EXTENDING GENOME-WIDE ASSOCIATION STUDY DATA THROUGH ANALYSIS OF SNPs IN THE *G6PC2* GENE

By

Devin August Baerenwald

Thesis

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

In

Molecular Physiology and Biophysics

May, 2012

Nashville, Tennessee

Approved:

Maureen Gannon, Ph.D.

Roland Stein, Ph.D.

ACKNOWLEDGEMENTS

The work presented in this document was made possible by the funding received by Dr. Richard O'Brien from the NIH, as well as the Molecular Endocrinology Training Program training grant. In addition to his money, I am also indebted to Dr. O'Brien for his time and patience in directing me through countless experiments as well as graduate school in general. I would also be remiss to neglect thanking Dr. Maureen Gannon, the chair of my thesis committee, and Dr. Roland Stein for their time and aid in preparing this document and assisting me throughout my time at Vanderbilt.

I must also acknowledge those with whom I have spent my time in the laboratory. Without them, my time at work would have failed to be anywhere near as fruitful, entertaining, or enjoyable, and their contributions to both my science and my life outside of graduate school should not go unacknowledged.

Finally, I must recognize and thank my friends and family. As a source of scientific assistance or social reprieve, my friends have helped me in innumerable ways. Most importantly, however, my parents Phil and Ginger and brother Derek have been supportive and helpful no matter the path I ultimately took, and it is with their help that I reached this point at all.

ACKNOWLEDGEMENT OF SUPPORT

Funding for this research has been provided to Dr. O'Brien through the National Institute of Health (grant number DA027002). I have been supported by the Molecular Endocrinology Training Grant (NIH 5T32 DK07563).

PRIOR PUBLICATIONS

Some of the material included in this thesis has been published. Analyses of rs13431652 and rs573225 in Chapter III were published in *Diabetes* (Bouatia-Naji N, Bonnefond A, Baerenwald DA, Marchand M, Bugliani M, Marchetti P, Pattou F, Printz RL, Flemming BP, Umunakwe OC, Conley NL, Vaxillaire M, Lantieri O, Balkau B, Marre M, Lévy-Marchal C, Elliott P, Jarvelin M, Meyre D, Dina C, Oeser JK, Froguel P, O'Brien RM. 2010 Oct; 59: 2662-2671).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Glucose Homeostasis and the Glucose-6-Phosphatase Enzyme System

Glucose is a molecule critical to myriad organ systems throughout the body. In order to maintain optimal function of these systems, glucose must circulate in the bloodstream within a rather narrow concentration range. Homeostatic mechanisms exist to return the concentration to the normal range when deviations occur. Significant contributions to maintaining this balance are made by the liver whether the body is fed (postprandial) or fasted (preprandial). In a postprandial state when glucose is abundant, the liver and muscle can simply store glucose in its long-chain form of glycogen, or it can be converted to glucose-6-phosphate (G6P) and metabolized via glycolysis or the pentose phosphate pathway. Such actions serve to keep circulating blood glucose levels from rising too high above normoglycemia. The opposite, low glucose abundance, is found in the preprandial state; here the liver must produce glucose to reintroduce to the bloodstream by degrading glycogen stores or from precursors to gluconeogenesis. These include glycerol, amino acids, and lactate, and without these liver homeostatic mechanisms, organ systems such as the brain and central nervous system would cease to operate properly.

These two processes, glycogenolysis and gluconeogenesis, are each tightly regulated by their own array of hormones and metabolites, but converge at G6P, which is hydrolyzed in the final step of glucose production in the liver to glucose and inorganic

phosphate. The responsible enzyme was first theorized due to the necessity for removal of the phosphate from G6P following glycogenolysis to create glucose. Glucose-6-phosphatase activity was eventually demonstrated in the endoplasmic reticulum (ER) of hepatocytes, but the enzyme was unable to be isolated at this time due to its incorporation in the ER membrane and the instability inherent in membrane-bound proteins [1, 2]. Finally, cloning of the cDNA encoding the glucose-6-phosphatase catalytic subunit (G6pc) was accomplished by screening a murine liver cDNA library with representative probes of mRNA populations from wild type or mutant mice possessing deletions at the locus for albinism associated with reduced gluconeogenic enzyme activity [3-5]. This resulted in isolation of murine G6pc cDNA [4, 6].

Demonstration of the hydrolysis of glucose-6-phosphate in the ER membrane of hepatocytes, paired with the discovery of the G6pc cDNA, led to the development of the currently accepted model of glucose-6-phosphatase as a multi-component enzyme system that comprises a catalytic subunit and transporters for both substrate and product (Fig. 1.1) [7, 8]. The catalytic subunit of this so-called Arion substrate-transport hypothesis, named after the group responsible for the microsomal analyses, lays in the membrane of the ER, with an inward, lumenal facing active site, while the opposite orientation is found in the fetus [9]. This model requires the presence of transporters to facilitate movement of substrate and products across the ER membrane; one such molecule, responsible for relocation of G6P to the ER lumen, was eventually identified and termed the G6P translocase (G6PT). Elucidation of this component of the glucose-6-phosphatase holoenzyme was made possible due to homology between mouse and human cDNAs and a known bacterial G6P transporter in *Escherichia coli*, UhpT [10]. Previous experiments

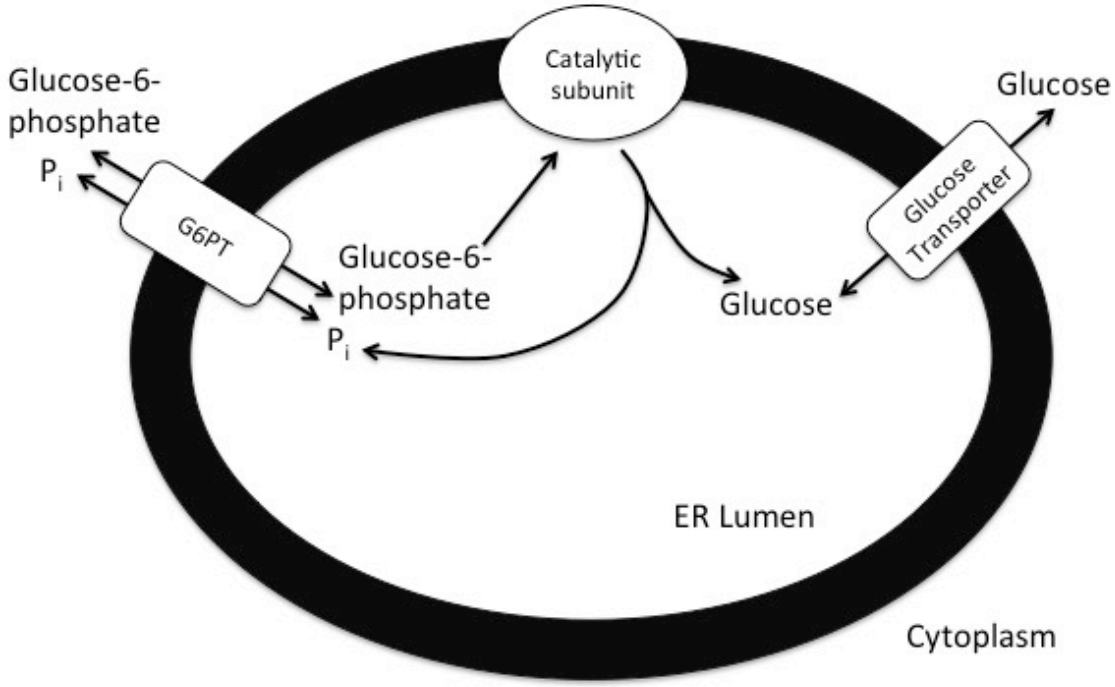


Figure 1.1. Arion model of the glucose-6-phosphatase multicomponent enzyme system

have shown antiporter activity by UhpT [11], as well, lending credence to the hypothesis that G6PT serves to export inorganic phosphate product in addition to introducing G6P to the lumen of the ER [12]. At this time, the final component of the model, the glucose transporter, has not been identified.

Disease States and Glucose-6-Phosphatase

In-depth investigation of this enzyme system is important due to association with human disease of either an increase or decrease in glucose-6-phosphatase activity. A deficiency in the activity of glucose-6-phosphatase results in multiple autosomal recessive disorders called glycogen storage disease (GSD) [13]. Persons suffering from GSD1a, caused by mutations in G6PC, experience fasted state hypoglycemia, increased glycogen deposition in liver and kidney, and other symptoms [14]. The remaining GSD variants, GSD1b and GSD1c, result from mutations in G6PT and demonstrate similar symptoms with an additional increased vulnerability to bacterial infection [15]. A basis for this susceptibility has been linked to dysfunctional neutrophils, which express G6PT but not G6PC, suggesting yet another role for the translocase.

A link between increased activity of glucose-6-phosphatase and disease has also been demonstrated, namely with type 2 and in poorly controlled cases of type 1 diabetes. These diseases are characterized by a multitude of symptoms, including elevated hepatic glucose production and ineffective peripheral glucose disposal, which result in hyperglycemia. Increased glucose-6-phosphatase activity in conjunction with increased mRNA levels has also been observed [16-18]. The reduced insulin signaling indicative of the diabetic state may be the cause, as an insulin responsive transcriptional unit has been

identified upstream of the transcriptional start site of *G6PC*, including a hepatocyte nuclear factor-1 (HNF-1) binding site from -231 to -199 [19] and two FOXO1 binding sites between -198 and -159 [20, 21]. While HNF-1 is considered to act as a stabilizer for FOXO1 binding [22, 23], direct targeting of FOXO1 by insulin signaling occurs [24-31]. In the context of *G6PC*, FOXO1 serves as a transcriptional activator, thus its exclusion from the nucleus following insulin activation results in decreased transcription of *G6PC*. Deficient signaling through the insulin receptor in type 2 diabetes allows for increased *G6PC* transcription and a worsened state of hyperglycemia.

Isoforms of the Glucose-6-Phosphatase Gene Family

Initial interest in the activity of glucose-6-phosphatase and its role on glucose production was focused on the liver, though it has now been shown to be expressed and influence gluconeogenesis in other tissues as well, such as the kidney and small intestine. Glucose-6-phosphatase activity has also been observed in microsomes isolated from the β-cells of the pancreas, but at roughly one tenth that of liver microsomes [32]. While experiments have shown that rat islets express a protein capable of being bound by antibodies specific to glucose-6-phosphatase [33], the islet isoform displays a dissimilar enzymatic profile, namely K_m value, pH optima, and inhibitor profile [32, 34]. These factors imply the presence of an islet-specific isoform or a tissue-specific regulatory factor capable of changing the kinetics of the liver enzyme. Upon the cloning of a cDNA from a mouse insulinoma library encoding a protein of similar weight and 50% identity to G6pc, it became apparent that a separate isoform existed. Further experiments established that this isoform was selectively expressed in the pancreatic Islets of Langerhans,

prompting the name islet-specific glucose-6-phosphatase catalytic subunit-related protein (IGRP), now commonly referred to as G6PC2 [32]. Based solely on its similarity in structure to G6PC, it would be predicted that G6PC2 would also be capable of hydrolyzing G6P; however, initial experiments using transient transfection of G6pc2 failed to validate this hypothesis [32, 35-37]. By slightly altering the protocol conditions to minimize the disturbance when permeabilizing the intracellular membrane to G6P [38], G6pc2 was shown to hydrolyze G6P, albeit at a lower K_m and V_{MAX} than G6pc (Fig. 1.2) [34].

The advent of computer-based genetics research and the sequencing of the human genome have played a massive role in the search for new genes of interest, and the glucose-6-phosphatase gene family has not been exempt. Using the sequence for *G6PC2* as the query, a search was carried out using the BLAST search engine from NCBI, returning yet a third isoform of the gene family sharing 36% identity with G6PC. Upon discovery of its widespread expression, the protein was first dubbed the ubiquitously expressed glucose-6-phosphatase catalytic subunit-related protein (UGRP), but is now known as G6PC3 [39]. Although demonstrating functional hydrolysis of G6P by G6PC3 also initially proved difficult [39], glucose-6-phosphatase activity was eventually demonstrated through overexpression of G6PC3 by stable [40] or adenoviral [41] transfection. The defining characteristics of each isoform are illustrated in Table 1.1.

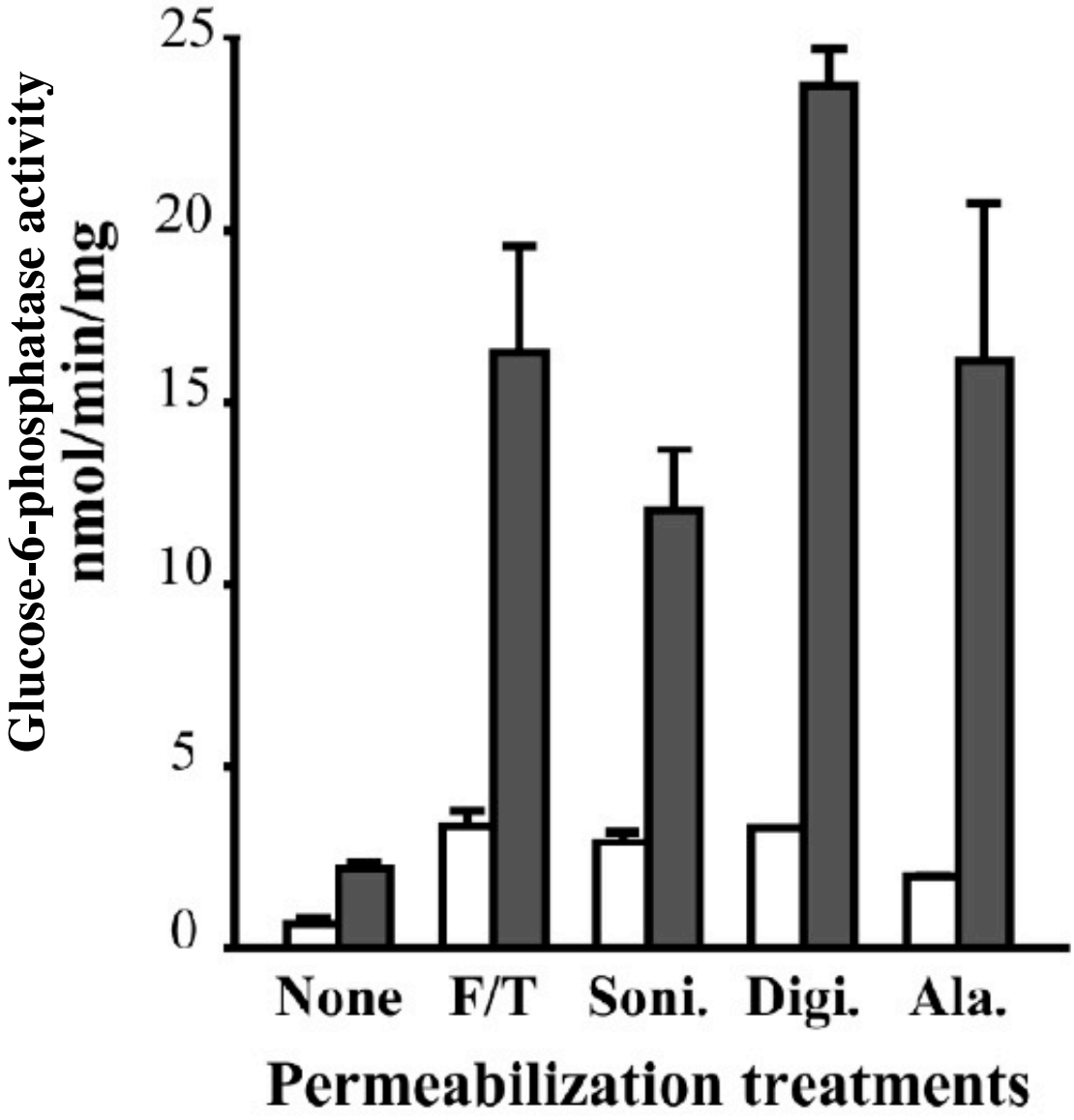


Figure 1.2. Glucose-6-phosphatase activity observed by G6PC2 in permeabilized cells

Using COS7 cells under a number of membrane permeabilization conditions, a measure of the glucose-6-phosphatase activity of G6PC2 was obtained. Cells were transfected with either a negative control pcDNA3.1 (white bars) or pcDNA3.1 containing the human G6PC2 (grey bars). Following transfection and a 48 hour incubation, the cells were harvested, pelleted, resuspended, and permeabilized with the listed technique. Each of these methods allows for limited disruption of the internal membranes of the cells, thus allowing for subsequent radiochemical assaying of glucose-6-phosphatase activity. F/T = freeze thawing; Soni. = sonication; Digi. = 0.1% digitonin for 15 min on ice; Ala. = 0.02% alamethicin for 15 min on ice. Results are means ± S.E.M. [38]

Table 1.1. The glucose-6-phosphatase catalytic subunit gene family

Gene	G6Pase	Islet-Specific G6Pase-Related Protein (IGRP)	Ubiquitous G6Pase Related Protein (UGRP)
	G6PC	G6PC2	G6PC3
Tissue	Liver	Islet β-cell	Ubiquitous
Size	357 aa	355 aa	346 aa
% Identity	100	50	36
Chromosome	17q21	2	17q21
Location	ER	ER	ER
#Transmembrane domains	9	9	9
Substrate	G6P	G6P	G6P

Islet-Specific Glucose-6-Phosphatase Activity

As was previously noted, expression and activity of a glucose-6-phosphatase have been detected in organs in addition to the liver, namely the β-cells of the pancreas [42]. The resounding perception regarding the β-cells of the pancreas prior to this discovery held glucokinase to be the primary glucose sensor [43-46]. A new model for glucose stimulated insulin secretion (GSIS) has now been developed, incorporating this islet-specific glucose-6-phosphatase activity as a regulator of insulin secretion [47]. GSIS occurs as the concentration of glucose in the blood rises above normoglycemia, usually in the postprandial state. As this occurs, the glucose concentration outside the β-cell rises above that of the cytosol, allowing passive transport of glucose into the β-cell through GLUT2. Glucose is then phosphorylated by glucokinase and continues metabolism by entering glycolysis and the tricarboxylic acid (TCA) cycle (Fig. 1.3). The resulting products include ATP, increasing the concentration of ATP in the β-cell as well as the ratio of ATP to ADP. This shift causes the ATP-sensitive potassium (K_{ATP}) channels to close, inhibiting potassium inflow and depolarizing the membrane in the process. In response to this depolarization, voltage-gated calcium channels open to allow influx of calcium, which, when combined with intracellular calcium stores, activates the exocytotic machinery and facilitates fusion of insulin vesicles with the plasma membrane for secretion.

Changes in insulin secretion must correspond appropriately to the varying concentrations of circulating glucose, and this process is inextricably linked due to the control exhibited by glucose metabolism on insulin exocytosis. As mentioned above, glucokinase [48] had been thought to act as the β-cell sensor for glucose with

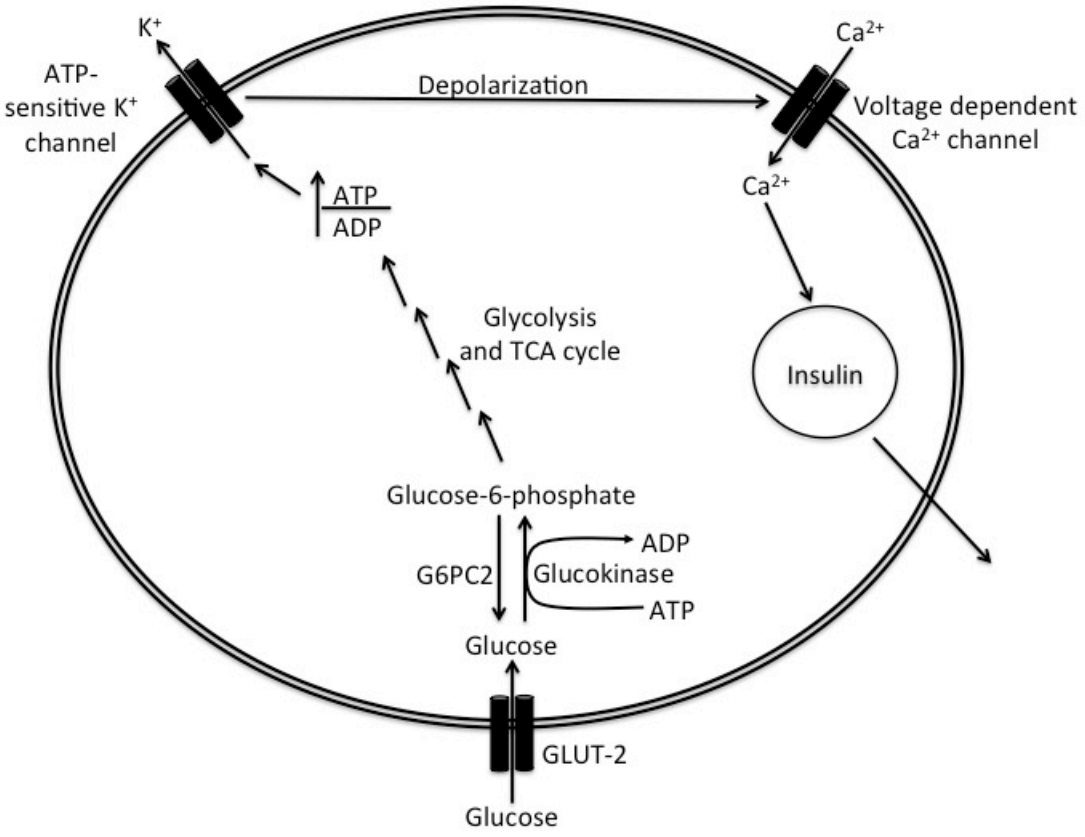


Figure 1.3. Futility of the glucose-6-phosphatase in the pancreatic β-cell
The opposition of the activity of glucokinase in the β-cell by a glucose-6-phosphatase in the form of G6PC2 creates a futile cycle of ATP hydrolysis without replenishment. Such a depletion would decrease insulin secretion and, in turn, increase the concentration of plasma glucose.

experimental evidence in multiple models supporting this [43-46]. Previous considerations of GLUT1 and GLUT2 as potential glucose sensors [49, 50] were dismissed after neither a significant increase in GLUT1 nor a reduction of GLUT2 expression by 90% affected GSIS in transgenic mice [45, 51]. However, the recent details discovered regarding the islet-specific glucose-6-phosphatase activity within the β-cells confound the supported model. The presence of an enzyme such as this counteracts the action of glucokinase and creates a futile cycle in the β-cell with a net product of hydrolyzed ATP between the two enzyme functions. By decreasing the pool of G6P available for GSIS, the ratio of ATP to ADP would also decrease and insulin secretion would be attenuated. The potential role of an islet-specific glucose-6-phosphatase as a negative component of the glucose sensor in the β-cell creates another point of regulation for maintaining normoglycemia through GSIS. Such a cycle has been seen to exist in an experimental setting using isolated islets from mice and rats, with varying percentages of glucose dephosphorylation occurring based on the characteristics of the animals from which the islets were isolated [52-54]. However, it has not yet been determined whether this result is indicative of variations in islet-specific glucose-6-phosphatase activity *in vivo*.

Exploring the Role of G6pc2 *in vivo*

In an attempt to clarify the role of an islet-specific glucose-6-phosphatase, phenotypic analyses of *G6pc2* knockout (KO) mice were carried out on a mixed genetic background [55], followed by pure genetic background analyses (Lynley Pound, Richard O'Brien, unpublished data). In both models, FPG was significantly reduced in both male

and female mice. A slight but significant reduction in fasting plasma cholesterol was also observed in the male and female KO mice on a pure C57BL/6J genetic background, consistent with a known association between levels of fasting glucose and cholesterol. A hypothesis was developed from these results, in which G6pc2 regulates GSIS as an inhibitory glucose sensor component in the β-cell. A futile substrate utilization cycle for ATP in this model would lead to a reduced ratio of ATP to ADP and decreased insulin secretion. Removal of G6pc2 from this system would thus allow unhindered GSIS to occur, resulting in a leftward shift of the $S_{0.5}$ of GSIS, without a change in V_{MAX}, and decreased FPG (Fig. 1.4).

Numerous experiments were carried out to analyze GSIS in *G6pc2* KO mice, including pancreas perfusions, intraperitoneal glucose tolerance tests (IPGTTs), and islet isolations (Lynley Pound, Richard O'Brien, unpublished data). Insulin secretion in response to various glucose concentrations was analyzed *in situ* from perfused pancreata. A sub-maximal glucose challenge of 6.5 mM led to secretion of ~252% more insulin in the pancreata of *G6pc2* KO mice than WT pancreata, data that is consistent with the previously mentioned leftward shift in the dose response curve for GSIS in *G6pc2* KO mice. In addition to the perfusion studies, IPGTTs were performed using *G6pc2* KO and WT mice injected with a submaximal dose of 0.75 g/kg body weight dose of glucose. Mice deficient for *G6pc2* had reduced blood glucose, but showed unchanged glucose tolerance and insulin secretion. Further support for a leftward shift in the GSIS dose response curve, but not a reduction in V_{MAX}, was observed in static incubations of islets isolated from *G6pc2* KO mice when compared to WT. Increases in GSIS of

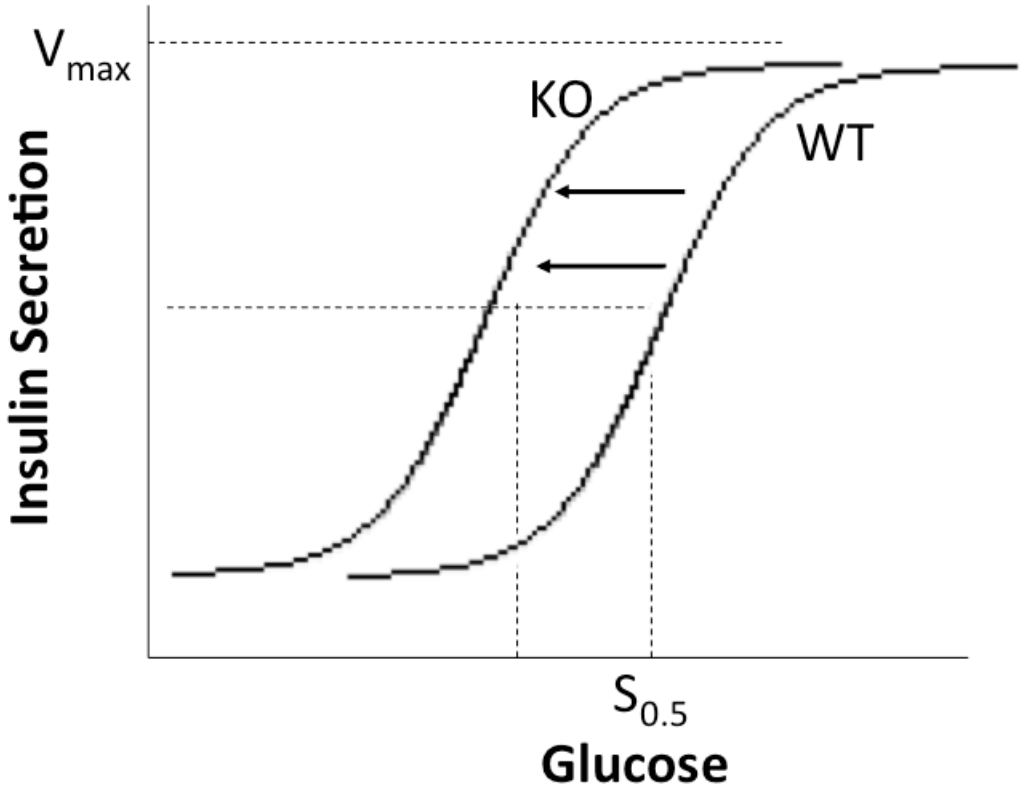


Figure 1.4. Hypothesized shift of GSIS response curve in animals lacking *G6pc2*
Due to the proposed role of G6PC2 as a negative component of the glucose sensor in β-cells, a leftward shift in the GSIS dose-response curve is expected to occur following deletion of *G6pc2* in mice. Support for this model was observed through pancreas perfusions, IPGTTs, and islet isolations in *G6pc2* KO mice.

approximately 2.2 fold were observed at both 5 mM and 11 mM glucose, while no difference in insulin secretion was observed at a maximal dose of 16.7 mM glucose.

To better describe this shift in the response curve of GSIS, the sequestration of calcium into the ER of mice lacking G6pc2 was assessed. It has been previously suggested that one of the products of G6P hydrolysis in the ER lumen, inorganic phosphate, may be critical to the calcium sequestration process by forming calcium phosphate [56]. As was previously mentioned, these intracellar calcium stores combined with the influx of calcium following membrane depolarization form the basis of the activation of insulin exocytosis. In order to test this hypothesis, islets were isolated from male *G6pc2* WT and KO mice and subsequently treated with thapsigargin (Lynley Pound, Richard O'Brien, unpublished data). Treatment with this compound results in the emptying of calcium from the sarco- and endoplasmic reticula of cells by inhibiting their corresponding calcium ATPases. Following thapsigargin addition, islets isolated from *G6pc2* KO mice exhibited increased intracellular calcium emptying, indicating that these mice may have diminished endoplasmic reticulum-specific calcium retention.

These results regarding retention of calcium suggest that deletion of *G6pc2* would negatively impact sequestration of calcium in the ER. However, it was reasoned that any deficiency in β-cell function because of loss of G6pc2 would only be observed at high glucose concentrations because of the role of the sarco-/endoplasmic reticulum calcium ATPase (SERCA) pump in sequestering calcium in the ER [57, 58]. Therefore, IPGTTs were performed following a 6 hr fast using a 2.0 g/kg body weight dose of glucose in WT and *G6pc2* KO mice, in which a slight impairment in glucose tolerance was observed in KO mice. Because negligible differences were observed in insulin sensitivity between

WT and *G6pc2* KO mice through insulin tolerance tests, these data are indicative of a defect in β-cell function in *G6pc2* KO mice that arises in a setting of high glucose concentration and leads to aberrant glucose tolerance. Support for this hypothesis was observed when measurement of insulin secretion in WT and *G6pc2* KO mice following intraperitoneal injection of 2.0 g/kg body weight glucose yielded no change, indicating that mice deficient for G6pc2 fail to secrete sufficient insulin to regulate blood glucose levels. These data may explain the paradoxical GWA study data associating SNPs within *G6PC2* with reduced FPG as well as a reduction in insulin secretion during glucose tolerance tests, despite a lack of consistent changes in insulin sensitivity [59-61]. Such a decrease in secretion of insulin may denote a change in calcium oscillation-driven pulsatility of insulin secretion [60, 62], as well as the previously mentioned deficiency in sequestration of calcium in the ER.

Characterization of mice lacking *G6pc2* has provided beneficial insight into the role G6pc2 plays in GSIS. Another study has also analyzed the converse protein status using a transgenic NOD mouse model that overexpresses human *G6PC2* via the rat insulin promoter (RIP). However, this model was not informative because increased *G6PC2* expression resulted in augmented ER stress, possibly due to the formation of unfolded proteins, causing cell death and eventually diabetes [63].

The Role of G6PC2 in Type 1 Diabetes

Also related to variations in FPG and G6PC2 are recent investigations of type 1 diabetes and the immunology of the disease in the β-cell. In a mouse model of human type 1 diabetes, the non-obese diabetic (NOD) mouse [64], researchers found that G6PC2

acts as an autoantigen [65]; this was also observed in humans [66, 67] and "humanized mice" [66, 68, 69]. As the pathophysiology of type 1 diabetes has been mostly unspecified, such a finding could be immensely beneficial to dissecting the nature of the disease and determining the primary responsible autoantigen. However, it has since been suggested that proinsulin may be the true primary autoantigen, whose T cell response occurs prior to that of G6PC2 [70]. Thus studies in mice showed that overexpression of proinsulin, but not G6PC2, in antigen-presenting cells abrogated the onset of diabetes [70]. This result was supported by studies showing that NOD mice expressing a form of proinsulin shown to be non-immunogenic did not develop type 1 diabetes [71], while deletion of *G6pc2* from NOD mice had no effect on diabetes onset [72]. These data suggest that G6PC2 autoimmunity is not a critical step in the inception of type 1 diabetes in mice but do not rule out its potential role in disease advancement.

Islet-Specific Expression and Transcriptional Regulation of *G6PC2*

Using immunofluorescent techniques, the expression pattern of *G6pc2* was localized primarily to β-cells of the pancreatic islets, though a small subset of α-cells were also identified that appeared to express the gene [73]. As this expression profile places *G6pc2* in a rather select family of islet-specific genes, much work has been focused on determining the molecular foundation of this specificity. Analysis of *G6PC*-CAT fusion genes in β-cell lines implicated a portion of the proximal promoter, -306 to +3, as adequate to drive maximal expression of a reporter gene [74]. Further experiments were carried out to determine how this promoter region would respond *in vivo*. Though islet-specific expression of the transgene was initiated at the proper time in mice, its

promoter-driven activation was not sustained into the adult, implying a lack of critical promoter elements in the -306 to +3 region [73]. Continued exploration to elucidate other domains critical to transcription of *G6pc2* is underway [75].

Though expression of the transgene failed to persist in adult mice, the proximal (-306/+3) sequence still comprises an islet-specific promoter. Work was thus begun to identify the transcription factors that bind this region, with the hope of discovering a novel factor critical for islet-specific gene expression. Putative factors were determined using bioinformatics software as well as *in situ* footprinting to search for binding sites of transcription factors in the proximal promoter [75, 76]. Binding of multiple factors, including Pdx-1, Pax-6, MafA, Foxa2, BETA-2, and USF, was confirmed in the insulinoma cell line βTC-3 via chromatin immunoprecipitation (ChIP) assay and mutational analysis [77-79]. These factors also regulate the transcriptional activation of the -306/+3 promoter, and are important for pancreatic expression of other genes, including insulin [80-82].

Splicing and Processing of pre-mRNA

Following gene activation by these and other transcription factors, the actual process of transcription generates pre-mRNA. Protein is derived from mRNA, thus pre-mRNA must undergo splicing to reach this state. Pre-mRNA contains not only the sequences critical to proper protein formation in the form of exons but also introns that span the junctions between exons; *G6PC2* contains 5 exons, separated by 4 introns of varying lengths [74]. Previously thought to be completely extraneous, introns have been shown to occasionally contain distinct coding regions for other proteins, enhancer

elements to promote activation of a gene, or noncoding RNA molecules [83]. Both introns and exons contain canonical splice sites that are recognized by components of the splicing machinery, termed the spliceosome. The splice site sequences utilized in the removal of the primary class of introns include the 5' splice site, 3' splice site, branch point element, and intronic polypyrimidine tract [84]. The spliceosome is comprised of 5 RNA-protein complexes that are termed small nuclear ribonucleoproteins (snRNPs), which form distinct complexes throughout the course of splicing. The first complex formed (E) includes binding of the U1 snRNP to the 5' splice site; U2AF binding to the polypyrimidine tract as well as the 3' splice site through its 65 and 35 kDa subunits, respectively; and interaction of SF1 with the branch point sequence. In the second complex (A), ATP-dependent recruitment of U2 to the branch point sequence occurs via the 65 kDa subunit of U2AF. This is followed by binding of the U4/U6-U5 tri-snRNP to form the B complex, all of which are subsequently rearranged to form the catalytic complex (C). Upon rearrangement, base pairing between U4/U6 is disrupted to allow interaction of U6 with both the 5' splice site and U2, displacing U1 and U4 in the process. It is at this point that U5, the most highly conserved snRNP, is finally able to bridge the 5' and 3' splice sites and the two transesterification reactions necessary for intron excision can occur. The intron is first detached from the upstream exon and reattached to the internal branch point sequence, forming an exon-intron lariat structure; this is followed closely by exon ligation and intron debranching. The detached intron lariat is then degraded and the components of the spliceosome recycled [85].

These splice sites are surprisingly less highly conserved in higher eukaryotes than lower eukaryotes such as *Saccharomyces cerevisiae*, and it is through this lack of

conservation that alternative splicing products arise [86, 87]. In many cases, the pre-mRNA of a single gene can be spliced into multiple mRNA molecules each capable of forming discrete functional proteins; however, in other circumstances, the presence of another semi-conserved splice site can cause aberrant splicing. This form of alternative splicing can lead to formation of dysfunctional transcripts destined for nonsense mediated decay, or, in the situation that an aberrant splicing product does not contain a premature stop codon, potentially harmful protein isoforms [87].

Genetic Variation in *G6PC2* and Human Disease Association

A recently developed powerful tool, the genome wide association (GWA) study, has become the primary technique for deciphering intricate disease states and the genes involved. The previously used linkage analysis method identified only rare genetic variants passed through families that contribute large effect sizes to the condition [88], while the alternate candidate gene approach only explored genes known to be associated with the diseased system [89]. The former lacked the power to elucidate variations in genes playing a smaller role in a disease while the latter excluded genes previously unknown to be involved.

The sequencing of the human genome and affordable automated SNP genotyping encouraged the development of the GWA study methodology, a technique that enables the association of a disease or a disease characteristic with single nucleotide polymorphisms (SNPs) in ones genetic code. This in turn allows for a more unbiased identification of genes and gene variants that may contribute less drastically to a disease state. Due to the nature of the GWA study method and the low odds ratios accompanying

each variant, this technique may not yield definitive answers regarding SNPs and a disease state. Instead, a positive association of a SNP and a condition may imply a higher predisposition to said condition, increasing the potential for contribution by other genes or environmental influences to the state. Additionally, a common limitation of a GWA study is founded in the simple task of naming each polymorphism, usually for the gene with which it is most closely associated [89]. This can confound the understanding of SNP influences because a SNP contributed to a gene in close proximity may actually be acting in a distal manner on a completely disparate gene. A polymorphism may also be associated with a disease characteristic without being causative, instead being in high linkage disequilibrium (LD) with the true causative SNP [89]. Because of this, the result of a GWA study must be followed up by secondary experiments before confident association between a SNP and a phenotype can be made. Such is the foundation of the experiments in this thesis.

As cases of diabetes rapidly increase worldwide, the importance of better understanding its genetic origins also increases, to which GWA studies have proved a significant boon. Through this experimental approach, several SNPs have been associated with type 2 diabetes while another subset of approximately 21 SNPs is associated with traits relevant to glycemia [89]. Studies designed to associate SNPs with variations in FPG by GWA study were performed in healthy individuals with normoglycemia [43, 90-92]. It is from this group that a small number of SNPs found in proximity to *G6PC2* were selected for further consideration; these are listed in Table 1.2. One of the strongest associations was found with SNP rs560887, located in the third intron of *G6PC2*. The presence of each minor T allele at this locus is associated with an approximate reduction

Table 1.2. Loci identified in *G6PC2* associated with variations in FPG

Marker	Location of SNP	Allele (Major/Minor)	Frequency of glucose raising allele	Effect (mmol/L/allele)
rs560887	Intron 3	C/T	0.69	0.071
rs13431652	Distal promoter	A/G	0.68	0.075
rs573225	Proximal promoter	A/G	0.66	0.073
rs2232316	Proximal promoter	G/A	0.12	0.05

in FPG of 1mg/dL [43]. This SNP has also been shown to associate with changes in hepatic glucose production and secretion of insulin after introduction of a glucose load [59-61]. The SNPs rs13431652, rs573225, and rs2232316, located in the *G6PC2* promoter, also associate with relatively similar variations in FPG as rs560887.

While a change in FPG of 1mg/dL may seem rather trivial, various studies have indicated that this can play a significant role in ones health. One such study, based in a European population, determined that a rise in FPG from <90mg/dL to between 99 and 108mg/dL is associated with an increased risk of 30% for cardiovascular associated mortality (CAM) [93]. The inverse was also demonstrated, such that another study of an Asian population illustrated a 25% reduction in the risk of CAM due to a decrease in FPG from 99mg/dL to 90mg/dL [94]. In addition to CAM, variations in FPG can also prove a risk factor for other disorders, including type 2 diabetes [95]. Such associations with a disease state would appear to solidify the importance of G6PC2 in regulating FPG and its correlating risk for CAM.

Hypothesis on the Role of SNPs in *G6PC2* Regulation

While numerous GWA studies have been performed to associate genetic variants with disease states, very few studies have extended this analysis to determine the functional effect of SNPs at a molecular level. The discovery of the *G6PC2* SNPs within areas of high homology led to further examination of these regions and the formation of hypotheses to explain the association between these SNPs and variations in FPG. It was hypothesized that the SNPs within the promoter of *G6PC2* associated with variations in FPG affect the binding of transcription factors critical to the activation of the gene. This

was tested through creation of fusion gene promoter constructs containing the various SNP alleles and subsequent transfection of cell culture lines to allow for reporter gene expression analyses, the data of which is included in Chapters III and IV. Additionally, the rs560887 SNP was theorized to affect constitutive splicing of *G6PC2*, causing increased formation of aberrantly spliced, nonfunctional isoforms. An *in vitro* splicing plasmid backbone was obtained from Dr. Thomas Cooper that allowed the splicing event and SNP in question to be analyzed. These results are described in detail in Chapter V.

CHAPTER II

MATERIALS AND METHODS

Fusion Gene Plasmid Construction

rs573225 and rs13431652

The construction of mouse *G6pc2*-chloramphenicol acetyltransferase (CAT) fusion genes containing the wild-type promoter sequence from -306 to +3, or the same sequence with a site-directed mutation (SDM) of the Foxa2 binding site have been previously described. The construction of a human *G6PC2*-CAT fusion gene containing the wild-type promoter sequence from -324 to +3 has also been previously described. A three-step PCR strategy was used to introduce the rs573225-G allele or a two base pair mutation into the Foxa2 binding site in the human *G6PC2* promoter within the context of the -324 to +3 promoter fragment. All promoter fragments were subcloned into the pGL3 MOD and/or pGL4 MOD luciferase expression vectors. pGL3 MOD and pGL4 MOD were generated by replacing the polylinker in the pGL3-Basic or pGL4.16 vectors (Promega) with a polylinker containing the following restriction endonuclease recognition sites: *Kpn*I, *Bam*H1, *Hin*DIII, *Xba*I, *Xho*I, and *Bgl*II.

An 8,574 base pair *Mfe*I DNA restriction fragment extending from -8563 to +11 relative to the human *G6PC2* gene transcription start site was isolated from the previously described Pac 294 clone. This *Mfe*I DNA fragment was ligated into the *Eco*RI site of the pGEM7 vector (Promega) and the resulting plasmid was designated *G6PC2*Pac294MfepGEM7. The alleles of rs573225 and rs13431652 present within the

Pac 294 derived -8,563 to +11 *G6PC2* promoter fragment were determined through DNA sequencing. Two additional *G6PC2* subclones were then generated in pGEM7 so as to change the sequences of rs573225 and rs13431652 to their alternative alleles by site-directed mutagenesis. One of these subclones, designated *G6PC2*Pac294SacShuttlepGEM7, was generated by isolating a *Sac*I fragment from *G6PC2*Pac294MfepGEM7 that contained human *G6PC2* promoter sequence from -8,563 to -4,390. The second subclone, designated *G6PC2*Pac294XmaShuttlepGEM7, was generated by isolating an *Xma*I fragment from *G6PC2*Pac294MfepGEM7 that contained human *G6PC2* promoter sequence from -2,803 to +11. The *G6PC2*Pac294SacShuttlepGEM7 plasmid was used as a template along with mutagenesis primers 5'-TGTCAGGCAGGCTGTGTCACTGGAGGGAAG-3' and 5'-CTTCCCTCCAGTGACACAGCCTGCCTGACA-3' to convert rs13431652 at position -4,405 from T to C and the *G6PC2*Pac294XmaShuttlepGEM7 plasmid was used as a template with mutagenesis primers 5'-GGAAATGAACTGTACAAAAAATTTCAAGCAAACATGATCCAACTGTTC-3' and 5'-GAACAGTTGGATCATGTTTGCTTGAAATTTTTTGTACAGTTCATTTCC-3' to convert rs573225 at position -259 from A to G. Mutagenesis was performed with these templates and primers using a QuikChange II Kit (Stratagene) as described by the manufacturer.

G6PC2-luciferase reporter plasmids, containing promoter sequence from -8,563 to +11, were generated in the pGL3-Basic and pGL4.16 vectors (Promega) by a two-step process. In the initial cloning step, a *Sac*I – *Hin*DIII fragment was isolated from the *G6PC2*Pac294MfepGEM7 plasmid, using a *Sac*I site in *G6PC2* at -4,395 and a *Hin*DIII

site within the pGEM7 vector, and ligated into the *Sac*I – *Hin*DIII digested pGL3-Basic and pGL4.16 vectors to generate fusion gene constructs that contain *G6PC2* promoter sequence from -4,395 to +11. In the second cloning step a *Sac*I fragment was isolated from the *G6PC2*Pac294SacShuttlepGEM7 plasmid and ligated into the *Sac*I digested -4,395 pGL3-Basic and pGL4.16 fusion gene plasmids to extend the *G6PC2* promoter sequences from -4,395 to -8,563.

A *Sac*I fragment from the *G6PC2*Pac294SacShuttlepGEM7 plasmid containing the alternate rs13431652 allele and an *Xma*I fragment from the *G6PC2*Pac294XmaShuttlepGEM7 plasmid with the alternate rs573225 allele were then exchanged with the corresponding *Sac*I and/or *Xma*I fragments that contain the original Pac294 sequences to generate pGL3-Basic and pGL4.16 plasmids with the alternate SNP alleles within the context of the *G6PC2* promoter region from -8,563 to +11.

rs2232316

The originally isolated promoter contained the rs2232316-G and rs573225-A alleles [36]. The construction of human *G6PC2*-CAT [36] and human *G6PC2*-luciferase [96] fusion genes, both containing promoter sequence from -324 to +3, have been previously described. A three-step PCR strategy [78] was used to introduce the alternate rs2232316-A and/or rs573225-G alleles into the human *G6PC2* promoter within the context of the -324 to +3 promoter fragment.

Human *G6PC2* promoter fragments from -386 to +3 containing the alternate alleles of rs2232315 were generated by PCR using the following 5' primers: 5'-CCCAAG(-386)CTTAAATTAAAATGCAGGGCTTTATGGGCTATTTGACATG-3';

5’- CCCAAG(-386)CTTAAATTAAAATGCAGGACTTTATGGGCTATTTGACATG-3’. *Hind*III sites used for cloning purposes are underlined, as are the rs2232315 alleles at -368. A ~9 kbp *Xho*I – *Xho*I fragment sub-cloned from PAC294 that contains the *G6PC2* promoter and exons 1-4 was used as the template [36].

The 3’ primer was the same as that previously used to isolate the human *G6PC2* promoter sequence from -324 to +3 [36]: 5’- AACTG*C*AGTGCTCTGATTCCCACCG-3’. A *Pst*I site used for cloning purposes is underlined. This 3’ primer incorporates a single base pair change (italics) at position -1 in the human promoter that generates a *Pst*I site such that the subsequent sub-cloning of the PCR fragments create fusion gene constructs with the same 3’ polylinker sequence between position +3 and the reporter gene as found in previously studied mouse and rat fusion gene constructs [36]. The PCR fragments were digested with *Hind*III and *Pst*I and sub-cloned into *Hind*III – *Pst*I digested pSP72 (Promega). The promoter fragments were then re-isolated from the pSP72 plasmid as *Hind*III – *Bam*HI fragments and ligated into *Hind*III – *Bgl*II digested pCAT(An) expression vector. The promoter fragments were then re-isolated from the pCAT(An) plasmids as *Hind*III – *Xho*I fragments and sub-cloned into the pGL3 MOD luciferase expression vector [35]. pGL3 MOD was generated by replacing the polylinker in the pGL3-Basic vector (Promega) with a polylinker containing the following restriction endonuclease recognition sites: *Kpn*I, *Bam*HI, *Hind*III, *Xba*I, *Xho*I, and *Bgl*II. Promoter fragments generated by PCR were completely sequenced to ensure the absence of polymerase errors. All plasmid constructs were purified by centrifugation through cesium chloride gradients [97].

rs560887, rs2232321, Exon 4 Splice Junction, and rs35259259

The RHCglo minigene vector was a generous gift from Dr. T. Cooper (Baylor College of Medicine). A *G6PC2* fragment containing exon 4 and ~300bp of 5' and 3' intronic sequence was isolated using PCR with the PAC 299 plasmid as the template in conjunction with the following primers: 5'-ACGCGTCGACGGTTCATTCTTGATCATTCTGG-3' and 5'-GCTCTAGACTCAGAGTCATTAAGTTCCCTC-3' (*Sal*I and *Xba*I cloning sites underlined). The PCR fragment was digested with *Sal*I and *Xba*I and ligated into the *Sal*I-*Xba*I digested RHCglo vector. The alternate rs560887-T allele was introduced into the resulting plasmid using the QuikChange II Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA) in conjunction with the following primers: 5'-GATCCAGTTTCTTTGCTTTTTATGCTTGTATCTATTCTTCCATCG-3' and 5'-CGATGGAAGAATAGATACAAGCATAAAAAGCAAAGAAACTGGATC-3' (mutated base underlined). The alternate rs2232321-A allele and the consensus exon 4 5' splice junction were then introduced into the resulting plasmid containing the rs560887-T allele using the following primers: rs2232321: 5'-GTGATCCAGTTTCTTTGCTTTTTATACTTGTATCTATTCTTCCATC-3' and 5'-GATGGAAGAATAGATACAAGTATAAAAAGCAAAGAAACTGGATCAC-3'; exon 4 junction: 5'-TATCTATTCTTCCATCGTAGGCTGACCTGGTCATTTCTTTG-3' and 5'-CAAAGAAATGACCAGGTCAGCCTACGATGGAAGAATAGATA-3' (mutated bases underlined). A *G6PC2* fragment containing ~1300bp of exon 5 and ~300bp of 5' intronic sequence was isolated using PCR with the PAC 299 plasmid as the template in conjunction with the following primers: 5'-

ACGCGTCGACCAAAGGCTAATTGGCTATGGC-3' and 5'-GCTCTAGACAAATATAGTAAGAAGCTGG-3' (*Sal* I and *Xba* I cloning sites underlined). The PCR fragment was digested with *Sal*I and *Xba*I and ligated into the *Sal*I-*Xba*I digested RHCglo vector (Fig. **2.1**). The alternate rs35259259 allele was introduced into the resulting plasmid using the QuikChange II Site-Directed Mutagenesis Kit in conjunction with the following primers: 5'-TCTTTCCAATCCTCAG-CATGCTGGTGGCAGAG-3' and 5'-CTCTGCCACCAGCATG-CTGAGGATTGGAAAGA-3' (deleted bases underlined).

Cell Culture

All cell lines, including the pancreatic islet-derived cell lines βTC-3, Min6, Hamster insulinoma tumor (HIT), and the 832/13 sub-clone of INS-1, as well as HeLa, were passaged as subconfluent cultures (150 cm^2 flasks) in Dulbecco's modified Eagle's medium (DMEM) supplemented with 100 U/mL penicillin and 100 μg/mL streptomycin. βTC-3 and Min6 cultures were also supplemented with 10% (vol/vol) fetal bovine serum, HIT cultures with 2.5% (vol/vol) fetal bovine serum and 15% (vol/vol) horse serum, and HeLa cultures with 10% (vol/vol) calf serum.

Transient Transfection

Cells were grown to approximately 70% confluence and were replated the day before use into 35 mm-diameter six-well dishes (one 150 cm^2 flask to 10 dishes). Attached cells were briefly incubated for 10 min in Opti-MEM reduced serum medium (GibcoBRL) before transfection using the lipofectamine reagent (GibcoBRL), per the

manufacturer's instructions. Separately, 2 μg of experimental plasmid DNA and 0.5 μg control plasmid DNA (*Renilla* (Promega) and firefly luciferase for the luciferase and CAT assays, respectively) were mixed with 100 μL Opti-MEM reduced serum medium. A solution containing 12.5 μL lipofectamine and 100 μL Opti-MEM reduced serum medium was then added dropwise to the DNA solution and incubated for 30 min. This DNA/lipofectamine mixture was then diluted with 800 μL of Opti-MEM reduced serum medium, and the resulting solution was added to the cells, which were returned to 37°C in a humidified 5% CO_2/95% air atmosphere in a Nuaire cell culture incubator. Following incubation for between 4 and 6 hours, the transfection medium was aspirated and replaced with the standard culture media prior to a final incubation for 18-20 hours.

CAT and Luciferase Assays

Transfected cells were harvested by trypsin digestion, suspended in 1X passive lysis buffer (Promega), and exposed to two cycles of freeze/thawing.

CAT Assays

Cell lysates for CAT assays were heated for 10 min at 65°C and cellular debris removed by centrifugation. From the resulting supernatant, a 100 μL aliquot was incubated in a final volume of 250 μL containing (at final concentrations) 200mM Tris□HCl (pH 7.8), 6mM $MgCl_2$, 75mM KCl, 0.4mM coenzyme A, 3mM ATP, 1mM chloramphenicol, 0.03 U of acetyl coenzyme A synthetase, and 0.34mM [^{3}H]acetate (0.5 Ci/mmol). This mixture was incubated for 2 h at 37°C, extracted with benzene, and transferred into scintillation vials. After removal of the benzene by evaporation, the [^{3}H]acetate incorporated into chloramphenicol was quantified by scintillation

spectroscopy. The counts per minute in a lysate-free blank were subtracted, and CAT activity was corrected for variations in protein concentration of the cell lysates, as measured by the Pierce bicinchoninic acid assay.

Luciferase Assays

After two cycles of freeze/thawing, firefly and *Renilla* luciferase activity were assayed sequentially in 10 μL of lysate using the dual luciferase assay kit (Promega). The activity of the *G6PC2* promoter was expressed as the ratio of the firefly/*Renilla* luciferase activities and is thus corrected for differences in transfection efficiency and cell viability.

Gel Retardation Assays

Labeled Probes

Sense and anti-sense oligonucleotides representing wild-type or mutant Foxa binding sites or the rs573225 variants were synthesized with *Bam*HI compatible ends. Oligonucleotides representing the rs13431652 variants of the NF-Y binding site were synthesized with *Hin*DIII compatible ends. Oligonucleotides were subsequently gel purified, annealed, and labeled with [α^{32}P]dATP using the Klenow fragment of *Escherichia coli* DNA Polymerase I to a specific activity of ~2.5 μCi/pmol.

Nuclear Extract Preparation

The preparation of βTC-3 and H4IIE nuclear extracts through extraction of nuclei with 800mmol/L NaCl (designated high salt) was as previously described [78]. The preparation of rat liver nuclear extract was also performed as previously described.

Binding Assays

Approximately 14 fmol of radiolabeled probe (~50,000 cpm) was incubated with the indicated nuclear extract in a final 20 μL reaction volume. Foxa2 binding reactions contained 1.0 μg βTC-3, 2.0 μg liver, or 2.0 μg H4IIE nuclear extract, 20mmol/L Hepes pH 7.9, 0.1mmol/L EDTA, 0.1mmol/L EGTA, 10% glycerol (v/v), 1mmol/L dithiothreitol, 1 μg poly(dI-dC)□poly(dI-dC), and 50mmol/L KCl. After incubation at room temperature for 20 minutes, samples were loaded onto a 6% polyacrylamide gel containing 1X TGE (25mmol/L Tris Base, 190mmol/L glycine, 1mmol/L EDTA) and 2.5% (v/v) glycerol. Samples were electrophoresed for 1.5 h at 150V in 1X TGE buffer before the gel was dried and then exposed to Kodak XB film with intensifying screens. Competition and supershift experiments were performed as previously described [78]. For supershift experiments, antisera raised against Foxa2 (HNF-3β) (sc-65540x) and NF-Y (sc-17753x) were obtained from Santa Cruz Biotechnology. A Pharos FX Plus Molecular Imager in conjunction with Imaging Screen and Cassette K and the Quantity One program (Bio-Rad) were used to quantitate ^{32}P associated with protein-DNA complexes.

Minigene Splicing Analyses

Human cervix-derived HeLa cells were grown in Dulbecco's modified Eagle's medium containing 10% (vol/vol) bovine serum. Cells were transiently transfected with 2 μg of the plasmids described above using lipofectamine as previously described. 18 hr following transfection cells were harvested by trypsin digestion (Bischof, 2001) and resuspended in Denaturation Solution (Ambion RNAqueous Kit) prior to RNA isolation according to the manufacturers' instructions. RNA (10 μg) was treated with Turbo-

DNase (Ambion Turbo DNA-*free*) per the manufacturers' instructions for 30 min at 37°C and then overnight with *Dpn* I to digest plasmid DNA. Reverse transcription (RT) reactions were carried out using the Bio-Rad iScript cDNA Synthesis Kit: 4 μL buffer mix, 5 μL treated RNA (1 μg), 10 μL water, and 1 μL reverse transcriptase were incubated at 25°C for 5 min, 42°C for 30 min, and then 85°C for 5 min. PCR reactions (final volume 100 μL) were performed using 0.1 μL of the RT reaction product with 5 U Taq polymerase (Applied Biosystems), 10 μL 10X PCR Buffer II (Applied Biosystems), 1.5 μM $MgCl_2$ (Applied Biosystems), 5 μL DMSO, 2 μCi [α-^{32}P]-dATP (Perkin Elmer), 200 nM dNTPs (200 nM each), and 100 nM of the following primers: 5'-CCCAAGCTTCATTCACCACATTGGTGTGC-3' and 5'-CCGCTCGAGGGGCTTTGCAGCAACAGTAAC-3' (*Hin*D III and *Xho* I cloning sites underlined). Reaction conditions were 24 cycles at 95°C for 1 min, 55°C for 1 min, and 72°C for 1 min, followed by one cycle at 72°C for 5 min. Reactions products were loaded on a 6% polyacrylamide gel containing 1X TGE (25 mM Tris Base, 190 mM glycine, 1 mM EDTA) and 2.5% (v/v) glycerol. Samples were electrophoresed for 1.5 hrs at 150V in 1X TGE buffer before the gel was dried and exposed to Kodak XB film with intensifying screens. For data quantitation, dried gels were exposed to a Bio-Rad Imaging Screen K overnight, which was then scanned in a Bio-Rad Molecular Imager FX System. PCR bands were quantitated using the "Volume Rectangle Tool" and "Mean Value" quantifier with the "Volume Analysis Report: of the Quantity One program. Background was measured with a separate volume and subtracted. To confirm the identity of splice junction reaction products were digested with *Hin*D III and *Xho* I and ligated into *Hin*D III – *Xho* I digested pGEM7 (Promega) for DNA sequencing.

Statistical Analysis

The transfection data were analyzed for differences from the control values, as specified in the figure legends. Statistical comparisons were calculated using an unpaired Student *t* test. The level of significance was $P < 0.05$ (two-sided test).

CHAPTER III

FUNCTIONAL DATA SUPPORT A POTENTIAL ROLE FOR rs13431652 AS A CAUSATIVE *G6PC2* SNP CONTRIBUTING TO VARIATION IN FPG

Introduction

The genetic association of multiple allelic variants of *G6PC2* with variations in FPG necessitates analysis of the potential functional contribution of each SNP. The intronic SNP rs560887 shows high LD (>0.80) with two promoter SNPs, rs13431652 and rs573225, according to data from HapMap phase III (r^2 = 0.88 and r^2 = 0.96, respectively). Their contributions to FPG, however, have not yet been examined. To ascertain whether these SNPs potentially affect glucose homeostasis, further genetic association analyses were first carried out by our collaborators Nabila Bouatia-Naji and Philippe Froguel. Following the demonstration of independent genetic linkage between rs13431652 and rs573225 and FPG, a biochemical approach was taken to evaluate the effects of these promoter variants on transcription factor binding and promoter activity.

Results

Association of promoter variants of *G6PC2* with FPG

Only two common SNPs (minor allele frequency > 0.05), namely rs13431652 and rs573225, are reported to be in high linkage disequilibrium (LD) with the previously identified intronic *G6PC2* SNP rs560887 ($r^2 > 0.80$, according to HapMap data [release XX]). rs13431652 is located in the distal *G6PC2* promoter, at -4405 relative to the transcription start site, whereas rs573225 is located in the proximal promoter at -259. We have assesse the effect of these two *G6PC2* promoter variants on FPG, adjusted for age, sex, and BMI in 9532 NG Europeans from four independent populations. In the meta-analyses, we show strong associations of both rs13431652 ($\beta = 0.075$, $P = 3.6\text{x}10^{-35}$) and rs573225 ($\beta = 0.073$, $P = 3.6\text{x}10^{-34}$) with FPG, in similar magnitude to that for rs560887 ($\beta = 0.075$, $P = 3.6\text{x}10^{-35}$) (Table 3.1). Conditioned regression model analyses to assess the independency of the three SNPs studied turn out to be noninformative (data not shown), as these analyses include a high variance inflation factor because of high LD that we observed between the SNPs in our populations.

To compensate for this limited analytical situation, we analyzed the association of FPG with multiple haplotypes of the promoter variants and rs560887, specifically individual contributions of each variant and contributions when variants were analyzed in pairs and when we constrained the effect of rs560887 to be null. Because of the LD discrepancies between populations, these haplotype analyses were conducted in the four poopulations separately. In DESIR, we found that rs560887 is significantly contributing to the haplotype association in combination with rs13431652 ($P = 7.58\text{x}10^{-5}$) and

Table 3.1. ***G6PC2*** **genetic variation and fasting glucose in four cohorts and one obese population**

		Meta-analysis (n = 9,532)		DESIR (n = 3,483)		NFBC86 (n = 4,372)		Haguenau (n = 1,201)		Obese children (n = 476)	
	EA (freq)	β (SE)	P	β (SE)	P	β (SE)	P	β (SE)	P	β (SE)	P
rs13431652	A (0.68)	0.075 (0.006)	3.6×10^{-35}	0.083 (0.010)	2.0×10^{-15}	0.069 (0.009)	1.1×10^{-14}	0.071 (0.015)	2.0×10^{-6}	0.086 (0.03)	0.005
rs573225	A (0.66)	0.073 (0.006)	3.6×10^{-34}	0.080 (0.010)	1.3×10^{-14}	0.066 (0.009)	6.8×10^{-14}	0.073 (0.015)	9.2×10^{-7}	0.093 (0.03)	0.002
rs560887	G (0.69)	0.071 (0.006)	1.2×10^{-31}	0.069 (0.011)	3.0×10^{-11}	0.072 (0.009)	8.7×10^{-16}	0.076 (0.015)	4.5×10^{-7}	0.062 (0.03)	0.04

Individual and meta-analyses data are displayed as regression coefficient β (SE) adjusted for age, sex, BMI, and associated P values.

rs573225 (P = 0.015). We also observed a significant contribution of rs13431652 (P = $4.09x10^{-7}$) and rs573225 (P = $1.21x10^{-8}$) to the haplotype associations, supporting independent effects of both promoter variants to the global association of the haplotype in combination with rs560887. In contrast, distinct results were obtained in the remaining cohorts in which we found that both promoter and intronic variants are not independently contributing to the haplotype associations, suggesting that promoter and intronic variants are indispensable to the two by two haplotype associations, with a modestly more significant contribution of rs560887 (P = 0.006 against rs13431652 and P = 0.07 against rs573225) in the NFBC 1986 cohort. We speculate that these distinct results in some way reflect the differences in LD within the different cohorts. Despite these differences, haplotype analyses support a substantial role of promoter variants to the association signal with FPG, and do not discard the contribution of rs560887.

We have also assessed the effect of the intronic variant rs853789 located in the 19^{th} intron of *ABCB11*, one of two SNPs, along with rs560887, that were reported by Chen et al. [91] to be associated with FPG. *ABCB11* (ATP-binding cassette family B 11) encodes the major bile salt export pump and is expressed predominantly in the liver [98] and was hypothesized to potentially drive the association with FPG observed at this locus [91]. We found a significant association between rs853789 and FPG (β = -0.062, P = $1.1x10^{-25}$). However, when we include either rs560887, rs13431652, or rs573225 in a regression conditional model that includes rs853789, age, sex, BMI, and cohort, the effect of rs853789 is highly reduced, whereas rs13431652 (P = $1.2x10^{-10}$), rs573225 (P = $1.8x10^{-9}$), and rs560887 (P = $1.3x10^{-7}$) remain strongly associated with FPG. These findings were confirmed by the haplotype analyses that did not support a frank,

independent contribution of rs853789 to the association in any of the cohorts studied. These genetic data suggest that in our populations, the *ABCB11* rs853789 effect on FPG is driven mainly by *G6PC2* variation through moderate LD, estimated between 0.42 and 0.78, depending on the population studied.

Expression Analyses

The expression profile of *G6PC2* was confirmed by quantitative real time PCR as restricted to pancreatic tissues. In humans this includes whole pancreas, islets, and β cells (Fig. 3.1A). These data had previously been reported for mouse [32, 43] and human [36] tissue through RNA blotting. No significant association of the minor alleles of rs13431652, rs573225, or rs560887 with high FPG and *G6PC2* gene expression were found (Fig. 3.1B). This was based on a small number of samples (N = 24) of pancreatic islet cDNA whose corresponding genomic DNA was available. Because of this, and the small sample size of the genetic analyses, a causative variant among the three SNPs was not determined.

Transcription factor binding at rs13431652 *in vitro*

To identify a *cis*-acting element whose corresponding binding factor was likely to be affected by allelic variation at rs13431652, MatInspector sequence analysis software was employed [99]. By analyzing the *G6PC2* promoter region encompassing this SNP, a CCAAT box was found (Fig. 3.2A). This DNA-binding element is known to bind several different transcription factors, including NF-Y [100, 101], and its sequence changes from CCAaT to CCAgT with the minor allele of rs13431652.

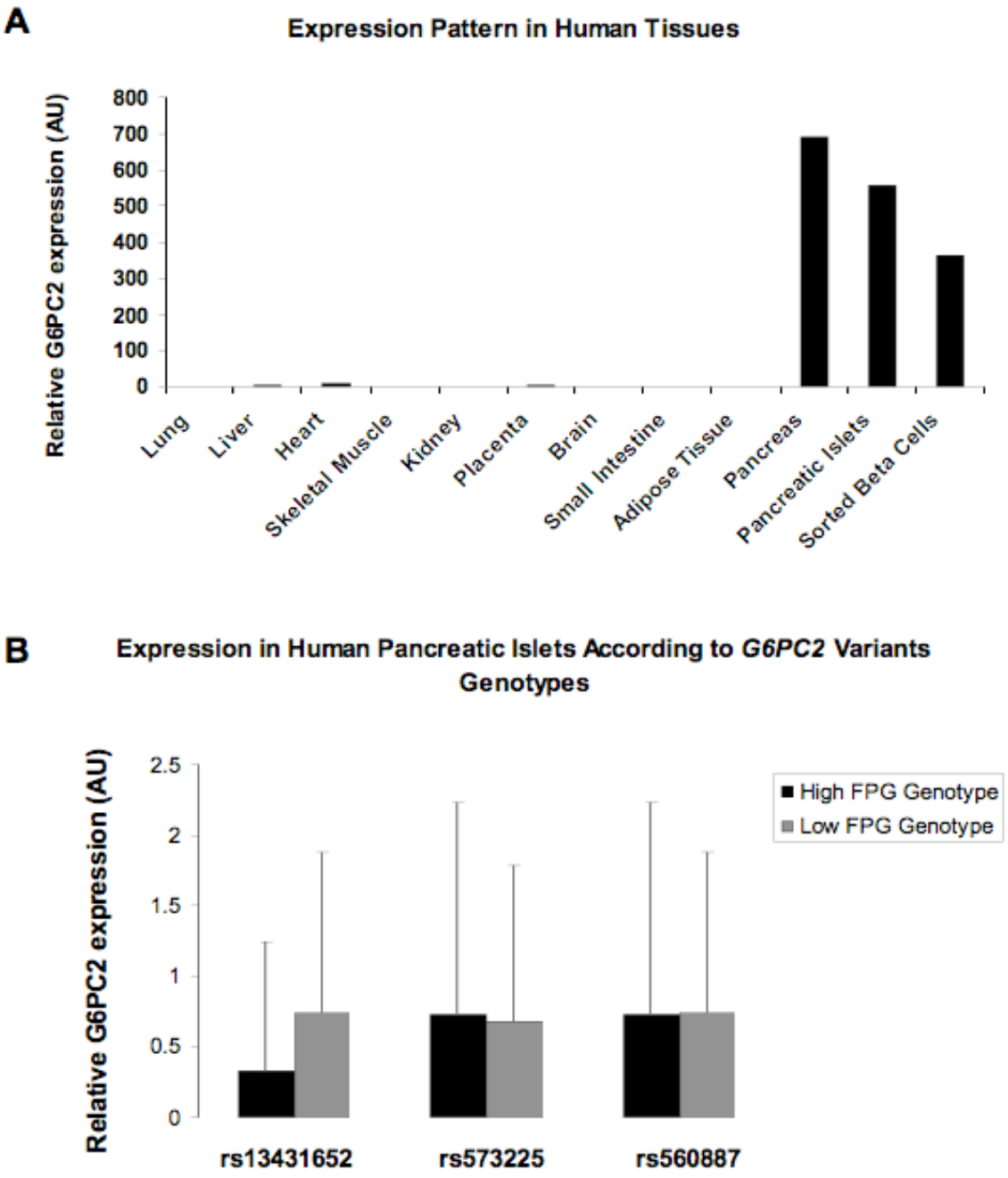


Figure 3.1. Analysis of *G6PC2* gene expression
(A) *G6PC2* expression was analyzed in human tissues by real-time PCR and data normalized according to expression levels of the house keeping gene *POLR2A*. (B) *G6PC2* expression was analyzed in pancreatic islets, relative to *G6PC2* genotypes, by real-time PCR and data is normalized according to expression levels of the house keeping gene *POLR2A*.

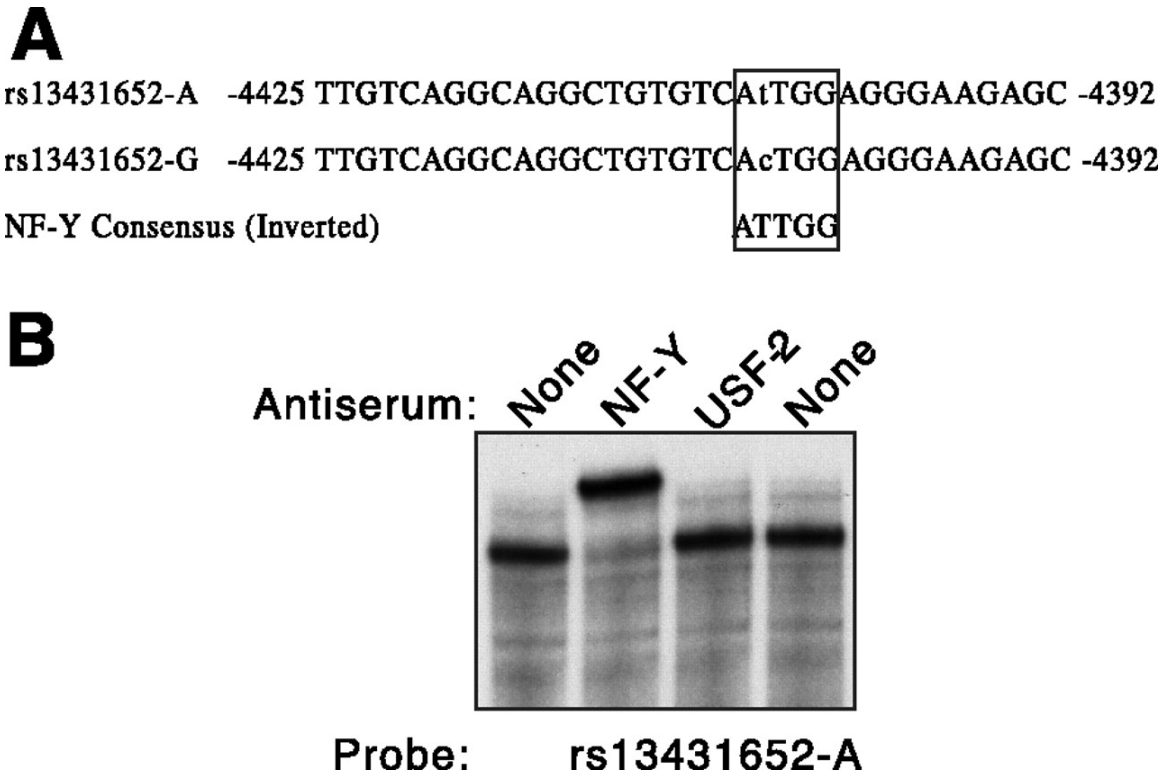


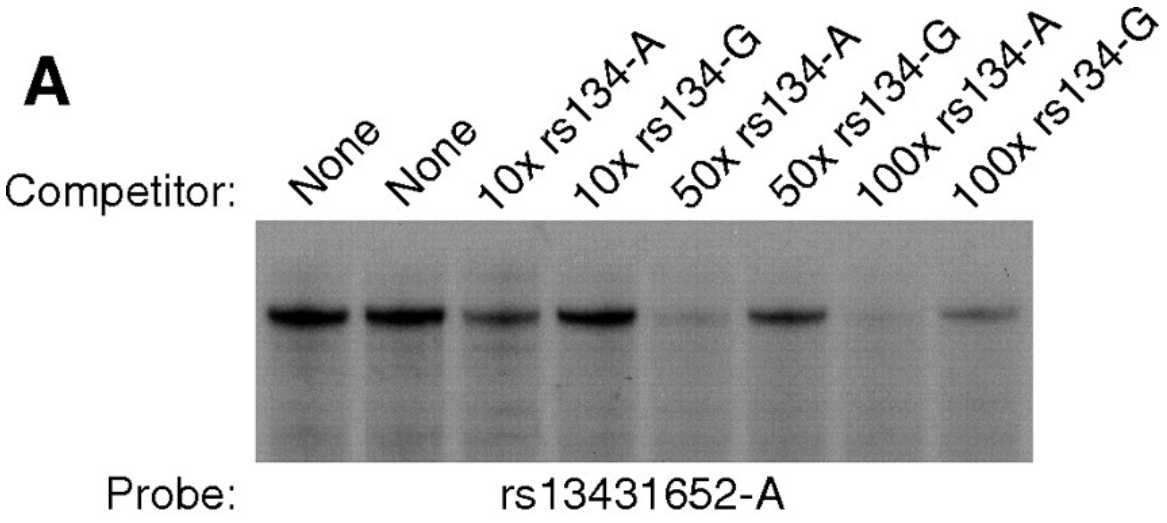
Figure 3.2. Binding of the -4425/-4392 promoter region of *G6PC2* by NF-Y *in vitro*
(A) The sense strand sequences of the oligonucleotides used in gel retardation assays are shown. SNP base pairs are shown in bold lower case letters. The consensus NF-Y binding motif, CCAAT, is taken from Quandt et al. (B) βTC-3 nuclear extract was incubated in the absence or presence of the indicated anti-serum for 10 min on ice. A labeled oligonucleotide representing the -4425/-4392 *G6PC2* promoter region and containing the rs13431652-A allele (rs13431652-A; Panel A) was then added and incubation continued for 20 min at room temperature. Protein binding was then analyzed using the gel retardation assay as described in Chapter II. In the representative autoradiograph shown, only the retarded complexes are visible and not the free probe, which was present in excess.

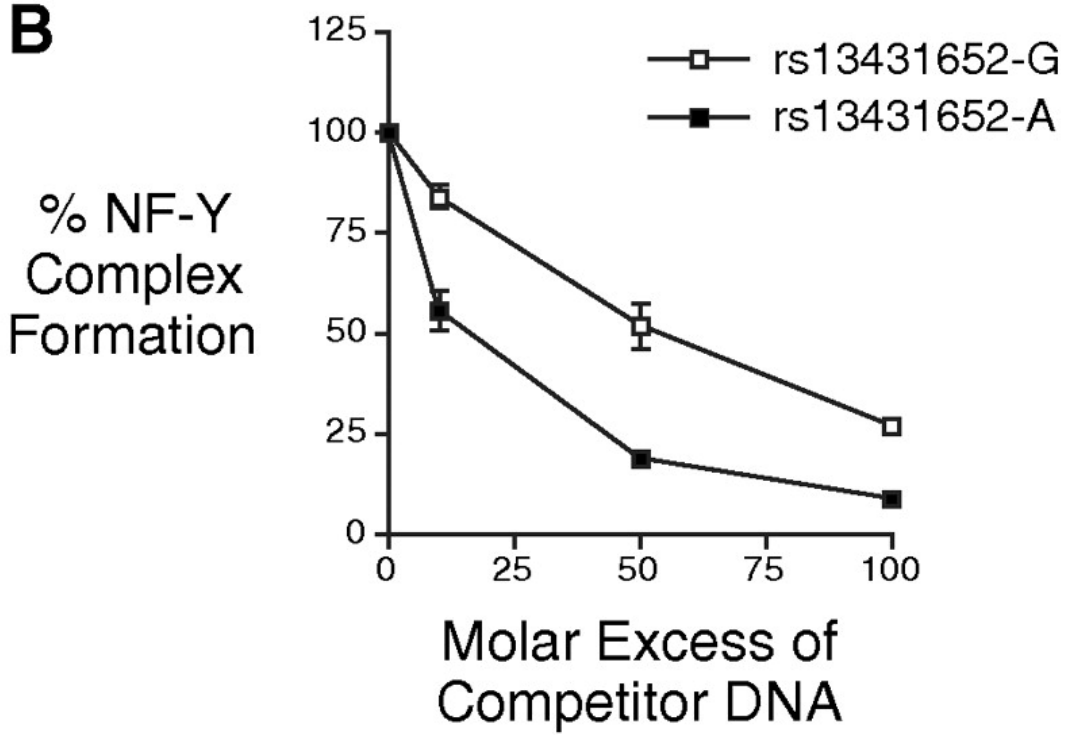
Binding analyses of NF-Y to the *G6PC2* promoter *in vitro* were carried out by gel retardation assay. A radioactively labeled double-stranded oligonucleotide, designated rs13431652-A, represented the *G6PC2* promoter sequence from -4425 to -4392 and included the rs13431652-A allele. Following incubation of rs13431652-A with nuclear extract prepared from βTC-3 cells, a single protein-DNA complex was detected by phosphorimaging (Fig. 3.2B). This binding factor was identified by gel retardation assay in which the βTC-3 cell nuclear extract was pre-incubated with control or experimental antisera for USF-2 or NF-Y, respectively. Addition of the antibodies specific for NF-Y resulted in a distinct supershift in the migration of the complex, while pre-incubation with antisera for USF-2 did not affect the complex migration (Fig. 3.2B). These results strongly suggest association of NF-Y in this complex.

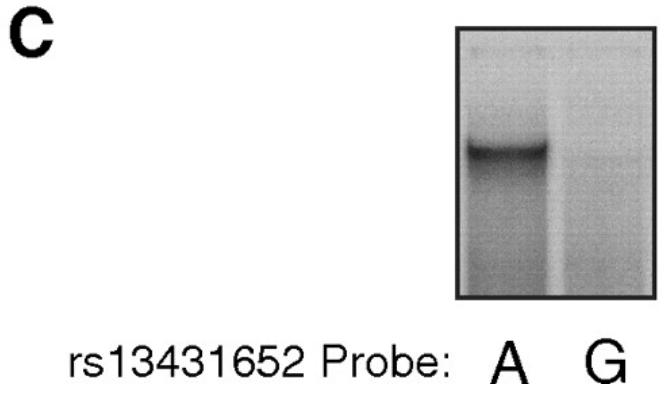
Comparison of the affinity of NF-Y binding to the variants of rs13431652 in the *G6PC2* promoter was carried out utilizing gel retardation competition experiments, in which an increasing molar excess of unlabeled DNA was included with the labeled major allele oligonucleotide—in this case rs13431652-A. It can be seen that the rs13431652-A oligonucleotide more effectively competed for the formation of the NF-Y-DNA complex than the rs13431652-G oligonucleotide (Fig. 3.3A). Quantification of this result from multiple experiments elucidated that the rs13431652-A oligonucleotide is bound by NF-Y with approximately five fold higher affinity (Fig. 3.3B). In contrast to the competition experiment data, however, direct analysis of NF-Y binding to the A and G alleles of rs13431652 using separate oligonucleotide probes labeled with equal specific activity indicated a greater difference in NF-Y binding affinity (Fig. 3.3C). This incongruity, observed prior in studies on FOXO1 binding [102], is hypothesized to arise when the

Figure 3.3. Comparison of the affinity of NF-Y binding to the rs13431652-A and rs13431652-G variants of the *G6PC2* NF-Y binding site *in vitro*

(A) The labeled rs13431652-A oligonucleotide (Fig. 3.1A) was incubated in the absence or presence of the indicated molar excess of the unlabeled human rs13431652-A or rs13431652-G oligonucleotide competitors (Fig. 3.1A) before the addition of βTC-3 cell nuclear extract. Protein binding was then analyzed using the gel retardation assay as described in Chapter II. In the representative autoradiograph shown, only the retarded complexes are visible and not the free probe, which was present in excess. (B) Protein binding was quantified by using a Packard Instant Imager to count ^{32}P associated with the retarded complex. The data represent the mean ± S.E.M. of three experiments. (C) The labeled rs13431652-A and rs13431652-G oligonucleotides (Fig. 3.1A) were incubated with βTC-3 cell nuclear extract, and protein binding was analyzed using the gel retardation assay as described in Chapter II. In the representative autoradiograph shown, only the retarded complexes are visible and not the free probe, which was present in excess.


A
Competitor:
None
None
10x rs134-A
10x rs134-G
50x rs134-A
50x rs134-G
100x rs134-A
100x rs134-G
Probe:
rs13431652-A


B
% NF-Y Complex Formation
125
100
75
50
25
0
0
25
50
75
100
rs13431652-G
rs13431652-A
Molar Excess of Competitor DNA


C
rs13431652 Probe:
A
G

rates of association and dissociation of a factor interacting with DNA are increased concurrently. Such an effect would yield a limited variation in K_D [103]; however, a significant difference in the dissociation of the protein-DNA complex upon separation of bound and free probe during electrophoresis would result.

Alteration of *G6PC2* fusion gene expression by rs13431652

The functional significance of altered NF-Y binding on human *G6PC2* promoter activity was next investigated. NF-Y can function as either an activator or repressor such that the effect of NF-Y on gene transcription is context dependent [104]. Fusion genes containing each of the alleles of rs13431652, generated in the context of the -8563 to +11 *G6PC2* promoter region, were analyzed by transient transfection of βTC-3 cells. Figure 3.4 shows that the rs13431652-G allele was associated with an approximately 25% decrease in promoter activity in comparison to that observed with the rs13431652-A allele. Figure 3.5 shows that this difference was only observed when the *G6PC2* promoter was analyzed in the context of the pGL4 but not the pGL3 vector; the former is an improved vector that lacks multiple transcription factor binding sites in the plasmid backbone and luciferase gene that are known to occasionally give rise to spurious data [79]. These results indicate that NF-Y acts as an activator in the context of the *G6PC2* promoter. More importantly, these data are consistent with the reduced FPG observed in *G6pc2* knockout mice [55] in that the rs13431652-A allele is associated with both elevated FPG and *G6PC2* promoter activity. As such, these functional data support a potential role for rs13431652 as a SNP linking *G6PC2* to variations in FPG.

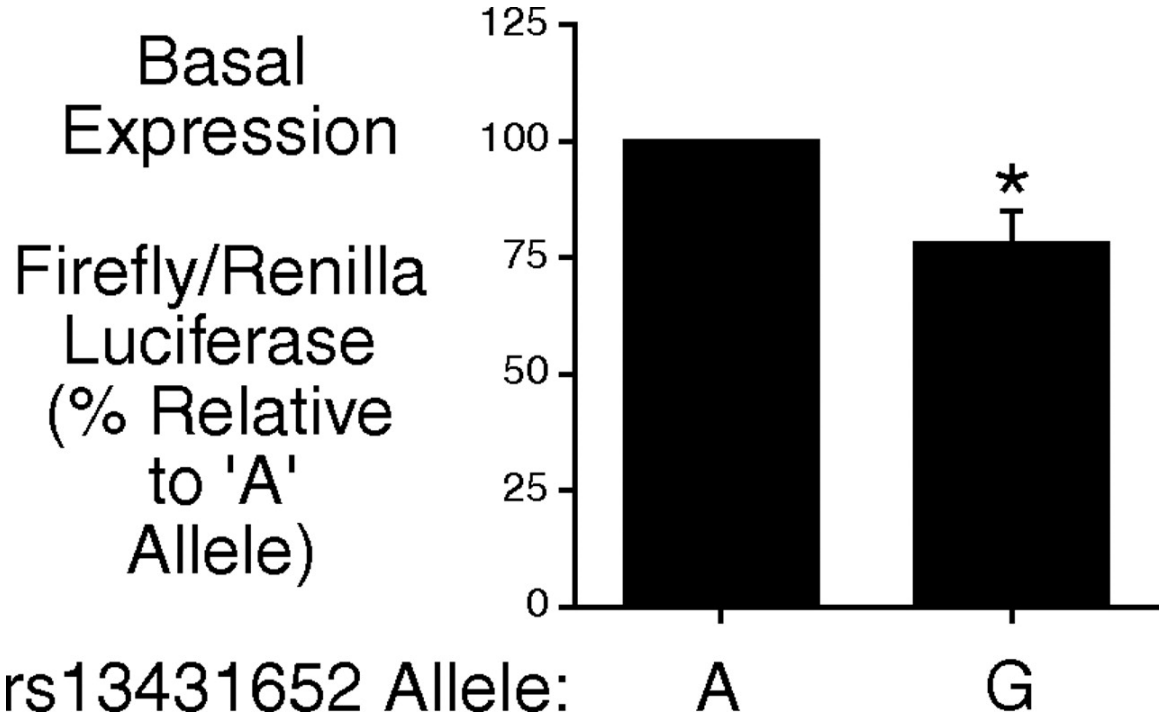


Figure 3.4. Decreased *G6PC2* promoter activity associated with the human *G6PC2* rs13431652-G allele relative to the rs13431652-A allele

βTC-3 cells were transiently cotransfected, as described in Chapter II, using a lipofectamine solution containing various *G6PC2*-luciferase fusion genes in the pGL4 MOD vector (2 μg) and an expression vector encoding *Renilla* luciferase (0.5 μg). The *G6PC2*-luciferase fusion genes represented the rs13431652-A or rs13431652-G alleles present in the context of the human *G6PC2* promoter sequence located between -8563 and +11. After transfection, cells were incubated for 18-20 h in serum-containing medium. The cells were then harvested, and firefly and *Renilla* luciferase activity were assayed as described in Chapter II. Results are presented as the ratio of firefly:*Renilla* luciferase activity, expressed as a percentage relative to the value obtained with the rs13431652-A allele, and represent the mean of 12 experiments ± S.E.M., each using an independent preparation of each fusion gene plasmid, assayed in triplicate. **P* < 0.05 vs. rs13431652-A allele.

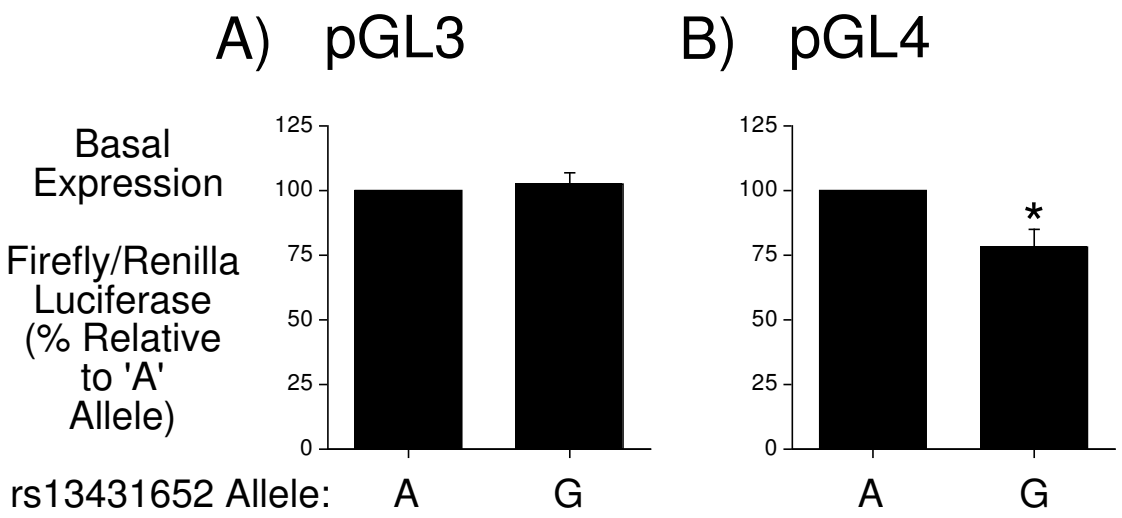


Figure 3.5. Decreased *G6PC2* promoter activity associated with the human *G6PC2* rs13431652-G allele relative to the rs13431652-A allele in pGL4 but not pGL3
βTC-3 cells were transiently cotransfected, as described in Chapter II, using a lipofectamine solution containing various *G6PC2*-luciferase fusion genes in the pGL3 MOD (Panel A) or pGL4 MOD (Panel B) vectors (2 μg) and an expression vector encoding *Renilla* luciferase (0.5 μg). The *G6PC2*-luciferase fusion genes represented the rs13431652-A or rs13431652-G alleles present in the context of the human *G6PC2* promoter sequence located between -8563 and +11. After transfection, cells were incubated for 18-20 h in serum-containing medium. The cells were then harvested, and firefly and *Renilla* luciferase activity were assayed as described in Chapter II. Results are presented as the ratio of firefly:*Renilla* luciferase activity, expressed as a percentage relative to the value obtained with the rs13431652-A allele, and represent the mean of three experiments ± S.E.M., each using an independent preparation of each fusion gene plasmid, assayed in triplicate. *$P < 0.05$ vs. rs13431652-A allele.

Transcription factor binding at rs573225 *in vitro*

The *G6PC2* promoter region that encompasses rs573225 is highly conserved in the mouse *G6pc2* promoter [36]. This conserved region was recently characterized in the context of the mouse *G6pc2* promoter, showing that it contains a Foxa2 binding site [79]. In addition, it was demonstrated by chromatin immunoprecipitation assay (ChIP) that Foxa2 binds the endogenous *G6pc2* promoter in mouse βTC-3 cells *in situ*. Figure 3.6 shows that this region of the human *G6PC2* promoter also binds Foxa2 *in vitro*. When labeled double-stranded oligonucleotides, designated rs573225-A and -G (Fig. 3.6A), representing the *G6PC2* promoter sequence from -265 and -246 and the rs573225-A and -G alleles, respectively, were incubated with rat liver nuclear extract a single major complex, designed A2, was detected with both alleles (Fig. 3.6B). When rat liver nuclear extract was pre-incubated with antisera specific for Foxa2 a clear supershift in the migration of complex A2 was observed (Fig. 3.6B), strongly suggesting that this complex represents Foxa2 binding. This experiment was performed using rat liver nuclear extract, because, like pancreatic islets [105], it contains both Foxa1 and Foxa2 [106]. In contrast, βTC-3 cells only express Foxa2 [79]. The results indicate that rs573225 does not differentially affect binding of these related factors; instead, both alleles strongly bind Foxa2.

Gel retardation competition experiments, in which a varying molar excess of unlabeled DNA was included with the labeled rs573225-A probe, were used to compare the affinity of Foxa binding to the rs573225 variants of the *G6PC2* Foxa binding site. Figure 3.7A shows that, when using rat liver nuclear extract, both the rs573225-A and -G oligonucleotides competed equally effectively for the formation of the Foxa2-DNA

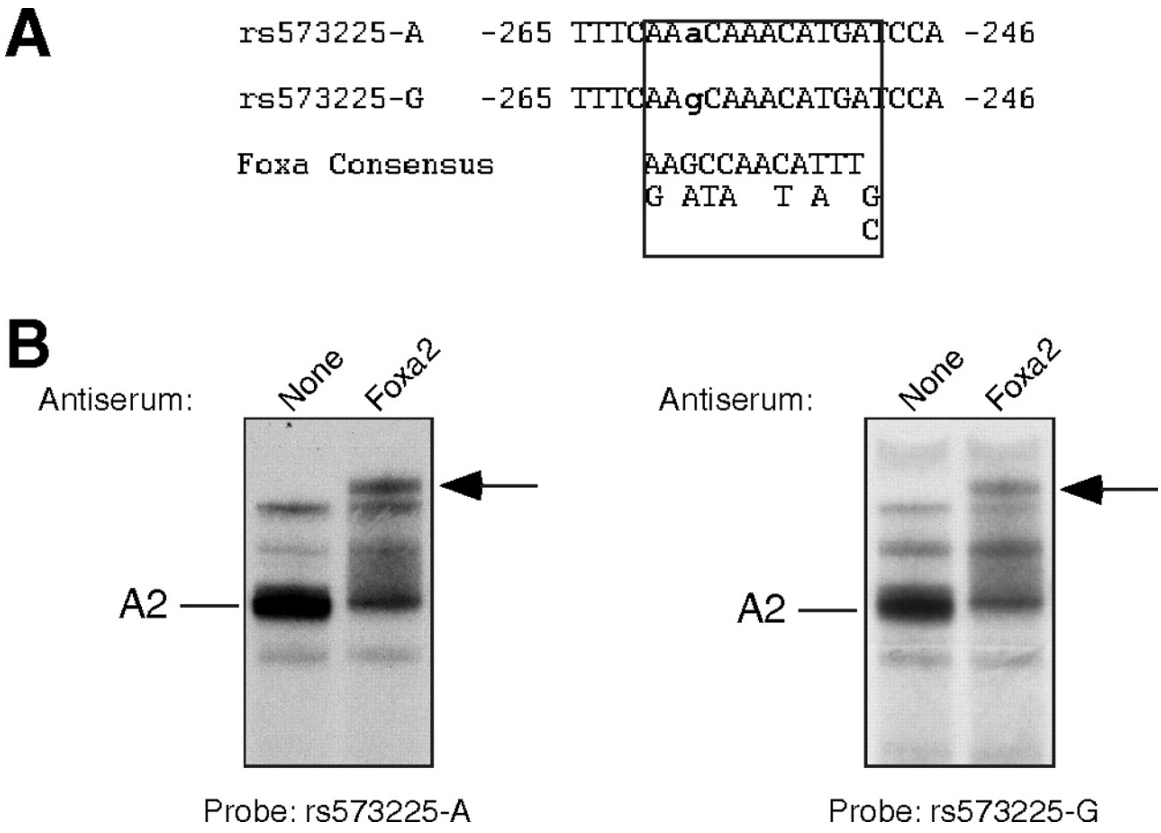
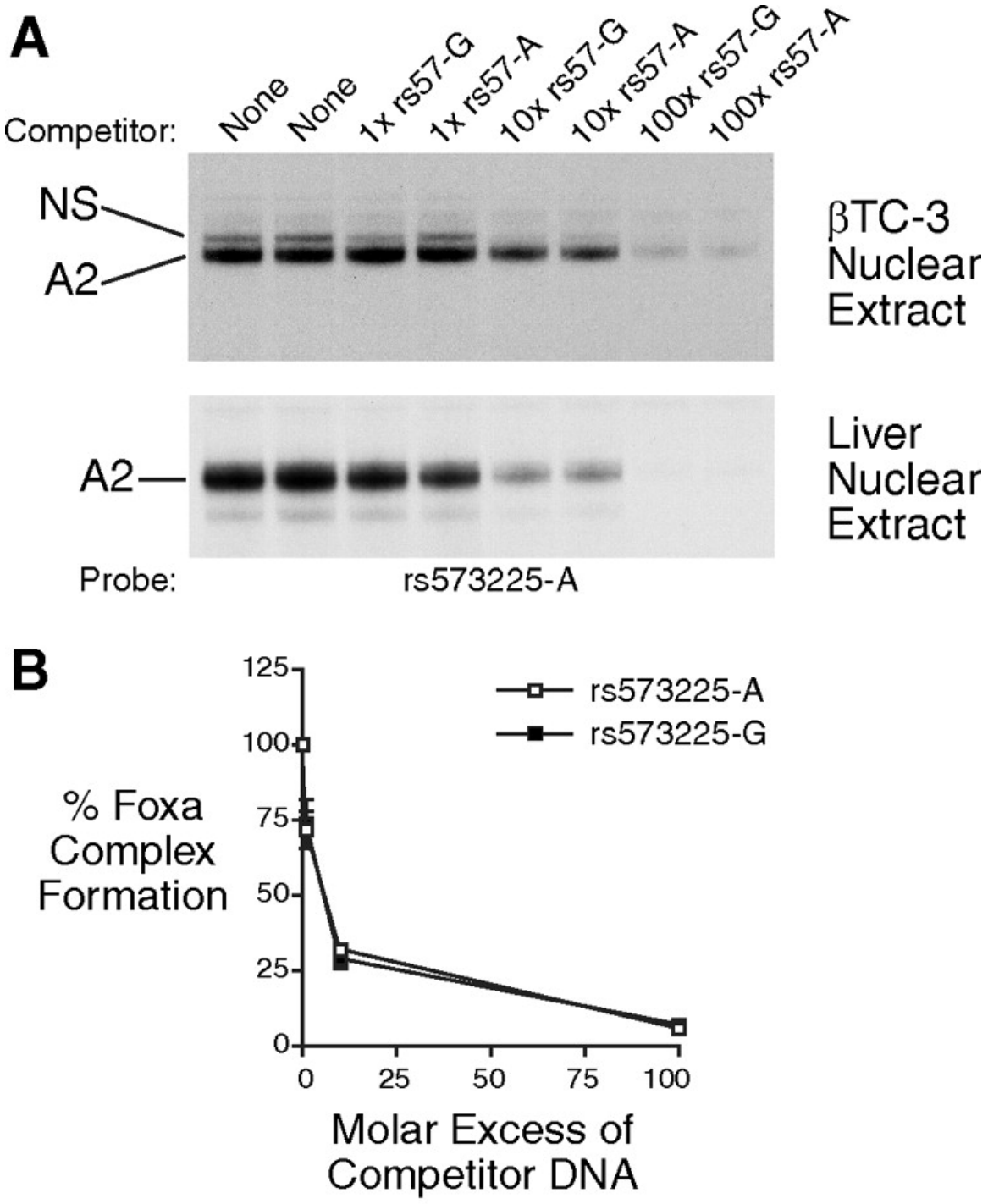


Figure 3.6. Binding of the -265/-246 promoter region of *G6PC2* by Foxa2 *in vitro*
(A) The sense strand sequences of the oligonucleotides used in gel retardation assays are shown. SNP base pairs are shown in bold lower case letters. The consensus Foxa binding motif, (A/G)A(G/A)(C/T)(C/A)AA(C/T)A(T/A)T(T/G/C), is taken from Overdier et al. (B) Liver nuclear extract was incubated in the absence (None) or presence of the indicated anti-serum for 10 min on ice. Labeled oligonucleotides representing the -265/-246 *G6PC2* promoter region and containing the rs573225-A or rs573225-G alleles (rs573225-A or rs573225-G; panel A) were then added and incubation continued for 20 min at room temperature. Protein binding was then analyzed using the gel retardation assay as described in Chapter II. In the representative autoradiograph shown, only the retarded complexes are visible and not the free probe, which was present in excess. A single major complex was detected (A2) that represents Foxa2 binding. The arrow indicates a supershifted complex.


A
Competitor:
None
None
1x rs57-G
1x rs57-A
10x rs57-G
10x rs57-A
100x rs57-G
100x rs57-A
NS
A2
βTC-3 Nuclear Extract
A2
Liver Nuclear Extract
Probe:
rs573225-A
B
125
100
75
50
25
0
% Foxa Complex Formation
rs573225-A
rs573225-G
0
25
50
75
100
Molar Excess of Competitor DNA

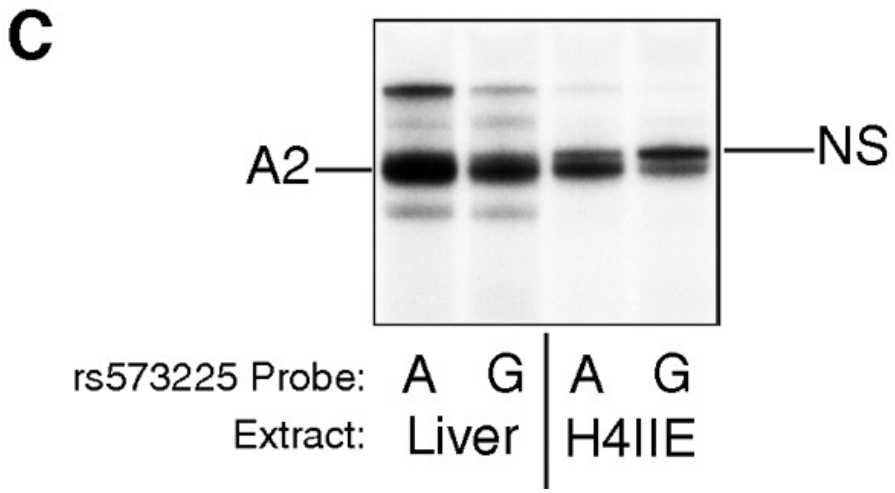


Figure 3.7. Comparison of the affinity of Foxa binding to the rs573225-A and rs573225-G variants of the *G6PC2* Foxa binding site *in vitro*

(A) The labeled rs573225-A oligonucleotides (Fig. 3.4A) was incubated in the absence or presence of the indicated molar excess of the unlabeled human rs573225-A or rs573225-G oligonucleotide competitors (Fig. 3.4A) before the addition of βTC-3 cell or rat liver nuclear extract. Protein binding was then analyzed using the gel retardation assay as described in Chapter II. In the representative autoradiograph shown, only the retarded complexes are visible and not the free probe, which was present in excess. With βTC-3 nuclear extract, complex A2 represents Foxa2 binding, and complex NS represents nonspecific binding (Martin 2008). With liver nuclear extract, complex A2 represents Foxa2 binding (O'Brien 1995 Hepatic nuclear factor 3- and hormone-regulated expression…). (B) Protein binding in experiments using liver nuclear extract was quantified by using a Packard Instant Imager to count ^{32}P associated with the retarded complex. The data represents the mean ± S.E.M. of three experiments. (C) The labeled rs573225-A and rs573225-G oligonucleotides (Fig. 3.4A) were incubated with either liver or H4IIE nuclear extract, and protein binding was analyzed using the gel retardation assay as described in Chapter II. In the representative autoradiograph shown, only the retarded complexes are visible and not the free probe, which was present in excess. NS, nonspecific.

complex. Quantitation of the results of several experiments confirmed that Foxa2 binds the rs573225-A and -G oligonucleotides with equal affinity (Fig. 3.7B). Identical results were obtained using mouse βTC-3 cell nuclear extract (Fig. 3.7A). It has previously been shown that the two complexes detected using βTC-3 cell nuclear extract represent the binding of Foxa2 (A2) and an unknown factor that either represents a non-specific (NS) interaction or binding to another region of the probe [79].

Interestingly, in contrast to the competition experiment data (Fig. 3.7A & B), the direct analysis of Foxa2 binding to the rs573225-A and -G oligonucleotides probes labeled with the identical specific activity showed that Foxa2 binds with slightly higher affinity to the rs573225-A allele (Fig. 3.6B & 3.7C). The identical observation was made using rat liver and rat H4IIE hepatoma cell nuclear extract (Fig. 3.7C). As explained above in the analysis of NF-Y binding (Fig. 3.3C), this suggests that rs573225 alters the association and dissociation rates of Foxa2 binding without affecting binding affinity.

Alteration of *G6PC2* fusion gene expression by rs573225

We next investigated the functional significance of altered Foxa2 binding on human *G6PC2* promoter activity. A block mutation in the *G6PC2* Foxa binding site that abolishes Foxa2 binding (Fig. 3.8) also markedly reduces fusion gene expression (Fig. 3.9) indicating that Foxa2 is an activator in the context of the human *G6PC2* promoter, as it is in the mouse *G6pc2* promoter [79]. Fusion genes containing each rs573225 variant, generated in the context of the -324 to +3 (Fig. 3.10) and -8563 to +11 (Fig. 3.11) *G6PC2* promoter regions, were analyzed by transient transfections of βTC-3 cells. Since the loss of Foxa2 binding reduces *G6PC2* promoter activity (Fig. 3.9) and the rs573225-G allele

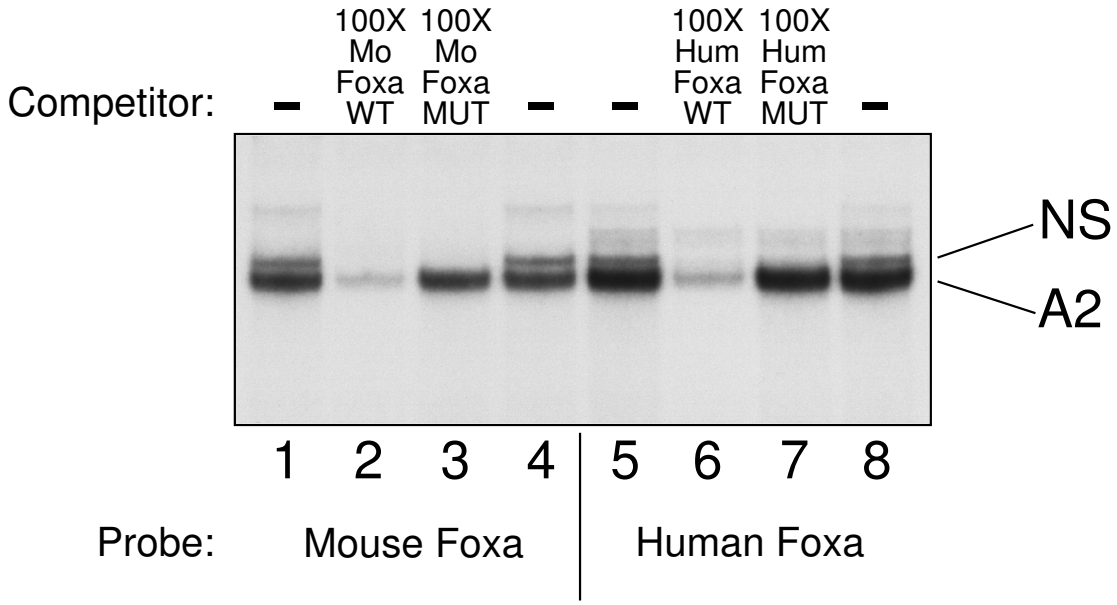


Figure 3.8. Comparison of Foxa2 binding to the labeled human *G6PC2* -265/-246 and mouse *G6pc2* -247/-228 promoter regions *in vitro*

Labeled oligonucleotides representing the wild-type (WT) human *G6PC2* -265/-246 or mouse *G6pc2* -247/-228 promoter regions were incubated in the absence or presence of a 100-fold molar excess of the indicated unlabeled WT or mutant (MUT) competitors. βTC-3 nuclear extract was then added and protein binding was analyzed using the gel retardation assay as described in Chapter II. In the representative autoradiograph shown, only the retarded complexes are visible and not the free probe, which was present in excess. The two major complexes detected represent Foxa2 (A2) and non-specific (NS) binding.

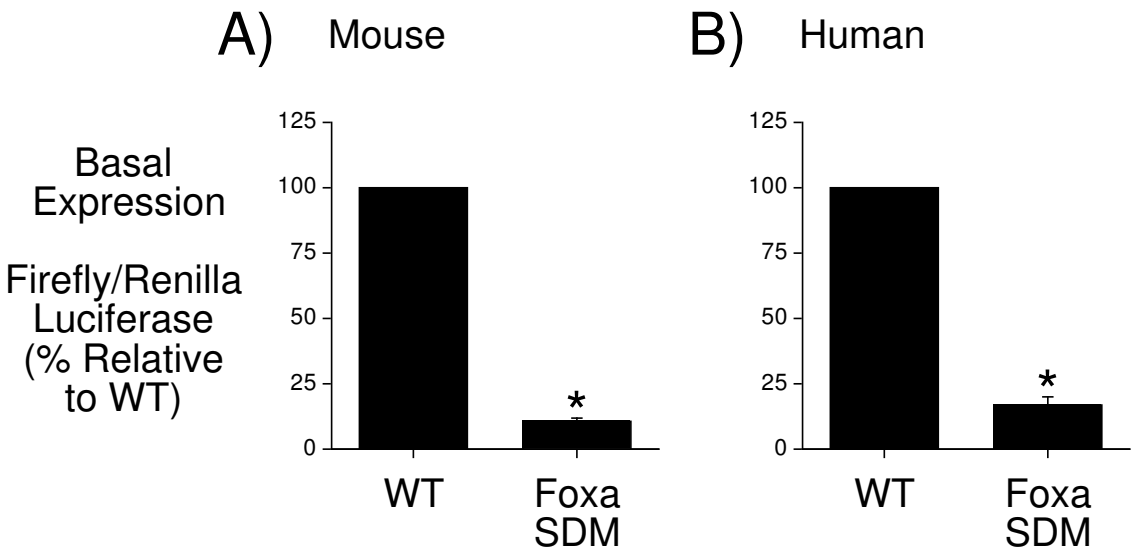


Figure 3.9. Reduction of human *G6PC2* and mouse *G6pc2* promoter activity by marked disruption of Foxa2 binding

βTC-3 cells were transiently cotransfected, as described in Chapter II, using a lipofectamine solution containing various *G6PC2*-luciferase fusion genes in the pGL3 MOD vector (2 μg) and an expression vector encoding *Renilla* luciferase (0.5 μg). The *G6PC2*-luciferase fusion genes represented either the wild-type (WT) mouse promoter sequence, located between -306 and +3 (Panel A), the WT human promoter sequence (rs573225-A allele), located between -324 and +3 (Panel B), or the same sequences with a site-directed mutation (SDM) in the Foxa2 binding site. After transfection, cells were incubated for 18-20 h in serum-containing medium. The cells were then harvested, and firefly and *Renilla* luciferase activity were assayed as described in Chapter II. Results are presented as the ratio of firefly:*Renilla* luciferase activity, expressed as a percentage relative to the value obtained with the WT fusion genes, and represent the mean of three experiments ± S.E.M., each using an independent preparation of each fusion gene plasmid, assayed in triplicate. *$P < 0.05$ vs. WT.

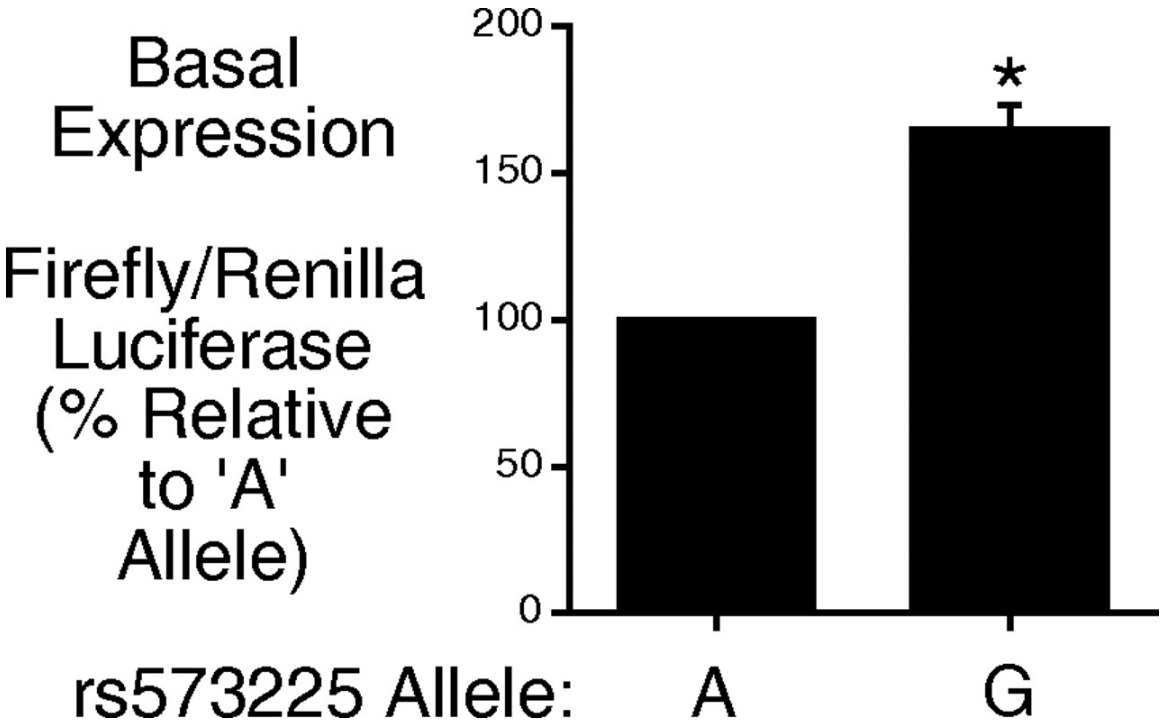


Figure 3.10. Increased *G6PC2* promoter activity associated with the human *G6PC2* rs573225-G allele relative to the rs573225-A allele

βTC-3 cells were transiently cotransfected, as described in Chapter II, using a lipofectamine solution containing various *G6PC2*-luciferase fusion genes in the pGL3 MOD vector (2 μg) and an expression vector encoding *Renilla* luciferase (0.5 μg). The *G6PC2*-luciferase fusion genes represented the rs573225-A or rs573225-G alleles present in the context of the human *G6PC2* promoter sequence located between -324 and +3. After transfection, cells were incubated for 18-20 h in serum-containing medium. The cells were then harvested, and firefly and *Renilla* luciferase activity were assayed as described in Chapter II. Results are presented as the ratio of firefly:*Renilla* luciferase activity, expressed as a percentage relative to the value obtained with the rs573225-A allele, and represent the mean of three experiments ± S.E.M., each using an independent preparation of each fusion gene plasmid, assayed in triplicate. *$P < 0.05$ vs. rs573225-A allele.

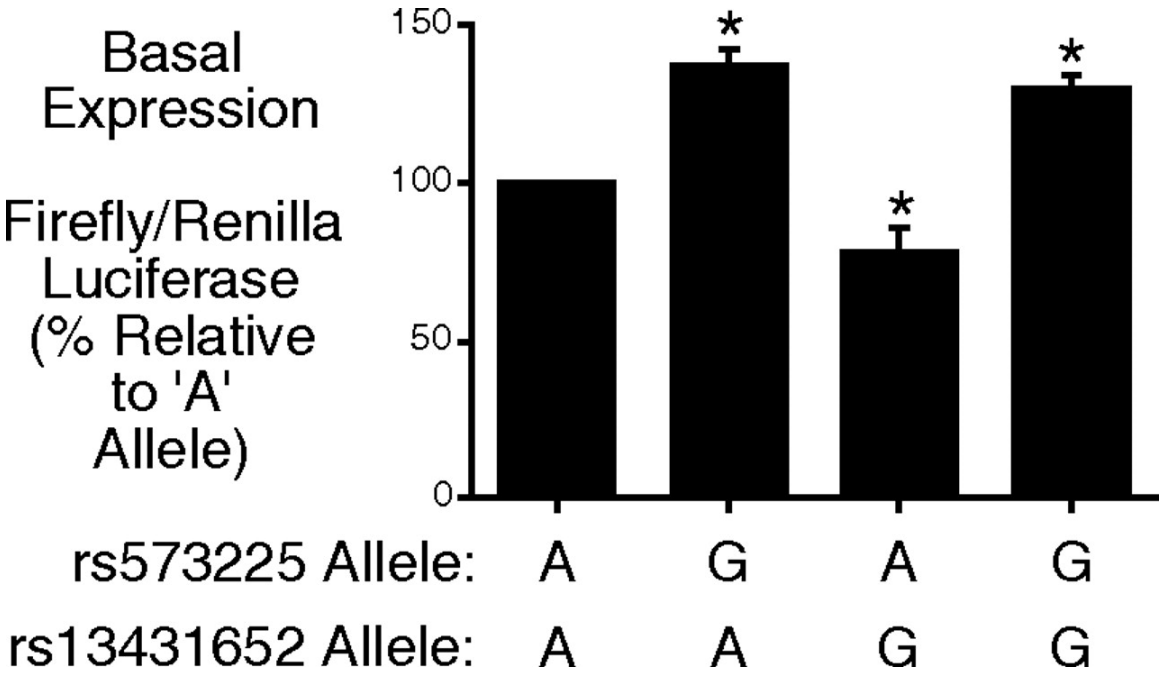


Figure 3.11. Interaction between the *G6PC2* rs13431652-G and rs573225-G alleles on *G6PC2* promoter activity

βTC-3 cells were transiently cotransfected, as described in Chapter II, using a lipofectamine solution containing various *G6PC2*-luciferase fusion genes in the pGL4 MOD vector (2 μg) and an expression vector encoding *Renilla* luciferase (0.5 μg). The *G6PC2*-luciferase fusion genes represented the rs573225-A or rs573225-G alleles and rs13431652-A or rs13431652-G alleles present in the context of the human *G6PC2* promoter sequence located between -8563 and +11. After transfection, cells were incubated for 18-20 h in serum-containing medium. The cells were then harvested, and firefly and *Renilla* luciferase activity were assayed as described in Chapter II. Results are presented as the ratio of firefly:*Renilla* luciferase activity, expressed as a percentage relative to the value obtained with the rs573225-A and rs13431652-A alleles, and represent the mean of eight experiments ± S.E.M., each using an independent preparation of each fusion gene plasmid, assayed in triplicate. *$P < 0.05$ vs. rs573225-A and rs13431652-A alleles.

decreases Foxa2 binding (Fig. 3.7C), we anticipated that the rs573225-G allele would also be associated with reduced *G6PC2* promoter activity. Surprisingly, Figures 3.10 and 3.11 show that the rs573225-G allele was actually associated with an increase in promoter activity in comparison to that observed with the rs573225-A allele. This result is not explained by the *de novo* creation of an activator binding site by the rs573225-G allele (Fig. 3.6B). Figures 3.12 and 3.13 show that a similar effect of the rs573225-G allele was also seen in the HIT and Min6 cell lines and when the *G6PC2* promoter was analyzed in the context of both the pGL4 and pGL3 vectors, respectively. Since the Foxa2 binding site incorporates the target sequence for the bacterial Dam methylase (GATC), we repeated this analysis using plasmids grown in SCS110 Dam and Dcm methylase deficient bacteria. Figure 3.14 shows that Dam methylation of the Foxa2 binding site did not alter the stimulatory effect of rs573225 on fusion gene expression. While the increased promoter activity observed with the rs573225-G allele was unexpected, given the reduction in Foxa2 binding, the key observation here is that the functional data obtained with rs573225 are not consistent with the reduced FPG observed in *G6pc2* knockout mice [55] in that the rs573225-A allele is associated with elevated FPG (Table 3.1) but reduced *G6PC2* promoter activity. As such, in contrast to the functional data obtained with rs13431652, these functional data do not support a potential role for rs573225 as a causative SNP linking *G6PC2* to variations in FPG.

Because endogenous *G6PC2* gene expression will reflect the combined effects of multiple SNPs we investigated the interaction between rs573225 and rs13431652 on *G6PC2* fusion gene expression (Fig. 3.11). The results show a trend toward the rs13431652-G allele blunting the elevated expression conferred by the rs573225-G allele,

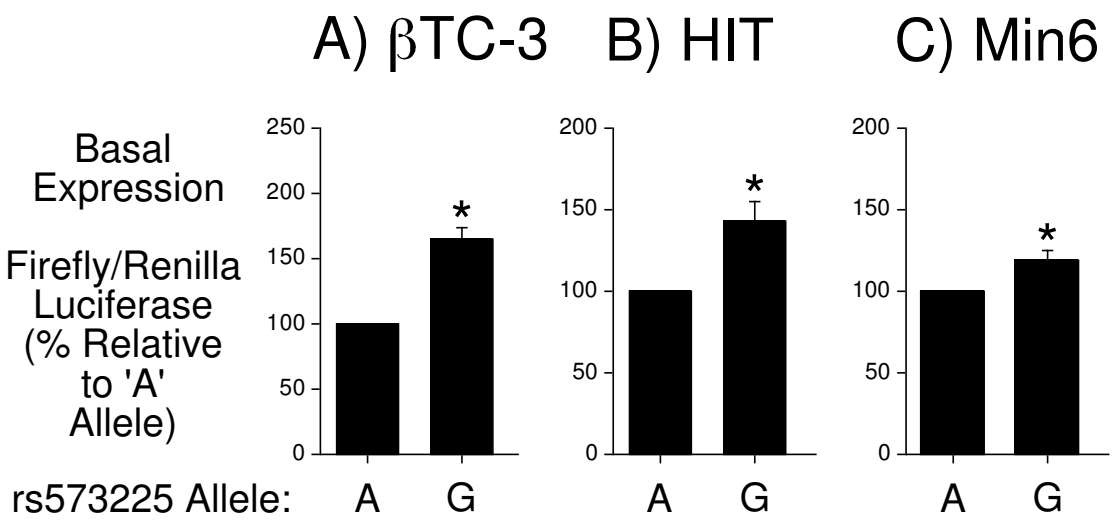


Figure 3.12. Increased *G6PC2* promoter activity associated with the human *G6PC2* rs573225-G allele relative to the rs573225-A allele in multiple cell lines

βTC-3 (Panel A), HIT (Panel B), and Min6 (Panel C) cells were transiently cotransfected, as described in Chapter II, using a lipofectamine solution containing various *G6PC2*-luciferase fusion genes in the pGL3 MOD vector (2 μg) and an expression vector encoding *Renilla* luciferase (0.5 μg). The *G6PC2*-luciferase fusion genes represented the rs573225-A or rs573225-G alleles alleles present in the context of the human *G6PC2* promoter sequence located between -324 and +3. After transfection, cells were incubated for 18-20 h in serum-containing medium. The cells were then harvested, and firefly and *Renilla* luciferase activity were assayed as described in Chapter II. Results are presented as the ratio of firefly:*Renilla* luciferase activity, expressed as a percentage relative to the value obtained with the rs573225-A allele, and represent the mean of three experiments ± S.E.M., each using an independent preparation of each fusion gene plasmid, assayed in triplicate. *$P < 0.05$ vs. rs573225-A allele.

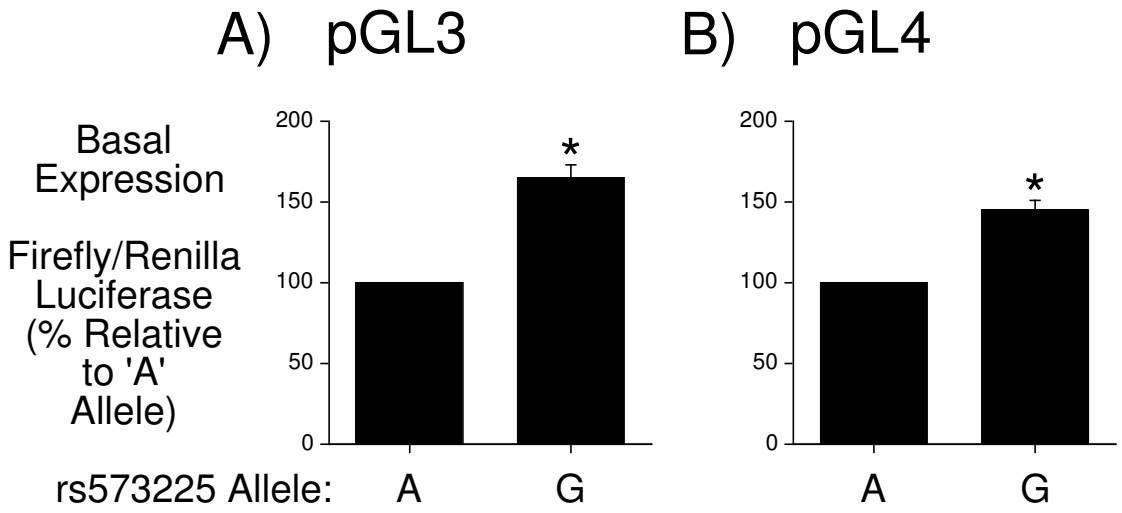


Figure 3.13. Increased *G6PC2* promoter activity associated with the human *G6PC2* rs573225-G allele relative to the rs573225-A allele in pGL3 and pGL4

βTC-3 cells were transiently cotransfected, as described in Chapter II, using a lipofectamine solution containing various *G6PC2*-luciferase fusion genes in the pGL3 MOD (Panel A) or pGL4 MOD (Panel B) vectors (2 μg) and an expression vector encoding *Renilla* luciferase (0.5 μg). The *G6PC2*-luciferase fusion genes represented the rs573225-A or rs573225-G alleles present in the context of the human *G6PC2* promoter sequence located between -324 and +3. After transfection, cells were incubated for 18-20 h in serum-containing medium. The cells were then harvested, and firefly and *Renilla* luciferase activity were assayed as described in Chapter II. Results are presented as the ratio of firefly:*Renilla* luciferase activity, expressed as a percentage relative to the value obtained with the rs573225-A allele, and represent the mean of three experiments ± S.E.M., each using an independent preparation of each fusion gene plasmid, assayed in triplicate. *$P < 0.05$ vs. rs573225-A allele.

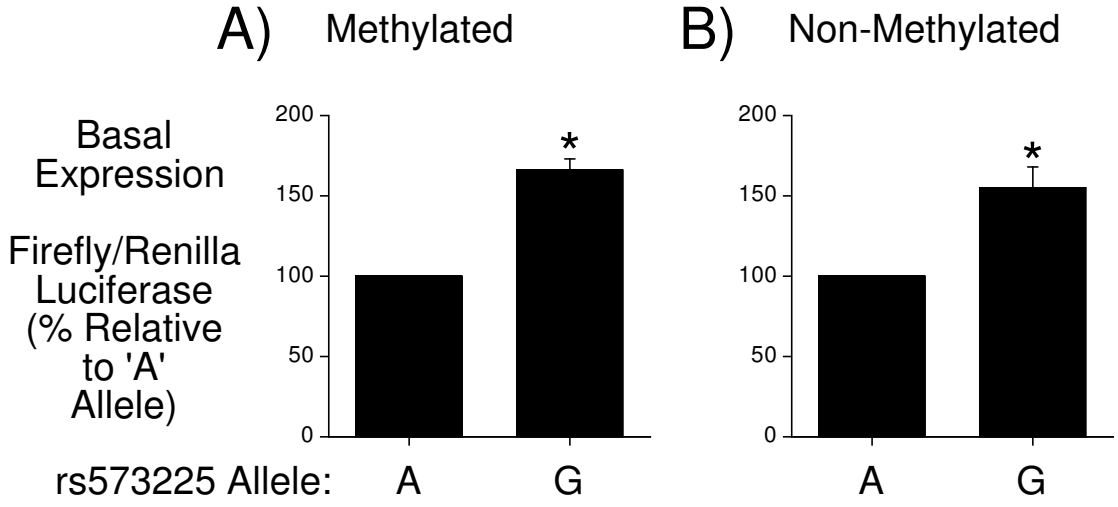


Figure 3.14. Increased *G6PC2* promoter activity associated with the human *G6PC2* rs573225-G allele relative to the rs573225-A allele independent of DNA methylation status

βTC-3 cells were transiently cotransfected, as described in Chapter II, using a lipofectamine solution containing various *G6PC2*-luciferase fusion genes in the pGL3 MOD vectors (2 μg) and an expression vector encoding *Renilla* luciferase (0.5 μg). The *G6PC2*-luciferase fusion genes represented the rs573225-A or rs573225-G alleles present in the context of the human *G6PC2* promoter sequence located between -324 and +3. The fusion gene plasmids were grown in methylase containing DH5α (Panel A) or methylase deficient SCS110 (Panel B) bacterial cells. After transfection, cells were incubated for 18-20 h in serum-containing medium. The cells were then harvested, and firefly and *Renilla* luciferase activity were assayed as described in Chapter II. Results are presented as the ratio of firefly:*Renilla* luciferase activity, expressed as a percentage relative to the value obtained with the rs573225-A allele, and represent the mean of three experiments ± S.E.M., each using an independent preparation of each fusion gene plasmid, assayed in triplicate. *$P < 0.05$ vs. rs573225-A allele.

though, because of the relatively small effects involved and the inherent variability in transient transfections, the decrease did not reach statistical significance.

Discussion

Genome-wide association studies have recently provided important new insights about the genetics of common forms of type 2 diabetes and its related quantitative traits, especially for FPG [43, 91, 107]. After this first discovery phase, attention is now beginning to focus on the study of the functional properties of the variants at the confirmed loci. In this study, using a combination of genetic and functional data we provide further genetic support for the primary contribution of *G6PC2* to variations in FPG, a conclusion supported by other genetic studies [90, 108-110]. Furthermore, it is specifically demonstrated that *1*) the 'A' allele of the common variant rs13431652 is strongly associated with elevated FPG, *2*) the rs13431652-A allele enhances NF-Y binding to the *G6PC2* promoter (Figs. 3.2 & 3.3) and *3*) increases fusion gene expression (Fig. 3.4). Although no significant correlation between endogenous *G6PC2* gene expression and rs13431652 genotypes was observed, probably due to the limited sample analyzed (N = 24), the *in vitro* data are consistent with the reduced FPG observed in *G6pc2* knockout mice [55] and the hypothesis that *G6PC2* expression would oppose the action of glucokinase leading to elevated FPG. These data suggest that rs13431652 is a strong candidate to be a causative SNP that contributes to the association signal between *G6PC2* and FPG, though it is important to note that our genetic data do not rule out a significant contribution of the intronic variant rs560887. Future studies will be designed to address our hypothesis that this SNP affects *G6PC2* mRNA splicing [43].

Our study does not give support to the contribution of intronic variant rs853789 located in the 19th intron of *ABCB11* to the association signal with FPG in our European cohorts. In a recent study conducted in Asian populations, a haplotype tagging variant

rs3755157 located in the 21st intron of *ABCB11* was identified as associated with FPG independently from rs560887, which may support the existence of two independent signals at this locus [111]. In this paper, limited attention was given to the striking differences in alleles frequencies between Asians and Europeans both at rs560887 (MAF_{EU} = 0.30 vs. MAF_{As} = 0.03) and rs3755157 (0.12 calculated in our French population and 0.09 in HapMap CEU population *vs*. MAF_{As} = 0.38). We believe that these differences may alter the efficiency of these two SNPs to tag putative causal variant(s), as reflected by the difference in the strength of the associations of these SNPs observed in Europeans vs. Asians. Nonetheless, we acknowledge that a more extensive fine mapping analysis, involving a larger number of variants in this locus, will be required to elucidate the contribution of *G6PC2* and *ABCB11* variants to the association signal and provide a more comprehensive genetic assessment of this important locus in the genetic determination of FPG.

Using a combination of genetic and functional data we show that rs573225-A allele is highly associated with elevated FPG (Table 3.1), enhanced Foxa2 binding (Figs. 3.6 & 3.7) but lower fusion gene expression (Fig. 3.10), a correlation that is inconsistent with the reduced FPG observed in *G6pc2* knockout mice [55]. As such, unlike the functional data obtained with rs13431652, these functional data do not support a potential role for rs573225 as a causative SNP linking *G6PC2* to variations in FPG. In contrast, Dos Santos et al. recently concluded that rs573225 was the causative SNP that explained the association signal between FPG and *G6PC2* [92]. As in our study they observed that the rs573225-A allele is associated with elevated FPG but lower fusion gene expression but they did not comment on the fact that these data are inconsistent with the function of

G6PC2. Dos Santos et al. also suggested that rs573225 is an epiSNP because the rs573225-G allele is located at a GpC dinucleotide within the Foxa2 binding site and methylation of the 'C' nucleotide affected Foxa2 binding [92]. However, in contrast to the well-studied methylation of CpG dinucleotides [112], we can find no reports of the existence of methylated GpC dinucleotides in mammals, though such a modification does exist in fish [113]. Finally, while our fusion gene data match that of Dos Santos et al. [92] our gel retardation results are the complete opposite. Dos Santos et al. report that the rs573225-G allele binds Foxa2 whereas the rs573225-A allele completely abolishes Foxa2 binding [92]. Their binding data therefore appear to correlate with their fusion gene data since the rs573225-G allele confers higher promoter activity. However, a detailed previous study by Overdier et al. showed that both 'G' and 'A' nucleotides at this location within the Foxa2 binding site support Foxa2 binding [114]. Our demonstration that Foxa2 binds with the same affinity to both alleles (Fig. 3.7A & B), though with different kinetics (Fig. 3.7C), is therefore consistent with the observations of Overdier et al. [114], but not those of Dos Santos et al. [92]. Moreover, given the importance of Foxa2 for *G6PC2* fusion gene expression (Fig. 3.9), if the rs573225-A allele were to abolish Foxa2 binding Dos Santos et al. should have detected a major difference in the level of reporter gene expression conferred by the promoters containing the rs573225-A and -G alleles instead of the ~15% difference reported [92].

Although the available data does not support a role for rs573225 as a causative SNP, a caveat with the experiments reported here is that the functional effect of rs573225 on endogenous *G6PC2* gene transcription *in vivo* may not be replicated by analyses involving the transient transfection of fusion genes in islet-derived cell lines. Thus, fusion

gene experiments have several limitations including *1*) the absence of chromatin structure in transient transfection assays [115], *2*) the use of a truncated fusion gene that lacks the five known *G6PC2* transcriptional enhancers [75], and *3*) the use of tissue culture cell lines whose gene expression profiles do not match that of islets *in vivo*. All of these factors have the potential to alter the affect of rs573225 on *G6PC2* gene transcription. Finally, Foxa2 is known to act as a pioneer factor opening chromatin to allow access of other transcription factors [116]. As such, rs573225 may have different effects on *G6PC2* gene expression early in development than in adults.

In summary, our study provides genetic and functional evidence supporting an important role for the promoter variant rs13431652 as a potentially causative SNP that contributes to the association signal between *G6PC2* and FPG but the data do not preclude a significant contribution of the intronic variant rs560887, or other additional unidentified variants.

CHAPTER IV

FUNCTIONAL DATA SUPPORT A POTENTIAL ROLE FOR rs2232316 AS A CAUSATIVE *G6PC2* SNP CONTRIBUTING TO VARIATION IN FPG

Introduction

The conflicting results regarding the rs573225 SNP in the previous set of experiments necessitated further exploration of this SNP, as well as another, rs2232316, located in its vicinity. Because the rs573225-A allele was associated with increased FPG but reduced promoter activity in fusion gene analyses [96], at odds with the putative function of G6PC2 in pancreatic islets, it is possible that the behavior of this SNP in tissue culture cells does not accurately reflect its effect on *G6PC2* gene transcription *in vivo*. Conversely, the consistencies seen between the genetic and promoter activity data for rs13431652 and the conclusions drawn from them in previous studies [96] may be incorrect in the unlikely event that human G6PC2 and mouse G6pc2 have opposite effects on FPG, in which case, the behavior of rs13431652 in tissue culture cells may not accurately portray the effect of this SNP on *G6PC2* gene transcription *in vivo*. To lend credence to one of these two possibilities, additional functional studies were performed on the role of rs573225. Furthermore, new studies were carried out to examine the effect of rs2232316 on transcription factor binding, *G6PC2* fusion gene expression, and its genetic association with FPG in humans.

Results

rs2232316 alters Foxa2 binding to the *G6PC2* promoter *in vitro*

The *G6PC2* promoter region that encompasses rs2232316, located at -238 relative to the transcription start site, was analyzed using MatInspector sequence analysis software [99] with the goal of identifying a *cis*-acting element whose binding of its cognate *trans*-acting factor was likely to be affected by the alternate rs2232316 alleles. This analysis identified a forkhead transcription factor binding motif [117].

While several members of the Fox family are expressed in the pancreas [117], members of the Foxa sub-group have been shown to play important roles in islet development and function [105]. Gel retardation assays were used to investigate whether any of the three Foxa isoforms [105] can bind to this *G6PC2* promoter region *in vitro*. When a labeled double-stranded oligonucleotide, designated rs2232316-A, representing the *G6PC2* promoter sequence from -250 to -226 and the rs2232316-A allele (Fig. 4.1A), was incubated with nuclear extract prepared from βTC-3 cells a single protein-DNA complex was detected (Fig. 4.1B). To identify the factor present in this complex a gel retardation assay was performed in which βTC-3 cell nuclear extract was pre-incubated with antisera specific for Foxa2 or, as a control, USF-2. Previous studies have shown that Foxa2 is the predominant Foxa isoform in βTC-3 cells [79]. As can be seen in Figure 4.1B, addition of antibodies recognizing Foxa2 resulted in a clear supershift in the migration of the complex whereas addition of antibodies recognizing USF-2 had no effect. This result strongly suggests that the complex represents Foxa2 binding.

Gel retardation competition experiments, in which a varying molar excess of

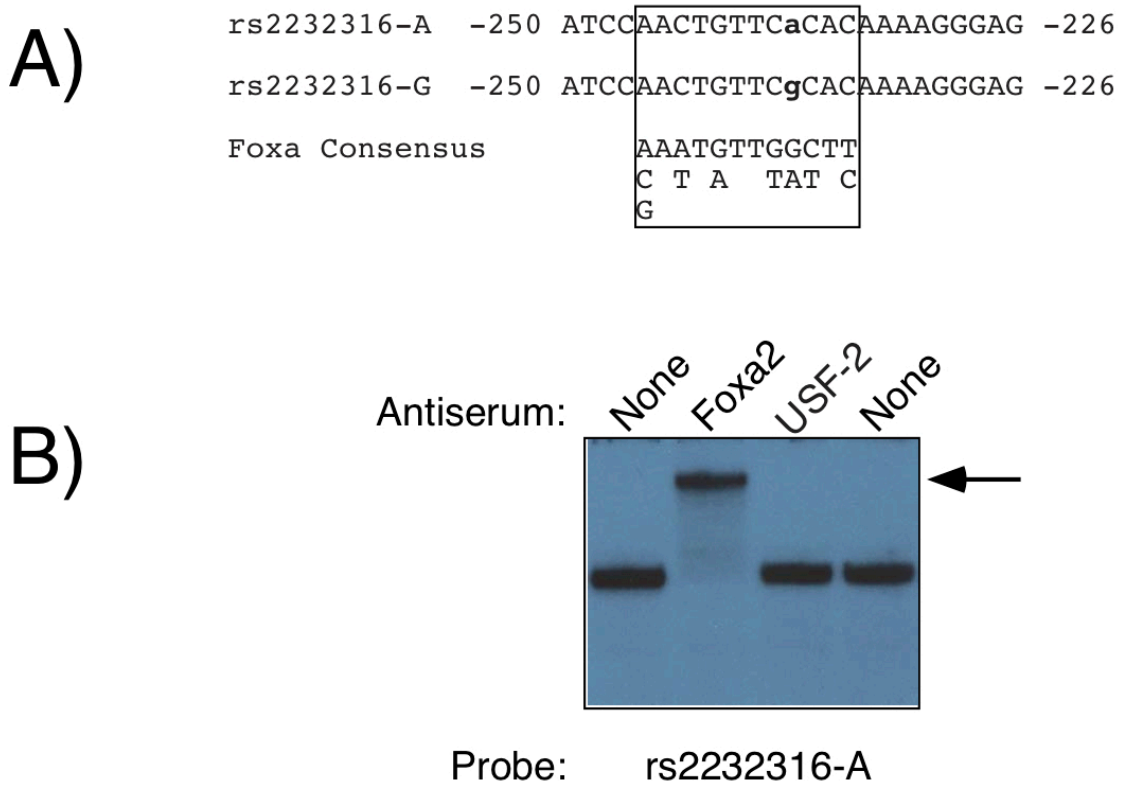
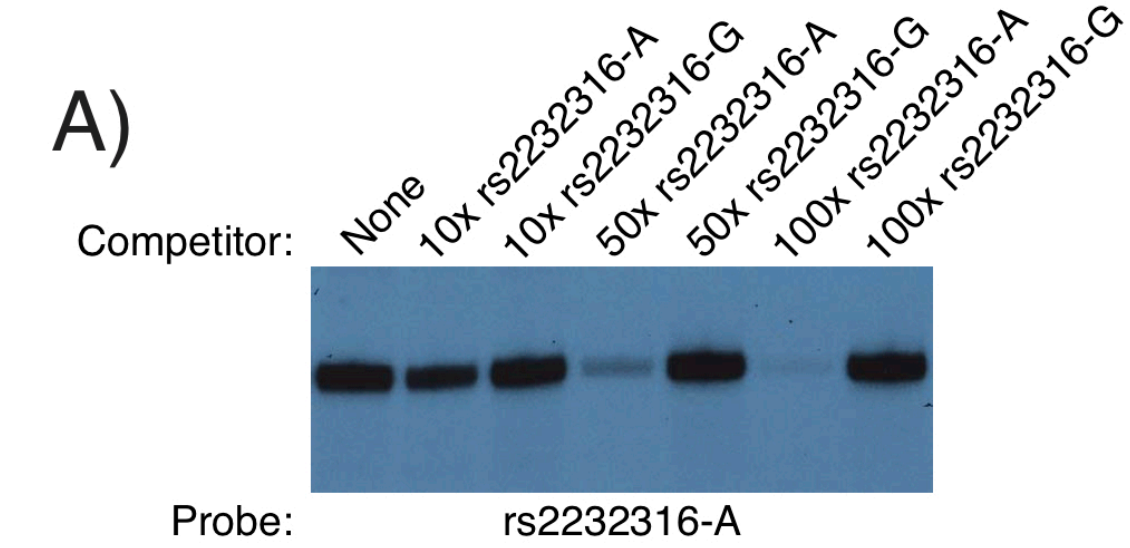


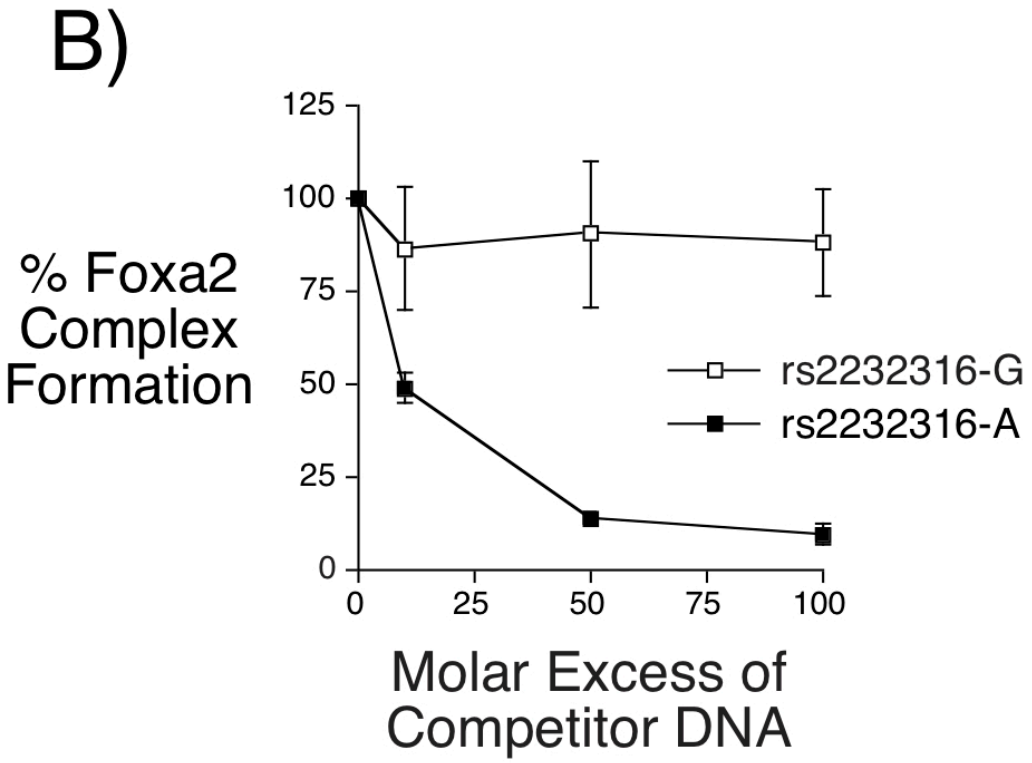
Figure 4.1. Binding of the -250/226 promoter region of *G6PC2* by Foxa2 *in vitro*
(A) The sense strand sequences of the oligonucleotides used in gel retardation assays are shown. SNP base pairs are shown in bold lower case letters. The consensus Foxa binding motif, (A/G)A(G/A)(C/T)(C/A)AA(C/T)A(T/A)T(T/G/C), is taken from Overdier et al. (B) βTC-3 nuclear extract was incubated in the absence (None) or presence of the indicated anti-serum for 10 minutes on ice. A labeled oligonucleotides representing the -250/-226 *G6PC2* promoter region and containing the rs2232316-A allele (rs2232316-A; Panel A) was then added and incubation continued for 20 min at room temperature. Protein binding was then analyzed using the gel retardation assay as described in Chapter II. In the representative autoradiograph shown, only the retarded complexes are visible and not the free probe, which was present in excess. A single major complex was detected that represents Foxa2 binding. The arrow indicates a supershifted complex.

unlabeled DNA was included with the labeled rs2232316-A oligonucleotide, were used to compare the affinity of Foxa2 binding to the rs2232316 variants of this *G6PC2* Foxa2 binding site. Figure 4.2 (Panels A and B) shows that the rs2232316-A oligonucleotide competed effectively for the formation of the Foxa2-DNA complex whereas the rs2232316-G oligonucleotide did not. Consistent with the competition experiment data (Fig. 4.2A and B), the direct analysis of Foxa2 binding to the rs2232316-A and -G oligonucleotide probes, labeled with the identical specific activity, showed a dramatic difference in Foxa2 binding affinity (Fig. 4.2C). Interestingly, the original consensus sequence for Foxa2 binding (Fig. 4.1A; Ref. [114]) indicates that either an 'A' or a 'G' nucleotide at the location of the rs2232316 SNP will support Foxa2 binding, but clearly this is not the case (Fig. 4.2C). However, this consensus was established using a sequential selection and amplification of binding sites protocol [114]. In contrast, an analysis of endogenous Foxa2 binding sites based on ChIP-Seq data [118] indicates that an 'A' nucleotide is preferred at the location of the rs2232316 SNP *in vivo*, consistent with the binding data (Fig. 4.2C). Both methods indicate that a 'G' or a 'T' nucleotide is preferred at the location immediately 5' of the rs2232316 SNP [114, 118], whereas in the *G6PC2* promoter a 'C' nucleotide is present at this location (Fig. 4.1A). We speculate that the presence of this 'C' nucleotide may have accentuated the preference for an 'A' nucleotide at the location of the adjacent rs2232316 SNP.

rs2232316 alters *G6PC2* fusion gene expression in βTC-3 cells

We next investigated the functional significance of altered Foxa2 binding on human *G6PC2* promoter activity. Fusion genes containing each of the rs2232316 alleles,


A)
Competitor:
None
10x rs2232316-A
10x rs2232316-G
50x rs2232316-A
50x rs2232316-G
100x rs2232316-A
100x rs2232316-G
Probe:
rs2232316-A


B)
% Foxa2 Complex Formation
125
100
75
50
25
0
0
25
50
75
100
rs2232316-G
rs2232316-A
Molar Excess of Competitor DNA

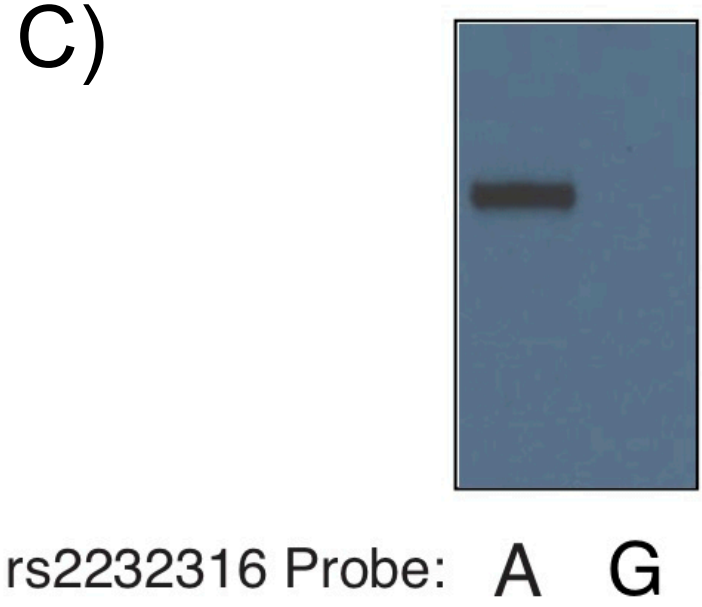


Figure 4.2. Comparison of the effect of the rs2232316-A and -G alleles on Foxa2 binding affinity *in vitro*
(A) The labeled rs2232316-A oligonucleotide (Fig. 4.1A) was incubated in the absence or presence of the indicated molar excess of the unlabeled human rs2232316-A or –G oligonucleotide competitors (Fig. 4.1A) prior to the addition of βTC-3 cell nuclear extract. Protein binding was then analyzed using the gel retardation assay as described in Chapter II. In the representative autoradiograph shown only the retarded complex is visible and not the free probe, which was present in excess. (B) Protein binding, specifically ^{32}P associated with the retarded complex, was quantified as described in Chapter II. The data represents the mean ± S.E.M. of three experiments. (C) The labeled rs2232316-A and –G oligonucleotides (Fig. 4.1A) were incubated with βTC-3 cell nuclear extract and protein binding was analyzed using the gel retardation assay as described in Chapter II. In the representative autoradiograph shown, only the retarded complexes are visible and not the free probe, which was present in excess. NS, non-specific.

generated in the context of the -324 to +3 *G6PC2* promoter region, were analyzed by transient transfection of βTC-3 cells. Figure 4.3 shows that the rs2232316-A allele was associated with an approximately 50% increase in promoter activity in comparison to that observed with the rs2232316-G allele. Figure 4.4 shows that a similar effect of the rs2232316-A allele was also seen in the HIT and Min6 islet-derived cell lines. In summary, the functional and genetic data for rs2232316 correlate, with the rs2232316-A allele being associated with both elevated *G6PC2* promoter activity (Fig. 4.3) and FPG (Table 4.1), thereby supporting a potential role for rs2232316 as a causative SNP linking *G6PC2* to variations in FPG.

rs2232316 does not influence the effect of rs573225 on *G6PC2* fusion gene expression in βTC-3 cells

We recently showed that the alternate alleles of rs573225 also alter Foxa2 binding to the *G6PC2* promoter *in vitro* [96]. The Foxa2 binding sites that encompass rs573225 (-261/-250) and rs2232316 (-246/-235) are located 3 bp apart in the *G6PC2* promoter and this region is highly conserved in the mouse *G6pc2* promoter [36]. Furthermore, chromatin immunoprecipitation (ChIP) assays show that Foxa2 binds this region of the endogenous *G6pc2* promoter in mouse βTC-3 cells *in situ* [79].

rs573225 has a paradoxical effect on *G6PC2* fusion gene expression. A block mutation in the *G6PC2* Foxa2 binding site that encompasses rs573225 abolishes Foxa2 binding and also markedly reduces fusion gene expression indicating that Foxa2 acts as an activator when bound to this site in the human *G6PC2* promoter [96], as it does when bound to the equivalent site in the mouse *G6pc2* promoter [79]. Since the loss of Foxa2 binding to this site reduces *G6PC2* promoter activity and because the rs573225-G allele

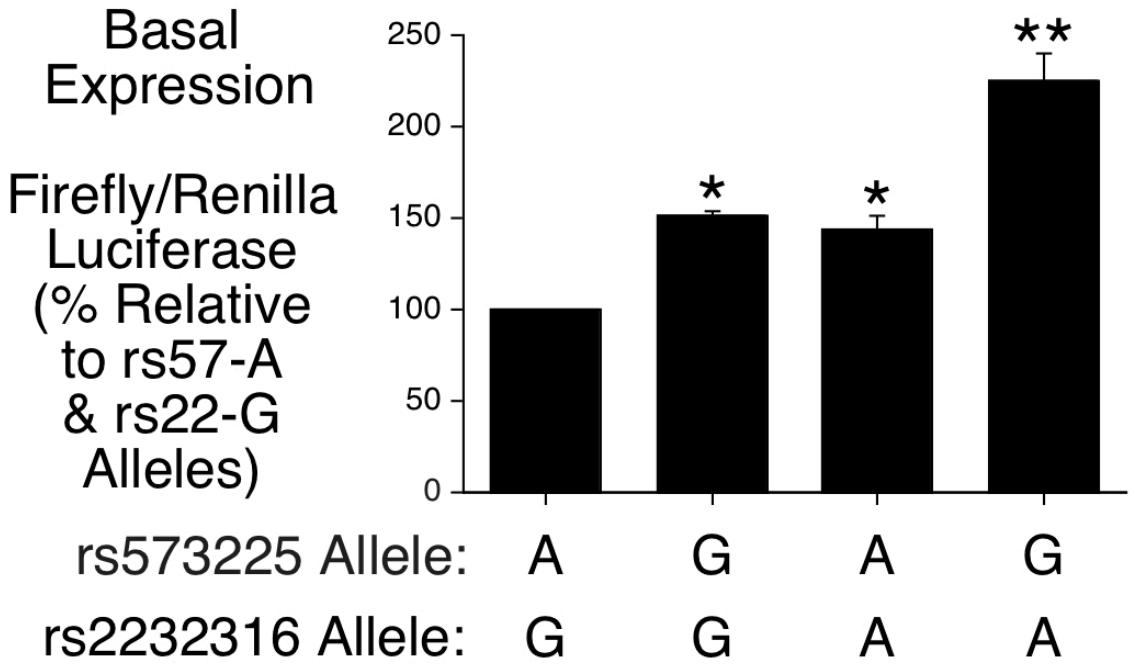


Figure 4.3. Increased *G6PC2* promoter activity associated with the human *G6PC2* rs2232316-A allele relative to the rs2232316-G allele in βTC-3 cells

βTC-3 cells were transiently co-transfected, as described in Chapter II, using a lipofectamine solution containing various *G6PC2*-luciferase fusion genes in the pGL3 MOD vector (2 μg) and an expression vector encoding *Renilla* luciferase (0.5 μg). The *G6PC2*-luciferase fusion genes represented the rs2232316-A or rs2232316-G alleles and rs573225-A or rs573225-G alleles present in the context of the human *G6PC2* promoter sequence located between -324 and +3. After transfection, cells were incubated for 18-20 h in serum-containing medium. The cells were then harvested, and firefly and *Renilla* luciferase activity were assayed as described in Chapter II. Results are presented as the ratio of firefly:*Renilla* luciferase activity, expressed as a percentage relative to the value obtained with the rs2232316-G and rs573225-A alleles, and represent the mean of three experiments ± S.E.M., each using an independent preparation of each fusion gene plasmid, assayed in triplicate. **P* < 0.05 vs. rs2232316-G and rs573225-A allele. ***P* < 0.05 vs. (rs2232316-G and rs573225-G) and (rs2232316-A and rs573225-A) alleles.

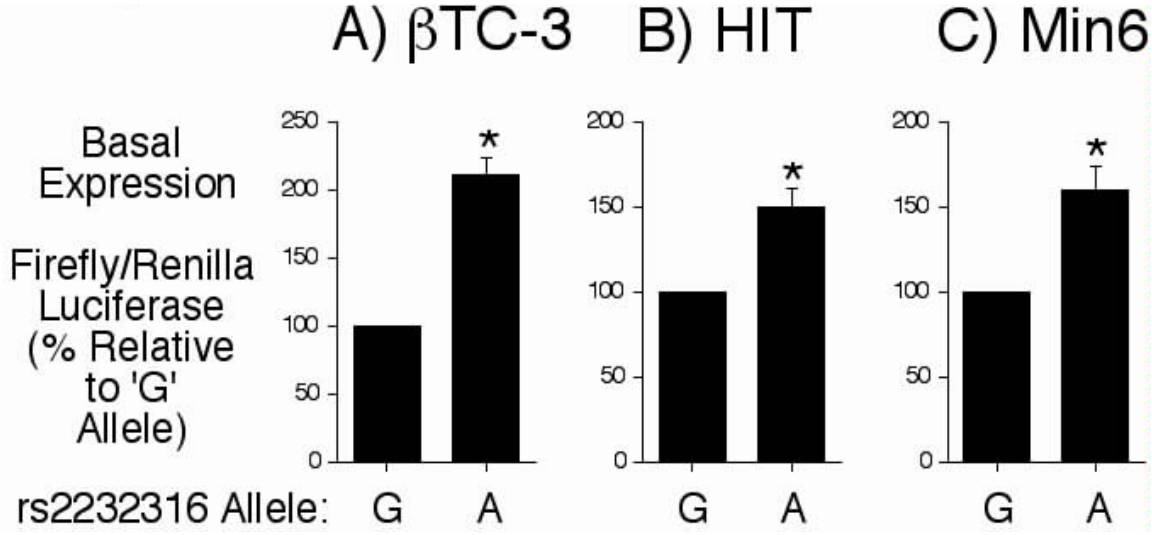


Figure 4.4. Increased *G6PC2* promoter activity associated with the human *G6PC2* rs2232316-A allele relative to the rs2232316-G allele in multiple cell lines
βTC-3 (Panel A), HIT (Panel B), and Min6 (Panel C) cells were transiently cotransfected, as described in Chapter II, using a lipofectamine solution containing various *G6PC2*-luciferase fusion genes in the pGL3 MOD vector (2 μg) and an expression vector encoding *Renilla* luciferase (0.5 μg). The *G6PC2*-luciferase fusion genes represented the rs2232316-A or rs2232316-G alleles alleles present in the context of the human *G6PC2* promoter sequence located between -324 and +3. After transfection, cells were incubated for 18-20 h in serum-containing medium. The cells were then harvested, and firefly and *Renilla* luciferase activity were assayed as described in Chapter II. Results are presented as the ratio of firefly:*Renilla* luciferase activity, expressed as a percentage relative to the value obtained with the rs573225-A allele, and represent the mean of three experiments ± S.E.M., each using an independent preparation of each fusion gene plasmid, assayed in triplicate. *$P < 0.05$ vs. rs2232316-G allele.

Table 4.1. Association with fasting plasma glucose in D.E.S.I.R. participants
The table shows the effect allele frequencies, genotype groups and numbers, mean fasting plasma glucose (with standard deviation per genotype group), regression coefficient β (with standard error) and associated p-value (*P*).

SNP	Effect allele (Freq.)	Genotype (N)	Mean FG (SD)	β (SE)	*P*
rs573225	A (0.66)	GG (471)	5.11 (0.02)		
		GA (1847)	5.16 (0.01)		
		AA (1865)	5.25 (0.01)	0.08 (0.01)	$3.7x10^{-17}$
rs2232316	A (0.11)	GG (3320)	5.18 (0.01)		
		GA (850)	5.24 (0.01)		
		AA (49)	5.21 (0.07)	0.05 (0.01)	$7.2x10^{-4}$

decreases Foxa2 binding [96] we anticipated that the rs573225-G allele would also be associated with reduced *G6PC2* promoter activity. Surprisingly, the rs573225-G allele is associated with an increase in promoter activity in comparison to that observed with the rs573225-A allele (Fig. 4.3) [96]. Because the Foxa2 binding sites that encompass rs573225 and rs2232316 are so close we hypothesized that Foxa2 binding at one element may modulate the action of Foxa2 at the other. The interaction between these SNPs on *G6PC2* fusion gene expression was therefore investigated (Fig. 4.3). The results show that the rs573225-G allele and the rs2232316-A allele have an additive and, therefore, apparently independent effect on reporter gene expression (Fig. 4.3).

The rs2232316 *G6PC2* promoter variant is associated with FPG but not independently from rs573225

We assessed the association of the promoter variant rs2232316 with FPG levels in 4,364 normal fasting glucose participants (FPG < 6.1 mmol/L) from the D.E.S.I.R cohort. In accord with the fusion gene data, we found that the rs2232316-A allele associates with increased FPG ($\beta = 0.05$, $P = 7.2\times10^{-4}$) (Table 4.1). This association was replicated ($\beta = 0.04$, $P = 7.3\times10^{-10}$) using *in silico* analysis computed in 45,688 non-diabetic participants of European descent in MAGIC (Meta-Analysis of Glucose and Insulin Consortium) [119]. Conditioned regression model analyses in D.E.S.I.R to assess the independency from rs573225 showed a lack of significance of rs2232316 ($P = 0.26$) when we took into account the effect of rs573225. Specifically no significant effect of rs2232316 was found in individuals where we fixed rs573225 (association of rs2232316 with FPG in rs573225-AA carriers: $\beta = 0.02$, $P = 0.17$; rs573225-AG carriers: $\beta = 0.01$, $P = 0.77$; and rs573225-GG carriers: $\beta = -0.39$, $P = 0.17$). However, despite a low correlation between the two

variants (r^2 = 0.04, HapMap III data, r^2 = 0.06 in D.E.S.I.R.) due to important differences in frequency of the effect alleles ($Freq_{rs573225\text{-}A}$ = 0.67 *vs*. $Freq_{rs2232316\text{-}A}$ = 0.11), we observed high linkage disequilibrium between the two variants (D' = 0.98) indicating that both associations represent a single signal. As such, this implies that both rs573225 and rs2232316 have the potential to be causative variants.

In summary the genetic data support a potential role for rs2232316 as a causative SNP linking *G6PC2* to variations in FPG. In the case of rs2232316, the functional and genetic data correlate with the association between the rs2232316-A allele with elevated *G6PC2* promoter activity (Fig. 4.3) and FPG (Table 4.1).

rs573225 and rs2232316 influence *G6PC2* fusion gene expression in an artificial cell system

Of the SNPs we have analyzed to date, the three that affected *G6PC2* promoter activity (rs13431652, rs573225 and rs2232316) all affected FPG (Table 4.2). Since it seems highly unlikely that mouse G6pc2 and human G6PC2 have opposite functions given the high level of sequence conservation [36], the data for rs13431652 and rs2232316 are consistent with the reduced FPG observed in *G6pc2* knockout mice [55] and therefore a role for these variants as potentially causative SNPs (Table 4.1). These data also imply that the behavior of rs573225 in several islet-derived cell lines (Fig. 3.12; Ref. [96]) does not accurately reflect its effect on *G6PC2* gene transcription *in vivo*.

Given this apparent limitation of islet-derived cell lines with respect to rs573225 we decided to examine the action of rs573225 and rs2232316 on *G6PC2* fusion gene expression in a previously described reconstituted cell system in which HeLa cells, a cervix-derived cell line [120], are co-transfected with *G6PC2* fusion genes and

Table 4.2. Summary of the functional and genetic properties of *G6PC2* promoter SNPs

SNP	Promoter Location	Effect on Binding	Effect on Transcription in βTC-3 Cells	Effect on Transcription in HeLa Cells	Effect on FPG
rs13431652-A	-4385	NF-Y Binding Up	Up	N.D.	Up
rs573225-A	-259	Foxa2 Binding Up	Down	Up	Up
rs2232316-A	-238	Foxa2 Binding Up	Up	Down	Up

expression vectors encoding the islet-enriched transcription factors Pdx-1, Pax-6, MafA, and NeuroD, with its heterodimerization partner E47 [79]. Individually, with the exception of NeuroD, none of these transcription factors markedly induce *G6PC2* fusion gene expression in HeLa cells whereas together they act synergistically to strongly induce expression [79]. The results show that in HeLa cells, in contrast to βTC-3 cells (Fig. 4.3), the rs573225-G allele is now associated with reduced *G6PC2* fusion gene expression (Fig. 4.5), consistent with the genetic data (Table 4.1). This effect was blunted when Foxa2 was also overexpressed (Fig. 4.6). Similarly in HeLa cells, in contrast to βTC-3 cells (Fig. 4.3), the rs2232316-A allele is now associated with reduced *G6PC2* fusion gene expression (Fig. 4.5), an effect that was reversed when Foxa2 was also overexpressed (Fig. 4.6). While these latter functional data with rs2232316 remain consistent with the genetic data, these HeLa cell experiments suggest that, under specific conditions, the rs573225-G allele is associated with reduced *G6PC2* fusion gene expression, consistent with the genetic data. These data also suggest that the action of rs573225 and rs2232316 on *G6PC2* gene expression *in vivo* will be critically dependent on Foxa2 levels.

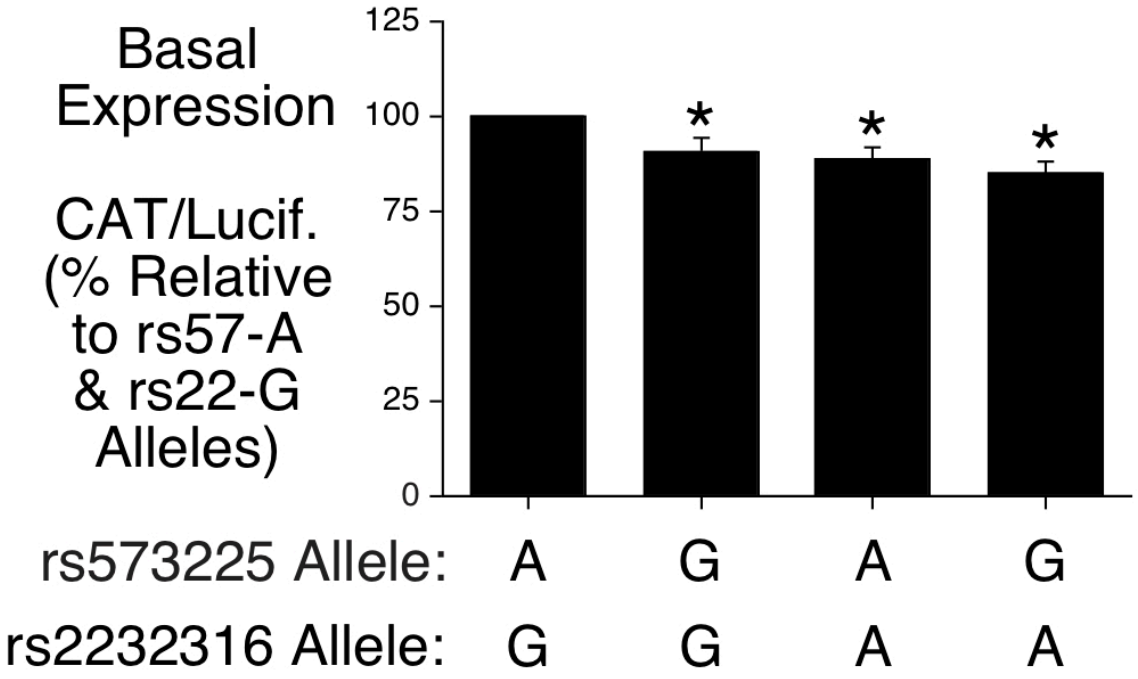


Figure 4.5. Increased *G6PC2* promoter activity associated with the human *G6PC2* rs573225-A allele relative to the rs573225-G allele in HeLa cells

HeLa cells were transiently co-transfected, as described in Chapter II, using a lipofectamine solution containing various *G6PC2*-CAT fusion genes (2 μg), a SV40-firefly luciferase expression vector (0.5 μg), and expression vectors (0.02 μg each) encoding Pdx-1, Pax-6, MafA, NeuroD and E47. The *G6PC2*-CAT fusion genes represented the rs2232316-A or –G alleles and rs573225-A or rs573225-G alleles present in the context of the human *G6PC2* promoter sequence located between -324 and +3. After transfection, cells were incubated for 18-20 h in serum-free medium. The cells were then harvested, and both CAT and luciferase activities were assayed as described in Chapter II. Results were calculated as the ratio of CAT and luciferase activities in the same sample and are expressed as a percentage relative to the value obtained in the presence of the rs2232316-G and rs573225-A alleles. Results represent the mean of three experiments ± S.E.M., each using an independent preparation of each fusion gene plasmid with each condition, assayed in triplicate. **P* < 0.05 vs. rs2232316-G and rs573225-A allele.

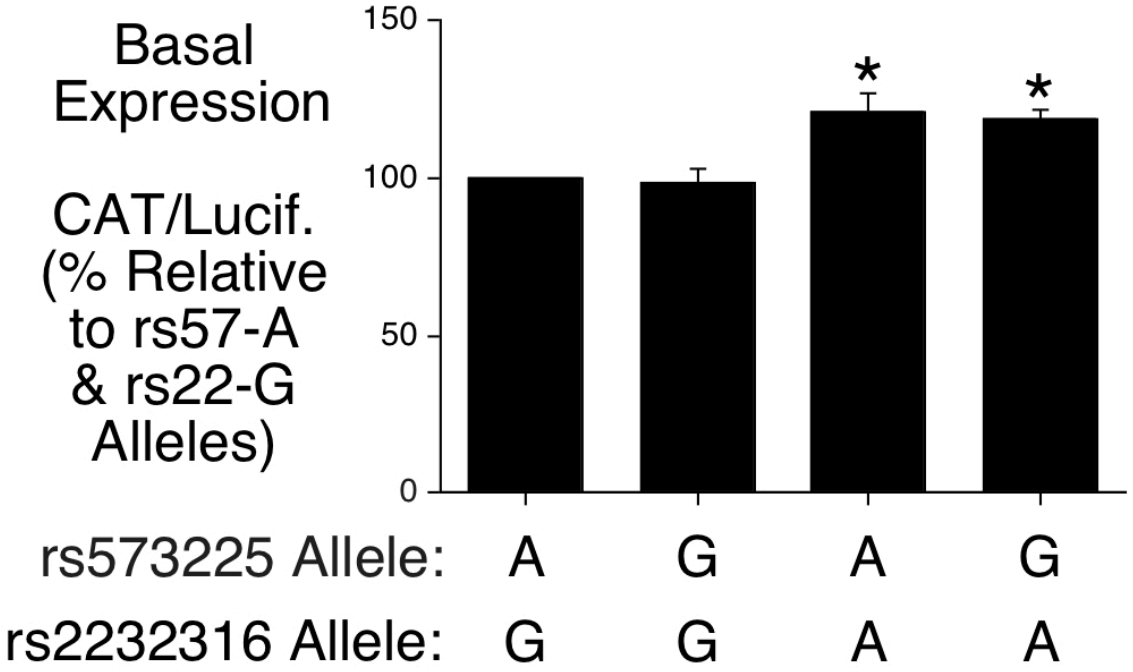


Figure 4.6. Abrogation of changes in *G6PC2* promoter activity with respect to human *G6PC2* rs573225 in HeLa cells overexpressing Foxa2

HeLa cells were transiently co-transfected, as described in Chapter II, using a lipofectamine solution containing various *G6PC2*-CAT fusion genes (2 μg), a SV40-firefly luciferase expression vector (0.5 μg), and expression vectors (0.02 μg each) encoding Pdx-1, Pax-6, MafA, NeuroD, E47, and Foxa2. The *G6PC2*-CAT fusion genes represented the rs2232316-A or –G alleles and rs573225-A or rs573225-G alleles present in the context of the human *G6PC2* promoter sequence located between -324 and +3. After transfection, cells were incubated for 18-20 h in serum-free medium. The cells were then harvested, and both CAT and luciferase activities were assayed as described in Chapter II. Results were calculated as the ratio of CAT and luciferase activities in the same sample and are expressed as a percentage relative to the value obtained in the presence of the rs2232316-G and rs573225-A alleles. Results represent the mean of three experiments ± S.E.M., each using an independent preparation of each fusion gene plasmid with each condition, assayed in triplicate. *$P < 0.05$ vs. rs2232316-G and rs573225-A allele.

Discussion

Genome-wide association studies have recently provided important new insights about the genetics of common forms of type 2 diabetes and some of its related quantitative traits such as FPG and HbA_{1C} [43, 90, 91, 107-110, 119, 121-124]. After this first discovery phase, attention is now beginning to focus on the analysis of the functional properties of the variants at the confirmed loci. Such studies can initially be driven by genetic data [96] or, as here, by the identification of a SNP with a marked functional effect. In either case, it is the combination of genetic and functional data together that are essential for assessing the importance of individual human genetic variants.

In a previous study we examined the contribution of two *G6PC2* promoter SNPs, rs13431652 and rs573225, to the association signal between *G6PC2* and FPG. The rs13431652-A allele was associated with increased FPG and elevated *G6PC2* promoter activity, whereas the rs573225-A allele was associated with increased FPG but reduced *G6PC2* promoter activity [96]. The rs13431652 data are consistent with the function of G6PC2 in pancreatic islets based on the analysis of *G6pc2* knockout mice [55] and the resulting hypothesis that G6PC2 opposes the action of glucokinase, such that elevated *G6PC2* expression would lead to elevated FPG. Since it is unlikely that human G6PC2 and mouse G6pc2 have opposite effects on FPG, the most plausible interpretation of these data is that the behavior of the rs573225 SNP in islet-derived cell lines does not accurately reflect its effect on *G6PC2* gene transcription *in vivo*. This hypothesis is consistent with the results of the present study in which, again using a combination of genetic and functional data, we show that in islet-derived cell lines the behavior of a second promoter SNP accurately reflects its effect on *G6PC2* gene transcription *in vivo*.

Thus the rs2232316-A allele is associated with elevated *G6PC2* promoter activity (Fig. 4.3) and FPG (Table 4.1). Furthermore, we show that regardless of the nature of the rs2232316 allele, the rs573225-A allele that is associated with increased FPG (Table 4.1) is associated with reduced *G6PC2* promoter activity (Fig. 4.3). Therefore while the genetic data support a potential role for rs13431652, rs2232316, and rs573225 as causative SNPs linking *G6PC2* to variations in FPG, the functional data suggest that only the action of rs13431652 and rs2232316 on *G6PC2* fusion gene expression in islet-derived cell lines mimics its effect on endogenous G6PC2 gene expression *in vivo*. Strikingly, in contrast to the results obtained using islet-derived cell lines, functional data from a HeLa cell reconstituted cell system (Fig. 4.5) showed that the rs573225-A allele was associated with elevated *G6PC2* promoter activity which correlates with the genetic data (Table 4.1) and is consistent with the function of G6PC2 in pancreatic islets, thereby supporting a role for rs573225 as a causative SNP. Ideally we would like to obtain additional support for these conclusions through the analysis of endogenous *G6PC2* expression in pancreatic islets; however, previous experience has shown that the limited sample sizes available from human cadavers are insufficient to detect significant correlations [96].

Dos Santos et al. [92] recently concluded that rs573225 was an epiSNP and the causative variant that explained the association signal between FPG and *G6PC2*. While our genetic data and theirs support a potential role for rs573225 as a causative SNP linking *G6PC2* to variations in FPG, as previously noted [96] their binding data were at odds with previous studies on DNA methylation in mammals [112] and Foxa2 binding [114] while their transfection data in an islet-derived cell line, as with ours, was at odds

with the role of G6PC2 in pancreatic islets [55]. While the functional analysis of rs573225 in a reconstituted cell system can resolve the latter discrepancy (Fig. 4.5 and Fig. 4.6), it remains to be determined why this SNP fails to function as expected in islet-derived cell lines. The HeLa cell experiments suggest that the action of rs573225 and rs2232316 on *G6PC2* gene expression *in vivo* will be critically dependent on Foxa2 levels. Whether Foxa2 levels differ between islet-derived cell lines and islets is unknown; however, Foxa2 levels are elevated in hepatoma cells as compared to liver [125]. Therefore the observed discrepancy may be limited to SNPs that affect Foxa2 binding. On the other hand, the observed discrepancy may reflect a broader problem with the use of islet-derived cell lines for such analyses. Interestingly, mouse *G6pc2* fusion genes that are active in islet-derived cell lines *in situ* are only active in fetal and newborn but not adult mouse islets *in vivo* [75], suggesting that the islet-derived cell lines used for these analyses do not accurately mimic the behavior of adult islet beta cells *in vivo*.

Since elevated FBG and HbA_{1C} are associated with an increased risk for the development of type 2 diabetes [95, 126, 127] and because *G6PC2* is associated with variations in fasting glucose and HbA_{1C} [43, 90, 91, 107-110, 119, 121, 122, 128], one would logically expect that *G6PC2* would also be associated with increased risk for the development of type 2 diabetes. Indeed some studies have now shown such an association [90, 110, 128] and two very recent reports suggest that elevated *G6PC2* expression may be specifically linked to the onset, rather than the progression, of beta cell failure [129] as well as a long term deterioration in glucose homeostasis [43, 130]. In contrast, other studies have shown that in some populations *G6PC2* is not associated with increased risk for the development of type 2 diabetes [119]. The molecular basis for these

differences is unclear. One possibility is that the reported association between *G6PC2* and type 2 diabetes is a false positive due to the use of relatively low sample sizes [90, 110, 128]. Another possibility is that modifier genes, whose actions are population-specific, determine whether variations in *G6PC2* will influence susceptibility to type 2 diabetes.

The observation that FPG is reduced in *G6pc2* knockout mice led to the hypothesis that the glucose-6-phosphatase activity of G6PC2 opposes the action of glucokinase, which catalyses the conversion of glucose to glucose-6-phosphate [55]. Glycolytic flux has been shown to determine the $S_{0.5}$ of glucose-stimulated insulin secretion (GSIS) and the existing paradigm in the islet field proposes that glucokinase alone is the beta cell glucose sensor [45, 131-133]. The significance of the *G6pc2* knockout mouse data is that they challenge this paradigm and suggest that G6PC2 is a fundamental inhibitory component of that sensor. Instead we hypothesize that a glucokinase/G6PC2 futile cycle acts as the beta cell glucose sensor determining glycolytic flux and the $S_{0.5}$ of GSIS. If correct, the simplest scenario would be that a reduction of *G6PC2* expression would result in a leftward shift in the concentration-response curve for GSIS in which the $S_{0.5}$ but not V_{max} of GSIS was reduced. However, in humans *G6PC2* SNPs that are associated with elevated FBG are associated with increased insulin secretion [59-61]. Because these SNPs are not associated with altered insulin sensitivity or glucose tolerance, altered pulsatility of insulin secretion has been invoked as a possible explanation for this paradoxical observation [60]. If correct, these data would imply that G6PC2 impairs the pulsatility of insulin secretion such that when *G6PC2* expression is decreased pulsatility and consequently the efficacy of insulin

signaling are enhanced, resulting in a lower requirement for insulin secretion. Future studies using *G6pc2* knockout mice should be able to resolve the impact of G6pc2 on the kinetic characteristics of GSIS and the pulsatility of insulin secretion.

In summary, our study provides genetic and functional evidence supporting an important role for the rs2232316 promoter variant, in addition to rs13431652 [96], as potentially causative SNPs that contribute to the association signal between *G6PC2* and FPG, though the data do not preclude a significant contribution of the intronic variant rs560887, or other additional unidentified variants. However, the *in situ* functional data suggest a major limitation in the use of islet-derived cell lines to extend GWAS data through the identification of causative SNPs that is specific to rs573225. This is a potentially more broadly relevant concern since many of the other genes implicated in GWAS relating to type 2 diabetes and its associated quantitative traits genes are involved in beta cell function [134].

CHAPTER V

THE rs560887 SINGLE NUCLEOTIDE POLYMORPHISM LINKED TO VARIATIONS IN FASTING PLASMA GLUCOSE ALTERS *G6PC2* RNA SPLICING

Introduction

The recent spate of GWA studies in humans, previously discussed in detail, have consistently identified a variant in *G6PC2* as the strongest common genetic determinant of FPG and HbA_{1C} levels in terms of effect size and significance [90, 107-110, 119, 121, 122, 128]. The rs560887 SNP, localized in the 3rd intron of *G6PC2*, may explain ~1% of the total variance in FPG within these genetic studies. In addition to its association with elevated FPG, the rs560887-G allele is also associated with decreased basal insulin secretion, as assessed by Homa%B, and increased risk of incidence of impaired fasting glucose over time using prospective data from the D.E.S.I.R. study [43]. The goal of the following experiments was to extend the data gleaned from GWA studies through molecular studies designed to assess the potential of rs560887 and two other SNPs, rs2232321 and rs35259259, to be causative, specifically by altering *G6PC2* RNA splicing. Additionally, the presence and contributions of a non-consensus splice junction at the 5' end of exon 4 was examined.

Results

rs560887 alters *G6PC2* RNA splicing

In higher eukaryotes, pre-mRNA splicing is mediated by multiple *cis*-elements in the pre-mRNA molecule. These elements comprise the branch point sequence, the polypyrimidine tract, the 5' and 3' splice sites and exonic/intronic splicing enhancers/silencers [85, 87, 135]. These elements bind a complex array of splicing factors that together form the spliceosome. rs560887 is located in the intron just 5' of the exon 4 splice junction and sequence analyses reveal that rs560887 lies precisely in a branch point sequence, defined as the position where intron lariats form [85, 87, 135]. The human branch point consensus sequence yUnAy, where y is C or U, and it is typically located 21-34 nucleotides upstream of the 5' end of an exon [136]. The putative branch point element 5' of exon 4 has the sequence UUUAU and it is located 26bp 5' of the exon. Both the sequence and location of this element are therefore consistent with this being a splicing branch point.

The SNP changes the sequence from UUUAU to UUUAC. While the change still matches the consensus [136] an analysis of the sequence of the RNA component of the small nuclear ribonucleoprotein (snRNP) U2, which binds the branch point element [135], demonstrates that the SNP nucleotide will form hydrogen bonds with a 'G' nucleotide in the U2 RNA. Since the 'G' will form three hydrogen bonds with the 'C' nucleotide and only two with the 'U' nucleotide, U2 binding should be stronger with the former. This then suggests that rs560887 has the potential to affect the strength of U2 binding and hence splicing. A minigene strategy, in which splicing of individual exons

can be assessed, was used to address this hypothesis [137, 138]. Specifically we utilized the RHCglo minigene plasmid vector, which contains the Rous sarcoma virus promoter, a 5' exon with a splice junction, an artificial exon surrounded by intron sequences from the human beta-globin gene and a 3' exon with a splice junction (Fig. 5.1) [139]. The artificial exon and surrounding human beta-globin gene intron sequences were replaced by *G6PC2* exon 4 and ~300bp of surrounding intronic sequences. Two minigenes were constructed containing the alternate alleles of rs560887.

These minigenes were expressed in HeLa cells using transient transfection, RNA isolated 18 hrs post-transfection and cDNA generated. Using primers A and primer B (Fig. 5.1) and a radiolabeled dATP to be incorporated through PCR, splicing between the vector 5' and 3' exons and the inserted G6PC2 exon was assessed. PCR reactions were performed under conditions where amplification was in the linear range (Fig. 5.2) and the products (X and Y; Fig. 5.3) were resolved on non-denaturing polyacrylamide gels as described [140]. DNA sequencing confirmed the identity of these products and the presence of the predicted splice junctions. Figure 5.3 shows that the alternate alleles of rs560887 alter the relative abundance of products X and Y, with the rs560887-C allele promoting inclusion of *G6PC2* exon 4. This is consistent with the genetic association between the rs560887-C allele, elevated FPG and the putative function of G6PC2 suggesting that rs560887 is a potentially causative SNP. The combination of these functional data with the genetic data supports a direct connection between *G6PC2* and FPG.

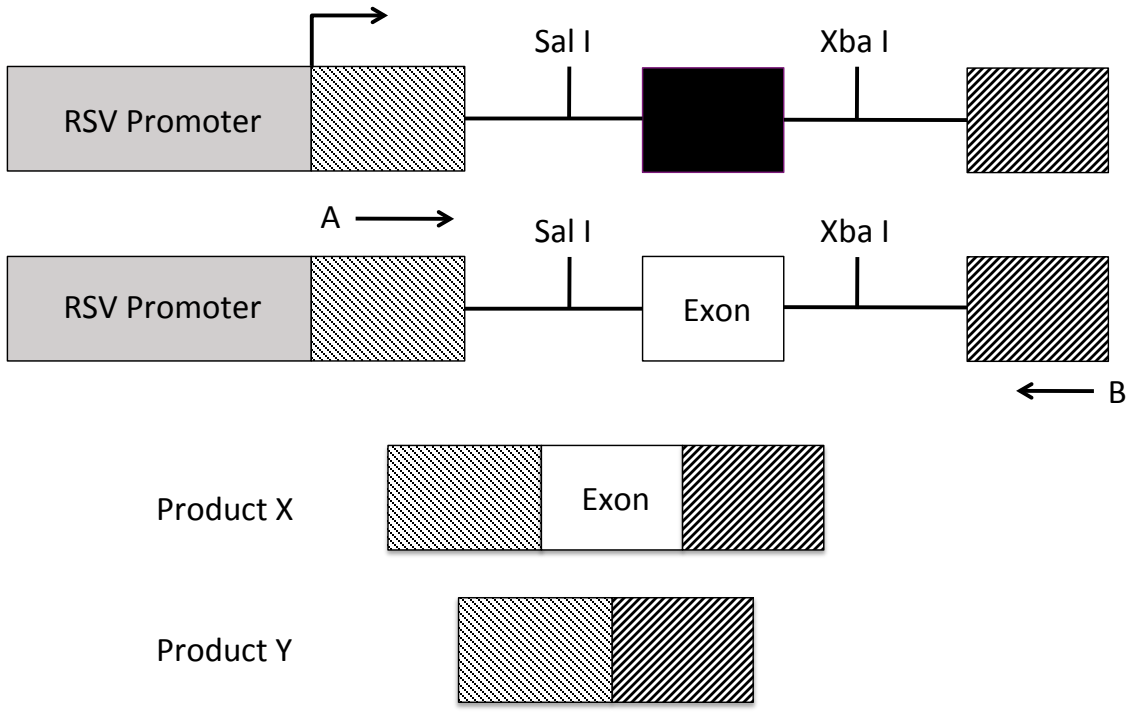


Figure 5.1. Insertion of *G6PC2* exons and products of the RHCglo minigene vector
The RHCglo minigene vector comprises the RSV promoter, a 5' exon, a 3' exon, and *Sal* I plus *Xba* I restriction enzyme sites into which *G6PC2* exon (Exon) 4 or 5 were ligated along with surrounding 5' and 3' intronic sequences (for exon 4) or just 5' intronic sequences (for exon 5). RT-PCR was used to generate cDNA from the isolated RNA products, and PCR was used to quantitate the relative abundance of splice products X and Y using primers A and B.

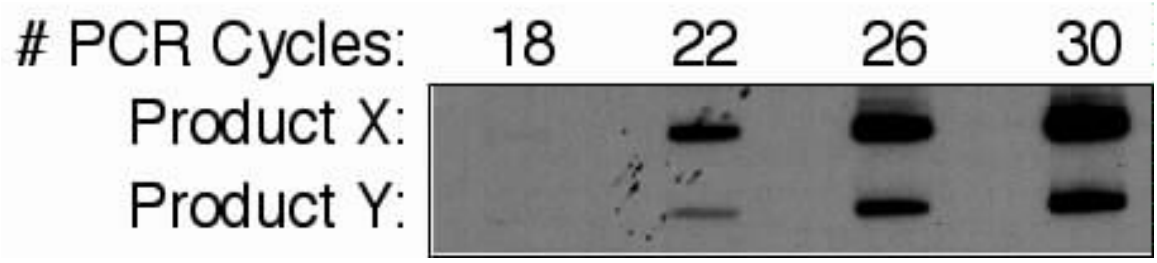


Figure 5.2. Linearity of PCR reaction for RHCglo minigene RNA products
A RHCglo minigene vector containing exon 4 and one allele of rs560887 in intron 3 was transiently transfected into HeLa cells. Following RNA isolation and cDNA synthesis as described in Chapter II, the linearity of the PCR reaction involving primers A and B (Fig. 5.1) was observed using increasing cycle numbers of 18, 22, 26, and 30.

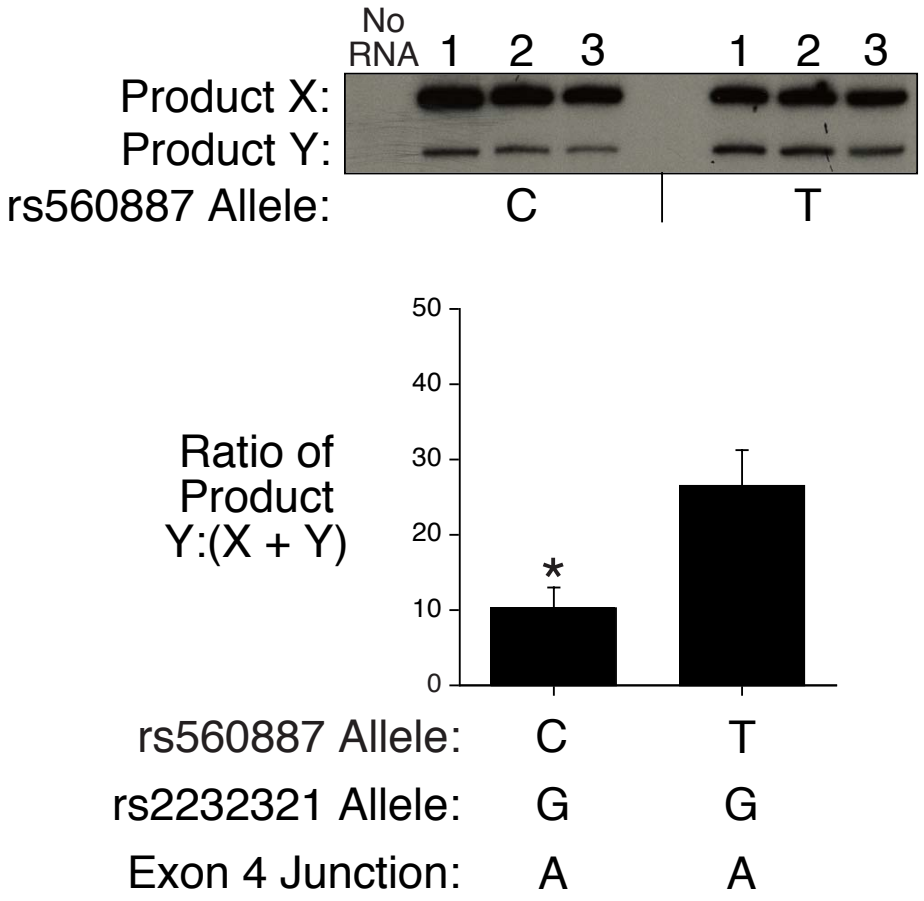


Figure 5.3. Alterations in *G6PC2* exon 4 splicing by rs560887
RHCglo minigene vectors containing the alternate alleles of rs560887 were transiently transfected into HeLa cells and RNA expression was quantitated as described in Chapter II. Results show the mean data from three experiments ± S.E.M. **P* < 0.05.

rs2232321 also alters *G6PC2* RNA splicing

rs2232321 is located immediately 3' of rs560887 and lies just outside the branch point element. However, by analogy with transcription factor binding elements where flanking sequence often affects binding affinity [141] we hypothesized that this SNP might also affect U2 binding and hence splicing. Using the same minigene strategy as described above, Figure 5.4 shows that the alternate alleles of rs2232321 alter the relative abundance of products X and Y, with the rs2232321-G allele promoting inclusion of *G6PC2* exon 4. The magnitude of the effect was smaller than that seen for rs560887, consistent with the fact that, unlike rs560887, rs2232321 does not alter the sequence of the branch point element (compare Fig. 5.3 and Fig. 5.4).

The effects of rs560887 and rs2232321 are dependent on the non-consensus exon 4 splice junction

We have previously shown that in both mice and humans the 5' exon 4 splice junction contains the non-consensus sequence tag/A [36, 74] instead of cag/G [142] (intronic sequence in lower case letters). We hypothesized that the presence of this non-consensus splice junction may result in splicing being more susceptible to the influence of rs560887 and rs2232321. This hypothesis was tested using the same minigene approach as discussed above. Figure 5.5 shows that when the sequence of the non-consensus 5' exon 4 splice junction is converted to tag/G the inclusion of *G6PC2* exon 4 is strongly enhanced, despite the presence of the sub-optimal rs560887-T allele. This indicates that the influence of the alternate alleles of rs560887 and rs2232321 on *G6PC2* RNA splicing is strongly dependent on the presence of the non-consensus exon 4 splice junction.

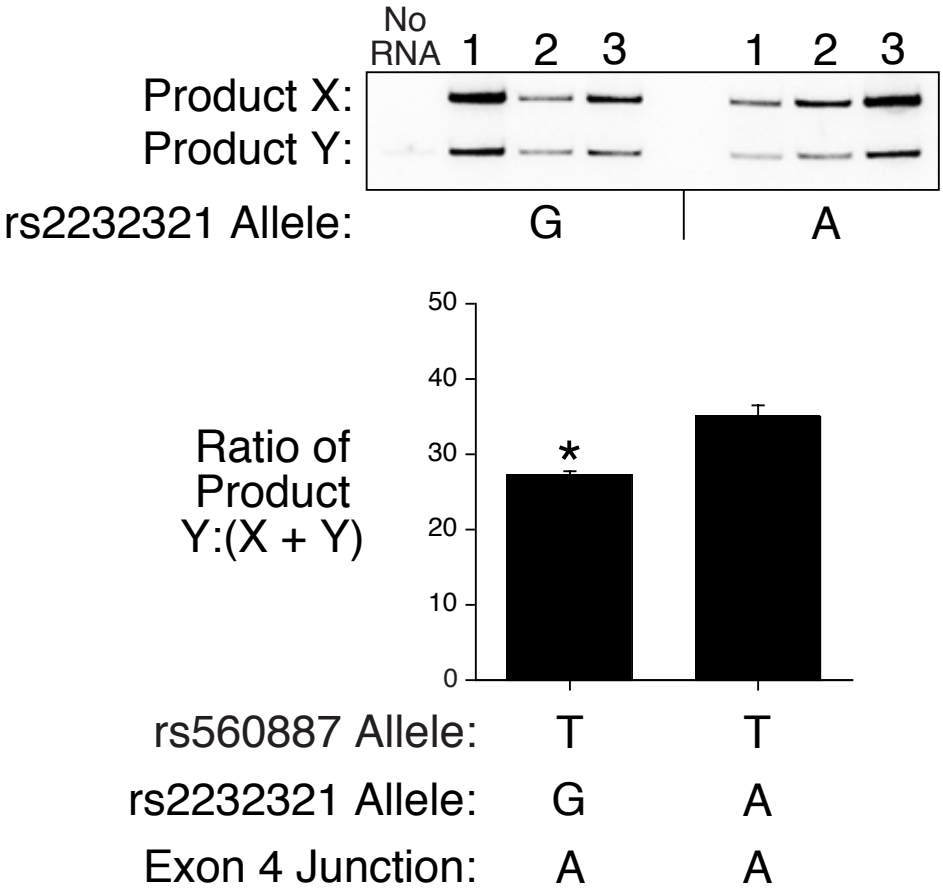


Figure 5.4. Alterations in *G6PC2* exon 4 splicing by rs2232321
RHCglo minigene vectors containing the alternate alleles of rs2232321 were transiently transfected into HeLa cells and RNA expression was quantitated as described in Chapter II. Results show the mean data from three experiments ± S.E.M. *$P < 0.05$.

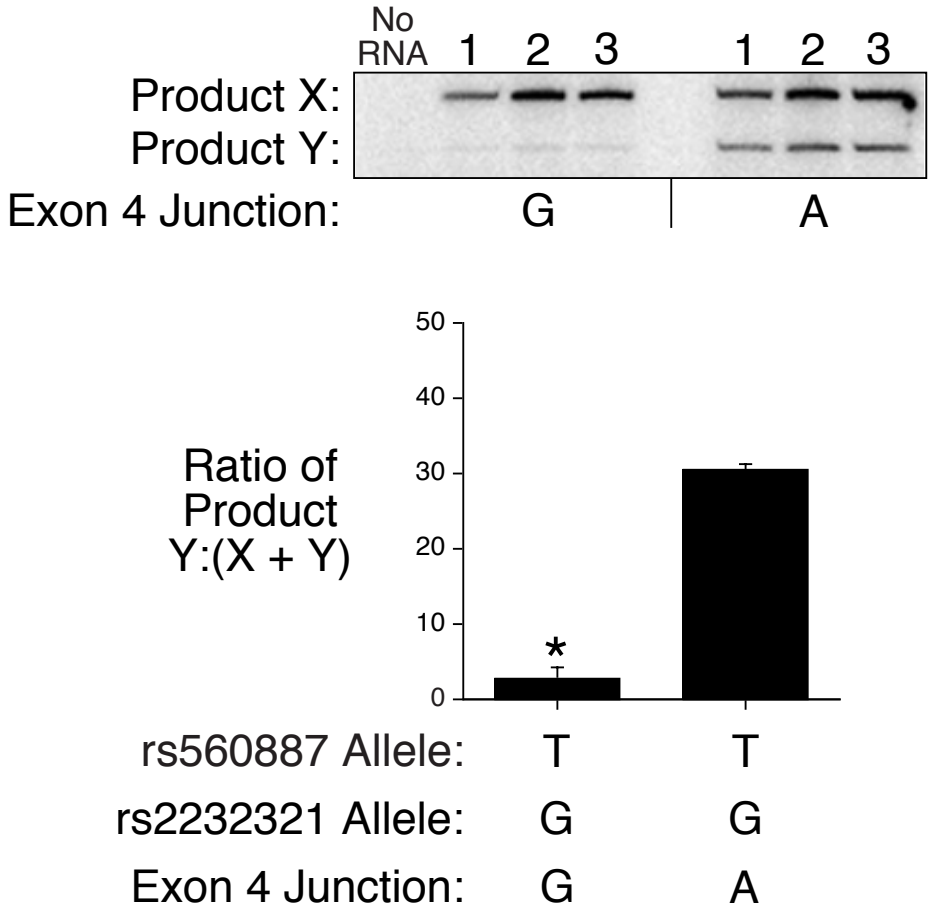


Figure 5.5. Dependency of the effects of rs560887 and rs2232321 on the non-consensus exon 4 splice junction

RHCglo minigene vectors containing the native (A) or consensus (G) exon 4 5' splice junction were transiently transfected into HeLa cells and RNA expression was quantitated as described in Chapter II. Results show the mean data from three experiments ± S.E.M. *$P < 0.05$.

The alternate alleles of rs35259259 result in a frame shift and affect *G6PC2* RNA splicing

rs35259259 is located at the 5' splice junction of exon 5 and represents a deletion SNP, specifically the presence or absence of the exon 5 splice junction 'G' acceptor nucleotide. This SNP changes the sequence of the splice junction from cag/GCAT to cag/CAT (intronic sequence in lower case letters), which no longer matches the consensus cag/G [142]. We hypothesized that the presence of this non-consensus splice junction would abolish splicing. Because exon 5 is the terminal exon in the *G6PC2* gene testing this hypothesis was more complex. However, we reasoned that the artificial minigene approach would still work if the minigene 3' exon spliced to a cryptic splice site within exon 5 (Fig. 5.1). Figure 5.6 shows that cryptic splicing indeed occurred resulting in the generation of product X (Fig. 5.1). DNA sequencing revealed that the cryptic 3' splice junction within exon 5 had been formed utilizing the *G6PC2* exon 5 sequence TC/gtga. This is only a partial match with the (A or T)G/gtaa consensus [142], which presumably results in inefficient splicing explaining why product Y > X (Fig. 5.6), in contrast to the exon 4 splicing experiments where product X >> Y (Fig. 5.2). Figure 5.6 shows that deletion of the exon 5 splice junction 'G' acceptor nucleotide further reduces the abundance of product X; however, splicing was not totally abolished. DNA sequencing revealed that the minigene 5' exon was now being spliced to *G6PC2* exon 5 using the non-consensus cag/CAT splice site described above. However, the product formed is now out of frame and would result in a truncated form of G6PC2 (207 versus 355 amino acids) lacking 5 out of 9 transmembrane domains [36], which is presumably non-functional.

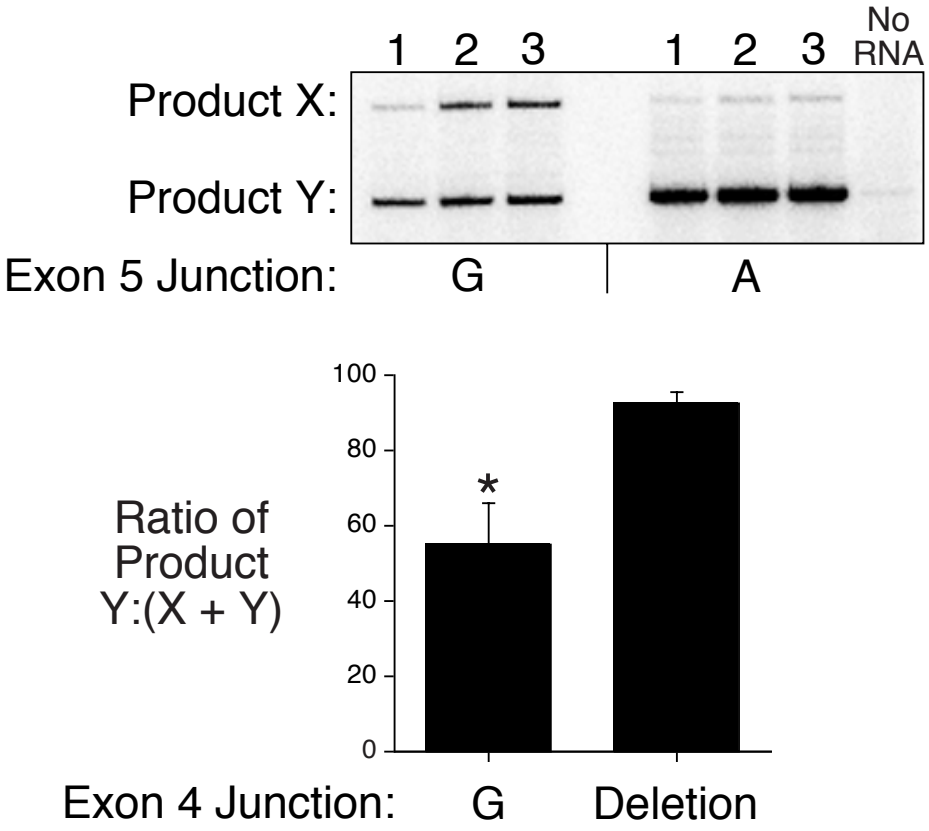


Figure 5.6. Frame shift effects of the variants of rs35259259 on *G6PC2* exon 5 splicing

RHCglo minigene vectors containing the alternate alleles of rs35259259 were transiently transfected into HeLa cells and RNA expression was quantitated as described in Chapter II. Results show the mean data from three experiments ± S.E.M. **P* < 0.05.

Discussion

Genome-wide association studies have recently provided important new insights concerning the genetics of quantitative traits related to type 2 diabetes, such as FPG and haemoglobin A1C [43, 90, 91, 107-110, 119, 121-124, 128]. These GWAS data showed that the *G6PC2* locus harbors the strongest common genetic determinant of FPG and HbA_{1C} levels in terms of effect size and significance. The studies described here represent a key step in extending these discovery phase GWAS data by establishing a functional link between the genetic variation in the *G6PC2* locus and variation in FPG. Specifically we used minigenes, in which splicing of individual exons can be assessed [137, 138], to show that a common variant, rs560887, that is located in the 3rd intron of *G6PC2* and which may explain ~1% of the total variance in FPG [90, 107, 109, 110, 113, 119, 121, 122, 128], modulates *G6PC2* RNA splicing (Fig. 5.1 and Fig. 5.2).

In a previous study we examined the contribution of two *G6PC2* promoter SNPs, rs13431652 and rs573225, to the association signal between *G6PC2* and FPG [96]. Genetic data support a potential role for both rs13431652 and rs573225 as causative SNPs linking *G6PC2* to variations in FPG. The rs13431652-A allele was associated with increased FPG and elevated promoter activity, consistent with the putative function of G6PC2 in pancreatic islets. In contrast, the rs573225-A allele was associated with increased FPG but reduced promoter activity, at odds with the putative function of G6PC2 in pancreatic islets [96]. One interpretation of these data is that human G6PC2 and mouse G6pc2 have opposite effects on FPG such that it is the behavior of the rs13431652 SNP in tissue culture cells that does not accurately reflect its effect on *G6PC2* gene transcription *in vivo*. The other interpretation of these data is that the

behavior of the rs573225 SNP in tissue culture cells does not accurately reflect its effect on *G6PC2* gene transcription *in vivo*. The studies described here strongly support the latter, simpler interpretation. Thus the rs560887-C allele promotes inclusion of *G6PC2* exon 4 and presumably higher G6PC2 protein expression. This would lead to increased G6P to glucose cycling and a reduction in glucose-stimulated insulin secretion. The splicing data are consistent with the genetic association between the rs560887-C allele and elevated FPG and the putative function of G6PC2. The functional data therefore suggest that rs560887 is a potentially causative SNP and the combination of these functional data with the genetic data support a direct connection between *G6PC2* and FPG.

In the initial characterization of mouse [32] and human [36] *G6PC2* mRNA, various splice forms were observed in islets. The most common human *G6PC2* variant lacked exon 4, an observation that could be explained by the fact that, in contrast to *G6PC* and *G6PC3*, the sequence of the 5' exon 4 splice junction of *G6PC2* does not match the consensus [35, 39]. Interestingly, the effect of rs560887 is dependent on the presence of this non-consensus junction (Fig. 5.5). Other *G6PC2* variants have also been detected that lacked exon 2, exons 3 plus 4, and exons 2, 3, and 4 together [36]. Human *G6PC2* splicing was subsequently studied in greater detail in thymus and spleen following the identification of *IGRP*/*G6PC2* as an autoantigen in type 1 diabetes [143]. Although the biological significance of the observation is unknown, it is fascinating that 80% of autoantigens show non-canonical splicing compared to 1% in the general population [144]. One intriguing possibility is that some parts of *G6PC2* may escape

central tolerance and thus contribute to autoimmunity and type 1 diabetes in the periphery [143].

In summary, our study provides functional evidence supporting an important role for the intronic rs560887 variant as a potentially causative SNP that contributes to the association signal between *G6PC2* and FPG and demonstrates the essential role of combining genetic and functional data in assessing the importance of individual human genetic variants. The genetic data do not preclude a contribution of additional variants. Indeed, we have previously shown that two *G6PC2* promoter SNPs, rs13431652 and rs573225, contribute to the association signal between *G6PC2* and FPG [96] and we demonstrate here that two additional SNPs, rs2232321 and rs35259259, affect *G6PC2* pre-mRNA splicing. Future genetic analyses will be required to establish the contribution of these SNPs to the association signal between *G6PC2* and FPG.

CHAPTER VI

SUMMARY AND FUTURE DIRECTIONS

Thesis Summary

The experiments described within this thesis served to elucidate the role of various *G6PC2* SNPs in the regulation of *G6PC2* transcription and pre-mRNA splicing. A number of these SNPs had been previously associated by GWA study with variations in FPG and HbA_{1C} [43, 90, 107-110, 119, 121, 122, 128]; however, the affect of these allelic variations had yet to be analyzed at a molecular level. The experiments were designed to determine what effect these SNPs have on *G6PC2* transcription and splicing.

Using a combination of transfection and reporter assays, it was demonstrated that the rs13431652 and rs573225 SNPs affect binding of NF-Y and Foxa2, respectively [96]. A change in *G6PC2* transcription was observed using fusion gene assays, with the introduction of the minor allele of rs13431652 resulting in a decrease in transcription and introduction of the minor allele of rs573225 resulting in an increase in fusion gene transcription. In the case of rs13431652, the fusion gene data show that the rs13431652-A allele increases transcription, which correlates with the GWA study association of this allele with elevated FPG [96], consistent with the function of G6PC2. However, the data obtained for rs573225 are at odds with this model. While the rs573225-A allele is highly associated with elevated FPG by GWA study and enhances Foxa2 binding in EMSAs [96], the lower fusion gene expression observed with this allele fails to support the function of G6PC2 and the reduced FPG seen in *G6pc2* knockout mice [55]. Thus, these

functional data support a causative role for rs13431652 but not rs573225.

In addition to rs13431652, the experiments of chapter IV provided both genetic and functional data supporting the role of rs2232316 as a potentially causative SNP. Conditioned regression model analyses were unable to assign independence of the signal from rs2232316 and rs573225 because the two variants are in high LD. This implies that the two associations actually represent a single signal and that both SNPs may be causative variants. Gel retardation studies elucidated not only specific binding of Foxa2 to the rs2232316 locus, but also a decrease in binding efficiency of the rs2232316-G allele for Foxa2 compared to the rs2232316-A allele. This variation was extended to fusion gene analysis in βTC-3 cells, where the A allele of rs2232316, in the context of the -324 to +3 *G6PC2* promoter region, was found to increase promoter activity by approximately 50% compared to the rs2232316-G allele. Thus the genetic and functional data for this particular SNP correlate, and, when taken as a whole, support the potential causative role of rs2232316 linking *G6PC2* to variations in FPG.

Because of the correlation between the fusion gene experiments and the genetic associations for rs2232316, two explanations for the contradictory results for rs573225 exist. In the instance that the fusion gene data for rs573225 truly matches the effect of the SNP *in vivo*, rs573225 could not be the causative variant in this promoter region as it simply does not correlate with the GWA study associations between the SNPs and FPG. Its association with variations in FPG would therefore have been due to its interaction with rs2232316, the action of which would serve to mask the variations in FPG contributed by rs573225. More likely, this SNP does not function correctly in our fusion gene experiments, as illustrated by the differing results found in the analyses in βTC-3

and HeLa cells.

Due to this potential limitation of islet-derived cell lines with respect to rs573225, it was decided that both rs573225 and rs2232316 would be re-analyzed in a reconstituted cell system comprised of HeLa cells co-transfected with the *G6PC2* fusion genes and expression vectors encoding various islet-enriched transcription factors. The results of these experiments showed the opposite effect for the rs573225-G allele, which was here associated with reduced, rather than increased, *G6PC2* fusion gene expression. A similarly opposite effect for rs2232316 was seen in this system compared to βTC-3 cells, with the rs2232316-A allele being associated with reduced *G6PC2* fusion gene expression in the reconstituted HeLa cell system. Overexpression of Foxa2 in the system served to blunt the effect of the alternate alleles of rs573225 but returned the results with the alternate rs2232316 alleles to those seen in βTC-3 cells. These results imply that, in specific circumstances, the rs573225-G allele fusion gene results correspond with the genetic data, but that the *in vivo* action of both rs573225 and rs2232316 on *G6PC2* expression hinges on the concentration of Foxa2.

While a number of these *G6PC2* SNPs have been associated with variations in FPG, the strongest association with FPG in multiple GWA studies was observed with rs560887. However, this SNP does not lie in the promoter or a proposed enhancer of *G6PC2*. Upon further examination, rs560887 was found to exist in a branch point sequence of intron 3. By utilizing a minigene plasmid developed by Dr. Thomas Cooper, the RHCglo plasmid vector, changes in the splicing of individual exons in response to allelic variation was assessed. A change from the rs560887-T allele to the -C allele resulted in increased inclusion of exon 4, possibly due to a stronger association of the U2

snRNP component of the spliceosome, which is consistent with the genetic association of this allele with elevated FPG and the proposed role of G6PC2. Another SNP, rs2232321, residing immediately 3' of rs560887, was also found to alter splicing, with the rs2232321-G allele increasing the product containing exon 4, albeit to a lower degree than that observed with the rs560887-C allele.

The presence of a non-consensus 5' exon 4 splice junction was hypothesized to influence the magnitude of splicing variation seen with both rs560887 and rs2232321. By using the same minigene approach, a change of the splice junction to the consensus tag/G resulted in a significant increase in exon 4 inclusion. This therefore implies that, while rs560887 and rs2232321 are able to influence *G6PC2* pre-mRNA splicing, their effects are amplified due to the presence of the non-consensus 5' exon 4 splice junction. Another SNP, rs35259259, lies at the 5' splice junction of exon 5, but rather than changing alleles this is a so-called "indel" SNP that affects the presence or absence of a base at this location. Such a change results in a frame shift mutation of exon 5, presumably creating another instance of a non-functional G6PC2 protein due to the loss of 5 out of 9 transmembrane domains. Through minigene analysis, it was found that a deletion at this SNP results in a frame shift and a truncated protein. However, due to the extreme rarity of this SNP, no genetic or metabolic associations have been made, though because of its nature, it is believed that a deletion at rs35259259 would also be associated with a significant reduction in FPG, as is the case with the rs560887-T allele.

As a whole, the experiments and data described within this document sought to elucidate the molecular roles of various SNPs within the *G6PC2* gene. Through these analyses, strong support was generated for a role of rs13431652 and rs2232316 as

causative variants, while the data derived for rs573225 likely indicates a limitation in the use of islet-derived cell lines and a causative label could not be applied. Additionally, the variant most strongly associated with changes in FPG, rs560887, and other intronic variants were revealed to affect the splicing of *G6PC2* pre-mRNA in a manner that further corroborates the role of G6PC2 as a regulator of FPG.

Future Directions

GWA studies performed more recently have elucidated a lack of association between FPG in a normoglycemic range (70-100mg/dL) and CAM [145], as well as a lack of influence by the SNPs in *G6PC2* on fasting glycemia in patients with elevated FPG [129]. These studies have created a link between FPG and CAM not in an entirely linear manner, but with a range of FPG on the low end of the scale that does not associate with an increased risk for CAM, before a linear increase in association is seen at higher FPG. The direct role of G6PC2 in CAM has yet to be studied, but such conflicting results necessitate further consideration of the results regarding this interaction.

The results described in this thesis have begun to elucidate the roles of multiple SNPs within the *G6PC2* gene, but while the specific SNPs analyzed have either been associated with variations in FPG (rs560887, rs573225, rs13431652, rs2232321), or occupy a location known to be critical for a cellular process (exon 4 splice junction, rs35259259), numerous other SNPs have been found in *G6PC2*. Though many of these are likely to prove non-causative, the potential exists that additional causative SNPs exist. In-depth genetic analysis and association studies have proven extremely useful in the search for relevant SNPs, and an increase in subject number for these studies, critical to

increase power, may continue to produce significant associations of different SNPs within *G6PC2* and FPG. Examination of the *G6PC2* promoter using a program such as MatInspector may uncover the presence of SNPs with the potential to affect transcription factor binding. The influence of these alleles could then be tested as described in chapters III and IV.

In addition to further analysis of promoter SNPs, other polymorphisms found within the open reading frame of *G6PC2* can be scrutinized for their effect on splicing efficiency or protein function. In the case that a SNP is purported to affect association or function of the spliceosome, an experimental approach similar to that of chapter V could be implemented to determine whether aberrant splicing occurs. As the glucose-6-phosphatase activity of G6PC2 has been successfully demonstrated *in vitro*, the assay could be used to examine whether non-synonymous SNPs affect enzyme activity. Thus, while the structure of G6PC2 and the glucose-6-phosphatase holoenzyme are unknown, SNPs causing a change in the amino acid sequence of G6PC2 may alter the ability of the enzyme to hydrolyze G6P. These changes could then be quantified by comparing the activities in the presence of either the major or minor allele of each SNP.

Further investigation of the *in vivo* nature of the SNPs discussed in this document, namely rs35259259, may also be pursued. As the rarity of this SNP has precluded previous metabolic analyses, a continuation of the search for individuals possessing this allelic deletion may provide vital information associating this locus with significant variations in FPG. Large databases, including BioVU at Vanderbilt University, now exist to provide access for researchers to the medical records and genomic sequences of de-identified patients at the Vanderbilt University Medical Center. By utilizing this resource,

a search for individuals possessing the deletion at rs35259259 will provide access to any medical information gleaned from their medical examination. Should this usage of BioVU be undertaken, information regarding this SNP and the correlating FPG values would prove invaluable.

REFERENCES

1. Mithieux, G., *New knowledge regarding glucose-6 phosphatase gene and protein and their roles in the regulation of glucose metabolism.* European journal of endocrinology / European Federation of Endocrine Societies, 1997. **136**(2): p. 137-45.
2. Foster, J.D., B.A. Pederson, and R.C. Nordlie, *Glucose-6-phosphatase structure, regulation, and function: an update.* Proceedings of the Society for Experimental Biology and Medicine. Society for Experimental Biology and Medicine, 1997. **215**(4): p. 314-32.
3. Lei, K.J., et al., *Mutations in the glucose-6-phosphatase gene that cause glycogen storage disease type 1a.* Science, 1993. **262**(5133): p. 580-3.
4. Shelly, L.L., et al., *Isolation of the gene for murine glucose-6-phosphatase, the enzyme deficient in glycogen storage disease type 1A.* The Journal of biological chemistry, 1993. **268**(29): p. 21482-5.
5. Gluecksohn-Waelsch, S., *Genetic control of morphogenetic and biochemical differentiation: lethal albino deletions in the mouse.* Cell, 1979. **16**(2): p. 225-37.
6. Ruppert, S., et al., *Two genetically defined trans-acting loci coordinately regulate overlapping sets of liver-specific genes.* Cell, 1990. **61**(5): p. 895-904.
7. van de Werve, G., et al., *New lessons in the regulation of glucose metabolism taught by the glucose 6-phosphatase system.* European journal of biochemistry / FEBS, 2000. **267**(6): p. 1533-49.
8. van Schaftingen, E. and I. Gerin, *The glucose-6-phosphatase system.* The Biochemical journal, 2002. **362**(Pt 3): p. 513-32.
9. Puskas, F., et al., *Conformational change of the catalytic subunit of glucose-6-phosphatase in rat liver during the fetal-to-neonatal transition.* The Journal of biological chemistry, 1999. **274**(1): p. 117-22.
10. Gerin, I., et al., *Sequence of a putative glucose 6-phosphate translocase, mutated in glycogen storage disease type Ib.* FEBS letters, 1997. **419**(2-3): p. 235-8.
11. Maloney, P.C., et al., *Anion-exchange mechanisms in bacteria.* Microbiological reviews, 1990. **54**(1): p. 1-17.
12. Chen, S.Y., et al., *The glucose-6-phosphate transporter is a phosphate-linked antiporter deficient in glycogen storage disease type Ib and Ic.* The FASEB journal : official publication of the Federation of American Societies for Experimental Biology, 2008. **22**(7): p. 2206-13.
13. Yang Chou, J. and B.C. Mansfield, *Molecular Genetics of Type 1 Glycogen Storage Diseases.* Trends in endocrinology and metabolism: TEM, 1999. **10**(3): p. 104-113.
14. Chou, J.Y. and B.C. Mansfield, *Molecular Genetics of Type 1 Glycogen Storage Diseases.* Trends in Endocrinology and Metabolism, 1999. **10**(3): p. 104-113.
15. Veiga-da-Cunha, M., et al., *The putative glucose 6-phosphate translocase gene is mutated in essentially all cases of glycogen storage disease type I non-a.* European journal of human genetics : EJHG, 1999. **7**(6): p. 717-23.

16. Mithieux, G., et al., *Glucose-6-phosphatase mRNA and activity are increased to the same extent in kidney and liver of diabetic rats.* Diabetes, 1996. **45**(7): p. 891-6.
17. Li, Y., M.C. Mechin, and G. van de Werve, *Diabetes affects similarly the catalytic subunit and putative glucose-6-phosphate translocase of glucose-6-phosphatase.* The Journal of biological chemistry, 1999. **274**(48): p. 33866-8.
18. Ashmore, J., A.B. Hastings, and F.B. Nesbett, *The Effect of Diabetes and Fasting on Liver Glucose-6-Phosphatase.* Proceedings of the National Academy of Sciences of the United States of America, 1954. **40**(8): p. 673-8.
19. Streeper, R.S., et al., *A multicomponent insulin response sequence mediates a strong repression of mouse glucose-6-phosphatase gene transcription by insulin.* The Journal of biological chemistry, 1997. **272**(18): p. 11698-701.
20. Barthel, A., et al., *Differential regulation of endogenous glucose-6-phosphatase and phosphoenolpyruvate carboxykinase gene expression by the forkhead transcription factor FKHR in H4IIE-hepatoma cells.* Biochemical and biophysical research communications, 2001. **285**(4): p. 897-902.
21. Schmoll, D., et al., *Regulation of glucose-6-phosphatase gene expression by protein kinase Balpha and the forkhead transcription factor FKHR. Evidence for insulin response unit-dependent and -independent effects of insulin on promoter activity.* The Journal of biological chemistry, 2000. **275**(46): p. 36324-33.
22. Streeper, R.S., et al., *Hepatocyte nuclear factor-1 acts as an accessory factor to enhance the inhibitory action of insulin on mouse glucose-6-phosphatase gene transcription.* Proceedings of the National Academy of Sciences of the United States of America, 1998. **95**(16): p. 9208-13.
23. Vander Kooi, B.T., et al., *The three insulin response sequences in the glucose-6-phosphatase catalytic subunit gene promoter are functionally distinct.* The Journal of biological chemistry, 2003. **278**(14): p. 11782-93.
24. Tang, E.D., et al., *Negative regulation of the forkhead transcription factor FKHR by Akt.* The Journal of biological chemistry, 1999. **274**(24): p. 16741-6.
25. Nakae, J., B.C. Park, and D. Accili, *Insulin stimulates phosphorylation of the forkhead transcription factor FKHR on serine 253 through a Wortmannin-sensitive pathway.* The Journal of biological chemistry, 1999. **274**(23): p. 15982-5.
26. Rena, G., et al., *Phosphorylation of the transcription factor forkhead family member FKHR by protein kinase B.* The Journal of biological chemistry, 1999. **274**(24): p. 17179-83.
27. Guo, S., et al., *Phosphorylation of serine 256 by protein kinase B disrupts transactivation by FKHR and mediates effects of insulin on insulin-like growth factor-binding protein-1 promoter activity through a conserved insulin response sequence.* The Journal of biological chemistry, 1999. **274**(24): p. 17184-92.
28. Brunet, A., et al., *Akt promotes cell survival by phosphorylating and inhibiting a Forkhead transcription factor.* Cell, 1999. **96**(6): p. 857-68.
29. Biggs, W.H., 3rd, et al., *Protein kinase B/Akt-mediated phosphorylation promotes nuclear exclusion of the winged helix transcription factor FKHR1.* Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(13): p. 7421-6.

30. Kops, G.J., et al., *Direct control of the Forkhead transcription factor AFX by protein kinase B.* Nature, 1999. **398**(6728): p. 630-4.
31. Takaishi, H., et al., *Regulation of nuclear translocation of forkhead transcription factor AFX by protein kinase B.* Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(21): p. 11836-41.
32. Arden, S.D., et al., *Molecular cloning of a pancreatic islet-specific glucose-6-phosphatase catalytic subunit-related protein.* Diabetes, 1999. **48**(3): p. 531-42.
33. Waddell, I.D. and A. Burchell, *The microsomal glucose-6-phosphatase enzyme of pancreatic islets.* The Biochemical journal, 1988. **255**(2): p. 471-6.
34. Petrolonis, A.J., et al., *Enzymatic characterization of the pancreatic islet-specific glucose-6-phosphatase-related protein (IGRP).* The Journal of biological chemistry, 2004. **279**(14): p. 13976-83.
35. Martin, C.C., et al., *Identification and characterization of a human cDNA and gene encoding a ubiquitously expressed glucose-6-phosphatase catalytic subunit-related protein.* Journal of molecular endocrinology, 2002. **29**(2): p. 205-22.
36. Martin, C.C., et al., *Cloning and characterization of the human and rat islet-specific glucose-6-phosphatase catalytic subunit-related protein (IGRP) genes.* The Journal of biological chemistry, 2001. **276**(27): p. 25197-207.
37. Shieh, J.J., et al., *The islet-specific glucose-6-phosphatase-related protein, implicated in diabetes, is a glycoprotein embedded in the endoplasmic reticulum membrane.* FEBS letters, 2004. **562**(1-3): p. 160-4.
38. Hutton, J.C. and R.M. O'Brien, *Glucose-6-phosphatase catalytic subunit gene family.* The Journal of biological chemistry, 2009. **284**(43): p. 29241-5.
39. Boustead, J.N., et al., *Identification and characterization of a cDNA and the gene encoding the mouse ubiquitously expressed glucose-6-phosphatase catalytic subunit-related protein.* Journal of molecular endocrinology, 2004. **32**(1): p. 33-53.
40. Guionie, O., et al., *Identification and characterisation of a new human glucose-6-phosphatase isoform.* FEBS letters, 2003. **551**(1-3): p. 159-64.
41. Shieh, J.J., et al., *A glucose-6-phosphate hydrolase, widely expressed outside the liver, can explain age-dependent resolution of hypoglycemia in glycogen storage disease type Ia.* The Journal of biological chemistry, 2003. **278**(47): p. 47098-103.
42. Khan, A., C. Hong-Lie, and B.R. Landau, *Glucose-6-phosphatase activity in islets from ob/ob and lean mice and the effect of dexamethasone.* Endocrinology, 1995. **136**(5): p. 1934-8.
43. Bouatia-Naji, N., et al., *A polymorphism within the G6PC2 gene is associated with fasting plasma glucose levels.* Science, 2008. **320**(5879): p. 1085-8.
44. Efrat, S., et al., *Ribozyme-mediated attenuation of pancreatic beta-cell glucokinase expression in transgenic mice results in impaired glucose-induced insulin secretion.* Proceedings of the National Academy of Sciences of the United States of America, 1994. **91**(6): p. 2051-5.
45. Efrat, S., M. Tal, and H.F. Lodish, *The pancreatic beta-cell glucose sensor.* Trends in biochemical sciences, 1994. **19**(12): p. 535-8.
46. Epstein, P.N., et al., *Expression of yeast hexokinase in pancreatic beta cells of transgenic mice reduces blood glucose, enhances insulin secretion, and decreases*

diabetes. Proceedings of the National Academy of Sciences of the United States of America, 1992. **89**(24): p. 12038-42.
47. Taljedal, I.B., *Presence, induction and possible role of glucose 6-phosphatase in mammalian pancreatic islets.* The Biochemical journal, 1969. **114**(2): p. 387-94.
48. Meglasson, M.D. and F.M. Matschinsky, *New perspectives on pancreatic islet glucokinase.* The American journal of physiology, 1984. **246**(1 Pt 1): p. E1-13.
49. Bell, G.I., et al., *Molecular biology of mammalian glucose transporters.* Diabetes care, 1990. **13**(3): p. 198-208.
50. Thorens, B., M.J. Charron, and H.F. Lodish, *Molecular physiology of glucose transporters.* Diabetes care, 1990. **13**(3): p. 209-18.
51. Tal, M., et al., *[Val12] HRAS downregulates GLUT2 in beta cells of transgenic mice without affecting glucose homeostasis.* Proceedings of the National Academy of Sciences of the United States of America, 1992. **89**(13): p. 5744-8.
52. Khan, A., et al., *Evidence for the presence of glucose cycling in pancreatic islets of the ob/ob mouse.* The Journal of biological chemistry, 1989. **264**(17): p. 9732-3.
53. Khan, A., et al., *Glucose cycling in islets from healthy and diabetic rats.* Diabetes, 1990. **39**(4): p. 456-9.
54. Sweet, I.R., et al., *Measurement and modeling of glucose-6-phosphatase in pancreatic islets.* The American journal of physiology, 1997. **272**(4 Pt 1): p. E696-711.
55. Wang, Y., et al., *Deletion of the gene encoding the islet-specific glucose-6-phosphatase catalytic subunit-related protein autoantigen results in a mild metabolic phenotype.* Diabetologia, 2007. **50**(4): p. 774-8.
56. Benedetti, A., R. Fulceri, and M. Comporti, *Calcium sequestration activity in rat liver microsomes. Evidence for a cooperation of calcium transport with glucose-6-phosphatase.* Biochimica et biophysica acta, 1985. **816**(2): p. 267-77.
57. Gericke, M., G. Droogmans, and B. Nilius, *Thapsigargin discharges intracellular calcium stores and induces transmembrane currents in human endothelial cells.* Pflugers Archiv : European journal of physiology, 1993. **422**(6): p. 552-7.
58. Worley, J.F., 3rd, et al., *Endoplasmic reticulum calcium store regulates membrane potential in mouse islet beta-cells.* The Journal of biological chemistry, 1994. **269**(20): p. 14359-62.
59. Ingelsson, E., et al., *Detailed physiologic characterization reveals diverse mechanisms for novel genetic Loci regulating glucose and insulin metabolism in humans.* Diabetes, 2010. **59**(5): p. 1266-75.
60. Li, X., et al., *Additive effects of genetic variation in GCK and G6PC2 on insulin secretion and fasting glucose.* Diabetes, 2009. **58**(12): p. 2946-53.
61. Rose, C.S., et al., *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, 2009. **52**(10): p. 2122-9.
62. Bertram, R., A. Sherman, and L.S. Satin, *Electrical bursting, calcium oscillations, and synchronization of pancreatic islets.* Advances in experimental medicine and biology, 2010. **654**: p. 261-79.

63. Shameli, A., et al., *Endoplasmic reticulum stress caused by overexpression of islet-specific glucose-6-phosphatase catalytic subunit-related protein in pancreatic Beta-cells.* The review of diabetic studies : RDS, 2007. **4**(1): p. 25-32.
64. Anderson, M.S. and J.A. Bluestone, *The NOD mouse: a model of immune dysregulation.* Annual review of immunology, 2005. **23**: p. 447-85.
65. Lieberman, S.M., et al., *Identification of the beta cell antigen targeted by a prevalent population of pathogenic CD8+ T cells in autoimmune diabetes.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(14): p. 8384-8.
66. Yang, J., et al., *Islet-specific glucose-6-phosphatase catalytic subunit-related protein-reactive CD4+ T cells in human subjects.* Journal of immunology, 2006. **176**(5): p. 2781-9.
67. Jarchum, I., T. Takaki, and T.P. DiLorenzo, *Efficient culture of CD8(+) T cells from the islets of NOD mice and their use for the study of autoreactive specificities.* Journal of immunological methods, 2008. **339**(1): p. 66-73.
68. Serreze, D.V., M.P. Marron, and T.P. Dilorenzo, *"Humanized" HLA transgenic NOD mice to identify pancreatic beta cell autoantigens of potential clinical relevance to type 1 diabetes.* Annals of the New York Academy of Sciences, 2007. **1103**: p. 103-11.
69. Takaki, T., et al., *HLA-A*0201-restricted T cells from humanized NOD mice recognize autoantigens of potential clinical relevance to type 1 diabetes.* Journal of immunology, 2006. **176**(5): p. 3257-65.
70. Krishnamurthy, B., et al., *Responses against islet antigens in NOD mice are prevented by tolerance to proinsulin but not IGRP.* The Journal of clinical investigation, 2006. **116**(12): p. 3258-65.
71. Nakayama, M., et al., *Prime role for an insulin epitope in the development of type 1 diabetes in NOD mice.* Nature, 2005. **435**(7039): p. 220-3.
72. Oeser, J.K., et al., *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, 2011. **60**(11): p. 2922-7.
73. Frigeri, C., et al., *The proximal islet-specific glucose-6-phosphatase catalytic subunit-related protein autoantigen promoter is sufficient to initiate but not maintain transgene expression in mouse islets in vivo.* Diabetes, 2004. **53**(7): p. 1754-64.
74. Ebert, D.H., et al., *Structure and promoter activity of an islet-specific glucose-6-phosphatase catalytic subunit-related gene.* Diabetes, 1999. **48**(3): p. 543-51.
75. Wang, Y., et al., *Long-range enhancers are required to maintain expression of the autoantigen islet-specific glucose-6-phosphatase catalytic subunit-related protein in adult mouse islets in vivo.* Diabetes, 2008. **57**(1): p. 133-41.
76. Bischof, L.J., et al., *Characterization of the mouse islet-specific glucose-6-phosphatase catalytic subunit-related protein gene promoter by in situ footprinting: correlation with fusion gene expression in the islet-derived betaTC-3 and hamster insulinoma tumor cell lines.* Diabetes, 2001. **50**(3): p. 502-14.
77. Martin, C.C., et al., *Upstream stimulatory factor (USF) and neurogenic differentiation/beta-cell E box transactivator 2 (NeuroD/BETA2) contribute to*

islet-specific glucose-6-phosphatase catalytic-subunit-related protein (IGRP) gene expression. The Biochemical journal, 2003. **371**(Pt 3): p. 675-86.
78. Martin, C.C., J.K. Oeser, and R.M. O'Brien, *Differential regulation of islet-specific glucose-6-phosphatase catalytic subunit-related protein gene transcription by Pax-6 and Pdx-1.* The Journal of biological chemistry, 2004. **279**(33): p. 34277-89.
79. Martin, C.C., et al., *Foxa2 and MafA regulate islet-specific glucose-6-phosphatase catalytic subunit-related protein gene expression.* Journal of molecular endocrinology, 2008. **41**(5): p. 315-28.
80. Sander, M. and M.S. German, *The beta cell transcription factors and development of the pancreas.* Journal of molecular medicine, 1997. **75**(5): p. 327-40.
81. Melloul, D., S. Marshak, and E. Cerasi, *Regulation of insulin gene transcription.* Diabetologia, 2002. **45**(3): p. 309-26.
82. Hay, C.W. and K. Docherty, *Comparative analysis of insulin gene promoters: implications for diabetes research.* Diabetes, 2006. **55**(12): p. 3201-13.
83. Rearick, D., et al., *Critical association of ncRNA with introns.* Nucleic acids research, 2011. **39**(6): p. 2357-66.
84. Konarska, M.M., et al., *Characterization of the branch site in lariat RNAs produced by splicing of mRNA precursors.* Nature, 1985. **313**(6003): p. 552-7.
85. Sharp, P.A., *Split genes and RNA splicing.* Cell, 1994. **77**(6): p. 805-15.
86. Cartegni, L., S.L. Chew, and A.R. Krainer, *Listening to silence and understanding nonsense: exonic mutations that affect splicing.* Nature reviews. Genetics, 2002. **3**(4): p. 285-98.
87. Solis, A.S., N. Shariat, and J.G. Patton, *Splicing fidelity, enhancers, and disease.* Frontiers in bioscience : a journal and virtual library, 2008. **13**: p. 1926-42.
88. Vaxillaire, M. and P. Froguel, *Genetic basis of maturity-onset diabetes of the young.* Endocrinology and metabolism clinics of North America, 2006. **35**(2): p. 371-84, x.
89. Billings, L.K. and J.C. Florez, *The genetics of type 2 diabetes: what have we learned from GWAS?* Annals of the New York Academy of Sciences, 2010. **1212**: p. 59-77.
90. Hu, C., et al., *A genetic variant of G6PC2 is associated with type 2 diabetes and fasting plasma glucose level in the Chinese population.* Diabetologia, 2009. **52**(3): p. 451-6.
91. Chen, W.M., et al., *Variations in the G6PC2/ABCB11 genomic region are associated with fasting glucose levels.* The Journal of clinical investigation, 2008. **118**(7): p. 2620-8.
92. Dos Santos, C., P. Bougneres, and D. Fradin, *A single-nucleotide polymorphism in a methylatable Foxa2 binding site of the G6PC2 promoter is associated with insulin secretion in vivo and increased promoter activity in vitro.* Diabetes, 2009. **58**(2): p. 489-92.
93. Khaw, K.T., et al., *Glycated haemoglobin, diabetes, and mortality in men in Norfolk cohort of european prospective investigation of cancer and nutrition (EPIC-Norfolk).* BMJ, 2001. **322**(7277): p. 15-8.
94. Lawes, C.M., et al., *Blood glucose and risk of cardiovascular disease in the Asia Pacific region.* Diabetes care, 2004. **27**(12): p. 2836-42.

95. Abdul-Ghani, M.A. and R.A. DeFronzo, *Plasma glucose concentration and prediction of future risk of type 2 diabetes.* Diabetes care, 2009. **32 Suppl 2**: p. S194-8.
96. Bouatia-Naji, N., et al., *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, 2010. **59**(10): p. 2662-71.
97. Sambrook, J., E.F. Fritsch, and T. Maniatis, *Molecular Cloning: A Laboratory Manual*1989, Plainview, New York: Cold Spring Harbor Laboratory Press.
98. Stieger, B., Y. Meier, and P.J. Meier, *The bile salt export pump.* Pflugers Archiv : European journal of physiology, 2007. **453**(5): p. 611-20.
99. Quandt, K., et al., *MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data.* Nucleic acids research, 1995. **23**(23): p. 4878-84.
100. Mantovani, R., *A survey of 178 NF-Y binding CCAAT boxes.* Nucleic acids research, 1998. **26**(5): p. 1135-43.
101. Mantovani, R., *The molecular biology of the CCAAT-binding factor NF-Y.* Gene, 1999. **239**(1): p. 15-27.
102. Onuma, H., et al., *Correlation between FOXO1a (FKHR) and FOXO3a (FKHRL1) binding and the inhibition of basal glucose-6-phosphatase catalytic subunit gene transcription by insulin.* Molecular endocrinology, 2006. **20**(11): p. 2831-47.
103. Limbird, L.E., *Cell Surface Receptors: A Short Course on Theory and Methods*. 3rd ed2004: Springer.
104. Ceribelli, M., et al., *The histone-like NF-Y is a bifunctional transcription factor.* Molecular and cellular biology, 2008. **28**(6): p. 2047-58.
105. Friedman, J.R. and K.H. Kaestner, *The Foxa family of transcription factors in development and metabolism.* Cell. Mol. Life Sci., 2006. **63**: p. 2317-2328.
106. O'Brien, R.M., et al., *Hepatic nuclear factor 3- and hormone-regulated expression of the phosphoenolpyruvate carboxykinase and insulin-like growth factor-binding protein 1 genes.* Molecular and cellular biology, 1995. **15**(3): p. 1747-58.
107. Bouatia-Naji, N., et al., *A variant near MTNR1B is associated with increased fasting plasma glucose levels and type 2 diabetes risk.* Nature genetics, 2009. **41**(1): p. 89-94.
108. Prokopenko, I., et al., *Variants in MTNR1B influence fasting glucose levels.* Nature genetics, 2009. **41**(1): p. 77-81.
109. Pare, G., et al., *Novel association of HK1 with glycated hemoglobin in a non-diabetic population: a genome-wide evaluation of 14,618 participants in the Women's Genome Health Study.* PLoS genetics, 2008. **4**(12): p. e1000312.
110. Reiling, E., et al., *Combined effects of single-nucleotide polymorphisms in GCK, GCKR, G6PC2 and MTNR1B on fasting plasma glucose and type 2 diabetes risk.* Diabetologia, 2009. **52**(9): p. 1866-70.
111. Takeuchi, F., et al., *Common variants at the GCK, GCKR, G6PC2-ABCB11 and MTNR1B loci are associated with fasting glucose in two Asian populations.* Diabetologia, 2010. **53**(2): p. 299-308.

112. Miranda, T.B. and P.A. Jones, *DNA methylation: the nuts and bolts of repression.* Journal of cellular physiology, 2007. **213**(2): p. 384-90.
113. Pontecorvo, G., B. De Felice, and M. Carfagna, *Novel methylation at GpC dinucleotide in the fish Sparus aurata genome.* Molecular biology reports, 2000. **27**(4): p. 225-30.
114. Overdier, D.G., A. Porcella, and R.H. Costa, *The DNA-binding specificity of the hepatocyte nuclear factor 3/forkhead domain is influenced by amino-acid residues adjacent to the recognition helix.* Molecular and cellular biology, 1994. **14**(4): p. 2755-66.
115. Smith, C.L. and G.L. Hager, *Transcriptional regulation of mammalian genes in vivo. A tale of two templates.* The Journal of biological chemistry, 1997. **272**(44): p. 27493-6.
116. Zaret, K.S., et al., *Pioneer factors, genetic competence, and inductive signaling: programming liver and pancreas progenitors from the endoderm.* Cold Spring Harbor symposia on quantitative biology, 2008. **73**: p. 119-26.
117. Hannenhalli, S. and K.H. Kaestner, *The evolution of Fox genes and their role in development and disease.* Nature reviews. Genetics, 2009. **10**(4): p. 233-40.
118. Tuteja, G., et al., *Extracting transcription factor targets from ChIP-Seq data.* Nucleic acids research, 2009. **37**(17): p. e113.
119. Dupuis, J., et al., *New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk.* Nature genetics, 2010. **42**(2): p. 105-16.
120. Jones, H.W., Jr., et al., *George Otto Gey. (1899-1970). The HeLa cell and a reappraisal of its origin.* Obstetrics and gynecology, 1971. **38**(6): p. 945-9.
121. Tam, C.H., et al., *Common polymorphisms in MTNR1B, G6PC2 and GCK are associated with increased fasting plasma glucose and impaired beta-cell function in Chinese subjects.* PloS one, 2010. **5**(7): p. e11428.
122. Soranzo, N., et al., *Common variants at 10 genomic loci influence hemoglobin A(C) levels via glycemic and nonglycemic pathways.* Diabetes, 2010. **59**(12): p. 3229-39.
123. Sparso, T., et al., *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, 2008. **51**(1): p. 70-5.
124. Orho-Melander, M., 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**(11): p. 3112-21.
125. Xu, L., et al., *Expression profiling suggested a regulatory role of liver-enriched transcription factors in human hepatocellular carcinoma.* Cancer research, 2001. **61**(7): p. 3176-81.
126. Edelman, D., et al., *Utility of hemoglobin A1c in predicting diabetes risk.* Journal of general internal medicine, 2004. **19**(12): p. 1175-80.
127. Abdul-Ghani, M.A., et al., *Minimal contribution of fasting hyperglycemia to the incidence of type 2 diabetes in subjects with normal 2-h plasma glucose.* Diabetes care, 2010. **33**(3): p. 557-61.
128. Hu, C., et al., *Effects of GCK, GCKR, G6PC2 and MTNR1B variants on glucose metabolism and insulin secretion.* PloS one, 2010. **5**(7): p. e11761.

129. Heni, M., et al., *The impact of genetic variation in the G6PC2 gene on insulin secretion depends on glycemia.* The Journal of clinical endocrinology and metabolism, 2010. **95**(12): p. E479-84.
130. Renstrom, F., et al., *Genetic predisposition to long-term nondiabetic deteriorations in glucose homeostasis: Ten-year follow-up of the GLACIER study.* Diabetes, 2011. **60**(1): p. 345-54.
131. Matschinsky, F.M., *Glucokinase as glucose sensor and metabolic signal generator in pancreatic beta-cells and hepatocytes.* Diabetes, 1990. **39**(6): p. 647-52.
132. Matschinsky, F.M., *Banting Lecture 1995. A lesson in metabolic regulation inspired by the glucokinase glucose sensor paradigm.* Diabetes, 1996. **45**(2): p. 223-41.
133. Iynedjian, P.B., *Molecular physiology of mammalian glucokinase.* Cellular and molecular life sciences : CMLS, 2009. **66**(1): p. 27-42.
134. van de Bunt, M. and A.L. Gloyn, *From genetic association to molecular mechanism.* Current diabetes reports, 2010. **10**(6): p. 452-66.
135. Kramer, A., *The structure and function of proteins involved in mammalian pre-mRNA splicing.* Annual review of biochemistry, 1996. **65**: p. 367-409.
136. Gao, K., et al., *Human branch point consensus sequence is yUnAy.* Nucleic acids research, 2008. **36**(7): p. 2257-67.
137. Stoss, O., et al., *The in vivo minigene approach to analyze tissue-specific splicing.* Brain research. Brain research protocols, 1999. **4**(3): p. 383-94.
138. Cooper, T.A., *Use of minigene systems to dissect alternative splicing elements.* Methods, 2005. **37**(4): p. 331-40.
139. Singh, G. and T.A. Cooper, *Minigene reporter for identification and analysis of cis elements and trans factors affecting pre-mRNA splicing.* BioTechniques, 2006. **41**(2): p. 177-81.
140. Cooper, T.A., *Muscle-specific splicing of a heterologous exon mediated by a single muscle-specific splicing enhancer from the cardiac troponin T gene.* Molecular and cellular biology, 1998. **18**(8): p. 4519-25.
141. Fry, C.J. and P.J. Farnham, *Context-dependent transcriptional regulation.* The Journal of biological chemistry, 1999. **274**(42): p. 29583-6.
142. Jackson, I.J., *A reappraisal of non-consensus mRNA splice sites.* Nucleic acids research, 1991. **19**(14): p. 3795-8.
143. Dogra, R.S., et al., *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, 2006. **49**(5): p. 953-7.
144. Ng, B., et al., *Increased noncanonical splicing of autoantigen transcripts provides the structural basis for expression of untolerized epitopes.* The Journal of allergy and clinical immunology, 2004. **114**(6): p. 1463-70.
145. Sarwar, N., et al., *Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies.* Lancet, 2010. **375**(9733): p. 2215-22.