Genetic Studies of Syndromes of Severe Insulin Resistance and Type 2 Diabetes: a candidate gene approach

Katherine A Fawcett

This dissertation is submitted for the
degree Doctor of Philosophy
October 2008

Trinity College
University of Cambridge

The Wellcome Trust Sanger Institute
Hinxton
Cambridge, UK

Declaration

I hereby declare that my dissertation contains material that has not been submitted for a degree or diploma or any other qualification at any other university. This thesis describes my own work and does not include the work that has been done in collaboration, except when specifically indicated in the text.

Katherine A Fawcett

Acknowledgements

This thesis would not have been possible were it not for my supervisor, Inês Barroso. Her intellectual support, generosity with her time, and her ability to remain enthusiastic, optimistic and affable despite a 20-hour working day, was indispensable. I am also obliged to all members of the Metabolic Disease Group at the Wellcome Trust Sanger Institute. Amanda Thompson for her help in the laboratory, for her careful management of our resources and information, and for putting up with my untidy bench/desk/freezer space! Our computer whiz, Allan Daly, for all the informatics support he has leant me over the past four years. Rosa-Maria Roccasecca for her supervision and encouraging pats. Bill Bottomley for his help in the lab, in particular during my rotation project, and Felicity Payne for her help with Taqman genotyping and her cheerful and sunny disposition. Last but not least, Eleanor Wheeler for her assistance with statistics on many of my projects.

I would also like to recognise the essential contribution of our various collaborators to the success of my PhD. Prof. Stephen O'Rahilly at Cambridge University was not only my second supervisor but established the severe insulin resistance cohort and collaborated with me on projects to elucidate their genetic aetiology. I worked with members of the MRC Epidemiology Unit on population-based and case-control studies of the genetics of type 2 diabetes and related traits, for which I would like to thank Prof. Nick Wareham, Dr Manjinder Sandhu and Dr Jian'an Luan for their statistics expertise, Dr Sally Debenham, and Dr Ruth Loos. Special thanks also go to Neil Grimsey and Dr Symeon Siniossoglou at the Cambridge Institute for Medical Research for their functional work on lipin 1, Satya Dash and Dr David Savage at Cambridge University for their work on AS160, and collaborators at the University of Southampton for their assistance with population-based studies of lipin 1 and PARL, including Holly Syddall and Prof. Cyrus Cooper. In addition, other researchers in the field of diabetes genetics in the UK have provided invaluable resources and expertise during the course of my PhD. I am also grateful for the support of the sequencing and genotyping teams at the Wellcome Trust Sanger Institute - in particular, Susannah Bumpstead and Andrew Keniry for lipin 1 genotyping.

Finally, I would like to express my deepest gratitude to my family and friends. My fellow Cambridge and Sanger PhD students have made the last four years an absolute pleasure, with their wit, good humour, exciting array of home and international backgrounds and propensity to socialise. I can't imagine the last few years without my beautiful, gentle, intelligent partner, Andy, with whom I have shared all the highs and lows of graduate life. I would like to thank him for his love and support and I look forward to our life together as non-students in the "real" world! Most importantly, I would like to thank my parents for giving me the best possible foundation from which to build a happy, healthy life. Words cannot express my appreciation of their love, selflessness, encouragement of interests in science, literature and politics, and for providing three wonderful siblings – John, Ian and Frances – without whom I wouldn't be who I am.

Dedication

Though I should dedicate this thesis to my younger self whose love of learning and naïve ambition has seen me through to a graduate degree, I would also like to mention my grandparents - two of whom, James Johnson and Joan Fawcett, are no longer with us. Despite their keen intelligence they were denied the educational opportunities that many of my generation now take for granted. May we never cease to value and improve the comprehensive education system.

Abstract

Insulin resistance plays a significant role in the pathogenesis of type 2 diabetes (T2D), a severe metabolic disease and major public health concern. Discovery of genes underlying insulin resistant syndromes can provide insights into pathophysiology and identify novel pathways for drug discovery. Moreover, these genes may also impact common T2D risk. The aim of this work was to investigate genetic variants for effects on insulin resistance, T2D and related continuous traits. Candidate genes studied included lipin family genes (Ch 3), components of the mTOR pathway (Ch 4), *PARL* (Ch 5) and genes involved in pancreatic β-cell function (Ch6), including *WFS1* (Ch 6 and 7).

The lipin 1 gene is responsible for two mouse models of lipodystrophy and insulin resistance and has been suggested to influence human insulin sensitivity and adiposity. I sequenced human *LPIN1*, *2* and *3* in insulin resistant patients to identify potential pathogenic mutations and tested for association of common variation in *LPIN1* with metabolic traits underlying T2D. These studies demonstrated that variants in the lipin family are unlikely to be common causes of severe insulin resistance, and that *LPIN1* common variants do not importantly contribute to risk of T2D.

Sequencing genes in the mTOR pathway revealed a number of rare variants in insulin resistant patients. Given that these genes are key players in the insulin signalling pathway some of these variants may be contributing to insulin resistance in patients. More detailed genetic and functional studies are needed to confirm this.

In the *PARL* gene the polymorphism Leu262Val had previously been reported to associate with fasting insulin levels. Despite a larger sample size (N=3666) I could not replicate this result in UK populations (P=0.79).

I contributed to a large candidate gene association study that investigated 1536 SNPs in 84 genes involved in pancreatic β-cell function for association with T2D. This study identified common variants in *WFS1*, a gene responsible for an autosomal recessive form of diabetes, impacting T2D risk. These results were confirmed in additional populations and updated meta-analysis (OR=0.89, *P*-value= 4.9×10^{-11}). In addition, I initiated re-sequencing and genotyping efforts to refine the association signal and investigated whether rare variants in *WFS1* also impact T2D risk. Further work is still required to identify the causal variant, and there was no evidence for association between rare variants and disease risk.

Recently, genome-wide association studies (GWAS), agnostic in terms of prior biological knowledge, have identified a number of genes underlying T2D and related traits. My work, which identified a novel T2D susceptibility gene not detected in initial GWAS results, *WFS1*, suggests that candidate gene studies can sometimes complement genome-wide approaches.

Contents

Figures and Tables

Chapter 1

Introduction

1.1 Overview

Diabetes mellitus is a group of metabolic diseases characterised by high blood glucose levels resulting from defects in insulin secretion, insulin action, or both (2008). The World Health Organisation diagnostic criteria for diabetes mellitus is 1) random plasma glucose > 11.1 mmol/l and symptoms of diabetes mellitus (polydipsia, polyuria or unexplained weight loss), or 2) fasting plasma glucose > 7.0 mmol/l, or 3) 2-hour plasma glucose > 11.1 mmol/l after a 75g glucose load. The vast majority of people with diabetes have one of two forms. Type 1 diabetes (T1D) accounts for 5-10% of diabetes cases and is characterised by absolute deficiency of insulin resulting from auto-immune destruction of the insulin-secreting ß-cells of the pancreas. By far the most prevalent form of diabetes is type 2 diabetes (T2D), accounting for ~90-95% of diabetes cases. My thesis will focus on T2D and related conditions. T2D includes individuals who have insulin resistance (a resistance of insulin-responsive tissues to the metabolic effects of insulin – Sections 1.2, 1.3 and 1.4), and pancreatic ß-cell dysfunction leading to relative insulin deficiency (Sections 1.2 and 1.4). More uncommon cases of diabetes and insulin resistance are single gene disorders causing severe defects in insulin action and ß-cell function (Section 1.5).

The global prevalence of diabetes is projected to rise from 171 million cases in 2000 to 366 million by 2025 (Wild et al. 2004), driven in large part by increasing rates of obesity. Most of this rise will take place in developing countries as a result of increased Westernisation and urbanisation, with numbers of people with diabetes estimated to grow from 84 million to 228 million in developing nations (Hossain et al. 2007). Furthermore, diabetes and obesity are associated with several major causes of morbidity and mortality. For example, T2D is among the most common causes of blindness in the elderly, (non-trauma related) lower limb amputation, and renal failure

requiring dialysis or kidney transplantation (O'Rahilly 1997). People with T2DM also have a highly increased risk of myocardial infarction and stroke. The human and economic cost of diabetes and its complications will be a massive burden on developed and developing countries and it is therefore critical to identify the underlying causes of diabetes to help prevention and treatment.

While most of the rise in diabetes prevalence over the last few decades can be explained by changes in lifestyle and diet, there is strong evidence for genetic predisposition (Section 1.6.1). T2D inheritance patterns are complex and many genes, interacting with environmental factors, are thought to affect disease susceptibility. This makes the identification of genetic risk factors challenging, though an explosion in genetic studies of T2D in recent years has yielded a number of susceptibility genes (Section 1.6). There is also increasing evidence that genes harbouring mutations causing monogenic or oligogenic forms of insulin resistance and ß-cell dysfunction can play a role in T2D predisposition (Section 1.6.7). Gaining insights into genes and pathways underlying monogenic and common diabetes will not only contribute to our basic knowledge of human genetics and metabolism, but also aid in the development of diagnostic and prognostic tools, and suggest targets for drugs and other interventions.

1.2 Insulin and glucose homeostasis

1.2.1 Glucose

Glucose is a critical substrate for energy production in the human body, with adverse effects when levels are either too low (hypoglycaemia) or too high (hyperglycaemia). Therefore, blood glucose levels in healthy humans are maintained within a tight physiological range between 4 and 7 mmol/l by a number of feedback mechanisms. In people with diabetes these mechanisms are no longer in balance and glucose homeostasis is disrupted (1.2.5).

1.2.2 Metabolic effects of insulin

Banting and Macleod received a Nobel prize in 1923 for the discovery of a peptide hormone, insulin, able to restore blood glucose homeostasis in people with type 1 diabetes. Insulin is secreted by the ß-cells of the pancreatic islets of Langerhans in response to elevated blood glucose (1.2.3) and regulates tissue development, growth and whole-body glucose homeostasis by regulating carbohydrate, lipid and protein metabolism (1.2.4).

1.2.3 Insulin secretion

There are approximately one million islets of Langerhans in a normal pancreas, each containing several types of endocrine cell, 60-80% of which are insulin-secreting ß-cells (Marchetti et al. 2008). Glucose enters the β-cells through GLUT2 transporters and is phosphorylated by glucokinase, preventing movement back across the plasma membrane. Metabolism of glucose results in a rise in the ATP:ADP ratio stimulating closure of ATP-sensitive K^+ (K_{ATP}) channels, depolarization of the β-cell plasma membrane, opening of voltage-gated Ca^{2+} channels, Ca^{2+} influx, a rise in cytoplasmic free Ca^{2+} concentration ($[Ca^{2+}]_i$), and activation of exocytosis of insulin secretory vesicles (Figure 1.1). Although glucose is the main insulin secretagogue, other

nutrients, hormones, neurotransmitters and drugs can induce or amplify insulin release.

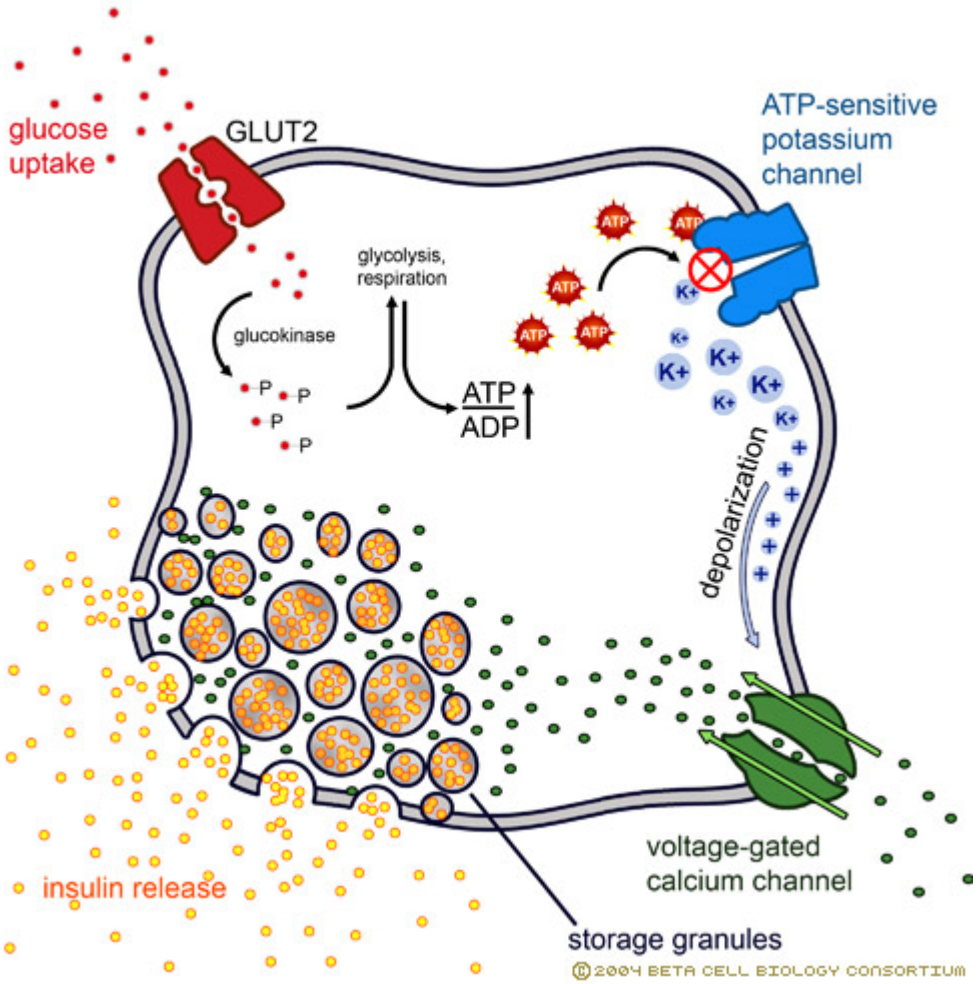


Figure 1.1 The insulin secretion pathway
Provided by the Beta Cell Biology Consortium (2004-2008) (http://www.betacell.org/content/articles/print.php?aid=1). Glucose enters pancreatic β-cells through GLUT2, is phosphorylated by glucokinase, and is used to generate ATP through glycolysis. A rise in the ratio of ATP to ADP results in closure of ATP-sensitive potassium (K_{ATP}) channels, depolarization of the plasma membrane, opening of voltage-gated Ca^{2+} channels, Ca^{2+} influx, and activation of exocytosis of insulin secretory vesicles.

1.2.4 Insulin action

Insulin travels in the bloodstream until it binds cell surface insulin receptors, which stimulates intracellular signalling cascades to downstream effectors of insulin's metabolic and mitogenic effects (1.3). Insulin regulates glucose homeostasis at

many sites. It promotes glucose uptake into muscle and adipose tissue by stimulating translocation of the GLUT4 glucose transporters to the cell surface (Birnbaum 1989; Cushman and Wardzala 1980; James et al. 1988; James et al. 1989; Suzuki and Kono 1980) and promotes storage of glucose as glycogen in muscle and liver by activating glycogen synthase (GS) (Parker et al. 1983). When glucose and circulating insulin levels are low, the liver is able to breakdown glycogen (glycogenolysis) and also to produce glucose from amino acids and glycerol (gluconeogenesis). Insulin reduces glucose output from the liver by inhibiting expression of gluconeogenic enzymes (Barthel and Schmoll 2003).

Insulin also promotes storage of glucose as fat by increasing lipid synthesis in liver and adipose tissue (lipogenesis), and suppresses release of fatty acids from triglycerides stored in adipose tissue and muscle (lipolysis). Consequently, impaired insulin secretion and action in type 2 diabetes results in multiple metabolic abnormalities including hyperglycaemia and abnormal blood lipid levels (dyslipidaemia). Chronic increases in blood glucose and lipids can further disrupt insulin secretion and action, and lead to other tissue damage (1.4).

1.2.5 Insulin action and secretion in type 2 diabetes

When tissues such as liver, muscle and adipose tissue become resistant to normal physiological levels of insulin, pancreatic β-cell mass increases to allow compensatory insulin secretion. This leads to higher circulating levels of insulin (hyperinsulinaemia). Type 2 diabetes develops when insulin resistance is accompanied by pancreatic β-cell dysfunction, which causes relative insulin deficiency and hyperglycaemia (Figure 1.2). Analysis of pancreatic tissue from patients with type 2 diabetes demonstrates diminished ß-cell mass and function (Butler et al. 2003; Del Guerra et al. 2005; Marchetti et al. 2004; Rahier et al. 1983;

Saito et al. 1978; Sakuraba et al. 2002; Westermark and Wilander 1978; Yoon et al. 2003).

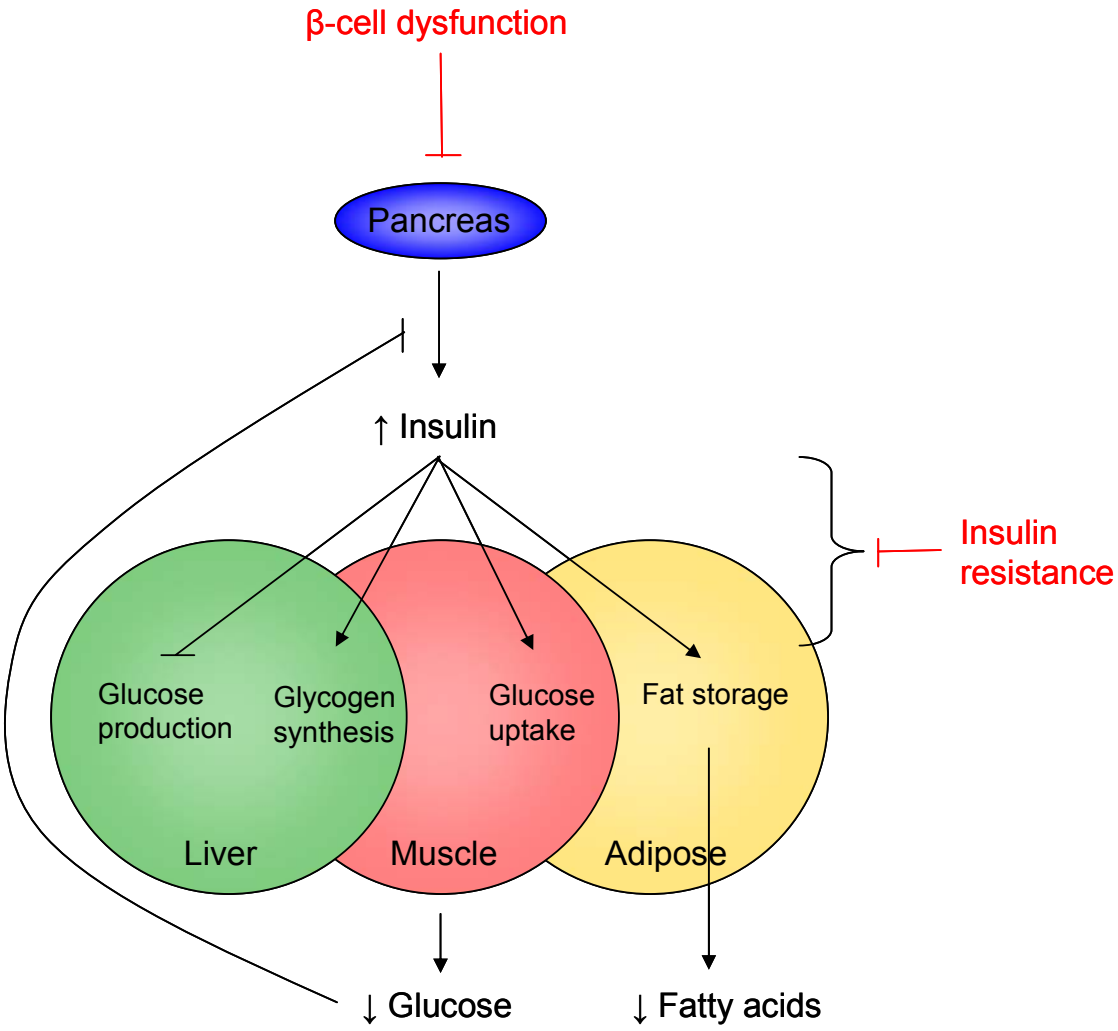


Figure 1.2 Insulin resistance and pancreatic ß-cell dysfunction in the pathogenesis of type 2 diabetes
When blood glucose levels rise the pancreatic β-cell secretes insulin. Insulin inhibits glucose production by the liver, and promotes glucose uptake by muscle and adipose tissue and storage of glucose as glycogen in liver and muscle. This results in a reduction in blood glucose levels and a reduction in insulin release. Type 2 diabetes results from resistance of target tissues to the metabolic effects of insulin and β-cell dysfunction leading to relative insulin deficiency.

1.3 Insulin signalling

Since the discovery of insulin and its effects on metabolic homeostasis, considerable insights into the mechanism of insulin action in insulin-responsive cells have been made (Figure 1.3).

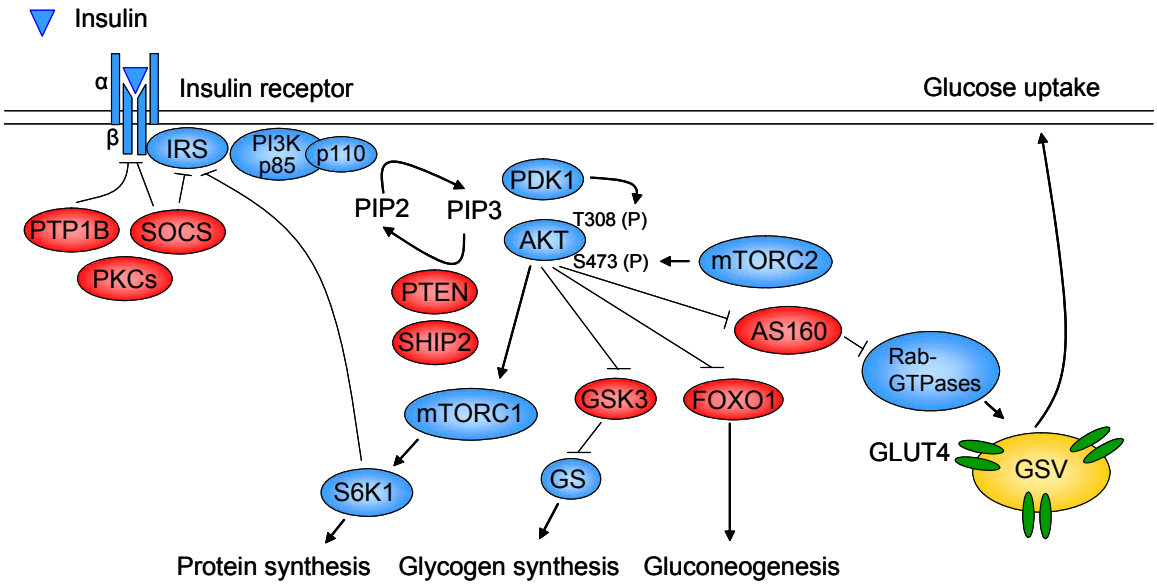


Figure 1.3 Regulation and metabolic effects of the insulin signalling pathway
Insulin binds to the α-subunits of the insulin receptor in target tissues resulting in phosphorylation of β-subunits and activation of insulin receptor substrate (IRS) proteins. The ensuing signaling cascade mediates the metabolic effects of insulin on cell growth and development (protein synthesis), glucose uptake, glucose production, and glycogen synthesis.

1.3.1 The insulin receptor

The insulin receptor belongs to a family of receptor tyrosine kinases (RTKs) (Becker and Roth 1990). It is a heterotetrameric protein, composed of two extracellular α subunits and two transmembrane ß subunits, in which the α subunits inhibit the protein tyrosine kinase activity intrinsic to the ß subunits. Insulin binding to the α subunits leads to a de-repression of the activity of the β subunits, resulting in transphosphorylation on intracellular tyrosine residues and a conformational change that further increases kinase activity.

The critical role of the insulin receptor in mediating the metabolic actions of insulin has been demonstrated by targeted disruption of the insulin receptor gene (Insr) in mice. Mice homozygous for the Insr null allele rapidly develop diabetic ketoacidosis after birth and die within a few days (Accili et al. 1996; Joshi et al. 1996). Hyperglycaemia is also accompanied by increased plasma levels of insulin, fatty acids, and triglycerides, and development of hepatic steatosis (fatty liver). Liver glycogen storage is also reduced.

Since 1996, tissue-specific disruption of the murine Insr gene has enabled researchers to dissect the metabolic role of insulin receptor signalling in different insulin-responsive tissues. Muscle is responsible for the vast majority of glucose disposal following a carbohydrate load, and muscle insulin resistance has been observed at the earliest phases of type 2 diabetes. Surprisingly, muscle-specific insulin receptor knockout (MIRKO) mice have normal glucose tolerance and sensitivity to the glucose lowering effects of exogenously administered insulin, despite markedly reduced insulin receptor expression, activation, and insulin-stimulated glucose transport in muscle *in vitro* (Bruning et al. 1998). Increased glucose transport into adipose tissue appears to compensate, at least in part, for reduced muscle glucose uptake, and indeed MIRKO mice have larger fat depots (Kim et al. 2000).

Liver-specific insulin receptor knock-out (LIRKO) mice are glucose intolerant and resistant to the glucose-lowering effects of exogenously administered insulin (Michael et al. 2000). Mild hyperglycaemia in LIRKO mice appears to result from failure of insulin to suppress hepatic glucose output, as insulin signalling in muscle and adipose tissue remains intact. However, overt diabetes does not develop due to compensatory insulin secretion by pancreatic β-cells and decreased insulin clearance by liver. In contrast, LIRKO mice with β-cell insulin receptor deficiency (βIRKO) develop early onset and progressive hyperglycaemia due to impaired proliferation of

β-cells and glucose-stimulated insulin release (Kulkarni et al. 1999; Okada et al. 2007). Taken together, these results show that insulin resistance accompanied by β-cell dysfunction leads to diabetes. Indeed, diet-induced insulin resistance also progresses to severe hyperglycaemia in β-cell insulin receptor knockout mice (Okada et al. 2007). Astonishingly, βIRKO-MIRKO mice show improved glucose tolerance and early insulin release compared to βIRKO mice, as well as normal glucose uptake into muscle and enhanced glucose uptake by adipose tissue and liver (Mauvais-Jarvis et al. 2000).

The role of adipose tissue insulin signalling has been investigated in fat insulin receptor knockout (FIRKO) mice (Bluher et al. 2002). Despite reduced insulin receptor expression and insulin-stimulated glucose uptake in adipose tissue, these mice are normoglycaemic and are protected against age- and hypothalamic lesion-induced obesity and glucose intolerance. The lean phenotype of these mice may be related to disrupted insulin regulation of lipogenesis and lipolysis. Furthermore, FIRKO mice exhibit increased levels of leptin, an insulin sensitiser secreted by adipose tissue in proportion to fat mass (Bluher et al. 2002). Finally, insulin is thought to be involved in long-term regulation of energy balance through signalling to the brain. Knock-down of insulin receptor in the hypothalamus of rats resulted in greater food intake, fat mass, and hepatic glucose production (Obici et al. 2002). These studies indicate a role for insulin signalling in the brain in the control of appetite and glucose metabolism.

1.3.2 Signalling downstream of the insulin receptor

Once activated, the insulin receptor can phosphorylate a number of adaptor proteins including members of the insulin receptor substrate (IRS) protein family. These substrates act as docking sites for effector molecules that trigger two major kinase cascades, the phosphatidylinositol 3-kinase (PI 3-kinase) and the mitogen-activated protein kinase (MAPK) pathways. The former appears to mediate most of the

metabolic effects of insulin whereas the latter is more predominant in growth and differentiation aspects of insulin action.

1.3.2.1 IRS proteins

IRS proteins have N-terminal pleckstrin-homology (PH) and phosphotyrosine-binding (PTB) domains and up to 20 potential tyrosine phosphorylation sites that can bind effector molecules containing Src-homology 2 (SH2) domains. IRS1 and IRS2 proteins are widely distributed and play distinct but overlapping roles in glucose homeostasis. Irs1 knock-out mice display insulin resistance and growth retardation due to IGF1 resistance (Araki et al. 1994; Tamemoto et al. 1994), as well as increased serum triglycerides (owing to impaired lipolysis) and blood pressure (Abe et al. 1998). Irs2 knock-out mice develop type 2 diabetes due to decreased pancreatic ß-cell mass and failure to compensate for hepatic insulin resistance (Kubota et al. 2000; Withers et al. 1998) and are also dyslipidaemic and hypertensive (Kubota et al. 2003). Altered growth is only seen in a few tissues in Irs2 knock-out mice such as certain neuronal cells (Kubota et al. 2004) and the islets of Langerhans (Withers et al. 1998). Hepatic knock-down of Irs1 increased expression of gluconeogenic enzymes and reduced expression of glucokinase, whereas hepatic deficiency in Irs2 appears to impact on insulin's regulation of lipogenesis through sterol regulatory element binding protein-1c (SREBP-1c) transcription (Taniguchi et al. 2005). Irs1 and Irs2 double knock-outs are early-fetal lethal (Withers et al. 1999), and hepatic knock-down of both causes insulin resistance, glucose intolerance and fatty liver (Taniguchi et al. 2005). Other IRS family members (IRS3 and IRS4) have more discrete patterns of expression and play limited roles in insulin signalling.

1.3.2.2 PI3K

The phosphatidylinositol 3-kinase (PI3K) enzyme is composed of a p85 regulatory subunit and a p110 catalytic subunit, both of which occur in several isoforms. SH2 domains on the regulatory subunit interact with specific phosphotyrosine motifs on

IRS proteins, enabling recruitment of PI3K to the plasma membrane and release of the catalytic subunit of PI3K from the inhibitory effect of the regulatory subunit (Yu et al. 1998). The catalytic subunit of PI3K then converts phosphatidylinositol-4,5-bisphosphate (PIP2) to phosphatidylinositol-3,4,5-triphosphate (PIP3) (Whitman et al. 1988). PIP3 binds proteins with PH domains, allowing their activation at the plasma membrane (Figure 1.3). Potent and specific inhibitors of PI3K or transfection of dominant negative constructs block many metabolic actions of insulin including stimulation of glucose transport, glycogen and lipid synthesis, and adipocyte differentiation (Kanai, 1993; Clarke, 1994; Cheatham, 1994). Knock-out of regulatory subunits of PI3K improves insulin sensitivity (Terauchi et al. 1999; Ueki et al. 2002) and rescues the diabetic phenotype of mice with reductions in insulin receptor and IRS1 (Mauvais-Jarvis et al. 2002). In contrast, deletion of the catalytic subunits produces glucose intolerance and hyperinsulinaemia (Brachmann et al. 2005). These studies suggest that insulin sensitivity is maintained by a balance between the p85 regulatory and p110 catalytic subunits.

1.3.2.3 AKT/PKB

A major target of PIP3 is the cAMP-dependent, cGMP-dependent protein kinase C (AGC) family of serine/threonine protein kinases, including PI-dependent kinase 1 (PDK1), AKT (also known as protein kinase B (PKB)) and some atypical forms of protein kinase C (PKC). AKT binds PIP3 through its PH domain, but insulin-induced activation of AKT requires phosphorylation of Thr308 and Ser473 (Alessi et al. 1996). Thr308 is phosphorylated by PDK1 (Alessi et al. 1997) and Ser473 is thought to be phosphorylated by a complex containing the mammalian target of rapamycin complex 2 (mTORC2) (Ali and Sabatini 2005; Hresko and Mueckler 2005; Sarbassov et al. 2005b) (discussed in more detail below). Activated AKT mediates most of the PI3K-mediated metabolic actions of insulin through phosphorylation of several substrates, including other kinases, signalling proteins and transcription factors (TFs). For example, AKT phosphorylates and deactivates glycogen synthase kinase 3 (GSK3),

thereby preventing inhibition of glycogen synthase and promoting glycogen synthesis (Cross et al. 1995). AKT also phosphorylates members of the winged helix or forkhead (FOX) class of TFs, leading to their exclusion from the nucleus. This prevents FOXO-1 from activating genes encoding enzymes involved in gluconeogenesis (Zhang et al. 2006). Finally, AKT promotes insulin-stimulated glucose uptake by phosphorylating and inhibiting the Rab-GTPase-activating protein, AS160 (AKT substrate of 160 kDa) (Sano et al. 2003). This triggers the activation of Rab small GTPases involved in cytoskeletal re-organisation required for translocation of the glucose transporter GLUT4 to the plasma membrane (Figure 1.3).

There are three different isoforms of AKT encoded by three different genes. AKT1 is ubiquitously expressed and knock-out of Akt1 in mice produces a global defect in growth but has no impact on glucose homeostasis (Cho et al. 2001b). Knock-out of Akt3, the predominant isoform in brain and testes, results in smaller brain size but, again, no defect in glucose homeostasis (Easton et al. 2005). However, knock-out of Akt2, which is expressed in the major metabolic tissues such as the pancreatic ß-cells and skeletal muscle, impairs insulin-stimulated glucose uptake in muscle and suppression of hepatic glucose production (Cho et al. 2001a). These mice are also glucose intolerant, hyperinsulinaemic and hypertriglyceridaemic with a reduction in pancreatic ß-cell and adipose tissue mass (Garofalo et al. 2003). Interestingly AKT2, but not AKT1 or AKT3, has been found co-localised with GLUT4-containing vesicles and phosphorylates proteins involved in docking and fusion of such vesicles (Calera et al. 1998; Yamada et al. 2005).

1.3.2.4 mTOR signalling pathway

The mammalian target of rapamycin (mTOR) exists in two distinct complexes, mTORC1 and mTORC2, which are downstream components of the insulin signalling pathway. As mentioned above, mTORC2 is responsible for phosphorylation of AKT

Ser473 (Ali and Sabatini 2005; Hresko and Mueckler 2005; Sarbassov et al. 2005b) and is also involved in organisation of the actin cytoskeleton (Jacinto et al. 2004; Sarbassov et al. 2004). mTORC1 is activated by insulin signalling through AKT (Manning and Cantley 2003; Tee et al. 2003) and positively regulates cell growth by modulating a number of major biological processes including translation, ribosome biogenesis, nutrient metabolism and macroautophagy (for review see (Wullschleger et al. 2006)). Activation of mTORC1 signalling and its effector p70 ribosomal S6 kinase (S6K1) is also involved in a negative feedback loop that inhibits insulin signalling through IRS proteins. mTORC complexes are discussed in more detail in Chapter 4.

1.3.3 Attenuation of the insulin signalling cascade

Negative regulation can occur at various points in the insulin signalling cascade (Figure 1.3). Autophosphorylation of the insulin receptor is reversed by protein tyrosine phosphatase-1B (PTP1B). Expression of PTP1B is elevated in insulin resistant humans and PTPB1 knockout mice have enhanced insulin sensitivity (Elchebly et al. 1999). Insulin receptor tyrosine kinase activity can also be inhibited by ß-subunit serine/threonine phosphorylation. Furthermore, tyrosine phosphorylation and activation of IRS proteins is opposed by serine phosphorylation on IRS proteins. There is evidence that the actions of insulin itself, as well as free fatty acids, cytokines, cellular stressors, and amino acids (via mTORC1), result in the activation of serine kinases such as protein kinase C (PKC) isoforms, suppressors of cytokine signalling (SOCS) proteins, Jun N-terminal Kinase (JNK) and S6K1. This could serve as a negative feedback mechanism for the insulin signalling pathway. The conversion of PIP2 to PIP3 can also be reversed by phosphatases, PTEN (phosphatase and tensin homologue on chromosome 10) and SHIP2 (SH2-domain-containing inositol 5-phosphate-2).

1.4 Mechanisms of insulin resistance and pancreatic β-cell dysfunction

1.4.1 Adipocyte dysfunction

Adipocytes within white adipose tissue store excess energy derived from food intake in the form of triglycerides, mostly in a single large lipid droplet. In times of caloric need, triglycerides, diglycerides and monoglycerides can be hydrolysed by adipose triglyceride lipase (ATGL), hormone sensitive lipase (HSL), and monoglyceride lipase respectively (lipolysis) to release fatty acids that can be oxidised by mitochondria in other tissues to generate energy. Increased adipose tissue mass (as in obesity) and absence or deficiency of adipose tissue (lipodystrophy), are both associated with insulin resistance. In both scenarios, the capacity of adipocytes to store excess lipids becomes saturated, resulting in lipid deposition in non-adipose tissue such as muscle, liver, and pancreas. Furthermore, obesity and lipodystrophy can cause aberrant secretion of adipokines, the collective term for the large number of hormones, cytokines, and growth factors secreted by adipose tissue to influence whole-body energy balance and nutrient metabolism.

1.4.1.1 Ectopic accumulation of lipids

High circulating fatty acids (as seen in obesity and lipodystrophy) lead to abnormally high lipid stores in liver (known as fatty liver or hepatic steatosis) and in muscle (muscle steatosis). When enhanced lipid accumulation is not balanced by an increase in fatty acid oxidation (the process of breaking down fatty acids to produce energy) lipid-derived metabolites such as diacylglycerol (DAG) can build up. Lipid intermediates are then thought to induce serine phosphorylation on IRS proteins through activation of serine kinases such as PKC isoforms, I kappa B kinase (IKK), and JNK (Gao et al. 2004; Itani et al. 2002). Insulin resistance in animal models and humans has been shown to correlate with intramyocellular lipid content (Pan et al. 1997; Perseghin et al. 1999; Phillips et al. 1996; Storlien et al. 1991) and hepatic lipid

content (Kotronen et al. 2008; Ryysy et al. 2000; Seppala-Lindroos et al. 2002). However, it remains controversial whether lipid accumulation in muscle and liver causes whole-body insulin resistance, or whether it is in fact merely a marker for insulin resistance. The lipogenic effects of insulin may cause, or at least exacerbate, lipid accumulation in ectopic tissues in states of hyperinsulinaemia (Savage et al. 2007).

High circulating fatty acids can also impact negatively on pancreatic β-function. Elevated fatty acid concentrations have been shown to increase basal insulin release and decrease glucose-stimulated insulin secretion in vitro (Sako and Grill 1990), in rats (Mason et al. 1999), and in humans (Paolisso et al. 1995). Fatty acids have also been shown to inhibit insulin gene expression *in vitro* (Kelpe et al. 2003). Fatty acids in concert with high glucose also promote β-cell death (El-Assaad et al. 2003). Many mechanisms have been suggested to mediate this effect including production of lipid intermediates and oxidative stress. More recently it has been shown that fatty acids can induce markers of ER stress and changes in ER morphology (discussed in 1.4.3), leading to β-cell death (Laybutt et al. 2007). There is strong evidence that the detrimental effects of fatty acids on β-cell function only occur in the presence of elevated glucose, which directs fatty acid partitioning away from oxidation and towards storage (Prentki and Corkey 1996).

1.4.1.2 Adipokines

The traditional view of adipose tissue as a passive organ for energy storage has been challenged by recent discoveries that adipocytes express and secrete a variety of adipokines. Many of these have been implicated in the regulation of systemic insulin sensitivity such as leptin, adiponectin, resistin, retinol binding protein (RBP)4, and various inflammatory factors. For example, genetic disruption of leptin and its receptor cause animal models of obesity and diabetes (Chen et al. 1996; Lee et al.

1996; Tartaglia et al. 1995; Zhang et al. 1994). Leptin acts in the hypothalamus to regulate food intake (Stephens et al. 1995), hepatic insulin sensitivity and glucose production (Cohen et al. 1996), and accumulation of triglycerides in peripheral tissues by inhibiting fatty acid synthesis and stimulating fatty acid oxidation (Minokoshi et al. 2002). There is evidence that the effect of leptin on lipogenesis and fatty acid oxidation is largely mediated by its effect on food uptake and other central mechanisms (Prieur et al. 2008). However, a direct effect on peripheral tissues cannot be excluded.

1.4.1.3 Animal models of lipodystrophy

Various animal models of lipodystrophy have demonstrated the importance of adipose tissue function to insulin sensitivity. Transgenic mice expressing a dominant negative protein, A/ZIP-F, which inhibits the function of transcription factors involved in adipose tissue development, had no white adipose tissue and exhibited increased food intake (hyperphagia), fatty liver, elevated plasma triglycerides and fatty acids, and hyperinsulinaemia. After 3 weeks hyperglycaemia developed as pancreatic β-cells were no longer able to compensate for insulin resistance (Moitra et al. 1998). Similarly, mice overexpressing constitutively active sterol response element binding protein (SREBP)-1c in adipose tissue, which results in lower expression of genes essential for adipose tissue differentiation such as peroxisome proliferator-activated receptor γ (*PPARγ*), were lipoatrophic (Shimomura et al. 1998). These mice also developed fatty livers, elevated plasma triglycerides, insulin resistance and diabetes. Both mouse models also exhibited low leptin levels, the importance of which was demonstrated by several experiments. First, the phenotype of A/ZIP-F transgenic mice was rescued by transplanting adipose tissue from wild-type mice into the transgenic mice (Gavrilova et al. 2000), but not by leptin-deficient adipose tissue from *ob/ob* mice unless combined with exogenous leptin administration (Colombo et al. 2002). Leptin also restored insulin sensitivity in SREBP-1c transgenic mice

(Shimomura et al. 1999). Overexpression of leptin from liver (Ogawa et al. 1999) generates lipodystrophic mice with lower ectopic deposititon of fat and increased liver and skeletal muscle insulin sensitivity. These results suggest that leptin rather than adipose tissue *per se* is essential for insulin sensitivity.

Finally, spontaneous mutations in the lipin 1 gene, Lpin1, cause lipodystrophy, hepatic steatosis, reduced leptin and insulin resistance in the fatty liver dystrophy (fld) mouse (Peterfy et al. 2001). Lipin family members are discussed in more detail in Chapter 3.

1.4.2 Mitochondrial dysfunction

Ectopic storage of lipids is not only a symptom of high circulating fatty acids in obesity and lipodystrophy. Impaired fatty acid oxidation due to mitochondrial dysfunction is also thought to predispose to intramyocellular lipid accumulation and insulin resistance. Consistent with this, evidence of reduced mitochondrial function has been observed in various insulin resistant states (Morino et al. 2006) (discussed in more detail in Chapter 5).

1.4.3 ER stress

The ER provides a unique oxidising environment and numerous protein chaperones for folding and assembly of membrane and secreted proteins in eukaryotic cells. Disruption of ER homeostasis and accumulation of misfolded proteins (ER stress) causes activation of the unfolded protein response (UPR), which aims to alleviate ER stress by attenuating protein translation to prevent further accumulation of unfolded proteins, inducing expression of ER chaperones and folding enzymes, and extruding irreversibly misfolded proteins for degradation (Eizirik et al. 2008). The UPR therefore is particularly important in pancreatic β cells to allow adaptation to the fluctuating physiological demand for insulin biosynthesis. If these mechanisms fail to correct protein folding defects, prolonged UPR activation will lead to apoptosis.

Prolonged exposure to hyperglycaemia and/or free fatty acids leads to hyperactivation of the UPR, β cell dysfunction and apoptosis. Indeed, ER stress markers are up-regulated in the islets of db/db mice and in pancreas sections from humans with type 2 diabetes (Laybutt et al. 2007) implying that ER stress plays a role in increased rates of β cell apoptosis in diabetes.

Genetic ablation of UPR components provides further evidence for the importance of ER stress responses for β cell survival and glucose homeostasis. For example, mice with deletion of double-stranded RNA-activated kinase (PKR)-like ER kinase (PERK) cannot phosphorylate eIF2α to inhibit protein synthesis and develop diabetes within a few weeks after birth due to progressive β-cell loss (Harding et al. 2001; Zhang et al. 2002). Similarly, inactivating mutations in the human ortholog of Perk, *EIF2K3*, cause a monogenic form of diabetes, Wolcott-Rallison syndrome, in humans. In addition, mice homozygous for a non-phosphorylable version of eIF2α die within 24 hours and show severe β-cell deficiency (Scheuner et al. 2001), whereas heterozygotes are more susceptible to diet-induced obesity and diabetes due to decreased islet insulin content and nutrient-stimulated insulin secretion (Scheuner et al. 2005). The ER in the β-cells of these mice showed evidence of delayed folding and/or misfolding of proinsulin in response to a high fat diet (Scheuner et al. 2005). Another component of the UPR, the Wolfram syndrome gene 1 (*WFS1*), is responsible for Wolfram syndrome, which is characterised by early-onset diabetes and progressive β-cell loss (Inoue et al. 1998; Strom et al. 1998). This gene is discussed in more detail in Chapters 6 and 7.

Recent studies have suggested that ER stress can also impact on insulin action in peripheral tissues. Molecular markers of ER stress are up-regulated in the liver and adipose tissue of diet-induced and ob/ob mouse models of obesity (Ozcan et al. 2004). Inducers of ER stress and genetic ablation of UPR components also reduced

insulin signalling in vitro and in vivo (Ozcan et al. 2004) and treatment of ob/ob mice and obese db/db mice with molecules that act as ER chaperones relieved ER stress in the liver and improved glucose tolerance and insulin sensitivity (Nakatani et al. 2005; Ozcan et al. 2006).

1.5 Genetic causes of insulin resistance and pancreatic β-cell dysfunction

1.5.1 Inherited syndromes of insulin resistance in humans

Human syndromes of severe insulin resistance (SIR) (Table 1.1) are characterised by resistance of muscle, adipose tissue, and liver to the metabolic effects of insulin. The hallmarks of such syndromes are high circulating levels of insulin to overcome insulin resistance, and resultant features such as the skin lesion, acanthosis nigricans. Other metabolic and morphological features can characterise specific syndromes.

1.5.1.1 Insulin receptor syndromes

Insulin receptor gene (*INSR*) defects produce a spectrum of insulin resistance syndromes with variable degrees of severity. Patients normally carry lesions in both copies of *INSR*, either as homozygotes or compound heterozygotes. The most severe is leprechaunism (Donahue syndrome), a rare disease usually resulting in mortality before one year of age (Krook and O'Rahilly 1996). Mutations are either in the extracellular domain or cause a premature stop giving rise to receptors that generally retain less than 10% of the insulin binding capacity of wild-type receptors (Longo et al. 2002). Additional features of the disease include intrauterine growth retardation, dysmorphic facies, lipoatrophy, distended abdomen, enlarged genitalia in males and polycystic ovary syndrome (PCOS) in females, which is characterised by infrequent or irregular ovulation, excess androgens, and cysts on the ovaries.

Patients with Rabson-Mendenhall syndrome are distinguished from Donohue syndrome patients by abnormal dentition and fingernails and thick, rapidly growing scalp hair. Often diabetes with ketoacidosis and chronic complications will also occur. Unlike leprechaunism, patients generally survive beyond one year of life as *INSR* mutations lead to receptors with up to 25% of normal insulin binding activity (Longo et al. 2002).

A milder form of insulin receptor dysfunction can cause Type A syndrome, in which patients survive into adulthood with insulin resistance, acanthosis nigricans, and hyperandrogenism in females, which includes hirsutism, menstrual disturbance, and masculinisation (Moller and Flier 1988). In some cases, inheritance follows an autosomal dominant pattern with mutant insulin receptors acting in a dominant negative manner by forming hybrids with wild-type receptors and inhibiting their function (Levy-Toledano et al. 1994).

1.5.1.2 Lipodystrophies

Lipodystrophies are a heterogeneous group of diseases characterised by loss of adipose tissue, and fat deposition in other organs that do not normally store fat. This is associated with metabolic abnormalities such as insulin resistance, acanthosis nigricans, dyslipidaemia, and hypertension. Females often suffer from hyperandrogenism and PCOS. A rare autosomal recessive form of lipodystrophy with near absence of adipose tissue at birth is congenital generalised lipoatrophy (CGL or Berardinelli-Siep syndrome). Muscle and liver are enlarged due to excess fat and glycogen deposition, and excess lipids in the bloodstream often cause pancreatitis and glucose intolerance or diabetes in early adolescence. Whole-genome linkage scans (Garg et al. 1999) followed by a positional candidate gene approach (Agarwal et al. 2002) identified the gene responsible for Berardinelli-Seip congenital lipodystrophy type 1 (BSCL1), which encodes 1-acylglycerol-3-phosphate 0-acetyltransferase (AGPAT2). This enzyme catalyses a key step in the synthesis of TAG and glycerophospholipids. Linkage and fine-mapping identified the *BSCL2* gene which encodes Seipin (Magre et al. 2001), an integral endoplasmic reticulum (ER) protein with a role in adipocyte differentiation, expression of lipogenic genes (Payne et al. 2008), and lipid droplet formation (Szymanski et al. 2007). More recently, mutations in the caveolin 1 (*CAV1*) gene have been shown to cause a third

type of CGL (Kim et al. 2008), as well as atypical partial lipodystrophy (Cao et al. 2008).

Familial partial lipodystrophies are characterised by diabetes mellitus, dyslipidaemia, acanthosis nigricans, and an abnormal distribution of subcutaneous fat. Familial partial lipodystrophy 1 (FPLD1) (Kobberling et al. 1975) is the most difficult to diagnose as fat loss is usually restricted to the limbs. No causative genes have yet been found. Patients with FPLD2 (Dunnigan variety) (Dunnigan et al. 1974) experience progressive subcutaneous fat loss during puberty resulting in atrophy of gluteal and truncal adipose depots. Female patients can experience hirsutism, menstrual disturbance and PCOS. Genome-wide linkage scans (Peters et al. 1998) and positional candidate gene approaches led to the identification of mutations in the lamin A/C (*LMNA*) gene (Cao and Hegele 2000; Shackleton et al. 2000). This gene is alternatively spliced to produce two proteins, lamin A and lamin C, which are essential scaffolding components of the nuclear envelope (Stuurman et al. 1998). This gene underlies over 11 different diseases referred to as laminopathies. Mutations affecting the C-terminal domain of lamin A in FPLD patients are associated with changes in nuclear morphology, aggregation of mutant unprocessed precursor lamin A, and are thought to cause lipodystrophy through disruption of interactions between lamin A/C and nuclear factors involved in adipocyte differentiation (Maraldi et al. 2007). *LMNA* mutations also cause mandibuloacral dysplasia (MAD) with type A lipodystrophy, which includes loss of subcutaneous fat at extremities (Novelli et al. 2002). Loss of fat in type B MAD is generalised and has been shown to result from mutations in an enzyme that plays a role in proteolytic cleavage of prolamin A to lamin A (Agarwal et al. 2003). Finally, a *LMNA* mutation has also been reported in a disease of generalised lipoatrophy with diabetes, hepatic steatosis, hypertrophic cardiomyopathy, and leukomelanodermic papules (LDHCP)) (Caux et al. 2003).

FPLD3 is characterised by loss of subcutaneous fat in limbs and buttocks, often accompanied by fatty liver, early-onset hypertension and diabetes mellitus (Agarwal and Garg 2002; Barroso et al. 1999; Hegele et al. 2002). A large candidate gene study in a cohort of severely insulin resistant patients identified dominant negative mutations in the ligand-binding domain of *PPARγ* which segregated with FPLD in two families (Barroso et al. 1999). PPARγ belongs to a family of nuclear receptors and functions as a transcriptional regulator of adipogenesis and triglyceride synthesis and storage.

PPARγ also causes a digenic disease of severe insulin resistance (Savage et al. 2002). Double heterozygotes for frameshift mutations in *PPARγ* and *PPP1R3A*, which is involved in muscle glycogen synthesis, exhibit acanthosis nigricans and hyperinsulinaemia (Savage et al. 2002).

There are many additional syndromes of insulin resistance, some of which are listed in Table 1.1. For a review of the genetics of syndromes of insulin resistance also see (Barroso 2005).

Table 1.1 Inherited syndromes of insulin resistance in humans
(adapted from (Barroso 2005))

Category	Condition	OMIM	Inheritance	Gene	Reference
Insulin receptor syndromes	Leprechaunism (Donahue)	246200	AR	*INSR*	(Kadowaki et al. 1988)
	Rabson-Mendenhall	262190	AR	*INSR*	(Kadowaki et al. 1990)
	Type A	610549	AR/AD	*INSR*	(Moller and Flier 1988)
Lipodystrophy syndromes	Berardinelli-Siep CGL1	608594	AR	*AGPAT2*	(Agarwal et al. 2002)
	Berardinelli-Siep CGL2	269700	AR	*BSCL2*	(Magre et al. 2001)
	Berardinelli-Siep CGL3	*	AR	*CAV1*	(Kim et al. 2008)
	FPLD1	608600	AD	Unknown	(Kobberling et al. 1975)
	FPLD2 (Dunnigan variety)	151660	AD	*LMNA*	(Cao and Hegele 2000)
	FPLD3	604367	AD	*PPARγ*	(Barroso et al. 1999)
	MADA	248370	AR	*LMNA*	(Novelli et al. 2002)
	MADB	608612	AR	*ZMPSTE24*	(Agarwal et al. 2003)
	LDHCP	608056	AD	*LMNA*	(Caux et al. 2003)
Other syndromes	SIR, acanthosis nigricans, and dyslipidaemia	125853	Digenic	*PPARγ* and *PPP1R3A*	(Savage et al. 2002)
	Alstrom	203800	AR	*ALMS1*	(Collin et al. 2002; Hearn et al. 2002)
	Myotonic dystrophy	160900	AD	*DMPK*	(Fu et al. 1992)
	Pseudoacromegaly and SIR	602511	Unknown	Unknown	(Flier et al. 1993)
	SIR and diabetes mellitus	125853	AD	*AKT2*	(George et al. 2004)
	Ataxia-telangiectasia (A-T)	208900	AR	*ATM*	(Savitsky et al. 1995)
	Werner Syndrome	277700	AR	*RECQL2* or *LMNA*	(Chen et al. 2003; Yu et al. 1996)
	Mulibrey nanism (MUL)	253250	AR	*TRIM37*	(Karlberg et al. 2005)

* Has not been assigned an OMIM identification yet

1.5.2 Inherited defects in ß-cell function

Monogenic forms of diabetes (Table 1.2) resulting from mutations that reduce ß-cell function account for 1-2% of diabetes cases and can be divided into four broad clinical situations (Murphy et al. 2008):

1.5.2.1 Diabetes diagnosed before 6 months of age

Neonatal diabetes mellitus is a rare condition characterised by neonatal hyperglycaemia and hypoinsulinaemia, and often low birth weight (Shield 2000). Diabetes resolves in approximately half of patients, sometimes returning during adolescence or early adulthood (Arthur et al. 1997). Most cases of transient neonatal diabetes mellitus (TNDM) are linked to imprinting abnormalities in the chromosome 6q24 region (Temple and Shield 2002). Permanent neonatal diabetes mellitus (PNDM) does not go into remission and is less common than the transient form. Half of patients with PNDM have mutations in *KCNJ11* (potassium inwardly rectifying channel, subfamily J, member 11 gene) (Gloyn et al. 2004) or *ABCC8* (ATP-binding cassette, subfamily C, member 8 gene) (Proks et al. 2006), which encode Kir6.2 and SUR1 subunits of the ATP-sensitive potassium channel (K_{ATP} channel). Mutations mostly reduce the sensitivity of the K_{ATP} channel to ATP, preventing channel closure and insulin secretion in the pancreatic ß-cells, explaining why insulin production is relatively low in these patients. *KCNJ11* and *SUR1* mutations have also been found in cases of TNDM (Babenko et al. 2006; Gloyn et al. 2005). Mutations in the insulin gene (*INS*) appear to account for 15-20% of PNDM (Stoy et al. 2007), whereas other known genetic causes are relatively rare. *KCNJ11* and *ABCC8* mutations also cause persistent hyperinsulinaemia hypoglycaemia in infancy (PHHI) (Thomas et al. 1996), and recently, mutations in the GLI-similar 3 (*GLIS3*) gene (which encodes a transcription factor expressed in pancreatic β-cells) were shown to cause a new neonatal diabetes syndrome associated with congenital hyperthyroidism (Senee et al. 2006).

1.5.2.2 Familial, mild fasting hyperglycaemia

Patients with mild fasting hyperglycaemia (5.5-8 mmol/l) that does not deteriorate with age (also known as monogenic diabetes of the young (MODY) 2) are asymptomatic and usually require no specific treatment. Linkage analysis (Froguel et al. 1992) and positional candidate gene screening (Hattersley et al. 1992; Vionnet et al. 1992) showed that patients are heterozygous for mutations in the glucokinase gene (*GCK*), which catalyses phosphorylation of glucose to glucose-6-phosphate and controls the rate limiting step of glycolysis. Glucokinase activity is therefore essential for hepatic glycogen synthesis and glucose sensing in the ß-cells (Matschinsky et al. 1993), but patients produce an adequate insulin response because the wild-type *GCK* allele is able to compensate. Homozygous mutations result in an insulin-dependent form of PNDM (Njolstad et al. 2001), and heterozygous activating mutations result in hyperinsulinaemia and hypoglycaemia (Glaser et al. 1998).

1.5.2.3 Familial, young-onset diabetes

Patients with familial, young-onset forms of diabetes have normal glucose levels at birth but experience deterioration in glucose tolerance due to progressive ß-cell failure. Diabetes develops before age 25 years, often accompanied by diabetic complications (Tattersall and Fajans 1975). Heterozygous mutations in transcription factors with important roles in pancreatic development and function account for most cases of disease.

Linkage analysis localised the genes responsible for two forms of familial, young-onset diabetes to chromosomes 20 (Bell et al. 1991) and 12 (Vaxillaire et al. 1995) and a combination of candidate gene and fine-mapping approaches (Yamagata et al. 1996a; Yamagata et al. 1996b) led to identification of causative mutations in *HNF4α* and *HNF1α*. Both genes encode transcription factors within the same complex regulatory network. Diabetes caused by *HNF1α* mutations (MODY3) is the most

common, accounting for 1-2% of all diabetes cases (Murphy et al. 2008). Whereas, diabetes caused *HNF4α* (MODY1) is much rarer.

Other rarer genetic causes of MODY are mutations in transcription factor genes insulin promoting factor 1 (*IPF1*) (MODY4) (Stoffers et al. 1997), neurogenic differentiation 1 (*NEUROD1*) (MODY6) (Malecki et al. 1999), Kruppel-like factor 11 (*KLF11*) (MODY7) (Neve et al. 2005), and paired box gene 4 (*PAX4*) (MODY9) (Plengvidhya et al. 2007), and the enzyme carboxyl ester lipase (*CEL*) (MODY8) (Raeder et al. 2006).

1.5.2.4 Diabetes with extrapancreatic features

Other inherited forms of monogenic diabetes include extrapancreatic features. For example, Wolfram Syndrome (also known as DIDMOAD) is characterised by diabetes insipidus, diabetes mellitus, optic atropy and deafness. Mutations in *WFS1*, which encodes a protein involved in the ER stress response in ß-cells, (Inoue et al. 1998; Strom et al. 1998) account for at least 90% of Wolfram Syndrome cases. A missense mutation in another ER membrane protein, the CDGSH iron sulphur domain protein 2, *CISD2*, which leads to aberrant RNA splicing, has also been show to cause Wolfram Syndrome in three consanguineous families of Jordanian descent (Amr et al. 2007). The eukaryotic initiation factor 2α kinase 3 (*EIF2AK3*) gene (component of the unfolded protein response) causes Wolcott-Rallison syndrome, which is characterised by early onset diabetes, epiphyseal dysplasia (a disorder of bone and cartilage development at the ends of the long bones in the arms and legs), renal impairment, acute hepatic failure and developmental delay (Brickwood et al. 2003; Delepine et al. 2000; Durocher et al. 2006).

Another form of monogenic diabetes featuring renal disease (MODY5) is caused by mutations in the transcription factor, *HNF1β*, which is part of the same regulatory

network as HNF1α and HNF4α (Horikawa et al. 1997). Patients with MODY5 are also more insulin resistant than *HNF1α* and *HNF4α* mutation carriers and female genital-tract malformations, gout and hyperuricaemia can also occur. Another form of diabetes with extrapancreatic features is maternally inherited diabetes and deafness (MIDD) (van den Ouweland et al. 1992), which is caused by mutations in mitochondrial DNA, particularly 3243A>G (Ciafaloni et al. 1992). This causes mitochondrial dysfunction, insulin deficiency due to disruption of ß-cell function and a decrease in ß-cell mass.

1.5.2.5 The clinical benefits of genetic analysis

Monogenic forms of diabetes have often been misdiagnosed as type 1 or early-onset type 2 diabetes, resulting in inappropriate treatment regimes. For example, patients with neonatal diabetes caused by mutations in the K_{ATP} channel have little or no endogenous insulin secretion and so lifelong insulin treatment was thought to be required. However, ~90% of such patients can achieve improved glycaemic control by transfering to sulphonylureas, which bind SUR1 subunits and close the K_{ATP} channels in an ATP-sensitive manner (Zung et al. 2004). Also, mild hyperglycaemia caused by heterozygous mutations in *GCK* does not usually require hypoglycaemic medication as patients do not usually develop complications. Identification of the underlying molecular genetic cause of diabetes is therefore important as it will help predict disease progression and indicate appropriate treatment strategies (Murphy et al. 2008).

Table 1.2 Monogenic forms of diabetes in humans

Category	Condition	OMIM	Inheritance	Gene(s)	Reference
Diabetes diagnosed before 6 months of age	PNDM	606176	AD/AR	*KCNJ11* or *ABCC8* or *INS* or *GCK*	(Gloyn et al. 2004; Njolstad et al. 2001; Proks et al. 2006; Stoy et al. 2007)
	TNDM1	601410	AD/AR	Chromosome 6q24	(Temple and Shield 2002)
	TNDM2	610374	AD/AR	*ABCC8*	(Babenko et al. 2006)
	TNDM3	610582	AD/AR	*KCNJ11*	(Gloyn et al. 2005)
	Neonatal diabetes mellitus and congenital hypothyroidism	610199	AR	*GLIS3*	(Senee et al. 2006)
Familial, mild fasting hyperglycaemia	MODY2	125851	AD	*GCK*	(Hattersley et al. 1992; Vionnet et al. 1992)
Familial, young-onset diabetes	MODY3	600496	AD	*HNF1α*	(Yamagata et al. 1996b)
	MODY1	125850	AD	*HNF4α*	(Yamagata et al. 1996a)
	MODY4	606392	AD	*IPF1*	(Stoffers et al. 1997)
	MODY6	606394	AD	*NEUROD1*	(Malecki et al. 1999)
	MODY7	610508	AD	*KLF11*	(Neve et al. 2005)
	MODY8	609812	AD	*CEL*	(Raeder et al. 2006)
	MODY9	612225	AD	*PAX4*	(Plengvidhya et al. 2007)
Diabetes with extrapancreatic features	Wolfram Syndrome	222300	AR	*WFS1*	(Inoue et al. 1998; Strom et al. 1998)
	Wolcott-Rallison syndrome	226980	AR	*EIF2AK3*	(Delepine et al. 2000)
	MODY5	137920	AD	*HNF1β*	(Horikawa et al. 1997)
	MIDD	520000	Maternally transmitted	Mitochondrial DNA	(Ciafaloni et al. 1992)

1.6 Genetics of common type 2 diabetes

1.6.1 Type 2 diabetes is a genetic disease

Though the rise in prevalence of type 2 diabetes over recent decades has been the result of rising obesity and changes in lifestyle and diet, there is evidence of a hereditary component to disease from a number of sources. The incidence of type 2 diabetes varies considerably between ethnic groups, from less than 1% in rural areas of developing countries, to up to half of populations on the pacific island of Nauru, the Aborigines of Australia, and American-Indian groups in the US (King and Rewers 1993). Environmental differences between these geographic populations will influence disease prevalence. For example, Pima Indians living in more traditional rural environments in Mexico have less than one-fifth of the incidence of type 2 diabetes seen in genetically similar Pima Indians living a more Westernised lifestyle in Arizona (Schulz et al. 2006). However, studies in admixed populations, in which alleles from two once geographically isolated populations unite, demonstrate the importance of genetic risk alleles to ethnic differences in disease risk. For example, the prevalence of type 2 diabetes in Pima Indians is inversely related to the extent of interbreeding with European Americans (Williams et al. 2000). Also, incidence of type 2 diabetes varies between ethnic groups living in a shared environment. For example, in the UK, individuals of African-Caribbean and South-East Asian decent have a higher risk of diabetes compared with individuals of European decent (Chaturvedi et al. 1993; Simmons et al. 1991).

Further evidence for a genetic component to type 2 diabetes comes from studies of disease incidence in families. Risk of type 2 diabetes has been shown to be higher in first degree relatives of patients with type 2 diabetes, compared to risk in more distant relatives and the general population. In a recent study the relative risk of developing type 2 diabetes in first-degree relatives of patients with type 2 diabetes was estimated

to be 2.24, compared to relative risks of 1.36 and 1.14 in second- and third-degree relatives (Weires et al. 2007). There is also a suggestion that relatives of type 2 diabetic probands are more likely to exhibit hyperinsulinaemia, insulin resistance, and glucose intolerance (Nauck et al. 2003; Tripathy et al. 2003). The higher prevalence of disease in family members is thought to be because of an increased number of shared genes, but could also be driven by shared environmental and cultural factors. To assess the extent to which familial aggregation can be accounted for by inherited genetic factors, twin studies have been used. As monozygotic (MZ) twins are genetically identical whereas dizygotic (DZ) twins share on average half of their genes, increased disease concordance rates in MZ compared to DZ twins indicate the presence of genetic factors contributing to disease predisposition. This conclusion assumes that DZ twins share the same amount of environmental factors as MZ twins, which may not be true. In particular, the intrauterine environment may be more similar between MZ twins than DZ twins (Poulsen and Vaag 2001). Though there is significant between study variability in concordance rates, the concordance between MZ twins was found to be higher than that in DZ twins in all studies (Barroso 2005; Condon et al. 2008). The incomplete concordance between MZ twins also supports a role for non-genetic factors in type 2 diabetes susceptibility. A final proof for a genetic component to type 2 diabetes is the fact that genes influencing diabetes risk have already been found (discussed in detail below).

1.6.2 The "geneticist's nightmare"

Eminent geneticist James V Neel referred to type 2 diabetes as the "geneticist's nightmare" in reference to the complex aetiology of the disease and consequent difficulty in identifying risk factors. Type 2 diabetes displays a complex inheritance pattern, and many genes, environmental factors, and the interactions between them are predicted to affect disease predisposition. Each of these predisposing factors is expected to have only modest effects on disease risk, meaning only large studies

have good statistical power to detect them (Risch and Merikangas 1996). At the molecular level, type 2 diabetes is likely to be a collection of many diseases with varying but overlapping aetiologies giving rise to similar phenotypes. Therefore, genetic variations are likely to contribute to disease to different extents in different populations making replication difficult (Barroso 2005). Furthermore, incomplete penetrance of susceptibility factors (where the penetrance of a given risk factor indicates the proportion of individuals exposed to the risk factor that exhibit the clinical manifestations of disease) means individuals carrying type 2 diabetes susceptibility factors will have varying degrees of disease severity, or no disease at all (Risch and Merikangas 1996). It is therefore difficult to define subpopulations with similar aetiological factors. Finally, type 2 diabetes has a variable age of onset therefore some apparently unaffected people will become affected later in life, reducing the power of case-control studies to identify susceptibility loci. Despite these challenges, a range of genetic study designs have been exploited to discover type 2 diabetes susceptibility loci. In particular, once the importance of large sample sizes was recognised studies had greater power to find and replicate loci.

1.6.3 Linkage studies

The aim of linkage analysis is to identify genomic regions, represented by polymorphic markers, which are shared by descent among relatives with disease in families. As recombination between alleles at two loci becomes increasingly unlikely with decreasing distance along a chromosome, cosegregating markers define an area of genome likely to be in close proximity to the disease locus (Borecki and Province 2008). As discussed above, this approach was employed successfully to identify disease loci underlying Mendelian syndromes of insulin resistance and diabetes. However, several factors decrease sharing by descent of complex disease risk alleles. Firstly, because predisposing alleles only cause a small increase in risk of disease, some unaffected relatives may carry the risk allele under study.

Secondly, due to the genetic heterogeneity of complex disease, some relatives may be affected because of other causes and may not carry the risk allele under study (Risch and Merikangas 1996). Given these limitations it is not surprising that, out of many type 2 diabetes linkage studies covering either candidate regions of the genome or the entire genome in a variety of populations, only a few regions have shown genome-wide significant evidence for linkage (Lander and Kruglyak 1995), and even fewer have shown strong evidence for linkage in multiple populations. This suggests that few susceptibility loci have a strong effect on type 2 diabetes risk in most populations (Barroso 2005). Alternatively these results suggest many of the potential loci may be false positives or suffer from the "winner's curse" (Lohmueller et al. 2003), that is the original study over-estimates the true effect of the loci on type 2 diabetes risk due to random fluctuations in disease parameters across studies. For this reason many of the replication studies may have been underpowered to replicate the initial finding given that larger samples sizes than the original study would have been needed to ensure sufficient power.

1.6.3.1 T2D genes found by linkage analysis

The first putative type 2 diabetes susceptibility gene to be found by genome-wide linkage and positional cloning was calpain-10 (*CAPN10*) (Horikawa et al. 2000), a member of the calpain-like cysteine protease family. This study detected association between SNPs and haplotypes in *CAPN10* with type 2 diabetes in a Mexican American and a Finnish population. The same markers replicated in some studies (del Bosque-Plata et al. 2004; Garant et al. 2002; Kang et al. 2006; Kifagi et al. 2008) but not in others (Chen et al. 2005; Elbein et al. 2002; Hegele et al. 2001; Rasmussen et al. 2002; Tsai et al. 2001). A number of reasons have been suggested to explain the inconsistency between results. Many studies lacked statistical power to detect the initial association due to small sample sizes and/or low frequency of *CAPN10* SNPs and haplotypes. Also, genetic heterogeneity between populations of different

ethnic background may account for the differences in SNPs and haplotypes associated with type 2 diabetes. Nevertheless, three meta-analyses succeeded in replicating an association between *CAPN10* and type 2 diabetes (Song et al. 2004; Tsuchiya et al. 2006; Weedon et al. 2003) and functional work has demonstrated a role for *CAPN10* in insulin action in muscle and liver and secretion in pancreatic islets (Brown et al. 2007; Marshall et al. 2005; Meier et al. 2007; Turner et al. 2007). Furthermore, *CAPN10* SNPs have been shown to be associated with *CAPN10* expression. The association between *CAPN10* SNPs and type 2 diabetes is still not statistically robust to the degree shown for established susceptibility genes and *CAPN10* was not detected in genome-wide association studies. Further evidence is required for *CAPN10* to be accepted as a susceptibility gene illustrating the difficulty in establishing genetic associations with complex disease.

A notable success of linkage approaches was the identification of the transcription factor 7 like-2 gene (*TCF7L2*) by fine-mapping a suggestive linkage peak on chromosome 10q (Table 1.3) (Grant et al. 2006; Reynisdottir et al. 2003), though the variants found to be associated with type 2 diabetes risk did not appear to explain the original linkage signal. Variants in *TCF7L2* have since been robustly replicated in a number of additional populations and studies, including different ethnic groups (Chandak et al. 2007; Lewis et al. 2008; Miyake et al. 2008; Ng et al. 2008a). The function of TCF7L2 and how it related to diabetic phenotypes was not known when it was discovered to impact risk of type 2 diabetes, demonstrating the advantage of a hypothesis-free genome-wide linkage approach as opposed to those studies focused on candidate regions of the genome. However, recent data suggests that TCF7L2 is an important component of the WNT signalling pathway and that it might influence the proliferation of pancreatic β-cells and the production of the incretin hormone, GLP-1 (Jin and Liu 2008). Still, no associated alleles have been shown to have a

direct functional impact on type 2 diabetes showing how hard it is to find underlying causal variants and the requirement for functional work.

1.6.4 Association studies

The purpose of association analysis is to test the correlation between a particular allele, genotype or haplotype of a genetic marker, and trait variation in a sample of individuals. Risch and Merikangas demonstrated that association studies have greater power than linkage studies to detect common variants with smaller effects on disease risk (Risch and Merikangas 1996). Associations can arise under three circumstances. Either the finding is a false positive association, the genetic variant is a true functional variant that directly effects disease risk, or the genetic variant is in linkage disequilibrium (LD) with one or more true functional variants (Barroso 2005).

1.6.4.1 Linkage disequilibrium

LD is a phenomenon whereby alleles at different loci will be inherited together at a disproportionately high rate given their frequencies in the population. When a new mutation arises it forms part of a haplotype, that is, a combination of alleles at multiple loci on the same chromosome that are inherited together. The linkage disequilibrium between these alleles is eroded by recombination or cross-over between homologous chromosomes with each generation (Figure 1.4).

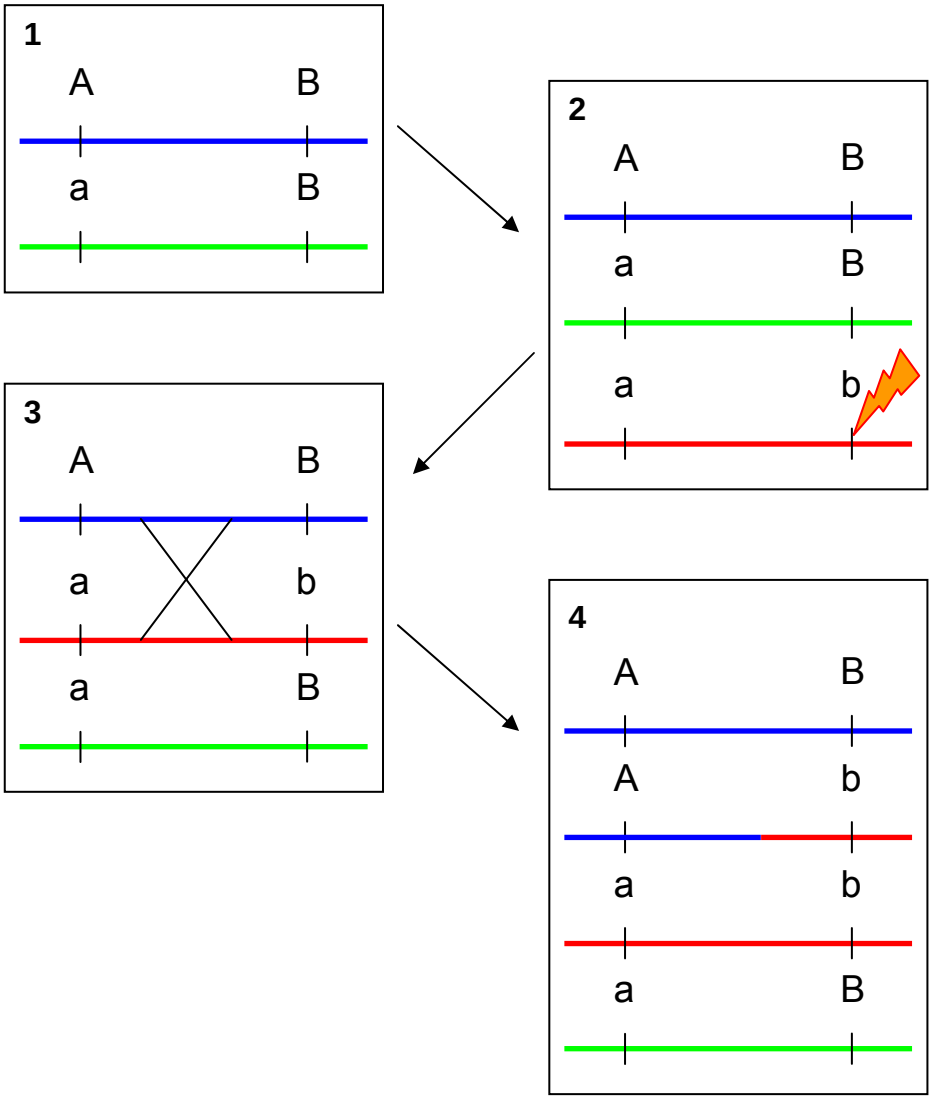


Figure 1.4 The erosion of LD by recombination, (adapted from (Ardlie et al. 2002)). In box 1 there is a polymorphic locus with alleles A and a. In box 2 a new mutation arises at a second locus changing allele B to b. This occurs on a chromosome carrying allele a, resulting in only three of four possible haplotypes in the population (or linkage disequilibrium between locus A/a and locus B/b). In box 3, crossovers between the two loci result in the presence of all four haplotypes in the population (as shown in box 4) and linkage disequilibrium declines as the frequency of the recombinant chromosome increases.

Correlation between alleles is expected to degrade with time (or number of generations between which recombination can take place) and the genetic distance (which is correlated with likelihood of recombination) between the alleles. For these reasons, LD is generally higher surrounding a locus that has arisen recently in evolution, and between loci in close proximity on a chromosome, as physical and genetic distance are related.

Linkage disequilibrium between two loci is quantified as *D*, *D'* or r^2. The maximum value of *D* depends on the allele frequencies at the loci in question. The maximum value of *D'* or r^2 is 1.0 representing perfect LD and meaning the information from both loci is completely redundant (knowing the genotype at one locus completely predicts the genotype at the other). Box 1 shows *D* for two alleles (A and B) at two different loci (P and Q respectively). It represents the difference between the observed haplotype frequency (X_{AB}) and the expected frequency if the alleles were segregating at random. Box 2 and Box 3 represent formulas for *D'* and r^2 respectively.

Box 1

$$D_{AB} = X_{AB} - P_A Q_B$$

Box 2

$$D' = \frac{D}{Dmax}$$

Box 3

$$r^2 = \frac{D^2}{P_A P_a Q_B Q_b}$$

Information about LD in a candidate region is useful for association studies because it allows investigators to select a subset of genetic markers that can predict the allelic status of other markers in high LD, without having to genotype these other markers themselves. These subsets of markers are referred to as “tagging” markers (Figure 1.5).

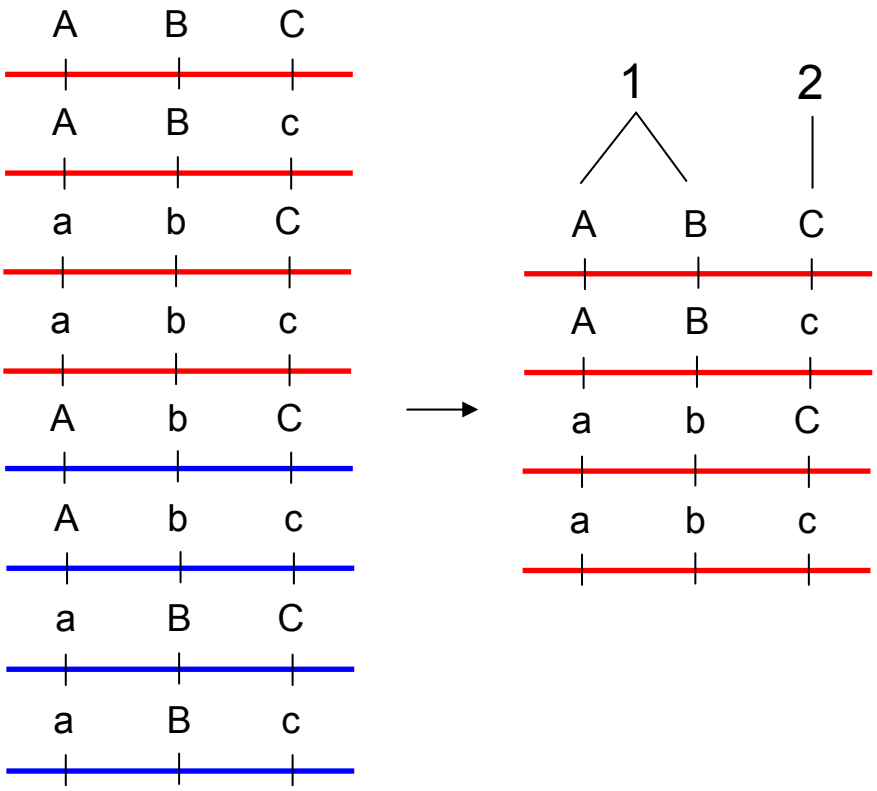


Figure 1.5 Selection of tagging markers based on pairwise correlation between three bi-allelic markers
To the left are all possible three-locus haplotypes of markers A/a, B/b, and C/c. However, only those in red are present at appreciable levels in the population. As A/a and B/b loci are in complete LD only one needs to be genotyped to predict the genotype at the other. Therefore, one requires two tagging SNPs to cover all three loci.

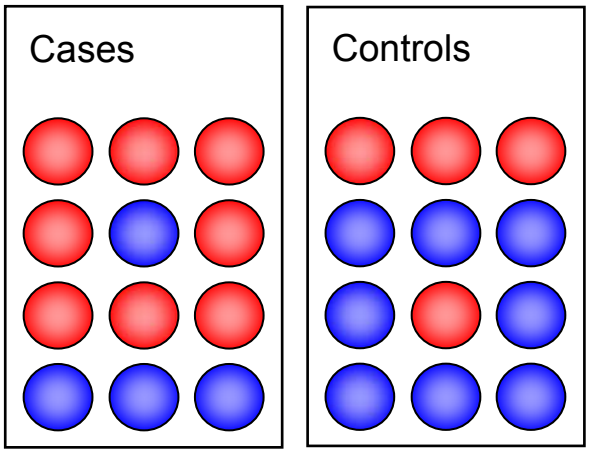
The genetic markers most commonly used for association studies are single nucleotide polymorphisms (SNPs) as millions of these have been catalogued and are easy to genotype. Patterns of LD among SNPs have been characterised in 270 DNA samples from four populations of European, African, and Asian (Chinese and Japanese) ancestry by the International HapMap Project (Sachidanandam et al. 2001), and found to be remarkably stable over different samples of individuals and between populations. These catalogues of SNPs and LD patterns enable researchers to select tagging SNPs and design association studies without the need to type every single SNP in the candidate gene or genome. Interestingly, HapMap and similar projects have noted the presence of blocks of LD separated by possible

hotspots of recombination (Daly et al. 2001). This means the density of markers required to cover a gene or genes will vary region to region.

1.6.4.2 Case-control studies

The traditional epidemiological case-control study was amongst the first approaches used to find susceptibility genes for type 2 diabetes. With this study design it is practical to recruit, phenotype and genotype the large numbers required to detect the modest effect sizes expected of genetic factors involved in complex disease (Borecki and Province 2008). Here, the frequency of a putative disease marker (or proxy (tag) SNP) is compared in individuals with type 2 diabetes (cases) and in unaffected individuals (controls) (Barroso 2005). Chi-squared statistics can then be employed to test the null hypothesis that there is no association between rows and columns of a 2 x 3 contingency table (Figure 1.6). Alternatively, a linear relationship between number of risk alleles and odds of disease can be tested using logistic regression, where the null hypothesis is no change (or a slope of 0) in odds of disease per test allele (Balding 2006). Indeed an additive model of gene action, where heterozygotes have a risk of disease intermediate between the two types of homozygote, is widely thought to apply in complex disease.

A



B

	AA	Aa	aa
Number of cases	**1**	**2**	**3**
Number of controls	**3**	**2**	**1**

C

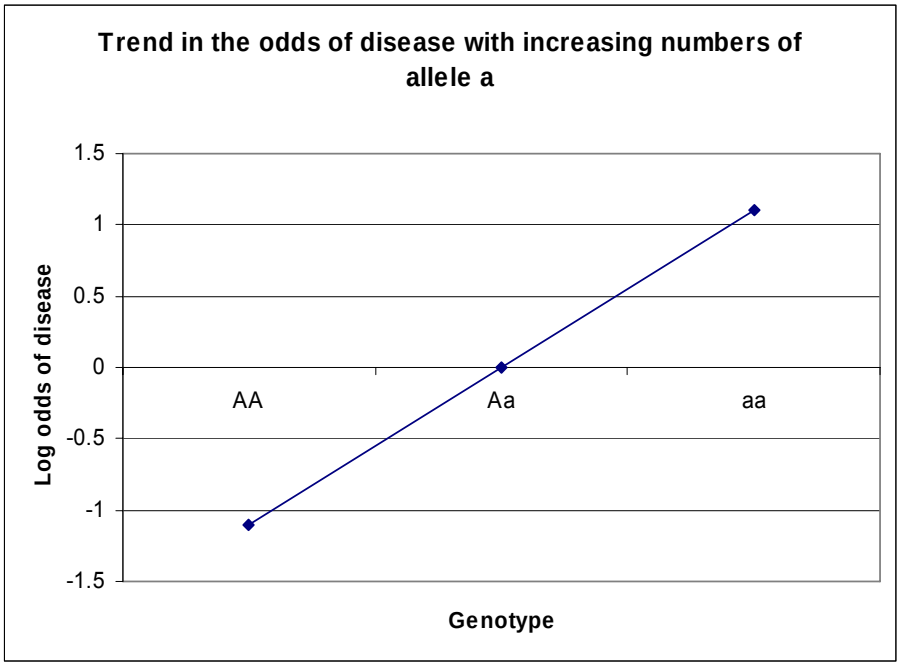


Figure 1.6 Association of a biallelic marker with complex disease, (adapted from (Balding 2006)). A, frequency of blue protective allele and red risk allele in cases and controls. B, 2 x 3 contingency table showing numbers of cases and controls in different genotypic groups. C, increase in odds of disease with increasing numbers of the red (a) risk allele.

In the early years of genetics association studies the literature became swamped with reports of genetic association with complex disease but, for several reasons, these proved difficult to replicate. Firstly, many published studies reported false positive associations. The shear number of early association studies meant that multiple hypotheses (SNPs/traits) were being tested. It is expected that 1/20 tests will show nominal statistical association ($P<0.05$) by chance. Failure to adjust for multiple testing led to over-interpretation of results, which was exacerbated by a publication bias towards those studies claiming statistical association (Cardon and Bell 2001). Secondly, due to the phenomenon of the "winner's curse" (described above), even reports of real associations often over-estimated the effect size of a given genetic factor on risk of disease. Replication studies (and indeed initial studies) were often performed on small sample sizes which were underpowered to detect the likely effect sizes of the alleles under investigation (Cardon and Bell 2001). Thirdly, some studies may have reported spurious associations due to poor sample selection – that is, significant differences in allele frequency between cases and controls may have been due to differences in other factors such as ethnicity, age and sex between cases and controls, rather than a real effect of the variant on disease risk (Cardon and Bell 2001). If an allele is at a particularly high frequency in one ethnic group, and people from this ethnic group are more common amongst the cases than controls, the allele will appear to associate with disease status even if it is neutral. This form of confounding is referred to as population stratification or population substructure (Figure 1.7).

Problems related to population stratification can be alleviated by careful selection of control samples to match cases in terms of ethnic origin, age, sex, and any known environmental factors that influence risk of disease (Cardon and Bell 2001). Finally, failure to replicate previously reported associations can reflect genetic and environmental differences between study populations. For example, patterns of LD

may differ between populations so that a tested variant is associated with a true disease susceptibility locus in one population but not in another.

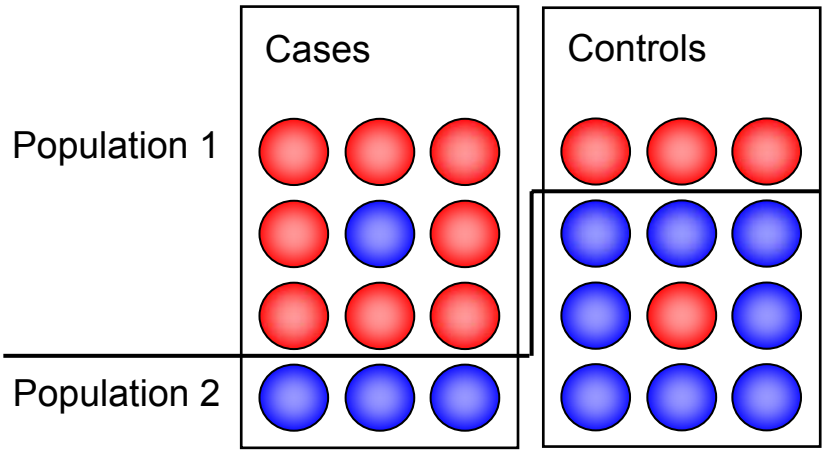


Figure 1.7 Spurious association of red alleles with disease due to population structure, (adapted from (Balding 2006). Population 1 has a higher frequency of red alleles compared to Population 2. As the frequency of people from population 1 is higher in cases than controls, red alleles are more prevalent in cases. This leads to a spurious association between the variant and disease risk.

1.6.4.3 T2D genes found by candidate gene association studies

Well-designed candidate gene association studies and meta-analyses to increase sample size and power to detect modest effects have succeeded in highlighting at least five true type 2 diabetes loci. This approach was used to fine-map associations under linkage peaks, as in *TCF7L2*, and to detect associations between type 2 diabetes risk and variation within biological candidate genes that encode known proteins in pathways influencing insulin action or secretion. For example, the common SNPs P12A in *PPARγ* (Altshuler et al. 2000) and E23K in *KCNJ11* (Gloyn et al. 2001; Gloyn et al. 2003; Hani et al. 1998; Nielsen et al. 2003) have been reproducibly associated with type diabetes risk in multiple populations (Table 1.3). More recently, a candidate gene association study investigating the impact of common variation in known MODY genes detected association between *TCF2* (*HNF1β*) and type 2 diabetes (Winckler et al. 2007). A few months later, I was involved in the detection of reproducible associations between SNPs in *WFS1* and

type 2 diabetes risk in a large candidate gene association study of genes involved in pancreatic β-cell function (Sandhu et al. 2007). Still, studies were very much limited by knowledge of biological pathways underlying type 2 diabetes and the ability to test only a small number of variants/genes.

1.6.5 Genome-wide association studies (GWAS)

Though candidate gene approaches identified some genes that impact on risk of type 2 diabetes, a definitive picture of the genetic contribution to disease remained elusive. During the course of my PhD it became feasible to test hundreds of thousands of SNPs across the entire genome for association with type 2 diabetes. The advantage of this study design is that, similar to whole-genome linkage scans, it is not biased by prior assumptions about the genes and pathways involved in disease, but retains the power of candidate gene association studies to detect modest effect sizes. Several break-throughs enabled this to happen. First, completion of the Human Genome Project in April 2003 (2004; Lander et al. 2001; Venter et al. 2001) provided a foundation for other technological advances necessary for GWAS. Second, an extensive catalogue of >10 million SNPs served as the basis for development of genome-wide LD maps in four populations described in public databases (International HapMap Project (Sachidanandam et al. 2001)). This meant that it was not necessary to genotype every SNP in the human genome as many would provide redundant information. Instead, a reduced number of SNPs could capture most of the genetic information of untested SNPs with only a slight loss of statistical power. The minimum correlation between tested and untested SNPs is usually set at $r^2 = 0.8$. Third, new genotyping technologies and the development of genome-wide genotyping arrays offering good coverage of HapMap Phase I and II data (2005; Frazer et al. 2007) by Illumina and Affymetrix reduced the cost and increased the speed of genotyping large numbers of SNPs (Gunderson et al. 2005; Kennedy et al. 2003; Steemers and Gunderson 2007). Fourth, the collection of large

sample sizes and the pooling of resources within large consortia improved power to detect modest genetic effects on complex disease risk, and to replicate initial findings and perform meta-analyses.

1.6.5.1 T2D genes found by GWAS

The first GWAS on T2D revealed four novel loci (Sladek et al. 2007), two of which (a non-synonymous SNP in *SLC30A8* and an intergenic SNP near *HHEX* and *IDE*) were successfully replicated by subsequent GWAS conducted by the Wellcome Trust Case Control Consortium (WTCCC) (2007), the Diabetes Genetics Initiative (DGI) (Saxena et al. 2007), the Finland-United States Investigation of Non-Insulin Dependent Diabetes Mellitus Genetics (FUSION) (Scott et al. 2007) and deCODE genetics (Steinthorsdottir et al. 2007) (Table 1.3). In addition, these studies uncovered 4 novel T2D susceptibility loci in or near the *CDKN2A/B*, *IGF2BP2*, *CDKAL1*, and *FTO* genes. Interestingly, *FTO* is only associated with diabetes through its effect on risk of obesity. This was noticed when the association of several SNPs in *FTO* with type 2 diabetes was abolished after adjustment for body mass index (BMI) in cases and controls (Frayling et al. 2007). For this reason, *FTO* was not detected in the DGI and FUSION studies which adjusted for BMI up front. These studies also replicated association of SNPs in *TCF7L2*, *PPARγ*, *KCNJ11*, and *WFS1* with type 2 diabetes. A genome-wide screen for variants involved in prostate cancer also confirmed the association between *TCF2* (*HNF1β*) and risk of type 2 diabetes (Gudmundsson et al. 2007). The combined sample size of the DGI, WTCCC, and FUSION studies was >10,000 enabling detection of variants conferring modest genetic effects on risk of type 2 diabetes (OR 1.10-1.20). To increase power still further, a meta-analysis of WTCCC, FUSION, and DGI genome-wide association data (including directly typed SNPs and those whose genotypes can be imputed from knowledge of correlation with typed SNPs) and replication in just under 54,000 independent samples yielded six more loci with genome-wide significance and very modest effects on disease risk (OR

1.09-1.15), *JAZF1*, *CDC123/CAMK1D*, *TSPAN8*, *THADA*, *ADAMTS9*, and *NOTCH2* (Table 1.3) (Zeggini et al. 2008). Another established type 2 diabetes risk locus is the potassium voltage-gated channel, KQT-like subfamily, member 1 (*KCNQ1*), first discovered by genome-wide association in Japanese cases and controls and replicated in a number of independent populations including those of East Asian and European ancestry (Unoki et al. 2008; Yasuda et al. 2008).

Between them these variants still explain only a small proportion of the heritable risk of type 2 diabetes (in a recent paper the combination of 18 type 2 diabetes loci only accounted for a sibling relative risk of 1.07, whereas the sibling relative risk for type 2 diabetes is ~3 (Lango et al. 2008)), indicating that more genetic loci are still to be uncovered. Furthermore, variants found to be associated with type 2 diabetes to date are not necessarily the true causal variants behind the association, but may be associated with disease by virtue of their correlation with the true causal variants. Therefore, true effect sizes of susceptibility loci may be larger than current estimations. Finally, genome-wide studies to date have been limited by the variants on genotyping arrays, which do not test rare or larger structural variants. These variations will be an intriguing subject of future work.

Table 1.3 Genomic regions associated with type 2 diabetes and the genetic study design used to discover them

Gene/Region	SNP	Consequence	Linkage	Candidate gene association	GWAS	Mendelian
PPARγ	rs1801282	P12A		✓	✓	✓
KCNJ11	rs5219	K23E		✓	✓	✓
TCF7L2	rs7903146	Intronic	✓	✓	✓	
WFS1	rs10010131	Intronic		✓	✓	✓
SLC30A8	rs13266634	R325W			✓	
HHEX/IDE	rs1111875	Intergenic			✓	
FTO	rs8050136	Intronic			✓	
CDKN2A/B	rs10811661	Intergenic			✓	
CDKN2A/B	rs564398	Downstream			✓	
CDKAL1	rs10946398	Intronic			✓	
IGF2BP2	rs4402960	Intronic			✓	
TCF2	rs757210	Intronic			✓	✓
JAZF1	rs864745	Intronic			✓	
CDC123/CAMK1D	rs12779790	Intergenic			✓	
TSPAN8/LGR5	rs7961581	Intronic			✓	
THADA	rs7578597	T1187A			✓	
ADAMTS9	rs4607103	Intergenic			✓	
NOTCH2/ADAM30	rs10923931	Intronic			✓	
KCNQ1	Various	Intronic			✓	

1.6.6 Intermediate phenotypes

Another approach to discovering susceptibility loci for complex disease is to study intermediate phenotypes that characterise the early disease process. Diabetes related traits are often continuous and include fasting and post-oral glucose tolerance test (OGTT) glucose, insulin, BMI and lipid levels. Linear regression is the most commonly used statistical test for assessing the contribution of genetic variants to quantitative traits. The null hypothesis is that there is no difference between the mean values of the trait between people in each genotypic class. These traits have been the subject of linkage and association approaches including genome-wide (Meigs et al. 2007) and have yielded a number of loci effecting variation in these traits – quantitative trait loci (QTLs). The *G6PC2* gene, encoding the glucose-6-phosphatase catalytic subunit-related protein which is predominantly expressed in pancreatic ß-cells, has been associated with fasting glucose levels and ß-cell function. However, there is no evidence for an impact on risk of type 2 diabetes (Bouatia-Naji et al. 2008; Chen et al. 2008). Polymorphisms in *GCK* have also been associated with fasting glucose (Weedon et al. 2006). More recently, MAGIC (Meta-Analyses of Glucose and Insulin-related traits Consortia) compared the top hits of meta-analyses of genome-wide association studies to find loci with consistent effects on fasting glucose across multiple studies. This analysis confirmed previous QTLs, *G6PC2* and *GCK*, and a new locus, *MTNR1B*, which was also associated with type 2 diabetes in a large meta-analysis of case-control studies (Prokopenko et al., in press). Though QTLs don't always overlap with type 2 diabetes risk loci, there is some suggestion that the robustly replicated genes might influence insulin and glucose traits (Grarup et al. 2008).

1.6.7 Monogenic insulin resistance and diabetes genes in common T2D

It is striking that of the 18 validated type 2 diabetes loci, at least four are involved in monogenic forms of insulin resistance and diabetes (Table 1.3). Mutations in the

ligand-binding domain of *PPARγ* cause FPLD3 characterised by partial lipodystrophy, insulin resistance and diabetes (Barroso et al. 1999), whereas common polymorphisms in this gene have been shown to influence risk of T2D (Altshuler et al. 2000; Barroso et al. 1999). Similarly, mutations in *KCNJ11* have been shown to cause persistent hyperinsulinaemia hypoglycaemia of infancy (PHHI), and transient and permanent neonatal diabetes, while polymorphisms increase susceptibility to T2D (Gloyn et al. 2001; Gloyn et al. 2004; Gloyn et al. 2003; Hani et al. 1998; Nielsen et al. 2003; Thomas et al. 1996). Common SNPs in *TCF2* (*HNF1β*) impact T2D risk whereas mutations are responsible for MODY5 (Gudmundsson et al. 2007; Horikawa et al. 1997). Hundreds of rare variants in *WFS1* have been shown to cause Wolfram Syndrome, while common SNPs associate with T2D (Inoue et al. 1998; Sandhu et al. 2007; Strom et al. 1998). This suggests that, in some cases, the same loci may be involved in both rare and common forms of disease, though the extent to which this is true remains to be elucidated. Genes responsible for rare, Mendelian forms of diabetes may be good candidate genes for more common complex forms of the disease and *vice-versa*.

1.7 Aims

The aim of this thesis was to contribute to our understanding of the genetic aetiology of syndromes of severe insulin resistance, common type 2 diabetes and related traits using a candidate gene resequencing and association approach. When this work began whole-genome association studies were not feasible and therefore I pursued candidate genes selected on the basis of their putative role in biological pathways impacting insulin action and/or secretion, and data from animal models and human phenotypes. My work focused on a number of different genes and pathways with the aim of:

1. Investigating the role of Lipin gene family members in insulin resistance syndromes and traits underlying type 2 diabetes (Chapter 3);
2. Investigating the role of components of the mTOR pathway in insulin resistance syndromes (Chapter 4);
3. Assessing the impact of a *PARL* (presenilins associated rhomboid-like) polymorphism on fasting insulin levels and BMI in UK populations (Chapter 5);
4. Performing a large scale candidate gene study to test association of 84 genes with putative roles in the pancreatic ß-cell with type 2 diabetes (Chapter 6);
5. Attempting to refine the association between *WFS1* and type 2 diabetes and assess the contribution of rare variants in *WFS1* to risk of type 2 diabetes (Chapter 7).

Chapter 2

Materials and Methods

2.1 Definition of cohorts

2.1.1 Severe Insulin Resistance (SIR) cohort

The severe insulin resistance (SIR) cohort was established by Professor Stephen O'Rahilly to identify genes that influence insulin resistance independently of obesity. All patients have severe insulin resistance, defined as fasting insulin above 150 pmol/l, or peak insulin on oral glucose tolerance testing above 1,500 pmol/l in non-diabetic patients. In complete insulin deficiency it was defined as an insulin requirement above 3U/kg/day. Most patients had a BMI <30 kg/m^2 and at least 58 had BMI>30. Those with partial beta cell decompensation and clinical features including acanthosis nigricans, and those with BMI >30 kg/m^2 were included at Professor O'Rahilly's discretion. All patients gave informed consent with approval of the local research ethics committee in Cambridge, U.K. and details (sex, year of birth, ethnic origin and possible clinical diagnosis) are given in the Appendix Table A1.

2.1.2 Control panels

2.1.2.1 CEPH

The Centre d'Etude du Polymorphisme Humain (CEPH) is a nonprofit research institute that makes available cellular DNA from cultured lymphoblastoid cells lines (LCLs) derived from each member of a reference panel of large nuclear families/pedigrees. These are white families from Utah, France, Venezuela and Pennsylvania. At the Wellcome Trust Sanger Institute we use a subset of 48 unrelated individuals from the CEPH families supplied by Coriell Cell Repositories (Dausset et al. 1990). These are control individuals of North and West European origin and are listed in the Appendix Table A2. Also highlighted are those 31 samples which overlap with the HapMap CEU trios.

2.1.2.2 HGDP-CEPH Human Genome Diversity Cell Line Panel

The HGDP-CEPH Human Genome Diversity Cell Line Panel is a resource of 1064 LCLs of individuals from 51 different world populations to provide DNA for studies of sequence diversity and the history of modern human populations. Corresponding milligram quantities of DNA for each cell line is deposited at the Foundation Jean Dausset in Paris. All samples used for this resource were collected with informed consent (Cann et al. 2002). Sample IDs, population and geographic origin are provided in the Appendix Table A3.

2.1.2.3 European-Indian control panel

This panel includes DNA samples from 47 white European individuals and 47 individuals of Asian Indian origin (Appendix Table A4). Samples were ordered from the European Collection of Cell Cultures (ECACC).

2.1.3 Case-control populations

2.1.3.1 Cambridgeshire case-control

This is a population based case-control study in which a total of 552 patients aged 45–76 years with T2D were randomly selected from general practitioner diabetes registers in Cambridgeshire, UK (Rathmann et al. 2001). Presence of T2D was based on clinical criteria; onset of diabetes after the age of 30 years without treatment with insulin in the first year after diagnosis. The controls were recruited at random from the same population sampling frames, and were individually matched to cases for age, sex and GP practice. Diabetes was excluded in controls by medical record search and by a glycated haemoglobin measurement of less than 6%. The study received ethical approval from the Cambridge Local Research Ethics Committee, and participants provided informed consent.

2.1.3.2 EPIC- Norfolk case-control study

The EPIC case-control study is nested within the EPIC - Norfolk Study, a population based cohort study of European men and women aged 40-78 years. Both the case-control (Harding et al. 2004) and full cohort (Day et al. 1999) study have been previously described in detail. Briefly, the case-control study consists of 417 incident type 2 diabetes cases and two sets of 417 controls, each matched in terms of age, sex, general practice, recruitment date, with one set additionally matched for BMI. A case was defined by a physician's diagnosis of type 2 diabetes, with no insulin prescribed within the first year following diagnosis, and/or HbA_{1c} > 7% at the health check. Controls were randomly selected from the EPIC-Norfolk cohort from among those without diabetes, cancer, stroke, or myocardial infarction at baseline and who had not developed diabetes by the time of selection. Potential controls with measured HbA_{1c} levels > 6% were excluded. The EPIC-Norfolk study was approved by the Norfolk Local Research Ethics Committtee.

2.1.3.3 Exeter case-control study

The diabetic subjects from Exeter came from two sources (i) a consecutive-case series of patients with T2D diagnosed before 45 years from North and East Devon (Owen et al. 2003). The patients were unrelated and recruited through questionnaires distributed though general practitioners (97% agreed to send out questionnaires, >70% return rate and >90% recruitment of those identified through the questionnaires). Validation of the diagnosis of diabetes was based on either current prescribed treatment with sulphonylureas, biguanides and/or insulin, or, in the case of individuals treated with diet alone, historical or contemporary laboratory evidence of hyperglycaemia (as defined by present WHO guidelines). All patients were off insulin for at least 1 year after diagnosis, and patients were excluded if they had pancreatic autoantibodies (GAD), first degree history of type 1 diabetes or clinical features (or DNA test results) suggestive of monogenic diabetes (Owen et al. 2003). (ii) Probands

from a collection of type 2 diabetes families that had either both parents available, or one parent and at least two siblings (Frayling et al. 1999). Only subjects collected in Exeter were used in this study. The sex matched controls are taken from the parents in the Exeter Family Study, a cohort study of newly born babies and both their parents (Knight et al. 2006). This study recruits from central Exeter so the controls come from a similar geographical region as the cases. Diabetes and hyperglycaemia were excluded by measuring fasting glucose and HbA1c. In total 601 cases and 610 controls were included in this study. Informed consent was obtained from all participants.

2.1.3.4 ADDITION case-control study

Cases were participants from the UK Cambridge arm of the ADDITION trial, which aims to evaluate whether screening for prevalent undiagnosed Type 2 diabetes is feasible, and whether subsequent optimised intensive treatment of diabetes is feasible and beneficial (Lauritzen et al. 2000). All cases were aged 40-69 and screen detected using OGTT and WHO diagnostic criteria. We used participants from the Ely study as controls—a population of white European men and women aged 35 to 79 years without diagnosed diabetes and from a similar population sampling frame as the Cambridge arm of the ADDITION study. The Medical Research Council (MRC) Ely Study (Wareham et al. 1998) is described in more detail below. All ELY participants were defined as cases or controls based on their OGTT (WHO diagnostic criteria). For this analysis, the ADDITION case-control study comprised 926 cases and 1497 controls. The Cambridge Research Ethics Committee approved both studies.

2.1.4 Population-based populations

2.1.4.1 MRC Ely study

The Medical Research Council (MRC) Ely Study is a prospective population-based cohort study of the aetiology and pathogenesis of type 2 diabetes and related metabolic disorders in the UK. Between 1990 and 1992 1122 white subjects of European ancestry aged 40-65 years were selected at random from a sampling frame of all adults without known diabetes and registered with a single general practice in the City of Ely (response rate 74%). Volunteers attended a clinical examination that comprised standard anthropometric tests, an oral glucose tolerance test (OGTT) that measured insulin and glucose concentrations at fasting, 30, 60, and 120 minutes, and a dietary and medical questionnaire (Ekelund et al. 2007; Wareham et al. 1999; Williams et al. 1995). Nine hundred and thirty seven of 1071 non-diabetic volunteers attended a re-screening at a mean follow-up time of 4.44 years, along with 183 newly recruited adults aged 30-40 years (phase 2). Between 2000 and 2004 participants were approached for a third time to attend a clinical examination along with 716 newly recruited volunteers from the original sampling frame. The phase 3 cohort analysed in my studies comprised 1721 men and women aged 35-79. Informed consent was obtained from all participants and ethical approval for the study was granted by the Cambridge Local Research Ethics Committee.

2.1.4.2 Hertfordshire cohort study

The Hertfordshire Cohort Study was established by David Barker and the Hertfordshire Cohort Study Group at the University of Southhampton and is now under the leadership of Cyrus Cooper at the MRC Epidemiology Resource Centre, University of Southhampton. The objective of the study is to evaluate interactions between the genome, intrauterine and early postnatal environment, and adult diet and lifestyle in the aetiology of chronic disorders. The cohort comprises men and women recruited from 7106 people born in Hertfordshire between 1931 and 1939 and

still alive and registered with a Hertfordshire GP in 1998. Permission to contact 6099 men and women by letter was obtained from their GPs, and 3225 of these agreed to a home interview with a trained nurse. Subsequently 2997 attended a clinic for detailed physiological investigations. The cohort details and measurements of metabolic traits analysed in my studies have been described previously (Syddall et al. 2005). Informed consent was obtained from all participants and ethical approval was granted by the Hertfordshire and Bedfordshire Local Research Ethics Committee.

2.2 Reagents

2.2.1 DNA preparation

20 X TE

200 mM Tris-HCL pH7.5

20 mM EDTA

2.2.2 Polymerase Chain Reaction (PCR) and sequencing

Loading buffer

50% glycerol	100 ml
5XTBE (Severn Biotech Limited)	20 ml
ddH_2O	80 ml
Bromophenol blue	2 mg

Reaction buffer

200 mM Tris-HCL pH8.0

100 mM $MgCl_2$

Dilution buffer

50 mM Tris-HCL pH8.0

Sequencing mix

BigDye (v3.1, Applied Biosystems, Foster City, CA, USA)	10 ml
Sanger BigDye reaction buffer	112.5 ml
ddH_2O	37.5 ml
dGTP BigDye (v3.0, Applied Biosystems)	3.2 ml

Precipitation mix

99.7-100% ethanol	771 ml
3M sodium acetate	16 ml
ddH_2O	189 ml

2.2.3 Sequenom reagents

10 X PCR Buffer

10 mM Tris-HCl

500 mM KCl

15 mM $MgCl_2$

0.01 % (w/v) gelatine

10 X Thermosequenase buffer

260 mM Tris-HCl pH: 9.5

65 mM $MgCl_2$

2.3 Protocols

2.3.1 DNA preparation

2.3.1.1 Whole-genome amplification

Genomic DNA from study participants was received from collaborators and randomly preamplified using the Genomiphi HY DNA Amplification kit (GE Healthcare UK, Chalfont St. Giles, UK). 2.5 µl of 4 ng/ µl genomic DNA was denatured in 22.5 µl

sample buffer (GE Healthcare) at 95°C for 3 minutes and then cooled on ice. 22.5 µl reaction buffer (GE Healthcare) and 2.5 µl enzyme mix (GE Healthcare) were then combined on ice and added to the cooled sample. The amplification reaction took place at 30°C for 6 hours, then the DNA polymerase was inactivated by heating at 65°C for 10 minutes. Finally samples were cooled to 4°C and stored at -20°C. Success of the reaction was confirmed by running a 1 µl of a 1:2 dilution of product on 0.75% agarose gel for approx. 30minutes at 200 watts and using Hind III digested lambda DNA ladder (see section 2.3.3 for details of gel electrophoresis). Samples were diluted in TE buffer to 250 ng/µl and 100 ng/µl stocks and stored at -80°C, while a working stock diluted to 20 ng/µl was stored at -20°C.

2.3.1.2 Quantification of DNA using Pico-Green

The amount of DNA produced by whole-genome amplification was measured using an Invitrogen Picogreen® dsDNA quantitation assay kit. A dilution series of Sigma Calf Thymus DNA was made using 1 X TE for calibration purposes, and 10 ng/µl, 5 ng/µl, and 2.5 ng/µl stocks were made as controls. Whole-genome amplified test DNA was diluted 1:50 prior to testing. 5 µl of DNA was added to a well of a black 96 well Greiner bio-one plate with 45 µl of 1 X TE and 50 µl of Picomix (made up using 0.26 µl Picogreen and 51.82 µl TE per reaction). Each sample, standard and control was done in duplicate on the plate and at least two wells were left blank as a negative control. The plates were sealed and vortexed gently, spun down and read on a DTX 800/880 Series Multimode Detector (Beckman Coulter). The Calf Thymus DNA readings were used to create a standard curve and calculate the DNA concentration of test samples and controls. If any standards failed the plate was repeated. If the concentration of test sample duplicates were not 10% of each other they were repeated.

2.3.2 PCR

Primers were designed using Primer3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) to cover all coding exons, splice junctions, 3'UTR, and other regions of interest (see Appendix for primer sequences and conditions). For *LPIN1*, PCRs were made up to a volume of 15 µl with H_20 and included 0.3 µl of primer mix containing 8.5 µM forward and 8.5 µM reverse primer diluted in water, 7.5 ng DNA, 1.5 µl 10X Thermo-Start® PCR buffer (ABgene), 0.56 µl of 25 mM $MgCl_2$, 1.5 µl of dNTP mix (each 1.25 mM), and 0.09 µl of 5U/µl Thermo-Start Taq polymerase (ABgene). Plates were heat-sealed, spun down, and run on MJ Thermocyclers. PCR conditions were:

1) 95°C denaturation for 15 minutes
2) 95°C for 30 seconds
3) 60°C (unless specified otherwise in Appendix 2) for 30 seconds
4) 72°C for 30 seconds,
5) Go to step 2 39 more times
6) Final extension at 72°C for 10 minutes

For *LPIN2*, *LPIN3*, *mTOR*, *Rictor*, *Raptor*, *GβL*, *MAPKAP1*, *AS160*, and *WFS1*, the PCR protocol is the same except that the 15 µl reactions contained 1.5 µl of 10X Thermo-Start® PCR buffer containing $MgCl_2$ (ABgene) and no separate addition of $MgCl_2$. Successful PCR for GC-rich amplicons often required the addition of deaza-dGTP (Roche) to the dNTP mix (final concentration = 0.5 mM). The amplicons requiring deaza-dGTP are indicated in the Appendix.

2.3.3 Gel electrophoresis

Products of the expected length were confirmed by gel electrophoresis which involved running 2µl of PCR product mixed with 3 µl loading buffer on 2.5% agarose (standard melting temperature, electrophoresis-grade from Invitrogen) gels made in

1XTris Borate EDTA (TBE) buffer (Severn Biotech Ltd, Worcestershire, UK) with ethidium bromide (10 mg/ml, Sigma-Aldrich). A 100 bp ladder (Hyperladder IV, Bioline) was also loaded onto the gel as a marker. DNA was visualised using a UV transilluminator and digitally photographed using LabWorks Image Acquisition and Analysis Software (UVP Bioimaging Systems).

2.3.4 DNA purification

PCR products were purified using 0.66 µl reaction buffer, 0.66 µl dilution buffer, 0.066 µl of 10U/µl exonuclease I and 0.66 µl of 1U/ul shrimp alkaline phosphatase (USB Corporation, Cleveland, OH, USA) by incubating at 37°C for 1 hour and 80°C for 15 minutes. In the case of *LPIN2*, *LPIN3*, *mTOR*, *Rictor*, *Raptor*, *GβL*, *MAPKAP1*, and *AS160* water was used instead of dilution buffer and samples were incubated at 80°C for 30 minute rather than 1 hour.

2.3.5 Sequencing

Bi-directional sequencing was performed using a DNA sequencing kit (Big Dye Terminator 3.1; Applied Biosystems). Each 5 µl sequencing reaction comprises 2 µl 15 ng/µl primer, 1 ul PCR product (diluted ½ in water) and 2 µl sequencing mix. The plates were heat sealed, centrifuged briefly to bring the contents of each well to the bottom of the well and placed on the MJ Thermocyclers. Cycling conditions were:

1) 96°C for 30 seconds
2) 92°C for 8 seconds
3) 50°C for 8 seconds
4) 60°C for 2 minutes
5) Go to step 2 44 more times

30 µl sequencing precipitation mix was added to each reaction and plates were centrifuged at 4000 rpm for 30 minutes at 4°C. Plates were inverted and liquid spun off onto filter pads at 400rpm for 30 seconds, then 30 µl 80% was added to each well.

Plates were spun at 4000rpm for 5 minutes, air-dried, and loaded onto ABI3730 capillary machines (Applied Biosystems).

2.3.6 Sequence analysis

LPIN1 sequences were analysed using Mutation Surveyor version.2.20 (SoftGenetics LLC, State College, PA, USA). All other genes were analysed by an automatic SNP caller, ExoTrace, developed at the Wellcome Trust Sanger Institute (Leonard, unpublished) and the results of SNP calling were displayed and manually reviewed and confirmed in a specific implementation of GAP4 (Staden Sequence Analysis Package software). SIFT (http://blocks.fhcrc.org/sift/SIFT.html), PolyPhen (http://genetics.bwh.harvard.edu/pph/) and PANTHER (http://www.pantherdb.org/tools/csnpScoreForm.jsp) were used to predict the functional impact of non-synonymous mutations. Multiple sequence alignments were performed using ClustalW (http://www.ebi.ac.uk/Tools/clustalw2/index.html).

2.3.7 Genotyping

2.3.7.1 Sequenom

2.3.7.1.1 MassArray Homogeneous MassEXTEND (hME) assay

Primers and probes were designed using Extend Primer Assay Design v3.0.1.1 (see Appendix 1 for sequences). The initial PCR mix per reaction comprised 0.75 µl 10X PCR buffer, 0.2 µl dNTP mix (5 mM each dNTP in H_20), 0.04 µl Titanium Taq polymerase (BD Biosciences - Clontech), 2 µl primer mix (each primer 375 nM), 2 µl (3.5 ng) DNA, and 0.01 µl H_20. PCR conditions were:

1) 95°C for 15 minutes
2) 95°C for 20 seconds
3) 56°C for 30 seconds
4) 72°C for 1 minute
5) Go to step 2 44 more times

6) 72°C for 3 minutes

Plates were centrifuged at 1000 rpm for 1 min. PCR products were purified with 0.2 µl 10x Thermosequenase buffer, 0.3 µl 1 U/µl Shrimp alkaline phosphatase (Amersham Biosciences), and 1.5 µl H_20 at 37°C for 20 minutes and then heat inactivated at 80°C for 5 minutes. Plates were centrifuged at 1000 rpm for 1 min. 2 µl hME reaction cocktail was added to the cleaned PCR product: 0.2 µl 10x Thermosequenase buffer, 0.018 µl 32 U/µl Thermosequanase DNA polymerase (Amersham Biosciences), 0.5 µl Extend-primer mix (10 µM each primer), 0.9 µl Stop-mix (500 µM each appropriate ddNTP or dNTP) and 0.382 µl H_20. The hME program comprised:

1) 94°C for 2 minutes
2) 94°C for 5 seconds
3) 52°C for 5 seconds
4) 72°C for 5 seconds
5) Go to step 2 54 more times

Plates were centrifuged at 1000 rpm for 1 min. 16 µl of water and 3 mg Clean Resin (Sequenom) were added and the plates sealed and rotated for 10 minutes., then centrifuged at 4000rpm for 4 minutes. Reactions were spotted using a SpectroPoint nanoliter sample dispensing instrument (Sequenom) onto SpectroCHIPs (Sequenom) and these were analysed by MALDI-TOF mass spectrometry. Automatic analysis of genotype clusters by Sequenom MassARRAY™ Typer 3.0.1 software was manually checked.

2.3.7.1.2 MassArray iPLEX assay

Primers and probes for the Sequenom MassArray iPLEX platform were designed using Extend Primer Assay Design v3.0.2.0 (see Appendix 1 for sequences). The

PCR mix comprised 2 µl of 5 ng/µl DNA, 0.75 µl Molecular Biology Grade water, 0.625 µl 10 X PCR buffer (Qiagen), 0.325 µl 25 mM $MgCl_2$, 0.1 µl 25 mM dNTPs, 1 µl primer mix (containing equal amounts of each sense and anti-sense primer in water), and 0.2 µl HotStar Taq (5U/µl) per reaction. The PCR cycling conditions were the same as hME above. Plates were centrifuged at 1000 rpm for 1 min. To purify the PCR product, 2 µl of a SAP cocktail comprising 1.53 µl HPLC water, 0.17 µl 10 X SAP buffer, and 0.3 µl SAP enzyme (1.7U/µl) was added to each well and plates were incubated at 37°C for 40 minutes followed by heat inactivation at 85°C for 5 minutes. Plates were centrifuged at 1000 rpm for 1 min. 2 µl of iPLEX cocktail was then added to each well. This comprised 0.619 µl HPLC water, 0.2 µl 10 X iPLEX Buffer Plus, 0.2 µl iPLEX Termination Mix, 0.94 µl extend primer mix (the concentration of each extension primer is related to its mass) and 0.041 iPLEX enzyme. The iPLEX extension programme was:

1) 94°C for 30 seconds
2) 94°C for 5 seconds
3) 52°C for 5 seconds
4) 80°C for 5 seconds
5) Steps 3-4 were repeated 5 times
6) Steps 2-5 were repeated 40 times
7) 72°C for 3 minutes

Plates were centrifuged at 1000 rpm for 1 min. 16 µl of water and 6 mg Clean Resin (Sequenom) were added and the plates sealed and rotated for 30 minutes, then centrifuged at 4000rpm for 6 minutes. Reactions were spotted using a SpectroPoint onto SpectroCHIP® Bioarrays and these were analysed by MALDI-TOF mass spectrometry using MassARRAY Workstation version 3.4 software. Automatic analysis of genotype clusters by Sequenom MassARRAY™ Typer 3.4.0 software was manually checked.

2.3.7.2 Taqman

Taqman® MGB chemistry (Applied Biosystems, Foster City, CA) was performed following the recommended protocol (Ranade et al. 2001). Primers and probes were designed by the Custom Taqman® Genomic Assays service based on requests submitted using File Builder v3.1 software. The assay was pre-screened on the HapMap CEU panel to determine the optimum DNA concentration and annealing temperature. 5 µl volume reactions comprising 0.06 µl primers/probes mix, 2.5 µl Taqman® Universal PCR mastermix, 0.44 µl water and 2 µl DNA (at 1 ng/µl or 4 ng/µl depending on optimisation) were prepared in 384-well clear PCR plates (Eppendorf). Cycling conditions were as follows:

1) 95°C for 10 minutes
2) 95°C for 10 seconds
3) 55/60°C for 1 minute
4) Go to step 2 39 more times

Plates were then spun down at 1000 rpm for a few seconds and loaded onto a 7900HT Fast Real-Time PCR system (Applied Biosystems). The SDS software (Applied Biosystems) was used for SNP calling and data visualisation.

2.3.8 Statistical analysis

Deviation of SNP genotypes from Hardy-Weinberg equilibrium was assessed using a goodness-of-fit χ^2 test. Linear regression analysis was used to assess the association between individual SNPs and quantitative traits using Stata v9 (Stata Corporation, Texas, USA). In case-control analyses, logistic regression in Stata was used to test for association between SNPs and risk of disease (where controls were coded as 0 and cases as 1). Chi-squared analysis was performed to test for significant differences ($P<0.01$) in call rate between cases and controls. For quality control measures, imputation, and permutation testing see results chapters.

Chapter 3

Investigation of lipin family genes for impact on syndromes of severe insulin resistance and metabolic traits in UK populations

3.1 Summary

Loss of Lpin1 activity causes lipodystrophy and insulin resistance in the fld (fatty liver dystrophy) mouse, and *LPIN1* expression and common genetic variation were recently suggested to influence adiposity and insulin sensitivity in humans. I conducted a comprehensive association study to clarify the influence of *LPIN1* common variation on adiposity, insulin sensitivity, and other metabolic traits previously associated with *LPIN1*, in UK populations. Twenty-two SNPs tagging *LPIN1* common variation were genotyped in the MRC Ely (N = 1709) and Hertfordshire (N = 2901) population-based cohorts. Where possible, data were meta-analysed with other in-house and publicly available datasets to increase power to detect modest effect sizes. No association was found between *LPIN1* SNPs and fasting insulin, but I report a nominal association between rs13412852 and BMI (P = 0.042) in a meta-analysis of 8504 samples. I also detected nominal associations between *LPIN1* SNPs and traits underlying metabolic syndrome, but these require replication in additional large cohorts.

To investigate the putative role of lipin family mutations in insulin resistance syndromes I sequenced *LPIN1*, *LPIN2*, and *LPIN3* exons, exon/intron boundaries and 3'UTR in 158 patients with idiopathic severe insulin resistance (including 23 lipodystrophic patients), and controls. Three rare nonsynonymous *LPIN1* variants (A353T, R552K and G582R) were detected but these did not co-segregate with disease in affected families and Lipin1 protein expression and phosphorylation in patient fibroblasts was indistinguishable from controls. Two rare nonsynonymous changes in *LPIN2* were predicted benign and not prioritised for further analysis. Two rare nonsynonymous *LPIN3* variants (G41S and W110C) were detected within the conserved N-terminal lipin domain in single individuals. W110C was also detected in a Druze control and was therefore considered unlikely to be pathogenic. G41S was

absent from controls but no family DNA was available for co-segregation analysis. Functional work will be required to evaluate pathogenicity.

In summary, large scale association and re-sequencing studies do not support a major effect of *LPIN1* common variation on metabolic traits and suggest that mutations in lipin family genes are not a common cause of lipodystrophy and insulin resistance in humans. *LPIN1* data was published (Fawcett et al. 2008).

3.2 Introduction

3.2.1 Lipin 1 null mutations cause lipodystrophy and insulin resistance in fatty liver dystrophy (fld) mice

Lipin 1 was identified by positional cloning as the gene responsible for two independent mutant mouse models, the fld and fld^{2J} models, both characterised by a triglyceride-filled fatty liver, lipodystrophy, insulin resistance, and progressive peripheral neuropathy (Langner et al. 1989; Peterfy et al. 2001; Reue et al. 2000). The fld mouse strain carries two copies of Lpin1 with gross structural abnormalities, while the phenotype of the fld^{2J} mouse strain results from a point mutation leading to substitution of arginine for glycine at residue 84 of the lipin 1 protein (Peterfy et al. 2001). The phenotype of these mouse models shares features with human lipodystrophies, and therefore *LPIN1* is a good candidate gene for human lipodystrophies. However, to my knowledge there has only been one study screening *LPIN1* for pathogenic mutations in human lipodystrophic patients (N=15), with no pathogenic mutation being reported (Cao and Hegele 2002).

3.2.2 Lipin 1 is required for the development of mature adipocytes

Using primary mouse embryonic fibroblasts (MEFs) from fld mice it was shown that Lpin1 is required for induction of adipogenic genes, peroxisome proliferator-activated receptor-γ (*PPARγ*) and CCAAT enhancer-binding protein-α (*C/EBPα*), and for adipocyte differentiation (Phan et al. 2004). The primary defect in fld mice is therefore likely to be reduced adipose tissue mass with ectopic deposition of lipids and/or aberrant adipokine signalling causing secondary characteristics such as insulin resistance. Lpin1 expression is induced at two time points during differentiation of the 3T3-L1 preadipocyte cell line. There is a transient spike at 10 hours into the differentiation process, which precedes induction of *PPARγ* expression at 20 hours. Lipin 1 levels then return to baseline and are induced at 2 days,

reaching a peak in mature lipid-loaded adipocytes (Peterfy et al. 2005; Phan et al. 2004).

3.2.3 Lipin 1 isoforms reveal distinct roles during adipocyte development

Two lipin 1 protein isoforms are generated by alternative splicing of Lpin1 mRNA (Peterfy et al. 2005). Lipin-1A levels diminish during the differentiation process whereas Lipin-1B, which includes 33 extra amino acids, is the predominant isoform during the transient spike at 10 hours and in mature adipocytes. Reintroduction of Lipin-1A to lipin-1-deficient MEFs induces adipogenic genes, whereas Lipin-1B expression leads to stronger induction of lipid synthesis and storage genes compared to Lipin-1A (Peterfy et al. 2005). The biphasic expression of lipin 1 and the different functions of its two isoforms suggest that lipin 1 has two distinct roles in adipocyte development: the induction of the adipogenic gene expression program, and lipid accumulation in mature adipocytes.

3.2.4 Lpin1 overexpression causes obesity in transgenic mice

Transgenic mice with adipose tissue-specific overexpression of Lpin1B exhibit diet-induced obesity and enhanced insulin sensitivity compared to wild-type littermates (Phan and Reue 2005). In these mice adipose tissue expression of lipid synthesis and storage genes diacylglycerol acyltransferase (*DGAT*), acetyl-CoA carboxylase-1 (*ACC-1*), and phospoenolpyruvate (*PEPCK*) is elevated, supporting a role for Lpin1 in lipid accumulation of mature adipocytes. Interestingly, Lpin1 appears to have distinct roles in different tissues as skeletal muscle-specific overexpression of Lpin1 results in more pronounced obesity, insulin resistance, and changes in whole-body energy expenditure and fuel utilisation (Phan and Reue 2005; Xu et al. 2006). This occurs with decreased expression of fatty acid oxidation genes such as carnitine palmitoyl transferase 1 (*CPT-1*) and acyl-CoA oxidase (*AOX*) in skeletal muscle.

3.2.5 Lipin 1 is a phospatidate phosphatase

The mechanism through which lipin 1 influences adipocyte development and fat accumulation is not entirely known. However, recent data shows that lipin 1 is a magnesium-dependent phospatidate phosphatase (PAP) responsible for catalysing the conversion of phosphatidate (PA) to DAG, the penultimate step in triacylglyceride synthesis (Figure 3.1). This could explain why Lpin1 deficient fld mice cannot accumulate fat in their limited number of mature adipocytes (Han et al. 2006; O'Hara et al. 2006). This is also the penultimate step in the synthesis of phospholipids, important components of cell membranes, which might explain how the budding yeast homolog of lipin 1 (Smp2) regulates nuclear membrane growth during the cell cycle by controlling phospholipid biosynthesis (Santos-Rosa et al. 2005) and mutation of the fission yeast homolog of lipin 1 (Ned1) causes aberrant nuclear shape (Tange et al. 2002).

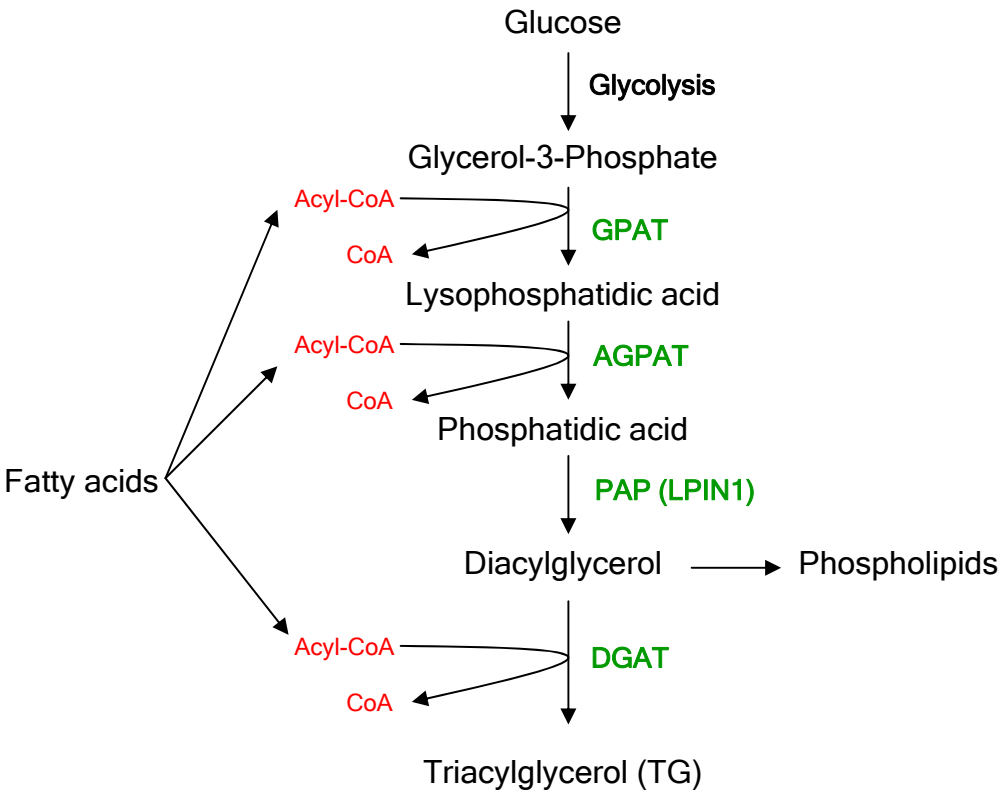


Figure 3.1 The role of *LPIN1* in triacylglycerol (or triglyceride) and phospholipid synthesis
Enzymes are shown in green and acylation in red. GPAT = glycerol-3-phosphatate acyltransferase, AGPAT = 1-acylglycerol-3-phosphate acyltransferase, PAP = phosphatidic acid phosphatase, and DGAT = diacylglycerol acyltransferase.

3.2.6 Lipin 1 is an inducible transcriptional coactivator

In addition to its role in TAG synthesis, lipin 1 appears to function as a transcriptional coactivator of *PGC1α/PPARα* target genes during fasting (Finck et al. 2006). Lipin 1 interacts directly with PGC1α and PPARα through an α-helical leucine-rich motif (LXXIL) to form a complex which then modulates gene transcription. Overexpression of lipin 1 in murine liver increases expression of *PGC1α/PPARα* target genes such as fatty acid oxidation genes and suppresses expression of genes involved in *de novo* fatty acid and TAG synthesis. Lipin 1 may therefore increase hepatic capacity for β-oxidation and help maintain hepatic lipid balance during increased lipid delivery under fasting conditions and diabetes (Finck et al. 2006).

3.2.7 Regulation of lipin 1

Regulation of lipin 1 activity occurs at the levels of mRNA transcription, mRNA splicing, protein phosphorylation, and subcellular localisation. Recent studies in mouse and rat cells have shown that glucocorticoids can induce Lpin1 transcription, increase Lpin1 expression and increase PAP acitivity (Manmontri et al. 2008; Zhang et al. 2008). Also, multiple sites on the lipin 1 protein are phosphorylated in response to insulin and amino acids and dephosphorylated in response to epinephrine and fatty acids (Harris et al. 2007). Insulin- or amino acid-stimulated phosphorylation of lipin 1 occurs in an mTOR (mammalian target of rapamycin) dependent manner (Huffman et al. 2002) in rat adipocytes. A study of murine lipin 1 showed that the phosphorylation status of lipin 1 does not appear to have any effect on its intrinsic PAP activity but does correlate with its subcellular localisation (Harris et al. 2007). Insulin increases the amount of lipin 1 and PAP activity in the soluable fraction of the cell, and decreases the amount of lipin 1 at intracellular membranes where PA is synthesised (Harris et al. 2007).

3.2.8 *LPIN1* expression and human adiposity and insulin resistance

In humans, *LPIN1* expression in adipose tissue appears to be inversely correlated with measures of adiposity such as BMI, and positively correlated with insulin sensitivity (Croce et al. 2007; Donkor et al. 2007b; Lindegaard et al. 2007; Suviolahti et al. 2006; van Harmelen et al. 2007; Yao-Borengasser et al. 2006). For example, *LPIN1* transcript levels from 19 Finnish human fat biopsies were inversely correlated with fasting plasma insulin, glucose, triglycerides and HOMA-IR (Suviolahti et al. 2006). In a US study, lipin 1 mRNA levels in adipose tissue from 36 women and 3 men were inversely correlated with obesity, BMI, percentage body fat and waist circumference, and positively correlated with insulin sensitivity and were higher in normal glucose tolerant subjects than in impaired glucose tolerant (IGT) subjects matched for BMI (Yao-Borengasser et al. 2006). In the same study lipin expression

was inversely correlated with intramyocellular lipids (IMCLs) independent of the fatty acid oxidative capacity of muscle. Therefore, lipin deficiency may lead to partitioning of lipids into muscle and insulin resistance in humans. Subjects treated with the drug pioglitazone, which increases adipose tissue mass and decreases IMCLs, also demonstrated increased expression of lipin and improved insulin sensitivity (Yao-Borengasser et al. 2006). These data suggest *LPIN1* genetic variation that influences expression of lipin 1 and/or lipin 1 function might impact upon human adiposity and insulin sensitivity.

3.2.9 *LPIN1* genetic variation and human adiposity and insulin resistance

There have been a number of studies evaluating the role of common variation in *LPIN1* on human metabolic phenotypes (Loos et al. 2007; Suviolahti et al. 2006; Wiedmann et al. 2007), but the results have been inconsistent across studies, and sometimes within the same study. For example, rs2716610 and a SNP in high linkage disequilibrium, rs2716609, were associated with BMI in a Finnish obesity case-control and in the Quebec Family Study (Loos et al. 2007; Suviolahti et al. 2006) but not in a German population-based cohort (the MONICA study) (Wiedmann et al. 2007). Moreover, *LPIN1* haplotypes were strongly associated with traits underlying metabolic syndrome in the MONICA study but these haplotypes often had the opposite effect on the same traits in a replication cohort (Wiedmann et al. 2007). This inconsistency suggests that further studies are needed to clarify the role of *LPIN1* variation on human metabolic traits.

3.2.10 *LPIN* family of genes

Multiple sequence alignments of lipin-related proteins in a broad range of eukaryotic organisms reveal some strongly conserved domains (Figure 3.2): the amino-terminal lipin domain (NLIP), a nuclear localisation signal (NLS), and a carboxy-terminal lipin domain (CLIP) which carries the DXDXT motif characteristic of a superfamily of magnesium-dependent phosphatases (Donkor et al. 2007a; Han et al. 2006; Peterfy

et al. 2001) and the LXXIL motif that mediates interaction with nuclear receptors (Finck et al. 2006) (see Figure 3.3). As expected, deletion of the CLIP domain abolishes PAP activity. However, insulin-induced phosphorylation of lipin 1 is more pronounced in the NLIP domain and this domain is also required for full PAP activity (Harris et al. 2007). These domains are shared by two other mammalian lipin 1-related proteins, lipin 2 and lipin 3. All three lipin family members in mouse and human possess PAP activity and exhibit unique but overlapping tissue distributions (Donkor et al. 2007a). Lipin 1 is the predominant form in mammalian adipose tissue and skeletal muscle. This is supported by the absence of PAP activity in adipose tissue and skeletal muscle of fld mice (Donkor et al. 2007a). Whereas lipin 2 predominates in brain and liver where fld mice have comparable PAP activity to wild-type (Donkor et al. 2007a). However, lipin 1 appears to account for most or all of the fasting- and glucocorticoid-induced PAP activity in liver (Finck et al. 2006; Harris et al. 2007; Manmontri et al. 2008). In humans *LPIN2* expression is also high in adipose tissue though this needs to be confirmed in other studies (Donkor et al. 2007a). In mice and humans lipin 3 is the main lipin family member in the gastro-intestinal tract (Donkor et al. 2007a).

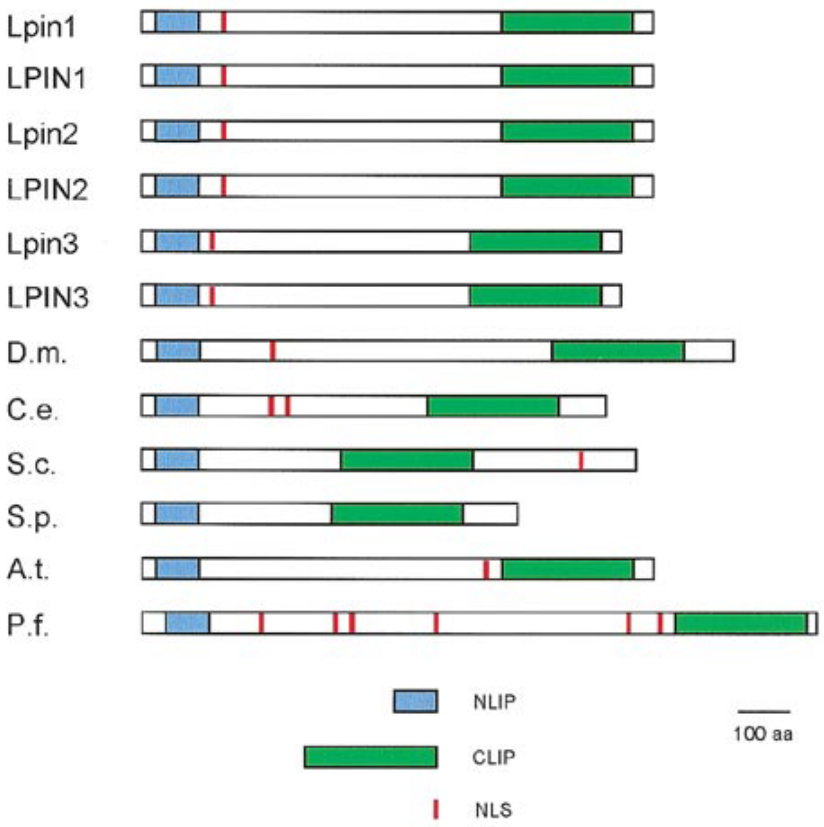


Figure 3.2 Evolutionary conservation of the lipin-protein family, (from (Peterfy et al. 2001)). Lipin homologs in mouse, human, *Drosophila* (D.m.), *C. elegans* (C.e.), *S. cerevisiae* (S.c.), *S. pombe* (S.p.), *A. thaliana* (A.t.) and *P. falciparum* (P.f.). Lpin1, Lpin2 and Lpin3 protein sequences were deduced from full-length cDNAs obtained by RACE cloning in this study. The *LPIN1*, *LPIN2*, *LPIN3*, *Drosophila* , *C. elegans*, *S. cerevisiae*, *S. pombe*, *A. thaliana* and *P. falciparum* protein sequences are based on predictions from EST and genomic sequences. NLIP (blue) and CLIP (green) domains, and predicted nuclear localization signals (NLS - red) are indicated.

Figure 3.3 Known functional motifs and disease mutations in lipin proteins, (adapted from (Reue and Zhang 2008)). G84R causes the phenotype of fld^{2J} mice and S734L is a *LPIN2* Majeed syndrome mutation. NLIP = amino-terminal lipin domain, NLS = nuclear localisation signal, and CLIP = carboxy-terminal lipin domain. DXDXT and LXXIL are peptide sequence motifs.

3.2.11 Genetic studies of *LPIN2* in humans

Mutations in *LPIN2* have been shown to cause Majeed Syndrome, a rare, autosomal recessive autoinflammatory disorder characterised by chronic recurrent multifocal osteomyelitis (CRMO), congenital dyserythropoietic anaemia (CDA), and inflammatory dermatosis (MIM no. #609628) (Al-Mosawi et al. 2007; Ferguson et al. 2006; Ferguson et al. 2005). There have been no reports of abnormalities in fat or sugar metabolism in these patients. However, given that different mutations in the same gene have been known to cause different phenotypes, for example, *LMNA* mutations lead to at least 11 different clinical syndromes (Worman and Bonne 2007), *LPIN2* may still be a good candidate for human insulin resistance and lipodystrophy. Indeed, a recent study reported a SNP within the 3'UTR of *LPIN2* associated with type 2 diabetes and measures of adiposity and insulin sensitivity in a Dutch population-based cohort (Aulchenko et al. 2007). This SNP requires replication but given time constraints I was unable to pursue this during the course of my work.

To my knowledge no genetic studies of *LPIN3* have yet been carried out so, given its likely role as a PAP in the gastro-intestinal tract, it may be worth screening *LPIN3* for pathogenic mutations causing dyslipidemia and lipodystrophy.

3.2.12 Aims of this study

1. To investigate, in UK populations (N=4610), the role of common genetic variation in *LPIN1* on insulin sensitivity, BMI, and other metabolic traits previously associated with *LPIN1* variation (Section 3.3.1);
2. To identify potentially pathogenic mutations in *LPIN1*, *2* and *3* in idiopathic insulin resistant subjects (N=158), including 23 lipodystrophic patients (Sections 3.3.2 and 3.3.3).

3.3 Results

3.3.1 Association studies of *LPIN1* tagging SNPs and metabolic traits

3.3.1.1 Identification and selection of common variants for association testing

To identify *LPIN1* common genetic variants for association studies I took two approaches (Figure 3.4). First, I used data available from the International HapMap consortium (Rel 20, PhaseII) (http://www.hapmap.org) to identify SNPs that were genotyped in CEPH Utah residents of Northern and Western European ancestry (CEU) within *LPIN1* and its flanking ~4 kb regions (Chromosome 2, coordinates 11800212 -11889941 (NCBI B36 coordinates). There were 46 SNPs with MAF>0.01 genotyped in HapMap within this region. Second, I sequenced *LPIN1* exons, exon-intron boundaries, and 3'UTR in 31 unrelated CEPH samples that overlapped with HapMap CEU samples. This approach identified 13 SNPs (including 4 novel variants - Table 3.1). Five of these SNPs overlapped with those in HapMap and two had call rates from my sequencing data <80% leaving 52 SNPs for association testing in UK population-based cohorts.

To avoid genotyping redundant SNPs I evaluated the extent of linkage disequilibrium (LD) between these SNPs (Figure 3.5) and used pairwise LD measures to select tagging SNPs (Section 3.4.2) using an r^2 cutoff of 0.8. Twenty -five SNPs selected using this processs tagged all SNPs in the full HapMap CEU panel except one, rs17603350, which was added to the tagging SNP set (Figure 3.4).

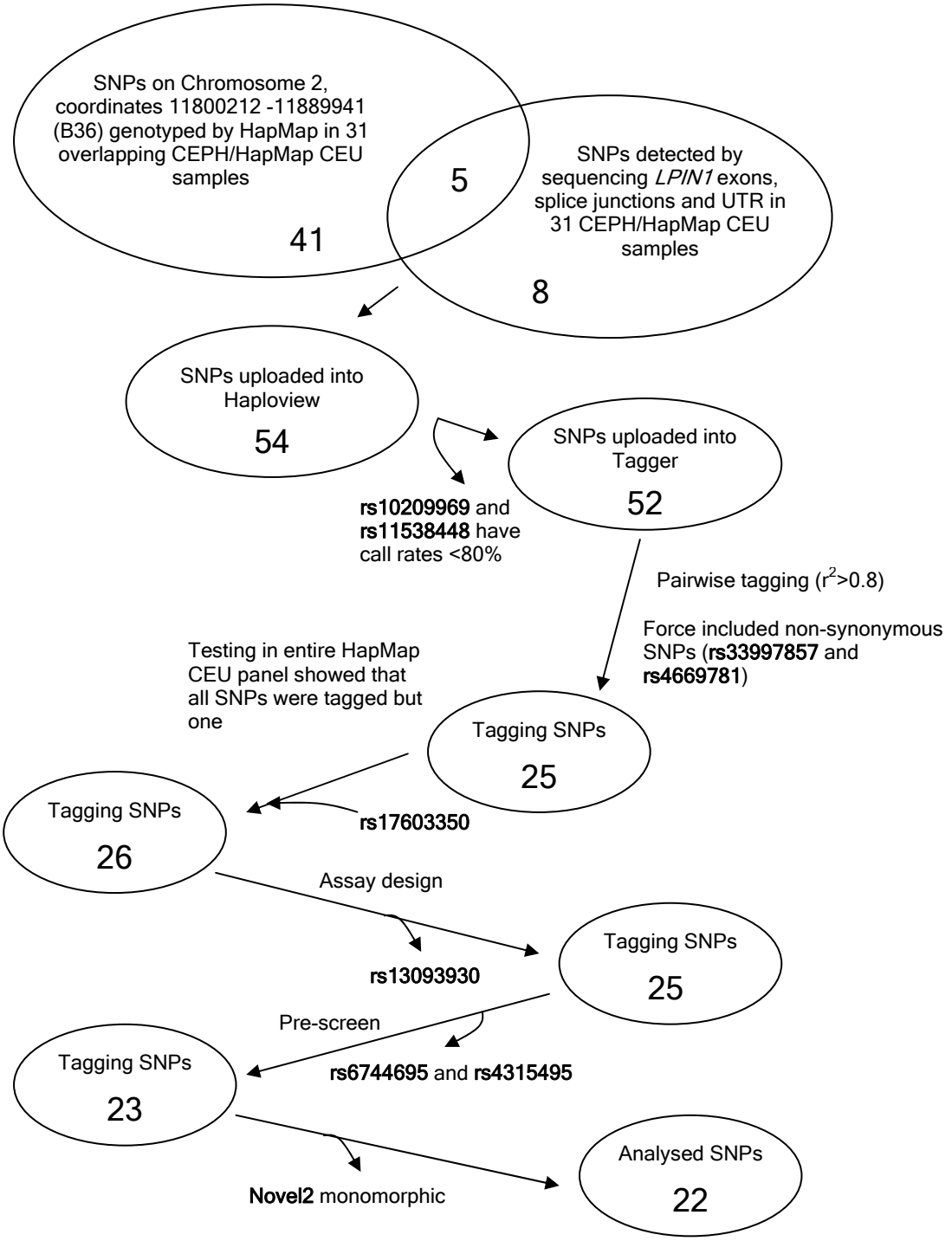


Figure 3.4 Flow chart showing the process of SNP identification and selection

Table 3.1 *LPIN1* sequence variants detected in 48 CEPH controls

Genic position	Genomic position	Nucleotide substitution	Protein consequence	MAF in CEPH	Rs number
Upstream	11789942-3	InsT		0.27	rs3214670
Intron 2	11823327	T>C		0.35	rs10209969
	11825293	C>T		0.28	Novel1
Exon 4	11829212	C>T	I184I	0.14	rs11538448
Intron 5	11836949	G>A		0.11	rs2289193
	11841434	C>T	D395D	0.016	Novel2
Intron 8	11842401	C>T		0.28	rs3795974
Exon 10	11844689	G>A	**V494M**	0.02	rs33997857
Intron 11	11849392-3	InsT		0.12	Indel1
Intron 12	11849624	G>T		0.44	rs7561070
Exon 14	11860533	C>T	**P610S**	0.04	rs4669781
Intron 15	11862198	DelG		0.14	Indel2
Exon 20	11882248	G>A	P851P	0.016	Novel3
3`UTR	11883265	C>T		0.3	rs1050800
	11883768	T>C		0.17	rs11524
	11884454	C>G		0.16	Novel4

Genomic coordinates correspond to NCBI Build 36. Non-synonymous changes are highlighted in bold. Ins = insertion. Del = deletion.

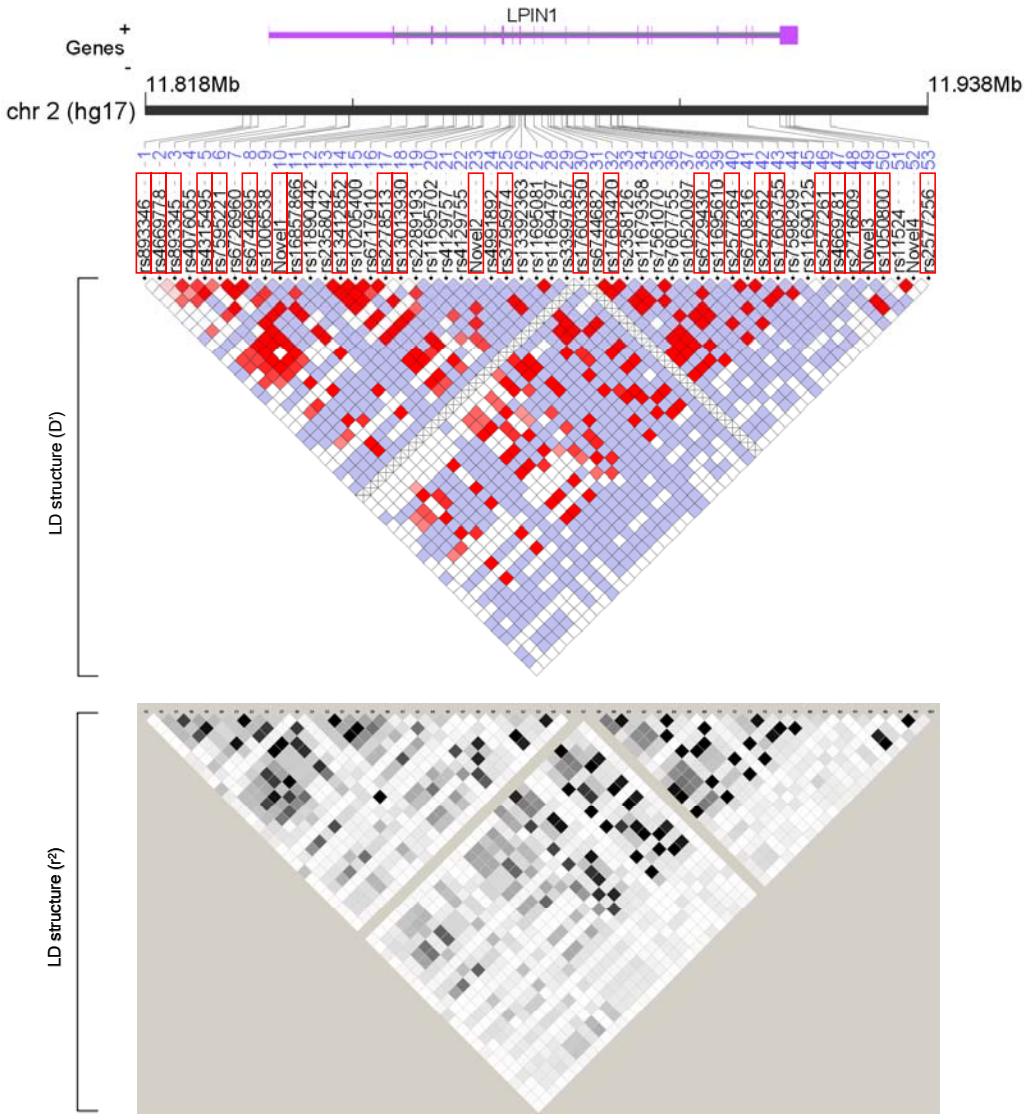


Figure 3.5 Pairwise linkage disequilibrium (LD) between SNPs in LPIN1 and the surrounding sequence

SNPs displayed are a combination of those detected from sequencing of 31 CEPH samples and those genotyped in overlapping samples of CEPH Utah trios in HapMap (release 20, phaseII Jan06, on NCBI B35 assembly, dbSNP B125) as well as one SNP (rs17603350) in CEU trios that was not present in the 31 overlapping samples. The upper plot presents LD as D' - see figure key for details. This figure was generated using Locusview (T. Petryshen, A. Kirby, M. Ainscow, unpublished software). In the lower plot LD is represented as r^2. Black diamonds represent an r^2 value of 1, grey diamonds represent intermediate values of r^2, and white diamonds indicate r^2 values of 0. This plot was generated using Haploview (Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005 Jan 15). SNPs in red boxes were selected as tagSNPs for this study using Tagger based on a pairwise r^2 threshold ≥0.8.

3.3.1.2 Association results of *LPIN1* tagging SNPs with fasting insulin and BMI

Of the 26 SNPs selected to tag the gene, one failed assay design, two failed pre-screening and one was monomorphic in the samples tested (section 3.5.4), leaving 22 SNPs for association testing (Figure 3.4 and Table 3.2). Of these 22 tagging SNPs with genotype data one, rs17603755, deviated from Hardy-Weinberg Equilibrium ($P > 0.01$) and was not tested for association. Tagging SNPs were successfully genotyped in >85% of samples from the MRC Ely and Hertfordshire cohort studies (call rates for each SNP are presented in Table 3.2) and the average call rate was 92.1%.

Association of tagging SNPs with fasting insulin levels and BMI was tested by linear regression analysis in the MRC Ely and Hertfordshire cohorts separately (Tables 3.3 and 3.4). In the MRC Ely cohort, the minor allele of rs13412852 is nominally associated with lower fasting insulin levels ($P = 0.041$) and the minor allele of rs17603350 is nominally associated with higher BMI ($P = 0.031$) (Table 3.3) but these associations are not replicated in the Hertfordshire cohort (Table 3.4). Conversely, in the Hertfordshire cohort, rs17603420 and rs2577261 are nominally associated with BMI ($P = 0.01$ and $P = 0.006$ respectively) (Table 3.4), but are not associated with BMI in the MRC Ely cohort (Table 3.3). To increase the statistical power of this study to detect modest effects of SNPs on insulin levels and BMI, I pooled data from Ely and Hertfordshire cohorts and performed a joint analysis (Table 3.5). No SNPs were associated with fasting insulin levels but rs13412852, rs17603420 and rs2577261 were nominally associated with BMI ($P \leq 0.05$). I performed 10,000 permutations of BMI to test for empirical significance of rs13412852, rs17603420 and rs2577261, which was confirmed in every case (P = 0.028, 0.006, and 0.005 respectively).

Table 3.2 Tagging SNPs in *LPIN1* analysed for association with metabolic traits

TagSNPs	Genomic position	Genic position	Protein consequence	Alleles (major/minor)	MAF in CEPHs	Call rate in Ely and Hertfordshire (%)
rs893346	11800212	5'		A/G	0.065	94
rs4669778	11801431	5'		T/C	0.371	92
rs893345	11802567	5'		G/A	0.387	86
rs7595221	11814426	Intron 1		A/G	0.371	93
Novel1	11825293	Intron 2		C/T	0.204	91
rs16857866	11828169	Intron 3		C/T	0.016	93
rs13412852	11832392	Intron 5		C/T	0.371	93
rs2278513	11835356	Intron 5		C/T	0.419	92
rs3795974	11842401	Intron 8		G/A	0.29	92
rs33997857	11844689	Exon 10	494V>M	G/A	0.016	94
rs17603350	11845874	Intron 10		C/T	0	91
rs17603420	11847191	Intron 11		G/A	0.5	93
rs6729430	11852207	Intron 12		C/T	0.016	94
rs2577264	11854249	Intron 13		C/T	0.306	93
rs2577262	11856397	Intron 13		G/A	0.339	93
rs17603755	11856469	Intron 13		G/A	0.125	87
rs2577261	11858238	Intron 13		G/T	0.145	91
rs4669781	11860533	Exon 14	610P>S	C/T	0.016	95
rs2716609	11877356	Intron 18		T/C	0.129	88
Novel3	11882248	Exon 20	851P>P	G/A	0.016	94
rs1050800	11883265	3'UTR		C/T	0.145	92
rs2577256	11889941	3'		A/G	0.468	93

Minor Allele Frequency (MAF) is provided for a subset of 31 CEPH samples that overlapped between the HapMap CEU trios and samples used for my own resequencing efforts. These frequencies compare well to those reported in HapMap. Genomic coordinates correspond to NCBI Build 36.

Table 3.3 Mean fasting insulin levels and mean BMI of study participants by *LPIN1* tagSNP genotype in the MRC Ely cohort

	Ely							
	Insulin (pmol/l)				BMI (kg/m^2)			
SNP	0	1	2	P value	0	1	2	P value
rs893346	49.18 ± 1.02	47.65 ± 1.05	57.57 ± 1.08	*0.609*	27.23 ± 0.13	27.5 ± 0.39	26.3 ± 1.7	*0.303*
rs4669778	49.99 ± 1.03	48.88 ± 1.02	47.23 ± 1.03	*0.216*	27.42 ± 0.24	27.41 ± 0.16	27.11 ± 0.27	*0.109*
rs893345	48.17 ± 1.03	48.69 ± 1.02	49.76 ± 1.03	*0.424*	27.27 ± 0.28	27.75 ± 0.32	25.87 ± 1.57	*0.967*
rs7595221	47.75 ± 1.03	50.08 ± 1.02	48.59 ± 1.04	*0.55*	27.11 ± 0.22	26.99 ± 0.24	26.09 ± 0.92	*0.815*
Novel1	48.46 ± 1.02	49.79 ± 1.03	49.67 ± 1.08	*0.409*	27.15 ± 0.15	27.49 ± 0.21	27.01 ± 0.64	*0.393*
rs16857866	49.1 ± 1.01	53.35 ± 1.08	47.7 ± 0	*0.398*	27.23 ± 0.12	28.02 ± 0.64	29.65 ± 0	*0.215*
rs13412852	50.18 ± 1.02	48.63 ± 1.02	45.1 ± 1.05	***0.042***	27.42 ± 0.18	27.36 ± 0.18	26.44 ± 0.35	*0.054*
rs2278513	47.41 ± 1.03	49.45 ± 1.02	48.82 ± 1.04	*0.37*	27.06 ± 0.22	27.28 ± 0.16	27.45 ± 0.26	*0.467*
rs3795974	48.59 ± 1.02	49.57 ± 1.02	49.93 ± 1.04	*0.441*	27.21 ± 0.19	27.76 ± 0.96	27.86 ± 3.19	*0.844*
rs33997857	48.95 ± 1.01	46.11 ± 1.1	56.18 ± 1.9	*0.627*	27.24 ± 0.12	27.7 ± 0.67	28.76 ± 0.57	*0.401*
rs17603350	48.99 ± 1.02	47.89 ± 1.05	99.5 ± 1.37	*0.986*	27.34 ± 0.13	26.3 ± 0.36	27.52 ± 1.28	***0.031***
rs17603420	50.63 ± 1.03	47.75 ± 1.02	49.14 ± 1.03	*0.271*	27.33 ± 0.2	27.39 ± 0.18	26.9 ± 0.25	*0.269*
rs6729430	48.95 ± 1.02	45.77 ± 1.1	49.99 ± 1.47	*0.61*	27.24 ± 0.12	27.39 ± 0.34	29.79 ± 1.28	*0.508*
rs2577264	48.61 ± 1.02	48.91 ± 1.02	50.6 ± 1.04	*0.368*	27.19 ± 0.2	27.37 ± 0.18	27.21 ± 0.3	*0.733*
rs2577262	49.25 ± 1.02	48.96 ± 1.02	48.72 ± 1.05	*0.639*	27.24 ± 0.17	27.61 ± 0.28	27.99 ± 0.86	*0.857*
rs2577261	49.52 ± 1.02	46.67 ± 1.03	48.09 ± 1.12	*0.15*	27.18 ± 0.12	27.38 ± 0.18	27.23 ± 0.3	*0.256*
rs4669781	48.91 ± 1.02	49.56 ± 1.04	43.59 ± 1.21	*0.844*	27.22 ± 0.12	27.19 ± 0.18	27.25 ± 0.29	*0.780*
rs2716609	48.99 ± 1.02	48.25 ± 1.05	44.85 ± 1.2	*0.61*	27.13 ± 0.15	27.34 ± 0.17	26.82 ± 0.24	*0.073*
Novel3	48.11 ± 1.02	50.44 ± 1.04	52.4 ± 1.17	*0.117*	27.25 ± 0.12	27.25 ± 1.26	0 ± 0	*0.987*
rs1050800	49.01 ± 1.01	47.15 ± 1.19	0 ± 0	*0.728*	27.21 ± 0.14	27.34 ± 0.23	26.55 ± 0.62	*0.993*
rs2577256	48.81 ± 1.02	49.32 ± 1.03	49.09 ± 1.09	*0.862*	27.02 ± 0.24	27.41 ± 0.19	26.94 ± 0.38	*0.208*

Data are means ± standard error. The *P* value indicates the results of a regression analysis assuming an additive model of gene action (nominally significant values, p<0.05, are highlighted in bold). For fasting insulin the analysis was performed on log-transformed data, and the table shows geometric means and standard errors. 0 = homozygous for the major allele (refer to Supplementary table 2), 1 = heterozygous, 2 = homozygous for the minor allele.

Table 3.4 Mean fasting insulin levels and mean BMI of study participants by *LPIN1* tagSNP genotype in the Hertfordshire cohort study

	Hertfordshire							
	Insulin (pmol/l)				BMI (kg/m^2)			
SNP	0	1	2	P value	0	1	2	P value
rs893346	70.6 ± 1.01	72.33 ± 1.04	78.17 ± 1.2	*0.486*	27.36 ± 0.09	27.43 ± 0.24	26.14 ± 1.08	*0.993*
rs4669778	71.28 ± 1.02	72.11 ± 1.02	69.06 ± 1.03	*0.409*	27.41 ± 0.16	27.37 ± 0.12	27.33 ± 0.16	*0.719*
rs893345	71.15 ± 1.03	71.73 ± 1.02	69.14 ± 1.03	*0.406*	27.21 ± 0.17	27.38 ± 0.12	27.49 ± 0.17	*0.234*
rs7595221	70.01 ± 1.02	71.45 ± 1.02	70.3 ± 1.03	*0.708*	27.34 ± 0.14	27.44 ± 0.12	27.23 ± 0.19	*0.734*
Novel1	71.71 ± 1.02	70.63 ± 1.02	71.25 ± 1.06	*0.575*	27.34 ± 0.1	27.52 ± 0.16	27.72 ± 0.45	*0.172*
rs16857866	70.96 ± 1.01	72.11 ± 1.06	0 ± 0	*0.734*	27.36 ± 0.09	27.8 ± 0.4	0 ± 0	*0.306*
rs13412852	70.57 ± 1.02	71.63 ± 1.02	69.18 ± 1.04	*0.89*	27.47 ± 0.13	27.36 ± 0.12	27.15 ± 0.2	*0.249*
rs2278513	69.47 ± 1.02	71.46 ± 1.02	72.34 ± 1.03	*0.234*	27.45 ± 0.14	27.32 ± 0.12	27.42 ± 0.2	*0.753*
rs3795974	71.04 ± 1.02	70.16 ± 1.02	72.91 ± 1.04	*0.602*	27.31 ± 0.13	27.41 ± 0.13	27.42 ± 0.22	*0.633*
rs33997857	70.82 ± 1.01	78.33 ± 1.07	73.83 ± 0	*0.166*	27.36 ± 0.08	27.91 ± 0.44	25.2 ± 0	*0.259*
rs17603350	71.27 ± 1.01	67.78 ± 1.05	60.16 ± 2.08	*0.282*	27.4 ± 0.09	27.15 ± 0.36	30.9 ± 4.96	*0.725*
rs17603420	70.43 ± 1.02	71.41 ± 1.02	70.26 ± 1.03	*0.971*	27.58 ± 0.16	27.42 ± 0.12	26.92 ± 0.17	***0.01***
rs6729430	70.98 ± 1.01	71.68 ± 1.08	0 ± 0	*0.849*	27.34 ± 0.08	28.37 ± 0.53	0 ± 0	*0.053*
rs2577264	71.27 ± 1.02	69.8 ± 1.02	73.51 ± 1.04	*0.644*	27.28 ± 0.13	27.39 ± 0.12	27.48 ± 0.23	*0.462*
rs2577262	70.76 ± 1.02	71.58 ± 1.02	69.77 ± 1.04	*0.95*	27.47 ± 0.12	27.37 ± 0.13	26.98 ± 0.25	*0.148*
rs2577261	71.08 ± 1.02	71.39 ± 1.03	73.5 ± 1.08	*0.719*	27.23 ± 0.09	27.67 ± 0.22	29.06 ± 0.72	***0.006***
rs4669781	71.03 ± 1.01	69.44 ± 1.03	66.11 ± 1.11	*0.426*	27.41 ± 0.09	27.02 ± 0.24	26.81 ± 2.07	*0.126*
rs2716609	71.19 ± 1.01	69.03 ± 1.04	78.39 ± 1.22	*0.597*	27.35 ± 0.1	27.42 ± 0.17	27.82 ± 0.66	*0.465*
Novel3	71.65 ± 1.02	68.38 ± 1.03	76.59 ± 1.08	*0.34*	27.38 ± 0.08	25.71 ± 0.72	0 ± 0	*0.103*
rs1050800	71.02 ± 1.01	69.3 ± 1.15	0 ± 0	*0.848*	27.4 ± 0.1	27.34 ± 0.16	27.14 ± 0.55	*0.585*
rs2577256	70.48 ± 1.02	71.77 ± 1.02	74.09 ± 1.07	*0.386*	27.19 ± 0.16	27.42 ± 0.12	27.47 ± 0.18	*0.219*

Data are means ± standard error. The *P* value indicates the results of a regression analysis assuming an additive model of gene action (nominally significant values, p<0.05, are highlighted in bold). For fasting insulin the analysis was performed on log-transformed data, and the table shows geometric means and standard errors. 0 = homozygous for the major allele (refer to Supplementary table 2), 1 = heterozygous, 2 = homozygous for the minor allele.

Table 3.5 Joint analysis of the association between *LPIN1* tagSNPs and fasting insulin or BMI in combined Ely and Hertfordshire datasets

	Fasting insulin		BMI	
SNP	β ± SE	P value	β ± SE	P value
rs893346	0.0086 ± 0.03	*0.755*	0.1339 ± 0.2	*0.494*
rs4669778	-0.0191 ± 0.01	*0.163*	-0.131 ± 0.1	*0.18*
rs893345	-0.0042 ± 0.01	*0.768*	0.084 ± 0.1	*0.401*
rs7595221	0.0085 ± 0.01	*0.532*	0.0005 ± 0.1	*0.996*
Novel1	0.0002 ± 0.02	*0.988*	0.1893 ± 0.12	*0.123*
rs16857866	0.0374 ± 0.05	*0.442*	0.5632 ± 0.35	*0.107*
rs13412852	-0.0183 ± 0.01	*0.203*	-0.2242 ± 0.1	***0.028***
rs2278513	0.0207 ± 0.01	*0.137*	0.0281 ± 0.1	*0.777*
rs3795974	0.0116 ± 0.01	*0.408*	0.0602 ± 0.1	*0.544*
rs33997857	0.0386 ± 0.05	*0.446*	0.4933 ± 0.36	*0.171*
rs17603350	-0.0299 ± 0.04	*0.428*	-0.4412 ± 0.27	*0.1*
rs17603420	-0.0091 ± 0.01	*0.502*	-0.2653 ± 0.1	***0.006***
rs6729430	-0.0121 ± 0.06	*0.839*	0.8156 ± 0.42	*0.053*
rs2577264	0.0121 ± 0.01	*0.39*	0.0873 ± 0.1	*0.381*
rs2577262	-0.0049 ± 0.01	*0.738*	-0.1236 ± 0.1	*0.237*
rs2577261	-0.0128 ± 0.02	*0.593*	0.4756 ± 0.17	***0.005***
rs4669781	-0.0237 ± 0.03	*0.434*	-0.1977 ± 0.21	*0.356*
rs2716609	0.003 ± 0.02	*0.881*	0.2383 ± 0.14	*0.093*
Novel3	-0.043 ± 0.12	*0.71*	-1.0324 ± 0.83	*0.214*
rs1050800	0.0144 ± 0.02	*0.447*	-0.0461 ± 0.13	*0.731*
rs2577256	-0.0094 ± 0.01	*0.503*	0.1761 ± 0.1	*0.074*

β (regression coefficient) is the mean change (± standard error) in fasting insulin or BMI per minor allele. Analysis for fasting insulin was performed on log-transformed data and all analyses included an adjustment for cohort term. Nominally significant SNPs are highlighted in bold.

3.3.1.3 Use of publicly available datasets to increase power

Two of these SNPs, rs13412852 and rs2577261, overlapped with SNPs on the Affymetrix 500k and Illumina 300k SNP chips, and rs17603420 could be imputed. Consequently I was able to further increase the power of my study to detect modest effects of these SNPs on BMI by performing meta-analyses with in-house data (EPIC-Obesity study, N = 2415) and, in the case of rs13412852 and rs2577261, data deposited by the WTCCC and WTSI and published online from the British 1958 DNA collection (N = 1479) (http://www.b58cgene.sgul.ac.uk/, accessed January 2008). SNPs rs2577261 and rs17603420 were not associated with BMI in the meta-analysis (*P* = 0.114 and 0.071 respectively). However, the association between rs13412852 and BMI remained marginally statistically significant (*P* = 0.042) in the meta-analysis (Figure 3.6).

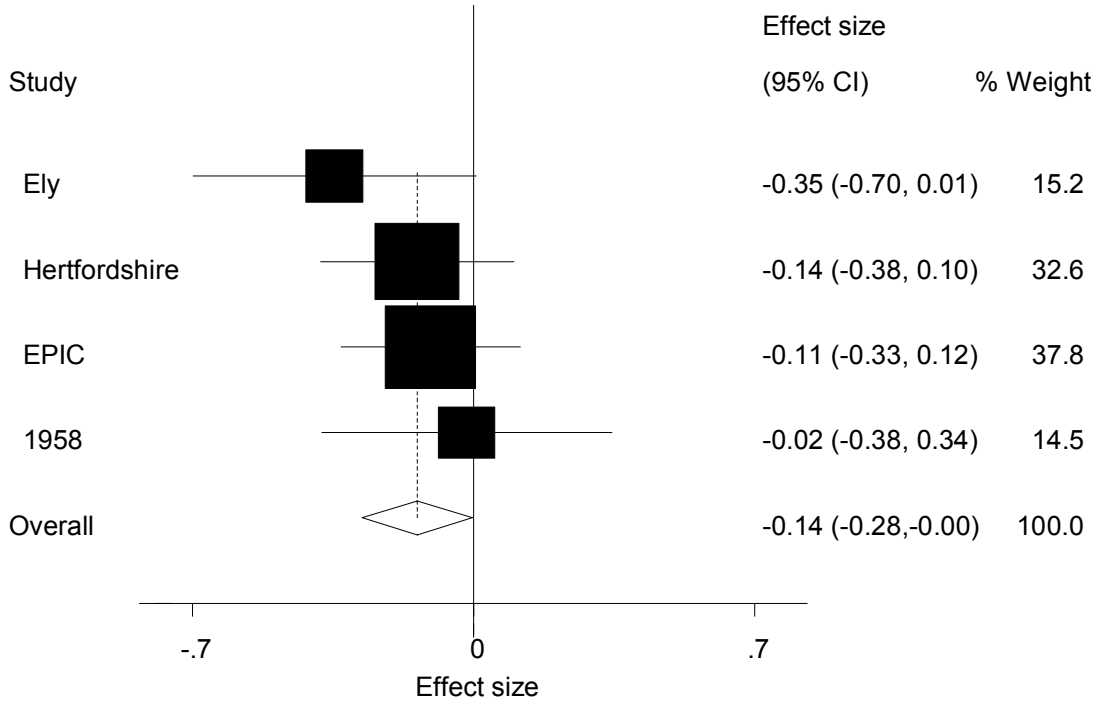


Figure 3.6 Association between rs13412852 and BMI in studies within the meta-analysis
Ely = MRC Ely study, Hertfordshire = Hertfordshire cohort study, EPIC = EPIC Obesity cohort, and 1958 = 1958 British Birth cohort. Overall = combined effect size, *P* = 0.045.

3.3.1.4 Testing for association of tagging SNPs with additional metabolic quantitative traits

In pooled data from Ely and Hertfordshire cohorts, linear regression analysis was used to test for association between *LPIN1* tagging SNPs and metabolic quantitative traits found to be associated with *LPIN1* variation in the MONICA study (Table 3.6A). A number of nominal associations (P <0.05) were detected, namely between *LPIN1* variation and systolic and diastolic blood pressure, low-density lipoprotein (LDL) and high-density lipoprotein (HDL) cholesterol, triglycerides and, in Ely only, glycosylated haemoglobin (HbA1c) levels. Two SNPs, rs2278513 and rs2577256, are present on the Affymetrix SNP chip and have been analysed as part of the DGI study at the Broad Institute. Consequently I was able to use publicly available summary statistics (http://www.broad.mit.edu/diabetes/scandinavs/metatraits.html) to perform meta-analyses with my data (see Table 3.6B for P-values).

3.3.1.5 Testing for association of tagging SNPs with hypertension and diabetes

In addition to continuous metabolic traits, the MONICA study reported association between *LPIN1* variation and risk of diabetes and hypertension. These traits were also tested in my pooled Ely and Hertfordshire data by logistic regression analysis. One SNP, rs2577256, was associated with diabetes and several SNPs were associated with hypertension status (Table 3.7). For rs2577256 meta-analysis was performed with publicly available WTCCC diabetes and hypertension data (see WTCCC and meta-analysis summary statistics in Table 3.7).

The MONICA study detected associations between metabolic traits and haplotypes of three *LPIN1* SNPs, rs33997857, rs6744682, and rs6708316. Only one of these SNPs, rs33997857, is within my SNP set but rs2577262 is highly correlated with the two other SNPs in the haplotype, rs6744682 and rs6708316 (r^2 = 1.0 and 0.96 respectively in HapMap CEU trios). To attempt replication of the MONICA study data

I tested haplotypes of rs33997857 and rs2577262 against metabolic traits in Ely and Hertfordshire but only found nominal associations with hypertension (Table 3.8).

Table 3.6A Statistically significant associations between traits underlying metabolic syndrome and *LPIN1* tagSNPs in Ely and Hertfordshire cohorts

SNP	Ely				Hertfordshire				Ely + Hertfordshire	
	0	1	2	P	0	1	2	P	β ± SE	P
	Systolic blood pressure (mmHg)									
rs2278513	130.30 ± 0.76	131.20 ± 0.59	133.33 ± 0.95	0.01	132.81 ± 0.67	133.09 ± 0.52	134.62 ± 0.83	0.074	0.009 ± 0.003	0.004
rs2716609	131.34 ± 0.51	130.17 ± 0.97	134.90 ± 3.29	0.824	132.56 ± 0.42	135.56 ± 0.76	132.49 ± 2.60	0.005	0.01 ± 0.004	0.023
	Diastolic blood pressure (mmHg)									
rs2278513	77.63 ± 0.46	78.62 ± 0.37	79.73 ± 0.61	0.004	70.39 ± 0.38	70.76 ± 0.30	71.24 ± 0.51	0.067	0.01 ± 0.003	0.001
rs2716609	78.64 ± 0.32	78.41 ± 0.62	80.29 ± 1.89	0.689	70.32 ± 0.25	72.03 ± 0.44	69.29 ± 1.51	0.011	0.012 ± 0.005	0.008
	HbA1c (%)									
rs4669778	5.45 ± 0.03	5.4 ± 0.02	5.34 ± 0.03	0.007	Not tested	Not tested	Not tested	N/A	-0.009 ± 0.004	0.007
rs13412852	5.44 ± 0.03	5.37 ± 0.02	5.33 ± 0.04	0.004	Not tested	Not tested	Not tested	N/A	-0.01 ± 0.004	0.004
rs3795974	5.36 ± 0.02	5.41 ± 0.02	5.46 ± 0.05	0.016	Not tested	Not tested	Not tested	N/A	0.008 ± 0.003	0.016
rs17603420	5.44 ± 0.03	5.4 ± 0.02	5.32 ± 0.03	0.004	Not tested	Not tested	Not tested	N/A	-0.01 ± 0.003	0.004
rs2577264	5.35 ± 0.02	5.42 ± 0.02	5.46 ± 0.05	0.004	Not tested	Not tested	Not tested	N/A	0.01 ± 0.003	0.004
Novel3	5.4 ± 0.01	5.05 ± 0.09		0.016	Not tested	Not tested	Not tested	N/A	-0.07 ± 0.029	0.016
	LDL cholesterol (mmol/l)									
rs2577256	3.43 ± 0.05	3.5 ± 0.03	3.57 ± 0.05	0.036	3.91 ± 0.04	3.93 ± 0.03	4.02 ± 0.04	0.02	0.019 ± 0.006	0.002
	HDL cholesterol (mmol/l)									
rs7595221	1.43 ± 0.02	1.49 ± 0.01	1.48 ± 0.02	0.011	1.53 ± 0.01	1.51 ± 0.01	1.57 ± 0.02	0.368	0.012 ± 0.005	0.019
rs893345	1.47 ± 0.02	1.48 ± 0.02	1.42 ± 0.02	0.032	1.56 ± 0.02	1.51 ± 0.01	1.54 ± 0.02	0.332	-0.011 ± 0.006	0.038
	Triglycerides (mmol/l)									
rs2577256	1.43 ± 0.04	1.41 ± 0.03	1.53 ± 0.05	0.04	1.59 ± 0.03	1.66 ± 0.02	1.67 ± 0.04	0.066	0.029 ± 0.01	0.007

Data for Ely and Hertfordshire cohorts separately are presented as means ± standard error. 0 = homozygous for the major allele (refer to Supplementary table 2), 1 = heterozygous, 2 = homozygous for the minor allele. β (regression coefficient) is the mean change (± standard error) in the metabolic trait per minor allele. The *P* values indicate the results of a regression analysis on log-transformed data that assumed an additive model of gene action. It should be noted that there is a high prevalence of treatment for hypertension in the Hertfordshire cohort (36%) which might render measured blood pressure values of limited use.

Table 3.6B Meta-analysis of Ely, Hertfordshire and Broad continuous trait data

SNP	Trait	Broad β ± SE	Meta P
rs2278513	Systolic BP	-0.570 ± 0.543	0.110
rs2278513	Diastolic BP	-0.860 ± 0.416	0.157
rs2577256	LDL cholesterol	-0.01 ± 0.027	0.051
rs2577256	Triglycerides	0 ± 0.27	0.006

Regression coefficients ± SE of publicly available data from the Broad Institute are represented in the penultimate column and *P* values for the meta-analysis of my UK data and the Broad data are represented in the last column.

Table 3.7 Statistically significant associations between *LPIN1* tagSNPs and risk of diabetes or hypertension in Ely and Hertfordshire cohorts, and in a pooled analysis of both cohorts and publicly available WTCCC data

SNP	Ely		Hertfordshire		Ely and Hertfordshire		WTCCC		Meta (+WTCCC)	
	OR ± SE	P value	OR ± SE	P value	OR ± SE	P value	OR ± SE	P value	OR ± SE	P value
	Diabetes (502 cases, 3688 controls)									
rs2577256	1.06 ± 0.17	0.716	1.29 ± 0.10	0.001	1.24 ± 0.09	0.002	1.02 ± 0.04	0.589	1.08 ± 0.04	0.038
	Hypertension (1603 cases, 2991 controls)									
rs3795974	1.05 ± 0.09	0.562	1.14 ± 0.06	0.023	1.11 ± 0.05	0.024				
rs17603420	1.09 ± 0.09	0.283	0.84 ± 0.05	0.002	0.91 ± 0.04	0.044				
rs2577264	1.06 ± 0.09	0.503	1.16 ± 0.07	0.009	1.13 ± 0.05	0.010				
rs2577262	0.90 ± 0.08	0.250	0.89 ± 0.05	0.058	0.90 ± 0.04	0.029				
rs4669781	0.73 ± 0.14	0.093	0.78 ± 0.10	0.052	0.76 ± 0.08	0.011				
rs1050800	0.84 ± 0.10	0.141	0.84 ± 0.07	0.025	0.84 ± 0.05	0.009				
rs2577256	1.10 ± 0.09	0.248	1.20 ± 0.07	0.001	1.17 ± 0.05	0.001	1.00 ± 0.04	0.998	1.07 ± 0.03	0.021

Data are odds ratios (giving the increase in disease risk per minor allele) ± standard error. *P* values indicate the outcome of logistic regression analyses. Included as diabetes cases were individuals with previously diagnosed diabetes, and also newly diagnosed diabetes based on glucose measurements taken 2 hours after an oral glucose tolerance test. The two columns headed WTCCC provide summary statistics for the publicly available Wellcome Trust Case Control Consortium (WTCCC). The two columns headed Meta (+WTCCC) describe effect size and significance of the association between rs2577256 and diabetes or hypertension in meta-analyses of Ely, Hertfordshire and the publicly available WTCCC data. The meta-analyses are based on diabetes data for 2426 cases and 6619 controls, and hypertension data for 3486 cases and 5796 controls.

Table 3.8 Statistical association of haplotypes of rs33997857 and rs2577262 with metabolic traits in Ely and Hertfordshire cohorts

	P values		
Trait	Haplotype 11	Haplotype 12	Haplotype 21
Systolic blood pressure	0.1421	0.1956	0.5611
Diastolic blood pressure	0.1492	0.2575	0.6469
HbA1c *	0.617	0.781	0.448
LDL cholesterol	0.9382	0.937	0.9806
HDL cholesterol	0.1022	0.08027	0.8017
Waist	0.8637	0.7763	0.6769
Triglycerides	0.9778	0.6594	0.1002
Diabetes	0.8743	0.9038	0.904
Hypertension	*0.02982*	*0.02752*	0.9864

P values are the result of meta-analysis of Ely and Hertfordshire summary statistics using the METAL program, except in the case of HbA1c data which is based on Ely only (*).

3.3.2 *LPIN1* mutation screening in the SIR cohort

3.3.2.1 *LPIN1* variation in SIR

A total of 44 variants were detected in insulin resistant or lipodystrophic patients (Table 3.9), eight of which were present in the coding sequence (Figure 3.7). Coding sequence variants that did not alter the amino acid sequence (shown below the schematic in Figure 3.7) and/or that were also present in CEPH controls were considered unlikely to be pathogenic. This left three rare nonsynonymous variants (A353T, R552K, and G582R - underlined above the schematic in Figure 3.7) that did not fall within any known functional domains within *LPIN1*.

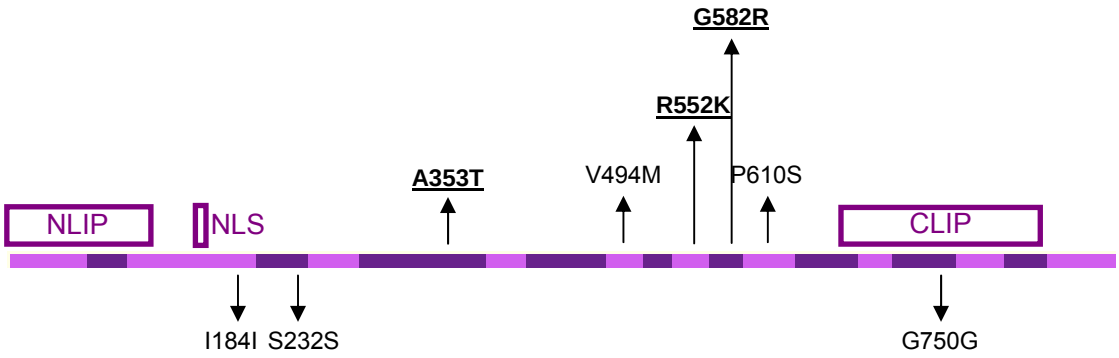


Figure 3.7 Coding *LPIN1* variants in the SIR cohort
Schematic of the lipin 1 protein showing exons in alternating bright and dark purple and known domains among lipin family proteins in boxes. Arrows indicate the location of coding SNPs detected in *LPIN1* by sequencing 23 patients with partial lipodystrophy and 135 patients with other syndromes of severe insulin resistance. Non-synonymous variants are above the schematic and synonymous variants are below the schematic. Non-synonymous mutations absent from control samples were considered potentially pathogenic (underlined). NLIP (amino acids 1-114) = N-terminal lipin domain, NLS (amino acids 153-158) = nuclear localisation signal, and CLIP (amino acids 674-830) = C-terminal lipin domain, also referred to as the LNS2 (Lipin/Ned1/Smp2) domain.

Table 3.9 *LPIN1* sequence variants detected in a cohort of severe insulin resistant patients and 48 CEPH controls

Genic position	Genomic position	Nucleotide substitution	Protein consequence	MAF in SIR	MAF in CEPH	Rs number
Upstream	11789942-3	InsT		0.27	0.27	rs3214670
Intron 2	11823327	T>C		0.25	0.35	rs10209969
	11823350	T>C		<0.01	0	
	11825297	G>T		<0.01	0	
	11825293	C>T		0.19	0.28	Novel1
Exon 4	11829212	C>T	I184I	0.14	0.14	rs11538448
Exon 5	11831296	G>C	S232S	<0.01	0	
Intron 5	11836949	G>A		0.12	0.11	rs2289193
Exon 7	11839985	G>A	**A353T**	<0.01	0	
	11841434	C>T	D395D	0	0.016	Novel2
Intron 8	11841548	A>T		<0.01	0	
	11842376	C>T		<0.01	0	
	11842401	C>T		0.43	0.28	rs3795974
Exon 10	11844689	G>A	**V494M**	0.03	0.02	rs33997857
Intron 10	11845874	C>T		0.02	0	rs17603350
	11845928	C>T		<0.01	0	
Intron 11	11846054	G>C		<0.01	0	
	11849353	C>T		<0.01	0	
	11849392-3	InsT		0.1	0.12	
	11849396	C>T		<0.01	0	
Exon 12	11849540	G>A	**R552K**	<0.01	0	
Intron 12	11849624	G>T		0.49	0.44	rs7561070
	11852911	A>G		<0.01	0	
	11852932	G>T		<0.01	0	
Exon 13	11853030	G>A	**G582R**	<0.01	0	
Intron 13	11853126	T>C		<0.01	0	
Exon 14	11860533	C>T	**P610S**	0.05	0.04	rs4669781
Intron 14	11860690	G>A		<0.01	0	
Intron 15	11862198	DelG		0.07	0.14	
	11862564	T>G		<0.01	0	
	11862577	T>G		<0.01	0	
Intron 16	11872591	T>C		<0.01	0	
Exon 17	11872773	G>C	G750G	0.01	0	
Intron 17	11872940	G>A		<0.01	0	
Exon 20	11882248	G>A	P851P	0	0.016	Novel3
3`UTR	11883265	C>T		0.23	0.3	rs1050800
	11883450	DelG		<0.01	0	
	11883469	C>T		<0.01	0	
	11883631	C>T		<0.01	0	
	11883712	DelATT		<0.01	0	
	11883768	T>C		0.13	0.17	rs11524
	11883798	C>T		<0.01	0	
	11884301	G>C		<0.01	0	
	11884454	C>G		0.08	0.16	Novel4
	11884670	G>T		0.02	0	
	11884739	T>C		<0.01	0	

Genomic coordinates correspond to NCBI Build 36. Non-synonymous changes are highlighted in red bold. Ins = insertion. Del = deletion.

3.3.2.2 Investigation of the A353T variation

A353T was detected in a female patient with a Pakistani father and British white mother. She presented with clinical features of severe insulin resistance at 8 years old, which worsened with weight gain in the second decade, before improving dramatically with weight loss in adult life. She had no evidence of lipodystrophy. Three methods were used to investigate whether A353T caused disease in the patient. First, I used a web-based program, PANTHER, which uses information on evolutionary sequence conservation to predict whether an amino acid substitution is likely to have a functional impact on the protein. An alanine to threonine change was predicted to have no functional impact on the lipin 1 protein. Figure 3.8 shows ClustalW alignments of orthologous lipin 1 peptide sequences from a variety of organisms. Highly conserved residues are more likely to have been subject to purifying selection and their substitution is more likely to have a functional impact on the protein (Miller and Kumar 2001). The alanine 353 residue is conserved in primates, mouse, rat, frog and chicken (Figure 3.8) but not opossum or zebrafish. Second, to determine if the variant co-segregates with disease in the family, DNA from the patient's mother, maternal aunts, and maternal grandparents was sequenced. Co-segregation analysis demonstrated that the A353T variant did not segregate with the hallmarks of insulin resistance in the family (Figure 3.9). Finally, A353T was also genotyped in 1064 participants of the HGDP panel to check for its presence in unaffected individuals but was not detected (data not shown).

```
H. sapiens          TLGAAAPLLPMI | KKGGRWWFSWRG | AHSTGEQPPQLSLATRVKHE
P. troglodytes      TLGAAAPLLPMI | KKGGRWWFSWRG | AHSTGEQPPQLSLATRVKHE
M. musculus         SLGAAAPPSPVA | KKGGRWWFSWRG | GHNTGEQPAQLGLATRIKHE
R. norvegicus       ALGAAAPPLSVA | KKGGRWWFSWRG | GHNTGEQPAQLGLTTRIKHE
M. domesticus       GQGGGGQALPGA | RKGGRWWFSWRG | GLVIGEQPAKLSIGTRMKEE
G. gallus           SAGAAVPSLPAN | KKGGRWWFSWRG | SRLKGEDSSQMTMANRIKDE
X. tropicalis       SLGAAAPPLPYD | KKGGRWWFSWRG | GPYSGGQPVGSSLENRIKDE
D. rerio            PISVVAH-LITE | KKGGRWWFSWRG | ESIRTG-----SVSSRLKDE
```

Figure 3.8 Multiple sequence alignments (using ClustalW) showing conservation of *LPIN1* amino acids A353, R552, and G582
Straight lines indicate hidden sequence.

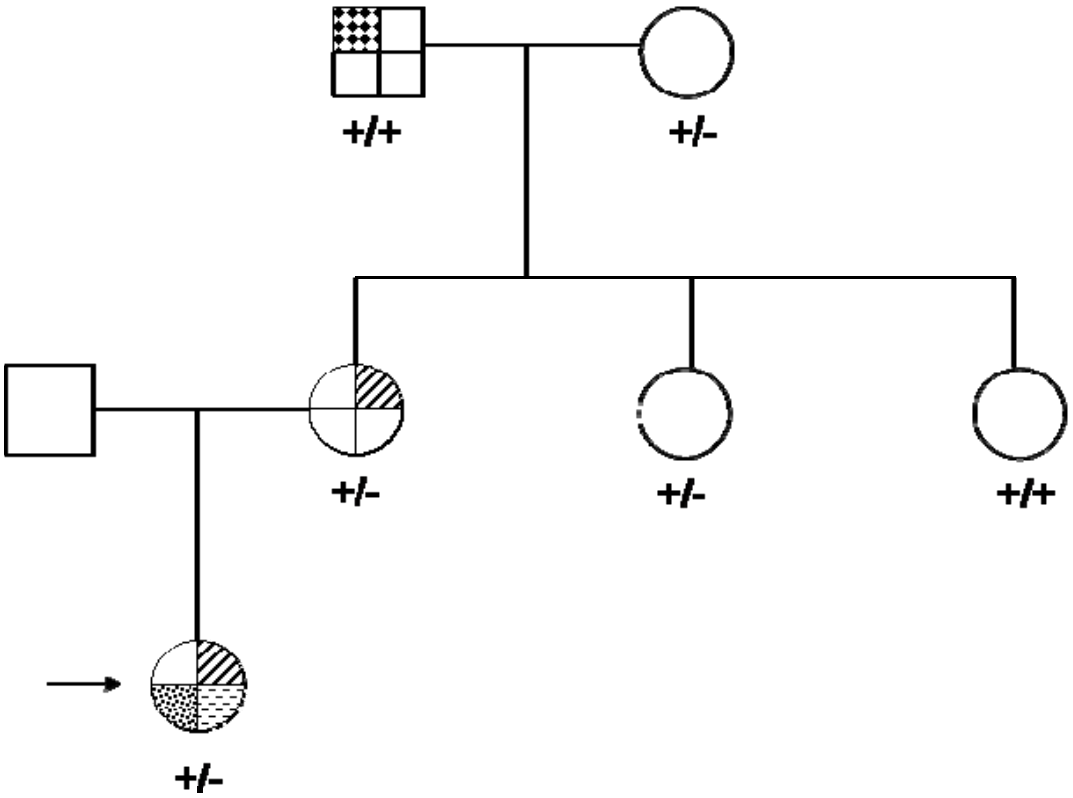


Figure 3.9 Family pedigree demonstrating that the A353T mutation does not segregate with disease in a fully penetrant manner
+/- represents a heterozygous genotype and +/+ represents the wild-type genotype. The patient (indicated by the arrow) has hyperinsulinaemia (diagonal stripes), hirsutism (spots) and acanthosis nigricans (dashes). The grandfather has diabetes (diamonds).

3.3.2.3 Investigation of the R552K variation

R552K was detected as a heterozygous change in two unrelated white European females, but not in 1064 controls from the Diversity Panel. The first proband presented with severe insulin resistance and femorogluteal lipodystrophy at 15 years old. The lipodystrophy progressed to become generalised in conjunction with the appearance of aggressive haemolytic anaemia and autoimmune liver disease. Liver failure led to her demise at 24 years old. The other proband was diagnosed with insulin resistant diabetes at 32 years old, and subsequently required in excess of 4U/day exogenous insulin to maintain satisfactory glycaemic control. She had no clinical evidence of lipodystrophy, and her BMI was sustained above 30 kg/m^2. R552 is within a highly conserved tract (Figure 3.8) and mutation to lysine is predicted by PANTHER to have deleterious effects on lipin 1 function. Family DNA was not available for co-segregation analysis for either patient.

3.3.2.4 Investigation of the G582R variation

G582R was identified as a heterozygous change in a white, European male with a complex syndrome. This included severe insulin resistance and severe, early onset sensorimotor neuropathy which confined him to a wheelchair, a combination highly reminiscent of the combined lipodystrophy/insulin resistance and neuropathy of the fld mouse. This patient also underwent allogeneic bone marrow transplantation in childhood for acute lymphoblastic leukaemia, and had a cerebral cavernous hemangioma. All genomic analyses were undertaken on DNA extracted from cultured skin fibroblasts. G582 is a well conserved residue within the protein (Figure 3.8), and mutation to arginine is predicted by PANTHER to have deleterious effects on lipin 1 function. Co-segregation analysis was performed using DNA from first-degree relatives of the patient (Figure 3.10). The father also carried the variant but although diagnosed with diabetes at age 69 years, he had no peripheral neuropathy, nor clinical or biochemical evidence of insulin resistance/lipodystrophy. This could be

because the wild-type copy of *LPIN1* can compensate for the effects of the G582R. As the patient is also heterozygous for G582R I sequenced patient cDNA to exclude the possibility that the patient only expresses the mutant copy of *LPIN1*. However, the patient's cDNA contained both alleles (Figure 3.11). Subsequently, the G582R variant was genotyped in the HGDP panel and detected in a Bedouin control individual from Nedev, Israel.

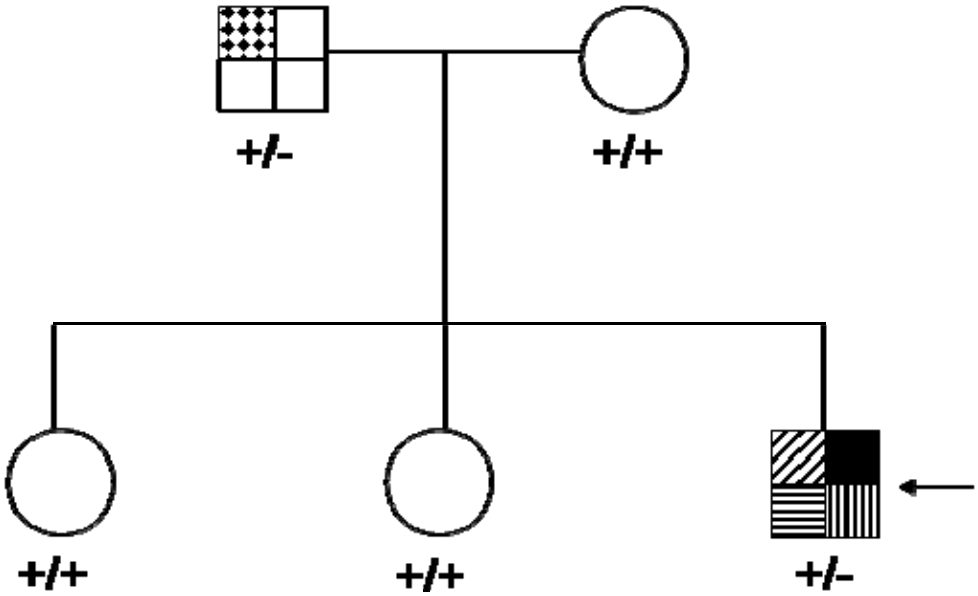


Figure 3.10 Family pedigree demonstrating that the G582R mutation does not segregate with disease in a fully penetrant manner
+/- represents a heterozygous genotype and +/+ represents the wild-type genotype. The patient (indicated by the arrow) has hyperinsulinaemia (diagonal stripes), severe peripheral neuropathy (black), previous bone marrow transplant for AML (horizontal stripes) and an intracerebral cavernous haemangioma (vertical stripes). His father has diabetes (diamonds).

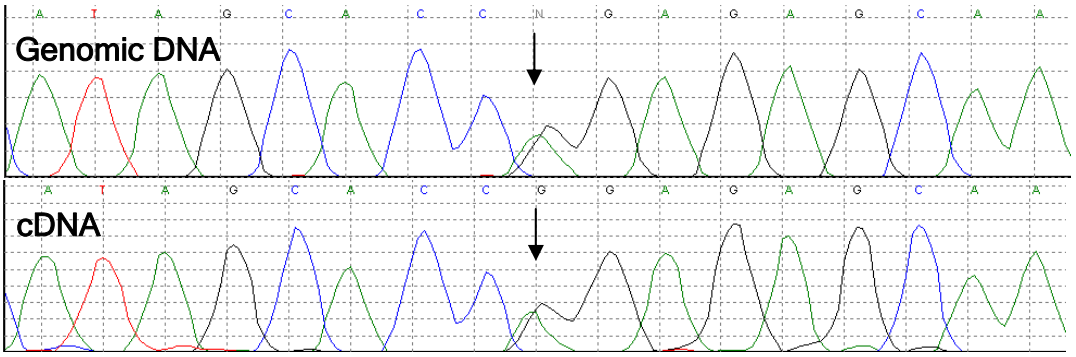


Figure 3.11 Sequences of genomic DNA and cDNA from the patient carrying the G582R variant (indicated by arrow)
This experiment shows that both alleles of G582R are expressed in patient fibroblasts.

In summary, I identified 3 rare missense variants in *LPIN1* in a cohort of patients with extreme insulin resistance. Two of these, G582R and R552K, are predicted to be deleterious to function by the PANTHER algorithm. Furthermore, the proband carrying the G582R variant had a syndrome including severe neuropathy and insulin resistance as seen in the fld mouse. Thus, despite the failure of this variant to co-segregate with the disease phenotype in the wider kindred, and despite the absence of available family members from the R552K kindred, Neil Grimsey and Symeon Siniossoglou at the Cambridge Institute for Medical Research investigated the impact of G582R and R552K on Lipin 1 function in primary skin fibroblasts from the probands. As A353T was predicted to be a benign change based on PANTHER and family co-segregation analysis, and as no fibroblasts were available, this variant was not investigated further.

3.3.2.5 Investigation of R552K and G582R in patient fibroblasts

To investigate the impact of R552K and G582R mutations on lipin 1 protein levels and phosphorylation status, total cell extracts from patient fibroblasts were probed with lipin 1, lipin 2, a nuclear pore marker (Mab414), and lamin B-specific primary antibodies to detect protein levels. This work was carried out by Neil Grimsey and Symeon Siniossoglou at the Cambridge Institute for Medical Research. The resulting

Western blot shown in Figure 3.12 shows similar intensities of all four proteins in patient fibroblasts compared to control cells. Immunocytochemistry was employed to detect changes in nuclear membrane morphology by staining a nuclear pore marker (Figure 3.13). There was no discernable difference in morphology between patient and control fibroblasts.

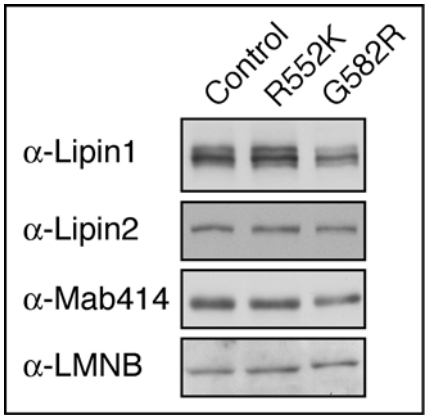


Figure 3.12 Western blot analysis of total cell extracts from cultured patient fibroblasts carrying the R552K and G582R *LPIN1* mutations (lanes 2 and 3 respectively) and control fibroblasts (lane 1)
Data from Neil Grimsey and Symeon Siniossoglou. The blots were probed with primary antibodies specific to lipin 1 (α-Lipin1), lipin 2 (α-Lipin2), a nuclear pore marker (α-Mab414) and lamin B (α-LMNB), and species-specific secondary antibodies. See Materials and Methods for details.

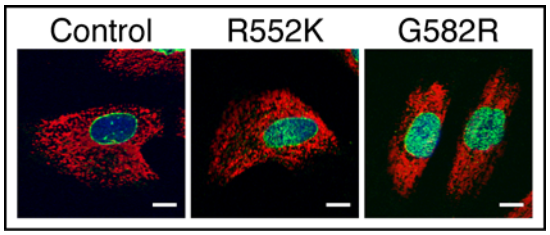


Figure 3.13 Immunofluoresence of control and patient fibroblasts showing Mab414, a nuclear pore marker (green), calreticulin, an ER calcium binding protein (red), and DNA (blue)
Data from Neil Grimsey and Symeon Siniossoglou
Bar = 5μm.

3.3.3 *LPIN2* and *LPIN3* mutation screening in the SIR cohort

3.3.3.1 *LPIN2* variation in SIR

Ninety-two *LPIN2* variants were detected in insulin resistant or lipodystrophic patients (Appendix Table A5), nine of which were present in the coding sequence (Table 3.10 and Figure 3.14). Coding sequence variants that did not alter the amino acid sequence (shown below the schematic in Figure 3.14) and/or that were also present in controls were considered unlikely to be pathogenic. This left two rare nonsynonymous variants (E497K and P626S - underlined above the schematic in Figure 3.14) that did not fall within any known functional domains within *LPIN2*, and were predicted benign by PANTHER. These were investigated no further.

Table 3.10 *LPIN2* coding variants detected in a cohort of severe insulin resistant patients and 11 Indian and 23 CEPH controls

Genic position	Genomic position	Minor|Major allele	Protein consequence	MAF in SIR	Detected in controls?
Exon 5	2930695	G|A	A202A	0.0027	
Exon 5	2930693	A|G	S203F	0.0081	Yes
Exon 10	2919124	T|C	**E497K**	0.0026	No
Exon 13	2916779	G|A	S579P	0.0026	Yes
Exon 14	2915359	T|C	E601K	0.0111	Yes
Exon 14	2915284	A|G	**P626S**	0.0027	No
Exon 16	2913790	A|G	L719L	0.0052	
Exon 20	2910387	G|A	S865S	0.0029	
Exon 20	2910372	A|G	S870S	0.0026	

Genomic coordinates correspond to NCBI Build 36.

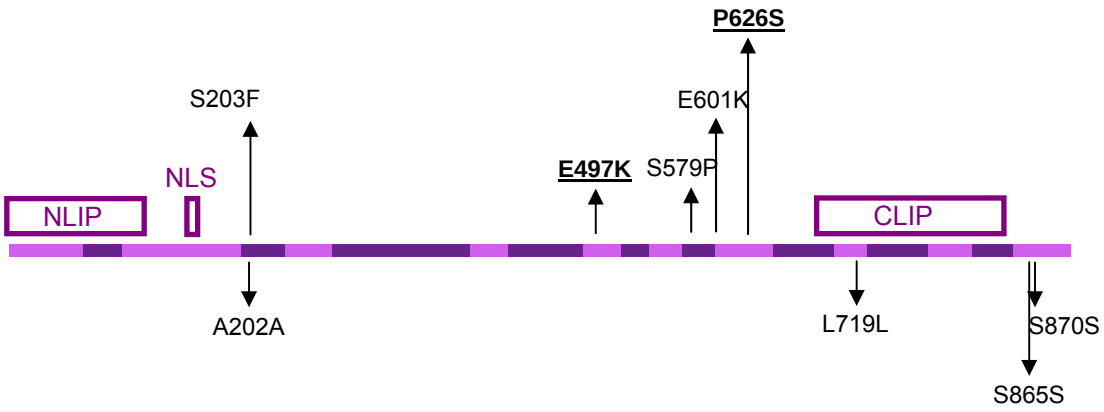


Figure 3.14 Coding *LPIN2* variants in the SIR cohort
Schematic of the lipin 2 protein showing exons in alternating bright and dark purple and known domains among lipin family proteins in boxes. Arrows indicate the location of coding SNPs detected in *LPIN2* by sequencing 23 patients with partial lipodystrophy and 135 patients with other syndromes of severe insulin resistance. Non-synonymous variants are above the schematic and synonymous variants are below the schematic. Non-synonymous mutations absent from control samples were considered potentially pathogenic (underlined). NLIP (amino acids 1-114) = N-terminal lipin domain, NLS (amino acids 153-158) = nuclear localisation signal, and CLIP (amino acids 685-841) = C-terminal lipin domain, also referred to as the LNS2 (Lipin/Ned1/Smp2) domain.

3.3.3.2 *LPIN3* variation in SIR

A total of 54 *LPIN3* variants were detected in the SIR cohort (Appendix Table A6), 17 of which were present in the coding sequence (Table 3.11 and Figure 3.15). Coding sequence variants that did not alter the amino acid sequence (shown below the schematic in Figure 3.15) and/or that were also present in controls were considered unlikely to be pathogenic. This left three rare nonsynonymous variants (G41S, W110C and E539K - underlined above the schematic in Figure 3.14), two of which fell within the highly conserved N-terminal domain (residues 1-114) common to all three members of the lipin family.

Table 3.11 *LPIN3* coding variants detected in a cohort of severe insulin resistant patients and 11 Indian and 23 CEPH controls

Genic position	Genomic position	Minor|Major allele	Protein consequence	MAF in SIR	Detected in controls?	rs ID
Exon 2	39407866	T|C	Y3Y	0.0027		
Exon 2	39407896	T|G	G13G	0.0026		rs16985673
Exon 2	39407916	A|G	R20Q	0.0132	Yes	
Exon 2	39407947	T|C	G30G	0.0026		
Exon 2	39407978	A|G	**G41S**	0.0026	No	
Exon 2	39407998	G|C	P47P	0.0165		
Exon 2	39408038	T|C	R61W	0.0239	Yes	
Exon 4	39410714	C|G	**W110C**	0.0027	No	
Exon 7	39412412	C|G	V355L	0.0161	Yes	
Exon 8	39413957	A|G	A395A	0.0026		
Exon 9	39414271	T|C	L454L	0.0052		
Exon 12	39416786	A|G	**E539K**	0.0026	No	
Exon 12	39416791	T|G	E540D	0.0026	Yes	
Exon 17	39419954	A|G	L686L	0.0315		rs2072969
Exon 17	39419963	A|G	S689S	0.0026		rs41277020

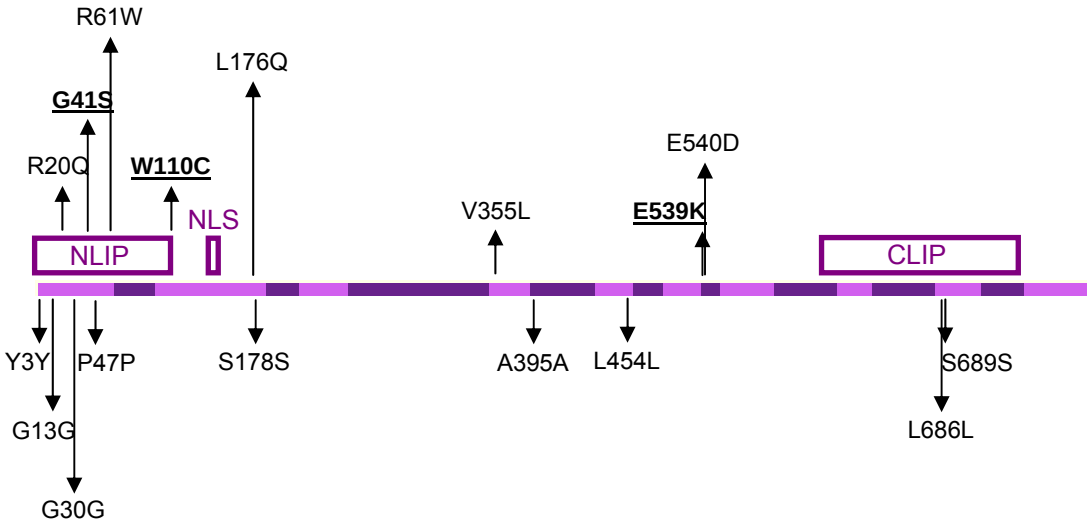


Figure 3.15 Coding *LPIN3* variants in the SIR cohort
Schematic of the lipin 3 protein showing exons in alternating bright and dark purple and known domains among lipin family proteins in boxes. Arrows indicate the location of coding SNPs detected in *LPIN3* by sequencing 23 patients with partial lipodystrophy and 135 patients with other syndromes of severe insulin resistance. Non-synonymous variants are above the schematic and synonymous variants are below the schematic. Non-synonymous mutations absent from control samples were considered potentially pathogenic (underlined). NLIP (amino acids 1-114) = N-terminal lipin domain, NLS (amino acids 141-148) = nuclear localisation signal, and CLIP (amino acids 640-796) = C-terminal lipin domain, also referred to as the LNS2 (Lipin/Ned1/Smp2) domain. HAD = haloacid dehalogenase domain.

3.3.3.3 Investigation of *LPIN3* G41S and W110C variants

I decided to prioritise those *LPIN2* and *LPIN3* variants most likely to have a functional impact upon the protein for further investigation. *LPIN3* non-synonymous variants G41S and W110C were selected as both these variants were present in single severe insulin resistant individuals, absent from controls, and mapped within a known functional domain. Both were also predicted to be highly likely to have a deleterious effect on protein function by PANTHER.

G41S was detected in an Asian female with partial lipodystrophy. The mutation was not present in 47 Indian and 47 white European controls on the CIN panel (Appendix Table 4) (data not shown). Unfortunately, family DNA was not available preventing further co-segregation analysis.

W110C was detected in an Arabic female with pseudoacromegaly and acanthosis nigricans. To investigate whether this variant was present in any unaffected individual of the same ethnic background, *LPIN3* exon 4 was sequenced in 173 Arabic controls from the HGDP panel. W110C was detected in one Druze control from Carmel, Israel, and is therefore unlikely to be pathogenic (data not shown).

3.4 Discussion

3.4.1 *LPIN1*

In this study I performed a comprehensive analysis of *LPIN1* variants and their effects on metabolic quantitative traits and syndromes of insulin resistance (including lipodystrophy). Analysis of *LPIN1* polymorphisms (MAF>0.01) in two UK population-based cohorts (N = 4610) revealed nominal significant associations with BMI, and rs13412582 remained marginally associated with BMI (P = 0.042) in a meta-analysis of UK population-based samples from in-house and publicly available genome-wide studies (N = 8504). I also detected nominal associations between my tagSNPs and metabolic traits previously reported to be associated with *LPIN1* variation (Wiedmann et al. 2007). Sequencing of 23 patients with lipodystrophy and 135 patients with syndromes of insulin resistance revealed that mutations in *LPIN1* are unlikely to be a common cause of these diseases in humans. This study has been published (Fawcett et al. 2008).

To my knowledge, neither rs13412582 nor any highly correlated SNPs have been tested in other association studies published to date. Further replication in larger cohorts will be required to confirm the BMI association.

Seven of my tagSNPs that passed quality control were directly genotyped in at least one of the other association studies (Loos et al. 2007; Suviolahti et al. 2006; Wiedmann et al. 2007). All analyses, including my own, found no association between rs4669781, rs1050800, and rs2577256 and insulin levels and measures of adiposity. However, results for the other four SNPs are inconsistent between studies. For example, rs2716610 was associated with obesity in a Finnish case-control population (Suviolahti et al. 2006). In contrast, I detected no association between rs2716609, which is in complete LD with rs2716610 (r^2 = 1) in HapMap CEU trios,

and obesity based on analysis of 1128 obese and 3601 lean individuals from Ely and Hertfordshire population-based cohorts (OR = 1.11 ± 0.08, P = 0.149).

SNP rs2716610 was also associated with BMI in lean Finnish men (Suviolahti et al. 2006). This association with quantitative measures of adiposity is supported by the Quebec Family study (Loos et al. 2007), which analysed parents (N = 335) and offspring (N = 377) from 172 French-Canadian families. Here, the highly correlated SNP rs2716609 was associated with skinfolds and waist circumference, and BMI showed the same trend. Given my sample size of 4130 individuals with full rs2716609 genotype and BMI data I had >80% power to detect the effect size observed in the Quebec Family study. Nevertheless, I did not replicate the association between rs2716609 and BMI in Ely and Hertfordshire cohorts (Table 3.3). Nor did I replicate the association with waist circumference, though the direction of the effect was consistent with the Quebec study (data not shown). My results agree with the MONICA study Augsburg (N = 1416), a German population-based cohort, which found no association between rs2716610 and BMI in men or women (Wiedmann et al. 2007).

Two other SNPs, rs893346 and rs2577262, were associated with BMI in lean Finnish men (Suviolahti et al. 2006) but showed no statistical association with BMI in 1873 lean men from Ely and Hertfordshire cohort studies (P = 0.631 and 0.253 respectively). Similarly, rs2278513 and rs2577262 were associated with BMI in obese Finnish men but not in obese men from the UK (P = 0.780 and 0.676 respectively). My data agree with the MONICA study which found no association of rs893346 and two SNPs highly correlated with rs2577262 in HapMap CEU trios (r^2 = 1.0 and 0.96 for rs6744682 and rs6708316 respectively) with BMI in men (Wiedmann et al. 2007).

In the Quebec Family study, rs2577262 and rs2577256 were associated with resting metabolic rate (RMR) in parents but not offspring. The MRC Ely study included data for mean resting energy expenditure which was not statistically

associated with rs2577262 or rs2577256 (P = 0.895 and 0.923 respectively) despite having >80% power to detect the magnitude of the effect size described previously.

The MONICA study reported strong associations between haplotypes of markers rs33997857, rs6744682 and rs6708316 and quantitative traits underlying metabolic syndrome including hypertension-, obesity-, and diabetes-related traits (Wiedmann et al. 2007). Several of the traits were also statistically associated with the same haplotypes in a replication study, but the effect was always in the opposite direction compared to the original cohort. I tested all my individual tagging SNPs as well as haplotypes of rs33997857 and rs2577262, which is highly correlated with rs6744682 and rs6708316, for association with systolic and diastolic blood pressure, HDL and LDL cholesterol, plasma triglycerides, waist circumference, HbA1c levels, and hypertension and diabetes status. I detected a number of nominal associations between these traits and individual SNPs in my study (Table 3.6, 3.7 and 3.8) but none of these associations reached statistical significance after adjustment of the P value threshold for multiple testing using the Bonferroni correction and thus require further replication.

There are several possible reasons why I could not replicate some previously reported associations between *LPIN1* variants and metabolic quantitative traits. Firstly, I may have reported false negative results. However, where effect sizes were reported in previously published studies I was able to calculate that my study had >80% power to detect them. Secondly, previous studies might have reported false positive results. In particular, as a consequence of multiple testing, detection of false positive associations becomes more likely when analyses are performed in subsets of samples and on many traits. Alternatively, the discrepancy in results between studies may be due to genetic and/or environmental differences between the populations genotyped. For example, the degree of linkage disequilibrium between *LPIN1* tag SNPs and the putative unmeasured true functional variant(s) may vary

between the cohorts. Also, *LPIN1* SNPs may interact with other genetic and/or environmental risk factors in different studies. Therefore, I cannot rule out the possibility of population-specific effects of *LPIN1* genotype on metabolic quantitative traits and hypertension, diabetes and obesity risk.

In the fld mouse model Lpin1 null mutations cause lipodystrophy, insulin resistance and peripheral neuropathy (Peterfy et al. 2001). However, of the three rare (MAF<0.01) nonsynonymous *LPIN1* variants detected within the cohort of patients with syndromes of severe insulin resistance, none are likely to be pathogenic in isolation in heterozygous form: family co-segregation analysis showed that A353T and G582R did not segregate with disease in a fully penetrant manner and G582R was also detected in one Bedouin control.

Western blotting of patient fibroblasts showed that G582R and R552K had no discernable impact on lipin 1 protein levels. Proteins orthologous to lipin 1 in yeast are proposed to be involved in nuclear membrane growth and morphology (Santos-Rosa et al. 2005; Siniossoglou et al. 1998; Tange et al. 2002). However, staining of a nuclear pore marker in patient fibroblasts with R552K and G582R variants revealed no abnormalities in membrane morphology compared to control fibroblasts.

To date, my study (N=23) and previously published work (N=15) (Cao and Hegele 2002) have demonstrated that *LPIN1* coding mutations are unlikely to be a common cause of human lipodystrophy. However, I cannot exclude the possibility that *LPIN1* mutations are rarer causes of these disorders or that rare variants in *LPIN1* interact with other genetic defects to cause severe insulin resistance. The methods I used to screen for mutations would not have detected copy number variations affecting large regions, such as whole exon deletions and duplications nor would they detect potential mutations affecting regulatory regions, therefore I cannot exclude these types of *LPIN1* variation as causes of human lipodystrophy and insulin resistance.

Furthermore, the *in vitro* assays used to assess the functional impact of *LPIN1* non-synonymous variants might have missed some functional effects, such as phosphatidic acid phosphatase (PAP) activity.

In conclusion, coding variants in *LPIN1* are not a common cause of lipodystrophy and severe insulin resistance in humans, and polymorphisms in *LPIN1* are unlikely to importantly contribute to insulin sensitivity in UK populations. SNP rs13412852 is nominally associated with BMI in UK cohorts but the effect size is very modest and requires confirmation. Nominal associations between *LPIN1* variants and blood pressure, cholesterol, triglycerides, HbA1c, and risk of hypertension need replicating in larger cohorts.

3.4.2 *LPIN2* and *LPIN3*

As far as I am aware I have performed the first screening of *LPIN2* and *LPIN3* for potentially pathogenic mutations in patients with syndromes of severe insulin resistance. Two rare, non-synonymous variants were detected in the N-terminal conserved lipin domain of *LPIN3*. One of these, W110C, was considered unlikely to be pathogenic as it was present in a Druze control. However, G41S is present in a patient with partial lipodystrophy and severe insulin resistance without family available for co-segregation analysis. The role of this mutation, if any, in the syndrome will need to be investigated further.

Two non-synonymous variants present in patients with syndromes of insulin resistance and absent from control samples were detected in *LPIN2*. However, these were not prioritised for further analysis as they were predicted to be benign changes and did not fall within any known functional domains. In future work more controls could be tested for these variants. If absent from controls and if DNA from family members is available, co-segregation analysis could be carried out.

Given that my sequencing approach would not detect large insertions and deletions, I cannot exclude the possibility that copy number variants affecting *LPIN2* and *LPIN3* impact insulin resistance. It is also plausible that combinations of rare variants in these genes cause syndromes of insulin resistance. Furthermore, I did not screen putative regulatory regions and therefore may have missed pathogenic mutations affecting *LPIN2* and *LPIN3* gene regulation.

In my study I did not evaluate the role of common variants in *LPIN2* and *LPIN3* on human metabolic traits. In this context it would be of interest to design a study to attempt replication of the recently published association between rs3745012 in the 3'UTR of *LPIN2* (Aulchenko et al. 2007) and type 2 diabetes and related metabolic traits. This SNP has not been genotyped in HapMap samples and is not present on Affymetrix or Illumina SNP chips, therefore genome-wide association study data will not provide enough data to replicate this association. Still, other SNPs within *LPIN2* and *LPIN3* are available on genome-wide arrays and may not need to be genotyped in extra cohorts.

In conclusion, coding variants in *LPIN2* and *LPIN3* are unlikely to be a common cause of lipodystrophy and severe insulin resistance in humans. A potentially pathogenic variant in *LPIN3* requires further investigation to establish whether it is causing insulin resistance and lipodystrophy in the patient.

3.5 Materials and Methods

3.5.1 Description of cohorts

Brief descriptions of cohorts used in this study follows. For more details see the corresponding sections of Chapter 2.

3.5.1.1 ELY Cohort

The Medical Research Council (MRC) Ely Study is a population-based cohort study of the aetiology and pathogenesis of type 2 diabetes and related metabolic disorders in the UK (Wareham et al. 1998). It comprises white men and women aged 35-79 years without diagnosed diabetes. Measurements of anthropometric and metabolic data analysed in this study have been described previously (Ekelund et al. 2007).

3.5.1.2 Hertfordshire Cohort

The Hertfordshire Cohort Study was recruited from the cohort of people born in Hertfordshire between 1931 and 1939. The cohort details and measurements of metabolic traits analysed in this study have been described previously (Syddall et al. 2005).

3.5.1.3 EPIC-Obesity Study

The EPIC-Obesity study is nested within the EPIC-Norfolk study, a population based cohort study of 25663 white European men and women aged 39-79 recruited in Norfolk, UK between 1993 and 1997 (Day et al. 1999). Height and weight were measured using standard anthropometric techniques (Day et al. 1999). For *LPIN1*, I analysed BMI data from a random subset of the entire EPIC-Norfolk study (N = 2415) with genome-wide association study data (see 3.5.4).

3.5.1.4 HGDP-CEPH

The HGDP-CEPH Human Genome Diversity Cell Line Panel is a resource of 1064 DNA samples from individuals distributed around the world and has been described previously (Cann et al. 2002).

3.5.1.5 CEPH

48 unrelated individuals from CEPH families supplied by Coriell Cell Repositories (Dausset et al. 1990) are control individuals of North and West European origin.

3.5.1.6 CIN

This panel includes DNA samples from the European Collection of Cell Cultures (ECACC), including 47 white European individuals and 47 individuals of Asian Indian origin.

3.5.1.7 Severe Insulin Resistance Cohort

All patients had severe insulin resistance, defined as fasting insulin > 150 pmol/l, or peak insulin on oral glucose tolerance testing > 1,500 pmol/l in non-diabetic patients. In complete insulin deficiency it was defined as an insulin requirement above 3U/kg/day. Acanthosis nigricans was also used as a marker of insulin resistance. Most patients had a BMI <30 kg/m^2 and at least 58 had BMI>30.

3.5.2 PCR and sequencing

Genomic DNA from 23 patients with lipodystophy, 135 patients with other syndromes of severe insulin resistance and, in the case of *LPIN2* and *LPIN3* sequencing, 11 Indian controls from the CIN panel and 23 CEPH controls, was whole-genome amplified (Chapter 2.3.1.1) prior to amplification with gene-specific primers covering all coding exons, splice junctions, and 3'UTR (see Appendix Table A7 for sequences and cycling conditions).

PCR was performed using standard conditions (Chapter 2.3.2) and products purified using exonuclease I and shrimp alkaline phosphatase (USB Corporation, Cleveland, OH, USA).

Bi-directional sequencing was performed using the Big Dye Terminator 3.1 kit (Applied Biosystems, Foster City, CA, USA). Sequencing reactions were run on ABI3730 capillary machines (Applied Biosystems) and analysed using Mutation Surveyor version.2.20 (SoftGenetics LLC, State College, PA, USA) (Chapter 2.3.5). All non-synonymous variants with MAF<0.01 were confirmed in a second PCR and sequencing reaction using patient genomic DNA. DNA from family members used for co-segregation analysis was genomic.

3.5.3 Tagging SNP selection

Tagging SNPs were selected from 52 known and novel *LPIN1* SNPs with a pairwise $r^2 \geq 0.8$ using Tagger (de Bakker et al. 2005) as a stand-alone program in Haploview (Barrett et al. 2005). Twenty-five SNPs were selected as tags and one was force included, giving a total of 26 SNPs for association testing.

3.5.4 Genotyping

Genotyping was performed by Susannah Bumpstead and Andrew Keniry in the genotyping facility within the Genetics of Complex Traits in Humans team at the Wellcome Trust Sanger Institute. *LPIN1* mutations A353T, R552K, and G582R were genotyped on the HGDP-CEPH Human Genome Diversity panel as stand-alone assays using the Sequenom MassArray hME platform according to the manufacturer's instructions (see Chapter 2 for details and Appendix Table A8 for primer sequences). Twenty-six SNPs were selected for genotyping in the MRC Ely population-based cohort (N = 1709) and the Hertfordshire cohort study (N = 2901) using the Sequenom MassArray iPLEX platform (Chapter 2.3.7.1.2). One SNP, rs13093930, failed Sequenom assay design but because it was a rare intronic SNP

that only tagged itself a replacement could not be designed. This left 25 tagging SNPs (Figure 3.5 red boxes). Twenty-three of these tagging SNPs passed pre-screening and all except the monomorphic loci Novel02 and rs17603755 (which did not pass HWE) were analysed for association with metabolic traits (Table 3.2). Primers, probes and conditions for the remaining twenty-two *LPIN1* tagging SNPs are presented in Appendix Table A8. All genotyped SNPs were tested for deviation from Hardy-Weinberg Equilibrium ($P > 0.01$) and call rates > 85%.

Genotyping of the EPIC-Obesity study using the Affymetrix GeneChip Human Mapping 500K Array Set has been described (Loos et al. 2008). This study contained genotype information for five of my *LPIN1* tagSNPs, each of which had a call rate >90%. In total, 2415 individuals with height and weight measures and quality-controlled genotype data were available for analyses.

3.5.5 Statistical analysis

Deviation of *LPIN1* tagSNP genotype from Hardy-Weinberg equilibrium was assessed using a goodness-of-fit χ^2 test. Linear regression analysis was used to assess the association between individual SNPs and BMI, log-transformed fasting plasma insulin, and log-transformed additional metabolic traits (systolic and diastolic blood pressure, HDL and LDL cholesterol, plasma triglycerides, waist circumference, and HbA1c levels) in Ely and Hertfordshire cohorts using Stata v9 (Stata Corporation, Texas, USA). All analyses were adjusted for age and sex and, in the case of the joint analysis, study. Logistic regression in Stata was used to test for association between *LPIN1* SNPs and risk of hypertension and diabetes. Chi-squared analysis was performed to test for significant differences ($P<0.01$) in call rate between cases and controls. Where nominally significant values were found in the Ely and Hertfordshire joint analysis, I performed 10,000 permutations of the dependent variable to test for empirical significance. This was performed in Stata.

The joint Ely and Hertfordshire cohort analysis of additional traits underlying metabolic syndrome comprised 189 tests so the *P* value threshold adjusted for multiple testing using the Bonferroni correction is 0.0003.

Fixed effect meta-analysis using the inverse variance method was performed by using the metan command in Stata (Bradburn et al. 1999). Heterogeneity among studies was assessed using the Q statistic. IMPUTE software (http://www.stats.ox.ac.uk/~marchini/software/gwas/impute.html) was used to impute genotypes for rs17603420 in the EPIC cohort.

Plink (http://pngu.mgh.harvard.edu/~purcell/plink/) was used to perform hapotype analysis (Purcell et al. 2007), and Ely and Hertfordshire cohorts were meta-analysed using METAL (http://www.sph.umich.edu/csg/abecasis/metal/index.html).

Power calculations were performed using Quanto v1.1.1. I calculated that we have >80% power to detect a per allele effect on BMI of >1.33 kg/m2 with MAF=0.01, and >0.27kg/m2 with MAF=0.5. For logged fasting insulin data this range is >1.04 to >1.22.

3.5.6 Western blotting

The following analyses of patient fibroblasts carrying the *LPIN1* mutations R552K and G582R mutations were performed by Neil Grimsey and Symeon Siniossoglou at the Cambridge Institute for Medical Research.

Patient fibroblast cells were maintained in DMEM medium supplemented with 10% fetal bovine serum (FBS) and 2mM L-glutamine in a humidified 37°C incubator with 5% CO_2. All cells were routinely assessed for and protected against mycoplasma

infection using VenorGeM® mycoplasma detection kit (Minerva biolabs, CamBio, VGM-025) and BM-cyclin (Roche; 799050) respectively.

Fibroblasts were collected by trypsin EDTA release, washed with PBS and then lysed in a 50mM HEPES pH7.4 buffer containing 150mM NaCl, 1% Triton X-100, 100µM AEBSF, Protease inhibitor cocktail 1, Phosphatase inhibitor cocktail II, 1µg/ml Dnase and 4mM MgCl2 chilled to 4°C. Each sample was homogensised by passing through a 25G needle 10 times. Insoluble debris was removed by a 16,000g centrifugation step at 4°C.

Sample protein concentration was measured by a comparative Bradford protein assay. Samples were then suspended in 1 x SDS sample buffer and boiled at 95°C for 5 minutes, loaded onto 7% SDS-PAGE, transferred onto nitrocellulose, and blocked in PBS with 1% TX-100 and 5% milk. These were then probed with protein specific primary antibodies: Lipin 1, Lipin 2, anti-Mab414 (Covance; MMS-120P), anti-laminB (Santa-Cruz; sc-6217). Lipin 1 and 2 polyclonal antibody production will be described elsewhere (Grimsey and Siniossoglou, in preparation). This was followed by species specific secondary antibodies coupled to Horse Radish Peroxidase (HRP): anti-rabbit IgG (Jackson immuno research; 211-032-171), anti-goat IgG (Novus Biologicals; NB 710-H), anti-mouse IgG (H & L) highly cross-adsorbed (Molecular probes; A11029). Proteins were then detected using standard electrochemiluminescence techniques (Amersham ECL-reagents).

3.5.7 Indirect immunofluorescence by confocal microscopy

The following analyses were performed by Neil Grimsey and Symeon Siniossoglou at the Cambridge Institute for Medical Research.

Fibroblasts were fixed with 3% Formaldehyde, permiablised with 0.1% Triton X-100, and blocked with 1mg/ml BSA in PBS. Each cover slip was labelled with primary mouse $-\alpha$Mab414 (nuclear pore marker) and secondary anti-mouse conjugated to FITC (green), primary rabbit $-\alpha$Calreticulin (Endoplasmic reticulum calcium binding protein) (Calbiochem; 208910), and secondary anti-rabbit conjugated to Alexa fluor 594 (red) (Molecular probes; A11037), DNA was stained with DAPI (blue). Each slide was mounted onto glass slides and then visualised with a 63X or 100X Plan Apochromat objective (numerical aperture,1.4) on a Ziess Axiovert 200M inverted microscope with an LSM 510 confocal laser Scanning attachment.

Chapter 4

Investigation of components of mTORC1 and mTORC2 complexes in severely insulin resistant patients

4.1 Summary

Components of the mammalian target of rapamycin (mTOR) pathway are key players in insulin signalling. mTOR exists in two distinct complexes: mTORC1, which contains mTOR, Raptor, and GβL, and mTORC2, which contains mTOR, Rictor, GβL, and MAPKAP1. Hyperactivation of mTORC1 and its downstream effectors is associated with serine phosphorylation of insulin receptor substrates and inhibition of insulin signalling. mTORC2 is required for insulin-stimulated phosphorylation of AKT, which goes on to repress AS160 to promote GLUT4-mediated glucose uptake. The aim of this study was to screen members of the mTOR pathway in human syndromes of insulin resistance for potentially pathogenic mutations. I sequenced exons, exon-intron boundaries and 3'UTR of *mTOR*, *Raptor*, *Rictor*, *GβL*, *MAPKAP1*, and *AS160* in 158 patients with syndromes of severe insulin resistance, as well as 11 Indian and 48 European controls. I detected 12 rare non-synonymous variants across all genes, 11 of which were considered unlikely to be pathogenic in a fully penetrant manner. However, a nonsense mutation in *AS160*, R636X, co-segregated with high peak-to-fasting insulin levels in a family with five affected individuals. The truncated protein significantly reduced insulin-stimulated GLUT4 translocation to the cell surface *in vitro* and was able to bind full length AS160, providing a possible mechanism for a dominant negative effect. These studies suggest that mutations in genes encoding components of mTOR complexes are not common causes of severe insulin resistance, whereas a stop mutation in *AS160* causes an inherited defect in insulin-stimulated glucose transporter translocation (Satya et al., manuscript submitted).

4.2 Introduction

4.2.1 The mammalian target of rapamycin

The mammalian target of rapamycin (mTOR) (also known as FKBP-rapamycin associated protein (FRAP1), rapamycin and FKBP target (RAFT), and rapamycin target (RAPT)) is an evolutionarily conserved large serine/threonine protein kinase belonging to the phosphatitidylinositol kinase-related kinase (PIKK) family (Abraham 2004). It is a central component of a complex signalling network that regulates cell growth and development in response to nutritional, hormonal, energy- and stress-related cues (Sarbassov et al. 2005a; Wullschleger et al. 2006). TOR genes were first identified in yeast during a screen for mutations conferring resistance to the growth inhibitor, rapamycin (Heitman et al. 1991). A complex of rapamycin and cofactor, FKBP12, bind mTOR at its FKBP-rapamycin binding (FRB) domain and inhibit its kinase activity.

4.2.2 mTOR complexes

mTOR exists in at least two distinct complexes, mammalian target of rapamycin complex 1 (mTORC1) and 2 (mTORC2), both of which play roles in insulin signalling (Figure 4.1). In mTORC1, mTOR is accompanied by the regulatory associated protein of TOR (Raptor) and GβL (also known as mLST8), both of which are required for mTORC1 signalling. The mTORC2 complex comprises mTOR, rapamycin-insensitive companion of TOR (Rictor), GßL, and mitogen-activated protein kinase-associated protein 1 (MAPKAP1 (also known as SIN1)). Rictor is a large protein with no obvious catalytic motifs and MAPKAP1 is a member of the poorly conserved stress kinase-interacting protein 1 (SIN1) family. Both are essential for mTORC2 assembly and function. Though this complex is resistant to acute exposure to rapamycin, long-term administration can block its assembly. This probably occurs as a result of decreased availability of free mTOR to bind components of mTORC2 (Barilli et al. 2008; Frias et al. 2006; Sarbassov et al. 2006; Varma et al. 2008).

4.2.3 The role of mTORC1 in insulin signaling

As decribed in Chapter 1, mTORC1 is an important mediator of insulin and nutrient-stimulated protein synthesis and cell growth (Figure 4.1). Another critical kinase within the insulin signalling pathway, AKT, indirectly activates mTORC1 by inhibiting the tuberous sclerosis protein complex (TSC1-TSC2) (reviewed in (Wullschleger et al. 2006)). TSC1-TSC2 negatively regulates mTORC1 by acting as a guanine triphosphatase (GTPase) activating protein (GAP) for Ras homolog enriched in brain (Rheb), which is able to activate mTORC1 activity in its GTP-bound form (Manning and Cantley 2003; Tee et al. 2003). mTORC1 is then able to activate mRNA translation through phosphorylation of p70 ribosomal S6 kinase (S6K1) and the eukaryotic translation initiation factor (eIF4E)-binding protein (4E-BP1). Recent studies have shown that activation of the mTORC1/S6K1 pathway is involved in a negative feedback mechanism that inhibits insulin signal transduction (Haruta et al. 2000; Tremblay and Marette 2001).

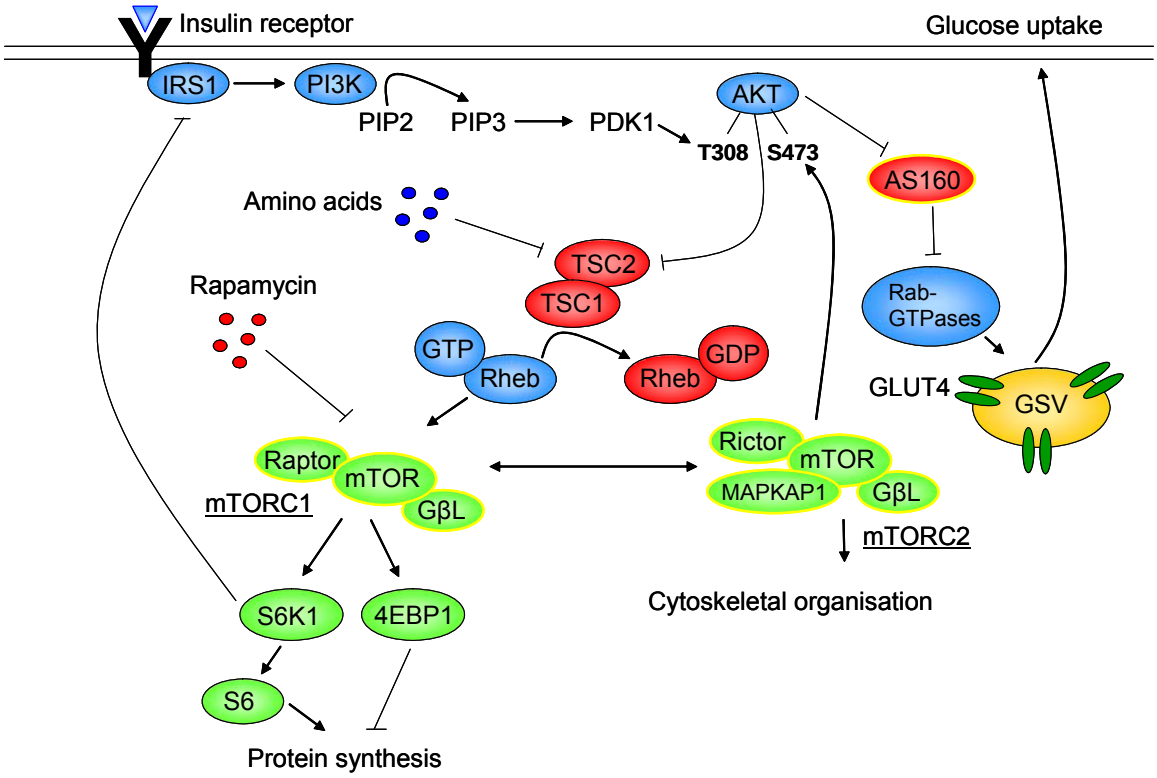


Figure 4.1 Insulin signalling through mTORC1 and mTORC2
Insulin binding to the insulin receptor in target cells triggers a signalling cascade leading to activation of AKT. Full activity of AKT requires phosphorylation by PDK1 on T308 and by mTORC2 on S473. AKT mediates insulin-induced stimulation of mTORC1 activity and protein synthesis, and prevents inhibition of glucose uptake by AS160. mTORC1 substrate S6K1 participates in a negative feedback loop to attenuate further insulin signalling. GSV = GLUT4 storage vesicle.

4.2.4 mTORC1/S6K1 pathway augments serine phosphorylation of IRS proteins

Serine phosphorylation of IRS proteins has emerged as a key event underlying the induction of insulin resistance (Zick 2001). There are several mechanisms by which serine phosphorylation inhibits IRS protein activity, including inducing release of IRS proteins from intracellular complexes that hold them in close proximity to the insulin receptor (Heller-Harrison et al. 1995), and inducing their degradation (Pederson et al. 2001). Activation of the mTORC1/S6K1 pathway by insulin or amino acids has been shown to augment serine phosphorylation on IRS1 and thereby suppress insulin signalling in myocytes and adipocytes (Haruta et al. 2000; Takano et al. 2001; Tremblay et al. 2005; Tremblay and Marette 2001). These effects were blocked by rapamycin treatment. Constitutively active mTORC1/S6K1 as a result of disruption of

TSC1/TSC2, or ectopic expression of Rheb, also caused phosphorylation and downregulation of insulin receptor substrates and insulin resistance *in vitro* (Harrington et al. 2004; Shah et al. 2004).

Both mTOR and S6K1 have been shown to directly catalyse serine phosphorylation of IRS1 *in vitro* (Harrington et al. 2004; Ozes et al. 2001). Conversely, siRNA-mediated knockdown of S6K1 decreased IRS1 serine phosphorylation and increased insulin-induced AKT phosphorylation (Khamzina et al. 2005; Um et al. 2004) and mice deficient for S6K1 or resistant to phosphorylation of its substrate, the 40S ribosomal protein S6, exhibited greater insulin sensitivity (Pende et al. 2000; Ruvinsky et al. 2005; Um et al. 2004). These mice were also glucose intolerant due to reduced pancreatic beta-cell size and consequent hypoinsulinaemia, indicating a role for S6K1 in pancreatic beta-cell growth.

4.2.5 mTORC1/S6K1 provides a mechanism for diet-induced insulin resistance

Given the nutrient-sensing role of the mTORC1/S6K1 it is plausible that enhanced stimulation of this pathway might cause insulin resistance under physiological conditions of over-nutrition, such as obesity. Indeed, enhanced mTOR/S6K1 activation, IRS1 serine phosphorylation, and reduced AKT activation have been observed in wild-type mice fed on a high fat diet and in rodent models of obesity (Khamzina et al. 2005; Um et al. 2004). In contrast, S6K1-deficient mice are also protected against diet-induced obesity and insulin resistance (Um et al. 2004). Furthermore, in humans, mTOR/S6K1 activation by amino acid infusion causes IRS1 serine phosphorylation and insulin resistance in a rapamycin-sensitive manner (Krebs et al. 2007; Tremblay et al. 2005).

4.2.6 mTORC1/S6K1 signaling and β-cell function

IRS2, the most abundant and functionally important member of the IRS family in pancreatic beta-cells, is also down-regulated by mTOR/S6K1 signaling (Briaud et al. 2005; Shah et al. 2004). Chronic exposure of pancreatic beta-cells to glucose or

IGF1 increased serine/threonine phosphorylation and proteasomal degradation of IRS2 leading to insulin resistance and apoptosis (Briaud et al. 2005). These effects were blocked by rapamycin suggesting that inhibition of mTORC1/S6K1 might alleviate insulin resistance and improve β-cell survival under conditions of chronic hyperglycaemia. However, long-term rapamycin treatment actually increased insulin resistance in kidney transplanted patients (Di Paolo et al. 2006) and in diabetic *P. Obesus* Israeli sand rats (Fraenkel et al. 2008). The latter also experienced a reduction in beta-cell survival and glucose-stimulated insulin secretion and biosynthesis, which interfered with compensatory insulin secretion to overcome insulin resistance and consequently worsened hyperglycaemia. The discrepancy between the effects of rapamycin in these and previous studies could reflect the difference in metabolic backgrounds of subjects, and the duration of rapamycin treatment. Interestingly, there is evidence to suggest that long-term rapamycin administration inhibits the function of the mTORC2 complex which, as described next, also plays a role in insulin signaling.

4.2.7 mTORC2 plays a role in cytoskeletal organisation and AKT phosphorylation

mTORC2 plays a critical role in organisation of the actin cytoskeleton (Jacinto et al. 2004; Sarbassov et al. 2004) and phosphorylation of AKT Ser473 (Ali and Sabatini 2005; Hresko and Mueckler 2005; Sarbassov et al. 2005b). Rictor-, GßL-, and MAPKAP1-deficient mouse embryos die at midgestation, most likely as a result of defective fetal vascular development (Guertin et al. 2006; Jacinto et al. 2006; Shiota et al. 2006). Furthermore, basal and insulin-stimulated AKT Ser473 phosphorylation is impaired in Rictor-, GßL-, and MAPKAP-deficient mouse embryonic fibroblasts (MEFs) (Guertin et al. 2006; Jacinto et al. 2006), Drosophila, and mammalian cells (Frias et al. 2006; Yang et al. 2006). AKT phosphorylation of Thr308 is also impaired suggesting that mTORC2 facilitates this phosphorylation.

4.2.8 mTORC2 plays a role in insulin-stimulated glucose uptake

Muscle-specific *Rictor* knock-out mice displayed decreased insulin-stimulated AKT Ser473 phosphorylation and, additionally, impaired insulin-stimulated glucose uptake (Kumar et al. 2008). Impaired insulin-stimulated glucose uptake could be mediated by disrupted downstream signalling of AKT, which regulates GLUT4 redistribution from intracellular sites to the cell surface in muscle and fat tissues. Indeed, Rictor-deficient muscles show diminished phosphorylation of AKT substrate of 160 kDa (AS160, also known as TBC1D4) (Kane et al. 2002), which is phosphorylated on several residues in response to insulin in a PI3K-dependent manner (Bruss et al. 2005; Sano et al. 2003). AS160 contains two phosphotyrosine binding domains and a C-terminal Rab-GAP domain, which is proposed to promote hydrolysis of Rab proteins on the GLUT4 storage vesicle (GSV). Insulin-stimulated AS160 phosphorylation inactivates its Rab-GAP activity allowing GTP-bound GSV-associated Rabs to promote processes that lead to GLUT4 translocation to the plasma membrane. In support of this, expression of phosphorylation-resistant forms of AS160 reduces insulin-stimulated GLUT4 translocation to the cell surface (Kramer et al. 2006; Sano et al. 2003; Thong et al. 2007), and RNAi-mediated knockdown of AS160 increases basal GLUT4 levels at the cell surface (Eguez et al. 2005; Larance et al. 2005) in adipocytes.

4.2.9 Genetic studies

Mutations in the insulin receptor signalling pathway can lead to syndromes of insulin resistance (George et al. 2004; Krook and O'Rahilly 1996), therefore it is reasonable to postulate that other genes downstream of the INSR may also be involved in these syndromes. Indeed variants in genes encoding downstream components of the insulin signalling cascade have been detected by our group (in collaboration with Steven O'Rahilly) in syndromes of insulin resistance, for example, *IRS1* (Berger et al. 2002), *IRS2* (Bottomley, submitted), *AKT2* (George et al. 2004; Tan et al. 2007),

GLUT4 (unpublished), and some of these variants have been shown to be causal of disease (George et al. 2004). However, to my knowledge, there have been no efforts to screen components of mTOR complexes or *AS160* for mutations involved in insulin resistant diseases. As mTORC1 is involved in a negative feedback loop that blunts insulin signalling by inhibition of IRS1 and IRS2, and as mTORC2 is one of two known kinases that target AKT, which phosphorylates AS160 to promote insulin-stimulated glucose uptake in muscle and fat, I selected genes encoding components of mTORC1 and mTORC2 complexes and AS160 as candidates for syndromes of insulin resistance.

4.2.10 Aims of this study

To sequence genes encoding components of mTORC1 and mTORC2, and AS160, in 158 patients with syndromes of severe insulin resistance to look for potentially pathogenic mutations.

4.3 Results

4.3.1 *mTOR* sequencing in the SIR cohort

A total of 129 variants were detected in *mTOR* in insulin resistant patients (Appendix Table A9), 17 of which were present in the coding sequence (Table 4.1 and Figure 4.2). Coding sequence variants that did not alter the amino acid sequence (shown below the schematic in Figure 4.2) and/or that were also present in controls were considered unlikely to be pathogenic. This left only one rare nonsynonymous variant, V455L, after reconfirmation in genomic DNA had been carried out (underlined above the schematic in Figure 4.2).

4.3.2.1 Investigation of the V455L variation

V455L is present in heterozygous form in an Iranian female with hyperandrogenism, insulin resistance and acanthosis nigricans, and absent from 48 Iranian controls on the HGDP panel. No family members were available for co-segregation analysis so I assessed the likelihood that the variant was pathogenic by its evolutionary conservation and predicted functional effect according to various bioinformatics programs. The valine residue at position 455 is conserved in chimpanzee, mouse, rat, chicken, and opossum but not zebrafish, fruit fly, or either TOR1 or TOR2 in budding yeast (Figure 4.3). The V455L change is predicted to be benign by web-based programs SIFT, PolyPhen and PANTHER. Given this data and the chemical similarity between valine and leucine, I decided not to pursue this variant further.

Table 4.1 *mTOR* coding sequence variants detected in a cohort of severe insulin resistant patients and 11 Indian and 23 CEPH controls

Genic position	Genomic position	Minor|Major allele	Protein consequence	MAF in SIR	Detected in controls?	rs ID
Exon 7	11230691	T|C	H296H	0.006		
Exon 9	**11225807**	**C|G**	**V455L**	**0.004**	**No**	
Exon 10	11224301	C|T	D479D	0.301		rs1135172
Exon 15	11216130	G|C	A778G	0.003	Yes	
Exon 15	11216051	G|A	E804E	0.003		
Exon 19	11211537	A|G	L935L	0.004		
Exon 19	11211345	C|T	N999N	0.253		rs1064261
Exon 23	11195055	C|G	R1154R	0.05		rs17036536
Exon 24	11193482	C|T	D1210D	0.005		
Exon 33	11127645	G|A	A1577A	0.004		rs1057079
Exon 39	11113317	T|C	A1832A	0.044		rs17848553
Exon 39	11113233	T|C	S1851S	0.281		rs2275527
Exon 43	11110729	A|G	T1984T	0.006		
Exon 47	11107180	C|T	L2208L	0.007		
Exon 48	11104650	T|C	L2261L	0.003		
Exon 49	11103914	G|A	L2303L	0.1735		rs1112169
Exon 56	11092007	T|C	D2485D	0.01		

Genomic coordinates correspond to NCBI Build 36.

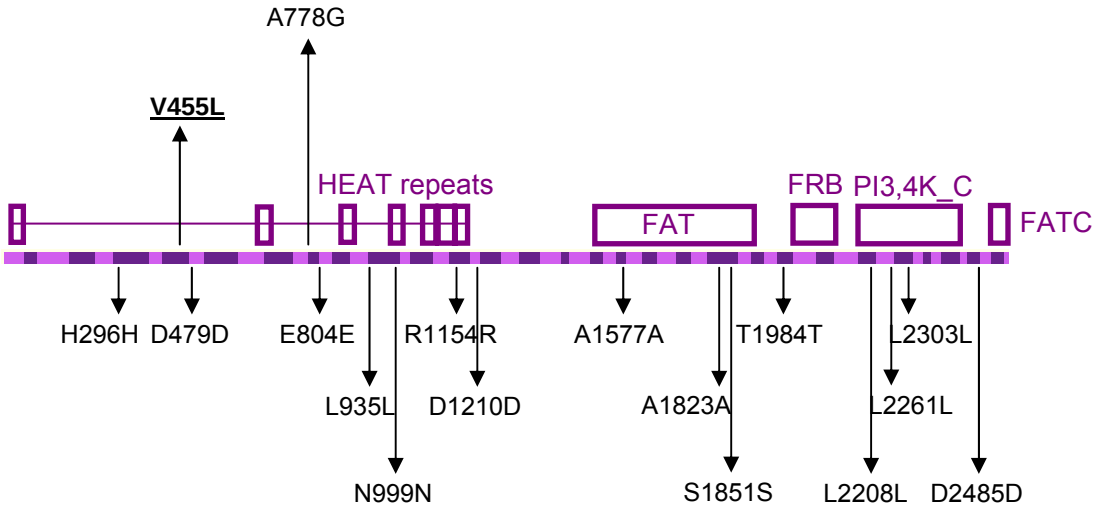


Figure 4.2 Coding *mTOR* variants in the SIR cohort
Schematic of the mTOR protein showing exons in alternating bright and dark purple and known functional domains in boxes. Arrows indicate the location of coding SNPs detected in *mTOR* by sequencing 158 patients with syndromes of severe insulin resistance. Non-synonymous variants are above the schematic and synonymous variants are below the schematic. Non-synonymous mutations absent from control samples were considered potentially pathogenic (underlined). FAT (amino acids 1513-1910) = FRAP-ATM-TRAPP domain, FRB (amino acids 2015-2114) = FKBP12-rapamycin binding, PI3,4K_C (amino acids 2181-2431) = PI3,4-kinase, catalytic domain, and FATC (amino acids 2517-2549) = FAT C-terminus domain.

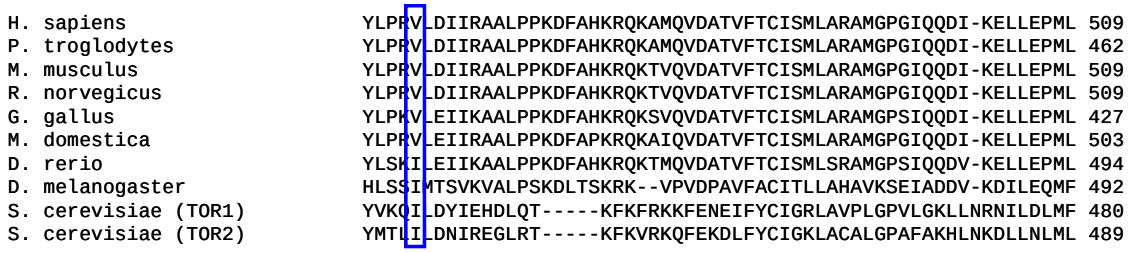


Figure 4.3 Multiple sequence alignments (using ClustalW) of *mTOR* V455 and flanking protein sequence.

4.3.2 *Rictor* sequencing in the SIR cohort

A total of 88 variants were detected in insulin resistant patients (Table A10), 10 of which were present in the coding sequence (Table 4.2 and Figure 4.4). Coding sequence variants that did not alter the amino acid sequence (shown below the schematic in Figure 4.4) and/or that were also present in controls were considered unlikely to be pathogenic. This left a homozygous non-synonymous variation, A3V.

4.3.2.1 Investigation of the A3V variation

A3V was detected in an Asian female with acanthosis nigricans and severe insulin resistance whose parents were first cousins. The variant was not present in 32 Indian samples and 28 European samples from the CIN panel. Family DNA was not available for co-segregation analysis but as alanine and valine residues have similar biochemical properties, the variant is predicted benign by PolyPhen and is not present in a conserved portion of the peptide sequence (Figure 4.5). I did not prioritise it for further analysis.

Table 4.2 *Rictor* coding sequence variants detected in a cohort of severe insulin resistant patients and 11 Indian and 23 CEPH controls

Genic position	Genomic position	Minor|Major allele	Protein consequence	MAF in SIR	Detected in controls?	rs ID
Exon 1	39110229	A|G	A3V	0.0054	No	
Exon 5	39038462	G|A	R108R	0.003		
Exon 17	38998810	C|T	I497I	0.003		
Exon 23	38994634	G|A	E745E	0.007		
Exon 23	38994622	C|T	N749N	0.007		
Exon 25	38993577	A|T	I811I	0.003		
Exon 26	38991553	T|C	S837F	0.299	Yes	rs2043112
Exon 31	38986533	C|T	S1058S	0.033		rs2115949
Exon 31	38986023	T|C	G1228G	0.003		
Exon 33	38982400	C|T	I1442I	0.002		
Exon 38	38978206	T|C	T1695I	0.017	Yes	

Genomic coordinates correspond to NCBI Build 36.

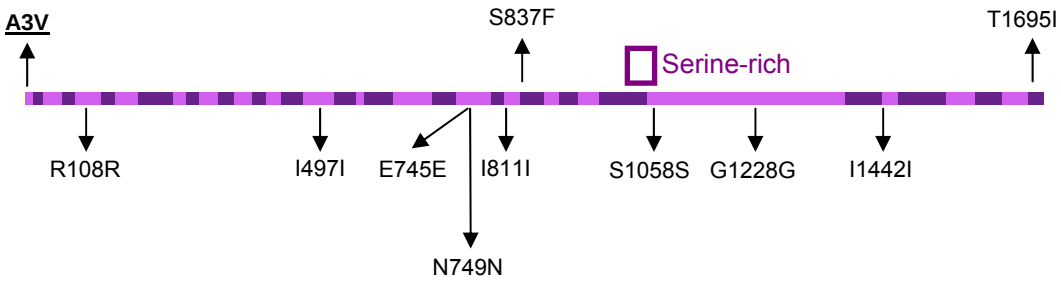


Figure 4.4 Coding *Rictor* variants in the SIR cohort
Schematic of the Rictor protein showing exons in alternating bright and dark purple and features of the peptide sequence in boxes. Arrows indicate the location of coding SNPs detected in *Rictor* by sequencing 158 patients with syndromes of severe insulin resistance. Non-synonymous variants are above the schematic and synonymous variants are below the schematic. One of the non-synonymous variants (above the schematic) was absent from controls and was therefore considered potentially pathogenic.

```
H. sapiens          -----MAAIGRGRSLKNLRVRGRND--SGEENVPLDLTREPSDNLREILQNVARLQGVSN
P. troglodytes      -----MAAIGRGRSLKNLRVRGRND--SGEENVPLDLTREPSDNLREILQNVARLQGVSN
M. musculus         -----MAAIGRGRSLKNLRIRGRND--SGEENVPLDLTREPSDNLREILQNVAKLQGVSN
R. norvegicus       ----------------------------------------EPSDNLREILQNVAKLQGVSN
M. domesticus       -----MAASVRGRSLKHLRIRGRND--SGEENVPLDLTREPYDNLREILQNVAKLQGVSN
G. gallus           -----LTLFGEGEKIESCEYRGNNNKAVQEDDLHLQLGAEPCDNMREILQNVAKLQGVSN
D. melanogaster     MASQHSSWRFGKRSKLQLRIKVSQD---PEDFYRLDPQRSAAENAFEIYS-MLCLEETRD
S. cerevisiae       ----------------MSIPHSAK-----QSSPLSSRRRSVTNTTPLLTPRHSRDNSS-
```

Figure 4.5 Multiple sequence alignment (using ClustalW) around *Rictor* A3 residue

4.3.3 *GβL* sequencing in the SIR cohort

Only ten variants were detected in insulin resistant patients (Appendix Table A11) as a result of *GβL* sequencing, 4 of which were present in the coding sequence (Table 4.3 and Figure 4.6). Two of the variants altered the amino acid sequence of the protein, A88V and E299D (underlined in Figure 4.6), and were absent from 47 Indian and 47 European controls on the CIN panel and a further 48 CEPH controls.

4.3.3.1 Investigation of the A88V variation

A88 is conserved in chimpanzee, mouse and rat but not opossum or any of the non-mammalian organisms investigated (Figure 4.7). A change from alanine to valine was detected as heterozygous in the same white, European male that carried the *LPIN1* G582R variant (see Chapter 3.3.2 for details). Briefly, this patient has a complex syndrome including severe insulin resistance and peripheral neuropathy. As DNA from first-degree relatives was available I sequenced family members for co-segregation analysis (Figure 4.8). The father (who has diabetes) and two unaffected sisters also carry the A88V variant so this change is unlikely to be causing disease in the patient. Furthermore, co-segregation analysis does not support an interaction between *LPIN1* G582R and *GβL* A88V resulting in disease as the father also carries both variants and does not share the features of the complex syndrome exhibited by the proband.

Table 4.3 *GβL* coding sequence variants detected in a cohort of severe insulin resistant patients and 11 Indian and 23 CEPH controls

Genic position	Genomic position	Minor|Major allele	Protein consequence	MAF in SIR	Detected in controls?	rs ID
Exon 4	2196583	T|C	A88V	0.006	No	
Exon 5	2197106	C|G	P137P	0.372		rs26862
Exon 9	2198768	A|G	S289S	0.007		rs11863256
Exon 9	2198798	C|G	E299D	0.006	No	

Genomic coordinates correspond to NCBI Build 36.

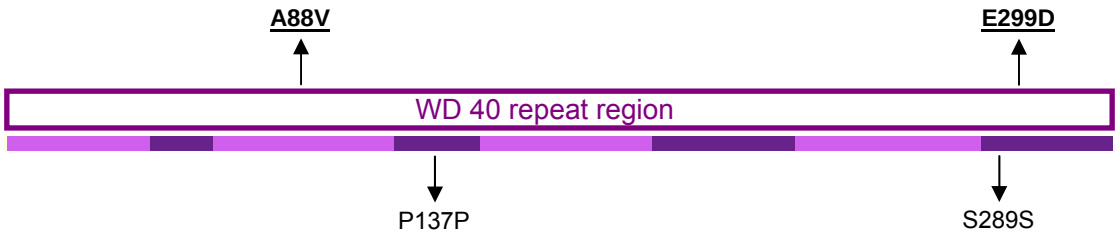


Figure 4.6 Coding *GβL* variants in the SIR cohort
Schematic of the GβL protein showing exons in alternating bright and dark purple and features of the peptide sequence in boxes. Arrows indicate the location of coding SNPs detected in *GβL* by sequencing 158 patients with syndromes of severe insulin resistance. Non-synonymous variants absent from controls (underlined) were considered potentially pathogenic.

```
H. sapiens          PIISYDGVNKNIASVGFHEDGRWMYTGGED   WCVETGEIKREYGGHQK
P. troglodytes      PIISYDGVNKNIASVGFHEDGRWMYTGGED   WCVETGEIKREYGGHQK
M. musculus         PIISYDGVSKNIASVGFHEDGRWMYTGGED   WCVETGEIKREYGGHQK
R. norvegicus       PIISYDGVSKNIASVGFHEDGRWMYTGGED   WCVETGEIKREYGGHQK
G. gallus           PVINYDGVSKNITSVGFHEDGRWMYTGGED   WCVETGEIKREYSGHQK
M. domestica        PVINYDGVSKNITSVGFHEDGRWMYTGGED   WCVDTGEIKREYGGHQK
D. rerio            PVINYDGVSKNITSVGFHEDGRWMYTGGED   WCVETGEIKREYSGHQK
D. melanogaster     PVINFDGVQKNVTRLGFQEDGNWMFTAGED   WKLQTKSSIRDYTGHTK
S. cerevisiae       PVASFEGHRGNVTSVSFQQDNRWMVTSSED   WDLSTREIVRQYGGHHK
```

Figure 4.7 Multiple sequence alignments (using ClustalW) around *GβL* amino acids A88 and E299
Straight lines indicate hidden sequence.

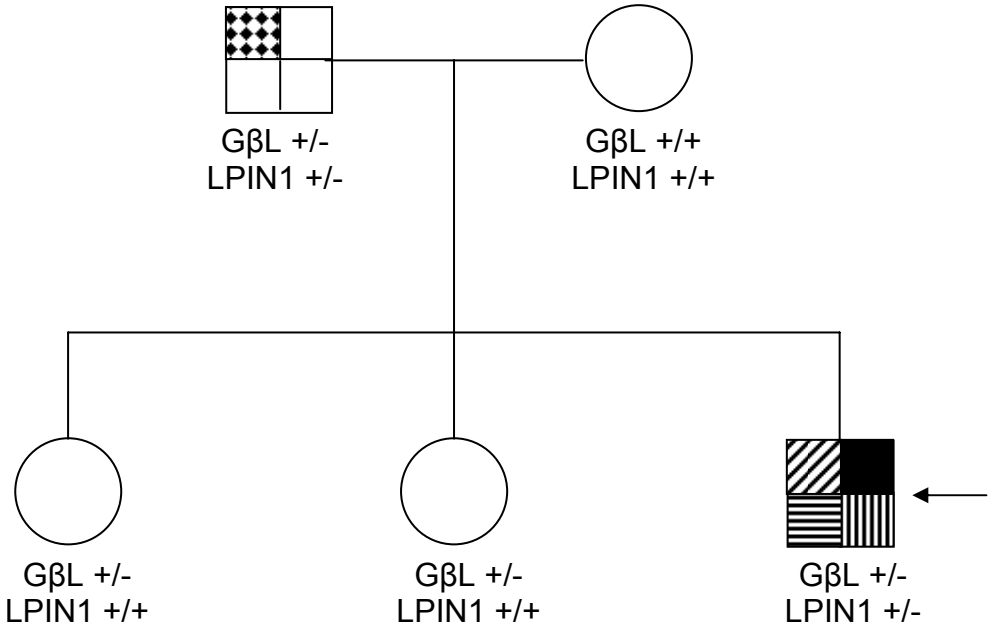


Figure 4.8 A family pedigree demonstrating that the *GβL* A88V and *LPIN1* G582R mutation do not segregate with disease in a fully penetrant manner
+/- represents a heterozygous genotype and +/+ represents the wild-type genotype. The patient (indicated by the arrow) has hyperinsulinaemia (diagonal stripes), severe peripheral neuropathy (black), previous bone marrow transplant for AML (horizontal stripes) and an intracerebral cavernous haemangioma (vertical stripes). His father has diabetes (diamonds).

4.3.3.2 Investigation of the E299D variation

E299D was detected in a female Asian patient with insulin resistant diabetes mellitus. It was also absent from the CIN control panel, which includes 47 Indian DNA samples and 47 European samples. No family were available for co-segregation analysis. Alignment of multiple *GβL* orthologs shows that E299 is not well conserved through evolution as although E is present in chimpanzee, mouse, rat, chicken and zebrafish, in opossum the corresponding residue is D (Figure 4.7). This, and the outcome of SIFT, PolyPhen, and PANTHER prediction programs (data not shown), suggest that this variant does not have a major deleterious impact on the protein and I decided not to follow this further.

4.3.4 *MAPKAP1* sequencing in the SIR cohort

A total of 37 *MAPKAP1* variants were detected in insulin resistant patients (Table A12), five of which were present in the coding sequence (Table 4.4 and Figure 4.9). Synonymous changes were considered unlikely to be pathogenic (shown below the schematic in Figure 4.9). There was one non-synonymous substitution, S260N, that reconfirmed in patient genomic DNA and was absent from 11 Asian and 23 CEPH controls.

4.3.4.1 Investigation of the S260N variation

S260N was present in two female patients, one white European with hyperandrogenism and acanthosis nigricans, and one of mixed ethnicity with pseudoacromegaly. Given that it was present in two of 158 patients I decided to sequence 96 regional controls. The variant was present in one Ely participant with fasting insulin of 42.1 pmol/l, fasting glucose of 5.2 mmol/l, and a BMI of 23.96. It is therefore likely that this variant is a rare, neutral polymorphism.

Table 4.4 *MAPKAP1* coding sequence variants detected in a cohort of severe insulin resistant patients and 11 Indian and 23 CEPH controls

Genic position	Genomic position	Minor|Major allele	Protein consequence	MAF in SIR	Detected in controls?	rs ID
Exon 5	127387762	T|C	D188D	0.005		
Exon 5	127387705	T|C	I207I	0.005		
Exon 6	**127361802**	**T|C**	**S260N**	**0.0161**	**No**	
Exon 7	127345211	T|C	K302K	0.0069		rs11542134
Exon 8	127286607	A|G	H345H	0.2173		rs2070113

Genomic coordinates correspond to NCBI Build 36.

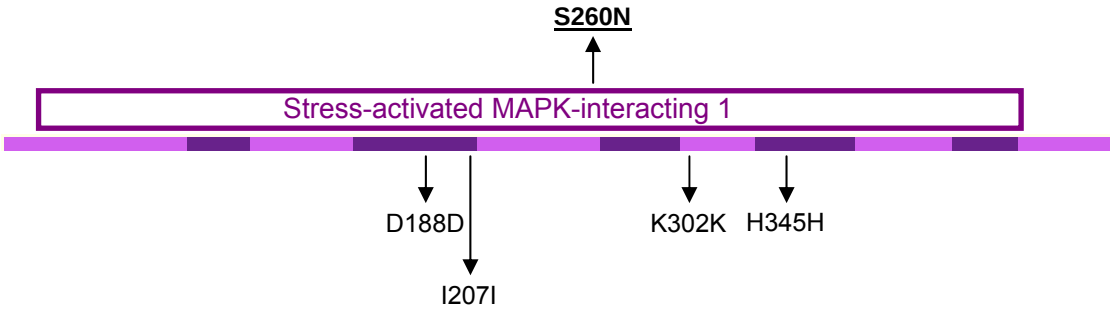


Figure 4.9 Coding *MAPKAP1* variants in the SIR cohort
Schematic of the MAPKAP1 protein showing exons in alternating bright and dark purple and the region conserved between members of the SIN1 family of proteins in boxes. Arrows indicate the location of coding SNPs detected in *MAPKAP1* by sequencing 158 patients with syndromes of severe insulin resistance. Non-synonymous variants absent from controls (underlined) were considered potentially pathogenic.

4.3.5 *AS160* sequencing in the SIR cohort

Sequencing of *AS160* was carried out in collaboration with Satya Dash at the Institute of Metabolic Science, University of Cambridge. Satya Dash sequenced a subset of the severe insulin resistance cohort with high post-prandial to fasting insulin ratios and found a point mutation in exon 3 that leads to replacement of an arginine residue with a premature stop codon at position 363.

I sequenced *AS160* exons and exon-intron boundaries in the entire SIR cohort and detected 28 variants (Appendix Table A13), 11 of which were located in coding regions (Table 4.5 and Figure 4.10). In addition to the R363X mutation I also identified eight non-synonymous changes, three of which were absent from controls (underlined above the schematic in Figure 4.10). The R363X was investigated further by Satya Dash and colleagues. I investigated the remaining three variants at the Wellcome Trust Sanger Institute.

Table 4.5 *AS160* coding sequence variants detected in a cohort of severe insulin resistant patients and 11 Indian and 23 CEPH controls

Genic position	Genomic position	Minor|Major allele	Protein consequence	MAF in SIR	Detected in controls?	rs ID
Exon 1	74953821	G|C	P28P	0.255		rs7327548
Exon 1	74953603	T|C	A101V	0.075	Yes	
Exon 2	**74834347**	**A|G**	**R299Q**	**0.0104**	**No**	
Exon 3	**74831989**	**T|C**	**R363X**	**0.0053**	**No**	
Exon 10	**74798404**	**T|A**	**N655Y**	**0.0052**	**No**	
Exon 13	74785004	G|A	T752A	0.0053	Yes	
Exon 13	**74784903**	**G|C**	**N785K**	**0.0106**	**No**	
Exon 14	74782217	G|A	V819I	0.14	Yes	rs1062087
Exon 16	74778517	A|G	K895K	0.005		
Exon 16	74774390	C|T	L967L	0.3333		rs2297208
Exon 19	74764285	T|C	T1147M	0.1	Yes	rs9600455

Genomic coordinates correspond to NCBI Build 36.

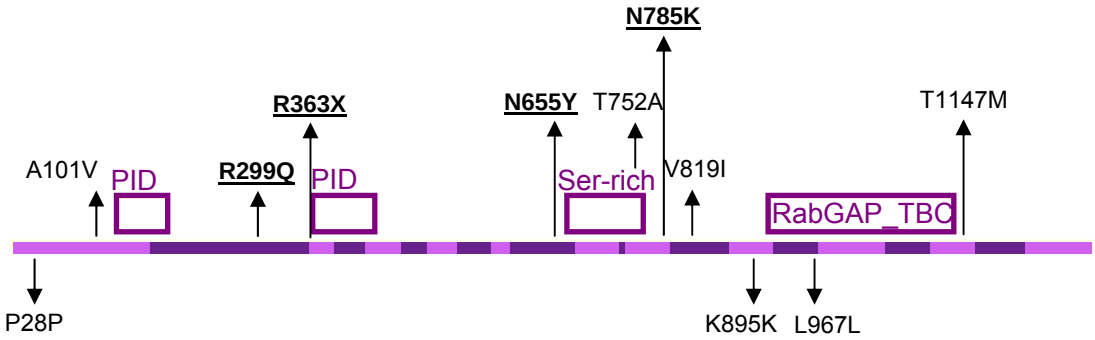


Figure 4.10 Coding *AS160* variants in the SIR cohort
Schematic of the AS160 protein showing exons in alternating bright and dark purple and known functional domains in boxes. Arrows indicate the location of coding SNPs detected in *AS160* by sequencing 158 patients with syndromes of severe insulin resistance. Non-synonymous variants absent from controls (underlined) were considered potentially pathogenic. PID (amino acids 130-188 and 366-438) = phosphotyrosine interaction domain and RabGAP_TBC (amino acids 915-1134) = Rab GTPase activator protein domain.

4.3.5.1 Investigation of the R363X variation

4.3.5.1.1 The patient

R363X was present in a female patient who, at age 11 years, presented with acanthosis nigricans and a BMI standard deviation score (SDS) +2.7. Her fasting glucose and insulin levels fell within age-/gender- and BMI-matched ranges but, following an oral glucose challenge, she displayed dramatic hyperinsulinaemia, with a peak to fasting insulin ratio of 62. By age 23 years her BMI standard deviation score had dropped to SDS 1.4 and her acanthosis nigricans resolved. Although her glucose tolerance had now normalized, she maintained an elevated peak to fasting insulin ratio of 14.

4.3.5.1.2 Family co-segregation analysis

The patient's half-sister presented with acanthosis nigricans and BMI SDS +3.5 at age 9 years and was shown to carry R363X. The mutation was also present in the patient's mother and maternal aunt, both of whom had no acanthosis nigricans but did display elevated peak to fasting insulin ratios at the time of testing (15 and 17

respectively), and the patient's maternal grandmother who was obese and had an elevated fasting glucose (118.9 mg/dL), rendering her peak to fasting insulin ratios uninterpretable. Another maternal aunt had normal glucose tolerance, a peak to fasting insulin ratio of 7, and did not carry R363X (Figure 4.11). Therefore, this mutation appears to segregate with elevated peak to fasting insulin ratios in the family.

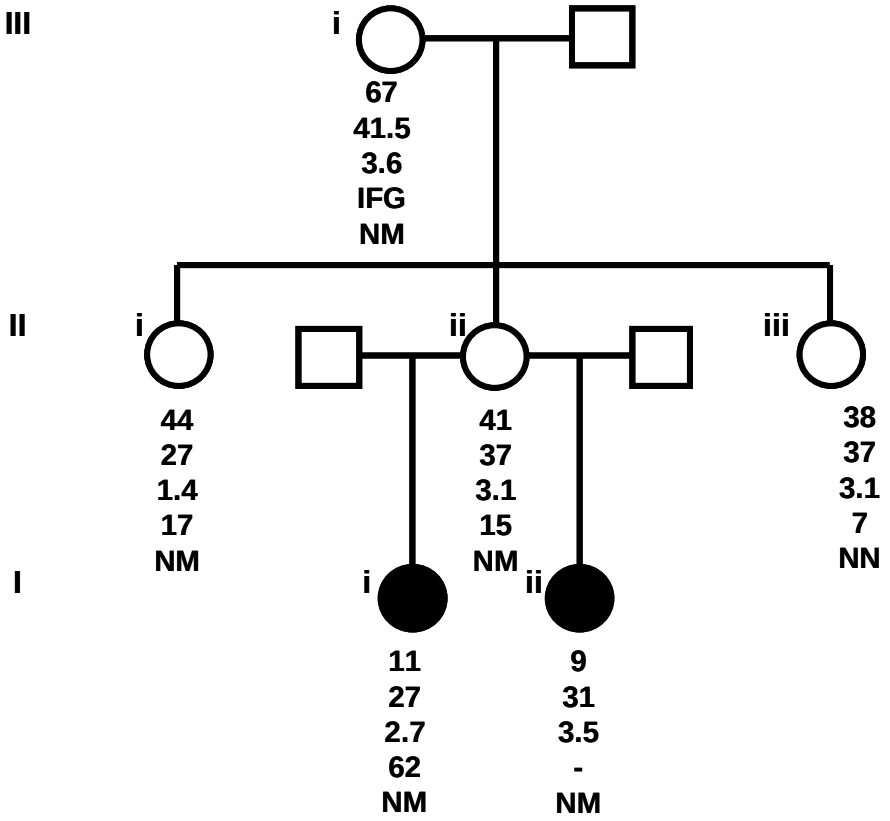


Figure 4.11 Pedigree demonstrating segregation of the *AS160* R363X variation with elevated fasting-to-peak insulin levels in the family of an insulin resistant patient (Ii), taken from (Dash et al., manuscript submitted). Black circles represent females with acanthosis nigricans. Underneath the symbols are shown age at measurement, BMI (kg/m^2), the BMI standard-deviation score (SDS), peak-to-fasting insulin ratios, and genotype (N = R363 and M = X363). IFG = impaired fasting glucose.

4.3.5.1.3 Functional analysis of R363X

Constructs carrying wild-type AS160, AS160 with substitution of a stop codon for an arginine codon (R363X), or a truncated version of AS160 which is missing the sequence from residue 363 (363Tr), were transfected by Satya Dash into 3T3L1

adipocytes. R363X and 363Tr constructs increased cell surface GLUT4 in the basal state and significantly inhibited insulin-stimulated GLUT4 translocation to the plasma membrane (Figure 4.12) compared to wild-type. Truncated protein and wild-type AS160 both bind to full length AS160 *in vitro* (Figure 4.13).

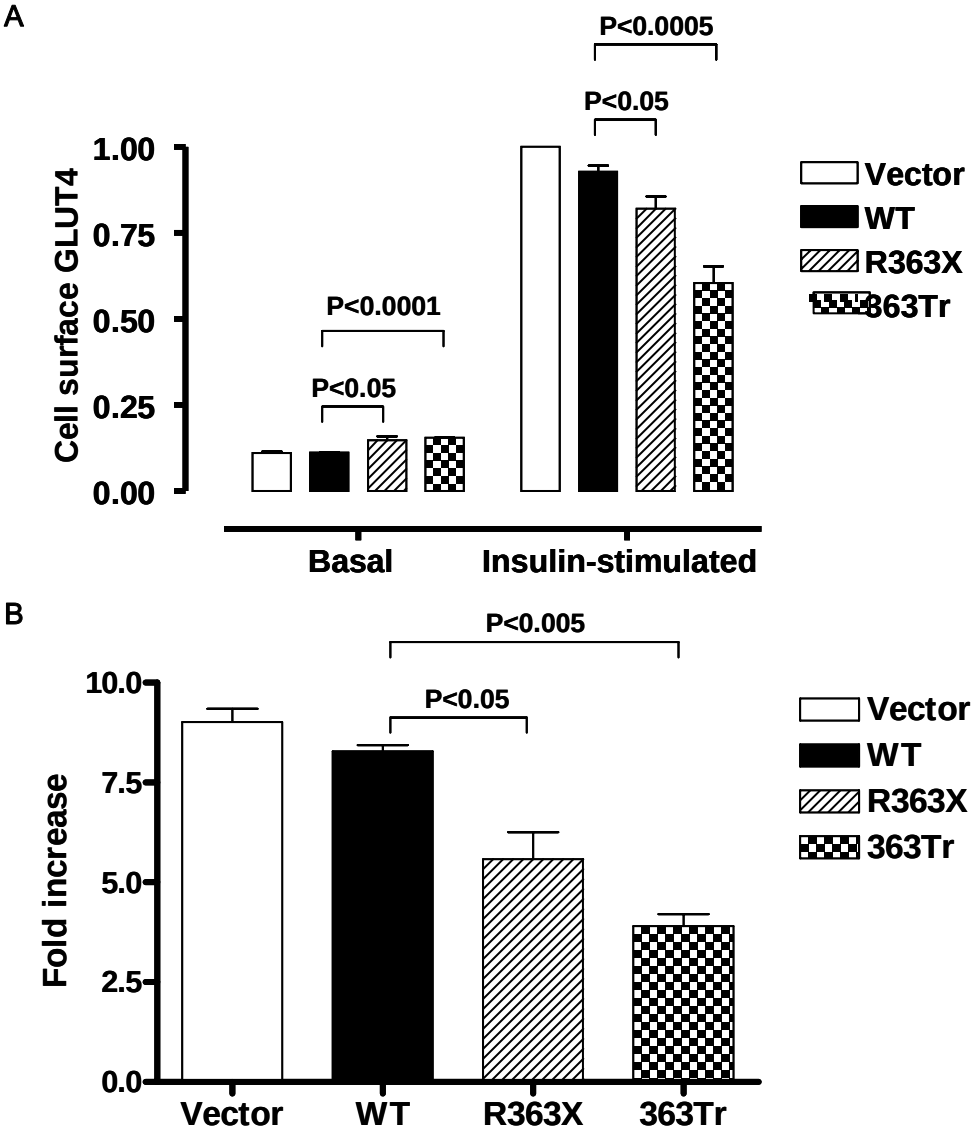


Figure 4.12 Differences in insulin-stimulated GLUT4 expression in cells carrying mutant vs wild-type *AS160*, taken from (Dash et al., manuscript submitted). **A)** Cell surface GLUT4 level, normalized to a value of 1.0 for the vector control in the insulin-stimulated state. Values are expressed as means ± SEs for 4-6 independent measurements of the Cy3/ GFP ratio in cells with and without 30 minutes treatment with 160 nM insulin and **B)** Data expressed as the ratio of cell surface GLUT4 in the insulin-stimulated state to that in the basal state. The R363X construct carries the R363X substitution, whereas the 363Tr carries the gene sequence up to residue 363.

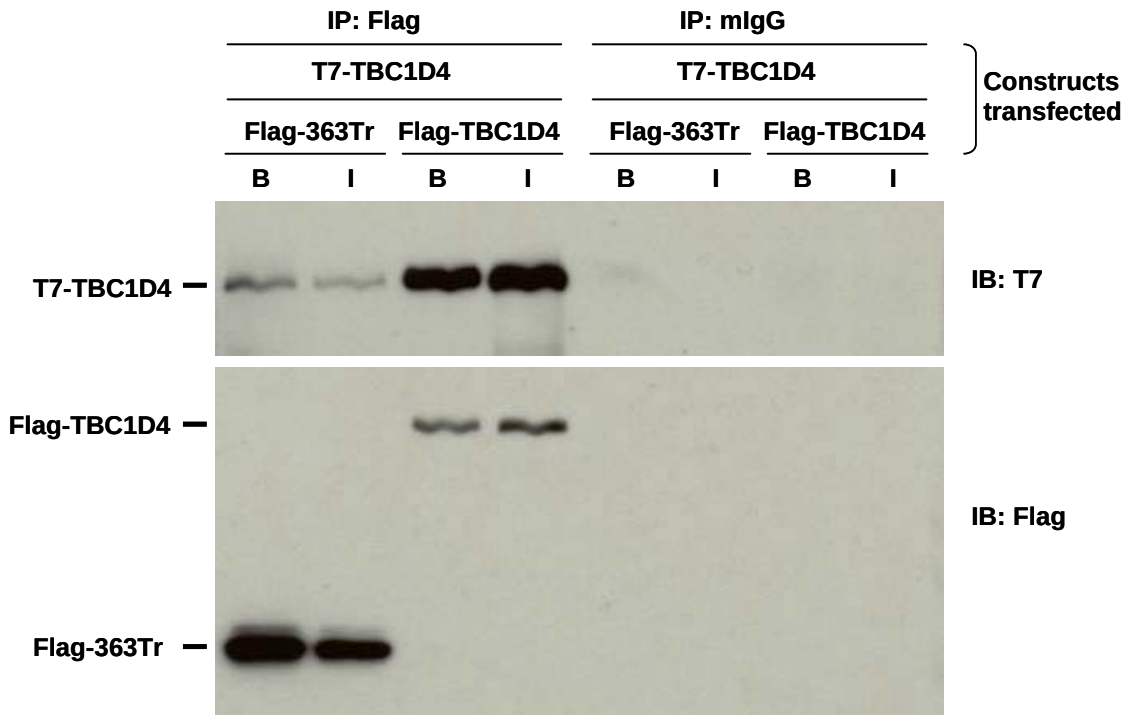


Figure 4.13 Immunoblots (IB) showing co-immunoprecipitation of T7-tagged AS160 (TBC1D4) from basal (B) and insulin-stimulated (I) 293 cells with Flag-tagged truncated AS160 (363Tr) or full-length AS160 (TBC1D4), taken from (Dash et al., manuscript submitted). To control for non-specific immunoprecipitation, lysates were also immunoblotted after immunoprecipitation with irrelevant mouse immunoglobulin and protein G-agarose (mIgG).

4.3.5.2 Investigation of the R299Q variation

R299Q is present in homozygous form in an Arabic female with pseudoacromegaly and acanthosis nigricans and is absent from 164 Arabic controls included in the HGDP panel. There was no family available for co-segregation analysis. Arginine is well conserved in a range of organisms, including chimpanzee, mouse, chicken, opossum, and frog (Figure 4.14) but a change to glutamine at residue 299 is predicted benign by SIFT, PolyPhen and PANTHER. As far as is known the patient is not the offspring of consanguineous union, therefore I would have expected to detect some heterozygotes. It is possible that the patient only appears homozygous because of a deletion in the other copy of the gene. Alternatively, the mutant copy of the gene may be preferentially amplified during PCR and sequencing. This will need to be investigated in future work.

4.3.5.3 Investigation of the N655Y variation

N655Y was detected in an obese, teenaged, white European female with acanthosis nigricans. The variant was present in the proband's obese mother (who is not insulin resistant), and was absent in her unaffected brother (Figure 4.15). Interestingly, a variant in β-melanocyte-stimulating hormone (*β-MSH*) has already been found to segregate with obesity in this family (Lee et al. 2006). It is possible that obesity-induced insulin resistance in the patient is exacerbated by the *AS160* N655Y mutation, though *β-MSH* and *AS160* together do not segregate with insulin resistance in the family. Furthermore, asparagine is not very well conserved at position 655 and there is a tyrosine residue at this position in rat (Figure 4.14), implying that this variant does not have a major deleterious effect.

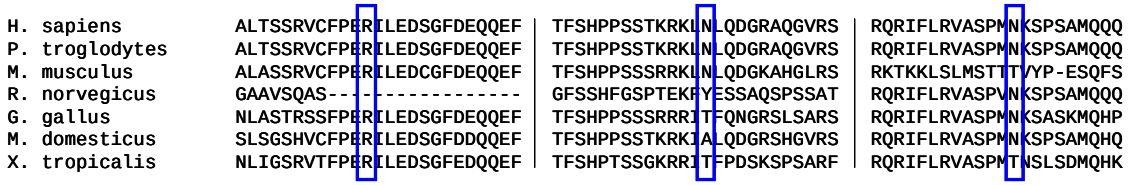


Figure 4.14 Multiple sequence alignments (using ClustalW) around *AS160* amino acids R299, N655 and N785
Straight lines indicate hidden sequence.

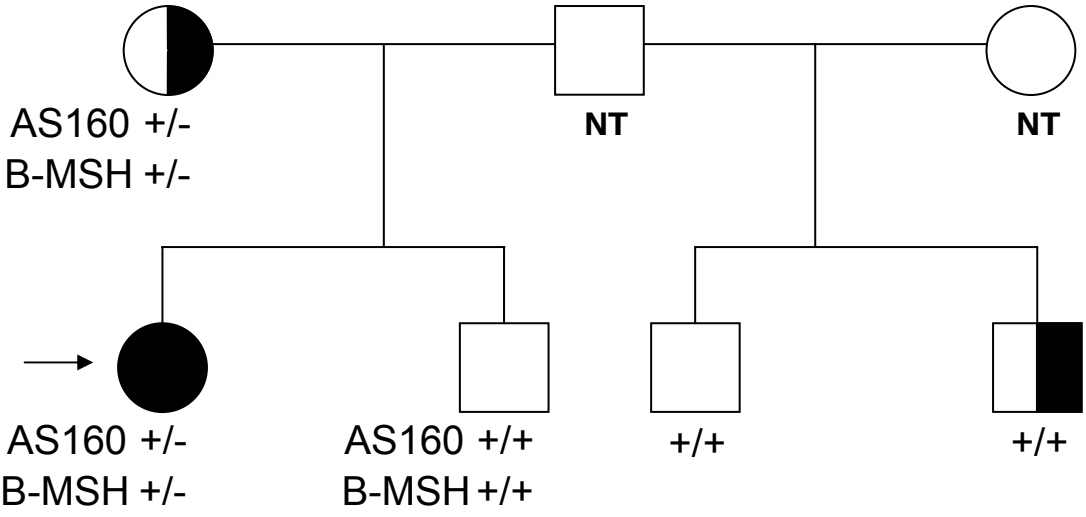


Figure 4.15 Presence of *AS160* N655Y and a previously reported *β-MSH* mutation in a family of a patient with obesity and insulin resistance
+/- represents a heterozygous genotype and +/+ represents the wild-type genotype. The patient (indicated by the arrow) is obese and insulin resistant. The mother and half brother with black and white symbols are also obese.

4.3.5.4 Investigation of the N785K variation

N785K was detected in an Asian male patient with early-onset diabetes and acanthosis nigricans, and by Satya Dash in an Asian female patient with hyperandrogenism, insulin resistance and acanthosis nigricans. Its presence in two Asian patients suggested that this variant might be an uncommon Asian polymorphism and indeed Satya Dash detected N785K in one of 95 Punjabi controls. In mouse and frog the asparagine is replaced by a threonine residue (Figure 4.14), and SIFT, PolyPhen, and PANTHER program predict that the change to lysine has no functional effect on the protein.

4.3.5.5 Summary of *AS160* mutation screening

In summary, four rare *AS160* variants that alter the protein sequence were detected in the SIR cohort, one of which lead to a premature stop codon early in the peptide sequence (R363X). R363X was extensively investigated by Satya Dash and his colleagues at the University of Cambridge. Three amino acid substitutions were also detected, none of which fell within known functional domains. R299Q is present in a patient with pseudoacromegaly and acanthosis nigricans but absence of DNA from family members prevented further analysis. N655Y was detected in an obese insulin resistant patient, in whose family obesity was likely caused by a *β-MSH* mutation. N785K was detected in two insulin-resistant Asian patients and a Punjabi control so this variant is likely to be a rare Asian polymorphism and was not investigated further.

4.3.6 *Raptor* sequencing in the SIR cohort

A total of 190 *Raptor* variants were detected in insulin resistant patients (Appendix Table A14), 29 of which were present in the coding sequence (Table 4.6 and Figure 4.16). None of the three non-synonymous variants were present in controls (underlined above the schematic in Figure 4.16). Due to time constraints, *Raptor* variants were prioritised for further study based on evolutionary conservation (Figure 4.17) and the predictions of bioformatics programs. G187S and A862T were predicted benign by PolyPhen and PANTHER and are not very well conserved (Figure 4.17). R997H is predicted to have a possible functional effect and is well conserved in vertebrates. For this reason R997H was investigated further.

4.3.6.1 Investigation of the R997H variation

R997H was detected in an Asian female with hyperandrogenism, insulin resistance, and acanthosis nigricans. The variant was also found in one CIN control and was therefore considered unlikely to be pathogenic.

Table 4.6 *Raptor* coding sequence variants detected in a cohort of severe insulin resistant patients and 11 Indian and 23 CEPH controls

Genic position	Genomic position	Minor|Major allele	Protein consequence	MAF in SIR	Detected in controls?	rs ID
Exon 1	76134114	C|T	F30F	0.0967		
Exon 2	76214157	G|C	T78T	0.2005		rs17848685
Exon 5	76319006	A|G	G187S	0.0026	No	
Exon 5	76319008	G|C	G187G	0.0053		
Exon 6	76342542	G|C	T264T	0.0026		
Exon 11	76434924	C|T	G423G	0.3617		rs3751945
Exon 11	76434969	A|G	Q438Q	0.1058		rs2589156
Exon 14	76468818	G|A	Q506Q	0.2526		rs2289759
Exon 15	76471873	A|G	T548T	0.0053		rs34848699
Exon 16	76472295	A|G	S590S	0.0058		
Exon 17	76473438	T|C	F626F	0.0028		
Exon 17	76473495	T|C	A645A	0.0028		
Exon 18	76480141	C|T	L670L	0.2526		rs2289764
Exon 18	76480225	G|A	A698A	0.1349		rs2289765
Exon 22	76511124	T|C	T842T	0.2919		rs2271603
Exon 22	76511182	A|G	A862T	0.0036	No	
Exon 24	76513818	A|G	T954T	0.0028		
Exon 25	76528905	T|C	H978H	0.0054		
Exon 25	76528961	A|G	R997H	0.0031	No	
Exon 26	76534153	T|C	A1039A	0.3521		rs1567962
Exon 28	76537938	A|G	T1122T	0.0026		
Exon 30	76548509	T|C	L1172L	0.0029		
Exon 31	76549792	T|C	R1203R	0.3411		rs9899178
Exon 31	76549870	C|T	S1229S	0.0088		

Genomic coordinates correspond to NCBI Build 36.

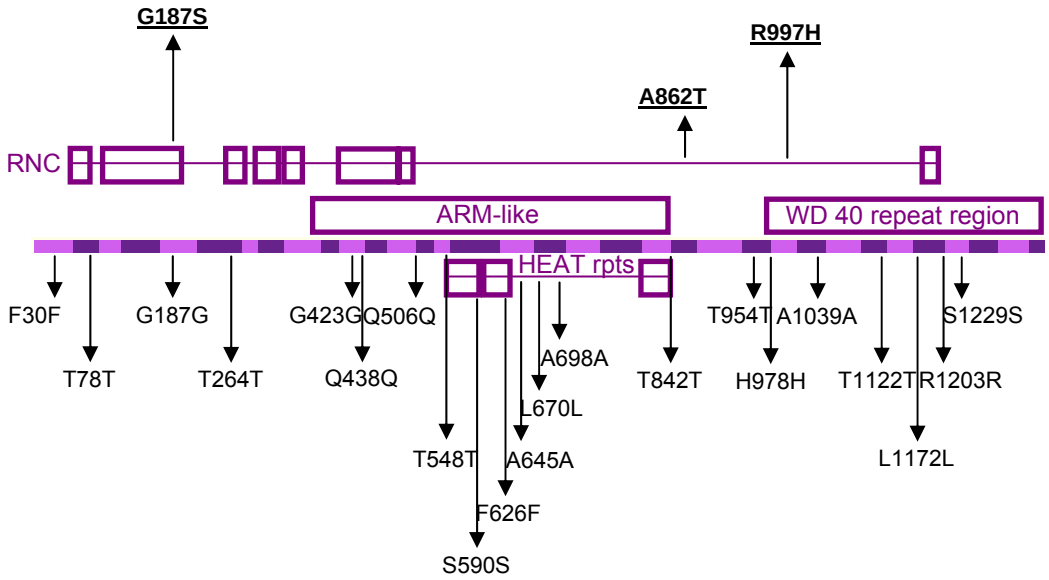


Figure 4.16 Coding *Raptor* variants in the SIR cohort
Schematic of the Raptor protein showing exons in alternating bright and dark purple and known functional domains in boxes. Arrows indicate the location of coding SNPs detected in *Raptor* by sequencing 158 patients with syndromes of severe insulin resistance. Non-synonymous variants absent from controls (underlined) were considered potentially pathogenic. RNC (amino acids 53-1196) = raptor N-terminal conserved domain and ARM-like (amino acids 372-839) = Armadillo repeat.

```
H. sapiens          QTWMGSPSIFVY | TQSAPASPTNKG | FLRNSRVRRQAQ
P. troglodytes      QTWMGSPSIFVY | TQSAPASPTNKG | FLRNSRVRRQAQ
M. musculus         QTWMGSPSIFVY | TQSAPASPTNKG | FLRNTRVRKQAQ
R. norvegicus       QTWMGSPSIFVY | TQSAPASPTNKG | FLRNTRVRKQAQ
G. gallus           QTWMGSPSIFVY | TQSAPASPTNKG | FLRNARVRKQAQ
M. domesticus       QTWMGSPSIFVY | TQSAPASPTNKG | FLRNTRVRKQAQ
X. tropicalis       QTWMGSPSIFVY | TQSAPASPTNKG | FLRNTRVRKQAQ
D. melanogaster     ITWMSAPSIYVY | AESPPVGAAASG | FLRNDFVRRHAG
S. cerevisiae       QTWLGAPCIFVY | SFSPMERVDNNA | RNRNETIIQETQ
```

Figure 4.17 Multiple sequence alignments (using ClustalW) around *Raptor* amino acids G187, A862, and R997
Straight lines indicate hidden sequence.

4.4 Discussion

mTOR complexes are important components of the insulin signalling pathway. mTORC1 modulates insulin sensitivity in response to growth factors, hormones, nutrients, energy status, and cellular stressors, while mTORC2 phosphorylates a critical kinase, AKT, required for insulin-stimulated glucose uptake. I have sequenced 158 patients with syndromes of severe insulin resistance for genes encoding components of mTORC1 and mTORC2, including *mTOR*, *GβL*, *Rictor*, *Raptor*, *MAPKAP1*, as well as a Rab-GAP protein (*AS160*), which is the substrate of AKT proteins and regulates GLUT4 trafficking to the cell surface. Eleven rare non-synonymous variants and a putative truncating mutation absent from a total of 59 controls were detected across these genes. A combination of analyses including family cosegregation analyses, sequencing of ethnically-matched controls, and prediction of functional effects based on evolutionary conservation indicated that the rare nonsynonymous variants in *mTOR*, *Rictor*, *Raptor*, *GBL*, *MAPKAP1* and *AS160* were not likely to be fully pathogenic in patients with severe insulin resistance. The stop mutation in *AS160* (R363X) segregated with high peak-to-fasting insulin ratios in a family with five affected members. *In vitro* studies demonstrated that this mutation impairs insulin-stimulated GLUT4 translocation to the plasma membrane and could disrupt activity of wild-type AS160.

The *AS160* variant R363X resulted in a premature stop codon after the first phosphotyrosine interaction domain (PID) but before the second PID and the Rab-GAP domain. This would be expected to disrupt the ability of the protein to hydrolyse Rab-GTP to Rab-GDP. Furthermore, all predicted AKT phosphorylation sites will be absent from the truncated AS160 protein potentially preventing its regulation of GLUT4 trafficking in response to insulin. Satya Dash at the Institute of Metabolic Science (IMS) in Cambridge demonstrated that the R363X co-segregated with higher

than average peak-to-fasting insulin levels in the family (Satya et al., unpublished). This phenotype seems consistent with the role of AS160 in insulin-stimulated GLUT4 trafficking. Disruption of the cell's ability to translocate GLUT4 to the membrane to facilitate glucose uptake would result in slower post-prandial glucose clearance and higher compensatory peak insulin levels, but might not be expected to affect basal insulin levels. Functional analyses undertaken at the IMS showed that the truncated protein significantly reduced insulin-stimulated GLUT4 translocation to the cell surface *in vitro*.

The proband and relevant family members are heterozygous for R363X, so this variant could be causing insulin resistance through haploinsufficiency or a dominant negative effect. In vitro studies demonstrated that AS160 exists as a dimer (unpublished observations of Cristinel Miinea and G.E.L., and Figure 4.12) and that truncated AS160 can bind the full length protein. This suggests that the mutation can act as a dominant negative mutation.

Given that the remaining eleven variants in mTORC/*AS160* genes are less likely to cause disease in a fully penetrant manner it would appear that point mutations and small insertion /deletions in components of mTORCs are not common causes of human severe insulin resistance. However, for several of these variants it was impossible to perform co-segregation analyses as DNA from family members of the patient was not available. It is difficult to establish pathogenicity in the absence of family members and, especially in smaller pedigrees, a variant can appear to segregate with disease by chance. Investigation of the candidate gene or variant in other similarly affected individuals and families might help build support for its involvement in disease. The same genetic factor might segregate with disease in multiple affected families, or affected individuals may be enriched for different rare functional variants in the same gene compared to controls. These studies would require large sample sizes of well-phenotyped cases of disease and control

individuals. Alternatively, a disease-causing variant may be unique to a particular individual/family (or in other words, the disease may be unique) rendering larger study sizes useless for establishing stronger evidence for pathogenicity. In these cases, functional studies may be the only way to establish whether and how a variant affects protein function and causes disease. As functional assays can be time-consuming to design and optimise, it is difficult to test every putative pathogenic mutation for a functional effect. Bioinformatics programs such as SIFT, PolyPhen, and PANTHER have been designed to assess the likelihood that an amino acid change is deleterious to protein function based on evolutionary conservation and the biochemistry of the residues involved. However, these programs should not be used as proof of a functional effect or lack thereof, and cannot explain how a functional variant may be impacting disease. These issues surrounding the investigation of Mendelian diseases will also be relevant to whole-genome resequencing studies, made possible by faster and cheaper sequencing technology (Ng et al. 2008b). My screening studies suggest that resequencing of all known exons in the genome will uncover many putative disease mutations. However, in the absence of large pedigrees and appropriate functional assays it may be difficult to determine the true functional variants behind Mendelian phenotypes.

As I only screened 158 patients with syndromes of severe insulin resistance, I cannot rule out variation in mTORC components as rarer causes of severe insulin resistant syndromes. Furthermore, I did not screen patients for mutations in promoter/enhancer/regulatory regions that may impact expression levels, or for large copy number changes such as deletions of whole exons. These changes cannot therefore be excluded as causes of severe insulin resistance in the cohort.

In conclusion, my screening efforts in components of mTOR complexes and *AS160* lead to the discovery of a stop mutation in *AS160* that impairs insulin-stimulated

GLUT4 translocation and segregates with high peak-to-fasting insulin levels in a family with five affected individuals. I recommend screening *AS160* in families with similar phenotypes for additional mutations, which would help corroborate these findings. Other putative functional variants were considered unlikely to be fully pathogenic but demonstrate that, in the absence of family members for co-segregation analyses and validated functional assays, it is difficult to establish the impact of rare genetic factors on Mendelian disease. These are issues that will affect future studies of Mendelian traits in the era of whole-genome sequencing.

4.5 Materials and Methods

4.5.1 Description of cohorts

See Chapter 2 for details of the cohorts. Briefly, the severe insulin resistance cohort comprises 158 patients with syndromes of severe insulin resistance, and the control cohorts comprise unaffected individuals of Northern and Western Europe (CEPH N=48), European-Indian (CIN panel N=94), and 164 individuls of Arabic ancestry (HGDP).

4.5.2 PCR and sequencing

Genomic DNA from 158 patients with syndromes of severe insulin resistance, 11 Indian controls (from the CIN panel, Appendix Table A4) and 23 CEPH controls was whole-genome amplified (Chapter 2.3.1.1). Primers were designed to amplify *mTOR*, *Rictor*, *GBL*, *MAPKAP1*, *AS160*, and *Raptor* exons, exon-intron boundaries and 3'UTR (see Appendix Table A15 for sequences and conditions). PCR and product purification were performed using the standard procedure outlined in Chapter 2.3.2. Bi-directional sequencing was performed with M13 primers using the Big Dye Terminator 3.1 kit (Applied Biosystems, Foster City, CA, USA) (see Chapter 2 for details). Sequencing reactions were run on ABI3730 capillary machines (Applied Biosystems) and analysed using Mutation Surveyor version.2.20 (SoftGenetics LLC, State College, PA, USA) or GAP4 (Staden Sequence Analysis Package software). All non-synonymous variants with MAF<0.01 were confirmed in a second PCR and sequencing reaction using patient genomic DNA. Three web-based programs were employed to predict the likelihood that point mutations had a functional impact on the protein (Chapter 2.3.6). DNA from family members used for co-segregation analysis was genomic.

4.5.3 In vitro studies of mutant AS160 function

Satya Dash and colleagues at the Institute of Metabolic Science, University of Cambridge, first identified and investigated the functional consequences of the R363X mutation in *AS160*. Human wild type *AS160* in the P3XFLAG CMV10 vector (Kane et al. 2002) was used to generate a Flag-tagged mutant construct using the QuikChangeTM site directed mutagenesis kit (Stratagene) according to the manufacturer's protocols. In order to evaluate the biological effects of the truncated mutant, a second construct was created in this vector with no coding DNA beyond the stop codon identified in the proband (designated 363Tr) (Dash et al., manuscript submitted).

The relative amount of GLUT4 at the cell surface in 3T3-L1 adipocytes was assayed by transfecting cells through electroporation with a plasmid for expression of HA-GLUT4-GFP and measuring the HA-GLUT4-GFP at the cell surface by quantitative single-cell immunofluorescence, as previously described (Sano et al. 2003). In this method, HA-GLUT4-GFP at the cell surface is labelled with anti-HA and Cy3-conjugated secondary antibody; the fluorescence intensities for Cy3 and GFP in individual cells are quantitated; and the relative amount of HA-GLUT4-GFP at the cell surface is expressed as the ratio of Cy3 to GFP, in order to correct for different levels of expression of the HA-GLUT4-GFP protein.

The association of AS160 with itself and with the 363Tr form was examined by co-transfecting human embryonic kidney 293E cells with plasmids for expression of N-terminal T7-tagged human AS160 and either N-terminal Flag-tagged AS160 or the 363Tr with Lipofectamine 2000. The cells were treated with insulin (1 µM for 10 minutes) or left unstimulated, then lysed in 40 mM Hepes, 150 mM NaCl, pH 7.4 containing 1.5 % nonionic detergent octaethyleneglycol dodecyl ether and protease and phosphatase inhibitors. The lysate was cleared by centrifugation at 12,000 x g for

15 minutes and the portions of the supernatant were immunoprecipitated with anti-Flag agarose or irrelevant mouse immunoglobulin and protein G-agarose. SDS samples of the immunoprecipitates were immunoblotted for the T7 and Flag epitopes.

4.5.4 Statistical Analysis

GLUT4 translocation data are expressed as means ± SE. Differences between vectors were compared with use of the unpaired Student's t-test. All reported P values are from two-sided tests, and P values of less than 0.05 were considered to indicate statistical significance.

Chapter 5

PARL Leu262Val and fasting insulin levels in UK populations

5.1 Summary

Mitochondrial dysfunction is associated with insulin resistance and type 2 diabetes, and increases with age. Expression levels of the presenilins-associated rhomboid-like protein (PARL), which plays a role in mitochondrial morphology and function, is associated with insulin sensitivity in Israeli sand rats and humans. The *PARL* gene maps to chromosome 3q27 within a quantitative trait locus (QTL) that influences components of the metabolic syndrome. Recently, an amino acid substitution (Leu262Val, rs3732581) in *PARL* was associated with fasting plasma insulin levels in a US white population (N = 1031). This variant was also found to modify the positive association between age and fasting insulin. The aim of this study was to test the replication of these findings in two UK population-based cohorts. Participants from the MRC Ely and Hertfordshire cohort studies were genotyped for this variant using a SNaPshot primer extension assay and Taqman assay respectively. Full phenotypic and genotypic data were available for 3666 study participants. Based on a dominant model, I found no association between the Leu262Val polymorphism and fasting insulin levels (p=0.79) or BMI (p=0.98). Nor did I observe the previously reported interaction between age and genotype on fasting insulin (p=0.14). Despite having greater statistical power, my data do not support the previously reported association between *PARL* Leu262Val and fasting plasma insulin levels, a measure of insulin resistance. My findings indicate that this variant is unlikely to importantly contribute to insulin resistance in UK populations (Fawcett et al. 2006).

5.2 Introduction

5.2.1 Mitochondrial dysfunction is associated with insulin resistance and type 2 diabetes (T2D)

Insulin resistance and the development of T2D have been associated with lower mitochondrial mass and function in key metabolic tissues. Mitochondrial oxidative capacity (Bjorntorp et al. 1967; Kelley et al. 2002; Petersen et al. 2004; Simoneau and Kelley 1997; Simoneau et al. 1999) and basal and insulin-stimulated ATP synthesis (Szendroedi et al. 2007) are reduced in skeletal muscle of patients with type 2 diabetes, their insulin-resistant offspring (Petersen et al. 2005), and in non-alcoholic fatty liver disease and hepatic insulin resistance (Perez-Carreras et al. 2003; Pessayre 2007; Zhang et al. 2007). Age-associated insulin resistance is accompanied by a reduction in mitochondrial oxidative and phosphorylation activity (Petersen et al. 2003). Smaller mitochondrial size and density have also been observed in skeletal muscle and adipose tissue of insulin-resistant, obese, and type 2 diabetic subjects (Kelley et al. 2002; Morino et al. 2005; Ritov et al. 2005; Ukropcova et al. 2007), suggesting impaired mitochondrial biogenesis. Indeed, the expression of genes involved in mitochondrial biogenesis and function, such as peroxisomal proliferator activator receptor γ coactivator 1α (*PGC1α*) and nuclear respiratory factor 1 (*NRF1*), have been shown to be reduced in skeletal muscle and adipose tissue of patients with insulin resistance, obesity and/or diabetes (Bogacka et al. 2005; Mootha et al. 2003; Patti et al. 2003; Semple et al. 2004). Furthermore, knock-down in pancreatic ß-cells of nuclear and mitochondrial genes involved in mitochondrial function results in reduced β-cell mass, impaired insulin secretion and development of overt diabetes (Silva et al. 2000; Soejima et al. 1996). These studies show that disruption of genes involved in mitochondrial function may contribute to insulin resistance and type 2 diabetes.

5.2.2 Expression of PARL correlates with insulin sensitivity

The importance of mitochondrial function in insulin action and secretion led Walder et al. (Walder et al. 2005) to use microarray analysis to screen for genes involved in mitochondrial function that were differentially expressed between skeletal muscle of lean, normal glucose tolerant, and obese type 2 diabetic *Psammomys obesus*. One gene, the presenilins-associated rhomboid-like (Parl) gene, was of particular interest for two reasons. First, expression of Parl was reduced by 50% in skeletal muscle of obese, type 2 diabetic animals compared to lean glucose tolerant animals. Second, the human homolog, *PARL*, maps to chromosome 3q27 within a region that has been linked to quantitative traits that are components of the metabolic syndrome, such as body mass index (BMI), waist and hip circumference, body weight, fasting plasma insulin and insulin/glucose ratio (Kissebah et al. 2000) (Figure 5.1). Parl expression was also negatively correlated with blood glucose and plasma insulin levels in Israeli sand rats (Walder et al. 2005). Consistent with these observations, in a Mexican American population, *PARL* expression was positively correlated with insulin sensitivity and citrate synthase activity, a marker of mitochondrial oxidative capacity (Walder et al. 2005). This lends support to the idea that *PARL* is required for mitochondrial function and thereby promotes insulin sensitivity.

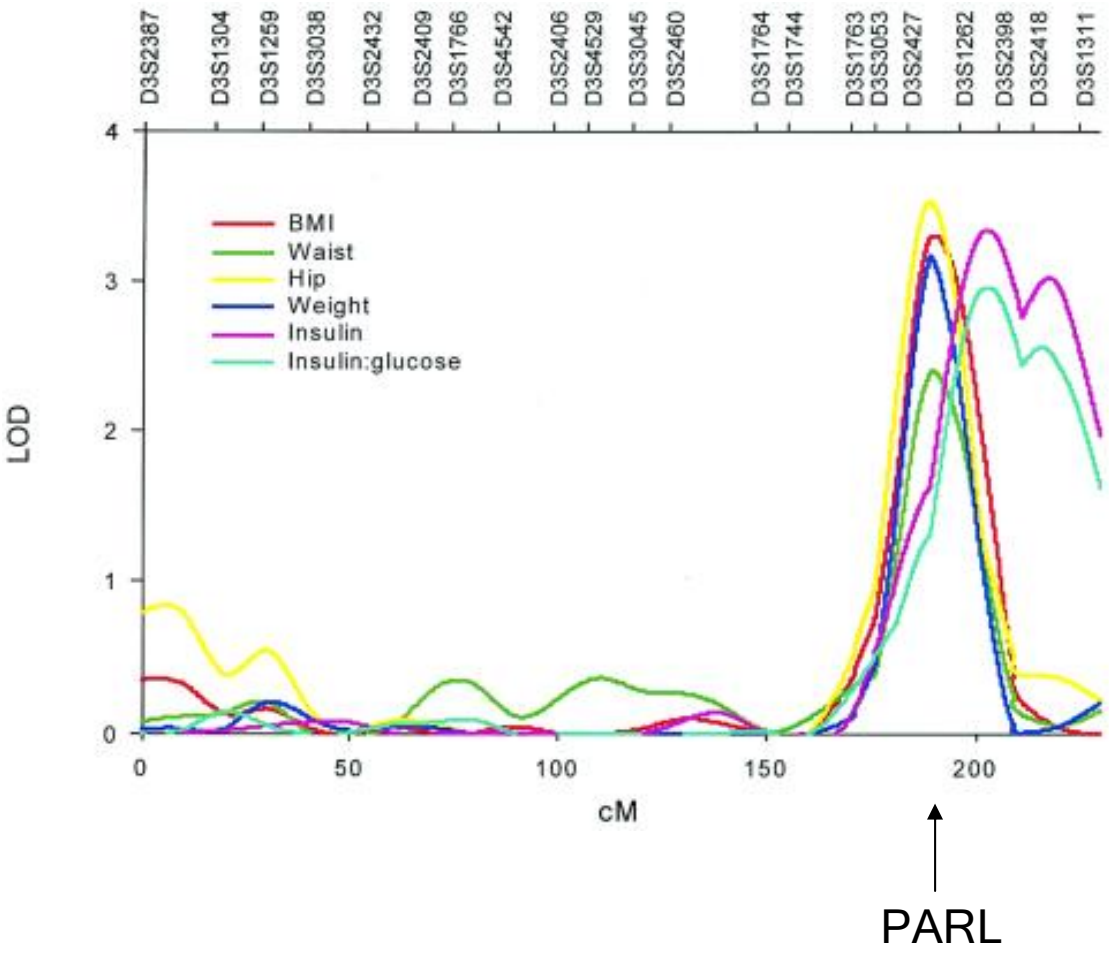


Figure 5.1 Evidence for linkage of the *PARL* locus to obesity and diabetes traits in a US study

5.2.3 A SNP in *PARL* is associated with fasting insulin levels in a US cohort

Based on its chromosomal position, and animal and human data suggesting that *PARL* expression is associated with insulin sensitivity, variation at this gene was previously investigated for association with fasting plasma insulin and BMI in a US population. A common SNP in exon 7 resulting in a leucine to valine substitution at amino acid residue 262 (Figure 5.2), was genotyped in 1031 US white individuals, a sub-group of the original population in which the QTL was discovered (Kissebah et al. 2000; Walder et al. 2005). In a dominant model, *PARL* genotype was associated with fasting plasma insulin levels and modified the positive association between age and fasting insulin. There was no statistically significant association between Leu262Val and BMI (Walder et al. 2005).

Figure 5.2 Schematic of the *PARL* protein
L262V = location of the SNP tested in this study. Exons are shown in alternating light and dark purple, and purple boxes indicate transmembrane domains. The protein is 379 amino acids long.

5.2.7 Aims of this study

The aim of my study was to investigate whether *PARL* Leu262Val genotype is associated with fasting plasma insulin levels in participants from two UK population-based cohorts.

5.3 Results

The *PARL* Leu262Val variant was in Hardy-Weinberg equilibrium in the combined Ely and Hertfordshire dataset (P = 0.94). Demographic, anthropometric and biochemical characteristics according to Leu262Val genotype for the separate cohorts are shown in Table 5.1. I did not detect any statistically significant association between genotype and fasting plasma insulin, either in the Ely (p = 0.49) or Hertfordshire (p = 0.23) populations. I also performed a meta-analysis of Ely and Hertfordshire studies adjusting for cohort to increase power to detect an effect of genotype on insulin levels, but no significant association was found (p = 0.79) (Table 5.2).

Table 5.1 Demographic, anthropometric, and biochemical characteristics of study participants by *PARL* Leu262Val genotype in Ely and Hertfordshire cohorts

				P value
	GG	GC	CC	
Ely				
N	412	832	364	
Women (%)	217 (52.7)	444 (53.4)	205 (56.3)	0.58
Age (Years)	62.2 ± 0.46	60.9 ± 0.31	60.8 ± 0.46	0.01
Fasting insulin (pmol/l)	60.1 ± 2.01	59.3 ± 1.30	55.4 ± 1.82	0.49*
BMI (kg/m2)	27.3 ± 0.24	27.3 ± 0.16	27.1 ± 0.27	0.65
Hertfordshire				
N	547	990	521	
Women (%)	257 (47.0)	476 (48.1)	245 (47.0)	0.77
Age (Years)	65.7 ± 0.12	65.6 ± 0.09	65.7 ± 0.12	0.91
Fasting insulin (pmol/l)	88.5 ± 2.58	92.1 ± 2.09	97.5 ± 3.64	0.23*
BMI (kg/m2)	27.2 ± 0.18	27.4 ± 0.14	27.0 ± 0.19	0.75

Data are means ± SE, unless otherwise indicated. The P value indicates the results of regression analysis assuming dominance of the C allele. * = analysis performed on log transformed data.

Table 5.2 Analysis of associations between *PARL* Leu262Val genotype and fasting insulin levels or BMI in pooled Ely and Hertfordshire cohorts

	n	B ± SE	P value
Fasting insulin	3666	-1.76 ± 2.26	0.79*
BMI	3666	-0.01 ± 0.17	0.95

B = Regression coefficient is the mean change (± SE) in the trait of interest when the C allele is present. * = analysis performed on log transformed data and included adjustment for cohort term.

I found no association between genotype and BMI in Ely (p = 0.65), Hertfordshire (p = 0.75) or in the combined dataset (p = 0.94) (Table 5.1). Furthermore, no age-dependent effects of genotype on fasting insulin levels were detected in either of the separate cohorts (p = 0.40 for Ely, and p = 0.21 for Hertfordshire) or in the pooled analysis adjusted for study (p = 0.14) (data not shown).

The statistically significant association between genotype and fasting insulin reported by the original study was based on a US cohort with higher mean BMI than our UK cohorts (Walder et al. 2005). Therefore I investigated the impact of BMI on the association between genotype and fasting insulin in the pooled dataset. In an analysis restricted to individuals in the upper tertile for BMI (mean BMI = 32.3, N = 1276), I found no statistically significant association between *PARL* genotype and fasting insulin (p = 0.23). Secondly, I wanted to test for an interaction between genotype and BMI on fasting insulin (in other words, I wanted to determine whether the way fasting insulin varies by genotype depends on BMI). To test this I fitted an interaction term (genotype*BMI) to the regression model. However, the coefficient of the interaction term was not statistically significant (the change in fasting insulin in response to genotype did not depend on BMI) (p = 0.88).

Finally, I assessed the association between this genetic variant and other quantitative metabolic traits (fasting glucose levels (pmol/l), two-hour glucose levels (pmol/l), and 30 min insulin increment (pmol/mmol) with *PARL* genotype (Table 5.3 and 5.4). Only one nominally significant association was detected with insulin increment in Ely subjects but this did not replicate in Hertfordshire participants and was not significant in a joint analysis (Table 5.4).

Table 5.3 Summary of glucose traits and insulin increment in Ely and Hertfordshire study participants by *PARL* Leu262Val genotype

				P value
	GG	GC	CC	
Ely				
Fasting glucose	5.12 ± 0.04	5.07 ± 0.03	5.12 ± 0.05	0.47
2-hour glucose	6.46 ± 0.13	6.44 ± 0.09	6.48 ± 0.14	0.94
Insulin increment	39.49 ± 1.52	36.44 ± 0.91	35.87 ± 1.25	0.03
Hertfordshire				
Fasting glucose	5.97 ± 0.05	5.89 ± 0.03	5.91 ± 0.04	0.15
2-hour glucose	7.68 ± 0.13	7.39 ± 0.09	7.48 ± 0.12	0.05
Insulin increment	603.09 ± 20.55	629.07 ± 18.0	621.20 ± 25.6	0.66

Data are means ± SE. The P value indicates the results of regression analysis assuming dominance of the C allele. All statistical analyses were performed on log transformed data.

Table 5.4 Analysis of associations between *PARL* Leu262Val genotype and glucose traits and insulin increment in pooled Ely and Hertfordshire cohorts

	n	B ± SE	P value
Fasting glucose	3715	-0.008 ± 0.005	0.12
2-hour glucose	3599	-0.02 ± 0.01	0.13
Insulin increment	2900	-0.05 ± 0.03	0.08

B = Regression coefficient is the mean change (± SE) in the trait of interest when the C allele is present. All analyses were performed on log transformed data and included adjustment for cohort term.

5.4 Discussion

Based on 3666 study participants with full genotypic and phenotypic data, I found no evidence of an association between the *PARL* Leu262Val SNP and fasting insulin levels. Nor did I replicate the genotype- age interaction on fasting insulin levels (Walder et al. 2005), suggesting that this genetic variant does not importantly contribute to variation in fasting insulin levels in UK populations.

There are several reasons that may explain the lack of agreement between my data and that which was previously published. The original study showed that plasma insulin levels were significantly higher in GG compared to GC/CC genotypic groups and that plasma insulin increased more rapidly with age in GG homozygotes (Walder et al. 2005). The study population consisted of extended families recruited from Take Off Pounds Sensibly (TOPS) membership and the mean BMI in this population was greater than the two UK cohorts included in this analysis (33.4 ± 7.2 kg/m^2 in the US cohort compared to 27.3 ± 4.5 kg/m^2 in the UK cohorts). To examine the possibility that *PARL* Leu262Val is more strongly associated with fasting plasma insulin in more overweight people, I repeated my analyses in the restricted population of Ely and Hertfordshire participants in the upper tertile of the distribution of BMI. No statistically significant impact of genotype on fasting insulin was detected, nor was there evidence for an interaction between BMI and genotype on fasting insulin levels. These data indicate that failure to replicate the previous result is unlikely to be due to population differences in BMI.

My study may have lacked the statistical power to detect the association of Leu262Val genotype with plasma insulin levels. However, I estimated that the pooled study would have had a 93% probability of detecting the previously reported effect size (a fasting insulin decrease of 7.84 pmol/l in GC and CC genotypic classes

compared to GG homozygotes), given the frequency of the C allele in my study (0.49) and my sample size of 3666 with a type 1 error rate of 0.05. The power of the previously reported study was less due to the smaller sample size. In our stratified analyses our power to detect an association of the previously reported effect size in the upper tertile of BMIs (N = 1276) was 64%. I therefore cannot exclude a false negative finding in this stratified analysis.

The difference in results could also be explained by other genetic and/or environmental differences between the US and UK study populations. For example, the degree of linkage disequilibrium between the *PARL* SNP and the putative unmeasured true functional variants may vary between the cohorts.

I detected a nominally significant difference in age between those Ely subjects with a C allele and those without (Table 5.1), which is probably a chance finding resulting from multiple testing as it is not apparent in the combined Ely and Hertfordshire dataset. Although there was a difference in average fasting plasma insulin between the Ely and Hertfordshire cohorts, most probably related to between-laboratory variation, this difference was not associated with genotype and would not have affected our results as we included an adjustment for study cohort in our pooled analysis.

The PARL data from Ely and Hertfordshire cohorts was published in 2006 (Fawcett et al. 2006). Since then, data from genome-wide association studies have become publicly available allowing me to evaluate the role of this polymorphism in the context of type 2 diabetes. Leu262Val is not on the Affymetrix 500k SNP chip but is present on Illumina 317k *P*-value based meta-analysis of WTCCC, DGI, and FUSION studies by the DIAGRAM consortium showed no statistical association of rs3732581 with type 2 diabetes ($P = 0.07$) (Zeggini et al. 2008).

In summary, my results suggest that *PARL* Leu262Val is unlikely to importantly contribute to insulin resistance in UK populations. Furthermore, meta-analyses of data from whole-genome association studies show no association of this SNP with type 2 diabetes risk.

5.5 Materials and Methods

5.5.1 Description of cohorts

The MRC Ely study and Hertfordshire cohort study are population-based cohorts comprised of white men and women of European ancestry. Participants underwent standard anthropometric measurements and a 75-g OGTT. These cohorts are described in more detail in Chapter 2.

5.5.2 Genotyping

The Leu262Val SNP (rs3732581) was genotyped in Ely using an ABI PRISM SNaPshot™ multiplex kit (Applied Biosystems, Foster City, CA). First, a~400 bp product containing the Leu262Val SNP locus was amplified by PCR (Chapter 2.3.2) (Forward primer: AGGAGAAGTATTCATGCCCG, reverse primer: ACTGCACCAAGTGATGGTCC) and purified as described in 2.3.4 except that 3.86 units SAP (USB) and 2.90 units Exo1 (USB) were used. Each SNaPshot extension reaction, containing 0.24 µM of probe (CAGGTGTTATTTCCAATTTTGTCAGTTAC) designed to complement sequence adjacent to the SNP, 2 µl PCR product, and 4 µl SNaPshot, was carried out by repeating the following cycle 40 times: 96°C for 10 seconds, 50°C for 10 seconds, and 60°C for 30 seconds. The extension product was incubated with 1 unit of SAP at 37°C for 1 hour then 80°C for 15 minutes. 0.5 µl of the product was mixed with 1 µl GeneScan 120LIZ™ Size Standard (Applied Biosystems) and 8.5 µl dimethylformamide Hi-Di, denatured at 96°C for 2 minutes, and then electrophoresed on an ABI 3100 genetic analyser (Applied Biosystems). Analysis was carried out using GeneMapper v3.0 software (Applied Biosystems, UK). One-hundred-and-twenty water blanks were included and 86 DNA samples (2.2%) were in duplicate on the plates. Genotyping success rate was 95% and there was no discordance between replicate samples. Hertfordshire cohort samples were genotyped using a custom TaqMan® SNP assay (Applied Biosystems, UK) on 10ng of DOP DNA by Matthew Sims at MRC epidemiology unit. Allele calling was

performed on the ABI PRISM® 7900HT Sequence Detection System (Applied Biosystems, UK). One-hundred-and-fifty-five water blanks and 91 duplicates were included. Genotyping success rate was 96% and there was no discordance between replicate samples.

5.5.3 Statistical analysis

Linear regression analysis in Stata v8 (College Station, TX) was used to assess the association between genotype and fasting plasma insulin or BMI in 1608 Ely and 2058 Hertfordshire cohort samples with full genotypic and phenotypic data. Based on the previous reported association, my primary analysis was for a dominant genetic model and I tested genotype–age interactions using log-likelihood ratio tests. In secondary exploratory analyses, I assessed the association between Leu262Val genotype and three further quantitative traits: fasting plasma glucose, 2-hour glucose, and 30min insulin increment in an OGTT test (the difference between 30 minute and fasting insulin concentrations divided by the 30 minute glucose concentration in an OGTT test), using regression analysis. I also investigated whether the association between genotype and fasting insulin was modified by BMI. To increase power I undertook all analyses on both cohorts together, adjusting for study.

Chapter 6

Large candidate gene association studies of pancreatic β-cell genes and risk of type 2 diabetes

6.1 Summary

Pancreatic β-cell dysfunction is a necessary component of type 2 diabetes and genes with important roles in pancreatic β-cell have been shown to influence disease susceptibility. I contributed to analysis of a large-scale candidate gene T2D association study of genes predicted to be involved in β-cell development, function and survival. One thousand five hundred and thirty-six non-redundant SNPs across 84 candidate genes were genotyped in three UK based case-control studies (1,484 cases and 1,856 controls) and in an Ashkenazi case-control study (930 cases and 461 controls). SNPs mapping within the Wolfram Syndrome 1 (*WFS1*) gene were associated with type 2 diabetes in both populations and in a pooled analysis of all four studies. Analysis of *WFS1* variation in HapMap samples revealed that the entire gene maps within a strong LD block extending 15 kb 5' and 3' of the gene. Further support for the association was sought by genotyping *WFS1* SNPs in three further UK studies (7119 cases and 9072 controls) and a Swedish case-control study (1296 cases and 1412 controls). An updated meta-analysis of all studies with data available from WTCCC, FUSION, and DGI genome-wide association studies demonstrated genome-wide significance of the association between rs10010131 and type 2 diabetes risk (OR = 0.89 (0.86-0.92), $P = 5.4 \times 10^{-11}$). Rare mutations in *WFS1* cause an autosomal recessive condition characterised in part by juvenile-onset non-autoimmune diabetes. These studies demonstrate that common variation in *WFS1* also predisposes to common T2D with complex inheritance patterns. The initial association studies (Sandhu et al. 2007) and Swedish replication study (Franks et al. 2008) have been published.

6.2 Introduction

Type 2 diabetes is a metabolic disease characterised by defects in insulin action and insulin secretion. Candidate gene linkage and association studies have therefore focused on genes involved in biological pathways affecting one or both of these processes. Recently, a large candidate gene association study from my group tested 152 SNPs in 71 genes for association with type 2 diabetes risk and underlying quantitative traits (Barroso et al. 2003). These genes were assigned to one of three categories: those involved in pancreatic ß-cell function, those primarily influencing insulin action and glucose metabolism in muscle, liver and adipose tissue, and those influencing other processes relevant to diabetes such as energy intake, energy expenditure and lipid metabolism. A number of SNPs showed nominal association with type 2 diabetes, most of which fell within genes involved in pancreatic ß-cell function. These results led us to hypothesise that insulin sensitivity may be more heterogeneous and influenced by environmental and other factors, while ß-cell function might have a stronger genetic component. Certainly, genes with important roles in pancreatic β-cells have been shown to influence susceptibility to type 2 diabetes, including *KCNJ11* (Gloyn et al. 2003; Nielsen et al. 2003), *HNF4α* (Love-Gregory et al. 2004; Silander et al. 2004), *TCF7L2* (Grant et al. 2006), and *TCF2* (*HNF1β*) (Winckler et al. 2007).

Hypothesis-free genome-wide association studies were not yet feasible when planning this project so we embarked on a large-scale candidate gene association study in three UK-based case-control studies and an independent replication case-control study of Ashkenazi Jewish ethnicity, prioritising genes involved in pancreatic ß-cell development, function, and survival as more likely to bestow detectable risk of type 2 diabetes. The aims of this study were:

1) To test for association between variation across candidate genes and risk of type 2 diabetes

2) To replicate any promising findings in further case-control studies

6.3 Results and Discussion

6.3.1 Type 2 diabetes association study of genes involved in pancreatic ß-cell function

This was a collaborative project involving groups lead by Dr Inês Barroso (Wellcome Trust Sanger Institute, Hinxton, UK), Prof. Nick Wareham (MRC Epidemiology Unit, Cambridge, UK), Prof. Andrew Hattersley (Peninsula Medical School, Exeter, UK), Prof. Alan Permutt (Washington University, St. Louis, USA), Prof. Benjamin Glaser (Hadassah Hebrew Medical Center, Israel) and Prof. Mark McCarthy (Wellcome Trust Centre for Human Genetics, Oxford). I contributed to the analysis of the 1536 SNPs, the genotyping of the replication studies and the writing of the paper (Sandhu et al. 2007).

6.3.1.1 Candidate gene and SNP selection

Eighty-four genes predicted to be involved in pancreatic ß-cell development, growth, function and survival were prioritised for testing in a type 2 diabetes association study by Inês Barroso (Wellcome Trust Sanger Institute, UK), Alan Permutt (Washington University, St. Louis, USA), and Andrew Hattersley (Peninsula Medical School, Exeter). All SNPs present in HapMap in November 2004 that mapped within these 84 transcripts and had been genotyped in CEU trios were selected for genotyping, except those that were in complete linkage disequilibrium ($r^2 = 1$) with a SNP already selected. In total 1536 SNPs were selected for genotyping (Table 6.1).

Table 6.1 Candidate genes and number of SNPs genotyped according to functional group

Group	Hugo Symbol	Common Name	Number of SNPs
Pancreas development/ transcription	*ACVR1*	activin A receptor, type I	12
	ACVR1b	activin A receptor, type IB	4
	ACVR2	activin A receptor, type II ; ACTRII	9
	BTC	betacellulin	5
	CDH1	cadherin 1, type 1, E-cadherin (epithelial)	10
	CDH2	cadherin 2, type 1, N-cadherin (neuronal)	80
	CDK2	cyclin-dependent kinase 2	4
	EGF	epidermal growth factor (beta-urogastrone)	7
	EGFR	epidermal growth factor receptor	37
	FGF10	fibroblast growth factor	5
	FGF7	fibroblast growth factor 7 (keratinocyte growth factor)	9
	FGFR1	fibroblast growth factor receptor 1 (fms-related tyrosine kinase 2, Pfeiffer syndrome)	4
	FGFR2	fibroblast growth factor receptor 2 (bacteria-expressed kinase, keratinocyte growth factor receptor, craniofacial dysostosis 1, Crouzon syndrome, Pfeiffer syndrome, Jackson-Weiss syndrome)	113
	FOXA3	HNF3γ	4
	FOXO1A	FKHR, FOXO1	10
	FOXO3A	FKHRL1	14
	HNF4γ	HNF4γ	9
	ISL1	ISL1 transcription factor, LIM/homeodomain, (islet-1)	1
	JAG1	jagged 1 (Alagille syndrome)	10
	JAG2	jagged 2	5
	LHX4	LIM homeobox 4	10
	LHX6	LHX6.1	8
	LMX1A	LIM homeobox transcription factor 1, alpha	29
	NEUROD1	neurod/Beta2	1
	Notch1	Notch homolog 1, translocation-associated (Drosophila)	9
	NOTCH2	Notch homolog 2 (Drosophila)	8
	NOTCH3	Notch homolog 3 (Drosophila)	8
	Notch4	Notch homolog 4 (Drosophila)	5
	NR5A2	Lrh1	30
	ONECUT1	HNF6	6
	PAX6	paired box gene 6 (aniridia, keratitis)	4
	PBX1	pre-B-cell leukaemia transcription factor 1	56
	PSEN2	presenilin 2 (Alzheimer disease 4)	5
	RARB	retinoic acid receptor, beta	95
	RORB	nuclear receptor RZR-beta	19
	TCF1	HNF1α	5
	TCF2	HNF1β	12
	TCF3	E47	3
	TGFBR1	transforming growth factor, beta receptor I (activin A receptor type II-like kinase, 53kDa)	5
	VIPR1	vasoactive intestinal peptide receptor 1	9
	VIPR2	vasoactive intestinal peptide receptor 2	9
β-cell death/ apoptosis	*BAX*	BCL2-associated X protein	3
	BCL2	B-cell leukaemia/lymphoma 2	53
	BID	BH3 interacting domain death agonist	16

Table 6.1 Candidate genes and number of SNPs genotyped according to functional group continued

Group	Hugo Symbol	Common Name	Number of SNPs
	BIK	BCL2-interacting killer (apoptosis-inducing)	4
	CASP9	caspase 9	4
	EIF2AK3	PERK	7
	GSK3B	glycogen synthase kinase 3 beta	5
	WFS1	Wolfram syndrome 1 (wolframin)	6
Insulin secretion	*ATP2B1*	ATPase, Ca++ transporting	6
	CACNA1C	calcium channel, voltage-dependent, L type, alpha 1C subunit; CaV1.2	133
	CACNA1D	calcium channel, voltage-dependent, L type, alpha 1D subunit	69
	CHGA	chromogranin A (secretogranin 2)	5
	CPE	carboxypeptidase E	19
	EIF2S1	eukaryotic translation initiation factor 2, subunit 1 alpha, 35kDa	4
	GLP1R	glucagon-like peptide 1 receptor	10
	KCNJ6	potassium inwardly-rectifying channel, subfamily J, member 6	103
	MAPK8IP1	islet-brain 1	3
	PCSK1	prohormone convertase 1	8
	PCSK2	prohormone convertase 2	45
	SGNE1	secretory granule, neuroendocrine prot 1	9
	SNAP25	SNAP25 synaptosomal-associated protein	16
	VTI1B	VTI1B vesicle transport through interaction with t-SNAREs homolog 1B (yeast) or vSNARE	4
Insulin signalling	*FGF2*	fibroblast growth factor 2 (basic); bFGF	9
	INSR	insulin receptor	27
	IRS1	insulin receptor substrate 1	5
	IRS2	insulin receptor substrate 2	7
	PIK3R1	phosphatidylinositol 3-kinase, regulatory subunit, polypeptide 1 (p85 alpha)	15
	PRKCE	PKC epsilon	109
	PRKCZ	PKC zeta	5
	PSEN1	presenilin 1 (Alzheimer disease 3)	8
	RPS6KB1	ribosomal protein S6 kinase, 70kDa, polypeptide 1	7
Enzymes and metabolism	*CAMK2A*	CAMK2A	16
	CAMK2B	CAMK2B	8
	CAMK2D	CAMK2D	28
	CAMK2G	calcium/calmodulin-dependent protein kinase (CaM kinase) II gamma	7
	CAPN3	calpain 3	11
	GCK	glucokinase	6
	GCKR	glucokinase regulatory protein	5
Inflammation	*IKBKB*	IKKb	4
	IL6	interleukin 6	2
	NFATC1	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 1	25
	NFKB1	NFKbeta; NFkB	10
	SOCS3	suppressor of cytokine signaling 3	2

6.3.1.2 Association testing of candidate gene SNPs in type 2 diabetes case-control studies

SNPs were genotyped by P. Deloukas's laboratory at the Wellcome Trust Sanger Institute in four case-control studies of type 2 diabetes. These included three UK-based studies: Cambridgeshire, EPIC, and Exeter (1484 cases and 1856 controls), and an Ashkenazi study (930 cases and 461 controls). Of 1536 genotyped SNPs, 1367 (89%) met the genotyping QC criteria (section 6.4.2) and were analysed in two stages. In the first stage, association with diabetes risk was tested in UK populations only (1484 cases and 1856 controls) (Figure 6.1). Eighteen (1.3%) were associated with diabetes risk at $P < 0.01$ (Table 6.2).

In the second phase of the analysis, these 18 SNPs were tested for association with diabetes risk in the Ashkenazim study set, an ethnically distinct founder population comprising 930 cases and 461 controls. Two of the originally associated SNPs (rs10010131 and rs6446482) were associated in this independent study at $P < 0.05$ (Table 6.2). These SNPs are both located in *WFS1* and are in high linkage disequilibrium ($r^2 = 0.98$) in our study populations.

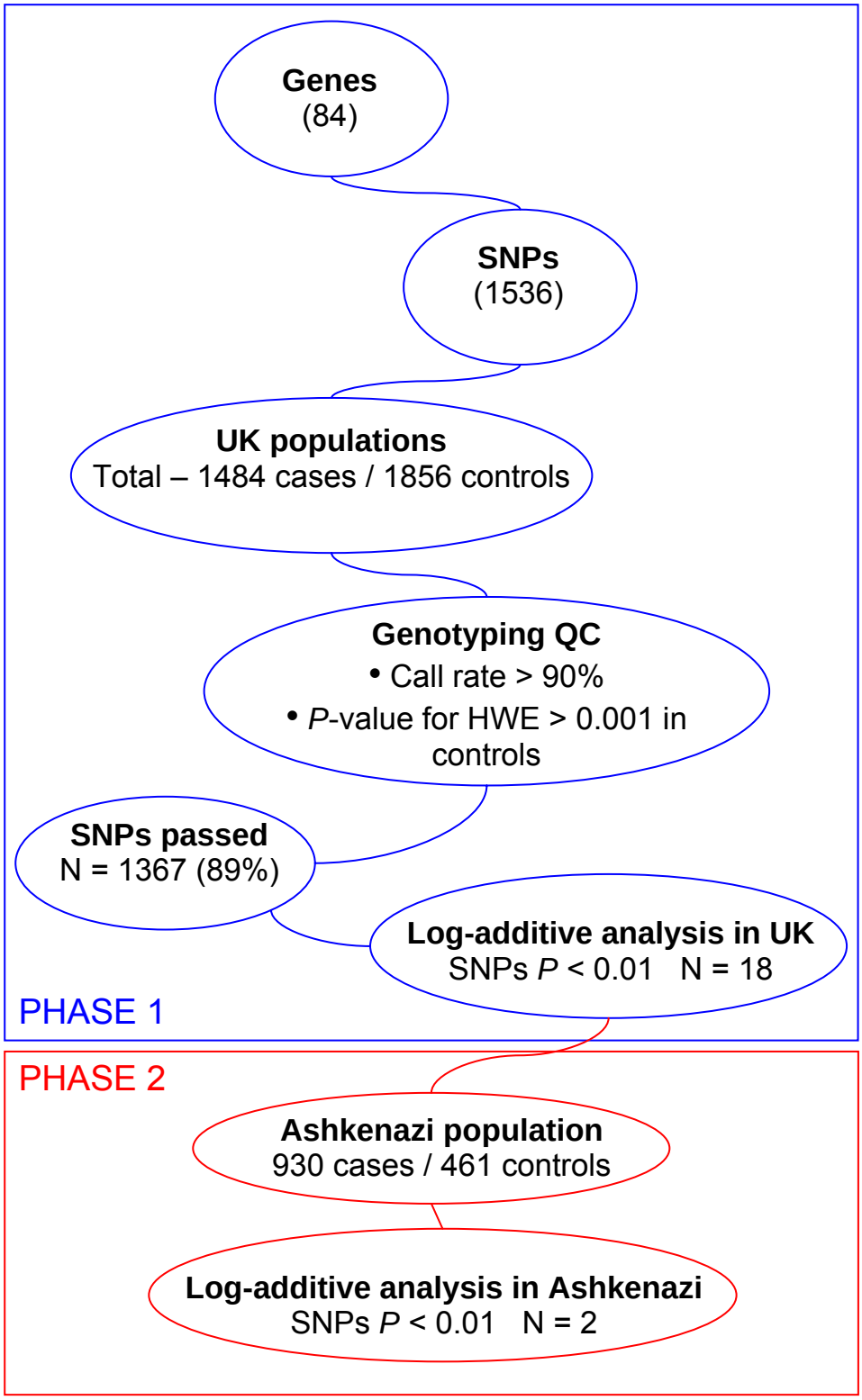


Figure 6.1 Flow chart of the procedure followed during the first and second phases of the candidate SNP analysis

Table 6.2 Statistical associations (P < 0.01) between SNPs in genes involved in beta-cell development, growth, function and survival and risk of type 2 diabetes in UK populations and for a study in an Ashkenazi population

Gene	SNP	Odds ratio (95 % CI) for stage 1: UK populations	P value for phase 1 (UK populations)*	Odds ratio (95 % CI) for Ashkenazi population	P value for r Ashkenazi population
CHGA	rs941584	0.80 (0.72-0.90)	2.8 x 10-4	1.03 (0.87-1.22)	0.75
NFATC1	rs643705	1.30 (1.12-1.52)	5.4 x 10-4	0.85 (0.71-1.02)	0.08
PAX6	rs628224	0.82 (0.71-0.93)	2.8 x 10-3	0.95 (0.74-1.22)	0.71
NFKB1	rs1609798	0.85 (0.77-0.95)	2.9 x 10-3	1.04 (0.87-1.24)	0.68
NFKB1	rs11722146	0.85 (0.76-0.95)	3.0 x 10-3	1.03 (0.86-1.22)	0.75
NFKB1	rs230498	0.86 (0.77-0.95)	3.0 x 10-3	1.03 (0.87-1.21)	0.74
WFS1	rs10010131	0.86 (0.78-0.95)	3.1 x 10-3	0.79 (0.66-0.94)	0.01
CACNA1D	rs4687736	1.18 (1.06-1.32)	3.3 x 10-3	1.00 (0.82-1.22)	0.97
EGFR	rs2075112	1.15 (1.04-1.27)	5.1 x 10-3	1.03 (0.87-1.22)	0.74
PBX1	rs7535186	0.87 (0.79-0.96)	5.3 x 10-3	0.97 (0.83-1.14)	0.76
WFS1	rs6446482	0.87 (0.79-0.96)	5.7 x 10-3	0.79 (0.66-0.95)	0.01
NFKB1	rs230539	0.87 (0.78-0.96)	5.9 x 10-3	1.03 (0.87-1.21)	0.77
TCF2	rs7501939	1.15 (1.04-1.27)	6.6 x 10-3	1.04 (0.89-1.22)	0.64
CACNA1D	rs3796347	1.15 (1.04-1.26)	7.0 x 10-3	0.96 (0.81-1.13)	0.60
CAMK2A	rs3822607	0.87 (0.78-0.96)	7.1 x 10-3	1.06 (0.89-1.26)	0.50
NFATC1	rs3826567	1.30 (1.07-1.58)	7.3 x 10-3	0.88 (0.69-1.11)	0.27
FOXA3	rs11669442	1.15 (1.04-1.28)	7.6 x 10-3	0.92 (0.77-1.08)	0.31
NUDT6	rs1048201	1.19 (1.05-1.36)	8.2 x 10-3	1.19 (0.98-1.45)	0.08

The UK population comprised up to 1,484 cases and 1,856 controls, and the Ashkenazi population comprised up to 930 cases and 461 controls. All SNPs were in Hardy-Weinberg equilibrium ($P > 0.01$ in controls). CI, confidence interval

* Based on a log additive model adjusting for study.

6.3.1.3 Association of *WFS1* SNPs with type 2 diabetes risk in three UK studies and one Ashkenazi study

Although only two of the original six variants typed across *WFS1* were associated with diabetes in the two-staged approach, in a pooled analysis of all four study sets, five SNPs showed statistical association with diabetes (Table 6.3). These SNPs were in high linkage disequilbirum (Table 6.4). As rs10010131 showed a stronger association than the other *WFS1* SNPs, likelihood ratio tests were used to assess whether rs10010131 explained all the observed associations in this region. We consecutively added the other SNPs in this region in a log additive form to a model containing rs10010131 (1df assuming no dominance at the test locus). This analysis demonstrated that none of the other SNPs in this region of high LD improved the model containing just rs10010131, and indicated that no additional genotyped SNP is independently contributing to T2D risk (Table 6.5). The reciprocal analysis, adding rs1001031 to models containing the other associated SNPs, showed that rs10010131 significantly improved the fit of all models (Table 6.5), with the exception of SNP rs6446482. The correlation between SNPs rs10010131 and rs6446482 is very high (r^2=0.98) making it difficult to distinguish between their effects on T2D risk. These analyses suggest that either rs10010131 or rs6446482 might be causal alleles, or that they are in LD with a causal allele, or both, and that the other SNPs do not independently contribute to disease risk.

There is also evidence that rs752854 might contribute to disease risk independently of rs10010131, as adding rs752854 (which was not statistically associated with disease) to the model containing rs10010131 substantially improved the fit of the model (P = 0.002, 1df) (Table 6.5).

Table 6.3 Association between SNPs located in the *WFS1* gene and risk of type 2 diabetes: all study populations comprising up to 2,414 cases and 2,317 controls

SNP	Odds ratio	95 % CI	P-value*
rs10010131	0.84	0.77-0.92	1.3×10^{-4}
rs6446482	0.85	0.78-0.93	2.7×10^{-4}
rs4689391	0.86	0.79-0.94	9.6×10^{-4}
rs3821943	0.89	0.81-0.96	5.0×10^{-3}
rs1801212	0.89	0.81-0.98	0.015
rs752854	0.94	0.85-1.03	0.164

* Based on a single locus log additive model adjusted for study. CI, confidence interval.

Table 6.4 Correlations among *WFS1* SNPs genotyped in all study populations: 2,317 controls

	rs10010131	rs6446482	rs4689391	rs3821943	rs1801212	rs752854
rs10010131	-					
rs6446482	0.98	-				
rs4689391	0.94	0.93	-			
rs3821943	0.88	0.87	0.92	-		
rs1801212	0.76	0.74	0.72	0.68	-	
rs752854	0.86	0.84	0.82	0.76	0.84	-

Correlations were not materially different in UK and Ashkenazi populations (data not shown).

Table 6.5 Log-likelihood ratio tests assessing the association among SNPs in the *WFS1* gene with risk of type 2 diabetes: all study populations comprising up to 2,414 cases and 2,317 controls

SNP	P value (1 df)*	P-value (1 df)**
rs10010131	-	-
rs6446482	0.254	0.053
rs4689391	0.524	0.047
rs3821943	0.290	5.1×10^{-3}
rs1801212	0.457	2.0×10^{-3}
rs752854	1.7×10^{-3}	4.3×10^{-6}

CI, confidence interval, all models include a study variable

* P value (log likelihood ratio test) for addition of SNP (log additive) to model containing SNP rs10010131 (2df)

** P value (log likelihood ratio test) for addition of SNP rs10010131 (log additive) to model containing SNP (2 df).

6.3.1.4 Selection of SNPs for replication in further UK cohorts

As the associations between T2D risk and SNPs rs10010131 and rs6446482 were indistinguishable given the high LD between these SNPs, and as rs752854 improves the fit of the model containing rs10010131, these three SNPs were prioritised for replication studies. We also endeavoured to discover other possible causative variants by examining variation in the region of the association signal. Using data from HapMap, and based on sequence spanning 63.4 kb (chromosome 4, 6374656-6438055), including 15 kb extending both 5' and 3' from *WFS1*, strong LD was detected across the region. The entire gene was defined by a single haplotype block of ~40 kb (Figure 6.2). Within this block, 53 SNPs had a minor allele frequency (MAF) of ≥1%. The six SNPs typed in this part of the study were all located in this region and together tagged 88% of the common variation in this block (47 of the 53 SNPs, MAF ≥1%, r^2 ≥0.8) with a mean r^2 of 0.97. One nonsynonymous SNP (R611H, rs734312) is highly correlated with SNPs rs10010131 (r^2 = 0.92) and rs6446482 (r^2 = 0.88) and thus may be a causal variant. It also showed suggestive association with type 2 diabetes in a previous study based on 479 cases and 509 controls (a subset of our samples) (Minton et al. 2002).

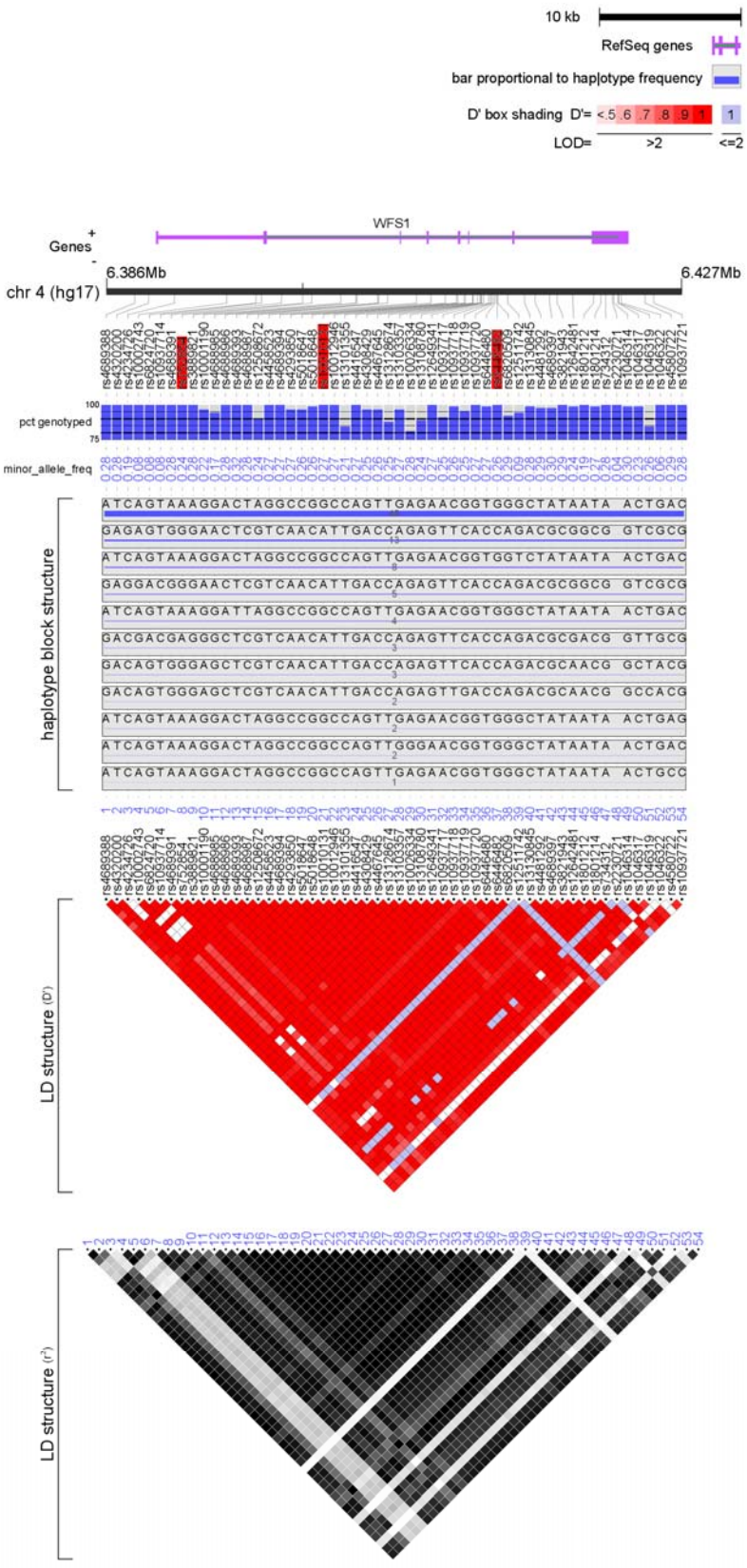


Figure 6.2 Feature map of the *WFS1* gene. Legend continued on page 181.

Figure 6.2 legend continued. The positions of the 54 SNPs genotyped in HapMap with a MAF≥ 0.01 are shown relative to the locus (purple) and chromosome 4 (black bar) (see text for details). The three SNPs typed in all studies are highlighted in red. Underlying each SNP is QC information and MAF in the CEPH samples. The middle part of the figure shows the 11 common haplotypes for the *WFS1* gene with frequency ≥ 0.01 in the CEPH samples. The thickness of the blue lines are proportional to the haplotype frequencies. The bottom of the figure depicts two LD plots for the *WFS1* locus with pairwise LD values presented for SNPs. The upper plot presents LD as D' - see figure key for details. The figure was generated using LocusView (T. Petryshen, A. Kirby, M. Ainscow, unpublished software, available from the Broad Institute, Cambridge, MA (http://www.broad.mit.edu/mpg/locusview/). In the lower plot, LD among SNPs is given as r^2. r^2 values of 1.0 are represented by black diamonds, intermediate r^2 values are shown in grey and r^2 values of 0 as white. This plot was generated using Haploview(Barrett et al. 2005), available from the HapMap website (http://www.broad.mit.edu/mpg/haploview/index.php).

6.3.1.5 Replication of *WFS1* SNPs rs10010131, rs6446482, rs752854 and rs734312 in three further UK-based case-control studies

To extend support for an association between variation at *WFS1* and diabetes risk, further replication of rs10010131, rs6446482, rs752854 and the highly correlated nonsynonymous SNP (rs734312) in three further case-control studies, ADDITION (926 cases and 1,497 controls), Warren 2 (2,465 cases and 3,843 controls) and Tayside (3,728 cases and 3,732 controls) was attempted. Independent evidence for association of rs10010131, rs6446482 and rs752854, was found in each study (Table 6.6).

Table 6.6 Association between SNPs located in the *WFS1* gene and risk of type 2 diabetes: replication studies and pooled analysis

SNP	Odds ratio (95% CI)	P-value
ADDITION study (926 cases and 1,497 controls)		
rs10010131	0.87 (0.77-0.98)	0.020
rs6446482	0.87 (0.77-0.98)	0.021
rs752854	0.86 (0.76-0.97)	0.013
rs734312	0.92 (0.82-1.03)	0.163
Warren 2 study (2,465 cases and 3,843 controls)		
rs10010131	0.91 (0.84-0.98)	0.011
rs6446482	0.92 (0.86-0.99)	0.027
rs752854	0.93 (0.86-1.00)	0.060
rs734312	0.93 (0.87-1.00)	0.061
Tayside study (3,728 cases and 3,732 controls)		
rs10010131	0.93 (0.87-0.99)	0.029
rs6446482	0.92 (0.87-0.99)	0.019
rs752854	0.93 (0.86-0.99)	0.032
rs734312	0.93 (0.87-0.99)	0.019
All seven pooled studies comprising up to 9,533 cases and 11,389 controls*		
rs10010131	0.90 (0.86-0.93)	1.4×10^{-7}
rs6446482	0.90 (0.87-0.94)	3.4×10^{-7}
rs734312	0.92 (0.88-0.95)	2.0×10^{-5}
rs752854	0.92 (0.88-0.96)	1.3×10^{-4}

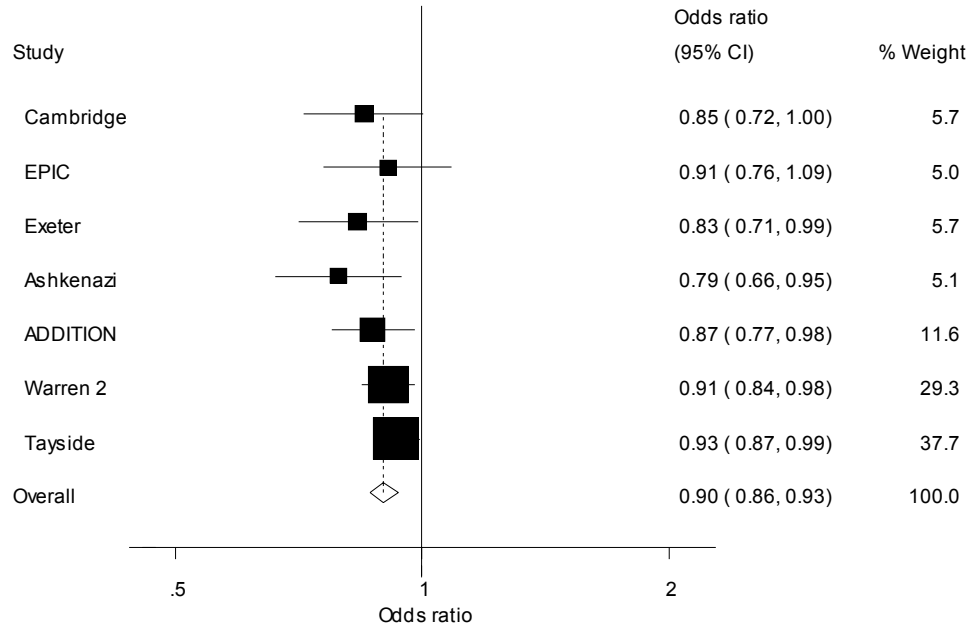
* Based on a single locus log additive model adjusted for study. CI, confidence interval.

The possible interdependency of rs10010131 and rs752854 was tested in a combined analysis of ADDITION, Warren 2 and Tayside studies. However, adding rs752854 did not improve the fit of the logistic regression model containing rs10010131 indicating that rs752854 does not interact with rs10010131 to effect risk of T2D.

6.3.1.6 Pooled analysis of *WFS1* SNPs rs10010131, rs6446482, rs752854 and rs734312 in all UK studies and the Ashkenazi study

As there was some evidence for association of the nonsynonymous SNP rs734312 in the ADDITION, Warren 2 and Tayside studies (Table 6.6) this variant was genotyped in the original four studies and a pooled analysis of all seven studies was conducted, comprising up to 9,533 cases and 11,389 controls (Table 6.6). In this analysis, rs734312 was associated with diabetes risk ($P = 2.0 \times 10^{-5}$). However, likelihood ratio tests showed that rs734312 did not contribute to a model including rs10010131 ($P = 0.88$), whereas rs10010131 substantially improved the fit of a model including rs734312 ($P = 4.9 \times 10^{-3}$), suggesting that rs734312 is unlikely to be the functional variant explaining these associations. There was also no consistent evidence for interdependency between rs10010131 or rs6446482 and rs752854 and diabetes risk in the combined study sets (data not shown). In the pooled analysis, rs10010131 (MAF = 40%) and rs6446482 (MAF = 41%) were strongly associated with diabetes risk at $P = 1.4 \times 10^{-7}$ and $P = 3.4 \times 10^{-7}$, respectively (Table 6.6). Furthermore, the magnitude of this association was highly consistent across studies (Figure 6.3), with no heterogeneity among studies (P (six degrees of freedom) = 0.59 and 0.68 for rs10010131 and rs6446482, respectively).

(a)



(b)

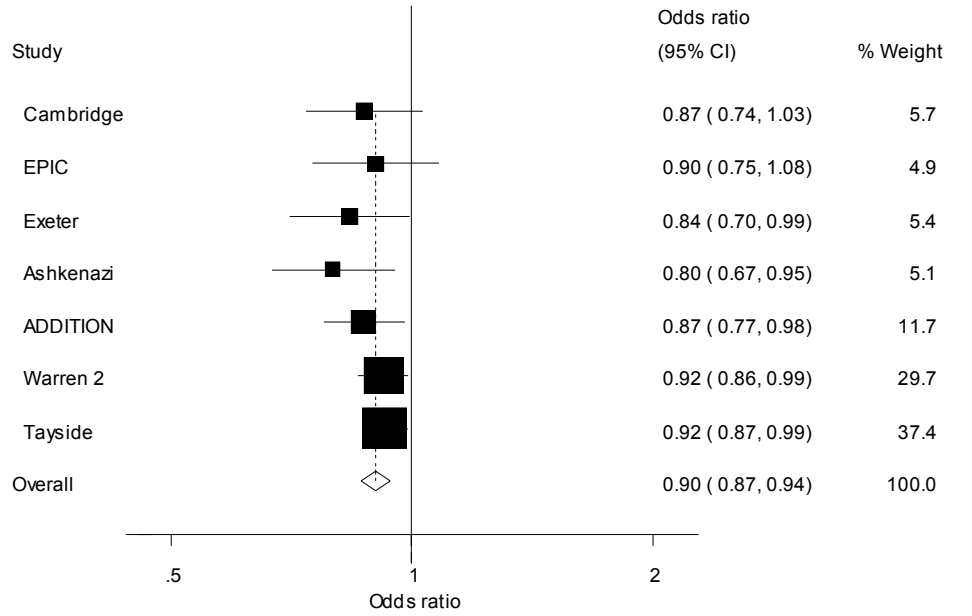


Figure 6.3 Meta-analysis of *WFS1* SNPs in all seven case-control studies, (a) Association between SNP rs10010131 and risk of type 2 diabetes in individual studies. (b) Association between SNP rs6446482 and risk of type 2 diabetes in individual studies.

6.3.1.7 Discussion

These analyses provide strong evidence for association of variation across *WFS1* and type 2 diabetes risk. Of 1536 SNPs across 84 candidate genes involved in β-cell function, two SNPs in *WFS1* were associated with type 2 diabetes in three UK-based case-control studies and in an Ashkenazi case-control study (P<0.01). In a pooled analysis of UK and Ashkenazi studies, five of the six *WFS1* SNPs tested in the candidate gene association study were significantly associated with type 2 diabetes risk (P<0.05). A meta-analysis of these studies with three further UK-based case-control studies (9533 cases and 11389 controls in total) demonstrated robust replication of the association signal in *WFS1*, the minor allele of the strongest SNP conferring a protective effect (OR = 0.90) from type 2 diabetes ($P = 1.4 \times 10^{-7}$).

Our study provides more robust statistical support for a previously reported association between rs734312 and type 2 diabetes risk. Previously, Minton et al. sequenced *WFS1* in 29 type 2 diabetic subjects and identified 5 non-synonymous SNPs, including rs734312 (R611H) and rs1801208 (R456H) (Minton et al. 2002). In 152 parent-offspring trios, R456 and H611 alleles, and the R456-H611 haplotype demonstrated borderline significant overtransmission to affected offspring from heterozygous parents. Furthermore, the H611 allele and the R456-H611 haplotype occurred significantly ($P = 0.06$ and $P = 0.023$ respectively) more frequently in 323 type 2 diabetic patients than in 357 normoglycaemic controls.

WFS1 encodes wolframin, an endoplasmic reticulum (ER) membrane protein with a role in the maintenance of ER calcium homeostasis (Takei et al. 2006). Homozygous and compound heterozygous inactivating mutations in *WFS1* cause Wolfram Syndrome (MIM222300), characterised by diabetes insipidus, young onset non-autoimmune insulin-dependent diabetes mellitus, optic atrophy and deafness (Inoue

et al. 1998; Strom et al. 1998). WFS1 deficiency also causes glucose intolerance and decreased β-cell mass in mice, which is thought to occur as a result of ER stress-induced apoptosis (Ishihara et al. 2004; Riggs et al. 2005; Yamada et al. 2006). This indicates that variation in *WFS1* might influence risk of type 2 diabetes through its effects on β-cell survival.

6.3.2 Replication of *WFS1* SNPs rs10010131, rs6446482, rs752854 and rs734312 in the Västerbotten type 2 diabetes case-control study

Although there was strong evidence for association between SNPs in *WFS1* and T2D risk in the β-cell candidate gene study (Sandhu et al. 2007) this did not reach current thresholds of genome-wide significance ($5x10^{-8}$ or 0.05 corrected for 10^6 independent tests) therefore we sought further support for the association from an independent type 2 diabetes case-control study from the county of Västerbotten in northern Sweden. In collaboration with Paul Franks' group at the Department of Public Health and Clinical Medicine, Umeå University Hospital, Umeå, Sweden, I genotyped *WFS1* SNPs rs10010131, rs6446482, rs752854 and rs734312 in 1296 Swedish cases and 1412 Swedish controls. All SNPs had genotype call rates >98%, were in Hardy-Weinberg equilibrium ($P > 0.1$), and genotyping concordance rates were 100%. The LD between SNPs in control participants was generally lower than the values reported in other populations (Figure 6.4).

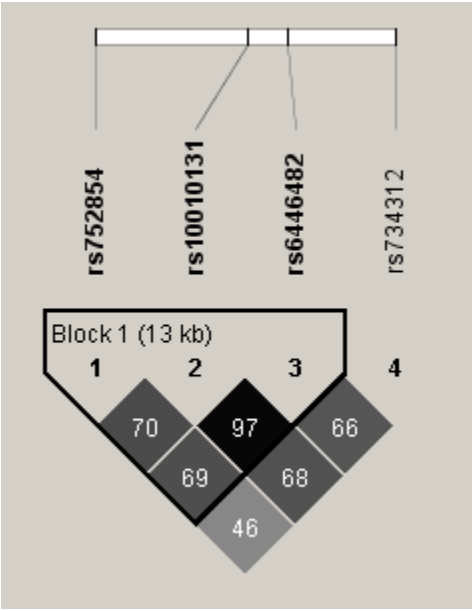


Figure 6.4 Linkage disequilibrium (r^2) between *WFS1* genotypes in the control samples of the Västerbottens type 2 diabetes case-control study (N=1,412)

SNPs rs10010131, rs6446482 and rs734312 showed borderline statistical association with type 2 diabetes (Table 6.7). The magnitude and direction of these associations were consistent with those found in UK and Ashkenazi study sets. SNP rs752854 was statistically associated with type 2 diabetes, with the minor allele conveying a protective effect (OR 0.85, 95% CI 0.75-0.96, *P* = 0.010; Table 6.7). This SNP was not the most strongly associated in UK and Ashkenazi samples, though this may reflect random statistical fluctuation rather than allelic heterogeneity between populations.

Table 6.7 Effect estimates for each of the *WFS1* SNPs in relation to glucose levels or type 2 diabetes in the Västerbottens type 2 diabetes case-control study

	Adjusted geometric means or Odds ratios (95% CI)		*P*-value
	Heterozygotes	Minor allele homozygotes	
rs10010131 (MAF = 0.43)			
Type 2 diabetes (yes vs. no)	0.87 (0.69, 1.01)	0.81 (0.63, 1.03)	0.083
rs6446482 (MAF = 0.44)			
Type 2 diabetes (yes vs. no)	0.93 (0.75, 1.16)	0.83 (0.65, 1.05)	0.098
rs752854 (MAF = 0.35)			
Type 2 diabetes (yes vs. no)	0.84 (0.64, 1.11)	0.72 (0.55, 0.96)	0.010
rs734312 (MAF = 0.48)			
Type 2 diabetes (yes vs. no)	0.81 (0.66, 1.00)	0.80 (0.64, 1.01)	0.066

Results are odds ratios (OR: type 2 diabetes) and 95% confidence intervals from additive genetic models. The frequency of T2D cases in participants homozygous for the major allele was used as the baseline risk and therefore the OR will be 1.00. Data are adjusted for age, sex and BMI. *P*-values are from tests for linear trend.

6.3.2.1 Meta-analysis of UK, Ashkenazi and Swedish type 2 diabetes case-control studies with data from genome-wide association studies

As well as data from the original UK and Ashkenazi study sets, data from three of the first five type 2 diabetes genome wide association scans were available for meta-analysis with the Swedish case-control study (Saxena et al. 2007; Scott et al. 2007; Sladek et al. 2007). We contacted relevant investigators of these GWA scans and requested summary statistics (odds ratio and 95% confidence intervals). SNP rs10010131 was available for analysis from the Finland-United States Investigation of Non-Insulin-Dependent Diabetes Mellitus Genetics (FUSION), but for Sladek et al. and the Diabetes Genetics Initiative (DGI) SNPs in high linkage disequilibrium with rs10010131 were used as proxies (Table 6.8). T2D data from the WTCCC was not included in the meta-analysis as this comprised the same samples already present in the Sandhu *et al.* paper.

Table 6.8 Details of GWA studies used in meta-analysis of *WFS1* SNP rs10010131

Study	Reference	Cases (N)	Controls (N)	SNP	Correlation with rs10010131 (r^2)*
FUSION	(Scott, Mohlke et al. 2007)	1,160	1,172	rs10010131	1.0
Sladek	(Sladek, Rocheleau et al. 2007)	686	669	rs4416547	1.0
DGI	(Saxena, Voight et al. 2007)	1,464	1,467	rs10012946	1.0

* = in the HapMap CEU panel.

The meta-analysis of the Västerbotten study, the original seven study sets, and the three additional genome-wide studies showed robust evidence for statistical association (OR = 0.89; $P = 5.4 \times 10^{-11}$; P for heterogeneity = 0.42 with 10 *df*, Figure 6.4).

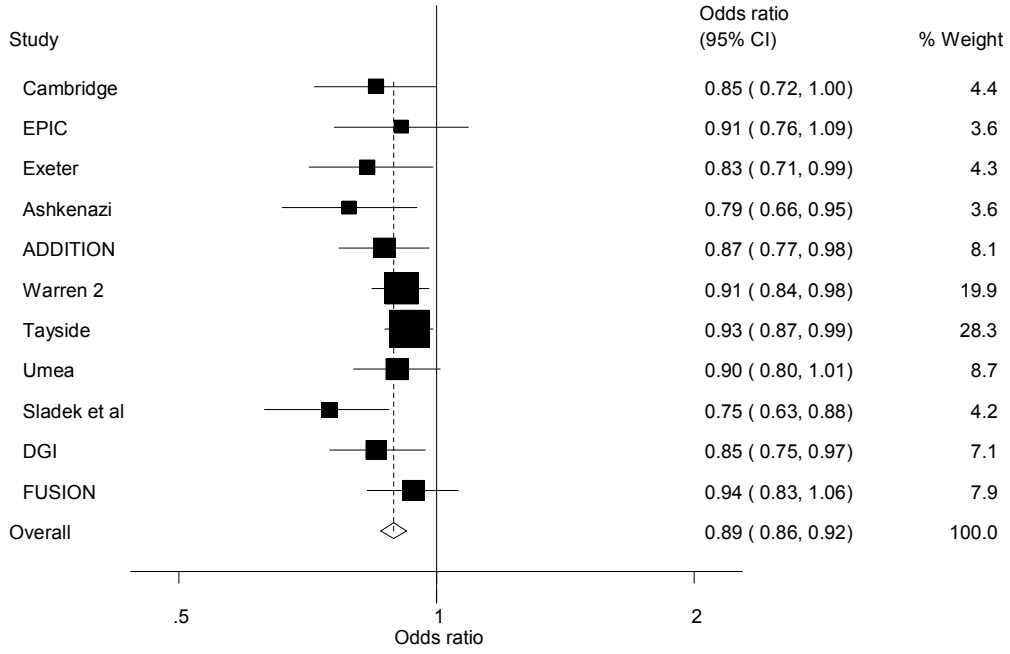


Figure 6.4 Association between *WFS1* SNPs and risk of type 2 diabetes in all seven original study sets, the Västerbotten study, and in data from three genome-wide association studies

Summary data for SNP rs10010131 was used in the meta-analysis with the exception of Sladek et al (Sladek et al. 2007) and DGI (Saxena et al. 2007) where rs4416547 and rs10012946 were substituted as proxy markers. These SNPs are perfectly correlated (r^2 = 1.0) with rs10010131 in HapMap samples.

6.3.2.2 Discussion

Four SNPs in *WFS1* were tested for association with T2D in the Västerbotten study (1296 Swedish cases and 1412 Swedish controls), and rs10010131 (or proxies in perfect LD) was tested for association with T2D in an updated meta-analysis comprising 11 different study cohorts of Northern and Western European and Ashkenazi Jewish ancestry. In the Västerbotten study, *WFS1* SNP rs752854 was statistically associated with T2D and effects of rs10010131, rs6446482 and rs734312 were similar in magnitude and direction to those previously reported (Sandhu et al. 2007). In the updated meta-analysis of data from the β-cell candidate gene study, the Västerbotten study, and three genome-wide association studies, the association of rs10010131 with T2D risk was replicated. Furthermore, the protective effect of the minor allele (OR = 0.89) reached genome-wide significance ($P = 5.4 \times 10^{-11}$). This work was published by Franks et al. (Franks et al. 2008).

The size and direction of the effects of *WFS1* SNPs rs752854, rs10010131, rs6446482 and rs734312 on T2D risk in the Västerbotten study were similar to those reported in the original β-cell candidate gene study (Sandhu et al. 2007). However, only rs752854 reached statistical significance (P<0.05). Power calculations indicate that this study had 48% probability of detecting the previously reported effect size of rs10010131 and rs6446482 (OR = 0.90) given MAF = 0.43 and 0.44 respectively, and 33% probability of detecting the effect of rs734312 (OR = 0.92) given MAF = 0.48. Therefore, the Västerbotten study on its own was underpowered to detect the previously reported effects of the *WFS1* variants.

There are several possible explanations for why rs734312 was the most strongly associated SNP in the Västerbotten study, while rs10010131 was the strongest SNP in the 6 UK and 1 Askenazi-based case-control studies. This discrepancy could simply be due to stochastic variation between these populations. Also, the difference

in statistical associations between studies could be attributable in part to different linkage disequilibrium patterns in the Västerbotten compared to the UK-based and Ashkenazi cohorts. This hypothesis is supported by the lower pair-wise LD between SNPs in the Västerbottens case-control study compared with the studies included in the original report. LD in the original study populations ranged between r^2=0.75-0.98 for pair-wise comparisons of the four SNPs, and the LD between the two SNPs showing the strongest statistical associations with type 2 diabetes were correlated at r^2=0.98. In the Västerbottens study, the LD between these two SNPs was similar (r^2=0.97), but the remaining pair-wise comparisons ranged between r^2=0.46-0.97.

By meta-analysing the Västerbottens study with the seven case-control cohorts from the Sandhu *et al.* study, and adding data from genome-wide association analyses in three additional populations, we increased power to detect effects of *WFS1* SNPs on diabetes risk. Identification of the true functional variant(s) behind the association between *WFS1* and T2D risk will require sequencing of *WFS1* to discover novel putative functional variants and further analyses of association with T2D.

6.3 Conclusion

The β-cell candidate gene study and replication study produced strong evidence for an association between *WFS1* variation and T2D risk (for rs10010131, OR = 0.89; P = 5.4 × 10−11). *WFS1* SNPs associated with T2D are highly correlated within a linkage disequilibrium block that covers the entire *WFS1* gene. It is therefore unknown whether these SNPs are individually contributing to disease risk or are simply associated with disease risk due to their correlation with tested or untested true causal variant(s). Fine-mapping of the association signal by resequencing of *WFS1* and further analysis in case-control studies will be required as a first step towards identifying functional variant(s).

6.4 Materials and Methods

6.4.1 Description of cohorts

More details of Cambridgeshire, EPIC, Exeter, ADDITION and Ely studies are provided in Chapter 2.

Discovery cohorts

6.4.1.1 Cambridgeshire case-control

This cohort comprised 552 patients aged 45–76 years with T2D and 552 controls matched for age and sex and general practice, recruited from general practitioner registers in Cambridgeshire, UK (Rathmann et al. 2001).

6.4.1.2 EPIC- Norfolk case-control study

This study is nested within the EPIC - Norfolk Study, a population based cohort study of white UK men and women aged 40-78 years and consists of 417 incident type 2 diabetes cases and two sets of 417 controls, each matched in terms of age, sex, general practice, recruitment date, with one set additionally matched for BMI.

6.4.1.3 Exeter case-control study

Six hundred and one individuals with type 2 diabetes were recruited from two sources: a consecutive-case series of patients with T2D diagnosed before 45 years from North and East Devon (Owen et al. 2003), and a collection of type 2 diabetes families that had either both parents available, or one parent and at least two siblings (Frayling et al. 1999). Six hundred and ten controls from the same geographic region were also used in this study.

6.4.1.4 Ashkenazi case-control study

Of the cases, 303 are from the multiplex-affected sibships that were ascertained for the genome scan described by Permutt et al, 2001 (Permutt et al. 2001). The cases were of Ashkenazi Jewish origin, defined as having all four grandparents born in Northern or Eastern Europe. Subjects with known or suspected Sephardic Jewish or non-Jewish ancestry were excluded. T2D was initially defined according to World Health Organization criteria (fasting glucose 140 mg/dl on two or more occasions, or random glucose 200 mg/dl on two or more occasions). Their average age at ascertainment was 60 years. Average age at diagnosis was 47 years and average duration of diabetes was 13 years (range 0-47). In this population, the incidence of type 1 diabetes is relatively low therefore anti-GAD or anti-islet cell antibody titers were not routinely measured. The additional 627 cases were ascertained as part of a study with the dual aim of looking for diabetes related genes and for genes related to the risk of developing diabetic complications. This group has an average age at ascertainment of 65.8 years, age of diagnosis of 46.8 years and duration of diabetes of 19.1 years. The Ashkenazi control samples consist of 149 elderly subjects (average age 76 years) with no personal or first-degree family history of T2D. The remaining 312 samples were obtained from The National Laboratory for the Genetics of Israeli Populations at Tel Aviv University, Israel. The institutional review boards of Washington University (St. Louis, MO) and Hadassah University Hospital (Jerusalem, Israel) approved the study.

Replication cohorts

6.4.1.5 Warren 2 case-control study

Informed consent was obtained from all participants. The subjects have been described in detail previously (Weedon et al. 2004). Briefly, all type 2 diabetes subjects were unrelated and of white UK origin who had diabetes defined either by WHO criteria (WHO Study Group 1999) or by being treated with medication for

diabetes, and were recruited from 3 sources: i) probands from type 2 diabetic sibships from the Warren 2 sibling pairs described previously (Minton et al. 2002; Wiltshire et al. 2001); ii) a collection of type 2 diabetes cases (Warren 2 cases) diagnosed between 35-65 years, but not selected on family history; iii) and probands from a collection of families that had either both parents available, or one parent and at least two siblings (Frayling et al. 1999).

Population control subjects were all white UK participants. These were recruited from 3 sources: i) the remaining parents from a consecutive birth cohort (Exeter Family Study) with normal (<6.0mmol/l) fasting glucose and/or normal HbA1c levels (< 6%; Diabetes Control and Complications trial corrected) (Minton et al. 2002); and ii) a nationally recruited population control sample of UK whites obtained from the European cell culture collection (ECACC), and iii) a follow-up study, that is ongoing, of all people born in Great Britain during one week in 1958 (National Child Development Study; http://www.cls.ioe.ac.uk/Cohort/Ncds/mainncds.htm). Cases and families where the proband had high GAD autoantibody levels (>99th percentile of the normal population) were excluded from the study. Known subtypes of diabetes (e.g. MODY) were excluded by clinical criteria and/or genetic testing.

6.4.1.6 ADDITION/Ely

The ADDITION case-control study comprised 926 cases of type 2 diabetes, aged 40-69, recruited from the UK Cambridge arm of the ADDITION trial (Lauritzen et al. 2000), and 1497 controls aged 35 to 79 years from the MRC Ely study (Wareham et al. 1998).

6.4.1.7 Tayside case-control study

The Tayside case-control study comprises 3,745 individuals with type 2 diabetes and 3,786 controls from the Wellcome Trust UK type 2 diabetes case-control collection (Go-DARTS2) which is a sub-study of Diabetes Audit and Research Tayside

(DARTS) (Doney et al. 2002; Doney et al. 2004a; Doney et al. 2003; Doney et al. 2005a; Doney et al. 2004b; Doney et al. 2005b; Morris et al. 1997). All cases were European with physician-diagnosed type 2 diabetes. They were recruited at primary or secondary care diabetes clinics or invited to participate from primary care registers from throughout the Tayside region of Scotland. The cases have not been characterised for GAD anti-bodies or MODY gene mutations. The controls were invited to participate through the primary care physicians or through their workplace occupational health departments. Controls did not have a previous physician-based diagnosis of diabetes. Control individuals with an HbA1c at recruitment of >6.2% or a fasting glucose of >=7 were removed from the analysis. All individuals in this ongoing study were recruited in Tayside between 1st October 2004 and 1st July 2006. For the purposes of this analysis, we excluded all participants below the age of 35 years, leaving 3,728 cases and 3,732 controls for analysis. The Tayside Medical Ethics Committee approved the study. Informed consent was obtained from all participants.

6.4.1.8 Västerbottens type 2 diabetes case-control study

Twelve-hundred-ninety-six adults with type 2 diabetes were identified through registries covering the county of Västerbotten in northern Sweden, and 1,412 non-diabetic individuals, group matched on age, sex, examination date and geographic region of residence, were selected from the Västerbotten Intervention Programme (VIP) as controls. Virtually all of these individuals were European whites. Type 2 diabetes was determined using the 1999 diagnostic criteria of the World Health Organization (World Health Organization: 1999). Participants with fasting capillary glucose concentrations <7.0mmol/l and no document history of diabetes underwent a 75g anhydrous oral glucose tolerance test. Accordingly, control subjects were those without a documented history of diabetes and with glucose concentrations below the thresholds for type 2 diabetes (World Health Organization: 1999). Type 2 diabetes in the case group was defined by clinical diagnoses. All living participants provided

written informed consent. Ethics permission was obtained from the Local Research Ethics Committee of Umeå University and approval for genetics investigations in this material was granted by the Swedish Data Inspection Board. Protocols for clinical measurements used in this study have been described previously (Franks et al. 2008).

6.4.2 Genotyping and Quality Control strategy

6.4.2.1 In Cambridgeshire, EPIC, Exeter and Ashkenazi case-control studies

1536 SNPs across 84 genes involved in pancreatic ß-cell function and survival (Table 6.1) were genotyped by P. Deloukas's laboratory at the Wellcome Trust Sanger Institute using Illumina's Golden Gate assay (Fan et al. 2003; Ke et al. 2004). DNA samples from all UK populations were whole-genome amplified by OmniPlexTM at Rubicon Genomics, Inc (Ann Arbor, MI, USA) and genomic DNA was used for the Ashkenazi samples. Samples with multiple SNP failures were repeated once and then excluded from clustering if their 50% locus Gene Call (GC) score was below 70%; these were thought to be samples of poor quality DNA. Clustering was performed on a per panel basis analysing no more than 500 samples at a time and used duplicate samples (two per 96-well plate) to improve clustering. We applied a locus cut off of 0.3 and cut off value in the genotype confidence score of 0.25. On a per panel basis we applied a call rate threshold ≥80% and removed markers that displayed more than 1 discrepancy per plate. Markers departing from Hardy Weinberg equilibrium ($\chi 2 > 10$) were flagged at this point. One thousand three hundred and sixty-seven (90%) of SNPs were polymorphic (minor allele frequency (MAF) of ≥ 0.1%), and met the following QC criteria: genotype call-rates greater than 90%; no significant difference in genotype call-rate between cases and controls; a P value ≥ 0.001 for tests of deviation from Hardy-Weinberg Equilibrium in controls. For those SNPs that survived QC where discordancy between replicate samples were found these samples were blanked out prior to analysis.

6.4.2.2 In ADDITION and Tayside studies

The ADDITION and Tayside studies were genotyped using a custom TaqMan® SNP assay (Applied Biosystems, UK) at Strangeways Research Laboratory, University of Cambridge, and at the Biomedical Research Centre, University of Dundee, respectively. Allele calling was done on the ABI PRISM® 7900HT Sequence Detection System (Applied Biosystems, UK). No SNPs deviated from Hardy-Weinberg equilibrium, tested in controls (P>0.05). Genotype call rates were >99%.

6.4.2.3 In the Warren 2 study

The Warren 2 study was genotyped by Kbiosciences (Herts, U.K.) using a KASPar assay system (details of the methods used can be found at http://www.kbioscience.co.uk). No SNPs deviated from Hardy-Weinberg equilibrium, tested in controls (P>0.05). Genotype call rates were >95%.

6.4.2.4 In the Västerbottens study

Västerbottens study genomic DNA samples were diluted to 4ng/µl. SNPs were assayed using the Taqman® MGB chemistry (Applied Biosystems, Foster City, CA) as described in Chapter 2. Genotyping success and concordance rates were >98% and 100% for all SNPs, respectively.

6.4.3 Statistical analysis

Statistical analyses for the β-cell candidate gene association study were conducted using Stata v8.2. Hardy-Weinberg was assessed using the $\chi 2$ statistic (1 df). Standard log likelihood ratio tests were used to assess the contribution of individual SNPs under a log additive model (1 df) to risk of type 2 diabetes in logistic regression analysis. An admixture maximum likelihood permutation method was used to correct for multiple testing (Tyrer et al. 2006). We also used log likelihood ratio tests to assess whether statistically associated SNPs independently contributed to risk of type 2 diabetes, comparing the log likelihood of a nested model (2 df) with that of the

full model (3 df). For the pooled analysis, we used logistic regression with study as categorical covariate (equivalent to a Mantel-Haenszel model weighted by study). Heterogeneity among studies was assessed using the Q statistic. Statistical analyses for the Västerbottens study were conducted using the SAS software v9.1 (SAS Inst., Carey, NC). Hardy Weinberg Equilibrium (HWE) was assessed using the likelihood ratio test with 1 df. Conditional logistic regression models were fitted to assess the associations between each of the *WFS1* genotypes and type 2 diabetes. Models were adjusted for age, sex, and BMI. Meta-analysis of studies was performed using STATA v8.2 using a fixed effects model and inverse-variance-weighted averages of log odds ratios to obtain a combined estimate of the overall odds ratio. Between-study heterogeneity was assessed using the $\chi 2$ statistic. In all studies, linkage disequilibrium (LD), expressed as r2, was calculated using Haploview v4.0 (http://www.broad.mit.edu/mpg/haploview) and power calculations were performed using Quanto v1.1.1 (http://hydra.usc.edu/gxe).

Chapter 7

Investigation of *WFS1* common and rare variation for association with type 2 diabetes

7.1 Summary

Homozygous and compound heterozygous mutations in the Wolfram Syndrome gene 1 (*WFS1*) are associated with a rare syndrome including juvenile-onset non-autoimmune diabetes. In addition, it was recently discovered that risk of common type 2 diabetes is associated with common variants in *WFS1*, which map within a ~40kb linkage disequilibrium block on chromosome 4. In this study I attempted to refine the association signal by resequencing *WFS1* exons, splice junctions, UTR and putative regulatory regions in a subset of type 2 diabetes cases and disease-free controls, and performing an association study in three UK-based case-control studies. I also aimed to assess the contribution of rare (MAF<0.01) variation in *WFS1* to type 2 diabetes risk by deep resequencing of *WFS1* in 1235 cases and 1668 controls. These studies demonstrated association between type 2 diabetes and five previously untested *WFS1* SNPs, of which rs1046322 was the strongest (P = 0.008). However, due to high correlation between previously tested and untested SNPs it was difficult to refine the association signal to a smaller region. There was no statistical difference between incidence of type 2 diabetes in carriers and non-carriers of rare *WFS1* missense and nonsense changes. Nor was there a difference between carriers and non-carriers of rare synonymous changes, or rare variants with a high likelihood of having a functional effect on the protein. This suggested that rare variation in *WFS1* does not have large (OR>1.46) effect on risk of type 2 diabetes.

7.2 Introduction

7.2.1 WFS1 deficiency in humans and animal models

Wolfram syndrome (MIM 222300) is an autosomal recessive disorder characterised by diabetes insipidus, young onset non-autoimmune insulin-dependent diabetes mellitus, optic atrophy and deafness (Wolfram DJ 1938). Most patients carry loss-of-function mutations in the Wolfram syndrome gene 1 (*WFS1*), which encodes wolframin (Inoue et al. 1998; Strom et al. 1998). Over 100 mutations, including missense, nonsense and frameshift mutations, distributed throughout the gene have been described in Wolfram syndrome patients thus far (Cano et al. 2007), which appear to cause loss of function through depletion of wolframin rather than dysfunction of the protein (Hofmann and Bauer 2006).

Wfs1 knock-out mice (Ishihara et al. 2004) or mice with pancreatic ß-cell-specific deletion of Wfs1 (Riggs et al. 2005) show glucose intolerance and progressive pancreatic ß-cell loss. This phenotype appears to result from activation of ER stress pathways, impaired cell cycle progression, and enhanced apoptosis (Riggs et al. 2005; Yamada et al. 2006).

7.2.2 WFS1 has a role in ER calcium homeostasis and stress response

Wolframin is an endoplasmic reticulum (ER) membrane protein with nine transmembrane segments (Takeda et al. 2001). The C-terminal domain is located in the ER lumen, while the N-terminal domain extends into the cytoplasm (Hofmann et al. 2003). There is evidence that Wolframin functions as an ion channel or regulator of existing channels on the ER membrane (Osman et al. 2003) and that it positively modulates ER calcium uptake (Takei et al. 2006).

As described in Chapter 1, ER stress and the unfolded protein response (UPR) play a role in pancreatic ß-cell adaptation to the physiological demand for insulin and the

pathophysiology of insulin resistance, ß-cell failure and diabetes. All three ER stress pathways (PERK, IRE1, and ATF6) are activated by WFS1 deficiency in pancreatic ß-cells (Fonseca et al. 2005; Yamada et al. 2006). WFS1 is also transcriptionally up-regulated by ER stress inducing agents (Ueda et al. 2005), and contains a conserved sequence in its promoter region similar to the ER stress response element (ERSE) found in other components of the UPR (Kakiuchi et al. 2006; Ricketts et al. 2006). It is plausible, therefore, that WFS1 deficiency causes ß-cell apoptosis and glucose intolerance in mice and humans by triggering ER stress responses as a result of impaired ER calcium homeostasis and perturbing consequent cellular survival mechanisms such as the UPR.

7.2.3 Genetic variation in *WFS1* and type 2 diabetes (T2D)

As discussed in Chapter 6, I was involved in studies that demonstrated convincing association of common SNPs in *WFS1* with type 2 diabetes risk (Franks et al. 2008; Sandhu et al. 2007). However, there have been no attempts to refine the association signal or uncover the underlying functional variants. *WFS1* SNPs associated with type 2 diabetes were present in a block of high LD. The size of the interval between recombination hotspots flanking this block is ~ 68 Kb, defining a region in which the search for causal variants should start.

Another limitation of the association analyses conducted to date is that the common SNPs typed will not act as good proxies for rare variation in *WFS1*, and yet recent studies suggest that rare genetic variation with effects on complex traits that are intermediate between the effect size seen for common SNPs (OR<1.4) and fully penetrant Mendelian disease mutations (OR>2) can explain a substantial proportion of heritability (Bodmer and Bonilla 2008). Several publications from Hobbs and Cohen have reported enrichment of nonsynonymous mutations in candidate genes at one extreme of the population distribution of plasma lipoprotein traits (Cohen et al.

2004; Cohen et al. 2006; Romeo et al. 2007). Another study found an enrichment for rare nonsynonymous changes in monogenic obesity genes in obese compared to lean individuals (Ahituv et al. 2007). More recently, mutations in genes involved in renal salt handling were associated with lower blood pressure and protection from hypertension in a population-based cohort (Ji et al. 2008). There is also some evidence that intermediate frequency polymorphisms (MAF 0.01-0.05) contribute increased risk of disease compared to more common alleles (MAF>0.05). An intermediate frequency polymorphism (MAF ~ 0.02) in ANGPTL4 was associated with 10-15% lower triglyceride levels in population-based cohorts (Romeo et al. 2007). There is some anecdotal evidence that obligate carriers of Wolfram Syndrome mutations are more susceptible to T2D (Fraser and Gunn 1977). However, to my knowledge, there has been no systematic investigation of rare variation across the entire WFS1 gene for association with type 2 diabetes risk.

7.2.4 Aims

1. To attempt to refine the *WFS1* association signal by resequencing putative functional regions in a subset of type 2 diabetes case-control samples from the Sandhu *et al.* study and genotyping of newly discovered variants in additional case-control individuals.

2. To investigate whether rare variants within the coding sequence, splice sites, UTRs and conserved non-coding regions of *WFS1* contribute to type 2 diabetes risk.

7.3 Results and Discussion

7.3.1 Fine-mapping of *WFS1*

7.3.1.1 Identifying possible *WFS1* regulatory regions

Before sequencing *WFS1* in a subset of type 2 diabetes cases and controls, I looked for potential *WFS1* regulatory regions by identifying conserved sequences upstream of *WFS1* and within *WFS1* introns. Using different informatics software Sally Debenham (MRC Epidemiology Unit, Cambridge, UK) and I undertook multiple species alignments to look for evidence of conserved regions. Two regions in intron 1 appear to be well conserved: NCBI B36 coordinates 6325012-6325193 (181 bp) and 6325875-6326013 (138 bp), as well as two regions upstream of *WFS1*: NCBI B36 coordinates 6321756-6321858 (102 bp) and 6322195-6322297 (102 bp) (Figure 7.1).

7.3.1.2 SNP discovery

I sequenced *WFS1* exons, exon-intron boundaries, UTRs, and conserved upstream and intronic sequences in a subset of 96 Cambridgeshire case-control samples in order to detect known and novel *WFS1* SNPs (Table 7.1). These regions were considered more likely than non-coding non-conserved inter- and intra-genic sequence to harbour a true causative variant underlying the association with type 2 diabetes. Using this approach I identified 58 variants (Table 7.1), none of which mapped within conserved non-coding sequences. Nine SNPs altered the amino acid sequence, all but two of which (V333I and A559T) were predicted to have a damaging impact on protein function by at least one of SIFT, PolyPhen or PANTHER.

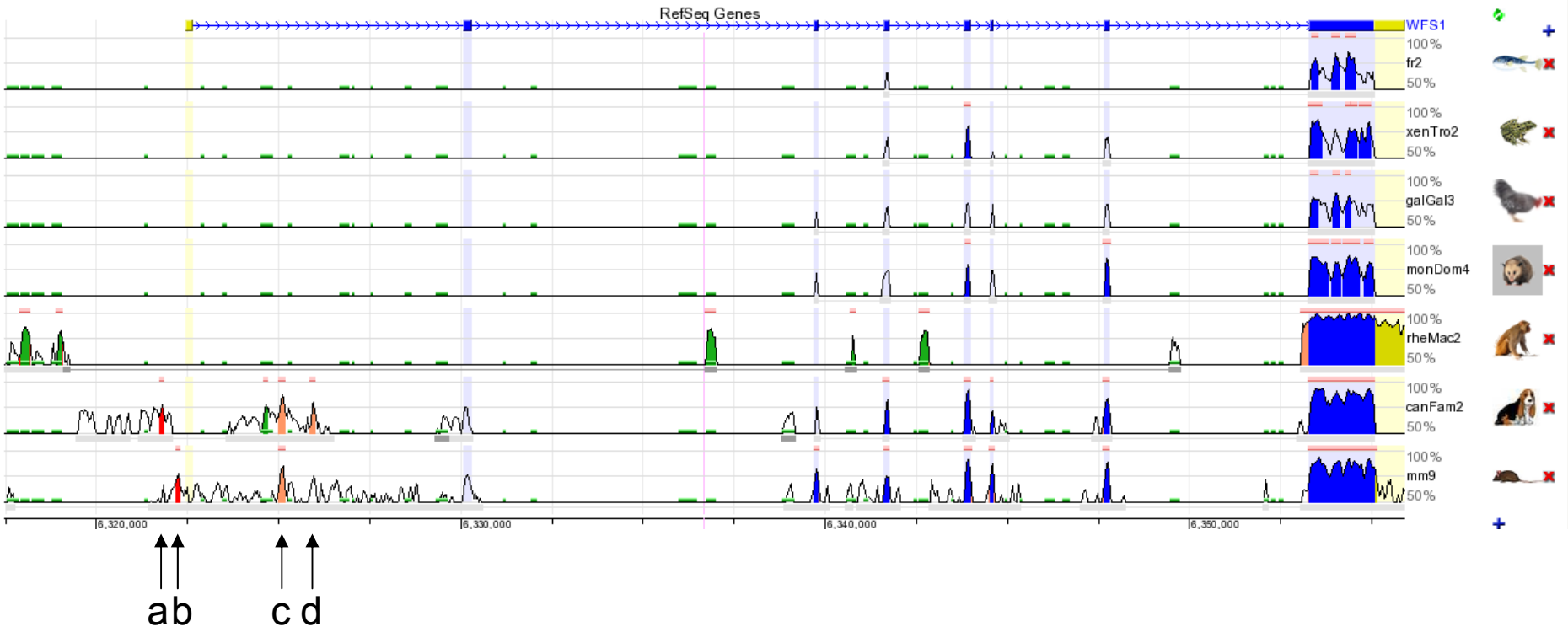


Figure 7.1 Evolutionary conserved regions (ECRs) in *WFS1* and 5000bp upstream
This figure was produced using the Dcode ECR browser (http://ecrbrowser.dcode.org/). Pink bars denote ECRs, blue bars denote exons and yellow bars denote UTR. The reference sequence is human and the graphs show sequence similarity to human in mouse, dog, monkey, opossum, chicken, frog, and fish (as shown on the right). Peak heights demonstrate the level of sequence similarity. White peaks indicate sequence with <80% similarity to human and <100bp in length, green peaks = transposons and simple repeats, blue peaks = exons, yellow peaks = UTR, salmon peaks = intron, and red peaks = intergenic sequence. Letters a and b = conserved upstream regions of 181 bp and 138 bp respectively, and c and d = conserved intronic regions of 102 bp.

Table 7.1 *WFS1* sequence variants detected in a subset of 96 Cambridgeshire case-control samples, with non-synonymous variants highlighted in blue

Genic position	Genomic position	Nucleotide substitution	Protein consequence	MAF in test samples	SNP ID
Upstream	6321944	T>A		0.33	rs4320200
Upstream	6321972	C>T		0.33	rs13107806
Upstream	6322051	G>C		0.34	rs13127445
Upstream	6322317	T>G		0.33	rs4273545
Intron 1	6324924	A>G		0.03	WFS1_1
Intron 1	6329948	T>C		0.20	rs10937714
Intron 2	6330405	A>G		0.50	rs28420833
Intron 3	6340039	A>T		0.02	WFS1_2
Intron 3	6341380	C>G		0.07	WFS1_3
Intron 3	6341421	G>A		0.01	WFS1_4
Intron 3	6341495	T>C		0.52	rs4688989
Intron 3	6341578	C>T		0.02	rs4688990
Intron 4	6341904	C>G		0.49	rs4689394
Intron 4	6343719	G>C		0.41	rs5018648
Intron 4	6343810	T>C		0.47	rs9998591
Intron 4	6343816	G>A		0.47	rs10010131
Exon 5	6343941	A>C	K193Q	0.01	WFS1_K193Q
Intron 5	6344138	G>C		0.47	rs9998835
Intron 5	6344253	C>T		0.47	rs10012946
Intron 5	6344351	C>T		0.47	rs13101355
Intron 5	6344378	G>A		0.48	rs13147655
Exon 6	6344597	G>C	R228R	0.30	rs7672995
Intron 6	6344703	T>C		0.12	rs7655482
Intron 6	6344739	G>A		0.40	rs11729672
Intron 6	6344746	G>A		0.01	WFS1_5
Intron 6	6344756	T>C		0.41	rs11725494
Intron 6	6344820	G>C		0.39	rs11725500
Intron 6	6344863	C>T		0.01	WFS1_6
Intron 6	6344868	A>G		0.39	rs4416547
Intron 6	6347348	G>T		0.07	rs12511742
Intron 6	6347438	G>A		0.02	WFS1_7
Exon 8	6353420	A>G	I333V	0.31	rs1801212
Exon 8	6353446	C>T	F341F	0.10	WFS1_F341F
Exon 8	6353608	T>C	V395V	0.45	rs1801206
Exon 8	6353717	C>G	L432V	0.01	rs35031397
Exon 8	6353731	C>T	T436T	0.01	WFS1_T436T
Exon 8	6353734	C>A	G437G	0.01	WFS1_G437G
Exon 8	6353790	G>A	R456H	0.04	rs1801208
Exon 8	6353923	T>C	N500N	0.42	rs1801214
Exon 8	6354098	G>A	A559T	0.01	WFS1_A559T
Exon 8	6354148	C>T	A575A	0.09	rs2230719
Exon 8	6354255	A>G	H611R	0.46	rs734312
Exon 8	6354475	G>A	A684A	0.01	WFS1_A684A
Exon 8	6354745	G>A	K774K	0.10	rs2230721
Exon 8	6354821	A>G	K800E	0.01	WFS1_K800E
Exon 8	6354856	G>A	K811K	0.48	rs1046314
Exon 8	6354875	C>T	R818C	0.01	rs35932623

Table 7.1 (continued). *WFS1* sequence variants detected in a subset of 96 Cambridgeshire case-control samples, with non-synonymous variants highlighted in blue

Genic position	Genomic position	Nucleotide substitution	Protein consequence	MAF in test samples	SNP ID
Exon 8	6354986	T>C	S855P	0.01	WFS1_S855P
Exon 8	6354988	A>G	S855S	0.36	rs1046316
Exon 8	6355034	G>A	V871M	0.02	WFS1_V871M
3'UTR	6355143	T>C		0.46	rs1046317
3'UTR	6355186	G>A		0.09	rs1802453
3'UTR	6355187	C>T		0.40	rs1046319
3'UTR	6355227	C>T		0.01	WFS1_8
3'UTR	6355245	G>A		0.50	rs1046320
3'UTR	6355349	G>A		0.12	rs1046322
3'UTR	6355370	A>G		0.03	rs1046325
3'UTR	6355384	C>T		0.01	WFS1_9

Genomic coordinates correspond to NCBI Build 36.

7.3.1.3 Selection of tagging SNPs

To assess whether any of the variants are highly correlated with the previously genotyped SNPs or whether any could act as proxies for one another, these 58 variants were uploaded into Haploview along with rs4689391, rs752854, rs6446482, and rs3821943 (which were genotyped in these samples as part of the original β-cell gene association study but were not covered by sequencing because they were outside the coding region and UTR). For reasons of power I decided to exclude rare SNPs (MAF<0.05) unless they altered the amino acid sequence, leaving 49 variants. Linkage disequilibrium between these 49 SNPs is indicated in Figure 7.2. Tagging SNPs were selected using an r^2 cut-off of 0.8, which generated 30 tagging SNPs. Seven of these had been genotyped as part of the original association study and so were removed from the selection (shown in red in Figure 7.2).

I then evaluated how well the selected tagging SNPs captured common variation in HapMap CEU trios within the linkage disequilibrium block containing the association signal (Rel 22/phase II April 2007) (Figure 7.3). This showed that 98% of common variants were captured and demonstrated that one additional SNP (rs12642481) was not well tagged. This SNP was force included into the tagging set (total number 24) to ensure all common variation was captured.

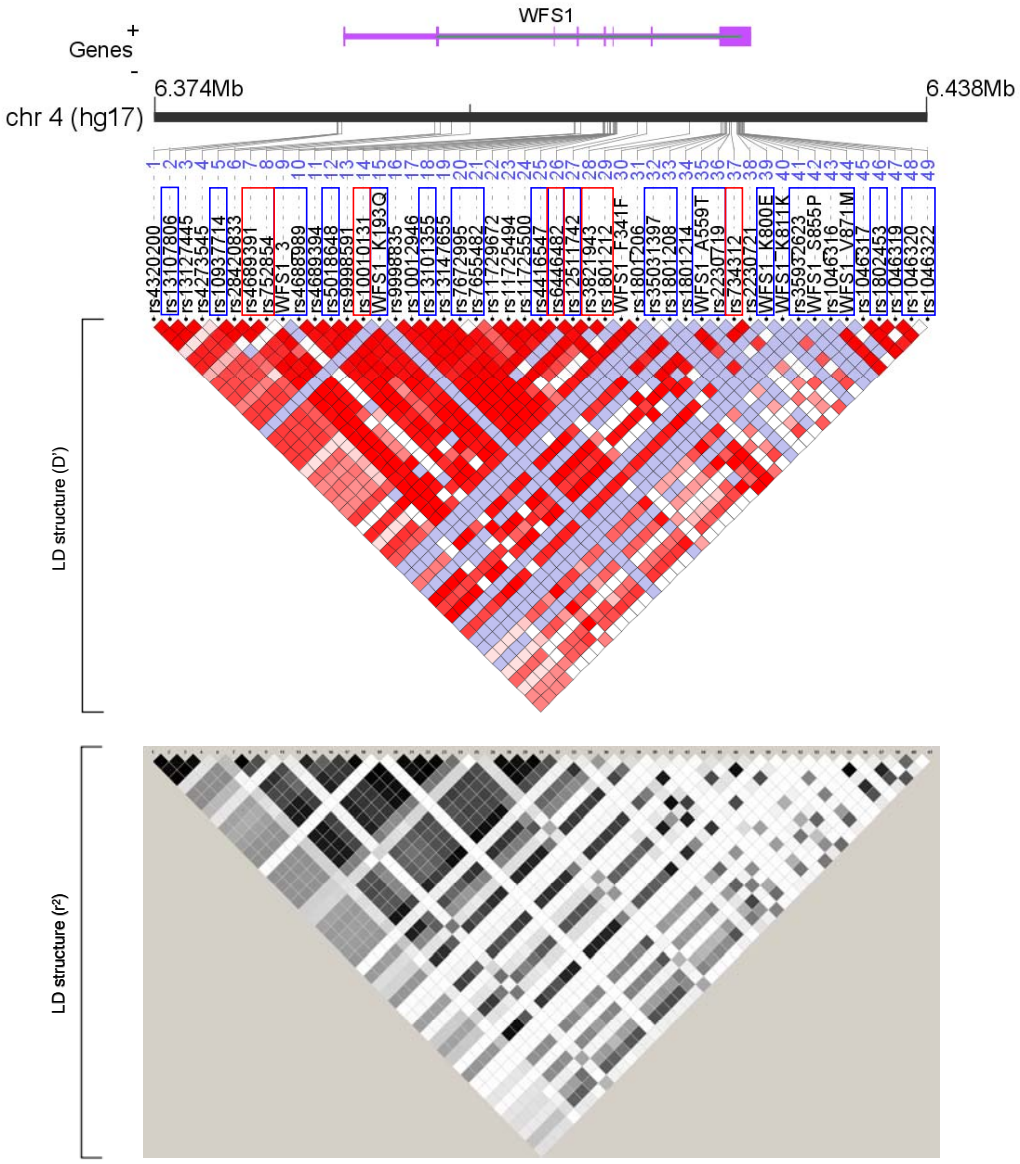


Figure 7.2 Feature map of the WFS1 gene showing SNPs discovered during resequencing and tagging SNPs

The positions of the 63 SNPs detected during sequencing of WFS1 in 96 Cambridgeshire cases and controls (including 4 additional SNPs genotyped during the original association study) are shown relative to the locus (purple) and chromosome 4 (black bar) (see text for details). The seven SNPs typed in the original studies are highlighted in red. Newly selected tagging SNPs are highlighted in blue. The bottom of the figure depicts two LD plots for the *WFS1* locus with pairwise LD values presented for SNPs. The upper plot presents LD as D' - see figure key for details. The figure was generated using LocusView (T. Petryshen, A. Kirby, M. Ainscow, unpublished software, available from the Broad Institute, Cambridge, MA (http://www.broad.mit.edu/mpg/locusview/). In the lower plot, LD among SNPs is given as r^2. r^2 values of 1.0 are represented by black diamonds, intermediate r^2 values are shown in grey and r^2 values of 0 as white. This plot was generated using Haploview (Barrett et al. 2005), available from the HapMap website (http://www.broad.mit.edu/mpg/haploview/index.php).

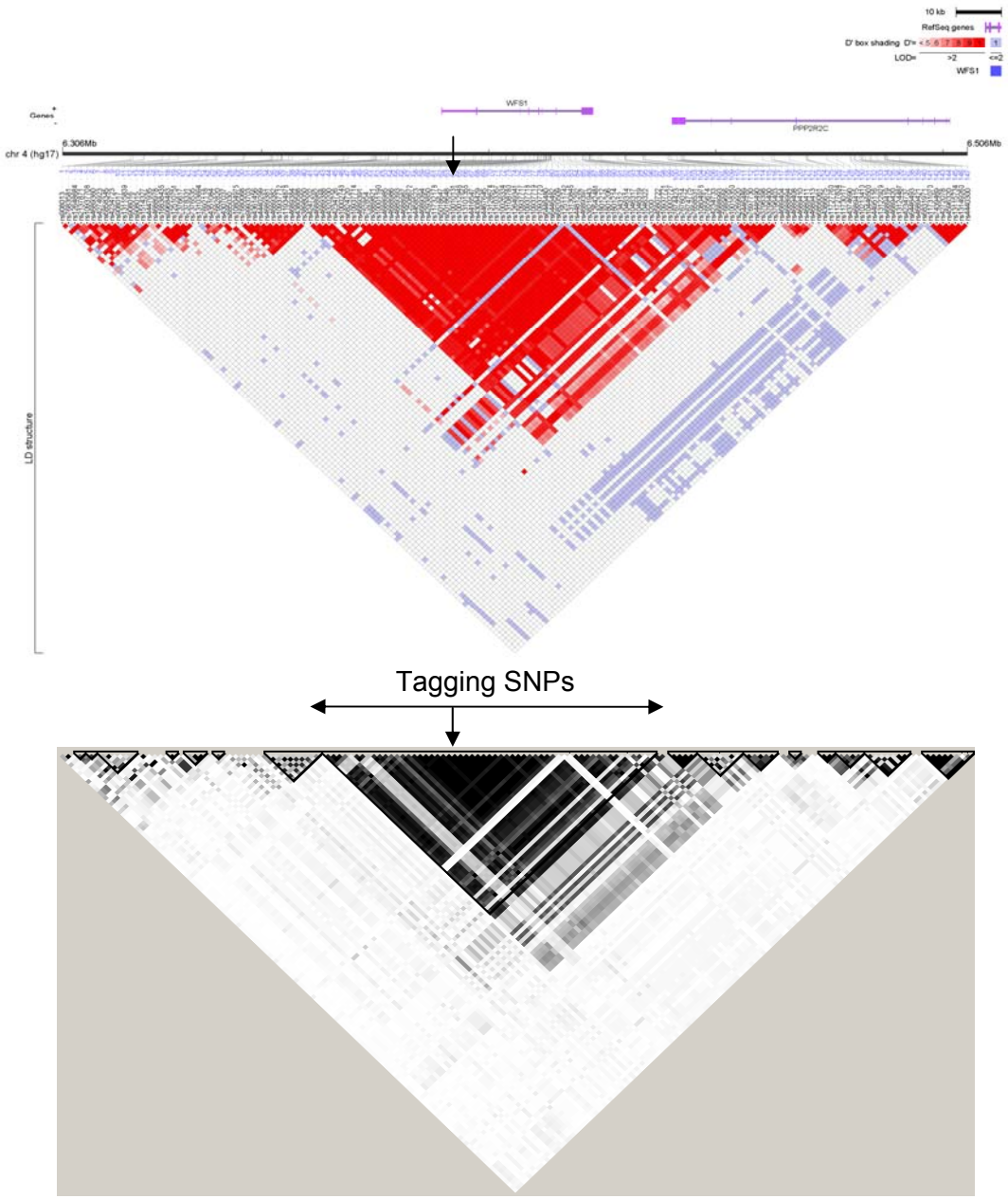


Figure 7.3 Feature map of the *WFS1* gene and flanking regions
The positions of SNPs genotyped in HapMap with a MAF≥ 0.05 are shown relative to known genes (purple) and chromosome 4 (black bar). The vertical arrows indicate the position of rs10010131, the SNP most significantly associated with risk of type 2 diabetes in the Sandhu *et al.* association study. The bottom of the figure depicts two LD plots for the *WFS1* locus with pairwise LD values presented for SNPs. The upper plot presents LD as D' - see figure key for details. The figure was generated using LocusView (T. Petryshen, A. Kirby, M. Ainscow, unpublished software, available from the Broad Institute, Cambridge, MA (http://www.broad.mit.edu/mpg/locusview/). In the lower plot, LD among SNPs is given as r^2. r^2 values of 1.0 are represented by black diamonds, intermediate r^2 values are shown in grey and r^2 values of 0 as white. This plot was generated using Haploview (Barrett et al. 2005), available from the HapMap website (http://www.broad.mit.edu/mpg/haploview/index.php). The horizontal arrows indicate the SNPs covered by my tagging SNP set.

7.3.1.4 Differences in call rate between cases and controls

Out of the 24 tagging SNPs selected for genotyping in Cambridgeshire, EPIC, and Exeter case-control studies, 21 passed clustering analysis (Table 7.2). Of these 21 SNPs, four had genotyping call rates that were statistically different between cases and controls. When each case-control study was analysed separately, it was clear that this disparity was driven in large part by Exeter genotyping, in which the call rate in cases was generally lower than in controls. Cases and controls were mixed on the plates so this cannot be due to differences in the genotyping quality between plates. The differentially called SNPs are spread throughout the gene, suggesting that a small regional duplication or deletion more common in Exeter cases and disrupting a primer/probe binding site is unlikely to account for all the observed discrepancies. The discrepancies were also across most DNA plates suggesting that it is unlikely to be a technical error concerning just a few DNA plates. In future studies, this could be confirmed by running the products of these PCR reactions in Exeter on gels to confirm success, and by sequencing across probe annealing sites. A large copy number variant (CNV) encompassing the whole region containing these SNPs could account for the lower call rates, however there are no annotated CNVs in the neighbourhood of *WFS1*. Another possible explanation for the differential call rates between cases and controls is a difference in the quality of the DNA, as cases and control samples were collected and extracted in separate studies. For this reason I decided to analyse only Cambridgeshire and EPIC studies (Cases = 854, Controls = 1242). A fresh supply of Exeter case samples may be required before repeating this genotyping.

Table 7.2 Quality control analyses of *WFS1* tagging SNP genotyping in UK case-control studies

						P value for the difference in call rate between cases and controls			
SNP	Genomic position	Protein consequence	MAF in all	HWE in controls	Call rate (All)	All	CCC	EPIC	Exeter
rs13107806	6321972		0.427	0.0403882	0.9335558	0.0009354	0.0855243	0.946636	0.000163
rs10937714	6329948		0.214	0.4157014	0.9262743	0.8371749	0.7272015	0.0409929	0.0048078
WFS1_3	6341380		0.05	0.5584238	0.9250607	0.0195131	0.0167841	0.8424461	0.0209893
rs4688989	6341495		0.399	0.1109792	0.9323422	0.0667898	0.016269	0.1618674	0.0009759
rs5018648	6343719		0.409	0.155748	0.9232403	0.1975619	0.0061325	0.2640687	0.017124
WFS1_K193Q	6343941	K193Q	0.005	*	0.964199	0.0001448	0.1067635	0.2083644	0.0013955
rs13101355	6344351		0.399	0.0495723	0.9123179	0.0000222	0.0046028	0.1793162	1.99E-09
rs7672995	6344597	R228R	0.317	0.2093614	0.9195995	0.0042431	0.5791705	0.8232458	1.09E-09
rs4416547	6344868		0.393	3.53E-06	0.8149272	0.019327	0.9772268	0.0370271	0.0976032
rs12511742	6347348		0.069	0.6957411	0.9535801	0.0095157	0.2939324	0.7612433	0.0004025
rs12642481	6351959		0.318	2.20E-07	0.8692355	0.3561777	0.1168826	0.6498275	0.0069747
rs35031397	6353717	L432V	0.004	*	0.9438714	0.6986788	0.6859223	0.0810481	0.0030927
rs1801208	6353790	R456H	0.049	0.0714386	0.9308252	0.2332388	0.3690197	0.0112531	0.0009366
WFS1_A559T	6354098	A559T	0.005	*	0.9611651	0.0004924	0.0681048	0.3805831	0.0015129
rs2230719	6354148	A575A	0.074	0.5240097	0.9414442	0.0775613	0.5540094	0.7915878	0.0056427
rs35932623	6354875	R818C	0.025	0.7604563	0.8658981	0.7940395	0.0352786	0.0136996	1.97E-14
WFS1_S855P	6354986	S855P	0.0003	*	0.8728762	0.0024014			
WFS1_V871M	6355034	V871M	0.024	1.12E-11	0.8834952	0.1891271	0.8149501	0.0295664	1.15E-06
rs1802453	6355186		0.092	0.9014183	0.9189927	0.9042388	0.7798427	0.2840468	0.0004025
rs1046320	6355245		0.414	0.4446617	0.8992718	0.3226095	0.9473739	0.7271306	0.0001631
rs1046322	6355349		0.122	0.9209425	0.9505461	0.2538266	0.0876172	0.0687599	0.0024334

Statistically significant *P* values are indicated in red. * = not applicable due to low frequency of minor allele (no rare homozygotes).

7.3.1.5 Association of *WFS1* SNPs with type 2 diabetes risk in Cambridgeshire and EPIC case-control studies

Of the 24 SNPs selected for genotyping 17 (71%) passed genotyping quality control in Cambridgeshire and EPIC and were taken forward for analysis, along with the seven SNPs genotyped as part of the original candidate gene association study. WFS1_S855P was only present in two individuals (MAF=0.0003), one case and one control, and was therefore excluded from further analysis. Given the high linkage disequilibrium across this region, the remaining 16 tagging SNPs that generated good quality genotypes (plus those genotyped as part of the original candidate gene study) captured 81% of the common (MAF>0.05) *WFS1* variation in the Cambridgeshire case-control samples used for SNP discovery. The 16 tagging SNPs covered 98% of the common *WFS1* variation in HapMap CEU trios, leaving only one intronic SNP (MAF = 0.24) untagged.

Eight SNPs were nominally associated with T2D risk (P<0.05) in a pooled analysis of Cambridgeshire and EPIC studies (Table 7.3). SNP rs10010131 is still the most statistically significant SNP of those seven genotyped as part of the original candidate gene association study (P = 0.024). However, four of my 16 new tagging SNPs show stronger association with T2D risk in Cambridgeshire and EPIC, rs1046320 being the most statistically significant (P = 0.008). SNP rs1046320 is in the 3'UTR of *WFS1* and is in high LD with the other nominally associated SNPs in this gene (Table 7.4). ClustalW mutliple sequence alignments showed that the nucleotide is only conserved in primates, not dog, cow, mouse or rat. Likelihood ratio tests demonstrated that adding SNP rs1046320 to logistic regression models containing one of the other seven statistically associated SNPs did not significantly improve the fit of these models (data not shown). Also, none of the statistically associated SNPs improved the fit of the simpler model containing only rs1046320. These SNPs are all correlated

in Cambridgeshire and EPIC samples (Table 7.4), indicating that they may all be linked to similar extents with the real causal allele(s) (which could be either untested or amongst them).

Table 7.3 Association of *WFS1* tagging SNPs with T2D risk in a pooled analysis of Cambridgeshire and EPIC case-control studies

SNP	Protein consequence	MAF	Odds ratio (95% CIs)	*P* odds ratio*
rs13107806	Conserved upstream	0.427	0.90 (0.79 - 1.02)	0.111
rs10937714	Intron 1	0.212	0.93 (0.79 - 1.09)	0.354
rs4689391	Intron 2	0.423	0.90 (0.79 - 1.03)	0.113
rs752854	Intron 2	0.344	0.87 (0.76 - 1.00)	***0.048***
WFS1_3	Intron 3	0.051	0.89 (0.66 - 1.21)	0.457
rs4688989	Intron 3	0.402	0.86 (0.75 - 0.98)	***0.025***
rs5018648	Intron 4	0.412	0.85 (0.74 - 0.97)	***0.014***
rs10010131	Intron 4	0.398	0.87 (0.76 - 0.98)	***0.024***
WFS1_K193Q	K193Q	0.004	1.00 (0.36 - 2.81)	0.997
rs13101355	Intron 5	0.4	0.85 (0.75 - 0.97)	***0.018***
rs7672995	R228R	0.316	0.84 (0.73 - 0.97)	***0.017***
rs6446482	Intron 6	0.405	0.87 (0.77 - 0.99)	***0.033***
rs12511742	Intron 6	0.072	0.93 (0.72 - 1.20)	0.584
rs3821943	Intron 7	0.457	0.91 (0.81 - 1.03)	0.146
rs1801212	I333V	0.28	0.90 (0.78 - 1.03)	0.137
rs35031397	L432V	0.004	1.10 (0.39 - 3.09)	0.856
rs1801208	R456H	0.046	1.25 (0.92 - 1.69)	0.152
WFS1_A559T	A559T	0.005	0.66 (0.25 - 1.75)	0.395
rs2230719	A575A	0.076	0.92 (0.72 - 1.18)	0.512
rs734312	H611R	0.455	0.93 (0.82 - 1.05)	0.247
rs1802453	3'UTR	0.089	0.93 (0.74 - 1.17)	0.539
rs1046320	3'UTR	0.419	0.83 (0.72 - 0.95)	***0.008***
rs1046322	3'UTR	0.119	1.01 (0.82 - 1.23)	0.948

* = the outcome of a logistic regression analysis. Bold text indicates significant *P*-values. Blue text highlights the most significantly associated SNP from the original study cohorts described in Chapter 6 (rs10010131) and the most significantly associated SNP in Cambridgeshire and EPIC case-control fine-mapping studies (rs1046320).

Table 7.4 Correlations among *WFS1* SNPs associated with T2D in the Cambridgeshire and EPIC case-control studies

	rs752854	rs4688989	rs5018648	**rs10010131**	rs13101355	rs7672995	rs6446482
rs752854							
rs4688989	0.71						
rs5018648	0.696	0.988					
rs10010131	0.717	0.963	0.967				
rs13101355	0.702	0.987	0.995	0.962			
rs7672995	0.59	0.7	0.699	0.686	0.7		
rs6446482	0.684	0.923	0.923	0.955	0.919	0.656	
rs1046320	0.655	0.932	0.939	0.92	0.939	0.666	0.883

LD values are r^2, where 1 denotes complete correlation and 0 denotes no correlation. Blue text highlights the SNPs from the original study described in Chapter 6. Bold text reveals the most significant SNPs in the original and fine-mapping studies.

7.3.1.6 Imputing untyped or failed SNPs

Using LD patterns between variants in the 96 sequenced Cambridgeshire case-control samples, I was able to impute genotypes of 25 additional variants detected during sequencing in all Cambridgeshire and EPIC samples. I also used LD patterns in HapMap CEU trios to impute HapMap SNPs in the interval between recombination hotspots flanking the association signal (Figure 7.4). SNP rs1046320 was still the most strongly associated SNP in Cambridgeshire and EPIC studies, except for one rare (MAF = 0.016) intronic SNP (rs7691824), imputed from HapMap (P = 0.0057). This SNP will need to be genotyped in Cambridgeshire and EPIC as imputation in this case is unlikely to be accurate considering the low frequency of the variant and its low correlation with typed SNPs.

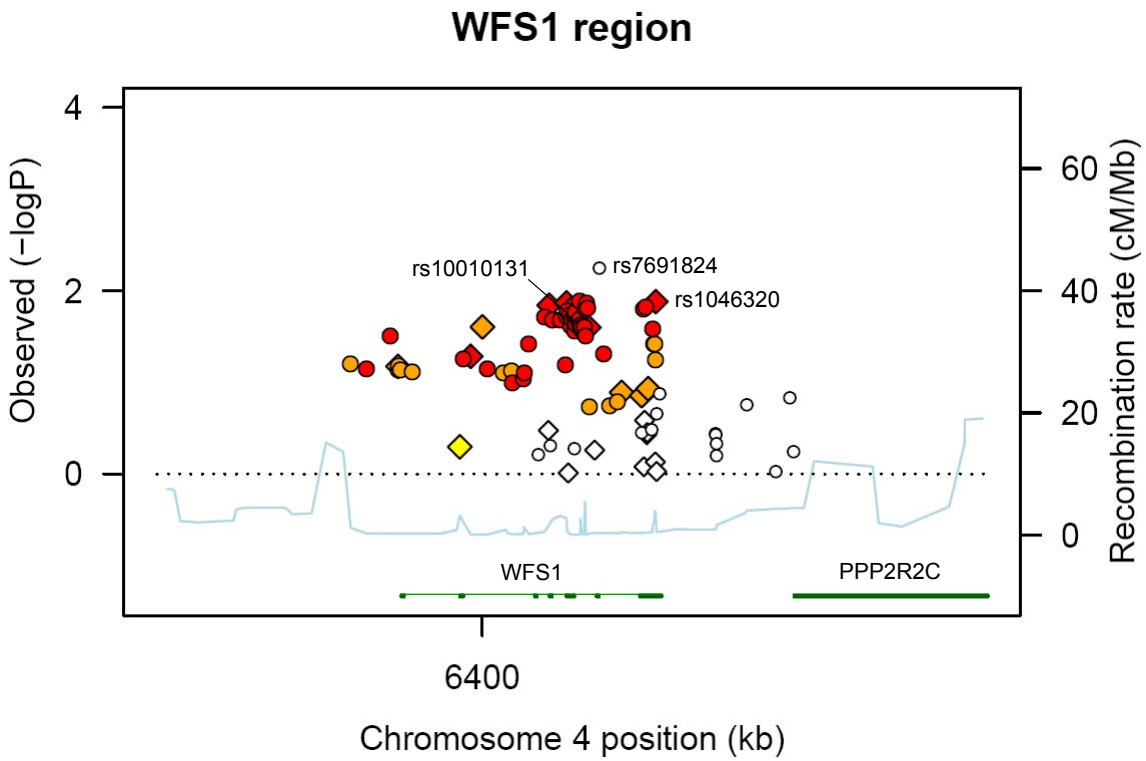


Figure 7.4 The statistical strength of the association of WFS1 tagging (diamonds) and imputed (circles) SNPs in the context of estimated recombination rates (blue line) and pairwise correlation between rs10010131 and surrounding markers
Red represents $r^2 > 0.85$, orange represents $0.5 < r^2 < 0.85$, yellow represents $0.2 < r^2 < 0.5$, and white represents $r^2 < 0.2$.

7.3.1.7 Combined analysis of rare variants in Cambridgeshire and EPIC

For very rare (MAF<0.005) non-synonymous variants WFS1_K193Q, rs35031397 (L432V), and WFS1_A559T, we had <80% power to detect effect sizes less than OR = 4.4, 4.5, and 3.7 respectively. Therefore, I tested whether the cumulative frequency of these variants influenced type 2 diabetes risk. In a combined analysis of WFS1_K193Q, rs35031397, and WFS1_A559T in Cambridgeshire and EPIC studies, there was no statistically significant difference in type 2 diabetes prevalence between carriers and wild-type individuals ($P = 0.709$).

7.3.1.8 Discussion

In this study I attempted to refine the previously reported association signal between *WFS1* variation and risk of T2D (Sandhu et al. 2007). I re-sequenced *WFS1* exons, splice junctions, UTR and putative regulatory regions in a subset of T2D cases (N=24) and controls (N=68) from the Cambridgeshire case-control study. I then selected tagging SNPs that covered common variation (MAF>0.05) and all non-synonymous variation detected by re-sequencing, as well as SNPs reported in HapMap CEU trios between recombination hotspots flanking the association signal. Tagging SNPs were genotyped and tested for association with T2D status in two UK case-control studies, Cambrideshire and EPIC case-control studies (854 cases and 1242 controls in total). Eight SNPs were nominally associated with T2D risk, five of which had not been tested in the Sandhu *et al.* study. Of these five previously untested SNPs, four showed stronger association with T2D than rs10010131. The strongest signal was from rs1046320 (P = 0.008). High correlation between these SNPs made it impossible to refine the association signal any further.

To test a denser set of variants across the region I imputed other variants discovered during sequencing of 96 Cambridgeshire case-control and other HapMap SNPs. Only one rare (MAF = 0.016) intronic SNP (rs7691824) imputed from HapMap showed stronger statistical association with T2D risk (P = 0.0057). This variant will need to be genotyped directly to confirm accurate imputation. The results from rs1046320 and rs7691824 need to be interpreted cautiously and repeated in other populations as, given that 89 variants were tested in the final analysis, the significance cut-off adjusted for multiple hypothesis testing using the Bonferroni correction is P = 0.000562.

This work demonstrates that while high linkage disequilibrium across regions of the genome is useful for minimising the amount of genotyping required to test the region for association with complex disease, it can compromise attempts to refine the association signal further. Discerning the underlying functional variants is particularly difficult when the surrounding variants are in nearly perfect linkage disequilibrium (r^2 > 0.9) because they give similar strengths of association. A more thorough approach could involve resequencing the entire interval between recombination hotspots (~68 Kb), rather than just those regions deemed most likely to harbour functional variation, and to analyse the sequence for copy number variations (CNVs) as well as SNPs. This would identify all possible genetic variants likely to impact disease risk. Also, as my search for conserved non-coding regions was restricted to *WFS1* intronic regions and 5 kb upstream and downstream of the gene, I will not have detected SNPs in potential regulatory regions towards the edges of the interval between recombination hotspots.

There is a risk that typing a more dense set of SNPs and CNVs may not add information due to high correlation between true functional variant(s) and other variants across the region. If this is so then studying populations with different and/or weaker patterns of linkage disequilibrium may help to refine the signal. For example, the LD block containing the *WFS1* gene is smaller in the HapMap samples of African descent, and correlation between SNPs is generally weaker (Figure 7.5). The LD between SNPs rs10010131 and rs1046320 is r^2 > 0.204 in YRI HapMap samples as opposed to r^2 > 0.92 in CEU samples. However, this study design carries certain caveats. The association signal in *WFS1* would need to be replicated in African populations, as the causal variants might not be present. Even if *WFS1* is a T2D susceptibility gene in Africans, the causal variant(s) (and those SNPs in LD with the causal variant(s)) may be different and would therefore be of limited value for refining the association in Europeans.

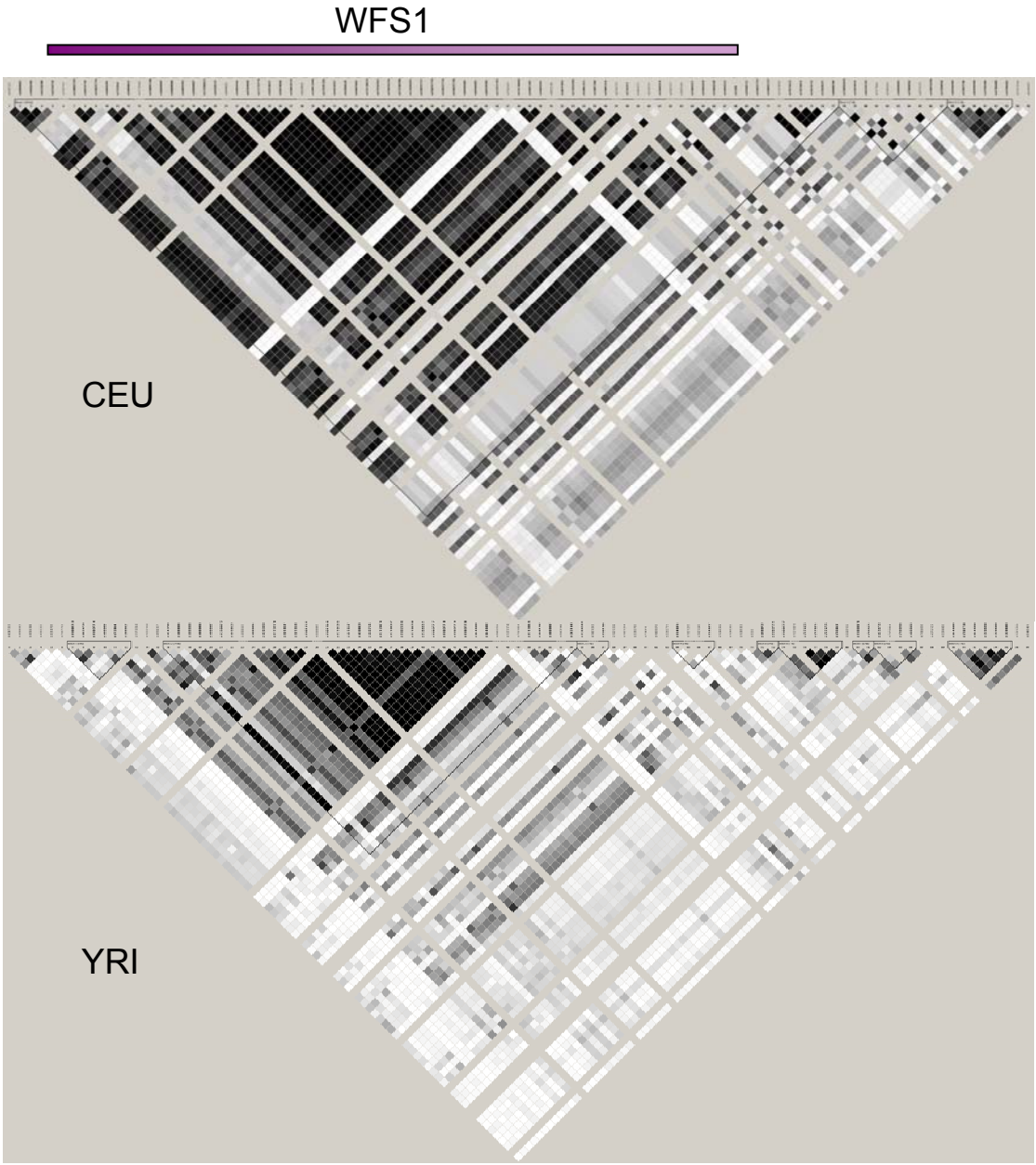


Figure 7.5 Patterns of linkage disequilibrium across the *WFS1* region in European (CEU) and African (YRI) samples
Presented are all SNPs in each population between NCBI build 36 coordinates 6315869 and 6379255. Gaps in the CEU and YRI LD plots represent SNPs not present in the respective samples. Linkage disequilibrium is measured by r2, with black diamonds representing high LD, white diamonds representing low LD, and grey diamonds representing intermediate levels of LD.

Another clue towards identifying true functional variants in *WFS1* would be the presence of eQTLs in the region – that is, genetic loci associated with changes in expression of *WFS1* or other genes in the region. However, no SNP or CNV within

the candidate interval has yet been found to be associated with gene expression variation in EBV-transformed lymphoblastoid cell lines from HapMap samples (GENEVAR http://www.sanger.ac.uk/humgen/genevar/).

Finally, haplotype analysis could be performed to test the joint actions of several SNPs across the *WFS1* region. It has been suggested that haplotype analyses may have better power than single SNP analyses to detect disease associations, as multiple SNPs in the haplotype may serve as better markers for the underlying risk allele(s). This approach might also help to focus resequencing efforts on individuals carrying a particular risk haplotype.

This study included several putative functional SNPs, including those that alter the amino acid sequence of Wolframin and those in highly conserved non-coding regions. However, none of the seven non-synonymous variants tested were associated with T2D risk in Cambridgeshire and EPIC studies. This could be because these variants do not affect risk of type 2 diabetes or the study could have been underpowered to detect their effect. I calculated that this study had <80% power to detect odds ratios less than 4.4, 4.5, 1.55 and 3.7 for SNPs WFS1_K193Q, rs35031397 (L432V), rs1801208 (R456H) and WFS1_A559T respectively. In a combined analysis of the very rarest non-synonymous SNPs (MAF≤0.005) I had <80% power to detect an odds ratio <2.55. Therefore, these variants cannot be ruled out as having a moderate impact on disease risk, but they are not causes of monogenic early-onset forms of diabetes as they were found in controls. Three SNPs found in an upstream conserved non-coding region by sequencing were tagged by rs13107806. However, this SNP was not statistically associated with T2D, indicating that upstream putative regulatory variants are not likely to contribute to risk of disease. The rs1046320 SNP is located in the 3'UTR and therefore could be affecting mRNA stability, processing and transport within the cell. Variants in the

3'UTR of genes have been found to impact disease. For example, a single nucleotide deletion in the 3'UTR of high mobility group A1 (*HMGA1*) gene reduces HMGA1 mRNA stability and expression and segregates with insulin resistance and type 2 diabetes in human subjects (Foti et al. 2005). More recently, it was suggested that SNPs in the 3'UTR of neurocalcin δ (*NCALD*) are associated decreased mRNA stability and risk of diabetic nephropathy (Kamiyama et al. 2007).

Four of the *WFS1* SNPs previously reported to be associated with T2D, rs4689391, rs3821943, rs1801212, and rs734312, did not reach statistical significance in this study. However, the direction and magnitude of their effects were similar. Given my sample size of 854 case-control pairs, I had between 28% and 35% power to detect an effect size OR 0.90 of SNPs with MAF between 0.28 and 0.48. Therefore, this study was statistically underpowered to detect the previously reported associations with these SNPs. Power could be improved by repeating the genotyping of the Exeter case-control study, as well as genotyping additional studies.

In conclusion, despite being statistically underpowered to detect the previously reported associations between *WFS1* SNPs and risk of T2D, I detected nominal associations with five previously untested SNPs, the strongest of which was rs1046320 (P = 0.008). Following imputation of HapMap SNPs, one imputed rare intronic SNP, rs7691824, was found to have even stronger association (P = 0.005). These SNPs will need to be genotyped in further case-control studies to investigate their impact on T2D risk.

7.3.2 Rare variant analysis

7.3.2.1 Resequencing of *WFS1*

I sequenced *WFS1* exons, exon-intron boundaries, UTRs, and conserved upstream and intronic sequences in the Cambridgeshire case-control study, the ADDITION study, and the MRC Ely cohort, which in total comprise 1668 controls and 1235 cases and 585 samples of unknown status (most of which were considered at high risk of developing type 2 diabetes). I detected 290 different sequence variants (Appendix Table A16) in these samples, 239 (82%) of which were novel. 235 (98%) of novel changes were rare (MAF<0.01) whereas only 15 (29%) of 51 known variants were rare, demonstrating the value of deep resequencing for identifying rare changes. 152 variants mapped within the coding region, of which 83 were non-synonymous, 66 were synonymous, and 3 were nonsense. There was a paucity of missense and nonsense changes with increasing minor-allele frequency, which is consistent with purifying selection acting on a significant fraction of such DNA sequence changes (Figure 7.6). Furthermore, there is an enrichment amongst low frequency variants for changes not detected in my sequencing of 96 Cambridgeshire case-control samples during the fine-mapping project (Figure 7.7). Though six variants were detected in non-coding regions with a high proportion of conserved residues, only two of these variants (in italics in Appendix Table A16) were actually conserved (the others falling between conserved nucleotides). Both were rare (MAF = 0.0003) and one was present only in cases and the other only in controls.

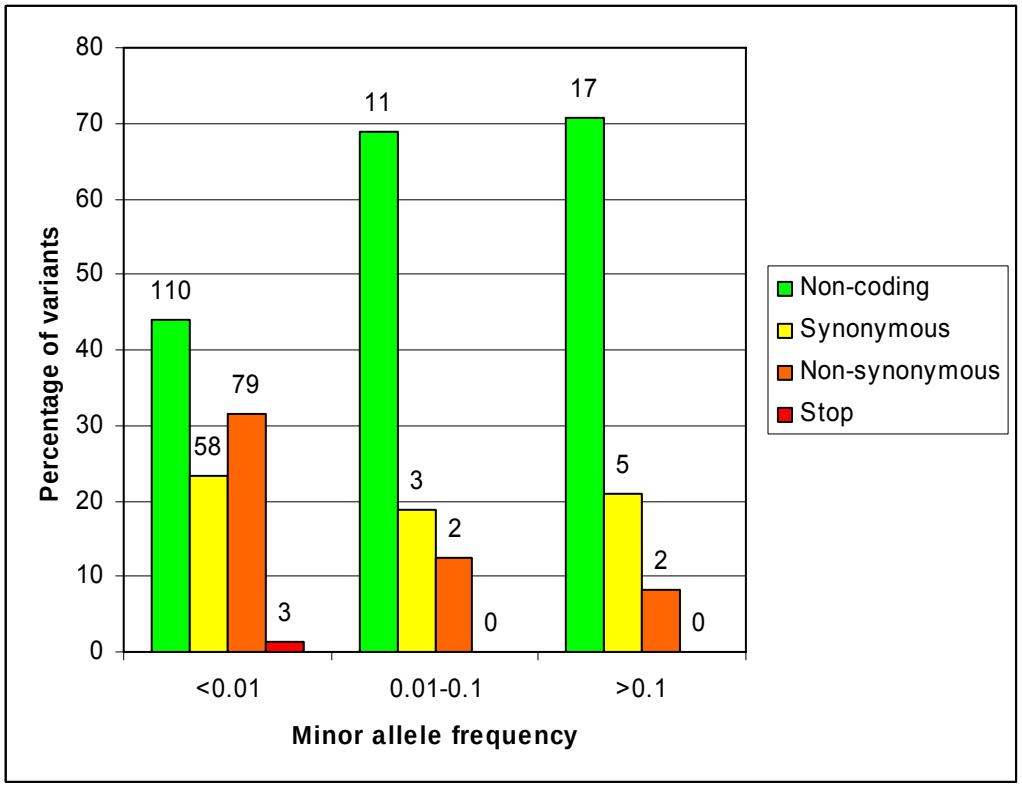


Figure 7.6 Distribution of types of *WFS1* variation discovered during resequencing of cases and controls at different minor allele frequency ranges
Numbers above the bars are the actual numbers found.

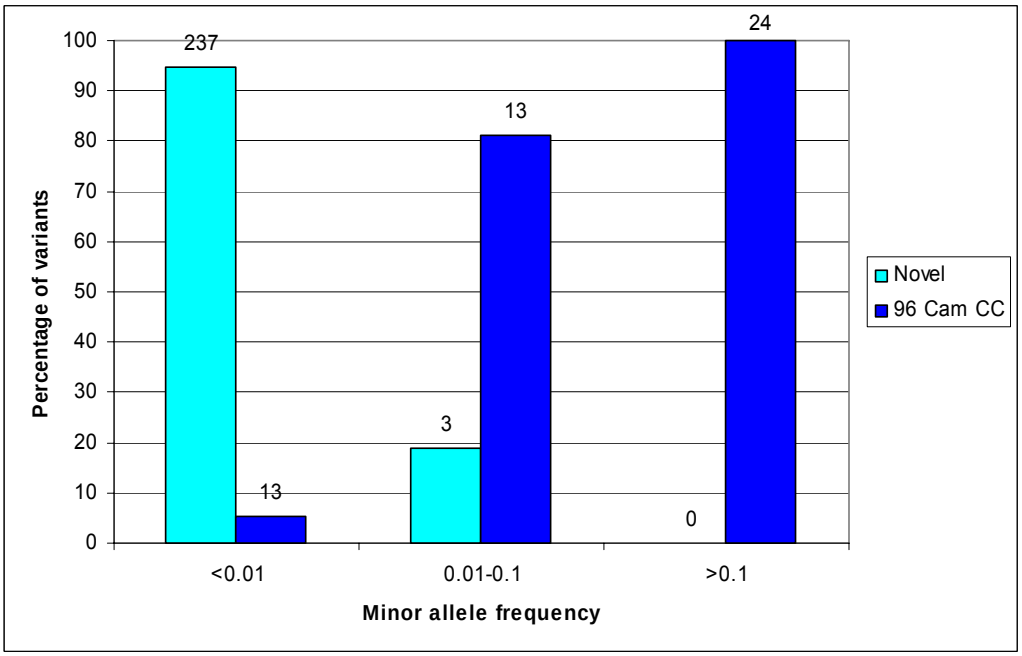


Figure 7.7 Distribution of novel and previously detected *WFS1* variation amongst different frequency ranges of changes discovered during resequencing of 1235 cases and 1668 controls
Actual numbers are shown above the bars.

7.3.2.2 Analysis of missense and nonsense variants with MAF<0.01

In my primary analysis I assessed the contribution of rare (MAF<0.01) missense and nonsense changes to risk of type 2 diabetes by comparing the odds of having type 2 diabetes in mutation carriers and non-carriers. A total of 82 missense and nonsense variants detected in our samples at a MAF<0.01 were included in this analysis (Appendix Table A17). Out of 2657 non-carriers, 1128 (42.45%) had type 2 diabetes, and out of 246 carriers, 107 (43.5%) had type 2 diabetes (Table 7.5). Therefore, there was no significant increase in risk of type 2 diabetes in carriers of rare missense and nonsense changes compared to non-carriers (OR = 1.04 (0.79-1.37), Fisher's exact $P = 0.788$).

Table 7.5 Number of cases and controls carrying missense or nonsense changes with MAF<0.01 vs wild-type

	Non-carriers	Carriers	Total
Controls	1,529	139	1,668
Cases	1,128	107	1,235
Total	2,657	246	2,903

OR = 1.04 (0.79-1.37), Fisher's exact $P = 0.788$.

As a small number of individuals carried more than one rare allele, I used logistic regression to assess the trend in the odds of disease with increasing number of mutations (Table 7.6). In this analysis, each additional mutation was associated with an extremely small and non-significant increase in risk of type 2 diabetes (OR = 1.01 ± 0.12, $P = 0.937$).

Table 7.6 Number of cases and controls carrying none, one, two, or three missense or nonsense changes with MAF<0.01

Number of mutations	0	1	2	3	Total
Controls	1,529	130	7	2	1,668
Cases	1,128	103	4	0	1,235
Total	2,657	233	11	2	2,903

OR = 1.01 ± 0.12, $P = 0.937$.

7.3.2.3 Analysis of synonymous variants with MAF<0.01

I decided to conduct a comparative study of synonymous variants, assumed to be functionally neutral, with MAF<0.01. This yielded similar results, though the effect sizes were larger. There was no significant change in odds of type 2 diabetes in carriers of at least one rare synonymous change compared to non-carriers (Table 7.7) (OR = 1.17 (0.81-1.68), P = 0.373), and no significant change in the odds per rare synonymous allele (Table 7.8) (OR = 1.32 ± 0.22, P = 0.089).

Table 7.7 Number of cases and controls carrying synonymous changes with MAF<0.01 vs wild-type

	Non-carriers	Carriers	Total
Controls	1,596	72	1,668
Cases	1,173	62	1,235
Total	2,769	134	2,903

OR = 1.17 (0.81-1.68), P = 0.373.

Table 7.8 Number of cases and controls carrying none, one, two, or three synonymous changes with MAF<0.01

Number of mutations	0	1	2	3	Total
Controls	1,596	72	0	0	1,668
Cases	1,173	55	6	1	1,235
Total	2,769	127	6	1	2,903

OR = 1.32 ± 0.22, P = 0.089.

7.3.2.4 Predicting variants with deleterious effects on the protein

Despite evidence that the majority of missense changes with MAF<0.01 have deleterious functional effects (Kryukov et al. 2007), I was concerned that I was diluting the effects of rare missense SNPs contributing to disease risk by analysing them with neutral missense changes. For this reason, I sought to identify non-synonymous and stop changes highly likely to impact on protein function and restrict the analysis to this group of variants (Table 7.9). I first looked for rare variants that had been shown biochemically to cause loss of function of Wolframin. R629W, W700X and P885L have all been shown to reduce the stability and half-life of wolframin (Hofmann and Bauer 2006) and have all been found in patients with

Wolfram Syndrome (Hardy et al. 1999; Hofmann and Bauer 2006; Kadayifci et al. 2001). Further variants with genetic evidence for involvement in Wolfram Syndrome include R558H (Colosimo et al. 2003), A559T and A671V (Smith et al. 2004), R708C (Tessa et al. 2001), E717K (Cryns et al. 2003), E776V (Smith et al. 2004) and R818C (Gomez-Zaera et al. 2001). Each missense variant was also entered into three different bioinformatics programs that predict functional impact based on sequence conservation and the biochemical properties of amino acids, SIFT, PolyPhen, and PANTHER. Furthermore, I carried out my own multiple sequence alignments to detect conservation of wild-type residues in monkey, mouse, rat, dog, chicken, frog, zebrafish, pufferfish and fruitfly. I noted that conservation in these multiple sequence alignments was not a good predictor of known inactivating *WFS1* mutations but biochemically proven mutations R629W and P885L were predicted damaging by all three bioinformatics programs. Therefore, I inferred 26 functionally important mutations based on a prediction of functional impact in SIFT, PolyPhen and PANTHER, and/or genetic/biochemical evidence for involvement in diabetes (Table 7.9).

Table 7.9 Known or inferred functional *WFS1* mutations

Chr:base	Variant	rs ID	Biochemical/genetic evidence	Pdel*	SIFT	PolyPhen	MAF in cases	MAF in controls	Conservation
4:6330207	R42X		Novel				0	0.0002998	
4:6343928	N188K		Novel	0.43033	affects protein	possibly damaging	0	0**	Low
4:6353402	L327F		Novel	0.55726	affects protein	possibly damaging	0.0004049	0	Vertebrate
4:6353502	C360Y		Novel	0.77597	affects protein	probably damaging	0.0008097	0	Vertebrate
4:6353739	F439C		Novel	0.59591	affects protein	probably damaging	0.0004049	0.0002998	Vertebrate
4:6353903	G494S		Novel	0.40523	affects protein	possibly damaging	0	0.0002998	Complete
4:6354096	R558H		Wolfram Syndrome (WS)	0.61604	affects protein	possibly damaging	0.0004049	0	Complete
4:6354098	A559T		WS and psychiatric disorders	0.34899	tolerated	benign	0.0048583	0.0029976	Low
4:6354105	I561S		Novel	0.44338	affects protein	possibly damaging	0	0.0005995	Low***
4:6354262	W613X		WS				0	0**	
4:6354306	T628M		Novel	0.74256	affects protein	possibly damaging	0.0004049	0	Vertebrate
4:6354308	R629W		WS & reduces half-life of wolframin	0.87775	affects protein	probably damaging	0.0004049	0	Low
4:6354435	A671V		WS and psychiatric disorders	0.21811	tolerated	benign	0.0012146	0	Low
4:6354443	G674R		Polymorphism	0.60981	affects protein	probably damaging	0.0004049	0	Low
4:6354449	R676C		Novel	0.75663	affects protein	probably damaging	0.0004049	0.0002998	Low
4:6354476	R685C		Polymorphism	0.84434	affects protein	probably damaging	0.0004049	0	Low***
4:6354522	W700X		WS & reduces half-life of wolframin				0	0**	
4:6354545	R708C		WS	0.77798	affects protein	probably damaging	0	0.0005995	Vertebrate
4:6354572	E717K		WS and psychiatric disorders	0.32653	tolerated	benign	0	0.0002998	Low
4:6354737	R772C		Psychiatric disorders	0.91098	affects protein	probably damaging	0	0.0008993	Low
4:6354750	E776V		WS	0.49302	affects protein	probably damaging	0.0040486	0.006295	Complete
4:6354792	S790W		Novel	0.71718	affects protein	possibly damaging	0	0.0002998	Low
4:6354875	R818C	rs35932623	WS and psychiatric disorders	0.68043	affects protein	possibly damaging	0.0048583	0.0053957	Low
4:6354917	R832C		Novel	0.75614	affects protein	probably damaging	0.0004049	0.0002998	Low***
4:6355061	D880N		Novel	0.49211	affects protein	possibly damaging	0	0.0002998	Vertebrate
4:6355077	P885L		WS & reduces half-life of wolframin	0.54691	affects protein	probably damaging	0	0.0002998	Complete

* Pdel score from PANTHER indicates the probability that an amino acid substitution will cause a deleterious effect on protein function based on alignment of evolutionarily related sequences (PANTHER classifies Pdel>0.38 as possibly deleterious) (continues on next page).

Table 7.9 legend continued.
** MAF = 0 in cases and controls shows that this variant was only found in samples of unknown disease status. N188K and W613X were found in ADDITION samples considered to be at high risk of developing diabetes (see Methods section), and W700X was detected in an Ely sample whose type 2 diabetes status was not recorded but quantitative trait data showed they had normal fasting glucose.
*** Amino acid with similar biochemical properties were conserved through evolution suggesting this locus may be of functional importance. Low = not well conserved. Vertebrate = conserved in all vertebrates. Complete = conserved in all species tested (monkey, mouse, rat, dog, chicken, frog, zebrafish, pufferfish and fruitfly).

7.3.2.5 Analysis of inferred functional variants with MAF<0.01

Out of 109 carriers of inferred functional mutations, 46 (42.2%) had type 2 diabetes, compared to 1189 cases (42.6%) in 2794 non-carriers (Table 7.10). This difference was not significant (OR = 0.99 (0.65-1.48), *P* = 1.00).

Table 7.10 Number of cases and controls carrying known and inferred functional *WFS1* mutations changes with MAF<0.01 vs wild-type

	Non-carriers	Carriers	Total
Controls	1,605	63	1,668
Cases	1,189	46	1,235
Total	2,794	109	2,903

OR = 0.99 (0.65-1.48), *P* = 1.00.

The trend in type 2 diabetes risk also decreased with increasing numbers of inferred mutations (Table 7.11).

Table 7.11 Number of cases and controls carrying none, one, two, or three known and inferred functional *WFS1* mutations with MAF<0.01

Number of mutations	0	1	2	3	Total
Controls	1,605	62	0	1	1,668
Cases	1,189	44	2	0	1,235
Total	2,794	106	2	1	2,903

7.3.2.6 Assessing association between disease status and a continuous measure of functionality of mutations

I decided to carry out an exploratory analysis to assess differences in the load of rare nonsynonymous variants between cases and controls, with mutations weighted by how likely they are to have deleterious effects on protein function. Instead of assigning carriers of mutations a score of 1 (as before), I weighted their score based on the PANTHER pdeleterious score for the mutation(s) they were carrying. In other words,

their score was now the sum total of the pdeleterious scores of all the rare (MAF<0.01) non-synonymous alleles they were carrying. To analyse differences between cases and controls I used a two sample T-test to assess the different in mean scores between case and control individuals. However, there was no significant difference between cases and controls (P = 0.5926).

7.3.2.7 The impact of intermediate frequency nonsynonymous SNPs (MAF 0.01-0.1) on risk of type 2 diabetes

Two nonsynonymous SNPs, V871M and R456H, had MAFs of 0.013 and 0.042 respectively. I had detected both SNPs during the fine-mapping study (Chapter 7.2.1) by sequencing a subset of 96 Cambridgeshire samples, but V871M failed genotyping and could not be imputed and R456H was not significantly associated with type 2 diabetes. I tested these in single SNP analyses in a pooled analysis of Cambridgeshire, ADDITION and Ely studies to assess association with type 2 diabetes risk. Neither V871M nor R456H were significantly associated with disease status (P = 0.132 and P = 0.249 respectively). This analysis had >80% power to detect effect sizes >1.93 and 1.45 for SNPs V871M and R456H respectively.

7.3.2.8 Discussion

Homozygous and compound heterozygous loss-of-function mutations in *WFS1* cause a Mendelian form of diabetes (Inoue et al. 1998; Strom et al. 1998), Wolfram Syndrome, and there is anecdotal evidence to suggest that obligate carriers of Wolfram Syndrome mutations have increased risk of type 2 diabetes (Fraser and Gunn 1977). Furthermore, numerous case-control studies have demonstrated association between polymorphisms in *WFS1* and risk of common type 2 diabetes (Franks et al. 2008; Sandhu et al. 2007), but were underpowered to detect moderate effect sizes of rare variants. Through deep resequencing of *WFS1* coding and conserved sequences in individuals with (N = 1235) and without (N = 1668) type 2 diabetes, I discovered 82 rare variants (MAF<0.01) which alter the amino acid sequence of Wolframin in 246 individuals. However, cases of type 2 diabetes were not significantly enriched amongst these rare variant carriers compared to non-carriers (P = 0.661). Carriers of rare variants deemed most likely to have a deleterious functional effect on the protein (based on web-based prediction algorithms and prior evidence for loss-of-function effects on the protein and/or co-segregation with Wolfram Syndrome) also did not have increased incidence of type 2 diabetes (P = 0.661). Nor was there a significantly different distribution of synonymous changes between cases and controls, as expected since these are predicted neutral.

Given the proportion of carriers in this case-control study (~8%), we had >80% power to detect OR>1.43 and >50% power to detect OR>1.29 (Power and Sample Size Program) (Dupont and Plummer 1990). This study was therefore well powered to detect previously reported effect sizes for rare variants on complex traits (the average being OR = 3.74) (Bodmer and Bonilla 2008). The impact of rare variants on risk of type 2 diabetes might have been diluted by pooling them with neutral rare variants for the analysis. Restricting

the analysis to those variants most likely to be functional reduced the frequency of the exposure (carrier status) to ~4%, but still retained >80% power to detect OR>1.65.

My study was underpowered to detect more modest effects (akin to those detected for common SNPs on T2D) of rarer variants. The Power and Sample Size Program (Dupont and Plummer 1990) indicates that a sample size of >22,000 cases-control pairs would be needed to have >80% power to detect effect sizes of rare missense and nonsense variants as low as OR = 1.09, assuming a similar proportion of carriers (8.5%) in the larger cohort. Also, studies with 80% power to detect modest effects (OR = 1.1) of SNPs with MAF=0.01-0.05 will require sample sizes between ~18,000-40,000 case-control pairs. However, it could be argued that such variants will not have an important impact on complex disease at a population-wide level.

In conclusion, I found no statistical enrichment for type 2 diabetes cases amongst individuals carrying at least one rare missense and/or nonsense change in *WFS1* compared to non-carriers. Given that my study was powered to detect effect sizes of OR>1.4, rare variants in *WFS1* are not likely to have an important impact on diabetes risk in UK populations.

7.4 Materials and Methods

7.4.1 Description of cohorts

7.4.2 Multiple sequence alignments

Sally Debenham used MultiPIP-maker (http://pipmaker.bx.psu.edu/pipmaker/) and VISTA MLAGAN (http://lagan.stanford.edu/lagan_web/index.shtml) to create alignments of the human *WFS1* genomic sequence and 5kb flanking regions and six other species (chimp, macaque, dog, cow, mouse and rat) to indicate the regions of conserved sequence. The human sequence was used as the reference sequence and was repeat masked. I identified two well conserved upstream regions using the Dcode ECR browser (http://ecrbrowser.dcode.org/).

7.4.3 PCR and sequencing

PCR, purification and sequencing of *WFS1* exons, exon-intron junctions, and UTR was performed using the standard protocol (Chapter 2.3.2). See Appendix Table A18 for primers and conditions. Sequencing of 96 samples from the Cambridgeshire case-control study were analysed, as part of the fine-mapping project, using Mutation Surveyor. Sequencing in the whole of the Cambridgeshire case-control study, ADDITION and MRC Ely studies was analysed using Gap4 (Chapter 2.3.6).

7.4.4 Genotyping

All 24 tagging SNPs passed assay design for genotyping on the Sequenom iPlex platform (Chapter 2.3.7.1.2). Primers and probes are listed in Appendix Table A19.

7.4.5 Quality control

Of 24 tagging SNPs genotyped in Cambridgeshire, EPIC and Exeter samples, three - rs7655482, rs1046316, and WFS1_K800E - were failed during manual confirmation of the genotype clusters. All remaining SNPs were checked for deviation from Hardy-Weinberg equilibrium (P<0.001), low call rates (N<85%) and significant discrepancy in call rate between cases and controls (P<0.001). Three SNPs, rs4416547, rs12642481, and WFS1_V871M, were not in Hardy-Weinberg and were not analysed. Except for rs35932623, all remaining tagging SNPs passed quality control in Cambridgeshire and EPIC case-control studies (Table 7.12). rs35932623 failed mostly in controls in Cambridgeshire and EPIC, but not Exeter. All failed SNPs except WFS1_K800E and WFS1_V871M were imputed, and WFS1_V871M was tested in Cambridgeshire and ADDITION/Ely samples as part of the rare variant analysis.

Table 7.12 QC in Cambridgeshire and EPIC samples

SNP	MAF	HWE in controls	Call rate	*P* difference*
rs13107806	0.427	0.04	0.918	0.02
rs10937714	0.212	0.98	0.902	0.557
rs4689391	0.423	0.06	0.921	0.473
rs752854	0.344	0.44	0.92	0.561
WFS1_3	0.051	0.12	0.904	0.027
rs4688989	0.402	0.14	0.914	0.291
rs5018648	0.412	0.11	0.902	0.333
rs10010131	0.398	0.02	0.998	0.521
WFS1_K193Q	0.004		0.959	0.005
rs13101355	0.4	0.05	0.898	0.072
rs7672995	0.316	0.09	0.905	0.771
rs6446482	0.405	0.17	0.997	0.711
rs12511742	0.072	0.53	0.949	0.275
rs3821943	0.457	0.08	0.996	0.65
rs1801212	0.28	0.22	1	0.407
rs35031397	0.004		0.93	0.664
rs1801208	0.046	0.06	0.908	0.06
WFS1_A559T	0.005		0.954	0.008
rs2230719	0.076	0.39	0.929	0.323
rs734312	0.455	0.08	0.988	0.739
rs35932623	0.027	1	0.85	**0.00001**
rs1802453	0.089	0.99	0.896	0.447
rs1046320	0.419	0.49	0.872	0.979
rs1046322	0.119	0.7	0.936	0.662

* between call rates in cases and controls.

SNPs and samples with call rate<0.9 were excluded from analysis of deep resequencing data. This lead to the elimination from analysis of 2 synonymous, 1 missense, and 2 non-coding variants out of a total of 322 variants detected. All these lost variants were rare (MAF<0.001). As manual editing of every common SNP call in sequence traces from ~3500 samples would have been too time consuming, only rare SNP calls were manually confirmed in raw sequence traces. All analysed SNPs were also tested for deviation from Hardy-Weinberg equilibrium and for statistically significant differences in call rate between cases and controls.

7.4.6 Statistical analysis

Statistical analyses were conducted using Stata v8.2. Hardy-Weinberg was assessed using the χ^2 statistic (1 df). Logistic regression was used to assess the contribution of individual SNPs under a log additive model (1 df) to risk of type 2 diabetes in the fine-mapping study, and to assess the trend in odds of type 2 diabetes in individuals with 0, 1, 2, and 3 rare variants in the rare variant analysis. Log likelihood ratio tests were also used to assess whether statistically associated SNPs independently contributed to risk of type 2 diabetes, comparing the log likelihood of a nested model (2 df) with that of the full model (3 df). The nested model contained only one SNP and the study cohort, whereas the full model contained an addition SNP to test if it contributes to disease independently of the variables in the nested model. For pooled analyses of Cambridgeshire and EPIC studies, and of Cambridgeshire, ADDITION and Ely studies, logistic regression with study as categorical covariate was carried out. The difference in odds of type 2 diabetes in carriers of rare variants vs non-carriers in the rare variant analysis was performed using Fisher's exact.

In all studies, linkage disequilibrium (LD), expressed as r^2, was calculated using Haploview v4.0 (http://www.broad.mit.edu/mpg/haploview) and power calculations were performed using Quanto v1.1.1 (http://hydra.usc.edu/gxe) and, for the analysis of type 2 diabetes in rare variant carriers vs non-carriers, the Power and Sample Size Program (Dupont and Plummer 1990).

7.4.7 Imputation

Imputation was performed by Eleanor Wheeler (Metabolic Disease Group, Wellcome Trust Sanger Institute). The best guess genotypes of SNPs were imputed with Cambridgeshire and EPIC separately using BIMBAM software

(http://stephenslab.uchicago.edu/software.html). SNPs that failed QC were not used to impute untyped SNPs but were instead imputed themselves.

Chapter 8

Discussion

8.1 Past

The work I describe in this thesis relied on a candidate gene approach to attempt to identify genetic variation involved in syndromes of insulin resistance and common complex type 2 diabetes and related traits. When I started my PhD in April 2005, genome-wide association studies (GWAS) were not feasible because of the prohibitive costs of large-scale genotyping and the absence of genome-wide genotyping arrays based on completed HapMap data. Instead two methods, linkage and candidate gene association, had already been used to identify disease genes. As described in Chapter 1, linkage studies in type 2 diabetes were disappointing as positive results did not generally replicate in other studies. This was also true across many other complex diseases and traits. For example, genomic loci identified through candidate gene or genome-wide linkage scans for coronary artery disease and myocardial infarction were largely non-overlapping and only one robustly replicated gene, *ALOX5AP*, has come out of these analyses (Hamsten and Eriksson 2008; Helgadottir et al. 2004). In type 1 diabetes the human leukocyte antigen (HLA) class II genes (thought to explain up to half of the heritability) were reproducibly linked to risk of disease, but linkage analysis was not successful in identifying loci with more modest effects (Smyth et al. 2006). These studies demonstrate that linkage analysis tended to overlook the very small effect sizes expected of complex disease genes.

By contrast, candidate gene association studies were suggested to have more statistical power to detect modest effects of genetic loci on complex disease (Risch and Merikangas 1996). Still, many reported susceptibility loci showed inconsistent evidence for association between studies, casting doubt on whether there were in fact many common causal alleles to be found. However, at the start of my PhD a growing number

of research groups were realising the importance of collecting large well-phenotyped sample sizes and of careful study design (such as appropriate matching of cases and controls to avoid the effects of population substructure) for replicating results and identifying novel loci. Also, awareness of the pitfalls of multiple hypothesis testing yielded more conservative interpretations of nominal significant findings. In type 2 diabetes, candidate gene association approaches succeeded in identifying and verifying several risk loci including *PPARγ* and *KCNJ11* (Altshuler et al. 2000; Gloyn et al. 2001). Association studies also helped to identify those type 1 diabetes susceptibility genes with small impacts on disease risk relative to the HLA locus. For example, the *INS*, *PTPN22*, and *CTLA4* genes were found to be associated with type 1 diabetes (Bell et al. 1984; Bottini et al. 2004; Nistico et al. 1996). Candidate gene studies also proved successful at identifying genes underlying more extreme phenotypes that demonstrated Mendelian patterns of inheritance, such as severe insulin resistance (Barroso et al. 1999; George et al. 2004; Savage et al. 2002).

Given the effectiveness of candidate gene studies for detecting genetic loci causing Mendelian disease and predisposing to common complex disease, and given the prohibitive costs of genome-wide approaches, I adopted a hypothesis-driven rather than a hypothesis-free study design to identify genes involved in severe insulin resistance and type 2 diabetes traits. I selected candidate genes based on their known or putative role in insulin action and/or secretion in animal models and human phenotypes.

In Chapter 3 I describe investigation of the lipin gene family, so chosen because Lpin1 is responsible for two independent mouse models of lipodystrophy and insulin resistance (Peterfy et al. 2001), and its expression levels correlate with insulin sensitivity and adiposity in mice and humans (Yao-Borengasser et al. 2006). I screened *LPIN1* in 158 patients with syndromes of severe insulin resistance, including 23 cases of

lipodystrophy, but detected no fully penetrant pathogenic mutations (Fawcett et al. 2008). *LPIN1* common variation was not statistically associated with insulin sensitivity in a population-based cohort but SNPs were nominally associated with BMI, blood pressure, cholesterol levels and risk of hypertension (Fawcett et al. 2008). These associations will need to be confirmed in further cohorts.

As presented in Chapter 4, I also screened members of the mTORC1 and mTORC2 complexes and *AS160*, which are important downstream components of the insulin signalling cascade, in insulin resistant patients. A nonsense mutation in *AS160* was shown to impair insulin-stimulated GLUT4 translocation and segregated with high peak-to-fasting insulin ratios in a pedigree of six genotyped individuals with five affected members. I recommend that *AS160* should be screened in families with similar syndromes of insulin action.

My interest in the *PARL* gene, which was identified during a screen for genes differentially expressed in obese, type 2 diabetic Israeli sand rats, was stimulated by reports of an association between a nonsynonymous SNP in *PARL* and plasma insulin levels in a US cohort (Walder et al. 2005). As described in Chapter 5, I did not replicate this association in UK populations (Fawcett et al. 2006). This demonstrates the importance of replication of association results to distinguish between true associations and statistical artefacts.

Finally, in Chapter 6 I described a large scale candidate gene association study of genes involved in pancreatic ß-cell function. This strategy lead to the discovery of another robustly replicated type 2 diabetes susceptibility gene, *WFS1* (Sandhu et al. 2007). Common variation within a linkage disequilibrium block encompassing most of the gene was associated with type 2 diabetes in different populations including UK, Ashkenazi and Swedish case-control studies (Franks et al. 2008; Sandhu et al. 2007).

8.2 Present

Four years later the field of complex disease genetics has undergone a revolution due to the availability of completed HapMap phase I and II data, relatively cheap genotyping technology, and the formation of large consortia able to pool samples to generate larger study sizes. Such advances made GWAS not only feasible but successful in identifying new complex disease loci. For example, an early GWAS identified a strong association between risk of age-related macular degeneration and an intronic SNP in the complement factor H (*CFH*) gene (Klein et al. 2005). Resequencing and fine-mapping lead to the identification of a non-synonymous SNP (Y402H) which was corroborated by two independent groups published in the same issue of Science (Edwards et al. 2005; Haines et al. 2005). Another early GWAS also identified a novel type 1 diabetes locus in the interferon-induced helicase (*IFIH1*) region (Smyth et al. 2006). As examined in Chapter 1, GWAS have also detected over a dozen type 2 diabetes susceptibility genes with robust evidence for replication. Towards the end of my PhD I was able to use publicly available data from GWAS to include in meta-analyses with my own candidate gene data. This not only increased the power of my studies to detect susceptibility loci but provided an independent dataset that could be analysed for evidence of replication.

Though GWAS have detected a number of reproducible susceptibility loci, this approach still has important limitations. In genome-wide studies, the hundreds of thousands of tests lead to substantial type 1 error but, on the other hand, adjustment for multiple tests results in a very stringent significance level ($P \sim 10^{-7}$), and inflation of the type 2 error rate. In other words, genome-wide studies will tend to capture the “low hanging fruit”, SNPs with larger effects on type 2 diabetes in the populations tested, but may ignore truly associated variants with more modest effects on disease risk. For example, the

candidate gene study described in Chapter 6 did not impose a very strict P value cut-off in discovery cohorts and consequently modest associations between WFS1 SNPs and type 2 diabetes were pursued in replication cohorts. However, WFS1 was not prioritised for replication in the first wave of GWAS. Another limitation of genome-wide approaches stems from the use of custom-made genotyping arrays, such as Affymetrix and Illumina SNP chips, which only cover between ~40-70% Phase II HapMap SNPs with $r^2 \geq 0.8$ (Dong et al. 2007). Therefore, many genes and functional non-coding regions will not be well covered in genome-wide association studies.

In contrast, more thorough characterisation of the variation in a given region can be achieved by a candidate gene study, which might resequence the gene to discover novel variation and population-specific patterns of LD. Though this can be an expensive and time-consuming study design for susceptibility gene discovery, it can act as a complimentary approach to GWAS. Indeed, candidate gene studies are beginning to be used to extend results from genome-wide analyses by focusing on the effects of known susceptibility genes in distinct subgroups, such as different ethnic groups or cohorts with data on metabolic quantitative traits, and on SNPs with less dramatic *P* values in GWAS.

8.3 And future

There are several outstanding challenges remaining in the field of complex disease genetics. Firstly, SNPs that have been associated with type 2 diabetes and other complex diseases to date are not necessarily causal variants but instead they represent genomic regions which are sometimes hundreds of kilobases away from known genes. The fine-mapping of association signals and the identification of true causal variants will be a necessary but potentially arduous task, especially when LD in the region of association is strong making it difficult to distinguish between the effects of different variants on disease risk. I experienced this difficulty while attempting to refine the

association signal between *WFS1* and type 2 diabetes (described in Chapter 7). I sequenced exons, splice junctions and conserved non-coding regions of *WFS1* in a subset of cases and controls to discover novel variation and genotyped tagging SNPs in 854 cases and 1242 controls. Though several previously untested SNPs were nominally associated with type 2 diabetes risk, high correlation between SNPs made it difficult to refine the signal any further. However, I did detect a nominal association between a previously untested 3'UTR variant and type 2 diabetes risk that was stronger than rs10010131 in Cambridgeshire and EPIC alone. This will require replication in further cohorts but could potentially impact protein function through mRNA stability, processing and transport. Identification of the true functional variant(s) involved in complex disease may require genotyping of variants across the region in populations characterised by different and/or weaker patterns of linkage disequilibrium and/or extensive resequencing efforts to cover the entire region between recombination hotspots flanking the association signal to make sure all putative disease variants are tested.

Secondly, an important goal of the complex disease genetics research community is to expand knowledge of the underlying biology of disease and non-disease states, and through this the identification of genes and pathways that could be targeted for therapeutic intervention. Type 2 diabetes susceptibility genes *PPARγ* and *KCNJ11* are proofs-of-principle as they are also targets for thiazolidinediones (insulin sensitising drugs) (Berger et al. 1996) and sulphonylureas (insulin secretagogues) (Sturgess et al. 1988) respectively. However, the biological function of certain predisposing genes and how they contribute to disease remains elusive. For example, very little was known about the fat mass and obesity associated gene, *FTO*, when it was detected in a GWAS. Bioinformatics analysis of the *FTO* sequence revealed that it shared motifs with members of the Fe(II)- and 2-oxoglutarate-dependent dioxygenase family, and appeared

to have a role in DNA methylation (Gerken et al. 2007; Sanchez-Pulido and Andrade-Navarro 2007). Studies of *FTO* expression showed its presence in human adipose tissue, where the protective genotype was associated with higher rates of lipolysis (Kloting et al. 2008; Wahlen et al. 2008). In rodent models *FTO* mRNA is also abundant in hypothalamic nuclei and its expression is regulated by nutrional status, suggesting a possible role in the regulation of energy balance (Fredriksson et al. 2008; Gerken et al. 2007). The story of FTO shows how results from GWAS can inspire new studies into the biological function of a gene, and potentially identify new pathways that can be targeted for drug discovery. Further studies in model organisms will of course be important in elucidating gene function, and may be required to discover the function of non-coding intergenic SNPs by chromosome engineering (Wallace et al. 2007).

A third challenge will be to elucidate the role of rare variants in complex disease. Candidate and genome-wide association study designs have thus far focused on common alleles (MAF>0.05) which do not effectively tag rarer variants. However, it is perfectly plausible that rare variants with moderate effects on disease risk that are somewhere between the effect size seen for common SNPs (OR<1.4) and fully penetrant Mendelian disease mutations, may collectively contribute a substantial portion of inherited susceptibility to complex disease. Such variants have already been found to influence risk of colorectal adenomas, reviewed in (Bodmer and Bonilla 2008), circulating lipid levels (Cohen et al. 2004; Cohen et al. 2006; Romeo et al. 2007), schizophrenia (Walsh et al. 2008) and blood pressure (Ji et al. 2008). These studies employed deep resequencing to compare the frequency of newly discovered and known rare variants in candidate genes between disease cases and controls as well as individuals at opposite extremes of continuous trait distributions. In a review of rare variant analyses in the literature, Bodmer and Bonilla show that odds ratios for rare variants are generally >2,

with an average of 3.74 (Bodmer and Bonilla 2008). As sequencing technology becomes faster and cheaper, future studies of this kind may test the entire genome rather than limiting themselves to candidate genes. I performed deep resequencing of *WFS1* in 1235 type 2 diabetes cases and 1668 controls and identified 83 rare (MAF<0.01) missense and nonsense changes. However, there was no statistically significant association between any single variant and disease status, and the cumulative frequency of these variants, weighted according to their likelihood of having deleterious effects on protein function, was not enriched in cases compared to controls. Furthermore, I found two nonsynonymous variants with moderate frequencies in the cohort (MAF>0.01 and <0.05) but these showed no statistical association with disease. Much larger sample sizes would be required to detect very modest effects (OR < 1.2) of rare variants on complex disease susceptibility. However, such variants are unlikely to importantly contribute to disease at a population-wide level and therefore such studies may not be cost-effective.

Finally, many studies of Mendelian disease and complex traits, including my own, have tended to focus on point mutations, SNPs and small insertions/deletions rather than larger structural genetic changes. CNVs are ubiquitous in the genome and are already reported to influence a few complex disease phenotypes such as familial breast cancer, autism and autoimmune diseases such as systemic lupus erythematosus (Fanciulli et al. 2007; Frank et al. 2007; Sebat et al. 2007; Willcocks et al. 2008). Structural variations have been shown to explain a substantial portion of the variability in gene expression in HapMap samples, providing a possible mechanism through which many CNVs might impact disease risk (Stranger et al. 2007). Global maps identifying CNV regions, such as the study finished in 2006 identifying ~1500 regions (Redon et al. 2006), will be important for future association studies of type 2 diabetes. Furthermore, a structural

variation analysis group will be looking for both common and rare (MAF<0.01) CNVs as part of the 1000 genomes project (Hayden 2008).

Copy number variations (CNVs) are also a well established cause of Mendelian disease. The development of the paired-end mapping technique for detection of structural variations will aid high-throughput screening of disease cases (Korbel et al. 2007).

Another form of variation understudied in my screening analyses and others are functional non-coding variants. The identification of functional non-coding regions for screening in disease phenotypes could therefore help to identify Mendelian disease mutations. Such studies should be greatly aided by completion of the ENCODE (ENCyclopedia Of DNA Elements) project which aims to catalogue structural and functional components of the genome, including non-coding and regulatory elements (Birney et al. 2007). This information could be used to screen non-coding regions of candidate genes in the severe insulin resistance cohort.

In conclusion, awareness of the caveats of candidate gene association studies and the introduction of genome-wide association studies in recent years has led to the detection of many type 2 diabetes susceptibility loci. Investigators in the field of complex disease genetics still face important challenges, namely the identification of causal variation in associated genomic regions, the investigation of rare variants for impact on disease risk, and the biological mechanisms behind disease association. With the help of new sequencing and genotyping technology, the forging of collaborations between research groups to share resources and expertise, and enthusiasm for elucidating the function of new susceptibility genes and variation, the next few years should see progress towards understanding the genetic architecture of common type 2 diabetes and Mendelian forms

of insulin resistance, as well as the biological pathways underlying their development. It is hoped that with this knowledge we can target therapies to effectively combat these devastating diseases.

References

2004. Finishing the euchromatic sequence of the human genome. *Nature* **431:** 931-945.
2005. A haplotype map of the human genome. *Nature* **437:** 1299-1320.
2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447:** 661-678.
2008. Diagnosis and classification of diabetes mellitus. *Diabetes Care* **31 Suppl 1:** S55-60.
Abe, H., N. Yamada, K. Kamata, T. Kuwaki, M. Shimada, J. Osuga, F. Shionoiri, N. Yahagi, T. Kadowaki, H. Tamemoto, S. Ishibashi, Y. Yazaki, and M. Makuuchi. 1998. Hypertension, hypertriglyceridemia, and impaired endothelium-dependent vascular relaxation in mice lacking insulin receptor substrate-1. *J Clin Invest* **101:** 1784-1788.
Abraham, R.T. 2004. PI 3-kinase related kinases: 'big' players in stress-induced signaling pathways. *DNA Repair (Amst)* **3:** 883-887.
Accili, D., J. Drago, E.J. Lee, M.D. Johnson, M.H. Cool, P. Salvatore, L.D. Asico, P.A. Jose, S.I. Taylor, and H. Westphal. 1996. Early neonatal death in mice homozygous for a null allele of the insulin receptor gene. *Nat Genet* **12:** 106-109.
Agarwal, A.K., E. Arioglu, S. De Almeida, N. Akkoc, S.I. Taylor, A.M. Bowcock, R.I. Barnes, and A. Garg. 2002. AGPAT2 is mutated in congenital generalized lipodystrophy linked to chromosome 9q34. *Nat Genet* **31:** 21-23.
Agarwal, A.K., J.P. Fryns, R.J. Auchus, and A. Garg. 2003. Zinc metalloproteinase, ZMPSTE24, is mutated in mandibuloacral dysplasia. *Hum Mol Genet* **12:** 1995-2001.
Agarwal, A.K. and A. Garg. 2002. A novel heterozygous mutation in peroxisome proliferator-activated receptor-gamma gene in a patient with familial partial lipodystrophy. *J Clin Endocrinol Metab* **87:** 408-411.
Ahituv, N., N. Kavaslar, W. Schackwitz, A. Ustaszewska, J. Martin, S. Hebert, H. Doelle, B. Ersoy, G. Kryukov, S. Schmidt, N. Yosef, E. Ruppin, R. Sharan, C. Vaisse, S. Sunyaev, R. Dent, J. Cohen, R. McPherson, and L.A. Pennacchio. 2007. Medical sequencing at the extremes of human body mass. *Am J Hum Genet* **80:** 779-791.
Al-Mosawi, Z.S., K.K. Al-Saad, R. Ijadi-Maghsoodi, H.I. El-Shanti, and P.J. Ferguson. 2007. A splice site mutation confirms the role of LPIN2 in Majeed syndrome. *Arthritis Rheum* **56:** 960-964.
Alessi, D.R., M. Andjelkovic, B. Caudwell, P. Cron, N. Morrice, P. Cohen, and B.A. Hemmings. 1996. Mechanism of activation of protein kinase B by insulin and IGF-1. *Embo J* **15:** 6541-6551.
Alessi, D.R., S.R. James, C.P. Downes, A.B. Holmes, P.R. Gaffney, C.B. Reese, and P. Cohen. 1997. Characterization of a 3-phosphoinositide-dependent protein kinase which phosphorylates and activates protein kinase Balpha. *Curr Biol* **7:** 261-269.
Ali, S.M. and D.M. Sabatini. 2005. Structure of S6 kinase 1 determines whether raptor-mTOR or rictor-mTOR phosphorylates its hydrophobic motif site. *J Biol Chem* **280:** 19445-19448.
Altshuler, D., J.N. Hirschhorn, M. Klannemark, C.M. Lindgren, M.C. Vohl, J. Nemesh, C.R. Lane, S.F. Schaffner, S. Bolk, C. Brewer, T. Tuomi, D. Gaudet, T.J. Hudson, M. Daly, L. Groop, and E.S. Lander. 2000. The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat Genet* **26:** 76-80.
Amr, S., C. Heisey, M. Zhang, X.J. Xia, K.H. Shows, K. Ajlouni, A. Pandya, L.S. Satin, H. El-Shanti, and R. Shiang. 2007. A homozygous mutation in a novel zinc-finger protein, ERIS, is responsible for Wolfram syndrome 2. *Am J Hum Genet* **81:** 673-683.

Araki, E., B.L. Haag, 3rd, and C.R. Kahn. 1994. Cloning of the mouse insulin receptor substrate-1 (IRS-1) gene and complete sequence of mouse IRS-1. *Biochim Biophys Acta* **1221:** 353-356.

Ardlie, K.G., L. Kruglyak, and M. Seielstad. 2002. Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet* **3:** 299-309.

Arthur, E.I., J. Zlotogora, I. Lerer, J. Dagan, K. Marks, and D. Abeliovich. 1997. Transient neonatal diabetes mellitus in a child with invdup(6)(q22q23) of paternal origin. *Eur J Hum Genet* **5:** 417-419.

Aulchenko, Y.S., J. Pullen, W.P. Kloosterman, M. Yazdanpanah, A. Hofman, N. Vaessen, P.J. Snijders, D. Zubakov, I. Mackay, M. Olavesen, B. Sidhu, V.E. Smith, A. Carey, E. Berezikov, A.G. Uitterlinden, R.H. Plasterk, B.A. Oostra, and C.M. van Duijn. 2007. LPIN2 is associated with type 2 diabetes, glucose metabolism, and body composition. *Diabetes* **56:** 3020-3026.

Babenko, A.P., M. Polak, H. Cave, K. Busiah, P. Czernichow, R. Scharfmann, J. Bryan, L. Aguilar-Bryan, M. Vaxillaire, and P. Froguel. 2006. Activating mutations in the ABCC8 gene in neonatal diabetes mellitus. *N Engl J Med* **355:** 456-466.

Balding, D.J. 2006. A tutorial on statistical methods for population association studies. *Nat Rev Genet* **7:** 781-791.

Barilli, A., R. Visigalli, R. Sala, G.C. Gazzola, A. Parolari, E. Tremoli, S. Bonomini, A. Simon, E.I. Closs, V. Dall'asta, and O. Bussolati. 2008. In human endothelial cells rapamycin causes mTORC2 inhibition and impairs cell viability and function. *Cardiovasc Res* **78:** 563-571.

Barrett, J.C., B. Fry, J. Maller, and M.J. Daly. 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21:** 263-265.

Barroso, I. 2005. Genetics of Type 2 diabetes. *Diabet Med* **22:** 517-535.

Barroso, I., M. Gurnell, V.E. Crowley, M. Agostini, J.W. Schwabe, M.A. Soos, G.L. Maslen, T.D. Williams, H. Lewis, A.J. Schafer, V.K. Chatterjee, and S. O'Rahilly. 1999. Dominant negative mutations in human PPARgamma associated with severe insulin resistance, diabetes mellitus and hypertension. *Nature* **402:** 880-883.

Barroso, I., J. Luan, R.P. Middelberg, A.H. Harding, P.W. Franks, R.W. Jakes, D. Clayton, A.J. Schafer, S. O'Rahilly, and N.J. Wareham. 2003. Candidate gene association study in type 2 diabetes indicates a role for genes involved in beta-cell function as well as insulin action. *PLoS Biol* **1:** E20.

Barthel, A. and D. Schmoll. 2003. Novel concepts in insulin regulation of hepatic gluconeogenesis. *Am J Physiol Endocrinol Metab* **285:** E685-692.

Becker, A.B. and R.A. Roth. 1990. Insulin receptor structure and function in normal and pathological conditions. *Annu Rev Med* **41:** 99-115.

Bell, G.I., S. Horita, and J.H. Karam. 1984. A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. *Diabetes* **33:** 176-183.

Bell, G.I., K.S. Xiang, M.V. Newman, S.H. Wu, L.G. Wright, S.S. Fajans, R.S. Spielman, and N.J. Cox. 1991. Gene for non-insulin-dependent diabetes mellitus (maturity-onset diabetes of the young subtype) is linked to DNA polymorphism on human chromosome 20q. *Proc Natl Acad Sci U S A* **88:** 1484-1488.

Berger, D., I. Barroso, M. Soos, G. Yeo, A.J. Schafer, S. O'Rahilly, and J.P. Whitehead. 2002. Genetic variants of insulin receptor substrate-1 (IRS-1) in syndromes of severe insulin resistance. Functional analysis of Ala513Pro and Gly1158Glu IRS-1. *Diabet Med* **19:** 804-809.

Berger, J., P. Bailey, C. Biswas, C.A. Cullinan, T.W. Doebber, N.S. Hayes, R. Saperstein, R.G. Smith, and M.D. Leibowitz. 1996. Thiazolidinediones produce a conformational change in peroxisomal proliferator-activated receptor-gamma:

binding and activation correlate with antidiabetic actions in db/db mice. *Endocrinology* **137:** 4189-4195.

Birnbaum, M.J. 1989. Identification of a novel gene encoding an insulin-responsive glucose transporter protein. *Cell* **57:** 305-315.

Birney, E. J.A. Stamatoyannopoulos A. Dutta R. Guigo T.R. Gingeras E.H. Margulies Z. Weng M. Snyder E.T. Dermitzakis R.E. Thurman M.S. Kuehn C.M. Taylor S. Neph C.M. Koch S. Asthana A. Malhotra I. Adzhubei J.A. Greenbaum R.M. Andrews P. Flicek P.J. Boyle H. Cao N.P. Carter G.K. Clelland S. Davis N. Day P. Dhami S.C. Dillon M.O. Dorschner H. Fiegler P.G. Giresi J. Goldy M. Hawrylycz A. Haydock R. Humbert K.D. James B.E. Johnson E.M. Johnson T.T. Frum E.R. Rosenzweig N. Karnani K. Lee G.C. Lefebvre P.A. Navas F. Neri S.C. Parker P.J. Sabo R. Sandstrom A. Shafer D. Vetrie M. Weaver S. Wilcox M. Yu F.S. Collins J. Dekker J.D. Lieb T.D. Tullius G.E. Crawford S. Sunyaev W.S. Noble I. Dunham F. Denoeud A. Reymond P. Kapranov J. Rozowsky D. Zheng R. Castelo A. Frankish J. Harrow S. Ghosh A. Sandelin I.L. Hofacker R. Baertsch D. Keefe S. Dike J. Cheng H.A. Hirsch E.A. Sekinger J. Lagarde J.F. Abril A. Shahab C. Flamm C. Fried J. Hackermuller J. Hertel M. Lindemeyer K. Missal A. Tanzer S. Washietl J. Korbel O. Emanuelsson J.S. Pedersen N. Holroyd R. Taylor D. Swarbreck N. Matthews M.C. Dickson D.J. Thomas M.T. Weirauch J. Gilbert J. Drenkow I. Bell X. Zhao K.G. Srinivasan W.K. Sung H.S. Ooi K.P. Chiu S. Foissac T. Alioto M. Brent L. Pachter M.L. Tress A. Valencia S.W. Choo C.Y. Choo C. Ucla C. Manzano C. Wyss E. Cheung T.G. Clark J.B. Brown M. Ganesh S. Patel H. Tammana J. Chrast C.N. Henrichsen C. Kai J. Kawai U. Nagalakshmi J. Wu Z. Lian J. Lian P. Newburger X. Zhang P. Bickel J.S. Mattick P. Carninci Y. Hayashizaki S. Weissman T. Hubbard R.M. Myers J. Rogers P.F. Stadler T.M. Lowe C.L. Wei Y. Ruan K. Struhl M. Gerstein S.E. Antonarakis Y. Fu E.D. Green U. Karaoz A. Siepel J. Taylor L.A. Liefer K.A. Wetterstrand P.J. Good E.A. Feingold M.S. Guyer G.M. Cooper G. Asimenos C.N. Dewey M. Hou S. Nikolaev J.I. Montoya-Burgos A. Loytynoja S. Whelan F. Pardi T. Massingham H. Huang N.R. Zhang I. Holmes J.C. Mullikin A. Ureta-Vidal B. Paten M. Seringhaus D. Church K. Rosenbloom W.J. Kent E.A. Stone S. Batzoglou N. Goldman R.C. Hardison D. Haussler W. Miller A. Sidow N.D. Trinklein Z.D. Zhang L. Barrera R. Stuart D.C. King A. Ameur S. Enroth M.C. Bieda J. Kim A.A. Bhinge N. Jiang J. Liu F. Yao V.B. Vega C.W. Lee P. Ng A. Shahab A. Yang Z. Moqtaderi Z. Zhu X. Xu S. Squazzo M.J. Oberley D. Inman M.A. Singer T.A. Richmond K.J. Munn A. Rada-Iglesias O. Wallerman J. Komorowski J.C. Fowler P. Couttet A.W. Bruce O.M. Dovey P.D. Ellis C.F. Langford D.A. Nix G. Euskirchen S. Hartman A.E. Urban P. Kraus S. Van Calcar N. Heintzman T.H. Kim K. Wang C. Qu G. Hon R. Luna C.K. Glass M.G. Rosenfeld S.F. Aldred S.J. Cooper A. Halees J.M. Lin H.P. Shulha X. Zhang M. Xu J.N. Haidar Y. Yu Y. Ruan V.R. Iyer R.D. Green C. Wadelius P.J. Farnham B. Ren R.A. Harte A.S. Hinrichs H. Trumbower H. Clawson J. Hillman-Jackson A.S. Zweig K. Smith A. Thakkapallayil G. Barber R.M. Kuhn D. Karolchik L. Armengol C.P. Bird P.I. de Bakker A.D. Kern N. Lopez-Bigas J.D. Martin B.E. Stranger A. Woodroffe E. Davydov A. Dimas E. Eyras I.B. Hallgrimsdottir J. Huppert M.C. Zody G.R. Abecasis X. Estivill G.G. Bouffard X. Guan N.F. Hansen J.R. Idol V.V. Maduro B. Maskeri J.C. McDowell M. Park P.J. Thomas A.C. Young R.W. Blakesley D.M. Muzny E. Sodergren D.A. Wheeler K.C. Worley H. Jiang G.M. Weinstock R.A. Gibbs T. Graves R. Fulton E.R. Mardis R.K. Wilson M. Clamp J. Cuff S. Gnerre D.B. Jaffe J.L. Chang K. Lindblad-Toh E.S. Lander M. Koriabine M. Nefedov K. Osoegawa Y. Yoshinaga B. Zhu and P.J. de Jong. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447:** 799-816.

Bjorntorp, P., T. Schersten, and S.E. Fagerberg. 1967. Respiration and phosphorylation of mitochondria isolated from the skeletal muscle of diabetic and normal subjects. *Diabetologia* **3:** 346-352.

Bluher, M., M.D. Michael, O.D. Peroni, K. Ueki, N. Carter, B.B. Kahn, and C.R. Kahn. 2002. Adipose tissue selective insulin receptor knockout protects against obesity and obesity-related glucose intolerance. *Dev Cell* **3:** 25-38.

Bodmer, W. and C. Bonilla. 2008. Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet* **40:** 695-701.

Bogacka, I., B. Ukropcova, M. McNeil, J.M. Gimble, and S.R. Smith. 2005. Structural and functional consequences of mitochondrial biogenesis in human adipocytes in vitro. *J Clin Endocrinol Metab* **90:** 6650-6656.

Borecki, I.B. and M.A. Province. 2008. Linkage and association: basic concepts. *Adv Genet* **60:** 51-74.

Bottini, N., L. Musumeci, A. Alonso, S. Rahmouni, K. Nika, M. Rostamkhani, J. MacMurray, G.F. Meloni, P. Lucarelli, M. Pellecchia, G.S. Eisenbarth, D. Comings, and T. Mustelin. 2004. A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. *Nat Genet* **36:** 337-338.

Bouatia-Naji, N., G. Rocheleau, L. Van Lommel, K. Lemaire, F. Schuit, C. Cavalcanti-Proenca, M. Marchand, A.L. Hartikainen, U. Sovio, F. De Graeve, J. Rung, M. Vaxillaire, J. Tichet, M. Marre, B. Balkau, J. Weill, P. Elliott, M.R. Jarvelin, D. Meyre, C. Polychronakos, C. Dina, R. Sladek, and P. Froguel. 2008. A polymorphism within the G6PC2 gene is associated with fasting plasma glucose levels. *Science* **320:** 1085-1088.

Brachmann, S.M., K. Ueki, J.A. Engelman, R.C. Kahn, and L.C. Cantley. 2005. Phosphoinositide 3-kinase catalytic subunit deletion and regulatory subunit deletion have opposite effects on insulin sensitivity in mice. *Mol Cell Biol* **25:** 1596-1607.

Bradburn, M.J., J.J. Deeks, and D.G. Altman. 1999. Metan - an alternative meta-analysis command. *Stata Technical Bulletin Reprints* **8:** 86-100.

Briaud, I., L.M. Dickson, M.K. Lingohr, J.F. McCuaig, J.C. Lawrence, and C.J. Rhodes. 2005. Insulin receptor substrate-2 proteasomal degradation mediated by a mammalian target of rapamycin (mTOR)-induced negative feedback down-regulates protein kinase B-mediated signaling pathway in beta-cells. *J Biol Chem* **280:** 2282-2293.

Brickwood, S., D.T. Bonthron, L.I. Al-Gazali, K. Piper, T. Hearn, D.I. Wilson, and N.A. Hanley. 2003. Wolcott-Rallison syndrome: pathogenic insights into neonatal diabetes from new mutation and expression studies of EIF2AK3. *J Med Genet* **40:** 685-689.

Brown, A.E., S.J. Yeaman, and M. Walker. 2007. Targeted suppression of calpain-10 expression impairs insulin-stimulated glucose uptake in cultured primary human skeletal muscle cells. *Mol Genet Metab* **91:** 318-324.

Bruning, J.C., M.D. Michael, J.N. Winnay, T. Hayashi, D. Horsch, D. Accili, L.J. Goodyear, and C.R. Kahn. 1998. A muscle-specific insulin receptor knockout exhibits features of the metabolic syndrome of NIDDM without altering glucose tolerance. *Mol Cell* **2:** 559-569.

Bruss, M.D., E.B. Arias, G.E. Lienhard, and G.D. Cartee. 2005. Increased phosphorylation of Akt substrate of 160 kDa (AS160) in rat skeletal muscle in response to insulin or contractile activity. *Diabetes* **54:** 41-50.

Butler, A.E., J. Janson, S. Bonner-Weir, R. Ritzel, R.A. Rizza, and P.C. Butler. 2003. Beta-cell deficit and increased beta-cell apoptosis in humans with type 2 diabetes. *Diabetes* **52:** 102-110.

Calera, M.R., C. Martinez, H. Liu, A.K. Jack, M.J. Birnbaum, and P.F. Pilch. 1998. Insulin increases the association of Akt-2 with Glut4-containing vesicles. *J Biol Chem* **273:** 7201-7204.

Cann, H.M., C. de Toma, L. Cazes, M.F. Legrand, V. Morel, L. Piouffre, J. Bodmer, W.F. Bodmer, B. Bonne-Tamir, A. Cambon-Thomsen, Z. Chen, J. Chu, C. Carcassi, L. Contu, R. Du, L. Excoffier, G.B. Ferrara, J.S. Friedlaender, H. Groot, D. Gurwitz, T. Jenkins, R.J. Herrera, X. Huang, J. Kidd, K.K. Kidd, A. Langaney, A.A. Lin, S.Q. Mehdi, P. Parham, A. Piazza, M.P. Pistillo, Y. Qian, Q. Shu, J. Xu, S. Zhu, J.L. Weber, H.T. Greely, M.W. Feldman, G. Thomas, J. Dausset, and L.L. Cavalli-Sforza. 2002. A human genome diversity cell line panel. *Science* **296:** 261-262.

Cano, A., C. Rouzier, S. Monnot, B. Chabrol, J. Conrath, P. Lecomte, B. Delobel, P. Boileau, R. Valero, V. Procaccio, V. Paquis-Flucklinger, and B. Vialettes. 2007. Identification of novel mutations in WFS1 and genotype-phenotype correlation in Wolfram syndrome. *Am J Med Genet A* **143A:** 1605-1612.

Cao, H., L. Alston, J. Ruschman, and R.A. Hegele. 2008. Heterozygous CAV1 frameshift mutations (MIM 601047) in patients with atypical partial lipodystrophy and hypertriglyceridemia. *Lipids Health Dis* **7:** 3.

Cao, H. and R.A. Hegele. 2000. Nuclear lamin A/C R482Q mutation in canadian kindreds with Dunnigan-type familial partial lipodystrophy. *Hum Mol Genet* **9:** 109-112.

Cao, H. and R.A. Hegele. 2002. Identification of single-nucleotide polymorphisms in the human LPIN1 gene. *J Hum Genet* **47:** 370-372.

Cardon, L.R. and J.I. Bell. 2001. Association study designs for complex diseases. *Nat Rev Genet* **2:** 91-99.

Caux, F., E. Dubosclard, O. Lascols, B. Buendia, O. Chazouilleres, A. Cohen, J.C. Courvalin, L. Laroche, J. Capeau, C. Vigouroux, and S. Christin-Maitre. 2003. A new clinical condition linked to a novel mutation in lamins A and C with generalized lipoatrophy, insulin-resistant diabetes, disseminated leukomelanodermic papules, liver steatosis, and cardiomyopathy. *J Clin Endocrinol Metab* **88:** 1006-1013.

Chandak, G.R., C.S. Janipalli, S. Bhaskar, S.R. Kulkarni, P. Mohankrishna, A.T. Hattersley, T.M. Frayling, and C.S. Yajnik. 2007. Common variants in the TCF7L2 gene are strongly associated with type 2 diabetes mellitus in the Indian population. *Diabetologia* **50:** 63-67.

Chaturvedi, N., P.M. McKeigue, and M.G. Marmot. 1993. Resting and ambulatory blood pressure differences in Afro-Caribbeans and Europeans. *Hypertension* **22:** 90-96.

Chen, H., O. Charlat, L.A. Tartaglia, E.A. Woolf, X. Weng, S.J. Ellis, N.D. Lakey, J. Culpepper, K.J. Moore, R.E. Breitbart, G.M. Duyk, R.I. Tepper, and J.P. Morgenstern. 1996. Evidence that the diabetes gene encodes the leptin receptor: identification of a mutation in the leptin receptor gene in db/db mice. *Cell* **84:** 491-495.

Chen, L., L. Lee, B.A. Kudlow, H.G. Dos Santos, O. Sletvold, Y. Shafeghati, E.G. Botha, A. Garg, N.B. Hanson, G.M. Martin, I.S. Mian, B.K. Kennedy, and J. Oshima. 2003. LMNA mutations in atypical Werner's syndrome. *Lancet* **362:** 440-445.

Chen, W.M., M.R. Erdos, A.U. Jackson, R. Saxena, S. Sanna, K.D. Silver, N.J. Timpson, T. Hansen, M. Orru, M. Grazia Piras, L.L. Bonnycastle, C.J. Willer, V. Lyssenko, H. Shen, J. Kuusisto, S. Ebrahim, N. Sestu, W.L. Duren, M.C. Spada, H.M. Stringham, L.J. Scott, N. Olla, A.J. Swift, S. Najjar, B.D. Mitchell, D.A. Lawlor, G.D. Smith, Y. Ben-Shlomo, G. Andersen, K. Borch-Johnsen, T. Jorgensen, J. Saramies, T.T. Valle, T.A. Buchanan, A.R. Shuldiner, E. Lakatta, R.N. Bergman, M. Uda, J. Tuomilehto, O. Pedersen, A. Cao, L. Groop, K.L. Mohlke, M. Laakso,

D. Schlessinger, F.S. Collins, D. Altshuler, G.R. Abecasis, M. Boehnke, A. Scuteri, and R.M. Watanabe. 2008. Variations in the G6PC2/ABCB11 genomic region are associated with fasting glucose levels. *J Clin Invest* **118:** 2620-2628.

Chen, Y., R. Kittles, J. Zhou, G. Chen, A. Adeyemo, R.K. Panguluri, W. Chen, A. Amoah, V. Opoku, J. Acheampong, K. Agyenim-Boateng, B.A. Eghan, Jr., A. Nyantaki, J. Oli, G. Okafor, E. Ofoegbu, B. Osotimehin, F. Abbiyesuku, T. Johnson, O. Fasanmade, T. Rufus, P. Furbert-Harris, H.I. Daniel, K.A. Berg, F.S. Collins, G.M. Dunston, and C.N. Rotimi. 2005. Calpain-10 gene polymorphisms and type 2 diabetes in West Africans: the Africa America Diabetes Mellitus (AADM) Study. *Ann Epidemiol* **15:** 153-159.

Cho, H., J. Mu, J.K. Kim, J.L. Thorvaldsen, Q. Chu, E.B. Crenshaw, 3rd, K.H. Kaestner, M.S. Bartolomei, G.I. Shulman, and M.J. Birnbaum. 2001a. Insulin resistance and a diabetes mellitus-like syndrome in mice lacking the protein kinase Akt2 (PKB beta). *Science* **292:** 1728-1731.

Cho, H., J.L. Thorvaldsen, Q. Chu, F. Feng, and M.J. Birnbaum. 2001b. Akt1/PKBalpha is required for normal growth but dispensable for maintenance of glucose homeostasis in mice. *J Biol Chem* **276:** 38349-38352.

Ciafaloni, E., E. Ricci, S. Shanske, C.T. Moraes, G. Silvestri, M. Hirano, S. Simonetti, C. Angelini, M.A. Donati, C. Garcia, and et al. 1992. MELAS: clinical features, biochemistry, and molecular genetics. *Ann Neurol* **31:** 391-398.

Cohen, B., D. Novick, and M. Rubinstein. 1996. Modulation of insulin activities by leptin. *Science* **274:** 1185-1188.

Cohen, J.C., R.S. Kiss, A. Pertsemlidis, Y.L. Marcel, R. McPherson, and H.H. Hobbs. 2004. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* **305:** 869-872.

Cohen, J.C., A. Pertsemlidis, S. Fahmi, S. Esmail, G.L. Vega, S.M. Grundy, and H.H. Hobbs. 2006. Multiple rare variants in NPC1L1 associated with reduced sterol absorption and plasma low-density lipoprotein levels. *Proc Natl Acad Sci U S A* **103:** 1810-1815.

Collin, G.B., J.D. Marshall, A. Ikeda, W.V. So, I. Russell-Eggitt, P. Maffei, S. Beck, C.F. Boerkoel, N. Sicolo, M. Martin, P.M. Nishina, and J.K. Naggert. 2002. Mutations in ALMS1 cause obesity, type 2 diabetes and neurosensory degeneration in Alstrom syndrome. *Nat Genet* **31:** 74-78.

Colombo, C., J.J. Cutson, T. Yamauchi, C. Vinson, T. Kadowaki, O. Gavrilova, and M.L. Reitman. 2002. Transplantation of adipose tissue lacking leptin is unable to reverse the metabolic abnormalities associated with lipoatrophy. *Diabetes* **51:** 2727-2733.

Colosimo, A., V. Guida, L. Rigoli, C. Di Bella, A. De Luca, S. Briuglia, L. Stuppia, D.C. Salpietro, and B. Dallapiccola. 2003. Molecular detection of novel WFS1 mutations in patients with Wolfram syndrome by a DHPLC-based assay. *Hum Mutat* **21:** 622-629.

Condon, J., J.E. Shaw, M. Luciano, K.O. Kyvik, N.G. Martin, and D.L. Duffy. 2008. A study of diabetes mellitus within a large sample of Australian twins. *Twin Res Hum Genet* **11:** 28-40.

Croce, M.A., J.C. Eagon, L.L. LaRiviere, K.M. Korenblat, S. Klein, and B.N. Finck. 2007. Hepatic lipin 1beta expression is diminished in insulin-resistant obese subjects and is reactivated by marked weight loss. *Diabetes* **56:** 2395-2399.

Cross, D.A., D.R. Alessi, P. Cohen, M. Andjelkovich, and B.A. Hemmings. 1995. Inhibition of glycogen synthase kinase-3 by insulin mediated by protein kinase B. *Nature* **378:** 785-789.

Cryns, K., T.A. Sivakumaran, J.M. Van den Ouweland, R.J. Pennings, C.W. Cremers, K. Flothmann, T.L. Young, R.J. Smith, M.M. Lesperance, and G. Van Camp. 2003.

Mutational spectrum of the WFS1 gene in Wolfram syndrome, nonsyndromic hearing impairment, diabetes mellitus, and psychiatric disease. *Hum Mutat* **22:** 275-287.

Cushman, S.W. and L.J. Wardzala. 1980. Potential mechanism of insulin action on glucose transport in the isolated rat adipose cell. Apparent translocation of intracellular transport systems to the plasma membrane. *J Biol Chem* **255:** 4758-4762.

Daly, M.J., J.D. Rioux, S.F. Schaffner, T.J. Hudson, and E.S. Lander. 2001. High-resolution haplotype structure in the human genome. *Nat Genet* **29:** 229-232.

Dausset, J., H. Cann, D. Cohen, M. Lathrop, J.M. Lalouel, and R. White. 1990. Centre d'etude du polymorphisme humain (CEPH): collaborative genetic mapping of the human genome. *Genomics* **6:** 575-577.

Day, N., S. Oakes, R. Luben, K.T. Khaw, S. Bingham, A. Welch, and N. Wareham. 1999. EPIC-Norfolk: study design and characteristics of the cohort. European Prospective Investigation of Cancer. *Br J Cancer* **80 Suppl 1:** 95-103.

de Bakker, P.I., R. Yelensky, I. Pe'er, S.B. Gabriel, M.J. Daly, and D. Altshuler. 2005. Efficiency and power in genetic association studies. *Nat Genet* **37:** 1217-1223.

del Bosque-Plata, L., C.A. Aguilar-Salinas, M.T. Tusie-Luna, S. Ramirez-Jimenez, M. Rodriguez-Torres, M. Auron-Gomez, E. Ramirez, M.L. Velasco-Perez, A. Ramirez-Silva, F. Gomez-Perez, C.L. Hanis, T. Tsuchiya, I. Yoshiuchi, N.J. Cox, and G.I. Bell. 2004. Association of the calpain-10 gene with type 2 diabetes mellitus in a Mexican population. *Mol Genet Metab* **81:** 122-126.

Del Guerra, S., R. Lupi, L. Marselli, M. Masini, M. Bugliani, S. Sbrana, S. Torri, M. Pollera, U. Boggi, F. Mosca, S. Del Prato, and P. Marchetti. 2005. Functional and molecular defects of pancreatic islets in human type 2 diabetes. *Diabetes* **54:** 727-735.

Delepine, M., M. Nicolino, T. Barrett, M. Golamaully, G.M. Lathrop, and C. Julier. 2000. EIF2AK3, encoding translation initiation factor 2-alpha kinase 3, is mutated in patients with Wolcott-Rallison syndrome. *Nat Genet* **25:** 406-409.

Di Paolo, S., A. Teutonico, D. Leogrande, C. Capobianco, and P.F. Schena. 2006. Chronic inhibition of mammalian target of rapamycin signaling downregulates insulin receptor substrates 1 and 2 and AKT activation: A crossroad between cancer and diabetes? *J Am Soc Nephrol* **17:** 2236-2244.

Doney, A., B. Fischer, D. Frew, A. Cumming, D.M. Flavell, M. World, H.E. Montgomery, D. Boyle, A. Morris, and C.N. Palmer. 2002. Haplotype analysis of the PPARgamma Pro12Ala and C1431T variants reveals opposing associations with body weight. *BMC Genet* **3:** 21.

Doney, A.S., B. Fischer, J.E. Cecil, K. Boylan, F.E. McGuigan, S.H. Ralston, A.D. Morris, and C.N. Palmer. 2004a. Association of the Pro12Ala and C1431T variants of PPARG and their haplotypes with susceptibility to Type 2 diabetes. *Diabetologia* **47:** 555-558.

Doney, A.S., B. Fischer, J.E. Cecil, P.T. Cohen, D.I. Boyle, G. Leese, A.D. Morris, and C.N. Palmer. 2003. Male preponderance in early diagnosed type 2 diabetes is associated with the ARE insertion/deletion polymorphism in the PPP1R3A locus. *BMC Genet* **4:** 11.

Doney, A.S., B. Fischer, S.P. Lee, A.D. Morris, G. Leese, and C.N. Palmer. 2005a. Association of common variation in the PPARA gene with incident myocardial infarction in individuals with type 2 diabetes: A Go-DARTS study. *Nucl Recept* **3:** 4.

Doney, A.S., B. Fischer, G. Leese, A.D. Morris, and C.N. Palmer. 2004b. Cardiovascular risk in type 2 diabetes is associated with variation at the PPARG locus: a Go-DARTS study. *Arterioscler Thromb Vasc Biol* **24:** 2403-2407.

Doney, A.S., S. Lee, G.P. Leese, A.D. Morris, and C.N. Palmer. 2005b. Increased cardiovascular morbidity and mortality in type 2 diabetes is associated with the glutathione S transferase theta-null genotype: a Go-DARTS study. *Circulation* **111:** 2927-2934.

Dong, C., Z. Qian, P. Jia, Y. Wang, W. Huang, and Y. Li. 2007. Gene-centric characteristics of genome-wide association studies. *PLoS ONE* **2:** e1262.

Donkor, J., M. Sariahmetoglu, J. Dewald, D.N. Brindley, and K. Reue. 2007a. Three mammalian lipins act as phosphatidate phosphatases with distinct tissue expression patterns. *J Biol Chem* **282:** 3450-3457.

Donkor, J., L.M. Sparks, H. Xie, S.R. Smith, and K. Reue. 2007b. Adipose tissue lipin-1 expression is correlated with ppar{alpha} gene expression and insulin sensitivity in healthy young men. *J Clin Endocrinol Metab*.

Dunnigan, M.G., M.A. Cochrane, A. Kelly, and J.W. Scott. 1974. Familial lipoatrophic diabetes with dominant transmission. A new syndrome. *Q J Med* **43:** 33-48.

Dupont, W.D. and W.D. Plummer, Jr. 1990. Power and sample size calculations. A review and computer program. *Control Clin Trials* **11:** 116-128.

Durocher, F., R. Faure, Y. Labrie, L. Pelletier, I. Bouchard, and R. Laframboise. 2006. A novel mutation in the EIF2AK3 gene with variable expressivity in two patients with Wolcott-Rallison syndrome. *Clin Genet* **70:** 34-38.

Easton, R.M., H. Cho, K. Roovers, D.W. Shineman, M. Mizrahi, M.S. Forman, V.M. Lee, M. Szabolcs, R. de Jong, T. Oltersdorf, T. Ludwig, A. Efstratiadis, and M.J. Birnbaum. 2005. Role for Akt3/protein kinase Bgamma in attainment of normal brain size. *Mol Cell Biol* **25:** 1869-1878.

Edwards, A.O., R. Ritter, 3rd, K.J. Abel, A. Manning, C. Panhuysen, and L.A. Farrer. 2005. Complement factor H polymorphism and age-related macular degeneration. *Science* **308:** 421-424.

Eguez, L., A. Lee, J.A. Chavez, C.P. Miinea, S. Kane, G.E. Lienhard, and T.E. McGraw. 2005. Full intracellular retention of GLUT4 requires AS160 Rab GTPase activating protein. *Cell Metab* **2:** 263-272.

Eizirik, D.L., A.K. Cardozo, and M. Cnop. 2008. The role for endoplasmic reticulum stress in diabetes mellitus. *Endocr Rev* **29:** 42-61.

Ekelund, U., P.W. Franks, S. Sharp, S. Brage, and N.J. Wareham. 2007. Increase in physical activity energy expenditure is associated with reduced metabolic risk independent of change in fatness and fitness. *Diabetes Care* **30:** 2101-2106.

El-Assaad, W., J. Buteau, M.L. Peyot, C. Nolan, R. Roduit, S. Hardy, E. Joly, G. Dbaibo, L. Rosenberg, and M. Prentki. 2003. Saturated fatty acids synergize with elevated glucose to cause pancreatic beta-cell death. *Endocrinology* **144:** 4154-4163.

Elbein, S.C., W. Chu, Q. Ren, C. Hemphill, J. Schay, N.J. Cox, C.L. Hanis, and S.J. Hasstedt. 2002. Role of calpain-10 gene variants in familial type 2 diabetes in Caucasians. *J Clin Endocrinol Metab* **87:** 650-654.

Elchebly, M., P. Payette, E. Michaliszyn, W. Cromlish, S. Collins, A.L. Loy, D. Normandin, A. Cheng, J. Himms-Hagen, C.C. Chan, C. Ramachandran, M.J. Gresser, M.L. Tremblay, and B.P. Kennedy. 1999. Increased insulin sensitivity and obesity resistance in mice lacking the protein tyrosine phosphatase-1B gene. *Science* **283:** 1544-1548.

Fan, J.B., A. Oliphant, R. Shen, B.G. Kermani, F. Garcia, K.L. Gunderson, M. Hansen, F. Steemers, S.L. Butler, P. Deloukas, L. Galver, S. Hunt, C. McBride, M. Bibikova, T. Rubano, J. Chen, E. Wickham, D. Doucet, W. Chang, D. Campbell, B. Zhang, S. Kruglyak, D. Bentley, J. Haas, P. Rigault, L. Zhou, J. Stuelpnagel, and M.S. Chee. 2003. Highly parallel SNP genotyping. *Cold Spring Harb Symp Quant Biol* **68:** 69-78.

Fanciulli, M., P.J. Norsworthy, E. Petretto, R. Dong, L. Harper, L. Kamesh, J.M. Heward, S.C. Gough, A. de Smith, A.I. Blakemore, P. Froguel, C.J. Owen, S.H. Pearce, L. Teixeira, L. Guillevin, D.S. Graham, C.D. Pusey, H.T. Cook, T.J. Vyse, and T.J. Aitman. 2007. FCGR3B copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. *Nat Genet* **39:** 721-723.

Fawcett, K.A., N. Grimsey, R.J. Loos, E. Wheeler, A. Daly, M. Soos, R. Semple, H. Syddall, C. Cooper, S. Siniossoglou, S. O'Rahilly, N.J. Wareham, and I. Barroso. 2008. Evaluating the role of LPIN1 variation in insulin resistance, body weight, and human lipodystrophy in U.K. Populations. *Diabetes* **57:** 2527-2533.

Fawcett, K.A., N.J. Wareham, J. Luan, H. Syddall, C. Cooper, S. O'Rahilly, I.N. Day, M.S. Sandhu, and I. Barroso. 2006. PARL Leu262Val is not associated with fasting insulin levels in UK populations. *Diabetologia* **49:** 2649-2652.

Ferguson, P.J., X. Bing, M.A. Vasef, L.A. Ochoa, A. Mahgoub, T.J. Waldschmidt, L.T. Tygrett, A.J. Schlueter, and H. El-Shanti. 2006. A missense mutation in pstpip2 is associated with the murine autoinflammatory disorder chronic multifocal osteomyelitis. *Bone* **38:** 41-47.

Ferguson, P.J., S. Chen, M.K. Tayeh, L. Ochoa, S.M. Leal, A. Pelet, A. Munnich, S. Lyonnet, H.A. Majeed, and H. El-Shanti. 2005. Homozygous mutations in LPIN2 are responsible for the syndrome of chronic recurrent multifocal osteomyelitis and congenital dyserythropoietic anaemia (Majeed syndrome). *J Med Genet* **42:** 551-557.

Finck, B.N., M.C. Gropler, Z. Chen, T.C. Leone, M.A. Croce, T.E. Harris, J.C. Lawrence, Jr., and D.P. Kelly. 2006. Lipin 1 is an inducible amplifier of the hepatic PGC-1alpha/PPARalpha regulatory pathway. *Cell Metab* **4:** 199-210.

Flier, J.S., D.E. Moller, A.C. Moses, S. O'Rahilly, R.L. Chaiken, F. Grigorescu, D. Elahi, B.B. Kahn, J.E. Weinreb, and R. Eastman. 1993. Insulin-mediated pseudoacromegaly: clinical and biochemical characterization of a syndrome of selective insulin resistance. *J Clin Endocrinol Metab* **76:** 1533-1541.

Fonseca, S.G., M. Fukuma, K.L. Lipson, L.X. Nguyen, J.R. Allen, Y. Oka, and F. Urano. 2005. WFS1 is a novel component of the unfolded protein response and maintains homeostasis of the endoplasmic reticulum in pancreatic beta-cells. *J Biol Chem* **280:** 39609-39615.

Foti, D., E. Chiefari, M. Fedele, R. Iuliano, L. Brunetti, F. Paonessa, G. Manfioletti, F. Barbetti, A. Brunetti, C.M. Croce, A. Fusco, and A. Brunetti. 2005. Lack of the architectural factor HMGA1 causes insulin resistance and diabetes in humans and mice. *Nat Med* **11:** 765-773.

Fraenkel, M., M. Ketzinel-Gilad, Y. Ariav, O. Pappo, M. Karaca, J. Castel, M.F. Berthault, C. Magnan, E. Cerasi, N. Kaiser, and G. Leibowitz. 2008. mTOR inhibition by rapamycin prevents beta-cell adaptation to hyperglycemia and exacerbates the metabolic state in type 2 diabetes. *Diabetes* **57:** 945-957.

Frank, B., J.L. Bermejo, K. Hemminki, C. Sutter, B. Wappenschmidt, A. Meindl, M. Kiechle-Bahat, P. Bugert, R.K. Schmutzler, C.R. Bartram, and B. Burwinkel. 2007. Copy number variant in the candidate tumor suppressor gene MTUS1 and familial breast cancer risk. *Carcinogenesis* **28:** 1442-1445.

Franks, P.W., O. Rolandsson, S.L. Debenham, K.A. Fawcett, F. Payne, C. Dina, P. Froguel, K.L. Mohlke, C. Willer, T. Olsson, N.J. Wareham, G. Hallmans, I. Barroso, and M.S. Sandhu. 2008. Replication of the association between variants in WFS1 and risk of type 2 diabetes in European populations. *Diabetologia* **51:** 458-463.

Fraser, F.C. and T. Gunn. 1977. Diabetes mellitus, diabetes insipidus, and optic atrophy. An autosomal recessive syndrome? *J Med Genet* **14:** 190-193.

Frayling, T., M. Walker, M. McCarthy, J. Evans, L. Allen, S. Lynn, S. Ayres, B. Millauer, C. Turner, R. Turner, M. Sampson, G. Hitman, S. Ellard, and A. Hattersley. 1999. Parent-offspring Trios: a resource to facilitate the identification of Type 2 diabetes genes. *Diabetes* **48:** 2475-2479.

Frayling, T.M., N.J. Timpson, M.N. Weedon, E. Zeggini, R.M. Freathy, C.M. Lindgren, J.R. Perry, K.S. Elliott, H. Lango, N.W. Rayner, B. Shields, L.W. Harries, J.C. Barrett, S. Ellard, C.J. Groves, B. Knight, A.M. Patch, A.R. Ness, S. Ebrahim, D.A. Lawlor, S.M. Ring, Y. Ben-Shlomo, M.R. Jarvelin, U. Sovio, A.J. Bennett, D. Melzer, L. Ferrucci, R.J. Loos, I. Barroso, N.J. Wareham, F. Karpe, K.R. Owen, L.R. Cardon, M. Walker, G.A. Hitman, C.N. Palmer, A.S. Doney, A.D. Morris, G.D. Smith, A.T. Hattersley, and M.I. McCarthy. 2007. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* **316:** 889-894.

Frazer, K.A. D.G. Ballinger D.R. Cox D.A. Hinds L.L. Stuve R.A. Gibbs J.W. Belmont A. Boudreau P. Hardenbol S.M. Leal S. Pasternak D.A. Wheeler T.D. Willis F. Yu H. Yang C. Zeng Y. Gao H. Hu W. Hu C. Li W. Lin S. Liu H. Pan X. Tang J. Wang W. Wang J. Yu B. Zhang Q. Zhang H. Zhao H. Zhao J. Zhou S.B. Gabriel R. Barry B. Blumenstiel A. Camargo M. Defelice M. Faggart M. Goyette S. Gupta J. Moore H. Nguyen R.C. Onofrio M. Parkin J. Roy E. Stahl E. Winchester L. Ziaugra D. Altshuler Y. Shen Z. Yao W. Huang X. Chu Y. He L. Jin Y. Liu Y. Shen W. Sun H. Wang Y. Wang Y. Wang X. Xiong L. Xu M.M. Waye S.K. Tsui H. Xue J.T. Wong L.M. Galver J.B. Fan K. Gunderson S.S. Murray A.R. Oliphant M.S. Chee A. Montpetit F. Chagnon V. Ferretti M. Leboeuf J.F. Olivier M.S. Phillips S. Roumy C. Sallee A. Verner T.J. Hudson P.Y. Kwok D. Cai D.C. Koboldt R.D. Miller L. Pawlikowska P. Taillon-Miller M. Xiao L.C. Tsui W. Mak Y.Q. Song P.K. Tam Y. Nakamura T. Kawaguchi T. Kitamoto T. Morizono A. Nagashima Y. Ohnishi A. Sekine T. Tanaka T. Tsunoda P. Deloukas C.P. Bird M. Delgado E.T. Dermitzakis R. Gwilliam S. Hunt J. Morrison D. Powell B.E. Stranger P. Whittaker D.R. Bentley M.J. Daly P.I. de Bakker J. Barrett Y.R. Chretien J. Maller S. McCarroll N. Patterson I. Pe'er A. Price S. Purcell D.J. Richter P. Sabeti R. Saxena S.F. Schaffner P.C. Sham P. Varilly D. Altshuler L.D. Stein L. Krishnan A.V. Smith M.K. Tello-Ruiz G.A. Thorisson A. Chakravarti P.E. Chen D.J. Cutler C.S. Kashuk S. Lin G.R. Abecasis W. Guan Y. Li H.M. Munro Z.S. Qin D.J. Thomas G. McVean A. Auton L. Bottolo N. Cardin S. Eyheramendy C. Freeman J. Marchini S. Myers C. Spencer M. Stephens P. Donnelly L.R. Cardon G. Clarke D.M. Evans A.P. Morris B.S. Weir T. Tsunoda J.C. Mullikin S.T. Sherry M. Feolo A. Skol H. Zhang C. Zeng H. Zhao I. Matsuda Y. Fukushima D.R. Macer E. Suda C.N. Rotimi C.A. Adebamowo I. Ajayi T. Aniagwu P.A. Marshall C. Nkwodimmah C.D. Royal M.F. Leppert M. Dixon A. Peiffer R. Qiu A. Kent K. Kato N. Niikawa I.F. Adewole B.M. Knoppers M.W. Foster E.W. Clayton J. Watkin R.A. Gibbs J.W. Belmont D. Muzny L. Nazareth E. Sodergren G.M. Weinstock D.A. Wheeler I. Yakub S.B. Gabriel R.C. Onofrio D.J. Richter L. Ziaugra B.W. Birren M.J. Daly D. Altshuler R.K. Wilson L.L. Fulton J. Rogers J. Burton N.P. Carter C.M. Clee M. Griffiths M.C. Jones K. McLay R.W. Plumb M.T. Ross S.K. Sims D.L. Willey Z. Chen H. Han L. Kang M. Godbout J.C. Wallenburg P. L'Archeveque G. Bellemare K. Saeki H. Wang D. An H. Fu Q. Li Z. Wang R. Wang A.L. Holden L.D. Brooks J.E. McEwen M.S. Guyer V.O. Wang J.L. Peterson M. Shi J. Spiegel L.M. Sung L.F. Zacharia F.S. Collins K. Kennedy R. Jamieson and J. Stewart. 2007. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449:** 851-861.

Fredriksson, R., M. Hagglund, P.K. Olszewski, O. Stephansson, J.A. Jacobsson, A.M. Olszewska, A.S. Levine, J. Lindblom, and H.B. Schioth. 2008. The obesity gene,

FTO, is of ancient origin, up-regulated during food deprivation and expressed in neurons of feeding-related nuclei of the brain. *Endocrinology* **149:** 2062-2071.

Frias, M.A., C.C. Thoreen, J.D. Jaffe, W. Schroder, T. Sculley, S.A. Carr, and D.M. Sabatini. 2006. mSin1 is necessary for Akt/PKB phosphorylation, and its isoforms define three distinct mTORC2s. *Curr Biol* **16:** 1865-1870.

Froguel, P., M. Vaxillaire, F. Sun, G. Velho, H. Zouali, M.O. Butel, S. Lesage, N. Vionnet, K. Clement, F. Fougerousse, and et al. 1992. Close linkage of glucokinase locus on chromosome 7p to early-onset non-insulin-dependent diabetes mellitus. *Nature* **356:** 162-164.

Fu, Y.H., A. Pizzuti, R.G. Fenwick, Jr., J. King, S. Rajnarayan, P.W. Dunne, J. Dubel, G.A. Nasser, T. Ashizawa, P. de Jong, and et al. 1992. An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science* **255:** 1256-1258.

Gao, Z., X. Zhang, A. Zuberi, D. Hwang, M.J. Quon, M. Lefevre, and J. Ye. 2004. Inhibition of insulin sensitivity by free fatty acids requires activation of multiple serine kinases in 3T3-L1 adipocytes. *Mol Endocrinol* **18:** 2024-2034.

Garant, M.J., W.H. Kao, F. Brancati, J. Coresh, T.M. Rami, C.L. Hanis, E. Boerwinkle, and A.R. Shuldiner. 2002. SNP43 of CAPN10 and the risk of type 2 Diabetes in African-Americans: the Atherosclerosis Risk in Communities Study. *Diabetes* **51:** 231-237.

Garg, A., R. Wilson, R. Barnes, E. Arioglu, Z. Zaidi, F. Gurakan, N. Kocak, S. O'Rahilly, S.I. Taylor, S.B. Patel, and A.M. Bowcock. 1999. A gene for congenital generalized lipodystrophy maps to human chromosome 9q34. *J Clin Endocrinol Metab* **84:** 3390-3394.

Garofalo, R.S., S.J. Orena, K. Rafidi, A.J. Torchia, J.L. Stock, A.L. Hildebrandt, T. Coskran, S.C. Black, D.J. Brees, J.R. Wicks, J.D. McNeish, and K.G. Coleman. 2003. Severe diabetes, age-dependent loss of adipose tissue, and mild growth deficiency in mice lacking Akt2/PKB beta. *J Clin Invest* **112:** 197-208.

Gavrilova, O., B. Marcus-Samuels, D. Graham, J.K. Kim, G.I. Shulman, A.L. Castle, C. Vinson, M. Eckhaus, and M.L. Reitman. 2000. Surgical implantation of adipose tissue reverses diabetes in lipoatrophic mice. *J Clin Invest* **105:** 271-278.

George, S., J.J. Rochford, C. Wolfrum, S.L. Gray, S. Schinner, J.C. Wilson, M.A. Soos, P.R. Murgatroyd, R.M. Williams, C.L. Acerini, D.B. Dunger, D. Barford, A.M. Umpleby, N.J. Wareham, H.A. Davies, A.J. Schafer, M. Stoffel, S. O'Rahilly, and I. Barroso. 2004. A family with severe insulin resistance and diabetes due to a mutation in AKT2. *Science* **304:** 1325-1328.

Gerken, T., C.A. Girard, Y.C. Tung, C.J. Webby, V. Saudek, K.S. Hewitson, G.S. Yeo, M.A. McDonough, S. Cunliffe, L.A. McNeill, J. Galvanovskis, P. Rorsman, P. Robins, X. Prieur, A.P. Coll, M. Ma, Z. Jovanovic, I.S. Farooqi, B. Sedgwick, I. Barroso, T. Lindahl, C.P. Ponting, F.M. Ashcroft, S. O'Rahilly, and C.J. Schofield. 2007. The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science* **318:** 1469-1472.

Glaser, B., P. Kesavan, M. Heyman, E. Davis, A. Cuesta, A. Buchs, C.A. Stanley, P.S. Thornton, M.A. Permutt, F.M. Matschinsky, and K.C. Herold. 1998. Familial hyperinsulinism caused by an activating glucokinase mutation. *N Engl J Med* **338:** 226-230.

Gloyn, A.L., Y. Hashim, S.J. Ashcroft, R. Ashfield, S. Wiltshire, and R.C. Turner. 2001. Association studies of variants in promoter and coding regions of beta-cell ATP-sensitive K-channel genes SUR1 and Kir6.2 with Type 2 diabetes mellitus (UKPDS 53). *Diabet Med* **18:** 206-212.

Gloyn, A.L., E.R. Pearson, J.F. Antcliff, P. Proks, G.J. Bruining, A.S. Slingerland, N. Howard, S. Srinivasan, J.M. Silva, J. Molnes, E.L. Edghill, T.M. Frayling, I.K. Temple, D. Mackay, J.P. Shield, Z. Sumnik, A. van Rhijn, J.K. Wales, P. Clark, S.

Gorman, J. Aisenberg, S. Ellard, P.R. Njolstad, F.M. Ashcroft, and A.T. Hattersley. 2004. Activating mutations in the gene encoding the ATP-sensitive potassium-channel subunit Kir6.2 and permanent neonatal diabetes. *N Engl J Med* **350:** 1838-1849.

Gloyn, A.L., F. Reimann, C. Girard, E.L. Edghill, P. Proks, E.R. Pearson, I.K. Temple, D.J. Mackay, J.P. Shield, D. Freedenberg, K. Noyes, S. Ellard, F.M. Ashcroft, F.M. Gribble, and A.T. Hattersley. 2005. Relapsing diabetes can result from moderately activating mutations in KCNJ11. *Hum Mol Genet* **14:** 925-934.

Gloyn, A.L., M.N. Weedon, K.R. Owen, M.J. Turner, B.A. Knight, G. Hitman, M. Walker, J.C. Levy, M. Sampson, S. Halford, M.I. McCarthy, A.T. Hattersley, and T.M. Frayling. 2003. Large-scale association studies of variants in genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes. *Diabetes* **52:** 568-572.

Gomez-Zaera, M., T.M. Strom, B. Rodriguez, X. Estivill, T. Meitinger, and V. Nunes. 2001. Presence of a major WFS1 mutation in Spanish Wolfram syndrome pedigrees. *Mol Genet Metab* **72:** 72-81.

Grant, S.F., G. Thorleifsson, I. Reynisdottir, R. Benediktsson, A. Manolescu, J. Sainz, A. Helgason, H. Stefansson, V. Emilsson, A. Helgadottir, U. Styrkarsdottir, K.P. Magnusson, G.B. Walters, E. Palsdottir, T. Jonsdottir, T. Gudmundsdottir, A. Gylfason, J. Saemundsdottir, R.L. Wilensky, M.P. Reilly, D.J. Rader, Y. Bagger, C. Christiansen, V. Gudnason, G. Sigurdsson, U. Thorsteinsdottir, J.R. Gulcher, A. Kong, and K. Stefansson. 2006. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet* **38:** 320-323.

Grarup, N., G. Andersen, N.T. Krarup, A. Albrechtsen, O. Schmitz, T. Jorgensen, K. Borch-Johnsen, T. Hansen, and O. Pedersen. 2008. Association testing of novel type 2 diabetes risk alleles in the JAZF1, CDC123/CAMK1D, TSPAN8, THADA, ADAMTS9, and NOTCH2 loci with insulin release, insulin sensitivity, and obesity in a population-based sample of 4,516 glucose-tolerant middle-aged Danes. *Diabetes* **57:** 2534-2540.

Gudmundsson, J., P. Sulem, V. Steinthorsdottir, J.T. Bergthorsson, G. Thorleifsson, A. Manolescu, T. Rafnar, D. Gudbjartsson, B.A. Agnarsson, A. Baker, A. Sigurdsson, K.R. Benediktsdottir, M. Jakobsdottir, T. Blondal, S.N. Stacey, A. Helgason, S. Gunnarsdottir, A. Olafsdottir, K.T. Kristinsson, B. Birgisdottir, S. Ghosh, S. Thorlacius, D. Magnusdottir, G. Stefansdottir, K. Kristjansson, Y. Bagger, R.L. Wilensky, M.P. Reilly, A.D. Morris, C.H. Kimber, A. Adeyemo, Y. Chen, J. Zhou, W.Y. So, P.C. Tong, M.C. Ng, T. Hansen, G. Andersen, K. Borch-Johnsen, T. Jorgensen, A. Tres, F. Fuertes, M. Ruiz-Echarri, L. Asin, B. Saez, E. van Boven, S. Klaver, D.W. Swinkels, K.K. Aben, T. Graif, J. Cashy, B.K. Suarez, O. van Vierssen Trip, M.L. Frigge, C. Ober, M.H. Hofker, C. Wijmenga, C. Christiansen, D.J. Rader, C.N. Palmer, C. Rotimi, J.C. Chan, O. Pedersen, G. Sigurdsson, R. Benediktsson, E. Jonsson, G.V. Einarsson, J.I. Mayordomo, W.J. Catalona, L.A. Kiemeney, R.B. Barkardottir, J.R. Gulcher, U. Thorsteinsdottir, A. Kong, and K. Stefansson. 2007. Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat Genet* **39:** 977-983.

Guertin, D.A., D.M. Stevens, C.C. Thoreen, A.A. Burds, N.Y. Kalaany, J. Moffat, M. Brown, K.J. Fitzgerald, and D.M. Sabatini. 2006. Ablation in mice of the mTORC components raptor, rictor, or mLST8 reveals that mTORC2 is required for signaling to Akt-FOXO and PKCalpha, but not S6K1. *Dev Cell* **11:** 859-871.

Gunderson, K.L., F.J. Steemers, G. Lee, L.G. Mendoza, and M.S. Chee. 2005. A genome-wide scalable SNP genotyping assay using microarray technology. *Nat Genet* **37:** 549-554.

Haines, J.L., M.A. Hauser, S. Schmidt, W.K. Scott, L.M. Olson, P. Gallins, K.L. Spencer, S.Y. Kwan, M. Noureddine, J.R. Gilbert, N. Schnetz-Boutaud, A. Agarwal, E.A. Postel, and M.A. Pericak-Vance. 2005. Complement factor H variant increases the risk of age-related macular degeneration. *Science* **308:** 419-421.

Hamsten, A. and P. Eriksson. 2008. Identifying the susceptibility genes for coronary artery disease: from hyperbole through doubt to cautious optimism. *J Intern Med* **263:** 538-552.

Han, G.S., W.I. Wu, and G.M. Carman. 2006. The Saccharomyces cerevisiae Lipin homolog is a Mg2+-dependent phosphatidate phosphatase enzyme. *J Biol Chem* **281:** 9210-9218.

Hani, E.H., P. Boutin, E. Durand, H. Inoue, M.A. Permutt, G. Velho, and P. Froguel. 1998. Missense mutations in the pancreatic islet beta cell inwardly rectifying K+ channel gene (KIR6.2/BIR): a meta-analysis suggests a role in the polygenic basis of Type II diabetes mellitus in Caucasians. *Diabetologia* **41:** 1511-1515.

Harding, A.H., N.E. Day, K.T. Khaw, S. Bingham, R. Luben, A. Welsh, and N.J. Wareham. 2004. Dietary fat and the risk of clinical type 2 diabetes: the European prospective investigation of Cancer-Norfolk study. *Am J Epidemiol* **159:** 73-82.

Harding, H.P., H. Zeng, Y. Zhang, R. Jungries, P. Chung, H. Plesken, D.D. Sabatini, and D. Ron. 2001. Diabetes mellitus and exocrine pancreatic dysfunction in perk-/- mice reveals a role for translational control in secretory cell survival. *Mol Cell* **7:** 1153-1163.

Hardy, C., F. Khanim, R. Torres, M. Scott-Brown, A. Seller, J. Poulton, D. Collier, J. Kirk, M. Polymeropoulos, F. Latif, and T. Barrett. 1999. Clinical and molecular genetic analysis of 19 Wolfram syndrome kindreds demonstrating a wide spectrum of mutations in WFS1. *Am J Hum Genet* **65:** 1279-1290.

Harrington, L.S., G.M. Findlay, A. Gray, T. Tolkacheva, S. Wigfield, H. Rebholz, J. Barnett, N.R. Leslie, S. Cheng, P.R. Shepherd, I. Gout, C.P. Downes, and R.F. Lamb. 2004. The TSC1-2 tumor suppressor controls insulin-PI3K signaling via regulation of IRS proteins. *J Cell Biol* **166:** 213-223.

Harris, T.E., T.A. Huffman, A. Chi, J. Shabanowitz, D.F. Hunt, A. Kumar, and J.C. Lawrence, Jr. 2007. Insulin controls subcellular localization and multisite phosphorylation of the phosphatidic acid phosphatase, lipin 1. *J Biol Chem* **282:** 277-286.

Haruta, T., T. Uno, J. Kawahara, A. Takano, K. Egawa, P.M. Sharma, J.M. Olefsky, and M. Kobayashi. 2000. A rapamycin-sensitive pathway down-regulates insulin signaling via phosphorylation and proteasomal degradation of insulin receptor substrate-1. *Mol Endocrinol* **14:** 783-794.

Hattersley, A.T., R.C. Turner, M.A. Permutt, P. Patel, Y. Tanizawa, K.C. Chiu, S. O'Rahilly, P.J. Watkins, and J.S. Wainscoat. 1992. Linkage of type 2 diabetes to the glucokinase gene. *Lancet* **339:** 1307-1310.

Hayden, E.C. 2008. International genome project launched. *Nature* **451:** 378-379.

Hearn, T., G.L. Renforth, C. Spalluto, N.A. Hanley, K. Piper, S. Brickwood, C. White, V. Connolly, J.F. Taylor, I. Russell-Eggitt, D. Bonneau, M. Walker, and D.I. Wilson. 2002. Mutation of ALMS1, a large gene with a tandem repeat encoding 47 amino acids, causes Alstrom syndrome. *Nat Genet* **31:** 79-83.

Hegele, R.A., H. Cao, C. Frankowski, S.T. Mathews, and T. Leff. 2002. PPARG F388L, a transactivation-deficient mutant, in familial partial lipodystrophy. *Diabetes* **51:** 3586-3590.

Hegele, R.A., S.B. Harris, B. Zinman, A.J. Hanley, and H. Cao. 2001. Absence of association of type 2 diabetes with CAPN10 and PC-1 polymorphisms in Oji-Cree. *Diabetes Care* **24:** 1498-1499.

Heitman, J., N.R. Movva, and M.N. Hall. 1991. Targets for cell cycle arrest by the immunosuppressant rapamycin in yeast. *Science* **253:** 905-909.

Helgadottir, A., A. Manolescu, G. Thorleifsson, S. Gretarsdottir, H. Jonsdottir, U. Thorsteinsdottir, N.J. Samani, G. Gudmundsson, S.F. Grant, G. Thorgeirsson, S. Sveinbjornsdottir, E.M. Valdimarsson, S.E. Matthiasson, H. Johannsson, O. Gudmundsdottir, M.E. Gurney, J. Sainz, M. Thorhallsdottir, M. Andresdottir, M.L. Frigge, E.J. Topol, A. Kong, V. Gudnason, H. Hakonarson, J.R. Gulcher, and K. Stefansson. 2004. The gene encoding 5-lipoxygenase activating protein confers risk of myocardial infarction and stroke. *Nat Genet* **36:** 233-239.

Heller-Harrison, R.A., M. Morin, and M.P. Czech. 1995. Insulin regulation of membrane-associated insulin receptor substrate 1. *J Biol Chem* **270:** 24442-24450.

Hofmann, S. and M.F. Bauer. 2006. Wolfram syndrome-associated mutations lead to instability and proteasomal degradation of wolframin. *FEBS Lett* **580:** 4000-4004.

Hofmann, S., C. Philbrook, K.D. Gerbitz, and M.F. Bauer. 2003. Wolfram syndrome: structural and functional analyses of mutant and wild-type wolframin, the WFS1 gene product. *Hum Mol Genet* **12:** 2003-2012.

Horikawa, Y., N. Iwasaki, M. Hara, H. Furuta, Y. Hinokio, B.N. Cockburn, T. Lindner, K. Yamagata, M. Ogata, O. Tomonaga, H. Kuroki, T. Kasahara, Y. Iwamoto, and G.I. Bell. 1997. Mutation in hepatocyte nuclear factor-1 beta gene (TCF2) associated with MODY. *Nat Genet* **17:** 384-385.

Horikawa, Y., N. Oda, N.J. Cox, X. Li, M. Orho-Melander, M. Hara, Y. Hinokio, T.H. Lindner, H. Mashima, P.E. Schwarz, L. del Bosque-Plata, Y. Oda, I. Yoshiuchi, S. Colilla, K.S. Polonsky, S. Wei, P. Concannon, N. Iwasaki, J. Schulze, L.J. Baier, C. Bogardus, L. Groop, E. Boerwinkle, C.L. Hanis, and G.I. Bell. 2000. Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat Genet* **26:** 163-175.

Hossain, P., B. Kawar, and M. El Nahas. 2007. Obesity and diabetes in the developing world--a growing challenge. *N Engl J Med* **356:** 213-215.

Hresko, R.C. and M. Mueckler. 2005. mTOR.RICTOR is the Ser473 kinase for Akt/protein kinase B in 3T3-L1 adipocytes. *J Biol Chem* **280:** 40406-40416.

Huffman, T.A., I. Mothe-Satney, and J.C. Lawrence, Jr. 2002. Insulin-stimulated phosphorylation of lipin mediated by the mammalian target of rapamycin. *Proc Natl Acad Sci U S A* **99:** 1047-1052.

Inoue, H., Y. Tanizawa, J. Wasson, P. Behn, K. Kalidas, E. Bernal-Mizrachi, M. Mueckler, H. Marshall, H. Donis-Keller, P. Crock, D. Rogers, M. Mikuni, H. Kumashiro, K. Higashi, G. Sobue, Y. Oka, and M.A. Permutt. 1998. A gene encoding a transmembrane protein is mutated in patients with diabetes mellitus and optic atrophy (Wolfram syndrome). *Nat Genet* **20:** 143-148.

Ishihara, H., S. Takeda, A. Tamura, R. Takahashi, S. Yamaguchi, D. Takei, T. Yamada, H. Inoue, H. Soga, H. Katagiri, Y. Tanizawa, and Y. Oka. 2004. Disruption of the WFS1 gene in mice causes progressive beta-cell loss and impaired stimulus-secretion coupling in insulin secretion. *Hum Mol Genet* **13:** 1159-1170.

Itani, S.I., N.B. Ruderman, F. Schmieder, and G. Boden. 2002. Lipid-induced insulin resistance in human muscle is associated with changes in diacylglycerol, protein kinase C, and IkappaB-alpha. *Diabetes* **51:** 2005-2011.

Jacinto, E., V. Facchinetti, D. Liu, N. Soto, S. Wei, S.Y. Jung, Q. Huang, J. Qin, and B. Su. 2006. SIN1/MIP1 maintains rictor-mTOR complex integrity and regulates Akt phosphorylation and substrate specificity. *Cell* **127:** 125-137.

Jacinto, E., R. Loewith, A. Schmidt, S. Lin, M.A. Ruegg, A. Hall, and M.N. Hall. 2004. Mammalian TOR complex 2 controls the actin cytoskeleton and is rapamycin insensitive. *Nat Cell Biol* **6:** 1122-1128.

James, D.E., R. Brown, J. Navarro, and P.F. Pilch. 1988. Insulin-regulatable tissues express a unique insulin-sensitive glucose transport protein. *Nature* **333:** 183-185.

James, D.E., M. Strube, and M. Mueckler. 1989. Molecular cloning and characterization of an insulin-regulatable glucose transporter. *Nature* **338:** 83-87.

Ji, W., J.N. Foo, B.J. O'Roak, H. Zhao, M.G. Larson, D.B. Simon, C. Newton-Cheh, M.W. State, D. Levy, and R.P. Lifton. 2008. Rare independent mutations in renal salt handling genes contribute to blood pressure variation. *Nat Genet* **40:** 592-599.

Jin, T. and L. Liu. 2008. The Wnt Signaling Pathway Effector TCF7L2 and Type II Diabetes Mellitus. *Mol Endocrinol*.

Joshi, R.L., B. Lamothe, N. Cordonnier, K. Mesbah, E. Monthioux, J. Jami, and D. Bucchini. 1996. Targeted disruption of the insulin receptor gene in the mouse results in neonatal lethality. *Embo J* **15:** 1542-1547.

Kadayifci, A., Y. Kepekci, Y. Coskun, and Y. Huang. 2001. Wolfram syndrome in a family with variable expression. *Acta Medica (Hradec Kralove)* **44:** 115-118.

Kadowaki, T., C.L. Bevins, A. Cama, K. Ojamaa, B. Marcus-Samuels, H. Kadowaki, L. Beitz, C. McKeon, and S.I. Taylor. 1988. Two mutant alleles of the insulin receptor gene in a patient with extreme insulin resistance. *Science* **240:** 787-790.

Kadowaki, T., H. Kadowaki, M.M. Rechler, M. Serrano-Rios, J. Roth, P. Gorden, and S.I. Taylor. 1990. Five mutant alleles of the insulin receptor gene in patients with genetic forms of insulin resistance. *J Clin Invest* **86:** 254-264.

Kakiuchi, C., M. Ishiwata, A. Hayashi, and T. Kato. 2006. XBP1 induces WFS1 through an endoplasmic reticulum stress response element-like motif in SH-SY5Y cells. *J Neurochem* **97:** 545-555.

Kamiyama, M., M. Kobayashi, S. Araki, A. Iida, T. Tsunoda, K. Kawai, M. Imanishi, M. Nomura, T. Babazono, Y. Iwamoto, A. Kashiwagi, K. Kaku, R. Kawamori, D.P. Ng, T. Hansen, P. Gaede, O. Pedersen, Y. Nakamura, and S. Maeda. 2007. Polymorphisms in the 3' UTR in the neurocalcin delta gene affect mRNA stability, and confer susceptibility to diabetic nephropathy. *Hum Genet* **122:** 397-407.

Kane, S., H. Sano, S.C. Liu, J.M. Asara, W.S. Lane, C.C. Garner, and G.E. Lienhard. 2002. A method to identify serine kinase substrates. Akt phosphorylates a novel adipocyte protein with a Rab GTPase-activating protein (GAP) domain. *J Biol Chem* **277:** 22115-22118.

Kang, E.S., H.J. Kim, M. Nam, C.M. Nam, C.W. Ahn, B.S. Cha, and H.C. Lee. 2006. A novel 111/121 diplotype in the Calpain-10 gene is associated with type 2 diabetes. *J Hum Genet* **51:** 629-633.

Karlberg, N., H. Jalanko, J. Kallijarvi, A.E. Lehesjoki, and M. Lipsanen-Nyman. 2005. Insulin resistance syndrome in subjects with mutated RING finger protein TRIM37. *Diabetes* **54:** 3577-3581.

Ke, X., S. Hunt, W. Tapper, R. Lawrence, G. Stavrides, J. Ghori, P. Whittaker, A. Collins, A.P. Morris, D. Bentley, L.R. Cardon, and P. Deloukas. 2004. The impact of SNP density on fine-scale patterns of linkage disequilibrium. *Hum Mol Genet* **13:** 577-588.

Kelley, D.E., J. He, E.V. Menshikova, and V.B. Ritov. 2002. Dysfunction of mitochondria in human skeletal muscle in type 2 diabetes. *Diabetes* **51:** 2944-2950.

Kelpe, C.L., P.C. Moore, S.D. Parazzoli, B. Wicksteed, C.J. Rhodes, and V. Poitout. 2003. Palmitate inhibition of insulin gene expression is mediated at the transcriptional level via ceramide synthesis. *J Biol Chem* **278:** 30015-30021.

Kennedy, G.C., H. Matsuzaki, S. Dong, W.M. Liu, J. Huang, G. Liu, X. Su, M. Cao, W. Chen, J. Zhang, W. Liu, G. Yang, X. Di, T. Ryder, Z. He, U. Surti, M.S. Phillips, M.T. Boyce-Jacino, S.P. Fodor, and K.W. Jones. 2003. Large-scale genotyping of complex DNA. *Nat Biotechnol* **21:** 1233-1237.

Khamzina, L., A. Veilleux, S. Bergeron, and A. Marette. 2005. Increased activation of the mammalian target of rapamycin pathway in liver and skeletal muscle of obese rats: possible involvement in obesity-linked insulin resistance. *Endocrinology* **146:** 1473-1481.

Kifagi, C., K. Makni, F. Mnif, M. Boudawara, N. Hamza, N. Rekik, M. Abid, A. Rebai, C. Granier, F. Jarraya, and H. Ayadi. 2008. Association of calpain-10 polymorphisms with type 2 diabetes in the Tunisian population. *Diabetes Metab* **34:** 273-278.

Kim, C.A., M. Delepine, E. Boutet, H. El Mourabit, S. Le Lay, M. Meier, M. Nemani, E. Bridel, C.C. Leite, D.R. Bertola, R.K. Semple, S. O'Rahilly, I. Dugail, J. Capeau, M. Lathrop, and J. Magre. 2008. Association of a homozygous nonsense caveolin-1 mutation with Berardinelli-Seip congenital lipodystrophy. *J Clin Endocrinol Metab* **93:** 1129-1134.

Kim, J.K., M.D. Michael, S.F. Previs, O.D. Peroni, F. Mauvais-Jarvis, S. Neschen, B.B. Kahn, C.R. Kahn, and G.I. Shulman. 2000. Redistribution of substrates to adipose tissue promotes obesity in mice with selective insulin resistance in muscle. *J Clin Invest* **105:** 1791-1797.

King, H. and M. Rewers. 1993. Global estimates for prevalence of diabetes mellitus and impaired glucose tolerance in adults. WHO Ad Hoc Diabetes Reporting Group. *Diabetes Care* **16:** 157-177.

Kissebah, A.H., G.E. Sonnenberg, J. Myklebust, M. Goldstein, K. Broman, R.G. James, J.A. Marks, G.R. Krakower, H.J. Jacob, J. Weber, L. Martin, J. Blangero, and A.G. Comuzzie. 2000. Quantitative trait loci on chromosomes 3 and 17 influence phenotypes of the metabolic syndrome. *Proc Natl Acad Sci U S A* **97:** 14478-14483.

Klein, R.J., C. Zeiss, E.Y. Chew, J.Y. Tsai, R.S. Sackler, C. Haynes, A.K. Henning, J.P. SanGiovanni, S.M. Mane, S.T. Mayne, M.B. Bracken, F.L. Ferris, J. Ott, C. Barnstable, and J. Hoh. 2005. Complement factor H polymorphism in age-related macular degeneration. *Science* **308:** 385-389.

Kloting, N., D. Schleinitz, K. Ruschke, J. Berndt, M. Fasshauer, A. Tonjes, M.R. Schon, P. Kovacs, M. Stumvoll, and M. Bluher. 2008. Inverse relationship between obesity and FTO gene expression in visceral adipose tissue in humans. *Diabetologia* **51:** 641-647.

Knight, B., B.M. Shields, and A.T. Hattersley. 2006. The Exeter Family Study of Childhood Health (EFSOCH): study protocol and methodology. *Paediatric And Perinatal Epidemiology* **20:** 172-179.

Kobberling, J., B. Willms, R. Kattermann, and W. Creutzfeldt. 1975. Lipodystrophy of the extremities. A dominantly inherited syndrome associated with lipatrophic diabetes. *Humangenetik* **29:** 111-120.

Korbel, J.O., A.E. Urban, J.P. Affourtit, B. Godwin, F. Grubert, J.F. Simons, P.M. Kim, D. Palejev, N.J. Carriero, L. Du, B.E. Taillon, Z. Chen, A. Tanzer, A.C. Saunders, J. Chi, F. Yang, N.P. Carter, M.E. Hurles, S.M. Weissman, T.T. Harkins, M.B. Gerstein, M. Egholm, and M. Snyder. 2007. Paired-end mapping reveals extensive structural variation in the human genome. *Science* **318:** 420-426.

Kotronen, A., A. Seppala-Lindroos, R. Bergholm, and H. Yki-Jarvinen. 2008. Tissue specificity of insulin resistance in humans: fat in the liver rather than muscle is associated with features of the metabolic syndrome. *Diabetologia* **51:** 130-138.

Kramer, H.F., C.A. Witczak, N. Fujii, N. Jessen, E.B. Taylor, D.E. Arnolds, K. Sakamoto, M.F. Hirshman, and L.J. Goodyear. 2006. Distinct signals regulate AS160 phosphorylation in response to insulin, AICAR, and contraction in mouse skeletal muscle. *Diabetes* **55:** 2067-2076.

Krebs, M., B. Brunmair, A. Brehm, M. Artwohl, J. Szendroedi, P. Nowotny, E. Roth, C. Furnsinn, M. Promintzer, C. Anderwald, M. Bischof, and M. Roden. 2007. The Mammalian target of rapamycin pathway regulates nutrient-sensitive glucose uptake in man. *Diabetes* **56:** 1600-1607.

Krook, A. and S. O'Rahilly. 1996. Mutant insulin receptors in syndromes of insulin resistance. *Baillieres Clin Endocrinol Metab* **10:** 97-122.

Kryukov, G.V., L.A. Pennacchio, and S.R. Sunyaev. 2007. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *Am J Hum Genet* **80:** 727-739.

Kubota, N., Y. Terauchi, K. Tobe, W. Yano, R. Suzuki, K. Ueki, I. Takamoto, H. Satoh, T. Maki, T. Kubota, M. Moroi, M. Okada-Iwabu, O. Ezaki, R. Nagai, Y. Ueta, T. Kadowaki, and T. Noda. 2004. Insulin receptor substrate 2 plays a crucial role in beta cells and the hypothalamus. *J Clin Invest* **114:** 917-927.

Kubota, N., K. Tobe, Y. Terauchi, K. Eto, T. Yamauchi, R. Suzuki, Y. Tsubamoto, K. Komeda, R. Nakano, H. Miki, S. Satoh, H. Sekihara, S. Sciacchitano, M. Lesniak, S. Aizawa, R. Nagai, S. Kimura, Y. Akanuma, S.I. Taylor, and T. Kadowaki. 2000. Disruption of insulin receptor substrate 2 causes type 2 diabetes because of liver insulin resistance and lack of compensatory beta-cell hyperplasia. *Diabetes* **49:** 1880-1889.

Kubota, T., N. Kubota, M. Moroi, Y. Terauchi, T. Kobayashi, K. Kamata, R. Suzuki, K. Tobe, A. Namiki, S. Aizawa, R. Nagai, T. Kadowaki, and T. Yamaguchi. 2003. Lack of insulin receptor substrate-2 causes progressive neointima formation in response to vessel injury. *Circulation* **107:** 3073-3080.

Kulkarni, R.N., J.C. Bruning, J.N. Winnay, C. Postic, M.A. Magnuson, and C.R. Kahn. 1999. Tissue-specific knockout of the insulin receptor in pancreatic beta cells creates an insulin secretory defect similar to that in type 2 diabetes. *Cell* **96:** 329-339.

Kumar, A., T.E. Harris, S.R. Keller, K.M. Choi, M.A. Magnuson, and J.C. Lawrence, Jr. 2008. Muscle-specific deletion of rictor impairs insulin-stimulated glucose transport and enhances Basal glycogen synthase activity. *Mol Cell Biol* **28:** 61-70.

Lander, E. and L. Kruglyak. 1995. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* **11:** 241-247.

Lander, E.S. L.M. Linton B. Birren C. Nusbaum M.C. Zody J. Baldwin K. Devon K. Dewar M. Doyle W. FitzHugh R. Funke D. Gage K. Harris A. Heaford J. Howland L. Kann J. Lehoczky R. LeVine P. McEwan K. McKernan J. Meldrim J.P. Mesirov C. Miranda W. Morris J. Naylor C. Raymond M. Rosetti R. Santos A. Sheridan C. Sougnez N. Stange-Thomann N. Stojanovic A. Subramanian D. Wyman J. Rogers J. Sulston R. Ainscough S. Beck D. Bentley J. Burton C. Clee N. Carter A. Coulson R. Deadman P. Deloukas A. Dunham I. Dunham R. Durbin L. French D. Grafham S. Gregory T. Hubbard S. Humphray A. Hunt M. Jones C. Lloyd A. McMurray L. Matthews S. Mercer S. Milne J.C. Mullikin A. Mungall R. Plumb M. Ross R. Shownkeen S. Sims R.H. Waterston R.K. Wilson L.W. Hillier J.D. McPherson M.A. Marra E.R. Mardis L.A. Fulton A.T. Chinwalla K.H. Pepin W.R. Gish S.L. Chissoe M.C. Wendl K.D. Delehaunty T.L. Miner A. Delehaunty J.B. Kramer L.L. Cook R.S. Fulton D.L. Johnson P.J. Minx S.W. Clifton T. Hawkins E. Branscomb P. Predki P. Richardson S. Wenning T. Slezak N. Doggett J.F. Cheng A. Olsen S. Lucas C. Elkin E. Uberbacher M. Frazier R.A. Gibbs D.M. Muzny S.E. Scherer J.B. Bouck E.J. Sodergren K.C. Worley C.M. Rives J.H. Gorrell M.L.

Metzker S.L. Naylor R.S. Kucherlapati D.L. Nelson G.M. Weinstock Y. Sakaki A. Fujiyama M. Hattori T. Yada A. Toyoda T. Itoh C. Kawagoe H. Watanabe Y. Totoki T. Taylor J. Weissenbach R. Heilig W. Saurin F. Artiguenave P. Brottier T. Bruls E. Pelletier C. Robert P. Wincker D.R. Smith L. Doucette-Stamm M. Rubenfield K. Weinstock H.M. Lee J. Dubois A. Rosenthal M. Platzer G. Nyakatura S. Taudien A. Rump H. Yang J. Yu J. Wang G. Huang J. Gu L. Hood L. Rowen A. Madan S. Qin R.W. Davis N.A. Federspiel A.P. Abola M.J. Proctor R.M. Myers J. Schmutz M. Dickson J. Grimwood D.R. Cox M.V. Olson R. Kaul C. Raymond N. Shimizu K. Kawasaki S. Minoshima G.A. Evans M. Athanasiou R. Schultz B.A. Roe F. Chen H. Pan J. Ramser H. Lehrach R. Reinhardt W.R. McCombie M. de la Bastide N. Dedhia H. Blocker K. Hornischer G. Nordsiek R. Agarwala L. Aravind J.A. Bailey A. Bateman S. Batzoglou E. Birney P. Bork D.G. Brown C.B. Burge L. Cerutti H.C. Chen D. Church M. Clamp R.R. Copley T. Doerks S.R. Eddy E.E. Eichler T.S. Furey J. Galagan J.G. Gilbert C. Harmon Y. Hayashizaki D. Haussler H. Hermjakob K. Hokamp W. Jang L.S. Johnson T.A. Jones S. Kasif A. Kaspryzk S. Kennedy W.J. Kent P. Kitts E.V. Koonin I. Korf D. Kulp D. Lancet T.M. Lowe A. McLysaght T. Mikkelsen J.V. Moran N. Mulder V.J. Pollara C.P. Ponting G. Schuler J. Schultz G. Slater A.F. Smit E. Stupka J. Szustakowski D. Thierry-Mieg J. Thierry-Mieg L. Wagner J. Wallis R. Wheeler A. Williams Y.I. Wolf K.H. Wolfe S.P. Yang R.F. Yeh F. Collins M.S. Guyer J. Peterson A. Felsenfeld K.A. Wetterstrand A. Patrinos M.J. Morgan P. de Jong J.J. Catanese K. Osoegawa H. Shizuya S. Choi and Y.J. Chen. 2001. Initial sequencing and analysis of the human genome. *Nature* **409:** 860-921.

Langner, C.A., E.H. Birkenmeier, O. Ben-Zeev, M.C. Schotz, H.O. Sweet, M.T. Davisson, and J.I. Gordon. 1989. The fatty liver dystrophy (fld) mutation. A new mutant mouse with a developmental abnormality in triglyceride metabolism and associated tissue-specific defects in lipoprotein lipase and hepatic lipase activities. *J Biol Chem* **264:** 7994-8003.

Lango, H., C.N. Palmer, A.D. Morris, E. Zeggini, A.T. Hattersley, M.I. McCarthy, T.M. Frayling, and M.N. Weedon. 2008. Assessing the combined impact of 18 common genetic variants of modest effect sizes on type 2 diabetes risk. *Diabetes*.

Larance, M., G. Ramm, J. Stockli, E.M. van Dam, S. Winata, V. Wasinger, F. Simpson, M. Graham, J.R. Junutula, M. Guilhaus, and D.E. James. 2005. Characterization of the role of the Rab GTPase-activating protein AS160 in insulin-regulated GLUT4 trafficking. *J Biol Chem* **280:** 37803-37813.

Lauritzen, T., S. Griffin, K. Borch-Johnsen, N.J. Wareham, B.H. Wolffenbuttel, and G. Rutten. 2000. The ADDITION study: proposed trial of the cost-effectiveness of an intensive multifactorial intervention on morbidity and mortality among people with Type 2 diabetes detected by screening. *Int J Obes Relat Metab Disord* **24 Suppl 3:** S6-11.

Laybutt, D.R., A.M. Preston, M.C. Akerfeldt, J.G. Kench, A.K. Busch, A.V. Biankin, and T.J. Biden. 2007. Endoplasmic reticulum stress contributes to beta cell apoptosis in type 2 diabetes. *Diabetologia* **50:** 752-763.

Lee, G.H., R. Proenca, J.M. Montez, K.M. Carroll, J.G. Darvishzadeh, J.I. Lee, and J.M. Friedman. 1996. Abnormal splicing of the leptin receptor in diabetic mice. *Nature* **379:** 632-635.

Lee, Y.S., B.G. Challis, D.A. Thompson, G.S. Yeo, J.M. Keogh, M.E. Madonna, V. Wraight, M. Sims, V. Vatin, D. Meyre, J. Shield, C. Burren, Z. Ibrahim, T. Cheetham, P. Swift, A. Blackwood, C.C. Hung, N.J. Wareham, P. Froguel, G.L. Millhauser, S. O'Rahilly, and I.S. Farooqi. 2006. A POMC variant implicates beta-

melanocyte-stimulating hormone in the control of human energy balance. *Cell Metab* **3:** 135-140.

Levy-Toledano, R., L.H. Caro, D. Accili, and S.I. Taylor. 1994. Investigation of the mechanism of the dominant negative effect of mutations in the tyrosine kinase domain of the insulin receptor. *Embo J* **13:** 835-842.

Lewis, J.P., N.D. Palmer, P.J. Hicks, M.M. Sale, C.D. Langefeld, B.I. Freedman, J. Divers, and D.W. Bowden. 2008. Association analysis in african americans of European-derived type 2 diabetes single nucleotide polymorphisms from whole-genome association studies. *Diabetes* **57:** 2220-2225.

Lindegaard, B., L.F. Larsen, A.B. Hansen, J. Gerstoft, B.K. Pedersen, and K. Reue. 2007. Adipose tissue lipin expression levels distinguish HIV patients with and without lipodystrophy. *Int J Obes (Lond)* **31:** 449-456.

Lohmueller, K.E., C.L. Pearce, M. Pike, E.S. Lander, and J.N. Hirschhorn. 2003. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* **33:** 177-182.

Longo, N., Y. Wang, S.A. Smith, S.D. Langley, L.A. DiMeglio, and D. Giannella-Neto. 2002. Genotype-phenotype correlation in inherited severe insulin resistance. *Hum Mol Genet* **11:** 1465-1475.

Loos, R.J. C.M. Lindgren S. Li E. Wheeler J.H. Zhao I. Prokopenko M. Inouye R.M. Freathy A.P. Attwood J.S. Beckmann S.I. Berndt K.B. Jacobs S.J. Chanock R.B. Hayes S. Bergmann A.J. Bennett S.A. Bingham M. Bochud M. Brown S. Cauchi J.M. Connell C. Cooper G.D. Smith I. Day C. Dina S. De E.T. Dermitzakis A.S. Doney K.S. Elliott P. Elliott D.M. Evans I. Sadaf Farooqi P. Froguel J. Ghori C.J. Groves R. Gwilliam D. Hadley A.S. Hall A.T. Hattersley J. Hebebrand I.M. Heid C. Lamina C. Gieger T. Illig T. Meitinger H.E. Wichmann B. Herrera A. Hinney S.E. Hunt M.R. Jarvelin T. Johnson J.D. Jolley F. Karpe A. Keniry K.T. Khaw R.N. Luben M. Mangino J. Marchini W.L. McArdle R. McGinnis D. Meyre P.B. Munroe A.D. Morris A.R. Ness M.J. Neville A.C. Nica K.K. Ong S. O'Rahilly K.R. Owen C.N. Palmer K. Papadakis S. Potter A. Pouta L. Qi J.C. Randall N.W. Rayner S.M. Ring M.S. Sandhu A. Scherag M.A. Sims K. Song N. Soranzo E.K. Speliotes H.E. Syddall S.A. Teichmann N.J. Timpson J.H. Tobias M. Uda C.I. Vogel C. Wallace D.M. Waterworth M.N. Weedon C.J. Willer Wraight X. Yuan E. Zeggini J.N. Hirschhorn D.P. Strachan W.H. Ouwehand M.J. Caulfield N.J. Samani T.M. Frayling P. Vollenweider G. Waeber V. Mooser P. Deloukas M.I. McCarthy N.J. Wareham I. Barroso K.B. Jacobs S.J. Chanock R.B. Hayes C. Lamina C. Gieger T. Illig T. Meitinger H.E. Wichmann P. Kraft S.E. Hankinson D.J. Hunter F.B. Hu H.N. Lyon B.F. Voight M. Ridderstrale L. Groop P. Scheet S. Sanna G.R. Abecasis G. Albai R. Nagaraja D. Schlessinger A.U. Jackson J. Tuomilehto F.S. Collins M. Boehnke and K.L. Mohlke. 2008. Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nat Genet* **40:** 768-775.

Loos, R.J.F., T. Rankinen, L. Pérusse, A. Tremblay, J.-P. Després, and C. Bouchard. 2007. Association of Lipin 1 Gene Polymorphisms with Measures of Energy and Glucose Metabolism. *Obesity* **15:** 2723-2732.

Love-Gregory, L.D., J. Wasson, J. Ma, C.H. Jin, B. Glaser, B.K. Suarez, and M.A. Permutt. 2004. A common polymorphism in the upstream promoter region of the hepatocyte nuclear factor-4 alpha gene on chromosome 20q is associated with type 2 diabetes and appears to contribute to the evidence for linkage in an ashkenazi jewish population. *Diabetes* **53:** 1134-1140.

Magre, J., M. Delepine, E. Khallouf, T. Gedde-Dahl, Jr., L. Van Maldergem, E. Sobel, J. Papp, M. Meier, A. Megarbane, A. Bachy, A. Verloes, F.H. d'Abronzo, E. Seemanova, R. Assan, N. Baudic, C. Bourut, P. Czernichow, F. Huet, F.

Grigorescu, M. de Kerdanet, D. Lacombe, P. Labrune, M. Lanza, H. Loret, F. Matsuda, J. Navarro, A. Nivelon-Chevalier, M. Polak, J.J. Robert, P. Tric, N. Tubiana-Rufi, C. Vigouroux, J. Weissenbach, S. Savasta, J.A. Maassen, O. Trygstad, P. Bogalho, P. Freitas, J.L. Medina, F. Bonnicci, B.I. Joffe, G. Loyson, V.R. Panz, F.J. Raal, S. O'Rahilly, T. Stephenson, C.R. Kahn, M. Lathrop, and J. Capeau. 2001. Identification of the gene altered in Berardinelli-Seip congenital lipodystrophy on chromosome 11q13. *Nat Genet* **28:** 365-370.

Malecki, M.T., U.S. Jhala, A. Antonellis, L. Fields, A. Doria, T. Orban, M. Saad, J.H. Warram, M. Montminy, and A.S. Krolewski. 1999. Mutations in NEUROD1 are associated with the development of type 2 diabetes mellitus. *Nat Genet* **23:** 323-328.

Manmontri, B., M. Sariahmetoglu, J. Donkor, M.B. Khalil, M. Sundaram, Z. Yao, K. Reue, R. Lehner, and D.N. Brindley. 2008. Glucocorticoids and cyclic AMP selectively increase hepatic lipin-1 expression, and insulin acts antagonistically. *J Lipid Res* **49:** 1056-1067.

Manning, B.D. and L.C. Cantley. 2003. Rheb fills a GAP between TSC and TOR. *Trends Biochem Sci* **28:** 573-576.

Maraldi, N.M., C. Capanni, E. Mattioli, M. Columbaro, S. Squarzoni, W.K. Parnaik, M. Wehnert, and G. Lattanzi. 2007. A pathogenic mechanism leading to partial lipodistrophy and prospects for pharmacological treatment of insulin resistance syndrome. *Acta Biomed* **78 Suppl 1:** 207-215.

Marchetti, P., S. Del Guerra, L. Marselli, R. Lupi, M. Masini, M. Pollera, M. Bugliani, U. Boggi, F. Vistoli, F. Mosca, and S. Del Prato. 2004. Pancreatic islets from type 2 diabetic patients have functional defects and increased apoptosis that are ameliorated by metformin. *J Clin Endocrinol Metab* **89:** 5535-5541.

Marchetti, P., F. Dotta, D. Lauro, and F. Purrello. 2008. An overview of pancreatic beta-cell defects in human type 2 diabetes: implications for treatment. *Regul Pept* **146:** 4-11.

Marshall, C., G.A. Hitman, C.J. Partridge, A. Clark, H. Ma, T.R. Shearer, and M.D. Turner. 2005. Evidence that an isoform of calpain-10 is a regulator of exocytosis in pancreatic beta-cells. *Mol Endocrinol* **19:** 213-224.

Mason, T.M., T. Goh, V. Tchipashvili, H. Sandhu, N. Gupta, G.F. Lewis, and A. Giacca. 1999. Prolonged elevation of plasma free fatty acids desensitizes the insulin secretory response to glucose in vivo in rats. *Diabetes* **48:** 524-530.

Matschinsky, F., Y. Liang, P. Kesavan, L. Wang, P. Froguel, G. Velho, D. Cohen, M.A. Permutt, Y. Tanizawa, T.L. Jetton, and et al. 1993. Glucokinase as pancreatic beta cell glucose sensor and diabetes gene. *J Clin Invest* **92:** 2092-2098.

Mauvais-Jarvis, F., K. Ueki, D.A. Fruman, M.F. Hirshman, K. Sakamoto, L.J. Goodyear, M. Iannacone, D. Accili, L.C. Cantley, and C.R. Kahn. 2002. Reduced expression of the murine p85alpha subunit of phosphoinositide 3-kinase improves insulin signaling and ameliorates diabetes. *J Clin Invest* **109:** 141-149.

Mauvais-Jarvis, F., A. Virkamaki, M.D. Michael, J.N. Winnay, A. Zisman, R.N. Kulkarni, and C.R. Kahn. 2000. A model to explore the interaction between muscle insulin resistance and beta-cell dysfunction in the development of type 2 diabetes. *Diabetes* **49:** 2126-2134.

Meier, M., H.H. Klein, J. Kramer, M. Drenckhan, and M. Schutt. 2007. Calpain inhibition impairs glycogen syntheses in HepG2 hepatoma cells without altering insulin signaling. *J Endocrinol* **193:** 45-51.

Meigs, J.B., A.K. Manning, C.S. Fox, J.C. Florez, C. Liu, L.A. Cupples, and J. Dupuis. 2007. Genome-wide association with diabetes-related traits in the Framingham Heart Study. *BMC Med Genet* **8 Suppl 1:** S16.

Michael, M.D., R.N. Kulkarni, C. Postic, S.F. Previs, G.I. Shulman, M.A. Magnuson, and C.R. Kahn. 2000. Loss of insulin signaling in hepatocytes leads to severe insulin resistance and progressive hepatic dysfunction. *Mol Cell* **6:** 87-97.

Miller, M.P. and S. Kumar. 2001. Understanding human disease mutations through the use of interspecific genetic variation. *Hum Mol Genet* **10:** 2319-2328.

Minokoshi, Y., Y.B. Kim, O.D. Peroni, L.G. Fryer, C. Muller, D. Carling, and B.B. Kahn. 2002. Leptin stimulates fatty-acid oxidation by activating AMP-activated protein kinase. *Nature* **415:** 339-343.

Minton, J.A., A.T. Hattersley, K. Owen, M.I. McCarthy, M. Walker, F. Latif, T. Barrett, and T.M. Frayling. 2002. Association studies of genetic variation in the WFS1 gene and type 2 diabetes in U.K. populations. *Diabetes* **51:** 1287-1290.

Miyake, K., Y. Horikawa, K. Hara, K. Yasuda, H. Osawa, H. Furuta, Y. Hirota, K. Yamagata, Y. Hinokio, Y. Oka, N. Iwasaki, Y. Iwamoto, Y. Yamada, Y. Seino, H. Maegawa, A. Kashiwagi, K. Yamamoto, K. Tokunaga, J. Takeda, H. Makino, K. Nanjo, T. Kadowaki, and M. Kasuga. 2008. Association of TCF7L2 polymorphisms with susceptibility to type 2 diabetes in 4,087 Japanese subjects. *J Hum Genet* **53:** 174-180.

Moitra, J., M.M. Mason, M. Olive, D. Krylov, O. Gavrilova, B. Marcus-Samuels, L. Feigenbaum, E. Lee, T. Aoyama, M. Eckhaus, M.L. Reitman, and C. Vinson. 1998. Life without white fat: a transgenic mouse. *Genes Dev* **12:** 3168-3181.

Moller, D.E. and J.S. Flier. 1988. Detection of an alteration in the insulin-receptor gene in a patient with insulin resistance, acanthosis nigricans, and the polycystic ovary syndrome (type A insulin resistance). *N Engl J Med* **319:** 1526-1529.

Mootha, V.K., C.M. Lindgren, K.F. Eriksson, A. Subramanian, S. Sihag, J. Lehar, P. Puigserver, E. Carlsson, M. Ridderstrale, E. Laurila, N. Houstis, M.J. Daly, N. Patterson, J.P. Mesirov, T.R. Golub, P. Tamayo, B. Spiegelman, E.S. Lander, J.N. Hirschhorn, D. Altshuler, and L.C. Groop. 2003. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* **34:** 267-273.

Morino, K., K.F. Petersen, S. Dufour, D. Befroy, J. Frattini, N. Shatzkes, S. Neschen, M.F. White, S. Bilz, S. Sono, M. Pypaert, and G.I. Shulman. 2005. Reduced mitochondrial density and increased IRS-1 serine phosphorylation in muscle of insulin-resistant offspring of type 2 diabetic parents. *J Clin Invest* **115:** 3587-3593.

Morino, K., K.F. Petersen, and G.I. Shulman. 2006. Molecular mechanisms of insulin resistance in humans and their potential links with mitochondrial dysfunction. *Diabetes* **55 Suppl 2:** S9-S15.

Morris, A.D., D.I. Boyle, R. MacAlpine, A. Emslie-Smith, R.T. Jung, R.W. Newton, and T.M. MacDonald. 1997. The diabetes audit and research in Tayside Scotland (DARTS) study: electronic record linkage to create a diabetes register. DARTS/MEMO Collaboration. *Bmj* **315:** 524-528.

Murphy, R., S. Ellard, and A.T. Hattersley. 2008. Clinical implications of a molecular genetic classification of monogenic beta-cell diabetes. *Nat Clin Pract Endocrinol Metab* **4:** 200-213.

Nakatani, Y., H. Kaneto, D. Kawamori, K. Yoshiuchi, M. Hatazaki, T.A. Matsuoka, K. Ozawa, S. Ogawa, M. Hori, Y. Yamasaki, and M. Matsuhisa. 2005. Involvement of endoplasmic reticulum stress in insulin resistance and diabetes. *J Biol Chem* **280:** 847-851.

Nauck, M.A., J.J. Meier, A.V. Wolfersdorff, H. Tillil, W. Creutzfeldt, and J. Kobberling. 2003. A 25-year follow-up study of glucose tolerance in first-degree relatives of type 2 diabetic patients: association of impaired or diabetic glucose tolerance with other components of the metabolic syndrome. *Acta Diabetol* **40:** 163-172.

Neve, B., M.E. Fernandez-Zapico, V. Ashkenazi-Katalan, C. Dina, Y.H. Hamid, E. Joly, E. Vaillant, Y. Benmezroua, E. Durand, N. Bakaher, V. Delannoy, M. Vaxillaire, T. Cook, G.M. Dallinga-Thie, H. Jansen, M.A. Charles, K. Clement, P. Galan, S. Hercberg, N. Helbecque, G. Charpentier, M. Prentki, T. Hansen, O. Pedersen, R. Urrutia, D. Melloul, and P. Froguel. 2005. Role of transcription factor KLF11 and its diabetes-associated gene variants in pancreatic beta cell function. *Proc Natl Acad Sci U S A* **102:** 4807-4812.

Ng, M.C., K.S. Park, B. Oh, C.H. Tam, Y.M. Cho, H.D. Shin, V.K. Lam, R.C. Ma, W.Y. So, Y.S. Cho, H.L. Kim, H.K. Lee, J.C. Chan, and N.H. Cho. 2008a. Implication of genetic variants near TCF7L2, SLC30A8, HHEX, CDKAL1, CDKN2A/B, IGF2BP2, and FTO in type 2 diabetes and obesity in 6,719 Asians. *Diabetes* **57:** 2226-2233.

Ng, P.C., S. Levy, J. Huang, T.B. Stockwell, B.P. Walenz, K. Li, N. Axelrod, D.A. Busam, R.L. Strausberg, and J.C. Venter. 2008b. Genetic variation in an individual human exome. *PLoS Genet* **4:** e1000160.

Nielsen, E.M., L. Hansen, B. Carstensen, S.M. Echwald, T. Drivsholm, C. Glumer, B. Thorsteinsson, K. Borch-Johnsen, T. Hansen, and O. Pedersen. 2003. The E23K variant of Kir6.2 associates with impaired post-OGTT serum insulin response and increased risk of type 2 diabetes. *Diabetes* **52:** 573-577.

Nistico, L., R. Buzzetti, L.E. Pritchard, B. Van der Auwera, C. Giovannini, E. Bosi, M.T. Larrad, M.S. Rios, C.C. Chow, C.S. Cockram, K. Jacobs, C. Mijovic, S.C. Bain, A.H. Barnett, C.L. Vandewalle, F. Schuit, F.K. Gorus, R. Tosi, P. Pozzilli, and J.A. Todd. 1996. The CTLA-4 gene region of chromosome 2q33 is linked to, and associated with, type 1 diabetes. Belgian Diabetes Registry. *Hum Mol Genet* **5:** 1075-1080.

Njolstad, P.R., O. Sovik, A. Cuesta-Munoz, L. Bjorkhaug, O. Massa, F. Barbetti, D.E. Undlien, C. Shiota, M.A. Magnuson, A. Molven, F.M. Matschinsky, and G.I. Bell. 2001. Neonatal diabetes mellitus due to complete glucokinase deficiency. *N Engl J Med* **344:** 1588-1592.

Novelli, G., A. Muchir, F. Sangiuolo, A. Helbling-Leclerc, M.R. D'Apice, C. Massart, F. Capon, P. Sbraccia, M. Federici, R. Lauro, C. Tudisco, R. Pallotta, G. Scarano, B. Dallapiccola, L. Merlini, and G. Bonne. 2002. Mandibuloacral dysplasia is caused by a mutation in LMNA-encoding lamin A/C. *Am J Hum Genet* **71:** 426-431.

O'Hara, L., G.S. Han, S. Peak-Chew, N. Grimsey, G.M. Carman, and S. Siniossoglou. 2006. Control of phospholipid synthesis by phosphorylation of the yeast lipin Pah1p/Smp2p Mg2+-dependent phosphatidate phosphatase. *J Biol Chem* **281:** 34537-34548.

O'Rahilly, S. 1997. Science, medicine, and the future. Non-insulin dependent diabetes mellitus: the gathering storm. *Bmj* **314:** 955-959.

Obici, S., Z. Feng, G. Karkanias, D.G. Baskin, and L. Rossetti. 2002. Decreasing hypothalamic insulin receptors causes hyperphagia and insulin resistance in rats. *Nat Neurosci* **5:** 566-572.

Ogawa, Y., H. Masuzaki, K. Hosoda, M. Aizawa-Abe, J. Suga, M. Suda, K. Ebihara, H. Iwai, N. Matsuoka, N. Satoh, H. Odaka, H. Kasuga, Y. Fujisawa, G. Inoue, H. Nishimura, Y. Yoshimasa, and K. Nakao. 1999. Increased glucose metabolism and insulin sensitivity in transgenic skinny mice overexpressing leptin. *Diabetes* **48:** 1822-1829.

Okada, T., C.W. Liew, J. Hu, C. Hinault, M.D. Michael, J. Krtzfeldt, C. Yin, M. Holzenberger, M. Stoffel, and R.N. Kulkarni. 2007. Insulin receptors in beta-cells are critical for islet compensatory growth response to insulin resistance. *Proc Natl Acad Sci U S A* **104:** 8977-8982.

Osman, A.A., M. Saito, C. Makepeace, M.A. Permutt, P. Schlesinger, and M. Mueckler. 2003. Wolframin expression induces novel ion channel activity in endoplasmic reticulum membranes and increases intracellular calcium. *J Biol Chem* **278:** 52755-52762.

Owen, K.R., A. Stride, S. Ellard, and A.T. Hattersley. 2003. Etiological investigation of diabetes in young adults presenting with apparent type 2 diabetes. *Diabetes Care* **26:** 2088-2093.

Ozcan, U., Q. Cao, E. Yilmaz, A.H. Lee, N.N. Iwakoshi, E. Ozdelen, G. Tuncman, C. Gorgun, L.H. Glimcher, and G.S. Hotamisligil. 2004. Endoplasmic reticulum stress links obesity, insulin action, and type 2 diabetes. *Science* **306:** 457-461.

Ozcan, U., E. Yilmaz, L. Ozcan, M. Furuhashi, E. Vaillancourt, R.O. Smith, C.Z. Gorgun, and G.S. Hotamisligil. 2006. Chemical chaperones reduce ER stress and restore glucose homeostasis in a mouse model of type 2 diabetes. *Science* **313:** 1137-1140.

Ozes, O.N., H. Akca, L.D. Mayo, J.A. Gustin, T. Maehama, J.E. Dixon, and D.B. Donner. 2001. A phosphatidylinositol 3-kinase/Akt/mTOR pathway mediates and PTEN antagonizes tumor necrosis factor inhibition of insulin signaling through insulin receptor substrate-1. *Proc Natl Acad Sci U S A* **98:** 4640-4645.

Pan, D.A., S. Lillioja, A.D. Kriketos, M.R. Milner, L.A. Baur, C. Bogardus, A.B. Jenkins, and L.H. Storlien. 1997. Skeletal muscle triglyceride levels are inversely related to insulin action. *Diabetes* **46:** 983-988.

Paolisso, G., A. Gambardella, L. Amato, R. Tortoriello, A. D'Amore, M. Varricchio, and F. D'Onofrio. 1995. Opposite effects of short- and long-term fatty acid infusion on insulin secretion in healthy subjects. *Diabetologia* **38:** 1295-1299.

Parker, P.J., F.B. Caudwell, and P. Cohen. 1983. Glycogen synthase from rabbit skeletal muscle; effect of insulin on the state of phosphorylation of the seven phosphoserine residues in vivo. *Eur J Biochem* **130:** 227-234.

Patti, M.E., A.J. Butte, S. Crunkhorn, K. Cusi, R. Berria, S. Kashyap, Y. Miyazaki, I. Kohane, M. Costello, R. Saccone, E.J. Landaker, A.B. Goldfine, E. Mun, R. DeFronzo, J. Finlayson, C.R. Kahn, and L.J. Mandarino. 2003. Coordinated reduction of genes of oxidative metabolism in humans with insulin resistance and diabetes: Potential role of PGC1 and NRF1. *Proc Natl Acad Sci U S A* **100:** 8466-8471.

Payne, V.A., N. Grimsey, A. Tuthill, S. Virtue, S.L. Gray, E. Dalla Nora, R.K. Semple, S. O'Rahilly, and J.J. Rochford. 2008. The human lipodystrophy gene BSCL2/seipin may be essential for normal adipocyte differentiation. *Diabetes* **57:** 2055-2060.

Pederson, T.M., D.L. Kramer, and C.M. Rondinone. 2001. Serine/threonine phosphorylation of IRS-1 triggers its degradation: possible regulation by tyrosine phosphorylation. *Diabetes* **50:** 24-31.

Pende, M., S.C. Kozma, M. Jaquet, V. Oorschot, R. Burcelin, Y. Le Marchand-Brustel, J. Klumperman, B. Thorens, and G. Thomas. 2000. Hypoinsulinaemia, glucose intolerance and diminished beta-cell size in S6K1-deficient mice. *Nature* **408:** 994-997.

Perez-Carreras, M., P. Del Hoyo, M.A. Martin, J.C. Rubio, A. Martin, G. Castellano, F. Colina, J. Arenas, and J.A. Solis-Herruzo. 2003. Defective hepatic mitochondrial respiratory chain in patients with nonalcoholic steatohepatitis. *Hepatology* **38:** 999-1007.

Permutt, M.A., J.C. Wasson, B.K. Suarez, J. Lin, J. Thomas, J. Meyer, S. Lewitzky, J.S. Rennich, A. Parker, L. DuPrat, S. Maruti, S. Chayen, and B. Glaser. 2001. A genome scan for type 2 diabetes susceptibility loci in a genetically isolated population. *Diabetes* **50:** 681-685.

Perseghin, G., P. Scifo, F. De Cobelli, E. Pagliato, A. Battezzati, C. Arcelloni, A. Vanzulli, G. Testolin, G. Pozza, A. Del Maschio, and L. Luzi. 1999. Intramyocellular triglyceride content is a determinant of in vivo insulin resistance in humans: a 1H-13C nuclear magnetic resonance spectroscopy assessment in offspring of type 2 diabetic parents. *Diabetes* **48:** 1600-1606.

Pessayre, D. 2007. Role of mitochondria in non-alcoholic fatty liver disease. *J Gastroenterol Hepatol* **22 Suppl 1:** S20-27.

Peterfy, M., J. Phan, and K. Reue. 2005. Alternatively spliced lipin isoforms exhibit distinct expression pattern, subcellular localization, and role in adipogenesis. *J Biol Chem* **280:** 32883-32889.

Peterfy, M., J. Phan, P. Xu, and K. Reue. 2001. Lipodystrophy in the fld mouse results from mutation of a new gene encoding a nuclear protein, lipin. *Nat Genet* **27:** 121-124.

Peters, J.M., R. Barnes, L. Bennett, W.M. Gitomer, A.M. Bowcock, and A. Garg. 1998. Localization of the gene for familial partial lipodystrophy (Dunnigan variety) to chromosome 1q21-22. *Nat Genet* **18:** 292-295.

Petersen, K.F., D. Befroy, S. Dufour, J. Dziura, C. Ariyan, D.L. Rothman, L. DiPietro, G.W. Cline, and G.I. Shulman. 2003. Mitochondrial dysfunction in the elderly: possible role in insulin resistance. *Science* **300:** 1140-1142.

Petersen, K.F., S. Dufour, D. Befroy, R. Garcia, and G.I. Shulman. 2004. Impaired mitochondrial activity in the insulin-resistant offspring of patients with type 2 diabetes. *N Engl J Med* **350:** 664-671.

Petersen, K.F., S. Dufour, and G.I. Shulman. 2005. Decreased insulin-stimulated ATP synthesis and phosphate transport in muscle of insulin-resistant offspring of type 2 diabetic parents. *PLoS Med* **2:** e233.

Phan, J., M. Peterfy, and K. Reue. 2004. Lipin expression preceding peroxisome proliferator-activated receptor-gamma is critical for adipogenesis in vivo and in vitro. *J Biol Chem* **279:** 29558-29564.

Phan, J. and K. Reue. 2005. Lipin, a lipodystrophy and obesity gene. *Cell Metab* **1:** 73-83.

Phillips, D.I., S. Caddy, V. Ilic, B.A. Fielding, K.N. Frayn, A.C. Borthwick, and R. Taylor. 1996. Intramuscular triglyceride and muscle insulin sensitivity: evidence for a relationship in nondiabetic subjects. *Metabolism* **45:** 947-950.

Plengvidhya, N., S. Kooptiwut, N. Songtawee, A. Doi, H. Furuta, M. Nishi, K. Nanjo, W. Tantibhedhyangkul, W. Boonyasrisawat, P.T. Yenchitsomanus, A. Doria, and N. Banchuin. 2007. PAX4 mutations in Thais with maturity onset diabetes of the young. *J Clin Endocrinol Metab* **92:** 2821-2826.

Poulsen, P. and A. Vaag. 2001. Glucose and insulin metabolism in twins: influence of zygosity and birth weight. *Twin Res* **4:** 350-355.

Prentki, M. and B.E. Corkey. 1996. Are the beta-cell signaling molecules malonyl-CoA and cystolic long-chain acyl-CoA implicated in multiple tissue defects of obesity and NIDDM? *Diabetes* **45:** 273-283.

Prieur, X., Y.C. Tung, J.L. Griffin, I.S. Farooqi, S. O'Rahilly, and A.P. Coll. 2008. Leptin regulates peripheral lipid metabolism primarily through central effects on food intake. *Endocrinology* **149:** 5432-5439.

Proks, P., A.L. Arnold, J. Bruining, C. Girard, S.E. Flanagan, B. Larkin, K. Colclough, A.T. Hattersley, F.M. Ashcroft, and S. Ellard. 2006. A heterozygous activating mutation in the sulphonylurea receptor SUR1 (ABCC8) causes neonatal diabetes. *Hum Mol Genet* **15:** 1793-1800.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M.A. Ferreira, D. Bender, J. Maller, P. Sklar, P.I. de Bakker, M.J. Daly, and P.C. Sham. 2007. PLINK: a tool set for

whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81:** 559-575.

Raeder, H., S. Johansson, P.I. Holm, I.S. Haldorsen, E. Mas, V. Sbarra, I. Nermoen, S.A. Eide, L. Grevle, L. Bjorkhaug, J.V. Sagen, L. Aksnes, O. Sovik, D. Lombardo, A. Molven, and P.R. Njolstad. 2006. Mutations in the CEL VNTR cause a syndrome of diabetes and pancreatic exocrine dysfunction. *Nat Genet* **38:** 54-62.

Rahier, J., R.M. Goebbels, and J.C. Henquin. 1983. Cellular composition of the human diabetic pancreas. *Diabetologia* **24:** 366-371.

Ranade, K., M.S. Chang, C.T. Ting, D. Pei, C.F. Hsiao, M. Olivier, R. Pesich, J. Hebert, Y.D. Chen, V.J. Dzau, D. Curb, R. Olshen, N. Risch, D.R. Cox, and D. Botstein. 2001. High-throughput genotyping with single nucleotide polymorphisms. *Genome Res* **11:** 1262-1268.

Rasmussen, S.K., S.A. Urhammer, L. Berglund, J.N. Jensen, L. Hansen, S.M. Echwald, K. Borch-Johnsen, Y. Horikawa, H. Mashima, H. Lithell, N.J. Cox, T. Hansen, G.I. Bell, and O. Pedersen. 2002. Variants within the calpain-10 gene on chromosome 2q37 (NIDDM1) and relationships to type 2 diabetes, insulin resistance, and impaired acute insulin secretion among Scandinavian Caucasians. *Diabetes* **51:** 3561-3567.

Rathmann, W., B. Haastert, A. Icks, G. Giani, S. Hennings, J. Mitchell, S. Curran, and N.J. Wareham. 2001. Low faecal elastase 1 concentrations in type 2 diabetes mellitus. *Scand J Gastroenterol* **36:** 1056-1061.

Redon, R., S. Ishikawa, K.R. Fitch, L. Feuk, G.H. Perry, T.D. Andrews, H. Fiegler, M.H. Shapero, A.R. Carson, W. Chen, E.K. Cho, S. Dallaire, J.L. Freeman, J.R. Gonzalez, M. Gratacos, J. Huang, D. Kalaitzopoulos, D. Komura, J.R. MacDonald, C.R. Marshall, R. Mei, L. Montgomery, K. Nishimura, K. Okamura, F. Shen, M.J. Somerville, J. Tchinda, A. Valsesia, C. Woodwark, F. Yang, J. Zhang, T. Zerjal, J. Zhang, L. Armengol, D.F. Conrad, X. Estivill, C. Tyler-Smith, N.P. Carter, H. Aburatani, C. Lee, K.W. Jones, S.W. Scherer, and M.E. Hurles. 2006. Global variation in copy number in the human genome. *Nature* **444:** 444-454.

Reue, K., P. Xu, X.P. Wang, and B.G. Slavin. 2000. Adipose tissue deficiency, glucose intolerance, and increased atherosclerosis result from mutation in the mouse fatty liver dystrophy (fld) gene. *J Lipid Res* **41:** 1067-1076.

Reue, K. and P. Zhang. 2008. The lipin protein family: dual roles in lipid biosynthesis and gene expression. *FEBS Lett* **582:** 90-96.

Reynisdottir, I., G. Thorleifsson, R. Benediktsson, G. Sigurdsson, V. Emilsson, A.S. Einarsdottir, E.E. Hjorleifsdottir, G.T. Orlygsdottir, G.T. Bjornsdottir, J. Saemundsdottir, S. Halldorsson, S. Hrafnkelsdottir, S.B. Sigurjonsdottir, S. Steinsdottir, M. Martin, J.P. Kochan, B.K. Rhees, S.F. Grant, M.L. Frigge, A. Kong, V. Gudnason, K. Stefansson, and J.R. Gulcher. 2003. Localization of a susceptibility gene for type 2 diabetes to chromosome 5q34-q35.2. *Am J Hum Genet* **73:** 323-335.

Ricketts, C., M. Zatyka, and T. Barrett. 2006. The characterisation of the human Wolfram syndrome gene promoter demonstrating regulation by Sp1 and Sp3 transcription factors. *Biochim Biophys Acta* **1759:** 367-377.

Riggs, A.C., E. Bernal-Mizrachi, M. Ohsugi, J. Wasson, S. Fatrai, C. Welling, J. Murray, R.E. Schmidt, P.L. Herrera, and M.A. Permutt. 2005. Mice conditionally lacking the Wolfram gene in pancreatic islet beta cells exhibit diabetes as a result of enhanced endoplasmic reticulum stress and apoptosis. *Diabetologia* **48:** 2313-2321.

Risch, N. and K. Merikangas. 1996. The future of genetic studies of complex human diseases. *Science* **273:** 1516-1517.

Ritov, V.B., E.V. Menshikova, J. He, R.E. Ferrell, B.H. Goodpaster, and D.E. Kelley. 2005. Deficiency of subsarcolemmal mitochondria in obesity and type 2 diabetes. *Diabetes* **54:** 8-14.

Romeo, S., L.A. Pennacchio, Y. Fu, E. Boerwinkle, A. Tybjaerg-Hansen, H.H. Hobbs, and J.C. Cohen. 2007. Population-based resequencing of ANGPTL4 uncovers variations that reduce triglycerides and increase HDL. *Nat Genet* **39:** 513-516.

Ruvinsky, I., N. Sharon, T. Lerer, H. Cohen, M. Stolovich-Rain, T. Nir, Y. Dor, P. Zisman, and O. Meyuhas. 2005. Ribosomal protein S6 phosphorylation is a determinant of cell size and glucose homeostasis. *Genes Dev* **19:** 2199-2211.

Ryysy, L., A.M. Hakkinen, T. Goto, S. Vehkavaara, J. Westerbacka, J. Halavaara, and H. Yki-Jarvinen. 2000. Hepatic fat content and insulin action on free fatty acids and glucose metabolism rather than insulin absorption are associated with insulin requirements during insulin therapy in type 2 diabetic patients. *Diabetes* **49:** 749-758.

Sachidanandam, R., D. Weissman, S.C. Schmidt, J.M. Kakol, L.D. Stein, G. Marth, S. Sherry, J.C. Mullikin, B.J. Mortimore, D.L. Willey, S.E. Hunt, C.G. Cole, P.C. Coggill, C.M. Rice, Z. Ning, J. Rogers, D.R. Bentley, P.Y. Kwok, E.R. Mardis, R.T. Yeh, B. Schultz, L. Cook, R. Davenport, M. Dante, L. Fulton, L. Hillier, R.H. Waterston, J.D. McPherson, B. Gilman, S. Schaffner, W.J. Van Etten, D. Reich, J. Higgins, M.J. Daly, B. Blumenstiel, J. Baldwin, N. Stange-Thomann, M.C. Zody, L. Linton, E.S. Lander, and D. Altshuler. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409:** 928-933.

Saito, K., T. Takahashi, N. Yaginuma, and N. Iwama. 1978. Islet morphometry in the diabetic pancreas of man. *Tohoku J Exp Med* **125:** 185-197.

Sako, Y. and V.E. Grill. 1990. A 48-hour lipid infusion in the rat time-dependently inhibits glucose-induced insulin secretion and B cell oxidation through a process likely coupled to fatty acid oxidation. *Endocrinology* **127:** 1580-1589.

Sakuraba, H., H. Mizukami, N. Yagihashi, R. Wada, C. Hanyu, and S. Yagihashi. 2002. Reduced beta-cell mass and expression of oxidative stress-related DNA damage in the islet of Japanese Type II diabetic patients. *Diabetologia* **45:** 85-96.

Sanchez-Pulido, L. and M.A. Andrade-Navarro. 2007. The FTO (fat mass and obesity associated) gene codes for a novel member of the non-heme dioxygenase superfamily. *BMC Biochem* **8:** 23.

Sandhu, M.S., M.N. Weedon, K.A. Fawcett, J. Wasson, S.L. Debenham, A. Daly, H. Lango, T.M. Frayling, R.J. Neumann, R. Sherva, I. Blech, P.D. Pharoah, C.N. Palmer, C. Kimber, R. Tavendale, A.D. Morris, M.I. McCarthy, M. Walker, G. Hitman, B. Glaser, M.A. Permutt, A.T. Hattersley, N.J. Wareham, and I. Barroso. 2007. Common variants in WFS1 confer risk of type 2 diabetes. *Nat Genet* **39:** 951-953.

Sano, H., S. Kane, E. Sano, C.P. Miinea, J.M. Asara, W.S. Lane, C.W. Garner, and G.E. Lienhard. 2003. Insulin-stimulated phosphorylation of a Rab GTPase-activating protein regulates GLUT4 translocation. *J Biol Chem* **278:** 14599-14602.

Santos-Rosa, H., J. Leung, N. Grimsey, S. Peak-Chew, and S. Siniossoglou. 2005. The yeast lipin Smp2 couples phospholipid biosynthesis to nuclear membrane growth. *Embo J* **24:** 1931-1941.

Sarbassov, D.D., S.M. Ali, D.H. Kim, D.A. Guertin, R.R. Latek, H. Erdjument-Bromage, P. Tempst, and D.M. Sabatini. 2004. Rictor, a novel binding partner of mTOR, defines a rapamycin-insensitive and raptor-independent pathway that regulates the cytoskeleton. *Curr Biol* **14:** 1296-1302.

Sarbassov, D.D., S.M. Ali, and D.M. Sabatini. 2005a. Growing roles for the mTOR pathway. *Curr Opin Cell Biol* **17:** 596-603.

Sarbassov, D.D., S.M. Ali, S. Sengupta, J.H. Sheen, P.P. Hsu, A.F. Bagley, A.L. Markhard, and D.M. Sabatini. 2006. Prolonged rapamycin treatment inhibits mTORC2 assembly and Akt/PKB. *Mol Cell* **22:** 159-168.

Sarbassov, D.D., D.A. Guertin, S.M. Ali, and D.M. Sabatini. 2005b. Phosphorylation and regulation of Akt/PKB by the rictor-mTOR complex. *Science* **307:** 1098-1101.

Savage, D.B., M. Agostini, I. Barroso, M. Gurnell, J. Luan, A. Meirhaeghe, A.H. Harding, G. Ihrke, O. Rajanayagam, M.A. Soos, S. George, D. Berger, E.L. Thomas, J.D. Bell, K. Meeran, R.J. Ross, A. Vidal-Puig, N.J. Wareham, S. O'Rahilly, V.K. Chatterjee, and A.J. Schafer. 2002. Digenic inheritance of severe insulin resistance in a human pedigree. *Nat Genet* **31:** 379-384.

Savage, D.B., K.F. Petersen, and G.I. Shulman. 2007. Disordered lipid metabolism and the pathogenesis of insulin resistance. *Physiol Rev* **87:** 507-520.

Savitsky, K., A. Bar-Shira, S. Gilad, G. Rotman, Y. Ziv, L. Vanagaite, D.A. Tagle, S. Smith, T. Uziel, S. Sfez, M. Ashkenazi, I. Pecker, M. Frydman, R. Harnik, S.R. Patanjali, A. Simmons, G.A. Clines, A. Sartiel, R.A. Gatti, L. Chessa, O. Sanal, M.F. Lavin, N.G. Jaspers, A.M. Taylor, C.F. Arlett, T. Miki, S.M. Weissman, M. Lovett, F.S. Collins, and Y. Shiloh. 1995. A single ataxia telangiectasia gene with a product similar to PI-3 kinase. *Science* **268:** 1749-1753.

Saxena, R., B.F. Voight, V. Lyssenko, N.P. Burtt, P.I. de Bakker, H. Chen, J.J. Roix, S. Kathiresan, J.N. Hirschhorn, M.J. Daly, T.E. Hughes, L. Groop, D. Altshuler, P. Almgren, J.C. Florez, J. Meyer, K. Ardlie, K. Bengtsson Bostrom, B. Isomaa, G. Lettre, U. Lindblad, H.N. Lyon, O. Melander, C. Newton-Cheh, P. Nilsson, M. Orho-Melander, L. Rastam, E.K. Speliotes, M.R. Taskinen, T. Tuomi, C. Guiducci, A. Berglund, J. Carlson, L. Gianniny, R. Hackett, L. Hall, J. Holmkvist, E. Laurila, M. Sjogren, M. Sterner, A. Surti, M. Svensson, M. Svensson, R. Tewhey, B. Blumenstiel, M. Parkin, M. Defelice, R. Barry, W. Brodeur, J. Camarata, N. Chia, M. Fava, J. Gibbons, B. Handsaker, C. Healy, K. Nguyen, C. Gates, C. Sougnez, D. Gage, M. Nizzari, S.B. Gabriel, G.W. Chirn, Q. Ma, H. Parikh, D. Richardson, D. Ricke, and S. Purcell. 2007. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316:** 1331-1336.

Scheuner, D., B. Song, E. McEwen, C. Liu, R. Laybutt, P. Gillespie, T. Saunders, S. Bonner-Weir, and R.J. Kaufman. 2001. Translational control is required for the unfolded protein response and in vivo glucose homeostasis. *Mol Cell* **7:** 1165-1176.

Scheuner, D., D. Vander Mierde, B. Song, D. Flamez, J.W. Creemers, K. Tsukamoto, M. Ribick, F.C. Schuit, and R.J. Kaufman. 2005. Control of mRNA translation preserves endoplasmic reticulum function in beta cells and maintains glucose homeostasis. *Nat Med* **11:** 757-764.

Schulz, L.O., P.H. Bennett, E. Ravussin, J.R. Kidd, K.K. Kidd, J. Esparza, and M.E. Valencia. 2006. Effects of traditional and western environments on prevalence of type 2 diabetes in Pima Indians in Mexico and the U.S. *Diabetes Care* **29:** 1866-1871.

Scott, L.J., K.L. Mohlke, L.L. Bonnycastle, C.J. Willer, Y. Li, W.L. Duren, M.R. Erdos, H.M. Stringham, P.S. Chines, A.U. Jackson, L. Prokunina-Olsson, C.J. Ding, A.J. Swift, N. Narisu, T. Hu, R. Pruim, R. Xiao, X.Y. Li, K.N. Conneely, N.L. Riebow, A.G. Sprau, M. Tong, P.P. White, K.N. Hetrick, M.W. Barnhart, C.W. Bark, J.L. Goldstein, L. Watkins, F. Xiang, J. Saramies, T.A. Buchanan, R.M. Watanabe, T.T. Valle, L. Kinnunen, G.R. Abecasis, E.W. Pugh, K.F. Doheny, R.N. Bergman, J. Tuomilehto, F.S. Collins, and M. Boehnke. 2007. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316:** 1341-1345.

Sebat, J., B. Lakshmi, D. Malhotra, J. Troge, C. Lese-Martin, T. Walsh, B. Yamrom, S. Yoon, A. Krasnitz, J. Kendall, A. Leotta, D. Pai, R. Zhang, Y.H. Lee, J. Hicks, S.J. Spence, A.T. Lee, K. Puura, T. Lehtimaki, D. Ledbetter, P.K. Gregersen, J. Bregman, J.S. Sutcliffe, V. Jobanputra, W. Chung, D. Warburton, M.C. King, D. Skuse, D.H. Geschwind, T.C. Gilliam, K. Ye, and M. Wigler. 2007. Strong association of de novo copy number mutations with autism. *Science* **316:** 445-449.

Semple, R.K., V.C. Crowley, C.P. Sewter, M. Laudes, C. Christodoulides, R.V. Considine, A. Vidal-Puig, and S. O'Rahilly. 2004. Expression of the thermogenic nuclear hormone receptor coactivator PGC-1alpha is reduced in the adipose tissue of morbidly obese subjects. *Int J Obes Relat Metab Disord* **28:** 176-179.

Senee, V., C. Chelala, S. Duchatelet, D. Feng, H. Blanc, J.C. Cossec, C. Charon, M. Nicolino, P. Boileau, D.R. Cavener, P. Bougneres, D. Taha, and C. Julier. 2006. Mutations in GLIS3 are responsible for a rare syndrome with neonatal diabetes mellitus and congenital hypothyroidism. *Nat Genet* **38:** 682-687.

Seppala-Lindroos, A., S. Vehkavaara, A.M. Hakkinen, T. Goto, J. Westerbacka, A. Sovijarvi, J. Halavaara, and H. Yki-Jarvinen. 2002. Fat accumulation in the liver is associated with defects in insulin suppression of glucose production and serum free fatty acids independent of obesity in normal men. *J Clin Endocrinol Metab* **87:** 3023-3028.

Shackleton, S., D.J. Lloyd, S.N. Jackson, R. Evans, M.F. Niermeijer, B.M. Singh, H. Schmidt, G. Brabant, S. Kumar, P.N. Durrington, S. Gregory, S. O'Rahilly, and R.C. Trembath. 2000. LMNA, encoding lamin A/C, is mutated in partial lipodystrophy. *Nat Genet* **24:** 153-156.

Shah, O.J., Z. Wang, and T. Hunter. 2004. Inappropriate activation of the TSC/Rheb/mTOR/S6K cassette induces IRS1/2 depletion, insulin resistance, and cell survival deficiencies. *Curr Biol* **14:** 1650-1656.

Shield, J.P. 2000. Neonatal diabetes: new insights into aetiology and implications. *Horm Res* **53 Suppl 1:** 7-11.

Shimomura, I., R.E. Hammer, S. Ikemoto, M.S. Brown, and J.L. Goldstein. 1999. Leptin reverses insulin resistance and diabetes mellitus in mice with congenital lipodystrophy. *Nature* **401:** 73-76.

Shimomura, I., R.E. Hammer, J.A. Richardson, S. Ikemoto, Y. Bashmakov, J.L. Goldstein, and M.S. Brown. 1998. Insulin resistance and diabetes mellitus in transgenic mice expressing nuclear SREBP-1c in adipose tissue: model for congenital generalized lipodystrophy. *Genes Dev* **12:** 3182-3194.

Shiota, C., J.T. Woo, J. Lindner, K.D. Shelton, and M.A. Magnuson. 2006. Multiallelic disruption of the rictor gene in mice reveals that mTOR complex 2 is essential for fetal growth and viability. *Dev Cell* **11:** 583-589.

Silander, K., K.L. Mohlke, L.J. Scott, E.C. Peck, P. Hollstein, A.D. Skol, A.U. Jackson, P. Deloukas, S. Hunt, G. Stavrides, P.S. Chines, M.R. Erdos, N. Narisu, K.N. Conneely, C. Li, T.E. Fingerlin, S.K. Dhanjal, T.T. Valle, R.N. Bergman, J. Tuomilehto, R.M. Watanabe, M. Boehnke, and F.S. Collins. 2004. Genetic variation near the hepatocyte nuclear factor-4 alpha gene predicts susceptibility to type 2 diabetes. *Diabetes* **53:** 1141-1149.

Silva, J.P., M. Kohler, C. Graff, A. Oldfors, M.A. Magnuson, P.O. Berggren, and N.G. Larsson. 2000. Impaired insulin secretion and beta-cell loss in tissue-specific knockout mice with mitochondrial diabetes. *Nat Genet* **26:** 336-340.

Simmons, D., D.R. Williams, and M.J. Powell. 1991. The Coventry Diabetes Study: prevalence of diabetes and impaired glucose tolerance in Europids and Asians. *Q J Med* **81:** 1021-1030.

Simoneau, J.A. and D.E. Kelley. 1997. Altered glycolytic and oxidative capacities of skeletal muscle contribute to insulin resistance in NIDDM. *J Appl Physiol* **83:** 166-171.

Simoneau, J.A., J.H. Veerkamp, L.P. Turcotte, and D.E. Kelley. 1999. Markers of capacity to utilize fatty acids in human skeletal muscle: relation to insulin resistance and obesity and effects of weight loss. *Faseb J* **13:** 2051-2060.

Siniossoglou, S., H. Santos-Rosa, J. Rappsilber, M. Mann, and E. Hurt. 1998. A novel complex of membrane proteins required for formation of a spherical nucleus. *Embo J* **17:** 6449-6464.

Sladek, R., G. Rocheleau, J. Rung, C. Dina, L. Shen, D. Serre, P. Boutin, D. Vincent, A. Belisle, S. Hadjadj, B. Balkau, B. Heude, G. Charpentier, T.J. Hudson, A. Montpetit, A.V. Pshezhetsky, M. Prentki, B.I. Posner, D.J. Balding, D. Meyre, C. Polychronakos, and P. Froguel. 2007. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **445:** 881-885.

Smith, C.J., P.A. Crock, B.R. King, C.J. Meldrum, and R.J. Scott. 2004. Phenotype-genotype correlations in a series of wolfram syndrome families. *Diabetes Care* **27:** 2003-2009.

Smyth, D.J., J.D. Cooper, R. Bailey, S. Field, O. Burren, L.J. Smink, C. Guja, C. Ionescu-Tirgoviste, B. Widmer, D.B. Dunger, D.A. Savage, N.M. Walker, D.G. Clayton, and J.A. Todd. 2006. A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (IFIH1) region. *Nat Genet* **38:** 617-619.

Soejima, A., K. Inoue, D. Takai, M. Kaneko, H. Ishihara, Y. Oka, and J.I. Hayashi. 1996. Mitochondrial DNA is required for regulation of glucose-stimulated insulin secretion in a mouse pancreatic beta cell line, MIN6. *J Biol Chem* **271:** 26194-26199.

Song, Y., T. Niu, J.E. Manson, D.J. Kwiatkowski, and S. Liu. 2004. Are variants in the CAPN10 gene related to risk of type 2 diabetes? A quantitative assessment of population and family-based association studies. *Am J Hum Genet* **74:** 208-222.

Steemers, F.J. and K.L. Gunderson. 2007. Whole genome genotyping technologies on the BeadArray platform. *Biotechnol J* **2:** 41-49.

Steinthorsdottir, V., G. Thorleifsson, I. Reynisdottir, R. Benediktsson, T. Jonsdottir, G.B. Walters, U. Styrkarsdottir, S. Gretarsdottir, V. Emilsson, S. Ghosh, A. Baker, S. Snorradottir, H. Bjarnason, M.C. Ng, T. Hansen, Y. Bagger, R.L. Wilensky, M.P. Reilly, A. Adeyemo, Y. Chen, J. Zhou, V. Gudnason, G. Chen, H. Huang, K. Lashley, A. Doumatey, W.Y. So, R.C. Ma, G. Andersen, K. Borch-Johnsen, T. Jorgensen, J.V. van Vliet-Ostaptchouk, M.H. Hofker, C. Wijmenga, C. Christiansen, D.J. Rader, C. Rotimi, M. Gurney, J.C. Chan, O. Pedersen, G. Sigurdsson, J.R. Gulcher, U. Thorsteinsdottir, A. Kong, and K. Stefansson. 2007. A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat Genet* **39:** 770-775.

Stephens, T.W., M. Basinski, P.K. Bristow, J.M. Bue-Valleskey, S.G. Burgett, L. Craft, J. Hale, J. Hoffmann, H.M. Hsiung, A. Kriauciunas, and et al. 1995. The role of neuropeptide Y in the antiobesity action of the obese gene product. *Nature* **377:** 530-532.

Stoffers, D.A., J. Ferrer, W.L. Clarke, and J.F. Habener. 1997. Early-onset type-II diabetes mellitus (MODY4) linked to IPF1. *Nat Genet* **17:** 138-139.

Storlien, L.H., A.B. Jenkins, D.J. Chisholm, W.S. Pascoe, S. Khouri, and E.W. Kraegen. 1991. Influence of dietary fat composition on development of insulin resistance in rats. Relationship to muscle triglyceride and omega-3 fatty acids in muscle phospholipid. *Diabetes* **40:** 280-289.

Stoy, J., E.L. Edghill, S.E. Flanagan, H. Ye, V.P. Paz, A. Pluzhnikov, J.E. Below, M.G. Hayes, N.J. Cox, G.M. Lipkind, R.B. Lipton, S.A. Greeley, A.M. Patch, S. Ellard, D.F. Steiner, A.T. Hattersley, L.H. Philipson, and G.I. Bell. 2007. Insulin gene mutations as a cause of permanent neonatal diabetes. *Proc Natl Acad Sci U S A* **104:** 15040-15044.

Stranger, B.E., M.S. Forrest, M. Dunning, C.E. Ingle, C. Beazley, N. Thorne, R. Redon, C.P. Bird, A. de Grassi, C. Lee, C. Tyler-Smith, N. Carter, S.W. Scherer, S. Tavare, P. Deloukas, M.E. Hurles, and E.T. Dermitzakis. 2007. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315:** 848-853.

Strom, T.M., K. Hortnagel, S. Hofmann, F. Gekeler, C. Scharfe, W. Rabl, K.D. Gerbitz, and T. Meitinger. 1998. Diabetes insipidus, diabetes mellitus, optic atrophy and deafness (DIDMOAD) caused by mutations in a novel gene (wolframin) coding for a predicted transmembrane protein. *Hum Mol Genet* **7:** 2021-2028.

Sturgess, N.C., R.Z. Kozlowski, C.A. Carrington, C.N. Hales, and M.L. Ashford. 1988. Effects of sulphonylureas and diazoxide on insulin secretion and nucleotide-sensitive channels in an insulin-secreting cell line. *Br J Pharmacol* **95:** 83-94.

Stuurman, N., S. Heins, and U. Aebi. 1998. Nuclear lamins: their structure, assembly, and interactions. *J Struct Biol* **122:** 42-66.

Suviolahti, E., K. Reue, R.M. Cantor, J. Phan, M. Gentile, J. Naukkarinen, A. Soro-Paavonen, L. Oksanen, J. Kaprio, A. Rissanen, V. Salomaa, K. Kontula, M.R. Taskinen, P. Pajukanta, and L. Peltonen. 2006. Cross-species analyses implicate Lipin 1 involvement in human glucose metabolism. *Hum Mol Genet* **15:** 377-386.

Suzuki, K. and T. Kono. 1980. Evidence that insulin causes translocation of glucose transport activity to the plasma membrane from an intracellular storage site. *Proc Natl Acad Sci U S A* **77:** 2542-2545.

Syddall, H.E., A. Aihie Sayer, E.M. Dennison, H.J. Martin, D.J. Barker, and C. Cooper. 2005. Cohort profile: the Hertfordshire cohort study. *Int J Epidemiol* **34:** 1234-1242.

Szendroedi, J., A.I. Schmid, M. Chmelik, C. Toth, A. Brehm, M. Krssak, P. Nowotny, M. Wolzt, W. Waldhausl, and M. Roden. 2007. Muscle mitochondrial ATP synthesis and glucose transport/phosphorylation in type 2 diabetes. *PLoS Med* **4:** e154.

Szymanski, K.M., D. Binns, R. Bartz, N.V. Grishin, W.P. Li, A.K. Agarwal, A. Garg, R.G. Anderson, and J.M. Goodman. 2007. The lipodystrophy protein seipin is found at endoplasmic reticulum lipid droplet junctions and is important for droplet morphology. *Proc Natl Acad Sci U S A* **104:** 20890-20895.

Takano, A., I. Usui, T. Haruta, J. Kawahara, T. Uno, M. Iwata, and M. Kobayashi. 2001. Mammalian target of rapamycin pathway regulates insulin signaling via subcellular redistribution of insulin receptor substrate 1 and integrates nutritional signals and metabolic signals of insulin. *Mol Cell Biol* **21:** 5050-5062.

Takeda, K., H. Inoue, Y. Tanizawa, Y. Matsuzaki, J. Oba, Y. Watanabe, K. Shinoda, and Y. Oka. 2001. WFS1 (Wolfram syndrome 1) gene product: predominant subcellular localization to endoplasmic reticulum in cultured cells and neuronal expression in rat brain. *Hum Mol Genet* **10:** 477-484.

Takei, D., H. Ishihara, S. Yamaguchi, T. Yamada, A. Tamura, H. Katagiri, Y. Maruyama, and Y. Oka. 2006. WFS1 protein modulates the free Ca(2+) concentration in the endoplasmic reticulum. *FEBS Lett* **580:** 5635-5640.

Tamemoto, H., T. Kadowaki, K. Tobe, T. Yagi, H. Sakura, T. Hayakawa, Y. Terauchi, K. Ueki, Y. Kaburagi, S. Satoh, and et al. 1994. Insulin resistance and growth retardation in mice lacking insulin receptor substrate-1. *Nature* **372:** 182-186.

Tan, K., W.A. Kimber, J. Luan, M.A. Soos, R.K. Semple, N.J. Wareham, S. O'Rahilly, and I. Barroso. 2007. Analysis of genetic variation in Akt2/PKB-beta in severe

insulin resistance, lipodystrophy, type 2 diabetes, and related metabolic phenotypes. *Diabetes* **56:** 714-719.

Tange, Y., A. Hirata, and O. Niwa. 2002. An evolutionarily conserved fission yeast protein, Ned1, implicated in normal nuclear morphology and chromosome stability, interacts with Dis3, Pim1/RCC1 and an essential nucleoporin. *J Cell Sci* **115:** 4375-4385.

Taniguchi, C.M., K. Ueki, and R. Kahn. 2005. Complementary roles of IRS-1 and IRS-2 in the hepatic regulation of metabolism. *J Clin Invest* **115:** 718-727.

Tartaglia, L.A., M. Dembski, X. Weng, N. Deng, J. Culpepper, R. Devos, G.J. Richards, L.A. Campfield, F.T. Clark, J. Deeds, C. Muir, S. Sanker, A. Moriarty, K.J. Moore, J.S. Smutko, G.G. Mays, E.A. Wool, C.A. Monroe, and R.I. Tepper. 1995. Identification and expression cloning of a leptin receptor, OB-R. *Cell* **83:** 1263-1271.

Tattersall, R.B. and S.S. Fajans. 1975. A difference between the inheritance of classical juvenile-onset and maturity-onset type diabetes of young people. *Diabetes* **24:** 44-53.

Tee, A.R., B.D. Manning, P.P. Roux, L.C. Cantley, and J. Blenis. 2003. Tuberous sclerosis complex gene products, Tuberin and Hamartin, control mTOR signaling by acting as a GTPase-activating protein complex toward Rheb. *Curr Biol* **13:** 1259-1268.

Temple, I.K. and J.P. Shield. 2002. Transient neonatal diabetes, a disorder of imprinting. *J Med Genet* **39:** 872-875.

Terauchi, Y., Y. Tsuji, S. Satoh, H. Minoura, K. Murakami, A. Okuno, K. Inukai, T. Asano, Y. Kaburagi, K. Ueki, H. Nakajima, T. Hanafusa, Y. Matsuzawa, H. Sekihara, Y. Yin, J.C. Barrett, H. Oda, T. Ishikawa, Y. Akanuma, I. Komuro, M. Suzuki, K. Yamamura, T. Kodama, H. Suzuki, K. Yamamura, T. Kodama, H. Suzuki, S. Koyasu, S. Aizawa, K. Tobe, Y. Fukui, Y. Yazaki, and T. Kadowaki. 1999. Increased insulin sensitivity and hypoglycaemia in mice lacking the p85 alpha subunit of phosphoinositide 3-kinase. *Nat Genet* **21:** 230-235.

Tessa, A., I. Carbone, M.C. Matteoli, C. Bruno, C. Patrono, I.P. Patera, F. De Luca, R. Lorini, and F.M. Santorelli. 2001. Identification of novel WFS1 mutations in Italian children with Wolfram syndrome. *Hum Mutat* **17:** 348-349.

Thomas, P., Y. Ye, and E. Lightner. 1996. Mutation of the pancreatic islet inward rectifier Kir6.2 also leads to familial persistent hyperinsulinemic hypoglycemia of infancy. *Hum Mol Genet* **5:** 1809-1812.

Thong, F.S., P.J. Bilan, and A. Klip. 2007. The Rab GTPase-activating protein AS160 integrates Akt, protein kinase C, and AMP-activated protein kinase signals regulating GLUT4 traffic. *Diabetes* **56:** 414-423.

Tremblay, F., A. Gagnon, A. Veilleux, A. Sorisky, and A. Marette. 2005. Activation of the mammalian target of rapamycin pathway acutely inhibits insulin signaling to Akt and glucose transport in 3T3-L1 and human adipocytes. *Endocrinology* **146:** 1328-1337.

Tremblay, F. and A. Marette. 2001. Amino acid and insulin signaling via the mTOR/p70 S6 kinase pathway. A negative feedback mechanism leading to insulin resistance in skeletal muscle cells. *J Biol Chem* **276:** 38052-38060.

Tripathy, D., E. Lindholm, B. Isomaa, C. Saloranta, T. Tuomi, and L. Groop. 2003. Familiality of metabolic abnormalities is dependent on age at onset and phenotype of the type 2 diabetic proband. *Am J Physiol Endocrinol Metab* **285:** E1297-1303.

Tsai, H.J., G. Sun, D.E. Weeks, R. Kaushal, M. Wolujewicz, S.T. McGarvey, J. Tufa, S. Viali, and R. Deka. 2001. Type 2 diabetes and three calpain-10 gene

polymorphisms in Samoans: no evidence of association. *Am J Hum Genet* **69:** 1236-1244.

Tsuchiya, T., P.E. Schwarz, L.D. Bosque-Plata, M. Geoffrey Hayes, C. Dina, P. Froguel, G. Wayne Towers, S. Fischer, T. Temelkova-Kurktschiev, H. Rietzsch, J. Graessler, J. Vcelak, D. Palyzova, T. Selisko, B. Bendlova, J. Schulze, U. Julius, M. Hanefeld, M.N. Weedon, J.C. Evans, T.M. Frayling, A.T. Hattersley, M. Orho-Melander, L. Groop, M.T. Malecki, T. Hansen, O. Pedersen, T.E. Fingerlin, M. Boehnke, C.L. Hanis, N.J. Cox, and G.I. Bell. 2006. Association of the calpain-10 gene with type 2 diabetes in Europeans: results of pooled and meta-analyses. *Mol Genet Metab* **89:** 174-184.

Turner, M.D., F.K. Fulcher, C.V. Jones, B.T. Smith, E. Aganna, C.J. Partridge, G.A. Hitman, A. Clark, and Y.M. Patel. 2007. Calpain facilitates actin reorganization during glucose-stimulated insulin secretion. *Biochem Biophys Res Commun* **352:** 650-655.

Tyrer, J., P.D. Pharoah, and D.F. Easton. 2006. The admixture maximum likelihood test: a novel experiment-wise test of association between disease and multiple SNPs. *Genet Epidemiol* **30:** 636-643.

Ueda, K., J. Kawano, K. Takeda, T. Yujiri, K. Tanabe, T. Anno, M. Akiyama, J. Nozaki, T. Yoshinaga, A. Koizumi, K. Shinoda, Y. Oka, and Y. Tanizawa. 2005. Endoplasmic reticulum stress induces Wfs1 gene expression in pancreatic beta-cells via transcriptional activation. *Eur J Endocrinol* **153:** 167-176.

Ueki, K., C.M. Yballe, S.M. Brachmann, D. Vicent, J.M. Watt, C.R. Kahn, and L.C. Cantley. 2002. Increased insulin sensitivity in mice lacking p85beta subunit of phosphoinositide 3-kinase. *Proc Natl Acad Sci U S A* **99:** 419-424.

Ukropcova, B., O. Sereda, L. de Jonge, I. Bogacka, T. Nguyen, H. Xie, G.A. Bray, and S.R. Smith. 2007. Family history of diabetes links impaired substrate switching and reduced mitochondrial content in skeletal muscle. *Diabetes* **56:** 720-727.

Um, S.H., F. Frigerio, M. Watanabe, F. Picard, M. Joaquin, M. Sticker, S. Fumagalli, P.R. Allegrini, S.C. Kozma, J. Auwerx, and G. Thomas. 2004. Absence of S6K1 protects against age- and diet-induced obesity while enhancing insulin sensitivity. *Nature* **431:** 200-205.

Unoki, H., A. Takahashi, T. Kawaguchi, K. Hara, M. Horikoshi, G. Andersen, D.P. Ng, J. Holmkvist, K. Borch-Johnsen, T. Jorgensen, A. Sandbaek, T. Lauritzen, T. Hansen, S. Nurbaya, T. Tsunoda, M. Kubo, T. Babazono, H. Hirose, M. Hayashi, Y. Iwamoto, A. Kashiwagi, K. Kaku, R. Kawamori, E.S. Tai, O. Pedersen, N. Kamatani, T. Kadowaki, R. Kikkawa, Y. Nakamura, and S. Maeda. 2008. SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations. *Nat Genet*.

van den Ouweland, J.M., H.H. Lemkes, W. Ruitenbeek, L.A. Sandkuijl, M.F. de Vijlder, P.A. Struyvenberg, J.J. van de Kamp, and J.A. Maassen. 1992. Mutation in mitochondrial tRNA(Leu)(UUR) gene in a large pedigree with maternally transmitted type II diabetes mellitus and deafness. *Nat Genet* **1:** 368-371.

van Harmelen, V., M. Ryden, E. Sjolin, and J. Hoffstedt. 2007. A role of lipin in human obesity and insulin resistance: relation to adipocyte glucose transport and GLUT4 expression. *J Lipid Res* **48:** 201-206.

Varma, S., A. Shrivastav, S. Changela, and R.L. Khandelwal. 2008. Long-term effects of rapamycin treatment on insulin mediated phosphorylation of Akt/PKB and glycogen synthase activity. *Exp Cell Res* **314:** 1281-1291.

Vaxillaire, M., V. Boccio, A. Philippi, C. Vigouroux, J. Terwilliger, P. Passa, J.S. Beckmann, G. Velho, G.M. Lathrop, and P. Froguel. 1995. A gene for maturity onset diabetes of the young (MODY) maps to chromosome 12q. *Nat Genet* **9:** 418-423.

Venter, J.C. M.D. Adams E.W. Myers P.W. Li R.J. Mural G.G. Sutton H.O. Smith M. Yandell C.A. Evans R.A. Holt J.D. Gocayne P. Amanatides R.M. Ballew D.H. Huson J.R. Wortman Q. Zhang C.D. Kodira X.H. Zheng L. Chen M. Skupski G. Subramanian P.D. Thomas J. Zhang G.L. Gabor Miklos C. Nelson S. Broder A.G. Clark J. Nadeau V.A. McKusick N. Zinder A.J. Levine R.J. Roberts M. Simon C. Slayman M. Hunkapiller R. Bolanos A. Delcher I. Dew D. Fasulo M. Flanigan L. Florea A. Halpern S. Hannenhalli S. Kravitz S. Levy C. Mobarry K. Reinert K. Remington J. Abu-Threideh E. Beasley K. Biddick V. Bonazzi R. Brandon M. Cargill I. Chandramouliswaran R. Charlab K. Chaturvedi Z. Deng V. Di Francesco P. Dunn K. Eilbeck C. Evangelista A.E. Gabrielian W. Gan W. Ge F. Gong Z. Gu P. Guan T.J. Heiman M.E. Higgins R.R. Ji Z. Ke K.A. Ketchum Z. Lai Y. Lei Z. Li J. Li Y. Liang X. Lin F. Lu G.V. Merkulov N. Milshina H.M. Moore A.K. Naik V.A. Narayan B. Neelam D. Nusskern D.B. Rusch S. Salzberg W. Shao B. Shue J. Sun Z. Wang A. Wang X. Wang J. Wang M. Wei R. Wides C. Xiao C. Yan A. Yao J. Ye M. Zhan W. Zhang H. Zhang Q. Zhao L. Zheng F. Zhong W. Zhong S. Zhu S. Zhao D. Gilbert S. Baumhueter G. Spier C. Carter A. Cravchik T. Woodage F. Ali H. An A. Awe D. Baldwin H. Baden M. Barnstead I. Barrow K. Beeson D. Busam A. Carver A. Center M.L. Cheng L. Curry S. Danaher L. Davenport R. Desilets S. Dietz K. Dodson L. Doup S. Ferriera N. Garg A. Gluecksmann B. Hart J. Haynes C. Haynes C. Heiner S. Hladun D. Hostin J. Houck T. Howland C. Ibegwam J. Johnson F. Kalush L. Kline S. Koduru A. Love F. Mann D. May S. McCawley T. McIntosh I. McMullen M. Moy L. Moy B. Murphy K. Nelson C. Pfannkoch E. Pratts V. Puri H. Qureshi M. Reardon R. Rodriguez Y.H. Rogers D. Romblad B. Ruhfel R. Scott C. Sitter M. Smallwood E. Stewart R. Strong E. Suh R. Thomas N.N. Tint S. Tse C. Vech G. Wang J. Wetter S. Williams M. Williams S. Windsor E. Winn-Deen K. Wolfe J. Zaveri K. Zaveri J.F. Abril R. Guigo M.J. Campbell K.V. Sjolander B. Karlak A. Kejariwal H. Mi B. Lazareva T. Hatton A. Narechania K. Diemer A. Muruganujan N. Guo S. Sato V. Bafna S. Istrail R. Lippert R. Schwartz B. Walenz S. Yooseph D. Allen A. Basu J. Baxendale L. Blick M. Caminha J. Carnes-Stine P. Caulk Y.H. Chiang M. Coyne C. Dahlke A. Mays M. Dombroski M. Donnelly D. Ely S. Esparham C. Fosler H. Gire S. Glanowski K. Glasser A. Glodek M. Gorokhov K. Graham B. Gropman M. Harris J. Heil S. Henderson J. Hoover D. Jennings C. Jordan J. Jordan J. Kasha L. Kagan C. Kraft A. Levitsky M. Lewis X. Liu J. Lopez D. Ma W. Majoros J. McDaniel S. Murphy M. Newman T. Nguyen N. Nguyen M. Nodell S. Pan J. Peck M. Peterson W. Rowe R. Sanders J. Scott M. Simpson T. Smith A. Sprague T. Stockwell R. Turner E. Venter M. Wang M. Wen D. Wu M. Wu A. Xia A. Zandieh and X. Zhu. 2001. The sequence of the human genome. *Science* **291:** 1304-1351.

Vionnet, N., M. Stoffel, J. Takeda, K. Yasuda, G.I. Bell, H. Zouali, S. Lesage, G. Velho, F. Iris, P. Passa, and et al. 1992. Nonsense mutation in the glucokinase gene causes early-onset non-insulin-dependent diabetes mellitus. *Nature* **356:** 721-722.

Wahlen, K., E. Sjolin, and J. Hoffstedt. 2008. The common rs9939609 gene variant of the fat mass- and obesity-associated gene FTO is related to fat cell lipolysis. *J Lipid Res* **49:** 607-611.

Walder, K., L. Kerr-Bayles, A. Civitarese, J. Jowett, J. Curran, K. Elliott, J. Trevaskis, N. Bishara, P. Zimmet, L. Mandarino, E. Ravussin, J. Blangero, A. Kissebah, and G.R. Collier. 2005. The mitochondrial rhomboid protease PSARL is a new candidate gene for type 2 diabetes. *Diabetologia* **48:** 459-468.

Wallace, H.A., F. Marques-Kranc, M. Richardson, F. Luna-Crespo, J.A. Sharpe, J. Hughes, W.G. Wood, D.R. Higgs, and A.J. Smith. 2007. Manipulating the mouse

genome to engineer precise functional syntenic replacements with human sequence. *Cell* **128:** 197-209.

Walsh, T., J.M. McClellan, S.E. McCarthy, A.M. Addington, S.B. Pierce, G.M. Cooper, A.S. Nord, M. Kusenda, D. Malhotra, A. Bhandari, S.M. Stray, C.F. Rippey, P. Roccanova, V. Makarov, B. Lakshmi, R.L. Findling, L. Sikich, T. Stromberg, B. Merriman, N. Gogtay, P. Butler, K. Eckstrand, L. Noory, P. Gochman, R. Long, Z. Chen, S. Davis, C. Baker, E.E. Eichler, P.S. Meltzer, S.F. Nelson, A.B. Singleton, M.K. Lee, J.L. Rapoport, M.C. King, and J. Sebat. 2008. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* **320:** 539-543.

Wareham, N.J., C.D. Byrne, R. Williams, N.E. Day, and C.N. Hales. 1999. Fasting proinsulin concentrations predict the development of type 2 diabetes. *Diabetes Care* **22:** 262-270.

Wareham, N.J., S.J. Hennings, C.D. Byrne, C.N. Hales, A.M. Prentice, and N.E. Day. 1998. A quantitative analysis of the relationship between habitual energy expenditure, fitness and the metabolic cardiovascular syndrome. *Br J Nutr* **80:** 235-241.

Weedon, M., K. Owen, B. Shields, G. Hitman, M. Walker, M. McCarthy, L. Love-Gregory, M. Permutt, A. Hattersley, and T.M. Frayling. 2004. Common variants of the HNF4alpha P2 promoter are associated with type 2 diabetes in the UK population. *Diabetes* **53:** 3002-3006.

Weedon, M.N., V.J. Clark, Y. Qian, Y. Ben-Shlomo, N. Timpson, S. Ebrahim, D.A. Lawlor, M.E. Pembrey, S. Ring, T.J. Wilkin, L.D. Voss, A.N. Jeffery, B. Metcalf, L. Ferrucci, A.M. Corsi, A. Murray, D. Melzer, B. Knight, B. Shields, G.D. Smith, A.T. Hattersley, A. Di Rienzo, and T.M. Frayling. 2006. A common haplotype of the glucokinase gene alters fasting glucose and birth weight: association in six studies and population-genetics analyses. *Am J Hum Genet* **79:** 991-1001.

Weedon, M.N., P.E. Schwarz, Y. Horikawa, N. Iwasaki, T. Illig, R. Holle, W. Rathmann, T. Selisko, J. Schulze, K.R. Owen, J. Evans, L. Del Bosque-Plata, G. Hitman, M. Walker, J.C. Levy, M. Sampson, G.I. Bell, M.I. McCarthy, A.T. Hattersley, and T.M. Frayling. 2003. Meta-analysis and a large association study confirm a role for calpain-10 variation in type 2 diabetes susceptibility. *Am J Hum Genet* **73:** 1208-1212.

Weires, M.B., B. Tausch, P.J. Haug, C.Q. Edwards, T. Wetter, and L.A. Cannon-Albright. 2007. Familiality of diabetes mellitus. *Exp Clin Endocrinol Diabetes* **115:** 634-640.

Westermark, P. and E. Wilander. 1978. The influence of amyloid deposits on the islet volume in maturity onset diabetes mellitus. *Diabetologia* **15:** 417-421.

Whitman, M., C.P. Downes, M. Keeler, T. Keller, and L. Cantley. 1988. Type I phosphatidylinositol kinase makes a novel inositol phospholipid, phosphatidylinositol-3-phosphate. *Nature* **332:** 644-646.

WHO Study Group. 1999. Report of a WHO Consultation: Part 1: Diagnosis and Classification of Diabetes Mellitus. World Health Organisation, Geneva.

Wiedmann, S., M. Fischer, M. Koehler, K. Neureuther, G. Riegger, A. Doering, H. Schunkert, C. Hengstenberg, and A. Baessler. 2007. Genetic variants within the LPIN1 gene, encoding lipin, are influencing phenotypes of the metabolic syndrome in humans. *Diabetes*.

Wild, S., G. Roglic, A. Green, R. Sicree, and H. King. 2004. Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. *Diabetes Care* **27:** 1047-1053.

Willcocks, L.C., P.A. Lyons, M.R. Clatworthy, J.I. Robinson, W. Yang, S.A. Newland, V. Plagnol, N.N. McGovern, A.M. Condliffe, E.R. Chilvers, D. Adu, E.C. Jolly, R. Watts, Y.L. Lau, A.W. Morgan, G. Nash, and K.G. Smith. 2008. Copy number of

FCGR3B, which is associated with systemic lupus erythematosus, correlates with protein expression and immune complex uptake. *J Exp Med* **205:** 1573-1582.

Williams, D.R., N.J. Wareham, D.C. Brown, C.D. Byrne, P.M. Clark, B.D. Cox, L.J. Cox, N.E. Day, C.N. Hales, C.R. Palmer, and et al. 1995. Undiagnosed glucose intolerance in the community: the Isle of Ely Diabetes Project. *Diabet Med* **12:** 30-35.

Williams, R.C., J.C. Long, R.L. Hanson, M.L. Sievers, and W.C. Knowler. 2000. Individual estimates of European genetic admixture associated with lower body-mass index, plasma glucose, and prevalence of type 2 diabetes in Pima Indians. *Am J Hum Genet* **66:** 527-538.

Wiltshire, S., A.T. Hattersley, G.A. Hitman, M. Walker, J.C. Levy, M. Sampson, S. O'Rahilly, T.M. Frayling, J.I. Bell, G.M. Lathrop, A. Bennett, R. Dhillon, C. Fletcher, C.J. Groves, E. Jones, P. Prestwich, N. Simecek, P.V. Rao, M. Wishart, R. Foxon, G.F. Bottazzo, S. Howell, D. Smedley, L.R. Cardon, S. Menzel, and M.I. McCarthy. 2001. A genomewide scan for loci predisposing to type 2 diabetes in a U.K. population (the Diabetes UK Warren 2 Repository): analysis of 573 pedigrees provides independent replication of a susceptibility locus on chromosome 1q. *Am J Hum Genet* **69:** 553-569.

Winckler, W., M.N. Weedon, R.R. Graham, S.A. McCarroll, S. Purcell, P. Almgren, T. Tuomi, D. Gaudet, K.B. Bostrom, M. Walker, G. Hitman, A.T. Hattersley, M.I. McCarthy, K.G. Ardlie, J.N. Hirschhorn, M.J. Daly, T.M. Frayling, L. Groop, and D. Altshuler. 2007. Evaluation of common variants in the six known maturity-onset diabetes of the young (MODY) genes for association with type 2 diabetes. *Diabetes* **56:** 685-693.

Withers, D.J., D.J. Burks, H.H. Towery, S.L. Altamuro, C.L. Flint, and M.F. White. 1999. Irs-2 coordinates Igf-1 receptor-mediated beta-cell development and peripheral insulin signalling. *Nat Genet* **23:** 32-40.

Withers, D.J., J.S. Gutierrez, H. Towery, D.J. Burks, J.M. Ren, S. Previs, Y. Zhang, D. Bernal, S. Pons, G.I. Shulman, S. Bonner-Weir, and M.F. White. 1998. Disruption of IRS-2 causes type 2 diabetes in mice. *Nature* **391:** 900-904.

Wolfram DJ, W.H. 1938. Diabetes mellitus and simple optic atophy among siblings: report of four cases. *Mayo Clin. Proc.* **13:** 715-718.

World Health Organization:. 1999. Definitions, Diagnosis and Classification of Diabetes Mellitus and its Complications: Part 1: Diagnosis & Classification of Diabetes Mellitus. . World Health Organization, Geneva.

Worman, H.J. and G. Bonne. 2007. "Laminopathies": a wide spectrum of human diseases. *Exp Cell Res* **313:** 2121-2133.

Wullschleger, S., R. Loewith, and M.N. Hall. 2006. TOR signaling in growth and metabolism. *Cell* **124:** 471-484.

Xu, J., W.N. Lee, J. Phan, M.F. Saad, K. Reue, and I.J. Kurland. 2006. Lipin deficiency impairs diurnal metabolic fuel switching. *Diabetes* **55:** 3429-3438.

Yamada, E., S. Okada, T. Saito, K. Ohshima, M. Sato, T. Tsuchiya, Y. Uehara, H. Shimizu, and M. Mori. 2005. Akt2 phosphorylates Synip to regulate docking and fusion of GLUT4-containing vesicles. *J Cell Biol* **168:** 921-928.

Yamada, T., H. Ishihara, A. Tamura, R. Takahashi, S. Yamaguchi, D. Takei, A. Tokita, C. Satake, F. Tashiro, H. Katagiri, H. Aburatani, J. Miyazaki, and Y. Oka. 2006. WFS1-deficiency increases endoplasmic reticulum stress, impairs cell cycle progression and triggers the apoptotic pathway specifically in pancreatic beta-cells. *Hum Mol Genet* **15:** 1600-1609.

Yamagata, K., H. Furuta, N. Oda, P.J. Kaisaki, S. Menzel, N.J. Cox, S.S. Fajans, S. Signorini, M. Stoffel, and G.I. Bell. 1996a. Mutations in the hepatocyte nuclear

factor-4alpha gene in maturity-onset diabetes of the young (MODY1). *Nature* **384:** 458-460.

Yamagata, K., N. Oda, P.J. Kaisaki, S. Menzel, H. Furuta, M. Vaxillaire, L. Southam, R.D. Cox, G.M. Lathrop, V.V. Boriraj, X. Chen, N.J. Cox, Y. Oda, H. Yano, M.M. Le Beau, S. Yamada, H. Nishigori, J. Takeda, S.S. Fajans, A.T. Hattersley, N. Iwasaki, T. Hansen, O. Pedersen, K.S. Polonsky, G.I. Bell, and et al. 1996b. Mutations in the hepatocyte nuclear factor-1alpha gene in maturity-onset diabetes of the young (MODY3). *Nature* **384:** 455-458.

Yang, Q., K. Inoki, T. Ikenoue, and K.L. Guan. 2006. Identification of Sin1 as an essential TORC2 component required for complex formation and kinase activity. *Genes Dev* **20:** 2820-2832.

Yao-Borengasser, A., N. Rasouli, V. Varma, L.M. Miles, B. Phanavanh, T.N. Starks, J. Phan, H.J. Spencer, 3rd, R.E. McGehee, Jr., K. Reue, and P.A. Kern. 2006. Lipin expression is attenuated in adipose tissue of insulin-resistant human subjects and increases with peroxisome proliferator-activated receptor gamma activation. *Diabetes* **55:** 2811-2818.

Yasuda, K., K. Miyake, Y. Horikawa, K. Hara, H. Osawa, H. Furuta, Y. Hirota, H. Mori, A. Jonsson, Y. Sato, K. Yamagata, Y. Hinokio, H.Y. Wang, T. Tanahashi, N. Nakamura, Y. Oka, N. Iwasaki, Y. Iwamoto, Y. Yamada, Y. Seino, H. Maegawa, A. Kashiwagi, J. Takeda, E. Maeda, H.D. Shin, Y.M. Cho, K.S. Park, H.K. Lee, M.C. Ng, R.C. Ma, W.Y. So, J.C. Chan, V. Lyssenko, T. Tuomi, P. Nilsson, L. Groop, N. Kamatani, A. Sekine, Y. Nakamura, K. Yamamoto, T. Yoshida, K. Tokunaga, M. Itakura, H. Makino, K. Nanjo, T. Kadowaki, and M. Kasuga. 2008. Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. *Nat Genet*.

Yoon, K.H., S.H. Ko, J.H. Cho, J.M. Lee, Y.B. Ahn, K.H. Song, S.J. Yoo, M.I. Kang, B.Y. Cha, K.W. Lee, H.Y. Son, S.K. Kang, H.S. Kim, I.K. Lee, and S. Bonner-Weir. 2003. Selective beta-cell loss and alpha-cell expansion in patients with type 2 diabetes mellitus in Korea. *J Clin Endocrinol Metab* **88:** 2300-2308.

Yu, C.E., J. Oshima, Y.H. Fu, E.M. Wijsman, F. Hisama, R. Alisch, S. Matthews, J. Nakura, T. Miki, S. Ouais, G.M. Martin, J. Mulligan, and G.D. Schellenberg. 1996. Positional cloning of the Werner's syndrome gene. *Science* **272:** 258-262.

Yu, J., Y. Zhang, J. McIlroy, T. Rordorf-Nikolic, G.A. Orr, and J.M. Backer. 1998. Regulation of the p85/p110 phosphatidylinositol 3'-kinase: stabilization and inhibition of the p110alpha catalytic subunit by the p85 regulatory subunit. *Mol Cell Biol* **18:** 1379-1387.

Zeggini, E., L.J. Scott, R. Saxena, B.F. Voight, J.L. Marchini, T. Hu, P.I. de Bakker, G.R. Abecasis, P. Almgren, G. Andersen, K. Ardlie, K.B. Bostrom, R.N. Bergman, L.L. Bonnycastle, K. Borch-Johnsen, N.P. Burtt, H. Chen, P.S. Chines, M.J. Daly, P. Deodhar, C.J. Ding, A.S. Doney, W.L. Duren, K.S. Elliott, M.R. Erdos, T.M. Frayling, R.M. Freathy, L. Gianniny, H. Grallert, N. Grarup, C.J. Groves, C. Guiducci, T. Hansen, C. Herder, G.A. Hitman, T.E. Hughes, B. Isomaa, A.U. Jackson, T. Jorgensen, A. Kong, K. Kubalanza, F.G. Kuruvilla, J. Kuusisto, C. Langenberg, H. Lango, T. Lauritzen, Y. Li, C.M. Lindgren, V. Lyssenko, A.F. Marvelle, C. Meisinger, K. Midthjell, K.L. Mohlke, M.A. Morken, A.D. Morris, N. Narisu, P. Nilsson, K.R. Owen, C.N. Palmer, F. Payne, J.R. Perry, E. Pettersen, C. Platou, I. Prokopenko, L. Qi, L. Qin, N.W. Rayner, M. Rees, J.J. Roix, A. Sandbaek, B. Shields, M. Sjogren, V. Steinthorsdottir, H.M. Stringham, A.J. Swift, G. Thorleifsson, U. Thorsteinsdottir, N.J. Timpson, T. Tuomi, J. Tuomilehto, M. Walker, R.M. Watanabe, M.N. Weedon, C.J. Willer, T. Illig, K. Hveem, F.B. Hu, M. Laakso, K. Stefansson, O. Pedersen, N.J. Wareham, I. Barroso, A.T. Hattersley, F.S. Collins, L. Groop, M.I. McCarthy, M. Boehnke, and D. Altshuler.

2008. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet* **40:** 638-645.

Zhang, D., Z.X. Liu, C.S. Choi, L. Tian, R. Kibbey, J. Dong, G.W. Cline, P.A. Wood, and G.I. Shulman. 2007. Mitochondrial dysfunction due to long-chain Acyl-CoA dehydrogenase deficiency causes hepatic steatosis and hepatic insulin resistance. *Proc Natl Acad Sci U S A* **104:** 17075-17080.

Zhang, P., B. McGrath, S. Li, A. Frank, F. Zambito, J. Reinert, M. Gannon, K. Ma, K. McNaughton, and D.R. Cavener. 2002. The PERK eukaryotic initiation factor 2 alpha kinase is required for the development of the skeletal system, postnatal growth, and the function and viability of the pancreas. *Mol Cell Biol* **22:** 3864-3874.

Zhang, P., L. O'Loughlin, D.N. Brindley, and K. Reue. 2008. Regulation of lipin-1 gene expression by glucocorticoids during adipogenesis. *J Lipid Res*.

Zhang, W., S. Patil, B. Chauhan, S. Guo, D.R. Powell, J. Le, A. Klotsas, R. Matika, X. Xiao, R. Franks, K.A. Heidenreich, M.P. Sajan, R.V. Farese, D.B. Stolz, P. Tso, S.H. Koo, M. Montminy, and T.G. Unterman. 2006. FoxO1 regulates multiple metabolic pathways in the liver: effects on gluconeogenic, glycolytic, and lipogenic gene expression. *J Biol Chem* **281:** 10105-10117.

Zhang, Y., R. Proenca, M. Maffei, M. Barone, L. Leopold, and J.M. Friedman. 1994. Positional cloning of the mouse obese gene and its human homologue. *Nature* **372:** 425-432.

Zick, Y. 2001. Insulin resistance: a phosphorylation-based uncoupling of insulin signaling. *Trends Cell Biol* **11:** 437-441.

Zung, A., B. Glaser, R. Nimri, and Z. Zadik. 2004. Glibenclamide treatment in permanent neonatal diabetes mellitus due to an activating mutation in Kir6.2. *J Clin Endocrinol Metab* **89:** 5504-5507.

Appendix

Table A1 Demographic and clinical description of SIR patients

	Sex	Year of birth	Ethnic origin	Diagnosis	BMI	A/N	C/H	M/H	L/A
1	F	1967	Asian	HAIR-AN		Y	Y	N	N
2	F	1962	Asian	IR-DM	28.8	N	Y	N	N
3	F	1951	European	PA		Y	Y	Y	N
4	F	1964	European	Type A	24.1	Y	Y	N	N
5	F	1974	European	Type A	30.2	Y	Y	N	N
6	F	1956	European	IR-DM	34.9	N	N	N	N
7	F	1951	European	IR-DM	29.4	N	N	N	N
8	F	1950	European	HAIR-AN?	36.9	Y	Y	N	N
9	F	1924	European	Lipodystrophy	28.2	N	N	Y	Y
10	F	1939	European	HAIR-AN?		Y	Y	N	N
11	F	1972	Asian	Type A?		Y			
12	M	1975	European	MIR-AN	26.7	Y	N	N	N
13	M	1975	Asian	PA	46.3	Y	N		
14	F	1970	European	Type A					
15	F	1973	Afro-Carribean	HAIR-AN?	47.5	Y	Y	N	N
16	F	1976	European	Type A	31.7	Y	Y	N	
17	F	1960	European	HAIR-AN	31.6	Y	Y	N	N
18	F	1968		Lipodystrophy	24.8	Y	Y	Y	Y
19	F	1948	Afro-Carribean	Lipodystrophy	25	Y	N	Y	Y
20	F	1958	Iranian	HAIR-AN	34.5	Y	Y	N	N
21	F	1974	Asian	Type A	14.9	N	Y	N	Y
22	F	1975	Asian	HAIR-AN	33.5	Y	Y	N	N
23	F	1958	European	Type A		Y	Y	N	N
24	M			Type A					
25	F	1982	Somali	Type A	15.8	Y	N	N	Y?

Appendix

	Sex	Year of birth	Ethnic origin	Diagnosis	BMI	A/N	C/H	M/H	L/A
26	F	1963	European	HAIR-AN	41.2	Y	Y	N	N
27		1976	Cypriot	MIR-AN		Y			
28	F	1980	European	InsRes/short stature		Y			
29	F	1969	European	HAIR-AN	37	Y		N	N
30	F	1959		HAIR-AN		Y	Y		
31	F	1974	Asian	HAIR-AN?		Y	Y	N	N
32	F	1975	Asian	HAIR-AN	30.2	Y	Y	N	N
33	M			MIR-AN					
34	F	1947	European	Type A?		Y?	Y	N	N
35	F	1971	European	PA?	37.3	N	Y?	Y?	N
36	F	1991	European	Leprechaunism?					
37	F	1961	Mixed	HAIR-AN?	47.2	Y	Y	N	N
38	F	1980	Mixed	Type A?	27.5	Y	N	N	
39	F	1959	European	HAIR-AN?	30.7	Y	Y	N	N
40	F	1977	Asian	HAIR-AN	31.5	Y	Y	?	N
41	F	1943	European	Partial lipodystrophy	27.1	N	N	N	Y
42	F	1956		Lipodystrophy	30.6	N	Y	N	N
43	F	1944	European	IR-DM	32.6	Y	Y	N	N
44	F	1965	European	HAIR-AN	41.8	Y	N	N	N
45	F	1975	Asian	HAIR-AN	46.7	Y	N	N	N
46	M	1976		PA	35.8	Y			
47	F	1973	Asian	HAIR-AN	39.7	Y	Y	N	N
48	M	1974	Mixed	MIR-AN	48.9	Y		N	N
49		1977	Asian						
50	F	1964	European	HAIR-AN	34.8	N	Y	N	N
51	F	1972	Somali	InsRes/short stature	17.3	Y	Y	N	N
52	F	1972		Type A?	39.9	Y			

	Sex	Year of birth	Ethnic origin	Diagnosis	BMI	A/N	C/H	M/H	L/A
53	F	1976	European	HAIR-AN	37	Y	Y		N
54	F	1975	European	HAIR-AN	42.5	Y	N	N	N
55	M	1972	European	IR-DM	23	N			
56	F	1974		Lipodystrophy	21.3	N	N	N	Y
57	F	1958	European	HAIR-AN					
58	F	1977	European	HAIR-AN	31.5	Y	Y	Y	
59	F	1957	European	HAIR-AN	35	Y	Y	N	
60	F	1976	Asian	HAIR-AN	>47.4	N	Y		N
61	M	1990	European						
62	F	1954	Asian	HAIR-AN	36.3	Y		N	N
63	F	1979		Type II CMT	27	Y			
64	F	1959	Asian	Type A	22.5	Y	N		N
65	F	1953	European	HAIR-AN	38.5	N	Y	N	N
66	F	1982	European	PA?	28.2	Y	N	N	N
67	F	1975	Asian?	Type A		Y	Y	N	N
68	F	1978		Nerve deafness & diabetes					
69				Werner Syndrome					
70	M	1974	European		23.6	Y	N	Y	
71	F	1982	European	Type A?	33.4	Y	N	N	N
72	F	1974	European	HAIR-AN	49.6	Y	Y	N	N
73	M	1970	European		28.4	N		N	N
74	F	1954	European	HAIR-AN		Y	Y	N	N
75									
76	F	1980	Mixed	PA	30.7	N	N	N	N
77	F	1976	Asian	InsRes/short stature	28	Y	N	N	N
78	M	1982	Mauritius	MIR-AN	29	Y	N	N	N
79	F		Arabic	PA	17.3	Y	Y	Y	
80	F		Arabic	PA		Y	N	N	N

	Sex	Year of birth	Ethnic origin	Diagnosis	BMI	A/N	C/H	M/H	L/A
81	F	1973		Type A?	37.2	Y			
82	M	1985	European	MIR-AN	29.4	Y		N	N
83	F	1939	European	HAIR-AN	32.1	Y	N	Y	N
84	F	1968	European	IR-DM	33.1	Y	Y		Y?
85				HAIR-AN					
86	F			HAIR-AN		Y			
87	M	1982		InsRes/short stature		Y			
88	F	1976			37	Y			
89	F	1986	European		35.7	Y		N	N
90	F	1985	European	HAIR-AN	41.7	Y	N	N	N
91	M	1977	European	MOPDII (Osteodyplastic primordial dwarfism of Majewski type 2)	29	Y		N	N
92	F			OB-IR		Y			
93	F	1936	European	HAIR-AN	45.6	Y	Y		N
94	F	1980			30.1	N	Y	N	Y
95	F	1967	European	PA	38.9	N			
96	F	1964		HAIR-AN	35.8	Y	Y	N	N
97		1984	Asian	Type A	19.7	Y	N	Y	Y
98	M	1943							
99	F		European	OB-IR	34.6	Y	N	N	N
100	F	1988	European	InsRes/short stature		Y	N	N	Y?
101	F	1905	Jewish	PA	25.5	Y	N		N
102				Leprechaunism					
103	F	1988	European		21.1	Y	N	N	N
104				IR-DM (Type 1)					
105	F	1968	European	IR-DM	22.7	Y	Y	N	N
106	M	1960		MIR-AN					
107	M		European						
108	F	1959	European	HAIR-AN	42.3	Y	Y	Y	N

	Sex	Date of birth	Ethnic origin	Diagnosis	BMI	A/N	C/H	M/H	L/A
109	F	1974	European	OB-IR	39.6	N		N	N
110	F	1986	European		40.1	Y	?	N	?
111	F	1983	Irish	Type A	17.5	Y		N	N
112	M	1951		IR-DM	49.3				
113	F	1969		OB-IR	51.9	Y	N	N	N
114	F	1987	European	HAIR-AN		Y	Y	N	N
115	F	1989	European	Hemihypertrophy					
116	F	1983	European	HAIR-AN	31	Y	Y	N	N
117	F	1983	European		23.6	N	Y	N	N
118	M	1981		OB-IR	47.6	Y			
119	F	1959	European	PA? (originally Type A)	47.3	Y			
120	F	1987	European	HAIR-AN	32	Y	N	N	N
121	F	1956		Lipodystrophy					
122	F	1994	European	RM		Y	N	N	N
123	F	1973	Asian (Pakistani)		25.2	N	N	N	N
124	F	1990	European	Type A?	16	Y	Y	N	N
125	F	1980		Lipodystrophy (partial)	28				
126	F	1983		Lipodystrophy (partial)		Y			
127	M	1963	European	Lipodystrophy (partial)					
128	F	1940		Lipodystrophy (partial)					
129	F	1976		Lipodystrophy	29.2	Y?			Y
130	F	1991		Type A					
131	F	1989	Asian	HAIR-AN	30.3	Y	Y	N	N
132	F	1970	Brazilian	IR-DM	29				
133	M	1990		Leprechaunism					
134	F	1997	Asian (Indonesian)						
135	F	1956	From Azores	Lipodystrophy (partial)					
136	F	1983	European	Type A	24.5	Y	Y		N

	Sex	Year of birth	Ethnic origin	Diagnosis	BMI	A/N	C/H	M/H	L/A
137	F	1987	European	Type A	29	Y	N	?	N
138	F	1951	European	HAIR-AN	38.7	N	N	Y	Y
139	F	1989	Asian	Complex syndrome including severe insulin resistance		Y	Y	N	N
140	F	1990	European	Lipodystrophy (partial)?	23.9	Y mild		Y	Y partial
141	F	1977	Asian	HAIR-AN	36	Y	Y	N	N
142	F	1947	Cauc/Afro Caribbean	Lipodystrophy partial?	29.5	Y			Y
143	F	1983	European	HAIR-AN		Y			
144	F	1981	European	Lipodystrophy		Y	Y		Y partial
145	F	1975	European	HAIR-AN		Y	Y	N	N
146	F	1983		Lipodystrophy partial	22.9	Y			
147	M	1985		Lipodystrophy complex	18.1	Y			
148	M	1986	Asian	IR-DM	24.1	Y		Y	N
149	F	1937	European	Lipodystrophy partial	25.4	N	N	N	Y
150	F	1951	European	Lipodystrophy partial (peripheral)	45.1	?	N	?	Y
151	F	1976	European	Lipodystrophy partial	24.1	Y	Y	Y	Y
152	F	1975	European	Lipodystrophy partial	28.4	Y	Y	Y	
153	F	1981	Asian	Partial lipodystrophy	29.6	Y	Y	Y	Y
154	F	1979	Asian	Type A	20.4	Y	N	N	N
155	F	1984	European	Type A		Y	N	N	N
156	M	1988	European	MIR-AN	21.9	Y		N	N
157	F	1989	Asian Bengali	Type A?	21.8	Y			
158	M	1975	European	MIR-AN	25.2	Y		N	N

A/N = acanthosis nigricans, C/H = clinical hyperandrogenism, M/H = muscular hypertrophy , L/A = lipoatrophy. HAIR-AN = hyperandrogenism, insulin resistance, and acanthosis nigricans. IR-DM = insulin resistance and diabetes mellitus. PA = pseudoacromegaly. MIR-AN = male insulin resistance and acanthosis nigricans. OB-IR = obesity and insulin resistance. RM = Rabson-Mendenhall.

Table A2 CEPH samples

SampleID	Sex	Family	Relationship	HapMap CEU?
NA06985	Female	1341-14	maternal grandmother	Yes
NA06993	Male	1341-13	maternal grandfather	Yes
NA06994	Male	1340-9	paternal grandfather	Yes
NA07000	Female	1340-10	paternal grandmother	Yes
NA07022	Male	1340-11	maternal grandfather	Yes
NA07034	Male	1341-11	paternal grandfather	Yes
NA07055	Female	1341-12	paternal grandmother	Yes
NA07056	Female	1340-12	maternal grandmother	Yes
NA07345	Female	1345-13	maternal grandmother	Yes
NA11881	Male	1347-14	maternal grandfather	Yes
NA11882	Female	1347-15	maternal grandmother	Yes
NA11992	Male	1362-13	paternal grandfather	Yes
NA11993	Female	1362-14	paternal grandmother	Yes
NA11994	Male	1362-15	maternal grandfather	Yes
NA11995	Female	1362-16	maternal grandmother	Yes
NA12003	Male	1420-9	paternal grandfather	Yes
NA12004	Female	1420-10	paternal grandmother	Yes
NA12005	Male	1420-11	maternal grandfather	Yes
NA12006	Female	1420-12	maternal grandmother	Yes
NA12043	Male	1346-11	paternal grandfather	Yes
NA12044	Female	1346-12	paternal grandmother	Yes
NA12144	Male	1334-10	paternal grandfather	Yes
NA12145	Female	1334-11	paternal grandmother	Yes
NA12146	Male	1334-12	maternal grandfather	Yes
NA12154	Male	1408-10	paternal grandfather	Yes
NA12155	Male	1408-12	maternal grandfather	Yes
NA12156	Female	1408-13	maternal grandmother	Yes
NA12236	Female	1408-11	paternal grandmother	Yes
NA12239	Female	1334-13	maternal grandmother	Yes
NA12248	Male	1416-11	paternal grandfather	Yes
NA12249	Female	1416-12	paternal grandmother	Yes
NA07007	Male	1331-12	paternal grandfather	No
NA07340	Female	1331-13	paternal grandmother	No
NA07016	Male	1331-14	maternal grandfather	No
NA07050	Female	1331-15	maternal grandmother	No
NA07049	Male	1333-11	paternal grandfather	No
NA07002	Female	1333-12	paternal grandmother	No
NA07017	Male	1333-13	maternal grandfather	No
NA07341	Female	1333-14	maternal grandmother	No
NA12045	Male	1346-13	maternal grandfather	No
NA11879	Male	1347-12	paternal grandfather	No
NA11880	Female	1347-13	paternal grandmother	No
NA12250	Male	1416-13	maternal grandfather	No
NA12251	Female	1416-14	maternal grandmother	No
NA11917	Male	1423-11	paternal grandfather	No
NA11918	Female	1423-12	paternal grandmother	No
NA11919	Male	1423-13	maternal grandfather	No
NA11920	Female	1423-14	maternal grandmother	No

Table A3 HGDP-CEPH Human Genome Diversity panel

Sample ID	Sex	Population	Geographic origin
HGDP01253	M	Mozabite	Algeria (Mzab)
HGDP01254	F	Mozabite	Algeria (Mzab)
HGDP01255	M	Mozabite	Algeria (Mzab)
HGDP01256	M	Mozabite	Algeria (Mzab)
HGDP01257	M	Mozabite	Algeria (Mzab)
HGDP01258	M	Mozabite	Algeria (Mzab)
HGDP01259	M	Mozabite	Algeria (Mzab)
HGDP01260	M	Mozabite	Algeria (Mzab)
HGDP01261	M	Mozabite	Algeria (Mzab)
HGDP01262	M	Mozabite	Algeria (Mzab)
HGDP01263	M	Mozabite	Algeria (Mzab)
HGDP01264	M	Mozabite	Algeria (Mzab)
HGDP01265	M	Mozabite	Algeria (Mzab)
HGDP01266	M	Mozabite	Algeria (Mzab)
HGDP01267	F	Mozabite	Algeria (Mzab)
HGDP01268	M	Mozabite	Algeria (Mzab)
HGDP01269	M	Mozabite	Algeria (Mzab)
HGDP01270	F	Mozabite	Algeria (Mzab)
HGDP01271	M	Mozabite	Algeria (Mzab)
HGDP01272	M	Mozabite	Algeria (Mzab)
HGDP01273	F	Mozabite	Algeria (Mzab)
HGDP01274	F	Mozabite	Algeria (Mzab)
HGDP01275	F	Mozabite	Algeria (Mzab)
HGDP01276	F	Mozabite	Algeria (Mzab)
HGDP01277	F	Mozabite	Algeria (Mzab)
HGDP01278	M	Mozabite	Algeria (Mzab)
HGDP01279	M	Mozabite	Algeria (Mzab)
HGDP01280	F	Mozabite	Algeria (Mzab)
HGDP01281	F	Mozabite	Algeria (Mzab)
HGDP01282	M	Mozabite	Algeria (Mzab)
HGDP00490	M	NAN Melanesian	Bougainville
HGDP00491	M	NAN Melanesian	Bougainville
HGDP00655	M	NAN Melanesian	Bougainville
HGDP00656	F	NAN Melanesian	Bougainville
HGDP00657	F	NAN Melanesian	Bougainville
HGDP00658	F	NAN Melanesian	Bougainville
HGDP00659	F	NAN Melanesian	Bougainville
HGDP00660	F	NAN Melanesian	Bougainville
HGDP00661	F	NAN Melanesian	Bougainville
HGDP00662	M	NAN Melanesian	Bougainville
HGDP00663	F	NAN Melanesian	Bougainville
HGDP00664	F	NAN Melanesian	Bougainville
HGDP00787	F	NAN Melanesian	Bougainville
HGDP00788	M	NAN Melanesian	Bougainville
HGDP00789	M	NAN Melanesian	Bougainville
HGDP00823	M	NAN Melanesian	Bougainville
HGDP00824	M	NAN Melanesian	Bougainville
HGDP00825	F	NAN Melanesian	Bougainville
HGDP00826	F	NAN Melanesian	Bougainville

Sample ID	Sex	Population	Geographic origin
HGDP00978	F	NAN Melanesian	Bougainville
HGDP00979	F	NAN Melanesian	Bougainville
HGDP01027	F	NAN Melanesian	Bougainville
HGDP00995	F	Karitiana	Brazil
HGDP00996	F	Karitiana	Brazil
HGDP00997	M	Karitiana	Brazil
HGDP00998	M	Karitiana	Brazil
HGDP00999	F	Karitiana	Brazil
HGDP01000	M	Karitiana	Brazil
HGDP01001	F	Karitiana	Brazil
HGDP01003	F	Karitiana	Brazil
HGDP01004	M	Karitiana	Brazil
HGDP01005	M	Karitiana	Brazil
HGDP01006	F	Karitiana	Brazil
HGDP01007	F	Karitiana	Brazil
HGDP01008	F	Karitiana	Brazil
HGDP01009	M	Karitiana	Brazil
HGDP01010	F	Karitiana	Brazil
HGDP01011	F	Karitiana	Brazil
HGDP01012	M	Karitiana	Brazil
HGDP01013	M	Karitiana	Brazil
HGDP01014	F	Karitiana	Brazil
HGDP01015	M	Karitiana	Brazil
HGDP01016	F	Karitiana	Brazil
HGDP01017	F	Karitiana	Brazil
HGDP01018	F	Karitiana	Brazil
HGDP01019	M	Karitiana	Brazil
HGDP00830	F	Surui	Brazil
HGDP00832	F	Surui	Brazil
HGDP00833	F	Surui	Brazil
HGDP00834	M	Surui	Brazil
HGDP00835	M	Surui	Brazil
HGDP00837	M	Surui	Brazil
HGDP00838	F	Surui	Brazil
HGDP00839	M	Surui	Brazil
HGDP00840	F	Surui	Brazil
HGDP00841	F	Surui	Brazil
HGDP00842	M	Surui	Brazil
HGDP00843	M	Surui	Brazil
HGDP00844	M	Surui	Brazil
HGDP00845	M	Surui	Brazil
HGDP00846	F	Surui	Brazil
HGDP00847	M	Surui	Brazil
HGDP00848	F	Surui	Brazil
HGDP00849	M	Surui	Brazil
HGDP00850	F	Surui	Brazil
HGDP00851	M	Surui	Brazil
HGDP00852	F	Surui	Brazil
HGDP00711	M	Cambodian	Cambodia
HGDP00712	F	Cambodian	Cambodia
HGDP00713	F	Cambodian	Cambodia

Sample ID	Sex	Population	Geographic origin
HGDP00714	M	Cambodian	Cambodia
HGDP00715	M	Cambodian	Cambodia
HGDP00716	M	Cambodian	Cambodia
HGDP00717	M	Cambodian	Cambodia
HGDP00718	F	Cambodian	Cambodia
HGDP00719	F	Cambodian	Cambodia
HGDP00720	M	Cambodian	Cambodia
HGDP00721	F	Cambodian	Cambodia
HGDP00448	M	Biaka Pygmies	Central African Republic
HGDP00461	M	Biaka Pygmies	Central African Republic
HGDP00464	M	Biaka Pygmies	Central African Republic
HGDP00465	M	Biaka Pygmies	Central African Republic
HGDP00466	M	Biaka Pygmies	Central African Republic
HGDP00469	M	Biaka Pygmies	Central African Republic
HGDP00470	M	Biaka Pygmies	Central African Republic
HGDP00472	M	Biaka Pygmies	Central African Republic
HGDP00473	M	Biaka Pygmies	Central African Republic
HGDP00475	M	Biaka Pygmies	Central African Republic
HGDP00477	M	Biaka Pygmies	Central African Republic
HGDP00451	M	Biaka Pygmies	Central African Republic
HGDP00452	M	Biaka Pygmies	Central African Republic
HGDP00453	M	Biaka Pygmies	Central African Republic
HGDP00454	M	Biaka Pygmies	Central African Republic
HGDP00455	M	Biaka Pygmies	Central African Republic
HGDP00457	M	Biaka Pygmies	Central African Republic
HGDP00458	M	Biaka Pygmies	Central African Republic
HGDP00459	M	Biaka Pygmies	Central African Republic
HGDP00460	M	Biaka Pygmies	Central African Republic
HGDP00479	M	Biaka Pygmies	Central African Republic
HGDP00980	F	Biaka Pygmies	Central African Republic
HGDP00981	M	Biaka Pygmies	Central African Republic
HGDP00985	M	Biaka Pygmies	Central African Republic
HGDP00986	M	Biaka Pygmies	Central African Republic
HGDP01084	F	Biaka Pygmies	Central African Republic
HGDP01085	F	Biaka Pygmies	Central African Republic
HGDP01086	M	Biaka Pygmies	Central African Republic
HGDP01087	M	Biaka Pygmies	Central African Republic
HGDP01088	M	Biaka Pygmies	Central African Republic
HGDP01089	M	Biaka Pygmies	Central African Republic
HGDP01090	M	Biaka Pygmies	Central African Republic
HGDP01091	M	Biaka Pygmies	Central African Republic
HGDP01092	M	Biaka Pygmies	Central African Republic
HGDP01093	M	Biaka Pygmies	Central African Republic
HGDP01094	M	Biaka Pygmies	Central African Republic
HGDP01307	M	Dai	China
HGDP01308	M	Dai	China
HGDP01309	M	Dai	China
HGDP01310	M	Dai	China
HGDP01311	M	Dai	China
HGDP01312	M	Dai	China
HGDP01313	M	Dai	China

Sample ID	Sex	Population	Geographic origin
HGDP01314	F	Dai	China
HGDP01315	F	Dai	China
HGDP01316	F	Dai	China
HGDP01213	M	Daur	China
HGDP01214	M	Daur	China
HGDP01215	F	Daur	China
HGDP01216	M	Daur	China
HGDP01217	M	Daur	China
HGDP01218	M	Daur	China
HGDP01219	F	Daur	China
HGDP01220	M	Daur	China
HGDP01221	M	Daur	China
HGDP01222	F	Daur	China
HGDP00774	M	Han	China
HGDP00775	M	Han	China
HGDP00776	F	Han	China
HGDP00777	M	Han	China
HGDP00778	M	Han	China
HGDP00779	M	Han	China
HGDP00780	M	Han	China
HGDP00781	F	Han	China
HGDP00782	M	Han	China
HGDP00783	F	Han	China
HGDP00784	F	Han	China
HGDP00785	M	Han	China
HGDP00786	M	Han	China
HGDP00811	F	Han	China
HGDP00812	F	Han	China
HGDP00813	F	Han	China
HGDP00814	F	Han	China
HGDP00815	M	Han	China
HGDP00817	F	Han	China
HGDP00818	F	Han	China
HGDP00819	M	Han	China
HGDP00820	F	Han	China
HGDP00821	M	Han	China
HGDP00822	M	Han	China
HGDP00971	M	Han	China
HGDP00972	F	Han	China
HGDP00973	M	Han	China
HGDP00974	F	Han	China
HGDP00975	F	Han	China
HGDP00976	F	Han	China
HGDP00977	M	Han	China
HGDP01021	F	Han	China
HGDP01022	F	Han	China
HGDP01023	F	Han	China
HGDP01024	F	Han	China
HGDP01287	F	Han	China
HGDP01288	M	Han	China
HGDP01289	M	Han	China

Sample ID	Sex	Population	Geographic origin
HGDP01290	M	Han	China
HGDP01291	F	Han	China
HGDP01292	M	Han	China
HGDP01293	M	Han	China
HGDP01294	M	Han	China
HGDP01295	M	Han	China
HGDP01296	M	Han	China
HGDP01233	M	Hezhen	China
HGDP01234	F	Hezhen	China
HGDP01235	M	Hezhen	China
HGDP01236	M	Hezhen	China
HGDP01237	M	Hezhen	China
HGDP01238	F	Hezhen	China
HGDP01239	F	Hezhen	China
HGDP01240	M	Hezhen	China
HGDP01241	M	Hezhen	China
HGDP01242	F	Hezhen	China
HGDP01317	M	Lahu	China
HGDP01318	M	Lahu	China
HGDP01319	M	Lahu	China
HGDP01320	M	Lahu	China
HGDP01321	M	Lahu	China
HGDP01322	M	Lahu	China
HGDP01323	F	Lahu	China
HGDP01324	F	Lahu	China
HGDP01325	F	Lahu	China
HGDP01326	M	Lahu	China
HGDP01189	M	Miaozu	China
HGDP01190	M	Miaozu	China
HGDP01191	M	Miaozu	China
HGDP01192	M	Miaozu	China
HGDP01193	M	Miaozu	China
HGDP01194	M	Miaozu	China
HGDP01195	M	Miaozu	China
HGDP01196	F	Miaozu	China
HGDP01197	F	Miaozu	China
HGDP01198	F	Miaozu	China
HGDP01223	F	Mongola	China
HGDP01224	M	Mongola	China
HGDP01225	M	Mongola	China
HGDP01226	M	Mongola	China
HGDP01227	M	Mongola	China
HGDP01228	M	Mongola	China
HGDP01229	M	Mongola	China
HGDP01230	M	Mongola	China
HGDP01231	F	Mongola	China
HGDP01232	F	Mongola	China
HGDP01337	M	Naxi	China
HGDP01338	M	Naxi	China
HGDP01339	M	Naxi	China
HGDP01340	M	Naxi	China

Sample ID	Sex	Population	Geographic origin
HGDP01341	M	Naxi	China
HGDP01342	M	Naxi	China
HGDP01343	M	Naxi	China
HGDP01344	M	Naxi	China
HGDP01345	F	Naxi	China
HGDP01346	F	Naxi	China
HGDP01203	M	Oroqen	China
HGDP01204	M	Oroqen	China
HGDP01205	M	Oroqen	China
HGDP01206	M	Oroqen	China
HGDP01207	M	Oroqen	China
HGDP01208	M	Oroqen	China
HGDP01209	F	Oroqen	China
HGDP01210	M	Oroqen	China
HGDP01211	F	Oroqen	China
HGDP01212	F	Oroqen	China
HGDP01327	M	She	China
HGDP01328	M	She	China
HGDP01329	M	She	China
HGDP01330	M	She	China
HGDP01331	M	She	China
HGDP01332	M	She	China
HGDP01333	M	She	China
HGDP01334	F	She	China
HGDP01335	F	She	China
HGDP01336	F	She	China
HGDP01347	M	Tu	China
HGDP01348	M	Tu	China
HGDP01349	M	Tu	China
HGDP01350	M	Tu	China
HGDP01351	M	Tu	China
HGDP01352	M	Tu	China
HGDP01353	M	Tu	China
HGDP01354	F	Tu	China
HGDP01355	F	Tu	China
HGDP01356	F	Tu	China
HGDP01095	M	Tujia	China
HGDP01096	M	Tujia	China
HGDP01097	M	Tujia	China
HGDP01098	F	Tujia	China
HGDP01099	M	Tujia	China
HGDP01100	M	Tujia	China
HGDP01101	M	Tujia	China
HGDP01102	M	Tujia	China
HGDP01103	M	Tujia	China
HGDP01104	M	Tujia	China
HGDP01297	M	Uygur	China
HGDP01298	M	Uygur	China
HGDP01299	M	Uygur	China
HGDP01300	M	Uygur	China
HGDP01301	M	Uygur	China

Sample ID	Sex	Population	Geographic origin
HGDP01302	M	Uygur	China
HGDP01303	M	Uygur	China
HGDP01304	M	Uygur	China
HGDP01305	F	Uygur	China
HGDP01306	F	Uygur	China
HGDP01243	M	Xibo	China
HGDP01244	M	Xibo	China
HGDP01245	M	Xibo	China
HGDP01246	M	Xibo	China
HGDP01247	M	Xibo	China
HGDP01248	M	Xibo	China
HGDP01249	M	Xibo	China
HGDP01250	M	Xibo	China
HGDP01251	F	Xibo	China
HGDP01179	M	Yizu	China
HGDP01180	M	Yizu	China
HGDP01181	M	Yizu	China
HGDP01182	M	Yizu	China
HGDP01183	M	Yizu	China
HGDP01184	M	Yizu	China
HGDP01185	M	Yizu	China
HGDP01186	M	Yizu	China
HGDP01187	M	Yizu	China
HGDP01188	F	Yizu	China
HGDP00702	F	Colombian	Colombia
HGDP00703	M	Colombian	Colombia
HGDP00704	F	Colombian	Colombia
HGDP00705	M	Colombian	Colombia
HGDP00706	F	Colombian	Colombia
HGDP00707	F	Colombian	Colombia
HGDP00708	F	Colombian	Colombia
HGDP00709	M	Colombian	Colombia
HGDP00710	M	Colombian	Colombia
HGDP00792	M	Colombian	Colombia
HGDP00793	F	Colombian	Colombia
HGDP00827	F	Colombian	Colombia
HGDP00970	F	Colombian	Colombia
HGDP00449	M	Mbuti Pygmies	Democratic Republic of Congo
HGDP00450	M	Mbuti Pygmies	Democratic Republic of Congo
HGDP00456	M	Mbuti Pygmies	Democratic Republic of Congo
HGDP00462	M	Mbuti Pygmies	Democratic Republic of Congo
HGDP00463	M	Mbuti Pygmies	Democratic Republic of Congo
HGDP00467	M	Mbuti Pygmies	Democratic Republic of Congo
HGDP00468	M	Mbuti Pygmies	Democratic Republic of Congo
HGDP00471	F	Mbuti Pygmies	Democratic Republic of Congo
HGDP00474	M	Mbuti Pygmies	Democratic Republic of Congo
HGDP00476	F	Mbuti Pygmies	Democratic Republic of Congo
HGDP00478	M	Mbuti Pygmies	Democratic Republic of Congo
HGDP00982	M	Mbuti Pygmies	Democratic Republic of Congo
HGDP00983	M	Mbuti Pygmies	Democratic Republic of Congo
HGDP00984	M	Mbuti Pygmies	Democratic Republic of Congo

Sample ID	Sex	Population	Geographic origin
HGDP01081	M	Mbuti Pygmies	Democratic Republic of Congo
HGDP00511	M	French	France
HGDP00512	M	French	France
HGDP00513	F	French	France
HGDP00514	F	French	France
HGDP00515	M	French	France
HGDP00516	F	French	France
HGDP00517	F	French	France
HGDP00518	M	French	France
HGDP00519	M	French	France
HGDP00520	F	French	France
HGDP00521	M	French	France
HGDP00522	M	French	France
HGDP00523	F	French	France
HGDP00524	F	French	France
HGDP00525	M	French	France
HGDP00526	F	French	France
HGDP00527	F	French	France
HGDP00528	M	French	France
HGDP00529	F	French	France
HGDP00530	M	French	France
HGDP00531	F	French	France
HGDP00532	F	French	France
HGDP00533	M	French	France
HGDP00534	F	French	France
HGDP00535	F	French	France
HGDP00536	F	French	France
HGDP00537	F	French	France
HGDP00538	M	French	France
HGDP00539	F	French	France
HGDP01357	M	French Basque	France
HGDP01358	M	French Basque	France
HGDP01359	M	French Basque	France
HGDP01360	M	French Basque	France
HGDP01361	M	French Basque	France
HGDP01362	M	French Basque	France
HGDP01363	F	French Basque	France
HGDP01364	M	French Basque	France
HGDP01365	F	French Basque	France
HGDP01366	F	French Basque	France
HGDP01367	F	French Basque	France
HGDP01368	F	French Basque	France
HGDP01369	F	French Basque	France
HGDP01370	M	French Basque	France
HGDP01371	M	French Basque	France
HGDP01372	M	French Basque	France
HGDP01373	F	French Basque	France
HGDP01374	M	French Basque	France
HGDP01375	M	French Basque	France
HGDP01376	M	French Basque	France
HGDP01377	M	French Basque	France

Sample ID	Sex	Population	Geographic origin
HGDP01378	M	French Basque	France
HGDP01379	M	French Basque	France
HGDP01380	F	French Basque	France
HGDP00557	F	Druze	Israel (Carmel)
HGDP00558	F	Druze	Israel (Carmel)
HGDP00559	F	Druze	Israel (Carmel)
HGDP00560	F	Druze	Israel (Carmel)
HGDP00561	F	Druze	Israel (Carmel)
HGDP00562	M	Druze	Israel (Carmel)
HGDP00563	F	Druze	Israel (Carmel)
HGDP00564	F	Druze	Israel (Carmel)
HGDP00565	F	Druze	Israel (Carmel)
HGDP00566	F	Druze	Israel (Carmel)
HGDP00567	F	Druze	Israel (Carmel)
HGDP00568	F	Druze	Israel (Carmel)
HGDP00569	F	Druze	Israel (Carmel)
HGDP00570	F	Druze	Israel (Carmel)
HGDP00571	F	Druze	Israel (Carmel)
HGDP00572	F	Druze	Israel (Carmel)
HGDP00573	F	Druze	Israel (Carmel)
HGDP00574	F	Druze	Israel (Carmel)
HGDP00575	F	Druze	Israel (Carmel)
HGDP00576	M	Druze	Israel (Carmel)
HGDP00577	F	Druze	Israel (Carmel)
HGDP00578	F	Druze	Israel (Carmel)
HGDP00579	F	Druze	Israel (Carmel)
HGDP00580	M	Druze	Israel (Carmel)
HGDP00581	F	Druze	Israel (Carmel)
HGDP00582	F	Druze	Israel (Carmel)
HGDP00583	F	Druze	Israel (Carmel)
HGDP00584	F	Druze	Israel (Carmel)
HGDP00585	F	Druze	Israel (Carmel)
HGDP00586	F	Druze	Israel (Carmel)
HGDP00587	F	Druze	Israel (Carmel)
HGDP00588	M	Druze	Israel (Carmel)
HGDP00589	F	Druze	Israel (Carmel)
HGDP00590	F	Druze	Israel (Carmel)
HGDP00591	F	Druze	Israel (Carmel)
HGDP00592	F	Druze	Israel (Carmel)
HGDP00594	M	Druze	Israel (Carmel)
HGDP00595	M	Druze	Israel (Carmel)
HGDP00597	M	Druze	Israel (Carmel)
HGDP00598	M	Druze	Israel (Carmel)
HGDP00599	M	Druze	Israel (Carmel)
HGDP00600	M	Druze	Israel (Carmel)
HGDP00601	F	Druze	Israel (Carmel)
HGDP00602	M	Druze	Israel (Carmel)
HGDP00603	M	Druze	Israel (Carmel)
HGDP00604	M	Druze	Israel (Carmel)
HGDP00605	M	Druze	Israel (Carmel)
HGDP00606	F	Druze	Israel (Carmel)

Sample ID	Sex	Population	Geographic origin
HGDP00675	M	Palestinian	Israel (Central)
HGDP00676	M	Palestinian	Israel (Central)
HGDP00677	M	Palestinian	Israel (Central)
HGDP00678	M	Palestinian	Israel (Central)
HGDP00679	F	Palestinian	Israel (Central)
HGDP00680	F	Palestinian	Israel (Central)
HGDP00681	F	Palestinian	Israel (Central)
HGDP00682	F	Palestinian	Israel (Central)
HGDP00683	F	Palestinian	Israel (Central)
HGDP00684	F	Palestinian	Israel (Central)
HGDP00685	F	Palestinian	Israel (Central)
HGDP00686	F	Palestinian	Israel (Central)
HGDP00687	F	Palestinian	Israel (Central)
HGDP00688	F	Palestinian	Israel (Central)
HGDP00689	F	Palestinian	Israel (Central)
HGDP00690	F	Palestinian	Israel (Central)
HGDP00691	F	Palestinian	Israel (Central)
HGDP00692	F	Palestinian	Israel (Central)
HGDP00693	F	Palestinian	Israel (Central)
HGDP00694	F	Palestinian	Israel (Central)
HGDP00695	F	Palestinian	Israel (Central)
HGDP00696	F	Palestinian	Israel (Central)
HGDP00697	F	Palestinian	Israel (Central)
HGDP00698	F	Palestinian	Israel (Central)
HGDP00699	F	Palestinian	Israel (Central)
HGDP00700	F	Palestinian	Israel (Central)
HGDP00722	M	Palestinian	Israel (Central)
HGDP00723	M	Palestinian	Israel (Central)
HGDP00724	M	Palestinian	Israel (Central)
HGDP00725	M	Palestinian	Israel (Central)
HGDP00726	M	Palestinian	Israel (Central)
HGDP00727	M	Palestinian	Israel (Central)
HGDP00728	M	Palestinian	Israel (Central)
HGDP00729	M	Palestinian	Israel (Central)
HGDP00730	M	Palestinian	Israel (Central)
HGDP00731	M	Palestinian	Israel (Central)
HGDP00732	M	Palestinian	Israel (Central)
HGDP00733	M	Palestinian	Israel (Central)
HGDP00734	M	Palestinian	Israel (Central)
HGDP00735	F	Palestinian	Israel (Central)
HGDP00736	F	Palestinian	Israel (Central)
HGDP00737	F	Palestinian	Israel (Central)
HGDP00738	F	Palestinian	Israel (Central)
HGDP00739	F	Palestinian	Israel (Central)
HGDP00740	F	Palestinian	Israel (Central)
HGDP00741	F	Palestinian	Israel (Central)
HGDP00742	F	Palestinian	Israel (Central)
HGDP00743	F	Palestinian	Israel (Central)
HGDP00744	F	Palestinian	Israel (Central)
HGDP00745	F	Palestinian	Israel (Central)
HGDP00746	F	Palestinian	Israel (Central)

Sample ID	Sex	Population	Geographic origin
HGDP00607	F	Bedouin	Israel (Negev)
HGDP00608	M	Bedouin	Israel (Negev)
HGDP00609	M	Bedouin	Israel (Negev)
HGDP00610	M	Bedouin	Israel (Negev)
HGDP00611	M	Bedouin	Israel (Negev)
HGDP00612	F	Bedouin	Israel (Negev)
HGDP00613	F	Bedouin	Israel (Negev)
HGDP00614	F	Bedouin	Israel (Negev)
HGDP00615	F	Bedouin	Israel (Negev)
HGDP00616	M	Bedouin	Israel (Negev)
HGDP00617	M	Bedouin	Israel (Negev)
HGDP00618	M	Bedouin	Israel (Negev)
HGDP00619	M	Bedouin	Israel (Negev)
HGDP00620	M	Bedouin	Israel (Negev)
HGDP00621	M	Bedouin	Israel (Negev)
HGDP00622	M	Bedouin	Israel (Negev)
HGDP00623	M	Bedouin	Israel (Negev)
HGDP00624	M	Bedouin	Israel (Negev)
HGDP00625	M	Bedouin	Israel (Negev)
HGDP00626	M	Bedouin	Israel (Negev)
HGDP00627	M	Bedouin	Israel (Negev)
HGDP00628	M	Bedouin	Israel (Negev)
HGDP00629	M	Bedouin	Israel (Negev)
HGDP00630	M	Bedouin	Israel (Negev)
HGDP00631	M	Bedouin	Israel (Negev)
HGDP00632	F	Bedouin	Israel (Negev)
HGDP00633	F	Bedouin	Israel (Negev)
HGDP00634	F	Bedouin	Israel (Negev)
HGDP00635	F	Bedouin	Israel (Negev)
HGDP00636	F	Bedouin	Israel (Negev)
HGDP00637	F	Bedouin	Israel (Negev)
HGDP00638	F	Bedouin	Israel (Negev)
HGDP00639	M	Bedouin	Israel (Negev)
HGDP00640	M	Bedouin	Israel (Negev)
HGDP00641	M	Bedouin	Israel (Negev)
HGDP00642	M	Bedouin	Israel (Negev)
HGDP00643	F	Bedouin	Israel (Negev)
HGDP00644	M	Bedouin	Israel (Negev)
HGDP00645	M	Bedouin	Israel (Negev)
HGDP00646	F	Bedouin	Israel (Negev)
HGDP00647	F	Bedouin	Israel (Negev)
HGDP00648	M	Bedouin	Israel (Negev)
HGDP00649	F	Bedouin	Israel (Negev)
HGDP00650	F	Bedouin	Israel (Negev)
HGDP00651	F	Bedouin	Israel (Negev)
HGDP00652	F	Bedouin	Israel (Negev)
HGDP00653	F	Bedouin	Israel (Negev)
HGDP00654	M	Bedouin	Israel (Negev)
HGDP00701	F	Bedouin	Israel (Negev)
HGDP00665	M	Sardinian	Italy
HGDP00666	M	Sardinian	Italy

Sample ID	Sex	Population	Geographic origin
HGDP00667	F	Sardinian	Italy
HGDP00668	M	Sardinian	Italy
HGDP00669	F	Sardinian	Italy
HGDP00670	M	Sardinian	Italy
HGDP00671	M	Sardinian	Italy
HGDP00672	F	Sardinian	Italy
HGDP00673	F	Sardinian	Italy
HGDP00674	M	Sardinian	Italy
HGDP01062	F	Sardinian	Italy
HGDP01063	M	Sardinian	Italy
HGDP01064	F	Sardinian	Italy
HGDP01065	F	Sardinian	Italy
HGDP01066	M	Sardinian	Italy
HGDP01067	M	Sardinian	Italy
HGDP01068	F	Sardinian	Italy
HGDP01069	M	Sardinian	Italy
HGDP01070	F	Sardinian	Italy
HGDP01071	M	Sardinian	Italy
HGDP01072	F	Sardinian	Italy
HGDP01073	M	Sardinian	Italy
HGDP01074	F	Sardinian	Italy
HGDP01075	M	Sardinian	Italy
HGDP01076	M	Sardinian	Italy
HGDP01077	M	Sardinian	Italy
HGDP01078	F	Sardinian	Italy
HGDP01079	M	Sardinian	Italy
HGDP01161	M	Tuscan	Italy
HGDP01162	M	Tuscan	Italy
HGDP01163	M	Tuscan	Italy
HGDP01164	M	Tuscan	Italy
HGDP01166	M	Tuscan	Italy
HGDP01167	M	Tuscan	Italy
HGDP01168	F	Tuscan	Italy
HGDP01169	F	Tuscan	Italy
HGDP01147	M	North Italian	Italy (Bergamo)
HGDP01149	M	North Italian	Italy (Bergamo)
HGDP01151	M	North Italian	Italy (Bergamo)
HGDP01152	M	North Italian	Italy (Bergamo)
HGDP01153	M	North Italian	Italy (Bergamo)
HGDP01154	M	North Italian	Italy (Bergamo)
HGDP01155	M	North Italian	Italy (Bergamo)
HGDP01156	F	North Italian	Italy (Bergamo)
HGDP01157	F	North Italian	Italy (Bergamo)
HGDP01171	F	North Italian	Italy (Bergamo)
HGDP01172	F	North Italian	Italy (Bergamo)
HGDP01173	M	North Italian	Italy (Bergamo)
HGDP01174	M	North Italian	Italy (Bergamo)
HGDP01177	F	North Italian	Italy (Bergamo)
HGDP00747	M	Japanese	Japan
HGDP00748	M	Japanese	Japan
HGDP00749	M	Japanese	Japan

Sample ID	Sex	Population	Geographic origin
HGDP00750	M	Japanese	Japan
HGDP00751	M	Japanese	Japan
HGDP00752	M	Japanese	Japan
HGDP00753	M	Japanese	Japan
HGDP00754	F	Japanese	Japan
HGDP00755	M	Japanese	Japan
HGDP00756	F	Japanese	Japan
HGDP00757	M	Japanese	Japan
HGDP00758	M	Japanese	Japan
HGDP00759	M	Japanese	Japan
HGDP00760	F	Japanese	Japan
HGDP00761	F	Japanese	Japan
HGDP00762	M	Japanese	Japan
HGDP00763	M	Japanese	Japan
HGDP00764	M	Japanese	Japan
HGDP00765	F	Japanese	Japan
HGDP00766	M	Japanese	Japan
HGDP00767	M	Japanese	Japan
HGDP00768	M	Japanese	Japan
HGDP00769	M	Japanese	Japan
HGDP00770	M	Japanese	Japan
HGDP00771	F	Japanese	Japan
HGDP00772	F	Japanese	Japan
HGDP00773	F	Japanese	Japan
HGDP00790	M	Japanese	Japan
HGDP00791	M	Japanese	Japan
HGDP00828	M	Japanese	Japan
HGDP01025	M	Japanese	Japan
HGDP01405	M	Bantu N.E.	Kenya
HGDP01406	M	Bantu N.E.	Kenya
HGDP01408	M	Bantu N.E.	Kenya
HGDP01411	M	Bantu N.E.	Kenya
HGDP01412	M	Bantu N.E.	Kenya
HGDP01413	M	Bantu N.E.	Kenya
HGDP01414	F	Bantu N.E.	Kenya
HGDP01415	M	Bantu N.E.	Kenya
HGDP01416	M	Bantu N.E.	Kenya
HGDP01417	M	Bantu N.E.	Kenya
HGDP01418	M	Bantu N.E.	Kenya
HGDP01419	M	Bantu N.E.	Kenya
HGDP00854	F	Maya	Mexico
HGDP00855	F	Maya	Mexico
HGDP00856	M	Maya	Mexico
HGDP00857	F	Maya	Mexico
HGDP00858	F	Maya	Mexico
HGDP00859	F	Maya	Mexico
HGDP00860	F	Maya	Mexico
HGDP00861	F	Maya	Mexico
HGDP00862	F	Maya	Mexico
HGDP00863	F	Maya	Mexico
HGDP00864	F	Maya	Mexico

Sample ID	Sex	Population	Geographic origin
HGDP00865	F	Maya	Mexico
HGDP00866	F	Maya	Mexico
HGDP00867	F	Maya	Mexico
HGDP00868	F	Maya	Mexico
HGDP00869	F	Maya	Mexico
HGDP00870	F	Maya	Mexico
HGDP00871	F	Maya	Mexico
HGDP00872	F	Maya	Mexico
HGDP00873	F	Maya	Mexico
HGDP00874	F	Maya	Mexico
HGDP00875	F	Maya	Mexico
HGDP00876	F	Maya	Mexico
HGDP00877	M	Maya	Mexico
HGDP00878	M	Maya	Mexico
HGDP01037	M	Pima	Mexico
HGDP01038	F	Pima	Mexico
HGDP01039	M	Pima	Mexico
HGDP01040	M	Pima	Mexico
HGDP01041	F	Pima	Mexico
HGDP01042	M	Pima	Mexico
HGDP01043	M	Pima	Mexico
HGDP01044	F	Pima	Mexico
HGDP01045	M	Pima	Mexico
HGDP01046	F	Pima	Mexico
HGDP01047	M	Pima	Mexico
HGDP01048	F	Pima	Mexico
HGDP01049	F	Pima	Mexico
HGDP01050	M	Pima	Mexico
HGDP01051	F	Pima	Mexico
HGDP01052	M	Pima	Mexico
HGDP01053	F	Pima	Mexico
HGDP01054	F	Pima	Mexico
HGDP01055	M	Pima	Mexico
HGDP01056	F	Pima	Mexico
HGDP01057	M	Pima	Mexico
HGDP01058	F	Pima	Mexico
HGDP01059	M	Pima	Mexico
HGDP01060	M	Pima	Mexico
HGDP01061	M	Pima	Mexico
HGDP00987	M	San	Namidia
HGDP00988	M	San	Namidia
HGDP00991	M	San	Namidia
HGDP00992	M	San	Namidia
HGDP01029	M	San	Namidia
HGDP01032	M	San	Namidia
HGDP01036	M	San	Namidia
HGDP00540	M	Papuan	New Guinea
HGDP00541	M	Papuan	New Guinea
HGDP00542	M	Papuan	New Guinea
HGDP00543	M	Papuan	New Guinea
HGDP00544	F	Papuan	New Guinea

Sample ID	Sex	Population	Geographic origin
HGDP00545	M	Papuan	New Guinea
HGDP00546	M	Papuan	New Guinea
HGDP00547	M	Papuan	New Guinea
HGDP00548	M	Papuan	New Guinea
HGDP00549	M	Papuan	New Guinea
HGDP00550	F	Papuan	New Guinea
HGDP00551	M	Papuan	New Guinea
HGDP00552	F	Papuan	New Guinea
HGDP00553	M	Papuan	New Guinea
HGDP00554	F	Papuan	New Guinea
HGDP00555	M	Papuan	New Guinea
HGDP00556	M	Papuan	New Guinea
HGDP00920	F	Yoruba	Nigeria
HGDP00921	F	Yoruba	Nigeria
HGDP00922	F	Yoruba	Nigeria
HGDP00923	M	Yoruba	Nigeria
HGDP00924	F	Yoruba	Nigeria
HGDP00925	F	Yoruba	Nigeria
HGDP00926	F	Yoruba	Nigeria
HGDP00927	M	Yoruba	Nigeria
HGDP00928	F	Yoruba	Nigeria
HGDP00929	M	Yoruba	Nigeria
HGDP00930	M	Yoruba	Nigeria
HGDP00931	M	Yoruba	Nigeria
HGDP00932	M	Yoruba	Nigeria
HGDP00933	F	Yoruba	Nigeria
HGDP00934	F	Yoruba	Nigeria
HGDP00935	F	Yoruba	Nigeria
HGDP00936	M	Yoruba	Nigeria
HGDP00937	M	Yoruba	Nigeria
HGDP00938	F	Yoruba	Nigeria
HGDP00939	F	Yoruba	Nigeria
HGDP00940	M	Yoruba	Nigeria
HGDP00941	M	Yoruba	Nigeria
HGDP00942	M	Yoruba	Nigeria
HGDP00943	M	Yoruba	Nigeria
HGDP00944	M	Yoruba	Nigeria
HGDP00794	F	Orcadian	Orkney Islands
HGDP00795	M	Orcadian	Orkney Islands
HGDP00796	F	Orcadian	Orkney Islands
HGDP00797	F	Orcadian	Orkney Islands
HGDP00798	M	Orcadian	Orkney Islands
HGDP00799	F	Orcadian	Orkney Islands
HGDP00800	F	Orcadian	Orkney Islands
HGDP00801	F	Orcadian	Orkney Islands
HGDP00802	F	Orcadian	Orkney Islands
HGDP00803	M	Orcadian	Orkney Islands
HGDP00804	M	Orcadian	Orkney Islands
HGDP00805	F	Orcadian	Orkney Islands
HGDP00806	F	Orcadian	Orkney Islands
HGDP00807	M	Orcadian	Orkney Islands

Sample ID	Sex	Population	Geographic origin
HGDP00808	M	Orcadian	Orkney Islands
HGDP00810	M	Orcadian	Orkney Islands
HGDP00052	M	Balochi	Pakistan
HGDP00054	M	Balochi	Pakistan
HGDP00056	M	Balochi	Pakistan
HGDP00057	M	Balochi	Pakistan
HGDP00058	M	Balochi	Pakistan
HGDP00060	M	Balochi	Pakistan
HGDP00062	M	Balochi	Pakistan
HGDP00064	M	Balochi	Pakistan
HGDP00066	M	Balochi	Pakistan
HGDP00068	M	Balochi	Pakistan
HGDP00070	M	Balochi	Pakistan
HGDP00072	M	Balochi	Pakistan
HGDP00074	M	Balochi	Pakistan
HGDP00076	M	Balochi	Pakistan
HGDP00078	M	Balochi	Pakistan
HGDP00080	M	Balochi	Pakistan
HGDP00082	M	Balochi	Pakistan
HGDP00084	M	Balochi	Pakistan
HGDP00086	M	Balochi	Pakistan
HGDP00088	M	Balochi	Pakistan
HGDP00090	M	Balochi	Pakistan
HGDP00092	M	Balochi	Pakistan
HGDP00094	M	Balochi	Pakistan
HGDP00096	M	Balochi	Pakistan
HGDP00098	M	Balochi	Pakistan
HGDP00001	M	Brahui	Pakistan
HGDP00003	M	Brahui	Pakistan
HGDP00005	M	Brahui	Pakistan
HGDP00007	M	Brahui	Pakistan
HGDP00009	M	Brahui	Pakistan
HGDP00011	M	Brahui	Pakistan
HGDP00013	M	Brahui	Pakistan
HGDP00015	M	Brahui	Pakistan
HGDP00017	M	Brahui	Pakistan
HGDP00019	M	Brahui	Pakistan
HGDP00021	M	Brahui	Pakistan
HGDP00023	M	Brahui	Pakistan
HGDP00025	M	Brahui	Pakistan
HGDP00027	M	Brahui	Pakistan
HGDP00029	M	Brahui	Pakistan
HGDP00031	M	Brahui	Pakistan
HGDP00033	M	Brahui	Pakistan
HGDP00035	M	Brahui	Pakistan
HGDP00037	M	Brahui	Pakistan
HGDP00039	M	Brahui	Pakistan
HGDP00041	M	Brahui	Pakistan
HGDP00043	M	Brahui	Pakistan
HGDP00045	M	Brahui	Pakistan
HGDP00047	M	Brahui	Pakistan

Sample ID	Sex	Population	Geographic origin
HGDP00049	M	Brahui	Pakistan
HGDP00336	F	Burusho	Pakistan
HGDP00338	F	Burusho	Pakistan
HGDP00341	M	Burusho	Pakistan
HGDP00346	M	Burusho	Pakistan
HGDP00351	M	Burusho	Pakistan
HGDP00356	F	Burusho	Pakistan
HGDP00359	M	Burusho	Pakistan
HGDP00364	M	Burusho	Pakistan
HGDP00371	F	Burusho	Pakistan
HGDP00372	M	Burusho	Pakistan
HGDP00376	M	Burusho	Pakistan
HGDP00382	M	Burusho	Pakistan
HGDP00388	M	Burusho	Pakistan
HGDP00392	M	Burusho	Pakistan
HGDP00397	M	Burusho	Pakistan
HGDP00402	M	Burusho	Pakistan
HGDP00407	M	Burusho	Pakistan
HGDP00412	M	Burusho	Pakistan
HGDP00417	M	Burusho	Pakistan
HGDP00423	M	Burusho	Pakistan
HGDP00428	M	Burusho	Pakistan
HGDP00433	M	Burusho	Pakistan
HGDP00438	M	Burusho	Pakistan
HGDP00444	F	Burusho	Pakistan
HGDP00445	M	Burusho	Pakistan
HGDP00099	M	Hazara	Pakistan
HGDP00100	M	Hazara	Pakistan
HGDP00102	M	Hazara	Pakistan
HGDP00103	M	Hazara	Pakistan
HGDP00104	M	Hazara	Pakistan
HGDP00105	M	Hazara	Pakistan
HGDP00106	M	Hazara	Pakistan
HGDP00108	M	Hazara	Pakistan
HGDP00109	M	Hazara	Pakistan
HGDP00110	M	Hazara	Pakistan
HGDP00111	M	Hazara	Pakistan
HGDP00112	M	Hazara	Pakistan
HGDP00113	M	Hazara	Pakistan
HGDP00115	M	Hazara	Pakistan
HGDP00116	M	Hazara	Pakistan
HGDP00118	M	Hazara	Pakistan
HGDP00119	M	Hazara	Pakistan
HGDP00120	M	Hazara	Pakistan
HGDP00121	M	Hazara	Pakistan
HGDP00122	M	Hazara	Pakistan
HGDP00124	M	Hazara	Pakistan
HGDP00125	M	Hazara	Pakistan
HGDP00127	M	Hazara	Pakistan
HGDP00128	M	Hazara	Pakistan
HGDP00129	M	Hazara	Pakistan

Sample ID	Sex	Population	Geographic origin
HGDP00267	M	Kalash	Pakistan
HGDP00274	F	Kalash	Pakistan
HGDP00277	M	Kalash	Pakistan
HGDP00279	M	Kalash	Pakistan
HGDP00281	M	Kalash	Pakistan
HGDP00285	M	Kalash	Pakistan
HGDP00286	F	Kalash	Pakistan
HGDP00288	M	Kalash	Pakistan
HGDP00290	M	Kalash	Pakistan
HGDP00292	M	Kalash	Pakistan
HGDP00298	F	Kalash	Pakistan
HGDP00302	M	Kalash	Pakistan
HGDP00304	F	Kalash	Pakistan
HGDP00307	M	Kalash	Pakistan
HGDP00309	M	Kalash	Pakistan
HGDP00311	M	Kalash	Pakistan
HGDP00313	M	Kalash	Pakistan
HGDP00315	M	Kalash	Pakistan
HGDP00319	M	Kalash	Pakistan
HGDP00321	M	Kalash	Pakistan
HGDP00323	F	Kalash	Pakistan
HGDP00326	M	Kalash	Pakistan
HGDP00328	M	Kalash	Pakistan
HGDP00330	M	Kalash	Pakistan
HGDP00333	M	Kalash	Pakistan
HGDP00130	M	Makrani	Pakistan
HGDP00131	M	Makrani	Pakistan
HGDP00133	M	Makrani	Pakistan
HGDP00134	M	Makrani	Pakistan
HGDP00135	M	Makrani	Pakistan
HGDP00136	M	Makrani	Pakistan
HGDP00137	M	Makrani	Pakistan
HGDP00139	M	Makrani	Pakistan
HGDP00140	M	Makrani	Pakistan
HGDP00141	M	Makrani	Pakistan
HGDP00143	M	Makrani	Pakistan
HGDP00144	M	Makrani	Pakistan
HGDP00145	M	Makrani	Pakistan
HGDP00146	M	Makrani	Pakistan
HGDP00148	M	Makrani	Pakistan
HGDP00149	M	Makrani	Pakistan
HGDP00150	M	Makrani	Pakistan
HGDP00151	F	Makrani	Pakistan
HGDP00153	F	Makrani	Pakistan
HGDP00154	F	Makrani	Pakistan
HGDP00155	F	Makrani	Pakistan
HGDP00157	F	Makrani	Pakistan
HGDP00158	M	Makrani	Pakistan
HGDP00160	M	Makrani	Pakistan
HGDP00161	M	Makrani	Pakistan
HGDP00213	M	Pathan	Pakistan

Sample ID	Sex	Population	Geographic origin
HGDP00214	M	Pathan	Pakistan
HGDP00216	M	Pathan	Pakistan
HGDP00218	M	Pathan	Pakistan
HGDP00220	M	Pathan	Pakistan
HGDP00222	M	Pathan	Pakistan
HGDP00224	M	Pathan	Pakistan
HGDP00226	M	Pathan	Pakistan
HGDP00228	M	Pathan	Pakistan
HGDP00230	M	Pathan	Pakistan
HGDP00232	F	Pathan	Pakistan
HGDP00234	M	Pathan	Pakistan
HGDP00237	F	Pathan	Pakistan
HGDP00239	F	Pathan	Pakistan
HGDP00241	M	Pathan	Pakistan
HGDP00243	M	Pathan	Pakistan
HGDP00244	F	Pathan	Pakistan
HGDP00247	F	Pathan	Pakistan
HGDP00248	M	Pathan	Pakistan
HGDP00251	M	Pathan	Pakistan
HGDP00254	M	Pathan	Pakistan
HGDP00258	M	Pathan	Pakistan
HGDP00259	M	Pathan	Pakistan
HGDP00262	M	Pathan	Pakistan
HGDP00264	M	Pathan	Pakistan
HGDP00163	M	Sindhi	Pakistan
HGDP00165	M	Sindhi	Pakistan
HGDP00167	M	Sindhi	Pakistan
HGDP00169	M	Sindhi	Pakistan
HGDP00171	M	Sindhi	Pakistan
HGDP00173	M	Sindhi	Pakistan
HGDP00175	M	Sindhi	Pakistan
HGDP00177	M	Sindhi	Pakistan
HGDP00179	M	Sindhi	Pakistan
HGDP00181	M	Sindhi	Pakistan
HGDP00183	M	Sindhi	Pakistan
HGDP00185	M	Sindhi	Pakistan
HGDP00187	M	Sindhi	Pakistan
HGDP00189	M	Sindhi	Pakistan
HGDP00191	M	Sindhi	Pakistan
HGDP00192	F	Sindhi	Pakistan
HGDP00195	F	Sindhi	Pakistan
HGDP00197	M	Sindhi	Pakistan
HGDP00199	M	Sindhi	Pakistan
HGDP00201	M	Sindhi	Pakistan
HGDP00203	M	Sindhi	Pakistan
HGDP00205	M	Sindhi	Pakistan
HGDP00206	F	Sindhi	Pakistan
HGDP00208	M	Sindhi	Pakistan
HGDP00210	F	Sindhi	Pakistan
HGDP00879	M	Russian	Russia
HGDP00880	M	Russian	Russia

Sample ID	Sex	Population	Geographic origin
HGDP00881	F	Russian	Russia
HGDP00882	M	Russian	Russia
HGDP00883	M	Russian	Russia
HGDP00884	F	Russian	Russia
HGDP00885	F	Russian	Russia
HGDP00886	M	Russian	Russia
HGDP00887	M	Russian	Russia
HGDP00888	M	Russian	Russia
HGDP00889	F	Russian	Russia
HGDP00890	M	Russian	Russia
HGDP00891	M	Russian	Russia
HGDP00892	M	Russian	Russia
HGDP00893	M	Russian	Russia
HGDP00894	M	Russian	Russia
HGDP00895	M	Russian	Russia
HGDP00896	M	Russian	Russia
HGDP00897	M	Russian	Russia
HGDP00898	F	Russian	Russia
HGDP00899	F	Russian	Russia
HGDP00900	M	Russian	Russia
HGDP00901	F	Russian	Russia
HGDP00902	F	Russian	Russia
HGDP00903	F	Russian	Russia
HGDP01381	F	Adygei	Russia Caucasus
HGDP01382	F	Adygei	Russia Caucasus
HGDP01383	M	Adygei	Russia Caucasus
HGDP01384	F	Adygei	Russia Caucasus
HGDP01385	M	Adygei	Russia Caucasus
HGDP01386	F	Adygei	Russia Caucasus
HGDP01387	F	Adygei	Russia Caucasus
HGDP01388	F	Adygei	Russia Caucasus
HGDP01396	M	Adygei	Russia Caucasus
HGDP01397	M	Adygei	Russia Caucasus
HGDP01398	F	Adygei	Russia Caucasus
HGDP01399	F	Adygei	Russia Caucasus
HGDP01400	F	Adygei	Russia Caucasus
HGDP01401	F	Adygei	Russia Caucasus
HGDP01402	M	Adygei	Russia Caucasus
HGDP01403	M	Adygei	Russia Caucasus
HGDP01404	M	Adygei	Russia Caucasus
HGDP00904	M	Mandenka	Senegal
HGDP00905	M	Mandenka	Senegal
HGDP00906	M	Mandenka	Senegal
HGDP00907	M	Mandenka	Senegal
HGDP00908	M	Mandenka	Senegal
HGDP00909	F	Mandenka	Senegal
HGDP00910	F	Mandenka	Senegal
HGDP00911	M	Mandenka	Senegal
HGDP00912	M	Mandenka	Senegal
HGDP00913	M	Mandenka	Senegal
HGDP00914	F	Mandenka	Senegal

Sample ID	Sex	Population	Geographic origin
HGDP00915	F	Mandenka	Senegal
HGDP00916	F	Mandenka	Senegal
HGDP00917	F	Mandenka	Senegal
HGDP00918	F	Mandenka	Senegal
HGDP00919	M	Mandenka	Senegal
HGDP01199	M	Mandenka	Senegal
HGDP01200	M	Mandenka	Senegal
HGDP01201	F	Mandenka	Senegal
HGDP01202	M	Mandenka	Senegal
HGDP01283	M	Mandenka	Senegal
HGDP01284	M	Mandenka	Senegal
HGDP01285	M	Mandenka	Senegal
HGDP01286	M	Mandenka	Senegal
HGDP00945	M	Yakut	Siberia
HGDP00946	M	Yakut	Siberia
HGDP00947	M	Yakut	Siberia
HGDP00948	M	Yakut	Siberia
HGDP00949	M	Yakut	Siberia
HGDP00950	M	Yakut	Siberia
HGDP00951	M	Yakut	Siberia
HGDP00952	M	Yakut	Siberia
HGDP00953	M	Yakut	Siberia
HGDP00954	M	Yakut	Siberia
HGDP00955	F	Yakut	Siberia
HGDP00956	F	Yakut	Siberia
HGDP00957	F	Yakut	Siberia
HGDP00958	M	Yakut	Siberia
HGDP00959	F	Yakut	Siberia
HGDP00960	M	Yakut	Siberia
HGDP00961	M	Yakut	Siberia
HGDP00962	M	Yakut	Siberia
HGDP00963	F	Yakut	Siberia
HGDP00964	M	Yakut	Siberia
HGDP00965	M	Yakut	Siberia
HGDP00966	F	Yakut	Siberia
HGDP00967	F	Yakut	Siberia
HGDP00968	M	Yakut	Siberia
HGDP00969	M	Yakut	Siberia
HGDP00993	M	Bantu S.E. Pedi	South Africa
HGDP00994	M	Bantu S.E. S.Sotho	South Africa
HGDP01030	M	Bantu S.E. Tswana	South Africa
HGDP01034	M	Bantu S.E. Tswana	South Africa
HGDP01033	M	Bantu S.E. Zulu	South Africa
HGDP01028	M	Bantu S.W. Herero	South Africa
HGDP01035	M	Bantu S.W. Herero	South Africa
HGDP01031	M	Bantu S.W. Ovambo	South Africa

Table A4 European-Indian (CIN) control panel

Sample ID	Ethnic Origin	Sample ID	Ethnic Origin
POD-AB41	European	POD-BD462	Indian
POD-AB44	European	POD-BD56	Indian
POD-AB54	European	POD-BD57	Indian
POD-AB56	European	POD-BD59	Indian
POD-AB79	European	POD-BD61	Indian
POD-AB80	European	POD-BD78	Indian
POD-AC315	Indian	POD-BD781	Indian
POD-AC332	European	POD-BD891	Indian
POD-AF124	Indian	POD-BO101	European
POD-AF15	European	POD-BO19	European
POD-AF236	Indian	POD-BO590	European
POD-AF24	European	POD-BO65	European
POD-AF3	European	POD-BO71	European
POD-AF50	European	POD-BO96	European
POD-AF6	European	POD-BQ63	Indian
POD-AG0033	European	POD-BQ68	European
POD-AG11	European	POD-BV278	Indian
POD-AG17	European	POD-BW152	Indian
POD-AG547	European	POD-BW156	Indian
POD-AG646	Indian	POD-BW158	Indian
POD-AG66	European	POD-BW160	Indian
POD-AJ10	European	POD-CB276	Indian
POD-AJ2	Indian	POD-CB464	Indian
POD-AK353	European	POD-CB874	Indian
POD-AK400	Indian	POD-CD59	Indian
POD-AK503	European	POD-CD68	Indian
POD-AM107	Indian	POD-CD79	Indian
POD-AM125	Indian	POD-DB104	Indian
POD-AM148	Indian	POD-GG1	European
POD-AM16	Indian	POD-HH15	European
POD-AM17	Indian	POD-HH170	European
POD-AM21	Indian	POD-HH176	European
POD-AM31	Indian	POD-HH195	European
POD-AM37	Indian	POD-HH198	European
POD-AM38	Indian	POD-HH30	European
POD-AM40	Indian	POD-HH74	European
POD-AM42	Indian	POD-HH8	European
POD-AM54	Indian	POD-MM16	European
POD-AM69	European	POD-QQ63	Indian
POD-AR55	European	POD-QQ65	Indian
POD-AR70	Indian	POD-QQ67	Indian
POD-AW165	Indian	POD-TT100	European
POD-AW53	Indian	POD-TT37	European
POD-AX17	European	POD-TT71	European
POD-BC45	Indian	POD-WW001	European
POD-BD238	Indian	POD-WW5	European
POD-BD245	European	POD-YY4	European

Table A5 *LPIN2* sequence variants detected in a cohort of severe insulin resistant patients and 11 Indian and 23 CEPH controls

Genic position	Genomic position	Minor|Major allele	Protein consequence	MAF in SIR	Detected in controls?	rs ID
Upstream	3002144	C|G		0.0055		rs3810061
Upstream	3002078	G|C		0.0989		rs3810062
Upstream	3001969	T|C		0.0026		
5prime_UTR	3001787	A|T		0.0026		
5prime_UTR	3001772	A|G		0.0026		
5prime_UTR	3001756	C|T		0.0026		
5prime_UTR	3001743	C|G		0.0078		
5prime_UTR	3001742	C|G		0.0078		
Intron 1	3001646	T|C		0.0026		
Intron 1	3001587	A|G		0.1		
Intron 1	3001563	C|T		0.1099		
Intron 1	3001480	T|C		0.0026		
Intron 1	3001440	G|T		0.006		
Intron 1	3001415	A|G		0.0186		rs12455800
Intron 1	2972876	C|T		0.0027		
Intron 1	2972733	T|A		0.0026		
Intron 3	2944448	A|G		0.0026		
Intron 3	2944439	T|C		0.3803		rs7226624
Intron 4	2940926	G|A		0.0026		
Exon 5	2930695	G|A	A202A	0.0027		
Exon 5	2930693	A|G	S203F	0.0081	Yes	
Intron 5	2929627	G|T		0.0026		
Intron 6	2929424	C|T		0.0026		
Intron 6	2929307	A|T		0.0026		
Intron 7	2924536	C|G		0.1219		rs3765622
Intron 9	2921225	C|T		0.0133		rs16944068
Intron 9	2921074	C|G		0.0081		
Intron 9	2919186	G|T		0.0026		
Exon 10	2919124	T|C	**E497K**	0.0026	No	
Intron 12	2917669	C|A		0.0026		
Intron 12	2917023	G|T		0.4274		rs602233
Exon 13	2916779	G|A	S579P	0.0026	Yes	
Exon 14	2915359	T|C	E601K	0.0111	Yes	
Exon 14	2915284	A|G	**P626S**	0.0027	No	
Intron 14	2915104	T|C		0.0028		
Intron 15	2914278	G|A		0.0026		
Intron 15	2913884	C|A		0.0052		
Exon 16	2913790	A|G	L719L	0.0052		
Intron 16	2913699	A|T		0.0026		
Intron 16	2912450	C|T		0.378		rs12458532
Intron 16	2912360	A|C		0.0026		
Intron 17	2911979	T|C		0.032		
Intron 17	2911777	C|T		0.2771		rs2282636
Intron 17	2911752	C|T		0.0027		
Intron 18	2911444	C|T		0.2771		rs2282635
Intron 18	2911433	C|A		0.0027		
Intron 19	2910725	T|A		0.1648		rs3737514

Genic position	Genomic position	Minor|Major allele	Protein consequence	MAF in SIR	Detected in controls?	rs ID
Intron 19	2910631	G|A		0.014		
Intron 19	2910482	A|G		0.0083		rs16944043
Exon 20	2910387	G|A	S865S	0.0029		
Exon 20	2910372	A|G	S870S	0.0026		
3prime_UTR	2910288	A|G		0.293		rs3745012
3prime_UTR	2910201	A|G		0.0108		
3prime_UTR	2910186	A|G		0.0026		
3prime_UTR	2910124	A|G		0.0026		
3prime_UTR	2909933	T|C		0.0052		
3prime_UTR	2909853	A|G		0.0052		
3prime_UTR	2909843	A|G		0.0026		
3prime_UTR	2909790	G|A		0.0629		rs35176958
3prime_UTR	2909734	T|C		0.05		
3prime_UTR	2909639	A|G		0.0026		
3prime_UTR	2909634	G|T		0.086		rs16944040
3prime_UTR	2909606	A|T		0.2783		rs607549
3prime_UTR	2909538	G|A		0.0026		
3prime_UTR	2909494	C|A		0.0026		rs3810064
3prime_UTR	2909478	G|T		0.0026		
3prime_UTR	2909377	G|A		0.0028		
3prime_UTR	2909341	G|T		0.0058		rs3810065
3prime_UTR	2909239	G|A		0.0364		
3prime_UTR	2909214	C|G		0.0027		
3prime_UTR	2909209	T|C		0.0026		
3prime_UTR	2909170	T|C		0.0058		
3prime_UTR	2908933	T|C		0.0026		
3prime_UTR	2908888	T|C		0.0026		
3prime_UTR	2908789	T|C		0.009		
3prime_UTR	2908699	C|T		0.0026		
3prime_UTR	2908669	G|A		0.0026		
3prime_UTR	2908426	A|C		0.0027		
3prime_UTR	2908215	T|G		0.0028		
3prime_UTR	2908188	T|C		0.0081		
3prime_UTR	2907884	T|C		0.0029		
3prime_UTR	2907874	A|G		0.0031		
3prime_UTR	2907798	C|G		0.037		rs8091401
3prime_UTR	2907756	G|A		0.0026		
3prime_UTR	2907738	T|C		0.0052		
3prime_UTR	2907357	G|A		0.3882		rs7980
3prime_UTR	2907345	T|C		0.0106		rs4781
3prime_UTR	2907276	G|A		0.0026		rs14916
3prime_UTR	2907223	A|T		0.2777		rs1985
3prime_UTR	2907185	A|C		0.0026		
3prime_UTR	2907152	C|T		0.2063		rs1164
Downstream	2906904	G|C		0.0846		

Genomic coordinates are NCBI build 36 (B36).

Table A6 *LPIN3* sequence variants detected in a cohort of severe insulin resistant patients and 11 Indian and 23 CEPH controls

Genic position	Genomic position	Minor|Major allele	Protein consequence	MAF in SIR	Detected in controls?	rs ID
Upstream	39402927	A|C		0.2807		rs6029636
Intron 1	39403067	C|A		0.0027		
Intron 1	39403147	T|C		0.0026		
Intron 1	39407730	G|A		0.2884		rs6129820
Exon 2	39407866	T|C	Y3Y	0.0027		
Exon 2	39407896	T|G	G13G	0.0026		rs16985673
Exon 2	39407916	A|G	R20Q	0.0132	Yes	
Exon 2	39407947	T|C	G30G	0.0026		
Exon 2	39407978	A|G	**G41S**	0.0026	No	
Exon 2	39407998	G|C	P47P	0.0165		
Exon 2	39408038	T|C	R61W	0.0239	Yes	
Intron 2	39408175	G|C		0.0029		
Intron 2	39408256	A|G		0.0058		
Intron 3	39409732	T|C		0.0026		
Exon 4	39410714	C|G	**W110C**	0.0027	No	
Intron 4	39410998	T|C		0.0225		rs6072349
Intron 4	39411052	C|G		0.0304		
Intron 5	39411282	G|A		0.0179		
Intron 6	39411978	T|G		0.2549		rs6029646
Intron 6	39412063	G|T		0.0235		
Exon 7	39412412	C|G	V355L	0.0161	Yes	
Intron 7	39412493	T|C		0.0026		
Intron 7	39412628	A|G		0.0026		
Exon 8	39413957	A|G	A395A	0.0026		
Exon 9	39414271	T|C	L454L	0.0052		
Intron 9	39414286	T|C		0.0263		rs6102366
Intron 9	39414375	C|G		0.0265		
Intron 9	39414593	T|A		0.0217		
Intron 11	39414984	T|C		0.0026		rs8119824
Intron 11	39416616	C|T		0.0026		
Exon 12	39416786	A|G	**E539K**	0.0026	No	
Exon 12	39416791	T|G	E540D	0.0026	Yes	
Intron 14	39419011	G|A		0.0263		rs6102369
Intron 14	39419066	T|C		0.0026		
Intron 15	39419261	C|T		0.1421		rs6072350
Exon 17	39419954	A|G	L686L	0.0315		rs2072969
Exon 17	39419963	A|G	S689S	0.0026		rs41277020
Intron 17	39420265	T|C		0.0494		rs41277022
Intron 17	39420272	A|G		0.0027		
Intron 19	39420739	T|C		0.0186		
3prime_UTR	39421026	A|G		0.1548		
3prime_UTR	39421362	A|G		0.1538		
3prime_UTR	39421519	C|T		0.0027		
3prime_UTR	39421553	C|G		0.2158		rs6072351
3prime_UTR	39421666	A|C		0.1587		rs6065338
3prime_UTR	39421761	T|C		0.0238		
3prime_UTR	39422072	C|T		0.0026		rs2235595

Genic position	Genomic position	Minor|Major allele	Protein consequence	MAF in SIR	Detected in controls?	rs ID
3prime_UTR	39422160	A|G		0.0216		rs12625592
3prime_UTR	39422218	C|T		0.0026		
3prime_UTR	39422426	A|G		0.0027		
3prime_UTR	39422490	A|C		0.0027		rs6102378
3prime_UTR	39422595	T|C		0.0376		rs2235594
Downstream	39422749	A|G		0.0246		
Downstream	39422867	C|A		0.0093		rs41278102

Genomic coordinates are NCBI build 36 (B36).

Table A7 *LPIN* family PCR primer sequences

Gene	Genic position	Forward	Reverse	Amplicon length	Annealing temperature
LPIN1	5'UTR/Exon 1	CAGGAACGCTAAGATAGGCG	GAACCACTTCCCGTAGAGGG	404	FAIL
	Exon 2	ATGGGGGTGGATAGAACACA	GTTCAAGACAGGTGGCATCA	435	60
	Exon 3	TTTGGGTTCCATCTTTCCTG	AAATCCCCCACACAGATCAA	481	53
	Exon 4	CTCTAGCTGGGCCATCTGC	CTGGTCCCTAGTGCCAAAAA	593	60
	Exon 5	TTAGAATTCGGCCAAAGCAT	CCGTGTTTTTCCTGGTCCTA	460	60
	Exon 6	GGGCCACTTCTTTTTCTCCT	CTGGCTCCTGGTAGCTGTGT	491	60
	Exon 7	CCAGGTGATTGTTGAGGCTT	TGGAAAGGACCTGTCATGTG	514	55
	Exon 8	ACTGTGCCCAAACCTCAAAG	CCCAGCTGTCTTGCACACT	418	60
	Exon 9	ATACAGGGAGATGCAGCCAG	GCACTCCTTGGAGAAGCCTA	492	60
	Exon 10	TAAAATGGTGCGGCCTTTAC	CATGGAACCCTGACCACAGT	414	60
	Exon 11	TGCCTAGGGGTGATATGGAG	GCCTTGATAAGGAAGAGGGG	435	60
	Exon 12	TCATGTAATTTTGGACGCCAT	GAGCTGGCTCATGGAAACTC	408	60
	Exon 13	CGTCATTGGGGAATGGTCTA	AAGGACTATTGCATGCCAGG	478	60
	Exon 14	GCAAGTGAATGAGGAGAGGC	GGGTACCCAGTTGAGGGATT	496	60
	Exon 15	AAGGGTGGGGAAAGAAAGAA	AGAATTCGAACCGTGCAAGT	413	60
	Exon 16	TCGACTTGGTTGCTTCTGTG	TGCTGGGACAGGAAAGATTC	463	60
	Exon 17	GGACTTTTGTATGTGGGGGA	AGGAGCCAATGCAGGTGTAG	456	60
	Exon 18	CTCGTGCTTTTCTGGGATGT	TGGCTCTTCACCAGGCTACT	407	60
	Exon 19	TTCACTGCACCACTTTGAGG	AGACCCAGAAGGGATCTGATT	446	60
	Exon 20	CGTGTCGATAAGTAGGCGGT	TGCTTAGAAATGTCAGCCCC	459	60
	3'UTR	TCCTCTGCGCTTGTTCACTA	CCAGGGCTCACATTGAACTT	371	60
	3'UTR	AGGCGGAGTTTTCAGAGGAT	TGTGCTTACAGTGGTGATCCA	489	60
	3'UTR	TTTCATTGGCATATCTCCCC	AGGCCCTTTCGATAGTCTAACAC	453	60
	3'UTR	TGGGAGCTCTCCAAAGAAAC	GGAAAATGTTTCAATCGTTGG	463	60
	3'UTR	TAACACCTGCCTTCTGGCTT	TTGCATCCAGTACAACACAATG	420	60
	3'UTR	ACTCAAATTAAGGGCAAGCG	GATGGCAGATCTGTTTGCCT	389	60
	3'UTR	AAATGGAACGGTTGAATGAAA	AACCGTCTGAACCTTTGCAG	473	60

Gene	Genic position	Forward	Reverse	Amplicon length	Annealing temperature
*LPIN2**	Exon 1	ATCAAAAGTTGGGCATGGAG	CACACAGTGCAGGGGATG	526	60
	Exon 1	TACAATGTAATTGGAGGGAAGAGC	ACGAGGCTCTAGACTTGGTGTC	450	60
	Exon 2	CATTGTCTCAACATCATTCAGG	CGTGTGTGGAAATGCGTAGA	540	60
	Exon 3	TGTCTTGCCAACAACAGTCC	TGGTTGTCCATCTCTCCCTC	419	60
	Exon 4	GAGCTCACGCCTACATCACA	TGGCATGTCAGTTCCTTCAG	524	60
	Exon 4	AACTGCCTTTGCTGCTTGAT	CTCAAGGAAAAGCTTCCTGC	460	60
	Exon 5	GATGGCTAAATTCCATGCTGA	ATAGTGGGTGGCATTGGTGT	525	60
	Exon 6	GAAATCCTTCGGTTGCAAAT	AATAATTGATGCTCCCTTCCA	467	60
	Exon 7	CAGCCATCACGTTATGTGGA	AAGGCAATTATTAGTTTTTCTCACTT	600	FAIL
	Exon 8	TTCCAAAACCTGAAGCCATC	GGCTGGGAACCAACCTAAG	401	60
	Exon 9	CACTACCTCAAAGCCACTCCA	CCCTATCTAAGAAAAGTTATGCCTGA	545	60
	Exon 10	ACCTTCGCTTCAAGCACAAC	TCTTTCTTGTCCTGGGGAAA	520	60
	Exon 11	CCAGAGTTTGGGCATGAGAT	AATTAGCCAGTGGGCAATGA	541	60
	Exon 12	ACCTGCACGTCCTAAGTGCT	ATGAGATTTGCCCACCTGTC	508	60
	Exon 13	GTCATCTGCTGAGTGGCACA	AGAAGGAAGCTGGGGACCTA	535	60
	Exon 14	AGGTGGGCCTGATAGCTCTT	GGATATGGGAGAATGGCAGA	468	60
	Exon 15	ACCATCTCCTCCATGCAAAA	CTGTGTCTAAGACCCCGGAA	467	60
	Exon 16	GACACAGGTTCGATGCCAAT	GACTCCGCTTTCAGAAGCAC	473	60
	Exon 17	ATTGACCCAGTGCAGGTAGC	TGGTTTTCACAGTAAATGAGAGGA	515	60
	Exon 17	CCCCAGTGAGTGGTCCTAAA	GCCAGAACATGTGGGCTAAC	543	60
	Exon 18	AGGACCTCCCCAGTTGCTAT	TGGTGAAGAAAGGCAGGTTC	516	60
	Exon 19	ACAGGGCTCAAACTCCCTCT	GCTCAGTTTTGGTGGCAACT	461	60
	Exon 19 and 20	GACAGGTCATCCAGGTCCAC	TGAACCCCAAGGGTGAATTA	533	60
	Exon 20	AGGAAACATGTGTGCGACC	TGAGTGACAGCTTCACAGCC	498	60
	3'UTR	CCAAAGAGGGCCAGGTTTA	GCTGACCTCAGATACCAGCC	549	60
	3'UTR	TTCACTGTGCTTCCCTGACA	GTCAGTGATGCTGGCTGAAG	512	60
	3'UTR	AGCAGCTTATGGCACACCTT	TTAAAACAGACCAAACCCCG	491	60
	3'UTR	ATGGCTGAAATCAGTCAGGG	GTGGAAGCAGCGTCAAAGA	543	60
	3'UTR	CAAGAGGACACACAGATGGC	GTGGTTCCTGAGCAGAGAGG	505	60

Gene	Genic position	Forward	Reverse	Amplicon length	Annealing temperature
	3'UTR	TCAGCAAGAGGGACTCCTTT	AAGCTTCCCCCAGCTTAAAT	536	60
	3'UTR	AGCACACAGACAGCAAAGCA	TTCTCTTTGGACACATGGGC	509	60
	3'UTR	TTCCCAGTCATCCAATCTCC	GAGATGCCCTCGAGTGTGTT	453	60
	3'UTR	GGCTGAACATAATTATTAAAAGAGCA	GTGCTCGTGTAGACCTGGAG	511	60**
	3'UTR	AGGTTAGTTCCTGGGGGTGT	TCCCGTGTTGTAGAGCTTCC	549	60
*LPIN3**	Exon 1	TCTTTCCAGGGAGTGGAGC	CGACGTCTCCTCTCCAGACT	437	FAIL
	Exon 2	TTTAGAGGCACGAGGCAACT	GAGCACTCACCACCTTCTCC	510	60
	Exon 2	AGCTTCTTGTGGAGGAGCTG	GGATGAGCCCTACTGGTCTG	490	60
	Exon 3	GATGGGTAGGGAGCTCAGTG	CATATCACGTGTGCCAGAGG	485	60
	Exon 4	CCACACCTGTTTTCCTGACC	GCCACAAGGAAATGGTCACT	595	FAIL
	Exon 5	CCAGGGTGTGTAAGGACCAA	GAAGATGGCAAAGTTGTAGCTG	637	60
	Exon 6	TTCTTACTGGATTGTCTTCTTCTTG	CCTTCAAGGACCACTGAGGA	528	60
	Exon 7	ACACCCCAGAGACAGAGGAA	ACCTGAGCTTGTGAGCTGGA	618	60
	Exon 8	TTCAGAGTCAGAGGGCACTT	GTCCACCACAGAGGGACAGT	451	60
	Exon 9	GCTTTACTTCCCCCAAAGGT	GGAGTGTGGACTGCACAGG	462	FAIL
	Exon 10	AGTCCCCATGAAGCAAAATG	TCTCCGTGCCCTGAGTCTAT	541	60
	Exon 11	CAGCACAGCGTCTCTTACCA	CTTCCCCTCTGGTTTGTCCT	509	60
	Exons 12 and 13	AGGTGAGCACACTTCCTTGG	ACAGGGTGCTGGCTCCTC	468	FAIL
	Exon 14	TCTCCAAGGCAGACACTGAG	GCTCCCAAGGGGATAGATTC	530	60
	Exon 15	ATGGATGTGTCTGTGCATGG	TTACTCCAAGGCAGCTCCAC	471	60
	Exon 16	GTCTTCAGCGTGACCACTCA	GGATGAGCCTTTCCCAGAAC	543	60
	Exon 17	GTCCAGCTGAGTCCACTGC	GGGCCAAAGTAGGGAAGAAG	487	60
	Exon 18	CTCTGCCCTCCACAGGTAAC	TAGCCTCACGTGGATTTGTG	532	60
	Exon 19	CCTGCCTGAGTGACATCCA	GAGGAGCTCGACCACTTCAC	481	60
	Exon 20	GCTAAACCCTGCCATGTTCC	GCTAATTGCCCCAGGACAG	550	60**
	3'UTR	AAACCCACTGAAGGGGAAG	AGGGCATAGAAAACCCTGC	500	FAIL
	3'UTR	GGGGCAATTAGCTTGTCATC	TGTACCACACTCCCACTGCT	497	FAIL
	3'UTR	CCTTGCAGGGTTTTCTATGC	CTGCTTCCCACAGCTCCTTA	476	FAIL
	3'UTR	CATGACACACACCCACACCT	AAAATGCATGTTCCTCAGGC	487	FAIL

Genic position	Genomic position	Minor|Major allele	Protein consequence	MAF in SIR	Detected in controls?
	3'UTR	GTGGAGGGCAGAATACCAAA	CCCCCACTCCAAATTTATCC	460	FAIL
	3'UTR	CCTCATTCAGGCCACTTTGT	TCAGCAACTGTGCAGAGACC	539	60**

* *LPIN2* and *LPIN3* primers are M13-tailed. Forward primer sequences are preceded by TGTAAAACGACGGCCAGT and reverse sequences are preceded by AGGAAACAGCTATGACCAT. ** 7-deaza-dGTP was added to dNTP mix.

Table A8 *LPIN1* sequenom primers

Plex	SNP_ID	2nd PCR primer	1st PCR primer	MassEXTEND primer
hME1	G582R	ACGTTGGATGGTAAGCCAGAGCAGTGCTTG	ACGTTGGATGATCCCAAGAAGCCTCCTACC	GCAAGGCCCATAGCACC
hME1	R552K	ACGTTGGATGAGGCCACTGTGGAATCTATC	ACGTTGGATGTTGTGGTGTTTCTTCCCCTC	GCCCAAAAAGGGAGGAA
hME1	A353T	ACGTTGGATGAGCCTCAGACAGAAATGCAG	ACGTTGGATGGTTTGAGCTCCTCGATCATG	AGACCTGGAGACCTTAGGA
iPlex1	Novel2	ACGTTGGATGAGGTATCCGAACAGGGAAAG	ACGTTGGATGGAGACTCTGTGAAGTAGTCG	TGGCTTCTTTTCAGCAA
iPlex1	rs33997857	ACGTTGGATGCTCCCAATCTTTACCACGAG	ACGTTGGATGGGAGCAAGCTGTGTCATATC	AATAGCGGGGTTGTCCA
iPlex1	rs2577264	ACGTTGGATGCAGCCAAACTCAGAATTCCC	ACGTTGGATGTGTGCTTCACAGGCACAAG	CCATACCCTGGTGCCTTCC
iPlex1	Novel1	ACGTTGGATGAGCGTATCTGTGGAGACTTG	ACGTTGGATGAACCTGAAGGGAAAACATGG	GTCCCCATGAGGTCTCCTGT
iPlex1	rs4669778	ACGTTGGATGTGCTCATTCTTCCCGCCTAC	ACGTTGGATGCGAGAACTCTGAAGTTGTCC	GAAGGTATTCTGGGCCTGTG
iPlex1	rs2577262	ACGTTGGATGTGGTGACAAGCTGTTTGCTC	ACGTTGGATGAGGTGTCAGAATGTTTTCC	CGGGGACTATTCTGTCAATCA
iPlex1	rs13412852	ACGTTGGATGTATGTCCCATGCCTGGTTTC	ACGTTGGATGATGCAGGGACTTTGTCACAG	CCTTCCAGCCAATACTGATGTCG
iPlex1	rs6744695	ACGTTGGATGAGAGAGCAGAGCAACGATTC	ACGTTGGATGTTCCAAGACAAAGGCTCCAC	CAACGATTCAGTTATTAAAGATGC
iPlex1	rs2278513	ACGTTGGATGGAGCAAGCTTTGTCTAAAGG	ACGTTGGATGTGGCTATAAGTTCTTTGCAG	GGATTTCATTTAGTGACAGTGAGA
iPlex1	rs6729430	ACGTTGGATGAAAGGACCGTAGATTTCCTC	ACGTTGGATGAGAGAGAAATGCCAAGGGAC	CCAGGACCGTAGATTTCCTCTCCTGA
iPlex2	rs17603420	ACGTTGGATGGCATGTGTTGTGTTTTGAGG	ACGTTGGATGAACCACTCTGGAGAACATGC	GTGTTTTGAGGACCGCA
iPlex2	rs2577256	ACGTTGGATGATGCTAATATCCCTTCCCCC	ACGTTGGATGGCTAAGCAGTCTCAAAAGCC	CTTCTCCAGGCACTTCAC
iPlex2	rs3795974	ACGTTGGATGACTGTCCCAGTTCAGGGTC	ACGTTGGATGATCTGTGCTGAGATGCAGCC	CCAGTTCAGGGTCCTACATC
iPlex2	rs893346	ACGTTGGATGAGCATAGCAGCCACTGTTTC	ACGTTGGATGACAGGGAAAGAGACCTCAAC	CCCATGGAGTTCTCTCCTCCC
iPlex2	rs16857866	ACGTTGGATGACGTGTGCTGCTGATCATTC	ACGTTGGATGCTGTGTGTCTCAGGTGTAG	GCTGATCATTCAAATAGGTTT
iPlex2	rs7595221	ACGTTGGATGGGTGTATTATCTGTTTGGTTC	ACGTTGGATGCATCCCACTGGAATACCATC	CTCCAACAGTTTCTAACCCAGAG
iPlex2	rs17603350	ACGTTGGATGCAAAAAACCGAGGCTCAGAC	ACGTTGGATGTGAGCTCTTTAAGGGCCTCG	GGTAAAGTAGTCAAGGAAAGCTC
iPlex2	rs1050800	ACGTTGGATGAGTTTCTTTGGAGAGCTCCC	ACGTTGGATGTGCTGACAGTGTGGACTTGA	CAGAAGTCAGTATCATTATCATTT
iPlex3	rs893345	ACGTTGGATGTAGATAGGAGCTGCCCCAAG	ACGTTGGATGGCTTGGAGAGCTTTGTCCAG	CACTTGACCAGATGGAC
iPlex3	rs2716609	ACGTTGGATGATCTCAAGCTTTGACCCACC	ACGTTGGATGTGCCCACATCTAAAGGCAAG	AACCTTCAAAGAACCCTGGA
iPlex3	rs4669781	ACGTTGGATGACATTAGGCAGAAGAGGGAG	ACGTTGGATGTGTGTGTTCCAGGGTAAAGC	GAGGTGGCCTGCGTTTGATG
iPlex3	rs17603755	ACGTTGGATGCCGAAGTCCCATAATTGAGC	ACGTTGGATGTTAGCACTTGTCAGTGTACG	TGTCACCATACAAATAATTAAAATA
iPlex4	rs2577261	ACGTTGGATGAACTTCCTCTCGCCTTACTG	ACGTTGGATGAAATGAACACTTATGGGGTC	ATCTCCTATGCAACCTAC
iPlex5	rs4315495	ACGTTGGATGACTGTAGCATTAATTGTAAT	ACGTTGGATGCCAGAGTCATACAGATATTA	GCATTAATTGTAATAGGGAAAAT
iPlex6*	rs4669781	ACGTTGGATGTGTGTGTTCCAGGGTAAAGC	ACGTTGGATGACATTAGGCAGAAGAGGGAG	AGCGCGCAGCTGCCAAG

Plex	SNP_ID	2nd PCR primer	1st PCR primer	MassEXTEND primer
iPlex6*	rs2577261	ACGTTGGATGAAATGAACACTTATGGGGTC	ACGTTGGATGAACTTCCTCTCGCCTTACTG	CTTATGGGGTCAGAAAGC
iPlex6*	rs17603755	ACGTTGGATGCCGAAGTCCCATAATTGAGC	ACGTTGGATGTTAGCACTTGTCAGTGTACG	TGTCACCATACAAATAATTAAAATA

* = Redesigned for Ely cohort only.

Table A9 *mTOR* sequence variants detected in a cohort of severe insulin resistant patients and 11 Indian and 23 CEPH controls

Genic position	Genomic position	Minor|Major allele	Protein consequence	MAF in SIR	Detected in controls?	rs ID
Upstream	11245215	C|A		0.118		rs2295080
Upstream	11245152	G|C		0.116		rs2295079
5prime_UTR	11245131	A|C		0.013		
Intron 1	11242153	A|G		0.0395		rs17229137
Intron 2	11241394	G|C		0.004		rs28730695
Intron 2	11241375	T|A		0.328		
Intron 2	11241350	G|C		0.003		rs2076657
Intron 3	11241106	A|C		0.008		
Intron 3	11239989	T|C		0.01		
Intron 3	11239920	A|G		0.43		
Intron 3	11239907	G|A		0.011		rs2092642
Intron 3	11239897	A|G		0.01		rs12121319
Intron 4	11239523	A|G		0.011		
Intron 4	11238929	A|G		0.011		
Intron 4	11238912	C|G		0.005		
Intron 4	11238849	T|-		0.005		
Intron 5	11238610	G|A		0.005		
Intron 5	11236649	T|C		0.0026		
Exon 7	11230691	T|C	H296H	0.006		
Intron 7	11230378	A|G		0.024		
Exon 9	**11225807**	**C|G**	**V455L**	**0.004**	**No**	
Intron 9	11225740	T|G		0.0195		rs12141961
Intron 9	11225674	A|G		0.0375		rs28730692
Intron 9	11224410	A|G		0.005		
Exon 10	11224301	C|T	D479D	0.301		rs1135172
Intron 11	11222925	C|T		0.017		rs17848533
Intron 11	11221377	A|G		0.2735		rs2024627
Intron 11	11221326	G|A		0.016		rs12132215
Intron 11	11221279	G|C		0.007		
Intron 12	11220965	A|C		0.007		
Intron 12	11220814	T|G		0.004		
Intron 13	11217061	A|T		0.247		rs718206
Intron 14	11216229	C|T		0.005		rs17229172
Intron 14	11216164	T|C		0.004		rs28730690
Exon 15	11216130	G|C	A778G	0.003	Yes	
Exon 15	11216051	G|A	E804E	0.003		
Intron 15	11215215	G|A		0.041		
Intron 16	11214996	T|C		0.006		
Intron 16	11214865	A|G		0.017		
Intron 17	11213751	-|TGT		0.016		
Intron 18	11213456	T|C		0.011		
Exon 19	11211537	A|G	L935L	0.004		
Exon 19	11211345	C|T	N999N	0.253		rs1064261
Intron 19	11211248	C|G		0.005		
Intron 19	11198940	T|G		0.0275		rs28730687

Genic position	Genomic position	Minor|Major allele	Protein consequence	MAF in SIR	Detected in controls?	rs ID
Intron 19	11198928	A|G		0.0275		rs28730686
Intron 20	11198706	T|-		0.0405		
Intron 21	11196007	C|A		0.008		rs1211695
Intron 21	11196005	C|A		0.026		rs12116957
Intron 21	11195578	T|C		0.004		
Intron 22	11195204	A|G		0.021		
Intron 22	11195180	G|A		0.007		
Exon 23	11195055	C|G	R1154R	0.05		rs17036536
Exon 24	11193482	C|T	D1210D	0.005		
Intron 25	11191890	G|T		0.011		
Intron 25	11187415	G|T		0.024		rs2273127
Intron 26	11182424	A|G		0.004		
Intron 26	11182417	G|T		0.005		rs11121697
Intron 30	11133037	A|C		0.0235		
Intron 30	11132980	G|A		0.0105		
Intron 30	11132924	C|T		0.01		
Intron 30	11132918	A|G		0.005		rs28730683
Intron 32	11129279	A|T		0.005		
Intron 32	11129277	A|T		0.034		rs2275942
Intron 32	11127797	T|C		0.002		
Exon 33	11127645	G|A	A1577A	0.004		rs1057079
Intron 35	11122165	A|C		0.2845		
Intron 35	11122145	A|G		0.011		
Intron 35	11122128	T|C		0.0263		rs17848548
Intron 35	11122105	G|A		0.05		rs28730682
Intron 36	11121736	G|A		0.0265		rs2275528
Intron 36	11121717	-|AT		0.021		
Intron 36	11121582	A|G		0.0455		rs17417751
Intron 36	11121571	T|C		0.006		
Intron 36	11117178	T|C		0.021		rs3730381
Intron 36	11117169	G|C		0.003		
Intron 37	11116856	A|C		0.022		
Intron 37	11115995	A|G		0.0175		rs12122483
Intron 37	11115993	A|G		0.0175		rs17235612
Intron 37	11115988	G|T		0.0215		rs12141233
Intron 38	11115543	C|T		0.019		rs2275526
Intron 38	11115495	C|T		0.0045		
Intron 38	11115376	C|G		0.007		
Intron 38	11115280	C|T		0.005		
Intron 38	11114926	A|G		0.016		
Intron 38	11114721	A|C		0.003		
Intron 38	11114509	G|A		0.235		rs1417131
Intron 38	11114072	G|C		0.005		
Intron 38	11113819	T|C		0.038		rs12732063
Exon 39	11113317	T|C	A1832A	0.044		rs17848553
Exon 39	11113233	T|C	S1851S	0.281		rs2275527
Intron 39	11113114	A|G		0.007		
Intron 39	11112638	A|T		0.046		
Intron 39	11112511	A|G		0.005		
Intron 40	11112314	G|A		0.005		

Genic position	Genomic position	Minor|Major allele	Protein consequence	MAF in SIR	Detected in controls?	rs ID
Intron 40	11111794	T|C		0.005		rs3730378
Intron 40	11111778	T|C		0.029		rs3730379
Intron 40	11111757	A|G		0.004		
Intron 40	11111664	G|A		0.016		rs3730380
Intron 41	11111486	A|G		0.005		
Intron 41	11111469	T|G		0.016		rs17848558
Intron 41	11111292	T|C		0.047		
Intron 41	11111215	A|G		0.004		
Intron 41	11111213	A|G		0.004		
Intron 42	11111081	A|G		0.005		
Exon 43	11110729	A|G	T1984T	0.006		
Intron 43	11110599	A|G		0.005		
Intron 45	11109484	C|T		0.032		rs3737611
Exon 47	11107180	C|T	L2208L	0.007		
Intron 47	11104816	G|A		0.005		
Exon 48	11104650	T|C	L2261L	0.003		
Intron 48	11104572	T|C		0.032		
Intron 48	11104044	A|C		0.0215		rs17235633
Exon 49	11103914	G|A	L2303L	0.1735		rs1112169
Intron 50	11098268	T|C		0.007		
Intron 50	11098129	G|T		0.005		
Intron 52	11097438	T|C		0.01		rs2275523
Intron 53	11096918	G|A		0.0227		rs17235654
Intron 53	11095790	G|A		0.005		rs17229256
Intron 54	11092376	C|T		0.005		
Intron 55	11092263	G|A		0.0705		rs2275525
Exon 56	11092007	T|C	D2485D	0.01		
Intron 56	11090972	G|T		0.005		
Intron 57	11090233	G|A		0.005		
3prime_UTR	11090104	T|C		0.005		
3prime_UTR	11089889	-|A		0.01		
3prime_UTR	11089831	A|G		0.005		
3prime_UTR	11089733	T|C		0.021		rs12139042
3prime_UTR	11089300	G|A		0.019		rs2536

Genomic coordinates are NCBI build 36 (B36).

Table A10 *Rictor* sequence variants detected in a cohort of severe insulin resistant patients and 11 Indian and 23 CEPH controls

Genic position	Genomic position	Minor|Major allele	Protein consequence	MAF in SIR	Detected in controls?	rs ID
Upstream	39110453	T|C		0.3966		rs6869095
Upstream	39110425	T|C		0.003		
Upstream	39110388	G|A		0.0027		
Upstream	39110376	A|G		0.0027		
Upstream	39110365	A|G		0.0027		
Upstream	39110328	C|A		0.0027		
Exon 1	39110229	A|G	A3V	0.0054	No	
Intron 1	39110173	C|T		0.0027		
Intron 1	39110053	C|G		0.3655		rs2303808
Intron 2	39057126	C|T		0.302		rs13177690
Intron 4	39039278	A/-		0.017		
Intron 4	39039245	G|C		0.003		
Intron 4	39038633	G|A		0.003		
Exon 5	39038462	G|A	R108R	0.003		
Intron 5	39032802	A|G		0.021		rs2548796
Intron 8	39017548	G|C		0.004		
Intron 8	39017167	A|G		0.003		
Intron 8	39017144	-|T		0.037		
Intron 8	39017022	G|A		0.006		
Intron 8	39016743	G|A		0.003		
Intron 8	39016629	A|G		0.003		
Intron 8	39016620	T|C		0.007		
Intron 8	39016444	T|G		0.004		
Intron 8	39014597	C|G		0.011		
Intron 9	39014386	T|G		0.001		
Intron 9	39014355	T|C		0.003		
Intron 9	39014354	G|A		0.003		
Intron 9	39011565	C|T		0.007		
Intron 9	39011487	T|C		0.003		
Intron 10	39008092	A|G		0.027		rs10440634
Intron 11	39007669	A|T		0.011		rs7732615
Intron 11	39007666	G|A		0.008		
Intron 11	39006764	A|G		0.003		
Intron 11	39006641	C|G		0.003		
Intron 11	39006506	A|G		0.002		
Intron 11	39006439	G|T		0.182		
Intron 11	39006390	G|A		0.139		
Intron 12	39003781	A|G		0.0026		
Intron 13	39003134	A|G		0.024		rs41271107
Intron 15	39000949	G|A		0.419		
Exon 17	38998810	C|T	I497I	0.003		
Intron 17	38998551	G|A		0.003		
Intron 18	38998300	T|C		0.017		rs7729745
Intron 21	38995551	A|G		0.021		rs2115946
Intron 21	38995205	G|A		0.007		

Genic position	Genomic position	Minor|Major allele	Protein consequence	MAF in SIR	Detected in controls?	rs ID
Exon 23	38994634	G|A	E745E	0.007		
Exon 23	38994622	C|T	N749N	0.007		
Intron 23	38994507	T|C		0.491		rs6868087
Intron 23	38994500	G|A		0.003		
Intron 23	38994406	G|A		0.003		
Intron 24	38993701	G|A		0.003		
Exon 25	38993577	A|T	I811I	0.003		
Intron 25	38991719	C|T		0.008		
Intron 25	38991591	T|G		0.003		
Exon 26	38991553	T|C	S837F	0.299	Yes	rs2043112
Intron 26	38991435	C|G		0.003		
Intron 26	38991401	T|C		0.001		
Intron 26	38991377	A|G		0.004		
Intron 26	38990751	A|G		0.018		rs9292726
Intron 27	38990582	A/-		0.011		
Intron 29	38988818	C|T		0.014		
Exon 31	38986533	C|T	S1058S	0.033		rs2115949
Exon 31	38986023	T|C	G1228G	0.003		
Intron 31	38985522	C|A		0.006		
Exon 33	38982400	C|T	I1442I	0.002		
Intron 34	38981000	C|T		0.004		
Intron 36	38978913	G|A		0.0107		
Intron 36	38978644	T|G		0.0292		
Intron 36	38978564	T|C		0.0026		
Exon 38	38978206	T|C	T1695I	0.017	Yes	
3prime_UTR	38978160	G|T		0.003		
3prime_UTR	38977943	A|C		0.017		
3prime_UTR	38977847	A|G		0.02		rs1975089
3prime_UTR	38977184	-|C		0.011		rs3214763
3prime_UTR	38977096	A|G		0.004		
3prime_UTR	38977035	T|C		0.015		
3prime_UTR	38976615	C|A		0.029		rs443039
3prime_UTR	38976567	A|C		0.014		
3prime_UTR	38975560	C|T		0.003		
3prime_UTR	38975492	T|C		0.019		rs16867885
3prime_UTR	38975361	T|C		0.028		rs637747
3prime_UTR	38975147	G|C		0.004		
3prime_UTR	38974902	T|-		0.006		
3prime_UTR	38974878	T|A		0.003		
3prime_UTR	38974664	T|C		0.003		
3prime_UTR	38974518	G|A		0.003		
3prime_UTR	38974421	C|T		0.053		rs10941413
Downstream	38973706	T|C		0.003		

Genomic coordinates are NCBI build 36 (B36).

Table A11 *GβL* sequence variants detected in a cohort of severe insulin resistant patients and 11 Indian and 23 CEPH controls

Genic position	Genomic position	Minor|Major allele	Protein consequence	MAF in SIR	Detected in controls?	rs ID
5prime_UTR	2196052	-|C		0.386		rs34390502
Intron 2	2196327	T|C		0.421		rs27699
Intron 2	2196331	A|C		0.029		
Intron 3	2196462	G|C		0.366		rs26863
Exon 4	**2196583**	**T|C**	**A88V**	**0.006**	**No**	
Intron 4	2196690	A|G		0.006		
Exon 5	2197106	C|G	P137P	0.372		rs26862
Intron 6	2198188	T|C		0.0277		
Exon 9	2198768	A|G	S289S	0.007		rs11863256
Exon 9	**2198798**	**C|G**	**E299D**	**0.006**	**No**	

Genomic coordinates are NCBI build 36 (B36).

Table A12 *MAPKAP1* sequence variants detected in a cohort of severe insulin resistant patients and 11 Indian and 23 CEPH controls

Genic position	Genomic position	Minor|Major allele	Protein consequence	MAF in SIR	Detected in controls?	rs ID
Intron 2	127472130	C|A		0.0058		rs7039012
Intron 2	127472128	C|T		0.0067		
Intron 2	127472119	G|A		0.0119		
Intron 2	127472034	A|G		0.0268		
Intron 3	127471810	A|G		0.0114		
Intron 4	127452497	G|A		0.4042		rs631287
Exon 5	127387762	T|C	D188D	0.005		
Exon 5	127387705	T|C	I207I	0.005		
Intron 5	127367562	G|T		0.0053		
Exon 6	127361802	T|C	S260N	0.0161	No	
Intron 6	127361648	T|A		0.0283		
Intron 6	127361592	T|C		0.3947		rs562004
Exon 7	127345211	T|C	K302K	0.0069		rs11542134
Intron 7	127323575	A|G		0.0058		
Intron 7	127308724	G|T		0.4843		rs7875713
Exon 8	127286607	A|G	H345H	0.2173		rs2070113
Intron 8	127286442	C|T		0.0054		
Intron 8	127273911	A|G		0.0111		rs10986771
Intron 8	127273723	T|C		0.0063		
Intron 8	127273602	A|G		0.0062		
Intron 8	127273601	C|A		0.0079		
Intron 8	127273327	A|G		0.271		rs7018948
Intron 8	127273302	T|C		0.0112		
Intron 8	127273091	G|A		0.008		
Intron 8	127270219	C|A		0.027		
Intron 9	127270063	C|A		0.009		
Intron 9	127269936	T|G		0.0119		
Intron 10	127246538	C|G		0.0633		rs907501
Intron 10	127246521	A|G		0.0063		
Intron 10	127246488	T|C		0.0053		
Intron 10	127246422	T|C		0.0163		
3prime_UTR	127240959	T|C		0.0053		
3prime_UTR	127240919	A|G		0.0054		rs7864593
3prime_UTR	127240660	C|T		0.0054		
3prime_UTR	127240169	A|G		0.009		
3prime_UTR	127240120	G|C		0.3908		rs12202
3prime_UTR	127239694	T|C		0.3674		rs1129

Genomic coordinates are NCBI build 36 (B36).

Table A13 *AS160* sequence variants detected in a cohort of severe insulin resistant patients and 11 Indian and 23 CEPH controls

Genic position	Genomic position	Minor|Major allele	Protein consequence	MAF in SIR	Detected in controls?	rs ID
Exon 1	74953821	G|C	P28P	0.255		rs7327548
Exon 1	74953603	T|C	A101V	0.075	Yes	
Intron 1	74953398	A|G		0.249		rs9573565
Intron 1	74953322	G|A		0.003		
Exon 2	**74834347**	**A|G**	**R299Q**	**0.0104**	**No**	
Intron 2	74832061	C|T		0.005		
Exon 3	**74831989**	**T|C**	**R363X**	**0.0053**	**No**	
Intron 4	74821464	C|A		0.0058		
Exon 10	**74798404**	**T|A**	**N655Y**	**0.0052**	**No**	
Intron 10	74797597	G|T		0.358		
Intron 10	74796597	G|T		0.3066		rs2297210
Intron 10	74796580	G|A		0.005		
Intron 12	74785128	A|G		0.4821		rs2274757
Exon 13	74785004	G|A	T752A	0.0053	Yes	
Exon 13	**74784903**	**G|C**	**N785K**	**0.0106**	**No**	
Intron 13	74782291	T|C		0.3947		rs2297203
Exon 14	74782217	G|A	V819I	0.14	Yes	rs1062087
Intron 14	74778682	T|C		0.022		rs17064121
Intron 14	74778669	G|A		0.005		
Intron 14	74778668	T|C		0.323		rs2297209
Exon 16	74778517	A|G	K895K	0.005		
Exon 16	74774390	C|T	L967L	0.3333		rs2297208
Intron 16	74771788	T|C		0.4107		rs2297207
Intron 16	74767163	G|A		0.003		
Intron 18	74766937	T|C		0.021		
Exon 19	74764285	T|C	T1147M	0.1	Yes	rs9600455
Intron 19	74764193	A|G		0.3		rs471822
Intron 19	74761316	G|C		0.1636		rs7332292

Genomic coordinates are NCBI build 36 (B36).

Table A14 *Raptor* sequence variants detected in a cohort of severe insulin resistant patients and 11 Indian and 23 CEPH controls

Genic position	Genomic position	Minor|Major allele	Protein consequence	MAF in SIR	Detected in controls?	rs ID
5prime_UTR	76133375	T|C		0.2838		rs11547301
5prime_UTR	76133439	A|G		0.0026		
5prime_UTR	76133741	T|C		0.0026		
5prime_UTR	76133764	A|G		0.1825		rs12602885
5prime_UTR	76133969	A|G		0.0135		
Exon 1	76134114	C|T	F30F	0.0967		
Intron 1	76213908	A|G		0.0687		rs17848677
Exon 2	76214157	G|C	T78T	0.2005		rs17848685
Intron 2	76214241	A|C		0.0027		
Intron 2	76231933	T|C		0.0355		
Intron 2	76231989	G|A		0.0053		
Intron 2	76231998	A|G		0.1078		
Intron 2	76232034	T|C		0.2921		rs6565469
Intron 2	76232105	A|T		0.0026		
Intron 3	76232235	G|C		0.013		
Intron 3	76232272	A|C		0.0026		
Intron 3	76296115	G|A		0.1326		rs9900445
Intron 3	76296185	A|G		0.3844		rs2306690
Intron 3	76296221	T|C		0.0026		
Intron 4	76296527	T|C		0.0026		
Intron 4	76297225	G|A		0.2052		rs12450876
Intron 4	76297351	A|G		0.2026		rs12453034
Intron 4	76297497	C|G		0.0894		rs34461761
Intron 4	76318494	A|G		0.307		rs9896771
Intron 4	76318562	T|C		0.0053		
Intron 4	76318707	A|G		0.0026		
Intron 4	76318751	C|T		0.2827		
Intron 4	76318835	C|A		0.2026		rs12946115
Exon 5	76319006	A|G	G187S	0.0026	No	
Exon 5	76319008	G|C	G187G	0.0053		
Intron 5	76319108	A|G		0.0026		
Intron 5	76319162	A|G		0.0027		
Intron 5	76319173	A|G		0.0027		
Intron 5	76342329	C|T		0.2159		rs12944923
Intron 5	76342351	T|G		0.0026		
Intron 5	76342387	A|G		0.0026		
Exon 6	76342542	G|C	T264T	0.0026		
Intron 6	76342685	A|G		0.0026		
Intron 6	76342690	T|A		0.0081		
Intron 6	76379675	C|A		0.0629		rs17848624
Intron 6	76379814	A|G		0.2708		rs4969429
Intron 7	76410463	G|A		0.3931		rs9911574
Intron 8	76410728	A|G		0.0106		
Intron 9	76411661	A|T		0.1382		rs17848690
Intron 9	76412007	T|C		0.1961		rs9895174

Genic position	Genomic position	Minor|Major allele	Protein consequence	MAF in SIR	Detected in controls?	rs ID
Intron 9	76426092	A|G		0.1082		
Intron 10	76434762	T|C		0.0502		
Intron 10	76434807	G|A		0.3783		rs3829572
Exon 11	76434924	C|T	G423G	0.3617		rs3751945
Exon 11	76434969	A|G	Q438Q	0.1058		rs2589156
Intron 11	76434991	T|C		0.0026		
Intron 11	76435091	T|C		0.0053		
Intron 12	76443970	A|G		0.2154		rs12943041
Intron 12	76444079	G|T		0.2608		rs9895380
Intron 12	76444126	T|C		0.206		rs17848633
Intron 12	76446117	C|T		0.0345		
Intron 13	76468651	G|A		0.4563		rs6420478
Intron 13	76468684	A|G		0.0027		
Intron 13	76468733	A|G		0.0026		
Intron 13	76468759	T|C		0.0026		
Exon 14	76468818	G|A	Q506Q	0.2526		rs2289759
Intron 14	76468905	C|T		0.0052		
Intron 14	76468911	T|C		0.2552		rs2289760
Intron 14	76468946	A|G		0.0026		
Intron 14	76468971	T|C		0.0052		
Intron 14	76468972	A|G		0.0026		rs2289761
Intron 14	76469072	T|C		0.0028		
Intron 14	76469082	A|G		0.4692		rs2248843
Intron 14	76469091	T|C		0.1075		rs12937510
Intron 14	76471640	G|A		0.0055		
Intron 14	76471697	C|T		0.0053		
Exon 15	76471873	A|G	T548T	0.0053		rs34848699
Intron 15	76471913	T|C		0.0027		
Exon 16	76472295	A|G	S590S	0.0058		
Intron 16	76472480	A|G		0.238		rs11656246
Intron 16	76473364	T|C		0.0085		
Exon 17	76473438	T|C	F626F	0.0028		
Exon 17	76473495	T|C	A645A	0.0028		
Intron 17	76473599	T|C		0.2926		rs2016817
Intron 17	76473704	T|C		0.0028		
Intron 17	76473705	A|G		0.2926		rs2289762
Intron 17	76479975	C|G		0.0078		
Intron 17	76480086	G|C		0.1349		rs2289763
Exon 18	76480141	C|T	L670L	0.2526		rs2289764
Exon 18	76480225	G|A	A698A	0.1349		rs2289765
Intron 18	76480328	T|C		0.0052		
Intron 19	76481354	G|C		0.0027		
Intron 20	76482307	T|C		0.0029		
Intron 20	76482328	G|A		0.0143		rs3751943
Intron 21	76497401	A|G		0.335		rs1468029
Intron 21	76497420	A|G		0.1492		rs12936933
Intron 21	76497435	G|A		0.3246		rs1468030
Intron 21	76511083	T|C		0.3345		rs2271602
Exon 22	76511124	T|C	T842T	0.2919		rs2271603
Exon 22	76511182	A|G	A862T	0.0036	No	

Genic position	Genomic position	Minor|Major allele	Protein consequence	MAF in SIR	Detected in controls?	rs ID
Intron 22	76511246	T|C		0.0037		
Intron 22	76511357	T|C		0.3219		rs2271605
Intron 22	76511363	T|C		0.1718		rs2271606
Intron 23	76512142	T|C		0.304		rs7217786
Intron 23	76512156	C|T		0.1715		rs6565491
Intron 23	76512182	A|G		0.0423		rs6565492
Intron 23	76513718	A|G		0.0084		
Exon 24	76513818	A|G	T954T	0.0028		
Intron 24	76513913	T|C		0.0112		
Intron 24	76528840	T|G		0.3453		rs17848659
Exon 25	76528905	T|C	H978H	0.0054		
Exon 25	76528961	A|G	R997H	0.0031	No	
Intron 25	76529088	T|C		0.003		rs8082065
Intron 25	76529118	T|C		0.0059		
Intron 25	76529148	C|T		0.0089		rs7211235
Intron 25	76529243	G|A		0.0031		
Intron 25	76529343	G|A		0.174		rs6420480
Intron 25	76529346	A|G		0.1836		rs908237
Intron 25	76529356	C|A		0.2659		rs17848664
Exon 26	76534153	T|C	A1039A	0.3521		rs1567962
Intron 26	76534210	T|C		0.0026		
Intron 26	76534246	A|G		0.0053		
Intron 26	76534432	C|G		0.3567		rs2138119
Intron 26	76534454	G|C		0.1729		rs2271610
Intron 26	76534487	A|G		0.4486		rs2271611
Intron 26	76534515	T|C		0.4805		rs2271612
Intron 27	76536814	C|T		0.25		rs2138123
Intron 27	76536867	T|C		0.0133		
Intron 27	76537027	A|G		0.0053		
Intron 27	76537072	T|C		0.2005		rs2138124
Intron 27	76537073	A|G		0.0054		
Intron 27	76537150	C|T		0.3005		rs8082382
Intron 27	76537179	G|T		0.2121		rs8082395
Intron 27	76537721	G|T		0.0267		rs35638552
Intron 27	76537773	T|C		0.3972		rs3751940
Exon 28	76537938	A|G	T1122T	0.0026		
Intron 28	76538065	G|T		0.0026		
Intron 28	76538155	A|G		0.0028		
Intron 28	76546009	G|A		0.0135		
Intron 29	76548346	T|C		0.0086		
Intron 29	76548432	A|G		0.0029		rs4436840
Exon 30	76548509	T|C	L1172L	0.0029		
Intron 30	76548629	C|G		0.0174		
Intron 30	76548683	A|G		0.0294		
Intron 30	76548735	A|G		0.0174		rs9889472
Intron 30	76548765	T|C		0.0029		
Intron 30	76549666	A|G		0.311		rs9897319
Intron 30	76549670	A|G		0.0058		
Intron 30	76549691	A|C		0.0087		
Intron 30	76549710	A|G		0.0087		

Genic position	Genomic position	Minor|Major allele	Protein consequence	MAF in SIR	Detected in controls?	rs ID
Intron 30	76549715	C|T		0.0029		
Exon 31	76549792	T|C	R1203R	0.3411		rs9899178
Exon 31	76549870	C|T	S1229S	0.0088		
Intron 31	76549896	T|G		0.0029		
Intron 32	76551241	C|T		0.0068		
Intron 32	76551247	T|C		0.0103		
Intron 32	76551300	T|C		0.0068		
Intron 33	76551484	T|C		0.0035		
Intron 33	76552592	A|C		0.0069		
Intron 33	76552606	A|C		0.1758		rs3751937
3prime_UTR	76552729	T|C		0.0172		
3prime_UTR	76552799	C|G		0.1938		rs3751936
3prime_UTR	76552818	T|C		0.0034		
3prime_UTR	76552901	A|G		0.0034		
3prime_UTR	76552977	T|C		0.0055		
3prime_UTR	76552980	A|G		0.1868		rs11547302
3prime_UTR	76553093	A|C		0.4093		rs3751934
3prime_UTR	76553106	A|G		0.0054		
3prime_UTR	76553134	A|G		0.0027		
3prime_UTR	76553167	A|G		0.0333		
3prime_UTR	76553168	T|A		0.0274		rs3751933
3prime_UTR	76553250	A|G		0.0054		
3prime_UTR	76553338	T|C		0.0053		
3prime_UTR	76553354	T|C		0.0026		
3prime_UTR	76553370	A|G		0.0054		
3prime_UTR	76553468	T|C		0.0053		
3prime_UTR	76553485	G|A		0.0026		
3prime_UTR	76553497	G|A		0.0053		
3prime_UTR	76553509	T|C		0.0053		
3prime_UTR	76553545	A|G		0.0026		
3prime_UTR	76553635	A|G		0.0113		
3prime_UTR	76553680	C|T		0.0027		
3prime_UTR	76553730	T|C		0.0026		
3prime_UTR	76553996	A|G		0.0081		
3prime_UTR	76554009	C|T		0.0882		rs3751932
3prime_UTR	76554056	A|G		0.0027		
3prime_UTR	76554120	G|A		0.0027		
3prime_UTR	76554236	T|C		0.0086		
3prime_UTR	76554312	G|A		0.0028		
3prime_UTR	76554452	C|T		0.4099		rs1062935

Genomic coordinates are NCBI build 36 (B36).

Table A15 mTORC PCR primer sequences

Gene	Genic position	Forward	Reverse	Amplicon length	Annealing temperature
mTOR	Exon 1	GCAATACCAGAGCTGATGGG	CGGGTCCTGAAGCTTCTCTT	476	60
	Exon 2	ATGGATGTGACAGGTTGGGT	CCTTTTTAGAATCCCCTGCC	442	60
	Exon 3	AAGCCACGGGCTTCTTAAA	TCAGATGATTGCAAGTTCAGG	415	60
	Exon 4	TTTCCATGACAACTGGGTCA	GTAGAATGGAGGGTTCACTCG	554	60
	Exon 4	CCACCACACCATGCTAGATTCC	TCATCACCTTTTGCCTTCCTC	440	60
	Exon 5	GAGAGCAAACCAAAACGTGA	TTGCTTTCCTTCCTCTGTCC	478	60
	Exon 6	CCTGGGCAATAGAGCAAGAC	CCCCTCCCCATCACTCTCTA	545	60
	Exon 7	CACTGGCACACCTGAGAGAG	CCACCCTTCACTTTGGAAGTA	499	60
	Exon 8	GCGTTGCTTCAAAGGAAAAG	GGTACAGCTCTCACCAAGGC	404	60
	Exon 9	GCAAAACTAGCAGATTCCCAGTA	TGAATTGCTTTATAGGACCTTTG	550	60
	Exon 10	TCCAGTTGAATGTGCAAACA	GTGGTAGCCCTGCCTGAGTA	343	60
	Exon 11	CTTTCCCTCAAACCATTTGC	CTTGCTGGGAGAGAAAGCTG	554	60
	Exon 12	TATCAGCTTTTTGAGCCGAA	GAGTTCCGGGTCAGTATCGT	475	60
	Exon 13	CGCTTCCTCAAGCAGTTCTC	CGTGTTAGGCATGGCTACTGT	497	60
	Exon 14	CTGAAACACTCCCATCTCCC	CTCTGCTTTAGAGAGGGCCA	472	60
	Exon 15	GCTGGGTTCCGCAATAAATA	TCACATTATGGTGCTTCCCA	433	60
	Exon 16	CTTCATAGGCAGTGGCCAAG	CGGCCTCCAAAATACTTTCT	481	60
	Exon 17	CAAAATTGGAGAGGGACAGG	CAGTGATTGTTGCAGGGTTG	461	60
	Exon 18	TTCCGGGATTCAAGAAGAAA	GGACGATGGCAAGAGAGAGA	499	60
	Exon 19	GCACAATTAAGAAGCTGAAGCA	GACCTTGAACTCCCAACAGTG	487	60*
	Exon 20	TCTCTTCAGCCTTCATGCCT	CCTCCAAACCTAGCCAAGC	470	60
	Exon 21	GGCATCAACCTGTCACTCAG	GAGGAGTTGATCTCGGCG	398	60
	Exon 22	AGAAAGATGGCCTGGGAACT	CCAGCAACGGATTCCTTATC	434	60
	Exon 23	TTCCACGTTCTCTGATGGTG	TGGATCCATGTCATCCAAAA	414	60
	Exon 24	GCCAAAAAGCTACACAAACTCC	ACAGAGCTAGCCACAAACCC	309	60
	Exon 25	GTTTTCGGTTGCCCAGTTTA	TTTGCTAGTTGCATGGTTGC	492	60
	Exon 26	CCAGCCCCTTGATTATTACTTC	CCAGCACAAAGGAGTTGTGA	460	60

Gene	Genic position	Forward	Reverse	Amplicon length	Annealing temperature
	Exon 27	TCTCACCCAGCAGAACAATG	GCACTGGTCTGTATAGCACTGG	480	60
	Exon 28	GTGACTCCAGTGAAGTGGCA	CAAGGTGAGACTTTTCCCCA	373	60
	Exon 29	TGAAGACTGGTAAGTCAGCATCA	CCACAGAGCAGCAAGATGAG	462	60
	Exon 30	AAGTGAGAACTCCGTGTGGG	AGCAGCACTCCTCCAGATTG	412	60
	Exon 31	TACTCTTCTGATGCGAGCCC	GCCCAGAATTCTTTGGCATA	486	60
	Exon 32	GCTTCTTTACCAAAGCACCG	ATAGGGGCTCTGCAGGATTT	360	60
	Exon 33	AAGCAAATCGCATTCCAAAC	TTCAGTTCTGGGAAAATGCC	438	60
	Exon 34	GGCAGGAAAAGCAAGTTGAG	GTTACAGTCGGGCATATGGG	447	60
	Exon 35	GCCTGTAATCCCCGCTACTC	GGGCACTCTTCCACATGT	528	60
	Exon 36	ACAGGAAGAGACTCCCTGACA	CCTGGCTCAAGTATGCAAGC	410	60
	Exon 36	CCCTCCTGGTTTCAGCATTA	TGTTGCTCCTGGGAGTTGAT	562	60
	Exon 37	CCTAGGAAAGCAAGTGGCAG	GGTCCTCTGTCCTATGGGGT	455	60
	Exon 38	ACTTGTAGGGGGAAATGCAG	TGAGAGACAAAGGTCAAGATGG	519	60
	Exon 39	AGCTCCATCTGGCACTTCAG	TCCTTGAAGTGTTTGCTTTATGG	499	60
	Exon 40	GGCCCAGTCAGCTGTTACTC	CCTGTTCAGGAAGCCACTGT	496	60
	Exon 41	GGACACTTGACACTGGGACC	GTGACCTGGGATGAAATGCT	469	60
	Exon 42	AGACTACCAGGAAGGGGCTC	GGGAAGGGAATGAACCATTT	435	60
	Exon 43	GTCAGAGGAAGTGCACAGCA	AGCCCCTTCCTGGTAGTCTC	363	60
	Exon 44	CTTGGGTCACGTCCTTTCAT	TTTTTCTTCCTCGCATTGGT	419	60
	Exon 45	CTGCCTCCAGGGAAGAATTT	ATGCTCTTGCTTTCTCCAGC	462	60
	Exon 46	TTTCAGAAGAGGGAAGGGGT	TGTGTTCCGACGAATCTCAA	496	60
	Exon 47	CATGACTACACGAGACAAATGTAGG	TCATTTCCATGCCTAACCGT	470	60
	Exon 48	TCTCCATATGGCCAGTGCTT	TCTGCCTGTGTTCTGAGCTG	506	60
	Exon 49	TGTCTTGCTCACCCATTTCA	TCTCCTTTGGCTATGATGGG	411	60
	Exon 51	ACAGTGTCAGAGGAGGGGG	TGATTCACACATAAGCCCACA	479	60
	Exon 52	TTTGGAGAACTGATGGGCTC	GTTCCCATGCCCTCTGAGTA	478	60
	Exon 53	CCAAATGGGATCAGGACAGT	CAAGCAACTCCTCTGCCTTG	557	60
	Exon 54	GAGCCTCTGTAACGTCCTGC	CCTTAGGGGATGTTCACCAA	472	60
	Exon 55	CCTTTTCTGCTCAAAGGCAG	TCCCCTTTAGGGTAGGTAGGG	322	60

Gene	Genic position	Forward	Reverse	Amplicon length	Annealing temperature
	Exon 56	TGCCAAAGCTCGTCACTAAC	TGCCTTTGAGCAGAAAAGGT	366	60
	Exon 57	AACAATGGGCACATGCAGTA	TCAAATCCAACTATGATGGTATCC	407	60*
	Exon 58	CCCATGTTGAGAGGAGCAAC	GTGGGCATTAGCCTTGACTG	570	60
	3'UTR	GTGATCCCCTCTGTGCATCT	GAAACACAAGGCTTGATTTGGT	488	60
	3'UTR	CTTCCCTGTGTTCAGCACCT	TGGGTACAGCAAACTCAGCA	437	60
	3'UTR	CACCCTTACGTCCTCTGTCA	GATGCACAGAGGGGATCACT	409	60
Rictor	Exon 1	CACACAGCCCCAATTAGGAT	GCATCAATGGGACTTTCCTC	476	60*
	Exon 3	TCCATGAGAAGCAGGAAAGAA	CCATGGGTACAGGGGAGTAG	439	60
	Exon 4	TTTGAACATTCACTTTGGTTCTTC	CTGGCAGGTCAAATGAGAAA	469	60
	Exon 5	ATGGCAGGCATGCTAAAAAC	CAGAAGGTAAGAAAGAACTCAATAACA	454	60
	Exon 6	TTGCCCTACCTATACCCAATCA	TGACAGATATTTATCCCTCGAGATG	466	60
	Exon 7	TTGACTACTATTGTCCCCTCTGG	TTCATTACTTGAAGAAAATTCTGACC	504	60
	Exon 8	TGTCGGCTAATTTAGGCTTGA	TTAAATATCAGGTTGTGGGGAAGT	486	60
	Exon 8	GCCCTGTTACTTTTGACTAACTGT	CGACAGTATGTGCGAGCTGA	548	60
	Exon 9	CAATGACCTTGGTTTCTGTCA	TCACCAAACATATTCCCTGATG	474	60
	Exon 10	TTTTGTACGAGTTCCTTTGGTTC	AATAGTTTTCAGGGAGCAGGAA	478	60
	Exon 11	AACAACTAATGAGGAAAACACATCA	TCTCCCGTGTTCTTACTCCC	494	60
	Exon 12	TGACAGACAATATGACTGCTGCT	AGAGCCAAGGTGATAAAGAAGAA	529	60
	Exon 13	TGCAGAGAGTATCAGTGCCAA	CCGAAGTCCCAAAATGAAGA	439	60
	Exon 14	GAGGCTTTCAGATGGCTTTG	GGGTTCAAGCAATTCTCGTG	509	60
	Exon 16	TTTCCCACCATAACCTATCCA	AACTGGGTACCACTTAAAACCATAA	488	60
	Exon 17	TGCATGAAGAGAAGTGTAAACTCA	CAATTGCATCTAAATAATCAGTGTGT	476	60
	Exon 18	ACATTTGGCCACTAGAAAAACA	ACATCTTAATTCCTTAACCACTGGA	488	60
	Exon 19	TCCAAAACTAAAAGCAGAATGTAAA	TTTTAGGATACAGAGGAAGCTCTTTT	500	60
	Exon 20	TCATTTCTTATGTTTCCAAATACGG	TTGCCATTTATCCCAATGGT	400	60
	Exon 21	GCTTTTTGGAAAGAAAGCATTTA	TGGATTAGAAAACAACAGGTACTCAC	531	60
	Exon 22	GCAGTTCCTTGGAGAAACTAAAA	TCTTCCCACTCCAGCATTC	499	60
	Exon 23	CCTCAGCAGGAGAAGCAAAC	CAGGCATTCTTCCCTTTTCA	497	60
	Exon 24	TTCAGCGTACTTCCCATTTG	CCTGCAGACTCTATGCAACAA	485	60

Gene	Genic position	Forward	Reverse	Amplicon length	Annealing temperature
	Exon 25	CAGTCAACAGTTCAGAATGAAAGAA	TGTTGTTGCAAAGATTGTGAAAA	498	60
	Exon 26	TGCCACTGAATTTTTGAGCA	CGTGATGGCTTGAAAATGTG	483	60
	Exon 27	ATGCATTGCAAGGCCATTA	AGGGAAGCCCAGCCTATATTT	471	60
	Exon 28	GAAACAAATTTAACCTCATCTCTGA	CAAGAAATTTTACATAAAATTATTTAG	537	55
	Exon 29	AACAGTTATGCATTGCACTTAATCAT	AGACTGGATTGTTCCTGCCA	461	60
	Exon 30	CCAAGAGGCTATTATTGCCAAA	CAACCATTTTGGTGGGTTTT	484	60
	Exon 31	TGGATGCATTCTTTGTTGCT	TCACAGAGAAAATACAAGCCGA	451	60
	Exon 31	GCAGCTGCAAGAATCCATTT	CCAAATCAAATTCGGTGTCC	479	60
	Exon 31	TCCCTCCTTTTGGATCACTG	CCTGTGCCATAGTGGGTCTT	492	60
	Exon 31	TCGGCTTGTATTTTCTCTGTGA	ATATTGGAGGATGACCGGTTT	465	60
	Exon 31	GTGGCGTCAAATAGTGGCTT	AAGACAAGCAACAGGCGAAT	445	60
	Exon 33	TAAAAATCATGCTGCCCCA	TGGTGACCTTTTACATAGCGG	494	60
	Exon 34	TTGATCCAAAAGTGAGTCTTCCTA	CCCAGGAGATACCCACAGAT	500	60
	Exon 35	GATCTGTCAGACTCAGGCCC	ATTTACTTATGCACAGGTCTTCCA	452	60
	Exon 36	AAATTGCTTCACTAAAATGGACG	TCTTTTCCACTGTGAACAGCC	499	60
	Exon 37	TGCAATGCAGTTTAGAACTACTGA	TTGGTTTTGTTTTGTTTTTCG	538	60
	Exon 38	CAGAATGTCTAAGTGTGTGACAGG	TCTTGAATTCCTGGGCTCAA	479	60
	Exon 39	GAGGTCAGGATTCAGCAGATG	GGTGTGGTGACTCATGCCTA	584	60
	Exon 39	TGCATATCCCATTATTGTTCCA	CCCTAGATGCATGAAGAAGCA	428	60
	3'UTR	AACGTGAAAGGGCCAAAGTT	CACTGTTCCGCAGTTACTGGT	410	60
	3'UTR	AGCGATGAGATGGAAAGTTGA	GGATATGCAGAAGCACCCTT	490	60
	3'UTR	ACCCACCTACCACTGCATTT	TCTTCATTTTGTTGCAATGTGTT	457	60
	3'UTR	TTGCTATGACTGTGTCAAAACTGA	CATTCCATGGGGAAAGAAGA	483	60
	3'UTR	TGAAGTGCAGTCAAAGAGGTG	TTGATGAGGGCAAGGTTTTT	464	60
	3'UTR	CCAAGACCAAAGACCACGAT	CCCATGATGCAGATCCTTTT	406	60
	3'UTR	ACAGGGATAGTGATGGTGGG	CAGATAAGGGCTGTAAGCTGC	424	60
	3'UTR	TGCTGATATCATTCCTAACCTCTTC	GACATTCATTTTAAACACTACCTCAG	521	60
	3'UTR	TTTCCCCCTTCCAATCTACC	TCTCCATGATTTCACTGACAGG	533	60
	3'UTR	TCTTACCATGCAAAAATAACTGC	TCAAGAACAGATTTCAGCCTCA	539	60

Gene	Genic position	Forward	Reverse	Amplicon length	Annealing temperature
	3'UTR	CACTGCTATTAATTTCACTGGTATGC	CATATCACAACGGCCAAGTG	527	60
	3'UTR	TGCAGCTTTAGGGTTTCTGA	GGAGCTATTTGGAATGAACCA	546	60
	3'UTR	TCAAGGTGAGCAAAACCATTT	CCACAGCATGCATAAATTGC	506	60
	3'UTR	CCCATTTTCCAAATCCCTTT	TTTGGAGGTTCTGGTGTTTG	530	60
Raptor	5'UTR	GAGACCGCTCTGAGGAGGTT	ACTTCAGGCCTCACGATGAC	518	60
	5'UTR	TGAGAGTGGCTGGAGGAGAC	ATTTCGGACTCCATCAGTGG	512	60
	Exon 1	TGAGTAAGGGTCTCCGCACT	GCATTTTCCATTTTCAACCAA	487	60
	Exon 2	GCCTGAGAATAGCCCTTCCT	GAGCGCATCGGAAATTACTC	522	60
	Exon 3	TTCACTTGCCCGTGATGTTA	GGACGCTGTCAAGACAACAC	507	FAIL
	Exon 4	CCTGGTTTTATAGATGGAGAAACTC	GCAGTCATTCCATGAACACG	514	60
	Intron 4	CCAGCTATGGGTCTCACTGC	CGCCTGGTGGTTCATTACAT	542	60
	Exon 5	GCAGCATCAGAGGAGATGTG	CAGCATTGGAGCAGTCGTAG	577	60
	Exon 5	GCCTTTTGTAACGGTGCAGT	CTGAGTGCCTCCCTAGCTCA	401	60
	Exon 6	CCAGGTTTCCAGGCTTACAG	AGAACCCAAACCTCCACCTG	540	60
	Exon 7	TCTCACGCCTTGACAGAATG	TGGCCCATGTTTGGTTATCT	459	60
	Exon 8	GAGCTGTGAGCTCATGGAGA	AGAACCTGGCGTGCAGTAAC	475	60
	Exon 9	GGTTGATTTGGGGTGTTTGT	CACCTGGAGGATCCTGAAAC	591	60
	Exon 9	TCAGACAGGACTTGCTGGTG	CGATCTTGCATGTGCTTCAT	490	60
	Intron 9	GCTCCCAGAGATCTCCACAC	TGCTTGTTAGGACAGCCACA	490	60
	Exon 10	TGACAAAATGGTGAGGATGC	AAACCCTGGTTGACTCGATG	505	60
	Exon 11	GGAGTCATCCTCAGCTGCTC	AGCCCAGTGACTCCTGATTG	541	60
	Exon 12	GGAAAGGAGCACCAGCTCTA	GGCCGACAACACACTCACT	513	60
	Exon 13	CAGTCAAGGCTGGCAGTATG	GGGACAGGACAGGTAGGTGA	539	60
	Exon 14	AGCCAAGGAAAGCATGTCTG	CCGGGTTAACTGAGCTCTTG	496	60
	Exon 15	GCTTCCAAGGAGCAGTCTGT	TCTTGGGAAGCCTTGATTTG	505	60
	Exon 16	GGCTCTGACCCTTCATCTAGC	CTCACCAAGGTCCCACTGTC	532	60
	Exon 17	TAGTGCGAGGAGAGCCTTG	CCAGGCAGGAAGGAGACTAC	543	60
	Exon 18	TAGCAGTGGGGAGGATTCTG	GGCAGGTGCCTAAAAAGATG	528	60
	Exon 19	CGGTTGTTGTTTGTCCAGTG	TCATCACCATGTTGGCAGAT	486	60

Gene	Genic position	Forward	Reverse	Amplicon length	Annealing temperature
	Exon 20	CGTCTCCCAGTAACCTGTGGT	CTGCTGGTGCTGAGGTTTC	598	60
	Exon 20	AAATATGTATCTCAGTCATGCCATT	TCTCACCCGTCTATGGAACA	550	60
	Exon 21	ACCCGTTGCCAGAATAACAA	GTAAATCCAGCACCCTCTGC	487	60
	Exon 22	CAGTGGATGCATGTCCAAGA	CAAGATGGGGAGAGCTCAAG	533	60
	Exon 23	ACCGAGCAGCATGGACTC	GTCTGCAAAGGCAGAGAGTG	532	60
	Exon 24	ACGTGGCTCACACTTCACAG	TCTGCGGTTCTACCTCAACC	488	60
	Exon 25	AAAGCCTTCAGCCTTTGAGA	GGGTGCCCTTACCCTTCTG	538	60
	Exon 25	GTTTCATCGCTGGGCTTAAC	CTCTCAGAGGCCTTCCCTTC	542	60
	Exon 26	GCTCCTGGATCCACTCTTGT	CTGGGGCACCTAGCATCTT	530	60
	Exon 27	CCTTAAGCTTTTGGGACTGG	GGCATCCTTCCCTGATTCTC	479	60
	Intron 27	TAGCATCAACCGATGTCGAA	TCAGGGCTTAGAATGGAGGA	526	60
	Exon 28	GAGATCTGGAACCACGGGTA	CAAGCCAAGGATGAAGGTTG	520	60
	Exon 29	GTCCCTAGCGGGAGCTCATA	TGCAGTTCACCTTCATCCAG	470	60
	Exon 30	CCTTTGATGCTTCCCTGGAT	TGTGCTCCTCCATAGGCTGT	531	60
	Exon 31	AGCTTCCATGTGGCTCAGAC	ATCTTTCTGGCCAGTTCCAG	513	60
	Exon 32	CAGTATGGGTGCATGTCTGG	GGTGGGCAGACATCTTTTGT	464	60
	Exon 33	TTCACTGCTGTGGGGACATA	AGGGGCCCTAGAGATCCTG	535	60*
	Exon 34	CCCTGCAGAGGACTGGTG	CACAGAACAGGAGAGGGACAG	532	FAIL
	3'UTR	CCTGCTACTCGCTTTTGTCTG	GCTTCCTGCTCAGTGATGCT	512	60
	3'UTR	GCAAGCAGGGACATTTCCTA	TTGCCTGTATGTGAGGAACG	498	60
	3'UTR	CTGTGGTCTCCATGCCTGTG	CATCCACCCAGAAGATGCTC	523	60
	3'UTR	GAGAGGTGAAGGAGCCAGGT	GCCTTTTGACAGGTGGTGAT	528	60
	3'UTR	GGTCCGGAAGGTGTAGAGAG	ACAGTGCCTGCCTGGAGAG	627	60
AS160	Exon 1	CGGCAGTTGGCACAGTTT	CTTCTCTCCCACTCCCAAGGT	775	60*
	Exon 2	GACTGCATGGAGAAGTTCAGC	CCTGGAAGAGCATGGTCCTGT	411	60
	Exon 3	TTGAGGCAGCTGATATTTGG	CGGGGTCAAGCCCACTTATTT	350	60
	Exon 4	GCCACTGCATGAGTTTGCTA	CCGCCTTCTTCTTTTCCCATA	383	60
	Exon 5	TCAACAAAGGCTCACATTGC	CGCAGAGATTGACAGCCCACT	382	60
	Exon 6 and 7	CCTGGATCTCTGCTGCTAGAC	CAACCTGAGAGTACAAAGGGAAGAA	598	60

Gene	Genic position	Forward	Reverse	Amplicon length	Annealing temperature
	Exon 8	CATCTTGGAAAGGGATGGTT	CGGAAATCAGTGAGCCTGTTGA	548	60
	Exon 9	GAGACCACCTGGAACTACCTTTT	CCCATATCGGTGCTGACTTTG	217	60
	Exon 10	AGCTCTGTGTTTTGGCAGGT	CCACAATCCACCTTGACTCCA	379	60
	Exon 11	TGTAAGATCCAGCCCTGAGC	CAAAAAGCTTTGAAACTACTGGTTGA	503	60
	Exon 12	TCCTGAATCCTGATTTGGACA	CGCATGCATTCAGGAGACAAA	197	60
	Exon 13	GTGGCCTGGGTGACTAGTGT	CGGAGTAGGCCCAACTACAGG	561	60
	Exon 14	AGCGGCTTTGTGACATTTTT	CTCCAAAACTGAGTGCGGATA	429	60
	Exon 15	AAAGCTTGGGACCCTTTCAT	CGCAAAATCTGAGCTCCACAA	422	60
	Exon 16	TTAGGGTGGCATGCAGTACA	CTCTGCATATGCCAACAGGTG	396	60
	Exon 17	GGGGTGAGAGAGGGAGAGAC	CATTGAGAAGAGGGCCAGGAT	479	60
	Exon 18	GTGTTTGATGCACATGCAGAT	CGGGGCATCATATGGCTAGAA	301	60
	Exon 19	GCGTGCTTTTTACAAGAGTGG	CCAAGCAAAGCTACATTTTTAGGC	321	60
	Exon 20	GAGCTATCTGTTTTTATGTTGACAGTG	CTTTGTAGCCCCAGAGCACTT	532	60
	Exon 21	CATCGTTTCCTGTAGGTTTTCA	CCATGCAGGCTCTTCCTCTCT	441	60
*MAPKAP1***	5'UTR	GAGGCTCATGGCAGAGTCAT	CTCCTCGACAGCCTCTTCAGT	500	60
	5'UTR	GAATTCATCAGCATGACCCC	CTTCCACCTGCATGTTCTCTG	433	56.4
	5'UTR	GCAGGAGGGACACTGAAGAG	CGTCACTTTCCCCTTCAGCAC	442	60
	5'UTR	GCGCTGTAAGGAGAAGAGGA	CATGATGGACCATTCTTTGGG	474	51
	5'UTR	GTGCTGAAGGGGAAAGTGAC	CTTGGCAGTGGCTTTGATAAG	472	60
	5'UTR	AACATAATCAATCCTCCGCC	CAGAAGGAGAAGAAATCCGGG	494	60
	Exon 1	TACCAGGTCAGCTGGCTTCT	CCTACCGCTCCTTGCTTTCAG	569	60
	Exon 2	CCCCATTAAGCCTATTCCAA	CAAGCTCCACAGAACAGGGAT	549	52.5
	Exon 3	CAATCCCTGTGACCCAGAGT	CTGGTATTTGCTTTTCAGGGG	469	60
	Exon 4	TGGAAACAATACTGAACCCAAA	CGTGGGGCTTTCTGCATGTAT	475	50
	Exon 5	GGCCAAAAGACCACAAGAAA	CAGAGAGGCACCCAGTGCTAA	570	60
	Exon 6	GCTTCACACAATGTGCCACT	CTCCTCAATTTGCTTGCTGTG	493	62.7
	Exon 7	CTGGGATCCTTTTCTTCGC	CTTGGCACCTCAGTATGCACTTA	550	60
	Exon 7	CTTCTCACATATGCCGCTGA	CATCTTGCAGTGAGCCGAGAT	497	60
	Exon 8	GCACTCAAGTGGGAAAAAGG	CCCAAGGACCTGCCATAGTGT	532	60

Gene	Genic position	Forward	Reverse	Amplicon length	Annealing temperature
	Exon 9	GGCACCATAATACAATCCCG	CACTTGTGTGTGCCATTTGGA	492	60
	Exon 10	AGGGTTATTGGGATTGGGAG	CTGGGAAGACTAGGTGCTGCT	492	60
	Exon 11	GCATCGCTTATCAAAGCCAC	CGACTCCTTCCACACGACAGC	491	60
	Vega isoform OTTHUMT00000054086, exon 4	GGAAACCTTTACCTTTGTCCC	CAACAATTGGCTCTAAGCTTGC	547	52
	Vega isoform OTTHUMT00000054091, exon 7	AACCAGACCAGTGAGCGTTT	CACTTGGTTTCGGGGAGAACT	429	56.4
GβL	Exon 2	AACCTACCTGCGCTTCTTGTC	CAGCAATCATGCTGCGGTC	500	56.4
	Exon 3	GCAGCACCAGGACTCCGTA	CCACCTGAGGTCCCAGATCCT	464	61.7
	Exon 4	GAATGCCTTGGAGGTCACAC	CAAGGGGAATAACAAGCTGGC	500	54.2
	Exon 5 and 6	AAGAGGCAGGTCCTTCAGC	CTCCAGGAATAAGCCCACCT	500	54.2
	Exon 7	TGGTGCCCTTCTCCTGTCTA	CACGTCCTCCAGATCTTGCAC	576	60
	Exon 8	ACTAAGATCCCTGCCCACAC	CCAGGTTGTCCGAGGAAGC	500	54.2
	Exon 9	ATCGTCACTGGTGAGCCC	CGTCCAGGAGTGTGCAGCC	499	58.9

Primers are M13-tailed. Forward primer sequences are preceded by TGTAAAACGACGGCCAGT and reverse sequences are preceded by AGGAAACAGCTATGACCAT. * 7-deaza-dGTP was added to dNTP mix. ** *MAPKAP1* comprises several isoforms. All exons numbers refer to those in OTTHUMT00000054090 except where specified.

Table A16 *WFS1* variants detected by sequencing 1235 type 2 diabetes cases and 1668 controls

Genic position	Genomic position	Nucleotide substitution (minor|major)	Protein consequence	MAF	SNP ID
Upstream	6316152	C|T		0.0722	rs11726771
Upstream	6316308	T|C		0.0001	
Upstream	6316341	A|C		0.0001	
Upstream	6316361	T|G		0.0063	
Upstream	6316417	A|T		0.0003	
Upstream	6316465	A|G		0.0001	
Upstream	6316475	T|C		0.0006	
Upstream	6316476	A|G		0.0001	
Upstream	6316614	A|G		0.0001	
Upstream	6316637	C|G		0.001	
Upstream	6321827	T|C		0.0004	
Upstream	6321914	G|T		0.0001	
Upstream	6321944	A|T		0.1856	rs4320200
Upstream	6321972	T|C		0.1835	rs13107806
Upstream	6321981	T|C		0.0001	
Upstream	6322051	C|G		0.1856	rs13127445
Upstream	6322203	G|T		0.0003	
Upstream	6322207	A|C		0.0001	
Upstream	6322317	G|T		0.1781	rs4273545
Upstream	6322384	G|A		0.0001	
Upstream	6322420	T|G		0.0001	
Upstream	6322429	A|C		0.0001	
Upstream	6322436	C|A		0.0001	
5'UTR	6322518	T|C		0.002	
5'UTR	6322519	G|A		0.0004	
5'UTR	6322527	T|C		0.0001	
5'UTR	6322576	G|A		0.0001	
5'UTR	6322580	T|C		0.0001	
5'UTR	6322593	G|C		0.0004	
5'UTR	6322609	A|G		0.0001	
Intron 1	6322673	T|G		0.0001	
Intron 1	6322727	T|C		0.0331	rs6830765
Intron 1	6324924	G|A		0.003	
Intron 1	6324986	A|G		0.0001	
Intron 1	6324997	C|T		0.0001	
Intron 1	6325007	G|A		0.0001	
Intron 1	*6325055*	*A|T*		*0.0003*	
Intron 1	*6325091*	*A|G*		*0.0003*	
Intron 1	6325206	G|T		0.0003	
Intron 1	6325258	G|T		0.0003	
Intron 1	6325330	A|G		0.0001	
Intron 1	6325386	A|G		0.0003	rs7657752
Intron 1	6325387	C|T		0.0024	
Intron 1	6325978	T|G		0.0003	
Intron 1	6326039	C|T		0.0003	
Intron 1	6326040	A|G		0.0001	

Genic position	Genomic position	Nucleotide substitution (minor|major)	Protein consequence	MAF	SNP ID
Intron 1	6326063	T|G		0.0003	
Intron 1	6326161	C|G		0.0001	
Intron 1	6326178	G|A		0.0003	
Intron 1	6326225	A|G		0.0001	
Intron 1	6326253	A|G		0.0023	
Intron 1	6326269	A|G		0.0001	
Intron 1	6326276	T|C		0.0001	
Intron 1	6326376	A|C		0.0001	
Intron 1	6329948	G|A		0.0087	rs10937714
Intron 1	6330010	A|G		0.0003	
Exon 2	6330104	A|C	P7P	0.0003	
Exon 2	6330124	C|T	Q14R	0.0001	
Exon 2	6330139	A|G	P19L	0.0003	
Exon 2	6330151	A|G	A23V	0.0001	
Exon 2	6330160	T|C	R26Q	0.0001	
Exon 2	6330207	A|G	R42X	0.0001	
Exon 2	6330212	A|T	A43A	0.0001	
Exon 2	6330213	T|G	P44T	0.0003	
Exon 2	6330215	T|G	P44P	0.0003	
Exon 2	6330314	A|G	T77T	0.0001	
Exon 2	6330315	T|C	G78R	0.0001	
Intron 2	6330358	T|C		0.0003	
Intron 2	6330360	G|A		0.0001	
Intron 2	6330363	A|G		0.0001	
Intron 2	6330375	T|C		0.0001	
Intron 2	6330405	C|T		0.3429	rs28420833
Intron 2	6330455	A|G		0.0001	
Intron 2	6330456	T|C		0.0001	
Intron 2	6330540	T|G		0.0001	
Intron 2	6339429	T|C		0.0003	
Intron 2	6339464	T|C		0.0026	
Intron 2	6339486	C|G		0.0003	
Intron 2	6339501	T|C		0.0001	
Intron 2	6339641	A|G		0.0006	
Intron 2	6339648	A|G		0.0001	
Intron 2	6339692	T|G		0.0001	
Intron 3	6339815	T|C		0.0001	
Intron 3	6339847	T|C		0.0001	
Intron 3	6339866	A|G		0.0001	
Intron 3	6339907	T|C		0.0105	
Intron 3	6339923	T|C		0.0001	
Intron 3	6339931	C|G		0.0001	
Intron 3	6339979	A|C		0.0001	
Intron 3	6340039	A|T		0.002	
Intron 3	6341451	C|T		0.0007	
Intron 3	6341495	A|G		0.3016	rs4688989
Intron 3	6341578	A|G		0.0054	rs4688990
Intron 3	6341579	T|C		0.0006	
Exon 4	6341699	T|C	A134T	0.0004	
Exon 4	6341701	T|C	A134A	0.0003	

Genic position	Genomic position	Nucleotide substitution (minor|major)	Protein consequence	MAF	SNP ID
Intron 4	6341787	A|G		0.0001	rs7688426
Intron 4	6341904	G|C		0.2982	rs4689394
Intron 4	6343810	G|A		0.2139	rs9998519
Intron 4	6343816	T|C		0.1497	rs10010131
Exon 5	6343835	A|G		0.0001	
Exon 5	6343846	T|C	R161Q	0.0001	
Exon 5	6343905	A|G	L181L	0.0001	
Exon 5	6343928	T|G	N188K	0.0001	
Exon 5	6343941	G|T	K193Q	0.0039	
Intron 5	6344085	C|G		0.0001	
Intron 5	6344132	C|T		0.0004	
Intron 5	6344138	G|C		0.3889	rs9998835
Intron 5	6344150	T|C		0.0001	
Intron 5	6344159	G|C		0.0001	
Intron 5	6344161	A|T		0.0001	
Intron 5	6344210	T|C		0.0001	
Intron 5	6344250	A|T		0.001	
Intron 5	6344251	A|G		0.2425	rs10012946
Intron 5	6344302	C|T		0.0001	
Intron 5	6344339	C|G		0.0011	
Intron 5	6344347	A|G		0.1577	rs13101355
Intron 5	6344375	T|C		0.3827	rs13147655
Intron 5	6344439	A|G		0.0001	
Intron 5	6344454	A|G		0.0001	
Intron 5	6344456	C|G		0.0001	
Intron 5	6344495	A|G		0.0001	
Exon 6	6344567	A|G	P218P	0.0001	
Exon 6	6344596	T|C	R228H	0.0009	
Exon 6	6344597	G|C	R228R	0.3081	rs7672995
Exon 6	6344608	T|C	R232H	0.0001	
Intron 6	6344641	T|C		0.0017	
Intron 6	6344703	G|A		0.112	rs7655482
Intron 6	6344717	A|G		0.0001	
Intron 6	6344730	T|C		0.0024	
Intron 6	6344739	T|C		0.0328	rs11729672
Intron 6	6344756	G|A		0.0641	rs11725494
Intron 6	6344806	A|G		0.0001	
Intron 6	6344820	G|C		0.0232	rs11725500
Intron 6	6344868	C|T		0.0082	rs4416547
Intron 6	6347348	A|C		0.0115	rs12511742
Intron 6	6347438	T|C		0.0066	
Intron 6	6347519	T|C		0.0001	
Intron 6	6347522	T|C		0.0004	
Intron 6	6347524	C|G		0.0001	
Intron 6	6347582	T|C		0.007	
Intron 6	6347602	T|C		0.0001	
Intron 6	6347603	A|G		0.0001	
Exon 7	6347682	A|G	I242I	0.0001	
Exon 7	6347684	A|G	A243V	0.0003	
Exon 7	6347685	T|C	A243A	0.0001	

Genic position	Genomic position	Nucleotide substitution (minor|major)	Protein consequence	MAF	SNP ID
Exon 7	6347755	T|C	D267N	0.0007	
Exon 7	6347810	T|C	R285H	0.0004	
Exon 7	6347813	G|A	L286P	0.0001	
Intron 7	6347845	T|C		0.0003	
Intron 7	6347856	A|G		0.0003	
Exon 8	6353353	G|A	A310A	0.0001	
Exon 8	6353386	C|G	T321T	0.0001	
Exon 8	6353398	T|C	N325N	0.0003	
Exon 8	6353400	T|C	A326V	0.0004	
Exon 8	6353402	T|C	L327F	0.0001	
Exon 8	6353408	A|T	F329I	0.0003	
Exon 8	6353420	G|A	V333I	0.2613	rs1801212
Exon 8	6353431	G|C	L336L	0.0001	
Exon 8	6353446	T|C	F341F	0.0718	
Exon 8	6353447	A|G	A342T	0.0003	
Exon 8	6353487	G|T	I355S	0.0001	
Exon 8	6353502	A|G	C360Y	0.0004	
Exon 8	6353548	T|C	R375R	0.0001	
Exon 8	6353557	A|C	T378T	0.0004	
Exon 8	6353571	A|G	R383H	0.0001	
Exon 8	6353576	A|G	E385K	0.0004	
Exon 8	6353608	C|T	V395V	0.3809	rs1801206
Exon 8	6353614	T|C	F397F	0.0003	
Exon 8	6353669	G|A	I416V	0.0001	
Exon 8	6353688	T|C	A422V	0.0004	
Exon 8	6353698	T|C	D425D	0.0001	
Exon 8	6353700	A|G	C426Y	0.0007	rs35218685
Exon 8	6353717	G|C	L432V	0.005	rs35031397
Exon 8	6353720	A|G	A433T	0.0001	
Exon 8	6353725	A|C	V434V	0.0001	
Exon 8	6353731	T|C	T436T	0.0083	
Exon 8	6353739	G|T	F439C	0.0003	
Exon 8	6353782	T|C	P453P	0.0001	
Exon 8	6353788	A|G	T455T	0.0001	
Exon 8	6353790	A|G	R456H	0.0417	rs1801208
Exon 8	6353794	T|G	R457S	0.0004	
Exon 8	6353795	C|G	A458P	0.0004	
Exon 8	6353815	T|C	T464T	0.0001	
Exon 8	6353822	T|C	L467L	0.0004	
Exon 8	6353893	T|C	T490T	0.0001	
Exon 8	6353903	A|G	G494S	0.0001	
Exon 8	6353918	T|C	L499F	0.0007	
Exon 8	6353923	C|T	N500N	0.3161	rs1801214
Exon 8	6353929	T|C	S502S	0.0001	
Exon 8	6353975	G|A	M518V	0.0001	
Exon 8	6354003	T|C	T527I	0.0001	
Exon 8	6354020	T|C	P533S	0.0014	
Exon 8	6354031	A|G	V536V	0.0001	
Exon 8	6354055	T|C	S544S	0.0001	
Exon 8	6354056	A|G	V545M	0.0001	

Genic position	Genomic position	Nucleotide substitution (minor|major)	Protein consequence	MAF	SNP ID
Exon 8	6354067	A|G	L548L	0.0001	
Exon 8	6354068	A|C	L549L	0.0001	rs1801211
Exon 8	6354080	A|G	G553S	0.0003	
Exon 8	6354096	A|G	R558H	0.0001	
Exon 8	6354098	A|G	A559T	0.0037	
Exon 8	6354105	G|T	I561S	0.0004	
Exon 8	6354106	T|C	I561I	0.0006	
Exon 8	6354107	A|G	G562S	0.0001	
Exon 8	6354148	T|C	A575A	0.0562	rs2230719
Exon 8	6354149	A|G	G576S	0.0006	rs1805069
Exon 8	6354151	A|C	G576G	0.0001	
Exon 8	6354166	T|C	G581G	0.0001	
Exon 8	6354183	A|G	R587Q	0.0001	
Exon 8	6354199	T|G	L592L	0.0001	
Exon 8	6354220	A|C	V599V	0.0004	
Exon 8	6354223	T|C	T600T	0.0001	
Exon 8	6354228	T|C	A602V	0.0004	rs2230720
Exon 8	6354229	A|G	A602A	0.0003	
Exon 8	6354255	G|A	R611H	0.2863	rs734312
Exon 8	6354262	A|G	W613X	0.0001	
Exon 8	6354306	T|C	T628M	0.0001	
Exon 8	6354308	T|C	R629W	0.0001	
Exon 8	6354346	A|G	T641T	0.0001	
Exon 8	6354435	T|C	A671V	0.0006	
Exon 8	6354436	A|G	A671A	0.0001	
Exon 8	6354442	T|C	C673C	0.0006	
Exon 8	6354443	A|G	G674R	0.0001	
Exon 8	6354449	T|C	R676C	0.0006	
Exon 8	6354450	A|G	R676H	0.0003	
Exon 8	6354451	T|C	R676R	0.0001	
Exon 8	6354475	A|G	A684A	0.0026	
Exon 8	6354476	T|C	R685C	0.0001	
Exon 8	6354477	A|G	R685H	0.0003	
Exon 8	6354522	A|G	W700X	0.0001	
Exon 8	6354530	T|C	R703C	0.0003	
Exon 8	6354545	T|C	R708C	0.0003	
Exon 8	6354547	T|C	R708R	0.0011	
Exon 8	6354560	A|G	D713N	0.0001	
Exon 8	6354571	T|C	A716A	0.0001	
Exon 8	6354572	A|G	E717K	0.0001	
Exon 8	6354580	T|C	A719A	0.0004	
Exon 8	6354581	G|A	I720V	0.0006	rs1805070
Exon 8	6354607	T|C	G728G	0.0006	
Exon 8	6354628	T|C	Y735Y	0.0001	
Exon 8	6354631	T|C	G736G	0.0003	
Exon 8	6354636	A|C	A738D	0.0001	
Exon 8	6354661	A|C	N746K	0.0003	
Exon 8	6354709	A|G	K762K	0.0001	
Exon 8	6354720	C|A	H766P	0.0001	
Exon 8	6354733	T|C	F770F	0.0001	rs34384569

Genic position	Genomic position	Nucleotide substitution (minor|major)	Protein consequence	MAF	SNP ID
Exon 8	6354737	T|C	R772C	0.0004	
Exon 8	6354745	A|G	K774K	0.0669	rs2230721
Exon 8	6354750	T|A	E776V	0.0046	
Exon 8	6354758	A|G	V779M	0.0006	
Exon 8	6354782	A|G	A787T	0.0001	
Exon 8	6354792	G|C	S790W	0.0001	
Exon 8	6354793	A|G	S790S	0.0003	
Exon 8	6354808	C|G	E795D	0.0001	
Exon 8	6354815	A|G	V798I	0.0001	
Exon 8	6354847	T|C	S808S	0.0003	
Exon 8	6354856	G|A	K811K	0.3936	rs1046314
Exon 8	6354875	T|C	R818C	0.0046	rs35932623
Exon 8	6354889	T|C	L822L	0.0001	
Exon 8	6354892	T|C	I823I	0.0006	rs1801215
Exon 8	6354917	T|C	R832C	0.0003	
Exon 8	6354923	A|G	G834S	0.0003	
Exon 8	6354986	C|T	S855P	0.0001	
Exon 8	6354987	T|C	S855L	0.0003	
Exon 8	6354988	A|G	S855S	0.2484	rs1046316
Exon 8	6355004	A|G	V861M	0.0001	
Exon 8	6355012	T|C	I863I	0.0011	
Exon 8	6355019	A|G	D866N	0.0001	rs3821945
Exon 8	6355034	A|G	V871M	0.0129	
Exon 8	6355045	T|C	A874A	0.0001	
Exon 8	6355046	A|G	V875M	0.0003	
Exon 8	6355054	T|C	F877F	0.0001	
Exon 8	6355061	A|G	D880N	0.0001	
Exon 8	6355077	T|C	P885L	0.0001	
Exon 8	6355089	T|C	A889V	0.0001	
3'UTR	6355100	G|T		0.0001	
3'UTR	6355126	T|C		0.0001	
3'UTR	6355143	T|C		0.3131	rs1046317
3'UTR	6355150	G|A		0.0003	
3'UTR	6355165	T|C		0.0001	
3'UTR	6355178	G|C		0.0004	
3'UTR	6355186	A|G		0.0688	rs1802453
3'UTR	6355187	C|T		0.2659	rs1046319
3'UTR	6355218	A|G		0.0001	
3'UTR	6355227	T|C		0.0199	
3'UTR	6355245	G|A		0.303	rs1046320
3'UTR	6355321	G|A		0.0003	
3'UTR	6355337	C|T		0.0006	
3'UTR	6355349	A|G		0.051	rs1046322
3'UTR	6355370	G|A		0.0526	rs1046325

Genomic coordinates are NCBI build 36 (B36). Intron 1 SNPs in italics are conserved and within highly conserved regions.

Table A17 Rare (MAF<0.01) nonsense and missense changes detected in *WFS1* by sequencing 1235 type 2 diabetes cases and 1668 controls

Chr:base	Variant	rs ID	Biochemical/genetic evidence	Pdel*	SIFT	PolyPhen	MAF in cases	MAF in controls	Conservation
4:6330124	Q14R		Unknown	0.16483	affects protein	benign	0.000405	0	Low
4:6330151	A23V		Novel	0.43978	affects protein	benign	0	0.0003	Low
4:6330160	R26Q		Novel	0.32096	affects protein	benign	0	0.0003	Low
4:6330207	R42X		Novel				0	0.0003	STOP
4:6330213	P44T		Novel	0.40278	affects protein	benign	0	0.0003	Low
4:6330315	G78R		Novel	0.46078	affects protein	benign	0.000405	0	Low
4:6341699	A134T		Novel	0.30575	affects protein	possibly damaging	0	0.0006	Complete
4:6343846	R161Q		Associated with hearing loss	0.12633	tolerated	possibly damaging	0	0	Low
4:6343928	N188K		Novel	0.43033	affects protein	possibly damaging	0	0	Low
4:6343941	K193Q		Associated with earing loss (HL)	0.30458	tolerated	possibly damaging	0.003239	0.003297	Complete
4:6344596	R228H		Novel	0.50948	tolerated	possibly damaging	0.000405	0.000899	Low
4:6344608	R232H		Novel	0.55092	tolerated	possibly damaging	0.000405	0	Low
4:6347684	A243V		Novel	0.3299	tolerated	benign	0	0.0003	Low
4:6347755	D267N		Novel	0.3789	tolerated	benign	0	0.001499	Low
4:6347810	R285H		Novel	0.50948	tolerated	benign	0.000405	0.0006	Low
4:6347813	L286P		Novel	0.31806	tolerated	benign	0	0	Low
4:6353400	A326V		Associated with psychiatric disorders	0.40939	affects protein	benign	0.000405	0	Vertebrate
4:6353402	L327F		Novel	0.55726	affects protein	possibly damaging	0.000405	0	Vertebrate

Chr:base	Variant	rs ID	Biochemical/genetic evidence	Pdel*	SIFT	PolyPhen	MAF in cases	MAF in controls	Conservation
4:6353408	F329I		Novel	0.2868	affects protein	probably damaging	0.00081	0	Vertebrate
4:6353447	A342T		Novel	0.67015	tolerated	benign	0.000405	0.0003	Vertebrate
4:6353487	I355S		Novel	0.41045	tolerated	probably damaging	0	0.0003	Low
4:6353502	C360Y		Novel	0.77597	affects protein	probably damaging	0.00081	0	Vertebrate
4:6353571	R383H		Novel	0.59158	tolerated	benign	0	0	Low
4:6353576	E385K		Polymorphism	0.36132	affects protein	possibly damaging	0.000405	0.0003	Vertebrate
4:6353669	I416V		Novel	0.12475	tolerated	benign	0.000405	0	Vertebrate
4:6353688	A422V		Novel	0.54956	tolerated	benign	0.000405	0.0006	Low
4:6353700	C426Y	rs35218685	Associated with psychiatric disorders	0.63713	tolerated	probably damaging	0.00081	0.0003	Low
4:6353717	L432V	rs35031397	Associated with psychiatric disorders and HL	0.42538	affects protein	benign	0.005263	0.005396	Complete
4:6353720	A433T		Novel	0.42841	tolerated	benign	0	0.0003	Low
4:6353739	F439C		Novel	0.59591	affects protein	probably damaging	0.000405	0.0003	Vertebrate
4:6353794	R457S		Unknown	0.56262	tolerated	possibly damaging	0.000405	0.0003	Low
4:6353795	A458P		Novel	0.53226	tolerated	possibly damaging	0.000405	0.0006	Vertebrate
4:6353903	G494S		Novel	0.40523	affects protein	possibly damaging	0	0.0003	Complete
4:6353918	L499F		Suicide	0.6993	tolerated	benign	0.000405	0.001199	Low
4:6353975	M518V		Novel	0.31466	affects protein	probably damaging	0	0	Complete
4:6354003	T527I		Novel	0.37986	tolerated	possibly damaging	0	0.0003	Low
4:6354020	P533S		Suicide	0.12494	affects protein	probably damaging	0.00081	0.002098	Complete
4:6354056	V545M		Novel	0.72875	tolerated	benign	0	0.0003	Vertebrate

Chr:base	Variant	rs ID	Biochemical/genetic evidence	Pdel*	SIFT	PolyPhen	MAF in cases	MAF in controls	Conservation
4:6354080	G553S		Novel	0.40523	tolerated	benign	0.000405	0.0003	Low
4:6354096	R558H		WS	0.61604	affects protein	possibly damaging	0.000405	0	Complete
4:6354098	A559T		WS and psychiatric disorders	0.34899	tolerated	benign	0.004858	0.002998	Low
4:6354105	I561S		Novel	0.44338	affects protein	possibly damaging	0	0.0006	Low
4:6354107	G562S		Unknown	0.28935	tolerated	benign	0.000405	0	Low
4:6354149	G576S	rs1805069	Associated with psychiatric disorders	0.38948	tolerated	benign	0.00081	0.0003	Vertebrate
4:6354183	R587Q		Novel	0.42654	tolerated	benign	0.000405	0	Low
4:6354228	A602V	rs2230720	Polymorphism	0.10473	tolerated	benign	0.00081	0.0003	Low
4:6354262	W613X		WS				0	0	STOP
4:6354306	T628M		Novel	0.74256	affects protein	possibly damaging	0.000405	0	Vertebrate
4:6354308	R629W		WS & reduces half-life of wolframin	0.87775	affects protein	probably damaging	0.000405	0	Low
4:6354435	A671V		WS and psychiatric disorders	0.21811	tolerated	benign	0.001215	0	Low
4:6354443	G674R		Polymorphism	0.60981	affects protein	probably damaging	0.000405	0	Vertebrate
4:6354449	R676C		Novel	0.75663	affects protein	probably damaging	0.000405	0.0003	Low
4:6354450	R676H		Novel	0.54195	tolerated	possibly damaging	0	0.0003	Low
4:6354476	R685C		Polymorphism	0.84434	affects protein	probably damaging	0.000405	0	Low
4:6354477	R685H		Novel	0.68545	affects protein	benign	0.000405	0	Low
4:6354522	W700X		WS & reduces half-life of wolframin				0	0	
4:6354530	R703C		Novel	0.10889	affects protein	probably damaging	0	0.0006	Low
4:6354545	R708C		WS	0.77798	affects protein	probably damaging	0	0.0006	Vertebrate
4:6354560	D713N		Novel	0.41237	tolerated	benign	0.000405	0	Low

Chr:base	Variant	rs ID	Biochemical/genetic evidence	Pdel*	SIFT	PolyPhen	MAF in cases	MAF in controls	Conservation
4:6354572	E717K		WS and psychiatric disorders	0.32653	tolerated	benign	0	0.0003	Low
4:6354581	I720V	rs1805070	Association with T1D	0.12475	tolerated	benign	0.001215	0.0003	Complete
4:6354636	A738D		Novel	0.49144	tolerated	benign	0	0	Low
4:6354661	N746K		Novel	0.24923	tolerated	possibly damaging	0.000405	0	Vertebrate
4:6354720	H766P		Novel	0.53195	tolerated	probably damaging	0.000405	0	Vertebrate
4:6354737	R772C		Associated with psychiatric disorders	0.91098	affects protein	probably damaging	0	0.000899	Low
4:6354750	E776V		WS	0.49302	affects protein	probably damaging	0.004049	0.006295	Complete
4:6354758	V779M		Associated with HL	0.65969	tolerated	benign	0.000405	0.000899	Low
4:6354782	A787T		Unknown	0.17367	tolerated	benign	0	0.0003	Low
4:6354792	S790W		Novel	0.71718	affects protein	possibly damaging	0	0.0003	Low
4:6354808	E795D		Novel	0.31774	tolerated	benign	0	0.0003	Low
4:6354815	V798I		Novel	0.13624	tolerated	benign	0.000405	0	Low
4:6354875	R818C	rs35932623	WS and psychiatric disorders	0.68043	affects protein	possibly damaging	0.004858	0.005396	Low
4:6354917	R832C		Novel	0.75614	affects protein	probably damaging	0.000405	0.0003	Low
4:6354923	G834S		Novel	0.44336	tolerated	possibly damaging	0.000405	0.0003	Vertebrate
4:6354986	S855P		Novel	0.69611	tolerated	benign	0	0.0003	Low
4:6354987	S855L		Novel	0.47396	tolerated	benign	0.000405	0.0003	Low
4:6355004	V861M		Novel	0.17054	affects protein	benign	0.000405	0	Vertebrate
4:6355019	D866N	rs3821945	Associated with psychiatric disorders	0.39863	tolerated	benign	0.000405	0	Low
4:6355046	V875M		Novel	0.3126	tolerated	benign	0.000405	0.0003	Low
4:6355061	D880N		Novel	0.49211	affects protein	possibly damaging	0	0.0003	Vertebrate

Chr:base	Variant	rs ID	Biochemical/genetic evidence	Pdel*	SIFT	PolyPhen	MAF in cases	MAF in controls	Conservation
4:6355077	P885L		WS & reduces half-life of wolframin	0.54691	affects protein	probably damaging	0	0.0003	Complete
4:6355089	A889V		Novel		affects protein	benign	0	0.0003	Vertebrate

Genomic coordinates are NCBI build 36 (B36). * = Probability that the variation is deleterious according to PANTHER.

Table A18 *WFS1* PCR primer sequences

Gene	Genic position	Forward	Reverse	Amplicon length	Annealing temperature
WFS1	Upstream	ATAGACCCTCGCCCTCATCT	CGCCAATCACTCCACATCTCC	604	
WFS1	Exon 1*	GGAGTGGTGGATGAAAGGTG	CACAGTGCCTCAGGGCTCTT		50
WFS1	Intron 1	GGACCTACCCTCAGTCCACA	CGGCAGGATTTAAAGCAGCAG	576	61.7
WFS1	Intron 1	GTCTGAGTGGATGCATGGTG	CCTGGTCATGGTTGGTCACAG	557	58.9-62.7
WFS1	Exon 2	TGCAAATGGAAGCTGTGAAG	CGTGAGTGTCCTCCCATGGTT	718	61.7
WFS1	Exon 3	AGTGTATGTTGAGGGGTGGC	CGGTCATGGGCTGTCAAACTT	790	58.9-64.7
WFS1	Exon 4	ACCCAGGATCGATGTCTCAG	CTACCAGTACCAGTCGGAGCC	703	54.2
WFS1	Exon 5	TCCTACGTTCTGGGGATCAT	CTCCCTGGTAACCAAGTCCTG	743	50
WFS1	Exon 6	CTGGCACATCATGGTTTCTG	CGCCCTAGTGGTGAGGTGTGT	728	50-54.2
WFS1	Exon 7	CCCGAGGACACATCCTTATG	CGCTGAAGAGGAGGGTGTCAG	703	54.2-56.4
WFS1	Exon 8	CTCGTTCCCACGTACCATCT	CCTTCAGGTAGGGCCAATTCA	614	61.7
WFS1	Exon 8	TCATCACCGGCTTCTTTACC	CACACCAGGATGAGCTTGACC	615	52.5
WFS1	Exon 8	GTCTGTAGTGTGCCCCTGCT	CGATGGTGCTGAACTCGATGA	678	50
WFS1	Exon 8	TTCGACCGCTACAAGTTTGA	CGGGAGGAGAGGGAATCTCAT	698	54.2

Primers are M13-tailed. Forward primer sequences are preceded by TGTAAAACGACGGCCAGT and reverse sequences are preceded by AGGAAACAGCTATGACCAT. * 7-deaza-dGTP was added to dNTP mix.

Table A19 *WFS1* Sequenom primer sequences

Plex	SNP_ID	2nd PCR primer	1st PCR primer	MassEXTEND primer
iPLEX1	WFS1_S855P	ACGTTGGATGTTCGAGCTCAAGGCCATCAG	ACGTTGGATGCCAGTCGTGCTCGATCTTCA	CTGCATGGCCCAGCTC
iPLEX1	rs4688989	ACGTTGGATGTAGCCTAGTGGACATGCCTG	ACGTTGGATGTTCACCCACTTCCCTTTGTC	GGTGTGACCCCATTTC
iPLEX1	VAR_010602	ACGTTGGATGAAGAGGAAGAGGAAGTAGCC	ACGTTGGATGTGTGCTTCATGTGGTGTGAG	AAGTAGCCGATGGAGG
iPLEX1	rs13101355	ACGTTGGATGACTGTGTCCATCACCAAGTG	ACGTTGGATGCACCTCTGTACCTCCAGTAG	GGGAGCACGCTACGTGG
iPLEX1	rs2230719	ACGTTGGATGACTTCCTCTTCCTCTTTGCC	ACGTTGGATGTGAGCTCCAGAGACGTGAAC	CCTCCCCATCCTGGTGGC
iPLEX1	WFS1_K800E	ACGTTGGATGCTCTTGAACTCGCTGCTGG	ACGTTGGATGTGGGCATGCCATTCAGCAG	CCCGCAGCACGATGTCCT
iPLEX1	rs7672995	ACGTTGGATGTGCCCAAGTCCCTGCAGAAG	ACGTTGGATGACTCACACTCGCTGCTGAC	CCTGCAGAAGCAGAGGCG
iPLEX1	WFS1_A684A	ACGTTGGATGACTGACCTGGCAGCAGTATG	ACGTTGGATGTCCAGGTGGCTGCAGAGGAT	GGAAGGAGACCAACATGGC
iPLEX1	rs12642481	ACGTTGGATGAAGGCCAAGAGCTCTTTCTG	ACGTTGGATGATAGTCAAGGCCAAAGGCTG	GAGCTCTTTCTGACCTTAGC
iPLEX1	WFS1_W648R	ACGTTGGATGTTGTAGACCTTCATGCCCTC	ACGTTGGATGTCAAGCTCATCCTGGTGTGG	TGAGCGGTACACATAGAACC
iPLEX1	rs12511742	ACGTTGGATGTCTCTGCCTAGCTCCTTATG	ACGTTGGATGTGGAGGTGCATGTTGTAAGG	CACACCTTTCACCCAGTTCAC
iPLEX1	rs1046322	ACGTTGGATGAATTGCATGCCATCTCCACC	ACGTTGGATGAAAGTGCTTTTCCAGACACC	ACCCTGAGCCTGACCTTTCTGA
iPLEX1	rs5018648	ACGTTGGATGTGGAAGGCTTTCGAGCGAGA	ACGTTGGATGCTTGAGTCAGATGTCCATGC	AGGCTTTCGAGCGAGACTCATA
iPLEX1	VAR_014995	ACGTTGGATGCATGTACTGGAAGCTCAACC	ACGTTGGATGTTGACCTGGCCGACATTCTC	GGAAGCTCAACCCCAAGAAGAAG
iPLEX1	WFS1_19903	ACGTTGGATGAACCTGTACCAGTACCAGTC	ACGTTGGATGTCAAAACACGGTGAGGAAGG	TACCAGTACCAGTCGGAGCCCGTGT
iPLEX1	rs13107806	ACGTTGGATGGCACTGATAGCAGTACTTGG	ACGTTGGATGTCCCCTAATAACAGCAGCAG	GCAGTACTTGGCTGCAGACACTCGTT
iPLEX2	rs1802453	ACGTTGGATGCAGTGCATGTTGCCATGAGG	ACGTTGGATGTGCAGCCACAGTCTGCACAC	GGCATGCACCAGTGCC
iPLEX2	rs4416547	ACGTTGGATGTTCTGTCCTGGGTCTGCCTT	ACGTTGGATGTGTTCTGGACGCTGGGAGTA	CTGCCTTCCCTCCAGTG
iPLEX2	WFS1_33750	ACGTTGGATGTGCAGACTGTGGCTGCAGA	ACGTTGGATGTTCAGAGTGGACAGAGCTAC	TGGCTGCAGAGACCTTG
iPLEX2	rs10937714	ACGTTGGATGAGCTGCCAGACACTCCATAC	ACGTTGGATGTTTCCTCCCTGGAAGCGGTG	TCCATACAGGATCTCGCT
iPLEX2	rs7655482	ACGTTGGATGTCTCACCCATGCCTCCCAG	ACGTTGGATGACGGGTGAGATAGGGCAGG	CCTCTCCTTCCTGTGCGAC
iPLEX2	rs35031397	ACGTTGGATGAGCTGGTCACGGTAAAGAAG	ACGTTGGATGTCATCTTCTCCTTCCCCATC	AAGCCGGTGATGACAGCCA
iPLEX2	rs1801208	ACGTTGGATGAGCCTGAGCACCCATGCAGA	ACGTTGGATGCAATTCAAGGGCATGGAGGG	CCATGCAGAGCCCTACACGC
iPLEX2	rs1046316	ACGTTGGATGTGCTCGATCTTCACGTGCC	ACGTTGGATGTCTTCGAGCTCAAGGCCATC	CTTCACGTGCCGCCTGGTGGG
iPLEX3	VAR_014996	ACGTTGGATGGTGAAGATCGAGCACGACTG	ACGTTGGATGAAGTCGAAGGCGAACTTCAC	CGACTGGCGCAGCACC
iPLEX3	rs35932623	ACGTTGGATGGGGCCAGCAGCGAGTTCAA	ACGTTGGATGTCCAGGATGGTGCTGAACTC	GCGTGCTGCTCAGCCTG
iPLEX3	WFS1_33768	ACGTTGGATGGTATTCAGAGTGGACAGAGC	ACGTTGGATGAGAGACCTTGCGACCATGTG	TCCCTTTGTCGGGGTCCA