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.

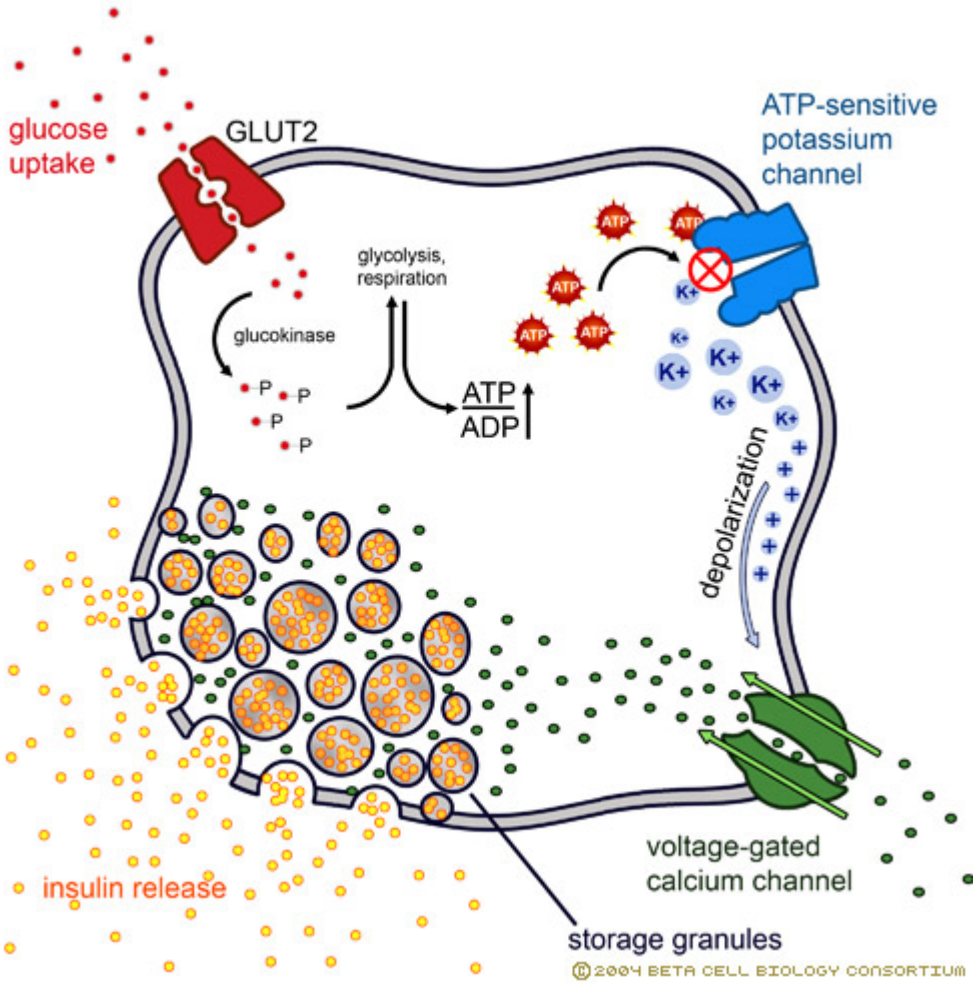


Figure 1.1 The insulin secretion pathway
Provided by the Beta Cell Biology Consortium (2004-2008) (http://www.betacell.org/content/articles/print.php?aid=1). Glucose enters pancreatic β-cells through GLUT2, is phosphorylated by glucokinase, and is used to generate ATP through glycolysis. A rise in the 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).

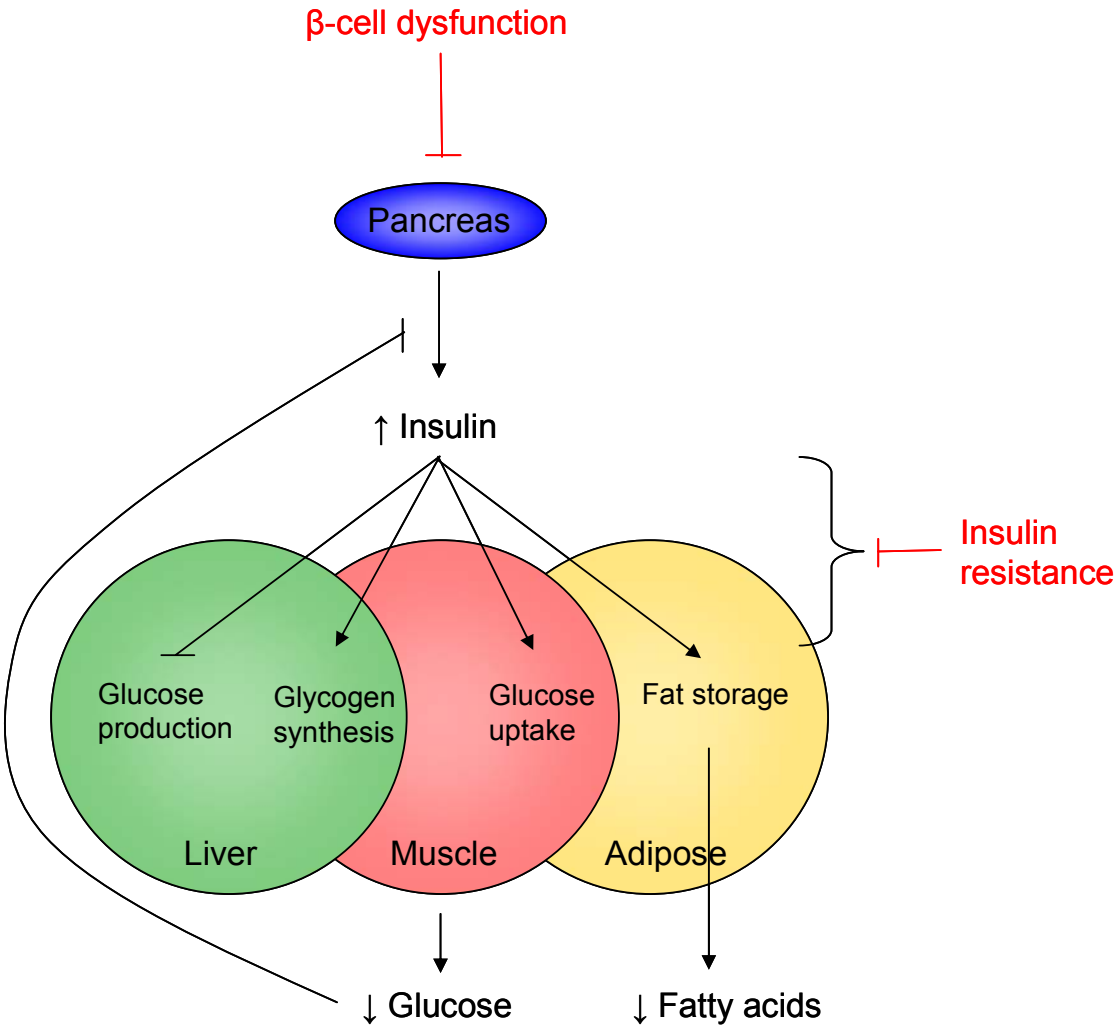


Figure 1.2 Insulin resistance and pancreatic ß-cell dysfunction in the pathogenesis of type 2 diabetes
When blood glucose levels rise the pancreatic β-cell secretes insulin. Insulin inhibits glucose production by the liver, and promotes glucose uptake by muscle and adipose tissue and storage of glucose 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).

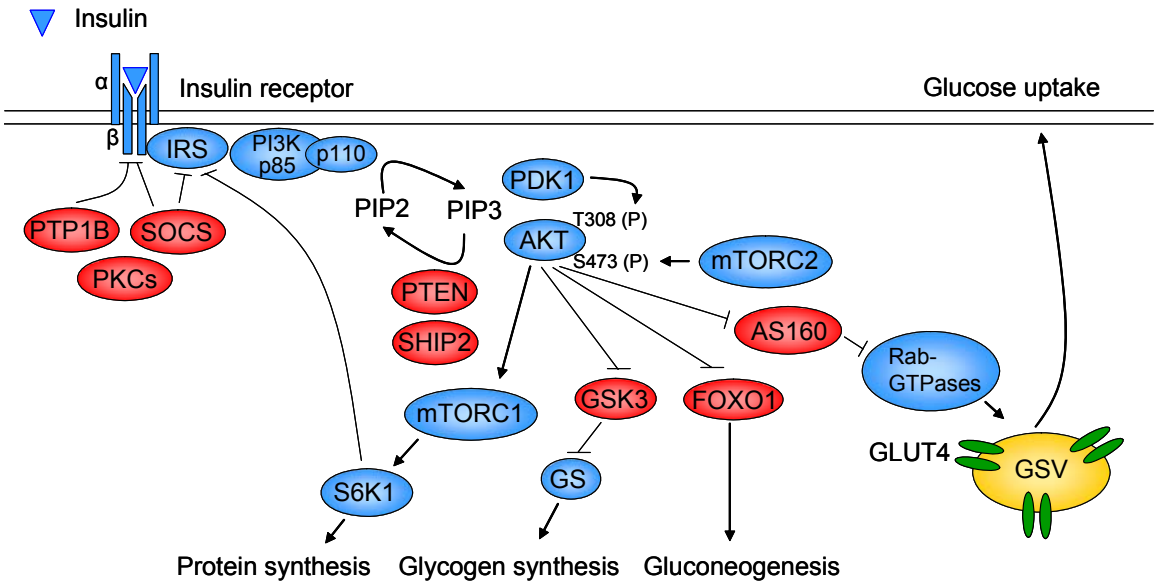


Figure 1.3 Regulation and metabolic effects of the insulin signalling pathway
Insulin binds to the α-subunits of the insulin receptor in target tissues resulting in phosphorylation of β-subunits and activation of insulin receptor substrate (IRS) proteins. The ensuing signaling cascade mediates the metabolic 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).

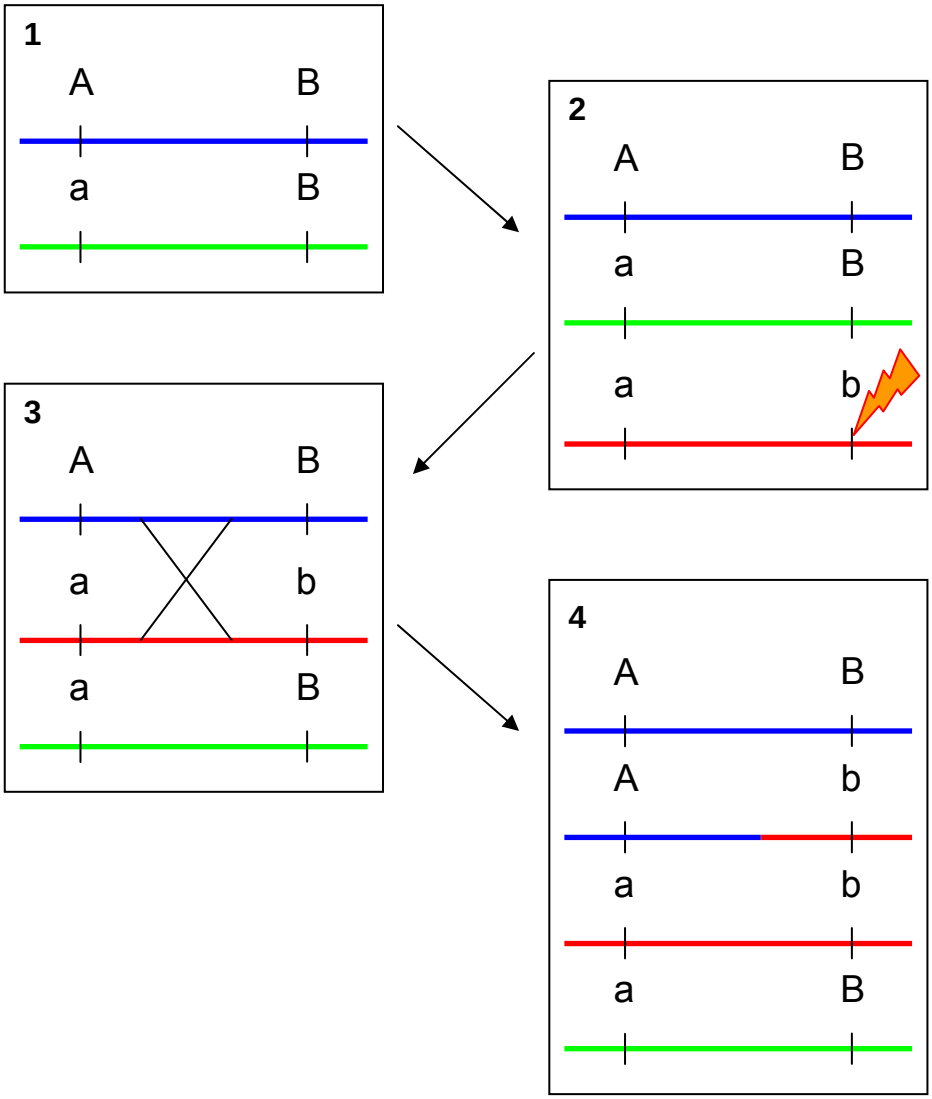


Figure 1.4 The erosion of LD by recombination, (adapted from (Ardlie et al. 2002)). In box 1 there is a polymorphic locus with alleles A and a. In box 2 a new mutation arises at a second locus changing allele B to b. This occurs on a chromosome carrying allele a, resulting in only three of four possible 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).

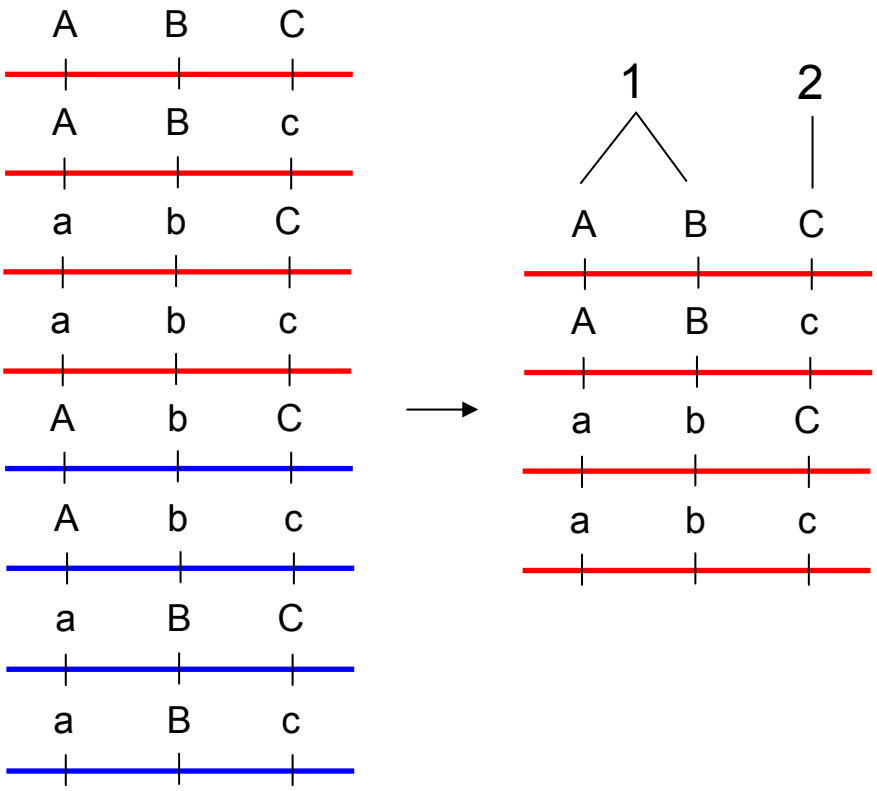


Figure 1.5 Selection of tagging markers based on pairwise correlation between three bi-allelic markers
To the left are all possible three-locus haplotypes of markers A/a, B/b, and C/c. However, only those in red are present at appreciable levels in the population. As A/a and B/b loci are in complete 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.

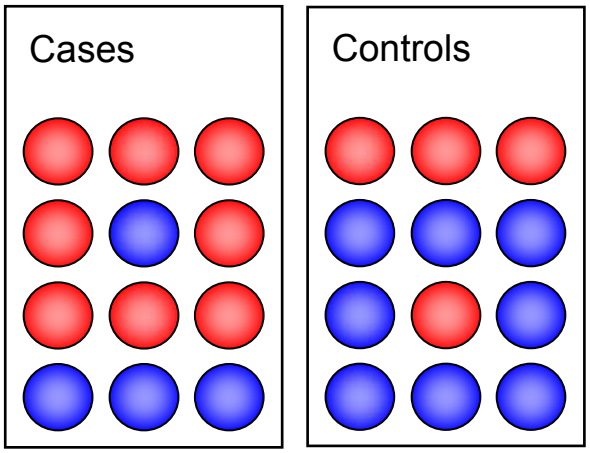
The genetic markers most commonly used for association studies are single nucleotide polymorphisms (SNPs) as millions of these have been catalogued and are easy to genotype. Patterns of LD among SNPs have been characterised in 270 DNA samples from four populations of European, African, and Asian (Chinese 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

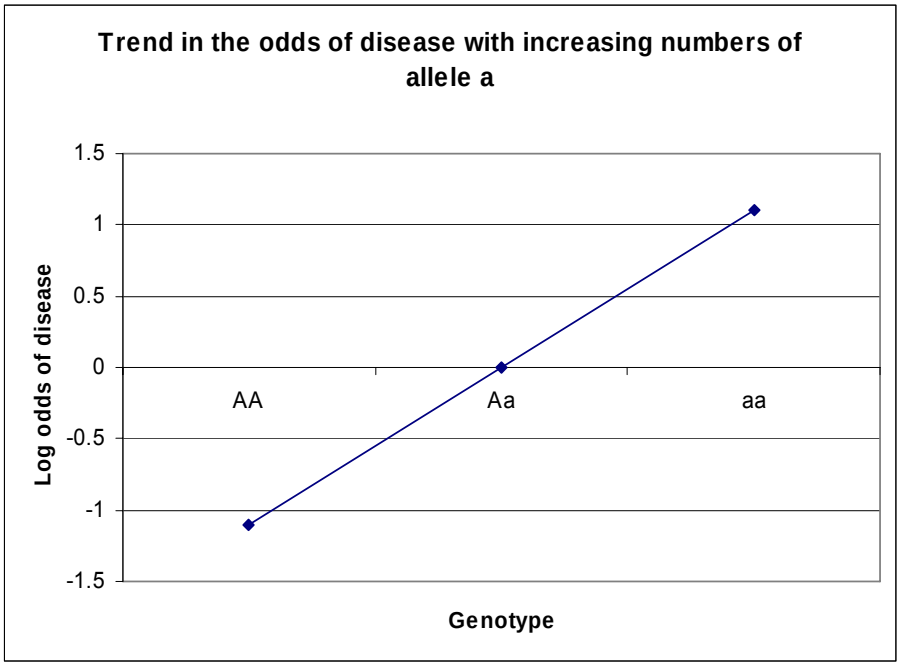


Figure 1.6 Association of a biallelic marker with complex disease, (adapted from (Balding 2006)). A, frequency of blue protective allele and red risk allele in cases and controls. B, 2 x 3 contingency table showing numbers of cases and controls in different genotypic groups. C, increase in odds of disease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

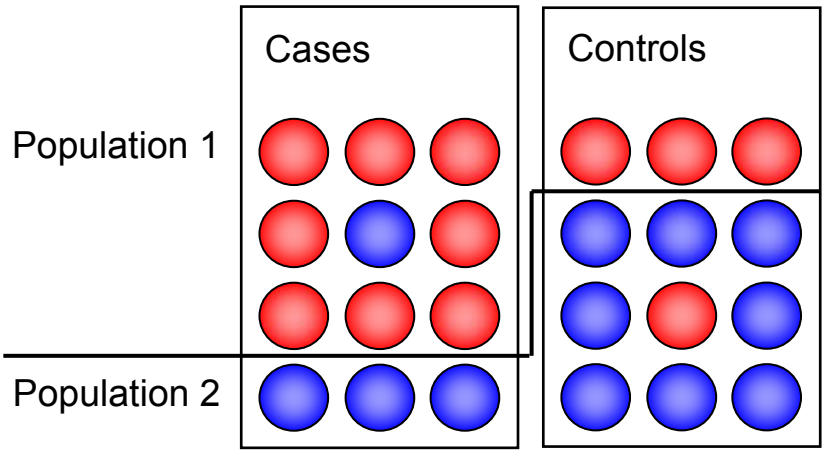


Figure 1.7 Spurious association of red alleles with disease due to population structure, (adapted from (Balding 2006). Population 1 has a higher frequency of red alleles compared to Population 2. As the frequency of people from population 1 is higher in cases than controls, red alleles are more prevalent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