The Role of Glucose-6-Phosphatase Catalytic Domain in Glucose Homeostasis

A thesis submitted for the degree of Doctor of Philosophy by

Natasha Hui Jin Ng

Balliol College, University of Oxford

and

Oxford Centre for Diabetes, Endocrinology & Metabolism

Radcliffe Department of Medicine, University of Oxford

Michaelmas Term 2016

Memorandum

The work in this thesis is the original work of the author and contributions from others are declared below. All experiments were performed at the Oxford Centre for Diabetes, Endocrinology & Metabolism (OCDEM) under the supervision of Professor Anna Gloyn and Professor Patrik Rorsman. Funding for the project was provided by the Wellcome Trust.

The evaluation of *G6PC2* coding variants and haplotype analysis (Chapter 3) were carried out in collaboration with the T2D-GENES/GoT2D consortia and has been published in PLoS Genetics (Mahajan et al. 2015). The manuscript is provided in the appendix.

The *in silico* and functional characterisations of *G6PC1* and *G6PC2* coding variants (Chapters 4 and 5) were based on exome array data from the MAGIC investigators and exome sequencing data from the T2D-GENES/GoT2D consortia. The collaborative effort with the MAGIC investigators is currently being written up as a manuscript.

The localisation studies of G6PC2 using electron microscopy techniques (Chapter 5) were performed with the assistance of Dr Anne Clark.

The analysis of *G6PC2* transcript expression in carriers of the rs560887 and rs146779637 (p.R283X) single nucleotide variants (Chapters 5 and 6) was performed by Dr Martijn van de Bunt using RNA sequence data from human islets from Oxford (n=69) and Edmonton (n=105).

Finally, the static incubation assays for analysis of insulin secretion (Chapter 6) were completed together with Soren Thomsen (OCDEM).

No part of this thesis has been submitted for any other degree at this or any other university.

This thesis is approximately 48,500 words.

Acknowledgements

My gratitude towards the people who have played a part in my DPhil experience and beyond over the past few years in Oxford cannot be adequately summed up in a single page. My scientific achievements and personal development would not have been possible without their presence. My first thank you goes to my primary supervisor Prof Anna Gloyn and secondary supervisor Prof Patrik Rorsman, who were the reasons why I chose to come to Oxford. Despite the challenges in the project, I have no regrets choosing this DPhil and am ever grateful to them for accepting me. I thank them for their guidance, resources, opportunities for growth, and most of all for their confidence and trust in my abilities and allowing me to conduct my research independently. I also thank both Anna and Prof Mark McCarthy for involving me in very active and productive local and international collaborations that enabled me to showcase my work.

I thank my colleagues from the Gloyn group, past and present, and others in OCDEM and WTCHG that I have shared the lab with. Their insightful discussions, companionship and patience with me have made OCDEM a place I never dreaded going to, and I only wish I had the chance to get to know them better. Thank you to Jana, Nicola, Mandy, Amy, Anne, Chris, Carla, Benoit, Vibe, Andrei and Anubha. Special mention goes to Nicola, who saw me through the progress of my DPhil and is my role model not just as a scientist but also as a person. I also thank Benoit and Anne Clark for spending time teaching me and giving me ideas. Your enthusiasm in your work and in helping others is inspiring. I would also like to thank Costas, Nellie, Katherine and Leanne who were always ready to provide a listening ear and a reassuring voice, even at times when I didn't know I needed them.

And of course, my DPhil experience would not have been complete without my fellow students, some of whom have graduated from DPhil school. Neelam – thank you for the very memorable chat before I even joined OCDEM, and for your very genuine advice and constant care about my life. Soren, Matthias, Mahesh, Sara, Katia, Mahdieh and Nik – thank you for sharing this journey with me and for your understanding even without words being said. I also thank my fellow co-founders of BioMe Oxford – Soren, Matthias, Aga, Patrick and Rasmus – for a very rewarding experience building this start up together and for spurring me on to step out of my comfort zone to explore something completely new and different.

Outside of work, I am extremely grateful to my 49 Ferry Road housemates Midori (who reads me to a tee and was with me throughout this journey), Jed, Adan, Alex and Laia for interesting and often deep conversations, work sessions, cookouts and activities that I will fondly miss. This also extends to other Balliol and non-Balliol friends Thomas, Zoubeir, Toli, Kyle, Amit, Johanna and Laura. You have opened my eyes to the world and helped me to immerse in the quirks of college experiences and I have learnt so much just being with you guys. I thank my other fellow postgrads who also call Oxford their second home but were always ready to give me a slice of Singapore. Glenn, Darren, Jeremy, Shu Yun, Yi En, Her Shuang, Kang Yee, Joy, Yun Ying and Pei Jin, thank you for noisy gatherings that never fail to involve great food. I also thank my second family in Oxford who I met through the charity Jacari – Audrey and her adorable kids Miche, David and Angelica – for the delicious home-cooked food and for the special bond we now share.

This thesis not only wraps up my DPhil but also concludes my time spent in the UK. The past seven years in this country and especially in Oxford have been rewarding and life-changing to say the least. I was blessed with this opportunity with the financial support from A*STAR Singapore. My previous mentors Prof Ren, Hui Ling and Ziwei, thank you for having guided me as I started navigating this world of scientific research and who led me through my early insecurities.

Last but not least, I thank my family – my dad, mum and my brother, who have watched me grow and quietly and patiently supported me and gave me more freedom than I could ever ask for to do what I wanted to do. Thank you for unconditionally taking care of me and the home and awaiting my every return. My final and biggest thank you goes to my fiancé Shiu Yuen for his unwavering support throughout my life journey abroad. Despite being on the other side of the world, you were the constant in my life and a pillar of strength. Thank you for your continuous trust and belief in me and for never giving up on us. No words would sufficiently describe my appreciation for having you in my life.

The Role of Glucose-6-Phosphatase Catalytic Domain in Glucose Homeostasis

Natasha Hui Jin Ng, Balliol College

Submitted in candidature for the degree of Doctor of Philosophy, Michaelmas Term 2016

Over the past decade, there has been unprecedented increase in the number of genetic loci associating with type 2 diabetes (T2D) risk and related glycemic traits, thanks to advances in sequencing technologies and access to large sample sizes. Identification of associated genetic variants across the frequency spectrum can provide valuable insight into disease pathophysiology. However, the translation into biological insights has been slow often due to uncertainties over the underlying effector transcripts. *G6PC2/ABCB11* is one locus characterised by common non-coding variants that are strongly associated with fasting plasma glucose (FG) levels in healthy adults. The work presented in this thesis aims to understand how protein-coding variants in glycemic trait loci such as *G6PC2* contribute to the variability of glycemic traits and in addition gain further insight into the physiological role of *G6PC2*.

To evaluate the role of coding variants in glycemic trait variation, an exome array genotyping study of non-diabetic European individuals (n=33,407) reported multiple coding variants in *G6PC2* that were independently associated with FG. I designed and conducted *in vitro* assays to functionally assess these variants and showed that they result in loss of function (LOF) due to reduced protein stability. This established *G6PC2* as the effector transcript influencing FG and highlighted a critical role for *G6PC2* (encoding the islet-specific glucose-6-phosphatase catalytic subunit) in glucose homeostasis.

To investigate the role of low frequency (MAF=1-5%) and rare (MAF<1%) coding variants in influencing glycemic traits, recent large-scale exome array meta-analyses and whole exome sequencing were carried out as part of MAGIC (n=144,060) and the T2D-GENES/GoT2D consortia (n=12,940) respectively. *G6PC1*, a gene homolog of *G6PC2* that primarily acts through the liver, was uncovered as a novel glycemic locus. My functional follow-up studies demonstrated that rare coding variants in *G6PC1* exhibited LOF to influence both FG and FI levels. As rare variation in *G6PC2* not previously identified could also affect G6PC2 function and modulate glycemic traits, I also functionally characterised a suite of rare non-synonymous *G6PC2* variants. Most of the variants tested exhibited markedly reduced protein levels and/or loss of glycosylation. Several variants were also found to impact on enzymatic activity through inactivating or activating mechanisms to influence FG levels.

Finally, to gain better understanding of the function of *G6PC2* I performed gene knockdown studies in the EndoC-βH1 human beta cell model followed by insulin secretion analyses. *G6PC2* knockdown resulted in increased insulin secretion at sub-threshold glucose stimulation levels, consistent with studies in knockout mouse models. In addition, expression of LOF G6PC2 variants were found to upregulate ER stress responses. These results warrant further studies of the precise roles that G6PC2, an ER-resident protein, plays in regulating insulin secretory function and ER homeostasis in the beta cell.

Overall, my work described multiple rare coding variants in both *G6PC1* and *G6PC2* that alter protein function to regulate glucose metabolism through diverse mechanisms in different tissues. Improved understanding of these effector transcripts will open up opportunities for the exploration of new therapeutic targets for glucose regulation and T2D.

Common abbreviations

2hGlu	2-hour plasma glucose levels after a glucose tolerance test
ER	Endoplasmic reticulum
FG	Fasting glucose
FI	Fasting insulin
G6P	Glucose-6-phosphate
G6Pase	Glucose-6-phosphatase
GCK	Glucokinase
GOF	Gain of function
GSD	Glycogen storage disease
GSD1a	Glycogen storage disease type 1a
GSIS	Glucose-stimulated insulin secretion
GWAS	Genome-wide association studies
HbA1c	Glycated haemoglobin
HGP	Hepatic glucose production
IFG	Impaired fasting glucose
IGT	Impaired glucose tolerance
kDa	Kilo Dalton
K_m	Michaelis constant of an enzyme, an inverse measure of affinity
KD	Knockdown
KO	Knockout
LD	Linkage disequilibrium
LOF	Loss of function
MAF	Minor allele frequency
NMD	Nonsense-mediated decay
Pi	Inorganic phosphate
PCR	Polymerase chain reaction
PTV	Protein-truncating variant
SDM	Site directed mutagenesis
SEM	Standard error of the mean
SNP	Single nucleotide polymorphism
T2D	Type 2 diabetes
V_{max}	Maximal velocity of an enzyme
WT	Wild type

Table of Contents

Chapter 1

General Introduction

The reference to the review that some of the ideas presented in this chapter contribute to is stated below. The complete manuscript is provided in the Appendices.

Ng, HJ and Gloyn, AL (2013) Bridging the gap between genetic associations and molecular mechanisms for type 2 diabetes. *Curr Diab Rep*, 13(6), 778-85.

1.1. Type 2 diabetes and metabolic disease

It is widely recognised that the global prevalence of diabetes will continue to rise rapidly in the next decades and be of substantial health burden all over the world, largely driven by changes in environmental and lifestyle factors (WHO 2016). Diabetes is classified into multiple subtypes including type 2 diabetes (T2D), a common chronic metabolic disorder accounting for more than 90% of diabetes cases in adults, and type 1 diabetes (T1D), an autoimmune disease during which the insulin-producing pancreatic islet beta cells are attacked by the body's immune cells (IDF 2015). There are also rare monogenic forms of diabetes, such as maturity-onset diabetes of the young (MODY) and neonatal diabetes, that are caused by single gene defects. T2D is characterised by high blood glucose levels (hyperglycemia) due to insufficient insulin production for the body's requirements that is caused by impaired insulin secretion and/or ineffective insulin action (Alberti and Zimmet 1998). The current recommended diagnostic criteria for T2D is a fasting plasma glucose (FG) level of ≥7.0 mmol/l, plasma glucose of ≥11.1 mmol/l two hours after a 75g oral glucose load (2hGlu), and HbA1c of ≥6.5% (WHO 2016). HbA1c, or glycated haemoglobin, offers a longer-term measurement of glucose control than FG as it reflects the average plasma glucose concentration over the preceding 2-3 months. Individuals with raised blood glucose levels not high enough for a diagnosis of T2D are additionally known to have impaired glucose tolerance (IGT) or impaired fasting glucose (IFG), which are features of prediabetes (Nathan et al. 2007). These individuals are likely to be at high risk for developing future T2D. The secondary complications resulting from prolonged high blood glucose is a major contributor to mortality and morbidity in patients with diabetes (WHO 2016).

1.1.1. Key mechanisms underlying glucose homeostasis

Blood glucose levels are tightly regulated both before and after food intake to ensure that they remain relatively stable and within the normoglycemic range of 4.0-6.0 mmol/l for FG and ≤7.8 mmol/l for 2hGlu. The rise in glucose acts as a signal to a variety of glucose-sensitive cells in organs such as the gut, brain, pancreas and liver, to activate glucose homeostatic mechanisms as part of a complex neuroendocrine network (Matschinsky et al. 2006).

The liver has a central role in maintaining tight control over glucose homeostasis, through a balance of glucose clearance after food intake and hepatic glucose production (HGP) in the post-absorptive state. These processes are primarily determined by the level of glucose supply and the relative actions of insulin and glucagon (Cherrington 1999). Briefly, glucose intake into the liver cells is mediated by the bidirectional glucose transporter 2 (GLUT2), and glucose is phosphorylated to glucose-6-phosphate (G6P) in an adenosine triphosphate (ATP)-requiring step by liver glucokinase (GCK), which makes up the large majority of total GCK in the body. G6P subsequently serves as the substrate for multiple metabolic pathways that lead to the generation of ATP, cholesterol and lipids through glycolysis, the Krebs cycle and pentose phosphate shunt (Agius 2008; Matschinsky 2009). G6P is also used for the synthesis of glycogen through glycogen synthase (Villar-Palasi and Guinovart 1997). At this point insulin inhibits HGP to prevent post-prandial elevation of blood glucose. In the fasted state, HGP is conversely turned on and glucose is produced by glycogen breakdown (glycogenolysis) and glucose synthesis from gluconeogenic precursors such as pyruvate and lactate (gluconeogenesis) to counteract the fall in blood glucose and maintain euglycemia. These processes are mediated by key enzymes such as glucose-6-phosphatase (opposing glucokinase) and glycogen phosporylase (opposing glycogen synthase) (Nordlie et al. 1999).

Another major point of glucose control involves the pancreatic islets of Langerhans, whose alpha and beta cells are highly sensitive to changes in circulating plasma glucose levels and secrete glucagon and insulin respectively. Insulin, secreted from the beta cells in response to elevation in plasma glucose, acts as the signal to drive glucose uptake from the bloodstream by tissues that utilise the sugar for energy generation or storage. It also inhibits glucose production from gluconeogenic tissues. In contrast, glucagon secreted from alpha cells in response to reduced plasma glucose levels stimulate glucose production. These hormonal signals are known to respond rapidly as part of a gluco-regulatory feedback loop to restore plasma glucose levels to its set point (Cherrington 1999).

The consensus model for glucose-stimulated insulin secretion (GSIS) in the pancreatic beta cell involves coupling glucose metabolism to exocytosis of insulin-containing secretory granules through a series of metabolic processes (summarised in Figure 1.1.1.1). Briefly, exposure to elevated glucose results in glucose uptake by the beta cell through glucose transporters. Glucose is first metabolised by the beta cell isoform of GCK to G6P which is then shunted through the rest of the glycolytic pathway, generating substrates for

mitochondrial oxidative phosphorylation and eventual synthesis of ATP. The increase in ATP/ADP ratio leads to the inhibition of ATP-sensitive K^+ channels (K_{ATP}), which are central to this triggering pathway of insulin secretion. The conductance of the K_{ATP} channels normally maintains cell membrane potential at a resting potential of -70mV. Closure of these channels depolarises the membrane leading to activation of voltage-sensitive Ca^{2+} channels, which open to allow Ca^{2+} influx and subsequent exocytosis of insulin granules (Ashcroft and Rorsman 2012; Henquin 2000). There are also amplifying pathways that enhance the secretory response in the presence of elevated Ca^{2+}, that are not directly dependent on K_{ATP} channels but mediated by secondary messengers (Henquin 2000). In human islets, glucose levels of 3 mM are sufficient to trigger insulin exocytosis, and ~15 mM glucose stimulates maximal insulin secretion (Henquin et al. 2006; Rorsman and Braun 2013). Various steps along the triggering pathway of insulin secretion may be specifically modified *in vitro* using various experimental conditions to investigate the precise role of each step along the pathway. For instance, agents that target the K_{ATP} channel are commonly used to enhance or antagonise its activity to dissociate glucose-dependent metabolism from beta cell electrical activity and insulin exocytosis.

In summary, glucose homeostasis requires the coordination of multiple hormonal and neuronal (though this has not been described here) circuits across multiple organs as well as complex metabolic signalling within cells. In particular, the critical mechanisms underlying the functions of the liver and pancreatic islets in glucose sensing and homeostasis have been described in this section.

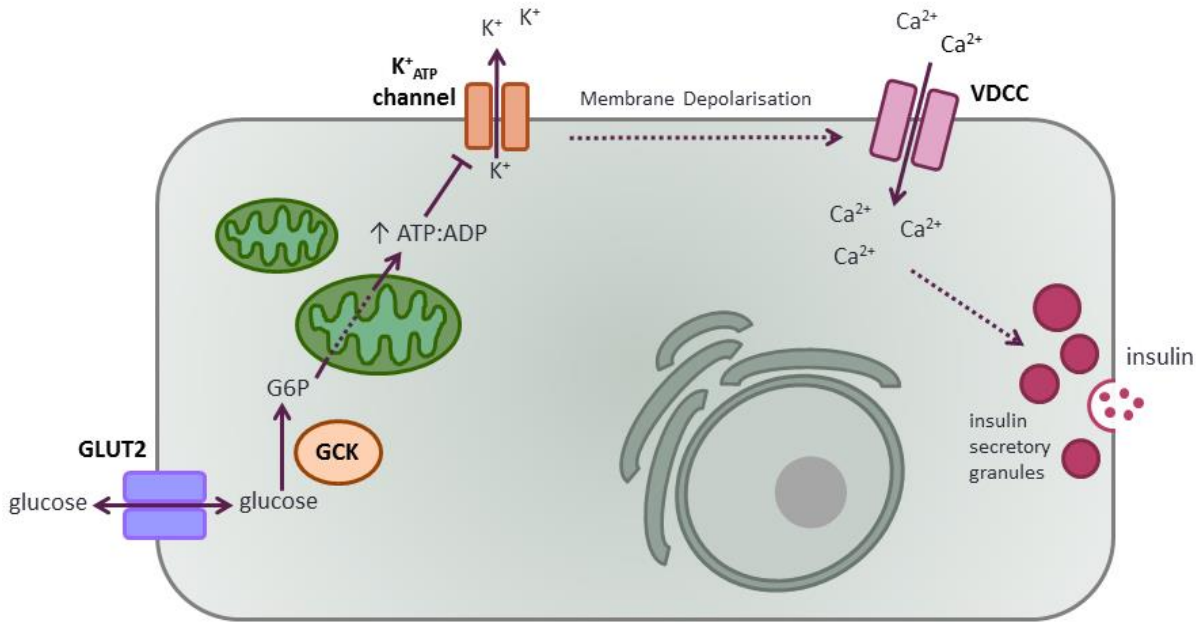


Figure 1.1.1.1. Consensus model of glucose-stimulated insulin secretion in the pancreatic beta cell. Glucose enters the cell through glucose transporters (GLUT2) and is phosphorylated to glucose-6-phosphate (G6P) by glucokinase (GCK). G6P enters glycolysis and mitochondrial oxidative phosphorylation which leads to the generation of ATP. The increase in ATP to ADP ratio inhibits K^{+}_{ATP} channel activity resulting in membrane depolarisation and subsequent opening of voltage-dependent Ca^{2+} channels (VDCCs). Ca^{2+} influx into the cell triggers the exocytosis of insulin-containing secretory granules.

1.1.2. Factors contributing to the pathogenesis of T2D

The causes of T2D are multifactorial and risk of disease is influenced by environmental factors such as urbanisation and exposure to unhealthy diets, and lifestyle changes such as age, pregnancy, lack of physical activity, and presence of other health conditions such as obesity (Ashcroft and Rorsman 2012). Although T2D is now known to have a strong genetic link, the multiple genetic loci that have so far been associated with T2D risk each only contribute a very small effect. Together, these factors make the elucidation of the precise causes of diabetes pathogenesis a great challenge. Nevertheless, beta cell dysfunction coupled with insulin resistance are recognised as the main culprits in polygenic forms of diabetes (Kahn 2003).

In insulin resistance, cells in peripheral tissues of the body are less sensitive to insulin signalling, such that glucose disposal in the skeletal muscle is reduced and endogenous HGP is not appropriately suppressed (Stumvoll et al. 2003; Stumvoll et al. 2005). As accelerated HGP is observed in patients with T2D or IFG (Basu et al. 2013; Rizza 2010; Weyer et al. 1999), the consequent increase in beta cell secretion from the pancreas in response to the high glucose levels may further fuel hyperinsulinemia and hyperglycemia. Obesity is highly associated with insulin resistance as increased concentrations of adipose tissue factors such as inflammatory cytokines and free fatty acids are known to adversely affect insulin signalling (Ravussin and Smith 2002; Stumvoll et al. 2005). The precise mechanisms underlying insulin resistance will not be discussed in detail here, but it is the reduction in insulin sensitivity without sufficient beta cell capacity to meet the required demands of insulin release that ultimately leads to T2D (Stumvoll et al. 2005).

There is accumulating evidence in the field that reduced beta cell function is a key contributor of T2D. Genetic evidence for T2D, which will be further discussed in section 1.2, largely point to genes that are involved in beta cell function (Dimas et al. 2014; Ingelsson et al. 2010). As disruptions in any step along the pathway of GSIS can alter beta cell function, and the majority of monogenic forms of diabetes have been found to be caused by single gene mutations that impair GSIS, knowledge of the genetic and molecular basis of these disorders have benefitted our understanding of beta cell dysfunction in the etiology of diabetes (Vaxillaire et al. 2012). A prominent example involves activating mutations in the Kir6.2 and SUR1 protein subunits of the K_{ATP} channels encoded by *KCNJ11* and *ABCC8*

respectively, that are known to be responsible for permanent and transient neonatal diabetes (Babenko et al. 2006; Flanagan et al. 2009; Gloyn et al. 2004). Experimental studies have shown that these mutants affect the sensitivity of K_{ATP} channels to ATP inhibition, resulting in lack of glucose-induced membrane depolarisation and inadequate insulin secretion (Babenko 2008; Koster et al. 2000). The translational impact of this work is seen when patients with these mutations were able to regain robust metabolic control through the use of oral sulfonylurea drugs that modulate K_{ATP} channel activity as opposed to insulin therapy (Pearson et al. 2006). Both genetic and molecular studies have not only shown clear evidence for association at these two genes with T2D risk, but also established the mechanisms that are likely to mediate their T2D associations, albeit with variable functional and clinical outcomes (Florez et al. 2004; Gloyn et al. 2003; Hamming et al. 2009; Riedel et al. 2003; Schwanstecher et al. 2002; Tarasov et al. 2008). Studies of neonatal diabetes and MODY also implicate other genes that are highly expressed in the beta cell such as *INS* and *GCK*, further highlighting the significance of beta cell defects that are relevant to T2D pathophysiology (Fajans et al. 2001; Vaxillaire et al. 2012).

For a heterogeneous common disease, multiple gene products and processes that regulate beta cell function, including the amplifying pathway of insulin secretion and beta cell health and mass, are likely to contribute to overall beta cell dysfunction. Increased understanding of the genetic and environmental determinants of both insulin secretion and action in the regulation of glycemic levels will be key to identifying individuals at risk of disease and development of targeted measures.

1.1.3. Glycemic traits in relation to T2D and other disease outcomes

The derangement of glucose levels in the fasting and post-prandial states are hallmarks of T2D. These conditions are implicated in IFG and IGT respectively even prior to any clinical diagnosis of T2D. The quality of glucose control is also related to longer term microvascular and macrovascular complications in T2D patients (Group et al. 2008; Holman et al. 2008; Ray et al. 2009). The evaluation of glycemic control takes on a number of forms, including FG, 2hGlu and HbA1c as mentioned, and also fasting insulin (FI) and proinsulin (a precursor of mature insulin), 1hGlu, and established measures of beta cell function (HOMA-B) and insulin resistance (HOMA-IR) (Matthews et al. 1985). These measures permit the assessment of beta cell function at basal conditions or in the presence of glucose challenge, and also the level of relative insulin resistance. Understanding the regulation (and dysregulation) of glycemic traits will help to reveal the biological mechanisms which maintain normal glucose homeostasis and which are involved in the development of T2D (Marullo et al. 2014).

As persistent hyperglycemia in T2D is widely known to increase the risk of cardiovascular disease (CVD) events (Beckman et al. 2002; Sarwar et al. 2010), variations in FG that do not reach diabetic levels can also independently influence cardiovascular health (Panzer et al. 2002). The risk of CVD has been shown to increase continuously with rising FG levels (Danaei et al. 2006; Singh et al. 2013). In fact, a recent Mendelian randomisation analysis of GWAS data for FG and coronary heart disease (CHD) was carried out to determine if FG-associated genetic variation is causally related to variation of CHD. The authors found that genetic risk variants that raised FG levels increased the risk of CHD, independently of the risk conferred by T2D (Merino et al. 2016). These findings support the notion that glycemia *per se* impacts on vascular health, and highlights the importance of the maintenance of glycemic levels within optimal ranges even in non-diabetic individuals.

1.2. Genetics of T2D and glycemic traits

1.2.1. The role of genetics in complex disease

The advent of genome-wide association studies (GWAS) has rapidly revolutionised the field of human genetics research of common complex diseases and phenotypes. Early genetic discoveries were obtained largely through family-based linkage studies that identified highly penetrant genetic mutations in monogenic disease, and small-scale candidate gene association studies that required prior knowledge of selected gene functions (Billings and Florez 2010; Marullo et al. 2014; Prasad and Groop 2015). Since then, GWAS, which stem from a hypothesis-free approach, have provided us with thousands of genetic loci to facilitate investigation of the genetic susceptibility of complex human traits. Some of the traits that have seen substantial yield in recent years include T2D (Cho et al. 2011; Fuchsberger et al. 2016; Mahajan et al. 2014; Morris et al. 2012), glycemic traits (Manning et al. 2012; Scott et al. 2012), cardiovascular outcomes (CARDIoGRAMplusC4D et al. 2013; Nikpay et al. 2015), obesity-related traits (Locke et al. 2015; Shungin et al. 2015) and even human height (Wood et al. 2014).

However, despite the rapid identification of genetic association signals, the identified signals only appear to account for a small portion of overall disease risk or trait variance. This was in part due to the majority of signals being characterised by common variants with very modest effect sizes. This led to the hypothesis that variants of lower allele frequency with moderate to large effect sizes, which tend not be well-captured by GWAS, potentially play a significant role in the yet-identified genetic susceptibility to disease (Manolio et al. 2009; McCarthy et al. 2008). Many recent and ongoing studies have been employing the strategy of probing for low-frequency (minor allele frequency [MAF]=1-5%) and rare (MAF<1%) variants using high throughput sequencing methods and large scale exome-targeted studies to address this question. Some of these efforts are discussed in section 1.2.4.

1.2.2. Genetics of T2D

GWAS of T2D patients and unaffected controls have so far identified over 120 genetic loci mostly characterised by common variant signals (MAF>5%) that are implicated in T2D risk (Prasad and Groop 2015). These findings have emphasised the role of islet beta cell function underlying most of the identified T2D genetic loci, and provided novel insight into fundamental biological pathways in T2D pathogenesis (Billings and Florez 2010; Dimas et al. 2014). Though many large scale GWAS and meta-analyses have been carried out in European-based cohorts, studies in other ethnic populations have helped to replicate known signals as well as contributed to the discovery of new risk loci or alleles. Notably, a study of individuals from Greenland, an isolated founder population, reported a novel association of a nonsense variant in *TBC1D4* (rs61736969, p.R684X) that had markedly greater impact on 2hGlu (β=3.8 mmol/l) and T2D risk (odds ratio 10.3), more than any other risk variants previously identified (Moltke et al. 2014). This variant is common in the Greenlandic Inuit population but not in any other populations tested, signifying the strength of conducting genetic association studies in diverse populations.

As the known associated loci so far only explained approximately ~20% of overall genetic risk, falling short of disease heritability estimates (Mahajan et al. 2014; Morris et al. 2012), much of the genetic contribution to disease remains to be discovered. To test if lower-frequency variants may explain a considerable proportion of the genetic contribution to T2D risk as previously suggested, a very recently published study made use of large-scale genome and exome sequencing in large sample sizes representing multiple ethnicities to assess associations with low frequency (MAF=0.5-5%) and rare (MAF<0.5%) variants. They however revealed little evidence for a prominent role for lower-frequency variants in T2D predisposition, and showed that most of the associations identified were still common variants in known GWAS regions (Fuchsberger et al. 2016).

Evidently, more extensive studies will be required to better understand the complex genetic architecture of T2D. Other factors that influence this genetic architecture such as parent-of-origin effects, sex differences, gene-gene and gene-environment interactions are likely to regulate the genetic susceptibility to disease (Mohlke and Boehnke 2015; Prasad and Groop 2015; Stancakova and Laakso 2016).

1.2.3. Genetics of glycemic traits

Dysregulation of quantitative glycemic traits such as FG and FI levels, as well as post-glucose glycemia is tightly linked to the development of T2D. The study of the genetic basis of these measurable continuous traits offers an alternative approach to identifying additional genetic factors and pinpointing key components of the machinery that regulate glucose homeostasis that may also impact on T2D predisposition and pathogenesis. These studies can be conducted in larger non-diabetic populations as opposed to a case-control study. In addition, as these traits have been linked to both prediabetes and cardiovascular events independent of diabetes, understanding the genetic etiology of glycemic traits may inform future development of disease preventative measures.

Indeed, one of the early GWAS published for T2D and related traits uncovered a signal for T2D at the *GCKR* locus (rs780094) that was also associated with FG, insulin resistance and serum triglyceride levels (Saxena et al. 2007). Subsequently, a slew of studies focusing on glycemic traits were carried out and further identified additional novel loci and also provided replication for known loci (Bouatia-Naji et al. 2009; Dupuis et al. 2010; Manning et al. 2012; Prokopenko et al. 2009; Prokopenko et al. 2014; Saxena et al. 2010; Scott et al. 2012; Soranzo et al. 2010; Strawbridge et al. 2011), bringing the total number of known glycemic trait loci to >80. Many of these discoveries were possible with the formation of the Meta-Analysis of Glucose and Insulin-related traits Consortium (MAGIC), an international collaboration bringing together data from multiple groups working on the genetics of glycemic traits in a meta-analysis, expanding the sample size and phenotypes studied (Prokopenko et al. 2009). Furthermore, detailed studies of the impact of glycemic/T2D loci on *in vivo* measures of beta cell function and insulin action provided substantial insight into whether these loci were specifically involved in glucose metabolism, beta cell secretory ability, insulin processing and secretion, or insulin sensitivity (Dimas et al. 2014; Ingelsson et al. 2010). In cases where the effector transcripts underlying the association signals were known, *in vitro* work was used to establish the biological processes through which associated genes and variants influenced glucose homeostasis. Some of these efforts are described in section 1.3.

However, given the logical physiological link between glucose homeostasis and T2D pathogenesis, only a small proportion of known glycemic trait loci have been shown to also

influence T2D risk. In fact, glycemic trait loci such as *G6PC2* and *MADD* which have been identified as strong genetic contributors of FG variation had no significant associations with T2D susceptibility; likewise, some T2D loci such as *TCF7L2* and *CDKN2A/B* which displayed the largest effects on T2D risk only showed modest effect sizes for FG levels (Dupuis et al. 2010; Scott et al. 2012). These observations confirmed that the mechanisms influencing glucose regulation within the normal physiological range may be distinct from those implicated in the pathophysiological state, at least based on current approaches (Marullo et al. 2014). There may be additional interacting factors that account for the complex relationship between homeostatic processes and disease endpoints, and at each of these loci fine-mapping and in-depth follow-up studies will be required to establish the functionally-relevant causal variants. As the currently identified glycemic trait loci together account for no more than 5% of variance for any one glycemic trait (Scott et al. 2012), improved approaches are also needed to identify additional genetic determinants of trait variability.

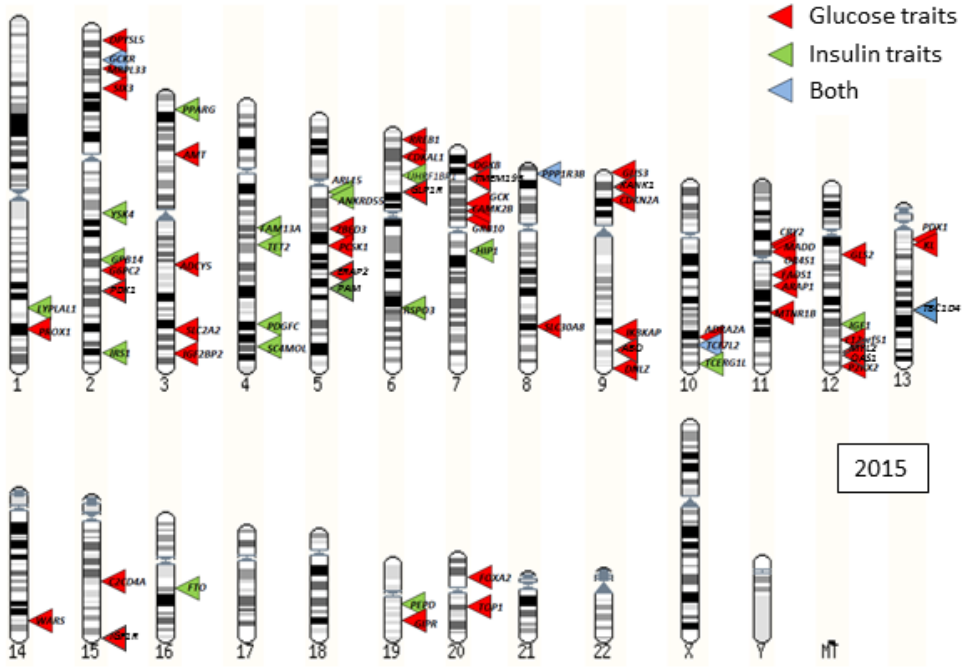


Figure 1.2.3.1. Diagram of regions in the genome associated with glucose (FG, 1hGlu, 2hGlu) and insulin (FI) related traits in published genome- and exome-wide association data. Figure obtained from Anubha Mahajan and modified. Not drawn to scale.

1.2.4. Major advancements in variant discovery

The success of GWAS was made possible by numerous developments. Association testing has primarily been based on large-scale genotyping of single nucleotide polymorphisms (SNPs) using array-based platforms and information from catalogues of human genome sequence variation such as the HapMap project (International HapMap 2005; International HapMap et al. 2007). However, GWAS regions tend to be large and due to linkage disequilibrium (LD) multiple possible candidate variants or even genes may be implicated. The reference set has now been extended through sequencing of larger number of individuals with wider representation from multiple ancestries, as part of the 1000 Genomes Project, to obtain denser genome coverage especially for lower-frequency SNPs to enable more comprehensive SNP imputation (1000 Genomes Project et al. 2012). These developments have also been coupled with rapid improvements in cost-effective genotyping arrays and high-throughput sequencing technologies.

The more recent efforts in variant discovery have focused on lower-frequency variation (MAF<5%) using whole genome and whole exome sequencing methods (WGS, WES) to decipher their role in complex phenotypes (Kiezun et al. 2012). WGS can identify both coding and regulatory variants, but studies focusing on exomes, though constituting only ~1% of the total genome, have garnered strong interest for a number of reasons. First, analysis of protein-coding regions cost substantially less than that for whole genome analyses and can therefore be carried out in larger sample sizes, enabling the identification of rarer variants that may confer larger effects than common variants. Second, coding variant signals are more likely to pinpoint precise effector transcripts that are directly involved in disease risk, facilitating the design of functional pipelines for studying protein perturbation.

One early study demonstrated how WES may benefit clinical diagnoses by identifying *de novo* heterozygous inactivating mutations in the *GATA6* gene that were likely to be disease-causing in a subset of individuals with unexplained pancreatic agenesis (Lango Allen et al. 2011). These findings highlighted the essential role of *GATA6* in human pancreas development. WES studies were also able to identify novel associations of low frequency and/or rare coding variations associated with common cardiometabolic phenotypes, though for these studies sample sizes were small (Albrechtsen et al. 2013; Lange et al. 2014). A recent study by the deCODE consortium used WGS with imputation data to perform exome

chip genotyping of large samples of Icelandic T2D cases and controls. The authors reported novel missense variants in *PAM* (one common and one rare) that conferred moderate risk for T2D, confirming *PAM* as the T2D-associated gene at the *PAM/PPIP5K2* locus (Steinthorsdottir et al. 2014). The common (MAF=5%) *PAM* variant (rs35658696, p.D563G) was previously only found to influence insulinogenic index in an exome array analysis, which had highlighted *PAM* as a novel locus for a T2D-related trait (Huyghe et al. 2013). Since then, functional follow-up has also demonstrated that these missense variants impact on protein function and activity, and that PAM has a previously unidentified role to play in beta cell function (Raimondo et al., unpublished).

To date, most variant discovery efforts have been restricted to specific populations, and the largest studies have focused on populations of European or North American origin. Undertaking studies in multiple ethnic groups can help to discover new associations and identify risk alleles that are shared between ancestry groups or likely to be ancestry-specific. A recent trans-ethnic meta-analysis of genome-wide data from five different ancestries – European, East Asian, South Asian, Mexican and Mexican American – identified seven new T2D susceptibility loci and helped to refine common variant association signals at established loci with the increased sample size and reduced overall LD structure (Mahajan et al. 2014). Since then, the field has continued to see ever-increasing study sizes through large scale consortia efforts. Large-scale association analyses or meta-analyses have greatly increased the power to detect more genetic variants associated with complex traits especially for variants with small effect sizes and of lower frequency. A pair of large collaborative projects – T2D Genetic Exploration by Next-generation sequencing in multi-Ethnic Samples (T2D-GENES) and Genetics of T2D (GoT2D) – generated large-scale genome and exome genotyping and sequencing datasets across multiple consortia to expand the discovery of genetic factors for T2D and its related traits and to more precisely identify causal variants.

Improvements in the availability and power of statistical analysis tools have added much needed value to genetic discovery in this post-GWAS era. Bioinformatics tools and *in silico* software are often used to functionally annotate nonsynonymous variants identified from sequencing studies to predict their pathogenicity, and to filter variants before carrying out laborious functional validation assays (Flannick et al. 2013; Juang et al. 2014). To test for associations between rare variants and complex traits, gene-based association tests are

used to aggregate the effects of rare variants within a gene, as classical single-marker association analyses lack power when evaluating individual variants of low MAF. Many sophisticated algorithms have been newly developed to increase the power and confidence of prioritising plausible rare causal variants amidst the large number of variants identified within the gene (Gamazon et al. 2015; Kircher et al. 2014; Lin 2016; Moutsianas et al. 2015).

Finally, to validate the associations between identified variants and the relevant traits, and to comprehend the mechanisms underlying these links, detailed functional assessments in the laboratory are necessary.

1.3. Translation of genetic association signals into biological discoveries

As the vast majority of variants identified through GWAS map to poorly annotated sequence in non-coding regions, there is often a lack of certainty of the exact transcripts through which the variants impact on phenotype. The observed signals may also be tagging other SNPs that are the real causal variants, and common variant signals with modest effect sizes make it difficult to detect differences between risk and non-risk alleles in functional assays. Thus, few variants or genetic loci have been attractive candidates to pursue in the laboratory (Ng and Gloyn 2013). Nevertheless, there are now many systematic pipelines in place to help transit from association signals to transcripts, molecular and cellular mechanisms.

1.3.1. Characterisation of coding variation in genes of known function

Variants located within the coding region provide a tractable route for characterisation and elucidation of functional mechanisms, especially those within biologically plausible candidate genes. A well-studied example is the common SNP rs1260326 (p.P446L) in *GCKR*, which has been reproducibly associated with T2D risk, FG levels, triglyceride levels and other metabolic traits (Dupuis et al. 2010; Orho-Melander et al. 2008; Saxena et al. 2007). *GCKR*, which encodes the glucokinase regulatory protein (GKRP), is a pivotal regulator of liver GCK activity and hence hepatic glucose metabolism. Validation of GKRP-P446L in microscopy, kinetic, and protein-protein interaction assays showed that the variant had decreased ability to inhibit GCK action. These data offered a plausible mechanism by which GKRP-P446L can

simultaneously reduce FG, protect against T2D risk, and result in an inverse effect on triglyceride levels (Beer et al. 2009; Rees et al. 2012a). Another study made use of the same suite of assays to analyse up to 19 rare nonsynonymous *GCKR* variants (some novel) identified from individuals at high risk of coronary artery disease, and showed that the majority of these variants affected protein function (Rees et al. 2012b). Critically, phenotypic associations with triglyceride and cholesterol levels were improved when wild type or variant carriers were reclassified based on results of the functional analysis. The authors also showed that mutation prediction algorithms, though useful for preliminarily assigning pathogenicity to large numbers of nonsynonymous variants, need to be validated by *in vitro* experiments.

Another gene for which investigation of coding variation has reaffirmed a functional link between a GWAS-identified susceptibility locus and T2D risk is *MTNR1B* (Bouatia-Naji et al. 2009; Prokopenko et al. 2009). It encodes melatonin receptor 1B, a G protein-coupled receptor known to be involved in the regulation of circadian physiology through its ligand melatonin. The group individually evaluated 40 nonsynonymous coding variants identified in a European cohort (mostly very rare with MAF<0.1%) for melatonin binding and signalling capability. Variants that displayed total loss of function (LOF) were strongly associated with increased T2D risk, while the other seemingly neutral variants were not (Bonnefond et al. 2012). Critically, these findings revealed an inverse relationship between MTNR1B function and T2D risk, contrary to what had been proposed based on previous associations observed between the common non-coding *MTNR1B* GWAS variant and T2D risk (Lyssenko et al. 2009). Such comprehensive functional variant profiling has since been performed on many other genes now, including T2D risk locus *PPARG* (Majithia et al. 2014) and myocardial infarction risk locus *LDLR* (Thormaehlen et al. 2015). In fact, an extended study of *PPARG* made use of a high throughput platform to simultaneously generate and evaluate all possible missense substitutions in the PPARG protein on CD36 expression, and this information was subsequently used to guide interpretation of variants of unknown significance identified from sequenced individuals (Majithia et al. 2016).

These studies have demonstrated the value of systematic *in vitro* studies in accurately identifying potentially disruptive variants and in providing an indication of the direction of effect of gene function on the phenotype. They not only generated valuable mechanistic insight into how rare genetic variation may contribute to complex phenotypes or Mendelian

disease, but are also beneficial for risk prediction, clinical diagnosis and therapy. Nonetheless, not all gene products are suitable for *in vitro* experimentation at scale, and specific strategies will have to be devised to develop relevant and scalable assays for genes of clinical significance.

1.3.2. Strategies for investigating non-coding association signals

For the majority of disease association signals that lie within non-coding regions, transition from signal to causal transcript is less straightforward. One approach is to determine whether genetic variants are associated with changes in levels of neighbouring transcripts. Such variants, termed expression quantitative trait loci (eQTL), may regulate other transcripts in *cis* or *trans*. Studies have used expression datasets in disease-relevant tissue such as adipose tissue or pancreatic islets to refine the association signal and ascertain effector transcripts (Kilpelainen et al. 2011; Rung et al. 2009; Small et al. 2011; van de Bunt et al. 2015). Links between an associated variant and regulatory motif may also be established using chromatin-conformation capture (3C) approaches coupled with sequencing, based on the premise that regulatory variants localised to enhancers interact with promoter regions to regulate gene transcription (Hughes et al. 2014; Lieberman-Aiden et al. 2009).

In line with establishing detailed transcriptional maps, much work has been done to improve sequence annotation by carrying out genome-wide high-resolution epigenomic profiling, to increase our understanding of the regulatory landscape in which genetic variants may act, in a context-specific manner. The ENCODE project is one such effort that has generated epigenomic maps across multiple human tissue types (Encode 2012), but does not include data from isolated pancreatic islets or beta cells. Subsequently, many groups have tackled this limitation by carrying out studies on human islets and sometimes sorted alpha and beta cell fractions to catalogue chromatin state, DNA hypersensitivity, histone marks, transcription factor and regulator binding sites (Ackermann et al. 2016; Bhandare et al. 2010; Claussnitzer et al. 2014; Dayeh et al. 2014; Gaulton et al. 2010; Nica et al. 2013; Pasquali et al. 2014; Stitzel et al. 2010). Other studies have also characterised non-coding RNA elements such as long non-coding RNAs (lncRNAs) and microRNAs (miRNAs), which are less often studied in human islets but believed to have diverse cellular roles in beta cell development

and function (Moran et al. 2012; van de Bunt et al. 2013). These elements will be important when considering the influence of non-coding genetic risk variants on T2D pathogenesis.

Since then, many T2D risk and related quantitative trait loci have been found to overlap with islet cell regulatory elements (Parker et al. 2013; Pasquali et al. 2014). These data together provide a vital resource for follow-up studies of risk variants that act through beta cell function. A recent study combining fine-mapping, genomic annotation and functional assays showed that the index non-coding GWAS SNP at established T2D locus *MTNR1B* (rs10830963) overlapped a FOXA2 transcription factor binding site. Importantly, the risk allele resulted in preferential NEUROD1 binding in islet-derived cells which may be responsible for increasing FOXA2-bound enhancer activity and thus increase T2D risk (Gaulton et al. 2015). This is just one of many examples that have harnessed genetic and genomic data in clarifying the biology underlying T2D susceptibility loci, which may harbour regulatory variants of unclear significance. Further discussions of such contributions to the field are well beyond the scope of this section. To enhance the utility of these tools, it is important for data to be made publicly accessible through various online genome browsers and databases such as RegulomeDB, which combines information from ENCODE and other datasets to enable functional annotation of non-coding loci (Boyle et al. 2012), and the Variant Effect Predictor of the Ensembl Genome Browser, which enables lookups of the consequences of putative regulatory variants on transcripts and proteins (McLaren et al. 2010). Integrative approaches are now critical for the move from association signals to causal regulatory variants and molecular mechanisms (Knight 2014). In my next sections I will zoom in on one of the earliest glycemic trait loci to be determined and what is currently known about its biology.

1.4. Genetic associations in the glycemic trait locus *G6PC2/ABCB11*

Two landmark papers published in 2008 that aimed to identify associations with glycemic traits both highlighted multiple common non-coding variant signals (rs560887 and rs563694) at the *G6PC2/ABCB11* locus (Bouatia-Naji et al. 2008; Chen et al. 2008). Subsequent GWAS meta-analysis replicated lead GWAS SNP rs560887, which is intronic to *G6PC2*, as among the strongest contributors to FG variation in terms of statistical significance and effect size per allele (Dupuis et al. 2010). Other studies also confirmed that the glucose-lowering allele at rs560887 was associated with lowered HbA1c and increased beta cell insulin secretion (HOMA-B model and insulinogenic index), respectively reflecting better long-term glucose control and beta cell function (An et al. 2014; Chen et al. 2014; Ingelsson et al. 2010).

As seen in Figure 1.4.1, multiple FG non-coding signals identified were in the same LD block as rs560887 and mapped within or near to both *G6PC2* and the adjacent gene *ABCB11*. At the time, it was unclear whether the effector transcript underlying the association signal was *G6PC2* or *ABCB11* (or both) as both genes appeared to be biologically-plausible candidates. *G6PC2* encodes the islet-specific glucose-6-phosphatase catalytic subunit; its function will be extensively described in subsequent sections. *ABCB11*, on the other hand, encodes an ATP-binding cassette transporter which acts as a bile salt export pump in liver hepatocytes. Changes in bile salt secretion and hence bile acid pools can affect lipid and glucose metabolism. Mutations in the gene are known to cause familial cholestatic liver disease (Lam et al. 2007; Strautnieks et al. 1998). *ABCB11* expression is known to be absent in human beta cells (Blodgett et al. 2015). Since both genes encode proteins that could be involved in glucose regulation, the possibility that *ABCB11* could be mediating variation in FG at this GWAS locus was not completely ruled out.

G6PC2 was not significantly associated with glucose tolerance, insulin sensitivity measures, risk of T2D or other diabetes-related and anthropometric traits in large-scale GWAS (Dupuis et al. 2010; Scott et al. 2012). Mutations in *G6PC2* have also not been implicated in monogenic forms of diabetes (Bonnefond et al. 2009; Edghill et al. 2009). However, in select population groups the glucose-raising allele at rs560887 in *G6PC2* was paradoxically associated with increased acute insulin response after oral or intravenous glucose load and increased insulinogenic index (Heni et al. 2010; Li et al. 2009; Rose et al. 2009). It was suggested that the effect of this allele on *G6PC2* function may disrupt insulin signalling

efficiency, which affects hepatic glucose sensitivity resulting in a compensatory increase in insulin secretion. The link between *G6PC2* and T2D risk has also been a matter of debate. A number of studies, some of which were performed in Asian cohorts, have found the FG-raising allele at rs560887 to be counterintuitively protective for T2D risk though this was with at best marginal statistical significance (Dupuis et al. 2010; Hu et al. 2010; Takeuchi et al. 2010; Tam et al. 2010; Wang et al. 2013). The contradictory findings from human genetic studies of *G6PC2* may be partially explained by the current poor understanding of the functional significance of rs560887, a potential regulatory variant, and its effects on transcript expression and function *in vivo*. GWAS regulatory variants might influence the function of more than one transcript (such as *G6PC2* and/or *ABCB11* expression) in the same or different tissues. It is also possible that there are inter-ethnic differences in allele frequency and different genetic determinants influencing the effects of regulatory *G6PC2* variants. These uncertainties hinder biological interpretation as it was unclear whether the association with FG at this locus was acting through the islet beta cells or the liver/gut.

The glucose-6-phosphatase enzyme (G6Pase) catalyses the reverse reaction mediated by the glucose-phosphorylating enzyme GCK, which is recognised as the beta cell glucose sensor (Matschinsky 1996, 2005). Therefore, G6PC2 has been suggested as a regulator of the glucose-sensing mechanism in beta cells. Human GWAS may provide additional clues to the significance of beta cell glucose phosphorylation in fasting glucose homeostasis. Variation in both *GCK* and *GCKR* is strongly associated with FG levels. While the former is linked to beta cell insulin secretory ability, the latter is likely to regulate insulin sensitivity through its action in the liver and is therefore unlikely to participate in the same glycolytic pathway involving *GCK* and *G6PC2* in the beta cells (Bouatia-Naji et al. 2008; Dupuis et al. 2010; Ingelsson et al. 2010; Orho-Melander et al. 2008; Rose et al. 2005; Saxena et al. 2007; Sparso et al. 2008). Analysis of the interaction between the *G6PC2* rs560887 and *GCK* rs1799884 lead GWAS SNPs showed that the glucose-lowering alleles at these SNPs had an additive effect on FG levels, underscoring the role of this pathway in glucose homeostasis in the normoglycemic population (Bouatia-Naji et al. 2008; Li et al. 2009).

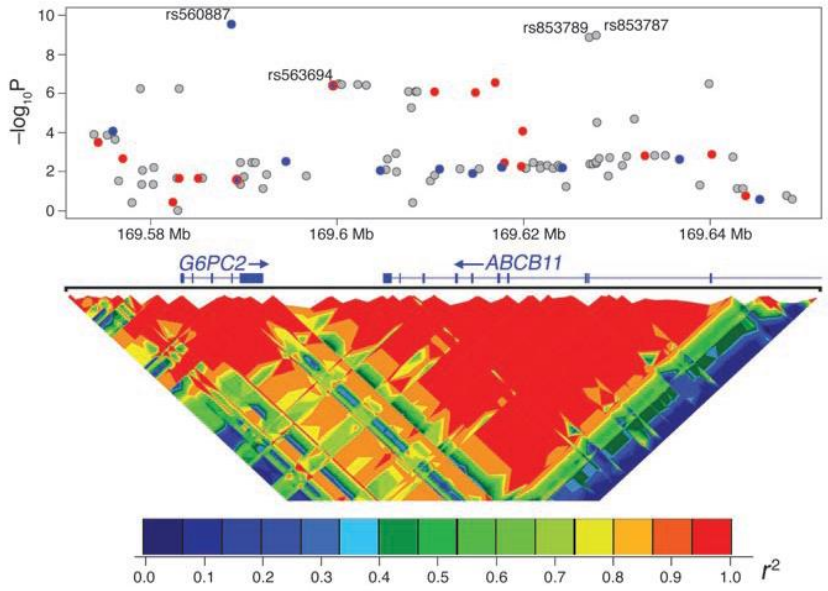


Figure 1.4.1. FG association plot at the *G6PC2/ABCB11* locus from Chen et al. Top panel shows level of association in -log(P value) against genomic position. Coloured dots indicate genotyped SNPs and grey dots indicate imputed SNPs in the data. Bottom panel shows the LD pattern (r^2) with the scale displaying magnitude of LD from low (blue) to high (red). Figure taken from Chen et al. 2008.

1.5. *G6PC2*, the islet-specific glucose-6-phosphatase catalytic subunit

1.5.1. Molecular structure and function of G6PC2 as currently understood

G6PC2 is expressed as a full-length transcript (comprising five exons) and alternatively spliced transcripts, including a short transcript which lacks exon 4 and has a truncation at exon 5 (as compared to the full-length transcript) due to a premature stop codon (Figure 1.5.1.1). The full-length transcript encodes a 355-amino acid protein while the short transcript which also appears to be abundantly expressed (de Jong et al. 2013; Dogra et al. 2006; Martin et al. 2001) gives rise to a 154-amino acid protein. There is no evidence that other splice variants are protein-coding. Based on currently unpublished RNA sequencing data of up to 174 human islets, expression levels of the full-length *G6PC2* isoform was comparable with that of the short isoform; both appear to be largely beta-cell specific (Martijn van de Bunt, personal communication). The functional significance of the short isoform is unknown but is likely to lack hydrolytic activity due to the removal of several active site residues (Figure 1.5.1.1).

G6PC2 is a highly hydrophobic nine-transmembrane glycoprotein localised to the endoplasmic reticulum (ER) of beta cells (Hutton and Eisenbarth 2003; Shieh et al. 2004). This is believed to be due to a consensus ER retention motif (amino acid sequence KKTK) for transmembrane-resident proteins at the C-terminal end (Jackson et al. 1990). It also has an established N-glycosylation site in the ER luminal domain that is conserved within the protein family (Pan et al. 1998a), and possesses a phosphatase sequence motif with conserved active site residues that form a catalytic centre in the luminal loop, suggesting conservation of enzymatic function (Arden et al. 1999). Figure 1.5.1.2 shows a schematic of the structure of G6PC2.

While GCK acts in the cytoplasm, G6PC2 is located in the ER and its action relies on transporters for shuttling substrates into the lumen and for products to be transported out of the ER. G6Pase is known to be part of a multi-component system which consists of the catalytic domain and transporters for G6P (encoded by *G6PT/SLC37A4*) and the reaction products glucose and inorganic phosphate (Pi), as illustrated in Figure 1.5.1.3. G6P hydrolysis is tightly coupled with G6PT activity but the identity of other transporters has not been characterised (Foster and Nordlie 2002; van Schaftingen and Gerin 2002). Therefore, the rate and substrate specificity of G6PC2 may be limited by G6PT. The metabolic actions of GCK and G6PC2 are likely to be tightly linked, but the precise molecular function of G6PC2 is less well established. As its mode of action antagonises GCK in glucose phosphorylation, the first and rate-limiting step in glycolysis, it is believed to be part of a futile substrate cycle in the beta cells. The concurrent phosphorylation of glucose to G6P and dephosphorylation of G6P to glucose is an ATP-consuming process. Glucose cycling between G6PC2 and GCK has been predicted to result in reduced glycolytic flux, ATP/ADP ratio and hence insulin secretion from the beta cell (Pound et al. 2013). The substrate cycle is an important regulatory process as it delivers sensitivity in the control of glucose metabolism by modifying enzyme activity to regulate substrate flux through a specific pathway (Curi et al. 2016; Katz and Rognstad 1978). Substrate cycles in the beta cell have been implicated in the regulation of GSIS (Jensen et al. 2008).

For a long time, glucose cycling rates and G6Pase activity detected in pancreatic islets have been inconsistent across studies and often at significantly lower rates than that observed in liver cells in rodents (Ashcroft and Randle 1968; Foster et al. 1997; Khan et al. 1989; Khan et al. 1990a; Khan et al. 1995; Perales et al. 1991; Sweet et al. 1997). Some studies have

attempted to detect phosphohydrolase activity directly from overexpressed G6PC2 protein but with little success (Arden et al. 1999; Martin et al. 2001; Martin et al. 2002). The differences in findings were attributed to differing assay conditions, while others were due to experiments carried out on rat islets, which have since been established to lack G6pc2 protein expression (Martin et al. 2001). Two groups did subsequently report detectable G6PC2 activity, though it was found to be ~20-fold lower than that of liver G6Pase encoded by *G6PC1* (Hutton and O'Brien 2009; Petrolonis et al. 2004).

A more recent study using a novel stable isotope method was able to demonstrate higher rates of glucose cycling than was previously reported, and critically demonstrated that G6pc2 was responsible for this as glucose cycling was abolished in islets from *G6pc2* knockout (KO) mice (Wall et al. 2015). Though the presence of G6P hydrolysis in mouse islets mediated by G6pc2 was confirmed, it would be informative if these experiments were conducted in human islets. Increased glucose cycling and G6Pase activity were found to be increased in animal models of T2D such as obese hyperglycemic mice (Chandramouli et al. 1991; Khan et al. 1990b; Khan et al. 1995) and non-obese Goto-Kakizaki (GK) diabetic rats (Ostenson et al. 1993). It is possible that increased glucose cycling, which requires ATP, may regulate insulin release and glycemia in these mice. Nonetheless, these results have generated controversy on whether the level of G6P hydrolysis in beta cells is enough to exert a meaningful impact on GSIS. The hepatic glucose cycle is well-established as a regulator of glucose homeostasis in the blood stream via its action in gluconeogenic tissues, but evidence of an equivalent pathway in the beta cell is less clear.

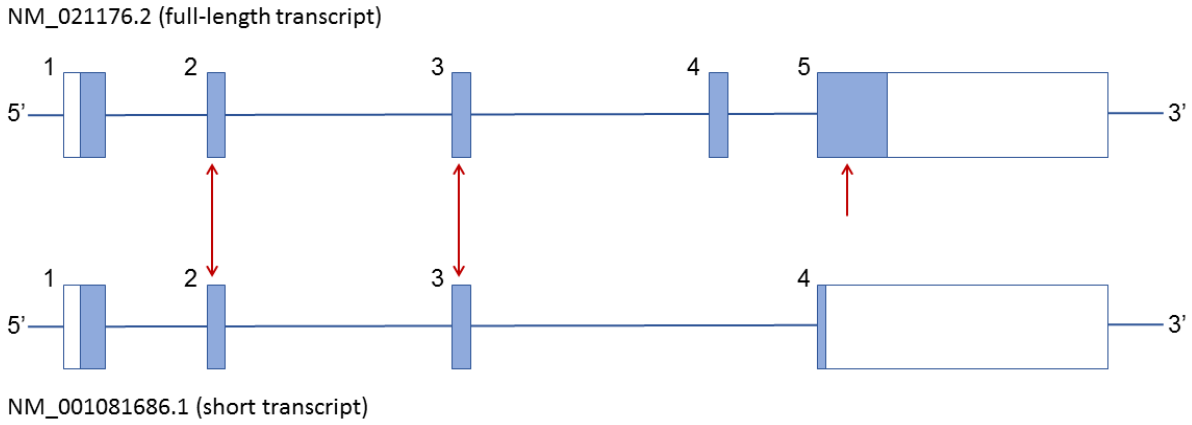


Figure 1.5.1.1. Known protein-coding transcripts of *G6PC2*. The exon/intron structure of the full-length transcript (top) and short transcript (bottom) are shown, corresponding to their respective reference sequences. Exons are labelled 1-5 with dark boxes representing coding regions and white boxes representing 5' and 3' untranslated regions. The red arrows denote regions that encode the conserved active site residues in the protein. Illustration is not drawn to scale.

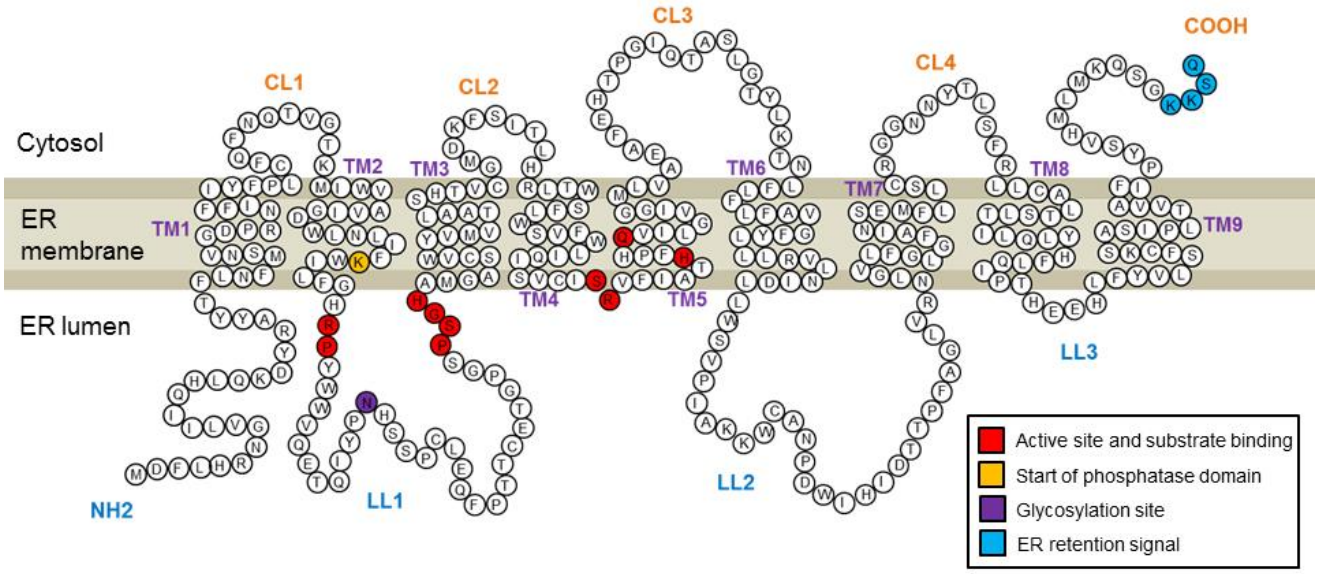


Figure 1.5.1.2. Illustration of the predicted structure of G6PC2 protein in the ER membrane. Amino acids are coloured as described in the legend. Transmembrane regions were annotated based on NCBI reference sequence NP_066999.1. NH2: amino-terminus; TM: transmembrane region; CL: cytoplasmic loop; LL: endoplasmic reticulum luminal loop; COOH: carboxy-terminus.

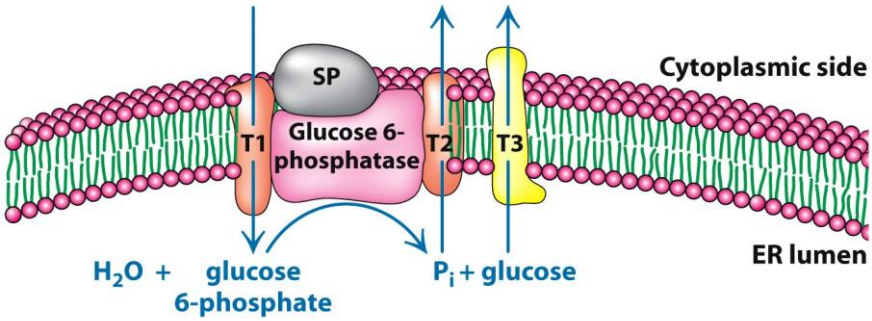


Figure 1.5.1.3. Illustration of the glucose-6-phosphatase system. The glucose-6-phosphatase catalytic subunit forms a complex with the G6P transporter (T1) and transporters for inorganic phosphate (T2) and glucose (T3). In this schematic it is also stabilised by a stabilising protein (SP). Source: Berg, Tymoczko & Stryer (2012) *Biochemistry (7th ed)* Basingstoke: W.H. Freeman.

1.5.2. Insights from the study of G6pc2 in rodent models

Rodent models are useful for investigations of beta cell biology and disease pathophysiology. The *G6PC2* transcript is highly expressed in both human and mouse beta cell fractions (Baron et al. 2016). Studies of *G6pc2* KO mice models by the O'Brien group have greatly contributed to the understanding of the biological role of *G6pc2*. G6pase activity in the islets of global *G6pc2* KO mice was abolished as compared to control mice. These KO mice also had ~15% decrease in FG levels without change in FI (Pound et al. 2013; Wang et al. 2007), consistent with human GWAS data. The authors hypothesised that G6pc2 action reduces glycolytic flux and thus shifts the dose response curve for GSIS leftwards (Pound et al. 2013). The model, as illustrated in Figure 1.5.2.1, shows that under fasting conditions *G6pc2* KO mice have reduced blood glucose levels; at sub-maximal glucose stimulation levels, glucose tolerance appears to be improved; at maximal glucose stimulation there is no change in glucose tolerance presumably due to the high rate of glycolysis mediated by Gck at saturation point. This model supports the counterintuitive lack of change in FI levels in *G6pc2* KO mice as maximal GSIS is unaffected, and plasma FI levels are an indicator of insulin sensitivity/resistance rather than glucose sensitivity (Laakso 1993; Olefsky et al. 1973).

In addition to potential effects on absolute insulin secretion at sub-threshold levels, G6PC2 has also been postulated to affect the pulsatility of insulin secretion, which could affect the efficacy of insulin signalling on other tissues (O'Brien 2013). Insulin release from the pancreatic islets occurs in pulses due to a combination of oscillations in metabolic activity, electrical activity and cytosolic Ca^{2+} concentrations (Lang et al. 1979; Merrins et al. 2010). Metabolic oscillations may be modulated from the early steps of glycolysis due to the ability of bifunctional enzymes to alter flux through the glycolytic cycle (Bertram et al. 2010; Merrins et al. 2012). Alternatively, G6PC2 may also regulate intracellular Ca^{2+} levels and hence Ca^{2+} signalling. In islets from *G6pc2* KO mice, cytoplasmic Ca^{2+} levels were found to be elevated (Pound et al. 2013). Conversely, increasing G6Pase activity (by expressing *G6pc1*) in MIN6 cells was capable of reducing ATP production, giving rise to a blunted intracellular Ca^{2+} response and a resulting inhibition of GSIS (Iizuka et al. 2000). Given that G6PC2 resides in the ER, which is a major storage site of calcium, G6P hydrolysis and generation of inorganic phosphates in the ER could impact on Ca^{2+} homeostasis (Wolf et al. 1986). ER calcium stores

not only regulate normal beta cell membrane potential through Ca^{2+} sequestration but also influence the Ca^{2+}-dependent triggering pathway of insulin secretion, which is dependent on both voltage-dependent Ca^{2+} channel activity and intracellular Ca^{2+} release (Gilon et al. 1999; Roe et al. 1993; Worley et al. 1994).

Finally, G6PC2 has also been implicated in autoimmune diabetes as G6pc2 was found to be a major autoantigen to a substantial proportion of diabetogenic CD8+ T cells in non-obese diabetic (NOD) mice, a model of T1D (Lieberman et al. 2003). Though the role of these G6pc2 peptides in the pathophysiology of autoimmune diabetes is unclear, they were deemed unlikely to be responsible for the onset of the disease in NOD mice (Krishnamurthy et al. 2006a; Oeser et al. 2011). In humans, epitopes in G6PC2 have also been identified as targets of autoreactive T cells in T1D subjects (Jarchum et al. 2008; Mallone et al. 2007; Ouyang et al. 2006) as well as in healthy subjects (Yang et al. 2006). Despite the prevalence of potentially autoreactive G6PC2-specific T cells, the significance of G6PC2 in immune-mediated diabetes is unclear. No association between *G6PC2* and T1D has also been observed in human GWAS (Polychronakos and Li 2011).

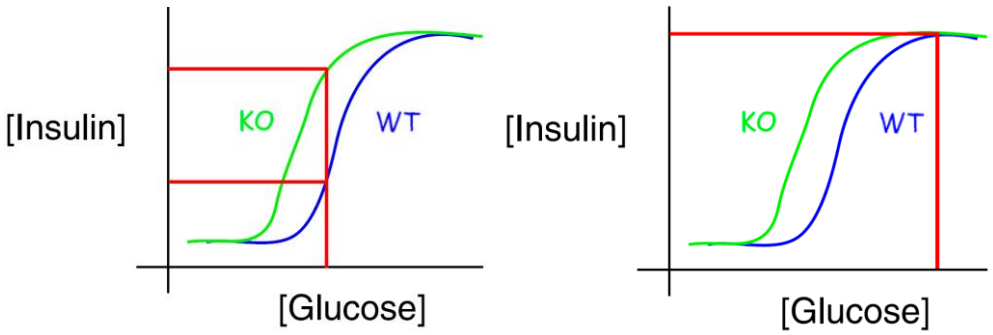


Figure 1.5.2.1. Hypothetical model showing that deletion of *G6pc2* in mice results in a leftward shift in the dose-response curve for GSIS. The model proposes that at sub-threshold glucose stimulation levels (left), KO mice have enhanced insulin secretion; at maximal glucose stimulation (right), there is no difference in insulin secretion between KO and WT mice. KO: knockout, WT: wild type. Figure taken from Pound et al. 2013.

1.6. The glucose-6-phosphatase catalytic subunit gene family

G6PC2 belongs to the same gene family that comprises two other genes, *G6PC1* and *G6PC3*, all of which encode G6Pase catalytic subunits with different tissue-specificities and roles in the body. Some of the properties of these proteins are briefly summarised in Table 1.6.1. Like *G6PC2*, both *G6PC1* and *G6PC3* encode proteins integral to the ER membrane and possess similar membrane topologies and phosphatase consensus sites. However, the G6PC3 protein lacks the consensus N-glycosylation site and ER retention signal unlike its other counterparts (Martin et al. 2002). All three protein isoforms also possess conserved active site regions, which are part of an extended sequence motif characteristic of mammalian lipid and acid phosphatases and bacterial vanadate-sensitive haloperoxidases (Hemrika et al. 1997; Stukey and Carman 1997). While *G6PC2* is found to be exclusively expressed in the pancreatic islets, *G6PC1* is predominantly expressed in the liver, kidney and intestine (Foster et al. 1997; Mithieux 1997). *G6PC3* is known to be ubiquitously expressed with high levels detected in the skeletal muscle, heart, brain and other tissues including the pancreas in both humans and mice (Boustead et al. 2004; Martin et al. 2002).

The distinct tissue expression patterns are due to differential transcriptional regulation. The promoter region of *G6PC2* is highly conserved in mice and humans and is substantially more active in rodent and hamster insulinoma cells than in other cell types (Ebert et al. 1999; Martin et al. 2001). Many of the key transcription factors that bind to the *G6pc2* promoter overlap with those that regulate insulin gene transcription, such as Pdx1, Pax6, Foxa2 and MafA (Hutton and O'Brien 2009). *G6pc1* expression in contrast appears to be regulated by transcription factors known to regulate hepatic genes such as the hepatocyte nuclear factors (HNFs), and other factors that affect hepatic glucose output such as insulin, glucocorticoids, glucose and glucagon (Hutton and O'Brien 2009; van Schaftingen and Gerin 2002). As for *G6PC3*, promoter sequences of both the human and mouse varieties appear to have putative binding sites for factors such as NFκB, Sp1 and E-box binding factors that are known to have wide-ranging tissue distributions (Boustead et al. 2004; Martin et al. 2002), but no detailed characterisation studies have been carried out.

G6PC1, the most well-characterised member of the family, is responsible for hepatic and renal glucose production through gluconeogenesis and glycogenolysis, and is therefore a key enzyme in the regulation of blood glucose levels. Evidence of its role in glucose

homeostasis will be further discussed in the next section. The role of *G6PC3* is much less well-defined *in vivo* even though it is found to be expressed in multiple tissues. Activity studies have shown that expressed G6PC3 can hydrolyse G6P like G6PC1, but with a ~6-fold lower maximal rate of reaction (Guionie et al. 2003; Shieh et al. 2003). This suggests that G6PC3 is capable of endogenous glucose production in other tissues, but its contribution to blood glucose homeostasis is unclear as it has not been associated with any metabolic traits. Instead, *G6PC3* deficiency is known to cause a rare autosomal recessive disorder called severe congenital neutropenia syndrome characterised by neutrophil and macrophage dysfunction. Patients with LOF mutations in *G6PC3* however display great phenotypic variability (Lin et al. 2015). The function of *G6PC3* will not be further discussed as it is beyond the scope of this thesis.

	G6PC1	**G6PC2**	**G6PC3**
Commonly used aliases	G6PC, G6Pase-α	IGRP	UGRP, G6Pase-β
Chromosomal location of gene	17q21	2q31	17q21
Length of protein (amino acids)	357	355	346
Tissue expression	Liver, kidney, intestines	Pancreatic islets	Ubiquitous
Mechanism of action	Glucose-6-phosphate + H_2O → Glucose + Pi		
Protein sequence identity/similarity[a] to G6PC1 (%)	-	51/67	36/64
Protein sequence identity/similarity[a] to G6PC2 (%)	51/67	-	37/54
Function	Gluconeogenesis and glycogenolysis (hepatic glucose production)	Potential regulator of glucose sensitivity and insulin secretion	Neutrophil function and activity
Phenotypic associations	Glycogen storage disease type Ia (autosomal recessive)	FG levels (GWAS) and T2D risk (GWAS)	Severe congenital neutropenia (autosomal recessive)

Table 1.6.1. Summary of known properties of G6PC1, G6PC2 and G6PC3 proteins. [a]Sequence identity is indicated by residues that are identical; sequence similarity is indicated by residues that have similar chemical characteristics.

1.6.1. The role of liver glucose-6-phosphatase in glucose homeostasis and disease

G6PC1 is the most active G6Pase catalytic subunit in the body and is expressed in major gluconeogenic tissues in the body, especially the liver, and is therefore pivotal to hepatic glycolytic flux and regulation of glycogen levels alongside liver GCK. Interest in G6PC1 first arose as inactivation of G6Pase activity was found to be linked to glycogen storage disease (GSD) in humans, a rare hereditary disease (Cori and Cori 1952). Inactivating mutations in *G6PC1* cause GSD type 1a (GSD1a), the most prevalent of the family of diseases and occurring in ~1 in 100 000 live births (Chen and Burchell 1995). GSD1a is an autosomal recessively-inherited monogenic disorder characterised by severe fasting hypoglycaemia due to lack of glucose production, and heptomegaly and nephromegaly due to excessive glycogen accumulation in the liver and kidneys. There is also a wide range of other clinical and biochemical abnormalities including renal dysfunction, growth retardation and hyperlipidemia (Chou 2001). Mouse models for GSD1a appear to replicate the metabolic disorders in humans and these functions may be restored by sustained expression of a functional G6Pase system in the liver and kidney (Lei et al. 1996; Sun et al. 2002). Current treatments of GSD1a are largely based on dietary regimens to prevent hypoglycaemia (Chen et al. 1984; Greene et al. 1976). There is unfortunately no cure for the disease and these therapies do not completely correct the metabolic derangements. Hence, long term complications often resulting from liver and renal disease are leading causes of mortality and morbidity in patients (Rake et al. 2002).

Other forms of glycogen storage disease exist and are caused by deficiencies in other components of the G6Pase complex (Burchell 1992; Chou and Mansfield 1999; Chou et al. 2002). GSD1b is caused by inactivating mutations in the G6P transporter *G6PT* and have identical metabolic manifestations as GSD1a, except with additional infectious complications including neutropenia, neutrophil dysfunction and inflammatory bowel disease (Chen et al. 2002; Dieckgraefe et al. 2002; Hiraiwa et al. 1999; Visser et al. 2000). As the expression of *G6PT* is more ubiquitous than *G6PC1*, defects in its activity are expected to impact on additional cell types. Other GSD sub-types have also been reported to correspond to defects in the putative phosphate transporter and glucose transporter, though the genes encoding these proteins are yet to be characterised (Chou and Mansfield 2008). Understanding the molecular basis of different types of GSDs sheds light on the roles of the

different components of the G6Pase system, and how dysregulation of these components give rise to diverse disease conditions.

Increased HGP in part due to activated G6Pase action has been linked to diabetes, though results have not been consistent across studies. *G6PC1* expression and activity was found to be increased in diabetic rat models in both liver and kidney as well as in glucose-conditioned human liver cell lines, likely due to the regulatory effects of insulin deficiency and hyperglycemia (Ashmore et al. 1954; Barzilai and Rossetti 1993; Li et al. 1999; Mithieux et al. 1996). Enhanced endogenous glucose production was also observed in human subjects with non-insulin-dependent diabetes (Mitrakou et al. 1990; Tappy 1995). *G6PC1* gene transcription is repressed by insulin signalling (Streeper et al. 1997; Vander Kooi et al. 2003), therefore altered insulin signalling may exacerbate the dysregulation of blood glucose levels in diabetes through G6PC1. However, a more recent study reported no increase in hepatic expression of *G6PC1* in a hyperglycemic rat model or in human T2D subjects despite the fasting hyperglycemia observed (Samuel et al. 2009). There are also no genetic associations found in *G6PC1* for T2D risk or glucose tolerance in human GWAS. Therefore, other cellular mechanisms may take precedence in the regulation of glucose homeostasis in diabetes.

1.7. Aims of thesis

In my introduction, I have outlined the major pathways known to be responsible for glucose homeostasis in the body, with emphasis on the liver and pancreatic islets, and how dysregulation of these mechanisms may lead to the derangement of glycemia and development of T2D. Human genetics research has vastly improved our understanding of the genetic basis of T2D and glycemic traits and facilitated the elucidation of novel functional mechanisms. These were possible due to technological advancements that enabled large-scale association testing and rigorous interpretation of genetic loci. As coding variants are typically easier to pursue in the laboratory than non-coding variants due to the implication of specific proteins, much effort has been put into combining coding variant discovery and *in vitro* functional validation. Nonetheless, most association signals are non-coding and hence the field is continuously seeing promising progress in comprehensive context-specific genomic and epigenomic profiling for evaluating the impact of non-coding regulatory variants.

My thesis focuses on the *G6PC2* glycemic trait locus, where the *G6PC2* gene is a plausible effector transcript for influencing glycemic trait levels due to knowledge gleaned from rodent models and current understanding of the role of the G6Pase gene family in the body. The overarching aims for each part of the thesis are as follows:

1. In Chapter 3 I aim to determine the effector transcript underlying the *G6PC2/ABCB11* FG locus through the identification and validation of coding variant associations in *G6PC2*. This work is part of an exome array analysis for glycemic traits by the T2D-GENES/GoT2D consortia.

2. In Chapters 4 and 5 I functionally characterise coding variants in both *G6PC1* and *G6PC2* across the allelic spectrum to establish a clearer relationship between protein function and phenotypic effects. This work is carried out in collaboration with the MAGIC investigators and T2D-GENES consortium who have amassed data from large-scale exome array meta-analyses and exome sequencing respectively.

3. Finally, in Chapter 6 I aim to shed light on the role of G6PC2 in the beta cell and how functional coding variants in G6PC2 may relate to phenotype through a number of exploratory studies.

Chapter 2

Materials and Methods

2.1. Construction of *G6PC1* and *G6PC2* vectors

2.1.1. Procurement of *G6PC1* and *G6PC2* vector constructs

Human *G6PC1* (NM_000151.3) and *G6PC2* cDNA (NM_021176.2) both within pCMV6-Entry vectors (with a C-terminal Myc-FLAG-tag) were purchased from OriGene Technologies. The vector maps for *G6PC1* and *G6PC2*, clones RC215623 and RC211146 respectively, are shown in Figure 2.1.1.1. The plasmids contain the kanamycin resistance gene (Kan^R) for the purpose of bacterial selection. Bacterial stock plates for each plasmid vector were made and the DNA upscaled for use as described in sections 2.3 and 2.4. A pCMV6-Entry empty vector was cloned for use as an empty vector control in experiments alongside the *G6PC1* and *G6PC2* vectors. This procedure is described in the next sections.

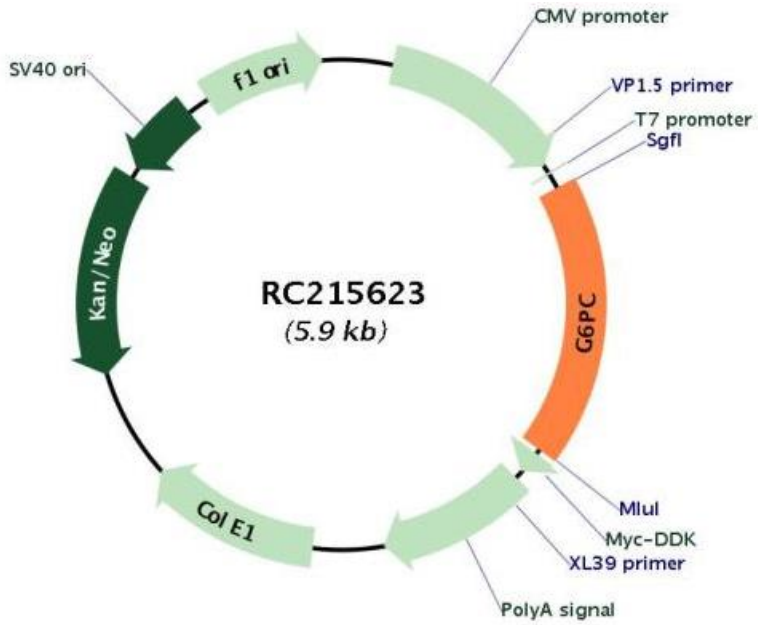
2.1.1. Restriction digest of DNA

To create an empty vector, the pCMV6-Entry-*G6PC2* vector was digested using the SgfI and MluI restriction enzymes (Promega) to excise the *G6PC2* open reading frame (ORF) according to a basic digestion protocol outlined in Table 2.1.2.1. Reactions were incubated at 37°C for 2-3 hours.

Reagent	Volume (µl)
DNA	7
Buffer C	2
BSA	0.5
Distilled H_2O	8.5
SgfI	1
MluI	1
Total	20

Table 2.1.2.1. DNA restriction digest protocol for one reaction. Reactions were carried out in 1.5 ml microcentrifuge tubes. Reagents were obtained from Promega.

A



B

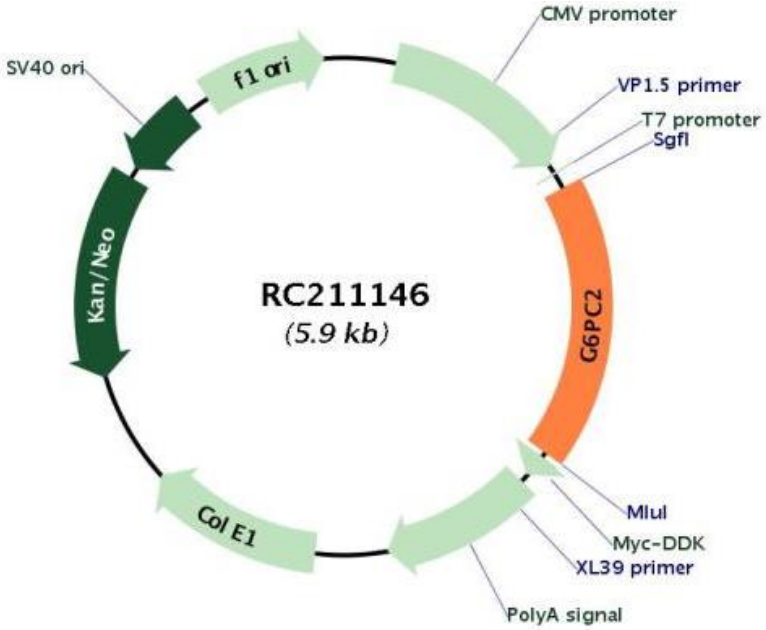


Figure 2.1.1.1. Vector maps for the human cDNA clones for (A) *G6PC1* and (B) *G6PC2* in a pCMV6-Entry vector backbone obtained from OriGene Technologies. Maps show the Myc-DDK (FLAG)-tagged ORFs and the flanking cloning sites SgfI and MluI, as well as a kanamycin resistance cassette in the vector. ORF expression is driven by a CMV promoter. Source: OriGene Technologies.

2.1.2. Visualisation of DNA by agarose gel electrophoresis

To confirm successful digestion of the vector, samples were run on an 0.8% agarose gel. The agarose gel was prepared by dissolving agarose (Sigma Aldrich) in Tris-acetate-ethylenediaminetetraacetic acid (EDTA) (TAE) buffer (Sigma Aldrich) using heat from a microwave. Ethidium bromide (Sigma Aldrich), an intercalator of DNA, was added to the agarose-TAE mixture as a nucleic acid stain to 0.005% of the final solution. Once the gel had set, the gel running tank was filled with 1X TAE buffer. DNA samples were mixed with 6X DNA gel loading dye (New England BioLabs) and loaded into the wells of the gel alongside a one kilobase (kb) plus ladder (New England BioLabs or Thermo Scientific). Gels were run at 120V for 1h (GeneFlow electrophoresis system) and viewed on a Gel Doc 2000 transilluminator system using the Quantity One v4.3.1 software (Bio-Rad).

2.1.3. Purification of digested vector by gel extraction

The DNA band corresponding to the empty linearised pCMV6-Entry vector, which was seen as a distinct band at ~5 kb, was excised from the rest of the gel using a scalpel under ultraviolet (UV) light and placed in a 1.5 ml microcentrifuge tube. The gel may be stored at -20°C until extraction. The QIAquick Gel Extraction kit (Qiagen) was used to extract the DNA from the agarose gel according to manufacturer's guidelines. Sodium acetate at 3 M, pH 5 (BDH Chemicals) was added at 10 µl per sample. The DNA was eluted in 30 µl of RNase-free water and quantified before use.

2.1.4. Blunt-end cloning of the linearised pCMV6-Entry empty vector

The cohesive ends of the digested pCMV6-Entry vector lacking an ORF were filled in to create blunt ends that allow the vector to self-ligate easily without requiring additional restriction sites. The overhangs were filled in by T4 DNA polymerase (Promega) at 5U per µg DNA in the presence of dNTPs according to Table 2.1.4.1. The reaction was carried out at 37°C for 5 min, and was terminated at 75°C for 10 min. The reaction was then purified using the QIAquick PCR Purification kit (Qiagen) according to the manufacturer's guidelines. The DNA was eluted in 30 µl of RNase-free water and quantified before use.

Ligation reactions were then set up for the blunted linearised vector (~4.9 kb) using the T4 DNA ligase and ligase buffer (Promega). Approximately 500 ng of vector and a higher concentration of ligase (2 µl) in a 10 µl reaction was used as compared to typical cohesive-end ligations to increase ligation efficiency. Ligation samples were incubated overnight at 16°C and collected the next day for transformation into competent *Escherichia coli (E. coli)* cells, the procedure of which is described in section 2.3. The original restriction sites of SgfI and MluI in this vector were lost and hence this vector may not be used for subsequent cloning using these sites. The plasmid vector was expanded using the midiprep protocol and the DNA sequence verified as described in section 2.4. This empty vector control was used for both *G6PC1* and *G6PC2* studies in Chapters 3 to 5.

Reagent	Volume (µl)
T4 DNA polymerase buffer (10X)	5
Purified digested DNA (2µg)	23
dNTP mix (10mM)	2
T4 DNA polymerase (10U/µl)	2
Distilled H_2O	18
Total	50

Table 2.1.4.1. Protocol for DNA end-blunting using T4 DNA polymerase for blunt-end cloning for one reaction. Reagents were obtained from Promega.

2.2. Cloning of N-terminal-tagged vectors

2.2.1. Design and annealing of tag oligonucleotides

A V5 epitope tag (amino acid sequence GKPIPNPLLGLDST) which was initially derived from an RNA polymerase α subunit of simian parainfluenza virus type 5 (Southern et al. 1991), was designed and inserted upstream of the *G6PC1* and *G6PC2* ORFs in their respective pCMV6-Entry wild type vectors, to enable the detection of truncated proteins lacking the C-terminal tags. Sense and anti-sense strands corresponding to the V5 sequence were designed by Dr Anne Raimondo, a post-doctoral research fellow in my group in OCDEM, to include additional SgfI sites on both the 5' and 3' ends for ease of cloning into the pCMV6-Entry vectors, as shown in Figure 2.2.1.1. Oligonucleotides were synthesised salt-free by Eurofins Genomics and resuspended to 100 µM in RNase-free water. These oligonucleotides were first phosphorylated to facilitate subsequent ligation, then annealed together to form a double-stranded sequence for restriction digestion and ligation. To do this, T4 polynucleotide kinase (New England BioLabs) was used in an ATP-containing T4 DNA ligase buffer (New England BioLabs) to add phosphates to the 5' ends of the single-stranded DNA. Phosphorylation reactions took place at 37°C for 30 min and were terminated by incubation at 65°C for 10 min. The phosphorylated oligonucleotides were diluted 1 in 2 with Tris-EDTA (TE) buffer solution, then used in annealing reactions by combining 50 µl of each phosphorylated oligonucleotide with 1.25 µl of 4M sodium chloride and incubated at 100°C for 5 min. The reactions were then cooled slowly to room temperature to allow for hybridisation. Annealed V5 oligonucleotides were diluted before use in ligation reactions.

5' SgfI site	START	V5 Tag sequence	3' SgfI site
CGCC	ATG	GGTAAGCCTATCCCTAACCCTCTCCTCGGTCTCGATTCTACG	GCGAT
TAGCGG	TAC	CCATTCGGATAGGGATTGGGAGAGGAGCCAGAGCTAAGATGC	CGC

Figure 2.2.1.1. V5 tag sequence designed for N-terminal tag cloning. SgfI sites flanking the tag are highlighted in yellow; the start codon for expression of the tag is highlighted in green; the tag sequence is in red.

2.2.2. Cohesive-end ligation of purified vector and V5 tag insert

The hybridised oligonucleotides and the pCMV6-Entry-*G6PC1/G6PC2* vectors were both digested separately using the SgfI restriction enzyme and buffer C (Promega) according to the basic protocol outlined in section 2.1.1. The digested vector was further treated with Thermosensitive Alkaline Phosphatase (TSAP) (Promega) to remove phosphate groups from the 5' sticky ends to prevent self-ligation of the vector. The TSAP-treated vector was run on a gel and purified as in sections 2.1.2 and 2.1.3. Ligation reactions were set up with the purified digested vector (~5.9 kb) and the digested annealed oligonucleotides ('insert' of 54 bp) in a 10 µl reaction, using the T4 DNA ligase and ligase buffer (Promega). The insert was added in excess in the ligation reaction typically using a vector to insert molar ratio of 1 to 3. Samples were incubated overnight at 4°C and collected the next day for transformation into *E. coli* cells as described in the next section.

2.3. Transformation of competent *Escherichia coli*

Plasmid DNA was introduced into competent *E. coli* cells by transformation. Two to five microlitres of ligation reaction or up to 100 ng of plasmid DNA was added to 40 µl of DH5α competent cells (Life Technologies) in 1.5 ml tubes on ice. Reactions were mixed by gentle flicking followed by incubation on ice for 30 min. Following incubation, reactions were subjected to heat shock at 42°C for 1 min, then immediately placed on ice for 2 min. Half a millilitre of super optimal broth (SOC) media (Invitrogen) was added to each reaction and the reaction incubated for 60 min at 37°C in an orbital shaking incubator set at 225 rpm (Stuart Orbital Incubator S150, Jencons Scientific Ltd). For bacteria transformed with plasmid DNA, 50 µl of the reaction was spread evenly across LB agar plates with the relevant antibiotic – 25 µg/ml of kanamycin (Sigma Aldrich) or 50 µg/ml of ampicillin (Sigma Aldrich). For bacteria transformed with a ligation reaction, the reaction was spun down in a microcentrifuge, and the pellet was resuspended in 100 µl of SOC and plated. Plates were incubated for 18 hours at 37°C. Only cells that have taken up the ligated vector such as with Kan^R will grow on LB/Kan plates.

2.4. DNA extraction from transformed competent bacteria

2.4.1. DNA extraction by miniprep

Bacterial colonies were picked from LB agar culture plates and each single colony was used to inoculate 10 ml LB medium (Sigma Alrich) with antibiotic (25-50 µg/mL) and cultured for 16h (overnight) at 37°C in the orbital shaking incubator at 250 rpm (Stuart Orbital Incubator S150, Jencons Scientific Ltd). After incubation, 500 µl of bacterial sample was first collected and added to an equal volume of 37% DMSO (Sigma Aldrich) made up in distilled H_2O in a 1.5 ml tube. These DMSO stocks were stored at -80°C and used in future for further plasmid preparation. The remaining cultured bacteria was pelleted at 13,000 rpm for 10 min in a Force 1618 benchtop microcentrifuge (Labnet International). The supernatant was discarded and the DNA from the pellet extracted using the QIAprep Spin Miniprep kit (Qiagen) or PureYield Plasmid Miniprep system (Promega), according to manufacturers' instructions. The DNA was eluted in 40 µl of RNase-free water and stored at -20°C.

2.4.2. DNA extraction by midiprep

The DMSO stock for the plasmid DNA to be extracted was scraped with a plastic pipette tip and used to inoculate 10 ml LB medium with antibiotic (25-50 µg/ml) and cultured for at least 5 hours at 37°C in the incubator shaking at 250 rpm (Stuart Orbital Incubator S150, Jencons Scientific Ltd). This starter culture was then added to 100 ml LB medium with antibiotic in 500 ml flasks and placed in the shaking incubator (Labnet 311DS, Labnet International) for 16 hours (overnight) at 37°C. After incubation, the culture was transferred to 50 ml Falcon tubes (Corning) and pelleted at 3,500 rpm for 10 min (Heraeus Multifuge 3SR, Thermo Scientific). The supernatant was discarded and the DNA from the pellet extracted using the QIAprep Spin Midiprep kit (Qiagen) or PureYield Plasmid Midiprep system (Promega), according to manufacturers' instructions. The DNA was eluted in at least 450 µl of RNase-free water and stored at -20°C.

2.4.3. DNA quantification

DNA was quantified using the NanoDrop 2000 spectrophotometer (Thermo Scientific), which measures the concentration of DNA based on sample absorbance at a wavelength of 260nm. Sample purity was assessed by measuring the 260nm/280nm absorbance ratio which provided an indication of any contamination by protein or organic compounds. DNA deemed as 'pure' had a 260/280 ratio of ~1.8-1.9. Only DNA midiprep samples with concentrations above 100 ng/µl and that were of ideal purity were subsequently used for transfections.

2.4.4. Sanger sequencing

All sequences in the pCMV6-Entry vector were verified by Sanger sequencing using forward primer T7 (5'- TAATACGACTCACTATAGGG -3') and reverse primer XL39 (5'-ATTAGGACAAGGCTGGTGGG -3'). Samples were sent to Source BioScience (UK) or Eurofins Genomics (Germany). For plasmids mutated by SDM, only the desired nucleotide changes were introduced.

2.5. Generation of variant constructs

2.5.1. Site directed mutagenesis

Site directed mutagenesis (SDM) was carried out on the pCMV6-Entry-*G6PC1* and pCMV6-Entry-*G6PC2* plasmids (unless otherwise specified) to introduce single nucleotide substitutions into the *G6PC1* or *G6PC2* coding sequence, to enable the study of the genetic variants described in Chapters 3, 4 and 5. Primers were designed using the Quikchange Primer Design (Agilent Technologies) software, which is optimised to design primers that are ideally 25 to 45 bases in length, of melting temperature ≥78°C, with minimum GC content of 40%, that terminate in one or more C or G bases, and that contain base mismatches located close to the middle of the primer. Primer sequences are found in the relevant Chapters 3, 4 and 5 and were all purchased as salt-free oligonucleotides from Eurofins Genomics (Germany). SDM reactions were performed using the Quikchange II Site-Directed Mutagenesis (Agilent Technologies) kit and the protocol adapted from the manufacturer's guidelines. Each reaction was set up with 50 ng of template DNA and 50 ng/µl each of forward and reverse primers and performed on a tetrad thermal cycler (DNA Engine, MJ Research). See Table 2.5.1.1 for the composition of the reaction mix and 2.5.1.2 for the thermal cycler conditions for the polymerase chain reaction (PCR). Upon completion of template amplification, reactions were cooled to 4°C for at least 2 min. Samples were then subjected to *Dpn I*-mediated digestion for 1h at 37°C which removed any template DNA that had not been mutagenised.

Reagent	Volume (µl)
Reaction buffer (10X)	5
Double-stranded DNA template (10 ng/µl)	2.5
Forward primer (50 ng/µl)	5
Reverse primer (50 ng/µl)	5
dNTP mix	1
Distilled H_2O	31.5
PfuUltra High Fidelity DNA polymerase (2.5U/µl)	1 (add last)
Total	50

Table 2.5.1.1. SDM reaction mix for one reaction. Primers were purchased from Eurofins Genomics. All other reagents were obtained from the Quikchange II Site-Directed Mutagenesis kit (Agilent Technologies).

Segment	Cycles	Temperature (°C)	Time (min)
DNA denaturation	1	95	1
Primer annealing	16	95	0.5
		55	1
Extension		68	6[a]
Extra extension	1	68	5
Time to add *Dpn I*	1	4	5
Dpn I digestion of template	1	37	60
End	1	4	10

Table 2.5.1.2. SDM thermal cycler conditions. Reactions were carried out on a thermal cycler. [a]Cycling times were optimised for the pCMV6-Entry-*G6PC1/G6PC2* vector (~6 kb).

2.5.2. Transformation of competent *E. coli* and DNA extraction

Plasmid DNA was introduced into competent *E. coli* cells by transformation, in a protocol similar to that described in section 2.3. Two microlitres of SDM reaction was added to 40 µl of XL1-Blue supercompetent cells (Agilent Technologies). For the heat shock step reactions were heated to 42°C for 45 seconds then placed on ice for at least 2 min. After incubation in SOC media, 100 µl of the reaction was plated on LB/Kan plates for overnight incubation at 37°C. The next day, at least three isolated colonies were picked per plate for DNA extraction by miniprep and the DNA sent for sequencing. Plasmid samples with the desired mutations were selected and expanded by midiprep as described in Section 2.4.2. All midiprep samples generated were verified by sequencing as described in Section 2.4.4.

2.6. General maintenance of cell lines

The functional assays evaluating *G6PC2* and its genetic variants as described in Chapters 3, 4 and 6 were performed in HEK293, INS-1 (832/13) and EndoC-βH1 cells. HEK293 (Graham et al. 1977), a human embryonic kidney cell line, was chosen as it lacks endogenous *G6PC2* and is a protein expression model with high transfection efficiency. INS-1 (832/13) is a sub-clone of INS-1, a rat insulinoma cell line, with improved glucose-stimulation insulin secretory response (Hohmeier et al. 2000). It provided a potentially more physiologically-relevant cellular environment even though it lacks a *G6PC2* orthologue at both mRNA and protein level (Martin et al. 2001). EndoC-βH1 is an immortalised human beta cell line (Ravassard et al. 2011) that expresses endogenous *G6PC2* and is a biologically-relevant cell model for the study of human *G6PC2*. However, it has poorer transfection efficiency compared to other cell lines used. The functional assays evaluating *G6PC1* and its genetic variants as described in Chapter 4 were performed in HEK293, HepG2 and Huh7 cells. HepG2 (Aden et al. 1979) and Huh7 (Nakabayashi et al. 1982) cells are human cell models for hepatocellular carcinoma (liver cancer); all three cell lines express endogenous *G6PC1*.

All tissue culture activity was carried out in biological safety cabinets in the tissue culture facilities at OCDEM. All cell lines were tested negative for mycoplasma contamination using the MycoAlert Assay kit (Lonza). Cells were maintained at 37°C and 5% CO_2. For cell maintenance, briefly, cells were washed with 1X phosphate buffered saline (PBS) (Sigma

Aldrich). In a T75 flask, 1 ml of 1X trypsin-EDTA solution (Life Technologies) was added to each flask for 3-5 min at 37°C to lift cells off. To neutralise the trypsin solution, 4-9 ml of culture medium (specific to cell line) was added to the flask. Cells may be collected and spun down at 200 g for 5 min in a centrifuge (Beckman GS-15R, Beckman Coulter), then resuspended in 5-10 ml of fresh culture medium. Cells were counted using the Cellometer Auto T4 cell counter (Nexcelom Bioscience). Table 2.6.1 shows the specific culturing conditions and procedures for each cell line used.

2.7. Transient transfections

Cells were seeded 18-24 hours prior to transfection. For protein expression and enzymatic activity experiments, cells were transiently transfected using the Lipofectamine 2000 transfection reagent (Life Technologies) following manufacturer's guidelines. Prior to transfection, the culture medium over cells was replaced with transfection medium (culture medium without FCS and antibiotics). This transfection medium was removed and replaced with fresh complete medium 5-7 hours after transfection. For microscopy experiments, cells were transiently transfected using the FuGene 6 transfection reagent (Promega) following manufacturer's guidelines. The plating density, amount of DNA and ratio of DNA to transfection reagent were optimised for each cell line and vector construct. These are summarised in Table 2.7.1 (for *G6PC2* studies) and Table 2.7.2 (*G6PC1* studies).

Cell line	Culture medium information	Culturing procedures
HEK293	Dulbecco's Modified Eagle's Medium (DMEM) (D6429, Sigma Aldrich), 10% (v/v) foetal bovine serum (FBS) (10500-064, Gibco by Life Technologies), 100 U/ml penicillin and 100 µg/ml streptomycin (15140122, Gibco by Life Technologies)	Cells were split in a 1:20 ratio twice weekly.
INS-1 (832/13)	Roswell Park Memorial Institute-1640 (RPMI-1640) media (R0883, Sigma Aldrich), 10% (v/v) FBS, 100 U/ml penicillin and 100 µg/ml streptomycin, 2 mM L-glutamine (25030081, Gibco by Life Technologies), 1 mM sodium pyruvate (S8636, Sigma Aldrich), 10 mM HEPES (H3537, Sigma Aldrich), 50 µM 2-mercaptoethanol (Gibco by Life Technologies)	Cells were split once a week with 8 x 10^5 cells seeded in a T75. The culture medium was replaced with fresh medium every 3-4 days.
EndoC-βH1	DMEM (31885, Gibco by Life Technologies), Bovine Serum Albumin (BSA) fraction V (10775835001, Roche), 100 U/ml penicillin and 100 µg/ml streptomycin, 2 mM L-glutamine, 50 µM 2-mercaptoethanol, 10 mM nicotinamide (Sigma Aldrich), 5.5 µg/ml transferrin (Sigma Aldrich), 6.6 ng/ml, sodium selenite (Sigma Aldrich)	All surfaces were coated with coating medium consisting of high glucose DMEM (41965, Gibco by Life Technologies), 100 U/ml penicillin and 100 µg/ml streptomycin with 2 µg/m fibronectin (Sigma Aldrich) and 1% (v/v) Extracellular Matrix (ECM) (Sigma Aldrich), at least 2h before plating. Cells were split once a week with 1.2 x 10^6 cells seeded in a T25.
HepG2	DMEM (D6429, Sigma Aldrich), 10% (v/v) FBS, 100 U/ml penicillin, 100 µg/ml streptomycin	Cells were split 1:10 twice weekly. Re-suspended cells were filtered through a 40 µm cell strainer (Corning) before seeding.
Huh7	DMEM (31885, Gibco by Life Technologies), 10% (v/v) FBS, 100 U/ml penicillin and 100 µg/ml streptomycin	Cells were split twice weekly with 1.1 x 10^6 cells seeded in a T75. The culture medium was replaced with fresh medium every 2-3 days.

Table 2.6.1. Cell line culture medium and cell-specific culturing procedures. Identical reagents were sourced from the same companies which was specified the first time the reagent was mentioned, unless otherwise stated.

	Transfection with Lipofectamine 2000 (at 80-90% cell confluency)			Transfection with FuGene 6 (at 40-50% cell confluency)		
Cell line	Number of cells plated	DNA (µg)	DNA:Lipofectamine	Number of cells plated	DNA (µg)	DNA:FuGene
HEK293	6.8×10^5	2.5	1:4	3.5×10^5	3	1:4
INS-1 (832/13)	1.2×10^6	2.5	1:4	8×10^5	3	1:4
EndoC-βH1	NA	NA	NA	5×10^5	3	1:6

Table 2.7.1. Recommended transfection conditions per well for *G6PC2* in a 6-well plate format. Lipofectamine 2000 was purchased from Life Technologies and FuGene 6 was purchased from Promega. Reagent quantities may be proportionately adjusted for different plate formats.

	Transfection with Lipofectamine 2000 (at 80-90% cell confluency)			Transfection with FuGene 6 (at 40-50% cell confluency)		
Cell line	Number of cells plated	DNA (µg)	DNA:Lipofectamine	Number of cells plated	DNA (µg)	DNA:FuGene
HEK293	6.8×10^5	2.5	1:2	3.5×10^5	3	1:2
HepG2	6.2×10^5	2.5	1:5	3×10^5	3	1:5
Huh7	6.0×10^5	2.5	1:5	3×10^5	3	1:5

Table 2.7.2. Recommended transfection conditions per well for *G6PC1* in a 6-well plate format. Lipofectamine 2000 was purchased from Life Technologies and FuGene 6 was purchased from Promega. Reagent quantities may be proportionately adjusted for different plate formats.

2.8. Inhibition of protein degradation or modification

To study the role of cell proteolysis in the regulation *G6PC2* variant proteins (Chapters 3 and 5), transfected cells were treated with 10 µM MG-132 in DMSO (Calchembio) or 100 µM chloroquine (Sigma Aldrich) in distilled H_2O for 15h, to inhibit the proteosomal and lysosomal pathways respectively (Goldstein et al. 1975; Palombella et al. 1994). For inhibition of N-linked protein glycosylation, cells were treated with 1 µg/ml tunicamycin (Sigma Aldrich) in DMSO for 15h, unless otherwise stated. DMSO alone (Sigma Aldrich) in equivalent volumes was used as a vehicle control for untreated cells.

2.9. Protein expression studies

2.9.1. Cell lysis

Total protein lysates were collected from cells transfected with each wild type (WT) or mutant *G6PC1*/*G6PC2* construct 36-48h after transfection. The protocol was performed on ice to minimise protein degradation. For the collection of HEK293 and INS-1 (832/13) cells, cells were first washed gently in PBS, then dislodged by pipetting solution up and down within the wells. Resuspended cells were collected in cold PBS in pre-chilled 1.5 ml microcentrifuge tubes and spun down at 1,200 rpm for 10 minutes at 4°C (Allegra 64R centrifuge, Beckman Coulter). The supernatant was removed without dislodging the cell pellet. Pellets were snap frozen on dry ice before cell lysis as the freeze-thaw step aids cell lysis. For the collection of HepG2 and Huh7 cells, cells were washed in PBS then incubated in 500 µl trypsin-EDTA solution (per well of a 6-well plate) for 2-3 min at 37°C. One millilitre of culture medium was added to each well to neutralise the trypsin solution and resuspended cells were transferred into pre-chilled 1.5 ml microcentrifuge tubes. Cells were spun down at 1,200 rpm for 10 min at 4°C. Cell pellets were washed once in cold PBS then snap frozen on dry ice.

Cell pellets were thawed on ice and homogenised by the addition of ice-cold lysis buffer. Each sample obtained from one well of a 6-well plate was lysed in 40-100 µl of lysis buffer. The recipe of the lysis buffer is shown in Table 2.9.1.1. Cells were lysed for 20 min, with vortexing every 5 min. HEK293 and HepG2 lysates were further sonicated on ice for short

pulses of 10s each (Jencons Scientific Ltd). Finally, lysates were centrifuged at 13,000 rpm for 20 min at 4°C (Allegra 64R, Beckman Coulter). The supernatant (post-nuclear fraction) was transferred into pre-chilled 1.5 ml microcentrifuge tubes and snap frozen on dry ice before long-term storage at -80°C.

Reagent (stock concentration)	Final concentration
Tris-HCl (1 M)	25 mM
Sodium fluoride (1 M)	50 mM
Sodium chloride (5 M)	100 mM
EGTA (0.25 M)	5 mM
EDTA (0.5 M)	1 mM
Sodium pyrophosphate decahydrate (0.1 M)	2 mM
Triton-X 100	1 % (v/v)
Sucrose	0.27 M
Sodium orthovanadate (0.5 M)	1 mM
β-mercaptoethanol	0.1 % (v/v)
Complete Mini EDTA-free Protease Inhibitor Cocktail (Roche)	1 tablet/10 ml buffer

Table 2.9.1.1. Cell lysis buffer composition for western blot analysis. The first seven reagents were made up in distilled H_2O as a 5X stock and stored at 4°C up to three months. A 1X lysis buffer was made up with the remaining reagents and stored at 4°C up to a week. The protease inhibitor was added fresh just before cell lysis. All reagents were obtained from Sigma Aldrich unless otherwise stated.

2.9.2. Protein quantification

Cell lysates were quantified using the Bradford assay reagent (Bio-Rad Laboratories). The assay detects changes in absorbance of the Coomassie Brilliant Blue dye. The blue form of the dye is stabilised when bound to protein and can be measured at 595nm, providing a measure of protein concentration. Two microlitres of each sample were aliquoted into glass assay tubes (Fisher Scientific) containing 38 µl 0.1 M NaOH (Sigma Aldrich). One millilitre of 1X Bradford reagent made up in distilled H_2O was added to each tube and mixed by brief vortexing and was incubated for 10 min. Each sample was measured in duplicate on a 96-well clear plate. Absorbance was read at 595nm on the Versamax microplate reader using the Softmax Pro v5.4 software (both Molecular Devices Ltd). Protein concentrations were calculated based on a standard curve made with bovine serum albumin (BSA, Sigma Aldrich).

2.9.3. Western blot analysis (Bio-Rad Criterion system)

Protein lysates were separated by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) on a Bio-Rad Criterion TGX stain-free 4-12% pre-cast gel (1.0 mm, 18-well). The 1X Tris/Glycine/SDS (TGS) gel running buffer consists of 25 mM Tris, 192 mM glycine and 0.1% SDS pH 8.3 (Bio-Rad Laboratories). Loading samples were prepared by adding 15 µl protein to 5 µl 4X Laemmli loading buffer (Bio-Rad Laboratories) and 1 µl β-mercaptoethanol (Sigma Aldrich). Samples were subject to mild heating at 37°C for a maximum of 20 min which improved the detection of membranous proteins. Denatured samples were spun down at 13,000 rpm for 1 min on a benchtop centrifuge (Force Micro 1618, Labnet International). A constant amount of protein (10 µg up to 80 µg depending on cell line) was loaded in each lane for every gel run, together with a molecular weight marker (Precision Plus All Blue standard, Bio-Rad Laboratories). Gels were run at 200V for 1h. Subsequently, the gel was activated and imaged using the ChemiDoc MP Imaging System and ImageLab 5 software (Bio-Rad Laboratories). Proteins were transferred to a 0.2 µm polyvinylidene difluoride (PVDF) membrane using the Trans-Blot Turbo transfer system (Bio-Rad Laboratories). Following transfer, the membrane was imaged to confirm protein transfer.

Membranes were washed briefly in wash buffer made up of 1 mM Tris-HCl pH 7.4, 5 M NaCl (both Sigma Aldrich), 0.05% Tween-20 (Fisher Scientific) (TBST), then blocked with 5% non-fat milk powder in TBST for 1h before primary antibody incubation. Primary antibodies used for determining recombinant G6PC1/G6PC2 protein expression and recommended dilutions are shown in Table 2.9.3.1. Dilutions were made in 5% milk in TBST and added to membranes for incubation overnight at 4°C with gentle shaking (Stuart see-saw rocker, Bibby Scientific Ltd). The following day, membranes were washed 3 times for 10 min each in TBST with agitation, followed by secondary antibody incubation for 1h with gentle shaking at room temperature. The secondary antibodies used and dilutions are shown in Table 2.9.3.2. After incubation membranes were washed 5 times for 5 min each in TBST with agitation. Membranes were then incubated in western enhanced chemiluminescence (ECL) reagent (Bio-Rad) for 5 min and visualised on the ChemiDoc imager (Bio-Rad Laboratories). For detection of a loading control such as tubulin, membranes were first stripped with a

western blot stripping buffer (Thermo Scientific) for 10 min with gentle shaking, then washed thoroughly in TBST before primary antibody incubation.

2.9.4. Western blot analysis (Invitrogen NuPage system)

All INS-1 (832/13) protein lysates were analysed using the Invitrogen NuPage SureLock Mini-Cell electrophoresis system (Life Technologies) as this allowed for more complete transfer of protein from gel to membrane via a wet transfer protocol, improving detection of low abundance proteins. Protein lysates were separated by SDS-PAGE on a NuPage 4-12% Bis-Tris pre-cast gel (1.0mm, 12-well). The gel running buffer was prepared from the NuPage MOPS SDS 20X stock buffer (Life Technologies) and consisted of 50 mM MOPS, 50 mM Tris Base, 0.1% SDS, 1 mM EDTA, pH 7.7 following dilution to 1X with distilled H_2O. NuPage anti-oxidant (Life Technologies) was added to the running buffer in the inner chamber as recommended in the manufacturer's guidelines. Loading samples were prepared as described in section 2.9.3 but using 4X Laemmli loading buffer from Life Technologies. A molecular weight marker was used with every run (SeeBluePlus2 prestained standard, Life Technologies). Gels were run at 150V for 15 min followed by 180V for 45 min. Proteins were transferred from the gel to a 0.45 µm PVDF membrane (Amersham, GE Healthcare) by wet transfer using the XCell II Blot Module (Life Technologies) at 35V for at least 150 min. The transfer buffer was prepared from the NuPage 20X transfer buffer (Life Technologies) with additional 20% methanol (v/v) (Sigma Aldrich). Prior to transfer, membranes were first activated in pure methanol for 10s then equilibrated in transfer buffer for at least 15 min. Successful transfer may be visualised by staining the membrane with Ponceau S stain (Thermo Scientific). The steps of washing and antibody incubation following transfer followed that outlined in section 2.9.3.

Primary antibodies for detecting G6PC2 were made up in 3% milk in TBST to improve protein detection. Protein bands were exposed on CL-Xposure film (Thermo Scientific) and developed on a film developer (SRX-101A, Konica Corporation) with appropriate X-ray developer and fixing solution (Tetenal).

2.9.5. Densitometric analysis by ImageJ

Western blots that were exposed on films were scanned at 600 dpi to obtain digital TIFF images (Epson). Protein bands were quantified by densitometry analysis using ImageJ. Values for G6PC1/G6PC2 bands were normalised to values for loading controls (tubulin or calnexin) on the same blot. Normalised values were expressed as densitometry units relative to WT.

Antibody	Recommended dilutions	Source
Anti-FLAG M2 (mouse monoclonal)	1:2500	F1804, Sigma Aldrich
Anti-V5 (mouse monoclonal)	1:3000	46-0705, Invitrogen
Anti-myc 4A6 (mouse monoclonal)	1:600	05-724, Millipore
Anti-β-tubulin (rabbit polyclonal)	1:2500	sc-9104, Santa Cruz
Anti-calnexin (rabbit polyclonal)	1:3000	sc-11397, Santa Cruz; ab10286, Abcam

Table 2.9.3.1. Primary antibodies used for western blot. All antibodies were stored at 4°C except for anti-FLAG which was stored at -20°C. Recommended dilutions were optimised for the *G6PC2* studies.

Antibody	Recommended dilutions	Source
Anti-mouse IgG (H+L) HRP (rabbit)	1:2500	31450, Thermo Scientific
Anti-rabbit IgG (H+L) HRP (goat)	1:5000	31460, Thermo Scientific
Anti-sheep IgG (H+L) HRP (rabbit)	1:2500	31480, Thermo Scientific

Table 2.9.3.2. Secondary antibodies used for western blot. All antibodies were made up in distilled H_2O according to manufacturer's instructions and stored at -20°C.

2.10. Cellular localisation studies

2.10.1. Immunofluorescence staining

Immunostaining was used to determine the subcellular localisation of WT or mutant G6PC1/G6PC2 variants expressed in cells. Cells were transfected in 4-chamber culture slides (Corning) and fixed with 4% paraformaldehyde (Sigma Aldrich) in PBS for 15 min, then permeabilised with PBS with 0.1% Triton-X (Sigma Aldrich) for 15 min. Cells were blocked in 10% BSA made up in PBS with 0.05% Tween-20 (PBST) for 1h. Cells were incubated in primary antibodies diluted in 10% BSA/PBST overnight at 4°C. The primary antibodies used are indicated in Table 2.10.1.1. Each well of a chamber slide required a minimum of 200 µl of incubation solution. The following day, cells were washed five times for 5 min each with PBST. Cells were then incubated in secondary antibody for 1h. Secondary antibodies used are indicated in Table 2.10.1.2. From this step on all procedures were carried out in the dark. Cells were washed in PBST for three times for 5 min each. To stain the nucleus, DRAQ5 (Life Technologies), a far-red cell-permeant dye, was made up in PBST and applied at 20 µm for 15 min at room temperature unless otherwise stated. Slides were mounted in Vectashield mounting medium (Vector Labs) or ProLong Gold Antifade Reagent (Life Technologies), covered with a glass coverslip (VWR), and allowed to cure in the dark at room temperature for 24h. Finally, slides were sealed using commercially available nail varnish and kept at 4°C for short-term storage or -20°C for long-term storage.

2.10.2. Confocal microscopy

Slides were visualised on a Bio-Rad Radiance 2100 confocal microscope with a 60X 1.0 N.A. water immersion objective, unless otherwise stated. Images for the three colour channels (green, red, far-red) were acquired consecutively using the LaserSharp 2000 software. Images were collected with different laser settings that were optimized for each sample and channel and therefore fluorescent intensities are not comparable across samples. Image files were exported with LSM Image Browser 4.2 (Zeiss) and arranged with Adobe Photoshop CS3 (Adobe Systems).

Antibody	Recommended dilutions	Source
Anti-FLAG M2 (mouse monoclonal)	1:2500	F1804, Sigma Aldrich
Anti-V5 (mouse monoclonal)	1:3000	46-0705, Invitrogen
Anti-calnexin (rabbit polyclonal)	1:100	sc-11397, Santa Cruz; ab10286, Abcam
Anti-calreticulin (rabbit polyclonal)	1:100	PA3-900, Thermo Scientific
Anti-TGN46 (rabbit polyclonal)	1:200	T7576, Sigma Aldrich

Table 2.10.1.1. Primary antibodies used for immunofluorescence staining. All antibodies were stored at -20°C except for anti-V5 and anti-calnexin which were stored at 4°C.

Antibody	Recommended dilutions	Source
Anti-mouse fluorescein (horse)	1:100	FI-2000, Vector Labs
Anti-rabbit TRITC (swine)	1:50	R0156, Dako
Anti-mouse Alexa Fluor 488 (donkey)	1:500	A21202, Life Technologies
Anti-rabbit Alexa Fluor 568 (goat)	1:500	A11011, Life Technologies
Anti-goat Alexa Fluor 647 (chicken)	1:500	A21469, Life Technologies

Table 2.10.1.2. Secondary antibodies used for immunofluorescence staining. All antibodies were stored at 4°C in the dark.

2.11. Enzymatic activity studies

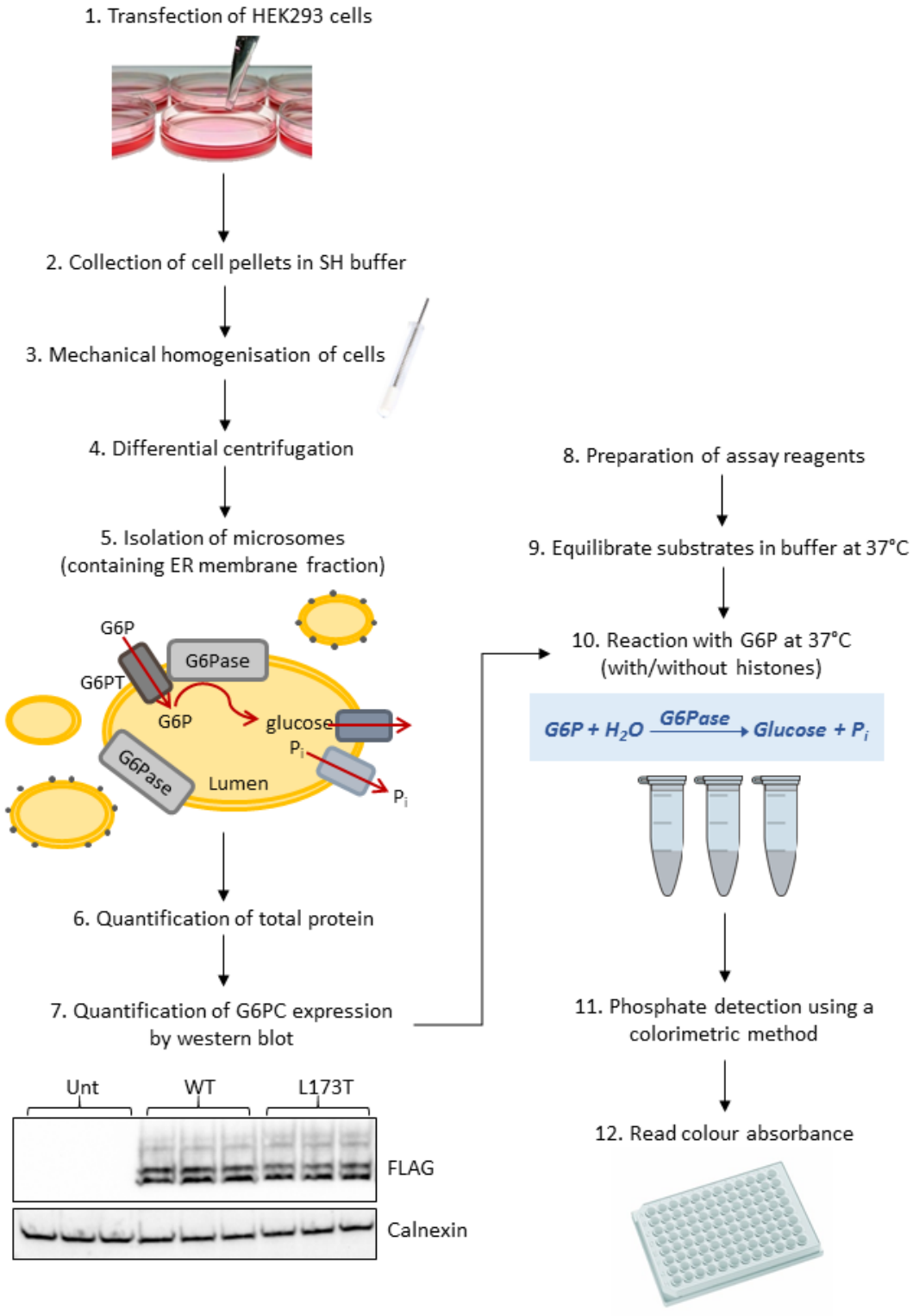


Figure 2.11.1. Summary of workflow for the investigation of glucose-6-phosphatase activity in microsomes of transfected HEK293 cells. Details of each step are outlined in the subsequent sections.

2.11.1. Collection of microsomes

HEK293 cells were transfected in 10 cm dishes for the collection of microsomes. At least two dishes of cells per condition were required to obtain sufficient microsomal protein. An untransfected control sample was collected for every experiment. All steps were carried out on ice and all reagents pre-chilled to minimise protein degradation. Cells were washed once gently in cold 0.25 M sucrose in 5 mM HEPES buffer (SH) then harvested with a cell scraper (Sarstedt) into fresh SH and transferred to 15 ml tubes. Cells were spun down at 1,800 rpm for 5 min at 4°C in a centrifuge (Heraeus Multifuge 3SR, Thermo Scientific). Cell pellets were resuspended in SH and pelleted again in the centrifuge. At least two of such washes in SH were carried out. Cells were then mechanically homogenised using a 1 ml Potter-Elvehjem glass tissue grinder and Teflon pestle (Anachem), followed by 12 passes through a 27-gauge needle (Sterican®, B.Braun Medical Ltd) and syringe (Codan). The homogenate was then transferred to 1.5ml microcentrifuge tubes and subjected to centrifugation at 10,000 g for 10 min (Allegra 64R, Beckman Coulter). The supernatant (post-nuclear fraction) was transferred to 1.5 ml polyallomer ultracentrifuge tubes (Beckman Coulter) and then centrifuged at 100,000 g for 1h with a TLA 100.4 rotor in a table top Optima TLX ultracentrifuge (CTX94L24, Beckman Coulter). A rigid pellet containing the microsomal fraction was obtained and resuspended thoroughly in approximately 120 µl SH. Microsomes were divided into single-use aliquots of 30 µl each in 0.5 ml tubes and snap frozen on dry ice for storage at -80°C.

2.11.2. Microsomal protein quantification

An aliquot of each microsomal sample was diluted in lysis buffer (for buffer composition see 2.9.1) and lysed for at least 20 min with vortexing every 5 min. Lysed protein was immediately used for quantification using the Bradford reagent (see 2.9.2). Western blot analysis of the crudely lysed microsomal protein was carried out using the Bio-Rad Criterion gel system (see 2.9.3); 10 µg of protein loaded per lane was sufficient for protein detection. The primary antibodies used to detect recombinant G6PC1 were anti-FLAG M2 or anti-V5 (see Table 2.9.3.1). An antibody against calnexin was used as the loading control as it is an ER-resident protein and is present in the microsomal fraction. Each sample was run in 2-3

lanes as technical replicates, quantified by densitometry analysis and mean values were taken to determine the relative levels of expression of recombinant G6PC1 WT and variant proteins. These values were required for calculating the input into enzymatic assays in order to match for the amount of G6PC1 protein used in the assay.

2.11.3. Glucose-6-phosphatase assay

For use in the enzymatic assay, microsomal samples were first diluted to the same protein concentration with cold SH on ice. Equal volumes of each sample was then aliquoted and further diluted with SH such that matched units of expressed G6Pase WT or variant protein (as determined from densitometric analysis) may be added to each reaction tube. The protocol for the glucose-6-phosphatase assay was modified from the enzymatic assay for Glucose-6-Phosphatase from Rabbit Liver from Sigma Aldrich (G5758) and the original adaptation was carried out with the assistance of Dr Jana Rundle, a previous post-doctoral research fellow in my group at OCDEM. The reaction components are listed in Table 2.11.3.1. Briefly, microsomal samples were incubated for 8 min at 37°C in a reaction mix containing ~100 mM MES pH 6.5 (Sigma Aldrich), and 0 mM to 20 mM D-glucose-6-phosphate (G6P) (Sigma Aldrich), in 1.5 ml tubes in a 37°C water bath (Grant Instruments Ltd). Approximately 1-5 µg of microsomal protein was added to each reaction. For reactions in the presence of histones, 10 mg/ml of histones from calf thymus (Sigma Aldrich) made up in distilled H_2O were added to the assay mix as stated in Table 2.11.3.1. The hydrolytic reaction was terminated by addition of 20% (v/v) trichloroacetic acid (TCA) (Sigma Aldrich). Negative control reactions were carried out in parallel whereby microsomal protein was only added after termination of the reaction by TCA. Thereafter, samples were centrifuged at 4,000 rpm for 10 min in a microcentrifuge (Force Micro 1618, Labnet International). The supernatant was then mixed in equal parts with a Taussky-Shorr colour reagent consisting of 5% (w/v) iron(II) sulphate heptahydrate and 1% (w/v) ammonium molybdate dissolved in 0.5 M H_2SO_4 (all Sigma Aldrich), in a 96-well clear optical plate (Corning) and incubated for 7.5 min for colour development to take place. Sample absorbance was measured at 660 nm on the Versamax microplate reader using the Softmax Pro v5.4 software (both Molecular Devices Ltd). The amount of phosphates detected in nmol/mg/min was calculated using a standard curve based on a phosphorus standard solution (Sigma Aldrich).

Results were expressed as normalised activity relative to the activity of WT at maximum substrate (20 mM G6P) for each experiment. Mean normalised activity rates were used for Michaelis-Menten enzyme kinetic analysis carried out on GraphPad Prism 6.0 for the calculation of V_{max} (maximum velocity or rate of reaction) and Michaelis constant K_m (substrate concentration required to achieve half of V_{max}).

Reagent	**Volume (µl)**			
	Without Histones		**With Histones**	
	+ Enzyme	**- Enzyme**	**+ Enzyme**	**- Enzyme**
100 mM MES, pH 6.5, 37°C	150	150	148	148
0 – 82 mM G6P (for final 0 – 20 mM in reaction)	50	50	50	50
Histones (10 mg/ml)	-	-	2	2
Enzyme (microsomal sample)	10	10 (after TCA)	10	10 (after TCA)
Total	210	210	210	210

Table 2.11.3.1. Glucose-6-phosphatase assay mix per reaction. Reactions were carried out in 1.5 ml clear microcentrifuge tubes. All reactions were terminated by addition of 45 µl 20% trichloroacetic acid (TCA). G6P, glucose-6-phosphate. All reagents were obtained from Sigma Aldrich.

2.12. Gene expression analysis

2.12.1. RNA extraction

RNA was extracted from cells using the TRIzol® reagent (Life Technologies) according to manufacturer's guidelines. After phase separation and isolation, RNA precipitation from the aqueous layer with isopropanol was carried out with overnight incubation at 4°C. Glycogen (R0551, Life Technologies) was also added during this incubation step to increase the efficiency of recovering RNA from small amounts of starting material. RNA pellets were resuspended in RNase-free H_2O (Ambion, Thermo Fisher Scientific) and quantified using a NanoDrop 8000 spectrophotometer (Thermo Scientific). RNA samples were stored at -20°C.

2.12.2. cDNA synthesis

cDNA was synthesised using the SuperScript III First-Strand Synthesis kit (18080-051, Life Technologies) according to manufacturer's guidelines, from RNA samples obtained from cells or tissue samples. The source of RNA samples is stated in the relevant experimental chapters. Volumes of enzymes used were downscaled due to low starting amounts of RNA. Reactions were carried out in 96-well PCR plates (4titude Ltd) in the tetrad thermal cycler (DNA Engine, MJ Research). Fifty nanograms to 1 µg of input RNA was used and primed with oligo(dT) or a mixture of oligo(dT) and random hexamers where stated. cDNA synthesis reactions were stored at 4°C for short-term or at -20°C for long-term storage.

2.12.3. Real time TaqMan® quantitative PCR

Quantitative PCR was performed on the Applied Biosystems 7900HT Fast Real-Time PCR system on the SDS v2.3 software using TaqMan® gene expression mastermix (Applied Biosystems) and commercially-available TaqMan® gene expression assays (Applied Biosystems). Details of the assays used are provided in the relevant experimental chapters. All samples were analysed in triplicate wells in 384-well plates (4titude Ltd) for the gene(s) of interest and at least one housekeeping gene. Where multiplex reactions were carried out, a FAM- and a VIC-labelled assay were combined. Non-template control (NTC) wells

containing H_2O were also included on each plate. All amplification curves were visually checked and threshold lines manually adjusted where necessary. Analysis was performed using the comparative Ct ($2^{-\Delta\Delta Ct}$) method on Microsoft Excel and graphs were plotted using GraphPad Prism 7.0.

2.13. *In silico* mutation analysis tools

In silico prediction tools were used to assess the impact of non-synonymous site changes in *G6PC2* and *G6PC1* identified in Chapters 3 to 5 on protein function. These tools were accessed online: Sorting Intolerant From Tolerant (SIFT) [http://sift.jcvi.org/] (Kumar et al. 2009), Polymorphism Phenotype 2 (PolyPhen2) [http://genetics.bwh.harvard.edu/pph2/] (Adzhubei et al. 2010), CONsensus DELeteriousness score of non-synonymous SNVs (Condel 2.0) [http://bg.upf.edu/fannsdb/] (Gonzalez-Perez and Lopez-Bigas 2011) and Genomic Evolutionary Rate Profiling (GERP) (Cooper et al. 2005).

PolyPhen2 makes predictions of the impact of non-synonymous SNPs on protein structure and function by assessing the sequence-based and 3D structural and functional features that may be affected due to the amino acid substitution. SIFT, similarly predicts functional significance of the allelic variant by evaluating the degree of conservation of amino acid residues based on sequence comparisons of closely related proteins. Condel integrates the results of multiple computational tools (including PolyPhen2, SIFT, MutationAssessor, FatHMM) and calculates an average weighted score based on the different methods to improve the performance of determining whether a missense variant is deleterious. Finally, GERP estimates the evolutionary rate of a given nucleotide position based on genomic sequence alignments across multiple mammalian species, as a result of the action of past purifying selection. GERP scores range from -12.3 to 6.17; higher GERP scores indicate greater conservation, therefore mutations at sites with relatively higher scores are expected to have a greater impact on protein function. These tools provide a preliminary understanding of the impact that a coding variant may have on protein function. The dates that these tools were accessed and details of other tools or databases used are stated in the relevant experimental chapters.

2.14. Gene knockdown by siRNA transfection

Gene silencing was carried out on EndoC-βH1 cells using small interfering RNA (siRNA) and lipid-based reverse transfection. siRNA complexes were prepared by combining ON-TARGETplus siRNA (Dharmacon, GE Healthcare) and Lipofectamine RNAiMAX (Life Technologies) in Opti-MEM reduced serum medium (31985, Gibco, Life Technologies) and incubated for 15-20 min. Transfection complexes were aliquoted into wells in 24- or 96-well cell culture plates (Corning) before resuspended cells in culture medium were added gently to the wells. The final concentration of siRNA was 25 nmol/L with 0.2 µl of RNAiMAX in a total final volume of 240 µl per well of a 96-well plate, unless otherwise indicated. After 24h of transfection, the medium over cells were replaced with 100 µl fresh culture medium. Where cells are transfected in other plate formats, reagents were proportionately adjusted. Details of the siRNAs used are provided in the relevant results Chapter 6.

2.15. Cell proliferation assay

Cells were plated and transfected as described in 2.14 but in clear-bottom black 96-well plates (Corning) to facilitate fluorescence detection. Viable cell count was measured using the CyQUANT Direct Cell Proliferation Assay kit (C35012, Thermo Scientific) following manufacturer's guidelines. Components in the kit were diluted in cell culture medium before use and cells were incubated in the detection reagent for 1h at 37°C in the dark. Sample fluorescence was measured at 485nm/530nm (excitation/emission) using the CytoFluor multi-well plate reader (PerSeptive Biosystems) with a bottom-read function for adherent cells. All cell count values were expressed as fluorescent units normalised to mean cell count at low (1 mmol/L) glucose levels.

2.16. Insulin secretion analysis

2.16.1. Glucose stimulation

For secretion experiments in EndoC-βH1 cells, glucose-free complete culture medium was made up which consisted of DMEM (11966, Gibco by Life Technologies) supplemented as stated in section 2.6 and additionally with glucose (Sigma Aldrich) up to the required amounts. Cells were placed in low glucose (2.8 mM) culture medium overnight, 72h or 96h post-transfection. The following day, cells were further starved in 0 mM glucose medium for 1h. To initiate stimulations, the starvation medium was replaced with pre-warmed fresh medium containing: 1 mM glucose, 6 mM glucose, 20 mM glucose, 20 mM glucose with 100 µM tolbutamide (Sigma Aldrich) or 20 mM glucose with 100 µM diazoxide (Sigma Aldrich). Each condition was carried out in triplicate or quadruplicate wells within each experiment. After 1h of static incubation at 37°C, three-quarters of the supernatants were carefully removed from the wells into 96-well plates and spun down at 800 g for 5 min at 4°C (Allegra 64R centrifuge, Beckman Coulter). A portion of the supernatant was further aliquoted and stored at -20°C until analysis. The remaining medium on the cells were discarded. If cells were to be assayed for cell proliferation as described in section 2.15, cells may be washed once in glucose-free culture medium first. Otherwise, cells were immediately extracted for the analysis of insulin content by addition of ice-cold HCl-ethanol (Sigma Aldrich) to the wells. These samples were stored at -20°C until analysis.

2.16.2. AlphaLISA assay

The concentration of insulin in supernatants and contents were measured using the human insulin AlphaLISA detection kit (AL204C, Perkin Elmer) according to manufacturer's guidelines. The immunoassay detects both proinsulin and insulin. Insulin supernatant samples were diluted 1:10 and insulin content samples were diluted 1:100 in 1X immunoassay buffer. Diluted samples were aliquoted into white 96-well ½-area microplates (Perkin Elmer) for the assay. Plates were read on the Enspire multimode plate reader (Perkin Elmer) using the AlphaScreen protocol. Insulin concentrations in the samples were calculated based on a 10-point standard curve generated using the insulin analyte (AL204S, Perkin Elmer) in every run.

2.17. Statistical analyses

In Chapters 3, 4 and 5, mean differences in protein expression levels between G6PC2/G6PC1 WT protein and each variant protein from 3-5 independent experiments were compared using two-tailed paired Students' t tests. In Chapters 4 and 5, mean differences in activity between G6Pase WT protein and each variant protein for the substrate G6P were compared using two-tailed unpaired Students' t tests of the determined kinetic constants V_{max} and K_m.

In Chapter 6, mean differences in gene expression between G6PC2 WT- and each set of variant-transfected cells were analysed using two-tailed unpaired Students' t tests. The same analysis was carried out for comparing gene expression data in *G6PC2* KO cells and control cells. For the analysis of insulin secretion data, mean differences between *G6PC2* KO cells and control cells for each condition or time point were compared using two-tailed unpaired Students' t tests. For the analysis of ER stress luciferase activity data, a two-way analysis of variance (ANOVA) was applied to compare mean fold difference between G6PC2 WT and each variant for every reporter. The two-way ANOVA enabled the use of a single test to compare the activity of multiple response elements in the presence of different G6PC2 constructs expressed in cells. Plotting of graphs and statistical analyses were carried out on GraphPad Prism 6.0 or 7.0. A P value <0.05 was considered significant.

Chapter 3

Coding variants in *G6PC2* influence fasting glucose levels

The reference to the publication that the results in this chapter contribute to is stated below. The complete manuscript is provided in the Appendices.

Mahajan A*, Sim X*, Ng HJ*, Manning A, Rivas MA, *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(1)*: e1004876. doi: 10.1371/journal.pgen.1004876

*joint first authors

3.1. Introduction

3.1.1. Challenges in ascertaining effector transcripts associated with glycemic traits

To date, over 80 genetic loci comprising an even larger number of variants have been reported to be linked to one or more glycemic traits (Mohlke and Boehnke 2015). In fact, with advances in sequencing and the amalgation of increasingly large datasets, this number will only continue to rise. As most of the variants that were identified during the GWAS era were located within non-coding sequence, as is typical for most complex traits, attempts to establish functional significance and assign causality to the variants underlying the association signals are less straightforward.

Such a challenge may be exemplified at the glycemic trait locus *G6PC2/ABCB11*, where multiple non-coding variants within and between the *G6PC2* and *ABCB11* genes have been associated with FG levels (Figure 3.1.1.1). These common variants, including lead GWAS SNP rs560887 which maps to intron 3 of *G6PC2*, and variants that are in high LD with rs560887, may be putative causal variants contributing to the association with FG. In addition, *G6PC2* and *ABCB11* are both plausible candidates acting through different tissues as described in chapter section 1.4, making it difficult to ascertain the effector transcript through which associated variants exert their impact, hindering efforts to design appropriate experiments to understand the biological mechanisms driving these associations.

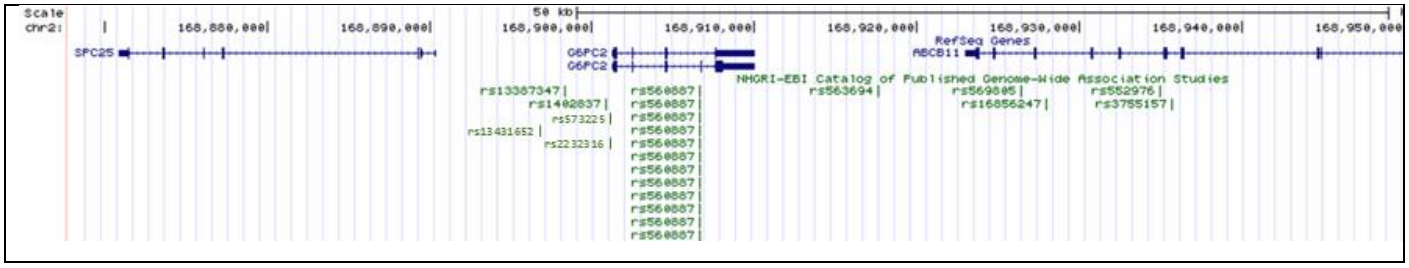


Figure 3.1.1.1. The *G6PC2/ABCB11* GWAS locus displaying common SNPs identified for glycemic trait associations. Image modified from the UCSC Genome Browser showing genomic regions based on the GRCh38/hg38 Assembly.

Recent molecular studies revealed that the glucose-lowering allele at GWAS SNP rs560887 led to a reduction in *G6PC2* expression by affecting pre-mRNA splicing (Baerenwald et al. 2013). The glucose-raising alleles at rs13431652 and rs2232316 in the promoter region of *G6PC2* were also found to affect promoter activity and hence upregulate *G6PC2* transcription (Baerenwald et al. 2013; Bouatia-Naji et al. 2010). These variants were suggested as functional causal variants contributing to the association signal at this locus. The studies also provided a direction of effect, revealing that reduced *G6PC2* function lowers FG levels. However, this hypothesis may be more precisely tested with the discovery of coding variant signals that may have direct impact on protein function.

To tackle some of these challenges, large collaborative efforts such as that of the T2D-GENES and GoT2D consortia used their extensive exome array genotyping dataset to identify genetic variants within the coding region that influence quantitative glycemic traits such as FG and FI levels. Coding variant signals are likely to display larger effect sizes but tend to be rarer. As known common variant signals from GWAS only account for a very small proportion of glycemic trait variance in the non-diabetic population – 4.8% for FG and 1.2% for FI (Scott et al. 2012) – variants within the rare and low frequency space may contribute substantially to the currently unexplained genetic variance. This chapter will introduce recent efforts in discovering the role of low-frequency (MAF=0.5-5%) and rare (MAF<0.5%) coding variants in influencing glycemic traits, in particular focusing on measures of FG and FI, and highlighting results from the study published in Mahajan et al. 2015 which formed the background of the work presented here.

3.1.2. Novel coding variants influence glycemic traits in exome array-based analysis

To identify novel coding variants influencing FG and FI levels, researchers from the T2D-GENES and GoT2D consortia carried out exome array genotyping in up to 33,407 normoglycemic European individuals across 14 studies (Mahajan et al. 2015). The custom-designed exome array covered mostly coding variants of lower allele frequency (MAF<5%) that tend not to be well-captured by traditional GWAS genotyping and imputation. Based on this analysis the authors identified a number of novel coding variants associated with fasting traits, including those that mapped to loci not previously associated with glycemic traits, and more that mapped to known GWAS loci, helping to highlight likely effector transcripts at these loci. These results are presented in Tables 3.1.2.1 and 3.1.2.2.

In the FG analysis, the authors identified a total of 12 novel coding variants (ten common and two low-frequency), all non-synonymous, across seven loci that were significantly associated with the trait (exome-wide significance threshold $P < 5 \times 10^{-7}$). This included one common variant at a novel locus not previously linked to FG, that is rs10305492 (p.A316T) at *GLP1R* (MAF = 1.5%, $P = 4.6 \times 10^{-7}$) that was associated with reduced FG. The same variant was identified in another exome array analysis with partially overlapping samples which showed that the FG-lowering allele was also associated with a modest decrease in T2D risk (Wessel et al. 2015). *GLP1R* encodes the glucagon-like peptide-1 receptor (GLP-1R), the receptor for glucagon-like peptide-1 (GLP-1), an incretin hormone released from the intestinal cells after food intake to stimulate insulin secretion and inhibit glucagon release. GLP-1R agonists are a class of drugs used to treat T2D by enhancing the incretin effect. The multiple *in silico* mutation analyses were not consistent in their predictions of the functional impacts of the threonine substitution at A316T. Though the evidence for the role of GLP-1R in post-prandial glycemic control is strong, the precise functional impact of A316T is yet to be investigated. A variant at the same site, A316G, was previously shown in *in vitro* expression assays to have similar pharmacological properties as wild type (WT) protein in response to ligands GLP-1 and exendin-4 (also a GLP-1R agonist) induction (Fortin et al. 2010).

All other loci except for *GLP1R* have been previously linked to FG regulation. Among the other variants identified, six mapped to known FG-associated loci but have not been identified before as causal variants at their respective loci. Two variants each mapped to

G6PC2 and *PCSK1*; one variant each mapped to *ZHX3* and *RREB1*. In particular, using conditional analysis the two coding variants in *G6PC2* were determined to be signals independent of each other and of the non-coding lead GWAS SNP rs560887, which was also included on the array. These findings, summarised in Table 3.1.2.1 for *G6PC2* and Table 3.1.2.2 for other variant signals, served as further evidence for pinpointing effector transcripts at GWAS loci that implicated multiple transcripts.

As for the analysis of FI, six non-synonymous variants (one rare, five common) mapping to four loci were associated at exome-wide significance. The only novel FI-associated locus was represented by a rare variant in *URB2* (E594V, MAF = 0.1%, P = 3.1×10^{-7}). All *in silico* algorithms used in the study predicted this variant to be neutral. *URB2* encodes ribosome biogenesis 2 homolog, which is known to interact closely with nuclear lamins (Vlcek et al. 2004). Conditions caused by defects in lamin genes include familial partial lipodystrophy, which can cause loss of adipose tissue, insulin resistance, and metabolic syndrome (Hegele et al. 2000; Mesa et al. 2007). However, the precise function of *URB2* itself remains to be understood and more studies will be required to investigate the role of *URB2* and the significance of the E594V variant.

This study has discovered additional coding variants contributing to glycemic trait variance and highlighted likely causal genes at multiple loci, including two novel loci, one significantly associated with FG and the other with FI. Nonetheless, further in-depth studies are needed to support the candidacy of the causal genes or variants at these loci. The next sections focus on the detailed analysis carried out on coding variants in *G6PC2*. My role in this study was to plan and execute *in vitro* follow-up studies of these coding variants, which is presented in the Results section of this chapter.

SNP	Variant	Minor/ Major allele	MAF (%)	N	Unconditional		Conditional	
					P	β (SE)	P	β (SE)
rs138726309	H177Y	T/C	0.8	32430	$3.1x10^{-8}$	-0.102 (0.020)	[a]$1.3x10^{-11}$	-0.125 (0.020)
							[b]$3.5x10^{-10}$	-0.115 (0.020)
rs492594	V219L	C/G	48.1	33231	$6.0x10^{-9}$	0.020 (0.004)	[a]$7.1x10^{-10}$	-0.034 (0.005)
							[c]$6.5x10^{-13}$	-0.032 (0.005)

Table 3.1.2.1. Novel coding variants in *G6PC2* significantly associated with FG levels in single variant unconditional and conditional analyses. Exome-wide significance at single variant level was defined as $P<5x10^{-7}$. Results reported were for analysis after adjustment for age, sex and BMI.

[a]After adjustment for common non-coding GWAS SNP rs560887.

[b]After adjustment for common non-coding GWAS SNP rs560887 and common non-synonymous variant V219L.

[c]After adjusting for common non-coding GWAS SNP rs560887 and low-frequency non-synonymous variant H177Y.

MAF: minor allele frequency; N: number of samples; β: regression coefficient estimates or effect size estimates reported for the minor allele in mmol/l; SE: standard error. Table adapted from Mahajan *et al*. 2015 Table 1.

Gene	Chr	SNP	Variant	Minor/ Major allele	MAF (%)	N	Unconditional	
							P	β (SE)
Fasting glucose								
PCSK1	5	rs6234	Q665E	C/G	27.9	33231	$3.0x10^{-8}$	-0.022 (0.004)
PCSK1	5	rs6235	S690T	G/C	27.9	33231	$4.1x10^{-8}$	-0.022 (0.004)
RREB1	6	rs35742417	S1554Y	A/C	21.1	33230	$8.4x10^{-9}$	-0.024 (0.004)
GLP1R	6	rs10305492	A316T	A/G	1.5	33230	$4.6x10^{-7}$	-0.073 (0.015)
ZHX3	20	rs17265513	N310S	C/T	23.8	33229	$3.9x10^{-7}$	0.022 (0.004)
Fasting insulin								
URB2	1	rs141203811	G594V	T/A	0.1	21130	$3.1x10^{-7}$	0.282 (0.066)

Table 3.1.2.2. Additional novel coding variants significantly associated with FG and FI levels in single variant analyses. Exome-wide significance at single variant level was defined as $P<5x10^{-7}$. Results reported were for analysis after adjustment for age, sex and BMI. Chr: chromosome; MAF: minor allele frequency; N: number of samples; β: regression coefficient estimates or effect size estimates reported for the minor allele in mmol/l for fasting glucose and pmol/l for fasting insulin; SE: standard error. Table adapted from Mahajan *et al*. 2015 Table 1.

3.1.3. Multiple coding variants in *G6PC2* were independently associated with FG levels

Two of the coding variants significantly associated with FG in the Mahajan et al. analysis were found within *G6PC2* at the *G6PC2/ABCB11* locus: V219L, a common variant ($P = 6.0 \times 10^{-9}$, MAF = 48.1%), and H177Y, a low-frequency variant ($P = 3.1 \times 10^{-8}$, MAF = 0.8%), as shown in Table 3.1.2.1. The signal for the known intronic lead GWAS SNP, rs560887, was also replicated in the study. To determine if these coding variants were indeed driving the association signal, conditional analyses were carried out to find out if the genotype of the lead GWAS SNP was accounting for the statistical significant association at the coding variant. Consequently, both coding variants remained associated with FG at exome-wide significance after conditioning on rs560887 (V219L: $P_{conditional} = 7.1 \times 10^{-10}$; H177Y: $P_{conditional} = 1.3 \times 10^{-11}$), and were hence determined to exert effects independent of the GWAS SNP. They were also found to be largely independent of one another as they remained significantly associated after conditioning on both rs560887 and the other associated coding variant. In fact, the signals for both coding variants were strengthened after conditional analysis.

G6PC2 was also the only gene with evidence of significant association with FG levels in gene-based tests across different variant masks. The study analysed a total of 15 low-frequency and rare non-synonymous variants in *G6PC2* that were in aggregate associated with FG at exome-wide significance (gene-level significance threshold $P < 2.5 \times 10^{-6}$). These variants and their single variant association results are listed in Table 3.1.3.1. A robust gene-based signal for *G6PC2* was also replicated by a concurrent study by the CHARGE consortium (Wessel et al. 2015). Step-wise conditional analyses, which re-evaluates significance after adjusting for each variant in turn, revealed that the gene-based signal was primarily driven by two variants: H177Y and Y207S (Table 3.1.3.2). The association was substantially diminished when the analysis was re-done conditioning on these two variants across all variant masks. V219L was not included in the gene-based tests due to its high allele frequency. Y207S is a low-frequency variant with a P value that fell only slightly below the significance threshold at single variant level ($P = 9.96 \times 10^{-7}$, MAF = 0.52%). It remained nominally significant after conditioning on lead GWAS SNP rs560887. Overall, there was clear evidence that there were at least four individual association signals within *G6PC2* – three coding (H177Y, V219L, Y207S) and one intronic (rs560887).

As both *G6PC2* and *ABCB11* have been considered potential effector transcripts at this GWAS locus, the authors looked for evidence of association in *ABCB11*. None of the 21 coding variants in *ABCB11* included in the analysis were significantly associated with FG levels, nor was there a gene-based association signal (P>0.05). Together, the genetic data confirmed *G6PC2* as the effector transcript for FG regulation at this GWAS locus, and that V219L, H177Y, and Y207S are likely causal coding variants.

SNP	Variant	Minor/major allele	MAF (%)	Minor allele counts	Direction of effect on FG	*P* value
rs138726309	H177Y	T/C	0.78	502	↓	5.98×10^{-8}
rs2232323	Y207S	C/A	0.52	338	↓	9.96×10^{-7}
rs2232326	S324P	C/T	0.087	56	↓	0.00044
rs145050507	I171T	C/T	0.13	81	↓	0.0018
rs142189264	S30F	T/C	0.036	23	↓	0.033
rs148689354	I273V	G/A	0.0031	2	↑	0.052
rs199682245	N68I	T/A	0.0031	2	↓	0.12
rs145217135	I230T	C/T	0.022	14	↓	0.12
rs150538801	F256L	C/T	0.014	9	↑	0.32
rs201561079	I63T	C/T	0.0093	6	↓	0.34
rs200336133	L310F	T/C	0.065	42	↑	0.44
rs2232322	I171V	G/A	0.0062	4	↑	0.54
rs146779637	R283X	T/C	0.11	71	↓	0.67
rs149874491	I38L	C/A	0.0016	1	↑	0.91
rs184807114	A119V	T/C	0.036	23	↓	0.97

Table 3.1.3.1. FG single variant association results for the rare and low-frequency *G6PC2* coding variants included in the gene-based tests. Variants are arranged in descending order of significance based on P values. The direction of effect is aligned to the minor allele. MAF: minor allele frequency. Table modified from Mahajan *et al*. Supplementary Table 6.

	P_{SKAT}	P_{BURDEN}
Unconditional for all 15 variants in PTV + missense	$3.6x10^{-12}$	$5.1x10^{-13}$
• Conditioning on H177Y	$1.6x10^{-8}$	$1.0x10^{-9}$
• Conditioning on Y207S	$2.2x10^{-9}$	$5.1x10^{-11}$
• Conditioning on H177Y and Y207S	$4.8x10^{-5}$	0.00019
• Conditioning on H177Y, Y207S and I171T	0.0024	0.012
• Conditioning on H177Y, Y207S and S324P	0.0034	0.015
• Conditioning on H177Y, Y207S, I171T and S324P	0.35	0.39
Unconditional for all 4 variants PTV + NS_{strict}	$3.6x10^{-12}$	$5.1x10^{-13}$
• Conditioning on H177Y	$7.0x10^{-7}$	$1.6x10^{-6}$
• Conditioning on Y207S	$3.9x10^{-8}$	$7.3x10^{-8}$
• Conditioning on H177Y and Y207S	0.58	0.44
Unconditional for all 12 variants PTV + NS_{broad}	$2.0x10^{-13}$	$1.2x10^{-17}$
• Conditioning on H177Y	$1.9x10^{-8}$	$2.9x10^{-11}$
• Conditioning on Y207S	$2.5x10^{-9}$	$1.8x10^{-12}$
• Conditioning on H177Y and Y207S	$4.5x10^{-5}$	$7.4x10^{-6}$
• Conditioning on H177Y, Y207S and I171T	0.0017	0.00089
• Conditioning on H177Y, Y207S and S324P	0.0026	0.0015
• Conditioning on H177Y, Y207S, I171T and S324P	0.27	0.10

Table 3.1.3.2. Conditional analysis of coding variants in *G6PC2* which in aggregate demonstrated exome-wide significance with FG levels in gene-based tests. SNP IDs for the variants specified are as follows: H177Y (rs138726309), Y207S (rs2232323), I171T (rs145050507) and S324P (rs2232326). SKAT and BURDEN are two different gene-based tests. Table modified from Mahajan *et al*. Supplementary Table 5.

3.1.4. Impact of *G6PC2* coding variants on other related quantitative traits

To find further evidence that may reveal the involvement of *G6PC2* in other metabolic or disease processes, the authors evaluated the impact of the common non-synonymous variant V219L (rs492594) on T2D and other glycemic measures and quantitative traits based on previously published GWAS data (reported in Supplementary Table 7 of Mahajan et al. 2015). The only notable finding was that the G allele encoding the valine residue of V219L, which had a FG-raising effect, showed a modest increase in T2D risk (P = 0.0011; odds ratio = 1.05; 95% CI 1.02-1.06) in a meta-analysis of exome array data from 28,344 T2D cases and 51,801 controls of European ancestry. The direction of effect on T2D at this variant was unchanged after conditioning on the lead GWAS SNP, unlike that observed for the FG conditional analysis. This finding was consistent with a small case-control study in 3,676 individuals of Chinese ancestry which reported the G allele at V219L to be nominally associated with T2D risk (P=0.0062), though this signal seemed to be tagging that of a non-coding signal downstream of the gene (Hu et al. 2009). No other coding variant in *G6PC2* had been shown to influence T2D risk in European populations. The glucose-raising allele at GWAS SNP rs560887 also showed little evidence of association with T2D (Dupuis et al. 2010; Scott et al. 2012). This study's finding has provided fresh evidence that *G6PC2* is critical not only for controlling FG levels within the physiological range, but that impairment of its function could contribute to T2D pathogenesis.

3.1.5. Experimental aims

The exome array-based association analysis published in Mahajan et al. had discovered new coding variants influencing glycemic traits, but more so evaluated the role of coding variants at known GWAS loci, thereby expecting to pinpoint causal transcripts. This knowledge facilitates future investigation of precise molecular mechanisms influencing glycemic traits and T2D susceptibility. The study established *G6PC2* as an effector transcript influencing FG levels, and uncovered variants in the coding region that were tractable for functional validation. Given the gene's excellent biological credentials and robust genetic evidence in current as well as prior studies, it was sensible to functionally evaluate its coding variants as an example of establishing the relationship between genetic risk variants and glycemic phenotypes.

The aims of my study were to design and implement *in vitro* characterisation pipelines to functionally validate coding variants in *G6PC2* that were associated with FG levels in the exome chip analysis of glycemic traits in >33,000 European individuals from the T2D-GENES/GoT2D consortia. I will do this by assessing protein expression and localisation of the coding variants. This effort will help to provide a clearer picture of the molecular mechanisms underlying coding variation in *G6PC2* and improve our understanding of the correlation between *G6PC2* function and regulation of FG levels in humans.

3.2. Methods

3.2.1. *In silico* mutation analysis

Four *in silico* tools were used to predict the functional effects of non-synonymous G6PC2 variants on protein integrity and function. These were SIFT, PolyPhen-2, Condel and GERP. All four analytical algorithms are described in Section 2.13.

3.2.2. Haplotype analysis

Haplotype analysis was carried out across selected *G6PC2* variants of interest using data from a subset of individuals in the study, in 4,442 individuals from the Oxford Biobank (Tan et al. 2006). The analysis was carried out by Anubha Mahajan and Andrew Morris who were part of the core group of analysts from the T2D-GENES/GoT2D exome array study. This is described in further detail in Mahajan et al. 2015.

3.2.3. Generation of *G6PC2* variant constructs

Variant constructs were generated for the study of G6PC2 coding variant proteins that have, in this current study, been individually associated with FG levels. All mutagenesis was carried out on the pCMV6-Entry-*G6PC2* wild type vector using the Quikchange II Site-Directed Mutagenesis kit (Agilent Technologies) and plasmid sequences verified using Sanger sequencing as described in Section 2.5. The primers designed and used for the PCR are listed in Table 3.2.3.1.

G6PC2 variant	Forward primer (5' to 3')	Reverse primer (5' to 3')
I171T	GTGTCTGCATCTCCAGAGTATTCA**C**AGCAACACATTTTC	GAAAATGTGTTGC**T**G**T**GAATACTCTGGAGATGCAGACAC
H177Y	GTATTCATAGCAACACATTTTCCT**T**ATCAAGTTATTCTTGGAGTAATTG	CAATTACTCCAAGAATAACTTGAT**A**AGGAAAATGTGTTGCTATGAATAC
Y207S	GCCAGTCTGGGCACAT**C**CCTGAAGACCAACC	GGTTGGTCTTCAGG**G**ATGTGCCCAGACTGGC
V219L	TTTCTCTTCCTGTTTGCA**C**TTGGCTTTTACCTGCTTC	GAAGCAGGTAAAAGCCAA**G**TGCAAACAGGAAGAGAAA
S324P	GAGCATTTATTTTATGTGCTG**C**CTTTTTGTAAAAGTGCATCC	GGATGCACTTTTACAAAAAGG**G**CAGCACATAAAATAAATGCTC

Table 3.2.3.1. Primer sequences used for SDM. The letter in bold **red** denotes the nucleotide that was newly incorporated into the DNA sequence during mutagenesis; the underlined triplet denotes the codon that was affected.

3.2.4. Protein expression studies by western blot analysis

HEK293 and INS-1 (832/13) cells were transfected with each wild type or mutant *G6PC2* construct using Lipofectamine 2000 (Invitrogen) in 6-well format, as described in Section 2.7. In protein degradation assays, cells were treated with MG-132 (Calchembio) or chloroquine (Sigma Aldrich) as described in Section 2.8. However, for INS-1 (832/13) experiments cells were treated with MG-132 or chloroquine for only 6h due to cell toxicity observed with longer treatment. Total cellular protein from both HEK293 and INS-1 (832/13) cells were collected, quantified and analysed by western blot similar to that described in Section 2.9. Briefly, all cell lysates were separated by 4-12% SDS-PAGE using the Invitrogen NuPage system (as this was the system available at the time) and transferred to nitrocellulose membranes. G6PC2 protein expression was determined by immunoblotting using a primary antibody against FLAG (Sigma, F1804) or myc (Millipore, 05-724), as indicated. A primary antibody against β-tubulin (Santa Cruz, sc-9104) was used as the loading control. Protein bands were detected using the Pierce ECL reagent (Thermo Fisher Scientific) and developed on film.

3.2.5. Cellular localisation studies by immunofluorescence microscopy

HEK293 cells were transfected on 4-chamber slides or glass cover slips using FuGene 6 transfection reagent (Promega) as described in Section 2.7 and prepared for microscopy as described in Section 2.10. Double immunostaining of cells was carried out using primary antibodies for FLAG (Sigma, F1804) and calnexin (Santa Cruz, sc-11397), followed by secondary antibodies anti-mouse fluorescein (Vector Labs) and anti-rabbit TRITC (Dako). DRAQ5 (Thermo Fisher Scientific) or NucRed 647 ReadyProbes reagent (Thermo Fisher Scientific) was used as the far-red DNA stain where stated. Slides were mounted with Vectashield mounting medium (Vector Labs) and analysed on a confocal microscope as described in 2.10.2. Where otherwise stated, some slides were visualised on a LSM 510 confocal microscope (Zeiss) with a 63X 1.4 N.A. oil objective using the LSM 5 software, due to microscope availability.

3.2.6. Statistical analysis

Two-tailed paired Students' t tests were used to compare mean relative protein expression levels of each variant to WT as described in 2.9.5 and 2.17. Statistical analyses and plotting of graphs were performed on GraphPad Prism 6.0. Results were considered statistically significant at $P<0.05$.

3.3. Results

3.3.1. *In silico* assessment of the impact of G6PC2 coding variants

To begin exploring the functional impact of missense G6PC2 variant proteins, all variants included in the gene-based tests in Mahajan et al. as well as common variant V219L were initially analysed using four sequence- and structure-based bioinformatic tools SIFT, PolyPhen-2, Condel and GERP. These tools predict the impact of missense substitutions on protein function based on different parameters that have been described in Sections 2.13 and 3.2.1. *In silico* analyses are often used to validate a large number of candidate variants to identify putative functional variants as this approach is less costly and time-consuming than *in vitro* studies.

As seen in Table 3.3.1.1, it is clear that different bioinformatic algorithms may not always agree, hence the need to utilise multiple tools which make use of different methods and databases to assess protein structure and function. Among the coding variants that had genetic evidence for influencing FG in Mahajan et al., both H177Y and Y207S were consistently predicted to be deleterious, indicative of LOF of the G6PC2 protein. However, V219L, was consistently predicted as a neutral variant despite its strong association with FG levels. This was one of the reasons why V219L had been excluded from the variant masks in the gene level analyses. Likewise, based on conservation analysis by GERP, V219L is situated at a site that is less conserved across species (lower GERP scores) as compared to H177Y and Y207S. *In silico* methods for annotating variant function are helpful but may not be fully accurate, and may miss out on identifying functional causal variants. They should therefore be supported by *in vitro* assays where possible to elucidate the true functional effects of the variants.

SNP	Protein change[a]	SIFT	PolyPhen-2	Condel	GERP
rs492594	V219L	Tolerated	Benign	Neutral	3.06
rs138726309	H177Y	Damaging	Probably damaging	Deleterious	5.9
rs2232323	Y207S	Damaging	Probably damaging	Deleterious	4.73
rs2232326	S324P	Damaging	Probably damaging	Deleterious	5.76
rs145050507	I171T	Tolerated	Benign	Deleterious	4.73
rs142189264	S30F	Damaging	Probably damaging	Deleterious	5.42
rs148689354	I273V	Damaging	Benign	Neutral	2.13
rs199682245	N68I	Damaging	Probably damaging	Deleterious	5.62
rs145217135	I230T	Damaging	Possibly damaging	Deleterious	5.66
rs150538801	F256L	Damaging	Benign	Deleterious	4.6
rs201561079	I63T	Damaging	Benign	Deleterious	4.46
rs200336133	L310F	Tolerated	Benign	Neutral	-0.69
rs2232322	I171V	Tolerated	Benign	Neutral	5.9
rs146779637	R283X	NA	NA	NA	-0.72
rs149874491	I38L	Tolerated	Benign	Deleterious	5.62
rs184807114	A119V	Tolerated	Benign	Deleterious	4.91

Table 3.3.1.1. *In silico* predictions of the impact of non-synonymous variants in G6PC2 on protein function and the evolutionary conservation of variants. [a]Protein change is based on Ensembl ID ENSP00000364512. Table includes common variant V219L in addition to all rare and low-frequency *G6PC2* coding variants in the gene-based tests in the Mahajan *et al*. study. These are arranged in descending order of significance based on P values like that in Table 3.1.3.1. Predictions for 'Damaging'/'Deleterious' changes are marked in red. GERP scores range from -12.3 to 6.17 with 6.17 indicating the most conserved site. All algorithms were accessed on 8th Aug 2016.

3.3.2. Haplotype analysis in *G6PC2*

A striking result that arose from the association analysis in Mahajan et al. was the observed switch in direction of effect of common variant V219L from an increase to decrease in FG levels, before and after conditional analysis on the GWAS SNP rs560887 (Tables 3.1.2.1 and 3.3.2.1[A]). As previous studies have suggested a regulatory role of rs560887 on *G6PC2* expression (Baerenwald et al. 2013), we reasoned that the observed phenotypic impact of these coding variants might be influenced by the action of the regulatory GWAS SNP on transcript levels. Our lead collaborators therefore carried out additional association analyses of the three coding variants associated with FG in the context of the GWAS variant in a subset of individuals (from the Oxford Biobank), to determine if regional haplotype structure could be playing a role. They found that the C allele encoding the leucine residue of V219L was carried exclusively in *cis* with the glucose-raising A/C allele at rs560887 (Figure 3.3.2.1B). As the estimated effect size of V219L was considerably smaller (β = 0.020 ± 0.004 mmol/l per allele) than rs560887 (β = 0.070 ± 0.004 mmol/l per allele), the effect of Leu219 would be masked by that of rs560887. Similarly, the glucose-lowering T allele of H177Y was carried together with the glucose-raising C allele of rs560887, thus explaining the increase in effect size of H177Y after conditioning on rs560887.

These observations together explained the reversal in direction of effect observed for V219L between unconditional and conditional analyses. Whilst Leu219 appears to have a glucose-raising effect in single-variant analyses, the molecular consequence of Leu219 is more likely to be glucose-lowering (negative β value). The minor alleles of the other two coding variants, H177Y and Y207S, consistently displayed glucose-lowering effects. Given the proposed role of *G6PC2* in islet beta cells it was predicted that these variants would be associated with a reduction of *G6PC2* function, which we would determine through *in vitro* studies.

A

Variant	MAF (%)	β (unconditional)	β (conditional)
GWAS (rs560887)	30.7	-0.080	-
H177Y (rs138726309)	0.3	-0.139	-0.174
Y207S (rs2232323)	0.6	-0.207	-0.148
V219L (rs492594)	45.3	0.019	-0.041

B

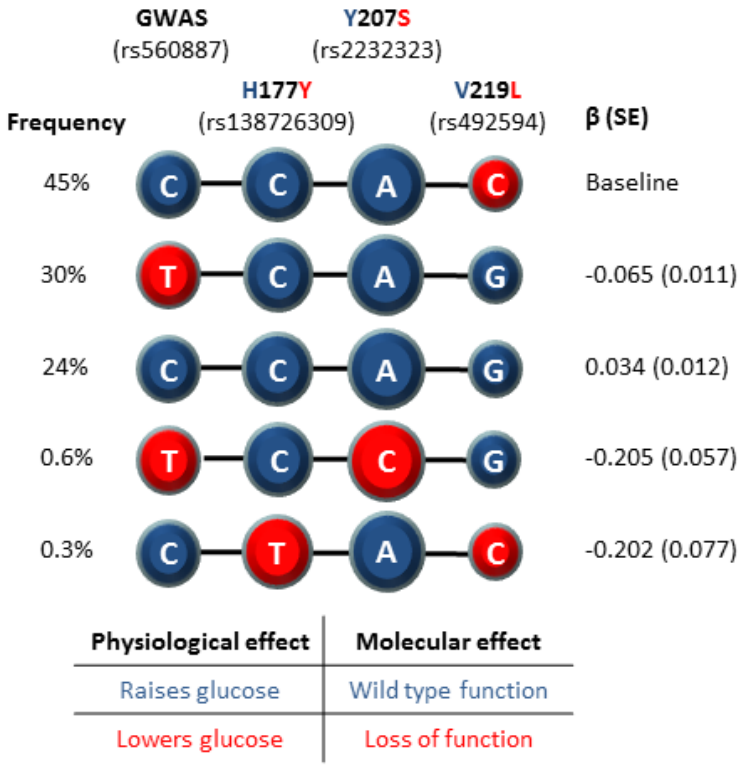


Figure 3.3.2.1. Haplotypes of four variants in *G6PC2* associated with FG in 4,442 unrelated individuals from the Oxford Biobank. (A) Effect size estimates (β) in mmol/l of FG reported for the minor allele of three *G6PC2* coding variants before and after conditioning on GWAS lead SNP (rs560887) in single variant analysis. MAF: minor allele frequency. (B) Haplotypes of the non-coding GWAS lead SNP and the three associated coding variants. The most frequent haplotype was used as baseline for the haplotype association analysis. Wild type, glucose-raising alleles are circled in blue and the mutant, glucose-lowering alleles are circled in red. Diameter of the circle is proportional to the effect size estimates (β) of individual alleles. SE: Standard error. Adapted from Mahajan *et al.* Figure 1.

3.3.3. Protein expression of *G6PC2* coding variants associated with FG

Given the hypothesis that the three glucose-lowering *G6PC2* coding variants that were identified to be likely causal variants for FG are associated with reduction of G6PC2 function, I set to test this with a series of *in vitro* assays. I first generated Myc-FLAG-tagged *G6PC2* variant constructs for H177Y, Y207S and V219L, and transiently transfected cells to investigate the impact of the variant proteins on protein expression. I utilised HEK293 cells, which is a robust protein expression cell model which does not express endogenous G6PC2, enabling the specific interrogation of over-expressed G6PC2. I also replicated results in the rat insulinoma cell line INS-1 (832/13), a more biologically-relevant cell model. Cells transfected with all three constructs showed a marked reduction in G6PC2 protein levels that was not caused by differences in transcript expression (Figure 3.3.3.1). In HEK293 cells, variant protein levels were significantly decreased by 99% for H177Y, 100% for Y207S, and 49% for V219L (all $P < 0.01$, Figure 3.3.3.3), when compared to WT protein (defined here as the major allele at each of these variant sites). Reduced protein expression was also confirmed in INS-1 (832/13) cells where the reduction, though less pronounced, was 76%, 97%, and 49% for each variant respectively (all $P < 0.01$ except for V219L).

Next, we used specific inhibitors of the ubiquitin-proteasomal and lysosomal pathways (MG-132 and chloroquine respectively) to investigate if G6PC2 variant protein levels were being regulated by either or both of the major cellular proteolytic pathways. Previous cellular studies showed that G6PC1 and G6PC2 mutant proteins were accumulated in COS-1 cells in the presence of a proteasome inhibitor lactacystin (Shieh et al. 2002; Shieh et al. 2004). Both lactacystin and MG-132 block proteasomal function via different molecular interactions with the proteasome (Fenteany et al. 1995; Myung et al. 2001). I first tested the length of time of inhibition required to observe an effect in transfected HEK293 cells in a time course experiment with WT and a selected variant. Figure 3.3.3.2 showed that 16h treatment with proteasomal inhibitor MG-132 was able to substantially increase expression of WT and H177Y, yet treatment with lysosomal inhibitor chloroquine for the same amount of time revealed no effect. Subsequently, we showed that protein expression of all variants tested could be rescued in the presence of MG-132 but not chloroquine (Figure 3.3.3.3), indicating that the variants were predominantly degraded through the ubiquitin-

proteasome pathway, an important cellular mechanism for clearing misfolded proteins from the ER (Brodsky and McCracken 1997; Wickner et al. 1999).

The extent of functional impact of the variant proteins mirrored our genetic observations – variant protein V219L, which only had a modest effect on FG, was more stably expressed in both HEK293 and INS-1 (832/13) cells than H177Y and Y207S, both of which have larger impacts on FG levels. The functional results for H177Y and Y207S are also concordant with data from *in silico* mutation assessment tools, which predicted these variants to have harmful effects (Table 3.3.1.1). In contrast, the common V219L variant was predicted to be benign, though *in vitro* characterisation clearly showed a significant effect on protein stability, demonstrating the importance of experimentally evaluating coding variants for functional consequences.

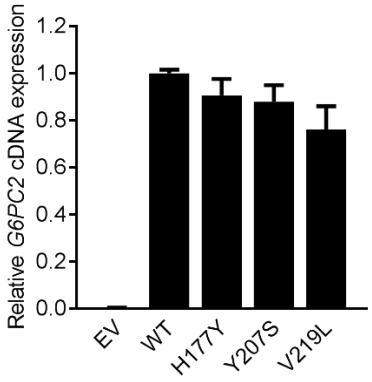


Figure 3.3.3.1. Transcript expression levels of *G6PC2* wild type and variant constructs in in HEK293 cells. Expression level of each variant was not significantly different from wild type based on unpaired Students' t tests (n=4).

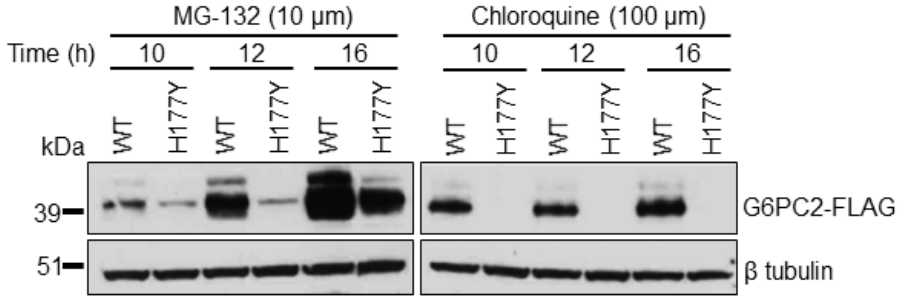


Figure 3.3.3.2. Protein expression levels of G6PC2 wild type and H177Y variant in HEK293 cells treated with MG-132 (proteasomal inhibitor) or chloroquine (lysosomal inhibitor) in a time course experiment. All cell lysates were collected at 48h post-transfection.

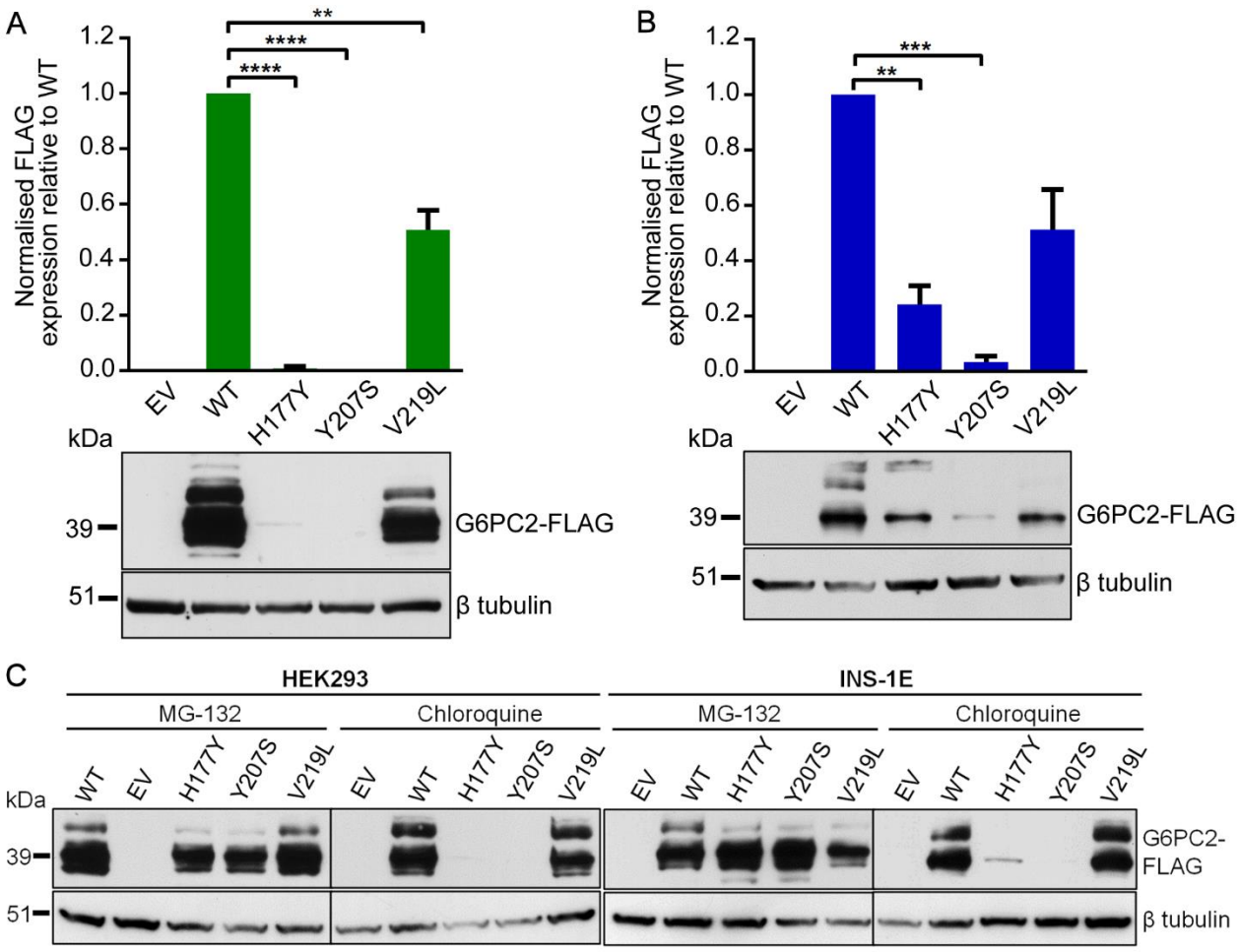


Figure 3.3.3.3. Protein expression levels of wild type and variant G6PC2 proteins in (A) HEK293 cells and (B) INS-1 (832/13) cells determined by western blot and densitometry analysis. (C) Protein expression levels in HEK293 and INS-1 (832/13) cells were assessed in the presence of proteasomal inhibitor MG-132 and lysosomal inhibitor chloroquine. The multiple bands on the western blots are likely to represent glycosylated G6PC2 protein products. Data are presented as mean ± standard error of the mean (SEM) for at least three independent experiments. Significant differences are indicated as ** P<0.01; ***P<0.001; **** P<0.0001. EV: empty vector; WT: wild type. Adapted from Mahajan *et al*. Figure 2.

3.3.4. Cellular localisation of *G6PC2* coding variants associated with FG

In addition, to determine if any *G6PC2* variants exerted their effects on FG due to protein mis-localisation, I used immunofluorescence confocal microscopy to assess subcellular localisation. As G6PC2-WT is expected to localise to the ER membrane and had been demonstrated as such in a previous study using a different cellular system (Shieh et al. 2004), I used calnexin, a protein-folding chaperone integral to the ER as a marker. G6PC2-WT when expressed in HEK293 cells appeared to partially co-localise with calnexin (Figure 3.3.4.1). The fluorescence signals marked the perinuclear space and were largely diffused throughout the cytoplasm, characteristic of ER localisation. When expressed in INS-1 (832/13) cells (Figure 3.3.4.2), G6PC2-WT also appeared to have the same localisation pattern, though co-staining with calnexin was not carried out due to technical difficulties.

Subsequently, variant proteins when expressed in HEK293 cells seemed to co-localise with calnexin, similar to that seen for WT (Figure 3.3.4.3). Even though variant proteins with these missense substitutions did not appear to impact on protein localisation, this approach was unable to definitively prove that the variants are expressed within the ER membrane, due to limited resolution of the current technique that is based on immunofluorescence confocal microscopy. Both G6PC2-WT and V219L also appeared to display additional aggregated signals close to the nucleus, an observation that requires further investigation.

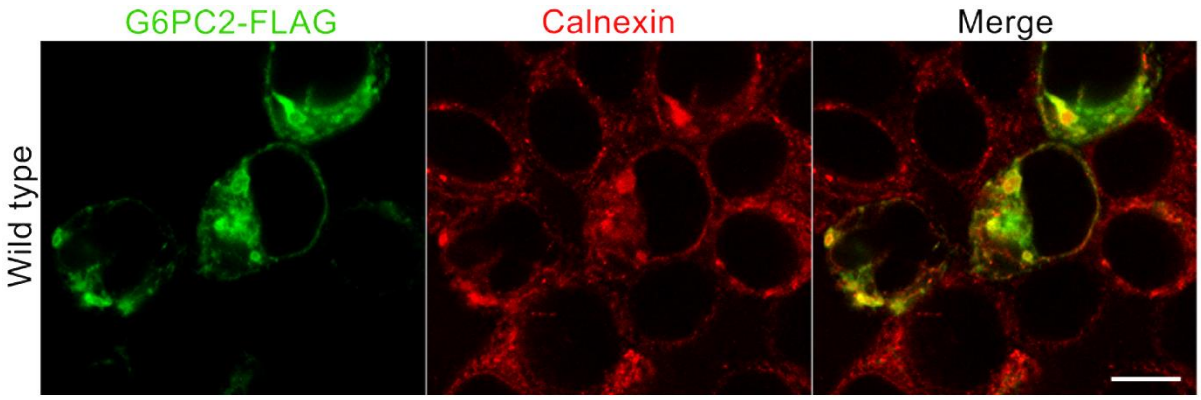


Figure 3.3.4.1. Co-localisation of wild type G6PC2 and ER protein calnexin in HEK293 cells determined by immunofluorescence microscopy. Cells were double immunostained for FLAG (fluorescein, green) and calnexin (TRITC, red) and visualised on the LSM 510 confocal microscope (Zeiss). Scale bar indicates 10µm.

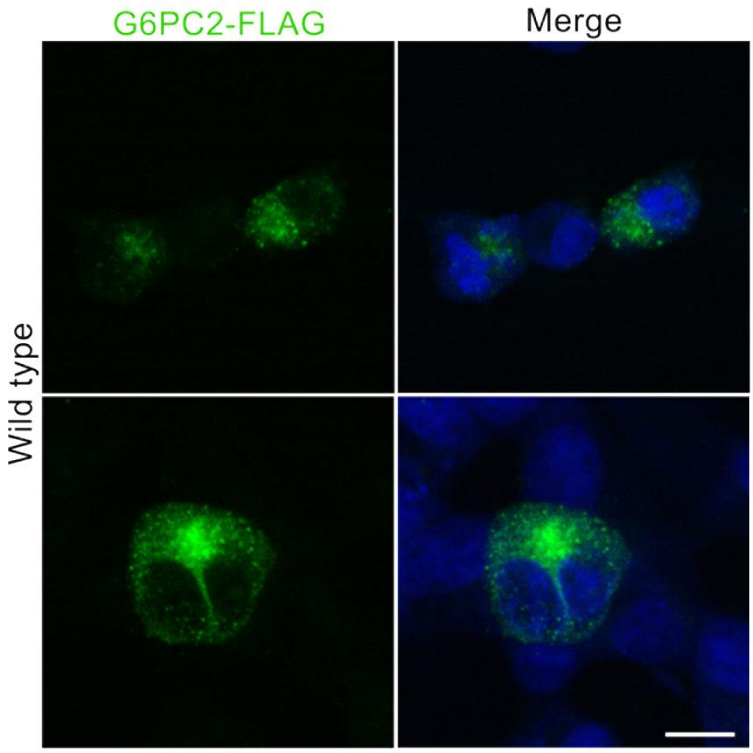


Figure 3.3.4.2. Localisation of wild type G6PC2 in INS-1 (832/13) cells determined by immunofluorescence microscopy. Cells were immunostained for FLAG (fluorescein, green) and images merged with staining of the nuclei with NucRed reagent (blue). Scale bar indicates 10µm.

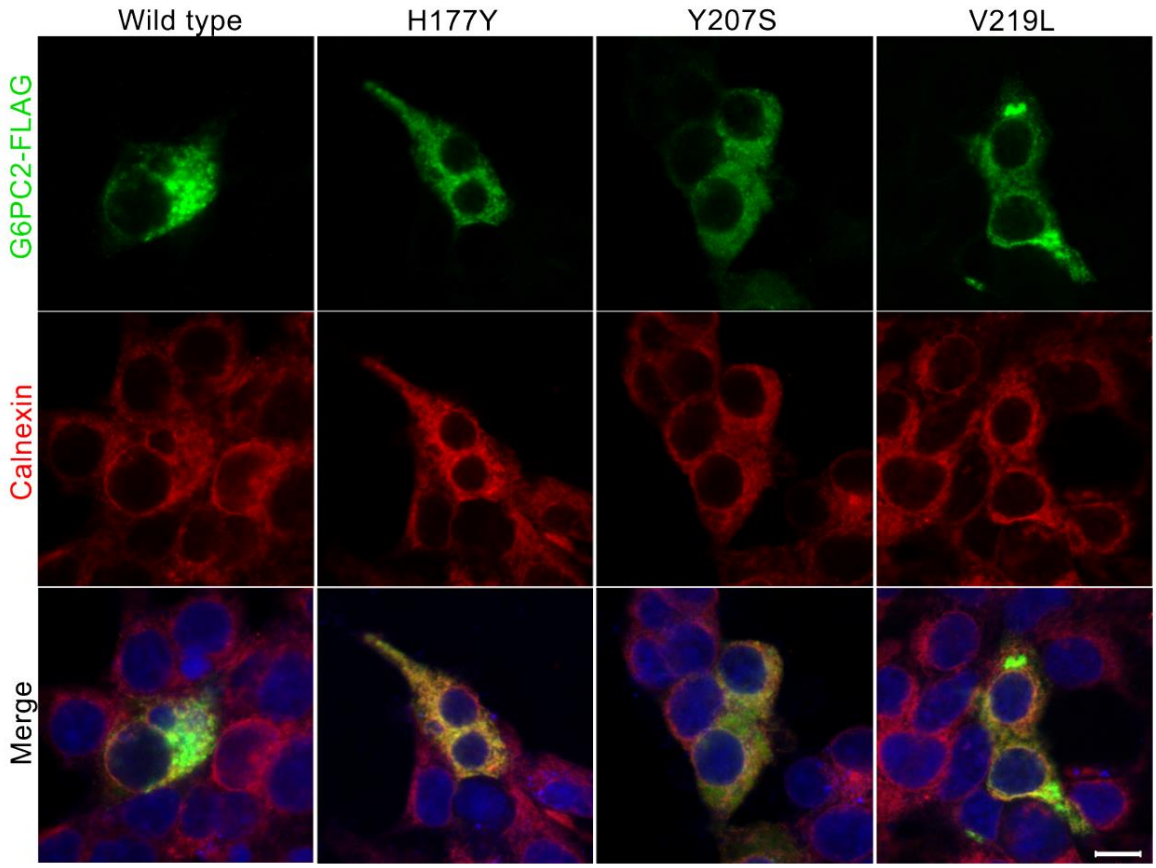


Figure 3.3.4.3. Cellular localisation of wild type and variant G6PC2 proteins expressed in HEK293 cells as assessed by immunofluorescence confocal microscopy. Cells were double immunostained for FLAG tag (green) and ER protein calnexin (red), and merged with staining of the nuclei (blue). Images were taken with laser settings that were optimised separately for each sample. Scale bar indicates 10 µm. Figure adapted from Mahajan *et al*. Figure 1D.

3.4. Discussion

The current study has identified coding variant associations in *G6PC2* that are linked to reduced protein function, highlighting the critical role of *G6PC2* in glucose homeostasis, and demonstrating that coding variants in the gene may impact on FG levels through altered G6Pase function in the pancreatic islets. It confirmed for the first time that *G6PC2* is the effector transcript mediating the reported GWAS signal for FG as no coding signals were detected in the neighbouring gene *ABCB11* in this study. This however does not imply that *ABCB11* is not involved in glucose regulation, only that it is unlikely to be underlying the observed association signal for FG at this locus. The lack of association between *G6PC2* coding signals and insulin traits also supports a beta cell- rather than liver-mediated mechanism. The ascertainment of causal transcripts and causal coding variants at associated genetic loci will greatly facilitate the investigation of biological mechanisms in the relevant *in vitro* systems.

Before the characterisation presented in this chapter was carried out, no functional work on coding variants in the gene had been reported. The identification of naturally-occurring variants in *G6PC2* that resulted in non-synonymous substitutions within the catalytic domain or the transmembrane domain prompted the exploration of *G6PC2* molecular function, and studies of protein activity will likely be needed in addition to protein stability studies. The use of selective inhibitors of major cellular proteolytic pathways also provided mechanistic insight into the regulation of G6PC2 proteins. Mutations in the protein, which is highly embedded in the ER membrane, may result in misfolding and subsequent turnover of unstable proteins primarily driven by the ubiquitin-proteasome system. Our finding that loss of *G6PC2* function leads to a reduction in FG levels in humans is consistent with rodent data, which showed that *G6pc2* knockout mice have decreased FG levels (Wang et al. 2007).

Our strategy of combining genetic and functional analyses has highlighted a number of important implications for the design and interpretation of association studies of coding variation. First, haplotype-based analysis showed that appreciation of regional haplotypes was essential for resolving apparent discrepancies in the direction of effects driven by each allele and our *in vitro* data (as exemplified by G6PC2-V219L). Variants may have to be studied in their relevant haplotype background especially in the context of other regulatory variants. Current experiments do not consider the possible interactions between regulatory

non-coding variants and coding variants (Lappalainen et al. 2011). Secondly, although the widely-used *in silico* prediction tools provided a quick means of assigning functional significance and filtering candidate variants, they are often inaccurate in evaluating functional impact, and may be biased towards variants with larger effects. Impact on protein function will need to be proved through experimental validation when possible. Thirdly, exome-array genotyping studies are limited as they do not capture the full spectrum of coding variation, which will only be achieved with sequencing studies. The novel FG- and FI- associated coding variants identified in Mahajan et al. were still largely common variants.

Given the independent coding variant associations observed in *G6PC2* in the current dataset, it is likely that there are additional rare coding variants in *G6PC2* that influence FG levels, but have not achieved statistically significant association due to their low allele frequencies. For instance, rare variants S324P and I171T were both found to have nominally significant associations with FG at single variant level as seen in Table 3.1.1.1 (S324P: $P=4.4 \times 10^{-4}$, MAF=0.09%; I171T: $P=1.8 \times 10^{-3}$, MAF=0.13%). In gene-based conditional analyses (Table 3.1.3.2), it was only after conditioning on four coding variants (H177Y, Y207S, I171T and S324P) that the gene-based association was completely lost (P>0.05). Therefore, it is possible that the 2 additional rare variants are also contributing to the gene-based signal. *In silico* analyses predicted S324P to have damaging effects, whilst I171T was predicted to be deleterious by only one of the algorithms (Table 3.3.1.1). Indeed, when evaluated for protein stability, S324P was found to result in complete loss of protein expression, whilst I171T was expressed at levels similar to WT (Figure 3.4.1[A]). Like the other variants studied, G6PC2-S324P was degraded uniquely via the proteasomal pathway (Figure 3.4.1[B]). Further *in vitro* characterisations such as localisation and activity assays will be required to follow up on I171T to elucidate additional functional mechanisms by which it may cause reduced FG levels in variant carriers.

To thoroughly evaluate the genetic contribution of rare variants in FG regulation and take greater advantage of gene-based tests, larger sample sizes and exome sequencing data will be required to obtain a more comprehensive assessment of the impact of coding variation and increase the power to detect associations. It is likely that there are additional rare variants of moderate to large effect sizes that impact on glycemic traits. Nonetheless, our study has provided a clear example of how analysis of coding variation can shed light on the

biology underlying GWAS findings, and in particular highlighted *G6PC2* as a glycemic trait locus that warrants further investigation.

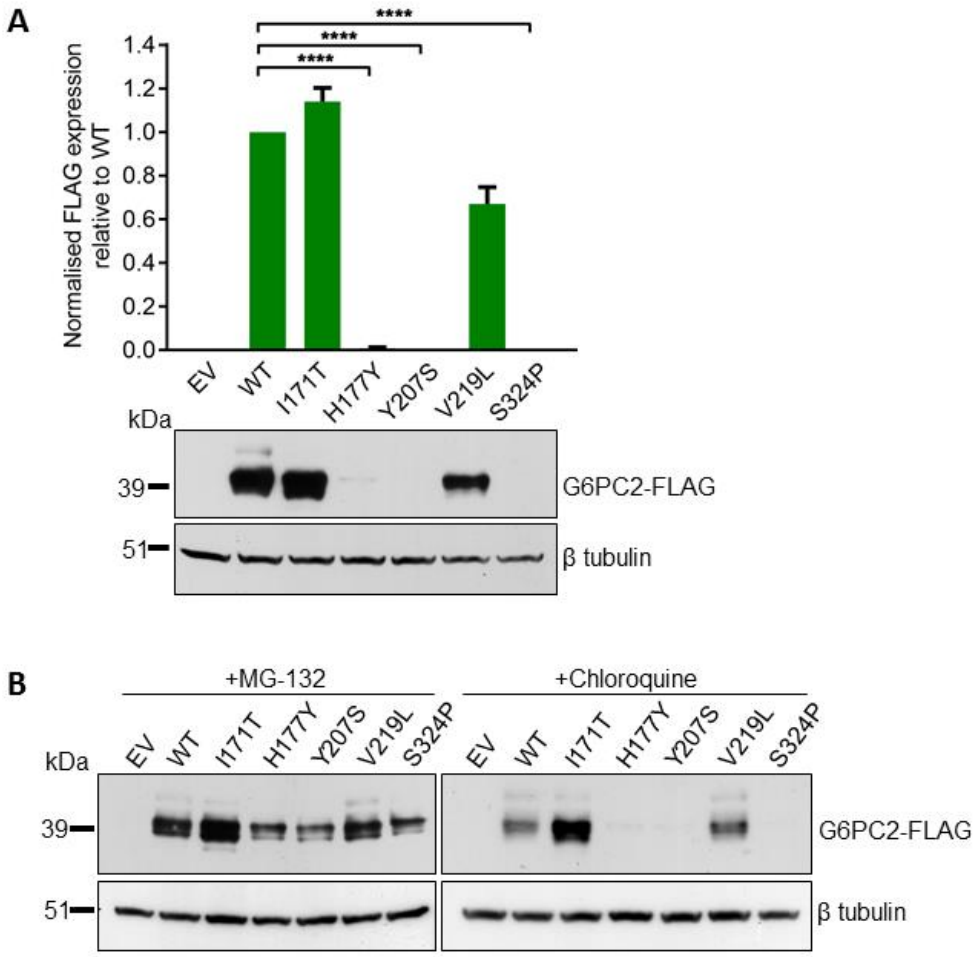


Figure 3.4.1. Protein expression of additional rare G6PC2 variant proteins in HEK293 cells.

(A) Expression levels of G6PC2 variants were determined by densitometric analysis (n=3). Data are presented as mean ± SEM. Significant difference is indicated as **** P<0.0001 based on paired t tests. Expression was also assessed in the presence of (B) proteasomal inhibitor MG-132 and lysosomal inhibitor chloroquine for 15h. Molecular weight in kilodalton (kDA) is indicated on the left of the western blot images. EV: empty vector; WT: wild type.

Chapter 4

Identification and functional characterisation of *G6PC1* coding variants influencing glycemic traits

---

The results presented in this chapter will form part of a MAGIC manuscript which is currently in preparation.

4.1. Introduction

4.1.1. Novel coding variant associations in *G6PC1* from large-scale exome array meta-analysis of glycemic traits

It is clear that studies of coding variation are advantageous as effects from the perturbation of proteins may be more directly predicted and functionally validated. Coding variants tend to be rare (frequency<1%) but rare variation is also known to be enriched for evolutionary deleterious, thus, functional variants (Kiezun et al. 2012; Marth et al. 2011). To facilitate the identification of likely causal variants and effector transcripts associated with glycemic traits, the MAGIC investigators carried out a large-scale exome array meta-analysis that investigated the association of 241,320 exome array variants across the allelic spectrum with four glycemic traits (FG, FI, 2hGlu and HbA1c levels), in up to 144,060 normoglycemic individuals. These individuals represented five major ancestries across 66 studies, of which majority were European (85% European, 6% African-American, 5% South Asian, 2% East Asian and 2% Hispanic). This study, to date the largest in sample size for the investigation of glycemic trait associations, provided increased statistical power to assess the significance of low-frequency (MAF=1-5%) and rare (MAF<1%) variants that underlie both novel and known association signals. Gene-based analyses based on a frequency-weighted burden test and sequence kernel association test (SKAT) were carried out to improve detection of rare variant associations within genes. The burden test assesses mutational load within the same gene due to accumulation of minor alleles of similar effect sizes (Magi et al. 2011); the SKAT approach works well for testing of rare alleles of different effects (Wu et al. 2011). Given the large number of rare variants present, many of which may be neutral, filters based on the NSbroad and NSstrict criteria were set as previously described (Purcell et al. 2014). Briefly, only rare variants were included in the gene-based tests and had to be a PTV or a missense variant predicted to be damaging by at least one of five (for NSbroad) or all five (for NSstrict) *in silico* functional prediction algorithms.

The gene-based analysis results revealed two novel loci that were significantly associated with glycaemic traits at exome-wide significance ($P<2.5 \times 10^{-6}$) – *G6PC1* was associated with FG ($P_{burden}=1.41 \times 10^{-6}$) and FI ($P_{burden}=1.62 \times 10^{-6}$) (Table 4.1.1.1); *TF* was associated with HbA1c ($P_{burden}=2.15 \times 10^{-6}$). Conditional analyses were carried out to identify variants within the set tested that were driving these aggregate signals. Firstly, within *TF*, five missense

variants were shown to drive the association signal for HbA1c (P>0.05 after conditioning on those variants). *TF* encodes transferrin, a circulating serum protein responsible for iron transport from sites of iron absorption and heme degradation to all proliferating cells in the body. Several polymorphisms in *TF* have been previously described and associated with iron homeostasis (Benyamin et al. 2009; Benyamin et al. 2014). Homozygous or compound heterozygous mutations in *TF* are also known to underlie a rare condition called atransferrinaemia, which is characterised by deficient plasma transferrin leading to microcytic hypochromic anemia and hepatic iron overload in humans (Beutler et al. 2000; Knisely et al. 2004). The variant signals for *TF* will not be discussed in detail here, but based on the lack of association observed at this locus for any of the other glycaemic traits studied in addition to the biological evidence discussed, it is likely that the observed effect of *TF* on HbA1c variability is due to non-glycemic pathways. Genetic variants from GWAS may influence HbA1c through both glycemic and non-glycemic pathways that implicate systemic glucose control or erythrocyte biology respectively (Soranzo 2011). In a previous study by the MAGIC investigators, many of the HbA1c-associated common variant signals identified had no evidence of association with other glycemic measures or T2D (Soranzo et al. 2010).

As for *G6PC1*, the gene-based association signal for FG was found to be driven by missense variants A204S (rs201961848) and R83C (rs1801175), while the signal for FI was again driven by A204S (rs201961848) in addition to PTV Q347X (rs80356487), as seen in Table 4.1.1.2. These variants were also the most strongly associated variants (though only nominally) among all the variants tested for their respective traits in single variant meta-analysis (Table 4.1.1.3). *G6PC1* encodes the major glucose-6-phosphatase catalytic subunit in the body responsible for glucose production from gluconeogenic tissues. Though mutations in this gene are already known to be recessively inherited in the rare metabolic disorder GSD1a, this is the first time that variants in *G6PC1* are shown to influence fasting glucose and insulin levels in the physiological range. Among the variants found to influence glycemic trait variance, R83C and Q347X are both known pathogenic variants for causing GSD1a and have been shown to be devoid of phosphohydrolase activity (Lei et al. 1993; Lei et al. 1994). Based on prior evidence we would expect these variants to lower FG and FI levels through LOF effects. On the other hand, the mechanisms through which A204S influences FG and FI levels are still unknown. No associations with T2D risk or other anthropometric traits were identified in *G6PC1* based on existing data from large consortia efforts.

Trait	Mask	No. of variants	Gene effect	P_{burden}	P_{SKAT}
FG	NSbroad	9	-0.14	**1.41×10^{-6}**	1.32×10^{-5}
	NSstrict	3	-0.25	1.41×10^{-3}	7.43×10^{-4}
FI	NSbroad	8	-0.14	**1.62×10^{-6}**	8.58×10^{-6}
	NSstrict	3	-0.07	1.85×10^{-3}	7.80×10^{-3}

Table 4.1.1.1. *G6PC1* gene-based association with FG and FI levels. NSbroad: PTV and missense variants predicted to be deleterious by at least one of five annotation algorithms (SIFT, Polyphen2-HumDiv, PolyPhen2-HumVar, LRT and MutationTaster); NSstrict: PTV and missense variants predicted to be deleterious by all five annotation algorithms. Gene effect results were obtained from the burden test analysis and are reported as mmol/l and pmol/l for FG and FI respectively. P values reflected were obtained from the burden test or SKAT analysis. P values that are exome wide significant, defined as $P<2.5 \times 10^{-6}$, are highlighted in bold. Data provided by Sara Willems, Hidetoshi Kitajima, Xueling Sim from the MAGIC investigators.

Trait	Condition on	P_{burden}
FG	Unconditional	1.41×10^{-6}
	(1) A204S	1.02×10^{-2}
	(2) R83C	3.06×10^{-5}
	(3) Q347X	1.16×10^{-5}
	(1) & (2)	0.204
	(1) & (3)	7.03×10^{-2}
FI	Unconditional	1.62×10^{-6}
	(1) A204S	4.41×10^{-2}
	(2) R83C	2.93×10^{-6}
	(3) Q347X	2.75×10^{-5}
	(1) & (2)	0.47
	(1) & (3)	8.87×10^{-2}

Table 4.1.1.2. Conditional analysis of variants in *G6PC1* highlight variants driving the gene-based association signals. Conditioned variants are the top three most strongly-associated variants in *G6PC1* at single variant level: rs201961848 (A204S), rs1801175 (R83C) and rs80356487 (Q347X). Data provided by Sara Willems, Hidetoshi Kitajima, Xueling Sim from the MAGIC investigators.

Variant SNP ID	Protein change	Alleles (effect/other)	EAF_{FG}	$Effect_{FG}$	P_{FG}	EAF_{FI}	$Effect_{FI}$	P_{FI}	NSbroad	NSstrict
rs201961848	A204S	T/G	3.33×10^{-3}	-0.065	**2.49×10^{-5}**	3.93×10^{-3}	-0.074	**1.05×10^{-5}**	1	0
rs1801175	R83C	T/C	2.89×10^{-4}	-0.179	**1.59×10^{-3}**	2.98×10^{-4}	-0.066	2.82×10^{-1}	1	1
rs80356487	Q347X	T/C	1.74×10^{-4}	-0.162	1.79×10^{-2}	1.55×10^{-4}	-0.336	**1.82×10^{-4}**	1	1
rs796136235	K141N	C/G	4.34×10^{-6}	-0.778	8.45×10^{-2}	-	-	-	1	0
rs181624619	R149Q	A/G	3.36×10^{-5}	+0.126	3.95×10^{-1}	2.64×10^{-5}	-0.030	8.68×10^{-1}	1	0
rs145172999	V232I	A/G	5.10×10^{-4}	-0.0261	4.81×10^{-1}	3.21×10^{-4}	+0.051	4.99×10^{-1}	1	0
rs140224951	L265V	G/C	9.45×10^{-5}	+0.0648	5.95×10^{-1}	1.08×10^{-4}	-0.007	9.88×10^{-1}	1	0
rs149486847	A331V	T/C	1.68×10^{-4}	-0.0513	7.25×10^{-1}	1.43×10^{-4}	-0.047	7.62×10^{-1}	1	0
rs143321486	P315A	G/C	2.08×10^{-4}	-0.0128	7.98×10^{-1}	1.65×10^{-4}	-0.096	3.35×10^{-1}	1	1

Table 4.1.1.3. FG and FI single variant association results of all non-synonymous *G6PC1* variants included in the gene-based analyses from MAGIC. Variants are arranged in descending order of level of significance for FG. All variants are missense except for one PTV (Q347X). Allele changes are reported for the full-length transcript NM_000151. EAF: effect allele frequency. Effect sizes are reported as mmol/l and pmol/l for FG and FI respectively. Variants that are driving the gene-based association signals for FG and FI are highlighted in bold (P > 0.05 after conditioning on those variants). For NSbroad and NSstrict classifications, 1 indicates inclusion in the mask while 0 indicates exclusion. Data provided by Sara Willems, Hidetoshi Kitajima, Xueling Sim from the MAGIC investigators.

4.1.2. Functional analyses of mutations in *G6PC1*

G6PC1 is known to be expressed as a full-length isoform consisting of five exons (reference NM_000151.3) encoding a protein product of 357 amino acids, and a short isoform with alternative splicing of exons 3 and 5 (reference NM_001270397.1) resulting in a protein product of 176 amino acids. The role of the short isoform has not been studied. Up to 84 mutations (missense, nonsense, indels, splice site mutations) in *G6PC1* have so far been identified in GSD1a patients and these have been comprehensively reviewed by Chou and Mansfield 2008. These nonsynonymous variants were found in both helical (transmembrane domains) and non-helical regions as well as in the active sites, based on the proposed topological structure of G6PC1 (Figure 4.1.2.1). Many have been functionally characterised by protein expression and enzymatic analyses to reveal their impact on protein stability and activity (Angaroni et al. 2004; Bruni et al. 1999; Lei et al. 1993; Lei et al. 1994; Shieh et al. 2002). These enzymatic assays were adapted from a sensitive technique for assessing hepatic G6Pase activity first developed by Burchell and colleagues (Burchell et al. 1988), enabling kinetic analysis from small amounts of liver homogenate.

The majority of the variants found in the helical regions which demonstrate significant LOF do so by destabilisation of the protein leading to reduced levels of G6PC1 protein, whilst variants in the non-helical regions are less likely to affect protein stability (Angaroni et al. 2004; Shieh et al. 2002). These studies showed that maintaining the sequence and structural integrity of the transmembrane domains is important for the correct folding and stability of the protein. The comprehensive analysis of such a large collection of variants has increased our understanding of the susceptibility of different domains to amino acid changes and whether these changes were more likely to affect protein stability or catalytic activity.

The mutations found in the active sites as expected completely abolished enzymatic activity when transiently expressed in cells. These were K76N, R83C, R83H, H119L, and R170Q (Lei et al. 1993; Shieh et al. 2002). The WT residues at these four sites have been identified as active site residues based on alignments with conserved phosphatase domain motifs (Hemrika et al. 1997; Stukey and Carman 1997). In addition, another residue at H176 was determined to be the phosphate acceptor during the dephosphorylation reaction, resulting in the generation of a histidine-phosphate intermediate (Ghosh et al. 2002; Pan et al.

1998b). R83 and R170 residues provide stabilisation of the transition state while H119 provides the proton needed to release the glucose moiety (Chou and Mansfield 2008; Lei et al. 1995a). All of these residues localise close to or within the ER lumen, as illustrated in Figure 4.1.3.1, confirming that the catalytic reaction takes place on the luminal side of the ER.

The two PTVs that have been functionally characterised so far, R170X and Q347X, were both enzymatically inactive (Lei et al. 1994; Takahashi et al. 2000). The former understandably so as it results in loss of multiple active site residues. However, the mechanism by which Q347X impacts on protein activity is less straightforward; the transcript is likely to be expressed as the variant occurs within the last exon, evading nonsense mediated decay. Detailed mutational analyses of the carboxyl end of the protein suggested that loss of the consensus ER retention signal (KKSL) was neither responsible for the loss of activity nor essential for ER localisation, as variants lacking this signal remained associated with microsomal membranes and retained some enzymatic activity (Lei et al. 1995a). Therefore, one or more residues prior to this signal must be critical for functional activity.

A number of the GSD1a causal variants reported are also ethnic-specific. In particular, missense variants R83C and Q347X were prevalent mutations identified in Caucasian and Jewish GSD1a subjects, and the splice site variant c.648G>T accounted for more than 54% of all mutations identified in Chinese, Japanese and Korean patients (Chou and Mansfield 2008). In fact, in an Ashkenazi Jewish population the R83C variant allele was found to have much higher MAF (1.4%) than all other populations, and all GSD1a patients in this population were found to be homozygous for R83C (Ekstein et al. 2004). This was consistent with a founder effect for this mutation and the increased frequency of GSD1a observed in the Ashkenazi Jewish population.

The current catalogue of known mutations in *G6PC1*, now linked with information on their molecular effects and disease phenotypes as well as ancestry-specific origins, have greatly benefited the diagnosis pipeline of the disease (Lei et al. 1995b). DNA-based diagnostic testing using targeted sequencing and mutation analysis is now readily available and reliable, which is a clear improvement over previous requirements to assess G6Pase activity in liver biopsy samples (Lei et al. 1994). Nonetheless, there is reported phenotypic heterogeneity exhibited by GSD1a patients and lack of clear genotype-phenotype correlation for some

GSD1a mutations (Chou and Mansfield 2008). The impact of other known (eg. G6PT) and unknown genetic modifiers and the adequacy of current experimental assays used to evaluate protein function will require further investigation.

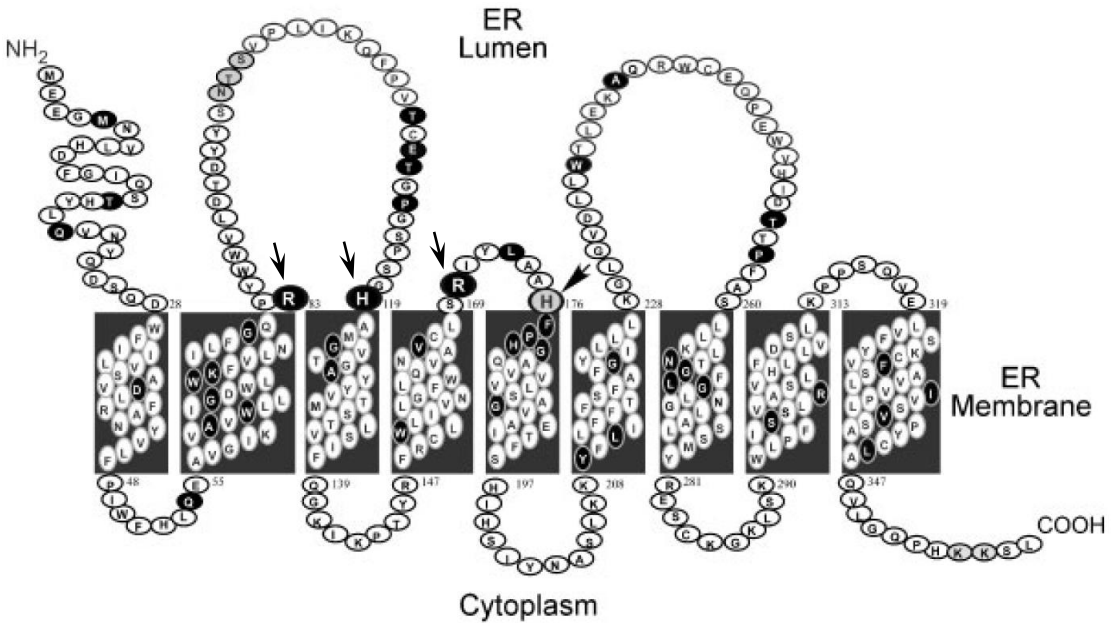


Figure 4.1.2.1. Predicted structure of G6PC1 embedded in the ER membrane. The residues that are found to be mutated in patients with GSD1a are marked in black. The residues that form the catalytic centre are denoted by the black arrows. Figure modified from Chou & Mansfield 2008.

4.1.3. Experimental aims

It is well-established that *G6PC1* has a critical role in hepatic glucose metabolism. Mutations in *G6PC1* are not only linked to a monogenic metabolic disorder in homozygous variant carriers, but new evidence from the MAGIC investigators have now also revealed that they influence a plethora of fasting glycemic traits in heterozygous carriers in a normoglycemic population. I aimed to further explore the relationship between coding variation in *G6PC1* and variations in glycemic traits by carrying out the following:

1. I evaluated the nine coding variants from the MAGIC exome array gene-based analyses for protein expression levels in relevant cellular models, to investigate the spectrum of effects these variants may have on protein stability.
2. I further characterised the three coding variants that were driving the association signals for FG and FI using cellular and enzymatic assays designed to study G6Pase action.

4.2. Methods

4.2.1. *In silico* mutation analysis and online database lookups

Four *in silico* prediction tools SIFT, PolyPhen2, Condel and GERP were used to evaluate the evolutionary conservation of identified *G6PC1* variants and their potential impacts on protein function. These methods were described in Chapter 2 section 2.13. The variants were also looked up in the Exome Aggregation Consortium (ExAC) browser [http://exac.broadinstitute.org/] (Lek et al. 2016), which contains exome sequencing data obtained from multiple large-scale sequencing projects of up to 60,706 unrelated subjects of diverse ancestries, including T2D and other polygenic disease cases. The dataset enables estimation of the allele frequencies of protein-coding genetic variants and aids interpretation of these variants, especially those of very low frequency. However, detailed phenotypic information is unavailable for ExAC samples and so limited disease inferences may be made.

4.2.2. Generation of *G6PC1* variant constructs

G6PC1 variants that were selected to be studied *in vitro* comprise of all variants included in the MAGIC exome array meta-analysis gene-based association tests. Missense variants were generated by site directed mutagenesis of the pCMV6-Entry-*G6PC1* vector backbone with C-terminal Myc and FLAG tags following the steps stated in Section 2.5. Constructs for protein-truncating variants were generated by mutagenesis of the cloned pCMV6-Entry-*G6PC1* vector with additional N-terminal V5 tag as described in Section 2.2. The primers designed and used for the PCR are listed in Table 4.2.2.1. All generated constructs were verified by DNA sequencing as in Section 2.4.4 to confirm that only the desired nucleotide changes were introduced.

G6PC1 variant	Forward primer (5' to 3')	Reverse primer (5' to 3')
R83C	GGATTCTCTTTGGACAG**T**GTCCATACTGGTGGGTT	AACCCACCAGTATGGAC**A**CTGTCCAAAGAGAATCC
K141N	TCCATCTTTCAGGGAA**C**ATAAAGCCGACCTACAG	CTGTAGGTCGGCTTTAT**G**TTTCCCTGAAAGATGGA
R149Q	AAGCCGACCTACAGATTTC**A**GTGCTTGAATGTCATTTTG	CAAAATGACATTCAAGCAC**T**GAAATCTGTAGGTCGGCTT
A204S	GCCACATCCACAGCATCTATAAT**T**CCAGCCTCAAGA	TCTTGAGGCTGG**A**ATTATAGATGCTGTGGATGTGGC
V232I	CTCAAGGGACTGGGT**A**TAGACCTCCTGTGGA	TCCACAGGAGGTCTA**T**ACCCAGTCCCTTGAG
L265V	GCCTCCTCAAGAAC**G**TGGGCACGCTCTT	AAGAGCGTGCCC**A**C**C**GTTCTTGAGGAGGC
P315A	GACTCCTTGAAACCC**G**CATCCCAAGTCGAGC	GCTCGACTTGGGATG**C**GGGTTTCAAGGAGTC
A331V	CTTCTGCAAGAGTG**T**GGTAGTGCCCCTGG	CCAGGGGCACTACC**A**CACTCTTGCAGAAG
Q347X	TACTGCCTCGCC**T**AGGTCCTGGGCC	GGCCCAGGACCT**A**GGCGAGGCAGTA

Table 4.2.2.1. Primer sequences used for SDM. The forward primer sequence is representative of the sense strand. The letter in bold **red** denotes the nucleotide that was newly incorporated into the DNA sequence during mutagenesis; the underlined triplet denotes the codon that was affected.

4.2.3. Gene expression analysis

RNA was collected from untransfected or transfected cells with TRIzol® reagent (Life Technologies). RNA of selected tissue samples were obtained from the Human Total RNA Master Panel II (Clontech) and the FirstChoice Human Total RNA Survey Panel (Ambion). A subset of RNA samples for HepG2, Hep3B and Huh7 cells were kindly provided by Nikolaos Nikolaou from the Tomlinson lab at OCDEM. cDNA synthesis was carried out using the SuperScript III First-Strand Synthesis kit (Life Technologies) and analysed for *G6PC1* expression using the TaqMan® real time quantitative PCR procedure as described in 2.12. The TaqMan® gene expression assays (Applied Biosystems) used are listed in Table 4.2.3.1. Ct values were first normalised to the sample with the lowest Ct value within each gene using the comparative Ct ($2^{-\Delta Ct}$) method. Values of the gene of interest were then normalised to that of the housekeeping genes (HKGs). Where more than one HKG was used, mean values of both HKGs were taken for normalisation.

Target gene	Assay number	Target region (RefSeq transcript)
G6PC1	Hs00609178_m1	Spans 3-4 exon boundary (NM_000151.3)
PPIA	Hs99999904_m1	Exon 4 (NM_021130.3)
TBP	Hs00427620_m1	Spans 2-3 exon boundary (NM_001172085.1) and 3-4 exon boundary (NM_003194.4)

Table 4.2.3.1. TaqMan® gene expression assays used for real-time PCR and their target amplification regions. All assays were obtained from Applied Biosystems.

4.2.4. Protein expression analysis

HEK293, HepG2 and Huh7 cells were transfected with each WT or mutant *G6PC1* construct or empty vector control using Lipofectamine 2000 (Invitrogen) in 6-well format, as described in Section 2.7. Total cellular protein from transfected cells were collected, quantified and analysed by western blot as described in Section 2.9. All G6PC1 samples were analysed using the Bio-Rad Criterion system and Trans-Blot Turbo dry transfer system (Bio-Rad Laboratories). Protein expression was determined by immunoblotting using a FLAG primary antibody (Sigma, F1804) for detecting C-terminal FLAG tags, and a V5 antibody (Invitrogen, 46-0705) for detecting N-terminal V5 tags of truncated proteins. An antibody for G6PC1 (Abcam, ab93857) was also used for detecting endogenous G6PC1 protein in cell lines and tissues. β-tubulin (Santa Cruz, sc-9104) was used as the loading control. At least three independent experiments were carried out for each cell line.

4.2.5. Cellular localisation

HEK293, HepG2 and Huh7 cells were transfected on 4-chamber culture slides (Corning) using FuGene 6 transfection reagent (Promega) as described in Section 2.7 and prepared for microscopy as described in Section 2.10. Where stated, double immunostaining of cells was carried out using primary antibodies for FLAG (Sigma Aldrich, F1804) or V5 (Invitrogen, 46-0705), together with calreticulin (Thermo Scientific, PA3-900) as the ER marker or TGN46 (Sigma Aldrich, T7576) as the marker for the golgi apparatus. This was followed by secondary antibody incubation with anti-mouse Alexa Fluor 488 and anti-rabbit Alexa Fluor

568 (both Life Technologies). DRAQ5 (Thermo Fisher Scientific) was used as a far-red DNA stain. Slides were mounted with ProLong Gold Antifade Reagent (Life Technologies) and analysed on a confocal microscope as in 2.10.2.

4.2.6. Enzymatic activity assay

Microsomal protein from transfected cells were collected, quantified and analysed for glucose-6-phosphatase activity as described in 2.11. Each experiment contained an untransfected HEK293 control and transfected G6PC1-WT sample. Data were fit to the Michaelis-Menten model using GraphPad Prism 7.0.

4.2.7. Statistical analysis

For the analysis of western blot densitometry data, two-tailed paired Students' t tests were carried out to compare mean relative expression levels of each variant to WT as described in 2.9.5 and 2.17. For the analysis of enzymatic activity, two-tailed unpaired Students' t tests of kinetic constants (V_{max} and K_m) were carried out to compare variant to WT, as described in 2.11.3 and 2.17. Compilation and transformation of data were performed on Microsoft Excel. Statistical analyses and plotting of graphs were performed on GraphPad Prism 7.0. Data were considered statistically significant at $P<0.05$.

4.3. Results

4.3.1. *In silico* mutation analyses of *G6PC1* coding variants

Gene-based association analyses from MAGIC showed that rare variants in *G6PC1* were in aggregate significantly associated with both FG and FI levels at exome-wide significance, as seen in Table 4.1.1.1. The tests comprised of nine rare nonsynonymous coding variants – eight missense and one PTV. The variants in the context of the gene and protein structure are illustrated in Figures 4.3.1.1 and 4.3.1.2 respectively. All the variant alleles that occur within the last exon implicate only the full-length isoform of *G6PC1*. At protein level, the variants map to different domains including the active site, transmembrane helices, cytosolic loops and ER luminal loops. All the variants analysed, except for A204S, were very rare with MAF<0.03%. Similar allele frequencies were found for the variants in the Exome Aggregation Consortium (ExAC) database (Lek et al. 2016). A number of variants appear to be ancestry-specific in ExAC, which has a greater representation of most non-European populations in terms of sample size as compared to MAGIC. These included R149Q identified in East Asians and A331V in South Asians. A204S, which was the only variant found to be driving both FG and FI association signals, appeared to be largely identified in Finnish individuals in the ExAC data. As the MAGIC data also contained overlapping Finnish cohorts, it is likely that the gene-based association is driven by the Finnish samples. *In silico* functional predictions based on a number of computational tools showed that A204S was predicted as likely to damaging by PolyPhen2 and Condel, but was annotated as a benign variant by SIFT (Table 4.3.1.3). Only two missense variants, R83C and P315A, were predicted consistently across all algorithms to be damaging.

The nine variants were included in the gene-based tests based on the NSbroad variant filtering criteria. Variants were also classified into an NSstrict criteria which additionally required any missense variant to be predicted as damaging by all five functional annotation algorithms, as a way of enriching for variants that are like to be causal (Purcell et al. 2014). However, the gene-based signals were lost when tests were performed using the NSstrict criteria, which comprised of only three variants. This indicated that this filtering strategy had missed out on variants that are likely to be functional and contributing to the associations with FG and FI, and highlighted the limitations of currently used *in silico* prediction tools in

accurately identifying variants of significance. To elucidate the true functional effects of the *G6PC1* variants identified in the MAGIC dataset, I investigated all nine variants in *in vitro* assays.

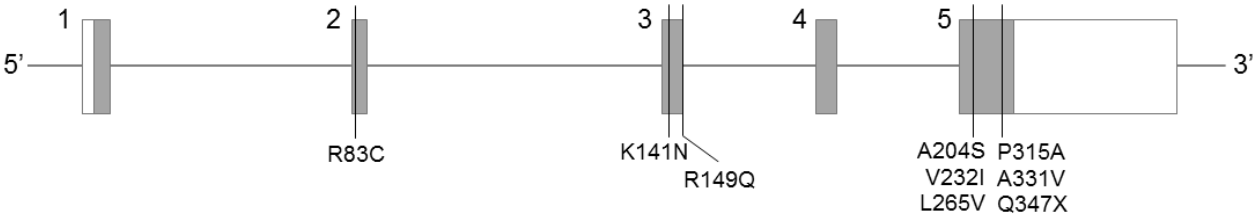


Figure 4.3.1.1. Schematic of *G6PC1* gene structure showing variants selected for *in vitro* functional analysis. The five exons are labelled 1-5 with dark boxes representing coding regions and white boxes representing 5' and 3' untranslated regions. Illustration corresponds to the full-length transcript NM_000151 (not drawn to scale).

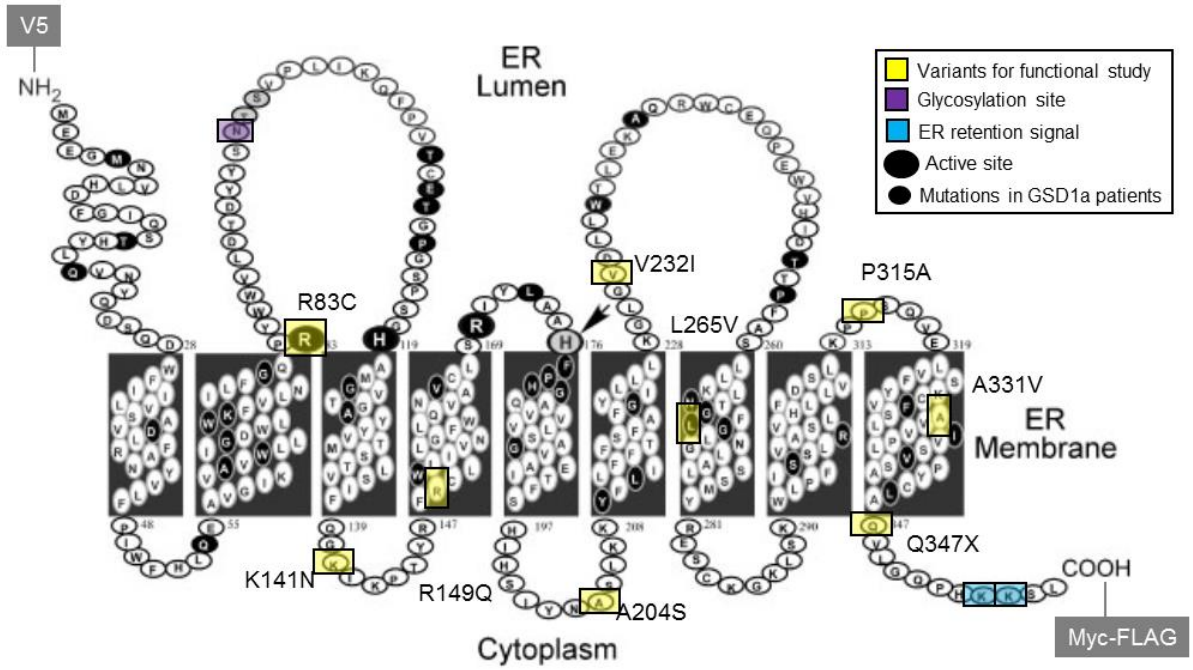


Figure 4.3.1.2. Non-synonymous G6PC1 variants selected for *in vitro* functional analysis in the context of the protein in the ER membrane (figure modified from Chou & Mansfield 2008). The illustration corresponds to the full-length protein NP_000142. Amino acid residues of interest are highlighted as described in the legend. As annotated in the original paper, residues that are mutated in patients with GSD1a are denoted by small black circles; active site residues are denoted by large circles and the arrow. The N-terminal V5 and C-terminal Myc-FLAG tags present in the expression constructs used are indicated.

	Variant SNP ID	Protein change	Consequence	% MAF in MAGIC[a]	% MAF in ExAC[b] (Ancestry[c])	SIFT	Polyphen2	Condel	GERP	Location	Previous reports
1	rs1801175	R83C	Missense	0.029	0.053 (EU)	Damaging	Probably damaging	Deleterious	4.93	LL1, active site residue	Prevalent GSD1a mutation (Lei et al. 1993)
2	rs796136235	K141N	Missense	0.00043	None	Tolerated	Benign	Deleterious	1.93	CL2	None
3	rs181624619	R149Q	Missense, splice site	0.0034	0.0079 (EA)	Tolerated	Benign	Deleterious	3.08	TM4	None
4	rs201961848	A204S	Missense	0.33	0.12 (EU Finnish)	Tolerated	Possibly damaging	Deleterious	3.90	CL3	None
5	rs145172999	V232I	Missense	0.051	0.069 (EU)	Tolerated	Benign	Neutral	2.74	LL3	Clinical significance likely benign (ClinVar 214457)
6	rs140224951	L265V	Missense	0.0095	0.0099 (EA, EU)	Tolerated	Benign	Deleterious	1.82	TM7	Multi-allelic site with L265P and L265Xfs (frameshift) variants reported in GSD1a patients (Seydewitz and Matern 2000; Trioche et al. 2000)
7	rs143321486	P315A	Missense	0.021	0.016 (EU)	Damaging	Probably damaging	Deleterious	5.07	LL4	Clinical significance likely pathogenic (ClinVar 214461)
8	rs149486847	A331V	Missense	0.017	0.062 (SA)	Tolerated	Possibly damaging	Deleterious	4.91	TM9	None
9	rs80356487	Q347X	Nonsense	0.017	0.027 (EU)	NA	NA	NA	-1.8	C-terminus	Prevalent GSD1a mutation (Lei et al. 1994)

Table 4.3.1.3. *In silico* analysis and look ups of *G6PC1* variants selected for functional analysis. Variants are arranged according to amino acid position. *In silico* analyses of variants were carried out based on the full-length protein sequence with Ensembl ID ENSP00000253801. Red boxes are used to highlight deleterious predictions. The algorithms and ExAC database were accessed on 13/9/16.

[a]n=Up to 144,060 individuals comprising 85% European, 6% African-American, 5% South Asian, 2% East Asian and 2% Hispanic

[b]n=Up to 60,706 individuals comprising 54% European (non-Finnish), 14% South Asian, 10% Latino, 9% African-American/African, 7% East Asian, 5% Finnish and 1% others

[c]Ancestry with the greatest allele frequency or allele counts observed for the variant

MAF: Minor allele frequency; GERP: Genomic Evolutionary Rate Profiling; TM: transmembrane region; CL: cytoplasmic loop; LL: endoplasmic reticulum luminal loop; C-terminus: carboxy terminal domain.

4.3.2. Protein expression analysis of G6PC1 variant proteins

To investigate the effects of *G6PC1* coding variants on protein expression levels, constructs for all nine variants were generated (Figure 4.3.2.1) and transiently transfected into HEK293 cells and two human hepatocellular carcinoma cell lines, Huh7 and HepG2. HEK293 cells were chosen as they were originally derived from human embryonic kidney cells grown in culture, and are a well-established fast-growing cell model for protein expression studies. Huh7 and HepG2, both commonly-used hepatocyte cell models, added to the value of studying *G6PC1* variants in a biologically-relevant context, as the liver is the major gluconeogenic organ in the body. I confirmed that strong endogenous *G6PC1* expression at the transcript level was present in selected human tissue samples, including the liver and kidney (Figure 4.3.2.2). The reason for lack of *G6PC1* expression in one of the kidney RNA samples is however not clear. Using the same assay, a panel of hepatocyte and kidney cell lines were also shown to express endogenous *G6PC1*, in contrast to a human pancreatic beta cell line EndoC-βH1 (Figure 4.3.2.2), and this was confirmed at protein level (Figure 4.3.2.4).

Western blot analysis was carried out to detect recombinantly-expressed G6PC1 protein using the C-terminal FLAG tags or N-terminal V5 tags of the missense variants and PTVs respectively. Results collected from all three cell lines are shown in Figure 4.3.2.3. However, as the HepG2 cells demonstrated poor transfection efficiency, thus reducing the reliability of densitometric analysis, the current analysis will focus on data from HEK293 and Huh7.

First, it was observed that most variants did not display severe effects on protein stability. In fact, seven of nine variants were expressed at least 50% that of WT levels in both HEK293 and Huh7 cells. Secondly, R83C and A204S were the only variants studied to consistently give rise to significantly reduced protein levels compared to WT in both cell models. R83C levels were reduced by 54% (P=0.002) and 58% (P=0.01) in HEK293 and Huh7 respectively, largely due to complete loss of the glycosylated form of the protein. The bands of higher molecular weight on the western blots are expected to represent glycosylated protein as G6PC1 is known to have at least one known glycosylation site at N96 (Pan et al. 1998a), and has in previous studies been shown to be present as 37 and 41 kDa proteins (Shieh et al. 2002). In Huh7 and HepG2 cells the glycosylated form of G6PC1 is the major form, and in

fact in Huh7 cells almost all protein is normally glycosylated. In addition to the lack of glycosylation, R83C is also an active site residue in the catalytic domain and the variant was previously shown to have completely abolished phosphatase activity at 10 mM G6P (Lei et al. 1993).

As for A204S, variant protein levels were reduced by 33% (P=0.01) and 30% (P=0.04) in HEK293 and Huh7 cells respectively. I attempted to corroborate these results using an antibody recognising a sequence within human G6PC1, however the antibody used only detected the 37 kDa unglycosylated form of G6PC1, hence effects on total protein expression cannot be analysed (Figure 4.3.2.4). A204S has not been previously studied, but was classified as a possibly damaging substitution by *in silico* prediction tools Polyphen2 and Condel respectively. The variant is located within a cytoplasmic loop and not known to be in close proximity to the catalytic sites. Only one variant that is localised to a cytoplasmic loop (Q54P) has so far been reported as a causal LOF variant for GSD1a (Shieh et al. 2002). Nonetheless, A204S is at least in part likely to influence FG and FI levels through loss of protein stability.

Finally, the only PTV in the analysis, Q347X, was shown to be expressed as a protein of smaller molecular weight compared to WT as shown in Figure 4.3.2.3 due to loss of 10 residues at its C-terminus. As the allele change resulting in a premature stop codon occurs within the last exon, it is predicted to evade nonsense-mediated decay (NMD) and be expressed in variant carriers, although this has not been specifically investigated. My cellular experiments demonstrated that Q347X had markedly reduced protein levels by 81% of WT (P<0.001) though this was only observed in Huh7 cells. Like R83C, it is also known as a functional causal variant in GSD1a that is prevalent in the Caucasian population, and previous studies have shown that it lacked any phosphatase activity in response to 10 mM G6P (Lei et al. 1994). As the truncation is expected to result in the loss of a conserved localisation signal to the ER (KKXX), I predicted that the variant may have LOF due to multiple functional mechanisms.

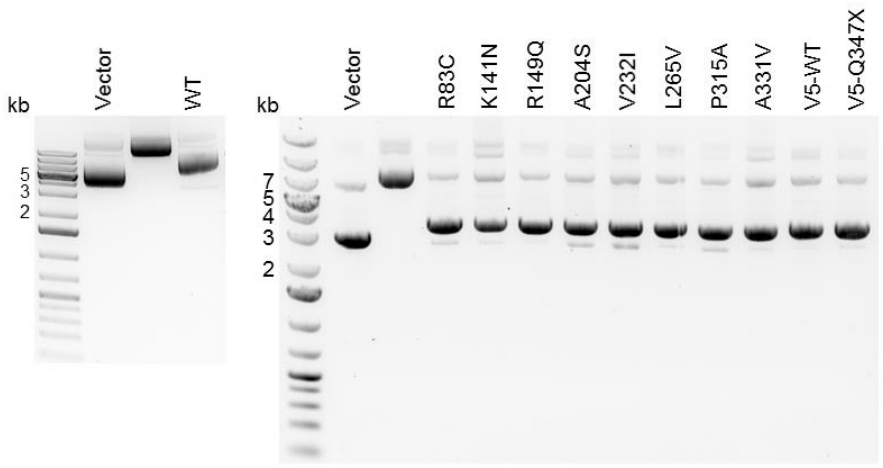


Figure 4.3.2.1. DNA gel of all pCMV6-G6PC1 WT and mutagenised plasmid constructs used in cellular experiments. 1 µg of midiprep DNA was loaded and run on a 0.8% agarose gel, alongside the 1kb plus DNA ladder (Thermo Scientific). kb: kilobase.

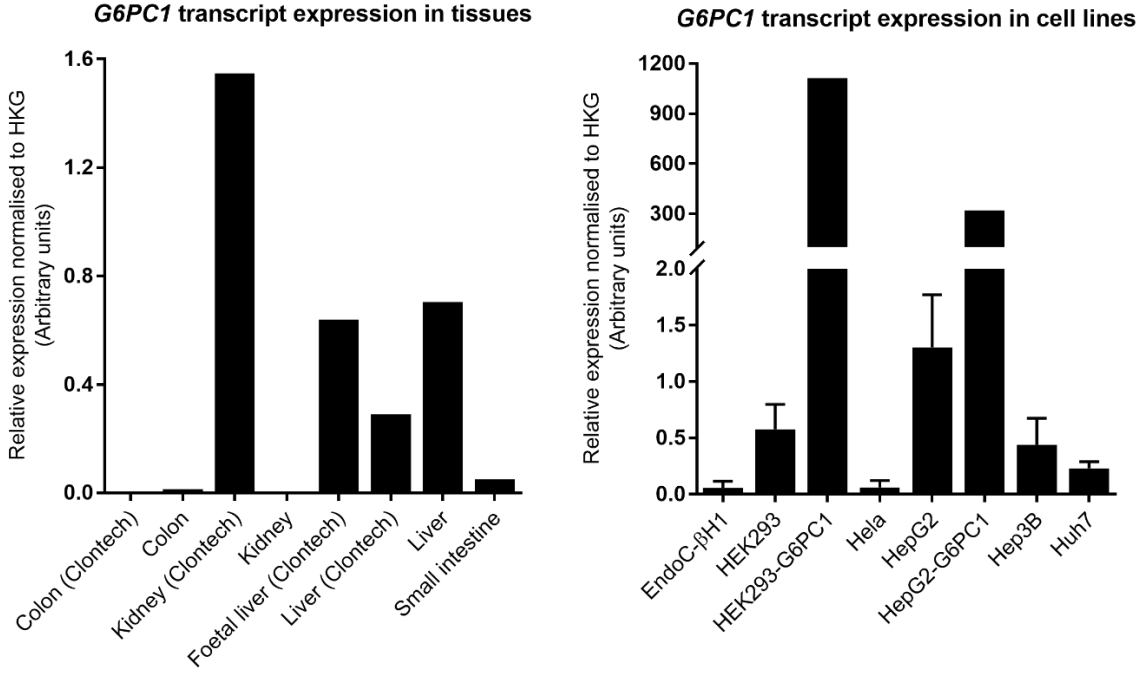


Figure 4.3.2.2. *G6PC1* mRNA expression across selected human tissues (left) and human cell lines (right). Expression levels were quantified by real-time PCR using commercially-available TaqMan® assays for *G6PC1* and housekeeping genes (HKGs) *TBP* and *PPIA*. Details of these assays are described in the Methods section. For the tissue panel (left), one experiment was performed with RNA samples obtained from Ambion or Clontech human RNA tissue panels. For the cell line panel (right), with the exception of cells transfected with *G6PC1* WT construct (HEK293-G6PC1 and HepG2-G6PC1), mean expression levels ± SEM are shown from multiple samples (EndoC-βH1, HEK293, Hela n=2; Hep3B, Huh7 n=3; HepG2 n=5).

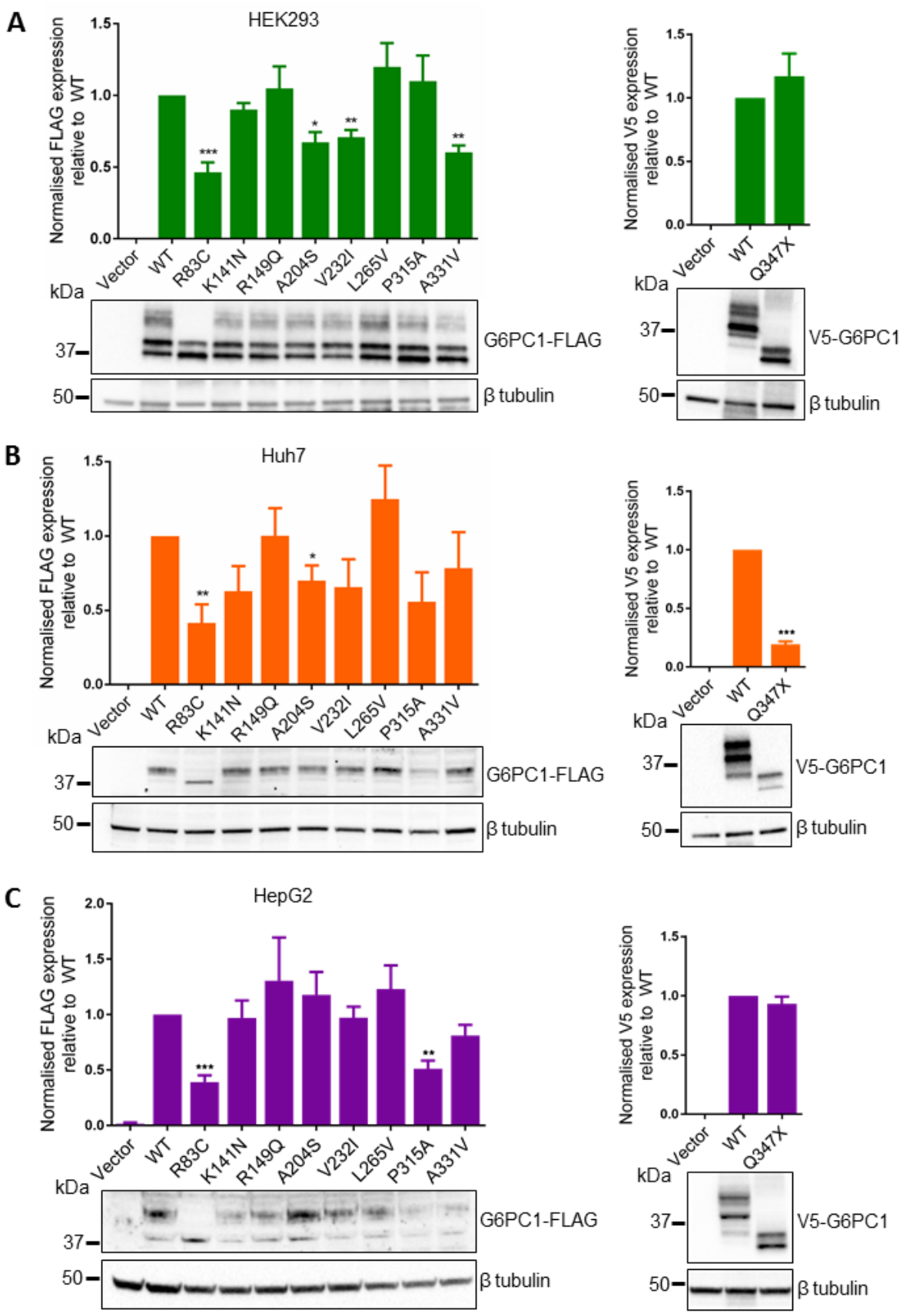


Figure 4.3.2.3. Functional characterisation of G6PC1 variant proteins in (A) HEK293, (B) Huh7 and (C) HepG2 cells by western blot analysis. Legend continued on next page.

Figure 4.3.2.3. Protein expression levels of non-synonymous G6PC1 variants were determined by densitometric analysis of tagged G6PC1 constructs relative to β tubulin control using a FLAG antibody for missense variants (left) and V5 antibody for PTV Q347X (right). Data in graphs are presented as mean ± SEM, with corresponding representative western blots. Data for missense variants were obtained from 4-5 independent experiments; data Q347X were obtained from 3-4 independent experiments. Statistical analysis of densitometry data was carried out using paired Students' t tests, which compares the expression level of each variant to WT. * P=0.01-0.05; ** P=0.001-0.01; *** P<0.001. Molecular weights in kilodalton (kDA) are indicated on the left of the western blot images. WT: wild type.

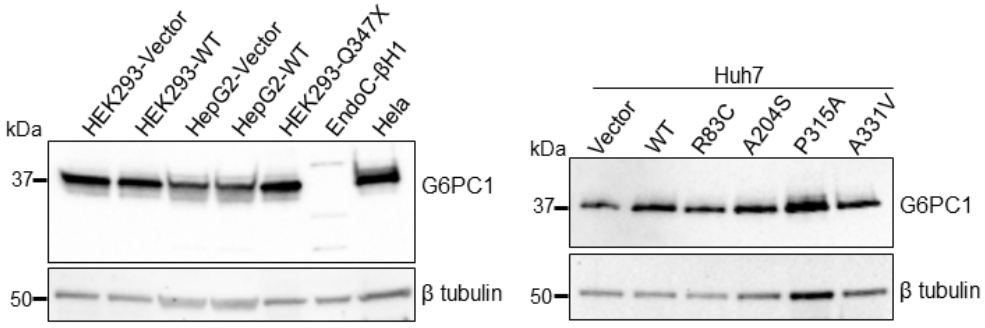


Figure 4.3.2.4. Western blots showing G6PC1 protein expression detected using an antibody against endogenous G6PC1 (Abcam, ab93857) in several untransfected and transfected cells (left) and in transfected Huh7 cells (right). Molecular weights in kilodalton (kDA) are indicated on the left of the western blot images. WT: wild type.

4.3.3. Subcellular localisation of G6PC1 WT and variant proteins

To compare the cellular localisations of G6PC1 WT and variant proteins, in particular Q347X, cells were transfected and co-immunostaining of expressed protein and ER marker calreticulin or Golgi apparatus marker TGN46 was performed. G6PC1 was previously shown to localise to the nuclear membrane and vesicular structures close to the nucleus, characteristic of the ER (Soty et al. 2012). In my studies, I confirmed an expression pattern characteristic of the ER but additionally observed that G6PC1 WT co-localises with the trans Golgi network (TGN). Staining for TGN46 was poor in HEK293 (Figures 4.3.3.1 and 4.3.3.4), but the overlapping fluorescent signals of G6PC1 and TGN46 were clear in Huh7 and HepG2 cells (Figures 4.3.3.2 and 4.3.3.3).

As Q347X loses an ER retention signal at the C-terminal end of the protein, I tested if the variant may be mislocalised in cells. Indeed, unlike WT protein Q347X did not co-localise with the Golgi network in all the cell lines tested (Figures 4.3.3.1, 4.3.3.2 and 4.3.3.3). Q347X localisation also appeared to be more diffused throughout the cytoplasm than WT, though due to limited resolution it is not possible to definitively differentiate between ER and cytoplasmic localisation with the current technique. On the other hand, missense variant A204S clearly displayed similar localisation patterns as WT (Figure 4.3.3.4).

Therefore, I have showed that G6PC1 normally localises to both the Golgi apparatus and the ER, where it is expected to be post-translationally processed and retained respectively. The absence of Q347X in the Golgi apparatus may result in lack of necessary post-translational modifications required for functional activity.

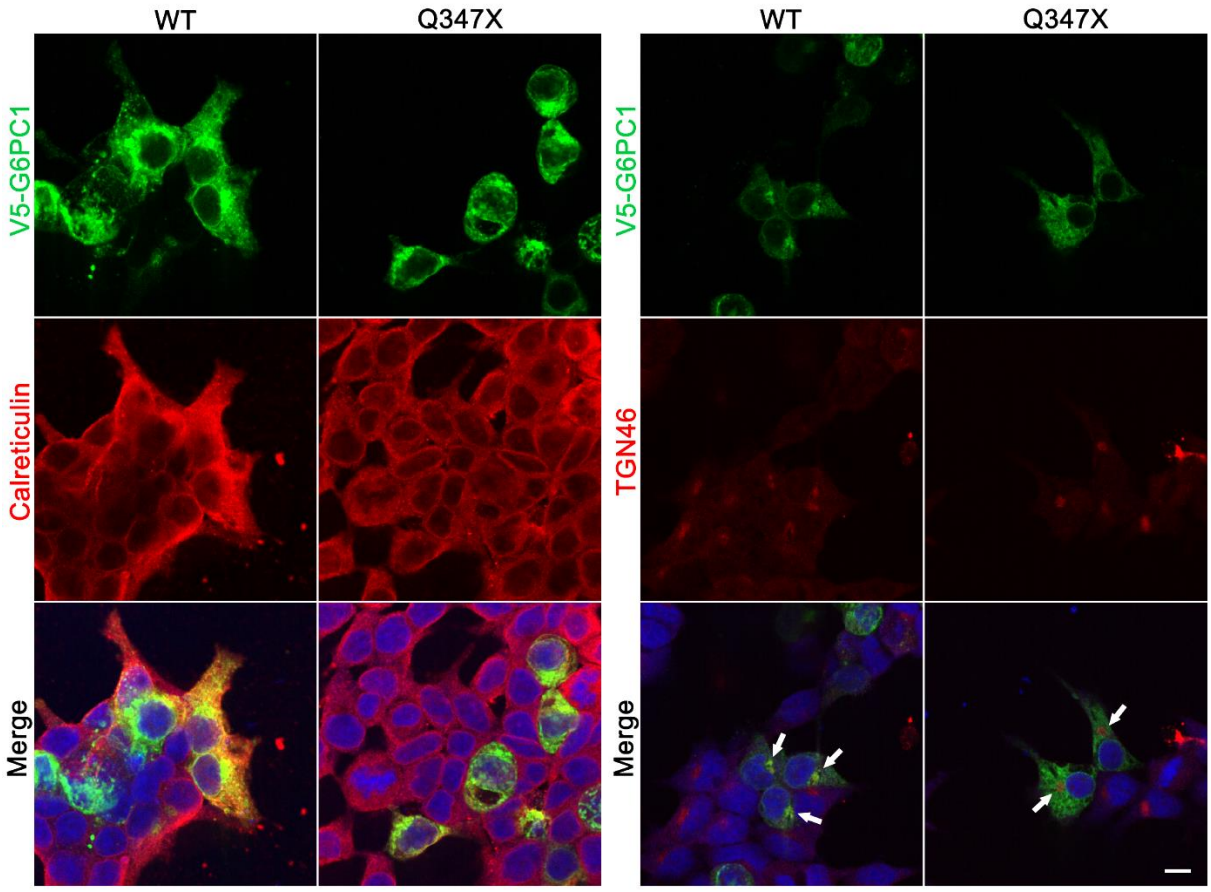


Figure 4.3.3.1. Cellular localisation of wild type G6PC1 and Q347X with respect to the ER (left) and Golgi network (right) in HEK293 cells was assessed by immunofluorescence confocal microscopy. Cells were double immunostained for FLAG tag (green) and markers for the ER, calreticulin, on the left, or trans-golgi network, TGN46, on the right (both in red). Images were merged with staining of the nuclei using DRAQ5 (blue). White arrows point to positions of the golgi apparatus. Scale bar indicates 10 µm.

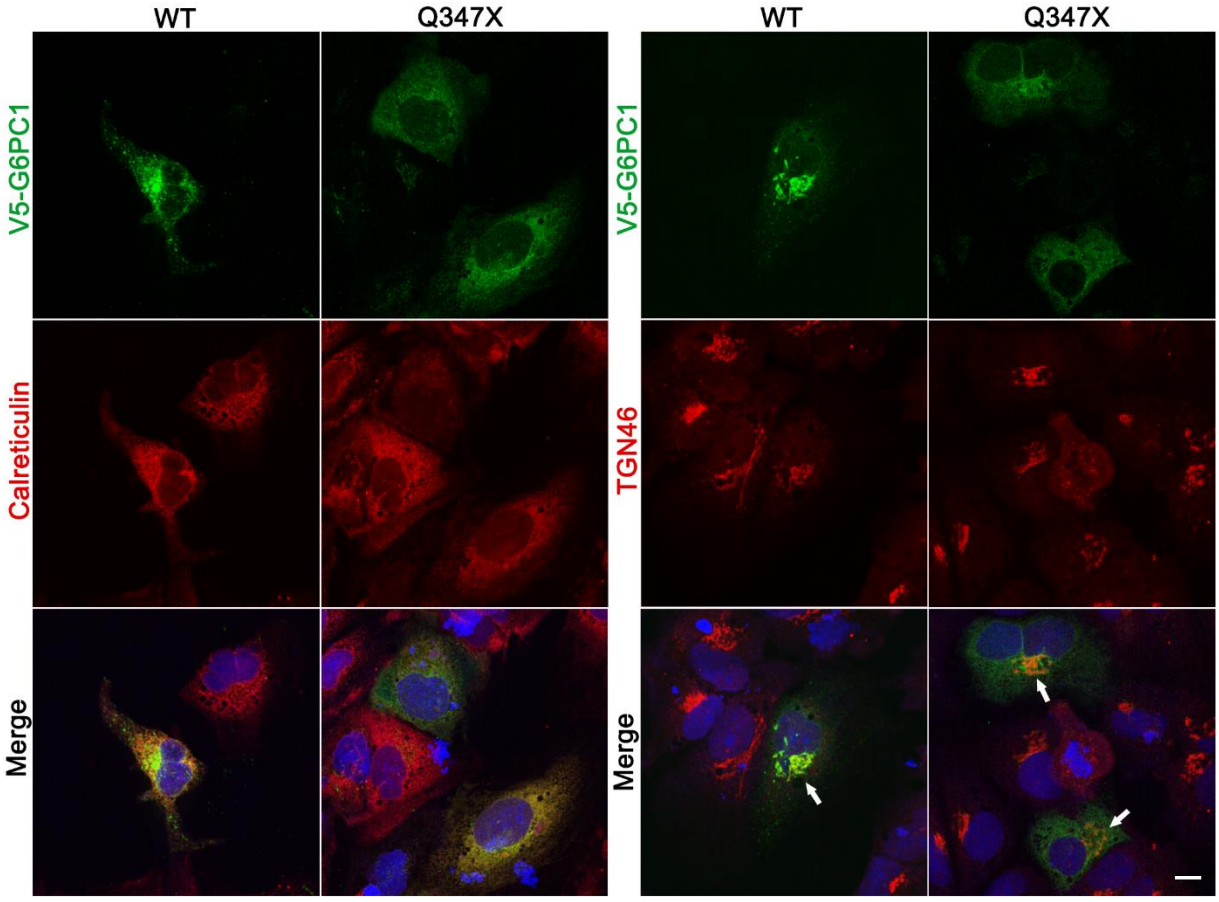


Figure 4.3.3.2. Cellular localisation of wild type G6PC1 and Q347X with respect to the ER (left) and Golgi network (right) in Huh7 cells was assessed by immunofluorescence microscopy. Cells were double immunostained for FLAG tag (green) and markers for the ER, calreticulin, on the left, or trans-golgi network, TGN46, on the right (both in red). Images were merged with staining of the nuclei using DRAQ5 (blue). White arrows point to positions of the golgi apparatus. Scale bar indicates 10 µm.

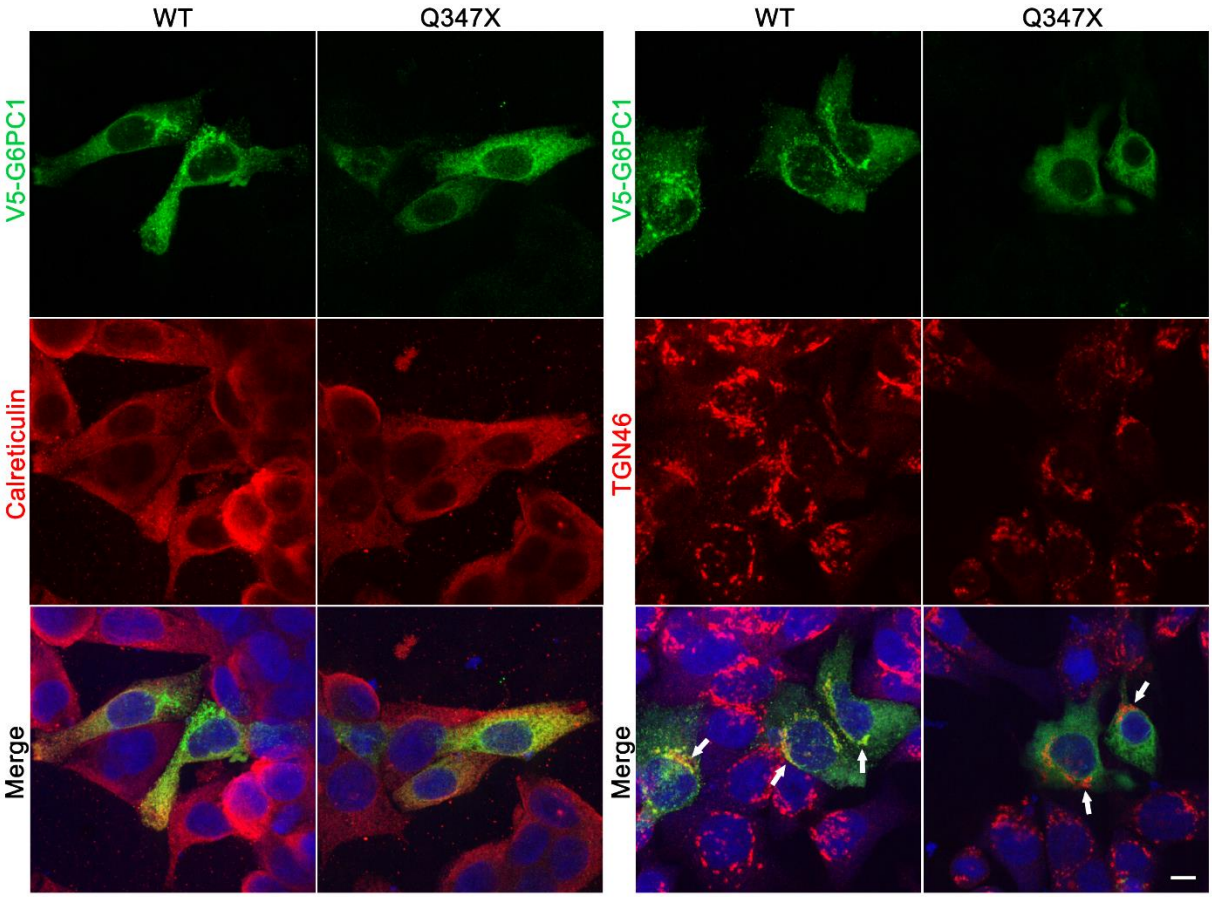


Figure 4.3.3.3. Cellular localisation of wild type G6PC1 and Q347X with respect to the ER (left) and Golgi network (right) in HepG2 cells was assessed by immunofluorescence microscopy. Cells were double immunostained for FLAG tag (green) and markers for the ER, calreticulin, on the left, or trans-golgi network, TGN46, on the right (both in red). Images were merged with staining of the nuclei using DRAQ5 (blue). White arrows point to positions of the golgi apparatus. Scale bar indicates 10 µm.

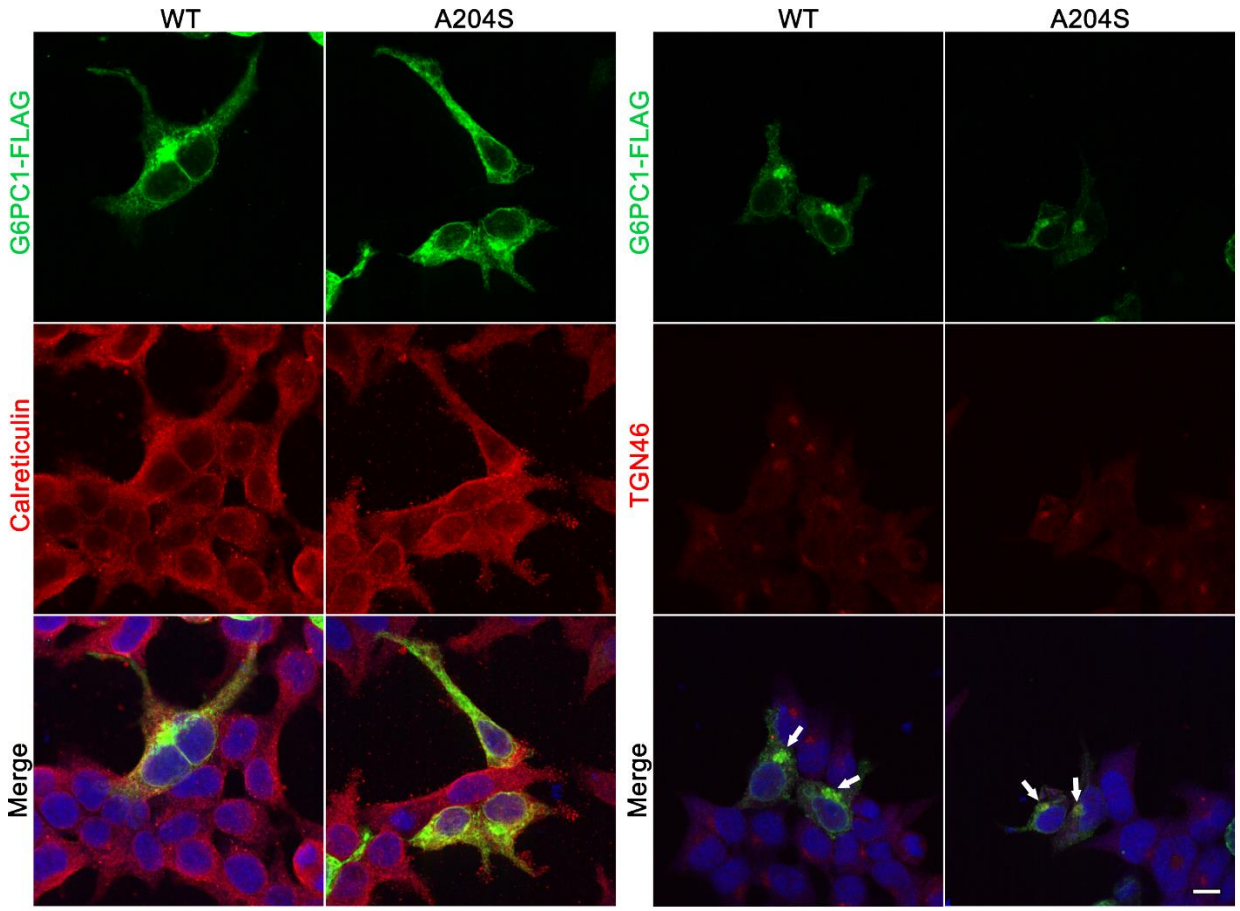


Figure 4.3.3.4. Cellular localisation of wild type G6PC1 and A204S with respect to the ER (left) and Golgi network (right) in HEK293 cells was assessed by immunofluorescence microscopy. Cells were double immunostained for FLAG tag (green) and markers for the ER, calreticulin, on the left, or trans-golgi network, TGN46, on the right (both in red). Images were merged with staining of the nuclei using DRAQ5 (blue). White arrows point to positions of the golgi apparatus. Scale bar indicates 10 µm.

4.3.4. Development of a glucose-6-phosphatase activity assay

Finally, to characterise the enzymatic activities of the G6PC1 variants that have been identified, an assay measuring G6Pase activity was designed based on detection of inorganic phosphate release from the catalytic reaction. The assay was first modified from a liver G6Pase protocol from Sigma Aldrich and downscaled to reduce reagent volumes. HEK293 cells were chosen as the protein expression system due to their high transfection efficiency and rapid growth, which was beneficial for the extraction of sufficient recombinant protein. HEK293 cells were first transiently transfected with G6PC1 WT construct, then mechanically homogenised and subjected to multiple centrifugation steps for the isolation of microsomes. Microsomes, which are small vesicles enclosed by membranes, most of which derive from fragmented ER, were extracted to enrich for G6PC1 protein and to remove other cytosolic enzymes. Western blot analysis of microsomal protein was carried out to confirm and quantify expression of recombinant FLAG-tagged G6PC1. Microsome samples were then added to reaction mixes containing 1 mM up to 40 mM G6P. As summarised in the scheme in Figure 4.3.4.1, the inorganic phosphates liberated from the hydrolytic reaction were quantified by the colorimetric method of Taussky-Shorr (Taussky and Shorr 1953). Colour intensity is directly proportional to phosphate concentration in the solution, which is in turn a measure of G6Pase activity. Reactions were also carried out in the presence of histone 2A proteins, which were reportedly used to improve the measurement of G6Pase activity such that access of enzyme to substrate is no longer a rate-limiting factor. This was either through permeabilisation of microsomal membranes or by changing the conformation of G6Pase in the ER membrane, though the precise mechanisms have not been ascertained (Blair and Burchell 1988; Pederson et al. 1998; St-Denis et al. 1994).

G6PC1 activity was analysed using the Michaelis-Menten model for enzyme kinetics to estimate kinetic constants V_{max} and K_m for a first order enzyme reaction in response to increasing doses of G6P (Freedland and Barnes 1963), as shown in Figure 4.3.4.2. Given the range of enzyme incubation times used in multiple published protocols, I tested several incubation times and found that 8.5 – 13 min reactions provided consistent results in our setup and subsequently followed this parameter in our experiments (Figure 4.3.4.3). The mean R^2 value indicating goodness-of-fit for the model was 0.96 for reactions with and

without histones (Figures 4.3.4.2 and 4.3.4.3). The activity rate detected at 10 mM G6P in our assay was similar to published data which reported 66 - 139 nmol/min/mg for G6PC1 WT (without histones) (Lei et al. 1995a; Shieh et al. 2002); our observations of the K_m of G6PC1 (2.5 – 7.3) were also comparable with the range of 0.4 – 5.2 reported in previous studies of human liver microsomes (Burchell et al. 1988). However, as the large error bars in our activity plots were due to inter-experimental variability, I subsequently normalised all measurements within each experiment to the activity observed for WT at maximum G6P concentration, prior to making comparisons between multiple samples.

Finally, even though HEK293 cells, being of kidney origin, are expected to possess endogenous G6Pase activity, our experiments showed low baseline activity with no increase in response to increasing G6P concentration in untransfected cells. This confirmed HEK293 as a suitable model for detecting G6Pase activity from recombinantly expressed G6PC1.

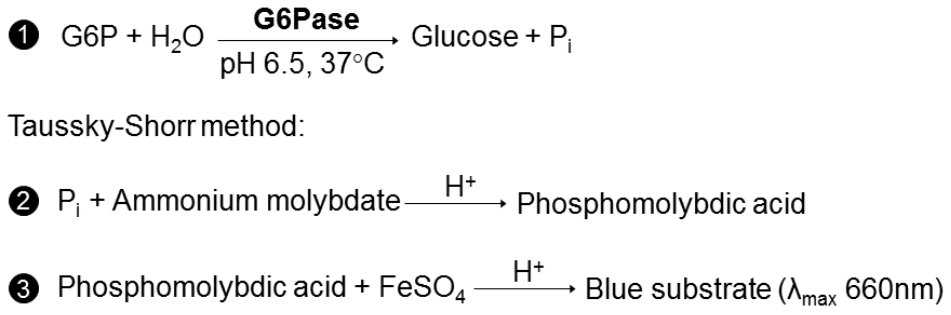


Figure 4.3.4.1. Reaction mechanism for enzymatic determination of glucose-6-phosphatase activity. Pi: inorganic phosphates.

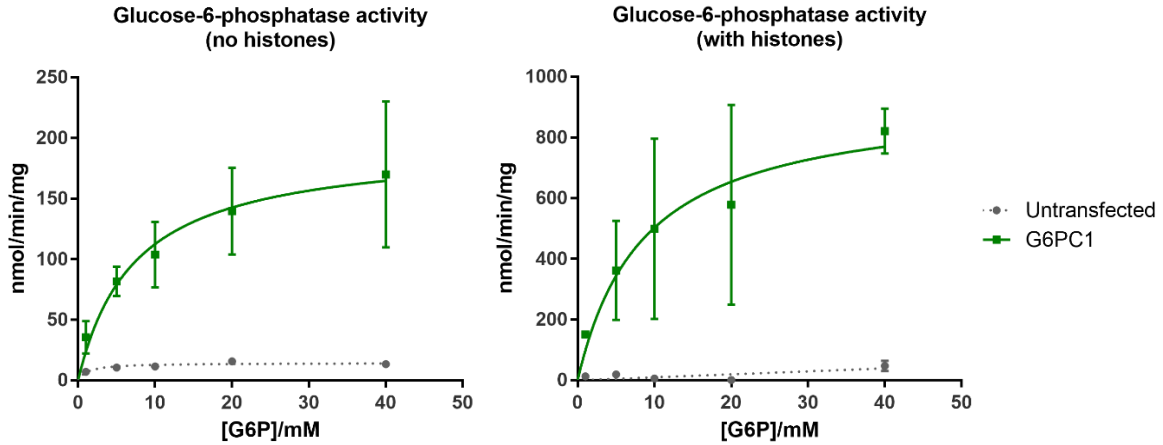


Figure 4.3.4.2. Glucose-6-phosphatase activity of microsomal preparations containing G6PC1 protein in HEK293 cells. Results obtained from two independent experiments (of two technical replicates each) presented as mean values ± SEM, with no normalisation within experimental replicates to show spread of reaction rates.

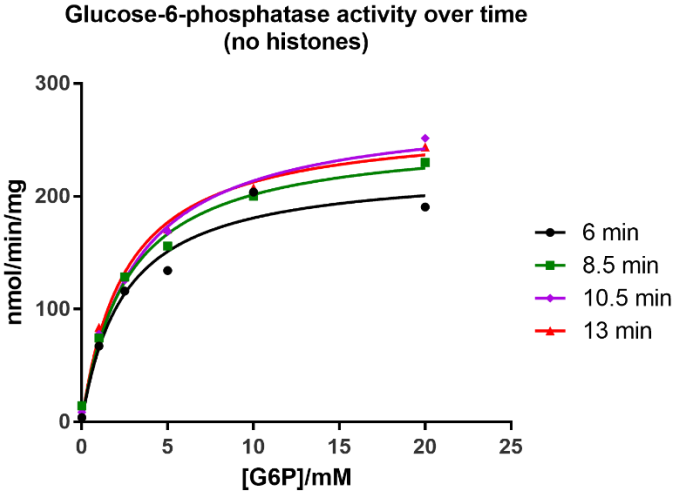


Figure 4.3.4.3. Effect of enzyme incubation time on activity detected. The R^2 value, which indicates the goodness-of-fit for the Michaelis-Menten model, was >0.98 for all plots. For subsequent experiments the enzyme reaction time was kept to 8.5 – 13 min. Results were obtained from one experiment.

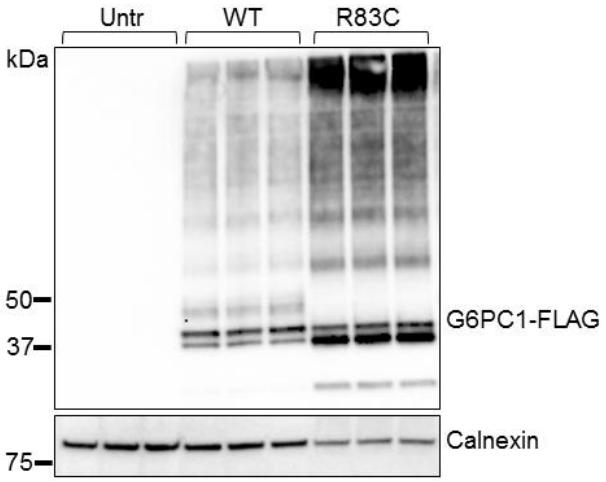
4.3.5. Phosphatase activity of G6PC1 variant proteins associated with FG and FI levels

With the phosphatase assay set up, I tested the three variants found to influence FG and FI levels in the MAGIC dataset – R83C, Q347X and A204S – for G6Pase activity. Both R83C and Q347X have previously been shown to have no detectable phosphatase activity at 10 mM G6P when expressed in COS-1 cells (Lei et al. 1993; Lei et al. 1994). This result was consistent with our activity studies where we observed no activity above background in HEK293 cells for both variants, when tested against a range of G6P concentrations up to 20 mM (Figures 4.3.5.1 and 4.3.5.2). I therefore confirmed that R83C and Q347X are complete LOF variants due to loss of G6Pase activity.

To evaluate if glycosylation is essential for G6Pase activity, unglycosylated protein was isolated from HEK293 cells treated with tunicamycin to inhibit N-linked glycosylation. The representative western blot in Figure 4.3.5.3(A) shows complete loss of the higher molecular weight bands (expected to be glycosylated proteins) in tunicamycin-treated cells. When quantified, total protein levels were also reduced compared to WT in untreated cells. When evaluated in phosphatase activity assays matched for protein input, unglycosylated WT protein appeared to retain some residual G6Pase activity, but its activity was downregulated by 10-14% at high G6P concentrations (10 and 20 mM), as shown in Figure 4.3.5.3(B). The kinetic constant V_{max} of unglycosylated WT was lower than that of total WT protein, though the difference was not significant. Previous studies by Shieh and colleagues reported that unglycosylated G6PC1 (obtained by mutating the N96 glycosylation site) possessed 37% of WT enzymatic activity (Pan et al. 1998a). Therefore, I confirmed that G6PC1 glycosylation is important, but not a requisite for functional activity.

A204S appears to possess G6Pase activity in our assays (data not shown), but as variant protein expression levels in the microsomal fraction were significantly reduced by 41% compared to WT ($P = 0.0107$) as seen in Figure 4.3.5.4, activity levels could not be accurately compared. Kinetic data for this variant is therefore not available. In this case, A204S is most likely exhibiting partial LOF as a result of reduced protein expression.

A



B

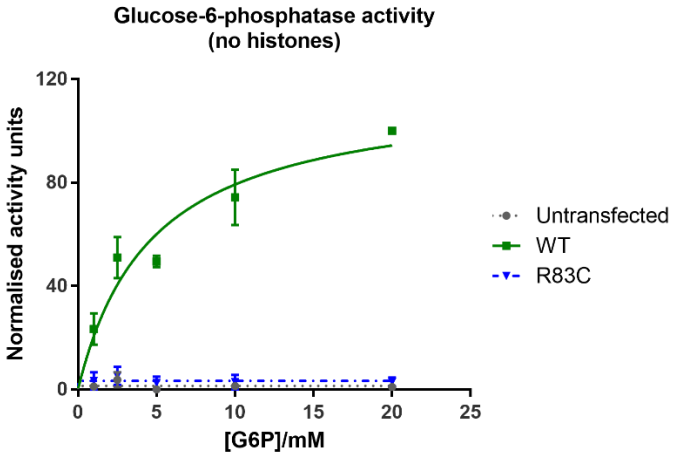
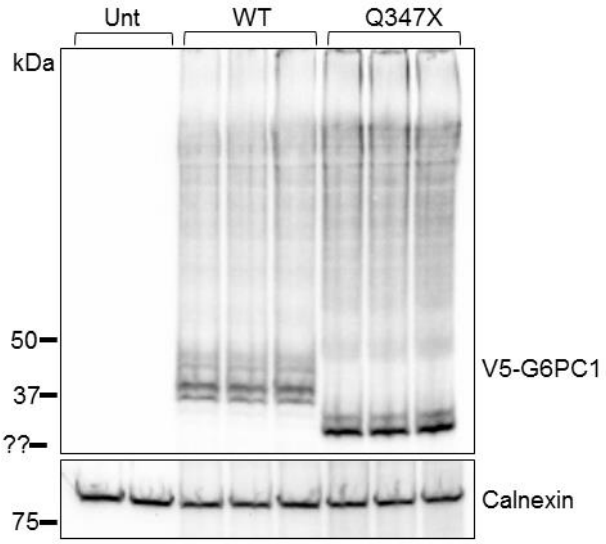


Figure 4.3.5.1. Glucose-6-phosphatase activity in microsomal preparations of G6Pase-R83C expressed in HEK293 cells. (A) Representative western blot of R83C in extracted microsomes with calnexin as the loading control. (B) Michaelis-Menten plot for G6Pase activity with varying concentrations of G6P. Results were from three individual experiments presented as mean values normalised to activity of WT at 20 mM G6P ± SEM. Untr: Untransfected control; WT: Wild type.

A



B

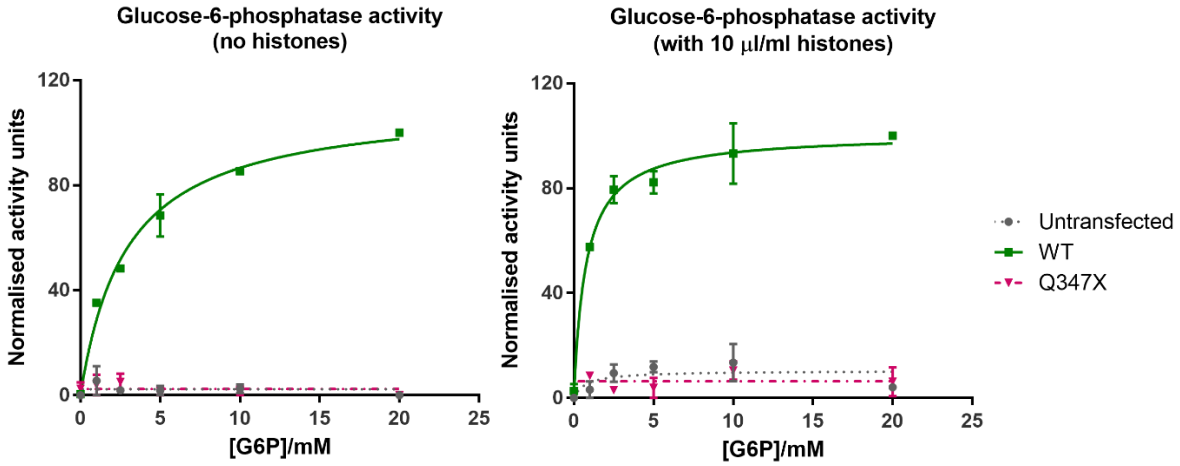
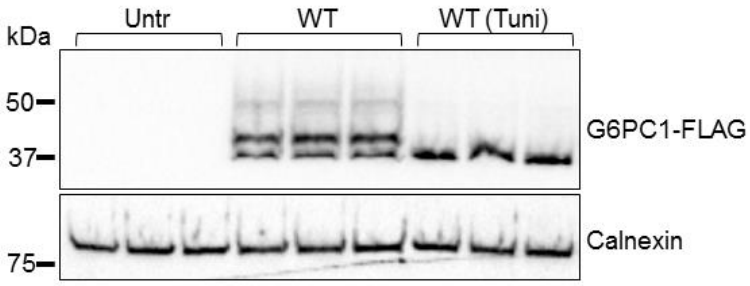
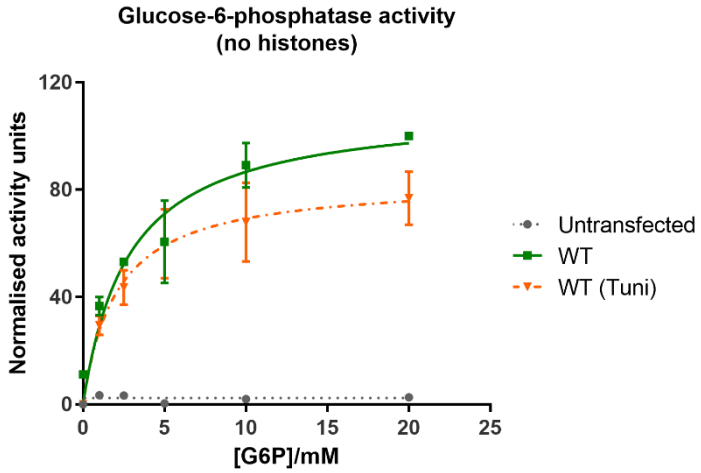


Figure 4.3.5.2. Glucose-6-phosphatase activity in microsomal preparations of G6Pase-Q347X expressed in HEK293 cells. (A) Representative western blot of Q347X in extracted microsomes with calnexin as the loading control. (B) Michaelis-Menten plots for G6Pase activity with varying concentrations of G6P. Results were from two individual experiments presented as mean values normalised to activity of WT at 20 mM G6P ± SEM. Unt: Untransfected control; WT: Wild type.

A



B



G6Pase	R^2	Vmax (SE)	P	Km (SE)	P
WT	0.88	111.10 (10.69)	0.19	2.82 (0.89)	0.59
WT (Tuni)	0.84	83.50 (9.02)		2.06 (0.82)	

Figure 4.3.5.3. Glucose-6-phosphatase activity in microsomal preparations of unglycosylated G6Pase-WT expressed in HEK293 cells. (A) Representative western blot of WT in microsomes from untreated and tunicamycin-treated cells with calnexin as the loading control. (B) Michaelis-Menten plot for G6Pase activity with varying concentrations of G6P. Results were from two individual experiments presented as mean values normalised to activity of WT at 20 mM G6P ± SEM. Unpaired t tests of the kinetic constants V_{max} and K_m were carried out to compare WT and treatment. R^2 indicates the goodness-of-fit for the non-linear regression. Untr: Untransfected control; WT: Wild type; Tuni: Tunicamycin-treated.

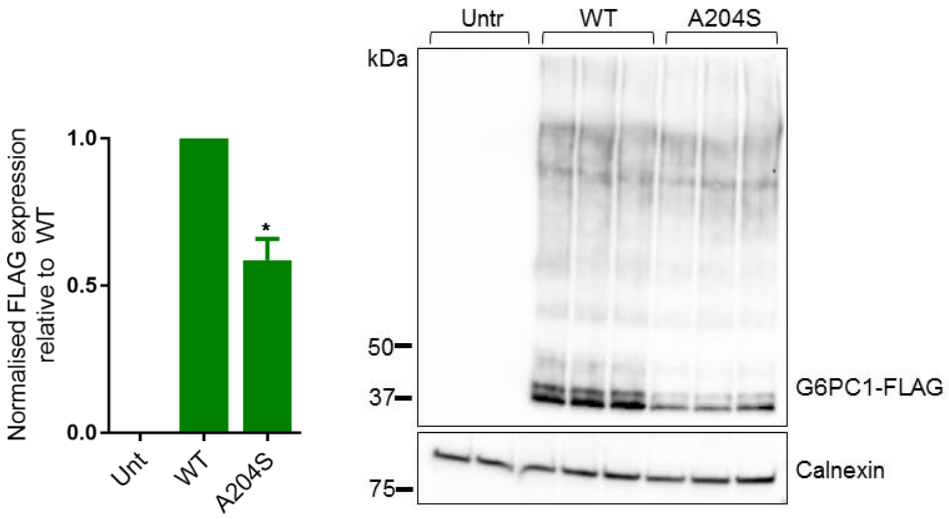


Figure 4.3.5.4. Expression levels of the A204S variant in microsomes isolated from HEK293 cells were determined by western blot analysis. Densitometric data of FLAG-tagged A204S relative to calnexin control in extracted microsomal protein were obtained from four independent experiments (left). Data presented as mean ± SEM and analysed using paired Students' t tests. *P=0.01. Representative western blot also shown (right). Unt: Untransfected; WT: Wild type.

4.4. Discussion

My functional studies together with large-scale genetic analyses for glycemic traits show direct support for *G6PC1* as a novel effector transcript influencing FG and FI levels, based on the identification of novel functional coding variants contributing to a gene-based association signal. None of the non-coding variants in the region that were tested were associated with glycemic traits. *In vitro* follow-up of rare non-synonymous coding variants included in the gene-based tests confirmed that A204S, R83C and Q347X, which exhibited loss of protein stability and/or activity, were indeed causal LOF variants at this locus. A204S displayed significantly reduced total protein expression in HEK293 and Huh7 cells, and the decrease was even more pronounced when variant protein levels were measured in the ER membrane-containing microsomal fraction. Earlier studies have shown that degradation of G6PC1 is mediated by proteasomes (Shieh et al. 2002), as I have also described for G6PC2 which is of the same protein family (Mahajan et al. 2015).

Both R83C and Q347X have been previously reported to be causal variants in the rare autosomal recessive disorder GSD1a and were the most highly prevalent mutations identified in cases in Caucasian and Jewish populations (Chou and Mansfield 2008). R83C not only localises to the active site but also resulted in complete loss of glycosylation in all cell lines tested. The precise mechanism for this is unclear since the only known glycosylation site is at residue N96. However, as both residues are on the luminal face of the ER membrane it is possible that R83 may play a role in stabilising the multi-step enzyme-mediated glycosylation process, which may in turn modulate overall enzymatic activity. I and others have showed that unglycosylated G6PC1 have reduced protein stability and that expressed protein retained residual, but reduced, G6Pase activity.

In addition to R83C, the truncating variant Q347X was also confirmed to have null phosphatase activity in our assay. The truncation resulted in Q347X losing the C-terminus consensus ER retention signal. Previous work had shown that the KK motif was neither crucial for ER residence nor phosphohydrolase activity of G6PC1, and that the loss in activity for Q347X may be instead explained by one or more amino acids within residues 347-354 (Lei et al. 1994). In my experiments, Q347X was indeed detected in cell microsomes and hence likely to remain associated with the ER membrane. Immunostaining however showed

that Q347X was not co-localised with the Golgi network unlike WT protein. This is also the first time that G6PC1 has been shown to be partially localised to the Golgi complex, an organelle primarily responsible for modifying, packaging and sorting newly-synthesised proteins to different cellular compartments. The nucleus, ER and Golgi complex are highly linked membrane-bound compartments that control protein trafficking along the secretory pathway. It is possible that G6PC1 is post-translationally modified within the Golgi network, before undergoing retrograde transport back to the ER mediated by signalling motifs and adaptor proteins that facilitate ER-Golgi interactions (Fujiwara et al. 2003; Tekirian 2002).

Overall, the genetic data and my variant characterisation results together revealed that rare disease-causing mutations in *G6PC1* exist in the heterozygous state to modulate fasting glycemic traits in normoglycemic asymptomatic individuals. A204S has a relatively milder effect on FG and FI compared to R83C and Q347X, consistent with its smaller impact on protein function. It is however the most prevalent variant among all the variants identified (an order of magnitude higher), therefore the power to detect trait associations is higher compared to rarer variants. In light of the critical role that this gene plays in glycemic regulation and the presence of multiple rare variants that alter protein function, it is possible that additional rare G6PC1 variants such as P315A may also influence FG and FI levels. P315A was consistently predicted as a deleterious variant by *in silico* algorithms, and variant protein levels were downregulated in the liver cell lines but not in HEK293 cells (Figure 4.3.2.3). Comprehensive *in vitro* assessment of all variants including P315A on phosphatase activity and possibly other functional readouts will be required to ascribe functional significance to the variants.

Despite the improved understanding of the genetic etiology of GSD1a facilitated by a comprehensive *G6PC1* mutation database, there is lack of clear genotype-phenotype correlation for most of the mutations identified, considerable phenotypic heterogeneity as well as little mechanistic insight into the development of secondary complications. There are also few alternatives to the current standard of care which are largely based on dietary therapies to correct metabolic abnormalities, but do not address the upstream pathological processes (Chou et al. 2010). Future studies will need to focus on studying the physiological mechanisms underlying disease such as using targeted knockout animal models to study the specific role of *G6pc1* in different tissues where it is expressed and to isolate organ-specific

effects (Rajas et al. 2015). Furthermore, it is important to study G6PC1 in the context of other interacting partners such as G6PT and other transporters that form the G6Pase complex. In fact, study of the ubiquitously expressed *G6PC3* may also provide further insight given that *G6PC3* deficiency gives rise to myeloid dysfunctions that are also common to other GSD sub-types such as GSD1b (Chou et al. 2015).

Chou and colleagues have led work on correction of GSD1a in G6Pase-deficient mice models using gene therapy. They showed that sustained hepatic and renal expression of the *G6pc1* transgene in these mice was able to reverse the underlying metabolic abnormalities and prevent disease progression (Ghosh et al. 2006; Yiu et al. 2010; Zingone et al. 2000). However, these studies also showed that only 11% of normal hepatic *G6pc1* activity was sufficient to maintain glucose homeostasis. Though still very much at the experimental stage, such strategies appear to be promising in the case of potential *G6PC1*-based gene therapy for GSD1a in humans, but less so for *G6PC1* as a therapeutic target for the modulation of fasting glycemic traits in normoglycemic individuals.

Chapter 5

Comprehensive identification and functional characterisation of *G6PC2* coding variants

---

The results presented in this chapter will form part of a MAGIC manuscript which is currently in preparation.

5.1. Introduction

5.1.1. Identification of rare coding variants in *G6PC2* influencing glycemic traits from large-scale exome array meta-analysis

The MAGIC exome array meta-analysis that was described in Chapter 4 provided the opportunity to identify rare, potentially causal, coding variants that influence glycemic traits. In addition to novel gene-based association signals that highlighted *G6PC1*, the study also discovered novel coding variant signals at both novel loci and loci previously linked to other glycemic traits or T2D. The analysis of approximately 144K non-diabetic individuals, the largest analysis of its kind, thereby greatly enhanced the power to detect associations and extended the discovery of rare coding variants (MAF<1%) in the known FG-associated locus *G6PC2*.

In *G6PC2*, associations with FG were replicated for five coding variants – common variant V219L and rare variants H177Y, Y207S, R283X and S324P – that were identified in previous studies of smaller scale (Mahajan et al. 2015; Wessel et al. 2015). Several of these (H177Y, Y207S and R283X) are novel coding variant associations for HbA1c levels in a directionally-consistent manner at exome-wide significance ($P < 2.21 \times 10^{-7}$), as shown in Table 5.1.1.1. Strong gene-based association signals comprising up to 18 rare variants for both FG and HbA1c were also observed at *G6PC2*. These were significant for all the variant masks tested ($P < 2.5 \times 10^{-6}$, listed in Table 5.1.1.2). Among these, H177Y, Y207S, R283X and S324P were again shown to be the main drivers of the gene-based signals based on conditional analysis. The significance of these variants was further confirmed when the strongest gene-based association signal for FG was observed for the variant mask consisting of only the four statistical drivers (Table 5.1.1.2).

Gene-based association analyses are frequently used in recent large-scale association studies as they consider the contribution of aggregated sets of variants within a locus to trait variance, since rarer variants are less likely to achieve statistical significance on their own due to extremely low MAF. The burden test assumes that all causal variants have the same magnitude and direction of effect, and therefore does not perform well in the presence of neutral variants or variants that have an opposite direction of effect. This is in contrast to the SKAT test, which accommodates the existence of deleterious, neutral and

protective variants in a gene (Basu and Pan 2011; Magi et al. 2011; Wu et al. 2011). The analyses were carried out on different variant masks that attempt to select for likely causal variants using *in silico* prediction algorithms (described in Chapter 4) to strengthen the association signal (Purcell et al. 2014). The results from these analyses shown in Table 5.1.1.2 provided a number of insights. First, both the NSbroad and NSstrict masks which included a large number of variants were strongly significant, indicating that additional rare variants with unestablished single variant level significance are likely to contribute to the aggregate signal at this gene. Second, the NSstrict mask performed better than the NSbroad mask for both FG and HbA1c, though this was observed only in the case of the burden test; the reverse was true for the SKAT test. This not only demonstrated that the NSsrict mask was successful in enriching for functional variants, but also implied that the NSbroad mask may include both neutral variants and variants that exert contrasting direction of effects. Several rare variants in *G6PC2* that display nominal significance ($P<0.05$) for association with FG and/or HbA1c that were not classified under NSstrict, may still be of functional importance and influence FG and HbA1c levels.

The rare coding variant association results derived from the MAGIC analysis provide an opportunity for investigating naturally-occurring mutations in humans that may affect G6PC2 function using *in vitro* validation pipelines. More pertinently, as functional rare coding variants tend to have larger effects on phenotype, they facilitate the understanding of the spectrum of molecular effects that can impact on protein function (from gain of function to loss of function) and how these relate to modulation of complex human traits, in this case FG and HbA1c levels. Such understanding becomes crucial for establishing the dose-dependent relationship between protein function and biological phenotype, which is important for the validation of G6PC2 as a potential therapeutic target (Plenge et al. 2013).

Variant SNP ID	Protein change	Alleles (effect/other)	EAF_{FG}	$Effect_{FG}$	P_{FG}	EAF_{HbA1c}	$Effect_{HbA1c}$	P_{HbA1c}	NSbroad	NSstrict
rs2232323	Y207S	C/A	6.31×10^{-3}	-0.129	**1.05×10^{-28}**	7.18×10^{-3}	-0.053	**3.25×10^{-13}**	1	1
rs138726309	H177Y	T/C	4.82×10^{-3}	-0.129	**1.70×10^{-25}**	3.37×10^{-3}	-0.062	**3.27×10^{-10}**	1	1
rs492594	V219L	C/G	0.451	0.016	**2.20×10^{-16}**	4.55×10^{-1}	0.003	2.91×10^{-2}	0	0
rs2232326	S324P	C/T	1.67×10^{-3}	-0.167	**7.70×10^{-15}**	2.75×10^{-3}	-0.039	5.23×10^{-4}	1	1
rs146779637	R283X	T/C	2.36×10^{-3}	-0.138	**1.78×10^{-12}**	2.34×10^{-3}	-0.074	**4.58×10^{-10}**	1	1
rs142189264	S30F	T/C	4.42×10^{-4}	-0.159	5.30×10^{-5}	5.18×10^{-4}	-0.084	1.35×10^{-3}	1	0
rs2232322	I171V	G/A	5.20×10^{-4}	0.131	5.68×10^{-3}	4.55×10^{-4}	0.093	1.72×10^{-2}	1	0
rs145050507	I171T	C/T	8.67×10^{-4}	-0.084	8.33×10^{-3}	6.05×10^{-4}	-0.007	9.26×10^{-1}	1	0
rs150538801	F256L	C/T	4.67×10^{-4}	-0.078	4.78×10^{-2}	3.99×10^{-4}	0.006	7.92×10^{-1}	1	0
rs137857125	P313L	T/C	1.66×10^{-5}	-0.510	4.79×10^{-2}	1.57×10^{-5}	-0.253	2.30×10^{-1}	1	0
rs199682245	N68I	T/A	1.02×10^{-4}	-0.174	9.67×10^{-2}	8.93×10^{-5}	-0.089	1.83×10^{-1}	1	1
rs201561079	I63T	C/T	1.66×10^{-4}	0.133	1.16×10^{-1}	1.53×10^{-4}	-0.042	4.67×10^{-1}	1	0
rs147360987	H250Y	T/C	7.04×10^{-5}	0.239	1.31×10^{-1}	5.10×10^{-5}	0.132	2.66×10^{-1}	1	0
rs184807114	A119V	T/C	1.21×10^{-4}	-0.103	2.01×10^{-1}	9.44×10^{-5}	-0.007	8.73×10^{-1}	1	0
rs149874491	I38L	C/A	2.09×10^{-4}	-0.077	3.83×10^{-1}	1.98×10^{-4}	-0.081	8.38×10^{-2}	1	0
rs187707963	Y124C	G/A	1.77×10^{-5}	0.247	5.19×10^{-1}	8.21×10^{-6}	-0.161	5.33×10^{-1}	1	1
rs200336133	L310F	T/C	2.40×10^{-4}	-0.032	5.94×10^{-1}	4.71×10^{-5}	0.075	4.65×10^{-1}	1	0
rs145217135	I230T	C/T	8.70×10^{-5}	-0.047	7.40×10^{-1}	7.47×10^{-5}	-0.036	4.36×10^{-1}	1	0
rs148689354	I273V	G/A	2.15×10^{-4}	-0.012	8.10×10^{-1}	1.45×10^{-4}	0.004	9.43×10^{-1}	1	0

Table 5.1.1.1: FG and HbA1c single variant association results for common variant V219L and all *G6PC2* variants included in the gene-based analyses from MAGIC. Variants are arranged in descending order of level of significance for FG. All variants are missense except for one PTV (R283X). Allele changes are reported for the full-length transcript NM_021176. EAF: effect allele frequency. Effect sizes are reported as mmol/l and % for FG and HbA1c respectively. Variants that are associated at exome-wide significance are highlighted in bold ($P<2.21 \times 10^{-7}$). For NSbroad and NSstrict classifications, 1 indicates inclusion in the mask while 0 indicates exclusion. Data provided by Sara Willems, Hidetoshi Kitajima, Xueling Sim on behalf of the MAGIC investigators.

Trait	Mask	No. of variants	Gene effect	P_{burden}	P_{SKAT}
FG	NSbroad	18	-0.12	4.09×10^{-67}	5.38×10^{-58}
	NSstrict	7	-0.11	7.80×10^{-69}	3.83×10^{-56}
	Statistical drivers	4	-0.22	**2.97×10^{-125}**	**6.42×10^{-94}**
HbA1c	NSbroad	18	-0.059	6.18×10^{-30}	**4.65×10^{-27}**
	NSstrict	7	-0.053	1.04×10^{-31}	1.92×10^{-26}
	Statistical drivers	4	-0.059	**5.65×10^{-32}**	1.95×10^{-26}

Table 5.1.1.2: *G6PC2* gene-based associations with FG and HbA1c. NSbroad: PTV and missense variants predicted to be deleterious by at least one of five annotation algorithms (SIFT, Polyphen2-HumDiv, PolyPhen2-HumVar, LRT and MutationTaster); NSstrict: PTV and missense variants predicted to be deleterious by all five annotation algorithms; Statistical drivers: Variants found to be driving the gene-based association (P>0.05 after conditioning on these variants). Gene effect results were obtained from the burden test analysis and are reported as mmol/l and % for FG and HbA1c respectively. P values reflected were obtained from the burden test or SKAT analysis. Most significant P values for each test per trait are highlighted in bold. Data provided by Sara Willems, Hidetoshi Kitajima, Xueling Sim on behalf of the MAGIC investigators.

5.1.2. Functional investigations of *G6PC2* variants associated with FG

To date, the common GWAS SNP rs560887 in intron 3 of *G6PC2* remains the strongest genetic determinant for FG in terms of level of significance, and accounts for 0.9% out of 1.1% of the total trait variance at this locus (Dupuis et al. 2010; Mahajan et al. 2015). The only functional analysis that has been performed on this SNP, which is located close to the 5' end of exon 4, suggested that the rs560887-G allele altered RNA splicing to promote the inclusion of exon 4 giving rise to expression of the full-length isoform, based on minigene assays in HeLa cells (Baerenwald et al. 2013). This is supported by human islet RNA sequencing data (n=174, unpublished) which showed that the rs560887-G allele was significantly associated with increased expression of the full-length *G6PC2* isoform (p=0.001) in variant carriers, confirming rs560887 as an eQTL (Martijn van de Bunt, personal communication). The results are consistent with the association between rs560887-G and elevated FG levels presumably due to increased *G6PC2* function. In light of recent studies, the effect size on FG of rs560887 is also relatively small (β=0.08 mmol/l per allele) compared to that observed for the rare coding variants (Table 5.1.1.1). As coding variants exert a direct impact on protein function, establishing a causative link between coding variant effects and FG variation, interpretation of their effects is likely to be more straightforward than that of regulatory variants.

A number of *G6PC2* coding variants in the MAGIC analysis have already been characterised previously. *In vitro* characterisations of H177Y, Y207S and V219L, which were presented in Chapter 3 and published in Mahajan et al. 2015 based on a smaller exome chip study of ~33k non-diabetic individuals, revealed for the first time that coding variants in *G6PC2* influence FG levels due to degradation of unstable variant proteins by the ubiquitin-proteasome pathway in both HEK293 and INS-1 cells. Notably, the expression of H177Y and Y207S variant proteins were almost completely lost, revealing a surprisingly large impact of missense substitutions at these sites on protein function.

A recently published study also showed that expression of H177Y, Y207S, and additional missense variants P313L and S324P were downregulated when expressed in COS-7 cells (Boortz et al. 2016a). The authors however did not observe any notable effects on expression of the common variant V219L in this cell line. The same study further analysed

the effect of a series of amino acid changes in human G6PC2 that were at residues conserved between human and mouse G6PC2 and G6PC1 proteins. They identified a number of variants, including S30F, N68I and F256L that markedly reduced mouse G6pc1 activity using an *in situ* luciferase-based reporter assay, and others that had no impact on either protein expression or activity. The study had provided a reference set of G6PC2 variants at highly conserved regions that conferred reduced expression or reduced enzymatic activity. However, the greatest caveat of this study was that variant activity was studied in the context of mouse G6pc1, which is both of a different species and encoded by a different gene though part of the same gene family, to circumvent experimental difficulties.

The only PTV highlighted in the MAGIC analysis, R283X (rs14677963), was previously shown by the CHARGE consortium to be strongly linked to FG levels (Wessel et al. 2015). However, this signal was not replicated by the T2D-GENES/GoT2D consortia (Mahajan et al. 2015), resulting in a discrepancy regarding the significance of this variant. As the minor allele for this variant introduces a stop mutation in the last exon of the gene, I initially hypothesised that the resulting transcript could evade nonsense-mediated decay and be expressed in variant carriers at the protein level (as is expected for PTV in the last exon). We confirmed this in unpublished human islet RNA sequencing data which showed no allelic imbalance between allele carriers and non-carriers (Martijn van de Bunt, personal communication). Though the sample size was small (only 3 heterozygous variant carriers in 174 samples), this result showed that the variant is indeed expressed and that functional relevance of this variant would have to be ascertained through *in vitro* follow-up studies. The impact of the truncation of 71 amino acids at the C-terminal end which results in the loss of the consensus ER retention signal also remains to be determined.

5.1.3. Cataloguing of coding variation in *G6PC2* using exome sequencing data

Since *G6PC2* has been established as an effector transcript for FG and multiple bodies of work have supported its critical role in glucose homeostasis (Mahajan et al. 2015; O'Brien 2013), there is interest in more fully defining the complete spectrum of alleles (from common to rare) in sequencing studies to explore the possibility of additional associations especially implicating previously untested rare variants. Specifically, exome sequencing data contain an abundance of rare coding variation, which in turn contains a higher proportion of functionally-significant variants that may have a wide range of effects on protein function mediated by both deleterious and protective alleles.

The current database of coding variation in *G6PC2* was expanded with the T2D-GENES/GoT2D exome sequencing efforts, which enabled us to extend the allelic series to facilitate more comprehensive studies of *G6PC2* variants. Whole exome sequencing data were collected from 12,940 individuals from five major ancestries (African American, East Asian, Hispanic, South Asian, European) that consisted of both T2D cases and controls. The dataset (now abbreviated as 13k exomes) revealed a total of 69 missense and nonsense variants in *G6PC2*, almost doubling the total number of variants reported by the CHARGE consortium in up to 7,452 individuals (Wessel et al. 2015). The full list of coding variants was provided by Anubha Mahajan from the T2D-GENES/GoT2D consortia and described in Table 5.3.1.1. The identification of such a comprehensive catalogue of coding variation in *G6PC2* will facilitate the investigation of coding variants at biologically interesting sites that localise to functional domains. One may expect these to represent the extremes of protein dysfunction.

Glycemic trait association analysis was also carried out on these samples by combining the 5,132 whole exome sequence data from controls (without diabetes) with the published exome array data from ~33k non-diabetic European individuals (Mahajan et al. 2015). The results from the combined analysis (unpublished) will not be discussed in this chapter, but the coding variant association signals identified in *G6PC2* were consistent with that from the MAGIC analysis, which was a much larger study.

5.1.4. Experimental aims

Based on association analysis results for glycemic traits in large exome array and sequencing datasets, we hypothesise that additional variants of lower frequency could also affect *G6PC2* function and modulate FG levels. *G6PC2* remains a promising locus to follow up on given the clues to its biological significance that have been informed by studies of human genetic and rodent models for G6pc2. I therefore aim to extend the investigation of coding variation in this gene which could harbour a series of causal alleles to examine the relationship between protein dysfunction and their phenotypic outcomes. The evidence from these studies will increase the understanding of the ability to therapeutically modulate the biological pathway involving G6PC2 to regulate glycemic levels.

To do this I set out to functionally characterise *G6PC2* variants across the allelic frequency spectrum using a suite of experimental pipelines:

1. I performed *in silico* functional analyses on all non-synonymous variants from exome array and sequencing data using computational tools that can functionally annotate large numbers of variants
2. I prioritised a subset of variants that were significantly or marginally associated with FG and HbA1c levels (single variant $P<0.05$) in the MAGIC analysis for *in vitro* follow-up to evaluate protein expression, localisation and activity, to identify potential causal variants
3. Using the above pipelines I characterised additional variants located at predicted functionally-important domains such as the catalytic sites to assess the molecular impact of variants at these sites

5.2. Methods

5.2.1. In silico mutation analysis and online databases

Four *in silico* prediction tools SIFT, PolyPhen2, Condel and GERP were used to evaluate the evolutionary conservation of identified *G6PC2* variants and their potential impacts on protein function. These methods were described in Chapter 2 section 2.13.

5.2.2. Generation of *G6PC2* variant constructs

For the *G6PC2* variants that overlap between this chapter and Chapter 3, the same plasmid constructs were used as in Chapter 3. Additional *G6PC2* variants to be studied *in vitro* were selected from the T2D-GENES/GoT2D 13k exome sequencing study and MAGIC exome array meta-analysis. Missense variants were generated by site directed mutagenesis of the pCMV6-Entry-*G6PC2* vector backbone with C-terminal Myc and FLAG tags following the steps stated in Section 2.5. Constructs for protein-truncating variants were generated by mutagenesis of the cloned pCMV6-Entry-*G6PC2* vector with additional N-terminal V5 tag that was described in Section 2.2. A number of variants were generated in the pCMV6-Entry-*G6PC1* vector for the activity assays, some of which required multiple nucleotide substitutions. The primers designed and used for the PCR are listed in Table 5.2.2.1. All generated constructs were verified by DNA sequencing as in Section 2.4.4 to confirm that only the desired nucleotide changes were introduced.

G6PC2 variant	Forward primer (5’ to 3’)	Reverse primer (5’ to 3’)
S30F	TACTACACTTTTCTAAATTTTATGT**T**CAATGTTGGAGACCCCAGG	CCTGGGGTCTCCAACATTG**A**ACATAAAATTTAGAAAAGTGTAGTA
F71V	CATTGGGGATTGGTTAAATCTTATA**G**TTAAATGGATATTATTTGGTCATCG	CGATGACCAAATAATATCCATTT**C**TATAAGATTTAACCAATCCCCAATG
W73R	AGTCATTGGGGATTGGTTAAATCTTATATTTAAA**C**GGATATTATTTGGTCATC	GATGACCAAATAATAT**G**TTTAAATATAAGATTTAACCAATCCCCAATGACT
R79Q	GGATATTATTTGGTCATC**A**ACCTTACTGGTGGGTCC	GGACCCACCAGTAAGGT**T**GATGACCAAATAATATCC
P80S	GATATTATTTGGTCATCGA**T**CTTACTGGTGGGTCCAAGA	TCTTGGACCCACCAGTAAG**A**TCGATGACCAAATAATATC
G114R	ACAGGTCCAGGAAGTCCATCT**C**GCCATGCAATG	CATTGCATGGC**G**AGATGGACTTCCTGGACCTGT
A119V	GCCATGCAATGGGCG**T**ATCCTGTGTCTGGTA	TACCAGACACAGGAT**A**CGCCCATTGCATGGC
I171V	CTGCATCTCCAGAGTATTC**G**TAGCAACACATTTTCCTCA	TGAGGAAAATGTGTTGCT**A**CGAATACTCTGGAGATGCAG
I171L	CTGCATCTCCAGAGTATTC**C**TAGCAACACATTTTCCTCA	TGAGGAAAATGTGTTGCTAGGAATACTCTGGAGATGCAG
H174P	CTCCAGAGTATTCATAGCAACAC**C**TTTCCTCATCAAGTTATTCTTG	CAAGAATAACTTGATGAG**G**AAAAGGTGTTGCTATGAATACTCTGGAG
F256L	CATTGACACCACGCCT**C**TTGCTGGACTCGTGAG	CTCACGAGTCCAGCAA**G**AGGCGTGGTGTCAATG
P313L	CCATTTCCTCCAGATCC**T**GACTCACGAAGAGCATT	AATGCTCTTCGTGAGTC**A**GGATCTGGAGGAAATGG
R283X	GTTCCTCCTGAGCTGC**T**GAGGGGGAAATAACTA	TAGTTATTTCCCCCTC**A**GCAGCTCAGGAGGAAC
G6PC1 variant	**Forward primer (5’ to 3’)**	**Reverse primer (5’ to 3’)**
L173I	TGTCTGTCACGAATCTAC**A**TTGCTGCTCATTTTCCTC	GAGGAAAATGAGCAGCAA**T**GTAGATTCGTGACAGACA
L173T	TCTGTCTGTCACGAATCTAC**AC**TGCTGCTCATTTTCCTCATC	GATGAGGAAAATGAGCAGC**AGT**GTAGATTCGTGACAGACAGA
L173V	GTCTGTCACGAATCTAC**G**TTGCTGCTCATTTTCCT	AGGAAAATGAGCAGCAA**C**GTAGATTCGTGACAGAC
F258L	TTGACACCACACCC**C**TTGCCAGCCTCCTC	GAGGAGGCTGGCAA**G**GGGTGTGGTGTCAA
R281X	CAACTCCAGCATGTAC**TA**GGAGAGCTGCAAGGGG	CCCCTTGCAGCTCTCC**TA**GTACATGCTGGAGTTG

Table 5.2.2.1. Primer sequences used for SDM. The forward primer sequence is representative of the sense strand. The letter in bold **red** denotes the nucleotide that was newly incorporated into the DNA sequence during mutagenesis; the underlined triplet denotes the codon that was affected.

5.2.3. Protein expression analysis

HEK293 and INS-1 (832/13) cells were transfected with each WT or mutant *G6PC2* construct or empty vector control using Lipofectamine 2000 (Invitrogen) in 6-well format, as described in Section 2.7. HEK293 cells that were treated with MG-132 (Calchembio), chloroquine (Sigma Aldrich) or tunicamycin (Sigma Aldrich) were treated for 15h as described in Section 2.8 unless otherwise stated. In INS-1 (832/13) experiments, cells were only treated for 6h due to cellular toxicity and were collected 36h post-transfection. Total cellular protein from both HEK293 and INS-1 (832/13) cells were collected, quantified and analysed by western blot as described in Section 2.9. G6PC2 protein expression in HEK293 cells was determined by immunoblotting using a FLAG primary antibody (Sigma Aldrich, F1804), whereas a primary antibody against Myc (Millipore, 05-724) was used for INS-1 (832/13) lysates due to detection issues with FLAG. The V5 antibody (Invitrogen, 46-0705) was used to detect expression of truncated proteins. β-tubulin (Santa Cruz, sc-9104) was used as the loading control. At least three independent experiments were conducted for each cell line for western blot analysis in untreated cells.

5.2.4. Cellular localisation by immunofluorescence microscopy

HEK293 and EndoC-βH1 cells were transfected on 4-chamber culture slides (Corning) or glass coverslips (VWR International) using FuGene 6 transfection reagent (Promega) as described in Section 2.7 and prepared for microscopy as described in Section 2.10. Where stated, double immunostaining of cells was carried out using primary antibodies for FLAG (Sigma Aldrich, F1804) or V5 (Invitrogen, 46-0705), together with calnexin (Abcam, ab10286) or calreticulin (Thermo Scientific, PA3-900), both of which are ER-associated proteins. TGN46 (Sigma Aldrich, T7576) was used as the marker for the Golgi apparatus. This was followed by secondary antibody incubation with anti-mouse Alexa Fluor 488 and anti-rabbit Alexa Fluor 568 (both Life Technologies). DRAQ5 (Thermo Fisher Scientific) was used as a far-red DNA stain. Slides were mounted with ProLong Gold Antifade Reagent (Life Technologies) and analysed on a confocal microscope as in 2.10.2.

5.2.5. Electron microscopy

HEK293 cells transfected with G6PC2 WT or R283X constructs for 48h in 10 cm dishes were harvested and loosely pelleted by gentle centrifugation. Cells were prepared for immunogold labelling and transmission electron microscopy under the guidance and assistance of Dr Anne Clark at OCDEM. First, cells were fixed in 2.5% paraformaldehyde/0.5% glutaraldehyde in 0.1M phosphate buffer at pH 7.2 for a minimum of 2h, washed in 0.1M acetate buffer and stained in 2% uranyl acetate for 1h. Cells were then washed again in acetate buffer, dehydrated in graded methanol at -20°C and incubated in a 1:1 mixture of methanol and London Resin Gold (LRG) embedding medium (Agar Scientific, UK) at -20°C overnight to prepare for embedding. Cells were embedded in pure LRG medium at -20°C for 4h. Fresh benzil (Agar Scientific, UK) in LRG at 0.1% w/v which acts as an activator was subsequently added in a clear air-free tube and cured under UV light at -20°C. LRG is an acrylic-based resin that provides better antigen preservation though poorer ultrastructural preservation as compared to epoxy resins. After polymerisation, the LRG blocks were cut into ultrathin sections of 70nm thickness using an ultramicrotome (Reichert-Jung Ultracut, AT) and placed onto 200 mesh Ni^{2+} grids (Agar Scientific, UK) by Dr Anne Clark at the Department of Physiology, Anatomy and Genetics (DPAG) of the University of Oxford. For immunogold labelling, grids were washed in distilled water and blocked in PBS with egg albumin (PBS/EA) for 30min at room temperature. Grids were incubated in a mouse monoclonal anti-V5 primary antibody (46-0705, Invitrogen) diluted 1:10 in PBS/EA for 1h, washed in PBS/EA, then incubated in a biotinylated anti-mouse secondary antibody (diluted 1:100, Dako, Agilent Technologies, US) for 1h followed by streptavidin-gold 15 nm conjugates (British Biocell, Cardiff, UK) for 1h for signal amplification. Grids were subsequently washed in PBS/EA and then distilled water. Finally, grids were contrasted with 2% uranyl acetate and lead citrate, washed with distilled water and dried for storage. Sections were examined in a Jeol 1010 transmission electron microscope with an accelerating voltage of 80 kV (Jeol, JP) fitted with a digital camera (Gatan, CA, US) at DPAG by Dr Anne Clark.

5.2.6. Enzymatic activity assay

Microsomal protein from transfected cells were collected and analysed as described in 2.11. With every experiment an untransfected control sample was also collected. Conditions for the G6Pase assay is stated in 2.11.3 unless otherwise stipulated. Variants in *G6PC2* that were studied for enzymatic activity were done so in the *G6PC1* vector backbone, as recombinant G6PC2 protein had no detectable activity based on the current assay design. As the *G6PC2* and *G6PC1* genes have sufficient sequence homology especially within the conservative regions, the equivalent amino acid residue could be mutated in *G6PC1* and used as a proxy for the study of the respective *G6PC2* variants. Nonetheless, this is based on the assumption that the relative effects of non-synonymous changes on G6Pase protein activity compared to WT protein is comparable in *G6PC2*- and *G6PC1*-encoded proteins. Sequence alignments for the relevant variants studied are shown in the Section 5.3.5.

5.2.7. Molecular visualisation and variant analysis

The crystal structure of a member of the phosphatidic acid phosphatase superfamily of enzymes, the peroxide form of the vanadium-containing chloroperoxidase from the fungus *Curvularia inaequalis*, was obtained from the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB ID: 1IDU). Molecular modelling including mutagenesis of selected residues within the active site were carried out using PyMOL v1.74.

5.2.8. Statistical analysis

For the analysis of western blot densitometry data, two-tailed paired Students' t tests were carried out to compare mean relative expression levels of each variant to WT as described in 2.9.5 and 2.17. For the analysis of enzymatic activity, two-tailed unpaired Students' t tests of kinetic constants (V_{max} and K_m) were carried out to compare variant to WT, as described in 2.11.3 and 2.17. Statistical analyses for evaluating differences between WT and variants and plotting of graphs were performed on GraphPad Prism 7.0. Data were considered statistically significant at $P<0.05$.

5.3. Results

5.3.1. *In silico* analyses of *G6PC2* variants identified from exome sequencing data

The identification of a comprehensive catalogue of coding variations from the T2D-GENES/GoT2D 13k exome sequencing project provided the opportunity to start interrogating the functional effects of rare variants in *G6PC2*. Most of the variants are rare (MAF<1%) and not included in previous exome array analyses, hence, knowledge of their phenotypic effects is limited. However, with this catalogue I carried out *in silico* mutation analyses on all 69 non-synonymous coding variants in *G6PC2* using previously described computational tools – SIFT, PolyPhen2, Condel and GERP – to predict their potential functional consequences (Table 5.3.1.1).

These tools, which are based on known protein sequence and structure information, not only provided a quick way of functionally annotating a large number of variants, but also provided information on the predicted tolerance of different protein domains to non-synonymous changes. In particular, transmembrane domain 2 (TM2) and ER luminal loop 2 (LL2) appear to possess sites of greater conservation (based on GERP scores, which range from -12.3 to 6.17 with 6.17 being the most conserved) and were more intolerant of changes as majority of variants that mapped to those domains were consistently predicted by all algorithms to be deleterious.

A number of variants map directly or close to known functionally-important regions such as the catalytic sites and consensus ER retention signal at the C-terminal end. The amino acid substitutions at the highly conserved active sites (R79Q, P80S, G114R, H174P) were, as expected, consistently predicted to be damaging. A number of variants also map adjacent to the catalytic domain and hence may be involved in the conformation of the active site. However, functional predictions were less conclusive and these variants have to be functionally analysed to determine their molecular roles.

As it was not feasible to use the current pipelines to characterise all the variants in the laboratory, I prioritised 17 coding variants for *in vitro* assessment based on two criteria. First, all variants that showed evidence of influencing glycemic traits in the MAGIC exome array meta-analysis were selected. These included the three coding variants that have been

characterised and presented in Chapter 3, which also acted as positive controls. Second, variants that mapped directly or close to the biologically interesting sites were also selected, as they may be expected to exert a direct measurable impact on protein function.

The variants selected map to all 5 exons of *G6PC2* (Figure 5.3.1.2). This means that variants in exon 4 and 5 will only implicate the full-length transcript of *G6PC2* and are irrelevant in the context of the expressed short isoform. Figure 5.3.1.3 shows the location of the variants in the predicted structure of G6PC2 in the ER membrane. Most of the variants map to the transmembrane domains, as well as a handful in the ER lumen and one in the cytosolic loop.

	Variant SNP ID	Protein change	Consequence	MAF[a]	SIFT	Polyphen2	Condel	GERP	Location	NSbroad[b]	NSstrict[b]
1	NA	Q16R	Missense	0.00024	Damaging	Probably damaging	Deleterious	5.62	N-terminus	1	0
2	rs142189264	S30F	Missense	0.00068	Deleterious	Probably damaging	Deleterious	5.42	TM1	1	0
3	NA	N37S	Missense	0.00033	Tolerated	Probably damaging	Deleterious	5.62	TM1	1	0
4	rs149874491	I38L	Missense	0.00011	Tolerated	Benign	Deleterious	5.62	TM1	0	0
5	rs371294159	T52I	Missense	0.00077	Tolerated	Benign	Deleterious	4.71	CL1	1	0
6	rs201561079	I63T	Missense	0.00047	Deleterious	Benign	Deleterious	4.46	TM2	1	0
7	NA	G64R	Missense	0.00023	Deleterious	Probably damaging	Deleterious	5.62	TM2	1	1
8	rs199682245	N68I	Missense	0.00078	Deleterious	Probably damaging	Deleterious	5.62	TM2	1	1
9	NA	I70T	Missense	0.00024	Deleterious	Probably damaging	Deleterious	5.4	TM2	1	0
10	NA	F71V	Missense	0.0019	Deleterious	Benign	Deleterious	5.62	TM2, adjacent to phosphatase start site	1	0
11	NA	W73R	Missense	0.00053	Deleterious	Probably damaging	Deleterious	5.62	TM2, adjacent to phosphatase start site	1	1
12	rs144254880	R79Q	Missense	0.00015	Deleterious	Probably damaging	Deleterious	5.58	LL1, active site residue	1	1
13	NA	P80S	Missense	0.00024	Deleterious	Probably damaging	Deleterious	5.58	LL1, active site residue	1	0
14	NA	S94X	Nonsense	0.0012	NA	NA	NA	5.58	LL1	1	1
15	NA	S95T	Missense	0.00024	Tolerated	Benign	Neutral	3.77	LL1	1	0
16	NA	P102S	Missense	0.00011	Tolerated	Probably damaging	Deleterious	5.58	LL1	1	0
17	NA	T103A	Missense	0.00012	Tolerated	Benign	Neutral	1.94	LL1	1	0
18	rs149663725	G114R	Missense	0.0026	Deleterious	Probably damaging	Deleterious	5.78	LL1, active site residue	1	0
19	NA	A119V	Missense	0.00036	Tolerated	Benign	Deleterious	4.91	TM3	1	0
20	rs187707963	Y124C	Missense	0.00023	Deleterious	Probably damaging	Deleterious	5.78	TM3	1	1
21	NA	Y124X	Nonsense	0.00023	NA	NA	NA	0.558	TM3	1	1
22	rs367930047	M126V	Missense	0.00048	Damaging	Possibly damaging	Neutral	5.78	TM3	1	0
23	NA	L148P	Missense	0.00023	Tolerated	Probably damaging	Neutral	5.9	TM4	0	0
24	NA	S155G	Missense	0.00011	Tolerated	Benign	Deleterious	4.73	TM4	0	0
25	NA	S155R	Missense	0.00011	Tolerated	Possibly damaging	Deleterious	5.9	TM4	0	0

	Variant SNP ID	Protein change	Consequence	MAF[a]	SIFT	Polyphen2	Condel	GERP	Location	NSbroad[b]	NSstrict[b]
26	NA	F157L	Missense	0.00047	Tolerated	Possibly damaging	Neutral	5.9	TM4	1	0
27	NA	W158X	Nonsense	0.00024	Na	Na	NA	5.9	TM4	1	1
28	NA	L159W	Missense	0.00024	Tolerated	Possibly damaging	Deleterious	5.9	TM4	1	0
29	rs2232322	I171V	Missense	0.00023	Tolerated	Benign	Neutral	5.9	TM5, close to active site	1	0
30	rs145050507	I171T	Missense	0.0013	Tolerated	Benign	Deleterious	4.73	TM5, close to active site	1	0
31	NA	H174P	Missense	0.00011	Deleterious	Probably damaging	Deleterious	5.9	TM5, active site residue	1	1
32	rs138726309	H177Y	Missense	0.0044	Deleterious	Probably damaging	Deleterious	5.9	TM5, close to active site	1	1
33	NA	L188V	Missense	0.00011	Tolerated	Benign	Neutral	5.88	TM5	0	0
34	NA	T201M	Missense	0.00048	Tolerated	Benign	Neutral	3.09	CL3	0	0
35	NA	S203R	Missense	0.00093	Tolerated	Benign	Deleterious	1.56	CL3	1	0
36	rs2232323	Y207S	Missense	0.006	Deleterious	Probably damaging	Deleterious	4.73	CL3	1	1
37	NA	N211T	Missense	0.00024	Tolerated	Benign	Neutral	4.73	CL3	1	0
38	rs139587795	F215S	Missense	0.00024	Tolerated	Benign	Deleterious	5.66	TM6	1	0
39	rs492594	V219L	Missense	0.49	Tolerated	Benign	Neutral	3.06	TM6	0	0
40	NA	L228H	Missense	0.00023	Deleterious	Probably damaging	Deleterious	5.86	TM6	1	0
41	rs145217135	I230T	Missense	0.00024	Deleterious	Possibly damaging	Deleterious	5.66	TM6	1	0
42	NA	W234C	Missense	0.00011	Deleterious	Probably damaging	Deleterious	5.86	LL2	1	1
43	NA	P246A	Missense	0.00011	Tolerated	Probably damaging	Deleterious	5.86	LL2	1	0
44	rs150435873	D247N	Missense	0.00024	Deleterious	Possibly damaging	Deleterious	5.86	LL2	1	0
45	rs147360987	H250Y	Missense	0.00072	Deleterious	Probably damaging	Deleterious	5.76	LL2	1	0
46	NA	I251T	Missense	0.00015	Deleterious	Possibly damaging	Deleterious	5.76	LL2	1	0
47	NA	T254M	Missense	0.00024	Deleterious	Probably damaging	Deleterious	5.76	LL2	1	1
48	rs150538801	F256L	Missense	0.00079	Deleterious	Benign	Deleterious	4.6	LL2	1	0
49	NA	L263P	Missense	0.00011	Deleterious	Probably damaging	Deleterious	5.76	TM7	1	1
50	NA	G270C	Missense	0.00023	Deleterious	Probably damaging	Deleterious	5.76	TM7	1	1

	Variant SNP ID	Protein change	Consequence	MAF[a]	SIFT	Polyphen2	Condel	GERP	Location	NSbroad[b]	NSstrict[b]
51	rs148689354	I273V	Missense	0.0022	Damaging	Benign	Neutral	2.13	TM7	0	0
52	NA	N274T	Missense	0.00052	Tolerated	Benign	Deleterious	4.6	TM7	1	0
53	NA	M277K	Missense	0.00011	Deleterious	Possibly damaging	Deleterious	5.76	TM7	1	0
54	NA	L279R	Missense	0.00023	Tolerated	Benign	Neutral	3.19	TM7	0	0
55	rs146779637	R283X	Nonsense	0.0008	NA	NA	NA	-0.722	CL4	1	1
56	NA	R283Q	Missense	0.00023	Tolerated	Benign	Neutral	1.61	CL4	0	0
57	NA	G285E	Missense	0.00024	Tolerated	Benign	Neutral	3.43	CL4	0	0
58	rs143686780	T289K	Missense	0.00022	Tolerated	Benign	Neutral	3.43	CL4	0	0
59	NA	R293Q	Missense	0.00024	Deleterious	Possibly damaging	Deleterious	4.93	CL4	1	0
60	NA	T299I	Missense	0.00027	Tolerated	Benign	Neutral	4.2	TM8	0	0
61	rs141041285	L301S	Missense	0.0014	Tolerated	Probably damaging	Deleterious	5.98	TM8	1	0
62	NA	L310F	Missense	0.00066	Tolerated	Benign	Neutral	-0.691	TM8	0	0
63	rs137857125	P313L	Missense	0.00048	Deleterious	Possibly damaging	Deleterious	5.11	TM8	1	0
64	NA	E316K	Missense	0.00068	Tolerated	Benign	Neutral	-10.7	LL3	0	0
65	NA	F320L	Missense	0.00011	Deleterious	Probably damaging	Deleterious	4.6	TM9	1	0
66	NA	Y321X	Nonsense	0.00027	NA	NA	NA	-6.31	TM9	1	1
67	rs2232326	S324P	Missense	0.0074	Deleterious	Probably damaging	Deleterious	5.76	TM9	1	0
68	rs2232328	S342C	Missense	0.14	Tolerated	Benign	Neutral	4.89	C-terminus	0	0
69	NA	Q355L	Missense	0.00068	Tolerated	Benign	Neutral	4.59	C-terminus, ER retention sequence	1	0

Table 5.3.1.1. *In silico* analyses of *G6PC2* variants identified from the T2D-GENES/GoT2D 13k exome sequencing study. Variants are arranged according to amino acid position. *In silico* analyses of variants were carried out based on the full-length protein sequence with Ensembl ID ENSP00000364512. Variants prioritised for functional study are highlighted in orange. The algorithms were accessed on 8/8/16. [a]T2D-GENES/GoT2D Consortia: n=Up to 12,886 of multiple ancestries (African American, East Asian, Hispanic, South Asian, European). [b]NSbroad and NSstrict classifications were obtained from the T2D-GENES/GoT2D gene-based analyses, with 1 indicating inclusion in the mask and 0 indicating exclusion. Variant list was provided by Anubha Mahajan. MAF: Minor allele frequency; GERP: Genomic Evolutionary Rate Profiling; TM: transmembrane region; CL: cytoplasmic loop; LL: endoplasmic reticulum luminal loop; C-terminus: carboxy terminal domain.

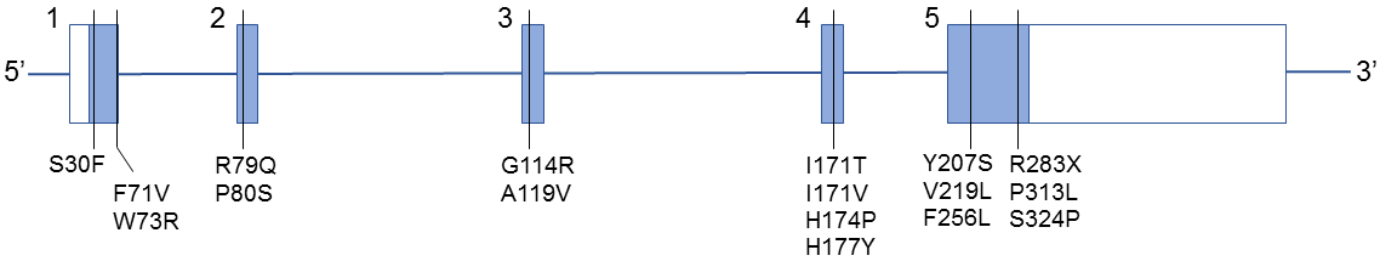


Figure 5.3.1.2. Schematic of *G6PC2* gene structure showing variants prioritised for *in vitro* functional analysis. The five exons are labelled 1-5 with dark boxes representing coding regions and white boxes representing 5' and 3' untranslated regions. Illustration corresponds to the full-length transcript NM_021176 (not drawn to scale).

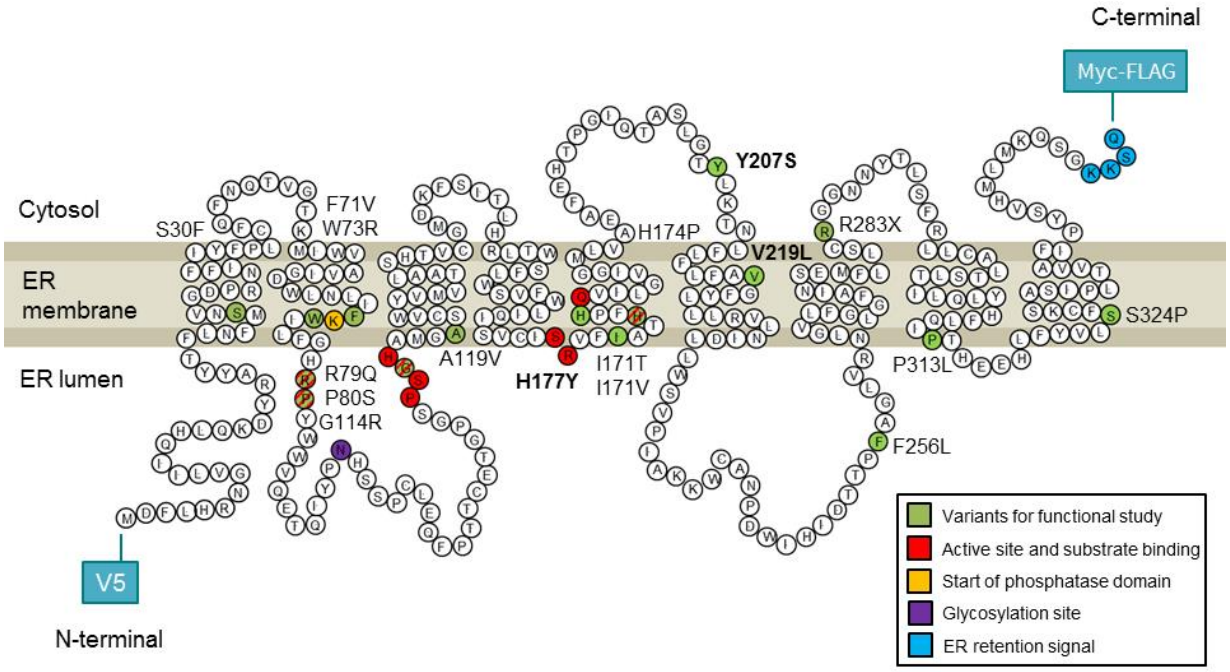


Figure 5.3.1.3. Illustration of the predicted structure of G6PC2 protein in the ER membrane, showing the variants prioritised for functional study. Amino acid residues are coloured as described in the legend. Variants selected for functional study, in green, are labelled. Those with a striped pattern indicate selected variants that are also active site residues. Variants in bold have been studied in Chapter 3. The N-terminal V5 and C-terminal Myc-FLAG tags present in the expression constructs are indicated. Transmembrane regions were annotated based on RefSeq NP_066999.1.

5.3.2. Protein expression of G6PC2 variant proteins

For the study of missense variants, site directed mutagenesis was carried out on the pCMV6-*G6PC2* vector with the C-terminal Myc-FLAG-tags. For the study of truncated variants where C-terminal tags would not be expressed, an N-terminal V5 tag was cloned into the *G6PC2* vector backbone as described in Section 2.1 and this construct was used for mutagenesis. All constructs generated for use in cellular assays (Figure 5.3.2.1) were sequence-checked.

As in previous characterisation studies, variant constructs were transiently expressed in HEK293 cells and INS-1 (832/13) cells and protein expression levels were analysed by western blot. As seen in Figure 5.3.2.2 and 5.3.2.3, the majority of all the variants studied (13 of 17) displayed markedly reduced protein levels as compared to WT in HEK293 cells ($P<0.01$). The results for previously studied variants were also replicated here. Consistent with previous findings, these variants were degraded by the ubiquitin-proteasome pathway as protein expression of all variants were rescued in the presence of proteasomal inhibitor MG-132 but not lysosomal pathway inhibitor chloroquine.

Western blot analysis was also carried out for INS-1 (832/13) samples, where recombinant G6PC2 protein was detected using an antibody against the C-terminal Myc tag, as technical difficulties were observed with the FLAG antibody at the time of experiment. Myc-based detection was however validated against previously-analysed samples to show that results produced were comparable to FLAG-based detection (Figure 5.3.2.5). All analyses of INS-1 (832/13) samples were subsequently carried out using the Myc antibody.

Results in the INS-1 (832/13) cells revealed that unlike in HEK293 cells, the variants exerted greater impact on protein glycosylation as opposed to total protein levels (Figures 5.3.2.6 and 5.3.2.7). The majority of the variants displayed partial to total loss of the glycosylated form of the protein, which were represented by bands of larger molecular weight on the western blot. These bands were confirmed to be the glycosylated form of G6PC2 as they were lost when transfected cells were treated with tunicamycin, an inhibitor of N-linked glycosylation and also an ER stress inducer, in both HEK293 cells (Figure 5.3.2.4) and INS-832/13 cells (Figure 5.3.2.6). When quantified, glycosylated protein levels of 13 of 16 of the missense variants expressed in INS-1 (832/13) cells were downregulated ($P<0.01$) in a trend

consistent with the HEK293 data. Though total protein levels were also regulated by the proteasomal pathway, the glycosylation defect was not reversed when cells were treated with either MG-132 or chloroquine. This may be more clearly demonstrated in side-by-side comparisons in Figure 5.3.2.8. These results revealed that the variants affected glycosylation levels in addition to influencing protein stability, depending on cellular background, highlighting the importance of functional investigations in multiple cell models which are likely to have different protein-processing machineries. As glycosylation may be important for binding or enzymatic properties of G6Pase (Shieh et al. 2004), changes to the relative proportions of glycosylated and unglycosylated protein will affect total protein activity.

Finally, the PTV R283X was expressed as a lower molecular weight protein as expected (Figure 5.3.2.3) and protein levels were reduced by 35% compared to WT in HEK293 cells (P=0.004). This effect was not quantified in INS-1 (832/13) cells due to poor transfection efficiency, but in one experiment R283X protein levels appeared to be downregulated and expression was rescued when the proteasomal pathway was blocked by administration of MG-132 (Figure 5.3.2.9). Nonetheless, as R283X was expressed I wanted to investigate the effects of loss of the ER retention signal on cellular localisation as well as the presence of any phosphatase activity. In addition, three missense variants (I171T, I171V, F256L) were also found to be stably expressed in one or both cellular models compared to WT. As residue I171 maps close to the conserved catalytic sites, I reasoned that the two variants observed at this multi-allelic site could impact on enzymatic activity, which I set out to test in G6Pase assays.

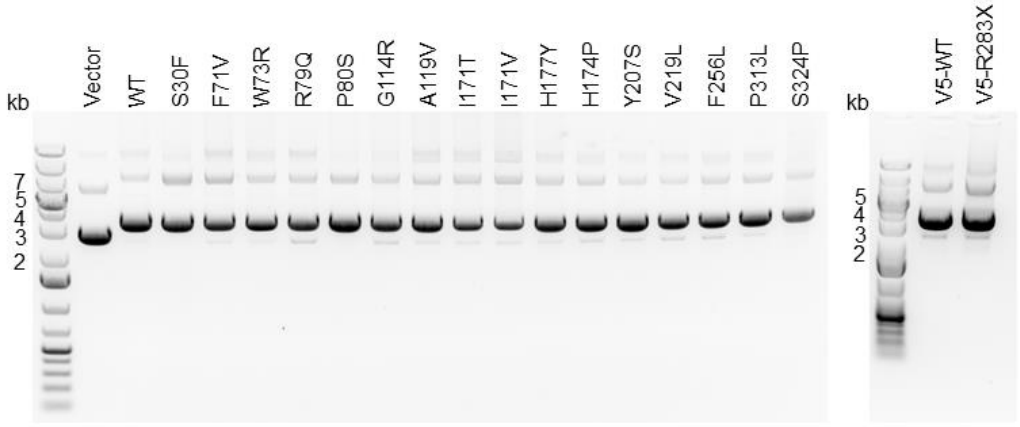


Figure 5.3.2.1. DNA gel of pCMV6-G6PC2 WT and mutagenised plasmid constructs used in cellular experiments. 1 µg of midiprep DNA was loaded and run on a 0.8% agarose gel, alongside the 1kb plus DNA ladder (Thermo Scientific). kb: kilobase.

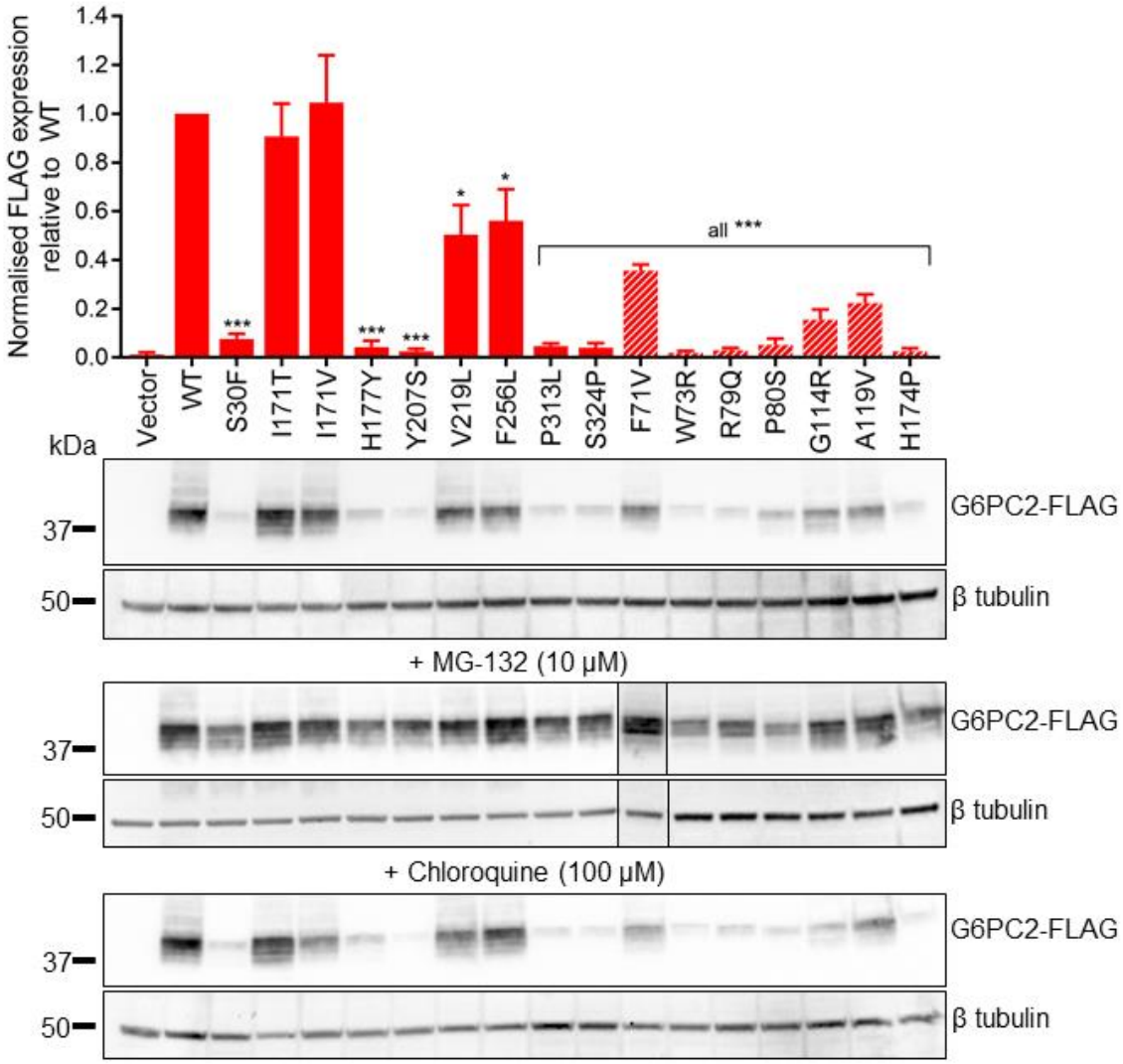


Figure 5.3.2.2. Functional characterisation of G6PC2 missense variants in HEK293 cells by western blot analysis. Protein expression levels were determined by densitometric analysis of tagged G6PC2 constructs relative to β tubulin control using a FLAG tag antibody (n=5). Striped bars indicate variants that were selected primarily based on biological interest. Data presented as mean ± SEM. Statistical analysis of densitometry data was carried out using paired t tests comparing each variant against WT. *** P<0.001; * P=0.01-0.05. Representative western blots are shown for untreated cells and cells treated with proteasomal inhibitor MG-132 or lysosomal inhibitor chloroquine. Protein bands in the spliced MG-132 image were re-arranged but were part of the same original image. Molecular weight in kilodalton (kDA) is indicated on the left of the western blot images. WT: wild type.

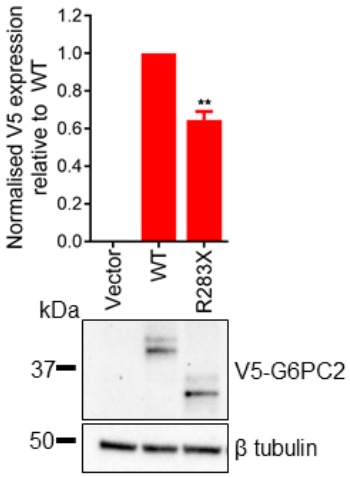


Figure 5.3.2.3. Functional characterisation of protein-truncating G6PC2 variant R283X in HEK293 cells by western blot analysis. Protein expression levels of R283X were determined by densitometric analysis using a V5 tag antibody (n=4). Data presented as mean ± SEM. Statistical analysis of densitometry data was carried out using paired t tests comparing the variant against WT. ** P=0.001-0.01.

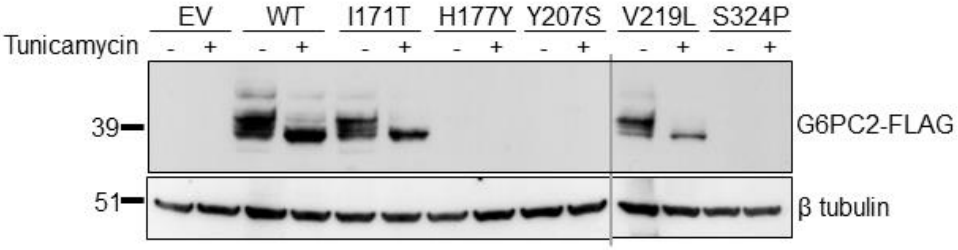


Figure 5.3.2.4. Effect of N-linked glycosylation inhibitor tunicamycin on WT and variant G6PC2 proteins in HEK293 cells. Cells were treated with DMSO vehicle control (-) or 1 µg/ml tunicamycin (+) for 15h. Samples from the same experiment were run on separate gels and spliced together as denoted by the vertical line. Molecular weight in kilodalton (kDA) is indicated on the left of the western blot images. EV: empty vector; WT: wild type.

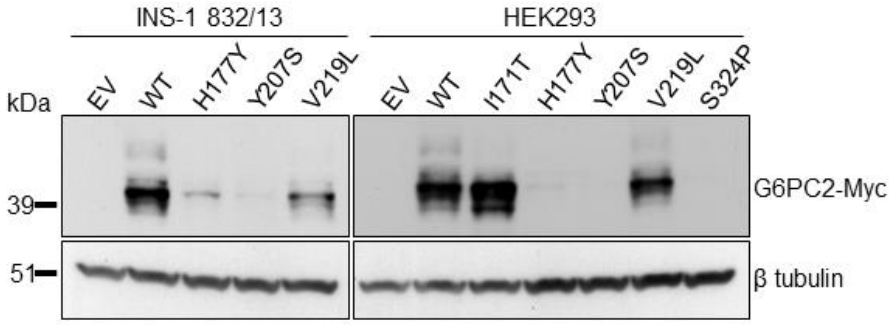


Figure 5.3.2.5. Western blot of WT and variant G6PC2 proteins using an antibody against the Myc tag (05-724, Millipore) in INS-1 (832/13) and HEK293 cells. All samples were previously analysed for FLAG expression in Chapter 3 (data not shown). Molecular weight in kilodalton (kDA) is indicated on the left of the western blot image. EV: empty vector; WT: wild type.

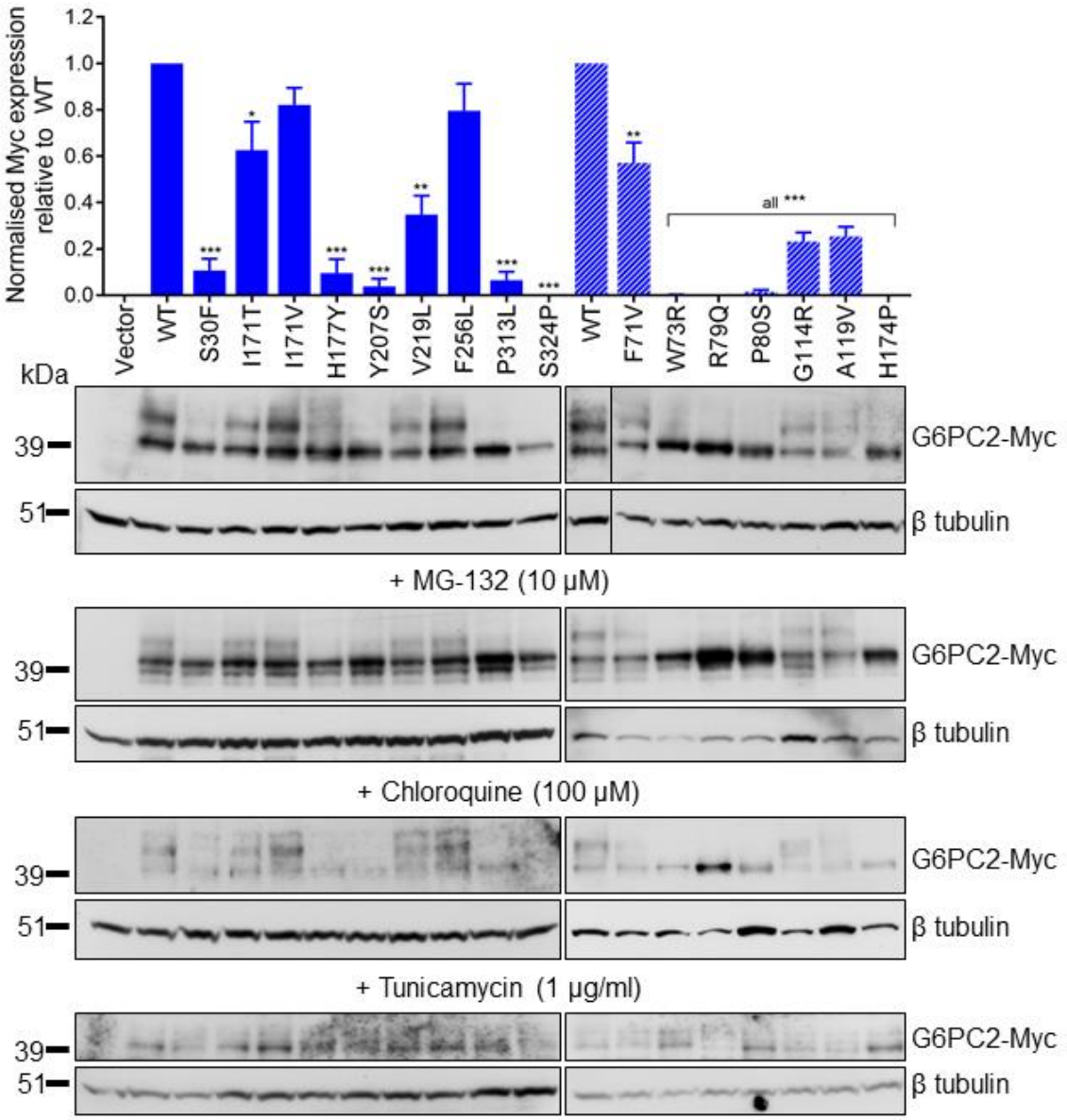


Figure 5.3.2.6. Functional characterisation of G6PC2 missense variants in INS-1 (832/13) cells by western blot analysis. Protein expression levels were determined by densitometric analysis of tagged G6PC2 constructs relative to β tubulin control using a Myc tag antibody (n=5). The glycosylated forms of the G6PC2 proteins (upper bands) were selectively quantified. Striped bars indicate variants that were selected primarily based on biological interest. Data presented as mean ± SEM. Statistical analysis of densitometry data was carried out using paired t tests comparing each variant to WT. *** P<0.001, ** P=0.001-0.01, * P=0.01-0.05. Representative western blots are shown for untreated cells and cells treated with proteasomal inhibitor MG-132, lysosomal inhibitor chloroquine or N-linked glycosylation inhibitor tunicamycin. Both lanes of WT were from the same sample. Protein bands were spliced together in the first image but were part of the same original image. Molecular weight in kilodalton (kDA) is indicated on the left of the western blot images. WT: wild type.

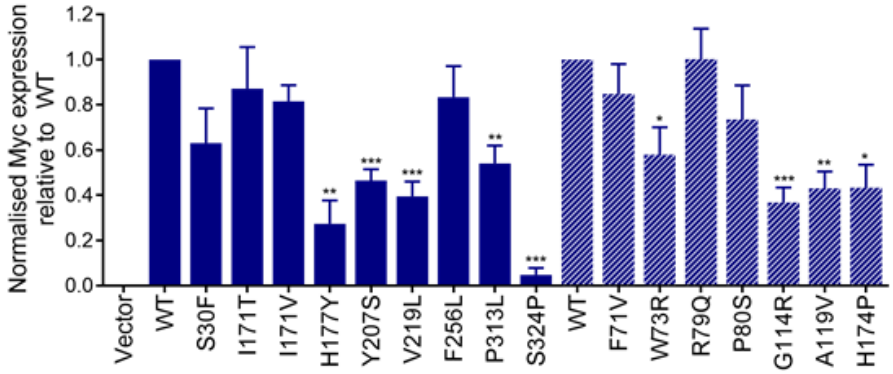


Figure 5.3.2.7. Total protein expression levels of expressed missense G6PC2 variants in INS-1 (832/13) cells by western blot analysis. Densitometric quantification was carried out on the same untreated samples as that in Figure 5.3.2.6, except that both upper and lower bands on the western blot were quantified. Data presented as mean ± SEM. *** P<0.001, ** P=0.001-0.01, * P=0.01-0.05 based on paired t tests comparing each variant to WT.

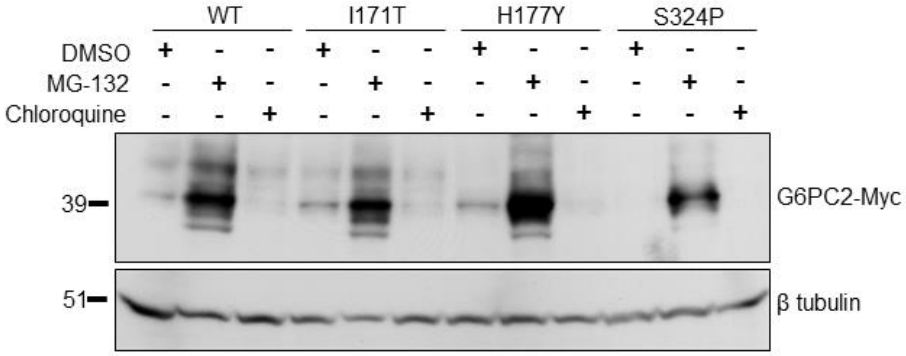


Figure 5.3.2.8. Effect of MG-132 (proteasomal pathway inhibitor) and chloroquine (lysosomal pathway inhibitor) on protein expression levels of G6PC2 WT and 3 selected variant after 6h treatment in INS-1 (832/13) cells. DMSO was used as a vehicle control. Molecular weight in kilodalton is indicated on the left of the western blot image. WT: wild type.

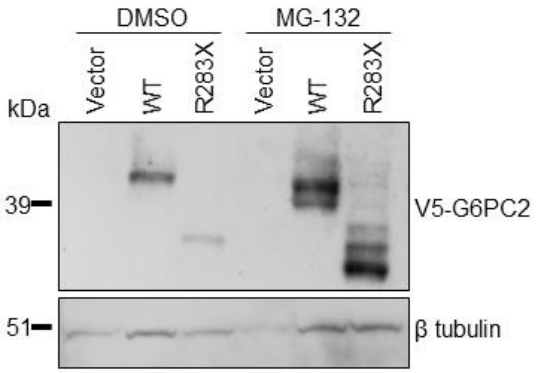


Figure 5.3.2.9. Western blot of R283X expressed in INS-1 (832/13) cells detected using an antibody against the V5 tag. Only one experiment was carried out due to technical difficulties in detecting expressed protein in cells not treated with MG-132. DMSO was used as a vehicle control. Molecular weight in kilodalton (kDa) is indicated on the left of the western blot image.

WT: wild type.

5.3.3. Subcellular localisation of G6PC2 variant proteins

In my previous studies in Chapter 3 I investigated the localisation of missense variants H177Y, Y207S and V219L in HEK293 cells and found that they displayed similar localisation patterns to WT and were at least in part localised to the ER. For this I had the opportunity to investigate cellular localisation in EndoC-βH1 cells, a human pancreatic beta cell line which is a physiologically-relevant cell model for the study of human G6PC2 (Ravassard et al. 2011). Immunofluorescence (IF) microscopy showed that WT protein and a number of missense variants studied appeared to also co-localise with the ER marker calnexin (Figure 5.3.3.1). However, like that observed in the HEK293 cells (Chapter 3 Figure 3.3.4.3), WT protein and the stably expressed variants such as I171T and V219L displayed aggregated signals close to the nucleus. No previous evidence of the protein localising to other cellular compartments has been reported, though the Golgi apparatus or cell aggresomes (aggregation of misfolded proteins) are both possibilities that would need to be tested.

I therefore carried out co-immunostaining with ER marker calreticulin or a marker for the trans Golgi network, TGN46, to investigate localisation to both organelles. The F256L variant protein was examined as it was expressed at levels similar to WT based on western blot analysis. Studies in HEK293 cells showed that F256L indeed displayed a similar expression pattern as WT and co-localised with the Golgi marker TGN46 (Figure 5.3.3.2). Given that I previously also showed that G6PC1 normally localises to the Golgi apparatus, it is possible that functional G6PC2 proteins also undergo post-translational modification within the Golgi apparatus before being shuttled back to the ER.

Next, given that the truncated variant R283X is expected to lack the consensus ER localisation signal, I hypothesised that it may not be retained in the ER. Strikingly, in both HEK293 (Figure 5.3.3.3) and EndoC-βH1 (Figure 5.3.3.4) cells R283X appeared to be more diffused throughout the cytoplasm than WT. R283X also failed to co-localise with TGN46 when assessed in the EndoC-βH1 cells (Figure 5.3.3.4). Due to the limited resolution of IF confocal microscopy, localisation to the ER could not be ascertained with absolute confidence. I therefore attempted to answer this question by carrying out immunogold labelling for electron microscopy with the assistance of Dr Anne Clark at OCDEM. HEK293

cells expressing WT or R283X were collected and processed for electron microscopy. We carried out immunogold staining using the V5 antibody and electron-dense gold particles to identify recombinantly expressed G6PC2. As seen in Figure 5.3.3.5(A-B), G6PC2 WT was found to localise to both the ER and Golgi apparatus. However, it also appears to be found in structures characteristic of autophagosomes which could be a result of clearance due to protein overexpression (Figure 5.3.3.5[C]). On the other hand, R283X was largely present in structures typical of autophagosomes/lysosomes (Figure 5.3.3.5[E-F]) possibly due to increased protein degradation. There was also some evidence of R283X localisation to the Golgi apparatus (Figure 5.3.3.5[D,F]) though this was inconsistent with my IF results. We were unable to find evidence of R293X in the ER. While electron microscopy has revealed the cellular localisation of G6PC2 WT and R283X proteins, it is important to acknowledge its limitations. First, an overexpression system may generate artefacts due to the 'unphysiological' expression of recombinantly tagged proteins. The HEK293 cells only served as one model system and as we were unable to observe clear ER structures in these cells we could not definitively determine if R283X may be localised to the ER. Additionally, detection of expressed protein proved to be a challenge due to the lack of antigens in the thin sections required for electron microscopy, therefore we were only able to collect a limited sample of images. Finally, we did not carry out double immunostaining for identifying the cellular structures observed.

Overall, we have established G6PC2 localisation for the first time using electron microscopy in addition to immunofluorescence microscopy, and showed that G6PC2 WT is present in both the ER as expected but also the Golgi apparatus. Additionally, based on IF we found that the PTV R283X did not co-localise with a marker for the Golgi network. We unable to determine if the C-terminal truncation of R283X affects localisation of the variant to the ER, but we observed the presence of the variant protein in autophagosomes/lysosomes when expressed in HEK293 cells.

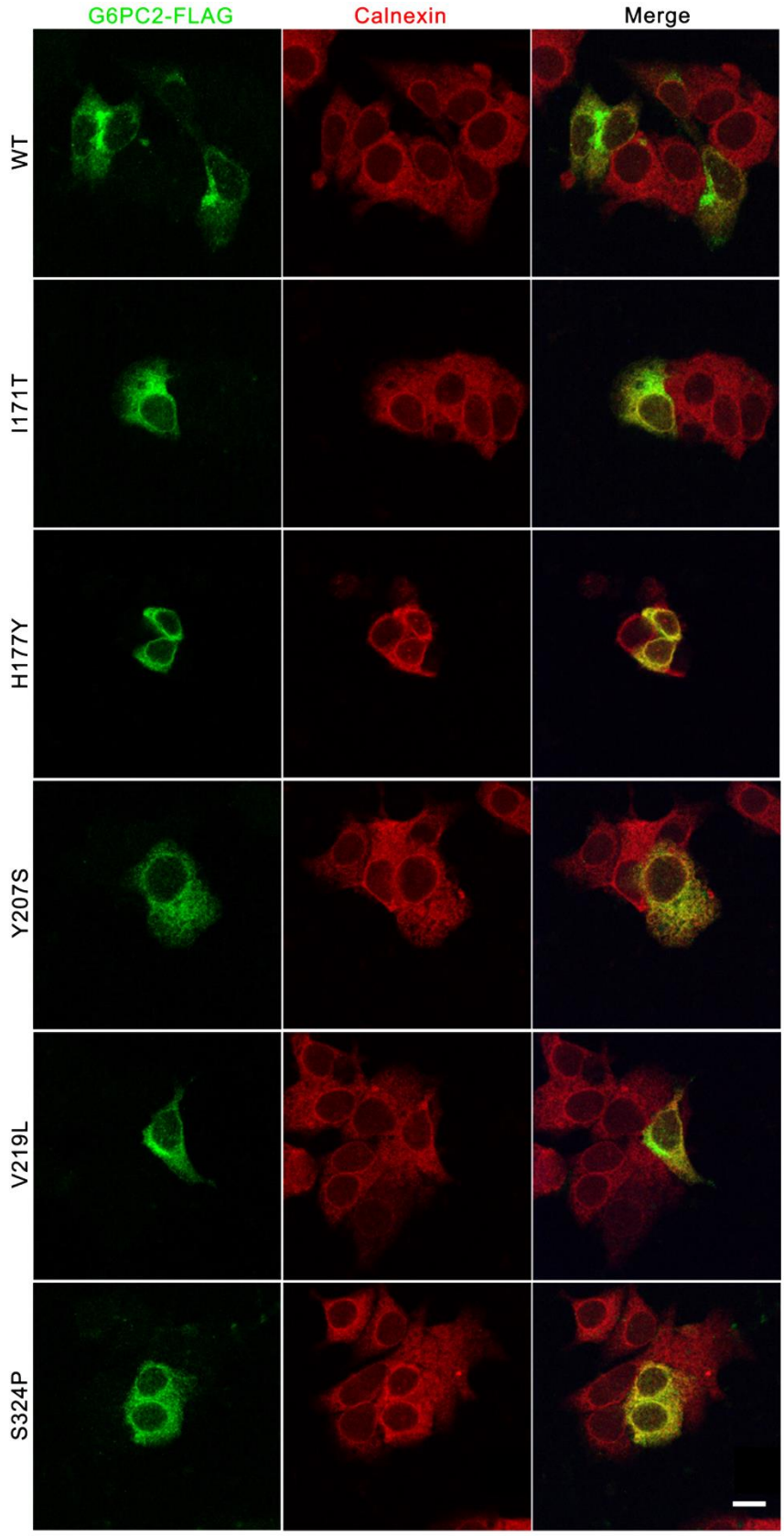


Figure 5.3.3.1. Cellular localisation of wild type and variant G6PC2 proteins expressed in EndoC-βH1 cells as assessed by immunofluorescence microscopy. Cells were double immunostained for FLAG (green) and ER protein calnexin (red), and merged. Images were taken with laser settings and gain intensities that were optimised separately for each sample. Scale bar indicates 10 µm.

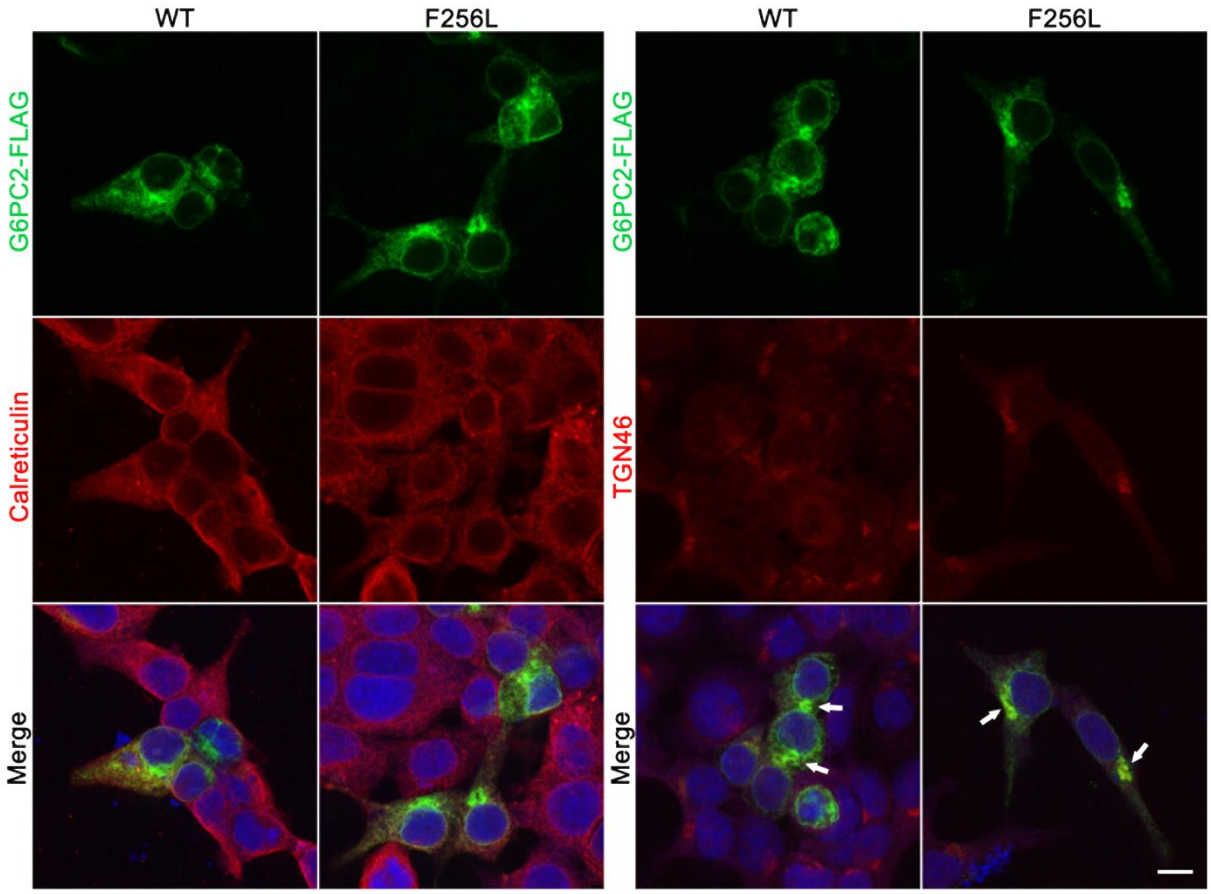


Figure 5.3.3.2. Cellular localisation of wild type G6PC2 and F256L with respect to the ER (left) and Golgi complex (right) in HEK293 cells was assessed by immunofluorescence confocal microscopy. Cells were double immunostained for FLAG tag (green) and markers for the ER, calreticulin, on the left, or trans-golgi network, TGN46, on the right (both in red). Images were merged with staining of the nuclei using DRAQ5 (blue). White arrows point to positions of the golgi apparatus in the merged images. Scale bar indicates 10 µm.

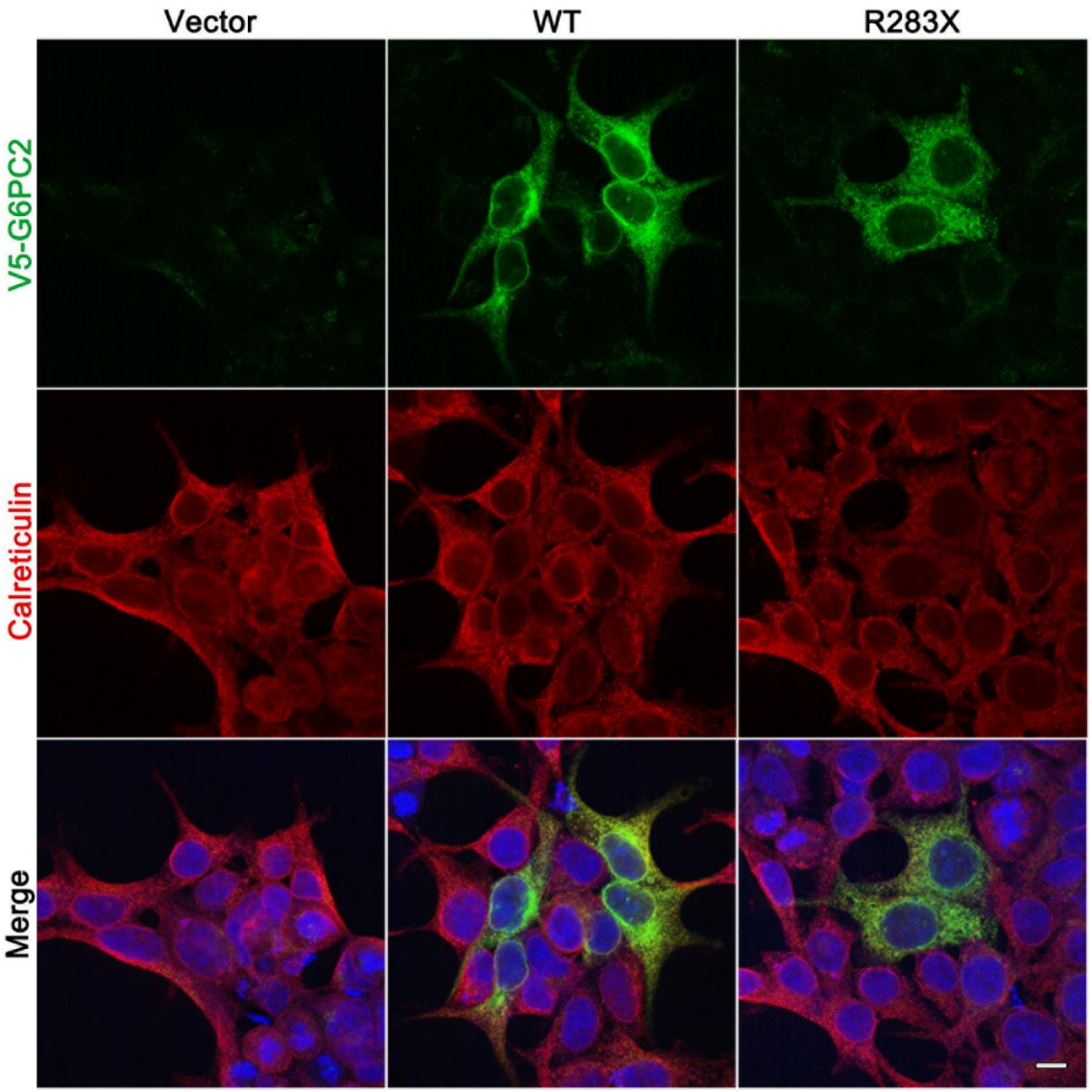


Figure 5.3.3.3. Cellular localisation of wild type G6PC2 and R283X with respect to the ER in HEK293 cells was assessed by immunofluorescence confocal microscopy. Cells were double immunostained for V5 tag (green) and ER protein calreticulin (red), and merged with staining of the nuclei using DRAQ5 (blue). Images were taken with laser settings that were optimised separately for each sample. Scale bar indicates 10 µm.

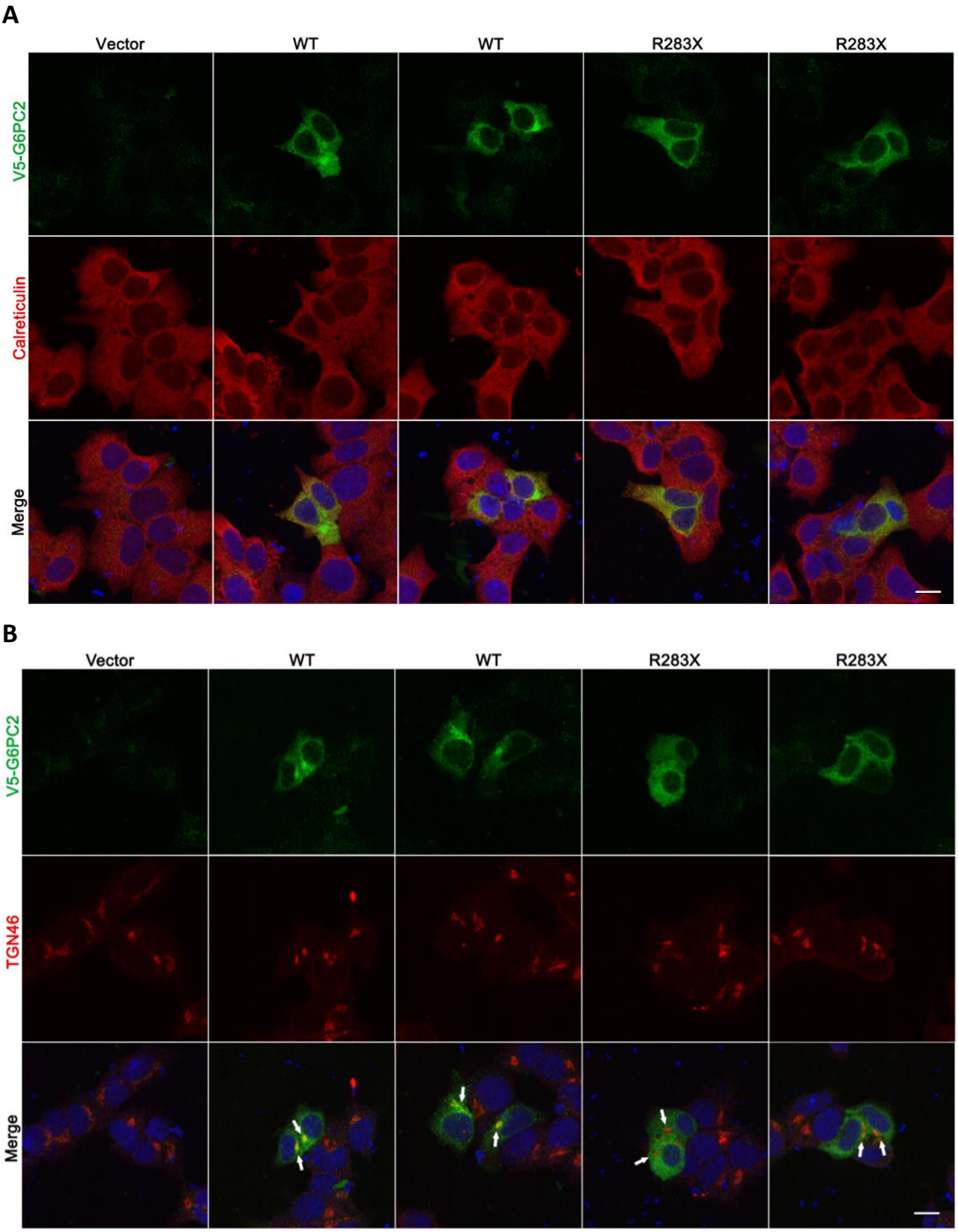


Figure 5.3.3.4. Cellular localisation of wild type G6PC2 and R283X with respect to the (A) ER and (B) Golgi complex in human EndoC-βH1 cells was assessed by immunofluorescence microscopy. Cells were double immunostained for V5 tag (green) and markers for the ER (calreticulin, red) or trans-golgi network (TGN46, red), and merged with staining of the nuclei using DRAQ5 (blue). (Legend continues on next page)

Figure 5.3.3.4. (Legend continued from previous page) Two representative images each for WT and R283X are shown. Images were taken with laser settings and gain intensities that were optimised separately for each sample. White arrows point to positions of the golgi apparatus in the merged images. Scale bar indicates 10µm.

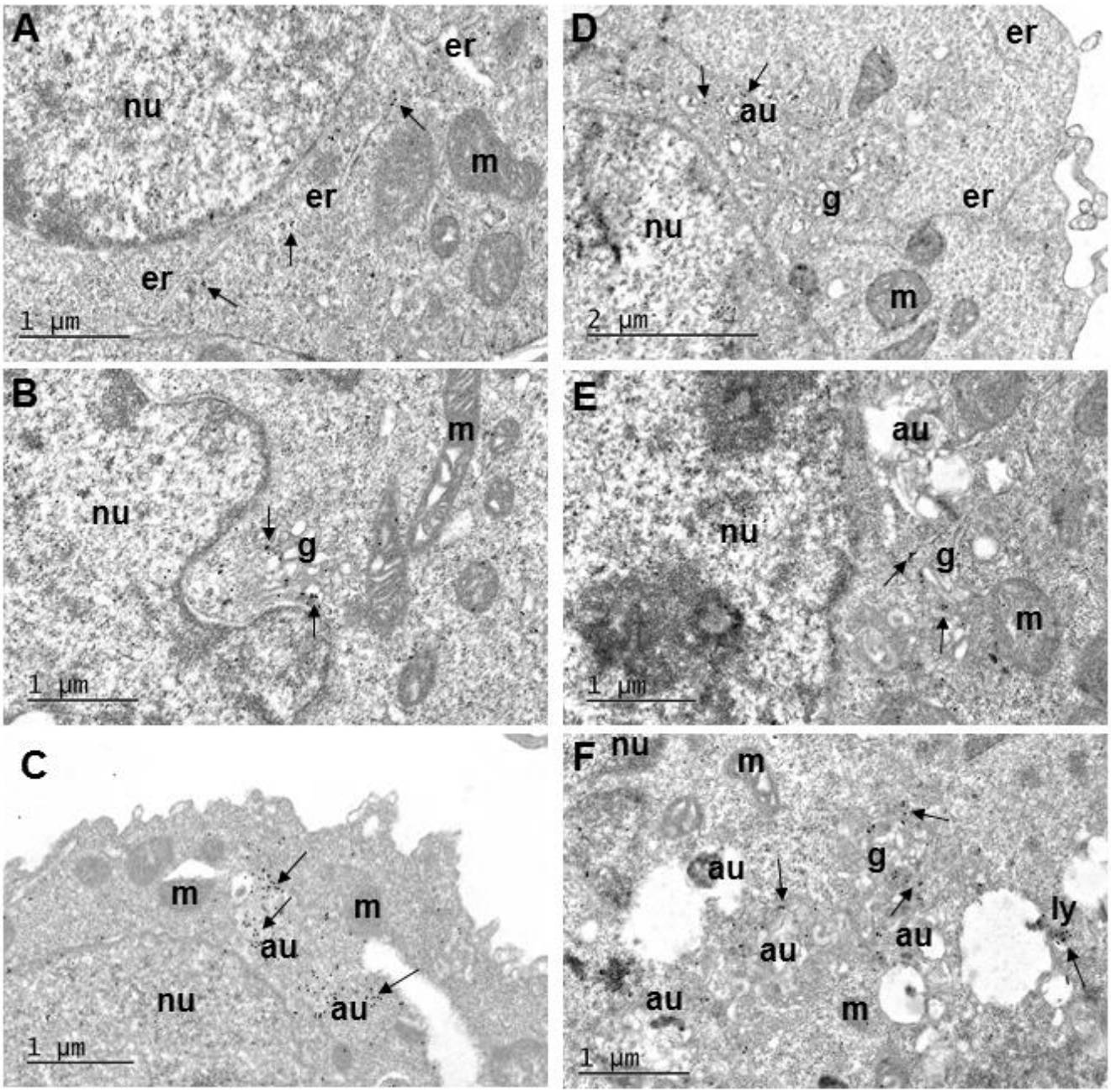


Figure 5.3.3.5. Electron micrographs of G6PC2 WT (A-C) and truncating variant R283X (D-F) expressed in HEK293 cells. Immunogold labelling of recombinant G6PC2 proteins are shown as solid black dots and some have been highlighted with black arrows. G6PC2-WT localises to both the ER (A) and Golgi apparatus (B). Protein overexpression may result in increased degradation in autophagosomes (C). R283X also appears to localise to the Golgi apparatus (E and F) but largely in autophagosomes or lysosomal bodies (D and F). nu, nucleus; m, mitochondrion; er, endoplasmic reticulum; g, Golgi apparatus; au, autophagosome; ly, lysosome. Scale bars as shown in each image.

5.3.4. Optimisation of a glucose-6-phosphatase assay for the study of G6PC2

To investigate the impact of coding variants on G6PC2 enzymatic activity, I first assessed the activity of G6PC2 WT using the G6Pase assay that I had set up for G6PC1 (described in Chapter 4). Briefly, microsomal protein was isolated from HEK293 cells expressing G6PC2 WT construct and incubated in the presence of the substrate G6P for enzymatic determination based on a phosphate detection method. 1 mM up to 40 mM of G6P was tested. However, no activity was detected from G6PC2 both in the absence and presence of histones (for unpermeabilised and permeabilised microsomes respectively), in stark contrast to G6PC1 as shown in Figure 5.3.4.1. To determine if this was due to different reaction times required for G6PC2, a time course experiment was set up to evaluate activity up to 60 min. Despite the continuous release of phosphates detected during the reaction, no specific activity above background was detected for G6PC2. On the contrary, G6PC1 activity peaked between 5-10 min (Figure 5.3.4.2).

Subsequently, a number of other factors were tested, including reaction buffer (Bis-Tris- or MES), pH, amount of protein input and colour development time, with little success in detection of G6PC2 activity (data not shown). I then tested other tissues or cells where endogenous G6PC2 is expressed and showed that G6Pase activity could be detected in mouse islet tissue and the mouse beta cell line MIN6 as compared to the low activity from HEK293 microsomes overexpressing G6PC2 or the human beta cell line EndoC-βH1. My unsuccessful attempt of studying specific G6PC2 activity based on the current cell model and assay design was consistent with previous failed attempts from other groups. The radioisotopic G6Pase assay reported in Pound et al. was however not attempted to avoid experiments involving radioactivity.

As a result, I came up with a different strategy for studying G6PC2 variants. As there is high conservation of the catalytic domains between G6PC1 and G6PC2, I decided to indirectly analyse the effect of G6PC2 variants that are located within the conservative sequences by examining the effect of the same variants on the highly-related and much more enzymatically-active G6PC1 protein. This is based on the assumption that variants in G6PC2 that affect enzyme function are likely to elicit the same consequence in G6PC1 because of the strong conservation of catalytically important sites and preserved topology of both

proteins. I proceeded to study selected variants of interest in G6PC2 within the conserved regions of the G6PC1 sequence.

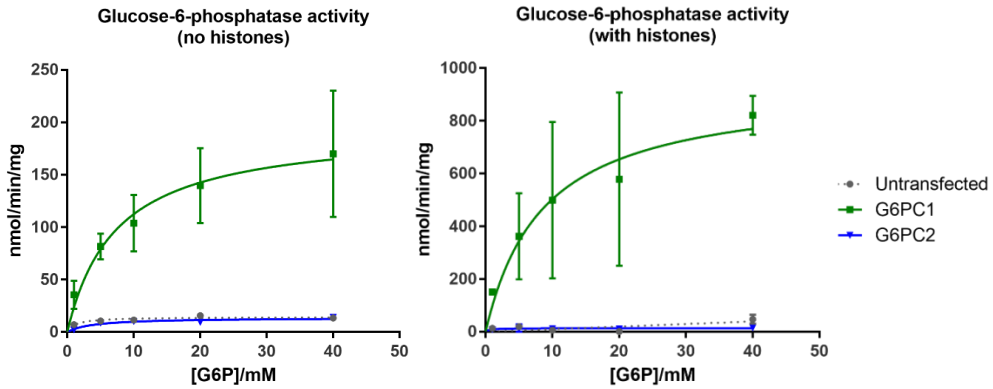
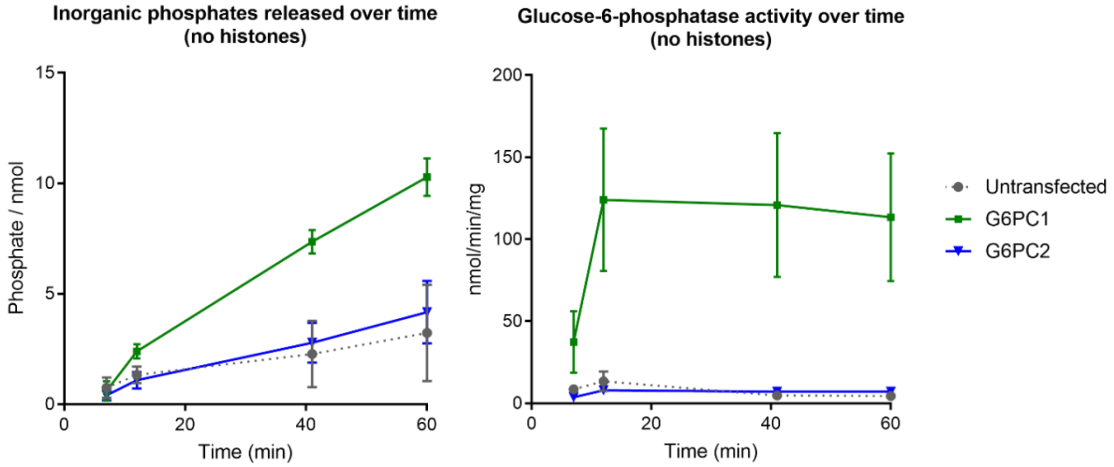


Figure 5.3.4.1. Glucose-6-phosphatase activity of microsomal preparations containing G6PC1 and G6PC2 proteins in HEK293 cells. Enzyme reaction time was kept to 7-9 min as recommended. Data collected from two independent experiments (of 2 technical replicates each) for each assay. Data presented as nmol/min/mg are not normalised within each experiment.

A



B

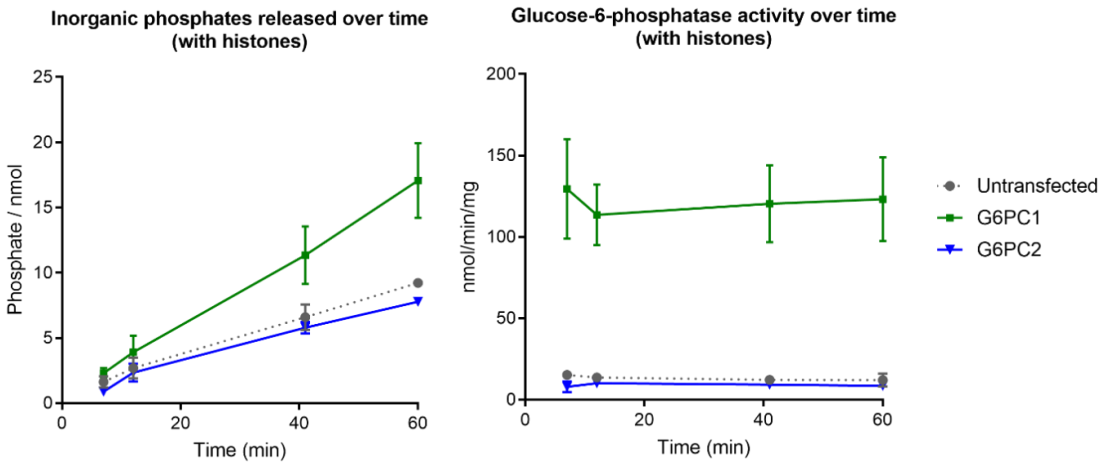


Figure 5.3.4.2. Time course experiment for the reaction of microsomal preparations containing G6PC1 and G6PC2 proteins with 5 mM G6P in the (A) absence or (B) presence of histones.

(Figure on previous page) Figure 5.3.4.2. Absolute phosphate release is shown on the left and calculated rate of reaction is shown on the right. Data collected from three independent experiments in (A) and two in (B).

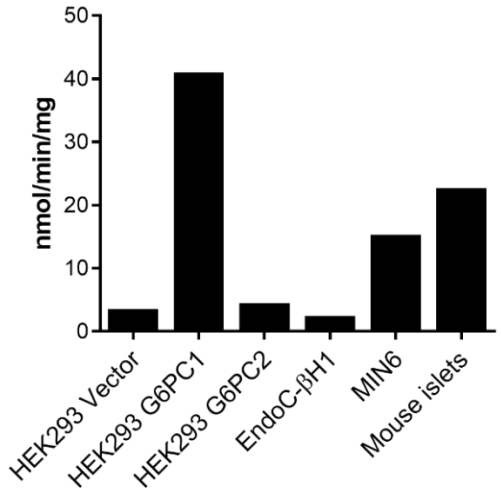


Figure 5.3.4.3. Comparison of glucose-6-phosphatase activity across different cell types. Reactions were carried out at 37°C for 10 min in the presence of 5 mM G6P. HEK293 cells were transfected with empty vector and human *G6PC1* and *G6PC2* constructs as indicated. Mouse islet samples were provided by Quan Zhang, OCDEM. Results were obtained from one experiment.

5.3.5. Phosphatase activity of G6PC2 variant proteins

Among the G6PC2 coding variants analysed for protein stability, I171T, I171V and F256L were found to be stably expressed and processed like WT in HEK293 and INS-1 (832/13) cells. We therefore set out to characterise their G6Pase activity. Variants at the associated sites were generated within the conserved sequences of the *G6PC1* construct by site directed mutagenesis. Based on the sequence homology comparison between the two isoforms where these variants are found (Figure 5.3.5.1), it was revealed that the I171 residue in G6PC2 mapped to L173 in G6PC1. Though one would expect the Ile and Leu residues to be biochemically similar, I first confirmed that both residues did not give rise to differential effects on protein expression in the G6PC1 or G6PC2 background (Figure 5.3.5.2). I also further demonstrated no difference in phosphatase activity between G6PC1-Leu173 (WT) and G6PC1-Ile173 (Figure 5.3.5.4).

I hypothesised that I171T and I171V may impact on enzymatic activity due to their proximity to the catalytic centre. Human G6Pase isoforms including G6PC2 belong to a superfamily of enzymes that contain a signature phosphatidic acid phosphatase motif (Stukey and Carman 1997). This superfamily also includes bacterial and fungal chloroperoxidase enzymes. As there is no crystal structure for G6Pase, I attempted to visualise the probable impact of mutations within the conserved catalytic domain of a vanadium-containing chloroperoxidase crystal structure (Messerschmidt et al. 1997). Figure 5.3.5.3 shows the physical proximity of the Leu493 residue (equivalent to Ile171 in G6PC2) to the ligand-binding site, suggesting that the amino acid substitution from the hydrophobic Leu residue to the polar Thr residue may affect active site conformation or ligand interactions. The Val residue remains a hydrophobic residue though it has a side chain similar in size to Thr. Microsomes from HEK293 cells expressing these variants (hence forth called G6Pase-L173T and G6Pase-L173V) were collected and quantified. Matched amounts of recombinant G6Pase protein (WT and variant) were added to assay mixes to account for any differences in expression, and evaluated for first order Michaelis-Menten enzymatic activity in response to 0 mM to 20 mM of G6P.

As seen in Figure 5.3.5.5, the Thr substitution at residue 173 resulted in a 20% decrease in activity as compared to the WT residue. V_{max} of G6Pase-WT was significantly higher than

that of G6Pase-L173T in both assays in the absence or presence of histones (P=0.01 and P=0.02 respectively). For reactions with histones, the K_m of G6Pase-L173T also appeared to have an upward trend compared to WT protein. To determine if K_m was truly affected in G6Pase-L173T, I performed assays matched for maximal G6Pase activity (at 20 mM G6P) and showed that the difference in K_m of WT and L173T was indeed enhanced though this was not significant probably due to small sample size, reflecting a possible decreased affinity of the variant for G6P (Figure 5.3.5.6). These results demonstrated that G6PC2-I171T is likely to reduce FG levels through reduced protein activity. Conversely, the valine substitution at the same site displayed enhanced phosphatase activity by ~20%. As seen in Figure 5.3.5.7, V_{max} of L173V was significantly higher than WT both in the absence and presence of histones (P=0.02 and P=0.03 respectively). In addition, K_m of L173V was also lower than that of WT (P<0.05 for reaction without histones). Therefore, I171V in G6PC2 is likely to be an activating mutation that results in increased G6PC2 activity leading to raised FG levels. These results together fully supported our observation from the genetic data that the two variants gave rise to opposite directions of effect on FG (β_{I171T}= -0.179 mmol/l; β_{I171V}= +0.263 mmol/l) and HbA1c (β_{I171T}= -0.007%; β_{I171V}= +0.093%).

As for the study of G6PC2-F256L, the G6Pase-F258L variant was evaluated for phosphatase activity. I observed that F258L had lower V_{max} as well as consistently higher K_m than WT, though only the impact on V_{max} was statistically significant in assays with histones (P=0.01, Figure 5.3.5.8). The activity of F258L also appeared to fit poorly to the Michaelis-Menten enzyme kinetics model (R^2=0.50). I have therefore showed that G6PC2-F256L is likely to be a deactivating variant, consistent with its glucose-lowering effects.

Finally, I investigated the phosphatase activity of G6Pase-R281X (G6PC2-R283X) and my results clearly showed that the truncated variant had completely abolished activity in reactions with and without histones, in response to increasing G6P concentrations (Figure 5.3.5.9). This was the case even after matching for the same amount of WT and variant protein added to the assay. Overall, through collaboration with the MAGIC investigators and T2D-GENES consortium we have reported further rare variants in G6PC2 that influence FG levels, and I have shown that a number of these variants have partial or complete loss of function likely through reduced G6Pase activity (I171T, F256L, R283X), as well as one variant with gain of function due to improved G6Pase activity (I171V).

```
G6PC1    152 nvilwlgfwavqlnvclsriylaahfphqvvagvlsgiavaetfshihsi    201
             ...||..||.:|::||:||:::|.||||||:.||:.|:.|||.|.|...|
G6PC2    150 wsflwsvfwliqisvcisrvfiathfphqvilgviggmlvaeafehtpgi    199

G6PC1    202 ynaslkkyflitfflfsfaigfylllkglgvdllwtlekaqrwceqpewv    251
             ..|||..|.....|||.||:|||||:.|.:||||::..|::||..|:|:
G6PC2    200 qtaslgtylktnlflflfavgfylllrvlnidllwsvpiakkwcanpdwi    249

G6PC1    252 hidttpfasllknlgtlfglglalnssmyresckgklskwlpfrlssiva    301
             ||||||||.|::|||.|||||.|:||.|:..||:|..:..|.|||...:.
G6PC2    250 hidttpfaglvrnlgvlfglgfainsemfllscrggnnytlsfrllcalt    299
```

Figure 5.3.5.1. Protein sequence alignment of G6PC2 (residues 150-299) and G6PC1 (residues 152-301) showing conservation of amino acids at residues I171 and F256 in G6PC2. Residues highlighted in yellow represent conserved active site residues in both proteins. Residues highlighted in blue represent residue I171 in G6PC2 and corresponding residue L173 in G6PC1. Residues highlighted in pink represent residue F256 in G6PC2 and corresponding residue F258 in G6PC1. Pairwise sequence alignment for G6PC1 (NP_000142) and G6PC2 (NP_066999) proteins were carried out using EMBOSS Water software (http://www.ebi.ac.uk/Tools/psa/emboss_water/). Vertical bars indicate identical residues, colons indicate similar residues, dots indicate dissimilar residues.

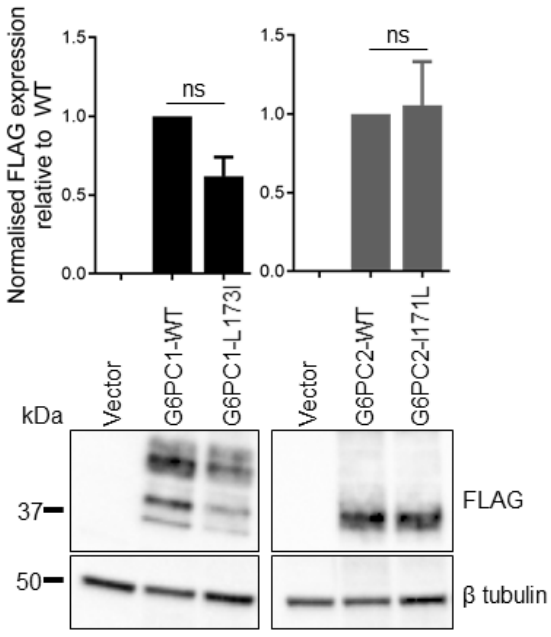


Figure 5.3.5.2. Western blot comparing expression levels of Leu173 (WT) with Ile173 in G6PC1 and Ile171 (WT) with Leu171 in G6PC2 in HEK293 cells. Graphs were plotted based on densitometry data obtained from three independent experiments each for G6PC1 and G6PC2. ns: not significant.

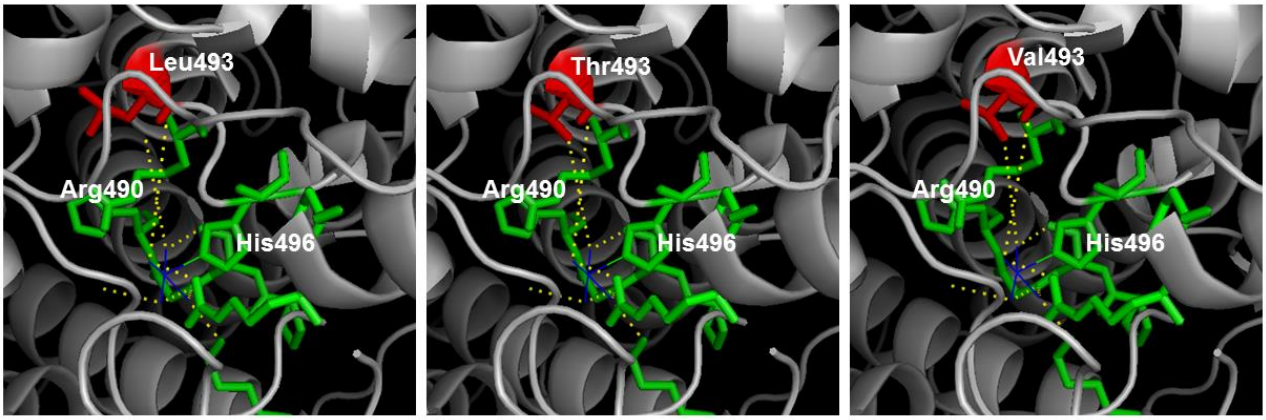
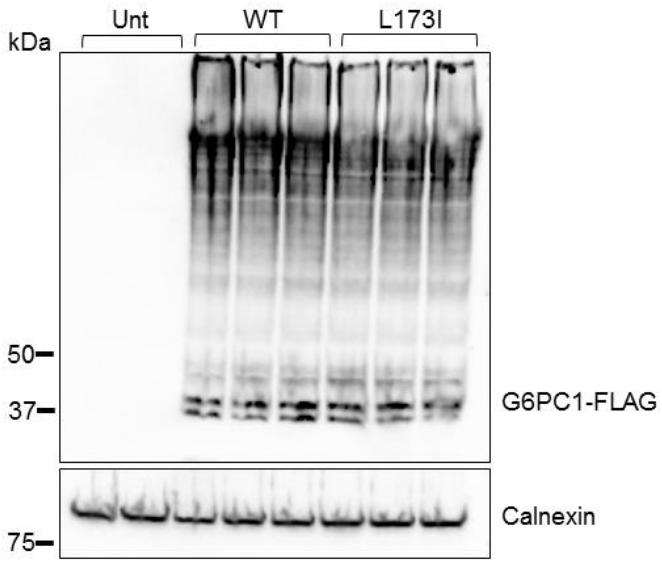


Figure 5.3.5.3. Visualisation of the amino acid substitutions at position 493 from leucine (WT, left) to threonine (middle) and valine (right) residues within the conserved catalytic domain of vanadium-containing chloroperoxidase. Residues in green (Arg490 and His496) are active site residues conserved in G6Pase proteins. They flank the residue marked in red (Leu493), which is physically located close to the binding pocket; mutated residues may affect active site conformation and catalysis. The ligand in blue is a vanadate ion. This crystal structure of the peroxide form of the vanadium-containing chloroperoxidase from *Curvularia inaequalis* was obtained from the Protein Data Bank, RCSB (PDB ID: 1IDU) and visualised using PyMOL.

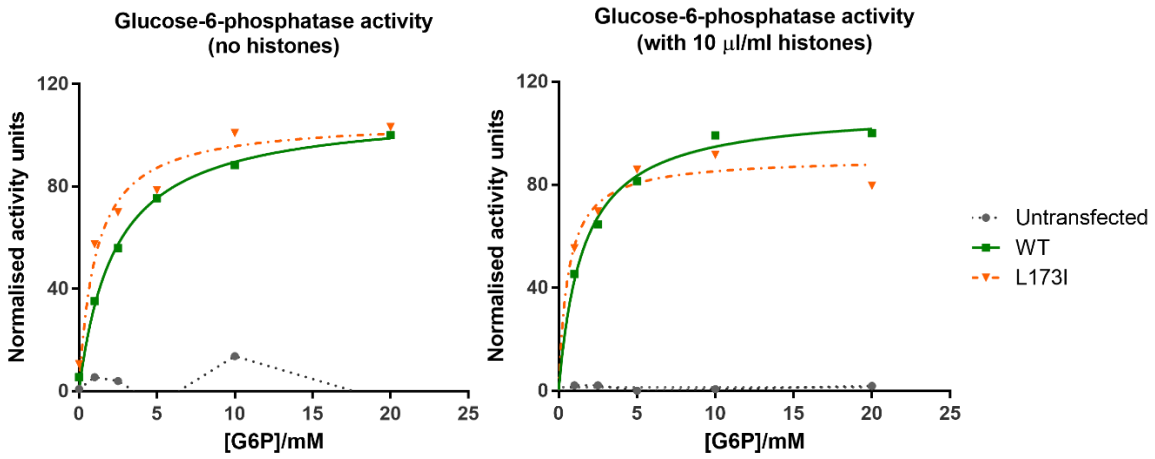
Figures for the representative microsomal western blots and activity assay plots for each variant are presented in the subsequent pages as follows:

- Figure 5.3.5.4. Results for G6Pase-L173I (G6PC2 WT)
- Figure 5.3.5.5. Results for G6Pase-L173T (G6PC2-I171T)
- Figure 5.3.5.6. Results for G6Pase-L173T (G6PC2-I171T) in matched activity assays
- Figure 5.3.5.7. Results for G6Pase-L173V (G6PC2-I171V)
- Figure 5.3.5.8. Results for G6Pase-F258L (G6PC2-F256L)
- Figure 5.3.5.9. Results for G6Pase-R281X (G6PC2-R283X)

A



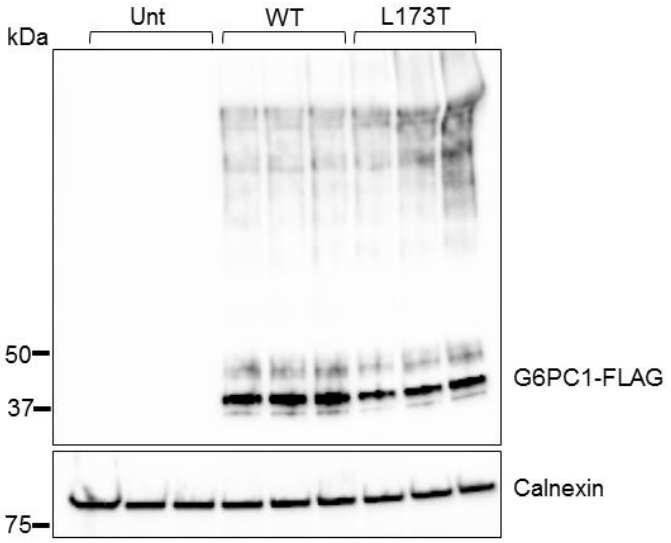
B



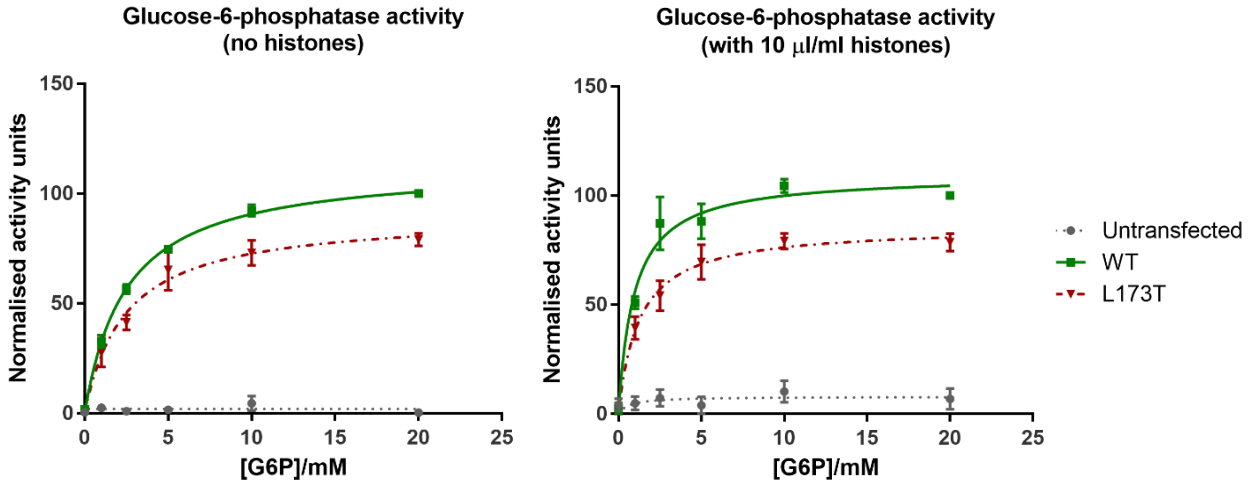
	No histones			With histones		
G6Pase mutation	**R^2**	**Vmax**	**Km**	**R^2**	**Vmax**	**Km**
WT	0.99	110.20	2.31	0.99	109.60	1.57
L173I	0.95	105.93	1.07	0.97	90.49	0.62

Figure 5.3.5.4. Glucose-6-phosphatase activity in microsomal preparations of G6Pase-L173I expressed in HEK293 cells. (A) Representative western blot of L173I in extracted microsomes with calnexin as the loading control. (B) Michaelis-Menten plot for G6Pase activity with varying concentrations of G6P. Results were presented as mean values normalised to activity of WT at 20 mM G6P ± SEM. Statistical tests were not carried out as data were obtained from one experiment. Unt: Untransfected control; WT: Wild type.

A

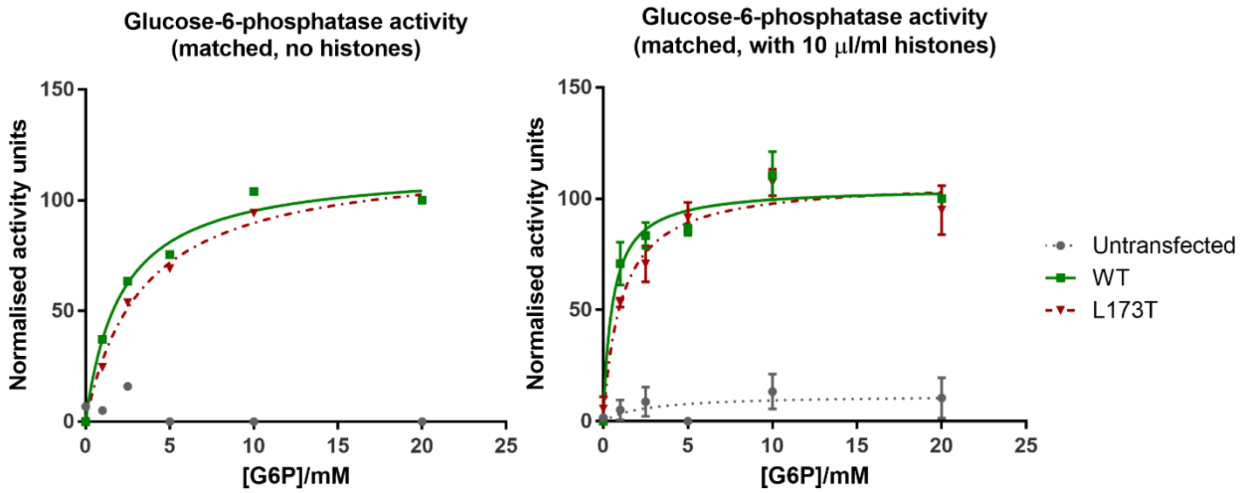


B



	No histones					With histones				
G6Pase mutation	**R^2**	**Vmax (SE)**	**P**	**Km (SE)**	**P**	**R^2**	**Vmax (SE)**	**P**	**Km (SE)**	**P**
WT	0.99	113.31 (2.32)	**0.0149**	2.49 (0.17)	0.9329	0.90	109.45 (5.51)	**0.0176**	0.95 (0.24)	0.4665
L173T	0.89	90.38 (6.38)		2.43 (0.61)		0.89	85.86 (4.75)		1.26 (0.31)	

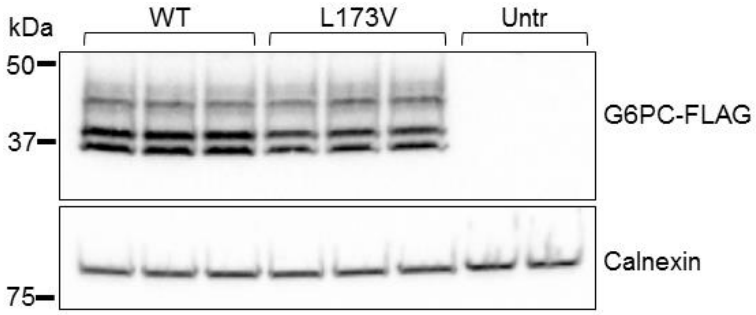
Figure 5.3.5.5. Glucose-6-phosphatase activity in microsomal preparations of G6Pase-L173T (G6PC2-I171T) in HEK293 cells. (A) Representative western blot of L173T in extracted microsomes with calnexin as the loading control. (B) Michaelis-Menten plot for G6Pase activity with varying concentrations of G6P for reactions with and without histones. Results were from 4 individual experiments presented as mean values normalised to maximum activity of WT at 20 mM G6P ± SEM. Unpaired t tests of the kinetic constants V_{max} and K_m were carried out to compare WT and variant. P values in bold indicate significant values ($P < 0.05$). R^2 indicates the goodness-of-fit for the non-linear regression. Unt: Untransfected control; WT: Wild type.



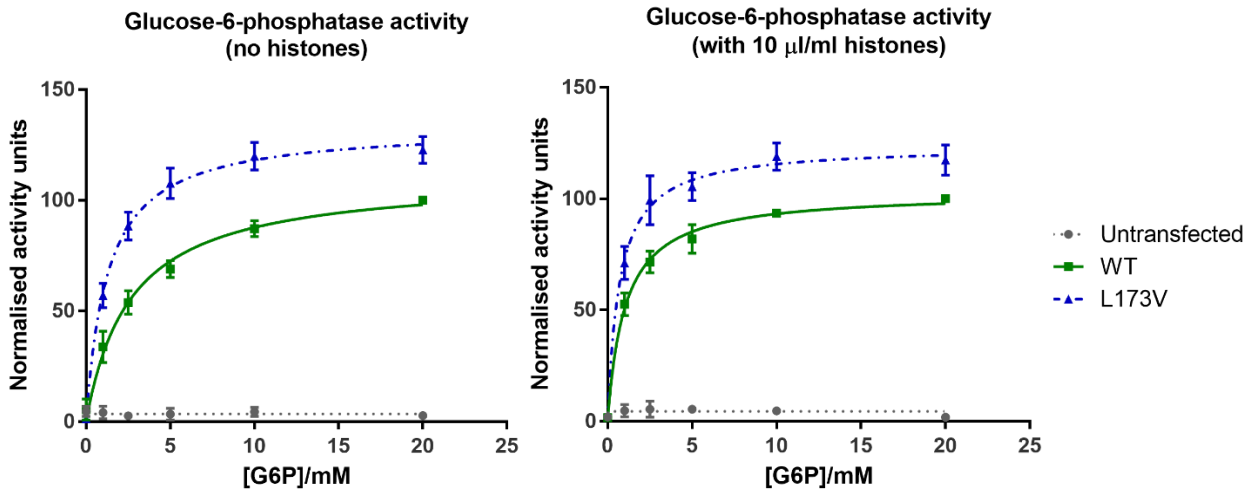
	No histones				With histones			
G6Pase mutation	**R^2**	**Vmax (SE)**	**P**	**Km (SE)**	**P**	**R^2**	**Vmax (SE)**	**P**	**Km (SE)**	**P**
WT	0.98	115.37 (6.65)		2.12 (0.44)		0.94	105.07 (5.27)	0.7374	0.55 (0.18)	0.2821
L173T	0.99	119.81 (5.23)		3.35 (0.45)		0.93	108.30 (6.55)		1.07 (0.31)	

Figure 5.3.5.6. Glucose-6-phosphatase activity in microsomal preparations of G6Pase-L173T (G6PC2-I171T) based on matched activity assays. In each experiment WT and variant proteins were added in different quantities such that maximum activity at 20 mM G6P for each sample were matched. Michaelis-Menten plot for G6Pase activity with varying concentrations of G6P for reactions with and without histones are shown. Results were from 1-2 experiments presented as mean values normalised to maximum activity of WT at 20 mM G6P ± SEM. Unpaired t tests of the kinetic constants Vmax and Km were carried out to compare WT and variant. R^2 indicates the goodness-of-fit for the non-linear regression. Unt: Untransfected control; WT: Wild type.

A



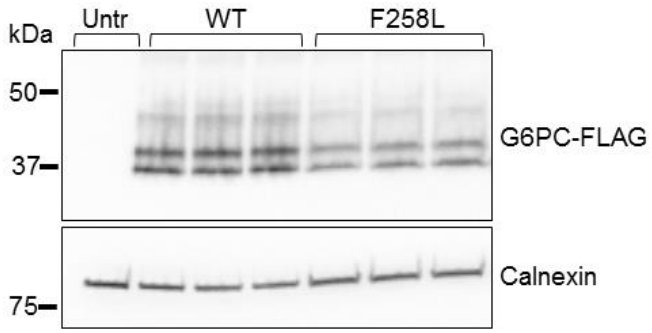
B



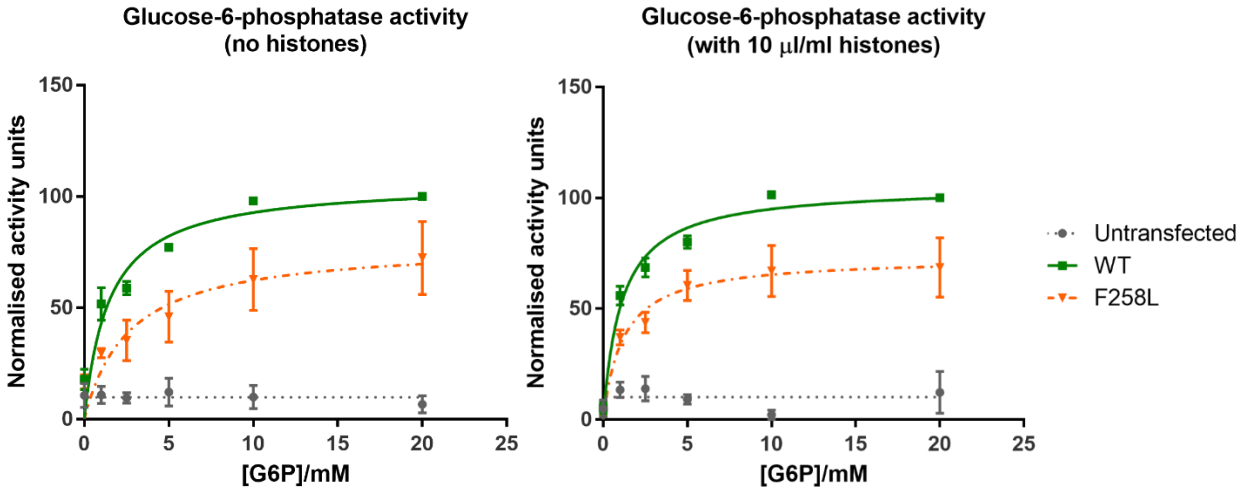
	No histones					With histones				
G6Pase mutation	**R^2**	**Vmax (SE)**	**P**	**Km (SE)**	**P**	**R^2**	**Vmax (SE)**	**P**	**Km (SE)**	**P**
WT	0.93	110.96 (5.87)	**0.0278**	2.65 (0.47)	**0.0378**	0.95	102.90 (3.42)	**0.0166**	1.04 (0.17)	0.2211
L173V	0.95	133.27 (5.03)		1.28 (0.21)		0.91	123.74 (5.33)		0.71 (0.18)	

Figure 5.3.5.7. Glucose-6-phosphatase activity in microsomal preparations of G6Pase-L173V (G6PC2-I171V) in HEK293 cells. (A) Representative western blot of L173V in extracted microsomes with calnexin as the loading control. (B) Michaelis-Menten plot for G6Pase activity with varying concentrations of G6P for reactions with and without histones. Results were from four individual experiments presented as mean values normalised to maximum activity of WT at 20 mM G6P ± SEM. Unpaired t tests of the kinetic constants Vmax and Km were carried out to compare WT and variant. P values in bold indicate significant values ($P < 0.05$). R^2 indicates the goodness-of-fit for the non-linear regression. Unt: Untransfected control; WT: Wild type.

A



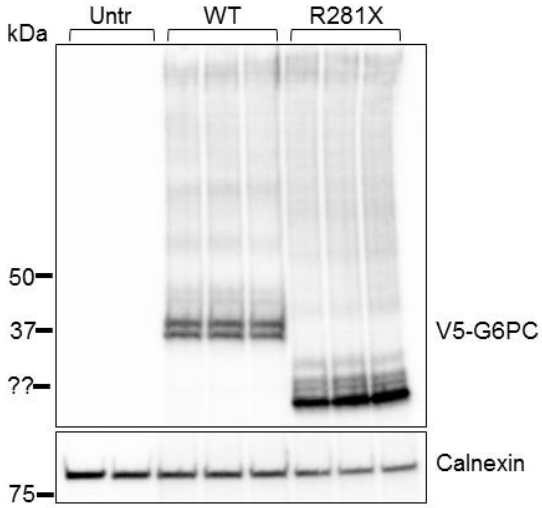
B



	No histones					With histones				
G6Pase mutation	**R^2**	**Vmax (SE)**	**P**	**Km (SE)**	**P**	**R^2**	**Vmax (SE)**	**P**	**Km (SE)**	**P**
WT	0.86	106.51 (6.63)	0.1447	1.48 (0.38)	0.5043	0.96	105.28 (3.71)	**0.0146**	1.09 (0.18)	0.8571
F258L	0.50	78.87 (13.77)		2.65 (1.55)		0.78	73.12 (6.86)		1.19 (0.51)	

Figure 5.3.5.8. Glucose-6-phosphatase activity in microsomal preparations of G6Pase-F258L (G6PC2-F256L) in HEK293 cells. (A) Representative western blot of F258L in extracted microsomes with calnexin as the loading control. (B) Michaelis-Menten plot for G6Pase activity with varying concentrations of G6P for reactions with and without histones. Results were from three individual experiments presented as mean values normalised to maximum activity of WT at 20 mM G6P ± SEM. Unpaired t tests of the kinetic constants Vmax and Km were carried out to compare WT and variant. P values in bold indicate significant values ($P < 0.05$). R^2 indicates the goodness-of-fit for the non-linear regression. Unt: Untransfected control; WT: Wild type.

A



B

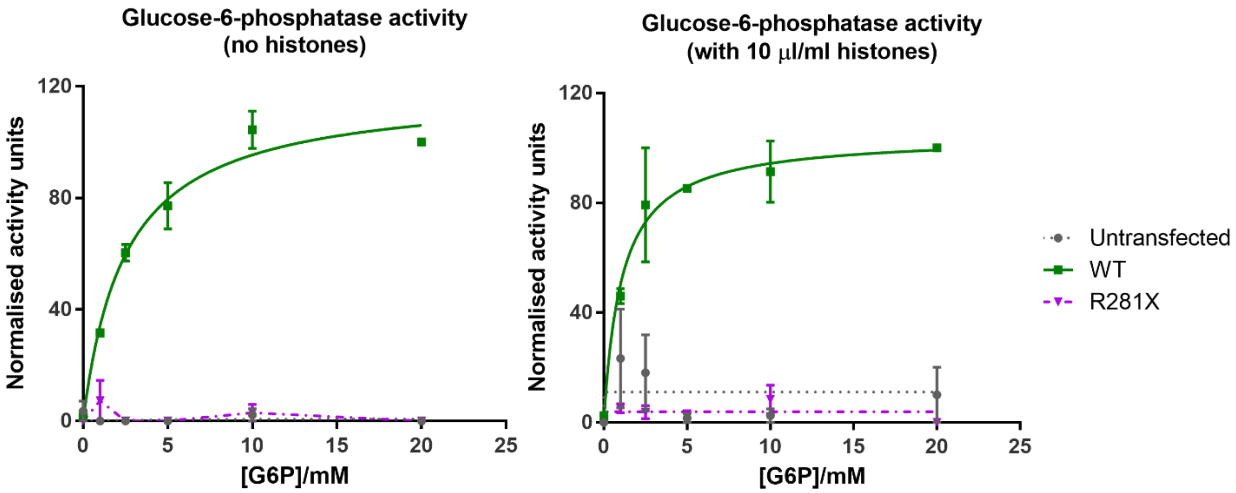


Figure 5.3.5.9. Glucose-6-phosphatase activity in microsomal preparations of G6Pase-R281X (G6PC2-R283X) in HEK293 cells. (A) Representative western blot of R281X in extracted microsomes with calnexin as the loading control. (B) Michaelis-Menten plot for G6Pase activity with varying concentrations of G6P for reactions with and without histones. Results were from two individual experiments presented as mean values normalised to maximum activity of WT at 20 mM G6P ± SEM. Unt: Untransfected control; WT: Wild type.

5.4. Discussion

This study is to date the most comprehensive characterisation of coding variants in the well-established glycemic trait locus *G6PC2*, as well as the first time variants are studied in the context of the EndoC-βH1 cells, a biologically-relevant cell model expressing endogenous *G6PC2*. From large-scale exome sequencing data we have obtained a more complete inventory of coding variation in *G6PC2* that provided me with the opportunity to follow up on rare coding variants that are likely to exhibit larger effects on protein function and hence glycemic trait levels, as compared to common non-coding GWAS variants.

I have functionally investigated a series of rare coding variants in *G6PC2* and showed that many of these altered protein function through diverse mechanisms, including both inactivating and activating mechanisms that resulted in reciprocal changes in FG and HbA1c levels. Specifically, a majority of the variants that had evidence for influencing FG and/or HbA1c levels in the MAGIC analysis exhibited LOF through loss of total protein stability (in HEK293 cells) or loss of glycosylation (in INS-1 [832/13] cells). For variants that did not give rise to reduced protein expression, I171T and F256L were found to exhibit partial loss of enzymatic activity, while I171V had a gain of activity. This is the first report of a catalytically-activating variant in *G6PC2* consistent with increasing FG and HbA1c. In addition, the only PTV to have been characterised, R283X, had completely abolished enzymatic activity and may also be mislocalised in cells.

My comprehensive characterisation studies have provided additional biological insight into the regulation of G6PC2. Most of the variants studied, including all the variants that map adjacent to or directly to conserved active sites in the non-helical domains, resulted in a LOF through protein instability. This highlighted the importance of the maintenance of structural integrity for G6PC2 that is tightly linked with functional activity. Conversely, my studies in G6PC1 showed that variants in G6PC1 are less likely to impact on protein stability, although a caveat here is that only a small number of variants were characterised (Chapter 4).

In both G6PC2 and G6PC1, glycosylation appeared to be important for function. However, whether glycosylation is a requisite for enzymatic activity has yet to be investigated in G6PC2. A previous study of recombinantly-expressed human G6PC2 in COS-1 cells showed that the non-glycosylated form of G6PC2 is normally less stable than the processed form

(Shieh et al. 2004). However, non-glycosylated G6PC1 is found to retain some residual activity (Chapter 4). The mechanism underlying the impact of missense G6PC2 variants on the glycosylation process in INS-1 (832/13) cells is also unclear, but could be due to disruption of structural integrity of the protein thus restricting access to the glycosylation site. G6PC2, like G6PC1, is now also found to normally localise to the Golgi apparatus, though the modifications that take place in the Golgi network is unknown as neither G6PC2 nor G6PC1 are secreted proteins. The importance of the C-terminal ER retention signal in G6PC2 ER localisation could not be definitively established due to technical difficulties with the current microscopy techniques and poor transfection efficiency of *G6PC2* in biologically-relevant cell lines. In addition, a recombinant G6PC2 protein construct with attached tags may not necessarily recapitulate the behaviour of the native protein in cells.

My functional studies have also been limited in other ways. First, G6PC2 activity itself could not be detected using the current transfection method and kinetic analysis, consistent with the difficulties that multiple groups have experienced previously as discussed earlier in chapter section 1.5. Though variants that localised within conservative sequences may be studied in the more enzymatically active G6PC1 isoform, this strategy only serves as an estimation of the probable effects of the variants on G6PC2 activity. There is currently no crystal structure for the G6Pase domain for accurate prediction of the impact of missense changes on active site conformation and interactions. However, advanced protein structure and variance modelling softwares are now widely available for modelling of the structural and functional effects of mutations (Kelley et al. 2015; Sillitoe et al. 2015); these analyses are not presented in this thesis. In my activity assays I have only investigated enzyme activity in response to G6P. There may however be other natural biological substrates of G6PC2 that provide diverse measures of function, though this will require evidence for the presence of transporters that transport these substrates into the ER lumen. Finally, due to the low throughput nature of the *in vitro* assays, which are both time-consuming and labour-intensive, it was not possible to comprehensively evaluate all variants to accurately assign functionality and differentiate between functional casual rare variants and neutral variants. Efforts to develop high throughput assays that have multiple phenotypic readouts will be valuable for such characterisation studies.

The study of an expanded allelic series in *G6PC2* has not only facilitated the discovery of novel functional mechanisms, but also increased our understanding of the range of effects (in terms of direction and magnitude) that genetic variation in this gene can exert on protein function to influence FG and HbA1c levels. From a therapeutic perspective, this is highly beneficial for the generation of a dose-response curve linking the extent of gene function to changes in FG levels, which can have a predictable effect on prediabetes.

Association analyses in large datasets have now revealed strong links between *G6PC2* variation and HbA1c, which is a measure of long term glucose control. The observation that *G6PC2* variants influence both FG and HbA1c levels but not other glycemic traits confirmed a strong beta cell effect. Based on the T2D-GENES exome array and sequencing analyses of T2D cases and controls, the FG-lowering alleles at rare variant S324P was found to be marginally associated with reduced T2D risk, in addition to common variant V219L (Anubha Mahajan, personal communication). This is consistent with the accepted dogma regarding the direction of effect of elevated FG and HbA1c levels on the risk of developing T2D. However, the associations were again not statistically significant, consistent with previous findings (Mahajan et al. 2015), indicating that even larger studies may be required to assess the role of functional rare coding variants in T2D susceptibility. Nonetheless, this confirmed that the mechanisms through which G6PC2 regulates beta cell function is most likely independent of processes that drive T2D pathophysiology. The precise role of G6PC2 on the regulation of pancreatic beta cell function and how it translates to influencing FG levels in humans remain of great interest. The investigation of *G6PC2* function in a human beta cell model will be discussed in Chapter 6.

Chapter 6

Investigating the role of G6PC2 in the pancreatic islet beta cells

6.1. Introduction

6.1.1. Current understanding of the role of G6PC2 in beta cell function

Recently published and unpublished large-scale exomes analyses and my functional characterisations of coding variants in *G6PC2* (presented in Chapters 3 and 5) have shed light on the relationship between *G6PC2* variation, FG levels and T2D risk. Reduced G6PC2 function in heterozygous carriers of LOF coding variants results in a reduction in FG and HbA1c levels, which is associated with protection against T2D susceptibility. However, in-depth mechanistic studies are needed to validate these observations and determine the direct/indirect roles of G6PC2 on beta cell function.

G6PC2 is among the genes that are most highly expressed in the beta cell population in published human islet transcriptomic data (Nica et al. 2013). *G6PC2* expression is also enriched in the beta cells of both humans and mice as compared to other islet cell types, though low levels may be found in subsets of alpha and gamma cells (Hutton and Eisenbarth 2003; Lawlor et al. 2017; Nica et al. 2013; Segerstolpe et al. 2016). *G6PC2* expression is not differentially expressed within the different subpopulations of beta cells identified in Segerstolpe et al, though this was the case in the two clusters of alpha cell populations. The various lines of evidence from genetic, molecular and rodent studies described in chapter sections 1.4-1.6 have pointed to a multitude of roles that G6PC2 may play in the regulation of beta cell function. The consensus in the field seems to be the generation of a G6PC2/GCK futile substrate cycle in the beta cell that acts as a regulatory checkpoint for modulating insulin secretion, as illustrated in Figure 6.1.1.1. Nonetheless, as G6Pase activity in islet cells is much lower than that detected in the liver and kidney, and overexpressed G6PC2 in transfected cells gave rise to low detectable phosphatase activity, if any, G6PC2 may be influencing beta cell function and hence FG through other mechanisms beyond glucose cycling and G6P hydrolysis (Marcolongo et al. 2013; O'Brien 2013).

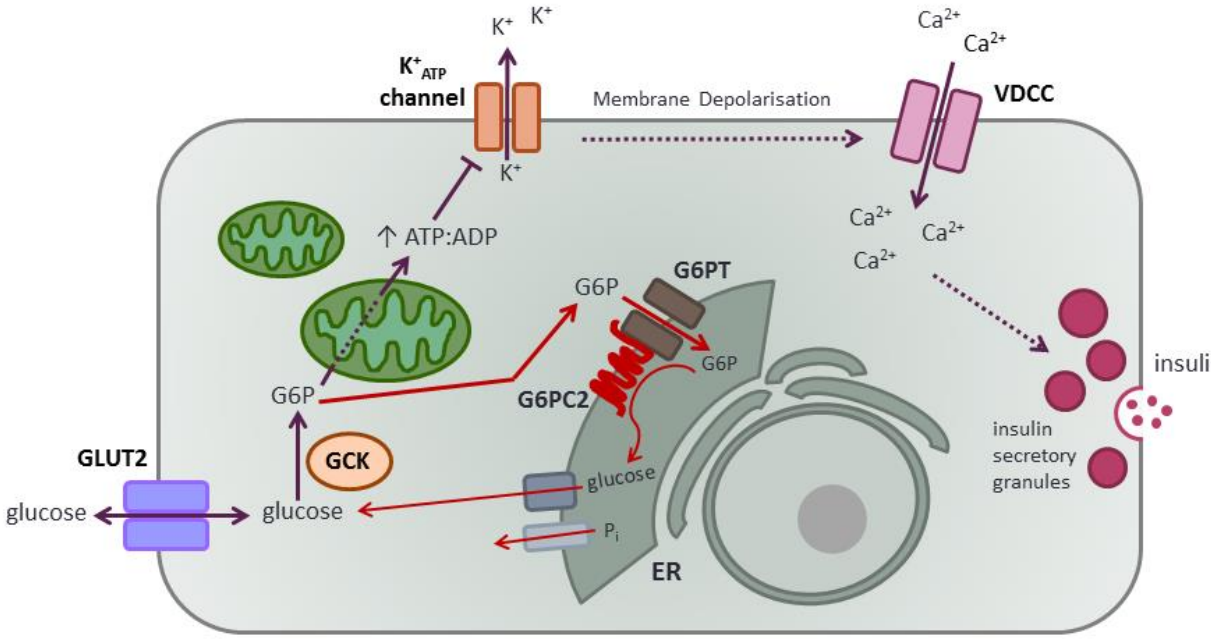


Figure 6.1.1.1. Schematic showing G6PC2 action in the context of glucose-stimulated insulin secretion in the beta cell. G6PC2 is localised to the ER membrane as part of the G6Pase complex also consisting of the G6P transporter (G6PT). G6P enters the ER lumen through G6PT and is hydrolysed to glucose and inorganic phosphates (P_i) which may be transported out of the ER through putative transporters. This is postulated to result in the G6PC2/GCK futile cycle.

6.1.2. The potential link between G6PC2 and ER homeostasis in the beta cells

Additional insight may be gained about the functional significance of G6PC2 through considering its role in the ER, given that beta cells, which are specialised secretory cells, are highly sensitive to ER stress. ER stress is a condition caused by cellular stressors such as oxidative stress, Ca^{2+} dysregulation or protein overexpression resulting in accumulation of inappropriately folded proteins in the ER. In an attempt to restore ER integrity, cells trigger a cellular response known as the unfolded protein response (UPR), which serves to alleviate ER stress through a number of different mechanisms including upregulating ER chaperones to enhance the protein-folding capacity of the ER, increasing ER-associated protein degradation (ERAD), reducing new protein synthesis, and promoting cell apoptosis when cells are unable to cope with the stress. As a result, the UPR helps the cell to adapt to stress and promote cell survival, or promote cell death when homeostasis cannot be restored (Kim et al. 2008; Ron and Walter 2007). G6PC2 may directly affect ER homeostasis such as through generation of unfolded or misfolded proteins in the ER lumen and/or generation of inorganic phosphates in the ER affecting calcium concentrations or redox status (Wang and Kaufman 2016).

Pancreatic islet beta cells are highly susceptible to ER stress due to the high rate of protein synthesis, in particular, that of insulin. Beta cells have a well-adapted ER to cope with the secretory workload but are sensitive to protein overload. For instance, the Akita mouse expresses a mutated form of the *Ins2* gene product, which causes accumulation of misfolded proinsulin 2 in the ER of its beta cells, leading to increased rates of beta cell death (Wang et al. 1999). Similarly, missense mutations in the insulin gene in humans that cause neonatal diabetes result in ER stress-induced apoptosis due to insulin misfolding (Colombo et al. 2008; Stoy et al. 2007). Another example whereby ER stress caused by protein defects may be responsible for disease pathophysiology involves LOF mutations in the *WFS1* gene (encoding the ER transmembrane protein wolframin), which are causal for the autosomal recessive disorder Wolfram syndrome. This disease is characterised by neurodegenerative conditions and a young-onset form of diabetes (Inoue et al. 1998; Strom et al. 1998). Perturbed *WFS1* function results in ER calcium depletion, disruption of ion homeostasis and protein folding mechanisms, promotes cell apoptosis and eventual death of the pancreatic

beta cells (Fonseca et al. 2005; Riggs et al. 2005; Takei et al. 2006; Yamada et al. 2006). These examples in addition to many others demonstrate the key role that ER stress plays in contributing to beta cell dysfunction, beta cell death and diabetes (Cnop et al. 2012; Eizirik et al. 2008; Hasnain et al. 2016). Therefore, maintenance of ER integrity is critical to the proper functioning of the insulin-secreting beta cells and in turn glucose homeostasis.

The ER stress response is often assessed by the activity of signalling pathways involved in the UPR. The UPR is known to be classified into three canonical signalling branches, each initiated by different signal transducers – protein kinase RNA (PKR)-like ER kinase (PERK), inositol-requiring protein-1 (IRE-1) and activating transcription factor-6 (ATF6). In the presence of unfolded or misfolded proteins in the ER, an important ER chaperone BiP binds to these proteins and dissociates from the three ER-resident stress signal transducers, which go on to activate the UPR in both distinct and overlapping ways (Back and Kaufman 2012; Ron and Walter 2007). The characteristic signalling markers of each of these different arms are summarised in Figure 6.1.2.1. Ultimately, the UPR seeks to restore ER homeostasis, but in the case of chronic stress such as misfolding of mutant proteins or ER Ca^{2+} dysregulation that remains unresolved, apoptosis is triggered to eradicate unhealthy cells (Back and Kaufman 2012; Cnop et al. 2012). Deficiencies in the UPR pathway, such as that caused by mutations in *PERK*, lead to ER dysfunction and can cause Wolcott-Rallinson syndrome, a rare autosomal recessive disorder characterised by insulin-dependent permanent neonatal diabetes (Delepine et al. 2000).

A recent study established a potential link between *G6pc2* levels and ER stress that is regulated through sorcin, a calcium sensor protein. Sorcin maintains ER Ca^{2+} stores and intracellular Ca^{2+} fluxes in rodent islets and beta cells (Marmugi et al. 2016; Noordeen et al. 2012). Importantly, in lipotoxic conditions, endogenous sorcin levels are reduced while *G6pc2* expression is increased. This result corresponded strongly with elevation of ER stress markers *CHOP* and *BiP*. The converse was also true in transgenic mice overexpressing sorcin (Marmugi et al. 2016). These findings suggested a possible regulation of *G6pc2* expression by sorcin that may together impact on GSIS through influencing calcium and ER homeostasis.

As G6PC2/G6pc2 peptides have also been identified as major autoantigens in human and mouse models of T1D as described in Chapter 1, a study of transgenic mice overexpressing human *G6PC2* in the pancreatic beta cells was conducted to investigate epitopes

responsible for driving immune responses. It unexpectedly uncovered new perspective by showing that islets from a number of transgenic lines exhibited signs of ER stress such as dilation of the ER lumen (a structural hallmark of the UPR), increased *CHOP* expression and beta cell necrosis leading to reduced insulin production and early onset diabetes that was independent of immune-mediated diabetes (Shameli et al. 2007).

Overall, a small number of studies have hinted at interactions between G6PC2 and ER stress response pathways, but few have directly examined the role of G6PC2 in regulating beta cell ER homeostasis, whether under normal or stressed conditions. From my previous work in Chapters 3 and 5 we also know that numerous coding variants in G6PC2 that influence FG levels in humans are characterised by protein instability and aberrant translational modification in terms of protein glycosylation. These LOF variants may provide us with an additional tool to begin interrogating the effects that variation in this ER transmembrane protein may exert on ER integrity.

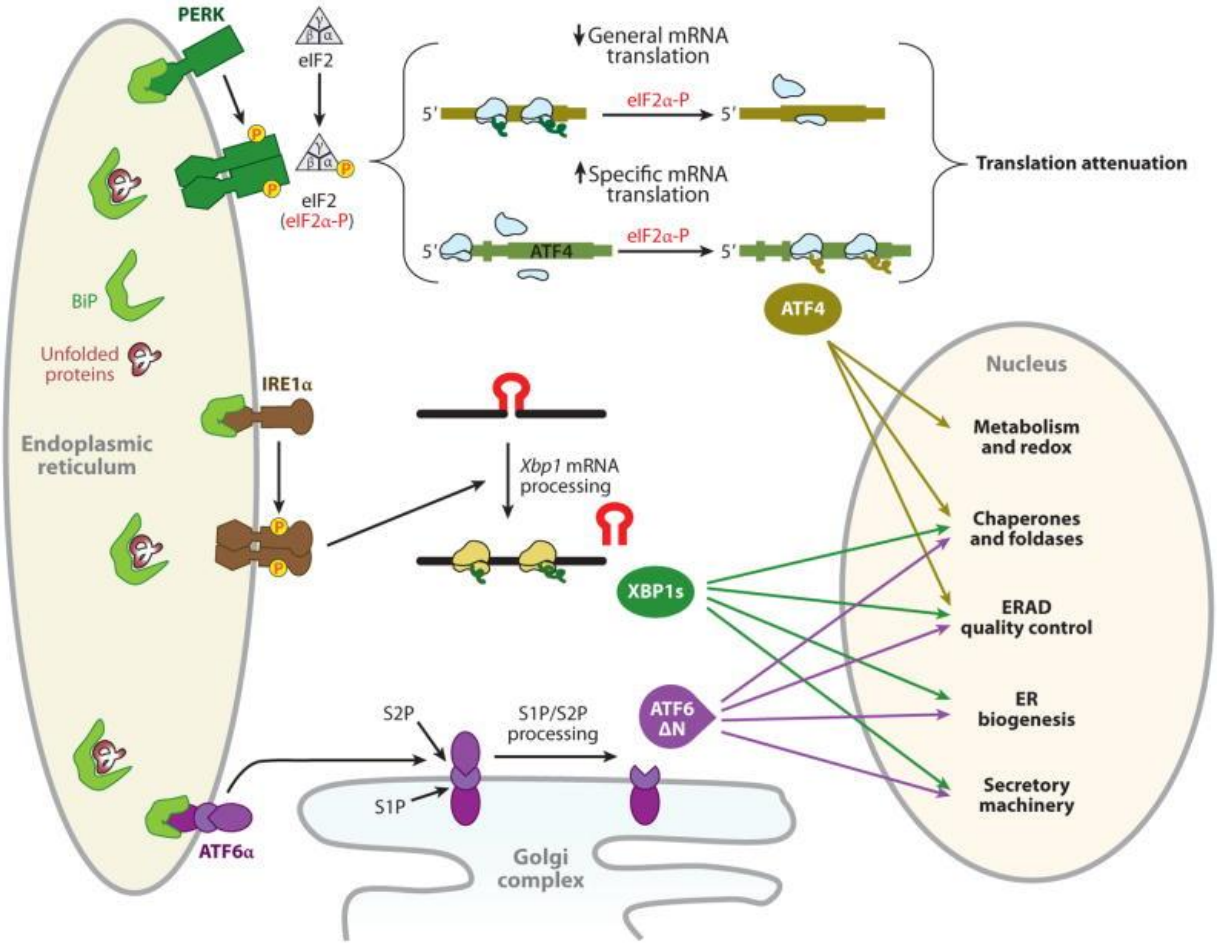


Figure 6.1.2.1. Schematic describing the canonical unfolded protein response (UPR) signalling pathways (figure from Back and Kaufman 2012). Under ER stress, ER chaperone BiP is sequestered through binding to unfolded/misfolded proteins thereby leading to the activation of UPR pathways initiated by three ER stress sensors. First (top of figure), activated PERK phosphorylates the α subunit of eIF2α which is a translation initiation factor. Phosphorylated eIF2α inhibits general mRNA translation to reduce ER workload and increases expression of transcription factor *ATF4* which goes on to promote the expression of various adaptive genes such as those involved in ER protein folding and the anti-oxidative stress response including *CHOP*. Next, activation of IRE-1 promotes the splicing of *XBP1* mRNA to express the potent transcription factor XBP1s, which upregulates the expression of ER chaperones and transporters, folding and enzymes and components of the ERAD quality control machinery. Finally, activated ATF6 is translocated to the Golgi complex where it is translationally processed to produce a cleavage product that then acts as a transcription factor to also promote expression of genes involved in ER expansion, ERAD and protein folding including *BiP* itself. Ultimately, these pathways result in gene expression changes that promote the restoration of ER integrity.

6.1.3. EndoC-βH1, a functional human beta cell model

Functional studies of beta cell function have largely been carried out on rodent cell models including both cell lines and primary tissue. Rodent models have enabled genetic manipulation in ways that cannot always be achieved in human studies and have provided valuable mechanistic insight into gene function and disease pathophysiology. However, rodent cells differ from human cells in many critical and sometimes lesser known ways, including islet architecture and the expression and activity of ion channels and transporters in the beta cells, affecting the metabolic regulation of insulin secretion. (Cabrera et al. 2006; McCulloch et al. 2011; Rorsman et al. 2012; Rorsman and Braun 2013). Therefore, hypotheses tested in rodent cells should be confirmed in human cells.

Sources of human cell models used for studying pancreatic islet and beta cell function include primary human islets isolated from cadaveric donors, induced pluripotent stem cells (iPSCs) as well as immortalised beta and other islet cell lines. Isolated human islets are often considered the gold standard for islet research, but these tissues are in limited supply and often of variable quality due to different donor characteristics and poor tolerance to isolation procedures and culture conditions. iPSCs are derived from easily accessible human somatic cells that can be differentiated using well-defined cocktails of growth factors into beta-like cells with presumably unlimited lifespan (Pagliuca et al. 2014; Rezania et al. 2014; Russ et al. 2015). These are becoming an increasingly attractive alternative with the development of robust protocols for differentiation, genome manipulation tools and the potential for disease modelling. Nonetheless, the generation of functional beta-like cells from iPSCs remains a complex, time-consuming and expensive process, and is therefore less efficient than clonal cell lines. Issues with heterogeneity in cellular phenotype also pose additional challenges. Numerous human beta cell lines have been developed but few were successful as reliable functional beta cell models and consequently became less often utilised (de la Tour et al. 2001; Levine et al. 1995; Narushima et al. 2005; Wang et al. 1997).

More recently, a promising human beta cell line (EndoC-βH1) was generated from derived human fetal pancreatic buds transduced with a SV40LT-expressing lentiviral vector under the control of the insulin promoter, that were transplanted into immune-compromised mice for *in vivo* maturation and expansion (Ravassard et al. 2011). Subsequent generations of this

cell line have since been developed (EndoC-βH2 and EndoC-βH3) to allow for conditional excision of the immortalizing transgenes to produce less proliferative cells that better recapitulate the functionality of adult beta cells (Benazra et al. 2015; Scharfmann et al. 2014).

Despite slow growth rates and finicky culturing conditions, the EndoC-βH1 human cell line displayed robust though modest insulin secretion in response to glucose and other insulin secretagogues. It has provided diabetes researchers with an unlimited source of cells for mechanistic studies of beta cell biology. Many fundamental studies carried out in this cell line have since been published, including detailed characterisations of glucose metabolism, calcium signalling, stimulus-coupled secretion as well as sensitivity to cytokine-induced toxicity (Andersson et al. 2015; Gurgul-Convey et al. 2015; Gurgul-Convey et al. 2016; Krishnan et al. 2015). More importantly, studies have also made use of this cell model for functional validation of genes or genetic regions that have been associated with T2D risk or monogenic forms of diabetes, to uncover possible causal mechanisms and gain new biological insights (Chandra et al. 2014; Collins et al. 2016; Gaulton et al. 2015; Pal et al. 2016; Thomsen et al. 2016b). In the case of the T2D risk locus *CDKN2A/B*, *CDKN2A* has been highlighted as a candidate effector transcript due to previous work in rodents that linked the tumour suppressor gene to glucose homeostasis and diabetes pathogenesis (Krishnamurthy et al. 2006b; Rane et al. 1999). The EndoC-βH1 cells provided a unique opportunity to investigate the significance of *CDKN2A* (encoding p16) on beta cell function independent of its role as a cell-cycle regulator, as the cells constitutively express the proto-oncogene SV40LT which masks the effects of p16 on cell cycle control. This study, which I had also contributed to with western blot and microscopy analysis, revealed that the pleiotropic gene *CDKN2A* is also directly involved in the regulation of GSIS in human beta cells (Pal et al. 2016).

As with all clonal cell lines, these cells also exhibit a few important differences from genuine adult beta cells. Not only are they continuously proliferating cells, the expression of some beta cell markers and the insulin content in these cells were significant lower than that in primary beta cells and other rodent cell lines, which may in part impact on the magnitude of secretory responses (Brandhorst et al. 1998; Ravassard et al. 2011; Weir and Bonner-Weir 2011). In addition, cell growth, morphology and function have to be closely monitored over

passages. Regardless, the EndoC-βH1 cells provide a good starting point for the investigation of the role of G6PC2 in the beta cell.

6.1.4. Experimental aims

From human genetic studies and detailed functional analyses we have learnt the importance of *G6PC2* in the regulation of glycemic traits, but the exact mechanisms linking *G6PC2* function to phenotype remains unclear. This chapter seeks to address this question through a series of exploratory studies to provide preliminary insight into the functional mechanisms underlying *G6PC2* in human cells. Prior studies have been primarily conducted in rodent models. I therefore set out to investigate the role of endogenous G6PC2 in the EndoC-βH1 cells, a glucose-responsive human beta cell model.

The aims of this study were to:

1. Establish the tissue expression profile of human *G6PC2* expression across a panel of human tissues and cell lines
2. Investigate the role of G6PC2 in the regulation of insulin and insulin release in response to stimulation in EndoC-βH1 cells
3. Investigate the impact of G6PC2 coding variants on ER homeostasis in a cellular expression model
4. Investigate the role of endogenous G6PC2 on ER homeostasis in EndoC-βH1 cells

6.2. Methods

6.2.1. Analysis of *G6PC2* gene expression

RNA was collected from cell lines or tissue samples for analysis of gene expression. RNA from cell lines were extracted using the TRIzol® reagent (Life Technologies) as described in 2.12.1. RNA of tissue samples were obtained from the commercially-available Human Total RNA Master Panel II (Clontech) and FirstChoice Human Total RNA Survey Panel (Ambion). Additional islet and pancreas tissue samples were obtained in-house. cDNA synthesis was carried out using the SuperScript III First-Strand Synthesis kit (Life Technologies) and analysed using the TaqMan® real time quantitative PCR procedure as described in 2.12. The TaqMan® gene expression assays (Applied Biosystems) used are listed in Table 6.2.1.1.

For the analysis of human tissue and cell panel for *G6PC2* expression, 1 µg of input RNA was primed with a mixture of oligo(dT) and random hexamers. Quantitative PCR data were analysed using the comparative Ct method ($2^{-\Delta\Delta Ct}$) – Ct values were first normalised to the sample with the lowest Ct value (calibrator sample) within each gene set, followed by subsequent normalisation of the gene of interest to that of housekeeping genes as indicated (HKGs). Where more than one HKG was used, mean values of both HKGs were calculated.

For the analysis of *G6PC2* knockdown in EndoC-βH1 cells, 50-250 ng of input RNA was used. *G6PC2* expression was measured using the FAM-labelled *G6PC2* primer-probe set spanning exons 2 and 3 (see Table 6.2.1.1) in multiplex with a VIC-labelled *TBP* assay.

6.2.2. RNA interference

siRNA-mediated RNA interference was carried out on EndoC-βH1 cells by Lipofectamine-based reverse transfection as described in section 2.14. siRNA complexes were prepared from ON-TARGETplus human siRNA (Dharmacon, GE Healthcare) to a final concentration of 25 nmol/L siRNA unless otherwise stated. The target sequences of the individual siRNAs are listed in Table 6.2.2.1. All sequences map to exons 1, 2 or 3, and hence target both NM_021176.2 and NM_001081686.1 *G6PC2* transcripts. The target sequence of the non-targeting control siRNA used has been verified to ensure no overlap with any known T2D-

associated gene locus that may affect insulin secretion (Martijn van de Bunt, personal communication).

Target gene	Assay number	Target region (RefSeq transcript)
G6PC2	Hs01549773_m1	Spans exon 2-3 boundary (NM_021176.2, NM_001081686.1)
G6PC2	Hs00221787_m1	Spans exon 3-4 boundary (NM_021176.2)
GAPDH	Hs99999905_m1	Exon 3 (NM_002046.4)
INS	Hs00355773_m1	Spans exon 2-3 boundary (NM_000207.2)
PPIA	Hs99999904_m1	Exon 4 (NM_021130.3)
TBP	Hs00427620_m1	Spans exon 2-3 boundary (NM_001172085.1) and exon 3-4 boundary (NM_003194.4)

Table 6.2.1.1. TaqMan® gene expression assays used for real time PCR and their target amplification regions. All assays were obtained from Applied Biosystems.

Target (name used in text)	Target sequence (5' to 3')	Catalogue No.
G6PC2 (siG6PC2-05)	UCCAAGAAACUCAGAUUUA	J-014015-05
G6PC2 (siG6PC2-06)	GUCCAUGCCUUGAACAGUU	J-014015-06
G6PC2 (siG6PC2-07)	CUACCGAGCUUACUACACU	J-014015-07
G6PC2 (siG6PC2-08)	GGAUGGAUAAGUUCUCUAU	J-014015-08
Non-targeting control #1	UGGUUUACAUGUCGACUAA	D-001810-01-05

Table 6.2.2.1. siRNAs used for *G6PC2* gene silencing experiments. siRNAs were purchased as ON-TARGETplus human siRNA from Dharmacon (GE Healthcare) as single oligonucleotides or pool of four oligonucleotides (SMARTpool). All *G6PC2* siRNAs were directed against the ORF.

6.2.3. Insulin secretion studies

Glucose-stimulated insulin secretion following *G6PC2* knockdown was assessed in EndoC-βH1 cells. Cells were cultured as described in section 2.6. For gene knockdown experiments 2 x 10^4 cells were seeded per well in a 96-well plate for siRNA transfection as described in 6.2.2. For constitutive secretion time course experiments transfected cells were maintained in the 5.5 mM glucose-containing culture medium and insulin samples were collected every 24h up to 120h post-transfection. For static incubations transfected cells were incubated with glucose in the absence or presence of chemical modulators in the respective conditions for 1h before insulin samples were collected, as described in 2.16.1. Cell counts were measured using the CyQUANT cell proliferation assay (Thermo Scientific) as described in section 2.15. Four independent experiments each with conditions tested in quadruplicate wells were carried out. For every experiment additional cells were collected for analysis of *G6PC2* knockdown efficiency.

Insulin was measured using the AlphaLISA assay (Perkin Elmer) as described in 2.16.2. Unknown sample values were interpolated from a 10-point insulin standard curve fit using five-parameter non-linear regression of log-transformed insulin count data. The regression analysis was performed on R using a script developed by Soren Thomsen at OCDEM. Insulin secretion values were expressed as insulin normalised to cell count for each well and subsequently normalised to siNT-transfected cells (at 1 mmol/L glucose in static incubation experiments or 48h knockdown in constitutive secretion experiments) of each experiment unless otherwise stated. Content-normalised insulin secretion values were expressed as insulin normalised to content instead of cell count for each well. The mean of normalised data was taken for statistical analyses.

6.2.4. ER stress response studies using gene expression analysis

For the study of *G6PC2* in ER homeostasis, *G6PC2* variant expression studies were carried out in HEK293 cells while *G6PC2* knockdown studies were carried out in EndoC-βH1 cells. HEK293 cells and EndoC-βH1 cells were seeded at 1.3 x 10^5 and 1.2 x 10^5 cells per well respectively in 24-well plates and transfected according to protocols described in sections 2.7 and 6.2.2 respectively. ER stress was induced in EndoC-βH1 cells by treatment with 1 µM thapsigargin (Sigma Aldrich) in DMSO or 5 µg/ml tunicamycin (Sigma Aldrich) in DMSO for 16h. DMSO was added in equivalent volumes to cells in separate wells as a negative vehicle control.

RNA extraction from transfected cells, cDNA synthesis and TaqMan® qPCR analysis were carried out as previously described in 2.12 and 6.2.1. Up to 500 ng of input RNA was primed with a mixture of oligo(dT) and random hexamers. Details of the TaqMan® gene expression assays used are shown in Table 6.2.1.1. above and Table 6.2.4.1 below. Results were expressed as fold change in the expression of the genes of interest (normalised to HKGs) relative to control cells as indicated.

Target gene	Assay number	Target region (RefSeq transcript)
ATF4	Hs00909569_g1	Spans exon 1-2 boundary (NM_001675.2)
BiP (HSPA5)	Hs99999174_m1	Spans exon 1-2 boundary (NM_005347.4)
CHOP (DDIT3)	Hs99999172_m1	Spans exon 1-2 boundary (NM_001195053.1, NM_001195057.1)
XBP1s	Hs03929085_g1	Exon 5 (NM_001079539.1)

Table 6.2.4.1. TaqMan® gene expression assays used in ER stress response studies. All assays were obtained from Applied Biosystems.

6.2.5. ER stress response studies using luciferase reporter-based assays

To further evaluate the level of ER stress in HEK293 cells transiently expressing *G6PC2* variants, a luciferase reporter-based assay was used to measure the activity of ER stress response cis-acting enhancer/promoter elements which were also co-expressed in the same cells. A Dual Luciferase Reporter system which makes use of the bioluminescent properties of the enzymes Firefly and Renilla luciferases, enable simultaneous expression and measurement of the activity of both reporters. First, four different consensus ER stress response elements were cloned into separate pGL3-Promoter vector constructs (Promega) carrying the Firefly luciferase gene by Dr Nicola Beer, a post-doctoral researcher in my group at OCDEM. These elements are known as ER stress elements (ERSE-I and ERSE-II) and UPR elements (UPRE-P and UPRE-W), which are activated by both distinct and overlapping mechanisms to activate downstream stress response pathways (Wang et al. 2000; Yamamoto et al. 2004). Binding of transcription factors and activators such as ATF6 and XBP1s to these elements will drive Firefly luciferase expression which is used as a proxy for the induction of stress response pathways. A summary of the experimental setup and sequences are shown in Figure 6.2.5.1 and Table 6.2.5.2. Second, the Renilla luciferase gene-containing vector pRL-CMV (Promega), which is co-transfected and constitutively expressed, acts as an internal control for technical variability including differences in cell plating densities, cell viability, transfection efficiency and transfer of samples.

All reporter assays were performed in HEK293 cells, which lack endogenous *G6PC2*. Cells were plated in 24-well plates at 7 x 10^4 and transfected the next day using the FuGene 6 transfection reagent (Promega). Each well of cells was transfected with a total of 250 ng of DNA comprising 124 ng of WT or mutant pCMV6-*G6PC2*, 124 ng of pGL3-Promoter carrying an ER stress response element (ERSE-I, ERSE-II, UPRE-P or UPRE-W), and 2 ng of pRL-CMV which was used as the internal transfection control vector. The Renilla construct pRL-CMV was selected as its strong promoter CMV (cytomegalovirus) generated luminescence values of similar magnitude to the pGL3-Promoter constructs. For negative control wells, an empty pGL3-promoter vector was used. Cells were transfected with a DNA to FuGene ratio of 1:4 in a total volume of 500 µl complete culture medium, with no media change required after transfection. Each condition was repeated in triplicate wells and at least three such

independent experiments were carried out. Cells were lysed 48h after transfection by removing the culture medium over cells and adding 100 µl of 1X passive lysis buffer (Promega) made up in distilled H_2O to each well. Culture plates were placed on an orbital shaker for 20 min at room temperature for cell lysis. Plates were then sealed with parafilm and stored at -80°C until analysis.

The luciferase assay protocol was adapted from manufacturer's instructions for the Dual Luciferase Assay System (Promega). Briefly, cell lysates were thawed and aliquoted into white 96-well ½-area microplates (Perkin Elmer) with 5 µl of lysate per well. Fifty microlitres each of Luciferase Assay Reagent (LAR II) and Stop & Glo Reagent was used per well, and luminescence was read on the Enspire multimode plate reader (Perkin Elmer). All firefly luciferase values were first normalised to Renilla values for each well, and results were calculated as average activity units relative to the activity of WT G6PC2 protein.

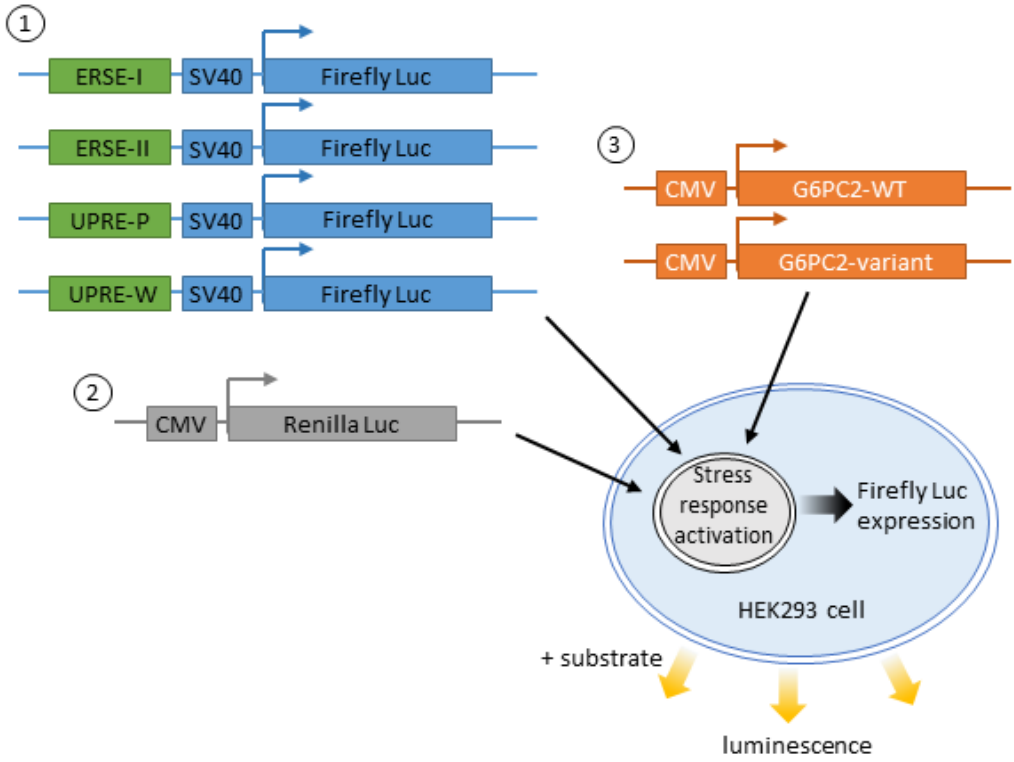


Figure 6.2.5.1. Schematic for luciferase reporter assay designed for studying G6PC2 variant expression on activation or ER stress response. Constructs carrying each stress response element luciferase plasmid (1), Renilla luciferase plasmid (2) and G6PC2 WT or variant plasmid (3) were co-transfected into HEK293 cells. After 48h transfection, cells were collected and ER stress response element activity was quantified by the Promega Dual Luciferase Assay according to the manufacturer's protocol. SV40 and CMV are promoters.

Element	Consensus sequence motif	Relevant features
ERSE-I	CCAATGGGCGGCAGCCACG	Binds cleaved ATF6 (preferential) in the presence of NF-Y binding. Controls ER chaperone genes such as *BiP*.
ERSE-II	ATTGGGCCACG	Binds cleaved ATF6 (preferential) in the presence of NF-Y binding and XBP1s in the absence of NF-Y. Controls ER chaperone genes such as *BiP*.
UPRE-P	TGACGTGG	Binds XBP1s (preferential) in the absence of NF-Y. May control expression of components of the ERAD machinery.
UPRE-W	TGACGTGG	As UPRE-P above.

Table 6.2.5.2. ER stress and UPR element consensus sequences used in the luciferase reporter assays. Each sequence motif (underlined) is repeated three times to form a trimer that is inserted into the plasmid for increased sensitivity. Luciferase reporters were cloned by Dr Nicola Beer. UPRE-P and UPRE-W have the same sequence motif but different flanking sequences. NF-Y is a general transcription factor.

6.2.6. Statistical analysis

For insulin secretion analyses unpaired two-tailed Students' t tests were carried out comparing normalised mean data of siG6PC2-transfected cells (KD) to siNT-transfected cells (controls) for each stimulation condition or time point. For analysis of ER stress qPCR data in HEK293 cells, unpaired two-tailed Students' t tests were carried out to compare each G6PC2 variant to WT. For analysis of *G6PC2* expression in these experiments, one-way ANOVA was carried out to assess for differences in transfection efficiency across transfected cells. For analysis of ER stress qPCR data in EndoC-βH1 cells, unpaired two-tailed Students' t tests comparing KD to control were carried out. Finally, for analysis of ER stress luciferase assay results luminescence values were first background-corrected and then normalized to WT for each experiment. Two-way ANOVA with Fisher's LSD post-hoc test was carried out to compare mean activities from G6PC2 WT- and variant-transfected cells for each luciferase reporter. Data analyses were carried out on Microsoft Excel or R; statistical tests and plotting of graphs were carried out on GraphPad Prism 7.0. Data were considered statistically significant at $P<0.05$. Further details are described in Chapter 2 Section 2.17.

6.3. Results

6.3.1. Specificity of *G6PC2* expression in islet beta cells

To confirm the tissue specificity of *G6PC2* expression, I measured mRNA levels across 28 human tissues and the human beta cell line EndoC-βH1. Commercially-available gene expression assays that were specific to the long *G6PC2* transcript (detecting exon 3-4 boundary) or to both the short and long transcripts (detecting exon 2-3 boundary) were used. Gene expression analysis as shown in Figure 6.3.1.1 showed that *G6PC2* was mostly highly expressed in the islets. *G6PC2* expression was also detected in the pancreas but at much lower levels than that in the islets, indicating that its expression was largely confined to islets than to exocrine cells of the pancreas. *G6PC2* expression was undetectable in all other tissues. *G6PC2* was also highly expressed in the EndoC-βH1 cell line. There was also no notable difference between the two gene assays used, indicating that the full-length transcript is likely to be the major transcript. My results were consistent with public RNA sequencing databases from the Human Protein Atlas (Ponten et al. 2008) and GTEx (Consortium 2013) for up to 32 tissues, which showed enrichment of *G6PC2* expression in the pancreas and but not in all other tissues. These databases do not contain islet cell fractions.

Unfortunately, G6PC2 expression at protein level could not be assessed as the commercially-available antibodies tested were not able to detect human G6PC2 protein (data not shown). Nonetheless, our results were consistent with previous reports of the islet-specificity of *G6PC2* expression (Martin et al. 2001). A previously published human islet transcriptomic database also showed that *G6PC2* expression is three times more enriched in the beta cell fraction than non-beta cell fraction (Nica et al. 2013). Together, these not only highlight a prominent role for *G6PC2* in the beta cell function, but also confirmed that regulation of FG levels by *G6PC2* occurs through the beta cell. In addition, the high levels of endogenous *G6PC2* expression in the EndoC-βH1 cells lend support to utilising these cells as a biological-relevant model for cellular studies.

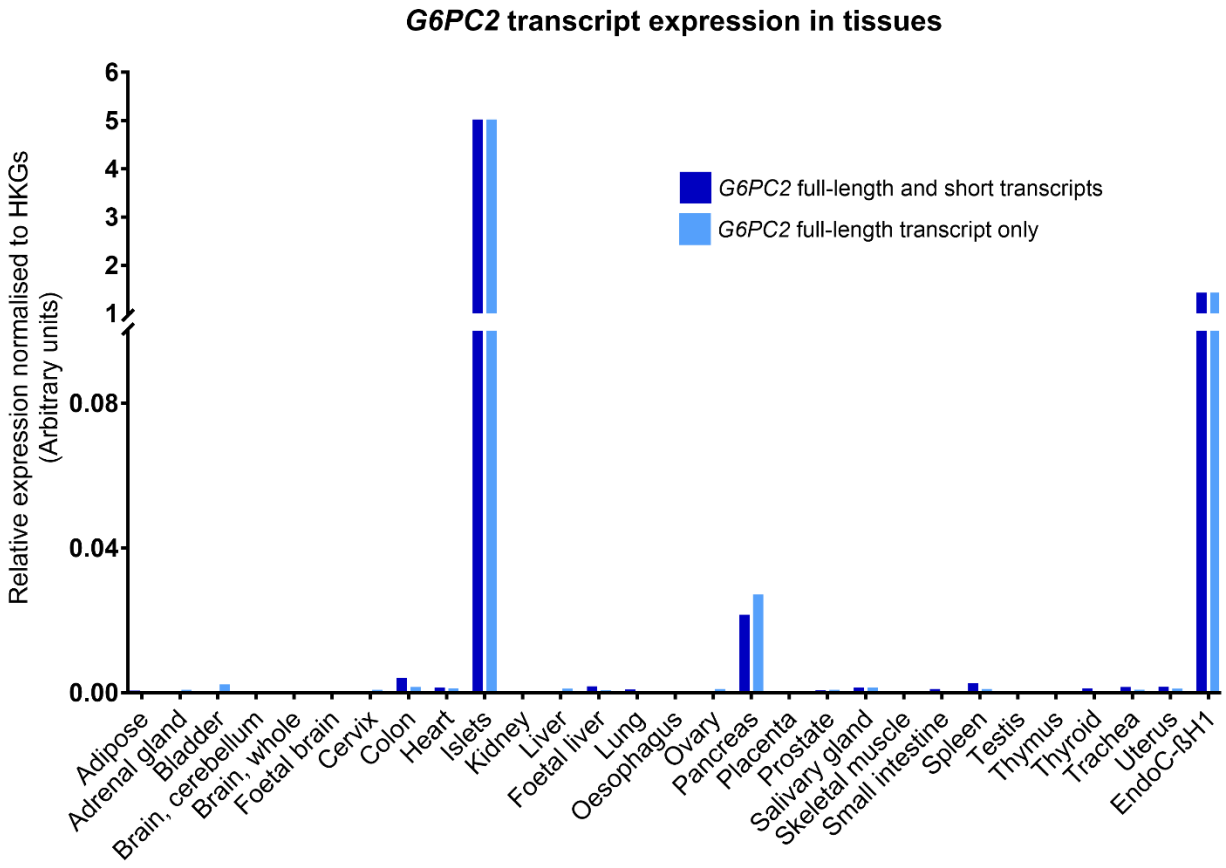


Figure 6.3.1.1. *G6PC2* mRNA expression across a human tissue panel and human beta cell line EndoC-βH1. *G6PC2* expression levels were quantified by real-time qPCR relative to housekeeping genes (HKGs) *GAPDH*, *TBP* and *PPIA* using commercially-available TaqMan® assays. Assay details are described in Methods section 6.2.1. RNA of all tissue samples were obtained from commercially-available tissue panels (Ambion and Clontech), with the exception of islets, pancreas and EndoC-βH1 samples which were sourced in-house. Data obtained from one experiment.

6.3.2. Effect of *G6PC2* knockdown on insulin content and secretion in EndoC-βH1 cells

The role of *G6PC2* has been investigated in various knockout mouse models as reviewed earlier in section 1.5.2, but not in any human cell models. I set out to study *G6PC2* in human beta cells by carrying out gene knockdown followed by glucose-stimulated insulin secretion in the EndoC-βH1 cell line. Knockdown was carried out using a pool of 4 siRNAs and liposome-based transfection, which gave robust gene downregulation of up to 87% (for 25 nmol/L siRNA, Figure 6.3.2.1[A]). Analysis of individual siRNAs from the pool showed that this result was fully supported by si-G6PC2-06 and si-G6PC2-07, both of which decreased *G6PC2* expression by ~87%, validating the knockdown effect (Figure 6.3.2.1[B]).

Given that basal glucose sensitivity was affected in *G6pc2* KO mouse islets as compared to WT islets (Pound et al. 2013), I first transfected EndoC-βH1 cells maintained at standard culture conditions (5.5 mM glucose) with non-targeting siRNA (siNT) or *G6PC2* siRNA (siG6PC2) and evaluated them for insulin secretion and content at basal glucose levels over time. At all time points robust and consistent levels of gene knockdown was observed with mean decrease in *G6PC2* expression of 70% ± 1% (Figure 6.3.2.2). First, siG6PC2-transfected cells were found to have decreased insulin content at all time points from 48h after knockdown as compared to siNT-transfected cells, with no differences in insulin secretion as shown in Figures 6.3.2.3(A) and (B). As *G6PC2* knockdown had significant impact on total insulin content, all insulin data were normalised to cell number instead of insulin content for taking into account variabilities in cell plating densities. Importantly, there was no change in insulin gene transcription (Figure 6.3.2.2) and therefore the reduction in total insulin content is likely to be an effect at the level of translation or protein turnover affecting total proinsulin and insulin levels.

To determine if GSIS is affected in *G6PC2*-deficient EndoC-βH1 cells, functional assays using static incubation were performed 96-120h after transient transfection with siRNA, at five different conditions: 1 mM glucose (no stimulation), 6 mM glucose (sub-maximal stimulation), 20 mM glucose (maximal stimulation), 20 mM glucose with tolbutamide (100 µM) and 20 mM glucose with diazoxide (100 µM). Tolbutamide is a sulfonylurea that stimulates insulin secretion by blocking the K_{ATP} channels and raising cytosolic Ca^{2+} concentration in beta cells, whereas diazoxide opens the K_{ATP} channels to inhibit insulin

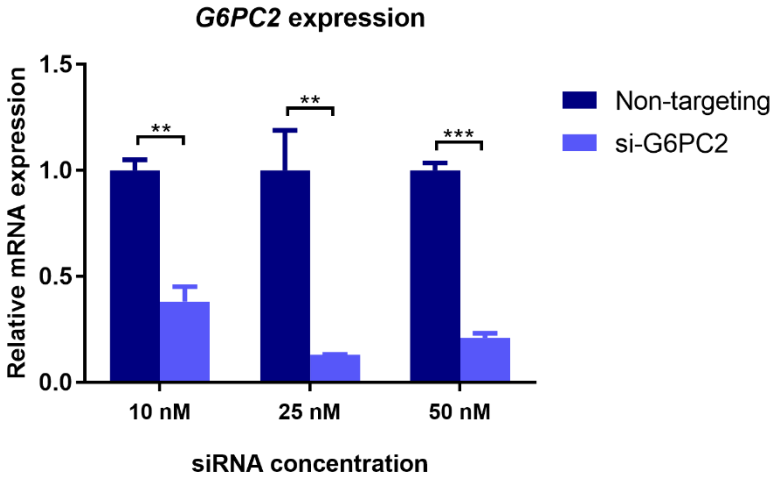
secretion. By uncoupling glucose and ATP metabolism from Ca^{2+}-dependent exocytosis, these chemical modulators can help to reveal parts of the triggering pathway of insulin secretion that may be affected by specific gene knockdown. In these experiments, gene knockdown was again confirmed and siG6PC2-transfected cells had a residual *G6PC2* expression of 17% ± 2% (Figure 6.3.2.4[A]). Consistent with previous findings, *G6PC2*-deficient cells had significantly decreased insulin content compared to control cells by mean of 15% ± 2% across all stimulatory conditions, without corresponding increase in insulin secretion in all conditions as shown in Figures 6.3.2.4(C) and (D). The reduction in total insulin content was significant under all conditions ($P<0.01$) except under stimulation with tolbutamide (likely due to slightly reduced insulin content in control cells treated with tolbutamide and variation in data). Nonetheless, as insulin content appeared to be reduced to similar levels in *G6PC2*-deficient cells regardless of the applied conditions, the effect on insulin content is more likely a consequence of *G6PC2* depletion itself. *G6PC2* action is also likely to be affecting the total pool of insulin rather than the maturation of insulin, though proinsulin to insulin ratio has not been directly assessed.

In addition, insulin secretion was modestly increased in *G6PC2*-deficient cells only at low and sub-maximal glucose stimulations measured at 1 mM and 6 mM glucose respectively, though this result was only statistically significant at low glucose levels ($P=0.04$ for 1G, $P=0.12$ for 6G; Figure 6.3.2.4[D]). This result suggests that *G6PC2* knockdown increases glucose responsiveness at sub-threshold levels of glucose. There was no difference in fold stimulation between control and siG6PC2-transfected cells despite the reduction in insulin content (Figure 6.3.2.4[E]). When expressed as a proportion of insulin content, insulin secretion appeared to be enhanced in *G6PC2*-deficient cells, though this was largely driven by the reduced content (Figure 6.3.2.4[F]). To put these results in context, the EndoC-βH1 cells secreted 3-5% of total insulin content at basal glucose levels in these experiments, or 1-2% when the CyQUANT assay reagent for cell counting is not used. Finally, *G6PC2* depletion in EndoC-βH1 cells gave rise to a small but consistent increase in cell growth across all experiments (Figures 6.3.2.3[D] and 6.3.2.4[B]).

In summary, my knockdown studies in EndoC-βH1 cells showed that *G6PC2* may have a direct/indirect role to play in the regulation of total insulin content (either through reduced translation or increased degradation), though the precise underlying mechanisms are

unclear. As the levels of other ER-processed proteins were not assessed, it is not clear whether there is a general translational block in these cells. Importantly, siG6PC2-transfected cells have a small but significant increase in basal insulin secretion but no effect at high glucose levels compared to control cells, indicating possible enhancement of glucose sensitivity at low glucose concentrations. The increase in glucose sensitivity may occur across sub-maximal glucose stimulation levels if *G6PC2*-deficient cells experience higher glycolytic flux, as long as GCK is not working at saturation point.

A



B

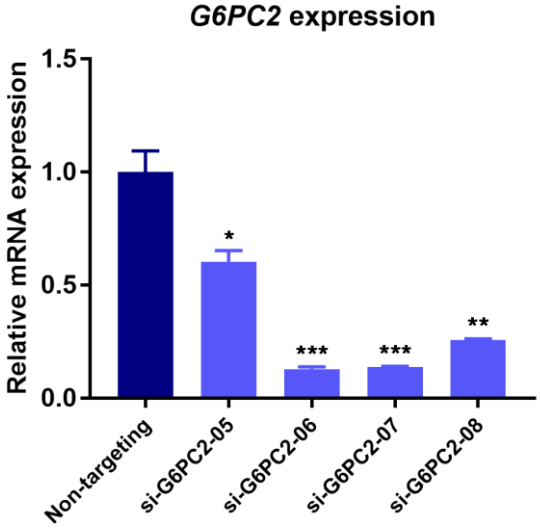


Figure 6.3.2.1. Knockdown efficiency of *G6PC2* in siRNA-transfected EndoC-βH1 cells at 96h. (A) *G6PC2* knockdown using SMARTpool siRNA at varying concentrations. (B) Validation of *G6PC2* knockdown using single siRNAs at 25 nM final concentration. All siRNAs were obtained from Dharmacon (GE Healthcare). Data collected from three independent replicates in one experiment are expressed as mean *G6PC2* mRNA levels relative to non-targeting control. *** P<0.001; ** P=0.001-0.01; * P=0.01-0.05 based on unpaired Students' t tests comparing each knockdown condition to non-targeting control. Sequences of individual siRNAs can be found under Methods section 6.2.2.

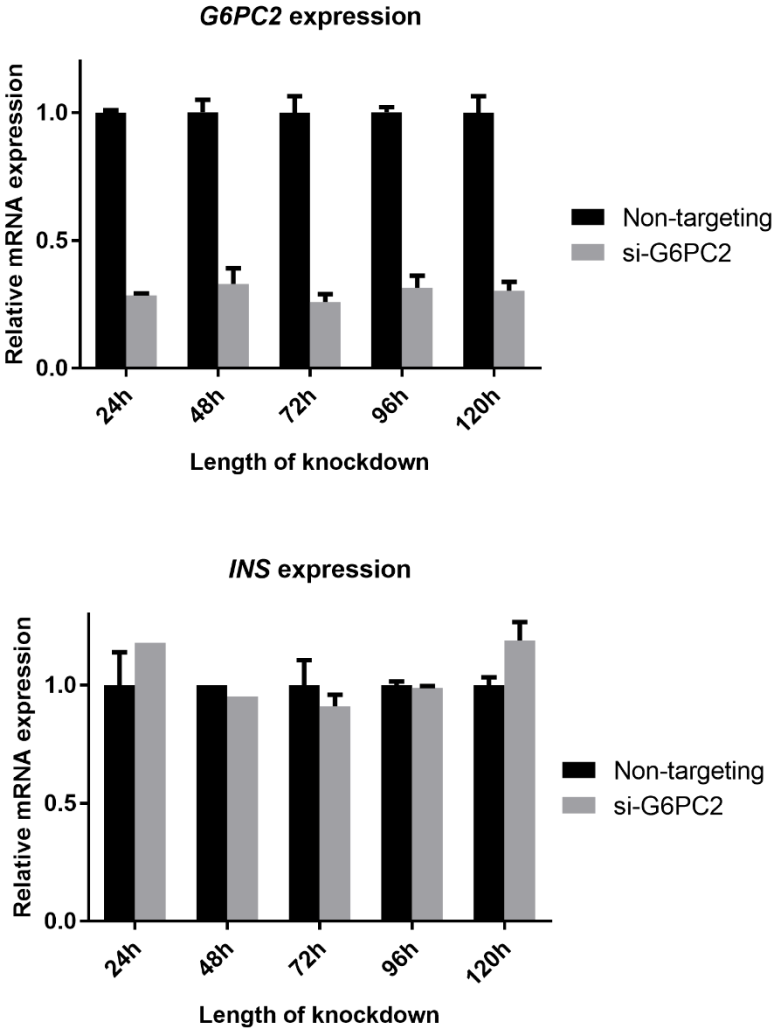


Figure 6.3.2.2. Gene expression levels of *G6PC2* and *INS* in a *G6PC2* knockdown time course experiment in EndoC-βH1 cells. *G6PC2* results were obtained from n=6 of three independent experiments; *INS* results were obtained from n=1-2 of one experiment. All bars represent mean ± standard error of the mean. *G6PC2* expression was significantly reduced in siG6PC2-transfected cells compared to siNT-transfected cells at all time points based on unpaired Students' t tests (P<0.001).

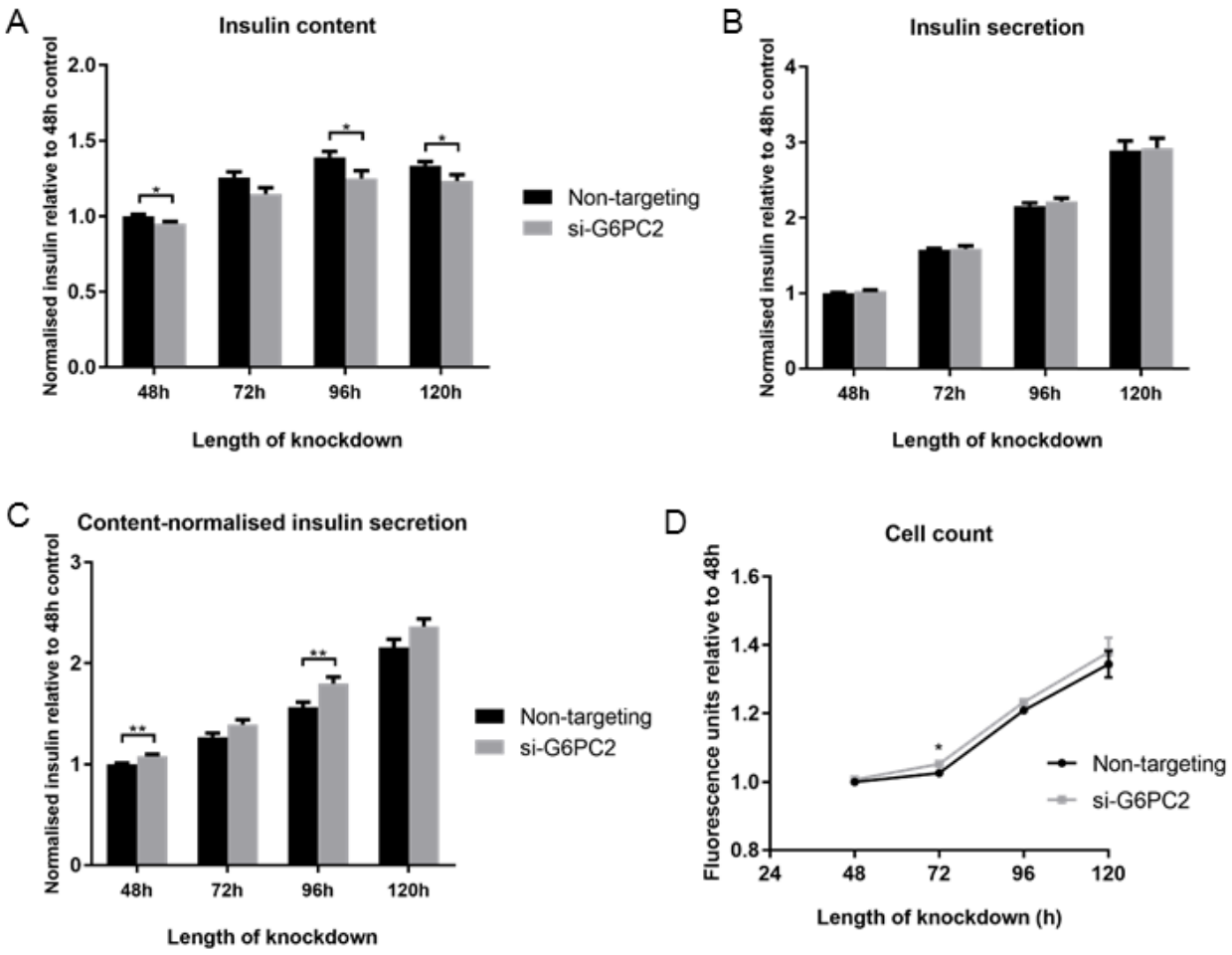


Figure 6.3.2.3. Effect of *G6PC2* knockdown on insulin content and secretion in EndoC-βH1 cells cultured at 5.5 mM glucose over time. (A) Insulin content, (B) insulin secretion, (C) content-normalised insulin secretion and (D) cell growth were assessed 48-120h following knockdown of cells. All data except for (C) were normalised to cell count assessed using the CyQUANT direct detection reagent. Data shown are relative to non-targeting control at 48h and representative of n=16 from four independent experiments. *G6PC2* knockdown efficiencies for these experiments are shown in Figure 6.3.2.2. All bars represent mean ± standard error of the mean. Statistical analyses were carried out using unpaired Students' t tests comparing *G6PC2* knockdown to control at each time point. ** P=0.001-0.01; * P=0.01-0.05.

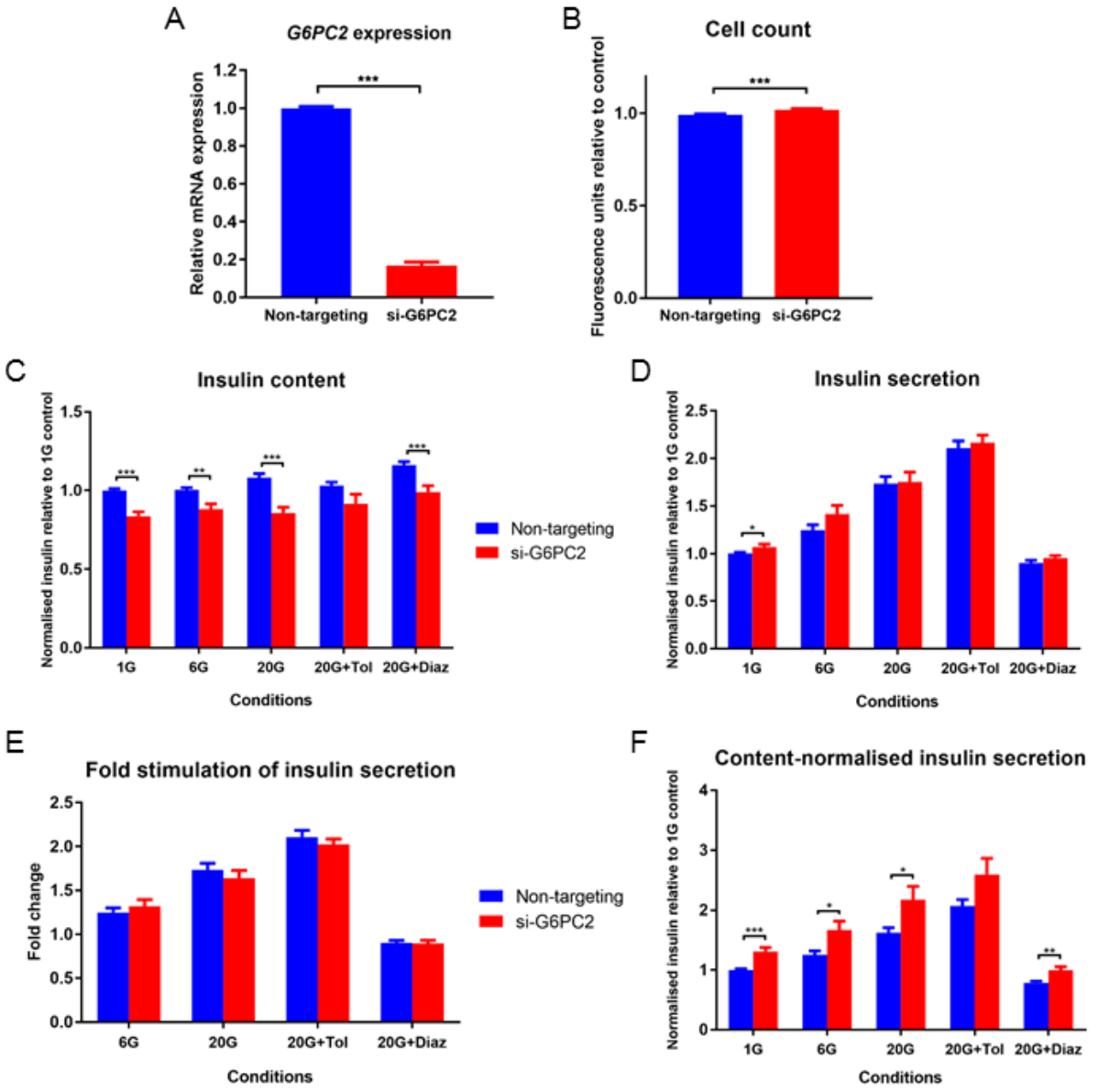


Figure 6.3.2.4. Effect of *G6PC2* knockdown on insulin content and secretion in EndoC-βH1 cells. (A) *G6PC2* expression in siG6PC2-transfected cells relative to control was quantified from n=6 across all four independent experiments presented here. (B) Mean cell count was measured using the CyQUANT direct detection reagent and are representative of n=80 each from four independent experiments. (C) Insulin content, (D) insulin secretion, (E) fold stimulation and (F) content-normalised insulin secretion were assessed at basal to high glucose conditions following 96-120h gene knockdown. All data except for (F) were normalised to cell count. Data are representative of n=16 from four independent static incubation experiments (in three experiments *G6PC2* SMARTpool siRNAs were transfected for 120h; in one experiment si-G6PC2-07 was transfected for 96h). In (C), (D) and (F), all data are relative to non-targeting control at 1G; in (E) data are relative to non-targeting control or si-G6PC2 at 1G respectively. All bars represent mean ± standard error of the mean. Statistical analyses were carried out using unpaired two-tailed Students' t tests comparing *G6PC2* knockdown to control for each condition. *** P<0.001; ** P=0.001-0.01; * P=0.01-0.05.

6.3.3. Effect of *G6PC2* genetic variation on ER homeostasis in a cell model system

I have previously shown in Chapters 3 and 5 that genetic variation in *G6PC2* results in the generation of unstable variant proteins. Protein misfolding can disturb ER homeostasis, causing ER stress and evoking the UPR. I set out to examine whether the presence of *G6PC2* variants elicits an ER stress response by assessing the gene expression of markers of the UPR. HEK293 cells were used as a cellular model system as poor transfection efficiency was observed in the more physiologically-relevant EndoC-βH1 cells which proved less reliable for overexpression studies (as further discussed in section 6.4). HEK293 cells were transiently transfected with *G6PC2* WT and selected variants H177Y, Y207S and V219L. H177Y and Y207S are both rare variants that gave rise to near-total loss of protein in HEK293 cells, while V219L is a common variant that retains a residual protein expression of ~50% compared to WT levels (shown in Chapter 3).

I first analysed *BiP* (*HSPA5*), *XBP1s* (spliced *XBP1*) and *CHOP* (*DDIT3*) levels in response to expression of G6PC2-WT or variants. Briefly, *BiP* encodes an ER chaperone involved in protein folding and assembly and is upregulated with activation of the UPR; splicing of *XBP1* is a surrogate of the IRE-1 branch of the UPR; *CHOP*, which promotes apoptosis, is a target of both PERK and IRE-1 activation. After 48h transfection, *BiP* levels were significantly increased by expression of the LOF variants H177Y and Y207S as compared to WT (Figure 6.3.3.1). Expression of these variants also appeared to give rise to mild elevation in *CHOP* levels though the effects were not significant. No differences were observed with respect to *XBP1s* levels. The expression of common variant V219L also did not induce ER stress gene expression over and above WT, at least for the genes tested. These results suggest that the presence of unstable and non-functional variants of *G6PC2* may stimulate transcription of a number of ER stress genes in cells as compared to WT protein (Figure 6.3.3.1). The elevation in ER stress response genes though significant, was modest, as a caveat with this analysis is that gene expression changes are transient and may not be well-captured at a single time point.

To validate the impact of LOF G6PC2 variants on the UPR, I utilised a more sensitive luciferase reporter-based assay which measured the activity of a number of canonical enhancer elements that represent activation of different arms of the ER stress response as

described in section 6.1.2 and 6.2.5. Triggering of the UPR will result in binding of transcription factors and co-factors to these enhancer sequences which will in turn drive luciferase gene transcription. The luciferase activity measured was a result of a cumulative effect of promoter activation over 48h. Firefly luciferase activity data were normalised to that of Renilla luciferase activity (internal transfection control), then corrected for background activity observed when empty pGL3-Promoter was transfected. In this experiment, additional *G6PC2* rare variants that give rise to varying effects on molecular function were tested. I171T and F256L are variants that were stably expressed like WT protein in HEK293 cells but have reduced enzymatic activity, as described in Chapter 5. The results shown in Figure 6.3.3.2 demonstrated that the three variants that give rise to unstable protein products (H177Y, Y207S, S324P) significantly increased the activity of all ER stress reporters after 48h expression, with up to ~3-fold increase (all $P<0.001$). I171T and F256L also significantly increased activity at ERSE II and UPRE though to a lesser extent. In contrast, the common variant V219L did not activate any of the reporters in comparison to WT. A small follow-up time course experiment demonstrated that the LOF variant H177Y was already able to significantly increase ER stress even after 24h expression (Figure 6.3.3.3).

Overall, I have showed that expression of *G6PC2* variants, especially variants that have severe LOF due to protein instability, triggered markedly greater ER stress responses as compared to expression of WT protein. The ER stress response seems to be propagated through both IRE-1- and ATF6-mediated pathways. While these experiments were able to provide insight into ER stress response in the presence of functional G6PC2 coding variants that influence FG in variant carriers, they were unable to address the question surrounding the involvement of endogenous G6PC2 protein in ER homeostasis in a physiologically-relevant beta cell model. I therefore extended my investigations to examine G6PC2 in EndoC-βH1 cells by carrying out gene silencing studies in the context of ER stress.

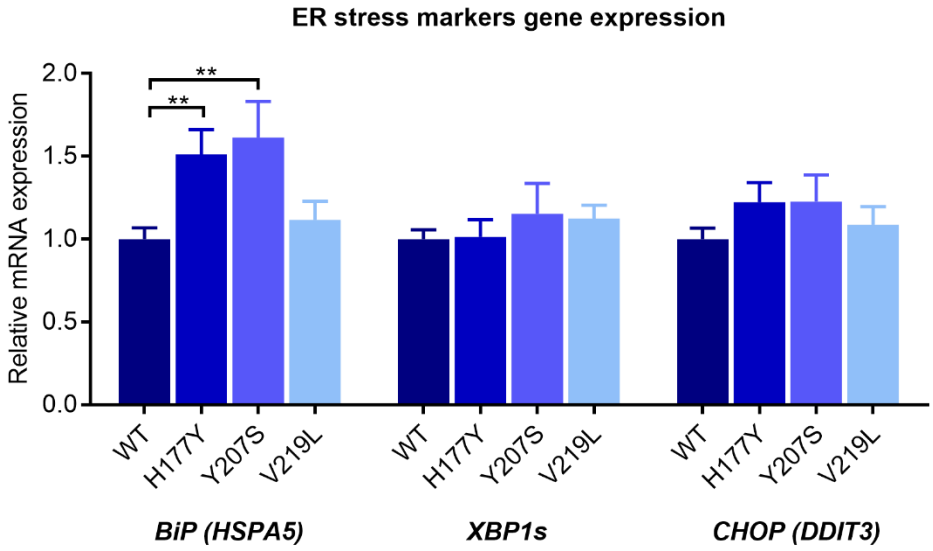


Figure 6.3.3.1. Effect of G6PC2 WT and variant protein expression on the expression of ER stress markers in HEK293 cells. Cells were transfected for 48h and gene expression data normalised to WT-transfected cells. Data from n=6 across three independent experiments were analysed by unpaired two-tailed Students' t tests, comparing each variant to WT. ** P=0.001-0.01.

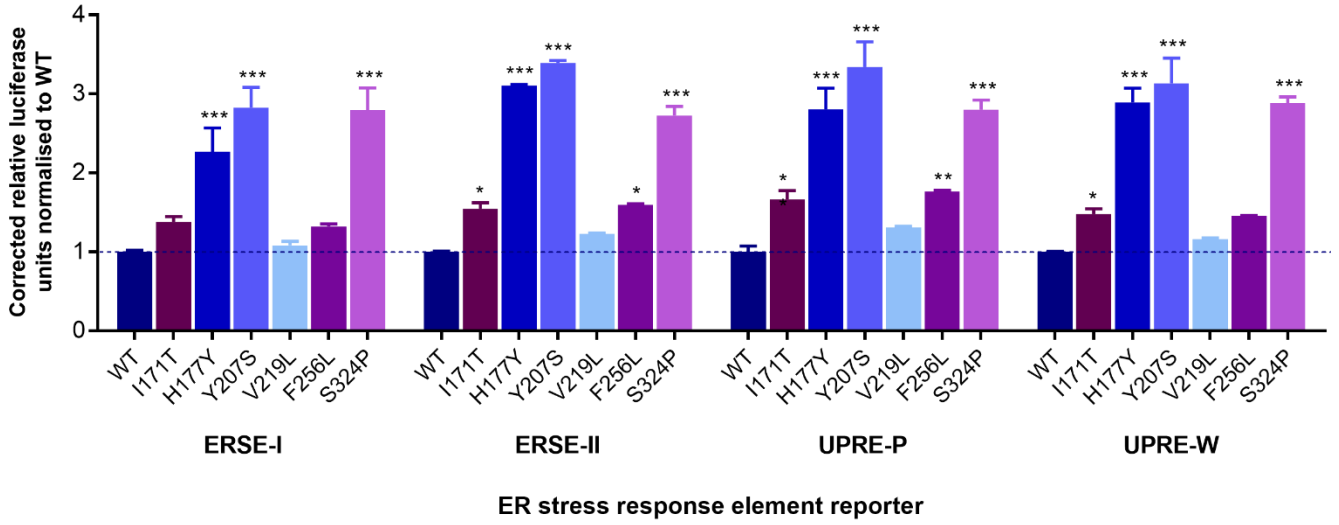


Figure 6.3.3.2. Effect of G6PC2 WT and variant protein expression on luciferase activity driven by ER stress response elements in HEK293 cells. Relative luciferase units corrected for background activity were normalised to WT for each reporter. All bars represent mean ± SEM. Statistical analyses were carried out on data from n=6 across two independent experiments (except for F256L, n=3 in one experiment) using two-way ANOVA with Fisher's LSD test comparing each variant to WT. *** P<0.001; ** P=0.001-0.01; * P=0.01-0.05.

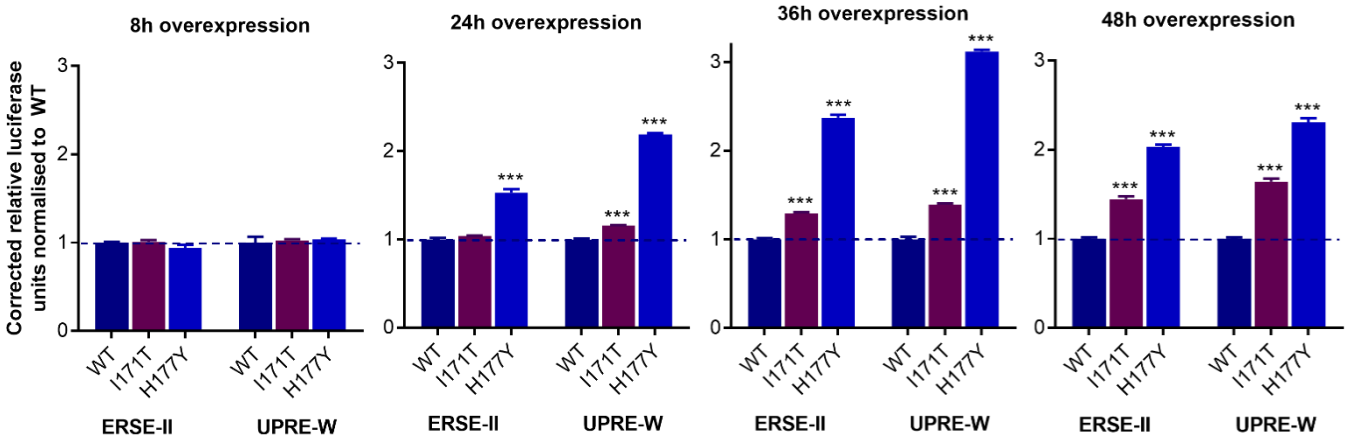


Figure 6.3.3.3. Effect of G6PC2 WT and variant protein expression on luciferase activity driven by ER stress response elements in HEK293 cells over time. Relative luciferase units corrected for background activity were normalised to WT for each reporter at each time point. All bars represent mean ± SEM. Statistical analyses were carried out on data from n=3 of one experiment using two-way ANOVA with Fisher's LSD test comparing each variant to WT. *** P<0.001.

6.3.4. Effect of *G6PC2* knockdown on ER homeostasis in EndoC-βH1 cells

I have shown that the expression of unstable LOF G6PC2 variants may contribute to ER stress responses in HEK293 cells. To investigate if the presence of functional WT G6PC2 could also play a role in ER homeostasis in an endogenous cellular system in the absence and presence of induced stress, a preliminary knockdown study of *G6PC2* followed by evaluation of ER stress gene expression in EndoC-βH1 cells treated with ER stress inducers thapsigargin (TG) or tunicamycin (TC) was carried out. TG is a sarco-endoplasmic reticulum Ca^{2+}-ATPase (SERCA) inhibitor that prevents the replenishment of ER Ca^{2+} stores, in turn increasing cytosolic Ca^{2+}. TC inhibits N-linked glycosylation resulting in unfolded or misfolded proteins. Thus TG and TC both induce ER stress via diverse mechanisms, through Ca^{2+} dysregulation and aberrant protein processing respectively. In addition to *XBP1s*, *CHOP* and *BiP*, *ATF4*, which is a transcription factor activated by the PERK signalling pathway, was also assessed.

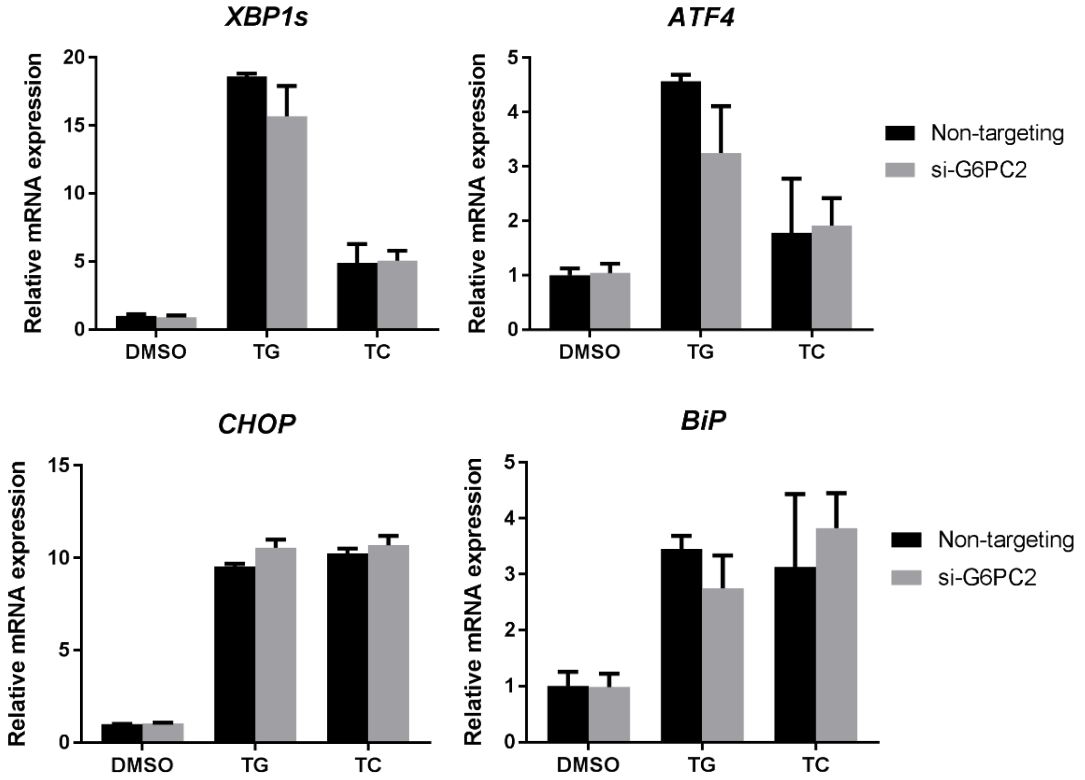
As shown in Figure 6.3.4.1(A), levels of all the ER stress genes tested were generally upregulated in the presence of both TG and TC as expected. In the presence of stress induced by TG, *G6PC2*-deficient cells tended to express lower levels of *XBP1s*, *ATF4* and *BiP* as compared to siNT-transfected cells, though the effects were not statistically significant. If these results were to be replicated, they suggest that a reduction in G6PC2 levels may reduce the magnitude of the UPR by mitigating ER stress in the face of TG-induced stress. These effects were however not observed in unstressed cells treated by DMSO or stressed cells induced by TC.

In all three conditions, knockdown efficiency was robust as *G6PC2* expression was reduced by mean of 75%, 67% and 74% in cells treated with DMSO, TG and TC respectively (Figure 6.3.4.1[B]). However, my analysis also revealed that *G6PC2* expression was increased in response to ER stress induced by TG treatment but not TC treatment (p=0.01 based on unpaired Students' t test comparing DMSO and TG treatment in non-targeting-transfected cells). This not only showed that G6PC2 may play a part in regulating ER stress signalling pathways triggered by disrupted Ca^{2+} homeostasis, but also that G6PC2 expression itself may be regulated indirectly by Ca^{2+} levels. In an attempt to corroborate these results, luciferase reporter assays for studying ER stress response which were previously carried out

in the HEK293 cells, were adapted for the EndoC-βH1 cells, but due to lack of consistency in transfection efficiencies reliable data could not be obtained.

Overall, my results, though very preliminary, suggested that depletion of *G6PC2* in EndoC-βH1 cells may regulate the susceptibility of cells to ER stress potentially through Ca^{2+}-dependent mechanisms, at least in the presence of stress induced by TG. It provided a starting point for the study of the role of G6PC2 in beta cell ER homeostasis, but more work remains to be done to address this question.

A



B

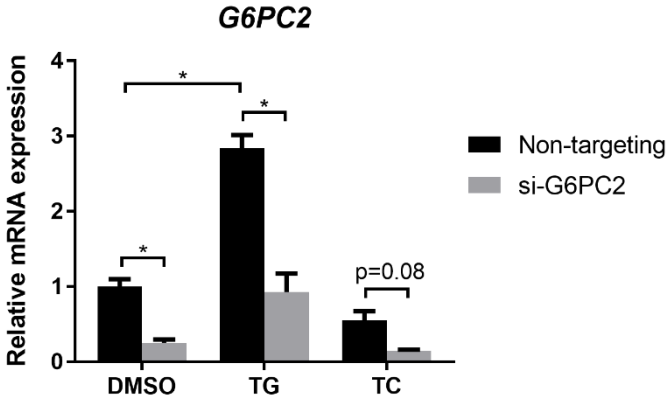


Figure 6.3.4.1. Effect of *G6PC2* knockdown on expression of ER stress markers in stress-induced EndoC-βH1 cells. Gene knockdown was carried out using si-G6PC2-07 for 48h during which cells were treated with DMSO vehicle, 1 µM thapsigargin (TG) or 5 µg/ml tunicamycin (TC) for 16h before collection. Gene expression levels of (A) ER stress markers *XBP1s*, *ATF4*, *CHOP* (*DDIT3*), *BiP* (*HSPA5*) and (B) *G6PC2* normalised to DMSO-treated non-targeting-transfected cells are shown. All bars represent mean ± SEM. Statistical analyses were carried out on data from n=2 of one experiment using unpaired two-tailed Students' t tests comparing siG6PC2 to control. * P=0.01-0.05.

6.4. Discussion

The results in this chapter have attempted to shed light on the function of G6PC2 in human beta cells, in particular in the context of GSIS and ER homeostasis in EndoC-βH1 cells. I first confirmed that *G6PC2* expression is largely confined to pancreatic islets and is also highly expressed in the EndoC-βH1 cells.

siRNA-based knockdown of *G6PC2* in EndoC-βH1 cells significantly decreased total insulin content and the effect was sustained when tested up to 120h post-knockdown. In secretion assays, insulin content was also significantly reduced in *G6PC2*-depleted cells without an expected corresponding increase in insulin secretion relative to control cells, regardless of glucose stimulation conditions or administration of the K_{ATP} channel modulators tolbutamide and diazoxide. As the insulin assay detects both proinsulin and insulin proteins, the reduction in insulin levels is likely to be caused by a defect in insulin translation (or stability) as opposed to insulin maturation, even though this was not directly tested. Insulin gene transcription was not affected. In addition, a modest enhancement in insulin secretion specifically at low (1 mM) and sub-maximal (6 mM) stimulatory glucose conditions was observed in *G6PC2*-deficient cells.

The secretion results are in agreement with the hypothesis that G6PC2 deficiency promotes glycolytic flux mediated through GCK effectively increasing insulin secretion and hence beta cell sensitivity at sub-threshold glucose concentrations. Insulin secretion in response to high glucose is unaffected by G6PC2 levels as GCK is likely to be saturated at that point and flux through glucose phosphorylation and subsequent metabolism is at its maximum rate. The mechanism behind the observed reduction in total insulin content in G6PC2-deficient cells is unclear, but could be related to changes in ER status that alter insulin production. As ER integrity and protein production are tightly linked, elevation of UPR activity may result in a general inactivation of protein translation, though insulin is the major protein product in beta cells. Evaluation of other ER proteins was not carried out.

As a logical follow-up of ER homeostasis, I made use of G6PC2 coding variants with known effects on protein function to investigate the impact of G6PC2 variation on ER stress. I used experimental pipelines that studied transient gene expression changes of key signalling markers (qPCR-based) and longer-term activation of ER stress response elements (luciferase

reporter-based). The results showed that expression of G6PC2 variants, especially LOF variants, strongly elicit the ER stress response through both IRE-1 and ATF6 signalling cascades, inducing the expression of ER chaperones, folding proteins and components of the ERAD machinery to relieve ER stress. We know that heterozygous carriers of these variants have only ~50% of normal G6PC2 function, which resembles a partial gene knockdown that may give rise to increased basal beta cell function and hence reduced FG levels. The presence of damaging variant proteins may however also have a collateral effect on promoting ER stress and increasing basal levels of UPR. Therefore, the balance between residual G6PC2 function and variant effects, and its overall impact on FG levels need to be considered, especially in a physiological human context.

The study of the impact of G6PC2 variant expression on ER stress response in HEK293 cells comes with several caveats. HEK293 cells do not provide the relevant physiology for studying ER stress responses in beta cells, which are highly sensitive to ER stress given their high translational load. I was unable to effectively use more biologically-relevant cell lines such as EndoC-βH1 and INS-1 for the expression studies as plasmid transfection efficiencies were poor and did not provide enough sensitivity for obtaining consistent results (data not shown). In addition, overexpression of proteins activates the UPR as cells attempt to cope with the increase in protein synthesis, therefore there is an inherent risk that the observed effects on ER stress genes are artefacts of an overexpression system. Nonetheless, the HEK293 model system provided the opportunity to study potential UPR activation caused by the presence of misfolded G6PC2 variants over and above WT protein within a cell expression model system. The results build on current knowledge of the molecular consequences of these variants in the same cellular system.

To address questions surrounding the role of endogenous WT G6PC2 in ER homeostasis, a pilot experiment was carried out in *G6PC2*-depleted EndoC-βH1 cells. Downregulation of G6PC2 appeared to reduce the activity of UPR signaling in the PERK and IRE-1 arms in response to TG-induced stress. As the regulation of the UPR was only revealed in TG-treated cells and not TC-treated cells, G6PC2 may potentially play a part in ER homeostasis through Ca^{2+}-dependent mechanisms. However, the observed trends warrant further exploration as there was lack of power to detect significant effects in this pilot experiment. An added

challenge stems from the potential heterogeneity in UPR gene expression signatures even within a beta cell population (Baron et al. 2016).

Some links have previously been drawn between G6PC2 and insulin. Transgenic mice expressing human G6PC2 in the islets have signs of increased beta cell ER stress resulting in necrosis, as well as significantly reduced insulin granules in the remaining beta cells, as compared to non-transgenic mice, hence the authors associated G6PC2 action with a defect in insulin production (Shameli et al. 2007). My experiments in EndoC-βH1 cells however showed the opposite effect – reduction of G6PC2 gave rise to reduced insulin. Again, the discrepancy could be due to different experimental models and physiological heterogeneity that the authors had observed in their transgenic lines. Another study recently added new perspective to show that lowered insulin production in mice relieves baseline ER stress, as exemplified by reduced expression of ER stress response genes in PERK- and IRE-1-mediated pathways (Szabat et al. 2016). Metabolomic and network analysis revealed that islets with reduced insulin also displayed increased levels of high energy phosphates such as ATP, without any increase (and in fact a slight decrease) in glycolytic or Krebs cycle activity, but instead had reduced energy-consuming metabolic pathways. Insulin production normally elicits a basal sub-threshold level of chronic ER stress in the beta cells (Back and Kaufman 2012). Perhaps beta cells with G6PC2 deficiency and reduced insulin may have lower biosynthetic burden and hence lower basal levels of ER stress.

Future studies that examine the UPR in greater depth may shed more light on the relationships between G6PC2 action, ER stress and insulin production. The relevance of the extent of these activities in a physiological context also remains a crucial question. A number of key signaling events of the UPR involve translational modification, such as phosphorylation of eIF2α of the PERK branch, cleavage of ATF6, as well as phosphorylation of pro-apoptotic signaling factors. More extensive studies of the UPR would require western blot analysis of these proteins as well as specific evaluation of cell apoptosis. EndoC-βH1 cells lacking G6PC2 appear to have small but consistently increased cell growth, and current observations with the pro-apoptotic marker *CHOP* were not clear, therefore follow-up experiments are needed to determine if G6PC2 is involved in apoptosis and beta cell loss. It is however worth noting that elevations in the UPR do not necessarily indicate a detrimental effect to cells. The UPR is instead a physiological process that cells undertake over a period

of time to restore ER homeostasis. It is also a complex process involving many feedback loops, context-dependent mechanisms and cross talks between multiple cellular pathways and organelles. This further emphasises the importance of getting the experimental time point and culture conditions right (Cnop et al. 2012).

Additional experiments should certainly investigate intracellular calcium levels and signaling dynamics in G6PC2-deficient EndoC-βH1 cells. Calcium imaging studies are commonly used as a proxy for stimulus-coupled secretion, and are capable of capturing dynamic changes in Ca^{2+} fluxes in response to stimulation. In fact, dynamic insulin secretion setups by perifusion improve the characterisation of GSIS which is oscillatory and biphasic in nature (Henquin et al. 2015). They are therefore more sensitive than static incubation assays. In preliminary studies that were recently set up to investigate Ca^{2+} signaling in EndoC-βH1 cells using the calcium dye Fura Red, *G6PC2* knockdown did not appear to affect Ca^{2+} response to high (20 mM) glucose (data not shown), consistent with what we have observed from static incubation assays. Future experiments would go on to characterise Ca^{2+} signaling in response to step-wise increases in glucose concentrations at sub-threshold glucose stimulatory levels, to enable systematic analysis of the effect of G6PC2 on glucose sensitivity. Specific assays for evaluating ATP/ADP ratio will also help to elucidate whether the effects are due to G6P hydrolysis or other pathways independent of metabolism.

Though the EndoC-βH1 cells have provided opportunities for the study of insulin secretion and ER homeostasis, there are several limitations with respect to their use as a beta cell model. It is a glucose-responsive cell line but its fold stimulation in response to glucose is small, which reduces the sensitivity of the system to detect small effects. In my experiments, the mean fold induction (based on insulin secreted) achieved in WT cells were 1.2 at 6 mM glucose and 1.7 at 20 mM glucose. The response was poorer than that published for this cell line and many fold lower than that observed in primary human islets (Ravassard et al. 2011). The EndoC-βH1 cells also demonstrated poor plasmid transfection efficiency and were not ideal for overexpression studies of G6PC2. By visual observation of immunofluorescence microscopy data transfection efficiency was estimated at <20% for *GFP* and no more than 1% for *G6PC2* plasmids (data not shown). One way to improve transfection efficiency of *G6PC2* constructs is to carry out codon optimisation (Boortz et al. 2016a). As subsequent generations of the EndoC-βH1 cell line have been developed and shown to display improved

insulin secretion upon conditional termination of cell proliferation (Benazra et al. 2015; Scharfmann et al. 2014), the use of these cell models may be considered.

More importantly, experiments should be replicated in primary human islet cells. The clinical islet isolation facility at OCDEM provides precious human islet samples for research, though their availability and quality can be unpredictable. Studies in human islets may provide valuable insight into unravelling the role of G6PC2 using many follow-up experiments suggested above. Progress in these efforts will help to confirm our findings observed in a beta cell line, complement what is already known from rodent models and answer some of the long-standing questions concerning the biological function of G6PC2.

Chapter 7

Discussion and Future Directions

7.1. Insights from the characterisation of G6PC1 and G6PC2 coding variant associations

The rapid progress in genetics studies of human complex disease has seen the discovery of a substantial number of susceptibility loci which have enhanced our understanding of the genetic architecture of disease and related traits. The ultimate goals, however, are to provide biological insight and to identify druggable targets or pathways relevant for clinical application. To do this, it is imperative to accumulate sufficient evidence from multiple preclinical models including human genetics and epidemiological findings as well as cellular and animal studies for target prioritisation and validation (Plenge et al. 2013). The work presented in this thesis complements the growing number of studies, many described in Chapter 1, that aim to bridge the gap between genetic associations and the underlying biological mechanisms. The focus on protein-coding regions of the genome (exomes) is one way of providing tractable targets for the design of *in vitro* studies to evaluate the consequences of perturbing gene and protein function.

Chapter 3 described the identification of low-frequency coding variants in the *G6PC2* locus from exome-wide association studies that influenced FG levels independently of the GWAS variant (Mahajan et al. 2015). *G6PC2* was an interesting locus to follow up on given prior knowledge of its role in glucose metabolism, yet inadequate understanding of how it regulates islet function. Protein expression assays performed on these coding variants revealed that FG levels in variant carriers were altered due to loss of G6PC2 protein function. This study also emphasised the value of haplotype-based analysis which took into account the impacts of neighbouring regulatory variants such as the non-coding variant rs560887 in *G6PC2*, suggested to affect transcript levels (Baerenwald et al. 2013), on the observed effects from coding variant alleles.

The exome array genotyping data presented in Chapter 3 focused on European populations. As an extension to this study, Chapter 4 and 5 described how collaborative large-scale exome-wide meta-analyses, an example of a rare variant association analysis (Zuk et al. 2014), increased sample size by at least 3-fold and made it possible to identify rare coding variants contributing to glycemic trait variance in multiple ancestry groups. The MAGIC investigators were able to identify *G6PC1*, a member of the same gene family as *G6PC2*, as a

novel locus for influencing both FG and FI levels. The gene-based conditional analyses together with my molecular studies described in Chapter 4 showed that multiple rare coding variants in *G6PC1* were driving these associations due to loss of protein stability or G6Pase activity.

Within the well-established glycemic trait locus *G6PC2* additional rare coding variants that altered protein function through multiple mechanisms were revealed, including both inactivating and activating mechanisms that resulted in reciprocal changes in FG levels. This work presented in Chapter 5 was the first systematic and comprehensive evaluation of *G6PC2* coding variants across the allelic frequency spectrum through a combination of *in silico* techniques, functional follow-ups, and gene-based association testing informed by the functional assessments. Indeed, based on the MAGIC dataset, the effect sizes of the causal rare coding variants (β_{FG}=0.08-0.51 mmol/L; β_{HbA1c}=0.006-0.25%, Chapter 5 Table 5.1.1.1) for glycemic trait associations were mostly larger than that of the common intronic variant rs560887 (β_{FG}=0.08 mmol/L; β_{HbA1c}=0.03%) and the common coding variant rs492594 (p.V219L β_{FG}=0.02 mmol/L; β_{HbA1c}=0.003%). The analysis of rare genetic variation may be best achieved with composite sequencing-genotyping studies that enable comprehensive and unbiased evaluation of rare variants. Exome sequencing data provided information on very rare (MAF<0.1%) or even private variants not included in the exome arrays that modulate protein function, as shown in Chapter 5. This recently published 13k exome sequencing dataset (Fuchsberger et al. 2016) has now doubled in sample size and will undoubtedly uncover even more rare variants of biological value. Greater representation from other ethnic populations in these datasets will further strengthen such analyses. However, new evidence has now shown that lower frequency variants play a limited role in T2D predisposition as compared to common variants of modest effect sizes (Fuchsberger et al. 2016), which may also be the case for glycemic traits based on current knowledge.

My results from Chapters 3 to 5 have revealed a number of important implications for variant characterisation studies. The study of coding variants in multiple cell types was critical as my characterisation experiments in INS-1 (832/13) cells revealed the impact of G6PC2 variants on protein glycosylation that was not observed in HEK293 cells. It is of no surprise that different cellular models are likely to have differences in protein processing machinery and degradation pathways. This is also pertinent for the characterisation of non-

coding risk variants, as their underlying causal mechanisms which may be mediated by context-specific regulatory factors at the genomic level will only be exposed in the relevant cell or disease context (Thomsen et al. 2016a). In my studies I have also utilised variant databases such as the ExAC browser which provides a reference set for interpretation of coding sequence changes (Lek et al. 2016). In fact, an even bigger resource from the Genome Aggregation Database (gnomAD, http://gnomad.broadinstitute.org/) for both whole exome and genome sequencing data in larger populations is now available. These resources include the use of functional prediction algorithms, which we know do not always predict functional consequences accurately. As these algorithms are often based on alignment to known structural and functional domains, I have in my G6PC1/G6PC2 characterisations observed that missense variants that are particularly disruptive tend to be more accurately predicted. Nonetheless, predictive functional annotations are commonly used for variant filtering in gene-based association analyses to enrich for subsets of variants that are more likely to be causal (Purcell et al. 2014). Improved ways of filtering variants, such as using *in vitro* results for defining functional and neutral alleles as shown in Chapter 5 and previous studies (Bonnefond et al. 2012; Rees et al. 2012b), have been shown to strengthen the gene-based tests and increase the possibility of identifying novel signals especially for loci that fall just below the significance threshold. This is however not always possible in the absence of any motivation to collect comprehensive *in vitro* data for these loci.

There are also limitations in my characterisation studies as it was impractical to evaluate all the coding variants identified to be able to completely ascribe pathogenicity or neutrality. The experiments conducted were too laborious for larger scale application, especially for the study of enzyme activity which required microsomal preparation and enzyme kinetic analysis. I have also not been able to evaluate phosphatase activity in relation to other substrates. Higher throughput functional assays to test large numbers of variants at the same time, such as that carried out for *PPARG* (Majithia et al. 2016), will have to be designed. It would also have been useful to evaluate G6PC1 and G6PC2 function in the context of the G6Pase complex, which includes substrate transporters. If the catalytic subunit were to form a tight complex with G6PT, this interaction may be important for catalytic function in a physiological context. Nonetheless, the current studies have increased

our understanding of the impact of mutations in different domains of G6PC2 on the modulation of protein function and phenotype, which is valuable for target validation in the context of drug development.

My study of G6PC1 and G6PC2 proteins based on the same experimental pipelines allowed me to make comparisons between the characteristics of both proteins. In contrast to G6PC1, the majority of LOF variant proteins in G6PC2 impacted on protein stability, reflecting the importance of amino acid interactions in maintaining correct protein folding and integrity. The characterisation of G6PC2 variants in the context of human or mouse G6PC1 will therefore result in misleading conclusions about variant effects (Boortz et al. 2016a). For both G6PC1 and G6PC2 I highlighted the potential importance of N-linked glycosylation in contributing to functional activity and that this process may be altered to elicit a phenotypic effect. However, G6PC3, the ubiquitous isoform, is found to possess detectable phosphatase activity though it lacks the same consensus glycosylation site (Guionie et al. 2003; Martin et al. 2002; Shieh et al. 2003). It is unclear whether other modification sites are present in G6PC3. G6PC1 and G6PC2, though homologous proteins, are encoded by two different genes and act through diverse mechanisms to influence glycemic traits. Importantly, their impacts on glycemic traits occur through different tissues. As G6PC1 is key to hepatic glucose metabolism it reasonably affects both fasting glucose and insulin traits, while G6PC2 impacts only on FG and HbA1c due to a beta cell effect. Future work will further our understanding of the precise mechanisms underlying the different G6Pase systems in the body.

7.2. Multiple insights surrounding the biological role of G6PC2

Chapter 6 focused on evaluating the function of G6PC2 in the human EndoC-βH1 beta cell line, specifically in insulin secretion and ER homeostasis. The latter was explored given considerations of G6PC2's roles beyond glucose and ATP metabolism. This chapter covered several pilot studies which warrant further studies. The *G6PC2* knockdown experiments can be related to the effects of *G6PC2* LOF variants, which result in *G6PC2* haploinsufficiency in variant carriers, even though knockdown studies represent a much more severe LOF effect. Individuals harbouring these variants may have increased glucose sensitivity and hence

insulin secretion at basal glucose levels, resulting in the lowered FG and HbA1c levels observed. The physiological significance of the elevation of beta cell ER stress responses in the presence of these LOF variants is unclear, especially since my cellular experiments only assessed acute effects. In fact, increased UPR signalling may not necessarily be associated with detrimental consequences but with favourable outcomes to restore ER homeostasis. The consequences of reduced insulin content observed in G6PC2-deficient EndoC-βH1 cells also remain to be established, but is unlikely to affect maximal GSIS. These experimental setups had aimed to provide preliminary insight into the function of G6PC2 which may not necessarily be representative of normal physiology.

The prediction of a leftward shift in the sigmoidal glucose dose-response curve with a loss of *G6pc2* function (Pound et al. 2013) may be likened to activation of GCK function. At one extreme end of the spectrum *GCK* activating mutations are known to cause hypoglycaemia as they result in increased glycolytic activity and inappropriate insulin secretion at low blood glucose levels (Osbak et al. 2009). Therefore, inactivation of *G6pc2* function would indirectly modulate insulin secretion through enhanced glycolytic flux. In-depth studies of the role of G6PC2 in calcium homeostasis will provide much needed insight into the mechanisms underlying its regulation of insulin secretion. As G6P hydrolysis generates inorganic phosphates (Pi) in the ER, this could alter ER calcium levels and result in redistribution of intracellular calcium pools, changes which may play a crucial role in the control of beta cell (dys)function (Gilon et al. 2014). Islets from *G6PC2* KO mice were found to have elevated basal cytosolic calcium levels (Pound et al. 2013). In addition, indirect support for potential calcium regulation arose from the observation that deletion of the *sorcin* gene, a regulator of ER calcium levels, resulted in increased *G6pc2* expression in mice (Marmugi et al. 2016). Several early studies that considered the role of G6PC1 in calcium regulation have showed that G6Pase activity or increased Pi concentrations in liver cells may stimulate calcium uptake into the ER (Benedetti et al. 1985; Fulceri et al. 1990). In the context of beta cells, ER calcium sequestration can in turn affect calcium-dependent mechanisms of insulin granule exocytosis (Wolf et al. 1986; Wolf et al. 1988).

It is well-established that G6PC1 is crucial to glycogen metabolism in the liver, and *G6PC1* deficiency in patients with GSD1a and in knockout mouse models have excessive accumulation of glycogen in the liver and kidney (Chou and Mansfield 1999; Lei et al. 1996).

The relationship between G6PC2 activity and glycogen metabolism in the beta cells has not been directly studied. Recently published work showed that hyperglycemia in mouse models of human diabetes resulted in substantial glycogen accumulation and increased beta cell apoptosis (Brereton et al. 2016). In the islets of these diabetic mice, changes in levels of key metabolic genes were observed, including the downregulation of *G6pc2*, implying that the accumulation of G6P in the islets was driving glycogen formation. Importantly, these gene expression and ultrastructural changes were reversed upon restoration of euglycemia. The involvement of *G6pc2* in glycogen metabolism may be a contributor to impaired beta cell metabolic function in hyperglycemia.

Like *G6PC1*, *G6PC2* expression has also been shown to be regulated by glucocorticoids through its ability to bind the promoters of both genes, and studies have proposed that *G6PC2* exists to modulate FG in response to glucocorticoid-related stress (Boortz et al. 2016c; Boortz et al. 2016b; Vander Kooi et al. 2005). In times of energy-requiring activities or stress, glucocorticoids are known to restrict insulin secretion, stimulate hepatic gluconeogenesis and raise blood glucose levels (Delaunay et al. 1997; Lambillotte et al. 1997; McMahon et al. 1988). Long-term glucocorticoid treatment however leads to insulin resistance, compensatory increases in beta cell insulin secretion and eventual disruption of glucose tolerance (Ogawa et al. 1992; Rafacho et al. 2014; van Raalte et al. 2009). There appears to be great inherent complexity in the effects of glucocorticoids on whole body glucose metabolism *in vivo* that is dependent on dose, duration and the tissue involved (Kuo et al. 2015; Rafacho et al. 2014). The complexity of glucocorticoid biology may be further exemplified by *HSD11B1*, which encodes an enzyme responsible for intracellular glucocorticoid generation. Studies in mice showed that while *HSD11B1* overexpression in the beta cell resulted in impaired basal beta cell secretory function, a modest 'sub-threshold' elevation of *HSD11B1* counterintuitively enhanced basal insulin secretion (Turban et al. 2012). Following this, recent studies in mice showed that increased glucocorticoid action due to stress or treatment with the synthetic glucocorticoid dexamethasone (Dex) increased *G6pc2* expression, which transiently increases FG and protects against low blood glucose, relative to mice with *G6pc2* deletion (Boortz et al. 2016c; Boortz et al. 2016b). In these KO mice, FG was comparatively lowered both in the absence and presence of Dex, and improved glucose tolerance was observed under Dex treatment. Therefore, the authors

hypothesised that *G6pc2* initially evolved to modulate GSIS sensitivity and provide beneficial effects under conditions of stress. These new findings support the notion that glucose-responsiveness of beta cells may be in part tuned by the degree of *G6pc2* activity. Like in the case of *HSD11B1*, this may be likened to the concept of 'disallowed' genes, which are genes that are actively repressed in specific cell types such as the pancreatic islets to achieve optimal metabolic control and protect against inappropriate insulin release (Quintens et al. 2008; Schuit et al. 2012).

Additionally, *G6PC2* has also been unexpectedly linked to body weight. The FG-lowering allele at GWAS SNP rs560887 is linked to reduced body fat and BMI (Li et al. 2009). *G6PC2* KO mice also had reduced body fat and weight, as well as cholesterol, though this was dependent on gender and genetic background (Boortz et al. 2017; Pound et al. 2013). However, *G6PC2* has not been associated with BMI and body fat composition in subsequent large-scale GWAS. It is unclear whether the observed effects on body weight were a result of long-term changes in glycemic status rather than changes in *G6PC2* expression *per se*.

7.3. Future outlook

It is clear that a multitude of factors such as glucose (glucotoxicity), glucocorticoids and other stress conditions may regulate *G6PC2* expression. Complex relationships also exist between G6PC2 activity and regulation of beta cell function through calcium-dependent pathways and insulin levels. Unfortunately, the time constraints and technical challenges related to my studies meant that I was unable to carry out thorough investigation of G6PC2 in the EndoC-βH1 cells or in primary human islets. Future work will hopefully be directed at improving the understanding of its biological function especially in the context of calcium and ER homeostasis in the absence and presence of stimulants. These experiments should be conducted in multiple beta cell models where possible, as the EndoC-βH1 cells do not necessarily respond to ER stress in a manner consistent with that in human islets (Brozzi et al. 2015; Oleson et al. 2015). Further studies will also take advantage of current cutting-edge genome editing technologies that efficiently modify endogenous genes such as the CRISPR-Cas9 system (Sander and Joung 2014) in relevant islet cell models. These techniques not only enable silencing of endogenous *G6PC2* but also permit the introduction of specific

mutations into non-coding or coding regions of the gene using targeted or high throughput methods, to evaluate their effects on cellular function and phenotype (Claussnitzer et al. 2015; Wang et al. 2014). They will also circumvent the problems that come with an 'artificial' protein overexpression system.

Finally, the relevance of G6PC2 to disease and its potential as a therapeutic target remains an open question. Human genetic studies have so far suggested at most marginal associations of *G6PC2* with T2D susceptibility despite robust associations with glycemic traits, unlike in the case of the *GCK* (Dupuis et al. 2010). However, the coding variant association studies as discussed in Chapter 5 revealed particularly large effects of rare LOF variant P313L (β_{FG}=0.51 mmol/L; β_{HbA1c}=0.25%) with magnitudes that may generate substantial therapeutic interest in the management of FG and HbA1c levels. Importantly, the effect of *G6PC2* modulation is restricted to beta cells and is unlikely to affect glucose tolerance and glycemic levels in the pathophysiological range. GCK has been an attractive therapeutic target given its central role in glucose homeostasis and GCK activators (GKAs) are under development for control of blood glucose levels in diabetes (Matschinsky 2009). However, GCK activation in the beta cell risks hypoglycaemia (Bonadonna et al. 2010; Meininger et al. 2011) while activation in the liver has been found to induce liver triglyceride accumulation (De Ceuninck et al. 2013), therefore the potential complications that may arise from GKA use remain of great concern.

Overall, my results have helped to further our understanding of G6PC1 and G6PC2 which both play important and diverse regulatory roles in glucose homeostasis. The findings build on the current knowledge required for validating potential targets for the development of more effective therapeutic strategies to target prediabetes or T2D development. The integration of genetics analyses and functional pipelines including molecular, cellular, animal models and human studies will boost efforts to unravel the molecular bases of glycemic trait variation and T2D predisposition and to formulate improved therapeutic strategies.

Bibliography

1000 Genomes Project, Consortium, et al. (2012), 'An integrated map of genetic variation from 1,092 human genomes', *Nature,* 491 (7422), 56-65.

Ackermann, A. M., et al. (2016), 'Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes', *Mol Metab,* 5 (3), 233-44.

Aden, D. P., et al. (1979), 'Controlled synthesis of HBsAg in a differentiated human liver carcinoma-derived cell line', *Nature,* 282 (5739), 615-6.

Adzhubei, I. A., et al. (2010), 'A method and server for predicting damaging missense mutations', *Nat Methods,* 7 (4), 248-9.

Agius, L. (2008), 'Glucokinase and molecular aspects of liver glycogen metabolism', *Biochem J,* 414 (1), 1-18.

Alberti, K. G. and Zimmet, P. Z. (1998), 'Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation', *Diabet Med,* 15 (7), 539-53.

Albrechtsen, A., et al. (2013), 'Exome sequencing-driven discovery of coding polymorphisms associated with common metabolic phenotypes', *Diabetologia,* 56 (2), 298-310.

An, P., et al. (2014), 'Genome-wide association study identifies common loci influencing circulating glycated hemoglobin (HbA1c) levels in non-diabetic subjects: the Long Life Family Study (LLFS)', *Metabolism,* 63 (4), 461-8.

Andersson, L. E., et al. (2015), 'Characterization of stimulus-secretion coupling in the human pancreatic EndoC-betaH1 beta cell line', *PLoS One,* 10 (3), e0120879.

Angaroni, C. J., et al. (2004), 'Glycogen storage disease type Ia in Argentina: two novel glucose-6-phosphatase mutations affecting protein stability', *Mol Genet Metab,* 83 (3), 276-9.

Arden, S. D., et al. (1999), 'Molecular cloning of a pancreatic islet-specific glucose-6-phosphatase catalytic subunit-related protein', *Diabetes,* 48 (3), 531-42.

Ashcroft, F. M. and Rorsman, P. (2012), 'Diabetes mellitus and the beta cell: the last ten years', *Cell,* 148 (6), 1160-71.

Ashcroft, S. J. and Randle, P. J. (1968), 'Glucose-6-phosphatase activity of mouse pancreatic islets', *Nature,* 219 (5156), 857-8.

Ashmore, J., Hastings, A. B., and Nesbett, F. B. (1954), 'The Effect of Diabetes and Fasting on Liver Glucose-6-Phosphatase', *Proc Natl Acad Sci U S A,* 40 (8), 673-8.

Babenko, A. P. (2008), 'A novel ABCC8 (SUR1)-dependent mechanism of metabolism-excitation uncoupling', *J Biol Chem,* 283 (14), 8778-82.

Babenko, A. P., et al. (2006), 'Activating mutations in the ABCC8 gene in neonatal diabetes mellitus', *N Engl J Med,* 355 (5), 456-66.

Back, S. H. and Kaufman, R. J. (2012), 'Endoplasmic reticulum stress and type 2 diabetes', *Annu Rev Biochem,* 81, 767-93.

Baerenwald, D. A., et al. (2013), 'Multiple functional polymorphisms in the G6PC2 gene contribute to the association with higher fasting plasma glucose levels', *Diabetologia,* 56 (6), 1306-16.

Baron, M., et al. (2016), 'A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure', *Cell Syst,* 3 (4), 346-60 e4.

Barzilai, N. and Rossetti, L. (1993), 'Role of glucokinase and glucose-6-phosphatase in the acute and chronic regulation of hepatic glucose fluxes by insulin', *J Biol Chem,* 268 (33), 25019-25.

Basu, R., et al. (2013), 'Pathogenesis of prediabetes: role of the liver in isolated fasting hyperglycemia and combined fasting and postprandial hyperglycemia', *J Clin Endocrinol Metab,* 98 (3), E409-17.

Basu, S. and Pan, W. (2011), 'Comparison of statistical tests for disease association with rare variants', *Genet Epidemiol,* 35 (7), 606-19.

Beckman, J. A., Creager, M. A., and Libby, P. (2002), 'Diabetes and atherosclerosis: epidemiology, pathophysiology, and management', *JAMA,* 287 (19), 2570-81.
Beer, N. L., 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 (21), 4081-8.
Benazra, M., et al. (2015), 'A human beta cell line with drug inducible excision of immortalizing transgenes', *Mol Metab,* 4 (12), 916-25.
Benedetti, A., Fulceri, R., and Comporti, M. (1985), 'Calcium sequestration activity in rat liver microsomes. Evidence for a cooperation of calcium transport with glucose-6-phosphatase', *Biochim Biophys Acta,* 816 (2), 267-77.
Benyamin, B., et al. (2009), 'Variants in TF and HFE explain approximately 40% of genetic variation in serum-transferrin levels', *Am J Hum Genet,* 84 (1), 60-5.
Benyamin, B., et al. (2014), 'Novel loci affecting iron homeostasis and their effects in individuals at risk for hemochromatosis', *Nat Commun,* 5, 4926.
Bertram, R., Sherman, A., and Satin, L. S. (2010), 'Electrical bursting, calcium oscillations, and synchronization of pancreatic islets', *Adv Exp Med Biol,* 654, 261-79.
Beutler, E., et al. (2000), 'Molecular characterization of a case of atransferrinemia', *Blood,* 96 (13), 4071-4.
Bhandare, R., et al. (2010), 'Genome-wide analysis of histone modifications in human pancreatic islets', *Genome Res,* 20 (4), 428-33.
Billings, L. K. and Florez, J. C. (2010), 'The genetics of type 2 diabetes: what have we learned from GWAS?', *Ann N Y Acad Sci,* 1212, 59-77.
Blair, J. N. and Burchell, A. (1988), 'The mechanism of histone activation of the hepatic microsomal glucose-6-phosphatase system: a novel method to assay glucose-6-phosphatase activity', *Biochim Biophys Acta,* 964 (2), 161-7.
Blodgett, D. M., et al. (2015), 'Novel Observations From Next-Generation RNA Sequencing of Highly Purified Human Adult and Fetal Islet Cell Subsets', *Diabetes,* 64 (9), 3172-81.
Bonadonna, R. C., et al. (2010), 'Piragliatin (RO4389620), a novel glucokinase activator, lowers plasma glucose both in the postabsorptive state and after a glucose challenge in patients with type 2 diabetes mellitus: a mechanistic study', *J Clin Endocrinol Metab,* 95 (11), 5028-36.
Bonnefond, A., et al. (2009), 'Mutations in G6PC2 do not contribute to monogenic forms of early infancy diabetes and beta cell dysfunction', *Diabetologia,* 52 (5), 982-5.
Bonnefond, A., et al. (2012), 'Rare MTNR1B variants impairing melatonin receptor 1B function contribute to type 2 diabetes', *Nat Genet,* 44 (3), 297-301.
Boortz, K. A., et al. (2016a), 'Functional Analysis of Mouse G6pc1 Mutations Using a Novel In Situ Assay for Glucose-6-Phosphatase Activity and the Effect of Mutations in Conserved Human G6PC1/G6PC2 Amino Acids on G6PC2 Protein Expression', *PLoS One,* 11 (9), e0162439.
Boortz, K. A., et al. (2016b), 'G6PC2 Modulates the Effects of Dexamethasone on Fasting Blood Glucose and Glucose Tolerance', *Endocrinology*, en20161678.
Boortz, K. A., et al. (2017), 'Effects of G6pc2 deletion on body weight and cholesterol in mice', *J Mol Endocrinol*.
Boortz, K. A., et al. (2016c), 'G6PC2 Modulates Fasting Blood Glucose In Male Mice in Response to Stress', *Endocrinology,* 157 (8), 3002-8.
Bouatia-Naji, N., et al. (2008), 'A polymorphism within the G6PC2 gene is associated with fasting plasma glucose levels', *Science,* 320 (5879), 1085-8.
Bouatia-Naji, N., et al. (2010), 'Genetic and functional assessment of the role of the rs13431652-A and rs573225-A alleles in the G6PC2 promoter that are strongly associated with elevated fasting glucose levels', *Diabetes,* 59 (10), 2662-71.
Bouatia-Naji, N., et al. (2009), 'A variant near MTNR1B is associated with increased fasting plasma glucose levels and type 2 diabetes risk', *Nat Genet,* 41 (1), 89-94.

Boustead, J. N., et al. (2004), 'Identification and characterization of a cDNA and the gene encoding the mouse ubiquitously expressed glucose-6-phosphatase catalytic subunit-related protein', *J Mol Endocrinol,* 32 (1), 33-53.
Boyle, A. P., et al. (2012), 'Annotation of functional variation in personal genomes using RegulomeDB', *Genome Res,* 22 (9), 1790-7.
Brandhorst, H., et al. (1998), 'Assessment of intracellular insulin content during all steps of human islet isolation procedure', *Cell Transplant,* 7 (5), 489-95.
Brereton, M. F., et al. (2016), 'Hyperglycaemia induces metabolic dysfunction and glycogen accumulation in pancreatic beta-cells', *Nat Commun,* 7, 13496.
Brodsky, J. L. and McCracken, A. A. (1997), 'ER-associated and proteasomemediated protein degradation: how two topologically restricted events came together', *Trends Cell Biol,* 7 (4), 151-6.
Brozzi, F., et al. (2015), 'Cytokines induce endoplasmic reticulum stress in human, rat and mouse beta cells via different mechanisms', *Diabetologia,* 58 (10), 2307-16.
Bruni, N., et al. (1999), 'Enzymatic characterization of four new mutations in the glucose-6 phosphatase (G6PC) gene which cause glycogen storage disease type 1a', *Ann Hum Genet,* 63 (Pt 2), 141-6.
Burchell, A. (1992), 'The molecular basis of the type 1 glycogen storage diseases', *Bioessays,* 14 (6), 395-400.
Burchell, A., Hume, R., and Burchell, B. (1988), 'A new microtechnique for the analysis of the human hepatic microsomal glucose-6-phosphatase system', *Clin Chim Acta,* 173 (2), 183-91.
Cabrera, O., et al. (2006), 'The unique cytoarchitecture of human pancreatic islets has implications for islet cell function', *Proc Natl Acad Sci U S A,* 103 (7), 2334-9.
CARDIoGRAMplusC4D, et al. (2013), 'Large-scale association analysis identifies new risk loci for coronary artery disease', *Nat Genet,* 45 (1), 25-33.
Chandra, V., et al. (2014), 'RFX6 regulates insulin secretion by modulating Ca2+ homeostasis in human beta cells', *Cell Rep,* 9 (6), 2206-18.
Chandramouli, V., et al. (1991), 'Quantification of glucose cycling and the extent of equilibration of glucose 6-phosphate with fructose 6-phosphate in islets from ob/ob mice', *Biochem J,* 278 (Pt 2), 353-9.
Chen, L. Y., et al. (2002), 'Structure-function analysis of the glucose-6-phosphate transporter deficient in glycogen storage disease type Ib', *Hum Mol Genet,* 11 (25), 3199-207.
Chen, P., et al. (2014), 'Multiple nonglycemic genomic loci are newly associated with blood level of glycated hemoglobin in East Asians', *Diabetes,* 63 (7), 2551-62.
Chen, W. M., et al. (2008), 'Variations in the G6PC2/ABCB11 genomic region are associated with fasting glucose levels', *J Clin Invest,* 118 (7), 2620-8.
Chen, Y-T. and Burchell, A. (1995), 'Glycogen storage diseases', in Charles R. Scriver, et al. (eds.), *The metabolic basis of inherited diseases (7th ed)* (7th edn.; New York: McGraw-Hill), 925-65.
Chen, Y. T., Cornblath, M., and Sidbury, J. B. (1984), 'Cornstarch therapy in type I glycogen-storage disease', *N Engl J Med,* 310 (3), 171-5.
Cherrington, A. D. (1999), 'Banting Lecture 1997. Control of glucose uptake and release by the liver in vivo', *Diabetes,* 48 (5), 1198-214.
Cho, Y. S., et al. (2011), 'Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians', *Nat Genet,* 44 (1), 67-72.
Chou, J. Y. (2001), 'The molecular basis of type 1 glycogen storage diseases', *Curr Mol Med,* 1 (1), 25-44.
Chou, J. Y. and Mansfield, B. C. (1999), 'Molecular Genetics of Type 1 Glycogen Storage Diseases', *Trends Endocrinol Metab,* 10 (3), 104-13.
Chou, J. Y. and Mansfield, B. C. (2008), 'Mutations in the glucose-6-phosphatase-alpha (G6PC) gene that cause type Ia glycogen storage disease', *Hum Mutat,* 29 (7), 921-30.

Chou, J. Y., Jun, H. S., and Mansfield, B. C. (2010), 'Glycogen storage disease type I and G6Pase-beta deficiency: etiology and therapy', *Nat Rev Endocrinol,* 6 (12), 676-88.
Chou, J. Y., Jun, H. S., and Mansfield, B. C. (2015), 'Type I glycogen storage diseases: disorders of the glucose-6-phosphatase/glucose-6-phosphate transporter complexes', *J Inherit Metab Dis,* 38 (3), 511-9.
Chou, J. Y., et al. (2002), 'Type I glycogen storage diseases: disorders of the glucose-6-phosphatase complex', *Curr Mol Med,* 2 (2), 121-43.
Claussnitzer, M., et al. (2015), 'FTO Obesity Variant Circuitry and Adipocyte Browning in Humans', *N Engl J Med,* 373 (10), 895-907.
Claussnitzer, M., et al. (2014), 'Leveraging cross-species transcription factor binding site patterns: from diabetes risk loci to disease mechanisms', *Cell,* 156 (1-2), 343-58.
Cnop, M., Foufelle, F., and Velloso, L. A. (2012), 'Endoplasmic reticulum stress, obesity and diabetes', *Trends in Molecular Medicine,* 18 (1), 59-68.
Collins, S. C., et al. (2016), 'Increased Expression of the Diabetes Gene SOX4 Reduces Insulin Secretion by Impaired Fusion Pore Expansion', *Diabetes,* 65 (7), 1952-61.
Colombo, C., et al. (2008), 'Seven mutations in the human insulin gene linked to permanent neonatal/infancy-onset diabetes mellitus', *J Clin Invest,* 118 (6), 2148-56.
Consortium, G. TEx (2013), 'The Genotype-Tissue Expression (GTEx) project', *Nat Genet,* 45 (6), 580-5.
Cooper, G. M., et al. (2005), 'Distribution and intensity of constraint in mammalian genomic sequence', *Genome Res,* 15 (7), 901-13.
Cori, G. T. and Cori, C. F. (1952), 'Glucose-6-phosphatase of the liver in glycogen storage disease', *J Biol Chem,* 199 (2), 661-7.
Curi, R., et al. (2016), 'Regulatory principles in metabolism-then and now', *Biochem J,* 473 (13), 1845-57.
Danaei, G., et al. (2006), 'Global and regional mortality from ischaemic heart disease and stroke attributable to higher-than-optimum blood glucose concentration: comparative risk assessment', *Lancet,* 368 (9548), 1651-9.
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 (3), e1004160.
De Ceuninck, F., 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 (2), 339-53.
de Jong, V. M., et al. (2013), 'Alternative splicing and differential expression of the islet autoantigen IGRP between pancreas and thymus contributes to immunogenicity of pancreatic islets but not diabetogenicity in humans', *Diabetologia,* 56 (12), 2651-8.
de la Tour, D., et al. (2001), 'Beta-cell differentiation from a human pancreatic cell line in vitro and in vivo', *Mol Endocrinol,* 15 (3), 476-83.
Delaunay, F., et al. (1997), 'Pancreatic beta cells are important targets for the diabetogenic effects of glucocorticoids', *J Clin Invest,* 100 (8), 2094-8.
Delepine, M., et al. (2000), 'EIF2AK3, encoding translation initiation factor 2-alpha kinase 3, is mutated in patients with Wolcott-Rallison syndrome', *Nat Genet,* 25 (4), 406-9.
Dieckgraefe, B. K., et al. (2002), 'Association of glycogen storage disease 1b and Crohn disease: results of a North American survey', *Eur J Pediatr,* 161 Suppl 1, S88-92.
Dimas, A. S., et al. (2014), 'Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity', *Diabetes,* 63 (6), 2158-71.
Dogra, R. S., et al. (2006), 'Alternative splicing of G6PC2, the gene coding for the islet-specific glucose-6-phosphatase catalytic subunit-related protein (IGRP), results in differential expression in human thymus and spleen compared with pancreas', *Diabetologia,* 49 (5), 953-7.

Dupuis, J., et al. (2010), 'New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk', *Nat Genet,* 42 (2), 105-16.
Ebert, D. H., et al. (1999), 'Structure and promoter activity of an islet-specific glucose-6-phosphatase catalytic subunit-related gene', *Diabetes,* 48 (3), 543-51.
Edghill, E. L., et al. (2009), 'Mutations in the third gene shown to alter fasting glucose levels in the population (G6PC2) are not a common cause of monogenic forms of pancreatic B-cell dysfunction', *Diabet Med,* 26 (1), 113-4.
Eizirik, D. L., Cardozo, A. K., and Cnop, M. (2008), 'The role for endoplasmic reticulum stress in diabetes mellitus', *Endocrine Reviews,* 29 (1), 42-61.
Ekstein, J., et al. (2004), 'Mutation frequencies for glycogen storage disease Ia in the Ashkenazi Jewish population', *Am J Med Genet A,* 129A (2), 162-4.
Encode, Project Consortium (2012), 'An integrated encyclopedia of DNA elements in the human genome', *Nature,* 489 (7414), 57-74.
Fajans, S. S., Bell, G. I., and Polonsky, K. S. (2001), 'Molecular mechanisms and clinical pathophysiology of maturity-onset diabetes of the young', *N Engl J Med,* 345 (13), 971-80.
Fenteany, G., et al. (1995), 'Inhibition of proteasome activities and subunit-specific amino-terminal threonine modification by lactacystin', *Science,* 268 (5211), 726-31.
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 (2), 170-80.
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 (11), 1380-5.
Florez, J. C., et al. (2004), 'Haplotype structure and genotype-phenotype correlations of the sulfonylurea receptor and the islet ATP-sensitive potassium channel gene region', *Diabetes,* 53 (5), 1360-8.
Fonseca, S. G., et al. (2005), 'WFS1 is a novel component of the unfolded protein response and maintains homeostasis of the endoplasmic reticulum in pancreatic beta-cells', *J Biol Chem,* 280 (47), 39609-15.
Fortin, J. P., et al. (2010), 'Pharmacological characterization of human incretin receptor missense variants', *J Pharmacol Exp Ther,* 332 (1), 274-80.
Foster, J. D. and Nordlie, R. C. (2002), 'The biochemistry and molecular biology of the glucose-6-phosphatase system', *Exp Biol Med (Maywood),* 227 (8), 601-8.
Foster, J. D., Pederson, B. A., and Nordlie, R. C. (1997), 'Glucose-6-phosphatase structure, regulation, and function: an update', *Proc Soc Exp Biol Med,* 215 (4), 314-32.
Freedland, R. A. and Barnes, J. K. (1963), 'Kinetic properties and adaptations of liver glucose-6-phosphatase', *Proc Soc Exp Biol Med,* 112, 995-7.
Fuchsberger, C., et al. (2016), 'The genetic architecture of type 2 diabetes', *Nature,* 536 (7614), 41-7.
Fujiwara, T., Misumi, Y., and Ikehara, Y. (2003), 'Direct interaction of the Golgi membrane with the endoplasmic reticulum membrane caused by nordihydroguaiaretic acid', *Biochem Biophys Res Commun,* 301 (4), 927-33.
Fulceri, R., et al. (1990), 'MgATP-dependent accumulation of calcium ions and inorganic phosphate in a liver reticular pool', *Biochem J,* 272 (2), 549-52.
Gamazon, E. R., et al. (2015), 'A gene-based association method for mapping traits using reference transcriptome data', *Nat Genet,* 47 (9), 1091-8.
Gaulton, K. J., et al. (2010), 'A map of open chromatin in human pancreatic islets', *Nat Genet,* 42 (3), 255-9.
Gaulton, K. J., et al. (2015), 'Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci', *Nat Genet,* 47 (12), 1415-25.
Ghosh, A., et al. (2002), 'The catalytic center of glucose-6-phosphatase. HIS176 is the nucleophile forming the phosphohistidine-enzyme intermediate during catalysis', *J Biol Chem,* 277 (36), 32837-42.

Ghosh, A., et al. (2006), 'Long-term correction of murine glycogen storage disease type Ia by recombinant adeno-associated virus-1-mediated gene transfer', *Gene Ther,* 13 (4), 321-9.
Gilon, P., et al. (2014), 'Calcium signaling in pancreatic beta-cells in health and in Type 2 diabetes', *Cell Calcium,* 56 (5), 340-61.
Gilon, P., et al. (1999), 'Uptake and release of Ca2+ by the endoplasmic reticulum contribute to the oscillations of the cytosolic Ca2+ concentration triggered by Ca2+ influx in the electrically excitable pancreatic B-cell', *J Biol Chem,* 274 (29), 20197-205.
Gloyn, A. L., et al. (2003), 'Large-scale association studies of variants in genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes', *Diabetes,* 52 (2), 568-72.
Gloyn, A. L., 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 (18), 1838-49.
Goldstein, J. L., Brunschede, G. Y., and Brown, M. S. (1975), 'Inhibition of proteolytic degradation of low density lipoprotein in human fibroblasts by chloroquine, concanavalin A, and Triton WR 1339', *J Biol Chem,* 250 (19), 7854-62.
Gonzalez-Perez, A. and Lopez-Bigas, N. (2011), 'Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel', *Am J Hum Genet,* 88 (4), 440-9.
Graham, F. L., et al. (1977), 'Characteristics of a human cell line transformed by DNA from human adenovirus type 5', *J Gen Virol,* 36 (1), 59-74.
Greene, H. L., et al. (1976), 'Continuous nocturnal intragastric feeding for management of type 1 glycogen-storage disease', *N Engl J Med,* 294 (8), 423-5.
Group, Advance Collaborative, et al. (2008), 'Intensive blood glucose control and vascular outcomes in patients with type 2 diabetes', *N Engl J Med,* 358 (24), 2560-72.
Guionie, O., et al. (2003), 'Identification and characterisation of a new human glucose-6-phosphatase isoform', *FEBS Lett,* 551 (1-3), 159-64.
Gurgul-Convey, E., Kaminski, M. T., and Lenzen, S. (2015), 'Physiological characterization of the human EndoC-betaH1 beta-cell line', *Biochem Biophys Res Commun,* 464 (1), 13-9.
Gurgul-Convey, E., et al. (2016), 'Sensitivity profile of the human EndoC-betaH1 beta cell line to proinflammatory cytokines', *Diabetologia,* 59 (10), 2125-33.
Hamming, K. S., 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 (10), 2419-24.
Hasnain, S. Z., Prins, J. B., and McGuckin, M. A. (2016), 'Oxidative and endoplasmic reticulum stress in beta-cell dysfunction in diabetes', *J Mol Endocrinol,* 56 (2), R33-54.
Hegele, R. A., et al. (2000), 'Association between nuclear lamin A/C R482Q mutation and partial lipodystrophy with hyperinsulinemia, dyslipidemia, hypertension, and diabetes', *Genome Res,* 10 (5), 652-8.
Hemrika, W., et al. (1997), 'From phosphatases to vanadium peroxidases: a similar architecture of the active site', *Proc Natl Acad Sci U S A,* 94 (6), 2145-9.
Heni, M., et al. (2010), 'The impact of genetic variation in the G6PC2 gene on insulin secretion depends on glycemia', *J Clin Endocrinol Metab,* 95 (12), E479-84.
Henquin, J. C. (2000), 'Triggering and amplifying pathways of regulation of insulin secretion by glucose', *Diabetes,* 49 (11), 1751-60.
Henquin, J. C., Dufrane, D., and Nenquin, M. (2006), 'Nutrient control of insulin secretion in isolated normal human islets', *Diabetes,* 55 (12), 3470-7.
Henquin, J. C., et al. (2015), 'Dynamics of glucose-induced insulin secretion in normal human islets', *Am J Physiol Endocrinol Metab,* 309 (7), E640-50.
Hiraiwa, H., et al. (1999), 'Inactivation of the glucose 6-phosphate transporter causes glycogen storage disease type 1b', *J Biol Chem,* 274 (9), 5532-6.

Hohmeier, H. E., et al. (2000), 'Isolation of INS-1-derived cell lines with robust ATP-sensitive K+ channel-dependent and -independent glucose-stimulated insulin secretion', *Diabetes,* 49 (3), 424-30.
Holman, R. R., et al. (2008), '10-year follow-up of intensive glucose control in type 2 diabetes', *N Engl J Med,* 359 (15), 1577-89.
Hu, C., et al. (2009), 'A genetic variant of G6PC2 is associated with type 2 diabetes and fasting plasma glucose level in the Chinese population', *Diabetologia,* 52 (3), 451-6.
Hu, C., et al. (2010), 'Effects of GCK, GCKR, G6PC2 and MTNR1B variants on glucose metabolism and insulin secretion', *PLoS One,* 5 (7), e11761.
Hughes, J. R., et al. (2014), 'Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment', *Nat Genet,* 46 (2), 205-12.
Hutton, J. C. and Eisenbarth, G. S. (2003), 'A pancreatic beta-cell-specific homolog of glucose-6-phosphatase emerges as a major target of cell-mediated autoimmunity in diabetes', *Proc Natl Acad Sci U S A,* 100 (15), 8626-8.
Hutton, J. C. and O'Brien, R. M. (2009), 'Glucose-6-phosphatase catalytic subunit gene family', *J Biol Chem,* 284 (43), 29241-5.
Huyghe, J. R., et al. (2013), 'Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion', *Nat Genet,* 45 (2), 197-201.
IDF (2015), *IDF Diabetes Atlas 7th edn,* Brussels, Belgium: International Diabetes Federation.
Iizuka, K., et al. (2000), 'Stable overexpression of the glucose-6-phosphatase catalytic subunit attenuates glucose sensitivity of insulin secretion from a mouse pancreatic beta-cell line', *J Endocrinol,* 164 (3), 307-14.
Ingelsson, E., et al. (2010), 'Detailed physiologic characterization reveals diverse mechanisms for novel genetic Loci regulating glucose and insulin metabolism in humans', *Diabetes,* 59 (5), 1266-75.
Inoue, H., et al. (1998), 'A gene encoding a transmembrane protein is mutated in patients with diabetes mellitus and optic atrophy (Wolfram syndrome)', *Nat Genet,* 20 (2), 143-8.
International HapMap, Consortium (2005), 'A haplotype map of the human genome', *Nature,* 437 (7063), 1299-320.
International HapMap, Consortium, et al. (2007), 'A second generation human haplotype map of over 3.1 million SNPs', *Nature,* 449 (7164), 851-61.
Jackson, M. R., Nilsson, T., and Peterson, P. A. (1990), 'Identification of a consensus motif for retention of transmembrane proteins in the endoplasmic reticulum', *EMBO J,* 9 (10), 3153-62.
Jarchum, I., et al. (2008), 'Identification of novel IGRP epitopes targeted in type 1 diabetes patients', *Clin Immunol,* 127 (3), 359-65.
Jensen, M. V., et al. (2008), 'Metabolic cycling in control of glucose-stimulated insulin secretion', *Am J Physiol Endocrinol Metab,* 295 (6), E1287-97.
Juang, J. M., et al. (2014), 'Utilizing multiple in silico analyses to identify putative causal SCN5A variants in Brugada syndrome', *Sci Rep,* 4, 3850.
Kahn, S. E. (2003), 'The relative contributions of insulin resistance and beta-cell dysfunction to the pathophysiology of Type 2 diabetes', *Diabetologia,* 46 (1), 3-19.
Katz, Joseph and Rognstad, Robert (1978), 'Futile cycling in glucose metabolism', *Trends in Biochemical Sciences,* 3 (3), 171-74.
Kelley, L. A., et al. (2015), 'The Phyre2 web portal for protein modeling, prediction and analysis', *Nat Protoc,* 10 (6), 845-58.
Khan, A., Hong-Lie, C., and Landau, B. R. (1995), 'Glucose-6-phosphatase activity in islets from ob/ob and lean mice and the effect of dexamethasone', *Endocrinology,* 136 (5), 1934-8.
Khan, A., et al. (1990a), 'Glucose cycling in islets from healthy and diabetic rats', *Diabetes,* 39 (4), 456-9.

Khan, A., et al. (1990b), 'Glucose cycling is markedly enhanced in pancreatic islets of obese hyperglycemic mice', *Endocrinology,* 126 (5), 2413-6.
Khan, A., et al. (1989), 'Evidence for the presence of glucose cycling in pancreatic islets of the ob/ob mouse', *J Biol Chem,* 264 (17), 9732-3.
Kiezun, A., et al. (2012), 'Exome sequencing and the genetic basis of complex traits', *Nat Genet,* 44 (6), 623-30.
Kilpelainen, T. O., et al. (2011), 'Genetic variation near IRS1 associates with reduced adiposity and an impaired metabolic profile', *Nat Genet,* 43 (8), 753-60.
Kim, I., Xu, W., and Reed, J. C. (2008), 'Cell death and endoplasmic reticulum stress: disease relevance and therapeutic opportunities', *Nat Rev Drug Discov,* 7 (12), 1013-30.
Kircher, M., et al. (2014), 'A general framework for estimating the relative pathogenicity of human genetic variants', *Nat Genet,* 46 (3), 310-5.
Knight, J. C. (2014), 'Approaches for establishing the function of regulatory genetic variants involved in disease', *Genome Med,* 6 (10), 92.
Knisely, A. S., Gelbart, T., and Beutler, E. (2004), 'Molecular characterization of a third case of human atransferrinemia', *Blood,* 104 (8), 2607.
Koster, J. C., et al. (2000), 'Targeted overactivity of beta cell K(ATP) channels induces profound neonatal diabetes', *Cell,* 100 (6), 645-54.
Krishnamurthy, B., et al. (2006a), 'Responses against islet antigens in NOD mice are prevented by tolerance to proinsulin but not IGRP', *J Clin Invest,* 116 (12), 3258-65.
Krishnamurthy, J., et al. (2006b), 'p16INK4a induces an age-dependent decline in islet regenerative potential', *Nature,* 443 (7110), 453-7.
Krishnan, K., et al. (2015), 'Calcium signaling in a genetically engineered human pancreatic beta-cell line', *Pancreas,* 44 (5), 773-7.
Kumar, P., Henikoff, S., and Ng, P. C. (2009), 'Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm', *Nat Protoc,* 4 (7), 1073-81.
Kuo, T., et al. (2015), 'Regulation of Glucose Homeostasis by Glucocorticoids', *Adv Exp Med Biol,* 872, 99-126.
Laakso, M. (1993), 'How good a marker is insulin level for insulin resistance?', *Am J Epidemiol,* 137 (9), 959-65.
Lam, P., et al. (2007), 'Levels of plasma membrane expression in progressive and benign mutations of the bile salt export pump (Bsep/Abcb11) correlate with severity of cholestatic diseases', *Am J Physiol Cell Physiol,* 293 (5), C1709-16.
Lambillotte, C., Gilon, P., and Henquin, J. C. (1997), 'Direct glucocorticoid inhibition of insulin secretion. An in vitro study of dexamethasone effects in mouse islets', *J Clin Invest,* 99 (3), 414-23.
Lang, D. A., et al. (1979), 'Cyclic oscillations of basal plasma glucose and insulin concentrations in human beings', *N Engl J Med,* 301 (19), 1023-7.
Lange, L. A., et al. (2014), 'Whole-exome sequencing identifies rare and low-frequency coding variants associated with LDL cholesterol', *Am J Hum Genet,* 94 (2), 233-45.
Lango Allen, H., et al. (2011), 'GATA6 haploinsufficiency causes pancreatic agenesis in humans', *Nat Genet,* 44 (1), 20-2.
Lappalainen, T., et al. (2011), 'Epistatic selection between coding and regulatory variation in human evolution and disease', *Am J Hum Genet,* 89 (3), 459-63.
Lawlor, N., et al. (2017), 'Single-cell transcriptomes identify human islet cell signatures and reveal cell-type-specific expression changes in type 2 diabetes', *Genome Res,* 27 (2), 208-22.
Lei, K. J., et al. (1993), 'Mutations in the glucose-6-phosphatase gene that cause glycogen storage disease type 1a', *Science,* 262 (5133), 580-3.
Lei, K. J., et al. (1994), 'Identification of mutations in the gene for glucose-6-phosphatase, the enzyme deficient in glycogen storage disease type 1a', *J Clin Invest,* 93 (5), 1994-9.

Lei, K. J., et al. (1995a), 'Structure-function analysis of human glucose-6-phosphatase, the enzyme deficient in glycogen storage disease type 1a', *J Biol Chem,* 270 (20), 11882-6.
Lei, K. J., et al. (1996), 'Glucose-6-phosphatase dependent substrate transport in the glycogen storage disease type-1a mouse', *Nat Genet,* 13 (2), 203-9.
Lei, K. J., et al. (1995b), 'Genetic basis of glycogen storage disease type 1a: prevalent mutations at the glucose-6-phosphatase locus', *Am J Hum Genet,* 57 (4), 766-71.
Lek, M., et al. (2016), 'Analysis of protein-coding genetic variation in 60,706 humans', *Nature,* 536 (7616), 285-91.
Levine, F., et al. (1995), 'Development of a cell line from the human fetal pancreas', *Transplant Proc,* 27 (6), 3410.
Li, X., et al. (2009), 'Additive effects of genetic variation in GCK and G6PC2 on insulin secretion and fasting glucose', *Diabetes,* 58 (12), 2946-53.
Li, Y., Mechin, M. C., and van de Werve, G. (1999), 'Diabetes affects similarly the catalytic subunit and putative glucose-6-phosphate translocase of glucose-6-phosphatase', *J Biol Chem,* 274 (48), 33866-8.
Lieberman-Aiden, E., et al. (2009), 'Comprehensive mapping of long-range interactions reveals folding principles of the human genome', *Science,* 326 (5950), 289-93.
Lieberman, S. M., et al. (2003), 'Identification of the beta cell antigen targeted by a prevalent population of pathogenic CD8+ T cells in autoimmune diabetes', *Proc Natl Acad Sci U S A,* 100 (14), 8384-8.
Lin, S. R., et al. (2015), 'Functional analysis of mutations in a severe congenital neutropenia syndrome caused by glucose-6-phosphatase-beta deficiency', *Mol Genet Metab,* 114 (1), 41-5.
Lin, W. Y. (2016), 'Beyond Rare-Variant Association Testing: Pinpointing Rare Causal Variants in Case-Control Sequencing Study', *Sci Rep,* 6, 21824.
Locke, A. E., et al. (2015), 'Genetic studies of body mass index yield new insights for obesity biology', *Nature,* 518 (7538), 197-206.
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 (1), 82-8.
Magi, R., Kumar, A., and Morris, A. P. (2011), 'Assessing the impact of missing genotype data in rare variant association analysis', *BMC Proc,* 5 Suppl 9, S107.
Mahajan, A., et al. (2015), 'Identification and functional characterization of G6PC2 coding variants influencing glycemic traits define an effector transcript at the G6PC2-ABCB11 locus', *PLoS Genet,* 11 (1), e1004876.
Mahajan, A., et al. (2014), 'Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility', *Nat Genet,* 46 (3), 234-44.
Majithia, A. R., 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 (36), 13127-32.
Majithia, A. R., et al. (2016), 'Prospective functional classification of all possible missense variants in PPARG', *Nat Genet*.
Mallone, R., et al. (2007), 'CD8+ T-cell responses identify beta-cell autoimmunity in human type 1 diabetes', *Diabetes,* 56 (3), 613-21.
Manning, A. K., et al. (2012), 'A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance', *Nat Genet,* 44 (6), 659-69.
Manolio, T. A., et al. (2009), 'Finding the missing heritability of complex diseases', *Nature,* 461 (7265), 747-53.
Marcolongo, P., et al. (2013), 'Multiple roles of glucose-6-phosphatases in pathophysiology: state of the art and future trends', *Biochim Biophys Acta,* 1830 (3), 2608-18.

Marmugi, A., et al. (2016), 'Sorcin Links Pancreatic beta-Cell Lipotoxicity to ER Ca2+ Stores', *Diabetes,* 65 (4), 1009-21.
Marth, G. T., et al. (2011), 'The functional spectrum of low-frequency coding variation', *Genome Biol,* 12 (9), R84.
Martin, C. C., et al. (2002), 'Identification and characterization of a human cDNA and gene encoding a ubiquitously expressed glucose-6-phosphatase catalytic subunit-related protein', *J Mol Endocrinol,* 29 (2), 205-22.
Martin, C. C., et al. (2001), 'Cloning and characterization of the human and rat islet-specific glucose-6-phosphatase catalytic subunit-related protein (IGRP) genes', *J Biol Chem,* 276 (27), 25197-207.
Marullo, L., El-Sayed Moustafa, J. S., and Prokopenko, I. (2014), 'Insights into the genetic susceptibility to type 2 diabetes from genome-wide association studies of glycaemic traits', *Curr Diab Rep,* 14 (11), 551.
Matschinsky, F. M. (1996), 'Banting Lecture 1995. A lesson in metabolic regulation inspired by the glucokinase glucose sensor paradigm', *Diabetes,* 45 (2), 223-41.
Matschinsky, F. M. (2005), 'Glucokinase, glucose homeostasis, and diabetes mellitus', *Curr Diab Rep,* 5 (3), 171-6.
Matschinsky, F. M. (2009), 'Assessing the potential of glucokinase activators in diabetes therapy', *Nat Rev Drug Discov,* 8 (5), 399-416.
Matschinsky, F. M., et al. (2006), 'The network of glucokinase-expressing cells in glucose homeostasis and the potential of glucokinase activators for diabetes therapy', *Diabetes,* 55 (1), 1-12.
Matthews, D. R., et al. (1985), 'Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man', *Diabetologia,* 28 (7), 412-9.
McCarthy, M. I., et al. (2008), 'Genome-wide association studies for complex traits: consensus, uncertainty and challenges', *Nat Rev Genet,* 9 (5), 356-69.
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 (4), 648-53.
McLaren, W., et al. (2010), 'Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor', *Bioinformatics,* 26 (16), 2069-70.
McMahon, M., Gerich, J., and Rizza, R. (1988), 'Effects of glucocorticoids on carbohydrate metabolism', *Diabetes Metab Rev,* 4 (1), 17-30.
Meininger, G. E., 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 (12), 2560-6.
Merino, Jordi, et al. (2016), 'Genetically Driven Hyperglycemia Increases Risk of Coronary Artery Disease Distinctly from Type 2 Diabetes', *Endocrine Society's 98th Annual Meeting and Expo*, CR14-CR14.
Merrins, M. J., et al. (2012), 'Phosphofructo-2-kinase/fructose-2,6-bisphosphatase modulates oscillations of pancreatic islet metabolism', *PLoS One,* 7 (4), e34036.
Merrins, M. J., et al. (2010), 'Metabolic oscillations in pancreatic islets depend on the intracellular Ca2+ level but not Ca2+ oscillations', *Biophys J,* 99 (1), 76-84.
Mesa, J. L., et al. (2007), 'Lamin A/C polymorphisms, type 2 diabetes, and the metabolic syndrome: case-control and quantitative trait studies', *Diabetes,* 56 (3), 884-9.
Messerschmidt, A., Prade, L., and Wever, R. (1997), 'Implications for the catalytic mechanism of the vanadium-containing enzyme chloroperoxidase from the fungus Curvularia inaequalis by X-ray structures of the native and peroxide form', *Biol Chem,* 378 (3-4), 309-15.
Mithieux, G. (1997), 'New knowledge regarding glucose-6 phosphatase gene and protein and their roles in the regulation of glucose metabolism', *Eur J Endocrinol,* 136 (2), 137-45.
Mithieux, G., et al. (1996), 'Glucose-6-phosphatase mRNA and activity are increased to the same extent in kidney and liver of diabetic rats', *Diabetes,* 45 (7), 891-6.

Mitrakou, A., et al. (1990), 'Contribution of abnormal muscle and liver glucose metabolism to postprandial hyperglycemia in NIDDM', *Diabetes,* 39 (11), 1381-90.
Mohlke, K. L. and Boehnke, M. (2015), 'Recent advances in understanding the genetic architecture of type 2 diabetes', *Hum Mol Genet,* 24 (R1), R85-92.
Moltke, I., et al. (2014), 'A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes', *Nature,* 512 (7513), 190-3.
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 (4), 435-48.
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 (9), 981-90.
Moutsianas, L., et al. (2015), 'The power of gene-based rare variant methods to detect disease-associated variation and test hypotheses about complex disease', *PLoS Genet,* 11 (4), e1005165.
Myung, J., Kim, K. B., and Crews, C. M. (2001), 'The ubiquitin-proteasome pathway and proteasome inhibitors', *Med Res Rev,* 21 (4), 245-73.
Nakabayashi, H., et al. (1982), 'Growth of human hepatoma cells lines with differentiated functions in chemically defined medium', *Cancer Res,* 42 (9), 3858-63.
Narushima, M., et al. (2005), 'A human beta-cell line for transplantation therapy to control type 1 diabetes', *Nat Biotechnol,* 23 (10), 1274-82.
Nathan, D. M., et al. (2007), 'Impaired fasting glucose and impaired glucose tolerance: implications for care', *Diabetes Care,* 30 (3), 753-9.
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 (6), 778-85.
Nica, A. C., et al. (2013), 'Cell-type, allelic, and genetic signatures in the human pancreatic beta cell transcriptome', *Genome Res,* 23 (9), 1554-62.
Nikpay, M., et al. (2015), 'A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease', *Nat Genet,* 47 (10), 1121-30.
Noordeen, N. A., et al. (2012), 'Glucose-induced nuclear shuttling of ChREBP is mediated by sorcin and Ca(2+) ions in pancreatic beta-cells', *Diabetes,* 61 (3), 574-85.
Nordlie, R. C., Foster, J. D., and Lange, A. J. (1999), 'Regulation of glucose production by the liver', *Annu Rev Nutr,* 19, 379-406.
O'Brien, R. M. (2013), 'Moving on from GWAS: functional studies on the G6PC2 gene implicated in the regulation of fasting blood glucose', *Curr Diab Rep,* 13 (6), 768-77.
Oeser, J. K., et al. (2011), 'Deletion of the G6pc2 gene encoding the islet-specific glucose-6-phosphatase catalytic subunit-related protein does not affect the progression or incidence of type 1 diabetes in NOD/ShiLtJ mice', *Diabetes,* 60 (11), 2922-7.
Ogawa, A., et al. (1992), 'Roles of insulin resistance and beta-cell dysfunction in dexamethasone-induced diabetes', *J Clin Invest,* 90 (2), 497-504.
Olefsky, J., Farquhar, J. W., and Reaven, G. (1973), 'Relationship between fasting plasma insulin level and resistance to insulin-mediated glucose uptake in normal and diabetic subjects', *Diabetes,* 22 (7), 507-13.
Oleson, B. J., et al. (2015), 'Distinct differences in the responses of the human pancreatic beta-cell line EndoC-betaH1 and human islets to proinflammatory cytokines', *Am J Physiol Regul Integr Comp Physiol,* 309 (5), R525-34.
Orho-Melander, M., 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 (11), 3112-21.
Osbak, K. K., et al. (2009), 'Update on mutations in glucokinase (GCK), which cause maturity-onset diabetes of the young, permanent neonatal diabetes, and hyperinsulinemic hypoglycemia', *Hum Mutat,* 30 (11), 1512-26.

Ostenson, C. G., et al. (1993), 'Abnormal insulin secretion and glucose metabolism in pancreatic islets from the spontaneously diabetic GK rat', *Diabetologia,* 36 (1), 3-8.
Ouyang, Q., et al. (2006), 'Recognition of HLA class I-restricted beta-cell epitopes in type 1 diabetes', *Diabetes,* 55 (11), 3068-74.
Pagliuca, F. W., et al. (2014), 'Generation of functional human pancreatic beta cells in vitro', *Cell,* 159 (2), 428-39.
Pal, A., et al. (2016), 'Loss-of-Function Mutations in the Cell-Cycle Control Gene CDKN2A Impact on Glucose Homeostasis in Humans', *Diabetes,* 65 (2), 527-33.
Palombella, V. J., et al. (1994), 'The ubiquitin-proteasome pathway is required for processing the NF-kappa B1 precursor protein and the activation of NF-kappa B', *Cell,* 78 (5), 773-85.
Pan, C. J., Lei, K. J., and Chou, J. Y. (1998a), 'Asparagine-linked oligosaccharides are localized to a luminal hydrophilic loop in human glucose-6-phosphatase', *J Biol Chem,* 273 (34), 21658-62.
Pan, C. J., et al. (1998b), 'Transmembrane topology of glucose-6-phosphatase', *J Biol Chem,* 273 (11), 6144-8.
Panzer, C., et al. (2002), 'Association of fasting plasma glucose with heart rate recovery in healthy adults: a population-based study', *Diabetes,* 51 (3), 803-7.
Parker, S. C., 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 (44), 17921-6.
Pasquali, L., et al. (2014), 'Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants', *Nat Genet,* 46 (2), 136-43.
Pearson, E. R., et al. (2006), 'Switching from insulin to oral sulfonylureas in patients with diabetes due to Kir6.2 mutations', *N Engl J Med,* 355 (5), 467-77.
Pederson, B. A., Foster, J. D., and Nordlie, R. C. (1998), 'Histone II-A activates the glucose-6-phosphatase system without microsomal membrane permeabilization', *Arch Biochem Biophys,* 357 (1), 173-7.
Perales, M. A., Sener, A., and Malaisse, W. J. (1991), 'Hexose metabolism in pancreatic islets: the glucose-6-phosphatase riddle', *Mol Cell Biochem,* 101 (1), 67-71.
Petrolonis, A. J., et al. (2004), 'Enzymatic characterization of the pancreatic islet-specific glucose-6-phosphatase-related protein (IGRP)', *J Biol Chem,* 279 (14), 13976-83.
Plenge, R. M., Scolnick, E. M., and Altshuler, D. (2013), 'Validating therapeutic targets through human genetics', *Nat Rev Drug Discov,* 12 (8), 581-94.
Polychronakos, C. and Li, Q. (2011), 'Understanding type 1 diabetes through genetics: advances and prospects', *Nat Rev Genet,* 12 (11), 781-92.
Ponten, F., Jirstrom, K., and Uhlen, M. (2008), 'The Human Protein Atlas--a tool for pathology', *J Pathol,* 216 (4), 387-93.
Pound, L. D., et al. (2013), 'G6PC2: a negative regulator of basal glucose-stimulated insulin secretion', *Diabetes,* 62 (5), 1547-56.
Prasad, R. B. and Groop, L. (2015), 'Genetics of type 2 diabetes-pitfalls and possibilities', *Genes (Basel),* 6 (1), 87-123.
Prokopenko, I., et al. (2014), 'A central role for GRB10 in regulation of islet function in man', *PLoS Genet,* 10 (4), e1004235.
Prokopenko, I., et al. (2009), 'Variants in MTNR1B influence fasting glucose levels', *Nat Genet,* 41 (1), 77-81.
Purcell, S. M., et al. (2014), 'A polygenic burden of rare disruptive mutations in schizophrenia', *Nature,* 506 (7487), 185-90.
Quintens, R., et al. (2008), 'Why expression of some genes is disallowed in beta-cells', *Biochem Soc Trans,* 36 (Pt 3), 300-5.
Rafacho, A., et al. (2014), 'Glucocorticoid treatment and endocrine pancreas function: implications for glucose homeostasis, insulin resistance and diabetes', *J Endocrinol,* 223 (3), R49-62.

Rajas, F., et al. (2015), 'Lessons from new mouse models of glycogen storage disease type 1a in relation to the time course and organ specificity of the disease', *J Inherit Metab Dis,* 38 (3), 521-7.
Rake, J. P., et al. (2002), 'Glycogen storage disease type I: diagnosis, management, clinical course and outcome. Results of the European Study on Glycogen Storage Disease Type I (ESGSD I)', *Eur J Pediatr,* 161 Suppl 1, S20-34.
Rane, S. G., et al. (1999), 'Loss of Cdk4 expression causes insulin-deficient diabetes and Cdk4 activation results in beta-islet cell hyperplasia', *Nat Genet,* 22 (1), 44-52.
Ravassard, P., et al. (2011), 'A genetically engineered human pancreatic beta cell line exhibiting glucose-inducible insulin secretion', *J Clin Invest,* 121 (9), 3589-97.
Ravussin, E. and Smith, S. R. (2002), 'Increased fat intake, impaired fat oxidation, and failure of fat cell proliferation result in ectopic fat storage, insulin resistance, and type 2 diabetes mellitus', *Ann N Y Acad Sci,* 967, 363-78.
Ray, K. K., et al. (2009), 'Effect of intensive control of glucose on cardiovascular outcomes and death in patients with diabetes mellitus: a meta-analysis of randomised controlled trials', *Lancet,* 373 (9677), 1765-72.
Rees, M. G., et al. (2012a), 'Cellular characterisation of the GCKR P446L variant associated with type 2 diabetes risk', *Diabetologia,* 55 (1), 114-22.
Rees, M. G., et al. (2012b), 'Correlation of rare coding variants in the gene encoding human glucokinase regulatory protein with phenotypic, cellular, and kinetic outcomes', *J Clin Invest,* 122 (1), 205-17.
Rezania, A., et al. (2014), 'Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells', *Nat Biotechnol,* 32 (11), 1121-33.
Riedel, M. J., et al. (2003), 'Kir6.2 polymorphisms sensitize beta-cell ATP-sensitive potassium channels to activation by acyl CoAs: a possible cellular mechanism for increased susceptibility to type 2 diabetes?', *Diabetes,* 52 (10), 2630-5.
Riggs, A. C., et al. (2005), 'Mice conditionally lacking the Wolfram gene in pancreatic islet beta cells exhibit diabetes as a result of enhanced endoplasmic reticulum stress and apoptosis', *Diabetologia,* 48 (11), 2313-21.
Rizza, R. A. (2010), 'Pathogenesis of fasting and postprandial hyperglycemia in type 2 diabetes: implications for therapy', *Diabetes,* 59 (11), 2697-707.
Roe, M. W., et al. (1993), 'Voltage-dependent intracellular calcium release from mouse islets stimulated by glucose', *J Biol Chem,* 268 (14), 9953-6.
Ron, D. and Walter, P. (2007), 'Signal integration in the endoplasmic reticulum unfolded protein response', *Nat Rev Mol Cell Biol,* 8 (7), 519-29.
Rorsman, P. and Braun, M. (2013), 'Regulation of insulin secretion in human pancreatic islets', *Annu Rev Physiol,* 75, 155-79.
Rorsman, P., Braun, M., and Zhang, Q. (2012), 'Regulation of calcium in pancreatic alpha- and beta-cells in health and disease', *Cell Calcium,* 51 (3-4), 300-8.
Rose, C. S., et al. (2005), 'A -30G>A polymorphism of the beta-cell-specific glucokinase promoter associates with hyperglycemia in the general population of whites', *Diabetes,* 54 (10), 3026-31.
Rose, C. S., et al. (2009), 'A variant in the G6PC2/ABCB11 locus is associated with increased fasting plasma glucose, increased basal hepatic glucose production and increased insulin release after oral and intravenous glucose loads', *Diabetologia,* 52 (10), 2122-9.
Rung, J., et al. (2009), 'Genetic variant near IRS1 is associated with type 2 diabetes, insulin resistance and hyperinsulinemia', *Nat Genet,* 41 (10), 1110-5.
Russ, H. A., et al. (2015), 'Controlled induction of human pancreatic progenitors produces functional beta-like cells in vitro', *EMBO J,* 34 (13), 1759-72.
Samuel, V. T., et al. (2009), 'Fasting hyperglycemia is not associated with increased expression of PEPCK or G6Pc in patients with Type 2 Diabetes', *Proc Natl Acad Sci U S A,* 106 (29), 12121-6.

Sander, J. D. and Joung, J. K. (2014), 'CRISPR-Cas systems for editing, regulating and targeting genomes', *Nat Biotechnol,* 32 (4), 347-55.
Sarwar, N., et al. (2010), 'Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies', *Lancet,* 375 (9733), 2215-22.
Saxena, R., et al. (2007), 'Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels', *Science,* 316 (5829), 1331-6.
Saxena, R., et al. (2010), 'Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge', *Nat Genet,* 42 (2), 142-8.
Scharfmann, R., et al. (2014), 'Development of a conditionally immortalized human pancreatic beta cell line', *J Clin Invest,* 124 (5), 2087-98.
Schuit, F., et al. (2012), 'beta-cell-specific gene repression: a mechanism to protect against inappropriate or maladjusted insulin secretion?', *Diabetes,* 61 (5), 969-75.
Schwanstecher, C., Meyer, U., and Schwanstecher, M. (2002), 'K(IR)6.2 polymorphism predisposes to type 2 diabetes by inducing overactivity of pancreatic beta-cell ATP-sensitive K(+) channels', *Diabetes,* 51 (3), 875-9.
Scott, R. A., et al. (2012), 'Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways', *Nat Genet,* 44 (9), 991-1005.
Segerstolpe, A., et al. (2016), 'Single-Cell Transcriptome Profiling of Human Pancreatic Islets in Health and Type 2 Diabetes', *Cell Metab,* 24 (4), 593-607.
Seydewitz, H. H. and Matern, D. (2000), 'Molecular genetic analysis of 40 patients with glycogen storage disease type Ia: 100% mutation detection rate and 5 novel mutations', *Hum Mutat,* 15 (1), 115-6.
Shameli, A., et al. (2007), 'Endoplasmic reticulum stress caused by overexpression of islet-specific glucose-6-phosphatase catalytic subunit-related protein in pancreatic Beta-cells', *Rev Diabet Stud,* 4 (1), 25-32.
Shieh, J. J., et al. (2003), 'A glucose-6-phosphate hydrolase, widely expressed outside the liver, can explain age-dependent resolution of hypoglycemia in glycogen storage disease type Ia', *J Biol Chem,* 278 (47), 47098-103.
Shieh, J. J., et al. (2004), 'The islet-specific glucose-6-phosphatase-related protein, implicated in diabetes, is a glycoprotein embedded in the endoplasmic reticulum membrane', *FEBS Lett,* 562 (1-3), 160-4.
Shieh, J. J., et al. (2002), 'The molecular basis of glycogen storage disease type 1a: structure and function analysis of mutations in glucose-6-phosphatase', *J Biol Chem,* 277 (7), 5047-53.
Shungin, D., et al. (2015), 'New genetic loci link adipose and insulin biology to body fat distribution', *Nature,* 518 (7538), 187-96.
Sillitoe, I., et al. (2015), 'CATH: comprehensive structural and functional annotations for genome sequences', *Nucleic Acids Res,* 43 (Database issue), D376-81.
Singh, G. M., et al. (2013), 'The age-specific quantitative effects of metabolic risk factors on cardiovascular diseases and diabetes: a pooled analysis', *PLoS One,* 8 (7), e65174.
Small, K. S., et al. (2011), 'Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes', *Nat Genet,* 43 (6), 561-4.
Soranzo, N. (2011), 'Genetic determinants of variability in glycated hemoglobin (HbA(1c)) in humans: review of recent progress and prospects for use in diabetes care', *Curr Diab Rep,* 11 (6), 562-9.
Soranzo, N., et al. (2010), 'Common variants at 10 genomic loci influence hemoglobin A(1)(C) levels via glycemic and nonglycemic pathways', *Diabetes,* 59 (12), 3229-39.
Soty, M., et al. (2012), 'New insights into the organisation and intracellular localisation of the two subunits of glucose-6-phosphatase', *Biochimie,* 94 (3), 695-703.

Southern, J. A., et al. (1991), 'Identification of an epitope on the P and V proteins of simian virus 5 that distinguishes between two isolates with different biological characteristics', *J Gen Virol,* 72 (Pt 7), 1551-7.
Sparso, T., et al. (2008), 'The GCKR rs780094 polymorphism is associated with elevated fasting serum triacylglycerol, reduced fasting and OGTT-related insulinaemia, and reduced risk of type 2 diabetes', *Diabetologia,* 51 (1), 70-5.
St-Denis, J. F., et al. (1994), 'A conformational model for the human liver microsomal glucose-6-phosphatase system: evidence from rapid kinetics and defects in glycogen storage disease type 1', *J Clin Endocrinol Metab,* 79 (4), 955-9.
Stancakova, A. and Laakso, M. (2016), 'Genetics of Type 2 Diabetes', *Endocr Dev,* 31, 203-20.
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 (3), 294-8.
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 (5), 443-55.
Stoy, J., et al. (2007), 'Insulin gene mutations as a cause of permanent neonatal diabetes', *Proc Natl Acad Sci U S A,* 104 (38), 15040-4.
Strautnieks, S. S., et al. (1998), 'A gene encoding a liver-specific ABC transporter is mutated in progressive familial intrahepatic cholestasis', *Nat Genet,* 20 (3), 233-8.
Strawbridge, R. J., et al. (2011), 'Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes', *Diabetes,* 60 (10), 2624-34.
Streeper, R. S., et al. (1997), 'A multicomponent insulin response sequence mediates a strong repression of mouse glucose-6-phosphatase gene transcription by insulin', *J Biol Chem,* 272 (18), 11698-701.
Strom, T. M., et al. (1998), 'Diabetes insipidus, diabetes mellitus, optic atrophy and deafness (DIDMOAD) caused by mutations in a novel gene (wolframin) coding for a predicted transmembrane protein', *Hum Mol Genet,* 7 (13), 2021-8.
Stukey, J. and Carman, G. M. (1997), 'Identification of a novel phosphatase sequence motif', *Protein Sci,* 6 (2), 469-72.
Stumvoll, M., Goldstein, B. J., and van Haeften, T. W. (2005), 'Type 2 diabetes: principles of pathogenesis and therapy', *Lancet,* 365 (9467), 1333-46.
Stumvoll, M., et al. (2003), 'Glucose allostasis', *Diabetes,* 52 (4), 903-9.
Sun, M. S., et al. (2002), 'Sustained hepatic and renal glucose-6-phosphatase expression corrects glycogen storage disease type Ia in mice', *Hum Mol Genet,* 11 (18), 2155-64.
Sweet, I. R., et al. (1997), 'Measurement and modeling of glucose-6-phosphatase in pancreatic islets', *Am J Physiol,* 272 (4 Pt 1), E696-711.
Szabat, M., et al. (2016), 'Reduced Insulin Production Relieves Endoplasmic Reticulum Stress and Induces beta Cell Proliferation', *Cell Metab,* 23 (1), 179-93.
Takahashi, K., et al. (2000), 'Heterogeneous mutations in the glucose-6-phosphatase gene in Japanese patients with glycogen storage disease type Ia', *Am J Med Genet,* 92 (2), 90-4.
Takei, D., et al. (2006), 'WFS1 protein modulates the free Ca(2+) concentration in the endoplasmic reticulum', *FEBS Lett,* 580 (24), 5635-40.
Takeuchi, F., et al. (2010), 'Common variants at the GCK, GCKR, G6PC2-ABCB11 and MTNR1B loci are associated with fasting glucose in two Asian populations', *Diabetologia,* 53 (2), 299-308.
Tam, C. H., et al. (2010), 'Common polymorphisms in MTNR1B, G6PC2 and GCK are associated with increased fasting plasma glucose and impaired beta-cell function in Chinese subjects', *PLoS One,* 5 (7), e11428.
Tan, G. D., et al. (2006), 'The in vivo effects of the Pro12Ala PPARgamma2 polymorphism on adipose tissue NEFA metabolism: the first use of the Oxford Biobank', *Diabetologia,* 49 (1), 158-68.
Tappy, L. (1995), 'Regulation of hepatic glucose production in healthy subjects and patients with non-insulin-dependent diabetes mellitus', *Diabete Metab,* 21 (4), 233-40.

Tarasov, A. I., et al. (2008), 'A rare mutation in ABCC8/SUR1 leading to altered ATP-sensitive K+ channel activity and beta-cell glucose sensing is associated with type 2 diabetes in adults', *Diabetes,* 57 (6), 1595-604.
Taussky, H. H. and Shorr, E. (1953), 'A microcolorimetric method for the determination of inorganic phosphorus', *J Biol Chem,* 202 (2), 675-85.
Tekirian, T. L. (2002), 'The central role of the trans-Golgi network as a gateway of the early secretory pathway: physiologic vs nonphysiologic protein transit', *Exp Cell Res,* 281 (1), 9-18.
Thomsen, S. K., McCarthy, M. I., and Gloyn, A. L. (2016a), 'The Importance of Context: Uncovering Species- and Tissue-Specific Effects of Genetic Risk Variants for Type 2 Diabetes', *Front Endocrinol (Lausanne),* 7, 112.
Thomsen, S. K., et al. (2016b), 'Systematic Functional Characterization of Candidate Causal Genes for Type 2 Diabetes Risk Variants', *Diabetes*.
Thormaehlen, A. S., et al. (2015), 'Systematic cell-based phenotyping of missense alleles empowers rare variant association studies: a case for LDLR and myocardial infarction', *PLoS Genet,* 11 (2), e1004855.
Trioche, P., et al. (2000), 'Genetic heterogeneity of glycogen storage disease type Ia in France: a study of 48 patients', *Hum Mutat,* 16 (5), 444.
Turban, S., et al. (2012), 'Optimal elevation of beta-cell 11beta-hydroxysteroid dehydrogenase type 1 is a compensatory mechanism that prevents high-fat diet-induced beta-cell failure', *Diabetes,* 61 (3), 642-52.
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 (1), e55272.
van de Bunt, M., et al. (2015), 'Transcript Expression Data from Human Islets Links Regulatory Signals from Genome-Wide Association Studies for Type 2 Diabetes and Glycemic Traits to Their Downstream Effectors', *PLoS Genet,* 11 (12), e1005694.
van Raalte, D. H., Ouwens, D. M., and Diamant, M. (2009), 'Novel insights into glucocorticoid-mediated diabetogenic effects: towards expansion of therapeutic options?', *Eur J Clin Invest,* 39 (2), 81-93.
van Schaftingen, E. and Gerin, I. (2002), 'The glucose-6-phosphatase system', *Biochem J,* 362 (Pt 3), 513-32.
Vander Kooi, B. T., et al. (2003), 'The three insulin response sequences in the glucose-6-phosphatase catalytic subunit gene promoter are functionally distinct', *J Biol Chem,* 278 (14), 11782-93.
Vander Kooi, B. T., et al. (2005), 'The glucose-6-phosphatase catalytic subunit gene promoter contains both positive and negative glucocorticoid response elements', *Mol Endocrinol,* 19 (12), 3001-22.
Vaxillaire, M., Bonnefond, A., and Froguel, P. (2012), 'The lessons of early-onset monogenic diabetes for the understanding of diabetes pathogenesis', *Best Pract Res Clin Endocrinol Metab,* 26 (2), 171-87.
Villar-Palasi, C. and Guinovart, J. J. (1997), 'The role of glucose 6-phosphate in the control of glycogen synthase', *FASEB J,* 11 (7), 544-58.
Visser, G., et al. (2000), 'Neutropenia, neutrophil dysfunction, and inflammatory bowel disease in glycogen storage disease type Ib: results of the European Study on Glycogen Storage Disease type I', *J Pediatr,* 137 (2), 187-91.
Vlcek, S., Foisner, R., and Wilson, K. L. (2004), 'Lco1 is a novel widely expressed lamin-binding protein in the nuclear interior', *Exp Cell Res,* 298 (2), 499-511.
Wall, M. L., et al. (2015), 'Novel stable isotope analyses demonstrate significant rates of glucose cycling in mouse pancreatic islets', *Diabetes,* 64 (6), 2129-37.
Wang, H., et al. (2013), 'Large scale meta-analyses of fasting plasma glucose raising variants in GCK, GCKR, MTNR1B and G6PC2 and their impacts on type 2 diabetes mellitus risk', *PLoS One,* 8 (6), e67665.

Wang, J., et al. (1999), 'A mutation in the insulin 2 gene induces diabetes with severe pancreatic beta-cell dysfunction in the Mody mouse', *J Clin Invest,* 103 (1), 27-37.
Wang, M. and Kaufman, R. J. (2016), 'Protein misfolding in the endoplasmic reticulum as a conduit to human disease', *Nature,* 529 (7586), 326-35.
Wang, S., et al. (1997), 'Isolation and characterization of a cell line from the epithelial cells of the human fetal pancreas', *Cell Transplant,* 6 (1), 59-67.
Wang, T., et al. (2014), 'Genetic screens in human cells using the CRISPR-Cas9 system', *Science,* 343 (6166), 80-4.
Wang, Y., et al. (2000), 'Activation of ATF6 and an ATF6 DNA binding site by the endoplasmic reticulum stress response', *J Biol Chem,* 275 (35), 27013-20.
Wang, Y., et al. (2007), 'Deletion of the gene encoding the islet-specific glucose-6-phosphatase catalytic subunit-related protein autoantigen results in a mild metabolic phenotype', *Diabetologia,* 50 (4), 774-8.
Weir, G. C. and Bonner-Weir, S. (2011), 'Finally! A human pancreatic beta cell line', *J Clin Invest,* 121 (9), 3395-7.
Wessel, J., et al. (2015), 'Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility', *Nat Commun,* 6, 5897.
Weyer, C., Bogardus, C., and Pratley, R. E. (1999), 'Metabolic characteristics of individuals with impaired fasting glucose and/or impaired glucose tolerance', *Diabetes,* 48 (11), 2197-203.
WHO (2016) *Global report on diabetes* [online text], World Health Organization, Geneva, Switzerland <http://www.who.int/diabetes/publications/grd-2016/en/>
Wickner, S., Maurizi, M. R., and Gottesman, S. (1999), 'Posttranslational quality control: folding, refolding, and degrading proteins', *Science,* 286 (5446), 1888-93.
Wolf, B. A., et al. (1986), 'Glucose 6-phosphate regulates Ca2+ steady state in endoplasmic reticulum of islets. A possible link in glucose-induced insulin secretion', *J Biol Chem,* 261 (35), 16284-7.
Wolf, B. A., et al. (1988), 'Regulation of Ca2+ homeostasis by islet endoplasmic reticulum and its role in insulin secretion', *Am J Physiol,* 254 (2 Pt 1), E121-36.
Wood, A. R., et al. (2014), 'Defining the role of common variation in the genomic and biological architecture of adult human height', *Nat Genet,* 46 (11), 1173-86.
Worley, J. F., 3rd, et al. (1994), 'Endoplasmic reticulum calcium store regulates membrane potential in mouse islet beta-cells', *J Biol Chem,* 269 (20), 14359-62.
Wu, M. C., et al. (2011), 'Rare-variant association testing for sequencing data with the sequence kernel association test', *Am J Hum Genet,* 89 (1), 82-93.
Yamada, T., et al. (2006), 'WFS1-deficiency increases endoplasmic reticulum stress, impairs cell cycle progression and triggers the apoptotic pathway specifically in pancreatic beta-cells', *Hum Mol Genet,* 15 (10), 1600-9.
Yamamoto, K., et al. (2004), 'Differential contributions of ATF6 and XBP1 to the activation of endoplasmic reticulum stress-responsive cis-acting elements ERSE, UPRE and ERSE-II', *J Biochem,* 136 (3), 343-50.
Yang, J., et al. (2006), 'Islet-specific glucose-6-phosphatase catalytic subunit-related protein-reactive CD4+ T cells in human subjects', *J Immunol,* 176 (5), 2781-9.
Yiu, W. H., et al. (2010), 'Complete normalization of hepatic G6PC deficiency in murine glycogen storage disease type Ia using gene therapy', *Mol Ther,* 18 (6), 1076-84.
Zingone, A., et al. (2000), 'Correction of glycogen storage disease type 1a in a mouse model by gene therapy', *J Biol Chem,* 275 (2), 828-32.
Zuk, O., et al. (2014), 'Searching for missing heritability: designing rare variant association studies', *Proc Natl Acad Sci U S A,* 111 (4), E455-64.

Appendices

Curr Diab Rep
DOI 10.1007/s11892-013-0429-1

GENETICS (TM FRAYLING, SECTION EDITOR)

Bridging the Gap Between Genetic Associations and Molecular Mechanisms for Type 2 Diabetes

Hui Jin Ng · Anna L. Gloyn

Abstract Type 2 diabetes is a global pandemic for which there is currently no disease-modifying treatment. New and targeted therapeutics are greatly needed, but progress in identifying novel targets for therapeutic intervention is severely hampered by poor understanding of disease pathogenesis. Over the past 6 years, the success of genome-wide association studies has led to an unprecedented increase in the number of loci robustly associating with type 2 diabetes risk. Each of these signals offers the opportunity to uncover biological insights into disease pathogenesis, which, if harnessed effectively, hold the promise to deliver new pathways for therapeutic intervention, strategies for patient stratification, and potentially, biomarkers for identifying those at greatest risk of developing diabetes. We review the progress that has been made and the approaches being adopted and discuss the inherent challenges in moving from association signals, which largely map to poorly annotated sequence, to transcripts, mechanisms, and disease biology.

Keywords Type 2 diabetes · Genome-wide association studies · Association signal · Molecular mechanisms · Beta cell · Pancreatic islet · Functional characterization · Targeted sequencing · eQTL · miRNA · DNA methylation · Knockout mice · Human model

Introduction

Human genetics offers a powerful tool for gaining insight into disease pathophysiology and the opportunity to uncover relevant biological pathways that are ripe for therapeutic intervention. Until relatively recently, our knowledge of the genetic landscape of type 2 diabetes (T2D) has been sparse, with only a handful of genetic loci showing robust association with disease. The advent of genome-wide association studies (GWAS) has revolutionized the field, taking the number of loci associated with T2D risk to approximately 70 and providing numerous opportunities to uncover biological insights into this disease [1]. Currently, one of the biggest challenges is translating these association signals into molecular mechanisms that can fuel target discovery, prioritization, and patient stratification.

Despite the rapid identification of signals for T2D, the pace at which they have been translated into disease mechanisms has been noticeably slower. This is in part due to the vast majority of variants mapping to poorly annotated sequence and, in many cases, a lack of certainty of the exact transcripts through which the variants mediate diabetes risk. Taken together with the very modest effect sizes (odds ratios 1.1–1.4), which can make it difficult to detect differences between risk and nonrisk alleles in functional assays, many variants have been unattractive candidates to pursue in the laboratory. Fortunately, advances in sequencing technology and sequence annotation are fueling ongoing efforts to provide detailed transcriptional maps for disease-relevant human tissues, such as pancreatic islets. In combination with publicly available data from the ENCODE project [2], we are now in a better position to transit from association signal to transcript and, consequently, to cellular and molecular mechanisms. This review will highlight examples of the approaches that have been adopted to capitalize on our increased knowledge of the genetic basis of T2D, which are leading to a greater mechanistic understanding of diabetes pathogenesis (Fig. 1).

H. J. Ng · A. L. Gloyn (✉)
Oxford Centre for Diabetes, Endocrinology & Metabolism, Churchill Hospital, University of Oxford, Oxford OX3 7LE, UK
e-mail: anna.gloyn@drl.ox.ac.uk

H. J. Ng
e-mail: natasha.ng@ocdem.ox.ac.uk

Published online: 16 October 2013

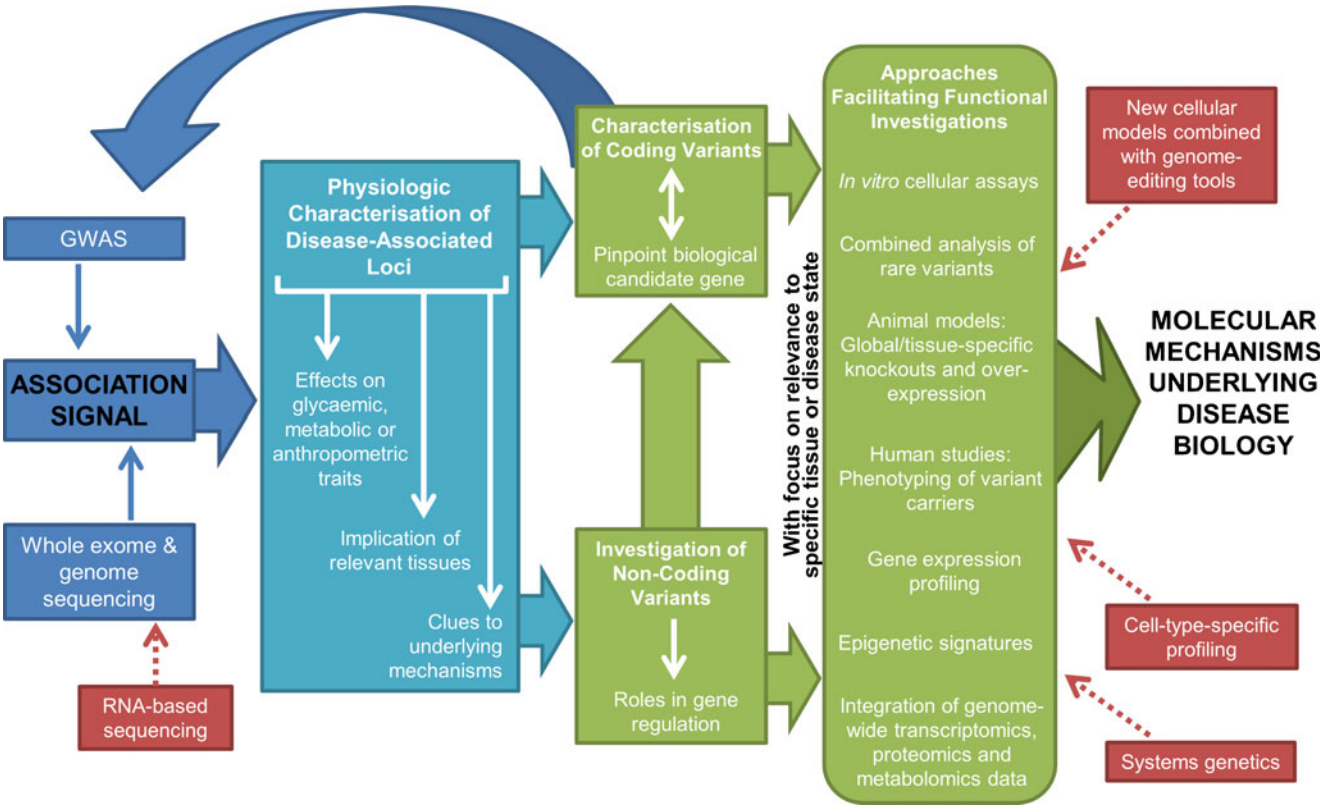


Fig. 1 Schematic representation illustrating potential research pipeline(s) for the transition from disease association signal to molecular mechanism for disease

Approaches to Uncovering Disease Biology

The ability to investigate the influence of T2D-associated variants on glycemic, metabolic, and anthropometric traits through existing meta-analyses has shed much needed light on the relevant tissues and likely mechanisms through which variants mediate diabetes risk [1, 3–7]. Understanding their effect on measures of insulin processing, secretion, and sensitivity can help pinpoint more precisely their role in the regulation of glucose control and insulin secretion and/or action [8•]. As a result, it has become clear that many T2D risk variants influence insulin secretion and, therefore, presumably islet cell function, placing the beta cell at center stage [5]. The ability to combine such information provides a strong steer toward appropriate cell biological investigations. For example, a number of variants (*ARAP1*, *TCF7L2*, *SLC30A8*, *VSP13C/C2CD4A/B*) are associated not only with T2D risk, but also with circulating proinsulin levels [7]. The proinsulin-raising alleles at a number of these loci (*TCF7L2*, *SLC30A8*, *and VPS13C/C2CD4A/B*) result in reduced beta cell function, suggesting that they cause defects in insulin processing and secretion distal to the first enzymatic step. In contrast, the T2D risk allele of *ARAP1* is associated with reduced proinsulin levels, suggesting a mechanism that results in generalized down-regulation of insulin secretion due to reduced functional beta cell mass [7].

Characterization of Coding Variants in Genes of Known Function

Refinement of an association signal to a particular transcript and identification of causal variants and disease mechanisms can be aided by the study of nonsynonymous coding variants, especially those within biologically plausible candidate genes, which are predicted to be deleterious on the basis of their likely impact on protein function. These opportunities provide a clear pathway for experimental investigation, especially in laboratories with domain expertise.

A well-studied example is the common P446L variant in *GCKR*, which has been reproducibly associated with T2D risk, fasting glucose, triglyceride levels, and other metabolically relevant quantitative traits [4, 9]. *GCKR* encodes the glucokinase regulatory protein (GKRP), a pivotal regulator of glucokinase (GCK) activity in the liver and hence, hepatic glucose uptake. The strong support for a role of GKRP in T2D pathogenesis was previously validated in functional studies that showed that P446L-GKRP had decreased intrinsic regulatory activity in *in vitro* kinetic assays [10]. Recently, Rees et al. carried out further analysis of P446L-GKRP, using fluorescence microscopy and fluorescence resonance energy transfer techniques, and found that the variant protein displayed decreased interaction with GCK and ability to sequester GCK into the nucleus [11•]. Reduced binding and

inhibition of GCK could, consequently, influence hepatic glucose metabolism. The data offer a plausible mechanism by which the P446L variant can simultaneously reduce fasting glucose levels, providing protection against T2D risk, as well as result in an inverse effect on triglyceride levels. This has implications for GCK activation as a potential target for T2D treatment [12].

Another gene that harbors a nonsynonymous variant (W325R) associated with T2D risk is *SLC30A8* [13]. It encodes ZnT8, a zinc transporter almost exclusively expressed in the pancreas and localized to beta cell insulin secretory granules. A few groups have studied the role of ZnT8 in rodent and cellular models, demonstrating that ZnT8 is required for the crystallization of insulin hexamers prior to secretion [14–17]. However, these studies were unable to demonstrate congruent effects on insulin secretion and glucose homeostasis, with a recent study concluding that loss of *Slc30A8* in mice has no effect on rodent physiology [18]. One study has investigated the physiological consequences of ZnT8 deletion in *Slc30a8* null mice, as well as the impacts of the *SLC30A8* W325R variant on protein function in rodent beta cell lines [16]. Cellular interrogation of R325-ZnT8 suggested that the T2D risk allele results in reduced zinc transport activity, which the authors proposed provides a potential mutational mechanism for the association with increased risk for T2D [16]. Ultimately, the study of glucose tolerance in humans harboring loss-of-function alleles in *SLC30A8,* assuming these alleles exist in nature, is required to settle the debate about whether a loss of this transporter increases the risk of developing glucose intolerance and diabetes.

Even with nonsynonymous variants in obvious candidate genes, it can be challenging to decipher precise mechanisms. This is illustrated by the lack of reproducibility seen in a number of functional studies involving the common E23K variant in *KCNJ11*, which shows robust evidence for association with T2D risk [19–22]. It is now clear that the reason for this could be attributed to a second coding variant (S1369A) in the adjacent gene *ABCC8*, which is in near perfect linkage disequilibrium with E23K [23]. The relevance of both variants to disease association is clearer when one considers that *KCNJ11* and *ABCC8*, respectively, encode the Kir6.2 and sulfonylurea receptor 1 (SUR1) protein subunits of the ATP-sensitive K^+ (K_{ATP}) channel, which is responsible for coupling membrane excitability and insulin secretion in the pancreatic beta cell. A recent study [24] examined the combined effect of both T2D risk alleles on K_{ATP} channel function and demonstrated that the K23/A1369 haplotype displayed decreased sensitivity to ATP inhibition, due to a direct effect of A1369 on K_{ATP} channel activity. It is now thought to be likely that both the E23K and the S1369A variants could be contributing to the T2D association at this locus [22, 24].

In contrast to common variation, low-frequency and rare variation is not well-captured by GWAS, but there are compelling arguments that it plays an important role in susceptibility to complex diseases, including T2D [25]. Advances in sequencing technologies enable many targeted, whole-genome and exome sequencing studies, but the interpretation of rare coding variants identified through these efforts is proving a major challenge. Two recent studies [26••, 27••] performed detailed functional studies to assign pathogenicity to rare nonsynonymous variants and then investigated their association with metabolic phenotypes. One study involved targeted sequencing of *GCKR* in individuals from the ClinSeq study, who were recruited on the basis of a high risk of coronary artery disease. Nineteen nonsynonymous variants were identified and analyzed for cellular localization, interaction with glucokinase, and impact on kinetic activity [26••]. All except the P446L variant had minor allele frequencies less than 2 %, and interestingly, the majority (79 %) were found to affect protein function. Critically, when results of the *in vitro* functional characterization were taken into account, there was an improvement in the phenotypic associations with both triglyceride and cholesterol levels [26••].

Another gene for which investigation of rare genetic variation has reaffirmed a functional link with T2D risk is *MTNR1B*. It encodes melatonin receptor 1B, a G protein-coupled receptor. Targeted resequencing of *MTNR1B* identified 40 nonsynonymous coding variants, the majority of which are again very rare [27••]. The group individually evaluated all variants for melatonin binding and signaling capability. Thirteen variants that display loss of function were strongly associated with increased T2D risk, while the other seemingly neutral variants were not. Again, this demonstrates the value of *in vitro* studies in accurately assigning pathogenicity and increasing the statistical power to detect associations [27••].

Using Gene Expression Data Sets to Identify Causal Transcripts for Noncoding Variants

For the majority of disease association signals that lie within non-protein-coding regions, transition from signal to causal transcript is not as straightforward. One way to approach this is to determine whether the risk allele alters expression levels of neighboring transcripts. Even though expression data sets are widely available from immortalized lymphoblastoid cell lines, data sets generated from disease-relevant tissues, the choice of which can be guided by the physiological trait implicated by the genetic association, are still the most informative for this approach. Genetic variants that are associated with changes in transcript levels are known as expression quantitative trait loci (eQTL), which can be present in either *cis* (on the same chromosome) or *trans* (anywhere in the genome).

One example in which the use of tissue expression data sets not only identified the likely causal transcript, but also uncovered a transcription factor network critical for adipose tissue function is the 7q32 locus, which is one of only two T2D susceptibility loci showing parental-origin-specific association [28]. This signal includes Kruppel-like factor 14 (*KLF14*), which encodes a maternally expressed transcription factor. Single-nucleotide polymorphisms (SNPs) near *KLF14* show strong maternal-specific association with *KLF14* expression levels in addition to robust association with T2D susceptibility [5, 28]. The eQTL was found to act as a master *trans* regulator of adipose gene expression, including many genes that are strongly associated with a variety of metabolic phenotypes [29•].

A further example of the value of expression data sets in refining association signals is at the *IRS1* locus, which is a T2D risk locus located upstream of the insulin receptor substrate 1 gene (*IRS1*) [30]. The risk allele, which is implicated in insulin resistance and hyperinsulinemia, was found to associate with lower basal levels of IRS1 protein and decreased IRS1-mediated phosphatidylinositol-3-OH kinase (PI3K) activity in skeletal muscle [30]. An additional eQTL near *IRS1* that is found to significantly associate with lower body fat percentage is also associated with reduced *IRS1* expression specifically in adipose tissue [31]. These findings convincingly establish the regulatory role of eQTLs near *IRS1* on *IRS1* expression in relevant tissues and demonstrate that *IRS1* is a contributor to the observed metabolic and diabetes risk traits.

Clues to Molecular Mechanisms for Noncoding Variants from Improved Sequence Annotation in Disease-Relevant Tissue

The ENCODE project has provided a wealth of annotation across multiple tissues, including the human pancreas. However, from an endocrine perspective, the critical components of the pancreas are the islets of Langerhans, which constitute only 1 %–2 % of total pancreatic mass. The majority of this tissue consists of cells involved in exocrine functions. Next-generation sequencing has been instrumental to laboratories with an interest in T2D pathogenesis in improving sequence annotation for the islet component of the tissue. Using an established technique called formaldehyde-assisted isolation of regulatory elements coupled with next-generation sequencing (FAIRE-seq), human islets were found to harbor clusters of open regulatory elements, sites that provide indications of regulatory activity that could be important to islet biology and disease [32••]. The importance of these regions was illustrated by the finding that an intronic variant (rs7903146) in *TCF7L2*, which is strongly associated with T2D risk, maps to an islet-selective open chromatin region and is likely to influence local chromatin structure and transcriptional regulation [32••]. Furthermore, the same data annotation was used to prioritize SNPs at the T2D-associated *JAZF1* locus for functional investigation in transcriptional activity assays on the basis of their localization to islet open chromatin regions. The authors found allelic differences in enhancer activity at rs1635852 in intron 1 and determined that the variant is part of a *cis*-regulatory transcriptional complex [33].

Improvements in islet cell annotation have not been limited to regions of open chromatin. A comprehensive genome-wide analysis of a suite of epigenetic markers, including DNase hypersensitivity, histone modification, and CCCTC factor binding, has been performed to generate a global snapshot of the epigenome in human islets [34••]. From a subset of GWAS-derived risk loci for T2D, 118 putative distal regulatory elements were identified, 6 of which harbored T2D-associated variants. Detailed functional analysis revealed that the lead SNP at the *TCF7L2* locus and a set of SNPs in a haplotype in *WFS1* exhibit allele-specific differences in enhancer activity, suggesting that alteration of islet gene expression contributes to the molecular mechanism underlying disease susceptibility at these loci [34••].

Another class of noncoding elements that has recently come into the limelight are long noncoding RNAs (lncRNAs). A number of lncRNAs, which are noncoding transcripts longer than 200 nucleotides, are known to have diverse cellular roles involving regulation of gene expression. However, most of them have an unknown function, and so their role in islets is poorly understood. Recently, a comprehensive transcriptome map of human islets and beta cells was reported after the integration of sequence-based transcriptional and chromatin data uncovered more than 1,100 islet lncRNA genes [35••]. These lncRNAs exhibit highly cell-type-specific expression, are dynamically regulated in islets, and are involved in beta cell differentiation and maturation. The potential importance of these regulatory elements in T2D pathogenesis was underscored by the fact that two lncRNAs (*KCNQ1OT1* and *HI-LNC45*) were significantly dysregulated in islets from T2D donors [35••]. Furthermore, several lncRNAs mapped to established T2D susceptibility loci, the most prominent of which were near *PROX1* and *WFS1*. These findings have uncovered a new class of genes that are likely to impact beta cell function and serve as candidates for dissecting the influence of noncoding variants on T2D risk.

At the other end of the spectrum of regulatory RNAs sit microRNAs (miRNAs), which are involved in posttranscriptional regulation by means of mRNA destabilization and translational repression. These short noncoding RNAs (~22 nucleotides in length) are implicated in a wide variety of diseases, including cancer [36], and have been found to be linked to glucose metabolism [37, 38]. Recently, the miRNA profile of human islets and enriched beta cell populations was described, providing evidence that islet-expressed miRNAs might be involved in T2D pathogenesis [39•]. The predicted

target sites of islet miRNAs were enriched for T2D association signals, and the strongest overlap detected was with a potentially causal variant (rs3802177) in the 3' UTR of the *SLC30A8* gene [39•]. While more in-depth studies are certainly required to completely evaluate the function of these miRNAs and their targets, this catalogue is a valuable resource for studying the role of miRNAs in islet biology and T2D predisposition.

The Impact of Noncoding Variants on DNA Methylation and Gene Expression

Another potential pathway by which noncoding variants could impact on diabetes risk is through epigenetic mechanisms impacting on DNA methylation and, subsequently, cellular phenotypes. For example, DNA methylation can be altered by SNPs that result in the introduction or removal of cytosine-phosphate-guanine (CpG) dinucleotides. As with expression studies, the choice of tissue is paramount for DNA methylation studies, and where possible, the use of disease-specific tissues is highly desirable.

Two recent studies have capitalized on relatively large human islet DNA and RNA bioresources to study the impact of selected T2D-associated variants on DNA methylation and gene expression [40, 41]. The first examined the impact of T2D-associated variants at the *KCNQ1* locus on regional DNA methylation and gene expression in adult islets and fetal pancreas [40]. The *KCNQ1* locus harbors several independent association signals likely to act through impaired islet function [5, 42, 43]. Consistent with the location of *KCNQ1* in an imprinted genomic region, association of alleles at two of the signals confers diabetes risk only when maternally inherited [28]. T2D risk genotype status at rs2237895 representing the intron 15 signal was shown by bisulphite sequencing to impact on regional DNA methylation in a developmentally flexible manner, but interestingly, these changes in methylation did not translate into changes in regional gene expression [40]. A lack of concordance between DNA methylation changes and gene expression has recently been reported in a large study of multiple cell types, suggesting that genetically driven changes in DNA methylation alone are unlikely to result in allele-specific gene expression [44].

The second study examined a panel of 40 variants associated with T2D risk, 19 of which mapped to CpG dinucleotide regions, and analyzed human islet DNA methylation levels according to genotype. Sixteen of the 19 CpG-SNPs were associated with differential DNA methylation of the CpG-SNP site in the islet samples, but again consistent with recent findings, only 2 were associated with changes in gene expression levels within 500-kb proximity of the SNP [41, 44]. Nonetheless, there was evidence that several of these CpG-SNPs led to altered RNA splicing events [41].

Using Rodent Models to Study Gene Function

A powerful and established approach to investigating gene function is to capitalize on the extensive tools available to manipulate the mouse genome and the ability to breed mice of a specific genotype on a homogeneous genetic background to perform both *in vivo* and *in vitro* phenotypic investigations. The International Knock-out Mouse Consortium (IKMC) systematically generates null mutations in mouse embryonic stem (ES) cells for every protein-coding gene in the mouse genome to generate global null alleles that can subsequently be converted into conditional alleles to give rise to global, temporal, and tissue-specific knockout mouse models [45]. The recently established International Mouse Phenotyping Consortium (IMPC) was funded to produce mouse lines from each targeted ES cell and to carry out basic phenotypic characterization of these mice. This resource will be enormously helpful in assigning function to genes and will provide a basis for follow-up investigations of genetic association signals.

Numerous groups have already used rodent models to tackle the function of genes mapping to GWAS signals for T2D (*SLC30A8, FTO, TCF7L2, CDKAL1*), and the likely molecular mechanisms for the causal genes are well-summarized in a recent review [46]. Two recent studies deserve specific mention, due to the mechanistic insights they have provided.

CDKAL1 was one of the first genes implicated in T2D risk through GWAS, and studies in humans support a defect in beta cell function [47–49]. However, the protein encoded by the gene, Cdk5 regulatory associated protein 1-like 1 (CDKAL1), was of relatively unknown function at that time, especially in the context of the pancreatic islets. Elegant studies on global and beta-cell-specific *Cdkal1* knockout mice have since shed much light on the role of Cdkal1 in islets and revealed that its deficiency gave rise to islet hypertrophy, reduced insulin secretion, and poor glucose tolerance [50••]. The molecular mechanism driving this phenotype is fascinating; the authors were able to demonstrate that Cdkal1 is a mammalian methylthiotransferase required for the accurate translation of lysine codons into proteins. Mice deficient in beta cell Cdkal1 incorporate fewer lysine residues correctly into their proinsulin molecules, resulting in reduced glucose-stimulated proinsulin synthesis. Moreover, this beta cell dysfunction is likely to be exacerbated by ER stress, presumably due to protein misfolding resulting from unfaithful protein translation in Cdkal1-deficient islets [50••].

One study in which the use of rodent models has presented unexpected findings involves the *TCF7L2* gene. Intronic variants in *TCF7L2* have the largest impact on T2D risk described to date, with a 30 % increase per risk allele [51]. *TCF7L2* encodes the transcription factor TCF4, which is known to regulate Wnt signaling, but its precise role in metabolic tissues remains unclear. The evidence thus far from

studies on human physiology, human islet culture, and rodent models has consistently shown that *TCF7L2* T2D risk alleles influence diabetes risk through aberrant beta cell function [52, 53, 54•]. However, a recent study by Boj and colleagues painted a different and more complicated picture. In their study, inducible beta-cell-specific knockout of *Tcf7l2* in mice unexpectedly had no impact on glucose-stimulated insulin secretion, whereas global or liver-specific inducible models of *Tcf7l2* deletion resulted in hypoglycaemia and reduced hepatic glucose production [55••]. In accordance with this observation, liver-specific overexpression of *Tcf7l2* leads to enhanced hepatic glucose production and glucose intolerance. Discrepancies in findings were reported in other studies, whereby pancreas-specific *Tcf7l2* null mice exhibited an islet phenotype characterized by reduced insulin secretion, blunted incretin response, and impaired beta cell expansion, which recapitulates the phenotype observed in human carriers of the T2D risk allele [54•]. The divergence in phenotypic effects could result from differences in the experimental models, since pancreas-specific knockouts clearly have more extensive disruption of transcription factor function across an array of islet cell types. The studies ultimately suggest that a complex interplay between metabolic tissues exists that may be crucial to the understanding of T2D.

Capitalizing on Human Models to Understand the Role of Disease-Related Genes

The study of individuals with rare and fully penetrant mutations is the human equivalent to the use of knockout mice. It is well-recognized that the study of monogenic forms of diabetes has contributed enormously to our understanding of the development and function of the human pancreatic beta cell with obvious direct translational relevance [56]. Very recently, the power of this approach was demonstrated by the detailed metabolic and anthropometric characterization of individuals with a rare cancer predisposition due to loss-of-function mutations in the tumor-suppressor phosphatase and tensin homologue (PTEN). Detailed physiological investigations in mutation carriers and well-matched healthy controls revealed for the first time that PTEN haploinsufficiency causes insulin sensitisation and increased risk of obesity in humans, while common genetic variation in the gene influences fasting insulin levels in the healthy population [57, 58]. An investigation of insulin signaling in adipose tissue from mutation carriers and matched controls demonstrated that the enhanced insulin sensitivity could be explained by increased insulin signaling through the PI3K–AKT pathway [57]. Although a series of studies in mice had shown metabolic effects of PTEN loss, there are important differences in the phenotypic consequences of PTEN loss in mice and humans that underpin the translational advantage of the study of human model systems.

Conclusions and Looking Ahead

As we move from the GWAS era into one largely focused on sequence-based discovery, the expectation is that coding variants of larger effect size will emerge as disease association signals, negating many of the obstacles currently faced for experimental follow-up. The research community will continue to benefit from high-throughput genomic methodologies such as RNA sequencing, which has most recently made the first characterization of the human beta cell transcriptome possible, giving rise to a pure beta-cell-specific expression signature [59••]. In parallel, authentic human pancreatic beta cell lines [60••] and patient-derived induced pluripotent stem cell lines [61, 62], combined with emerging genome-editing technologies such as transcription activator-like effector nucleases (TALENs) [63, 64••] and clustered regularly interspaced short palindromic repeat (CRISPR) associated systems [65••], hold the promise of boosting the current research pipeline (Fig. 1) and providing an improved arsenal of tools to enable researchers to truly bridge the gap between association signals and molecular mechanisms. These improved resources should allow us to fully capitalize on the substantial investments that have been made in the field of human genetics to illuminate the molecular basis of diabetes and formulate improved disease treatment strategies.

Compliance with Ethics Guidelines

Conflict of Interest Hui Jin Ng and Anna L. Gloyn declare that they have no conflict of interest.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Papers of particular interest, published recently, have been highlighted as:

- • Of importance
- •• Of major importance

1. Morris AP, Voight BF, Teslovich TM, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet. 2012;44:981–90.
2. Rosenbloom KR, Sloan CA, Malladi VS, et al. ENCODE data in the UCSC genome browser: year 5 update. Nucleic Acids Res. 2013;41:D56–63.
3. Willer CJ, Speliotes EK, Loos RJF, et al. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. Nat Genet. 2009;41:25–34.
4. Dupuis J, Langenberg C, Prokopenko I, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet. 2010;42:105–16.

5. Voight BF, Scott LJ, Steinthorsdottir V, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet. 2010;42:579–89.
6. Saxena R, Hivert MF, Langenberg C, et al. Genetic variation in *GIPR* influences the glucose and insulin responses to an oral glucose challenge. Nat Genet. 2010;42:142–8.
7. Strawbridge RJ, Dupuis J, Prokopenko I, et al. Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. Diabetes. 2011;60:2624–34.
8. • Ingelsson E, Langenberg C, Hivert MF, et al. Detailed physiologic characterization reveals diverse mechanisms for novel genetic Loci regulating glucose and insulin metabolism in humans. Diabetes. 2010;59:1266–75. *An investigation of the effects of genetic variants associated with T2D risk on measures of insulin processing, secretion, and sensitivity*.
9. Orho-Melander M, Melander O, Guiducci C, et al. 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. 2008;57:3112–21.
10. Beer NL, Tribble ND, McCulloch LJ, et al. 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. 2009;18:4081–8.
11. • Rees MG, Wincovitch S, Schultz J, et al. Cellular characterisation of the *GCKR* P446L variant associated with type 2 diabetes risk. Diabetologia. 2012;55:114–22. *A functional study of the common coding variant P446L in GKRP that revealed the potential cellular mechanism by which the variant influences T2D risk*.
12. Rees MG, Gloyn AL. Small molecular glucokinase activators: has another new anti-diabetic therapeutic lost favour? Br J Pharmacol. 2013;168:335–8.
13. Sladek R, Rocheleau G, Rung J, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature. 2007;445:881–5.
14. Pound LD, Sarkar SA, Benninger RK, et al. Deletion of the mouse *Slc30a8* gene encoding zinc transporter-8 results in impaired insulin secretion. Biochem J. 2009;421:371–6.
15. Lemaire K, Ravier MA, Schraenen A, et al. 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. 2009;106:14872–7.
16. Nicolson TJ, Bellomo EA, Wijesekara N, 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. 2009;58:2070–83.
17. Wijesekara N, Dai FF, Hardy AB, et al. Beta cell-specific Znt8 deletion in mice causes marked defects in insulin processing, crystallisation and secretion. Diabetologia. 2010;53:1656–68.
18. Pound LD, Sarkar SA, Ustione A, et al. The physiological effects of deleting the mouse *Slc30a8* gene encoding zinc transporter-8 are influenced by gender and genetic background. Plos One. 2012;7:e40972.
19. Sakura H, Wat N, Horton V, et al. Sequence variations in the human Kir6.2 gene, a subunit of the beta-cell ATP-sensitive K-channel: no association with NIDDM in white Caucasian subjects or evidence of abnormal function when expressed in vitro. Diabetologia. 1996;39:1233–6.
20. Schwanstecher C, Meyer U, Schwanstecher M. Kir6.2 polymorphism predisposes to type 2 diabetes by inducing overactivity of pancreatic β-cell ATP-Sensitive K^+ channels. Diabetes. 2002;51:875–9.
21. Gloyn AL, Weedon MN, Owen KR, et al. Large-scale association studies of variants in genes encoding the pancreatic beta-cell K-ATP channel subunits Kir6.2 (*KCNJ11*) and SUR1 (*ABCC8*) confirm that the *KCNJ11* E23K variant is associated with type 2 diabetes. Diabetes. 2003;52:568–72.
22. Riedel MJ, Boora P, Steckley D, et al. Kir6.2 polymorphisms sensitize beta-cell ATP-sensitive potassium channels to activation by acyl CoAs: a possible cellular mechanism for increased susceptibility to type 2 diabetes? Diabetes. 2003;52:2630–5.
23. Florez JC, Burtt N, de Bakker PI, et al. Haplotype structure and genotype-phenotype correlations of the sulfonylurea receptor and the islet ATP-sensitive potassium channel gene region. Diabetes. 2004;53:1360–8.
24. Hamming KS, Soliman D, Matemisz LC, et al. 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. 2009;58:2419–24.
25. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. Nature. 2009;461:747–53.
26. •• Rees MG, Ng D, Ruppert S, et al. Correlation of rare coding variants in the gene encoding human glucokinase regulatory protein with phenotypic, cellular, and kinetic outcomes. J Clin Invest. 2012;122:205–17. *A comprehensive functional analysis of common and rare nonsynonymous variants in GCKR identified through targeted exon sequencing*.
27. •• Bonnefond A, Clement N, Fawcett K, et al. Rare *MTNR1B* variants impairing melatonin receptor 1B function contribute to type 2 diabetes. Nat Genet. 2012;44:297–301. *A functional analysis of common and rare nonsynonymous variants in MTNR1B identified through large-scale exon resequencing, which reaffirmed a functional link between the gene and T2D risk*.
28. Kong A, Steinthorsdottir V, Masson G, et al. Parental origin of sequence variants associated with complex diseases. Nature. 2009;462:868–74.
29. • Small KS, Hedman AK, Grundberg E, et al. Identification of an imprinted master trans regulator at the *KLF14* locus related to multiple metabolic phenotypes. Nat Genet. 2011;43:561–4. *An association study that showed that an eQTL of KLF14 trans-regulates adipose gene expression and, hence, established links between expression levels of several genetic loci and metabolic phenotypes*.
30. Rung J, Cauchi S, Albrechtsen A, et al. Genetic variant near *IRS1* is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. Nat Genet. 2009;41:1110–5.
31. Kilpelainen TO, Zillikens MC, Stancakova A, et al. Genetic variation near *IRS1* associates with reduced adiposity and an impaired metabolic profile. Nat Genet. 2011;43:753–60.
32. •• Gaulton KJ, Nammo T, Pasquali L, et al. A map of open chromatin in human pancreatic islets. Nat Genet. 2010;42:255–9. *A landmark study that provided a profile of open chromatin sites in human islets that can be used to identify regulatory roles of variants, such as that of TCF7L2, that underlie disease susceptibility*.
33. Fogarty MP, Panhuis TM, Vadlamudi S, et al. Allele-specific transcriptional activity at type 2 diabetes-associated single nucleotide polymorphisms in regions of pancreatic islet open chromatin at the *JAZF1* locus. Diabetes. 2013;62:1756–62.
34. •• Stitzel ML, Sethupathy P, Pearson DS, et al. Global epigenomic analysis of primary human pancreatic islets provides insights into type 2 diabetes susceptibility loci. Cell Metab. 2010;12:443–55. *A genome-wide analysis that generated a profile of the epigenome of human islets to enhance the understanding of gene regulation in islets and the role of noncoding variants in disease susceptibility*.
35. •• Moran I, Akerman I, van de Bunt M, et al. Human beta cell transcriptome analysis uncovers lncRNAs that are tissue-specific, dynamically regulated, and abnormally expressed in type 2 diabetes. Cell Metab. 2012;16:435–48. *A study describing a comprehensive transcriptome map of human islets and beta cells that first revealed the role of lncRNAs in beta cell physiology and disease*.
36. Esteller M. Non-coding RNAs in human disease. Nat Rev Genet. 2011;12:861–74.
37. Poy MN, Eliasson L, Krutzfeldt J, et al. A pancreatic islet-specific microRNA regulates insulin secretion. Nature. 2004;432:226–30.
38. Zhu H, Shyh-Chang N, Segre AV, et al. The Lin28/let-7 axis regulates glucose metabolism. Cell. 2011;147:81–94.

39. • van de Bunt M, Gaulton KJ, Parts L, et al. The miRNA profile of human pancreatic islets and beta-cells and relationship to type 2 diabetes pathogenesis. PLoS One. 2013;8:e55272. *A study that used small RNA sequencing to generate an miRNA profile in human islets and beta cells to evaluate the role of islet-expressed miRNAs and their targets in diabetes pathogenesis*.
40. Travers ME, Mackay DJG, Nitert MD, et al. Insights into the molecular mechanism for 2 diabetes susceptibility at the *KCNQ1* locus from temporal changes in imprinting status in human islets. Diabetes. 2013;62:987–92.
41. Dayeh TA, Olsson AH, Volkov P, et al. Identification of CpG-SNPs associated with type 2 diabetes and differential DNA methylation in human pancreatic islets. Diabetologia. 2013;56:1036–46.
42. Unoki H, Takahashi A, Kawaguchi T, et al. SNPs in *KCNQ1* are associated with susceptibility to type 2 diabetes in East Asian and European populations. Nat Genet. 2008;40:1098–102.
43. Jonsson A, Isomaa B, Tuomi T, et al. A variant in the *KCNQ1* gene predicts future type 2 diabetes and mediates impaired insulin secretion. Diabetes. 2009;58:2409–13.
44. Gutierrez-Arcelus M, Lappalainen T, Montgomery SB, et al. Passive and active DNA methylation and the interplay with genetic variation in gene regulation. Elife. 2013;2:e00523.
45. Skarnes WC, Rosen B, West AP, et al. A conditional knockout resource for the genome-wide study of mouse gene function. Nature. 2011;474:337–42.
46. McMurray F, Moir L, Cox RD. From mice to humans. Curr Diab Rep. 2012;12:651–8.
47. Steinthorsdottir V, Thorleifsson G, Reynisdottir I, et al. A variant in *CDKAL1* influences insulin response and risk of type 2 diabetes. Nat Genet. 2007;39:770–5.
48. Zeggini E, Weedon MN, Lindgren CM, et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science. 2007;316:1336–41.
49. Groenewoud MJ, Dekker JM, Fritsche A, et al. Variants of *CDKAL1* and *IGF2BP2* affect first-phase insulin secretion during hyperglycaemic clamps. Diabetologia. 2008;51:1659–63.
50. •• Wei FY, Suzuki T, Watanabe S, et al. Deficit of tRNA(Lys) modification by Cdkal1 causes the development of type 2 diabetes in mice. J Clin Invest. 2011;121:3598–608. *An elegant study of Cdkal1 in beta-cell-specific knockout mice that uncovered both the molecular function of Cdkal1 in protein translation and the mechanism by which Cdkal1 deficiency may lead to development of T2D*.
51. McCarthy MI, Rorsman P, Gloyn AL. *TCF7L2* and diabetes: a tale of two tissues, and of two species. Cell Metab. 2013;17:157–9.
52. Lyssenko V, Lupi R, Marchetti P, et al. Mechanisms by which common variants in the *TCF7L2* gene increase risk of type 2 diabetes. J Clin Invest. 2007;117:2155–63.
53. Le Bacquer O, Kerr-Conte J, Gargani S, et al. *TCF7L2* rs7903146 impairs islet function and morphology in non-diabetic individuals. Diabetologia. 2012;55:2677–81.
54. • Xavier GD, Mondragon A, Sun G, et al. Abnormal glucose tolerance and insulin secretion in pancreas-specific Tcf7l2-null mice. Diabetologia. 2012;55:2667–76. *A study characterising pancreas-specific Tcf7l2 knockout mice that showed that selective loss of Tcf7l2 results in an islet phenotype through the impairment of insulin secretion and glucose homeostasis*.
55. •• Boj SF, van Es JH, Huch M, et al. Diabetes risk gene and Wnt effector Tcf7l2/TCF4 controls hepatic response to perinatal and adult metabolic demand. Cell. 2012;151:1595–607. *A comprehensive study of global and tissue-specific inducible mouse models that, surprisingly, showed that depletion of Tcf7l2 primarily disrupts liver metabolism, providing new insight into the role of TCF4 in diabetes development*.
56. Klupa T, Skupien J, Malecki MT. Monogenic models: what have the single gene disorders taught us? Curr Diab Rep. 2012;12:659–66.
57. Pal A, Barber TM, van de Bunt M, et al. PTEN mutations as a cause of constitutive insulin sensitivity and obesity. N Engl J Med. 2012;367:1002–11.
58. Scott RA, Lagou V, Welch RP, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nat Genet. 2012;44:991–1005.
59. •• Nica AC, Ongen H, Irminger JC et al. Cell-type, allelic and genetic signatures in the human pancreatic beta cell transcriptome. Genome Res. 2013; In press. *The first characterization of the human beta cell transcriptome based on deep RNA sequencing that will shed light on cell-specific expression signatures and aid the study of diabetes-associated loci involved in beta cell gene expression.*
60. •• Ravassard P, Hazhouz Y, Pechberty S, et al. A genetically engineered human pancreatic beta cell line exhibiting glucose-inducible insulin secretion. J Clin Invest. 2011;121:3589–97. *A study describing the generation of a stable and functional human beta cell line that can be used as a cellular model to probe human beta cell physiology and as a preclinical model for developing therapeutics for diabetes*.
61. Bar-Nur O, Russ HA, Efrat S, et al. Epigenetic memory and preferential lineage-specific differentiation in induced pluripotent stem cells derived from human pancreatic islet beta cells. Cell Stem Cell. 2011;9:17–23.
62. Bellin M, Marchetto MC, Gage FH, et al. Induced pluripotent stem cells: the new patient? Nat Rev Mol Cell Biol. 2012;13:713–26.
63. Christian M, Cermak T, Doyle EL, et al. Targeting DNA double-strand breaks with TAL effector nucleases. Genetics. 2010;186:757–61.
64. •• Joung JK, Sander JD. INNOVATION TALENs: a widely applicable technology for targeted genome editing. Nat Rev Mol Cell Biol. 2013;14:49–55. *A review article describing the use of TALENs for targeted genome alteration in a wide range of cell types and organisms, as well as highlighting the exciting prospects that will emerge from the development of TALE technology*.
65. •• Mali P, Yang LH, Esvelt KM, et al. RNA-guided human genome engineering via Cas9. Science. 2013;339:823–6. *A report of a novel RNA-guided genome-editing tool engineered from the bacterial CRISPR system that promises efficient gene targeting in mammalian cells in a robust and multiplexable manner*.

PLOS | GENETICS

RESEARCH ARTICLE

Identification and Functional Characterization of *G6PC2* Coding Variants Influencing Glycemic Traits Define an Effector Transcript at the *G6PC2-ABCB11* Locus

Anubha Mahajan[1‡], Xueling Sim[2‡], Hui Jin Ng[3‡], Alisa Manning[4], Manuel A. Rivas[1], Heather M. Highland[5], Adam E. Locke[2], Niels Grarup[6], Hae Kyung Im[7], Pablo Cingolani[8,9], Jason Flannick[4,10], Pierre Fontanillas[4], Christian Fuchsberger[2], Kyle J. Gaulton[1], Tanya M. Teslovich[2], N. William Rayner[1,3,11], Neil R. Robertson[1,3], Nicola L. Beer[3], Jana K. Rundle[3], Jette Bork-Jensen[6], Claes Ladenvall[12], Christine Blancher[13], David Buck[13], Gemma Buck[13], Noël P. Burtt[4], Stacey Gabriel[4], Anette P. Gjesing[6], Christopher J. Groves[3], Mette Hollensted[6], Jeroen R. Huyghe[2], Anne U. Jackson[2], Goo Jun[2], Johanne Marie Justesen[6], Massimo Mangino[14], Jacquelyn Murphy[4], Matt Neville[3], Robert Onofrio[4], Kerrin S. Small[14], Heather M. Stringham[2], Ann-Christine Syvänen[15], Joseph Trakalo[13], Goncalo Abecasis[2], Graeme I. Bell[16], John Blangero[17], Nancy J. Cox[18], Ravindranath Duggirala[17], Craig L. Hanis[19], Mark Seielstad[20,21], James G. Wilson[22], Cramer Christensen[23], Ivan Brandslund[24,25], Rainer Rauramaa[26], Gabriela L. Surdulescu[14], Alex S. F. Doney[27], Lars Lannfelt[28], Allan Linneberg[29,30,31], Bo Isomaa[32,33], Tiinamaija Tuomi[33,34], Marit E. Jørgensen[35], Torben Jørgensen[31,36], Johanna Kuusisto[37,38], Matti Uusitupa[39], Veikko Salomaa[40], Timothy D. Spector[14], Andrew D. Morris[41], Colin N. A. Palmer[42], Francis S. Collins[43], Karen L. Mohlke[44], Richard N. Bergman[45], Erik Ingelsson[1,46], Lars Lind[47], Jaakko Tuomilehto[48,49,50,51], Torben Hansen[6,52], Richard M. Watanabe[53,54,55], Inga Prokopenko[1,3,56], Josee Dupuis[57,58], Fredrik Karpe[3,59], Leif Groop[12], Markku Laakso[37,38], Oluf Pedersen[6], Jose C. Florez[4,60,61,62], Andrew P. Morris[1,63,64], David Altshuler[4,10,60,61,65,66], James B. Meigs[67], Michael Boehnke[2], Mark I. McCarthy[1,3,59], Cecilia M. Lindgren[1,4*‡], Anna L. Gloyn[3,59*‡], On Behalf of the T2D-GENES consortium and GoT2D consortium[¶]

1 Wellcome Trust Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom, **2** Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan, United States of America, **3** Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, United Kingdom, **4** Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, United States of America, **5** Human Genetics Center, The University of Texas Graduate School of Biomedical Sciences at Houston, The University of Texas Health Science Center at Houston, Houston, Texas, United States of America, **6** The Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, **7** Department of Health Studies, Biostatistics Laboratory, The University of Chicago, Chicago, Illinois, United States of America, **8** School of Computer Science, McGill University, Montreal, Quebec, Canada, **9** McGill University and Génome Québec Innovation Centre, Montreal, Quebec, Canada, **10** Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts, United States of America, **11** Department of Human Genetics, Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, United Kingdom, **12** Department of Clinical Sciences, Diabetes and Endocrinology, Lund University Diabetes Centre, Malmö, Sweden, **13** High Throughput Genomics, Oxford Genomics Centre, Wellcome Trust Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom, **14** Department of Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom, **15** Department of Medical Sciences, Molecular Medicine and Science for Life Laboratory, Uppsala University, Uppsala, Sweden, **16** Departments of Medicine and Human Genetics, The University of Chicago, Chicago, Illinois, United States of America, **17** Department of Genetics, Texas Biomedical Research Institute, San Antonio, Texas, United States of America,

OPEN ACCESS

Citation: Mahajan A, Sim X, Ng HJ, Manning A, Rivas MA, Highland HM, 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(1): e1004876. doi:10.1371/journal.pgen.1004876

Editor: Martin D. Tobin, University of Leicester, UNITED KINGDOM

Received: June 25, 2014

Accepted: November 4, 2014

Published: January 27, 2015

Data Availability Statement: This is a meta-analysis that was conducted on summary level results. Individual level data was not shared amongst the authors of the manuscript, and the corresponding authors are not in a position to make the individual level data available. For most of the samples included, individual level data deposition is precluded by existing consents, and other issues related to individual privacy. Summary level data from the meta-analysis are available from the DIAGRAM (http://www.diagram-consortium.org/Mahajan_2014_ExomeChip/) and

LocusZoom (http://csg.sph.umich.edu/locuszoom/) website.

Funding: HJN was supported by the Agency for Science, Technology and Research in Singapore. CF was supported by the Austrian Science Fund (FWF) grant J-3401. JCF was supported by MGH Research Scholars Award. APM is a Wellcome Trust Senior Research Fellow in Basic and Biomedical Science. JBM was supported by NIDDK R01 DK078616 and K24 DK080140. MIM is a Wellcome Trust Senior Investigator (WT098381); and a National Institute of Health Research Senior Investigator. CML is a Wellcome Trust Research Career Development Fellow (086596/Z/08/Z). ALG is a Wellcome Trust Senior Research Fellow in Basic and Biomedical Science (095010/Z/10/Z). Funding for the GoT2D and T2D-GENES studies was provided by grants NIH U01s DK085526, DK085501, DK085524, DK085545, and DK085584 (Multiethnic Study of Type 2 Diabetes Genes) and DK088389 (Low-Pass Sequencing and High-Density SNP Genotyping for Type 2 Diabetes). Analysis and genotyping of the British UK cohorts was supported by Wellcome Trust funding 090367, 098381, 090532, 083948, 085475, MRC (G0601261), EU (Framework 7) HEALTH-F4-2007-201413, and NIDDK DK098032. The Oxford Biobank is supported by the Oxford Biomedical Research Centre and part of the National NIHR Bioresource. The PIVUS/ULSAM cohort was supported by Wellcome Trust Grants WT098017, WT064890, WT090532, Uppsala University, Uppsala University Hospital, the Swedish Research Council and the Swedish Heart-Lung Foundation. GoDARTS study was funded by The Wellcome Trust Study Cohort Wellcome Trust Functional Genomics Grant (2004-2008) (Grant No: 072960/2/03/2) and The Wellcome Trust Scottish Health Informatics Programme (SHIP) (2009-2012) (Grant No: 086113/Z/08/Z). TwinsUK study was funded by the Wellcome Trust; European Community's Seventh Framework Programme (FP7/2007–2013). The study also receives support from the National Institute for Health Research (NIHR) BioResource Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London. TDS is holder of an ERC Advanced Principal Investigator award. The Danish studies were supported by the Lundbeck Foundation (Lundbeck Foundation Centre for Applied Medical Genomics in Personalised Disease Prediction, Prevention and Care (LuCamp); http://www.lucamp.org/) and the Danish Council for Independent Research. The Novo Nordisk Foundation Center for Basic Metabolic Research is an independent Research Center at the University of Copenhagen, partially funded by an unrestricted donation from the Novo Nordisk Foundation (http://www.metabol.ku.dk/). The METSIM study was supported by the Academy

18 Department of Medicine, Section of Genetic Medicine, The University of Chicago, Chicago, Illinois, United States of America, **19** Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, Texas, United States of America, **20** Blood Systems Research Institute, San Francisco, California, United States of America, **21** Department of Laboratory Medicine & Institute for Human Genetics, University of California, San Francisco, San Francisco, California, United States of America, **22** Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, Mississippi, United States of America, **23** Department of Internal Medicine and Endocrinology, Vejle Hospital, Vejle, Denmark, **24** Department of Clinical Biochemistry, Vejle Hospital, Vejle, Denmark, **25** Institute of Regional Health Research, University of Southern Denmark, Odense, Denmark, **26** Foundation for Research in Health, Exercise and Nutrition, Kuopio Research Institute of Exercise Medicine, Kuopio, Finland, **27** Division of Cardiovascular and Diabetes Medicine, Medical Research Institute, Ninewells Hospital and Medical School, Dundee, United Kingdom, **28** Department of Public Health and Caring Sciences, Geriatrics, Uppsala University, Uppsala, Sweden, **29** Department of Clinical Experimental Research, Glostrup University Hospital, Glostrup, Denmark, **30** Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, **31** Research Centre for Prevention and Health, Glostrup University Hospital, Glostrup, Denmark, **32** Department of Social Services and Health Care, Jakobstad, Finland, **33** Folkhälsan Research Centre, Helsinki, Finland, **34** Department of Endocrinology, Helsinki University Central Hospital, Helsinki, Finland, **35** Steno Diabetes Center, Gentofte, Denmark, **36** Faculty of Medicine, University of Aalborg, Aalborg, Denmark, **37** Faculty of Health Sciences, Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland, Kuopio, Finland, **38** Kuopio University Hospital, Kuopio, Finland, **39** Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland, **40** National Institute for Health and Welfare, Helsinki, Finland, **41** Clinical Research Centre, Centre for Molecular Medicine, Ninewells Hospital and Medical School, Dundee, United Kingdom, **42** Pat Macpherson Centre for Pharmacogenetics and Pharmacogenomics, Medical Research Institute, Ninewells Hospital and Medical School, Dundee, United Kingdom, **43** Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, United States of America, **44** Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America, **45** Cedars-Sinai Diabetes and Obesity Research Institute, Los Angeles, California, United States of America, **46** Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden, **47** Department of Medical Sciences, Uppsala University, Uppsala, Sweden, **48** Diabetes Research Group, King Abdulaziz University, Jeddah, Saudi Arabia, **49** Instituto de Investigacion Sanitaria del Hospital Universario LaPaz (IdiPAZ), University Hospital LaPaz, Autonomous University of Madrid, Madrid, Spain, **50** Center for Vascular Prevention, Danube University Krems, Krems, Austria, **51** Diabetes Prevention Unit, National Institute for Health and Welfare, Helsinki, Finland, **52** Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark, **53** Department of Physiology & Biophysics, Keck School of Medicine, University of Southern California, Los Angeles, California, United States of America, **54** Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California, United States of America, **55** Diabetes and Obesity Research Institute, Keck School of Medicine, University of Southern California, Los Angeles, California, United States of America, **56** Department of Genomics of Common Disease, School of Public Health, Imperial College London, London, United Kingdom, **57** National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Massachusetts, United States of America, **58** Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, United States of America, **59** Oxford NIHR Biomedical Research Centre, Oxford University Hospitals Trust, Oxford, United Kingdom, **60** Diabetes Research Center (Diabetes Unit), Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, United States of America, **61** Department of Medicine, Harvard Medical School, Boston, Massachusetts, United States of America, **62** Center for Human Genetic Research, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, United States of America, **63** Department of Biostatistics, University of Liverpool, Liverpool, United Kingdom, **64** Estonian Genome Centre, University of Tartu, Tartu, Estonia, **65** Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America, **66** Department of Genetics, Harvard Medical School, Boston, Massachusetts, United States of America, **67** General Medicine Division, Massachusetts General Hospital and Department of Medicine, Harvard Medical School, Boston, Massachusetts, United States of America

¶ Members of both consortia are listed in the Acknowledgments.
‡ These authors contributed equally to this work.
* celi@well.ox.ac.uk (CML); anna.gloyn@drl.ox.ac.uk (ALG)

of Finland (contract 124243), the Finnish Heart Foundation, the Finnish Diabetes Foundation, Tekes (contract 1510/31/06), and the Commission of the European Community (HEALTH-F2-2007-201681), and the US National Institutes of Health grants DK093757, DK072193, DK062370, and 1Z01 HG000024. The FUSION study was supported by DK093757, DK072193, DK062370, and 1Z01 HG000024. Genotyping of the METSIM and DPS studies was conducted at the Genetic Resources Core Facility (GRCF) at the Johns Hopkins Institute of Genetic Medicine. VS is funded by the Finnish Foundation for Cardiovascular Research and the Academy of Finland (grant # 139635). The FIN-D2D 2007 study has been financially supported by the hospital districts of Pirkanmaa, South Ostrobothnia, and Central Finland, the Finnish National Public Health Institute (current National Institute for Health and Welfare), the Finnish Diabetes Association, the Ministry of Social Affairs and Health in Finland, the Academy of Finland (grant number 129293), Commission of the European Communities, Directorate C-Public Health (grant agreement no. 2004310) and Finland's Slottery Machine Association. The DPS has been financially supported by grants from the Academy of Finland (117844 and 40758, 211497, and 118590 (MU); The EVO funding of the Kuopio University Hospital from Ministry of Health and Social Affairs (5254), Finnish Funding Agency for Technology and Innovation (40058/07), Nordic Centre of Excellence on 'Systems biology in controlled dietary interventions and cohort studies, SYSDIET (070014), The Finnish Diabetes Research Foundation, Yrjö Jahnsson Foundation (56358), Sigrid Juselius Foundation and TEKES grants 70103/06 and 40058/07. The DR's EXTRA Study was supported by grants to RR by the Ministry of Education and Culture of Finland (627;2004-2011), Academy of Finland (102318; 123885), Kuopio University Hospital, Finnish Diabetes Association, Finnish Heart Association, Päivikki and Sakari Sohlberg Foundation and by grants from European Commission FP6 Integrated Project (EXGENESIS); LSHM-CT-2004-005272, City of Kuopio and Social Insurance Institution of Finland (4/26/2010). The Broad Genomics Platform for genotyping of the FIN-D2D 2007, FINRISK 2007, DR'sEXTRA, FUSION, and PPP studies. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Genome wide association studies (GWAS) for fasting glucose (FG) and insulin (FI) have identified common variant signals which explain 4.8% and 1.2% of trait variance, respectively. It is hypothesized that low-frequency and rare variants could contribute substantially to unexplained genetic variance. To test this, we analyzed exome-array data from up to 33,231 non-diabetic individuals of European ancestry. We found exome-wide significant ($P<5\times10\text{-}7$) evidence for two loci not previously highlighted by common variant GWAS: *GLP1R* (p.Ala316Thr, minor allele frequency (MAF)=1.5%) influencing FG levels, and *URB2* (p.Glu594Val, MAF = 0.1%) influencing FI levels. Coding variant associations can highlight potential effector genes at (non-coding) GWAS signals. At the *G6PC2/ABCB11* locus, we identified multiple coding variants in *G6PC2* (p.Val219Leu, p.His177Tyr, and p Tyr207Ser) influencing FG levels, conditionally independent of each other and the non-coding GWAS signal. *In vitro* assays demonstrate that these associated coding alleles result in reduced protein abundance via proteasomal degradation, establishing *G6PC2* as an effector gene at this locus. Reconciliation of single-variant associations and functional effects was only possible when haplotype phase was considered. In contrast to earlier reports suggesting that, paradoxically, glucose-raising alleles at this locus are protective against type 2 diabetes (T2D), the p.Val219Leu *G6PC2* variant displayed a modest but directionally consistent association with T2D risk. Coding variant associations for glycemic traits in GWAS signals highlight *PCSK1*, *RREB1*, and *ZHX3* as likely effector transcripts. These coding variant association signals do not have a major impact on the trait variance explained, but they do provide valuable biological insights.

Author Summary

Understanding how FI and FG levels are regulated is important because their derangement is a feature of T2D. Despite recent success from GWAS in identifying regions of the genome influencing glycemic traits, collectively these loci explain only a small proportion of trait variance. Unlocking the biological mechanisms driving these associations has been challenging because the vast majority of variants map to non-coding sequence, and the genes through which they exert their impact are largely unknown. In the current study, we sought to increase our understanding of the physiological pathways influencing both traits using exome-array genotyping in up to 33,231 non-diabetic individuals to identify coding variants and consequently genes associated with either FG or FI levels. We identified novel association signals for both traits including the receptor for GLP-1 agonists which are a widely used therapy for T2D. Furthermore, we identified coding variants at several GWAS loci which point to the genes underlying these association signals. Importantly, we found that multiple coding variants in *G6PC2* result in a loss of protein function and lower fasting glucose levels.

Introduction

Large-scale GWAS of non-diabetic individuals have successfully identified > 60 loci associated with FG and FI levels, many of which are also implicated in susceptibility to T2D [1, 2, 3, 4].

Despite these successes, lead SNPs at GWAS loci have modest effects and cumulatively explain only a small proportion of the trait variance in non-diabetic individuals. By design, GWAS have focused predominantly on the interrogation of common variants, defined here to have MAF > 5%. Most of the identified variants are non-coding, complicating attempts to establish the molecular consequences of these GWAS loci. We therefore chose to extend discovery efforts to coding variants, particularly those of lower frequency that have not been well captured by GWAS genotyping and imputation. We aimed both to identify novel coding loci for FG and FI, and to evaluate the role of coding variants at known GWAS loci, thereby expecting to highlight causal transcripts and to facilitate characterization of the molecular mechanisms influencing glycemic traits and T2D susceptibility.

Results

We analyzed 33,231 (FG) and 30,825 (FI) non-diabetic individuals from 14 studies of European ancestry, all genotyped with the Illumina HumanExome BeadChip (**see URLs**). Characteristics of the contributing studies and study participants are summarized in S1-S2 Tables. Body mass index (BMI) adjustment has been shown to increase power to detect association with these glycemic traits [4], and in our study samples, BMI accounted for 6.1% and 24.6% of phenotypic variance of FG and FI, respectively. Consequently, within each study, we calculated residuals for both traits after adjustment for BMI and other study-specific covariates (S1 Table). Study-specific inverse-rank normalized residuals were tested for single-variant association using a linear mixed model to account for relatedness and fine-scale genetic population sub-structure [5]. We also repeated the analysis using the untransformed residuals to obtain allelic effect sizes. We then combined the association summary statistics across studies using fixed-effect meta-analysis. We restricted our single-variant analysis to 106,489 variants that pass quality-control and are polymorphic in more than one study. We declared a single-variant trait association as exome-wide significant at $P < 5\times10^{-7}$, corresponding to Bonferroni correction for the ~100,000 polymorphic variants. We also carried out gene-based meta-analysis [6, 7] by using the sequence kernel association test (SKAT) [8] and a frequency-weighted burden test [9] applying four alternate variant masks which combine functional annotation and allele frequency thresholds. Full details of the variant masks are provided in the Methods. Gene-based tests take into account overall variant-load within a specified locus and therefore may have greater power than single-variant tests to detect associations with multiple rare and low-frequency causal alleles. We note that this advantage is likely to be less in exome-array analysis compared with the more complete ascertainment of variants possible with exome sequencing [10]. We declared gene-based association as exome-wide significant at $P < 2.5\times10^{-6}$, corresponding to Bonferroni correction for ~20,000 protein-coding genes in the genome.

Coding variants influencing FG levels

Through single-variant analysis, we identified 12 coding variants (all of which were non-synonymous changes) associated with FG levels at exome-wide significance, two low-frequency and ten common (S3 Table). These variants mapped to seven loci, six of them previously implicated in FG regulation. The signals at known loci included previously-reported common coding variants driving GWAS signals at *GCKR* (p.Pro446Leu, MAF = 36.9%, $P = 5.3\times10^{-18}$) and *SLC30A8* (p.Arg325Trp, MAF = 35.7%, $P = 2.5\times10^{-10}$) [2, 11]. Three additional common coding variants associated with FG (in *C2orf16*, *GPN1*, and *SLC5A6*) were in moderate linkage disequilibrium (LD; $r^2 = 0.2–0.4$) with the p.Pro446Leu *GCKR* variant. Their associations were eliminated after conditioning on p.Pro446Leu, the known functional GWAS variant in this region [12, 13], indicating no causal role for these additional variants on FG regulation (S3 Table).

The sixth variant influencing FG levels at exome-wide significance was a low-frequency non-synonymous change which did not map to any previously known FG-influencing locus: p.Ala316Thr at *GLP1R* (MAF = 1.5%, $P = 4.6\times10^{-7}$; Table 1 and S1B Fig.). This variant showed modest association ($P = 1.3\times10^{-4}$) in a previous GWAS meta-analysis of FG [3], which partially overlaps with the present study, but now achieves exome-wide significance. The alanine residue at p.Ala316Thr is conserved across vertebrates, and the threonine substitution is predicted to be "possibly damaging" by *in silico* mutation analysis (S4 Table). GLP-1R (glucagon-like peptide 1 receptor) is the receptor for the incretin hormone glucagon-like peptide 1 (GLP1), which is released from enteroendocrine cells after food ingestion and potentiates insulin secretion. GLP1 receptor agonists are an established treatment for T2D [14].

The six remaining coding variants all map to known FG-associated GWAS loci, but have not previously been implicated as playing a causal role. These include two variants in the islet-specific glucose-6-phosphatase catalytic subunit (*G6PC2*) gene at the *G6PC2*/*ABCB11* locus: p.Val219Leu, a common variant ($P = 6.0\times10^{-9}$, MAF = 48.1%), and p.His177Tyr, a low-frequency variant ($P = 3.1\times10^{-8}$, MAF = 0.8%; S2A Fig.). These variants remained significantly associated (p.Val219Leu, $P_{conditional} = 7.1\times10^{-10}$ and p.His177Tyr, $P_{conditional} = 1.3\times10^{-11}$) with FG after conditioning on the intronic lead GWAS SNP, rs560887 (Table 1 and S2B Fig.). Conversely, conditioning on the coding variants did not completely abolish the association signal at the lead GWAS SNP ($P_{\text{unconditional}} = 6.4\times10^{-78}$; $P_{conditional\ on\text{His177Tyr}} = 3.1\times10^{-55}$, $P_{conditional\ on\text{Val219Leu}} = 1.2\times10^{-58}$, and $P_{conditional\ on\text{His177Tyr and Val219Leu}} = 2.1\times10^{-83}$), confirming that the effect of the coding variants were largely independent of the lead GWAS SNP. Furthermore, the coding variants each remained associated with FG at exome-wide significance even after conditioning on both the lead GWAS SNP and the other coding variant, providing clear evidence of at least three association signals at this locus (Table 1 and S2C-S2D Fig.). These results are consistent with a recent study in Finnish individuals that reported a FG association signal at p.His177Tyr [15], but we extend that finding by demonstrating exome-wide significant association of multiple coding variants after conditioning on other associated variants in the region.

G6PC2 was also the only one of the 14,465 genes with multiple exome-array variants to demonstrate evidence of significant association with FG levels with any mask in gene-based tests (Table 2). In fact, for this gene we observed significant association with FG levels for all the masks encompassing multiple variants. These gene-based associations remained significant even after Bonferroni correction for testing of four different variant masks. Step-wise conditional analyses, adjusting for each variant included in the respective masks, revealed that the gene-based signals were primarily driven by two variants: p.His177Tyr and p.Tyr207Ser (S5 Table). The protein-truncating variant (PTV) p.Arg283∗, which introduces a stop mutation in the last exon of the gene, showed no association with FG levels in single-variant or gene-based analyses. The most likely explanation is that this variant evades nonsense mediated decay (as is usual for PTV in the last exon) and that the truncated protein (missing only the terminal 72 amino acid residues) retains normal functional activity [16]. Due to its high allele frequency, and annotation as benign across multiple annotation algorithms, p.Val219Leu was not included in any of the gene-based variant masks (S6 Table). However, based on the single-variant it also has independent effects, with the result that we have three coding variants in *G6PC2* (p.His177Tyr, p.Tyr207Ser, and p.Val219Leu) influencing FG levels. These three variants explain an additional 0.2% of the phenotypic variance in FG beyond that explained by the GWAS variant rs560887, bringing the total variance explained by this locus to 1.1%. However, we recognize that this estimate has been obtained in the discovery cohort and consequently might be inflated.

Both *G6PC2* and *ABCB11* have been considered strong biological candidate genes for glucose regulation, and advocated as potential effector transcripts at this GWAS locus [17, 18].

Table 1. Coding variants associated with FG and FI levels at exome-wide significance.

SNP	Gene	Chr	Variant	Minor/ Major allele[a]	MAF[b] (%)	N	I^2	Cochran's Q	*P*_het	Unconditional		Conditional	
										P[c]	$\hat{\beta}$ (SE)	*P*	$\hat{\beta}$ (SE)
Fasting glucose													
rs138726309	*G6PC2*	2	p.His177Tyr	T/C	0.8	32,430	27.2	16.49	0.17	3.1×10^{-8}	-0.102 (0.020)	[d] 1.3×10^{-11}	-0.125 (0.020)
												[e] 3.5×10^{-10}	-0.115 (0.020)
rs492594	*G6PC2*	2	p.Val219Leu	C/G	48.1	33,231	0	6.68	0.92	6.0×10^{-9}	0.020 (0.004)	[d] 7.1×10^{-10}	-0.034 (0.005)
												[f] 6.5×10^{-13}	-0.032 (0.005)
rs6234	*PCSK1*	5	p.Gln665Glu	C/G	27.9	33,231	0	8.58	0.80	3.0×10^{-8}	-0.022 (0.004)	-	-
rs6235	*PCSK1*	5	p.Ser690Thr	G/C	27.9	33,231	0	8.65	0.80	4.1×10^{-8}	-0.022 (0.004)	-	-
rs35742417	*RREB1*	6	p.Ser1554Tyr	A/C	21.1	33,230	0	12.28	0.51	8.4×10^{-9}	-0.024 (0.004)	-	-
rs10305492	*GLP1R*	6	p.Ala316Thr	A/G	1.5	33,230	0	11.29	0.59	4.6×10^{-7}	-0.073 (0.015)	-	-
rs17265513	*ZHX3*	20	p.Asn310Ser	C/T	23.8	33,229	0	12.01	0.53	3.9×10^{-7}	0.022 (0.004)	-	-
Fasting insulin													
rs141203811	*URB2*	1	p.Glu594Val	T/A	0.1	21,130	59.9	9.98	0.04	3.1×10^{-7}	0.282 (0.066)	-	-

Chr: chromosome. MAF: minor allele frequency. N: number of samples analyzed. I^2: heterogeneity measure in %. *P*_het: P-value for Cochran's Q statistic.

$\hat{\beta}$: regression coefficient estimates. SE: standard error.

[a]Alleles are aligned to the forward strand of NCBI Build 37.

[b]Sample-size weighted average minor allele frequency percentage across all studies.

[c]Sample-size Z-score weighted *P* values are obtained with derived inverse normalized residuals of mmol/L of fasting glucose and pmol/L of natural log-transformed fasting insulin after adjustment for age, sex, and BMI.

Effect size estimates are reported for the minor allele in mmol/L of fasting glucose and pmol/L of natural log-transformed fasting insulin after adjustment for age, sex, and BMI.

[d]After adjusting for the common lead non-coding GWAS SNP rs560887.

[e]After adjusting for the common lead non-coding GWAS SNP rs560887 and the common non-synonymous variant p.Val219Leu.

[f]After adjusting for the common lead non-coding GWAS SNP rs560887 and the low-frequency non-synonymous variant p.His177Tyr.

doi:10.1371/journal.pgen.1004876.t001

However, none of the 21 coding variants in *ABCB11* that passed quality control (twelve rare, eight low-frequency, and one common) were significantly associated with FG levels, nor was there an aggregate association signal ($P>0.05$). Together, our genetic data provide compelling evidence that *G6PC2* is an effector gene for FG regulation at this GWAS locus, with p.His177Tyr, p.Tyr207Ser, and p.Val219Leu as likely causal coding variants.

Loss of G6PC2 function leads to a reduction in FG levels

The phenotypic impact of these coding variants might, we reasoned, be influenced by the action of the non-coding GWAS variants on *G6PC2* transcript regulation. Haplotype analysis of the four variants (rs560887, p.His177Tyr, p.Tyr207Ser, and p.Val219Leu) in a subset of 4,442 individuals revealed that the C allele encoding Leu219 was carried exclusively in *cis* with the glucose-raising allele at the GWAS SNP (Fig. 1). The estimated effect size of p.Val219Leu was considerably smaller ($\hat{\beta}$ =0.020 ± 0.004 mmol/L per allele) than rs560887 ($\hat{\beta}$ =0.070 ± 0.004 mmol/L per allele). These observations together explain the reversal in direction of effect of the Leu219 allele between conditional and unconditional analyses and indicate that, whilst Leu219

Table 2. *G6PC2* gene-based association with FG levels using SKAT and BURDEN test

Mask	Number of variants	Average allele frequency (%)	P_{SKAT}	P_{BURDEN}
Mask 1: PTV-only	0	-	-	-
Mask 2: PTV + missense	15	0.10	1.8×10^{-13}	4.1×10^{-16}
Mask 3: PTV + NS_{strict}	4	0.28	3.6×10^{-12}	5.1×10^{-13}
Mask 4: PTV + NS_{broad}	12	0.12	2.0×10^{-13}	1.2×10^{-17}

PTV: protein-truncating variant; NS: non-synonymous variants; NS_{strict}: missense variant predicted to be deleterious by all five annotation algorithms (Polyphen2-HumDiv, PolyPhen2-HumVar, LRT, MutationTaster, and SIFT); NS_{broad}: missense variant predicted to be deleterious by at least one of the five annotation algorithms.

Mask 1: "PTV-only" encompassed only one variant.

Mask 2 consists of PTVs and all missense variants with MAF<1%; Mask 3 consists of PTVs and missense variants predicted to be deleterious by five annotation algorithms without any upper MAF bounds; Mask 4 consists of variants in Mask 3 plus any missense variant with MAF<1% predicted to be deleterious by at least one of the five annotation algorithms.

Allele counts of the 15 non-synonymous variants at *G6PC2*: p.Ser324Pro (56), p.Leu310Phe (42), p.Arg283* (71), pIle273Val (2), p.Phe256Leu (9), p. Ile230Thr (14), p. Tyr207Ser (338), p.His177Tyr (502), p.Ile171Thr (81), p.Ile171Val (4), p.Ala119Val (23), p.Asn68Ile (2), p.Ile63Thr (6), p.Ile38Leu (1), p.Ser30Phe (23).

doi:10.1371/journal.pgen.1004876.t002

appears to have a glucose-raising effect in single-variant analyses, the molecular consequence of Leu219 is likely to be to lower glucose (Table 1). The minor alleles of the other two coding variants (p.His177Tyr, p.Tyr207Ser) also displayed glucose-lowering effects in conditional analyses. Given the role of G6PC2 in beta-cells we predicted that these would also be associated with reduced G6PC2 function, a hypothesis that we set out to test with a series of *in vitro* studies.

First, we generated recombinant FLAG-tagged *G6PC2* constructs to investigate the impact of these variants on protein expression and subsequent cellular localization. Transient transfection of all three constructs showed a marked reduction in G6PC2 protein levels. In HEK293

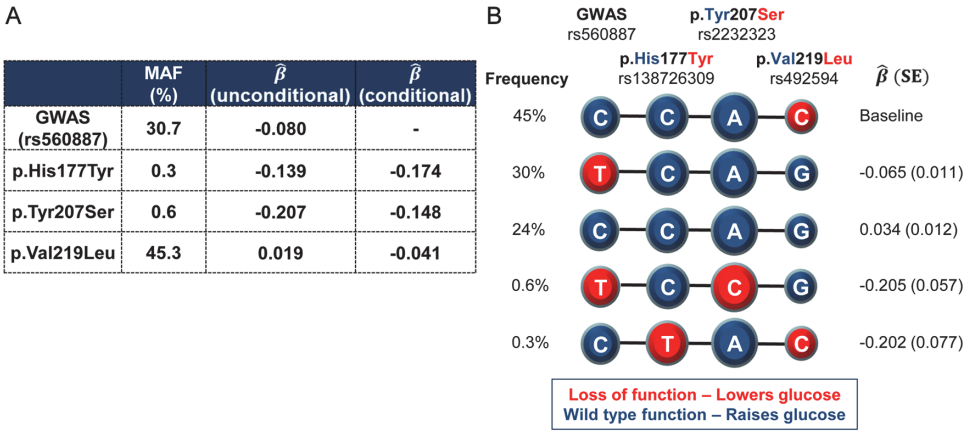


Figure 1. Haplotypes of the lead non-coding GWAS SNP rs560887 and the three coding variants. rs138726309 (p.His177Tyr), rs2232323 (p.Tyr207Ser), and rs492594 (p.Val219Leu), obtained from 4,442 unrelated individuals from the Oxford Biobank. (A) Percentage minor allele frequency (MAF) and effect size estimates ($\hat{\beta}$) of the four variants reported for the minor allele in mmol/L of FG after adjustment for age, sex, and BMI. (B) Haplotypes of the four associated variants in G6PC2 revealed that the glucose-lowering Leu219 allele was carried exclusively in cis with the glucose-raising allele at the GWAS SNP. Wild-type, glucose-raising alleles are circled in blue and the mutant, glucose-lowering alleles are circled in red. Diameter of the circle is proportional to the effect size estimates. Haplotype association was performed with FG derived residuals (after adjustment for age, sex, and BMI) using the most frequent haplotype as baseline.

doi:10.1371/journal.pgen.1004876.g001

cells, protein levels were decreased by 99% (p.His177Tyr), 100% (p.Tyr207Ser), and 49% (p.Val219Leu), when compared to wild type (defined for this study as the major allele for each of these variants) G6PC2 (Fig. 2A). Reduced protein expression was also confirmed in the rat pancreatic β-cell line INS-1E (76%, 97%, and 49% reduction, respectively; Fig. 2B). The functional impact of the variant proteins mirrored our genetic observations; the variant protein with the p.Val219Leu substitution, which has only a modest effect on FG, was present at higher levels in both HEK293 and INS-1E cells than the p.His177Tyr and p.Tyr207Ser proteins, both of which have larger impacts on FG levels.

Our functional results for p.His177Tyr and p.Tyr207Ser are concordant with data from the *in silico* mutation assessment tools SIFT and PolyPhen-2, which predicted these variants as deleterious (S4 Table). In contrast, the common p.Val219Leu variant was predicted to be benign, whereas *in vitro* characterization clearly shows a significant reduction in protein expression, demonstrating the importance of experimentally evaluating coding variants for functional consequences. The observed decrease in G6PC2 protein levels caused by the FG-associated variants is also directionally consistent with the evidence indicating that the non-coding GWAS SNP rs560887 may have effects on pre-mRNA splicing [19].

Next, we used specific inhibitors of the proteasomal and lysosomal pathways (MG-132 and chloroquine, respectively) to demonstrate that the three G6PC2 variant proteins with p.His177-Tyr, p.Tyr207Ser, and p.Val219Leu substitutions were predominantly degraded through the ubiquitin-proteasome pathway, an important cellular mechanism for clearing misfolded proteins. Protein expression could be rescued in the presence of MG-132 but not chloroquine (Fig. 2C). We also evaluated the impact of the variants on cellular localization of G6PC2 to the endoplasmic reticulum (ER) using calnexin as an ER marker. The variant proteins displayed similar localization patterns to wild type G6PC2 (Fig. 2D). Our finding that loss of G6PC2 function leads to a reduction in FG levels in humans is consistent with rodent data, which show that *G6pc2* knockout mice have a ~15% decrease in FG levels [20]. Hence, our data, linking genetic associations with reduced protein function, indicate that normal G6PC2 function is critical for glucose homeostasis and that these variants most likely impact FG levels through altered intracellular catalysis of glucose-6-phosphate to glucose and inorganic phosphate in pancreatic β-cells.

Impact of the functional *G6PC2* variants on other related quantitative traits and disease outcomes

We evaluated the impact of these functional *G6PC2* variants on other related quantitative traits and disease outcomes to gain further insights into the metabolic processes involved (S7 Table). None of the variants showed any evidence of association with FI levels ($P>0.1$) in 30,825 non-diabetic individuals analyzed in the present study. No association of the lead *G6PC2* GWAS variant with T2D risk has previously been shown in Europeans [2, 21, 22]. However, in a meta-analysis of exome-array data from 28,344 T2D cases and 51,801 controls of European ancestry (including the 33,231 controls used in the present meta-analyses), the common coding variant, p.Val219Leu, showed modest association with T2D risk ($P = 0.0011$; odds ratio = 1.05, 95% confidence interval 1.02–1.06). The G allele encoding Val219, which displayed a glucose-raising effect, conferred an increased risk of disease. In contrast to the observations in FG single-variant analysis, the direction of effect on T2D at this variant was unchanged, even after conditioning on the GWAS variant. This is consistent with a small case-control study in 3,676 individuals of Chinese ancestry where a nominal association (P = 0.0062) with T2D risk was reported [23]. Our finding provides evidence that *G6PC2* is critical not only for controlling FG levels in the physiological range, but that impairment of its function contributes to T2D pathogenesis. The effect on T2D risk is modest given the impact of the p.Val219Leu variant on FG levels but the effect size is

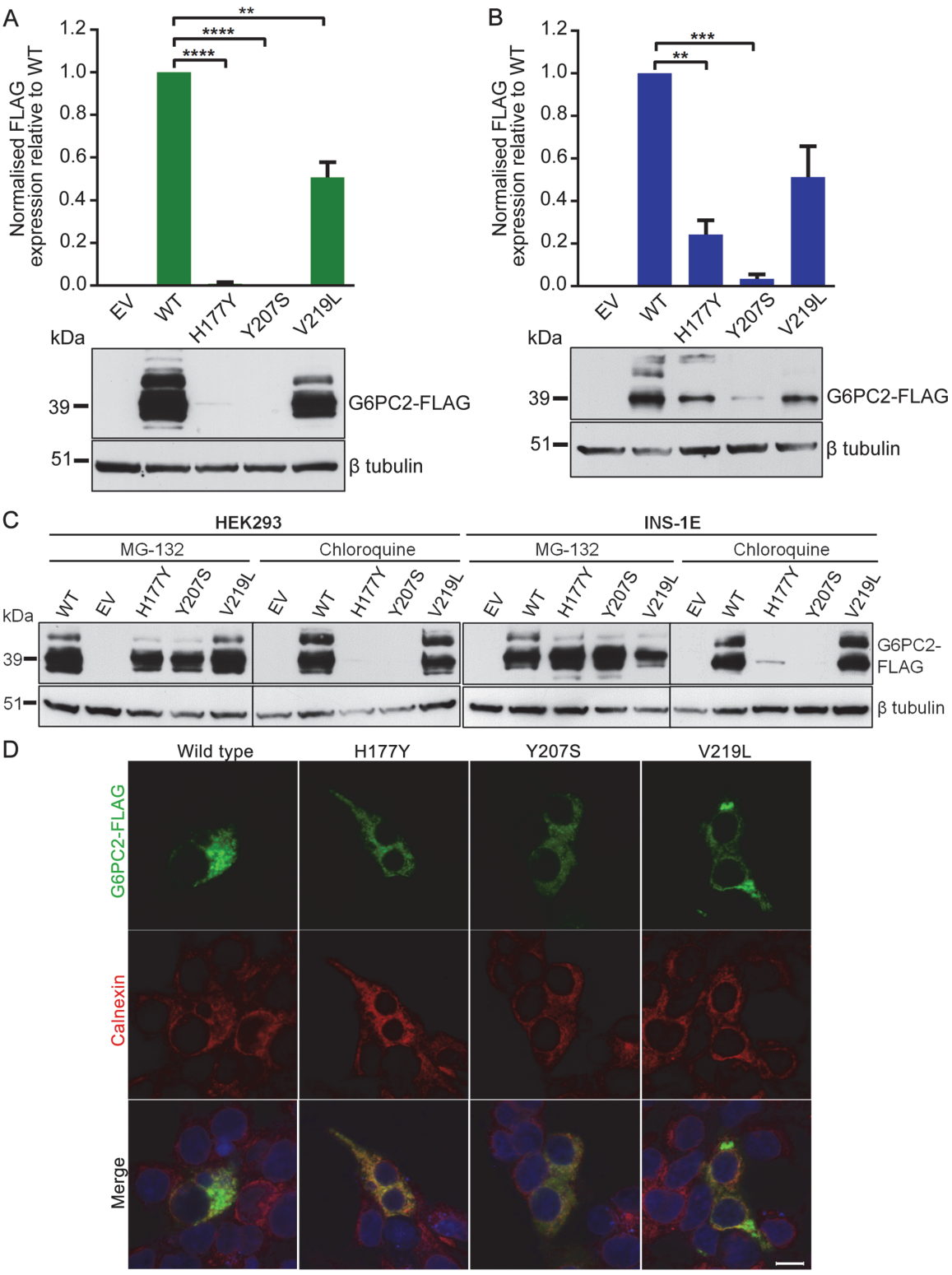


Figure 2. Functional characterization of wild type and variant G6PC2 proteins. (A) Expression levels in HEK293 and (B) INS-1E cells were determined by western blot and densitometry analysis. The multiple bands on the western blot are likely to represent glycosylated G6PC2 protein products. Data are presented as mean ± standard error of the mean for at least three independent experiments. Significant differences are indicated as ** $P<0.01$; *** $P<0.001$; **** $P<0.0001$. EV, empty vector; WT, wild type. (C) Expression levels in HEK293 and INS-1E cells in the presence of proteasomal inhibitor MG-132 or lysosomal inhibitor chloroquine were determined by western blot. (D) Cellular localization in HEK293 cells was assessed by immunofluorescence

microscopy. Cells were double immunostained for FLAG tag (green) and calnexin (red), and merged images with a DNA stain (blue) are shown. Images were taken with laser settings that were optimized separately for each sample. Scale bar, 10μm.

doi:10.1371/journal.pgen.1004876.g002

similar to that reported for the common non-coding variant GWAS signal at *GCK* [2]. *GCK* encodes the glycolytic enzyme glucokinase which catalyzes the reverse reaction to G6PC2. These observations are consistent with defects in glucose sensing rather than β-cell function.

Likely effector transcripts at established FG GWAS loci

Of the 12 coding variants significantly influencing FG levels, we have discussed eight above. The four remaining signals mapped to three additional established FG-associated GWAS regions and support the candidacy of particular effector transcripts at these loci. Two map to *PCSK1* (p.Gln665Glu, MAF = 27.9%, $P = 3.0\times10^{-8}$; p.Ser690Thr, MAF = 27.9%, $P = 4.1\times10^{-8}$), and have previously been proposed as likely causal variants at this GWAS locus because of strong LD with the lead non-coding GWAS SNP (rs4869272) in Europeans ($r^2 = 0.81$) [3]. These two variants are in complete LD with each other ($r^2 = 1$) and both have residual association signals after adjusting for the lead GWAS SNP (p.Gln665Glu, $P_{conditional} = 8.1\times10^{-3}$; p.Ser690Thr, $P_{conditional} = 0.01$) (Table 1 and S3 Table). Conversely, conditioning on the coding variants completely abolished the association signal at the GWAS SNP ($P_{unconditional} = 7.2\times10^{-7}$; $P_{conditional\ on\ Gln665Glu} = 0.24$, $P_{conditional\ on\ Ser690Thr} = 0.25$, and $P_{conditional\ on\ Gln665Glu\ and\ Ser690Thr} = 0.34$). Although we were unable to statistically distinguish between these two coding variants, there is suggestive *in vitro* evidence that the p.Gln665Glu variant may decrease PCSK1 activity whilst p.Ser690Thr behaved as wild-type [24], supporting the former as the most likely causal variant at the *PCSK1* locus. *PCSK1* encodes prohormone convertase 1/3 (PC1/3), a calcium-dependent serine endoprotease, which is essential for the conversion of a variety of prohormones, including proinsulin and proglucagon, to their bioactive forms.

The penultimate variant was in *ZHX3* (p.Asn310Ser, MAF = 23.8%, $P = 3.9\times10^{-7}$) which maps to the *TOP1* GWAS locus influencing FG levels [4]. Adjustment for p.Asn310Ser in a conditional analysis eliminated the association signal at the non-coding GWAS lead SNP rs6072275 ($P_{unconditional} = 2.5\times10^{-5}$ and $P_{conditional} = 0.33$), supporting *ZHX3* as the plausible causal gene at this locus (Table 1 and S3 Table). *ZHX3* (zinc fingers and homeoboxes 3) encodes a transcriptional repressor which belongs to a protein family known to regulate gene expression in the kidney podocytes and plays roles in both lipoprotein metabolism and triglyceride regulation in mice [25, 26].

The final variant was in *RREB1* (p.Ser1554Tyr, MAF = 21.1%, $P = 8.4\times10^{-9}$) which resides in the *RREB1*/*SSR1* GWAS locus influencing FG levels [4] and T2D risk [27] (Table 1). We were unable to explore the relationship between this association signal and the lead FG (rs17762454) or T2D (rs9502570) GWAS SNPs at this locus as neither the variants themselves, nor a close proxy ($r^2>0.80$), was present on the exome-array. Based on European haplotypes from the 1000 Genomes Project, the coding variant was more strongly correlated ($r^2 = 0.59$) with rs9502570 as compared to rs17762454 ($r^2 = 0.08$), which might indicate multiple FG association signals in this region. These data point to a likely functional role for *RREB1* at this locus, although further studies are needed to confirm this hypothesis.

Coding variants influencing FI levels

Turning to the analysis of FI, we identified six non-synonymous variants associated at exome-wide significance (one rare and five common). These mapped to four loci, three of them previously implicated in FI regulation (S3 Table). The single novel FI- influencing locus was

represented by a rare variant in *URB2* (p.Glu594Val, MAF = 0.1%, $P = 3.1\times10^{-7}$; Table 1 and S1 Fig.). The variant allele was observed in individuals across Europe, including UK, Denmark, Finland, and Sweden, and genotype calls across all cohorts passed visual inspection for clustering accuracy. Heterozygous genotypes (n = 6) were 100% concordant for 3,999 individuals genotyped on the array that had also been exome sequenced (S8 Table). This variant has a relatively large effect, with each copy of the minor allele increasing FI level by 32%. *URB2* encodes ribosome biogenesis 2 homolog, which interacts with nuclear lamins [28]. The precise mechanistic role of *URB2* is still poorly understood, but conditions caused by defects in lamin genes (laminopathies), including familial partial lipodystrophy, can cause loss of adipose tissue, insulin resistance, and metabolic syndrome [29, 30].

The remaining five coding variants implicated in the FI analysis, included three previously-reported FI-associated common variants in *GCKR* (p.Pro446Leu, $P = 8.1\times10^{-11}$), *PPARG* (p.Pro12Ala, MAF = 14.7%, $P = 1.3\times10^{-7}$), and *COBLL1* (p.Asn939Asp, MAF = 11.3%, $P = 6.7\times10^{-8}$) [3, 4]. The final two variants (in *C2orf16* and *GPN1*) also mapped within the *GCKR* FG/FI-associated GWAS locus, and their associations were eliminated after conditioning on the *GCKR* p.Pro446Leu variant as seen in FG analysis (S3 Table).

Discussion

In the current study, we have examined the contribution of coding variants in influencing glycemic trait levels and identified significant associations at ten loci, highlighting multiple functional genes.

We have established a functional role for *G6PC2* in FG homeostasis and have identified two novel loci associated with glycemic traits: *GLP1R* and *URB2*. In three additional GWAS regions for FG, we have identified potentially causal coding variants, highlighting likely effector transcripts (*PCSK1*, *RREB1*, and *ZHX3*). In line with earlier observations, very few of the FG- and FI-associated loci identified impact T2D risk [4]. The exception in this study is the p.Val219Leu *G6PC2* variant, which showed a modest but directionally consistent association with T2D risk, in contrast to earlier reports suggesting that, paradoxically, glucose-raising alleles at this GWAS locus are protective against T2D [2].

Our results have a number of important implications for the design, analysis, and interpretation of association studies of coding variation. First, we have demonstrated that an appreciation of regional haplotypes is fundamental to understanding the relationship between genetic variation and gene function, phenotype, and disease. In this study, haplotype analysis was essential in resolving the apparent discrepancy between the observed phenotypic effect driven by the *G6PC2* C allele at the variant encoding Leu219 and our *in vitro* data. From a clinical genomics perspective, this example illustrates how interpretation of single-variant analyses can be misleading and that the explicit phase relationships of variants are important for the correct interpretation of allelic effects. Furthermore, they highlight the importance of considering the possible interactions between regulatory and coding variants on protein levels [31]. The haplotype analysis also allows us to define specific regional diploid haplotypes that differ markedly in glucose levels. This is particularly relevant to efforts to follow up these variants in human data, whether for *in vivo* physiology, or for cellular studies on biopsies or patient derived stem cells.

Second, we have highlighted some limitations of widely used *in silico* prediction programs in accurately evaluating the impact of genomic variation on protein function. Earlier studies have shown that filtering functional variants based on *in vitro* data can substantially improve the power to detect causal genes [32, 33]. When such data are not available, *in silico* predictions can be used to help identify variants of functional significance [10, 34]. However, as we have

demonstrated with the *G6PC2* p.Val219Leu variant, current *in silico* predictions are not always accurate and should be validated with *in vitro* functional analysis wherever possible.

We find little support for a widespread impact of rare or low-frequency coding variants in accounting for known common SNP FG- and FI-associated GWAS signals. However, given that exome-array genotyping does not capture the complete repertoire of coding variation throughout the genome, exome sequencing is still required to achieve a complete assessment of their impact on glycemic traits at all associated loci. It is also likely that analysis of non-coding regulatory variation will be necessary to elucidate the causal genes and functional roles of association signals observed in prior GWAS. However, our study provides clear examples of how the analysis of coding variation can aid interpretation of GWAS findings where biology was previously unclear, and highlights the promise of this approach to provide insights into the pathophysiology of common complex disease.

Data availability

Summary statistics of single-variant and gene-based analyses are available at http://www.diagram-consortium.org/Mahajan_2014_ExomeChip/

Materials and Methods

Ethics statement

All human research was approved by the relevant institutional review boards, and conducted according to the Declaration of Helsinki and all patients provided written informed consent. FIN-D2D 2007, DPS, DR's EXTRA, FINRISK 2007, FUSION, and METSIM were approved by the University of Michigan Health Sciences and Behavioral Sciences Institutional Review Board (ID: H03-00001613-R2). The Danish studies (Health 2006, Inter99, and Vejle Biobank) were approved by the local Ethical Committees of Capital Region (approval # H-3-2012-155, KA 98155 and KA-20060011) and Region of Southern Denmark (approval # S-20080097). The GoDARTS study was approved by EoS REC 09/S1402/44. The Twins UK study was approved by EC04/015. The OBB study was a approved by South Central, Oxford C, 08/H0606/107+5, IRAS project 136602. The PIVUS study is approved by 00–419 and ULSAM study by 251/90 and 2007/338. The PPP study was approved by the Committee On the Use of Humans as Experimental Subjects at MIT (IRB 0912003615).

Phenotypes

Study participants. FG was measured in mmol/L and FI in pmol/L. Individuals were excluded from the analysis if they had a physician diagnosis of diabetes, were on diabetes treatment (oral or insulin), or had a fasting plasma glucose concentration $\geq$7 mmol/L or $\geq$11.1 mmol/L following a 2-hour oral glucose tolerance test. Individual studies applied further sample exclusions, including for pregnancy, nonfasting measurements, and type 1 diabetes (S1 Table). Measures of fasting glucose made in whole blood were corrected to approximate plasma level by multiplying by 1.13 [35].

Trait transformations and adjustment. To achieve approximate normality of the traits within each study, FG and natural logarithm–transformed FI levels were adjusted for age, sex, BMI, and study specific covariates followed by inverse normalization of the residuals. Inverse normalized residuals were used as the dependent quantitative trait in genetic association models to calibrate type 1 error. Effect estimates were obtained using untransformed FG and natural logarithm–transformed FI levels after adjusting for age, sex, BMI, and study specific covariates.

Genotyping and quality control

The Illumina HumanExome Beadchip array was genotyped by individual studies. This custom array was designed to facilitate large-scale genotyping of 247,870 mostly rare (MAF<0.5%) and low-frequency (MAF 0.5–5%) protein altering variants selected from sequenced exomes and genomes of ~12,000 individuals (see URLs). Details of genotype calling and quality control are presented in S1 Table. To confirm the genotyping quality of all the variants discussed in the manuscript, we compared the genotype calls in 2,312 to 4,000 individuals for which we have both exome-array genotypes and exome-sequence data (S8 Table). The results are highly concordant for all variants (99% heterozygous genotype concordance and 100% concordance observed for non-reference homozygotes). Call rate and Hardy-Weinberg equilibrium p-values for each variant are provided in S9 Table.

Association analysis

Single-variant analysis. We tested single variants for association with FG- and FI-derived inverse normalized residuals assuming an additive genetic model using a linear mixed model to account for relatedness with EMMAX [5]. Study-specific QQ plots and genomic lambdas are shown in the S3 Fig. We repeated single-variant association analyses with untransformed FG and natural logarithm–transformed FI residuals to obtain effect estimates. We then combined the association summary statistics across studies by using a fixed-effects meta-analysis (sample size Z-score weighted) using METAL [36]. Genotype cluster plots of all variants described here were inspected visually in all studies.

Single-variant conditional analysis. Conditional analysis was performed to identify additional association signals at known or novel loci. The analysis included the genotypes of the lead variant(s) at the conditioning loci as covariate(s) in the regression analysis in EMMAX. We then performed meta-analysis of the association summary statistics across studies by using a fixed-effects meta-analysis (sample size Z-score weighted).

Gene-based analysis. For gene-based testing, we first computed single-variant score statistics and their covariance matrices for all polymorphic markers within each study. We then combined the single-variant score statistics from all studies using the Cochran-Mantel-Haenszel method and computed corresponding variance-covariance matrices centrally [6]. These variance-covariance matrices were used to compute gene-level statistics. We applied a frequency-weighted burden test which assumes variants have similar effect sizes and SKAT, a dispersion test that performs well when both protective and deleterious variants are present [8, 9]. Test-specific asymptotic distributions were used to evaluate significance. For gene-based analyses, we used only unrelated individuals (n = 32,223 for FG and n = 29,848 for FI) and included principal components as covariates to adjust for population structure. We generated four variant lists using frequency data and functional annotation. Variants were mapped to transcripts in Ensembl 66 (GRCh37.66). Using annotations from CHAoS v0.6.3, SnpEff v3.1, and VEP v2.7, we identified variants predicted to be protein-truncating (PTV; e.g. nonsense, frameshift, essential splice site) or protein-altering (e.g. missense, in-frame indel, non-essential splice site) in at least one mapped transcript (by at least one of the three algorithms) [37, 38]. We additionally used the procedure described by Purcell *et al.* to identify subsets of missense variants meeting "strict" or "broad" criteria for being deleterious, using annotation predictions from Polyphen2-HumDiv, PolyPhen2-HumVar, LRT, MutationTaster, and SIFT [10]. Masks 1 ("PTV-only") and 3 ("PTV + NS_{strict}") are restricted to variants with predicted major effect on protein function, and, as a result, disproportionately favor inclusion of rare variants, whilst masks 2 ("PTV + missense") and 4 ("PTV + NS_{broad}") are more permissive. In total, we tested 1,028, 14,465, 4,603, and 13,093 genes having at least two such variants in the above four

categories respectively. For gene-based tests reaching exome-wide significance, if the conditional analysis showed that the signal was driven by a single variant, we required the variant to achieve exome-wide significance in the single-variant test as well.

Gene-based conditional analysis. We performed conditional gene-level analysis to evaluate whether rare or low-frequency variants associated with the trait in single-variant analysis could account for or were due to a gene-based test association signal [6]. The genotypes of the variant(s) at the conditioning locus were included as covariate(s) in this analysis.

Estimating phenotypic variance explained by SNPs. We used a subset of Finnish samples (FIN-D2D 2007, DPS, DR's EXTRA, FINRISK 2007, and METSIM; n = 10,266) to calculate variance explained by *G6PC2*. We ran a model regressing BMI adjusted FG on the four *G6PC2* variants: intronic GWAS lead SNP rs560887, p.His177Tyr (rs138726309), p.Tyr207Ser (rs2232323), and p.Val219Leu (rs492594), assuming an additive model (and adjusting for sex, age, age^2, and study origin). A separate model was run excluding the GWAS SNP to determine the additional variance captured with the three coding variants.

Power calculations. We had >99.9% power to identify variants that explain >0.3% of the phenotypic variance and 80% power to detect coding variants that explain >0.1% of the phenotypic variance. To achieve >80% power for variants with MAF<0.05% would require effect sizes of at least 1 SD unit of residuals of mmol/L for FG and pmol/L for FI.

When we estimate power to detect association for aggregation tests, we make many assumptions: i) proportion of variants contributing to trait variation, ii) direction of effects, iii) number of variants aggregated, and iv) allele frequency distribution of the variants [34]. In this study, assuming 100% of the variants contribute to trait variation, we had >99.9% power to detect association for genes that explain >0.25% of the phenotypic variance and 80% to detect genes that explain >0.1% of the phenotypic variance using a burden test and >0.5% of the phenotypic variance using SKAT. In a less favorable scenario, for example assuming a gene explains 0.25% of the phenotypic variance, 25% of the variants contribute to trait variation, different directions of effect, 20 variants tested, and variants sampled from the reported allele frequency distribution in the exome-array design, power to detect association in this study may be near 0% for a burden test and 63% for SKAT [34].

Haplotype analysis. In 4,442 individuals from the Oxford Biobank, we used an expectation-maximization (EM) algorithm to obtain the posterior distribution of haplotypes consistent with the observed genotypes at four *G6PC2* variants: intronic GWAS lead SNP rs560887, p.His177Tyr (rs138726309), p.Tyr207Ser (rs2232323), and p.Val219Leu (rs492594). Haplotype association with FG- and FI-derived residuals (after adjustment for age, sex, and BMI) was tested in a linear regression framework, as a function of haplotype dosage from posterior distribution, and including principal components as covariates to account for population structure using the most frequent haplotype as baseline.

***In-silico* mutation analysis.** SIFT [39], PolyPhen-2 [40], and Condel [41] algorithms were used to predict the functional effects of the associated non-synonymous variants on protein function. Genomic Evolutionary Rate Profiling (GERP) [42] scores were calculated to indicate the degree of evolutionary conservation at a given human nucleotide based on multiple genomic sequence alignments and were measured as site-specific 'rejected substitutions': higher scores indicate greater conservation.

Functional studies

Site directed mutagenesis. Human *G6PC2* cDNA (NM_021176.2) within a pCMV6-Entry vector (with a C-terminal FLAG-tag) was purchased from OriGene (RC211146). We generated non-synonymous variants in the *G6PC2* coding sequence of the clone using Quikchange II

Site-Directed Mutagenesis (Agilent). All mutations were verified by Sanger sequencing; in each case, only the desired nucleotide substitutions was introduced.

Western blot analyses. HEK293 and INS-1E 832–13 cells were transfected with each FLAG-tagged wild type or mutant G6PC2 construct using Lipofectamine 2000 (Invitrogen). In protein degradation assays, cells were treated with 10μM MG-132 (Calchembio) or 100μM chloroquine (Sigma) for 15h. At 48h after transfection, cells were collected and homogenized by sonication in lysis buffer. Total cellular protein was separated by 4–12% SDS-PAGE (Invitrogen) and transferred to nitrocellulose membranes. We determined G6PC2 expression by immunoblotting using a mouse monoclonal FLAG M2 antibody (Sigma, F1804). A rabbit polyclonal antibody against β tubulin (Santa Cruz, sc-9104) was used as a loading control. Secondary antibodies specific to mouse or rabbit IgG were obtained from Thermo Fisher Scientific. Protein bands were detected using the ECL reagent (Pierce Thermo Fisher Scientific). Western blots were quantified by densitometry analysis using ImageJ and paired t tests of densitometric data were carried out in GraphPad Prism.

Immunofluorescence. HEK293 cells were transfected with each FLAG-tagged G6PC2 construct using FuGene 6 transfection reagent (Promega) in 4-well chamber slides (BD Biosciences). After 48h, cells were fixed with 4% paraformaldehyde in PBS for 15 min. Cells were permeabilized with 0.05% Triton X-100 in PBS for 15 min, and blocked for 1 h with 10% BSA in PBS-Tween 20. We carried out double immunostaining of cells using FLAG M2 (Sigma, F1804) and calnexin (Santa Cruz, sc-11397) primary antibodies followed by anti-mouse fluorescein (Vector Labs) and anti-rabbit TRITC (Dako). DRAQ5 fluorescent probe (Thermo Fisher Scientific) was applied at 20μM as a far-red DNA stain. Slides were mounted with Vectashield mounting medium (Vector Labs) and visualized on a BioRad Radiance 2100 confocal microscope with a 60× 1.0 N.A. objective. Images were acquired with different laser settings that were optimized for each sample and therefore fluorescent intensities are not comparable across samples.

URLs

Exome-array design,
http://genome.sph.umich.edu/wiki/Exome_Chip_Design

EPACTS,
http://genome.sph.umich.edu/wiki/EPACTS

METAL,
http://www.sph.umich.edu/csg/abecasis/Metal/download/

RareMETALS,
http://genome.sph.umich.edu/wiki/RareMETALS

The 1000 Genomes Project,
www.1000genomes.org

Supporting Information

S1 Fig. Forest plots of association results with fasting glucose (FI) and fasting insulin (FG) levels. Individual study and meta-analysis effects of A) *URB2* coding variant rs141203811 (p.Glu594Val) on FI; B) *GLP1R* coding variant rs10305492 (p.Ala316Thr) on FG; and C), D), & E) *G6PC2* coding variants rs138726309 (p.His177Tyr), rs2232323 (p.Tyr207Ser), rs492594 (p.Val219Leu) on FG.
(PDF)

S2 Fig. Regional association plots for FG at *G6PC2* region on chromosome 2. (A) Unconditional association results highlight the previously known non-coding lead SNP rs560887. (B) Association results after conditioning on rs560887 highlight two non-synonymous coding variants rs138726309 (p.His177Tyr) and rs492594 (p.Val219Leu), both largely independent from the signal from rs560887. (C) Association results conditioning on rs560887 (GWAS SNP) and rs138726309 (p.His177Tyr) highlights rs492594 (p.Val219Leu) as an independently associated variant at *G6PC2*. (D) Association results conditioning on rs560887 (GWAS SNP) and rs492594 (p.Val219Leu) highlights rs138726309 (p.His177Tyr) as the second independently associated coding variant at *G6PC2*
(PDF)

S3 Fig. Quantile-quantile plots (qq-plot) for each study included in the meta-analysis for FG and FI. Genomic control (GC) is given in each qq-plot.
(PDF)

S1 Table. Characteristics of the study cohorts.
(DOCX)

S2 Table. Sample characteristics of the study cohorts.
(DOCX)

S3 Table. Coding variants achieving exome-wide significant association with FG and FI levels in single-variant analysis. MAF: minor allele frequency. $\hat{\beta}$: regression coefficient estimates. SE: standard error. N: number of samples analyzed. I^2: heterogeneity measure in %. *P*_het: P-value for Cochran's Q statistic. [a]Sample-size weighted average minor allele frequency percentage across all studies.
(DOCX)

S4 Table. *In silico* predictions of the impact of the non-synonymous changes on protein function and the evolutionary conservation of variants. GERP: Genomic Evolutionary Rate Profiling
(DOCX)

S5 Table. Conditional analysis of genes associated with FG and FI levels at exome-wide significance in gene-based tests using SKAT and BURDEN. ∗ Although *AKT2* reached exome-wide significance, the signal was driven by a single variant which showed only suggestive significance in single-variant analysis (P=9.3×10^{-7}). As described in the methods, we required the single-variant test to be exome-wide significant (P<5×10^{-7}) in such a scenario
(DOCX)

S6 Table. FG single-variant association results of the rare and low-frequency G6PC2 coding variants included in the gene-based tests based on unrelated samples. MAF: minor allele frequency; Allele counts: Minor allele counts. [a]Alleles are aligned to the forward strand of NCBI Build 37. [b]Sample-size weighted average minor allele frequency percentage across all studies. *P* values are obtained with derived inverse normalized residuals of mmol/L of fasting glucose after adjustment for age, sex, and BMI. Direction of effect is aligned to the minor allele.
(DOCX)

S7 Table. Association with T2D, birth weight and diabetes-related quantitative traits. m: Minor allele; M: Major allele; BMI: Body mass index; WHR: Waist-hip ratio; FG: Fasting glucose level; FI: Fasting insulin level, adjusted for BMI; HbA1c: Hemoglobin A1-C level; HOMA-B: Homeostasis model assessment-B score; HOMA-IR: Homeostasis model assessment-insulin resistance; SBP: Systolic blood pressure; DBP: Diastolic blood pressure; TG:

Triglycerides; HDL-C: HDL cholesterol; LDL-C: LDL cholesterol; TC: Total cholesterol; BW: Birth weight. (a) Trait increasing allele in bold. (b) Samples contributing to T2D case-control analysis include a subset of the non-diabetic samples contributing to the current analysis. (c) All published data is "unconditioned". As seen in our analysis, direction of effect switches after conditioning on rs560887 at least for FG. For exome-chip FG and T2D we have provided results from conditional analysis and G is the trait increasing allele.
∗ rs2232323 (*G6PC2*), rs138726309 (*G6PC2*), rs35742417 (*RREB1*), and rs141203811 (*URB2*) have not been investigated in earlier GWAS
(DOCX)

S8 Table. Genotype concordance of all the variants discussed in the manuscript. Concordance is computed for the variants with non-missing calls in both exome-array and exome sequence data. N: Number of samples.
(DOCX)

S9 Table. Study-wise call rate and Hardy-Weinberg equilibrium P-values of all the variants discussed in the manuscript. CR: call rate. HWE: Hardy-Weinberg equilibrium p-value.
(DOCX)

Acknowledgments

We gratefully acknowledge the contribution of all participants from the various studies that contributed to this work.

Members of T2D-GENES consortium and GoT2D consortium
Hanna E Abboud[1], Uzma Afzal[2], David Aguilar[3], Rector Arya[1], Gil Atzmon[4], Tin Aung[5, 6], Eric Banks[7], Inês Barroso[8, 9], Nir Barzilai[4], Jennifer E Below[10], Dwaipayan Bharadwaj[11], Thomas W Blackwell[12], Lori L Bonnycastle[13], Don Bowden[14–16], Jason Carey[7], Mauricio O Carneiro[7], John C Chambers[2, 17, 18], Edmund Chan[19], Juliana Chan[20–22], Giriraj R Chandak[23], Peng Chen[24], Yuhui Chen[25], Han Chen[26, 27], Ching-Yu Cheng[5, 6, 24, 28], Kee Seng Chia[24], Yoon Shin Cho[29], Adolfo Correa[30], Joanne E Curran[31], Mark J Daly[32], Aaron G Day-Williams[8], Ralph A DeFronzo[1], Mark DePristo[7], Peter J Donnelly[25, 33], Shah B Ebrahim[34], Paul Elliott[2, 35], Tõnu Esko[7, 36–38], João Fadista[39], Yossi Farjoun[7], Andrew J Farmer[40], Vidya S Farook[31], Timothy Fennell[7], Teresa Ferreira[25], Tasha Fingerlin[41], Tom Forsén[42, 43], Sharon P Fowler[1], Paul W Franks[44–46], Timothy M Frayling[47], Barry I Freedman[48], Philippe Froguel[49], Eric R Gamazon[50], Christian Gieger[51], Benjamin Glaser[52], Min Jin Go[53], Jacqueline I Goldstein[7, 32], Harald Grallert[54–56], George Grant[7], Todd Green[7], Michael Griswold[57], Daniel Esten Hale[58], Bok-Ghee Han[53], Christopher Hartl[7], Andrew T Hattersley[59], Pamela J Hicks[14–16], Dylan Hodgkiss[60], Momoko Horikoshi[25, 61], Martin Hrabé de Angelis[62], Cheng Hu[63], Frank B Hu[44, 64], Iksoo Huh[65], Mohammad Kamran Ikram[5, 6], Thomas Illig[55, 66], Kathleen A Jablonski[67], Christopher P Jenkinson[1, 68], Weiping Jia[63], Hyun Min Kang[12], Chiea-Chuen Khor[5, 6, 24, 69, 70], Yongkang Kim[65], Young Jin Kim[53], Bong-Jo Kim[53], Leena Kinnunen[71], Jaspal Singh Kooner[72], Jasmina Kravic[39], Jennifer Kriebel[55, 56], Ashish Kumar[25, 73], Satish Kumar[31], Teemu Kuulasmaa[74], Min-Seok Kwon[75], Claudia Langenberg[76], Torsten Lauritzen[77], Selyeong Lee[65], Jaehoon Lee[65], Juyoung Lee[53], Jong-Young Lee[78], Donna M Lehman[1], Benjamin Lehne[2], Jonathan C Levy[61], Jiang Li[14], Liming Liang[27, 64], Wei Yen Lim[24], Keng-Han Lin[12], Jianjun Liu[24, 70], Marie Loh[2, 79], Ronald C W Ma[20–22], Clement Ma[12], Reedik Mägi[37], Jared Maguire[7], Taylor J Maxwell[10], Gilean McVean[25], Christa Meisinger[54], Thomas Meitinger[80–82], Olle Melander[83], Andres Metspalu[37], Evelin Mihailov[37], Lili Milani[37], Loukas Moutsianas[25], Martina Müller-Nurasyid[51, 80, 84, 85], Solomon K Musani[86], Yoshihiko Nagai[87–89], Narisu Narisu[13], Benjamin M Neale[7, 32],

Maggie C Y Ng[14, 15], Peter Nilsson[90], Stephen P O'Rahilly[9], Marju Orho-Melander[91], Katharine R Owen[61, 92], Nicholette D Palmer[14–16], Taesung Park[65, 75], Dorota Pasko[47], Richard D Pearson[25], John R B Perry[25, 47, 76], Annette Peters[54, 56, 80], Toni I Pollin[93], Ryan Poplin[7], Dorairaj Prabhakaran[34], Sobha Puppala[31], Shaun Purcell[7, 94, 95], Lu Qi[44, 96], Qibin Qi[44, 97], Michael Roden[98], Olov Rolandsson[45], Anders H Rosengren[39], Manjinder Sandhu[8, 99], Thomas Schwarzmayr[82], Laura J Scott[12], Robert A Scott[76], James Scott[72], William R Scott[2], Jobanpreet Sehmi[17, 72], Khalid Shakir[7], Rob Sladek[87, 88, 100], Joshua D Smith[101], Alena Stančáková[74], Konstantin Strauch[51, 85], Tim M Strom[81, 82], Amy Swift[13], E Shyong Tai[19, 24, 102], Juan Fernandez Tajes[25], Sian-Tsung Tan[17, 72], Nikhil Tandon[103], Herman A Taylor Jr[30], Yik Ying Teo[24, 104, 105], Farook Thameem[1], Barbara Thorand[54, 56], Martijn van de Bunt[25, 61], Tibor V Varga[46], Mark Walker[106], Nicholas J Wareham[76], Ryan P Welch[12], Thomas Wieland[82], Gregory Wilson Sr[107], Tien Yin Wong[5, 6], Andrew R Wood[47], Joon Yoon[75], Eleftheria Zeggini[8], Weihua Zhang[2, 17]

1. Department of Medicine, University of Texas Health Science Center, San Antonio, Texas, USA.
2. Department of Epidemiology and Biostatistics, Imperial College London, London, UK.
3. Cardiovascular Division, Baylor College of Medicine, Houston, Texas, USA.
4. Departments of Medicine and Genetics, Albert Einstein College of Medicine, New York, USA.
5. Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, National University Health System, Singapore.
6. Singapore Eye Research Institute, Singapore National Eye Centre, Singapore.
7. Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, USA.
8. Department of Human Genetics, Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, UK.
9. Metabolic Research Laboratories, Institute of Metabolic Science, University of Cambridge, Cambridge, UK.
10. Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, Texas, USA.
11. Functional Genomics Unit, CSIR-Institute of Genomics & Integrative Biology (CSIR-IGIB), New Delhi, India.
12. Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan, USA.
13. Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA.
14. Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA.
15. Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA.
16. Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA.

17. Department of Cardiology, Ealing Hospital NHS Trust, Southall, Middlesex, UK.
18. Imperial College Healthcare NHS Trust, Imperial College London, London, UK.
19. Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, National University Health System, Singapore.
20. Hong Kong Institute of Diabetes and Obesity, The Chinese University of Hong Kong, Hong Kong, China.
21. Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong, China.
22. Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Hong Kong, China.
23. CSIR-Centre for Cellular and Molecular Biology, Hyderabad, Andhra Pradesh, India.
24. Saw Swee Hock School of Public Health, National University of Singapore, National University Health System, Singapore.
25. Wellcome Trust Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK.
26. Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA.
27. Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, USA.
28. Centre for Quantitative Medicine, Office of Clinical Sciences, Duke-NUS Graduate Medical School Singapore, Singapore.
29. Department of Biomedical Science, Hallym University, Chuncheon, Republic of Korea.
30. Department of Medicine, University of Mississippi Medical Center, Jackson, Mississippi, USA.
31. Department of Genetics, Texas Biomedical Research Institute, San Antonio, Texas, USA.
32. Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA.
33. Department of Statistics, University of Oxford, Oxford, UK.
34. Centre for Chronic Disease Control, New Delhi, India.
35. MRC-PHE Centre for Environment and Health, Imperial College London, London, UK.
36. Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA.
37. Estonian Genome Centre, University of Tartu, Tartu, Estonia.
38. Division of Endocrinology, Boston Children's Hospital, Boston, Massachusetts, USA.
39. Department of Clinical Sciences, Diabetes and Endocrinology, Lund University Diabetes Centre, Malmö, Sweden.
40. Department of Primary Care Health Sciences, University of Oxford, Oxford, UK.
41. Department of Epidemiology, Colorado School of Public Health, University of Colorado, Aurora, Colorado, USA.

42. Department of General Practice and Primary Health Care, University of Helsinki, Helsinki, Finland.

43. Vasa Health Care Center, Vaasa, Finland.

44. Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts, USA.

45. Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden.

46. Department of Clinical Sciences, Lund University Diabetes Centre, Genetic and Molecular Epidemiology Unit, Lund University, Malmö, Sweden.

47. Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter, UK.

48. Department of Internal Medicine, Section on Nephrology, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA.

49. Genomics and Molecular Physiology, CNRS (Institut de Biologie de Lille), Lille, France.

50. Department of Medicine, Section of Genetic Medicine, The University of Chicago, Chicago, Illinois, USA.

51. Institute of Genetic Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany.

52. Endocrinology and Metabolism Service, Hadassah-Hebrew University Medical Center, Jerusalem, Israel.

53. Center for Genome Science, Korea National Institute of Health, Chungcheongbuk-do, Republic of Korea.

54. Institute of Epidemiology II, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany.

55. Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany.

56. German Center for Diabetes Research (DZD), Neuherberg, Germany.

57. Center of Biostatistics and Bioinformatics, University of Mississippi Medical Center, Jackson, Mississippi, USA.

58. Department of Pediatrics, University of Texas Health Science Center, San Antonio, Texas, USA.

59. University of Exeter Medical School, University of Exeter, Exeter, UK.

60. Department of Twin Research and Genetic Epidemiology, King's College London, London, UK.

61. Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, UK.

62. Institute of Experimental Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany.

63. Department of Endocrinology and Metabolism, Shanghai Diabetes Institute, Shanghai Jiao Tong University Affiliated Sixth People's Hospital, Shanghai, China.

64. Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA.

65. Department of Statistics, Seoul National University, Seoul, Republic of Korea.

66. Hannover Unified Biobank, Hannover Medical School, Hannover, Germany.

67. The Biostatistics Center, The George Washington University, Rockville, Maryland, USA.

68. Research, South Texas Veterans Health Care System, San Antonio, Texas, USA.

69. Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore, National University Health System, Singapore.

70. Division of Human Genetics, Genome Institute of Singapore, A∗STAR, Singapore.

71. Diabetes Prevention Unit, National Institute for Health and Welfare, Helsinki, Finland.

72. National Heart and Lung Institute, Cardiovascular Sciences, Hammersmith Campus, Imperial College London, London, UK.

73. Chronic Disease Epidemiology, Swiss Tropical and Public Health Institute, University of Basel, Basel, Switzerland.

74. Faculty of Health Sciences, Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland, Kuopio, Finland.

75. Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul, Republic of Korea.

76. MRC Epidemiology Unit, Institute of Metabolic Science, University of Cambridge, Cambridge, UK.

77. Department of Public Health, Section of General Practice, Aarhus University, Aarhus, Denmark.

78. Ministry of Health and Welfare, Seoul, Republic of Korea.

79. Institute of Health Sciences, University of Oulu, Oulu, Finland.

80. Deutsches Forschungszentrum für Herz-Kreislauferkrankungen (DZHK), Partner Site Munich Heart Alliance, Munich, Germany.

81. Institute of Human Genetics, Technische Universität München, Munich, Germany.

82. Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany.

83. Department of Clinical Sciences, Hypertension and Cardiovascular Disease, Lund University, Malmö, Sweden.

84. Department of Medicine I, University Hospital Grosshadern, Ludwig-Maximilians-Universität, Munich, Germany.

85. Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, Neuherberg, Germany.

86. Jackson Heart Study, University of Mississippi Medical Center, Jackson, Mississippi, USA.

87. McGill University and Génome Québec Innovation Centre, Montreal, Quebec, Canada.

88. Department of Human Genetics, McGill University, Montreal, Quebec, Canada.

89. Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada.

90. Department of Clinical Sciences, Medicine, Lund University, Malmö, Sweden.

91. Department of Clinical Sciences, Diabetes and Cardiovascular Disease, Genetic Epidemiology, Lund University, Malmö, Sweden.

92. Oxford NIHR Biomedical Research Centre, Oxford University Hospitals Trust, Oxford, UK.

93. Department of Medicine, Program in Personalized and Genomic Medicine, University of Maryland, Baltimore, Maryland, USA.

94. Center for Human Genetic Research, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA.

95. Department of Psychiatry, Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, USA.

96. Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA.

97. Department of Epidemiology and Population Health, Albert Einstein College of Medicine, New York, USA.

98. Institute of Clinical Diabetology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University, Düsseldorf, Germany.

99. Department of Public Health and Primary Care, Institute of Public Health, University of Cambridge, Cambridge, UK.

100. Division of Endocrinology and Metabolism, Department of Medicine, McGill University, Montreal, Quebec, Canada.

101. Department of Genome Sciences, University of Washington School of Medicine, Seattle, Washington, USA.

102. Cardiovascular & Metabolic Disorders Program, Duke-NUS Graduate Medical School Singapore, Singapore.

103. Department of Endocrinology and Metabolism, All India Institute of Medical Sciences, New Delhi, India.

104. Department of Statistics and Applied Probability, National University of Singapore, Singapore.

105. Life Sciences Institute, National University of Singapore, Singapore.

106. The Medical School, Institute of Cellular Medicine, University of Newcastle, Newcastle, UK.

107. College of Public Services, Jackson State University, Jackson, Mississippi, USA.

Author Contributions

Conceived and designed the experiments: AMah XS HJN AMan MAR HMH AEL NG HKI GA GIB JB NJC RD CLH MS JGW RMW IP JD JCF APM DA JBM MB MIM CML ALG. Performed the experiments: AMah XS HJN AMan NG NWR NRR NLB JKR PC JF PF CF KJG

TMT JBJ CL CB DB GB NPB SG APG CJG MH JRH AUJ GJ JMJ MM JM MN RO KSS HMS ACS JTr CC IB RR GLS ASFD LLa AL BI TT MEJ TJ JK MU VS TDS ADM CNAP FSC KLM RNB EI LLi JTu TH FK LG ML OP APM DA MB MIM CML ALG. Analyzed the data: AMah XS HJN AMan MAR HMH AEL NG HKI APM JBM MB MIM CML ALG. Wrote the paper: AMah XS HJN AEL APM JBM MB MIM CML ALG. Steering committee overseeing the consortia: GA GIB JB NJC RD CLH KLM MS JGW JCF DA MB MIM.

References

1. Prokopenko I, Langenberg C, Florez JC, Saxena R, Soranzo N, et al. (2009) Variants in MTNR1B influence fasting glucose levels. Nat Genet 41: 77–81. doi: 10.1038/ng.290 PMID: 19060907
2. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, et al. (2010) New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet 42: 105–116. doi: 10.1038/ng.520 PMID: 20081858
3. Manning AK, Hivert M-F, Scott RA, Grimsby JL, Bouatia-Naji N, et al. (2012) A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nat Genet 44: 659–669. doi: 10.1038/ng.2274 PMID: 22581228
4. Scott RA, Lagou V, Welch RP, Wheeler E, Montasser ME, et al. (2012) Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nat Genet 44: 991–991005. doi: 10.1038/ng.2385 PMID: 22885924
5. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong S-Y, et al. (2010) Variance component model to account for sample structure in genome-wide association studies. Nat Genet 42: 348–354. doi: 10.1038/ng.548 PMID: 20208533
6. Liu DJ, Peloso GM, Zhan X, Holmen OL, Zawistowski M, et al. (2014) Meta-analysis of gene-level tests for rare variant association. Nat Genet 46: 200–204. doi: 10.1038/ng.2852 PMID: 24336170
7. Feng S, Liu D, Zhan X, Wing MK, Abecasis GR (2014) RAREMETAL: fast and powerful meta-analysis for rare variants. Bioinformatics. doi: 10.1093/bioinformatics/btu367 PMID: 24894501
8. Wu MC, Lee S, Cai T, Li Y, Boehnke M, et al. (2011) Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet 89: 82–93. doi: 10.1016/j.ajhg.2011.05.029 PMID: 21737059
9. Magi R, Kumar A, Morris AP (2011) Assessing the impact of missing genotype data in rare variant association analysis. BMC Proc 5 Suppl 9. doi: 10.1186/1753-6561-5-S9-S107 PMID: 22373025
10. Purcell SM, Moran JL, Fromer M, Ruderfer D, Solovieff N, et al. (2014) A polygenic burden of rare disruptive mutations in schizophrenia. Nature 506: 185–190. doi: 10.1038/nature12975 PMID: 24463508
11. Vaxillaire M, Cavalcanti-Proenca C, Dechaume A, Tichet J, Marre M, et al. (2008) The common P446L polymorphism in GCKR inversely modulates fasting glucose and triglyceride levels and reduces type 2 diabetes risk in the DESIR prospective general French population. Diabetes 57: 2253–2257. doi: 10.2337/db07-1807 PMID: 18556336
12. Beer NL, Tribble ND, McCulloch LJ, Roos C, Johnson PRV, 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. doi: 10.1093/hmg/ddp357 PMID: 19643913
13. Rees MG, Wincovitch S, Schultz J, Waterstradt R, Beer NL, et al. (2012) Cellular characterisation of the GCKR P446L variant associated with type 2 diabetes risk. Diabetologia 55: 114–122. doi: 10.1007/s00125-011-2348-5 PMID: 22038520
14. Drucker DJ, Nauck MA (2006) The incretin system: glucagon-like peptide-1 receptor agonists and dipeptidyl peptidase-4 inhibitors in type 2 diabetes. Lancet 368: 1696–1705. doi: 10.1016/S0140-6736(06)69705-5 PMID: 17098089
15. Service SK, Teslovich TM, Fuchsberger C, Ramensky V, Yajnik P, et al. (2014) Re-sequencing expands our understanding of the phenotypic impact of variants at GWAS loci. PLoS genetics 10: e1004147. doi: 10.1371/journal.pgen.1004147 PMID: 24497850
16. Khajavi M, Inoue K, Lupski JR. (2006) Nonsense-mediated mRNA decay modulates clinical outcome of genetic disease. Eur J Hum Genet 14: 1074–81. doi: 10.1038/sj.ejhg.5201649 PMID: 16757948
17. Chen W-M, Erdos MR, Jackson AU, Saxena R, Sanna S, et al. (2008) Variations in the G6PC2/ABCB11 genomic region are associated with fasting glucose levels. J Clin Invest 118: 2620–2628. doi: 10.1172/JCI34566 PMID: 18521185
18. Bouatia-Naji N, Rocheleau G, Van Lommel L, Lemaire K, Schuit F, et al. (2008) A polymorphism within the G6PC2 gene is associated with fasting plasma glucose levels. Science 320: 1085–1088. doi: 10.1126/science.1156849 PMID: 18451265

19. Baerenwald DA, Bonnefond A, Bouatia-Naji N, Flemming BP, Umunakwe OC, et al. (2013) Multiple functional polymorphisms in the G6PC2 gene contribute to the association with higher fasting plasma glucose levels. Diabetologia 56: 1306–1316. doi: 10.1007/s00125-013-2875-3 PMID: 23508304

20. Pound LD, Oeser JK, O'Brien TP, Wang Y, Faulman CJ, et al. (2013) G6PC2: a negative regulator of basal glucose-stimulated insulin secretion. Diabetes 62: 1547–1556. doi: 10.2337/db12-1067 PMID: 23274894

21. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segre AV, et al. (2012) Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet 44: 981–990. doi: 10.1038/ng.2383 PMID: 22885922

22. Wang H, Liu L, Zhao J, Cui G, Chen C, et al. (2013) Large scale meta-analyses of fasting plasma glucose raising variants in GCK, GCKR, MTNR1B and G6PC2 and their impacts on type 2 diabetes mellitus risk. PLoS One 8: e67665. doi: 10.1371/journal.pone.0067665 PMID: 23840762

23. Hu C, Zhang R, Wang C, Ma X, Fang Q, et al. (2009) A genetic variant of G6PC2 is associated with type 2 diabetes and fasting plasma glucose level in the Chinese population. Diabetologia 52: 451–456. doi: 10.1007/s00125-008-1241-3 PMID: 19082990

24. Pickett LA, Yourshaw M, Albornoz V, Chen Z, Solorzano-Vargas RS, et al. (2013) Functional consequences of a novel variant of PCSK1. PLoS One 8: e55065. doi: 10.1371/journal.pone.0055065 PMID: 23383060

25. Liu G, Clement LC, Kanwar YS, Avila-Casado C, Chugh SS (2006) ZHX proteins regulate podocyte gene expression during the development of nephrotic syndrome. J Biol Chem 281: 39681–39692. doi: 10.1074/jbc.M606664200 PMID: 17056598

26. Gargalovic PS, Erbilgin A, Kohannim O, Pagnon J, Wang X, et al. (2010) Quantitative trait locus mapping and identification of Zhx2 as a novel regulator of plasma lipid metabolism. Circ Cardiovasc Genet 3: 60–67. doi: 10.1161/CIRCGENETICS.109.902320 PMID: 20160197

27. 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. doi: 10.1038/ng.2897 PMID: 24509480

28. Vlcek S, Foisner R, Wilson KL (2004) Lco1 is a novel widely expressed lamin-binding protein in the nuclear interior. Exp Cell Res 298: 499–511. doi: 10.1016/j.yexcr.2004.04.028 PMID: 15265697

29. Hegele RA, Anderson CM, Wang J, Jones DC, Cao H (2000) Association between nuclear lamin A/C R482Q mutation and partial lipodystrophy with hyperinsulinemia, dyslipidemia, hypertension, and diabetes. Genome Res 10: 652–658. doi: 10.1101/gr.10.5.652 PMID: 10810087

30. Mesa JL, Loos RJF, Franks PW, Ong KK, Luan Ja, et al. (2007) Lamin A/C polymorphisms, type 2 diabetes, and the metabolic syndrome: case-control and quantitative trait studies. Diabetes 56: 884–889. doi: 10.2337/db06-1055 PMID: 17327461

31. Lappalainen T, Montgomery SB, Nica AC, Dermitzakis ET (2011) Epistatic selection between coding and regulatory variation in human evolution and disease. Am J Hum Genet 89: 459–463. doi: 10.1016/j.ajhg.2011.08.004 PMID: 21907014

32. Rees MG, Ng D, Ruppert S, Turner C, Beer NL, 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. doi: 10.1172/JCI46425 PMID: 22182842

33. Bonnefond A, Clement N, Fawcett K, Yengo L, Vaillant E, et al. (2012) Rare MTNR1B variants impairing melatonin receptor 1B function contribute to type 2 diabetes. Nat Genet 44: 297–301. doi: 10.1038/ng.1053 PMID: 22286214

34. Zuk O, Schaffner SF, Samocha K, Do R, Hechter E, et al. (2014) Searching for missing heritability: designing rare variant association studies. Proc Natl Acad Sci U S A 111: 455–464. doi: 10.1073/pnas.1322563111 PMID: 24443550

35. D'Orazio P, Burnett RW, Fogh-Andersen N, Jacobs E, Kuwa K, et al. (2005) Approved IFCC recommendation on reporting results for blood glucose (abbreviated). Clin Chem 51: 1573–1576. doi: 10.1373/clinchem.2005.051979 PMID: 16120945

36. Willer CJ, Li Y, Abecasis GR (2010) METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26: 2190–2191. doi: 10.1093/bioinformatics/btq340 PMID: 20616382

37. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, et al. (2010) Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics 26: 2069–2070. doi: 10.1093/bioinformatics/btq330 PMID: 20562413

38. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, et al. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 6: 80–92. doi: 10.4161/fly.19695

39. Kumar P, Henikoff S, Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4: 1073–1081. doi: 10.1038/nprot.2009.86 PMID: 19561590

40. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, et al. (2010) A method and server for predicting damaging missense mutations. Nat Methods 7: 248–249. doi: 10.1038/nmeth0410-248 PMID: 20354512

41. Gonzalez-Perez A, Lopez-Bigas N (2011) Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. Am J Hum Genet 88: 440–449. doi: 10.1016/j.ajhg.2011.03.004 PMID: 21457909

42. Cooper GM, Stone EA, Asimenos G, Green ED, Batzoglou S, et al. (2005) Distribution and intensity of constraint in mammalian genomic sequence. Genome Res 15: 901–913. doi: 10.1101/gr.3577405 PMID: 15965027