function (**e**). Some perturbations, such as protein-truncating variants, have predictable effects, while most alleles require extensive experimental follow-up to reliably ascertain their functional impacts. If an allelic series has been established, their phenotypic associations can be used to generate the genetic equivalent of a dose–response curve (**e**). The therapeutic window (TW) marks the range of perturbations that produce a suitable ratio between desirable effects (i.e. type 2 diabetes protection) and adverse effects (e.g. raised lipid levels). In cases where a potential treatment is not predicted to result in a net patient benefit, the target is considered unsuitable and the process can be repeated for a different candidate. However, if an appropriate TW has been identified, the target can be taken forward for drug development on the basis of this human genetic validation

a drug mechanism that has been corroborated by genetic evidence since first being discovered. TZDs are a class of commonly used drugs that act primarily through activation of the peroxisome proliferator-activated receptor γ (PPARγ) to improve insulin sensitivity [20]. Within a few years after obtaining market approval in 1996, the gene encoding PPARγ (*PPARG*) was found to contain a missense variant (Pro12Ala) that associates with type 2 diabetes susceptibility [21–23]. Though the functional impact of this common variant remains uncertain, the subsequent discovery of rare, loss-of-

function variants associated with disease risk have established direction-of-effect at this locus [24]. Genetic evidence thus points to a therapeutic benefit of PPARγ agonists, fully consistent with the clinical effects observed from TZDs. The Pro12Ala association has also since been replicated by GWAS for type 2 diabetes risk, despite the relatively small effect size of the risk allele (OR 1.16) [25]. As illustrated by this case, the measured effect size of a single genetic variant is not necessarily a useful predictor of therapeutic opportunities.

Sulfonylureas This lesson is further reinforced by insights from genetic studies on the ATP-sensitive potassium channel (K_{ATP}), which couples glucose metabolism to insulin secretion in pancreatic beta cells. As early as 1942, sulfonylureas inhibiting the channel were found to display hypoglycaemic effects in animal studies [26]. Around 60 years later, genetic studies in humans identified a type 2 diabetes association signal that overlaps two genes, *KCNJ11* and *ABCC8*, which encode subunits of the K_{ATP} (OR 1.1–1.2) [27–29]. Subsequent molecular studies have confirmed the risk haplotype to produce a channel that is less sensitive to ATP inhibition, thus reducing insulin secretion [30]. In contrast, sulfonylureas promote closure of the channel to depolarise the beta cell and mobilise insulin granules [31]. Thus, these findings demonstrate how genetic discovery can successfully predict the therapeutic potential of a known target based on genetic variants with moderate effect.

ZnT8 modulation Although no validated drug targets have emerged from type 2 diabetes GWAS to date, recently identified coding variants have highlighted plausible candidates. One candidate that has been the focus of particular interest is the *SLC30A8* gene, encoding the zinc transporter 8 (ZnT8), which is expressed in insulin secretory granules. Initially, common risk variants of unknown functional importance had spurred commercial interest in the development of agonists, based on the assumption of a negative correlation between activity levels of ZnT8 and diabetes risk [32, 33]. This notion was challenged by a more recent study that focused on protein-truncating variants in *SLC30A8* to determine the effect of loss-of-function on type 2 diabetes susceptibility [34]. Strikingly, the study found that carriers that were haploinsufficient for ZnT8 were protected from type 2 diabetes, with a 65% reduction in disease risk. These observations provide strong evidence in favour of a therapeutic strategy based on ZnT8 inhibition. More broadly, this example also illustrates the value in considering use of an extended allelic series to more fully explore the effects of target modulation at various levels of inhibition and/or activation.

Predicting adverse effects of new therapies

The suitability of a drug candidate is ultimately dependent on whether the therapeutic effect is expected to outweigh any on- and off-target adverse effects. These can be difficult to predict, especially if caused by unintended drug promiscuity, but attempts can be made to anticipate on-target side effects. In an analogous fashion to the use of GWAS for target identification, phenome-wide association studies (PheWAS) provide a tool for determining the long-term consequences of manipulating a target [35]. Rather than seeking to identify variants associated with a particular disease, a PheWAS is designed to systematically identify the diseases or traits associated with a particular variant (or multiple variants within a gene of interest). In the same way that genetic perturbations can pinpoint target proteins, the detected phenotypes are a consequence of life-long experiments of nature. Any pleiotropy thus raises the possibility of additional on-target effects from long-term target modulation (Fig. 1).

As for GWAS, there are a number of practical and conceptual limitations that apply to the PheWAS paradigm in the context of target validation. First, it is clear that the identified phenotypes (both therapeutic benefits and on-target adverse effects) may be restricted to a perturbation that is imposed over many years or is present at a specific stage of disease progression. For instance, in the case of type 1 diabetes, the identified association signals have primarily uncovered genes implicated in the immune system. Nevertheless, modulating immune function in individuals with type 1 diabetes is unlikely to be an effective therapy, since autoimmune beta cell destruction has already occurred. For such diseases, the therapeutic pathways for treating symptoms (e.g. insulin or beta cell replacement for type 1 diabetes) may be different from the susceptibility pathways (uncovered by genetics) that are relevant to preventing disease. More generally, the lifetime exposure of a genetic defect might also produce long-term secondary effects (e.g. through compensatory mechanisms) that are not directly predictive of acute therapeutic interventions.

A second limitation of PheWAS is the requirement for access to diverse, deeply phenotyped cohorts or electronic medical records with genotyping information. Though population-wide biobanks and large, industry-led cohorts with sequencing data are now taking form, systematic PheWAS have previously been impractical [35]. Studies of this nature have, therefore, been more akin to traditional candidate gene association studies, focusing on a specific hypothesis concerning a target gene and a selected outcome. Nonetheless, recent examples of this approach being applied to the development of new treatments for type 2 diabetes provide insights into the potential value of PheWAS.

Glucokinase and glucokinase regulatory protein Glucokinase (encoded by *GCK*) is a key glycolytic enzyme involved in sensing the energy status of the body's major organs. The protein is regulated in the liver by glucokinase regulatory protein (GKRP; encoded by *GCKR*), which sequesters glucokinase during fasting [36]. Genetic variation in both *GCK* and *GCKR* has been implicated in type 2 diabetes susceptibility, and the proteins are both targets of ongoing drug development efforts to modulate

this pathway [37–39]. While increasing glucokinase activity (e.g. through GKRP inhibition or allosteric activation) could lower plasma glucose to reduce the risk of type 2 diabetes, genetic evidence also points to the possibility of likely adverse effects [40–42]. Several studies of deleterious variants in *GCKR* have found increased risk of hypertriacylglycerolaemia, probably as a consequence of elevated substrate availability for hepatic lipogenesis [43–46]. Interestingly, in clinical trials of one glucokinase activator, mild dyslipidaemia was reported in treatment groups, providing preliminary confirmation of this potential adverse effect [47]. Similar results were reported across different classes of glucokinase activators in rodents, arguing for an effect that is independent of the specific chemical compound [48]. In light of corroborating genetic and molecular data, it is clear that monitoring lipid levels for therapies targeting glucokinase/GKRP is essential.

Sodium–glucose cotransporter 2 In a similar way, genetic evidence has been able to shine light on the clinical use of sodium–glucose cotransporter 2 (SGLT2) inhibitors, an emerging class of glucose-lowering drugs that act through increased renal clearance of glucose [49]. A naturally occurring inhibitor of SGLT2 (phlorizin) had been known for some time, spurring the development of synthetic analogues for use in humans [50]. Nevertheless, the discovery that familial renal glycosuria is caused by genetic variants in the gene encoding SGLT2 (S*LC5A2*) provided an opportunity to test for any side effects of long-term perturbations [51, 52]. Individuals carrying loss-of-function alleles in S*LC5A2* have reduced ability to reabsorb glucose in the kidney but display otherwise normal renal function and no or few additional clinical features (www.omim.org/entry/233100, accessed 1 March 2017). These observations suggest that selective targeting of SGLT2, even for prolonged periods of time, is not associated with any significant complications.

PTEN Another rare genetic disorder, known as Cowden's syndrome, has offered new clues into a possible link between type 2 diabetes, obesity and cancer, as initially suggested by epidemiological data [53]. The majority of patients suffering from the cancer predisposition syndrome carry germline loss-of-function mutations in the *PTEN* gene [54]. The protein encoded by *PTEN* (phosphatase and tensin homolog; PTEN) is a known tumour suppressor and a critical inhibitor of the phosphatidylinositol (3,4,5)-trisphosphate (PIP3) branch of insulin signalling. On this basis, individuals would be expected to display improved insulin sensitivity, with a concomitant increase in cell growth and metabolism. Indeed, a recent study found individuals with Cowden's syndrome to be profoundly insulin sensitive, even in the face of obesity [55]. This provides a dramatic example of the sometimes overlapping effects of intracellular signalling pathways involved in the regulation of metabolic and cell cycle-related processes.

Mendelian randomisation as a tool for predicting adverse effects of risk factor modulation Among studies within genetic epidemiology, a subset are based on a particular design known as Mendelian randomisation. The aim of Mendelian randomisation studies is to establish causal relationships between an environmental exposure and disease status [56]. More specifically, genetic variants are used as proxies for a modifiable exposure, which in turn may influence the outcome phenotype. As for other genetic association studies, the Mendelian randomisation design rests on the assumptions that genetic variants are fixed in time (not prone to reverse causation) and subject to independent assortment at conception (hence, more likely to produce unbiased estimates of a causal effect). In addition, Mendelian randomisation studies require that the selected variants influence disease status exclusively through the exposure of interest, and that they are not in linkage disequilibrium with any variants that could confound results [57]. If these conditions are satisfied, the paradigm can provide a powerful tool for causal inference without many of the confounding influences of conventional observational epidemiology. Most obviously, Mendelian randomisation studies can be used to define the role of environmental influences in disease aetiology, and thereby determine behavioural or molecular traits that can be modified to minimise risk.

Within a framework of drug target validation, Mendelian randomisation can be a useful strategy for exploring possible adverse effects of a proposed treatment. Unlike conventional GWAS/PheWAS, which seek to predict the side effects of drugs that modify a particular target, the aim of Mendelian randomisation is in doing so for any intervention that targets a particular risk factor. Recently, for example, Mendelian randomisation studies were used to delineate a clinically relevant link between treatments for cardiovascular disease and type 2 diabetes risk [58, 59]. Alleles in the genes *PCSK9* and *HMGCR*, known to predispose individuals to lower plasma LDL-cholesterol, were associated with the expected protective effect against cardiovascular events, but also with an inverse effect on type 2 diabetes risk. As the variants have no (known) pleiotropic effects, the results indicate a causal role of reduced LDL-cholesterol in type 2 diabetes susceptibility (among individuals that already have impaired glucose tolerance). Thus, the findings not only have implications for our understanding of current therapies targeting *HMGCR* (statins) and *PCSK9* (proprotein convertase subtilisin/kexin type 9 [PCSK9] inhibitors) but they also show that the same undesirable effect may turn out to be a general feature of any treatment that lowers LDL-cholesterol.

Finding therapeutic indications based on pharmacogenomics and precision medicine

When balancing the expected effects of a drug candidate (both adverse and beneficial) any therapeutic hypothesis should be

formulated in the context of an intended target population. Since not all individuals will benefit equally from a given treatment, identifying the most appropriate indication is critical to success in clinical trials. Clearly, the genetic associations identified during target discovery can be used to immediately propose broad indications for a candidate drug. By extension, PheWAS data can be used to search a larger phenotype space for any association that indicates a likely therapeutic or adverse effect.

The application of this principle extends beyond novel therapeutics and could be a powerful method for repositioning existing drugs for new indications [60]. One report overlapped known drug mechanisms with GWAS associations for each target [61]. Interestingly, around 40% of the associated traits matched the corresponding drug indication (e.g. the use of statins for hypercholesterolaemia is accurately predicted by *HMGCR* variants that are associated with LDL-cholesterol). Though this type of analysis will necessarily be limited to studying drugs that have known targets harbouring genetic variants associated with disease, the substantial overlap provides a validation of the approach and adds confidence to those indications corroborated by genetic evidence [61–63]. Still, more than half of the studied targets were associated with a different GWAS trait from that suggested by the indication for the drug. Some of these are likely a consequence of methodological limitations (e.g. difficulties translating GWAS signals), but the mismatches also highlight examples with additional supporting evidence. These represent plausible drug-repositioning opportunities.

Pharmacogenomics: application in monogenic vs polygenic diabetes The indications proposed by genetic associations are generally broad phenotypic labels. Within the field of type 2 diabetes, the heterogeneous nature of the disorder is often alluded to, sometimes with the implication that genetics could be used to inform more precise diagnostic categories. If clinically meaningful subtypes did exist, such diagnostic labels could likely improve treatment efficacy. Certainly, there are individuals carrying mutations with high, if not complete penetrance in specific genes. These individuals may either suffer from a monogenic form of diabetes or exist somewhere on the spectrum between complex type 2 diabetes and a Mendelian disorder [10, 15, 64]. Since disease progression in such individuals is determined by perturbations in a very limited number of pathways, genetic testing could in theory enable precision medicine.

Proof of concept has been provided by a life-changing treatment for individuals with permanent neonatal diabetes mellitus (PNDM). Genetic studies on PNDM has led to the realisation that a subset of individuals harbour mutations in the genes encoding the K_{ATP} channel [65]. Similar to the type 2 diabetes risk haplotype at this locus, the mutations were found to promote opening of the channel, suggesting that sulfonylureas could provide a disease-modifying therapy. This was confirmed in follow-up studies that demonstrated sustained efficacy in individuals with PNDM [66, 67]. Remarkably, most participants were able to discontinue insulin treatment, switching to oral therapy with improved metabolic control. Sulfonylureas have also been successful in disease management for certain forms of MODY. It was found that individuals with MODY carrying mutations in the *HNF1A* or *HNF4A* genes are sensitive to low-dose sulfonylureas, though the mechanism is incompletely understood [68–70]. The examples above, all of which are diseases with a defined genetic aetiology, provide compelling demonstrations that taking a pharmacogenetics approach can improve quality of life.

An interesting question pertains to whether such pharmacogenomic principles can be generalised to more complex forms of diabetes. In other words, can genetic testing identify subgroups of individuals with type 2 diabetes that are more likely to benefit from particular treatments than others? This would likely be the case if the underlying reality of diabetes was a collection of distinct subtypes, each dominated by defects in different pathways. As mentioned, it is clear that some individuals with type 2 diabetes do carry genetic variants with intermediate to high effect sizes that may be suggestive of increased sensitivity to drugs targeting the particular pathways affected. However, available evidence from genetic studies has shown that such individuals are in the minority and that the bulk of the genetic susceptibility for type 2 diabetes is carried by a very large number of common variants, each with small effect sizes. Equally, non-genetic factors, though less well understood, appear to be characterised by pervasive environmental perturbations. Individual risk profiles are thus dominated by exposures that are mostly common and widely shared, arguing against a model for disease architecture based on a set of distinct pathologies.

Emerging from the notion that existing disease models may be poorly suited for our current understanding of diabetes, an alternative taxonomy has recently been proposed [13, 71]; referred to as the ‘palette’ model (as opposed to a subtype-oriented model), it posits that diabetes is caused by a large number of small perturbations (environmental and genetic) across the component pathways of disease (e.g. beta cell function, insulin sensitivity, autoimmunity). Individually, the phenotypic impact of each perturbation is limited, but in aggregate will push a person on a path away from metabolic homeostasis. By analogy to colours combined in different hues and saturation, the palette taxonomy proposes an unlimited spectrum of disease manifestations. Monogenic and autoimmune forms of diabetes are represented in the extremes of this continuum [72]. It is thus an implication of this model that the majority of individuals are not dominated by defects in single or few processes [71]. These individuals cannot meaningfully be categorised into subtypes and, thus, attempts at delivering

precision medicine will be challenging. It may be that biomarkers for specific processes can be used to glean insights into the pathways that are driving disease progression at any given time [71, 73]. As more process-modulating therapies become available, these could be used to encourage individuals along an appropriate trajectory, towards health. In the near future, however, targeting people with high-impact mutations (those at the extremes of the diabetes spectrum) are likely to be a more tractable aim for precision medicine.

Experimental challenges for drug target validation using human genetics

A key aspect of translating GWAS signals into target validation naturally centres on the identification of the causal genes (or 'effector transcripts') driving disease susceptibility (Fig. 1). Despite advances in broadly understanding molecular and regulatory mechanisms involved in type 2 diabetes pathogenesis, progress on individual loci has been slow. A minority of the >100 independent association signals for disease risk have produced a single high-confidence candidate gene. As a result, the therapeutic value of GWAS for target discovery is still limited by this experimental bottleneck, especially since follow-up studies have tended to focus on the 'low-hanging fruits' supported by existing lines of evidence [74]. In the last few years, a number of different approaches have been taken to tackle this issue, providing complementary lines of evidence to enhance our understanding of causal mechanisms. The methods used for identifying effector transcripts broadly fall into three categories:

1. **The identification of coding risk variants to directly pinpoint effector transcripts** This approach has been facilitated by a recent shift in the attention of GWAS efforts towards low frequency and rare variants with higher penetrance [13, 75]. Even in regions with existing regulatory variants, coding variants can be used to direct experimental efforts towards particular candidates. This is illustrated by the *G6PC2/ABCC11* locus, which contained two strong candidate causal genes near a non-coding association signal identified for fasting glucose (an intermediate trait for type 2 diabetes susceptibility) [76, 77]. A more recent effort to map coding variation for glycaemic control found coding variants within the *G6PC2* gene [78, 79]. Follow-up experimental studies have since explored the effect of the variants to show a functional impact on the encoded glucose-6-phosphatase subunit [78]. An added benefit of finding causal variants in coding regions is the offer of a more straight-forward interpretation for therapeutic strategies. Non-coding variation is subject to the context-dependent activity of cis-regulation and the effects can be restricted to specific tissues or developmental stages [80]. As a consequence, drugs that target the affected gene could cause unexpected adverse effects by producing a more global phenotype. Though coding variants can also be subject to such context-dependency (e.g. through tissue-specific isoforms), the effector transcripts are often affected more widely [80, 81].

2. **Integration of genetic and genomic data to establish direct links between regulatory variation, genomic annotation and regional genes** One powerful approach within this category attempts to identify variants that affect the expression level of nearby genes, so-called cis-expression quantitative trait loci (cis-eQTLs). In cases where the association signal overlaps a cis-eQTL in a disease-relevant tissue, this can reveal both the target gene and the direction of effect for the risk variant. While many cis-eQTLs are shared across tissues, others appear to show more restricted effects that are specific to one or more tissues [82]. Since physiological characterisations of carriers of type 2 diabetes risk variants have implied a central role for islet dysfunction in disease susceptibility, several cis-eQTL studies have focused on pancreatic islets [83, 84]. Though the power to detect associations has been limited by the availability of islets from donors, the approach has successfully highlighted candidate effector transcripts with previously unknown roles in disease pathogenesis [85, 86]; this is the case for the poorly characterised *ZMIZ1* gene that was identified in a recent study [86]. In vitro work subsequently confirmed a role for *ZMIZ1* in islet function following functional studies.

   Intersecting human genetics with genomic annotation can also be used to define common regulatory themes that underlie causal mechanisms at multiple loci. A recent study, for example, demonstrated an enrichment of islet and liver binding sites for the forkhead box protein A2 (FOXA2) transcription factor among type 2 diabetes association signals [87]. These results suggest a shared role of FOXA2 across a subset of risk loci and highlight the potential to identify specific causal variants. In one instance, at the *MTNR1B* locus, where the association signal has been collapsed to a single variant through fine-mapping, the FOXA2 binding event was shown to be a marker for binding of another transcription factor, neurogenic differentiation 1 (NEUROD1). It was found that the risk allele creates a NEUROD1 binding site, leading to increased expression of *MTNR1B* in beta cells. This is in line with a cis-eQTL that was previously identified for this variant in islets, and adds support to a mechanism for this non-coding risk allele being mediated via elevated melatonin receptor 1B (MTRN1B; encoded by *MTNR1B*) activity [86, 88].

   Interestingly, a different direction of effect for the *MTRN1B* gene has been proposed by coding loss-of-function variants, which have also been associated with elevated risk of type 2 diabetes [89]. One potential explanation

is suggested by the observation that the regulatory variant appears to exhibit tissue-specificity in its activity [87]. It is thus possible that the discrepancy between coding and non-coding variants could reflect differences between global and local roles of *MTRN1B*. Other explanations are possible and it remains to be seen whether increased *MTNR1B* transcript levels translate into higher protein expression. MTNR1B, a G-protein-coupled receptor, has received considerable attention as a potentially 'druggable' target. Addressing the inconsistencies in genetic data will thus provide insights into the suitability of MTNR1B as a drug target and inform any potential therapeutic strategies.

3. **Indirect prioritisation of genes based on known biology** Last among the methods for identifying causal mechanisms, a third category aims to indirectly prioritise genes based on known biology. For instance, a number of type 2 diabetes loci harbour genes implicated in monogenic forms of diabetes. Given the overlapping aetiologies between the diseases, monogenic diabetes genes can also be prioritised as causal for complex diabetes. Though this type of evidence is circumstantial, it could be a useful method for limiting the search space of genes to be studied. However, the number of loci for which current evidence favours one candidate over others is limiting and tends to be biased towards previously studied genes. One recently developed method aimed to sidestep this limitation by performing high-throughput functional characterisation of positional candidates for type 2 diabetes GWAS signals [90]. The screen successfully replicated known mechanisms of beta cell dysfunction and pointed to several unknown candidate causal genes. While any such study will be limited to a particular cell state, focusing on those tissues with high expected relevance to disease are likely to be most informative. Emerging genetic tools, such as clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR associated protein 9 (Cas9) and induced pluripotent stem cells, will make surveying a multitude of relevant phenotypes across tissue types and developmental stages an increasingly tractable goal.

Conclusions

Available models for preclinical target validation have limited ability to assess causal relationships with direct relevance to humans. Advances in genetics and genomics hold the promise to bring down the cost of industry research and development pipelines by complementing these approaches. Through genotype–phenotype associations, 'experiments of nature' can, in principle, facilitate drug target validation. It is still too early to assess the impact of GWAS findings on prospective target discovery for diabetes treatment but the genes identified to date have successfully predicted known therapeutic mechanisms. These encouraging findings suggest that translating uncharacterised loci into pathophysiological mechanisms could point to novel drug targets. Increasingly, the uncovered therapeutic mechanisms may enable modes of precision medicine in diabetes for individuals with moderate- to high-penetrance mutations. For more common forms of diabetes, the extent to which pharmacogenetics will prove a relevant paradigm is still uncertain. Recent genetic insights have argued against a subtype-oriented taxonomy of disease, and more precise indications for type 2 diabetes therapies may not be a realistic target for the near future. Even so, human genetics could pave the way for new disease-modifying treatments that can benefit both common and rare forms of diabetes.

Acknowledgements This review article was based on the Minkowski 2014 lecture delivered by ALG at the 49th EASD Annual Meeting in Vienna and consequently focuses predominantly on her contribution to the field. It has been updated by ALG and SKT to reflect additional relevant studies reported since the delivery of the lecture. ALG would like to thank her mentors, colleagues, research team and all study participants over the years for their support and guidance and the opportunity to be part of international collaborative science.

Funding Work in the authors' laboratories has been supported by the Wellcome Trust, Medical Research Council, National Institute of Health Research Oxford Biomedical Research Centre, Diabetes UK, European Union and National Institute of Diabetes and Digestive and Kidney Diseases. ALG is a Wellcome Trust Senior Fellow in Basic Biomedical Science (095101/Z/10/Z and 200837/Z/16/Z) and SKT is a Radcliffe Department of Medicine Scholar.

Duality of interest The authors declare that there is no duality of interest associated with this manuscript.

Contribution statement All authors were responsible for drafting the article and revising it critically for important intellectual content. All authors approved the version to be published.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. IDF (2015) International diabetes federation diabetes atlas, 7th edn. International Diabetes Federation, Belgium. Available from http://www.diabetesatlas.org. Accessed 1 March 2017
2. DiMasi JA, Feldman L, Seckler A, Wilson A (2010) Trends in risks associated with new drug development: success rates for investigational drugs. Clin Pharmacol Ther 87:272–277
3. DiMasi JA, Grabowski HG, Hansen RW (2016) Innovation in the pharmaceutical industry: new estimates of R&D costs. J Health Econ 47:20–33

4. Arrowsmith J, Miller P (2013) Trial watch: phase II and phase III attrition rates 2011-2012. Nat Rev Drug Discov 12:569
5. Cook D, Brown D, Alexander R et al (2014) Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework. Nat Rev Drug Discov 13:419–431
6. Wehling M (2009) Assessing the translatability of drug projects: what needs to be scored to predict success? Nat Rev Drug Discov 8:541–546
7. Plenge RM, Scolnick EM, Altshuler D (2013) Validating therapeutic targets through human genetics. Nat Rev Drug Discov 12:581–594
8. Barrett JC, Dunham I, Birney E (2015) Using human genetics to make new medicines. Nat Rev Genet 16:561–562
9. Hindorff LA, Sethupathy P, Junkins HA et al (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A 106:9362–9367
10. MacArthur DG, Manolio TA, Dimmock DP et al (2014) Guidelines for investigating causality of sequence variants in human disease. Nature 508:469–476
11. Kryukov GV, Pennacchio LA, Sunyaev SR (2007) Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. Am J Hum Genet 80:727–739
12. Goldstein DB, Allen A, Keebler J et al (2013) Sequencing studies in human genetics: design and interpretation. Nat Rev Genet 14:460–470
13. Fuchsberger C, Flannick J, Teslovich TM et al (2016) The genetic architecture of type 2 diabetes. Nature 536:41–47
14. Zuk O, Schaffner SF, Samocha K et al (2014) Searching for missing heritability: designing rare variant association studies. Proc Natl Acad Sci U S A 111:E455–E464
15. Flannick J, Beer NL, Bick AG et al (2013) Assessing the phenotypic effects in the general population of rare variants in genes for a dominant Mendelian form of diabetes. Nat Genet 45:1380–1385
16. Begg CB (2002) On the use of familial aggregation in population-based case probands for calculating penetrance. J Natl Cancer Inst 94:1221–1226
17. Zhou K, Pedersen HK, Dawed AY, Pearson ER (2016) Pharmacogenomics in diabetes mellitus: insights into drug action and drug discovery. Nat Rev Endocrinol 12:337–346
18. Price AL, Spencer CC, Donnelly P (2015) Progress and promise in understanding the genetic basis of common diseases. Proc R Soc B 282:20151684
19. Nelson MR, Tipney H, Painter JL et al (2015) The support of human genetic evidence for approved drug indications. Nat Genet 47:856–860
20. Hauner H (2002) The mode of action of thiazolidinediones. Diabetes Metab Res Rev 18(Suppl 2):S10–S15
21. Altshuler D, Hirschhorn JN, Klannemark M et al (2000) The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. Nat Genet 26:76–80
22. Deeb SS, Fajas L, Nemoto M et al (1998) A Pro12Ala substitution in PPARgamma2 associated with decreased receptor activity, lower body mass index and improved insulin sensitivity. Nat Genet 20: 284–287
23. Yen CJ, Beamer BA, Negri C et al (1997) Molecular scanning of the human peroxisome proliferator activated receptor gamma (hPPAR gamma) gene in diabetic Caucasians: identification of a Pro12Ala PPAR gamma 2 missense mutation. Biochem Biophys Res Commun 241:270–274
24. Majithia AR, Flannick J, Shahinian P et al (2014) Rare variants in PPARG with decreased activity in adipocyte differentiation are associated with increased risk of type 2 diabetes. Proc Natl Acad Sci U S A 111:13127–13132
25. Mahajan A (2014) Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nat Genet 46:234–244
26. Janbon M, Chaptal J, Vedel A, Schaap J (1942) Accidents hypoglycémiques graves par un sulfamidothiodiazol (le VK 57 ou 2254 RP). Montp Med 441:21–22
27. Gloyn AL, Weedon MN, Owen KR 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
28. Hani EH, Boutin P, Durand E et al (1998) Missense mutations in the pancreatic islet beta cell inwardly rectifying K+ channel gene (*KIR6.2/BIR*): a meta-analysis suggests a role in the polygenic basis of type II diabetes mellitus in Caucasians. Diabetologia 41:1511–1515
29. Gloyn AL, Hashim Y, Ashcroft SJ, Ashfield R, Wiltshire S, Turner RC (2001) Association studies of variants in promoter and coding regions of beta-cell ATP-sensitive K-channel genes SUR1 and Kir6.2 with type 2 diabetes mellitus (UKPDS 53). Diabet Med 18:206–212
30. Hamming KS, Soliman D, Matemisz LC et al (2009) Coexpression of the type 2 diabetes susceptibility gene variants KCNJ11 E23K and ABCC8 S1369A alter the ATP and sulfonylurea sensitivities of the ATP-sensitive K(+) channel. Diabetes 58:2419–2424
31. Proks P, Reimann F, Green N, Gribble F, Ashcroft F (2002) Sulfonylurea stimulation of insulin secretion. Diabetes 51(Suppl 3):S368–S376
32. Sladek R, Rocheleau G, Rung J et al (2007) A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 445:881–885
33. Rutter GA, Chimienti F (2015) *SLC30A8* mutations in type 2 diabetes. Diabetologia 58:31–36
34. Flannick J, Thorleifsson G, Beer NL, Jacobs SB (2014) Loss-of-function mutations in *SLC30A8* protect against type 2 diabetes. Nat Genet 46:357–363
35. Bush WS, Oetjens MT, Crawford DC (2016) Unravelling the human genome-phenome relationship using phenome-wide association studies. Nat Rev Genet 17:129–145
36. Shiota C, Coffey J, Grimsby J, Grippo JF, Magnuson MA (1999) Nuclear import of hepatic glucokinase depends upon glucokinase regulatory protein, whereas export is due to a nuclear export signal sequence in glucokinase. J Biol Chem 274:37125–37130
37. Dupuis J, Langenberg C, Prokopenko I et al (2010) New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet 42:105–116
38. Lloyd DJ, St Jean DJ Jr, Kurzeja RJ et al (2013) Antidiabetic effects of glucokinase regulatory protein small-molecule disruptors. Nature 504:437–440
39. Matschinsky FM (2009) Assessing the potential of glucokinase activators in diabetes therapy. Nat Rev Drug Discov 8:399–416
40. Saxena R, Voight BF, Lyssenko V et al (2007) Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 316:1331–1336
41. Orho-Melander M, Melander O, Guiducci C et al (2008) Common missense variant in the glucokinase regulatory protein gene is associated with increased plasma triglyceride and C-reactive protein but lower fasting glucose concentrations. Diabetes 57:3112–3121
42. Johansen CT, Wang J, Lanktree MB et al (2010) Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. Nat Genet 42:684–687
43. Beer NL, Tribble ND, McCulloch LJ et al (2009) The P446L variant in GCKR associated with fasting plasma glucose and triglyceride levels exerts its effect through increased glucokinase activity in liver. Hum Mol Genet 18:4081–4088
44. Rees MG, Wincovitch S, Schultz J et al (2012) Cellular characterisation of the *GCKR* P446L variant associated with type 2 diabetes risk. Diabetologia 55:114–122
45. Rees MG, Ng D, Ruppert S et al (2012) Correlation of rare coding variants in the gene encoding human glucokinase regulatory protein with phenotypic, cellular, and kinetic outcomes. J Clin Invest 122: 205–217

46. Rees MG, Raimondo A, Wang J et al (2014) Inheritance of rare functional GCKR variants and their contribution to triglyceride levels in families. Hum Mol Genet 23:5570–5578
47. Meininger GE, Scott R, Alba M et al (2011) Effects of MK-0941, a novel glucokinase activator, on glycemic control in insulin-treated patients with type 2 diabetes. Diabetes Care 34:2560–2566
48. De Ceuninck F, Kargar C, Ilic C et al (2013) Small molecule glucokinase activators disturb lipid homeostasis and induce fatty liver in rodents: a warning for therapeutic applications in humans. Br J Pharmacol 168:339–353
49. Chao EC (2014) SGLT-2 inhibitors: a new mechanism for glycemic control. Clin Diabetes 32:4–11
50. Ehrenkranz JR, Lewis NG, Kahn CR, Roth J (2005) Phlorizin: a review. Diabetes Metab Res Rev 21:31–38
51. Kanai Y, Lee WS, You G, Brown D, Hediger MA (1994) The human kidney low affinity Na+/glucose cotransporter SGLT2. Delineation of the major renal reabsorptive mechanism for D-glucose. J Clin Invest 93:397–404
52. van den Heuvel LP, Assink K, Willemsen M, Monnens L (2002) Autosomal recessive renal glucosuria attributable to a mutation in the sodium glucose cotransporter (SGLT2). Hum Genet 111:544–547
53. Tancredi M, Rosengren A, Svensson AM et al (2015) Excess mortality among persons with type 2 diabetes. N Engl J Med 373:1720–1732
54. Marsh DJ, Dahia PL, Caron S et al (1998) Germline PTEN mutations in Cowden syndrome-like families. J Med Genet 35:881–885
55. Pal A, Barber TM, Van de Bunt M et al (2012) PTEN mutations as a cause of constitutive insulin sensitivity and obesity. N Engl J Med 367:1002–1011
56. Smith GD, Ebrahim S (2003) 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol 32:1–22
57. Smith GD, Ebrahim S (2004) Mendelian randomization: prospects, potentials, and limitations. Int J Epidemiol 33:30–42
58. Schmidt AF, Swerdlow DI, Holmes MV et al (2016) *PCSK9* genetic variants and risk of type 2 diabetes: a mendelian randomisation study. Lancet Diabetes Endocrinol 5:97–105
59. Ference BA, Robinson JG, Brook RD et al (2016) Variation in PCSK9 and HMGCR and risk of cardiovascular disease and diabetes. N Engl J Med 375:2144–2153
60. Rastegar-Mojarad M, Ye Z, Kolesar JM, Hebbring SJ, Lin SM (2015) Opportunities for drug repositioning from phenome-wide association studies. Nat Biotechnol 33:342–345
61. Sanseau P, Agarwal P, Barnes MR et al (2012) Use of genome-wide association studies for drug repositioning. Nat Biotechnol 30:317–320
62. Sanseau P, Agarwal P, Barnes MR et al (2013) Reply to rational drug repositioning by medical genetics. Nat Biotechnol 31:1082
63. Wang ZY, Zhang HY (2013) Rational drug repositioning by medical genetics. Nat Biotechnol 31:1080–1082
64. Althari S, Gloyn AL (2015) When is it MODY? Challenges in the interpretation of sequence variants in MODY genes. Rev Diabet Stud 12:330–348
65. Gloyn AL, Pearson ER, Antcliff JF et al (2004) Activating mutations in the gene encoding the ATP-sensitive potassium-channel subunit Kir6.2 and permanent neonatal diabetes. N Engl J Med 350:1838–1849
66. Pearson ER, Flechtner I, Njolstad PR et al (2006) Switching from insulin to oral sulfonylureas in patients with diabetes due to Kir6.2 mutations. N Engl J Med 355:467–477
67. Sagen JV, Raeder H, Hathout E et al (2004) Permanent neonatal diabetes due to mutations in KCNJ11 encoding Kir6.2: patient characteristics and initial response to sulfonylurea therapy. Diabetes 53:2713–2718
68. Shepherd M, Pearson ER, Houghton J, Salt G, Ellard S, Hattersley AT (2003) No deterioration in glycemic control in HNF-1alpha maturity-onset diabetes of the young following transfer from long-term insulin to sulphonylureas. Diabetes Care 26:3191–3192
69. Pearson ER, Pruhova S, Tack CJ et al (2005) Molecular genetics and phenotypic characteristics of MODY caused by hepatocyte nuclear factor 4alpha mutations in a large European collection. Diabetologia 48:878–885
70. Pearson ER, Starkey BJ, Powell RJ, Gribble FM, Clark PM, Hattersley AT (2003) Genetic cause of hyperglycaemia and response to treatment in diabetes. Lancet 362:1275–1281
71. McCarthy MI (2017) Painting a new picture of personalised medicine for diabetes. Diabetologia 60:793–799
72. Gale EAM (2006) Declassifying diabetes. Diabetologia 49:1989–1995
73. Franks PW, McCarthy MI (2016) Exposing the exposures responsible for type 2 diabetes and obesity. Science 354:69–73
74. Thomsen SK, Gloyn AL (2014) The pancreatic beta cell: recent insights from human genetics. Trends Endocrinol Metab 25:425–434
75. Steinthorsdottir V, Thorleifsson G, Sulem P (2014) Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. Nat Genet 46:294–298
76. Bouatia-Naji N, Rocheleau G, Van Lommel L et al (2008) A polymorphism within the G6PC2 gene is associated with fasting plasma glucose levels. Science 320:1085–1088
77. Chen WM, Erdos MR, Jackson AU et al (2008) Variations in the G6PC2/ABCB11 genomic region are associated with fasting glucose levels. J Clin Invest 118:2620–2628
78. Mahajan A, Sim X, Ng HJ et al (2015) Identification and functional characterization of G6PC2 coding variants influencing glycemic traits define an effector transcript at the G6PC2-ABCB11 locus. PLoS Genet 11:e1004876
79. Wessel J, Chu AY, Willems SM et al (2015) Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility. Nat Commun 6:5897
80. Thomsen SK, McCarthy MI, Gloyn AL (2016) The importance of context: uncovering species- and tissue-specific effects of genetic risk variants for type 2 diabetes. Front Endocrinol 7:112
81. Moltke I, Grarup N, Jorgensen ME et al (2014) A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes. Nature 512:190–193
82. GTEx Consortium (2015) Human genomics. The genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348:648–660
83. Voight BF, Scott LJ, Steinthorsdottir V et al (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet 42:579–589
84. Dimas AS, Lagou V, Barker A et al (2014) Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. Diabetes 63:2158–2171
85. Fadista J, Vikman P, Laakso EO 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
86. van de Bunt M, Manning Fox JE, Dai X 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 Genet 11:e1005694
87. Gaulton KJ, Ferreira T, Lee Y et al (2015) Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. Nat Genet 47:1415–1425
88. Tuomi T, Nagorny CL, Singh P et al (2016) Increased melatonin signaling is a risk factor for type 2 diabetes. Cell Metab 23:1067–1077
89. Bonnefond A, Clement N, Fawcett K et al (2012) Rare MTNR1B variants impairing melatonin receptor 1B function contribute to type 2 diabetes. Nat Genet 44:297–301
90. Thomsen SK, Ceroni A, van de Bunt M et al (2016) Systematic functional characterization of candidate causal genes for type 2 diabetes risk variants. Diabetes 65:3805–3811

Aparna Pal,[1] Thomas P. Potjer,[2] Soren K. Thomsen,[1] Hui Jin Ng,[1] Amy Barrett,[1] Raphael Scharfmann,[3] Tim J. James,[4] D. Timothy Bishop,[5] Fredrik Karpe,[1,6] Ian F. Godsland,[7] Hans F.A. Vasen,[8] Julia Newton-Bishop,[5] Hanno Pijl,[9] Mark I. McCarthy,[1,6,10] and Anna L. Gloyn[1,6]

Loss-of-Function Mutations in the Cell-Cycle Control Gene *CDKN2A* Impact on Glucose Homeostasis in Humans

Diabetes 2016;65:527–533 | DOI: 10.2337/db15-0602

At the *CDKN2A/B* locus, three independent signals for type 2 diabetes risk are located in a noncoding region near *CDKN2A*. The disease-associated alleles have been implicated in reduced β-cell function, but the underlying mechanism remains elusive. In mice, β-cell–specific loss of *Cdkn2a* causes hyperplasia, while overexpression leads to diabetes, highlighting *CDKN2A* as a candidate effector transcript. Rare *CDKN2A* loss-of-function mutations are a cause of familial melanoma and offer the opportunity to determine the impact of *CDKN2A* haploinsufficiency on glucose homeostasis in humans. To test the hypothesis that such individuals have improved β-cell function, we performed oral and intravenous glucose tolerance tests on mutation carriers and matched control subjects. Compared with control subjects, carriers displayed increased insulin secretion, impaired insulin sensitivity, and reduced hepatic insulin clearance. These results are consistent with a model whereby *CDKN2A* loss affects a range of different tissues, including pancreatic β-cells and liver. To test for direct effects of *CDKN2A*-loss on β-cell function, we performed knockdown in a human β-cell line, EndoC-bH1. This revealed increased insulin secretion independent of proliferation. Overall, we demonstrated that *CDKN2A* is an important regulator of glucose homeostasis in humans, thus supporting its candidacy as an effector transcript for type 2 diabetes–associated alleles in the region.

Noncoding genetic signals at the *CDKN2A/B* locus have been associated with increased risk of developing type 2 diabetes (1,2). One signal is contained within a long noncoding RNA (*ANRIL*), while two distinct signals map to a region located further upstream of *CDKN2A* and *CDKN2B*. Physiological characterizations of normoglycemic carriers have demonstrated that the risk alleles are associated with reduced β-cell function, yet the underlying "effector" transcript driving these effects has not been established (3,4).

CDKN2A encodes the alternatively spliced proteins $p16^{INK4a}$ and $p14^{ARF}$, which are known tumor suppressors acting via distinct signaling pathways (5,6). $p16^{INK4a}$ is a cyclin-dependent kinase inhibitor involved in the regulation of cell-cycle progression through inhibition of CDK4 and CDK6 (7). $p14^{ARF}$, in contrast, prevents the degradation of the cell-cycle regulator p53 by forming a stable complex with Mdm2 in the nucleus (8).

Rodent studies have linked *Cdkn2a* to glucose homeostasis, pointing to the gene as a plausible candidate effector transcript at the *CDKN2A/B* locus. In a β-cell–specific knockout mouse, *Cdkn2a* deficiency was found to increase

[1]Oxford Centre for Diabetes, Endocrinology & Metabolism, University of Oxford, Oxford, U.K.
[2]Department of Clinical Genetics, Leiden University Medical Center, Leiden, the Netherlands
[3]INSERM U1016, Institut Cochin, Université Paris Descartes, Paris, France
[4]Department of Clinical Biochemistry, John Radcliffe Hospital, Oxford, U.K.
[5]Leeds Institute of Cancer and Pathology, University of Leeds, Leeds, U.K.
[6]Oxford National Institute for Health Research Biomedical Research Centre, Churchill Hospital, Oxford, U.K.
[7]Diabetes, Endocrinology and Metabolism, Department of Medicine, Imperial College London, London, U.K.
[8]Department of Gastroenterology and Hepatology, Leiden University Medical Center, Leiden, the Netherlands
[9]Leiden University Medical Centre, Department of Internal Medicine, Leiden, the Netherlands
[10]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, U.K.
Corresponding author: Anna L. Gloyn, anna.gloyn@drl.ox.ac.uk.

Received 6 May 2015 and accepted 1 November 2015.
This article contains Supplementary Data online at http://diabetes.diabetesjournals.org/lookup/suppl/doi:10.2337/db15-0602/-/DC1.
A.P, T.P.P., and S.K.T. are joint first authors.

© 2016 by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered.

β-cell proliferation and conferred resistance to chemically induced diabetes (9). Overexpression, in contrast, reduced β-cell proliferation in both young and old mice. This is consistent with the effect of *Cdk4* loss, which has been shown to result in a reduced number of pancreatic β-cells and insulin-deficient diabetes (10). More recent mouse studies have also established a role for *Cdkn2a* and *Cdk4* in hepatic glucose production, demonstrating cell cycle–independent effects on gluconeogenesis under fasted and fed conditions (11,12).

While rodent studies have provided critical clues into the contribution of *Cdkn2a* to diabetes pathogenesis, less is known about the role of *CDKN2A* in glucose homeostasis in humans. The machinery regulating the G1/S transition in adult human β-cells differs from that of mouse β-cells, which do not express CDK6 (13–15). Individuals heterozygous for germline loss-of-function mutations in the *CDKN2A* gene have a high risk of developing (multiple) cutaneous melanoma, a condition known as familial atypical multiple mole melanoma syndrome (FAMMM) (16,17). These subjects provide a unique opportunity to study the effect of *CDKN2A* haploinsufficiency on glucose homeostasis in humans. The current study tested the hypothesis that mutation carriers show improved β-cell function compared with noncarriers.

RESEARCH DESIGN AND METHODS

Study Participants

Thirty-one case subjects diagnosed with FAMMM due to *CDKN2A* mutations were recruited from centers in the U.K. and the Netherlands. Twenty-eight had been cancer free for at least 2 years, and the remaining three case subjects had presented with melanoma between 4 and 12 months prior to inclusion in the study (Supplementary Table 1). For a control group of 31 participants, unaffected first-degree relatives or spouses of carriers were chosen when available and additional control subjects were recruited from the Oxford Biobank (www.oxfordbiobank.org.uk). Two control subjects were subsequently excluded based on 2-h oral glucose tolerance test (OGTT) glucose levels that diagnosed diabetes (serum glucose >11 mmol/L). All remaining participants were aged 18–80 years, not suffering from diabetes, and not taking any medication that could interfere with glucose tolerance.

Baseline Clinical Characteristics and OGTT

All participants underwent a 75-g OGTT after a 12-h fast. Blood samples were collected at 0, 15, 30, 60, 90, and 120 min after the oral glucose load to assay plasma glucose, serum insulin, and (for a subset of 12 mutation carriers and 12 control subjects) C-peptide. Insulin and C-peptide were measured using chemiluminescence immunoassays. Measures of insulin sensitivity, β-cell function, and hepatic clearance derived from the OGTT were calculated according to the formulas in Supplementary Table 2.

Intravenous Glucose Tolerance Test

Intravenous glucose tolerance tests (IVGTTs) were performed on a subset of the U.K. subjects (eight) who had attended for OGTT and consented to undergo an IVGTT. Control subjects, matched for age, sex, BMI, and activity, were recruited from the Oxford Biobank. After a 12-h fast, a dose of 50% dextrose (calculated based on weight 0.5 mg/kg) was given over 3 min. Blood samples were then taken at 0, 2, 4, 6, 8, 10, 15, 20, 30, 45, 60, 75, 90, 120, 150, and 180 min. These samples were batch analyzed for insulin, glucose, and C-peptide. Data were then analyzed using a minimal model approach, according to an algorithm designed to maximize precision and identification success rate (18).

Cellular Assays Using the EndoC-bH1 Cell Line

The EndoC-bH1 cell line was cultured and passaged as previously described (19). Reverse transfections were performed by adding preformed small interfering (si)RNA complexes prepared from ON-TARGETplus siRNA SMARTpools (Dharmacon) at a final concentration of 10 nmol/L siRNA. For gene expression analysis, RNA was extracted and quantitative PCR performed using the TaqMan gene expression kit and assays (Applied Biosystems) on oligo-dT primed cDNA. Seventy-two hours after transfection, cells were starved overnight in 2.8 mmol/L glucose followed by 1 h in 0 mmol/L glucose medium. Static insulin secretion assays were then initiated by adding glucose-free growth medium supplemented with the indicated amounts of glucose and isobutylmethylxanthine (IBMX). After 1 h, aliquots of supernatants were removed for later analysis and ice-cold acid ethanol was added to extract insulin content from cells. Sample analysis was performed using the AlphaLISA Human Insulin Immunoassay (PerkinElmer).

For protein kinase A (PKA) activity assays, cells were harvested after knockdown, as described above, and washed in PBS. After matching input for number of cells, the samples were then processed according to the manufacturer's instructions for the PepTag nonradioactive PKA assay (Promega) and visualized using the ChemiDoc MP system.

Statistical Analysis

Statistical analysis was performed using R 3.0.2. *P* values were determined by Welch *t* test, except for sex differences, where the χ^2 test was used, and for analysis of the IVGTT data, where the Mann-Whitney *U* test was used.

RESULTS

We recruited 31 participants carrying inherited *CDKN2A* loss-of-function mutations (Supplementary Table 1) and 31 control subjects matched as a group for age (P = 0.99), sex (P = 0.43), and BMI (P = 0.97) (Table 1). To test our hypothesis that *CDKN2A*-loss leads to improved β-cell function, we first performed a 120-min OGTT in all subjects (Fig. 1*A* and *B*). While no difference in glucose levels was detected, insulin levels

Table 1—Baseline characteristics of study participants

	Mutation carriers	Noncarriers	*P*
n	31	31	NA
BMI (kg/m^2)	27.1 (19; 38)	27.1 (19; 36)	0.97
Age (years)	51.8 (21; 71)	51.8 (25; 84)	0.99
Sex (% male)	45	32	0.43

Data are mean and range (minimum; maximum) unless otherwise indicated. *P* values are from Welch *t* test except for sex distribution where the χ^2 test was performed. NA, not applicable.

were significantly increased in carriers throughout the test (P = 0.01 for insulin area under curve [AUC_{ins}]) (Table 2).

Using these data, we derived standard indices of β-cell function and insulin sensitivity (Table 2). This revealed increased β-cell function in carriers compared with noncarriers, both assessed using a dynamic measure of acute insulin response (P = 0.03 for BIGTT-acute insulin response [AIR]) and in the fasted state (P = 0.05 for iHOMA2 of β-cell function). Corresponding measures of insulin sensitivity, BIGTT-S, and iHOMA2 of insulin sensitivity were also both found to be lower in carriers (P = 0.04 and P = 0.05, respectively). Other standard measures, the Belfiore and Matsuda insulin sensitivity indices (ISIs), confirmed the observed reduction in insulin sensitivity of carriers (P = 0.02 and P = 0.02, respectively). As a result, the disposition index, which is an aggregate measure of β-cell function relative to glucose sensitivity, remained unaffected compared with control subjects (P = 0.98). These results were not significantly altered by exclusion of three carriers that had presented with melanoma within 2 years prior to inclusion in the study (Supplementary Table 3).

To explore whether the observed phenotype was driven by underlying effects on $p16^{INK4a}$, $p14^{ARF}$, or both, we reanalyzed the data with carriers grouped by mutation status (Supplementary Fig. 1). Of the mutations identified, 26 affected both $p16^{INK4a}$ and $p14^{ARF}$, while five were located in regions affecting $p16^{INK4a}$ exclusively. No differences were observed between these two groups in insulin or glucose levels (P = 1.00 for AUC_{ins} and P = 0.49 for $AUC_{glucose}$, respectively), suggesting that the observed metabolic phenotype of mutation carriers may be driven either solely by effects on $p16^{INK4a}$ or by effects of similar magnitude on both proteins.

For a subset of participants (12 carriers and 12 control subjects), C-peptide measurements were obtained during the OGTT. Despite a tendency toward increased C-peptide levels in the fasted state (P = 0.48), the total response was not different for this subset of individuals (P = 1.00 for AUC). Indices of hepatic insulin clearance, derived from the ratio between C-peptide and insulin levels, however, showed significantly decreased hepatic clearance in mutation carriers (P = 0.03) (Table 2) (20).

To confirm these findings, we performed IVGTTs on eight case and eight control subjects available for follow-up studies (Supplementary Fig. 2). None of the measures derived from this test reached statistical significance, but directions of effect were confirmed for both insulin secretion (P = 0.14 for AUC_{ins} 10–180 min) and hepatic insulin clearance (P = 0.21) (Supplementary Table 4). The insulin response was found to be 66% and 110% higher for carriers during the first and second phase of secretion, respectively. In contrast, the C-peptide response (which is unaffected by hepatic clearance) was ~30% higher during

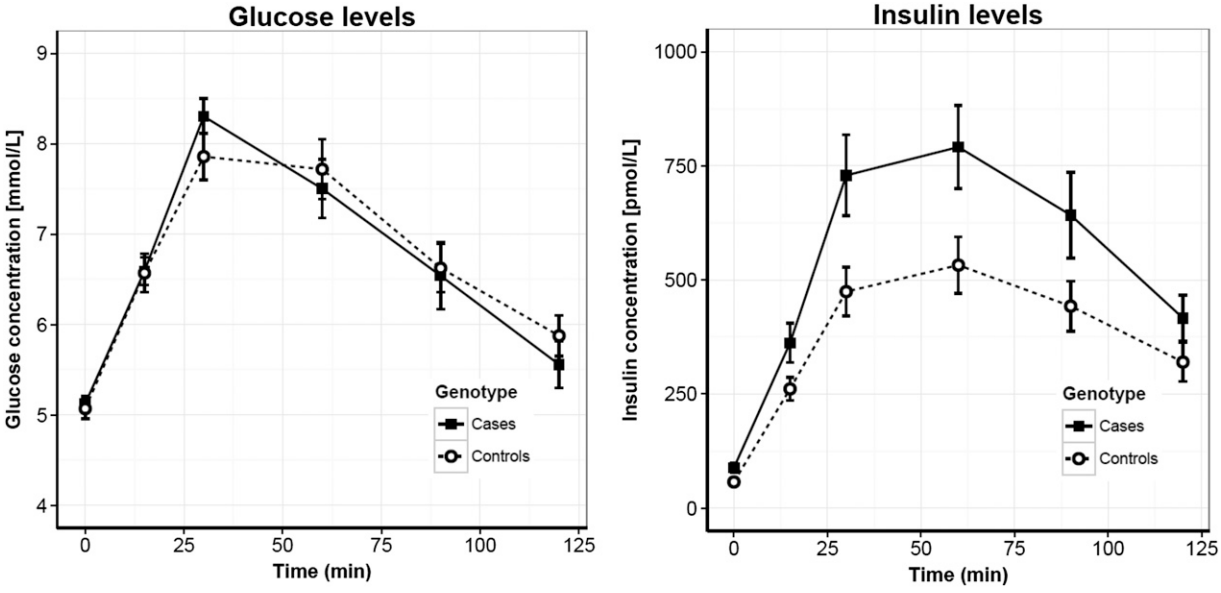


Figure 1—Serum glucose and insulin levels during a 120-min OGTT in 31 carriers and 31 control subjects. Data shown as mean ± SEM.

both phases of secretion, indicating a direct contribution of improved β-cell function to the elevated circulating insulin levels of carriers.

Finally, we sought to establish the extent to which cell cycle–independent effects of *CDKN2A* on the regulation of insulin secretion could contribute to the phenotype of mutation carriers. Recent work in rodent hepatocyte models has suggested a role of *Cdkn2a* in the regulation of PKA signaling (12). Given the well-characterized effects of PKA on potentiation of insulin secretion, we speculated that such signaling events could have a direct effect on β-cell function. To test this hypothesis, we performed knockdown and secretion studies in the human pancreatic β-cell line, EndoC-bH1. This cell line was transformed by Ravassard et al. (19) using the proto-oncogene SV40LT, which acts on the Retinoblastoma (Rb) protein, thereby masking effects of $p16^{INK4a}$ on cell-cycle control.

We first confirmed expression of $p16^{INK4a}$ by immunofluorescence and found that, consistent with previous reports, the protein localized to both the nucleus and cytoplasm (21,22) (Supplementary Fig. 3). siRNA-mediated silencing of *CDKN2A* was then performed and efficient knockdown observed both at the mRNA and protein levels (Fig. 2*A* and *B* and Supplementary Fig. 4). Ninety-six hours after gene silencing, cells treated with *CDKN2A* or nontargeting siRNAs were incubated under different conditions to assess the glucose responsiveness of the cells. In addition to basal and high-glucose conditions, the effect of the phosphodiesterase inhibitor (IBMX) on insulin secretion was tested. For all three conditions, *CDKN2A* knockdown was found to increase insulin secretion as a fraction of total content (basal, $P = 0.02$; high, $P = 0.01$; high glucose with IBMX, $P = 0.04$) (Fig. 2*C* and *D*), and as expected, no effect on proliferation was detected. We also observed a small but significant reduction in the total insulin content per cell ($P < 0.01$) (Supplementary Fig. 5). Finally, we performed PKA activity assays to directly assess the effect of *CDKN2A* silencing on the potentiating pathway of insulin secretion. Consistent with an increase in insulin secretion, this revealed a corresponding 23% increase in the activity of PKA after *CDKN2A* knockdown ($P = 0.02$) (Supplementary Fig. 6).

DISCUSSION

Individuals carrying heterozygous loss-of-function mutations in the *CDKN2A* gene provide a unique opportunity to study the role of $p16^{INK4a}$ and $p14^{ARF}$ in glucose homeostasis in humans. Through OGTT and IVGTT, we found that carriers displayed significantly increased insulin levels compared with matched control subjects. In a subset of individuals, measurements of C-peptide levels established a contribution of both decreased hepatic insulin clearance and increased β-cell function to the elevated circulating insulin. Further, grouping carriers by mutation status showed the effects to be driven either by $p16^{INK4a}$ exclusively or through similar effects on both $p16^{INK4a}$ and $p14^{ARF}$.

Overall, these results are consistent with a combination of two nonmutually exclusive mechanisms underlying the

Table 2—OGTT-derived measures of β-cell function, insulin sensitivity, and hepatic clearance

	Mutation carriers	Noncarriers	*P*
Fasting glucose (mmol/L)	5.2 (4.3; 6.3)	5.1 (3.2; 6.4)	0.65
Fasting insulin (pmol/L)	87 (15; 337)	55 (22; 150)	0.01
Fasting C-peptide (nmol/L)	0.44 (0.24; 0.81)	0.39 (0.20; 0.60)	0.48
iHOMA-B	124 (38.1; 452.4)	96 (45; 236.7)	0.05
iHOMA-S	91 (18; 328.7)	120 (35; 235.6)	0.05
BIGTT-AIR ($*10^3$)	6.4 (0.9; 28)	3.0 (1.2; 12)	0.03
BIGTT-S	5.8 (0.4; 12.8)	7.8 (1.1; 17.8)	0.04
Belfiore ISI	0.78 (0.17; 1.35)	0.97 (0.35; 1.77)	0.02
Matsuda ISI	4.3 (0.8; 11.1)	6.3 (1.5; 20.9)	0.02
$AUC_{glucose}$	839 (563; 1,449)	829 (502; 1,086)	0.79
AUC_{ins} ($*10^4$)	7.3 (2.4; 25)	4.7 (1.1; 15)	0.01
$AUC_{C\text{-}peptide}$	212 (106; 333)	212 (115; 260)	1.00
Insulinogenic index	203 (39; 561)	152 (53; 360)	0.08
C-peptidogenic index	0.45 (0.15; 1.12)	0.45 (0.17; 1.78)	1.00
Disposition index	2.3 (1.1; 3.8)	2.3 (1.0; 3.7)	0.98
Fasting insulin clearance	0.66 (0.12; 1.27)	0.88 (0.65; 1.29)	0.07
Insulin clearance	0.35 (0.16; 0.51)	0.56 (0.32; 1.05)	0.03

Data are mean and range (minimum; maximum). All indices based on C-peptide measurements are based on data from a subset of individuals only ($n = 12$ carriers and $n = 12$ control subjects; all U.K.). Details on definitions of physiological measures are listed in Supplementary Table 2. iHOMA-B, interactive HOMA of β-cell function; iHOMA-S, interactive HOMA of insulin sensitivity.

phenotype of carriers: *1*) primary β-cell hyperfunction driving progressive insulin resistance and/or *2*) primary insulin resistance triggering a compensatory increase in insulin levels (Supplementary Fig. 7). While both explanations are consistent with our data, existing evidence strongly supports a role for *CDKN2A* in β-cell function (9). Chronic hyperinsulinemia is known to result in a gradual downregulation of both insulin receptors and postreceptor signaling efficiency, thereby causing general insulin resistance and reduced insulin clearance (23). Our data are therefore in agreement with the expected physiological adaption to chronic hyperinsulinemia. However, due to limitations on the design of our clinical study, we cannot conclusively address the cause and effect between hyperinsulinemia and insulin resistance in mutation carriers. The IVGTT is well validated against clamp-based techniques, but power calculations based on our results suggest that an impractically high number of 50–60 individuals would be required to establish significant differences. Given the rarity of the disease, this exceeds the number of carriers available in the U.K. and Dutch cohorts recruited for our study.

To test for a cell cycle–independent role of *CDKN2A* in the regulation of insulin secretion, we performed knockdown studies in the human β-cell line, EndoC-bH1. This identified cell cycle–independent increases in insulin secretion under three conditions. These changes were found to be accompanied by increased PKA activity, in agreement with previous studies establishing such an effect of *CDKN2A* knockdown in liver (11,12). This suggests a possible contribution of the PKA-dependent potentiating pathway to the secretory effects observed in the EndoC-bH1 cell line. Taken in combination with existing data, our clinical and cellular studies indicate that the phenotype of carriers may arise out of a complex interplay between both cell cycle–independent and –dependent roles

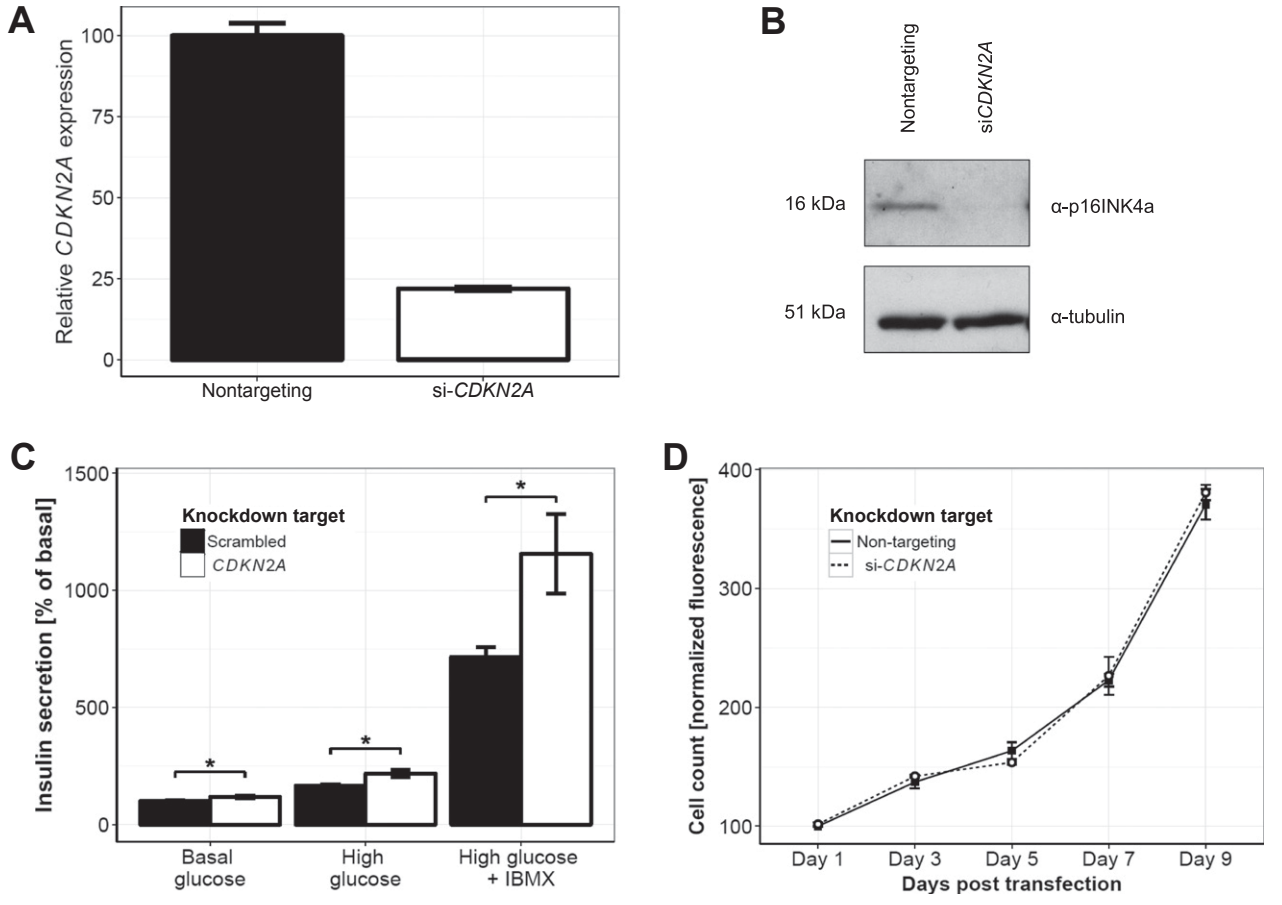


Figure 2—*CDKN2A* knockdown in the human β-cell line, EndoC-bH1. Silencing of *CDKN2A* and nontargeting sequence was performed using pools of siRNA (10 nmol/L) and knockdown confirmed both at the mRNA level (*A*) and at the protein level (*B*) after 72 h. Static insulin secretion assays were performed under culturing conditions of basal glucose (2.8 mmol/L glucose), high glucose (20 mmol/L), and high glucose with IBMX (100 μmol/L) (*C*). All secretion results were normalized to total insulin content per well. Cellular proliferation (*D*) after treatment with si-*CDKN2A* and nontargeting siRNA was measured using the CyQUANT Direct Cell Proliferation assay and normalized to the respective counts for day 1. Bars represent means for $n = 11$ (si-*CDKN2A*) and 18 (nontargeting) generated in three independent experiments. Data points for cellular proliferation are means for $n = 3$, and error bars are SEM. *$P < 0.05$.

of *CDKN2A* in a range of tissues (9,12) (Supplementary Fig. 7).

Upstream of the *CDKN2A* and *CDKN2B* genes, several independent association signals for type 2 diabetes risk have been identified. The underlying effector transcript and disease mechanism has remained elusive, and prior studies have not reported any *cis*-expression quantitative trait loci effects for these alleles (24). Our study has shown that both coding *CDKN2A* mutations and the noncoding type 2 diabetes variants are associated with effects on measures of β-cell function. This provides a link between *CDKN2A* and the common genome-wide association study alleles and thus points to the gene as a likely effector transcript at this locus.

Interestingly, type 2 diabetes–associated variants at the *CDKN2A/B* locus have consistently been linked to a more "classic" β-cell phenotype than that observed for carriers of coding mutations in our study, with no evidence for an impact on measures of insulin resistance (3,4). We speculated that any *cis*-regulatory effect exerted on *CDKN2A* could achieve a more restricted β-cell phenotype through tissue-specific regulation of gene expression. To address this hypothesis, we interrogated existing genome annotations and found that the noncoding disease-associated variants map to a cluster of islet enhancer activity and open chromatin. Specifically, the association signals overlap a strong enrichment for islet- and melanocyte-specific FOXA-2 binding (25) (Supplementary Fig. 8). This highlights a possible mechanism for the more specific β-cell phenotype caused by common disease-associated variants compared with carriers of coding variants. Based on the direction of effect on measures of β-cell function, the noncoding risk alleles would be predicted to increase expression of *CDKN2A* (3,4). No *cis*-expression quantitative trait loci effects have previously been reported in islets for this region, but larger studies currently underway may be able to shed further light on this hypothesis (26).

Taken together, our data establish *CDKN2A* as an important regulator of glucose homeostasis in humans. We have shown that our data are consistent with loss-of-function mutations in *CDKN2A* affecting a range of tissues, including both pancreatic β-cells and liver. Our study thus supports the candidacy of *CDKN2A* as the effector transcript of the type 2 diabetes–associated alleles in the region, and we have proposed a mechanism to account for the apparent tissue specificity of the β-cell dysfunction caused by diabetes risk alleles.

Acknowledgments. The authors thank Linda Whitaker (University of Leeds) and Beryl Barrow (University of Oxford) for valuable assistance with recruiting FAMMM patients in the U.K.; Marja Dijk, Bep Ladan, and Petra Beckers (Leiden University Medical Centre) for assistance with performing the OGTTs in the Netherlands; and Jonathan Levy (University of Oxford) for advice on designing the study. The authors thank all volunteers in the U.K. and the Netherlands for participation. The Oxford Biobank (www.oxfordbiobank.org.uk), National Institute for Health Research Oxford Biomedical Research Centre, is part of the National Institute for Health Research National Bioresource, which supported the recalling process of the volunteers.

Funding. This study was funded by the Wellcome Trust (095101/Z/10Z and 098381), the Medical Research Council (G0800467 and MR/L020149/1), the National Institute for Health Research Oxford Biomedical Research Centre, Cancer Research UK (C588/A19167), and the ZOLEON foundation (12.09). M.I.M. is a Wellcome Trust Senior Investigator. A.L.G. is a Wellcome Trust Senior Fellow in Basic Biomedical Science.

Duality of Interest. No potential conflicts of interest relevant to this article were reported.

Author Contributions. A.P., S.K.T., J.N.-B., M.I.M., and A.L.G. conceived and designed the study. A.P., T.P.P., S.K.T., H.J.N., and A.B. performed the experiments. R.S. and D.T.B. provided protocols and clinical data. A.P., S.K.T., H.J.N., and I.F.G. analyzed data. A.P., T.P.P., S.K.T., F.K., I.F.G., H.F.A.V., H.P., M.I.M., and A.L.G. interpreted data. S.K.T. and A.L.G. wrote the first draft of the manuscript. A.P., T.P.P., H.P., and M.I.M. edited the manuscript. A.P., T.P.P., S.K.T., H.J.N., A.B., R.S., T.J.J., D.T.B., F.K. I.F.G., H.F.A.V., J.N.-B., H.P., M.I.M., and A.L.G. approved the final manuscript. A.L.G. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Prior Presentation. Parts of this study were presented in abstract form at the 75th Scientific Sessions of the American Diabetes Association, Boston, MA, 5–9 June 2015.

References

1. Voight BF, Scott LJ, Steinthorsdottir V, et al.; MAGIC investigators; GIANT Consortium. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet 2010;42:579–589

2. Morris AP, Voight BF, Teslovich TM, et al.; Wellcome Trust Case Control Consortium; Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC) Investigators; Genetic Investigation of ANthropometric Traits (GIANT) Consortium; Asian Genetic Epidemiology Network–Type 2 Diabetes (AGEN-T2D) Consortium; South Asian Type 2 Diabetes (SAT2D) Consortium; DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet 2012;44:981–990

3. Grarup N, Rose CS, Andersson EA, et al. Studies of association of variants near the HHEX, CDKN2A/B, and IGF2BP2 genes with type 2 diabetes and impaired insulin release in 10,705 Danish subjects: validation and extension of genome-wide association studies. Diabetes 2007;56:3105–3111

4. Dimas AS, Lagou V, Barker A, et al.; MAGIC Investigators. Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. Diabetes 2014;63:2158–2171

5. Kamb A, Gruis NA, Weaver-Feldhaus J, et al. A cell cycle regulator potentially involved in genesis of many tumor types. Science 1994;264:436–440

6. Kamijo T, Zindy F, Roussel MF, et al. Tumor suppression at the mouse INK4a locus mediated by the alternative reading frame product p19ARF. Cell 1997;91:649–659

7. Serrano M, Hannon GJ, Beach D. A new regulatory motif in cell-cycle control causing specific inhibition of cyclin D/CDK4. Nature 1993;366:704–707

8. Weber HO, Samuel T, Rauch P, Funk JO. Human p14(ARF)-mediated cell cycle arrest strictly depends on intact p53 signaling pathways. Oncogene 2002; 21:3207–3212

9. Krishnamurthy J, Ramsey MR, Ligon KL, et al. p16INK4a induces an age-dependent decline in islet regenerative potential. Nature 2006;443:453–457

10. Rane SG, Dubus P, Mettus RV, et al. Loss of Cdk4 expression causes insulin-deficient diabetes and Cdk4 activation results in beta-islet cell hyperplasia. Nat Genet 1999;22:44–52

11. Lee Y, Dominy JE, Choi YJ, et al. Cyclin D1-Cdk4 controls glucose metabolism independently of cell cycle progression. Nature 2014;510:547–551

12. Bantubungi K, Hannou SA, Caron-Houde S, et al. Cdkn2a/p16Ink4a regulates fasting-induced hepatic gluconeogenesis through the PKA-CREB-PGC1α pathway. Diabetes 2014;63:3199–3209
13. Martín J, Hunt SL, Dubus P, et al. Genetic rescue of Cdk4 null mice restores pancreatic beta-cell proliferation but not homeostatic cell number. Oncogene 2003;22:5261–5269
14. Fiaschi-Taesch N, Bigatel TA, Sicari B, et al. Survey of the human pancreatic beta-cell G1/S proteome reveals a potential therapeutic role for cdk-6 and cyclin D1 in enhancing human beta-cell replication and function in vivo. Diabetes 2009;58:882–893
15. Fiaschi-Taesch NM, Salim F, Kleinberger J, et al. Induction of human beta-cell proliferation and engraftment using a single G1/S regulatory molecule, cdk6. Diabetes 2010;59:1926–1936
16. Cannon-Albright LA, Goldgar DE, Meyer LJ, et al. Assignment of a locus for familial melanoma, MLM, to chromosome 9p13-p22. Science 1992;258: 1148–1152
17. Goldstein AM. Familial melanoma, pancreatic cancer and germline CDKN2A mutations. Hum Mutat 2004;23:630
18. Godsland IF, Agbaje OF, Hovorka R. Evaluation of nonlinear regression approaches to estimation of insulin sensitivity by the minimal model with reference to Bayesian hierarchical analysis. Am J Physiol Endocrinol Metab 2006; 291:E167–E174
19. 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;121:3589–3597
20. Horwitz DL, Starr JI, Mako ME, Blackard WG, Rubenstein AH. Proinsulin, insulin, and C-peptide concentrations in human portal and peripheral blood. J Clin Invest 1975;55:1278–1283
21. McKenzie HA, Fung C, Becker TM, et al. Predicting functional significance of cancer-associated p16(INK4a) mutations in CDKN2A. Hum Mutat 2010;31:692–701
22. Keller-Melchior R, Schmidt R, Piepkorn M. Expression of the tumor suppressor gene product p16INK4 in benign and malignant melanocytic lesions. J Invest Dermatol 1998;110:932–938
23. Shanik MH, Xu Y, Skrha J, Dankner R, Zick Y, Roth J. Insulin resistance and hyperinsulinemia: is hyperinsulinemia the cart or the horse? Diabetes Care 2008; 31(Suppl. 2):S262–S268
24. Hannou SA, Wouters K, Paumelle R, Staels B. Functional genomics of the CDKN2A/B locus in cardiovascular and metabolic disease: what have we learned from GWASs? Trends Endocrinol Metab 2015;26:176–184
25. Pasquali L, Gaulton KJ, Rodríguez-Seguí SA, et al. Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. Nat Genet 2014;46:136–143
26. Fadista J, Vikman P, Laakso EO, et al. Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism. Proc Natl Acad Sci U S A 2014;111:13924–13929

ARTICLES
https://doi.org/10.1038/s41588-018-0173-1
nature genetics

Type 2 diabetes risk alleles in *PAM* impact insulin release from human pancreatic β-cells

Soren K. Thomsen[1,7,9], Anne Raimondo[1,8,9], Benoit Hastoy[1,9], Shahana Sengupta[1,2], Xiao-Qing Dai[3], Austin Bautista[3], Jenny Censin[4,5], Anthony J. Payne[5], Mahesh M. Umapathysivam[1], Aliya F. Spigelman[3], Amy Barrett[1], Christopher J. Groves[1], Nicola L. Beer[1], Jocelyn E. Manning Fox[3], Mark I. McCarthy[1,5,6], Anne Clark[1], Anubha Mahajan[5], Patrik Rorsman[1,6], Patrick E. MacDonald[3] and Anna L. Gloyn[1,5,6*]

The molecular mechanisms underpinning susceptibility loci for type 2 diabetes (T2D) remain poorly understood. Coding variants in peptidylglycine α-amidating monooxygenase (*PAM*) are associated with both T2D risk and insulinogenic index. Here, we demonstrate that the T2D risk alleles impact negatively on overall PAM activity via defects in expression and catalytic function. PAM deficiency results in reduced insulin content and altered dynamics of insulin secretion in a human β-cell model and primary islets from cadaveric donors. Thus, our results demonstrate a role for PAM in β-cell function, and establish molecular mechanisms for T2D risk alleles at this locus.

Human genetics has the potential to identify novel mechanisms for disease, which in turn present opportunities for clinical translation. Over the past 10 years, genome-wide association studies (GWAS) have identified over 400 genomic regions robustly associated with type 2 diabetes (T2D) risk, and have highlighted a central role for pancreatic β-cell dysfunction in T2D pathogenesis[1–4]. However, translating GWAS signals into molecular mechanisms for disease has been slow, primarily because of uncertainty over the transcripts through which non-coding association signals operate[5]. Recently, studies focused on identifying coding variant association signals have identified non-synonymous alleles in peptidylglycine α-amidating monooxygenase (*PAM*) independently that are associated with a risk of T2D (rs78408340, NM_000919.3, c.1616C > G, odds ratio (OR) = 1.47, minor allele frequency (MAF) = 7.3×10^{-3} in Europeans; rs35658696, NM_000919.3, c.1688A > G, OR = 1.23, MAF = 0.045), providing a direct link to the effector transcript[3,6,7]. Although the low-frequency variant at rs35658696 was in linkage disequilibrium with a coding variant in a nearby gene (*PPIP5K2*; rs36046591), recent fine-mapping efforts have confirmed rs35658696 as the causal variant[4]. Both T2D risk alleles are also associated with a reduced insulinogenic index (rs78408340, β = −8.42; rs35658696, β = −1.96)—a measure of glucose-stimulated insulin secretion—suggesting that their effects are mediated via altered β-cell function[6,8,9].

PAM encodes PAM, an enzyme in neuroendocrine cells that modifies peptides with a C-terminal glycine residue to create peptide amides[10–12]. Amidation can dramatically increase the biological potency of a peptide relative to its unmodified glycine-extended conjugate[13]. PAM is localized to the Golgi complex, where it is packaged with other endocrine proteins into nascent granules[10,14]. The functional enzyme exists in both integral membrane and luminal forms; the latter is co-secreted with the endocrine peptide(s)[10,11,15].

Despite reports of PAM expression in pancreatic islets, a functional role in β-cells has not yet been described[16]. Insulin itself is not a PAM substrate, so the association of *PAM* variants with an insulinogenic index must be mediated via other peptide(s) or mechanisms. We hypothesized that T2D-associated *PAM* missense alleles reduce PAM function, affecting the amidation of peptides that is critical for insulin secretion. We demonstrate that both diabetes risk alleles negatively impact on PAM expression and/or activity, and elucidate an endogenous role for PAM in insulin granule packaging and release from β-cells. We also show that PAM amidates the granule packaging factor chromogranin A (CgA), and identify this neuroendocrine peptide as a likely downstream effector for PAM in β-cells. Our results are consistent with the direction and magnitude of effects for T2D-associated risk alleles in *PAM*, and establish molecular mechanisms for their impact on disease susceptibility.

Results

T2D-associated *PAM* alleles cause PAM loss-of-function. PAM is a bifunctional enzyme, possessing two contiguous catalytic domains: peptidylglycine α-hydroxylating monooxygenase (PHM) and peptidyl α-hydroxyglycine alpha-amidating lyase (PAL)[11]. Both T2D risk variants in *PAM* encode mutations located within the PAL domain (rs78408340, p.Ser539Trp; rs35658696, p.Asp563Gly) and are predicted by in silico tools (SIFT (Sorting Intolerant From Tolerant), PolyPhen2 (Polymorphism Phenotyping v2)) to be damaging, suggesting that they could affect enzymatic activity. To test this, we produced recombinant luminal (non-integral

[1]Oxford Centre for Diabetes, Endocrinology & Metabolism, University of Oxford, Oxford, UK. [2]MRC Harwell Institute, Harwell Campus, Oxfordshire, UK. [3]Department of Pharmacology and Alberta Diabetes Institute, University of Alberta, Edmonton, Alberta, Canada. [4]Big Data Institute at the Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, UK. [5]Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK. [6]Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, UK. [7]Present address: Vertex Pharmaceuticals Europe Ltd, Milton Park, Abingdon, UK. [8]Present address: National Health and Medical Research Council, Canberra, Australia. [9]These authors contributed equally: Soren K. Thomsen, Anne Raimondo and Benoit Hastoy. *e-mail: anna.gloyn@drl.ox.ac.uk

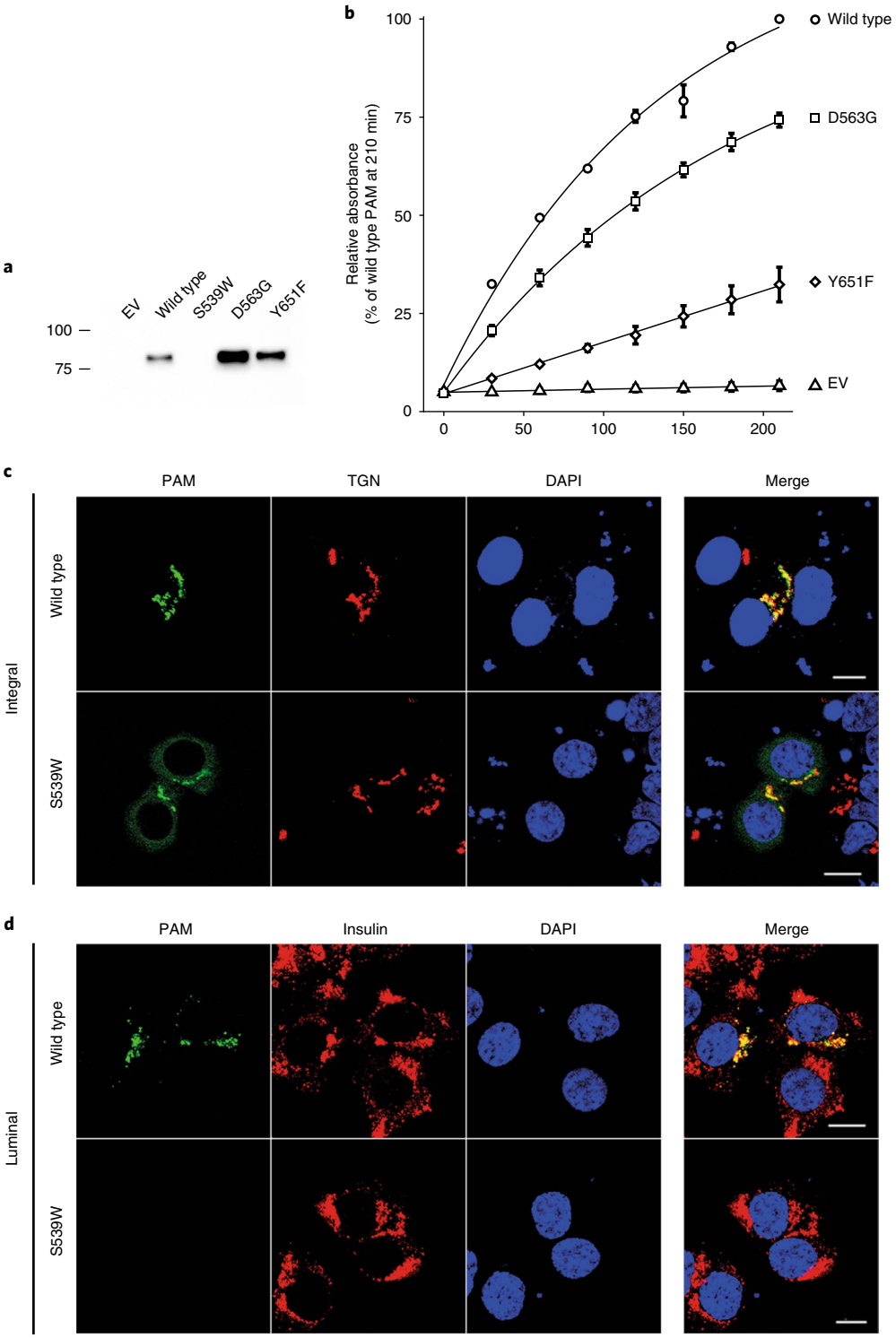


Fig. 1 | Analysis of wild-type and variant PAM function and expression. a, Western blot analysis of recombinant PAM protein production in supernatant from HEK293 stable cells. Size markers indicate protein mass in kilodaltons. **b**, Amidating activity of wild-type PAM (circles), p.Asp563Gly PAM (squares), p.Tyr651Phe PAM (diamonds) or empty vector (EV) (triangles) in vitro. The graph shows the means of four independent experiments; the error bars are the standard error of the mean. **c,d**, EndoC-βH1 cells transfected with expression vectors for integral membrane (**c**) or luminal (**d**) wild-type or variant PAM, and labeled for PAM (green), the trans-Golgi network (TGN) (red in **c**) or insulin (red in **d**). DAPI (blue) was used as a nuclear marker. Scale bar, 2 µm. Results in **a**,**c** and **d** are representative of three independent experiments.

membrane) PAM protein for in vitro amidation assays using human embryonic kidney 293 (HEK293) cells as a suitable human expression system. In line with previous observations, PAM was constitutively released into the supernatant (Fig. 1a)[17]. Wild-type PAM and p.Asp563Gly PAM were both robustly produced. Additionally, a catalytically inactive mutant protein, p.Tyr651Phe PAM, was also produced, which was used as a control[18]. Interestingly, we could not detect p.Ser539Trp PAM expression

(Fig. 1a). This was observed across three independently derived cell lines, and was not due to cellular retention of p.Ser539Trp PAM (data not shown).

We subsequently developed a cell-free kinetic assay capable of measuring PAM amidating activity via spectrophotometric detection of converted glyoxylate, a by-product of the amidation reaction[19,20]. Matching each reaction for PAM input, we observed reduced amidating activity for p.Asp563Gly (D563G) PAM ($P=1.0\times10^{-5}$) and p.Tyr651Phe (Y651F) PAM ($P=4.1\times10^{-6}$) (Fig. 1b). In agreement with its lack of expression, supernatant from p.Ser539Trp (S539W) PAM-transfected cells was inactive in this assay (Supplementary Fig. 1a). Further analysis showed no significant difference in substrate affinity between wild-type PAM and p.Asp563Gly PAM (substrate concentration which gives half maximal velocity of the enzymatic reaction [K_m] 0.95 mmol.l^{-1} versus 1.02 mmol.l^{-1}, $P=0.44$), suggesting that the p.Asp563Gly substitution affects the number of substrate molecules converted over time (K_{cat})(Supplementary Fig. 1b). These results demonstrate that the T2D-associated missense alleles in *PAM* decrease PAM function via a combination of defective expression and/or reduced catalysis.

PAM localizes to the β-cell secretory pathway. Having established the direction of effect for T2D risk alleles in *PAM*, we next explored a role for PAM in physiologically relevant tissues. Transcript expression profiling detected *PAM* in multiple tissue types, with the highest expression in human islets followed by the heart and salivary glands (Supplementary Fig. 2a). In mouse tissues, *Pam* expression was highest in the pituitary gland (not included in the human panel) followed by the heart and islets (Supplementary Fig. 2a). These results are consistent with the published association between *PAM* risk alleles and reduced measures of insulin secretion, and suggest a direct role for PAM in β-cell function. We confirmed broadly similar expression patterns in the publicly available Genotype Tissue Expression (GTEx) database, although data for pancreatic islets are not available from this project[21]. Published RNA sequencing data also showed *PAM* to be equally expressed in enriched fractions of β-cells and non-β-cells (Supplementary Fig. 2a)[22,23]. We verified this expression pattern at the protein level by performing immunofluorescence and electron microscopy of primary human islets, which showed co-localization of endogenous PAM in both insulin and glucagon vesicles (Fig. 2a,b). We also confirmed PAM localization to the insulin secretory pathway of the glucose-responsive human β-cell model EndoC-βH1 (Fig. 2c)[24,25]. These observations are in agreement with early studies that established PAM expression and activity in endocrine cells of human and murine islets[26–29].

We then performed detailed analysis of the subcellular localization of wild-type and variant PAM for both integral and luminal forms of the protein in EndoC-βH1 cells. Integral membrane wild-type PAM localized to the trans-Golgi network (TGN), where secretory granules originate, whereas luminal wild-type PAM displayed a punctate staining pattern and co-localized with endogenous insulin (Fig. 1c,d). Similar results were observed for p.Asp563Gly PAM and p.Tyr651Phe PAM (Supplementary Fig. 2b). As in HEK293 cells, luminal p.Ser539Trp PAM expression was undetectable (Fig. 1d); however, integral membrane p.Ser539Trp PAM displayed an aberrant staining pattern that localized to the endoplasmic reticulum as well as the TGN (Fig. 1c and Supplementary Fig. 2b). These results confirm that PAM localizes to components of the β-cell secretory pathway, in agreement with studies in other neuroendocrine cell types[10,14,30–32]. The p.Ser539Trp substitution may interfere with PAM protein folding, so that its translocation from the endoplasmic reticulum to the TGN is prevented or delayed.

Impact of risk variants on PAM expression. Given that luminal isoforms of PAM are known to be co-secreted with neuroendocrine peptides, we next wondered whether the impact of the p.Ser539Trp substitution would be reflected in lower circulating PAM levels in the bloodstream. Analyzing publicly available data from a human plasma proteome study by Sun et al.[33], we confirmed a significant decrease in the concentration of PAM for rs78408340 risk allele carriers (β = −1.53 per risk allele, 95% confidence interval (CI) (−1.84, −1.22), $P=5.6\times10^{-22}$, in standard deviation units of PAM levels), consistent with our in vitro observations in both HEK293 and EndoC-βH1 cells.

Interestingly, we also found that the T2D risk allele at rs35658696 was associated with reduced PAM protein levels in plasma (β = −0.97 per risk allele, 95% CI (−1.09, −0.85), $P=1.5\times10^{-67}$). To determine whether this effect could be mediated via regulatory effects on *PAM* gene expression, we analyzed RNA sequencing and genotype data from the largest available cohort of islets, which consisted of a combined collection of islets from Edmonton, Canada and Oxford, UK[34]. Out of a total of 118 donors, 11 individuals were identified as heterozygous for rs35658696. After correcting for potential mapping biases using WASP[35], allele-specific expression revealed a significant imbalance ($P=1.76\times10^{-4}$). The overall difference was relatively modest (53.2% versus 46.8%), but the imbalance was biased toward higher expression of the reference allele (Supplementary Fig. 3). This is directionally consistent with the negative functional impact of the minor allele on PAM amidation, highlighting a potential exacerbation of the reduction in PAM activity in risk allele carriers.

Finally, to identify a possible mechanism underlying this apparent difference in regulatory activity at rs35658696, we searched for genomic annotations overlapping the risk variant in reported

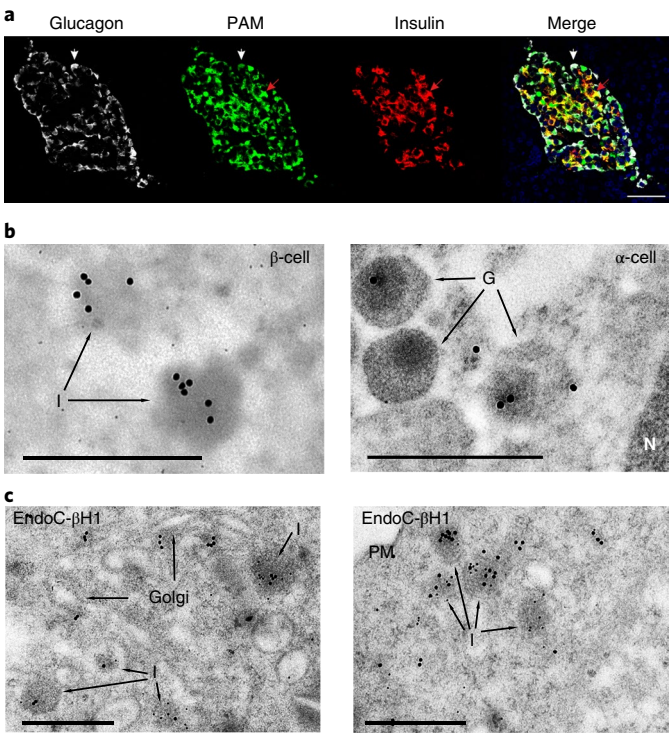


Fig. 2 | PAM expression in human pancreatic islets and EndoC-βH1 cells. **a**, Immunolocalization of endogenous PAM in a histological section of human pancreas; glucagon (white), PAM (green) and insulin (red). Red and white arrows indicate co-localization of PAM with insulin and glucagon, respectively; scale bar, 50 µm. **b**, Electron micrograph of PAM localization in human β- and α-cells. PAM immunogold-labeled (15 nm gold) was localized to insulin vesicles (I) and to glucagon vesicles (G). Scale bar, 500 nm. **c**, PAM localization in EndoC-βH1 cells. Insulin (10 nm), PAM (15 nm). PM, plasma membrane. N, nucleus. Scale bar, 500 nm.

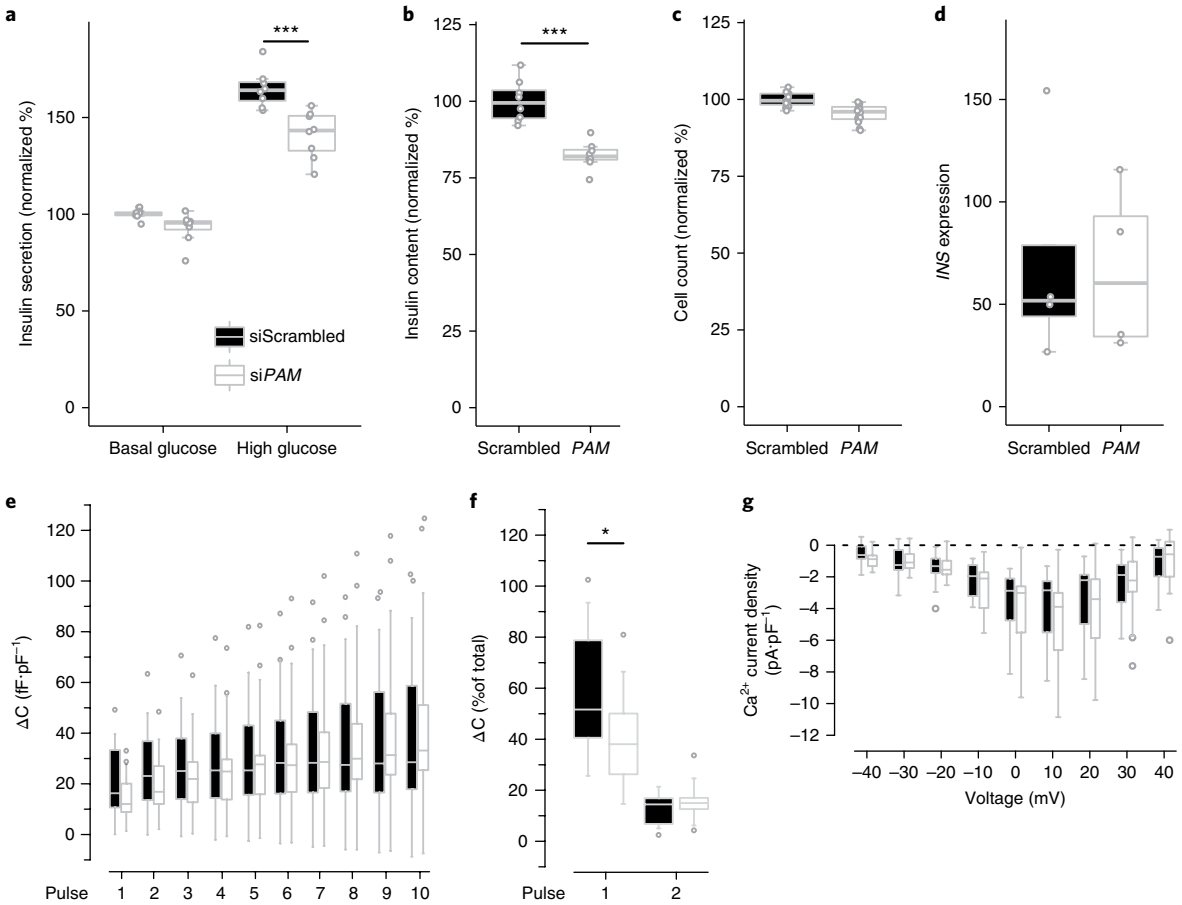


Fig. 3 | Effects of endogenous PAM on β-cell function. a–g, EndoC-βH1 cells were transfected with scrambled or *PAM* siRNA and then measured for insulin secretion ($n=8$ biologically independent samples) (**a**), cellular insulin content ($n=8$) (**b**), cell numbers ($n=16$) (**c**), insulin gene (*INS*) expression ($n=4$ independent experiments) (**d**), granule exocytosis measured as depolarization-evoked cumulative increase in membrane capacitance ($n=15$ scramble or $n=16$ *PAM* siRNA transfected cells) (**e**,**f**), and calcium current density ($n=15$ cells per treatment) (**g**). *P* values * < 0.05, ** < 0.01 and *** < 0.001 for two-tailed Student's *t*-tests where a single independent variable was being tested (**b–d**), or for two-way analysis of covariance (ANCOVA) followed by Tukey's HSD post-hoc test where two experimental variables were being tested (**a**,**e–g**). All box plots display the interquartile range (IQR) with medians, whiskers indicating 1.5 × the IQR, and outliers (for $n>10$) or all data points (for $n<10$).

chromatin states for human pancreatic islets[36]. This was extended to consider all variants in linkage disequilibrium ($r^2>0.8$) with the lead SNP. The predicted causal variant (rs35658696) overlapped annotations indicative of gene enhancer activity; two variants in strong linkage disequilibrium ($r^2>0.97$) with the causal variant overlapped weak enhancer annotations (Supplementary Table 1). These observations provide a framework for understanding the effect of the T2D risk allele on *PAM* transcript levels, although a more detailed mechanistic interrogation is outside the scope of this study.

Taken together, these results confirm the association of the rs78408340 (p.Ser539Trp) risk allele with reduced PAM protein levels and point to a regulatory component of the causal mechanism underlying the rs35658696 (p.Asp563Gly) risk allele. The latter effect may be relatively smaller than the functional impact on catalytic activity. Nevertheless, the effects are directionally aligned, thus compounding the negative effect of the risk allele through a dual mechanism of deficiencies in both catalytic function and expression.

PAM regulates glucose-responsive insulin secretion. In light of the negative impact of both T2D risk alleles on overall PAM function, we next determined the effects of reduced PAM levels on β-cell activity. Small interfering RNA (siRNA)-mediated knockdown of *PAM* in EndoC-βH1 cells caused a 79% reduction in *PAM* transcript levels (Supplementary Fig. 4a). This was associated with significantly decreased insulin secretion in high glucose (−14.4%; $P=5.6\times10^{-5}$), but not in basal glucose (−6.9%; $P=0.39$) (Fig. 3a). Furthermore, *PAM* knockdown cells displayed a significantly blunted response to glucose (ratio of high to basal secretion, −7.7%; $P=0.03$), which is indicative of a greater impact on secretion under stimulated conditions. Cellular insulin content was also reduced in *PAM* knockdown cells (−17.6%; $P=4.9\times10^{-5}$) despite normal viability and insulin (*INS*) gene expression (Fig. 3b–d). As a result, secretion measures could be normalized by adjusting for insulin content (that is, expressing secretion as a percentage of content), with no significant defects for small interfering *PAM* (si*PAM*) cells (Supplementary Fig. 5a–d). We also noted that content-adjusted rates of insulin secretion were not further impacted by stimulation with either depolarizing KCl or the 3′, 5′-cyclic adenosine monophosphate (cAMP)-elevating agent forskolin (Supplementary Fig. 6a). This indicates that the reductions in insulin secretion following *PAM* knockdown are not caused by a defect in calcium entry or total granule exocytosis.

We next performed quantitative electron microscopy using immunogold labeling of PAM and insulin. The results of this analysis confirmed successful knockdown of PAM at the protein level, with a 47% reduction in the number of particles per vesicle (ppv)

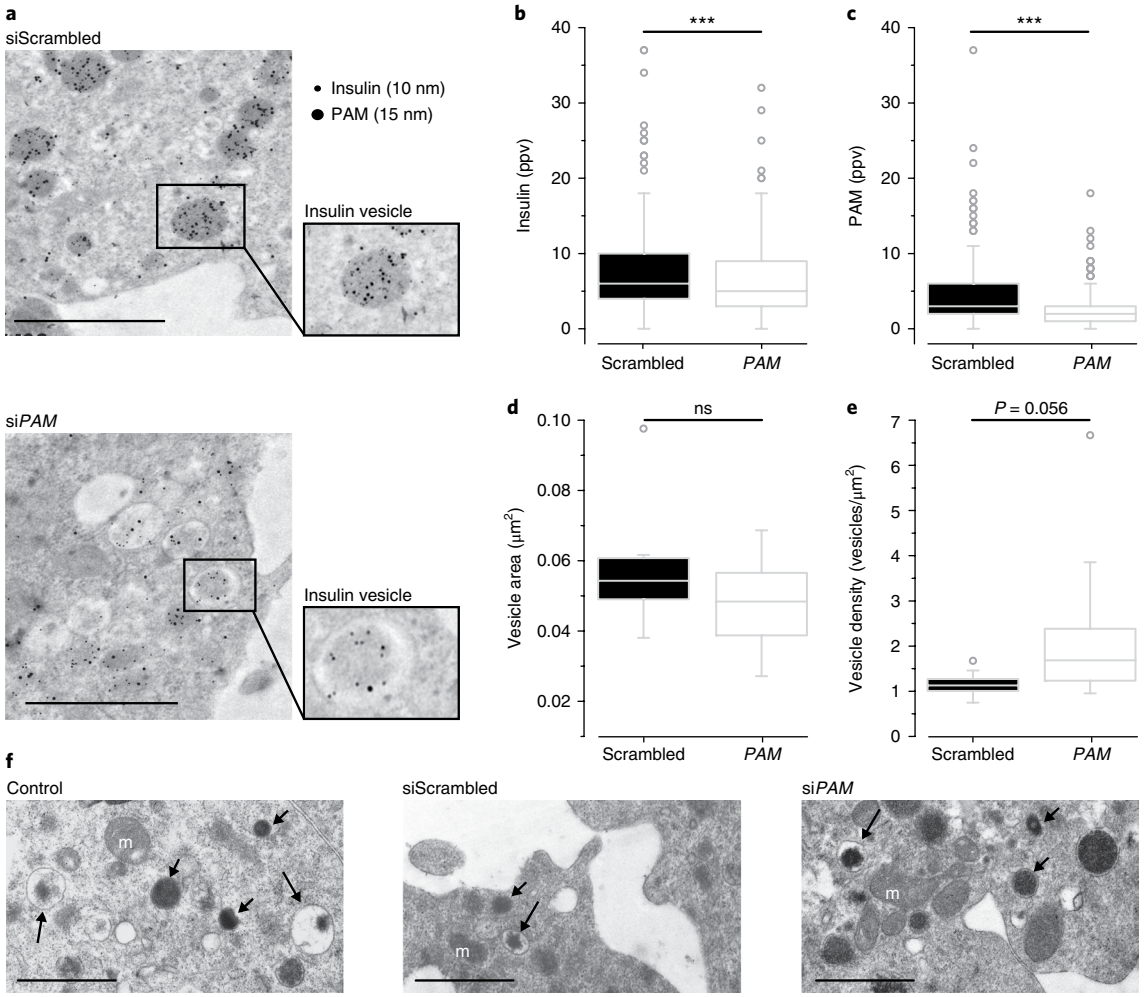


Fig. 4 | Analysis of β-cell ultrastructural features following PAM silencing. a, Electron micrographs from scrambled (siScrambled, upper panel) and *PAM* siRNA-transfected (si*PAM*, lower panel) EndoC-βH1 cells immunogold-labeled for insulin (10 nm) and PAM (15 nm). Scale bar, 2 μm. **b**, Number of immunogold insulin particles per vesicle (ppv). **c**, Number of immunogold PAM ppv. **d**, Vesicle area. **e**, Vesicle density (vesicle per μm^2). Results shown are the average of $n=235$ (scrambled) or $n=290$ (*PAM*) vesicles, except for **e**, where $n=10$ cells per treatment. **f**, Electron micrographs from EndoC-βH1 cells either non-transfected or siRNA-transfected, as indicated. Immature and mature insulin vesicles are highlighted by short and long arrows respectively. m, mitochondrion. *P* values * < 0.05, ** < 0.01, *** < 0.001 for two-tailed Student's *t*-tests. All box plots display the interquartile range (IQR) with medians, whiskers indicating 1.5 × the IQR, and outliers.

(Fig. 4a,c). si*PAM*-treated cells showed a 23% decrease in immunogold-labeled insulin per vesicle compared with control cells (6.26 ± 0.28 versus 8.1 ± 0.41 ppv; $P=1.77\times10^{-4}$) (Fig. 4a,b), thus corroborating the pooled content measures described earlier. There were no significant changes in cross-sectional vesicle area ($0.048\pm0.003\,\mu m^2$ versus $0.056\pm0.005\,\mu m^2$; $P=0.18$) or vesicle density (2.30 ± 0.55 vesicles/μm^2 versus 1.16 ± 0.09 vesicles/μm^2; $P=0.056$) in *PAM* knockdown cells (Fig. 4d,e). No qualitative differences in insulin crystallization were observed (Fig. 4f).

To probe further the effect of *PAM* silencing on insulin secretion, we performed high-resolution, single-cell capacitance measurements. Following stimulation of individual cells with a train of ten depolarizations, an increase in plasma membrane surface area due to vesicle exocytosis could be detected at each step. The cumulative increase in exocytosis was comparable between *PAM* knockdown and control cells (43.9 ± 8.6 fF pF^{-1} versus 38.0 ± 8.1 fF pF^{-1}; $P=0.64$) (Fig. 3e), and there was no difference in calcium entry (peak current 4.5 ± 0.7 pA pF^{-1} versus 3.6 ± 0.6 pA pF^{-1} for *PAM* knockdown versus control cells; $P=0.27$) (Fig. 3g). Moreover, the exocytotic response increased similarly with the calcium current amplitude in both control and *PAM* knockdown cells (Supplementary Fig. 6b), corroborating the normal calcium sensitivity inferred from secretion assays.

In contrast, the kinetics of vesicle release were significantly different; in controls cells, nearly 60% of the exocytotic response was elicited by the first pulse, compared with <40% in *PAM* knockdown cells ($P=0.013$) (Fig. 3f). The increase in cell capacitance during the first two depolarizations provides an estimate for the size of the readily releasable pool (RRP) of secretory granules, whereas subsequent pulses represent mobilization of granules from the more intracellular reserve pool[37]. Thus, our results indicate that PAM is an endogenous regulator of the exocytotic response in pancreatic β-cells, with a specific effect on the availability of the RRP.

Effects of the rs35658696 risk variant on primary islet function. Having established a role for PAM in a human β-cell model, we sought to corroborate our results by examining the effects of the low-frequency *PAM* variant (rs35658696) in primary islets. First, we analyzed insulin secretion and content measures for intact human islets isolated from cadaveric donors. In the Edmonton collection of functionally characterized islets, we identified 16 individuals heterozygous for the T2D risk allele ("carriers") and

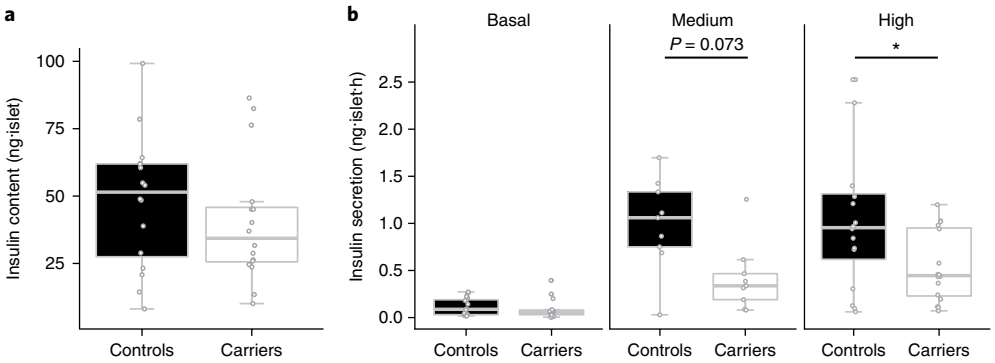


Fig. 5 | Effects of rs35658696 genotype status on insulin secretion measures from intact human islets. a,b, Primary islets isolated from individuals heterozygous for the rs35658696 risk variant (Carriers) and matched individuals homozygous for the reference allele (Controls) were analyzed for total insulin content (**a**; $n=16$ donors in each group) and insulin secretion (**b**) in basal (1 mM; $n=16$), medium (10 mM; $n=9$) and high glucose (16.7 mM; $n=16$). *P* values * < 0.05, ** < 0.01, *** < 0.001 for two-way analysis of covariance (ANCOVA) followed by Tukey's HSD post-hoc test. All box plots display the interquartile range (IQR) with medians, whiskers indicating 1.5 × the IQR, and individual data points.

16 age-, sex-, and body mass index (BMI)-matched donors homozygous for the reference allele ("controls"). All individuals were analyzed for insulin content, as well as secretion under basal (1 mM) and high (16.7 mM) glucose in sequential incubations. In a subset of nine individuals, insulin secretion was also assessed for an intermediate glucose level (10 mM).

We detected no significant change in insulin content (−16.3%; $P=0.37$), but we detected a highly significant reduction in overall insulin secretion for risk carriers compared with controls ($P=7.81\times10^{-4}$ by analysis of covariance (ANCOVA)) (Fig. 5). The effect on secretion showed a positive interaction with glucose levels ($P=0.041$), highlighting a greater difference between risk carriers and controls under stimulated conditions. Thus, testing for pairwise significant differences, we identified no effect of carrier status in basal glucose (−23.8%; $P=1.00$), but a trend toward a reduction in intermediate glucose (−58%; $P=0.073$) and significantly decreased secretion in high glucose (−49%; $P=0.016$). Importantly, these defects in raw insulin secretion could be fully rescued by normalizing to insulin content ($P=0.88$ by ANCOVA), with no significant pairwise differences ($P=1.00$ for basal; $P=1.00$ for intermediate; and $P=0.99$ for high glucose) (Supplementary Fig. 7). Overall, our results establish a defect in intact islets from risk allele carriers highly concordant with that observed following *PAM* knockdown in the EndoC-βH1 model.

Next, we analyzed exocytosis in single β-cells from dispersed human islets, with capacitance measurements following a train of ten depolarizations in either 1 or 10 mM glucose. Seven individuals heterozygous for the T2D risk allele were compared to seven matched controls (Fig. 6a). As with the insulin secretion measures from intact islets, we observed a statistically significant interaction between glucose level and genotype status ($P=2.83\times10^{-4}$ by ANCOVA), with a greater negative effect of the risk allele under stimulated conditions ($P=1.22\times10^{-3}$) than in basal glucose ($P=0.48$). To specifically test for differences in the RRP, we calculated the total increase in capacitance during the first two depolarizations (Fig. 6b). Although no differences were detected in low glucose, a significant decrease in capacitance was seen in risk allele carriers at 10 mM glucose (−26.5%; $P=0.04$), which is indicative of reduced exocytosis from the RRP. Our results are therefore consistent with an effect of the rs35658696 genotype on the kinetics of granule release that manifests predominantly under stimulated conditions.

Chromogranin A is a substrate and plausible downstream effector of PAM in β-cells. To elucidate the PAM-regulated pathway in β-cells, we finally sought to identify possible target proteins that could act as downstream effectors of PAM activity. While any C-terminal glycine-extended peptide could be a substrate for PAM, we reasoned that a key downstream mediator of the observed effects in β-cells would need to be co-expressed with PAM along the secretory pathway. Therefore, we searched a published catalog of the islet secretome for potential targets of PAM amidation[38]. This highlighted two proteins, chromogranin A (CgA, encoded by *CHGA*) and islet amyloid polypeptide (IAPP, encoded by *IAPP*), both of which are highly expressed in β-cells and have plausible links with PAM function. CgA is a granular packaging protein that self-aggregates to condense large numbers of neuroendocrine peptides into nascent secretory granules[39–41]. It can also be cleaved to produce smaller peptides with additional biological roles[42]. The role of IAPP in β-cells is less well understood, although it is known to be stored in secretory granules and co-released with insulin[43].

CHGA silencing (Supplementary Fig. 4b) resulted in a phenotype similar to that observed for *PAM*, with significant reductions in both insulin content and secretion (Fig. 7a–c and Supplementary Fig. 5e–h). Using an antibody specific to the glycine-extended (non-amidated) form of CgA (CgA-Gly), we also observed an accumulation of CgA-Gly in EndoC-βH1 cells treated with *PAM* siRNA by western blot analysis (Fig. 7d,e and Supplementary Fig. 8 for full blots). Treatment of cells with 4-phenyl-3-butenoic acid (4P3BA), an irreversible inhibitor of PAM catalytic activity, produced similar results, showing that this effect is not due to off-target effects. In both cases, the ratio of non-amidated CgA to total CgA was significantly increased following perturbation of PAM (+226%, $P=0.016$ for *PAM* knockdown; +282%, $P=0.043$ for PAM inhibition) (Fig. 7d)[44,45]. Thus, CgA is a plausible PAM substrate in β-cells and a strong candidate for mediating the effects of PAM on β-cell function. In contrast, silencing of *IAPP* in EndoC-βH1 cells had no effect on insulin content or secretion (Supplementary Fig. 4c–f).

Finally, we attempted to detect patterns of CgA amidation in primary human islets by immunostaining pancreatic sections from a subset of six risk allele carriers and six matched controls (a subset of the donors analyzed by electrophysiology). By co-staining for insulin and either CgA-Gly or total CgA, we successfully identified CgA in the endocrine component of the pancreas (Supplementary Fig. 9a). Using a mask to focus exclusively on insulin-positive β-cells, we then quantified average intensities and derived a ratio of non-amidated to total CgA. The results were highly variable, with more than a 15-fold discrepancy between the lowest (6.89%) and highest (109%) ratios. Thus, no overall difference was detected between carriers and controls ($P=0.90$) (Supplementary Fig. 9b); larger studies may be required to detect any amidation defects in risk carriers.

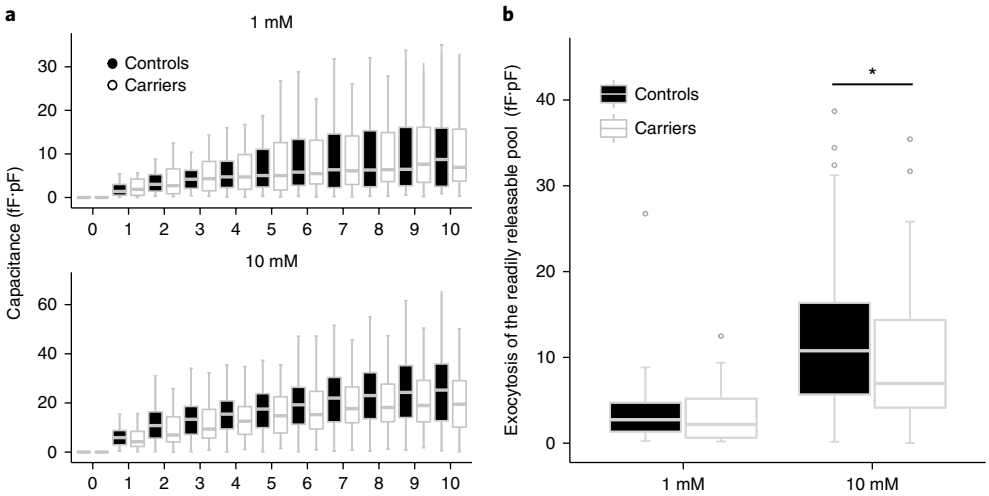


Fig. 6 | Effects of rs35658696 genotype status on exocytosis measurements in dispersed human islets. Capacitance measurements for primary β-cells from seven individuals heterozygous for the rs35658696 risk variant (Carriers; $n = 45$ and 41 cells analyzed in 1 and 10 mM glucose, respectively) and seven individuals homozygous for the reference allele (Controls; $n = 42$ and 40 cells). **a**, Cumulative capacitance measurements after each of ten depolarizations. **b**, Cumulative increase in capacitance for the first two depolarizations. *P* values * < 0.05, ** < 0.01 and *** < 0.001 for two-way analysis of covariance (ANCOVA) followed by Tukey's HSD post-hoc test. All box plots display the interquartile range (IQR) with medians; the whiskers indicate 1.5 × the IQR.

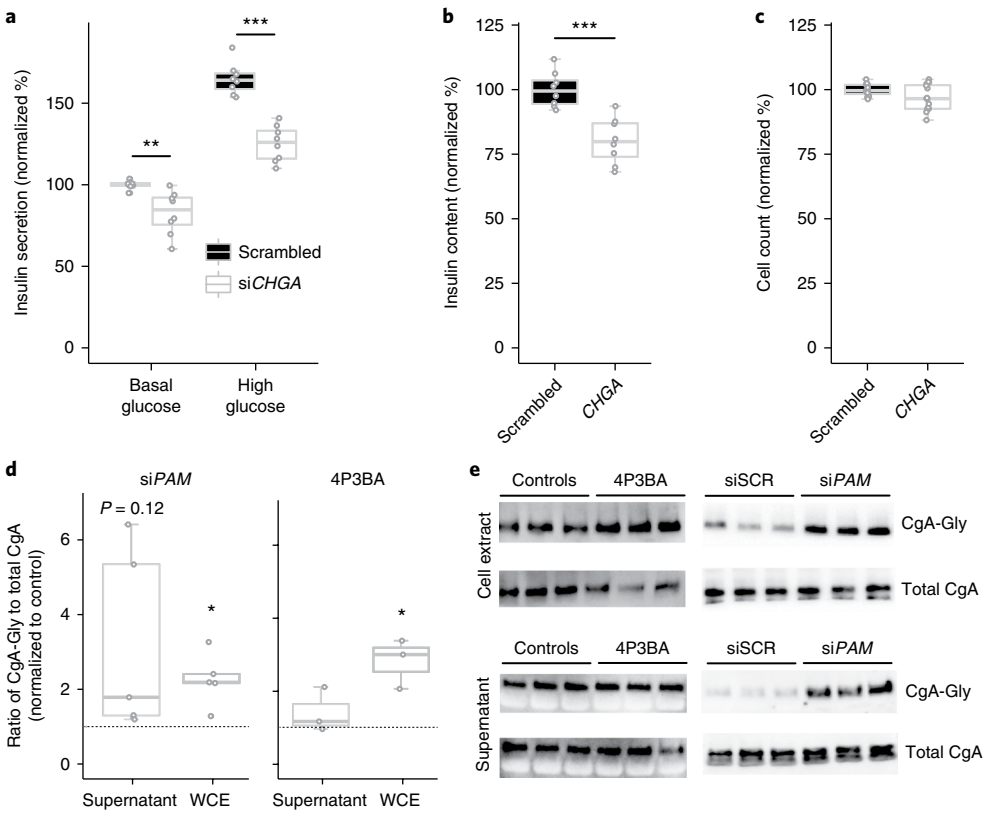


Fig. 7 | Effects of endogenous CgA on β-cell function. EndoC-βH1 cells were transfected with scrambled (siSCR) or *CHGA* siRNA then measured for **a** insulin secretion ($n = 8$ independent biological samples), **b** cellular insulin content ($n = 8$) and **c** cell numbers ($n = 16$). **d**–**e**, EndoC-βH1 cells were transfected with either scrambled or *PAM* siRNA ($n = 5$ independent experiments), or treated with either vehicle (Control) or 4P3BA ($n = 3$ independent experiments), then analyzed by western blot using antibodies specific for non-amidated CgA (CgA-Gly) and total CgA. Control levels are indicated in **d** by the dashed grey line. Results in **e** are representative of $n = 5$ independent experiments for *PAM* silencing, and $n = 3$ independent experiments for PAM inhibition. Corresponding unedited blots are provided in Supplementary Fig. 8. **b,c**, *P* values * <0.05, ** <0.01, *** <0.001 for paired, two-tailed Student's *t*-tests where a single independent variable was being tested. **a,d**, Two-way analysis of covariance (ANCOVA) followed by Tukey's HSD post-hoc test where two experimental variables were being tested. All box plots display the interquartile range (IQR) with medians, whiskers indicating 1.5 × the IQR and individual data points. WCE, whole-cell extract.

Discussion

Recent efforts to catalog coding variants associated with T2D risk have identified two independent association signals in the *PAM* gene[3,6,8]. The coding risk alleles are associated with a decreased insulinogenic index, a physiological measure of insulin secretion in response to a glucose challenge, indicating a likely endogenous role for PAM in β-cell function. Here, we provide evidence that both T2D risk alleles reduce overall PAM activity. While the risk allele at rs78408340 causes defective PAM expression, our results demonstrate that the risk allele at rs35658696 acts through a dual mechanism of reduced catalytic function and, albeit to a lesser extent, decreased expression.

We have further shown that reduced PAM activity causes β-cell dysfunction through multiple effects on insulin availability and the dynamics of granule exocytosis. Importantly, these defects preferentially affect insulin secretion under stimulated conditions, leading to a blunted glucose response. Taken together, our findings support a molecular mechanism whereby T2D risk alleles in *PAM* decrease PAM function, directly impacting on the ability of pancreatic β-cells to mobilize insulin in response to glucose.

This novel mechanism provides an explanation for the relatively larger effect size of rs78408340 (p.Ser539Trp) compared with rs35658696 (p.Asp563Gly), as reported in two independent GWAS for T2D risk and insulinogenic index[6,8]. Mislocalization or misexpression of p.Ser539Trp PAM would presumably compromise function more severely than the combined reduction in amidation activity and expression observed for p.Asp563Gly PAM. This is particularly true for cell types in which the luminal form predominates over the integral form of PAM, since misexpressed p.Ser539Trp PAM may retain some residual function. The precise cellular dynamics that yield luminal PAM in β-cells have not been elucidated; however, it appears generally to be more predominant in neuroendocrine cell types[17]. This raises the intriguing possibility of tissue-specific differences in PAM activity for rs78408340 risk allele carriers.

To explore the molecular mechanism underlying β-cell dysfunction in *PAM* risk allele carriers, we performed detailed characterizations in primary islets and the human β-cell line EndoC-βH1. The observed defects were broadly similar, with significant reductions in insulin secretion driven by decreased insulin content. Further, the size of the RRP was diminished in both primary β and EndoC-βH1 cells. However, total exocytosis remained unaffected, which is consistent with the lack of effect on content-normalized insulin secretion. Taken together, these results highlight a defect in insulin mobilization mediated through granule biogenesis and trafficking.

Interestingly, the phenotype caused by reduced PAM activity was consistently more pronounced under stimulated conditions. In EndoC-βH1 cells, insulin secretion was not affected in low glucose and the stimulation index (ratio of high to basal secretion) was significantly reduced. In primary β-cells, defects in exocytosis were apparent only in 10 mM glucose, with normal levels in basal glucose. Moreover, insulin secretion from intact islets was significantly reduced under stimulated conditions only and with a significant interaction between genotype status and glucose levels. *PAM* loss-of-function in β-cells thus manifests preferentially under conditions resembling the fed state in vivo, in line with the association of the T2D risk alleles with a reduced insulinogenic index.

Our molecular studies in EndoC-βH1 cells also confirmed that CgA is a substrate for PAM amidation in β-cells[45]. CgA is a granular packaging protein known to influence vesicle composition in other neuroendocrine cell types by facilitating the condensation of secreted proteins within nascent secretory vesicles[39–41,46–48]. Previous studies using global *Chga* knockout mouse models have attempted to investigate a role for CgA in pancreatic islets. However, they have been confounded by compensatory increases in other granin proteins, including chromogranin B[49,50]. In mice, knockdown of chromogranin B leads to strong inhibition of insulin secretion by mechanisms that have not been resolved[51]. Our results suggest a model whereby amidation of CgA by PAM enhances its ability to aggregate and package insulin efficiently in granules. In support of this hypothesis, amidation of numerous biological peptides has been shown to promote protein structure and self-aggregation due to improved electrostatic interactions[52–54]. C-terminal cleavage of CgA also increases protein stability in pituitary cells[55,56]. More detailed studies in primary islets is required to confirm any defective amidation of CgA in risk allele carriers.

Although our tissue expression analyses showed that *PAM* is highly expressed in pancreatic islets, the results also revealed *PAM* to be widely expressed across a range of other cell types. It is therefore possible that *PAM* could indirectly modulate β-cell function via amidated signals released from other tissues[23,57]. Glucagon-like peptide 1 (GLP-1) is an amidated peptide released from intestinal L cells that amplifies insulin secretion in response to a meal[58]. Moreover, GLP-1 (7–36) amide has an increased half-life compared to GLP-1 (7–37) in vitro, suggesting that amidation could lead to elevated GLP-1 levels in the circulation[59]. Therefore, PAM may have endocrine effects on β-cell function via decreased GLP-1 amidation and/or secretion, with possible implications for genotype-specific responses to incretin-based therapies.

Within the β-cell, it remains likely that additional PAM substrates may exist, some of which could further modulate insulin mobilization. We also cannot exclude the possibility of pleiotropic effects at the level of the T2D risk alleles; nevertheless, the presence of two independent coding variants in *PAM* strongly favor a primary causal mechanism mediated through *PAM*. This is further supported by our experimental findings, which establish direct impacts on PAM function and expression.

In conclusion, we have shown that the amidating enzyme PAM is a critical component of the regulated secretory pathway in β-cells, consistent with the observed effects of *PAM* missense variants on T2D risk and the insulinogenic index. Our results establish molecular mechanisms for T2D-associated *PAM* risk alleles and show multiple effects of reduced PAM function on insulin granule packaging and release. Our study illustrates the importance of coding variants implicated in disease risk for the identification of candidate causal genes. Further work is required to explore the possible contribution of other tissues to the effect of *PAM* risk alleles, particularly those that have implications for incretin-based therapies.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41588-018-0173-1.

Received: 21 September 2017; Accepted: 6 June 2018;
Published online: 27 July 2018

References

1. Dimas, A. S. et al. Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. *Diabetes* **63**, 2158–2171 (2014).
2. Morris, A. P. et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* **44**, 981–990 (2012).
3. Fuchsberger, C. et al. The genetic architecture of type 2 diabetes. *Nature* **536**, 41–47 (2016).
4. Mahajan, A. et al. Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes. *Nat. Genet.* **50**, 559–571 (2018).
5. Thomsen, S. K. & Gloyn, A. L. Human genetics as a model for target validation: finding new therapies for diabetes. *Diabetologia* **60**, 960–970 (2017).
6. Steinthorsdottir, V. et al. Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. *Nat. Genet.* **46**, 294–298 (2014).

7. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
8. Huyghe, J. R. et al. Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. *Nat. Genet.* **45**, 197–201 (2013).
9. Phillips, D. I., Clark, P. M., Hales, C. N. & Osmond, C. Understanding oral glucose tolerance: comparison of glucose or insulin measurements during the oral glucose tolerance test with specific measurements of insulin resistance and insulin secretion. *Diabet. Med.* **11**, 286–292 (1994).
10. Milgram, S. L., Mains, R. E. & Eipper, B. A. Identification of routing determinants in the cytosolic domain of a secretory granule-associated integral membrane protein. *J. Biol. Chem.* **271**, 17526–17535 (1996).
11. Oyarce, A. M. & Eipper, B. A. Neurosecretory vesicles contain soluble and membrane-associated monofunctional and bifunctional peptidylglycine alpha-amidating monooxygenase proteins. *J. Neurochem.* **60**, 1105–1114 (1993).
12. Eipper, B. A., Milgram, S. L., Husten, E. J., Yun, H. Y. & Mains, R. E. Peptidylglycine alpha-amidating monooxygenase: a multifunctional protein with catalytic, processing, and routing domains. *Protein Sci.* **2**, 489–497 (1993).
13. Merkler, D. J. C-terminal amidated peptides: production by the *in vitro* enzymatic amidation of glycine-extended peptides and the importance of the amide to bioactivity. *Enzym. Microb. Technol.* **16**, 450–456 (1994).
14. Milgram, S. L., Kho, S. T., Martin, G. V., Mains, R. E. & Eipper, B. A. Localization of integral membrane peptidylglycine alpha-amidating monooxygenase in neuroendocrine cells. *J. Cell Sci.* **110**, 695–706 (1997).
15. Eipper, B. A. et al. Alternative splicing and endoproteolytic processing generate tissue-specific forms of pituitary peptidylglycine alpha-amidating monooxygenase (PAM). *J. Biol. Chem.* **267**, 4008–4015 (1992).
16. Garmendia, O., Rodriguez, M. P., Burrell, M. A. & Villaro, A. C. Immunocytochemical finding of the amidating enzymes in mouse pancreatic A-, B-, and D-cells: a comparison with human and rat. *J. Histochem. Cytochem.* **50**, 1401–1416 (2002).
17. Tausk, F. A., Milgram, S. L., Mains, R. E. & Eipper, B. A. Expression of a peptide processing enzyme in cultured cells: truncation mutants reveal a routing domain. *Mol. Endocrinol.* **6**, 2185–2196 (1992).
18. De, M., Bell, J., Blackburn, N. J., Mains, R. E. & Eipper, B. A. Role for an essential tyrosine in peptide amidation. *J. Biol. Chem.* **281**, 20873–20882 (2006).
19. Maeda-Nakai, E. & Ichiyama, A. A spectrophotometric method for the determination of glycolate in urine and plasma with glycolate oxidase. *J. Biochem.* **127**, 279–287 (2000).
20. Carpenter, S. E. & Merkler, D. J. An enzyme-coupled assay for glyoxylic acid. *Anal. Biochem.* **323**, 242–246 (2003).
21. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
22. Nica, A. C. et al. Cell-type, allelic, and genetic signatures in the human pancreatic beta cell transcriptome. *Genome Res.* **23**, 1554–1562 (2013).
23. Blodgett, D. M. et al. Novel observations from next-generation RNA sequencing of highly purified human adult and fetal islet cell subsets. *Diabetes* **64**, 3172–3181 (2015).
24. Ravassard, P. et al. A genetically engineered human pancreatic β cell line exhibiting glucose-inducible insulin secretion. *J. Clin. Invest.* **121**, 3589–3597 (2011).
25. Tsonkova, V. G. 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.* **8**, 144–157 (2018).
26. Scharfmann, R. et al. Persistence of peptidylglycine alpha-amidating monooxygenase activity and elevated thyrotropin-releasing hormone concentrations in fetal rat islets in culture. *Endocrinology* **123**, 1329–1334 (1988).
27. Zhou, A. & Thorn, N. A. Evidence for presence of peptide alpha-amidating activity in pancreatic islets from newborn rats. *Biochem. J.* **267**, 253–256 (1990).
28. Maltese, J. Y. et al. Ontogenetic expression of peptidyl-glycine alpha-amidating monooxygenase mRNA in the rat pancreas. *Biochem. Biophys. Res. Commun.* **158**, 244–250 (1989).
29. Martinez, A. et al. Immunocytochemical localization of peptidylglycine alpha-amidating monooxygenase enzymes (PAM) in human endocrine pancreas. *J. Histochem. Cytochem.* **41**, 375–380 (1993).
30. Gurgul-Convey, E., Kaminski, M. T. & Lenzen, S. Physiological characterization of the human EndoC-βH1 β-cell line. *Biochem. Biophys. Res. Commun.* **464**, 13–19 (2015).
31. El Meskini, R., Mains, R. E. & Eipper, B. A. Cell type-specific metabolism of peptidylglycine alpha-amidating monooxygenase in anterior pituitary. *Endocrinology* **141**, 3020–3034 (2000).
32. Andersson, L. E. et al. Characterization of stimulus-secretion coupling in the human pancreatic EndoC-βH1 beta cell line. *PLoS One* **10**, e0120879 (2015).
33. Sun, B.B. et al. Consequences of natural perturbations in the human plasma proteome. *bioRxiv* https://doi.org/10.1101/134551 (2017).
34. van de Bunt, M. 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.* **11**, e1005694 (2015).
35. van de Geijn, B., McVicker, G., Gilad, Y. & Pritchard, J. K. WASP: allele-specific software for robust molecular quantitative trait locus discovery. *Nat. Methods* **12**, 1061–1063 (2015).
36. Thurner, M. et al. Integration of human pancreatic islet genomic data refines regulatory mechanisms at Type 2 Diabetes susceptibility loci. *Elife* **7**, e31977 (2018).
37. Gillis, K. D., Mossner, R. & Neher, E. Protein kinase C enhances exocytosis from chromaffin cells by increasing the size of the readily releasable pool of secretory granules. *Neuron* **16**, 1209–1220 (1996).
38. Yang, Y. H. et al. Paracrine signalling loops in adult human and mouse pancreatic islets: netrins modulate beta cell apoptosis signalling via dependence receptors. *Diabetologia* **54**, 828–842 (2011).
39. Kim, T., Tao-Cheng, J. H., Eiden, L. E. & Loh, Y. P. Chromogranin A, an "on/off" switch controlling dense-core secretory granule biogenesis. *Cell* **106**, 499–509 (2001).
40. Colomer, V., Kicska, G. A. & Rindler, M. J. Secretory granule content proteins and the luminal domains of granule membrane proteins aggregate *in vitro* at mildly acidic pH. *J. Biol. Chem.* **271**, 48–55 (1996).
41. Yoo, S. H. & Albanesi, J. P. Ca^{2+}-induced conformational change and aggregation of chromogranin A. *J. Biol. Chem.* **265**, 14414–14421 (1990).
42. Helle, K. B., Corti, A., Metz-Boutigue, M. H. & Tota, B. The endocrine role for chromogranin A: a prohormone for peptides with regulatory properties. *Cell. Mol. Life Sci.* **64**, 2863–2886 (2007).
43. Rorsman, P. & Braun, M. Regulation of insulin secretion in human pancreatic islets. *Annu. Rev. Physiol.* **75**, 155–179 (2013).
44. Katopodis, A. G. & May, S. W. Novel substrates and inhibitors of peptidylglycine alpha-amidating monooxygenase. *Biochemistry* **29**, 4541–4548 (1990).
45. Simpson, P. D. et al. Striking oxygen sensitivity of the peptidylglycine alpha-amidating monooxygenase (PAM) in neuroendocrine cells. *J. Biol. Chem.* **290**, 24891–24901 (2015).
46. Yoo, S. H. & Lewis, M. S. Dimerization and tetramerization properties of the C-terminal region of chromogranin A: a thermodynamic analysis. *Biochemistry* **32**, 8816–8822 (1993).
47. Mosley, C. A. et al. Biogenesis of the secretory granule: chromogranin A coiled-coil structure results in unusual physical properties and suggests a mechanism for granule core condensation. *Biochemistry* **46**, 10999–11012 (2007).
48. Bandyopadhyay, G. K. & Mahata, S. K. Chromogranin A regulation of obesity and peripheral insulin sensitivity. *Front. Endocrinol. (Lausanne)* **8**, 20 (2017).
49. Wollam, J. et al. Chromogranin A regulates vesicle storage and mitochondrial dynamics to influence insulin secretion. *Cell Tissue Res.* **368**, 487–501 (2017).
50. Bartolomucci, A. et al. The extended granin family: structure, function, and biomedical implications. *Endocr. Rev.* **32**, 755–797 (2011).
51. Obermüller, S. et al. Defective secretion of islet hormones in chromogranin-B deficient mice. *PLoS One* **5**, e8936 (2010).
52. Nanga, R. P., Brender, J. R., Vivekanandan, S. & Ramamoorthy, A. Structure and membrane orientation of IAPP in its natively amidated form at physiological pH in a membrane environment. *Biochim. Biophys. Acta* **1808**, 2337–2342 (2011).
53. Shalev, D. E., Mor, A. & Kustanovich, I. Structural consequences of carboxyamidation of dermaseptin S3. *Biochemistry* **41**, 7312–7317 (2002).
54. Sforça, M. L. et al. How C-terminal carboxyamidation alters the biological activity of peptides from the venom of the eumenine solitary wasp. *Biochemistry* **43**, 5608–5617 (2004).
55. Kim, T. & Loh, Y. P. Protease nexin-1 promotes secretory granule biogenesis by preventing granule protein degradation. *Mol. Biol. Cell* **17**, 789–798 (2006).
56. Koshimizu, H., Cawley, N. X., Kim, T., Yergey, A. L. & Loh, Y. P. Serpinin: a novel chromogranin A-derived, secreted peptide up-regulates protease nexin-1 expression and granule biogenesis in endocrine cells. *Mol. Endocrinol.* **25**, 732–744 (2011).
57. Benner, C. 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 Genom.* **15**, 620 (2014).
58. MacDonald, P. E. et al. The multiple actions of GLP-1 on the process of glucose-stimulated insulin secretion. *Diabetes 51* **Suppl 3**, S434–S442 (2002).
59. Wettergren, A., Pridal, L., Wøjdemann, M. & Holst, J. J. Amidated and non-amidated glucagon-like peptide-1 (GLP-1): non-pancreatic effects (cephalic phase acid secretion) and stability in plasma in humans. *Regul. Pept.* **77**, 83–87 (1998).

Acknowledgements

We acknowledge sharing of data from the GoT2D and T2D-GENES consortia before publication. We thank J. Galvanovskis (University of Oxford) for microscopy assistance, J. Lyon (Alberta Diabetes Institute IsletCore) for his work on human islet isolations, and J. Buteau and Y. Wang (both Alberta Diabetes Institute) for their assistance with imaging human pancreatic sections. We also thank the Human Organ Procurement and Exchange Program (Edmonton) and the Trillium Gift of Life Network (Toronto) and other organ

procurement agencies for their efforts in obtaining human pancreata for research. A.L.G. is a Wellcome Trust Senior Fellow in Basic Biomedical Science. P.E.M. holds a 2016–2017 Killam Annual Professorship. M.I.M. is a Wellcome Senior Investigator. S.K.T. is a Radcliffe Department of Medicine Scholar. S.S. is funded by the Medical Research council (MC_ST_15019 [2015 DTG/DTA]). J.C. is funded by the Oxford – Medical Research Council Doctoral Training Partnership and the Nuffield Department of Clinical Medicine. This work was funded by the Wellcome Trust (095101 (A.L.G.), 200837 (A.L.G.), 098381 (M.I.M.), 106130 (A.L.G., M.I.M.), 203141 (M.I.M.), 090531 (P.R.)), Medical Research Council (MR/L020149/1) (M.I.M., A.L.G., P.R.), European Union Horizon 2020 Programme (T2D Systems) (A.L.G.), and National Institutes of Health (U01-DK105535; U01-DK085545) (M.I.M., A.L.G.). Human islet isolation and phenotyping was supported by funding from the Alberta Diabetes Foundation (P.E.M.F.). 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, NIHR or the Department of Health.

Author contributions

S.K.T., A.R., B.H., and A.L.G. conceived the study. A.R., B.H., S.S., M.M.U., A. Barrett, C.J.G., N.L.B., P.R., and A.L.G. performed kinetic and cellular characterization of T2D-associated alleles. S.K.T., A.R., B.H., S.S., A.C., P.R., and A.L.G. performed the characterization of PAM knockdown in β-cells. S.K.T., X.-Q.D., A. Bautista, A.F.S., J.E.M.F., P.E.M., and A.L.G. performed characterization of primary human islets. A.J.P., M.I.M., and A.L.G. performed islet transcriptomics. J.C., M.I.M., and A.M. performed plasma proteomics. A.R., S.K.T., and A.L.G. wrote the manuscript. All authors approved the final draft of the manuscript.

Competing interests

S.K.T. is now an employee of Vertex Pharmaceuticals, and N.L.B. is now an employee of Novo Nordisk, although all experimental work was carried out under employment at the University of Oxford. A.L.G. has received research funding and honoraria from Novo Nordisk. M.I.M. serves on advisory panels for Pfizer, Novo Nordisk and Zoe Global; has received honoraria from Pfizer, Novo Nordisk and Eli Lilly; has stock options in Zoe Global; and has received research funding from AbbVie, Astra Zeneca, Boehringer Ingelheim, Eli Lilly, Janssen, Merck, Novo Nordisk, Pfizer, Roche, Sanofi Aventis, Servier and Takeda.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0173-1.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to A.L.G.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Ethics statement. All studies were approved by the University of Oxford's Tropical Research Ethics Committee (OxTREC reference 2–15) or the Oxfordshire Regional Ethics Committee B (REC reference 09/H0605/2). Human pancreatic tissue was obtained from subjects at postmortem examination according to local and national ethics permissions. Human pancreatic islets were isolated, with ethical approval and clinical consent, at the Diabetes Research and Wellness Foundation Human Islet Isolation Facility (Oxford Centre for Diabetes, Endocrinology and Metabolism, Oxford, UK). All animal experiments were conducted in accordance with the UK Animals (Scientific Procedures) Act 1986 and the University of Oxford ethical guidelines, and were approved by the local ethical committees.

Cloning of PAM expression plasmids. We purchased a pCMV6 expression vector encoding integral membrane *Homo sapiens PAM* (NM_138821.2) with an in-frame C-terminal Myc-Flag tag (PAM-Myc-Flag) from Origene. Luminal functional PAM-Myc-Flag encompassing the first 710 amino acids was generated via PCR using the primers luminal (F) and luminal (R) (Supplementary Table 2). Puromycin-resistant PAM-Myc-Flag expression vectors were generated via BglII/XmaI digestion, blunt ending with T4 DNA polymerase and subcloning into HpaI-digested pIRES puro 2 (Takara Bio USA). Protein alterations were introduced via site-directed mutagenesis using the Stratagene QuikChange II kit (Agilent Biotechnologies), according to manufacturer's instructions, with the following primers: SDM_S539W_F and SDM_S539W_R for generating p.Ser539Trp PAM; SDM_D563G_F and SDM_D563G_R for generating p.Asp563Gly PAM; and SDM_Y651F_F and SDM_Y651F_R for generating p.Tyr651Phe PAM (Supplementary Table 2). All plasmids were verified by sequencing.

Cell culture and transfection. Primary human islets were freshly isolated from deceased donors at the Oxford Centre for Islet Transplantation in Oxford, UK or at the Alberta Diabetes Institute IsletCore as previously described[60,61]. Islets were then cultured in CMRL medium (Thermo Fisher Scientific) at 37 °C or kept in UW cold storage solution (Viaspan) at 4 °C for short-term storage before being processed. HEK293 cells were routinely passaged in DMEM 6429 (Sigma-Aldrich) supplemented with 10% fetal calf serum, 50 international units ml^{-1} penicillin and 50 µg ml^{-1} streptomycin. EndoC-βH1 cells were routinely passaged in growth medium supplemented with 5.5 mM glucose[24]. All cells were maintained at 37 °C and 5% CO_2.

To generate PAM stable cell lines, HEK293 cells were transfected using FuGENE 6 (Promega Corporation), according to the manufacturer's instructions. Cells were routinely passaged and maintained them in 1 µg ml^{-1} puromycin. For fluorescence microscopy, EndoC-βH1 cells were transfected in chamber slides (BD Biosciences) using FuGENE 6. For siRNA experiments, EndoC-βH1 cells were seeded in Opti-MEM reduced serum-free medium (Thermo Fisher Scientific) containing 25 nM ON-TARGET *plus* SMARTpool siRNA complexes (Dharmacon) preformed with Lipofectamine RNAiMAX transfection reagent (Thermo Fisher Scientific), according to the manufacturer's instructions. Growth medium was replaced 24 h later and cells were incubated for a further 48 h. For PAM inhibitor experiments, seeded cells were treated with 500 µM 4P3BA or dimethylsulfoxide vehicle (Sigma-Aldrich) for 72 h.

Recombinant PAM protein production. HEK293 stable cell lines were seeded into Nunc TripleFlasks (Thermo Fisher Scientific). The medium was replaced with DMEM 1145, supplemented with 1 mM sodium pyruvate and incubated for 3–5 days. Supernatant containing luminal recombinant PAM protein was harvested directly from the flasks, filtered into a sterile storage bottle (Corning GmbH) and pH corrected to 5.5 with HCl. The filtered supernatant was concentrated using a Centricon Plus-70 Centrifugal Filter Device (Millipore Limited) with a 30 kDa molecular weight cutoff, according to the manufacturer's instructions.

Quantitation of recombinant PAM. For the relative quantitation of recombinant PAM proteins, 3 × 1 µl aliquots of concentrated supernatant were separated on a 4–20% Criterion TGX Stain-Free Precast Gel (Bio-Rad Laboratories, Inc) and transferred to polyvinylidene difluoride using a Trans-blot Turbo Transfer Pack (Bio-Rad Laboratories), according to the manufacturer's instructions. Membranes were incubated overnight in 4A6 anti-Myc tag antibody (Millipore Limited) at 4 °C and developed using the Clarity Western ECL Blotting Substrate (Bio-Rad Laboratories), according to manufacturer's instructions. Proteins were visualized on a ChemiDoc MP (Bio-Rad Laboratories). Lanes and bands were drawn in the Image Lab version 5.0 for Windows software; the average band volume (intensity) was quantified for each triplicate of samples. The amount of recombinant variant PAM protein per µl was then calculated relative to wild-type PAM.

Kinetic analysis of PAM amidating activity. We used the spectrophotometric detection of chemically converted glyoxylate to measure the amidating activity of PAM[20]. An assay mix of 100 mM MES pH 5.5, 2 µM copper sulfate, 10 mM L-ascorbic acid, 0.1 mg ml^{-1} bovine liver catalase and 20 mM hippuric acid (Sigma-Aldrich) was split into glass vials. Equimolar amounts of recombinant PAM were added to each vial and incubated in a 37 °C water bath. Every 30 min, samples were removed to a conical 96-well plate (4titude Limited) containing 40 mM EDTA pH 8.0. The glyoxylate produced was converted to glyoxylate phenylhydrazone by adding 0.1% phenylhydrazine-HCl (Sigma-Aldrich) to the plate, incubating it at room temperature for 15 min and saturating it with concentrated (37%) HCl. The contents of the entire plate were then aspirated and dispensed into a flat-bottomed 96-well plate (Corning GmbH) containing 0.2% potassium ferricyanide using an i-Pipette Automated Pipettor (Apricot Designs). We measured the absorbance of 1,5-diphenylformazan at 515 nm using a VersaMax Microplate Reader (Molecular Devices, LLC).

Kinetic analysis of variant PAM substrate affinity. For the K_m calculations, a 12-point, twofold serial dilution of hippuric acid in which the maximum concentration was 20 mM was prepared in the assay mix (100 mM MES pH 5.5, 2 µM $CuSO_4$, 10 mM L-ascorbic acid, 0.1 mg ml^{-1} bovine liver catalase). Recombinant PAM or empty vector control supernatant was then added. The amount of p.Asp563Gly PAM added to the assay was increased so that it displayed an equal V_{max} to wild-type PAM. The volume of empty vector supernatant matched the volume of wild-type PAM used in that experiment. Vials were incubated in a 37 °C water bath for 1 h. Samples were then removed into a conical 96-well plate containing 40 mM EDTA pH 8.0.

Gene expression analyses. Gene expression was measured with TaqMan real-time PCR assays. RNA was purified from TRIzol (Thermo Fisher Scientific) homogenates or sourced from commercially available RNA tissue panels: Human Total RNA Master Panel II (Takara Bio USA); FirstChoice Human Total RNA Survey Panel (Thermo Fisher Scientific); and mouse Total RNA Master Panel (Thermo Fisher Scientific). Pooled human adult pancreas RNA from three donors, and pooled human adult islet RNA from three donors, was sourced from the Oxford Islet Biobank[34]. Pooled mouse islet RNA was sourced from four 12–28-week-old Naval Medical Research Institute mice. Hypothalamic and pituitary RNA were commercially sourced from AMS Biotechnology Ltd. Complementary DNA was generated using the GoScript Reverse Transcription System (Promega Corporation), according to the manufacturer's instructions, and analyzed with the 7900HT Fast Real-Time PCR System (Thermo Fisher Scientific). C_t values were normalized to the combined average of the housekeeping genes peptidylprolyl isomerase A (*PPIA*) and TATA-box binding protein (*TBP*). *PAM* expression in sorted β-cells and non-β-cells was analyzed based on previously published RNA sequencing data[22].

Allele-specific expression was performed using publicly available RNA sequencing data from the previously described islets of 118 individuals[34]. In brief, 40 samples were freshly isolated at the Oxford Centre for Islet Transplantation as described[60]. The other samples, from Edmonton, Canada, were either extracted from the long-term cryopreserved biobank and thawed as described[62] ($n = 65$) or freshly isolated from donor pancreas ($n = 13$). Poly-A selected libraries were prepared from total RNA at the Oxford Genomics Centre, multiplexed (3 samples per lane) and clustered; 100 nucleotide paired-end reads were sequenced using Illumina TruSeq v3 chemistry on the Illumina HiSeq 2000 Sequencing System (Illumina, Inc). DNA was extracted from the spleen or the exocrine fraction of the islet for most donors, or from the islets when no other tissue was available. All samples were genotyped on the Illumina Omni2.5 Exome-8 version 1.3 BeadChip array.

In total, 11 out of 118 individuals were heterozygous at rs35658696 and were thus used for allele-specific expression analysis. Reads were mapped with the WASP pipeline[35] combined with the Spliced Transcripts Alignment to a Reference (STAR) aligner[63] and based on human genome assembly GRCh37.75[64]. The WASP pipeline removes reads with evidence of mapping bias. Final allele counts were then calculated with an allele counter[65] and a binomial test for allelic imbalance was performed in R against a null allelic proportion of 0.5.

Associations for rs35658696 (p.Asp563Gly) and rs78408340 (p.Ser539Trp) with PAM protein plasma levels in humans were looked up in the supplementary material to the article by Sun et al.[33]. Association analyses in their study had been performed within a linear regression framework under an additive model, adjusting for age, sex, time between sampling and processing, and three principal components. Effect directions were then aligned to correspond to the T2D risk alleles.

Immunofluorescent labeling and microscopy. Human pancreatic specimens fixed in 10% formalin and embedded in wax were used to localize PAM in islets. Sections of 4 µm were de-waxed, rehydrated and heated to 95 °C in 10 mmol l^{-1} Tris + 1 mmol l^{-1} EDTA pH 6.0 for 10 min for antigen retrieval. Tissue sections were incubated overnight in anti-PAM antibody (ab109175; Abcam), guinea pig anti-insulin antibody (in-house antibody) and mouse anti-glucagon antibody (G2654; Sigma-Aldrich). PAM immunoreactivity was enhanced using a tyramide signal amplification step (Thermo Fisher Scientific) and visualized with Alexa Fluor 488 (Thermo Fisher Scientific). Insulin and glucagon were visualized with Alexa Fluor 633 (Thermo Fisher Scientific) and anti-mouse tetramethylrhodamine (Sigma-Aldrich), respectively.

EndoC-βH1 cells transfected with *PAM* expression vectors were fixed in 4% paraformaldehyde, permeabilized in 0.1% (v/v) Triton X-100 and blocked in 10% (w/v) bovine serum albumin. Primary antibodies (4A6 anti-Myc, Belfast anti-insulin, TGN46 anti-trans Golgi (Sigma-Aldrich) and H-70 anti-calnexin (sc-11397; Santa Cruz Biotechnology, Inc)) were incubated overnight at 4 °C.

Slides were then incubated in Alexa Fluor-conjugated secondary antibodies (Thermo Fisher Scientific) and mounted in VECTASHIELD Mounting Medium with DAPI (Vector Laboratories Ltd). Cells were visualized with a Zeiss LSM 510 META confocal laser scanning module arranged on an Axiovert 200 microscope and a Plan-Apochromat 63×/1.4 oil immersion objective (Carl Zeiss). An argon laser was used to excite Alexa Fluor 488 at λ = 488 nm; helium–neon lasers 543 and 633 were used to excite Alexa Fluor 543 at λ = 543 nm and Alexa Fluor 633 at λ = 633nm, respectively. The signals were collected between 640 and 694 nm for Alexa Fluor 633, between 512 and 533 nm for Alexa Fluor 488, and between 587 and 619 nm for Alexa Fluor 543. The main dichroic beam splitter (HFT 488/543/633) was combined with a secondary one, either the NFT490 or NFT545 or NFT635vis. DAPI was excited in two-photon mode using the 740-nm line of an infrared light Chameleon laser combined with an HFT KP 650 filter (Carl Zeiss).

Gene silencing in EndoC-βH1. To perform gene silencing by RNA interference, a master mix of Opti-MEM reduced serum-free medium (Thermo Fisher Scientific) was pre-mixed with 0.4% RNAiMAX transfection reagent (Thermo Fisher Scientific) and 50 nM siRNA (commercially available SMARTpools from Dharmacon against *PAM*, *CHGA* or *IAPP* as indicated). The solution was then incubated at room temperature for 15 min to form siRNA–lipid complexes before 50 µl of the master mix was added to each well on a 96-well black OPTI-plate (Corning GmbH). Cells resuspended in standard culturing media were added in a 1:1 volume until a final density of 75,000 cells cm^{-2} and siRNA concentrations of 25 nM were reached. After 24 h, the transfection media were replaced with complete culturing media. After a further 2 days, cells were taken forward for functional assays or gene expression analyses as required.

Insulin secretion assays in EndoC-βH1. Seventy-two hours after gene silencing, cells prepared as described earlier were starved overnight in standard cell culture medium supplemented with glucose to a final concentration of 2.8 mM. Following overnight incubation, cells were further incubated for 1 h in culture medium with 0 mM glucose. The medium was then replaced with the appropriate conditions to initiate secretion in static incubation. After 1 h, supernatants were harvested in a staggered fashion, leaving a residual volume of supernatant (~30 µl in 96-well format) to minimize disturbances of the adherent cells. Supernatant samples were then centrifuged at 300 *g* in a pre-chilled centrifuge for 5 min and a small volume of the supernatant was transferred to a new plate.

Meanwhile, adherent cells were processed for cell counting based on the CyQUANT Direct Cell Proliferation Kit (Thermo Fisher Scientific), according to the manufacturer's recommendations. Briefly, residual medium was replaced with medium supplemented with 0.2% direct nucleic acid dye and 1% background suppressor dye. Cells were then incubated for 1 h at 37 °C in the dark and fluorescence was measured on the EnSpire Multimode Plate Reader (PerkinElmer) using fluorescein isothiocyanate filter settings (ex480/em535 nm). Cell counts were expressed as relative fluorescence units (RFU) normalized to the earliest time point. Following the CyQUANT assays, cells were washed once in standard media, before adding ice-cold acid-ethanol buffer (1.5% concentrated HCl, 75% ethanol and 23.5% deionized water) to extract the intracellular insulin content. Plates were then covered with parafilm to prevent evaporation and stored at −20 °C for short-term storage or −80 °C for long-term storage.

To analyze the insulin samples, the Insulin (human) AlphaLISA Detection Kit (PerkinElmer) was used to measure the concentrations of both supernatants and cellular contents. Samples were diluted in 1 × AlphaLISA immunoassay buffer (1:10 for supernatants and 1:200 for contents) and then analyzed in white 1/2 AreaPlate-96 (PerkinElmer) on the EnSpire using the AlphaScreen protocol (PerkinElmer), according to the manufacturer's recommendations. Normalized secretion values were calculated on a per-well basis as the ratio of total secreted insulin to total intracellular insulin content (%), or as the ratio of total secreted insulin to cell counts (pg/RFU), which were averaged across replicates. Corrected values were then normalized to basal insulin secretion rates for negative control on a per-experiment basis to eliminate variability in baseline secretion rates across independent experiments.

Insulin secretion assays in primary human islets. Insulin secretion assays were performed on intact human islets isolated from 16 risk allele carriers and 16 controls, matched for age ($P = 0.88$), sex, BMI ($P = 0.88$) and time of isolation. The experiments were performed at 37 °C in Krebs–Ringer buffer (KRB) (NaCl 115 mM; KCl 5 mM; $NaHCO_3$ 24 mM; $CaCl_2$ 2.5 mM; $MgCl_2$ 1 mM; HEPES 10 mM; 0.1% bovine serum albumin, pH 7.4) with glucose concentrations as indicated. For each donor, triplicate groups of 15 handpicked human islets were pre-incubated for 2 h with 1 mM glucose KRB. Islets were subsequently incubated for 1 h in 1 mM glucose KRB, after which supernatants were collected and replaced with 10 mM or 16.7 mM glucose KRB. After another hour, supernatants were again collected, and total insulin content extracted from the islet pellet using acid-ethanol. Samples were stored at −20 °C and assayed for insulin via electrochemiluminescence (Meso Scale Discovery, Rockville, MD, USA or ALPCO, Salem, NH, USA). Insulin secretion measurements at 1 mM and 16.7 mM glucose were performed for all individuals ($n = 16$ per group); measurements at 10 mM glucose were performed for a subset hereof ($n = 9$ per group). Averages were calculated after excluding any outliers using Grubb's test, and either displayed as raw values or normalized to insulin content on a per-individual basis.

Membrane capacitance assays in EndoC-βH1. Electrophysiological measurements were performed using an HEKA EPC10 Patch Clamp Amplifier (HEKA Elektronik Dr. Schulze GmbH) at 32 °C using the standard whole-cell configuration as previously described[66]. The pipette solution (intracellular medium) contained 125 mM Cs-glutamate, 20 mM CsCl, 15 mM NaCl, 1 mM $MgCl_2$, 5 mM HEPES, 0.05 mM EGTA, 0.1 mM cAMP and 3 mM Mg adenosine triphosphate (pH 7.15 with CsOH). The extracellular solution contained 1 mM glucose, 118 mM NaCl, 5.6 mM KCl, 1.2 mM $MgCl_2$, 5 mM HEPES and 2.6 mM $CaCl_2$ (pH 7.4 with NaOH). Tetraethylammonium chloride (20 mM) was added to block residual potassium currents. Calcium sensitivity of the vesicles was inferred from the relationship between calcium current density and exocytosis. The calcium current was measured at the first pulse, after inactivation of the sodium current (5 ms).

Analysis of exocytosis in primary human β-cells. Human pancreatic islets were isolated and prepared at the Alberta Diabetes Institute IsletCore as previously described; these had already been genotyped[34,61]. Single-cell capacitance responses were monitored as described earlier for EndoC-βH1 cells (including 0.1 mM cAMP), with an additional pre-incubation period of 1 h at 1 mM glucose. Following recordings, cells were positively identified by insulin immunostaining. The kinetics of exocytosis were examined at 1 and 10 mM glucose for seven risk allele carriers and seven controls, matched as groups for age ($P = 0.52$), sex and BMI ($P = 0.68$). The total number of cells per individual were [3; 5; 6; 6; 7; 8; 10] (low glucose, controls), [2; 3; 6; 7; 7; 7; 9] (high glucose, controls), [3; 4; 4; 5; 6; 8; 12] (low glucose, carriers), and [4; 4; 4; 6; 6; 8; 8] (high glucose, carriers). Mean capacitance was calculated per condition for each depolarization step.

Electron microscopy. Primary islets or EndoC-βH1 cells were fixed in 2.5% glutaraldehyde, post-fixed in 2% uranyl acetate, dehydrated in graded methanol and embedded in London Resin Gold (Agar Scientific Ltd). Ultrathin sections cut onto nickel grids were immunolabeled for PAM (ab109175; Abcam) followed by anti-rabbit biotin (Vector Laboratories Ltd) and streptavidin gold 15 nm (British Biocell International). For EndoC-βH1 cells, insulin was immunolabeled with an in-house antibody followed by anti-guinea pig gold 10 nm (British Biocell International). Sections were viewed on a JEOL JEM-1010 microscope (accelerating voltage 80 kV) with a digital camera (Gatan, Inc.). To estimate PAM concentration in secretory vesicles, quantification of immunogold labeling was performed by a blinded observer using the ImageJ software. The following parameters were determined: cross-sectional vesicle area; vesicle density (number of vesicles per cytoplasmic area); and insulin or PAM labeling per vesicle (number of immunogold particles per vesicle).

Western blot analysis of CgA amidation. To mimic the effect of gene silencing, cells were cultured for 24 h with the irreversible *PAM* inhibitor 4P3BA (Sigma-Aldrich) at a concentration of 500 µM[45]. Whole-cell extracts from siRNA or inhibitor-treated EndoC-βH1 cells were prepared as previously described[67]. Supernatant was harvested directly from cells and filtered with a 0.22-µm filter. Protein yields were determined with the Bio-Rad Protein Assay (Bio-Rad Laboratories), according to the manufacturer's instructions. For western blotting, 10 µg whole-cell extract and 0.5% total supernatant volume were analyzed with denaturing SDS–PAGE. Antibodies specific for CgA-Gly (ab52983; Abcam)[45] and total CgA (ab15160; Abcam) were used to calculate the ratio of amidated to non-amidated CgA using the Image Lab, version 5.0 for Windows (Bio-Rad Laboratories).

Immunohistochemistry of human pancreatic sections. Paraffin-embedded human pancreas biopsy sections were rehydrated, and citrate buffer antigen retrieval was performed. Consecutive 5-µm sections were stained for either total CgA (ab15160, 1:100; Abcam) or CgA-Gly (ab52983, 1:100; Abcam), plus insulin (1:500; Dako Canada) at 4 °C overnight. Species-specific secondary antibodies were Alexa Fluor 488 or 594 conjugates (Thermo Fisher Scientific) diluted 1:400 in 2% goat serum. Slides were cover-slipped with ProLong Gold Antifade Mountant with DAPI (Thermo Fisher Scientific). Images were acquired with an Olympus VS120 slide scanner (Olympus Canada Inc., Richmond Hill, Ontario, Canada), utilizing excitation filter wavelengths of 432 nm for DAPI, 515 nm for AlexaFluor 488 and 595 nm for AlexaFluor 594. Images were analyzed with the Olympus VS120-S6 analysis software (Olympus Corporation, Shinjuku, Tokyo, Japan. Olympus VS120-S6-ASW 2.9 [Build 13741] for Windows). and the ratio of mean total CgA and CgA-Gly signal intensity of paired insulin-positive regions was calculated.

Statistics. All statistical analyses were performed in R version 3.3.1 for Windows, except electron microscopy data and electrophysiological measurements in EndoC-βH1 cells, which were analyzed with the OriginPro software, version 8.5.1 for Windows (OriginLab Corporation). For kinetic data, rates were found to be dependent on substrate concentration; exponential models were fitted. Enzyme kinetics, electron microscopy, electrophysiological and cellular assay data were analyzed with two-tailed Student's *t*-tests where a single independent variable was being tested, and with two-way analysis of covariance (ANCOVA) followed by

Tukey's HSD post-hoc test in cases where two experimental variables were being tested (that is, glucose level and genotype status or knockdown). Western blot data were analyzed with paired, two-tailed Student's *t*-tests to eliminate differences in baseline intensity between experiments. Graphs show means of the indicated number of biological replicates, and error bars are s.e.m. Box plots display the interquartile range (IQR), whiskers indicating 1.5x the IQR, and outliers.

URLs. GTEx portal, https://www.gtexportal.org/home/gene/PAM.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Data availability. Previously published data sets analyzed during the current study can be accessed as follows: transcriptomic data from sorted human islets are available from the European Genome-phenome Archive under accession EGAS00001000442; tissue expression data for *PAM* from the GTEx consortium are available through the GTEx portal (see URLs); genotype data for primary human islets used in the analysis of exocytosis data are available from the European Genome-phenome Archive under accession EGAD00001001601; and secreted islet factors were analyzed based on data available from the supplementary information of PMID 21212933[68]. Genetic associations for plasma PAM expression levels are available in the supplementary material to the article by Sun et al.[33]. Data generated during this study are included in the published article (and its supplementary information files) or are available from the corresponding author on reasonable request.

References

60. Cross, S. E., Hughes, S. J., Clark, A., Gray, D. W. & Johnson, P. R. Collagenase does not persist in human islets following isolation. *Cell Transplant.* **21**, 2531–2535 (2012).
61. Lyon, J. et al. Research-focused isolation of human islets from donors with and without diabetes at the Alberta Diabetes Institute IsletCore. *Endocrinology* **157**, 560–569 (2016).
62. Manning Fox, J. E. et al. Human islet function following 20 years of cryogenic biobanking. *Diabetologia* **58**, 1503–1512 (2015).
63. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
64. McPherson, J. D. et al. A physical map of the human genome. *Nature* **409**, 934–941 (2001).
65. Castel, S. E., Levy-Moonshine, A., Mohammadi, P., Banks, E. & Lappalainen, T. Tools and best practices for data processing in allelic expression analysis. *Genome Biol.* **16**, 195 (2015).
66. Chandra, V. et al. RFX6 regulates insulin secretion by modulating Ca^{2+} homeostasis in human β cells. *Cell Rep.* **9**, 2206–2218 (2014).
67. Lees, M. J. & Whitelaw, M. L. Multiple roles of ligand in transforming the dioxin receptor to an active basic helix-loop-helix/PAS transcription factor complex with the nuclear protein Arnt. *Mol. Cell. Biol.* **19**, 5811–5822 (1999).
68. Yang, Y. H. et al. Paracrine signalling loops in adult human and mouse pancreatic islets: netrins modulate beta cell apoptosis signalling via dependence receptors. *Diabetologia* **54**, 828–842 (2011).

Soren K. Thomsen,[1] Alessandro Ceroni,[2] Martijn van de Bunt,[1,3] Carla Burrows,[1] Amy Barrett,[1] Raphael Scharfmann,[4] Daniel Ebner,[2] Mark I. McCarthy,[1,3,5] and Anna L. Gloyn[1,3,5]

Systematic Functional Characterization of Candidate Causal Genes for Type 2 Diabetes Risk Variants

Diabetes 2016;65:3805–3811 | DOI: 10.2337/db16-0361

Most genetic association signals for type 2 diabetes risk are located in noncoding regions of the genome, hindering translation into molecular mechanisms. Physiological studies have shown a majority of disease-associated variants to exert their effects through pancreatic islet dysfunction. Systematically characterizing the role of regional transcripts in β-cell function could identify the underlying disease-causing genes, but large-scale studies in human cellular models have previously been impractical. We developed a robust and scalable strategy based on arrayed gene silencing in the human β-cell line EndoC-βH1. In a screen of 300 positional candidates selected from 75 type 2 diabetes regions, each gene was assayed for effects on multiple disease–relevant phenotypes, including insulin secretion and cellular proliferation. We identified a total of 45 genes involved in β-cell function, pointing to possible causal mechanisms at 37 disease-associated loci. The results showed a strong enrichment for genes implicated in monogenic diabetes. Selected effects were validated in a follow-up study, including several genes (*ARL15*, *ZMIZ1*, and *THADA*) with previously unknown or poorly described roles in β-cell biology. We have demonstrated the feasibility of systematic functional screening in a human β-cell model and successfully prioritized plausible disease-causing genes at more than half of the regions investigated.

Type 2 diabetes risk is determined by a complex interplay between environmental and genetic factors, with heritability estimates ranging from 20% to 80% (1). Over the past decade, genome-wide association studies (GWAS) of ever-increasing size have discovered more than 100 regions of the genome (loci) associated with type 2 diabetes risk (2). Studies in individuals with diabetes have demonstrated that a large number of these association signals exert their effects on disease susceptibility through pancreatic islet dysfunction (3).

Despite these advances, progress in translating genetic findings into disease biology has been relatively slow. The majority of risk variants are located in noncoding regions of the genome and pinpointing the underlying causal genes or "effector transcripts" has proved challenging (4). Recent efforts have focused on identifying structural or functional links between association signals and regional genes (5,6). A complementary strategy uses candidate gene biology to prioritize genes located near association signals. High-throughput screening could facilitate the identification of genes implicated in β-cell function and thereby highlight potential effector transcripts at type 2 diabetes GWAS loci. To date, such approaches have been limited by the inadequacies of available human cellular models and the high cost of insulin immunoassays (~$2 per data point), the gold standard for measuring insulin. To circumvent these issues, previous studies have relied on rodent β-cell models and either used reporter assays as a proxy for insulin measurements or focused on cellular proliferation (7–11).

Recently, the first glucose-responsive human β-cell line, EndoC-βH1, was generated (12,13). The line is derived from

[1]Oxford Centre for Diabetes, Endocrinology & Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, U.K.
[2]Target Discovery Institute, Nuffield Department of Medicine, University of Oxford, Oxford, U.K.
[3]Wellcome Trust Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, U.K.
[4]INSERM U1016, Institut Cochin, Université Paris Descartes, Paris, France
[5]National Institute for Health Research Oxford Biomedical Research Centre, Churchill Hospital, Oxford, U.K.

Corresponding author: Anna L. Gloyn, anna.gloyn@drl.ox.ac.uk.

Received 17 March 2016 and accepted 18 August 2016.

This article contains Supplementary Data online at http://diabetes.diabetesjournals.org/lookup/suppl/doi:10.2337/db16-0361/-/DC1.

© 2016 by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered. More information is available at http://www.diabetesjournals.org/content/license.

See accompanying article, p. 3541.

fetal pancreatic buds matured in vivo and displays modest but robust induction of insulin secretion in response to glucose and secretagogues. Detailed characterizations have shown the cell line to be an authentic model system for studying stimulus-coupled secretion (14–16).

To accelerate the discovery of causal genes for type 2 diabetes, the current study performed and validated a genetic screen in the EndoC-βH1 cell line. We identified genes at half of the type 2 diabetes–associated loci studied (37/75) where small interfering RNA (siRNA)-mediated silencing resulted in β-cell dysfunction. This demonstrates the feasibility of performing systematic screening for insulin secretion in a human β-cell model, with implications for both high-throughput genetic and chemical compound screening. Our results can be integrated with existing lines of evidence to prioritize effector transcripts at GWAS loci and highlight potential roles for *ARL15*, *ZMIZ1*, and *THADA* in the regulation of insulin secretion.

RESEARCH DESIGN AND METHODS

RNA-seq

The EndoC-βH1 cell line was cultured as previously described and grown to near confluency (12). RNA was then TRIzol extracted and sequenced at the Oxford Genomics Centre (Wellcome Trust Centre for Human Genetics, University of Oxford) (see Supplementary Fig. 4 for details). The raw sequencing data have been deposited at the European Nucleotide Archive (http://www.ebi.ac.uk/ena) under accession number PRJEB15283.

Cellular Assays

Cellular phenotypes were adapted for automated screening on a PerkinElmer Janus liquid handling workstation based on previously described assays (Supplementary Fig. 1*A*) (17). Briefly, 20,000 cells/well were reverse transfected in 96-well format at final siRNA concentrations of 25 nmol/L preincubated with 0.2 μL RNAiMAX in Opti-MEM. Custom libraries of siRNAs (ON-TARGETplus SMARTpools [Dharmacon] for the primary screen and Silencer Select [Thermo Fisher Scientific] for follow-up validation) were designed based on criteria described in Supplementary Table 2. In each case, nontargeting (NT) sequences based on the same chemistries were used as negative controls. Three days after transfection, cells were starved overnight in complete media containing 2.8 mmol/L glucose followed by 1-h starvation in 0 mmol/L media. Static insulin secretion assays were then performed for 1 h in complete media under the indicated conditions, after which cells were counted as described below.

Sample Analysis

Following secretion assays, supernatants were analyzed for insulin content using AlphaLISA Human Insulin Immunoassays (PerkinElmer) on a PHERAstar FS plate reader (BMG Labtech). Supernatants (50–250 nL) and beads prediluted in water (500 nL) were dispensed into 384–shallow well microplates with an Echo 550 (Labcyte) acoustic liquid handler before manual addition of immunoassay buffer to a final volume of 5 μL. Cell counts were measured using the CyQUANT Direct Cell Proliferation kit (Thermo Fisher Scientific) on an EnVision plate reader (PerkinElmer). All responses were normalized as indicated (see relevant figure legends) and expressed as a percentage of NT control for each phenotype. Effect sizes are given as the percentage difference from NT ($Response_{Gene} - Response_{NT}$) and the absolute values hereof ($|Response_{Gene} - Response_{NT}|$).

Statistical Analysis

Data analysis was performed using R 3.0.2. To identify significant responses, cell counts and normalized insulin secretion measurements for each gene were compared with NT control using Student two-sample *t* test. The false discovery rate (FDR) was controlled at 5% by applying the Benjamini-Hochberg procedure to produce adjusted *P* values (q values) for each phenotype. The Z-factor measuring the control response for each phenotype was calculated as

$$Z' = 1 - \frac{3(\sigma_{INS} + \sigma_{NT})}{|\mu_{INS} - \mu_{NT}|}$$

RESULTS

We first developed an automated assay for disease-relevant phenotypes in the human β-cell line EndoC-βH1 (Supplementary Fig. 1*A*). Selected targets were silenced in a parallel format using RNA interference. Cells were then assessed for effects on cell number and insulin secretion under four different conditions: low glucose (1 mmol/L), high glucose (20 mmol/L), and high glucose with the sulfonylurea tolbutamide (100 μmol/L) or with the phosphodiesterase inhibitor IBMX (100 μmol/L). Low- and high-glucose conditions were included to provide information on the effect of gene silencing under conditions representing the fasted and fed states in vivo. Tolbutamide and IBMX act on the depolarizing and the potentiating pathways of insulin secretion, respectively, and were included to provide additional mechanistic insights through modulation (e.g., synergy or pharmacological rescue) of any primary defects observed in low or high glucose.

To reduce the cost of sample analysis, we made use of acoustic liquid handling to miniaturize insulin immunoassays. This generalizable method enabled us to maintain high sensitivity for insulin measurements (coefficient of variation <3%) (Supplementary Fig. 2*A* and *B*), while obtaining a 10-fold reduction in the cost of sample analysis ($0.20 per data point). Using the insulin gene (*INS*) as a positive control, we confirmed that we were able to robustly detect effects of gene silencing on the phenotypes of our assay (mean Z′ = 0.6 across conditions) (Supplementary Fig. 3).

On the basis of this combined analysis and assay pipeline, we designed a primary screen to assess the role of positional candidate genes for type 2 diabetes GWAS loci in β-cell

function (Supplementary Fig. 1*B*). For target selection, we considered all protein-coding genes located within 1 Mb of a type 2 diabetes association signal. To exclude genes not expressed in our cellular model, we performed whole-genome RNA sequencing of the EndoC-βH1. Our expression data strongly correlated with published sequencing data for enriched primary β-cells (ρ = 0.78) (Supplementary Fig. 4) and showed robust expression of key β-cell genes (12,18) (Supplementary Table 1). We included only genes expressed in both EndoC-βH1 and primary β-cells (Supplementary Table 2), resulting in inclusion of 300 positional candidates from 75 type 2 diabetes GWAS loci.

We next performed our primary screen in triplicate and derived standardized scores for each phenotype. Knockdown was visibly confirmed using *PLK1*, an essential gene, which caused extensive cell death across all conditions. In a representative subset of 16 genes, we assessed knockdown efficiency at the transcript level and found the median residual expression to be 43% (Supplementary Fig. 5), roughly equivalent to monoallelic loss of function. To account for differences in plating efficiency and proliferation, cell counts were used to normalize insulin secretion data on a per-well basis. Two criteria were then applied to identify robust effects ("hits"): *1*) an FDR-adjusted q value <0.05 and *2*) an absolute effect size among the top 5% (Supplementary Fig. 6). This identified a total of 67 hits (15 for cell count and 52 for insulin secretion phenotypes) between 45 genes at 37 loci (Table 1).

For cell numbers, effect sizes for each gene were estimated based on 12 independently plated replicates (four conditions in triplicate) and therefore likely represent true differences in cellular proliferation and/or viability rather than random plating effects. Aside from *KIF11*, a gene with a known role in cell division, the largest effect sizes compared with NT control (coefficient of variation = 4% for cell numbers) were observed for *ZMIZ1* (−15.2%; q value = 6.5×10^{-5}) and *PRDX3* (+16.6%; q value = 9.2×10^{-5}).

For the insulin secretion data, we first performed an enrichment analysis for genes implicated in maturity-onset diabetes of the young (MODY). MODY describes a collection of monogenic subtypes of diabetes characterized by insufficient release or production of insulin. As would be expected for a set of bona fide regulators of β-cell function, we observed a strong enrichment of MODY genes among the significant hits (Fisher exact test, $P = 5.5 \times 10^{-9}$). Aggregating absolute effect sizes for MODY and non-MODY genes revealed this enrichment to be driven by altered insulin secretion and not by effects on cell numbers (Fig. 1).

Further validating our secretion data, we observed strong positive correlations between the normalized responses across conditions ($P < 2.2 \times 10^{-16}$) (Fig. 2) and found 10 of 35 genes to cause significant effects under two or more conditions. This included four known MODY genes and *ZMIZ1*, which, independently of the effect on cell numbers, was one of the strongest hits for reduced insulin secretion (q value <0.01 for low and high glucose). Knockdown of the *ABCC8* gene, which encodes a subunit of the ATP-sensitive potassium channel, was found to significantly increase insulin secretion under low-glucose and IBMX stimulation. As expected, the depolarization caused by this was masked under high glucose (as cells are already partially depolarized) and fully rescued by tolbutamide (due to pharmacological depolarization of the cells). The pattern of modulation by secretion conditions can thus be used to pinpoint specific biological pathways affected by gene silencing. To explore the relationship between conditions in greater detail, we performed clustering analysis on Z-scores derived from the normalized secretion values. This revealed high glucose and tolbutamide to be most similar in terms of modulating knockdown effects, with low glucose and tolbutamide being most dissimilar (Supplementary Fig. 7).

Finally, we assessed the contribution of off-target effects by performing a small-scale validation experiment using siRNAs designed with an alternate algorithm. The sequences were confirmed to be different from those of the primary screen and could thus be used to establish the biological relevance of positive hits. We selected eight target genes, representing hits for both positive and negative defects across the four conditions, and confirmed that knockdown efficiency was satisfactory (median residual expression = 19.3%) (Supplementary Fig. 8). Compared with insulin secretion results from the primary screen, we observed an excellent linear correlation (ρ = 0.85, $P = 6.7 \times 10^{-10}$) (Supplementary Fig. 9) and 88% directional consistency in normalized responses. The validated hits included several genes with limited prior evidence of a role in the regulation of β-cell function, including *ARL15* and *ZMIZ1*, which were found to significantly reduce insulin secretion across conditions (q values <0.05) (Fig. 3*A* and *B*), and *THADA,* which modestly elevated insulin secretion across three conditions, though the effect under low glucose was not observed in the primary screen (q value = 5.6×10^{-3}) (Fig. 3*C*). Interestingly, gene silencing of the known MODY gene *HNF4A* was confirmed to cause a paradoxical increase in insulin secretion across all four conditions tested (q values <0.001) (Fig. 3*D*).

DISCUSSION

High-throughput screens for β-cell dysfunction offer the opportunity to systematically characterize the role of genes in a disease-relevant tissue for type 2 diabetes. Previous efforts have focused on nonhuman model systems (7–10), reporter-based proxy measurements for insulin (7,8), and/or phenotypes not directly related to insulin production and secretion (10,11). Here, we report a genetic screening strategy for the interrogation of multiple disease–relevant phenotypes in the human β-cell line EndoC-βH1. In a primary screen of 300 positional candidates, we successfully identified 15 genes regulating cell number (proliferation and/or viability) and 35 genes regulating insulin secretion. This is, to our knowledge, the first systematic,

Table 1—Effects of significant hits identified in a primary screen for β-cell dysfunction

Gene	Locus	Low glucose	High glucose	IBMX	Tolbutamide	Cell count
ABCC8	*KCNJ11*	48.2*	24.0	26.7*	1.8	−1.4
ADAMTS9	*ADAMTS9*	6.3	−4.8	2.8	−8.0	12.2*
ADIPOQ	*ST64GAL1*	87.0*	23.2	23.1	8.8	−7.3
ARL15	*ARL15*	−5.5	−25.9*	−2.1	−15.5	−1.5
BCAR1	*BCAR1*	5.9	25.2	28.5*	9.5	−7.0
BCL6	*LPP*	−20.7*	−8.9	−1.0	−12.0	5.6
BMP8B	*MACF1*	7.8	16.5	9.5	26.9*	−1.0
CCNT2	*TMEM163*	−32.6*	1.8	4.7	5.0	−2.8
CDKAL1	*CDKAL1*	2.4	1.7	−10.6	−23.5*	7.5
DGKQ	*MAEA*	3.3	17.5	20.4	33.6*	−9.8
DMRTA2	*FAF1*	32.3*	24.5	13.1	22.7	−0.3
ELAVL4	*FAF1*	−3.6	9.1	21.2	22.8	−11.4*
ETV5	*IGF2BP2*	−12.4	−25.6*	−10.9	−12.6	−2.5
FAH	*ZFAND6*	−23.1*	−20.6*	−16.9	−27.2	−2.3
FBXW7	*TMEM154*	45.2*	9.6	8.7	11.7	9.9*
GINS4	*ANK1*	−16.1	−14.2	−9.5	−17.0	8.7*
GLIS3	*GLIS3*	−13.0	−10.6	6.4	−9.3	−10.3*
HEYL	*MACF1*	29.2	29.2*	0.4	20.0	−7.2
HMGA2	*HMGA2*	16.5	4.1	14.2	24.1*	−1.2
HNF1A	*HNF1A*	21.7	38.3*	23.1	25.5*	5.2
HNF4A	*HNF4A*	36.7*	66.9*	74.9	92.9*	−8.8
IGF2	*DUSP8*	−26.3*	−10.4	1.2	0.1	−1.3
INS	*DUSP8*	−53.5*	−44.8*	−48.7	−38.7*	0.9
KCNK17	*KCNK16*	−9.4	−17.3	−2.5	−1.2	9.8*
KCTD15	*PEPD*	21.7	9.4	12.5	0.8	−10.9*
KIF11	*HHEX/IDE*	45.4*	35.0*	26.9	55.4*	−40.1*
LINGO1	*HMG20A*	0	19.1	−12.4	−14.9	7.9*
MFGE8	*AP3S2*	30.0*	3.4	2.8	−1.7	−5.7
MIER3	*ANKRD55*	5.9	36.5*	5.8	18.2	1.9
NDUFS4	*ARL15*	1.6	−4.9	3.9	−1.7	10.8*
PABPC1L	*HNF4A*	−9.7	−12.3	−10.4	−28.8*	−1.2
PHF23	*SLC16A11*	−25.6*	−1.7	−5.9	3.5	−0.9
PLA2R1	*RBMS1*	8.6	1.6	10.8	1.6	9.1*
PRDX3	*GRK5*	24.0	31.7*	23.4	9.6	16.6*
PTHLH	*KLHDC5*	−2.8	−6.5	−0.5	−25.0*	−5.9
RND3	*RND3*	−8.7	−6.1	0	−3.0	−14.9*
SLC2A4	*SLC16A11*	14.5	0	27.2*	18.2	−1.6
SOCS7	*HNF1B*	3.9	−18.5*	11.2	−14.7	−1.6
SPPL3	*HNF1A*	−11.9	−21.9*	−6.1	−10.0	−10.8*
STK38L	*KLHDC5*	15.2	40.9*	4.7	25.2*	−3.0
THADA	*THADA*	−1.9	6.5	27.5	24.8*	−10.3
TLE1	*TLE1*	4.3	−5.0	−23.0*	16.2	4.6
TM6SF2	*CILP2*	−22.6*	−8.3	−0.8	−12.0	−8.7
UPF2	*CDC123*	7.5	−12.5	4.7	−24.9*	−3.2
ZMIZ1	*ZMIZ1*	−29.5*	−21.4*	−19.8	−16.8	−15.2*

The table lists effect sizes (% deviation from NT control) for each gene with a least one significant effect across the five phenotypes measured. All insulin secretion measurements were normalized on a per-well basis to cell counts, and the mean percentage deviations from NT control were then calculated for each condition. For cell counts, values were median-normalized for interplate differences, and the mean percentage deviations from NT control were calculated across conditions. *$q < 0.05$ by Student *t* test (FDR-adjusted).

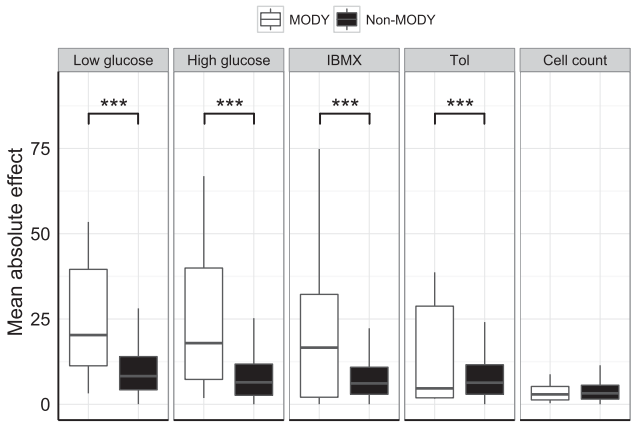


Figure 1—Comparing mean absolute effect sizes for MODY and non-MODY genes. Box plots of mean absolute effect sizes for MODY genes and non-MODY genes (excluding controls) across the five phenotypes measured. Effect sizes were calculated as described for Table 1, and the absolute values were then averaged for the two categories of genes. Among 14 identified MODY genes, 8 fulfilled criteria for inclusion in the screen: *HNF4A*, *GCK*, *HNF1A*, *HNF1B*, *PAX4*, *INS*, *ABCC8*, and *KCNJ11*. Tol, tolbutamide. Box plots show median and interquartile ranges for groups of n = 8 and 292 data points. ***q value <0.001 by Student t test (FDR-adjusted).

large-scale effort to identify genes involved in insulin secretion. Importantly, the identified hits can be used to prioritize novel effector transcripts for type 2 diabetes GWAS loci and may shed further light on the mechanisms underlying genes previously implicated in β-cell dysfunction.

The known MODY gene *HNF4A* was unexpectedly observed to cause a consistent increase (>40%) in insulin secretion across all conditions. *HNF4A* encodes the transcription factor HNF4α and is mutated in about 10% of all MODY cases (19). *HNF4A* loss-of-function mutations that cause monogenic diabetes later in life have also been associated with increased birth weight (indicative of increased fetal insulin secretion) and congenital hyperinsulinism in early infancy (20). The underlying reason for this switch from elevated to reduced insulin secretion is unknown, but it has been speculated that gradual β-cell exhaustion or, alternatively, a shift in the modulating cofactors of HNF4α may underlie this phenomenon (21,22).

Among the hits with limited prior evidence of a role in β-cell function, we independently validated *ZMIZ1*, *ARL15*, and *THADA*. Overexpression and knockdown of *ZMIZ1*, encoding ZMIZ1, has recently been shown to negatively impact on insulin secretion in primary human islets (6). Moreover, a nearby type 2 diabetes association signal overlaps a *cis*–expression quantitative trait loci for the gene, supporting its candidacy as the regional effector transcript (6). *ARL15* encodes ARL15, a relatively uncharacterized member of the ARF family of proteins involved in regulation of vesicle trafficking and biogenesis. The gene is highly expressed in β-cells and located downstream of an islet-active enhancer bound by key β-cell transcription factors (18,23) (Supplementary Fig. 10*A*). *THADA* encodes the protein THADA and contains a coding disease-association signal that has also been associated with reduced β-cell function (24) (Supplementary Fig. 10*C*). Consistent with the directionality of our findings, expression profiling has shown the gene to be more highly expressed in patients with type 2 diabetes compared with control subjects (25). All three genes thus emerge as strong candidates for future studies.

While successfully enabling unbiased functional characterization, our current screening strategy has a number of limitations. False negatives (i.e., true causal genes not identified as hits) could arise as a result of primary effects of the causal gene on non–β-cell tissues or through effects on genes expressed at different developmental stages. Likewise, overexpression or greater knockdown efficiency may in some cases be required to expose a disease-relevant phenotype. Among the targets analyzed for silencing efficiency, a variable range of knockdown was observed (34%–88%), and some genes might remain undetected due to insufficient silencing. Conversely, false-positive effects (i.e., non–β-cell regulators identified as hits) also cannot be excluded, and unexpected findings should be further functionally validated (e.g., *SLC2A4* effect on IBMX-stimulated insulin secretion). Though the EndoC-βH1 cell line has been found to recapitulate many aspects of β-cell function, it remains a possibility that some findings would not translate directly into human physiology.

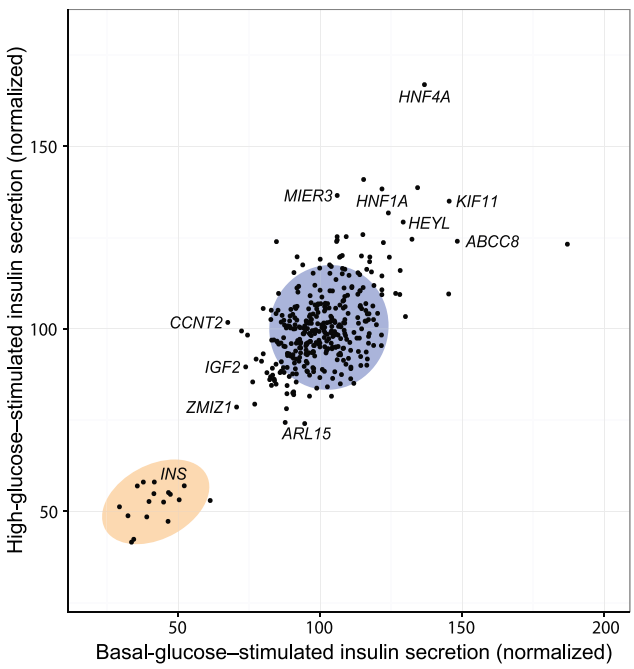


Figure 2—Comparison of insulin secretion data for high and low glucose. Normalized insulin secretion responses under high glucose vs. low glucose, with selected hits annotated. The blue circle indicates the 95% confidence contour for NT control, and the orange circle indicates the 95% confidence contour for *INS*-positive controls. All measurements were normalized on a per-well basis to cell counts, and averages for each condition were then subsequently normalized to the mean of NT control. Data points are mean of n = 3 and shown as percentage of NT control.

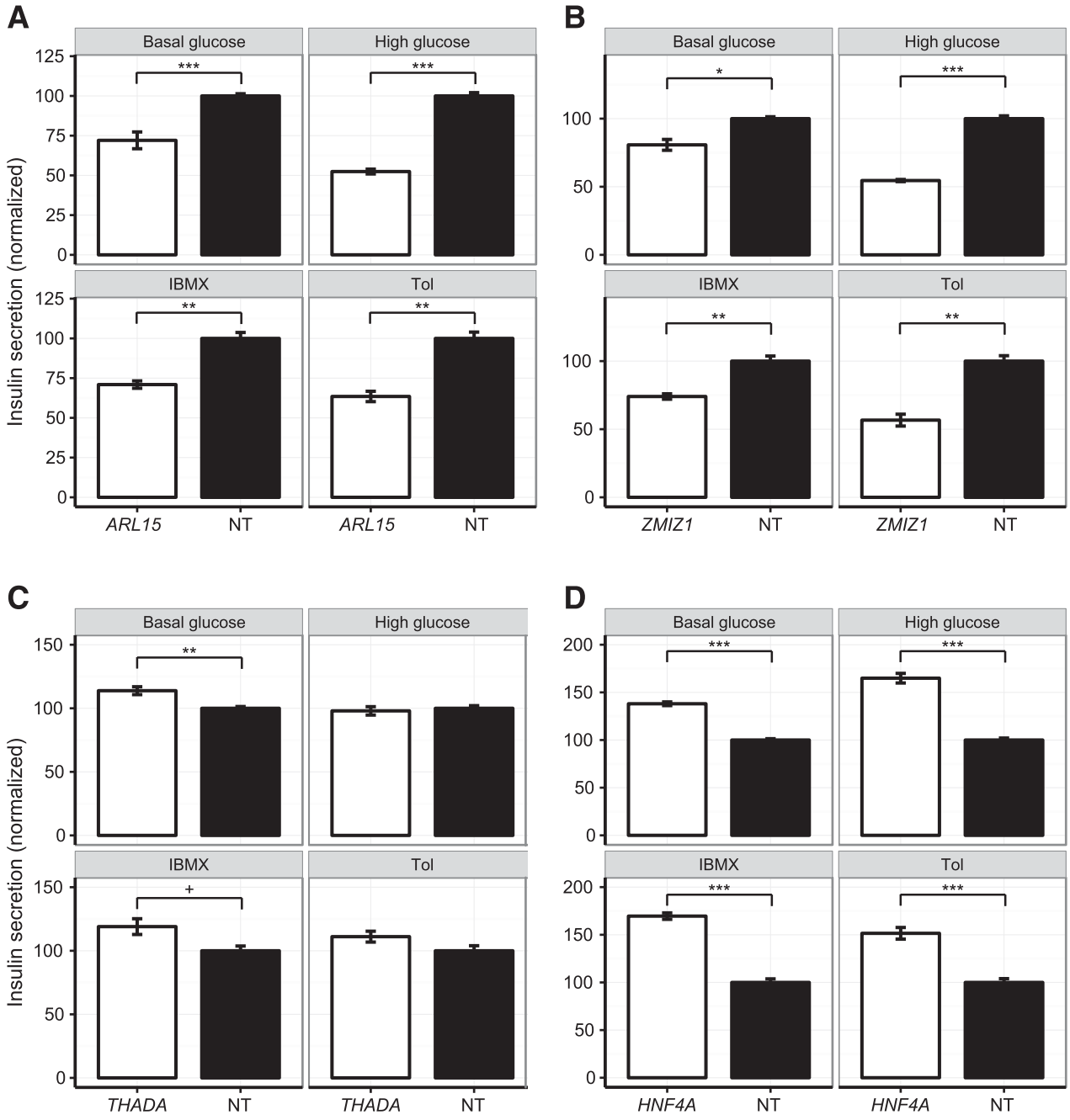


Figure 3—Insulin secretion data for selected genes in a follow-up validation experiment. Insulin secretion for *ARL15* (*A*), *ZMIZ1* (*B*), *THADA* (*C*), and *HNF4A* (*D*) (white bars) vs. NT (black bars) negative control under the indicated conditions. Measurements were processed as described for Fig. 2 and shown as percentage of NT control. Tol, tolbutamide. Data points are the mean of $n = 6$ for NT and $n = 3$ for other genes, and error bars are SEM. +q value <0.1, *q value <0.05, **q value <0.01, ***q value <0.001 by Student *t* test (FDR-adjusted).

Finally, a subset of the identified hits may represent true β-cell regulators that are independent of any disease risk variants and, though still of biological importance, not genuine effector transcripts for type 2 diabetes. In addition to the possibility of more than a single effector transcript per locus, this phenomenon likely also contributes to the relatively high proportion of multihit loci observed in the primary screen (8/37).

Despite these limitations, our screening strategy successfully replicated well-established biological mechanisms and identified genes involved in β-cell function at half of the loci investigated. This demonstrates, for the first time, the feasibility of performing scalable screens for insulin secretory defects in human pancreatic β-cells and opens up the possibility for not only large-scale genetic manipulations but also compound high-throughput screening to

therapeutically manipulate human β-cells. Insights from this and subsequent functional screens can be integrated with complementary lines of evidence from exome-wide association studies, chromatin conformation capture, and *cis*–expression quantitative trait loci studies to prioritize genes for follow-up studies. Ultimately, this could accelerate the translation of genetic association signals into molecular mechanisms for β-cell dysfunction, insulin insufficiency, and type 2 diabetes.

Funding. This study was funded in Oxford by the Wellcome Trust (095101/Z/10/Z and 098381). S.K.T. is a Radcliffe Department of Medicine Scholar. M.I.M. is a Wellcome Trust Senior Investigator. A.L.G. is a Wellcome Trust Senior Fellow in basic biomedical science.

Duality of Interest. M.v.d.B. is supported by a Novo Nordisk postdoctoral fellowship run in partnership with the University of Oxford. No other potential conflicts of interest relevant to this article were reported.

Author Contributions. S.K.T., A.C., D.E., M.I.M., and A.L.G. conceived and designed the study. S.K.T., M.v.d.B., M.I.M., and A.L.G. analyzed and interpreted the data. S.K.T., A.C., C.B., and A.B. performed the experiments. R.S. provided protocols. S.K.T., M.I.M., and A.L.G. wrote the manuscript. S.K.T., A.C., M.v.d.B., C.B., A.B., R.S., D.E., M.I.M., and A.L.G. edited and approved the manuscript. A.L.G. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Prior Presentation. Parts of this study were presented at the 52nd European Association for the Study of Diabetes Annual Meeting, Munich, Germany, 12–16 September 2016.

References

1. Ali O. Genetics of type 2 diabetes. World J Diabetes 2013;4:114–123
2. Prasad RB, Groop L. Genetics of type 2 diabetes-pitfalls and possibilities. Genes (Basel) 2015;6:87–123
3. Dimas AS, Lagou V, Barker A, et al.; MAGIC Investigators. Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. Diabetes 2014;63:2158–2171
4. Thomsen SK, Gloyn AL. The pancreatic β cell: recent insights from human genetics. Trends Endocrinol Metab 2014;25:425–434
5. Fadista J, Vikman P, Laakso EO, et al. Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism. Proc Natl Acad Sci U S A 2014;111:13924–13929
6. van de Bunt M, Manning Fox JE, Dai X, 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:e1005694
7. Burns SM, Vetere A, Walpita D, et al. High-throughput luminescent reporter of insulin secretion for discovering regulators of pancreatic beta-cell function. Cell Metab 2015;21:126–137
8. Ku GM, Pappalardo Z, Luo CC, German MS, McManus MT. An siRNA screen in pancreatic beta cells reveals a role for Gpr27 in insulin production. PLoS Genet 2012;8:e1002449
9. Wu W, Shang J, Feng Y, et al. Identification of glucose-dependent insulin secretion targets in pancreatic beta cells by combining defined-mechanism compound library screening and siRNA gene silencing. J Biomol Screen 2008;13:128–134
10. Lee JA, Chu S, Willard FS, et al. Open innovation for phenotypic drug discovery: the PD2 assay panel. J Biomol Screen 2011;16:588–602
11. Walpita D, Hasaka T, Spoonamore J, et al. A human islet cell culture system for high-throughput screening. J Biomol Screen 2012;17:509–518
12. 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;121:3589–3597
13. Weir GC, Bonner-Weir S. Finally! A human pancreatic β cell line. J Clin Invest 2011;121:3395–3397
14. 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:e0120879
15. Krishnan K, Ma Z, Björklund A, Islam MS. Calcium signaling in a genetically engineered human pancreatic β-cell line. Pancreas 2015;44:773–777
16. Gurgul-Convey E, Kaminski MT, Lenzen S. Physiological characterization of the human EndoC-βH1 β-cell line. Biochem Biophys Res Commun 2015;464:13–19
17. Pal A, Potjer TP, Thomsen SK, et al. Loss-of-function mutations in the cell-cycle control gene CDKN2A impact on glucose homeostasis in humans. Diabetes 2016;65:527–533
18. Nica AC, Ongen H, Irminger JC, et al. Cell-type, allelic, and genetic signatures in the human pancreatic beta cell transcriptome. Genome Res 2013;23:1554–1562
19. Shields BM, Hicks S, Shepherd MH, Colclough K, Hattersley AT, Ellard S. Maturity-onset diabetes of the young (MODY): how many cases are we missing? Diabetologia 2010;53:2504–2508
20. Pearson ER, Boj SF, Steele AM, et al. Macrosomia and hyperinsulinaemic hypoglycaemia in patients with heterozygous mutations in the HNF4A gene. PLoS Med 2007;4:e118
21. Glaser B. Type 2 diabetes: hypoinsulinism, hyperinsulinism, or both? PLoS Med 2007;4:e148
22. Rahman SA, Nessa A, Hussain K. Molecular mechanisms of congenital hyperinsulinism. J Mol Endocrinol 2015;54:R119–R129
23. Pasquali L, Gaulton KJ, Rodríguez-Seguí SA, et al. Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. Nat Genet 2014;46:136–143
24. Simonis-Bik AM, Nijpels G, van Haeften TW, et al. Gene variants in the novel type 2 diabetes loci CDC123/CAMK1D, THADA, ADAMTS9, BCL11A, and MTNR1B affect different aspects of pancreatic beta-cell function. Diabetes 2010;59:293–301
25. Parikh H, Lyssenko V, Groop LC. Prioritizing genes for follow-up from genome wide association studies using information on gene expression in tissues relevant for type 2 diabetes mellitus. BMC Med Genomics 2009;2:72