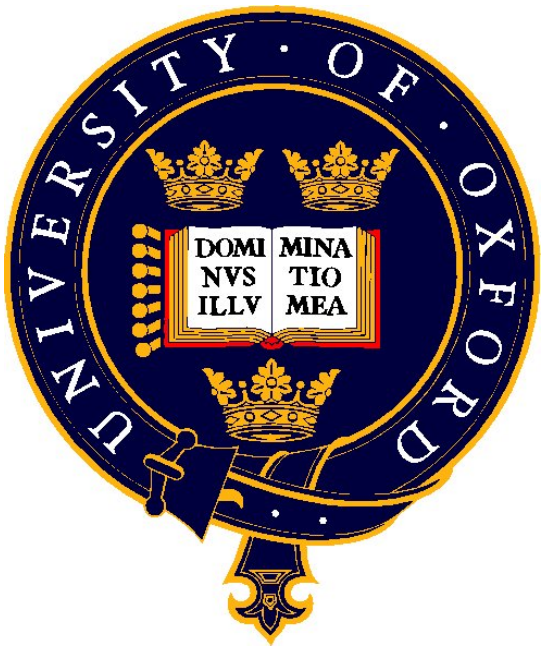
The evaluation of the contribution of low frequency, intermediate penetrance sequence variants to the pathogenesis of Type 2 Diabetes

Thesis submitted to the Faculty of Clinical Medicine
for the Degree of Doctor of Philosophy
University of Oxford



By
Bahram Jafar-Mohammadi
BMedSc MB BCh BAO (NUI) MRCPI

St Edmund Hall,
University of Oxford

Oxford Centre for Diabetes, Endocrinology and Metabolism,
Nuffield Department of Clinical Medicine,
University of Oxford

Trinity Term 2012

Acknowledgments

I would firstly like to thank my supervisors Professor Mark McCarthy and Dr Anna Gloyn for their support and the opportunity to conduct my research project at the Oxford Centre for Diabetes, Endocrinology and Metabolism. Their help and advice throughout my project has been invaluable.

I would also like to thank Mr Christopher Groves for his help with the technical aspects of the laboratory work involved in this project and also for his friendship. Dr Blanca Herrera and Mr William Rayner were invaluable in the analysis of the genotyping and high throughput sequencing data. Dr Katharine Owen, Dr Aparna Pal, Dr Gaya Thanabalasingham and Dr Nicholas Tribble were an immense source of support throughout my fellowship and thesis and I am grateful to them for this and for their friendship.

I will always be grateful to my parents, Maryam and Mohammad, for their love, sacrifice and support throughout my life. They have always provided a helping hand or a push of encouragement when one was needed.

I would especially like to thank my wife, Helen, whose patience, understanding and encouragement throughout my Fellowship/ D.Phil have been a constant source of support and I will forever be grateful to her. She inspires me daily and I will never forget her sacrifices that have allowed me to complete my DPhil. I am also grateful to her for proof reading my thesis. I am also thankful to our daughter, Tara, whose arrival 3 weeks prior to my viva has filled our lives with bliss.

I would like to thank the funding bodies that have supported my project, without which this project would not have been a possibility. The Irish Committee on Higher Medical Training and Sanofi Aventis provided a bursary, which together with the 2007 Ainsworth Scholarship from University College Cork, Ireland, funded the first year of my clinical training fellowship. Diabetes UK, generously funded the remainder of my clinical training fellowship. The National Institute of Health Research (NIHR) funded various aspects of the consumables of my project.

Table of Contents

Acknowledgments ii
Table of Contents iii
List of Figures vi
List of Tables viii
List of publications resulting from work carried out in this thesis xi
Glossary of common abbreviations xii
Abstract xiv

CHAPTER 1 Introduction 1
1.1 Introduction to Diabetes Mellitus 2
1.2 Type 2 Diabetes Mellitus 3
1.3 Pathogenesis of Type 2 Diabetes 4
1.3.1 Environmental factors in the pathogenesis of Type 2 Diabetes 4
1.3.1.1 Obesity and the risk of developing Type 2 Diabetes 5
1.3.1.2 Intrauterine environment 6
1.3.2 Epigenetic factors in the pathogenesis of Type 2 Diabetes 7
1.3.3 Genetic component in the pathogenesis of Type 2 Diabetes 8
1.4 Monogenic Diabetes Mellitus 9
1.4.1. Maturity onset Diabetes of the young 10
1.4.2 Neonatal Diabetes Mellitus 13
1.5 Genetics of Type 2 Diabetes Mellitus 15
1.5.1 Genome-wide linkage studies 16
1.5.2 Candidate gene studies 17
1.5.3 Genome Wide Association Studies 18
1.5.4 Clinical implications of T2D signals identified to date. 22
1.6 A role for low frequency intermediate penetrance variants in Type 2 Diabetes pathogenesis 22
1.7 High Throughput Sequencing 28
1.8 Strategy employed 29
1.9 Aims of thesis 30

CHAPTER 2 Materials and Methods 31
2.1 Subjects included 32
2.1.1 Group 1 samples 32
2.1.2 Group 2 samples 35
2.2 Methods 41
2.2.1 DNA and PCR product quantification 41
2.2.2 DNA pooling 42
2.2.3 Polymerase chain reaction amplification 43
2.2.4 Agarose gel electrophoresis 44
2.2.5 Capillary Sanger sequencing 45

2.2.5.1 Clean-up of PCR products 46
2.2.5.2 Sequencing reaction 46
2.2.5.3 Clean-up of samples post sequencing reaction 47
2.2.5.4 Analysis 47
2.2.6 Genotyping 48
2.2.6.1 Analysis of genotyping data 54
2.2.7 Blunt-ended ligation 55
2.2.8 Targeted sequencing- Genome Analyzer II 56
2.2.8.1 Data analysis 58

CHAPTER 3 Evaluation of the contribution of low frequency variants in the exons only encoding isoform A of *HNF1A* to susceptibility to Type 2 Diabetes 60
3.1 Introduction 61
3.2 Subjects & Methods: 63
3.2.1 Subjects 63
3.2.2 Methods 65
3.3 Results 67
3.4 Discussion 69

CHAPTER 4 Evaluation of the role of candidate coding, low frequency, incompletely penetrant variants in Type 2 Diabetes pathogenesis 71
4.1 Introduction 72
4.1.1 A98V variant of *HNF1A* 72
4.1.2 D76N variant of *PDX1* 74
4.1.3 The I88I variant of *KCNJ15* 75
4.1.4 T130I and V255M variants of *HNF4A* 77
4.2 Methods 79
4.2.1 Subjects studied 79
4.2.2 Genotyping and Quality Control 83
4.2.3 Statistical analysis 86
4.3 Results 89
4.3.1 Association analysis of the A98V variant of *HNF1A* 89
4.3.2 Association analysis of the D76N Variant of *PDX1* 91
4.3.3 Association analysis of the I88I variant of *KCNJ15* 91
4.3.5 Association analysis of the V255M and T130I variants of *HNF4A* 94
4.4 Discussion 98

CHAPTER 5 Genetic association analysis of low frequency variants in *LARS2* with Type 2 Diabetes 101
5.1 Introduction 102
5.2 Materials and methods 103
5.2.1 Study samples 103
5.2.2 Genotyping and quality control 108
5.2.3 Statistical analysis 110
5.3 Results 112

5.3.1 H324Q variant of *LARS2* 112
5.3.2. Geographical stratification of the H324Q variant 113
5.3.3 Association testing of rs9825041 with T2D risk 116
5.4 Discussion 118

CHAPTER 6 Targeted resequencing of exonic regions of genes implicated in Type 2 Diabetes pathogenesis 120
6.1 Introduction 121
6.2 Samples included 122
6.3 Pilot targeted sequencing project 124
6.3.1 Sequence alignment and SNP calling 125
6.3.2 Results of pilot project 125
6.3.3 Validation 127
6.4 Protocol Development 129
6.5 Main targeted sequencing project 132
6.5.1 DNA sample preparation 132
6.5.2 Sample preparation for the Genome Analyser II platform 132
6.5.3 Analysis and evaluation 135
6.5.4 Quality control 137
6.5.5 Results of the main high throughput resequencing project 140
6.5.6 Functional evaluation 143
6.5 Discussion 147

CHAPTER 7 General Conclusions and Discussion 151

Appendices 161
References 182

List of Figures

Figure 2.1 Figure depicting the stepwise dilution of the DNA samples used in the targeted sequencing project and final pooling step 43

Figure 2.2 Electropherogram illustrating the heterozygous c.1768+44C>T variant of *HNF1A* (detected in sequencing of the latter 3 exons of *HNF1A*) highlighted in blue box. 48

Figure 2.3 Schematic Representation of the TaqMan Genotyping Chemistry. 50

Figure 2.4 An example of the genotyping plot obtained using the ABI7900HT system in the analysis of the T130I variant of *HNF4A* (Chapter 4), demonstrating common homozygote, heterozygote and rare homozygote genotypes. 51

Figure 2.5 Schematic representation of the of the KASPAR genotyping chemistry 52

Figure 2.6 Schematic representation of the Genome Analyzer 2 platform including the cluster station where samples were hybridised to a flow cell.....58

Figure 3.1 *HNF1A* isoform structure and expression profile 62

Figure 3.2 Electropherogram illustrating the heterozygous c.1545G>A; T515T variant of *HNF1A* 66

Figure 4.1 Meta-analysis of all published studies of the A98V variant of *HNF1A* 90

Figure 4.2 Meta-analysis of the association studies of the T130I variant to Type 2 diabetes.....96

Figure 5.1 Linkage disequilibrium plot of the 14 low-frequency SNPs in *LARS2* included in the DIAGRAM scan. rs9825041 was used as a proxy for the block and is shown in the box.....104

Figure 5.2 Meta-analysis of H324Q variant of *LARS2* 113

Figure 5.3 Geographic stratification in the UK for the H324Q variant (MAF %). The MAF of the H324Q variant based on the region of the ascertainment the samples 115

Figure 5.4. Geographic stratification in other European samples (MAF %). The MAF of the H324Q in each of the different regions in Europe (other than UK) that samples in this study were ascertained. 116

Figure 5.5 Meta-analysis of the rs9825041 variant of *LARS2* 117

Figure 6.1. Figure depicting the design of the pilot and main targeted sequencing projects.....125

Figure 6.2 Comparison of the MAF of the 22 SNPs captured in this study (X axis) in the same 48 CEPH individuals that were also included in the HapMap project (Y-axis) 128

Figure 6.3 Sample preparation workflow for the targeted sequencing project.....134

Figure 6.4 Figure depicting the power to detect a singleton across the 20 lanes of the GAII run. Power to detect a singleton was ≥ 60% in approximately 75% of all targeted regions. Each line represents a separate pool (and hence lane used on the GAII platform). 138

Figure 6.5 Comparison of the MAF of the seven SNPs captured in this study (X axis) in the same individuals that were also included in the WTCCC study (Y-axis) .140

List of Tables

Table 1.1 Diagnostic criteria for the diagnosis of Diabetes Mellitus, impaired fasting glucose and impaired glucose tolerance based on the World Health Organisation and American Diabetes Association Guidelines 1 3

Table 3.1 Characteristics of subjects with T2Dincluded in the sequencing of the latter 3 exons of *HNF1A* 64

Table 3.2 Primer sequences for exons 8, 9 and 10 of *HNF1A* 65

Table 3.3 Variants identified using Sanger sequencing of exons 8-10 of *HNF1A* 67

Table 3.4 Probes used for genotyping of the variants detected in sequencing the latter 3 exons of *HNF1A* 68

Table 3.5 Allele frequencies of variants identified using Sanger sequencing of exons 8-10 of HNF1A 69

Table 4.1 Study samples included in each category for each candidate LFIP variant studied 81

Table 4.2 Characteristics of study subject included in the genotyping of V255M and T130I variants of *HNF4A,* A98V variant of *HNF1A,* D76N variant of *PDX1* and I88I variant of *KCNJ15.* 84

Table 4.3 Genotyping probes used in this study of V255M and T130I variants of *HNF4A*, A98V variant of *HNF1A*, D76N variant of *PDX1* and I88I variant of *KCNJ15* ... 86

Table 4.4 Power calculations for assocatin testing of V255M and T130I variants of *HNF4A,* A98V variant of *HNF1A*, D76N variant of *PDX1* and I88I variant of *KCNJ15* based on previously reported OR's 88

Table 4.5 Power calculations to evaluate the effect size which this project had 90% power to detect for the candidate LFIP variants V255M and T130I variants of *HNF4A,* A98V variant of *HNF1A*, D76N variant of *PDX1* and I88I variant of *KCNJ15* 88

Table 4.6 Association testing of the A98V (rs1800574) variant of *HNF1A* – Category 1 samples 89

Table 4.7 Association analysis of the D76N variant of *PDX11* 91

Table 4.8 Test for geographical stratification of the I88I variant of *KCNJ15* in the UK population 92

Table 4.9 Association study of I88I variant *KCNJ15* 93

Table 4.10 Association testing of the T130I variant of *HNF4* 95

Table 4.11 Association testing of the V255M variant of *HNF4A* 97

Table 5.1 Study Subject Characteristics included in the study of the *LARS2* variants..109

Table 5.2 Genotyping probes used in the study of rs982541 and H324Q of *LARS2* 110

Table 5.3 Association Testing of the H32Q variant in the UK samples.....113

Table 5.4 Association testing of the rs9825041 variant of *LARS2* in UK samples 117

Table 6.1 Genes prioritised for targeted sequencing 121

Table 6.2 Non-synonymous variants identified in pilot targeted sequencing project .126

Table 6.3 Synonymous variants identified in pilot targeted sequencing project.....127

Table 6.4 Table depicting the minor allele frequency bracket of variants previously reported in European samples and the number of samples called and missed in this experiment. A nominal false negative rate was calculated 137

Table 6.5 Overall summary of variants identified in the high throughput sequencing run and the C-Alpha statistic and C-Alpha P-Value for each gene (comparing the burden of rare variants in cases vs. controls) included in the main targeted sequencing project 141

Table 6.6 Non-synonymous polymorphisms with MAF >0≤0.05 with SIFT and PolyPhen predictions with regard to the amino acid change affecting the protein function sites that were called with “high confidence” by the syzygy software package. A p-value of the chi-square testing comparing the 500 cases vs. 500 control samples in this study has also been included for each variant. 142

Table 6.7 Polymorphisms predicted to affect Methylation sites that were called with “high confidence” by the syzygy software package. A p-value of the chi-square testing comparing the 500 cases vs. 500 control samples in this study has also been included for each variant 143

Table 6.8 Polymorphisms that are predicted to affect splice site that were called with "high confidence" by the syzygy software package. A p-value of the chi-square testing comparing the 500 cases vs. 500 control samples in this study has also been included for each variant 145

Table 6.9 Variants only seen in singleton cases, predicted to be damaging by PolyPhen and SIFT software packages and called with "high confidence" by the syzygy software package. The number of instances that each variant was detected has also been included.....147

Table S1.1 Loci reaching genome-wide significance (p<5x10-8) association with Type 2 Diabetes – arranged by chromosome number. The gene in closest proximity to the associated loci has also been quoted. The odds ratio quoted has been obtained from the referenced article in the last column**.** 161

Table S6.1 Primers utilised in the amplification of the exonic regions of the Genes included in targeted high throughput sequencing 165

Table S6.2 Variants detected with high confidence in the main targeted high throughput sequencing project. The P-value for the comparison of the 500 cases vs. 500 controls has been included or each value. The C-alpha statistic P-value for the burden of rare variants in cases vs. controls for the amplicon in which the variant is located is also included.....170

List of publications resulting from work carried out in this thesis

1). **Jafar-Mohammadi B**, Groves CJ, Gjesing AP, Herrera BM, Winckler W, Stringham HM, Morris AP, Lauritzen T, Doney AS, Morris AD, Weedon MN, Swift AJ, Kuusisto J, Laakso M, Altshuler D, Hattersley AT, Collins FS, Boehnke M, Hansen T, Pedersen O, Palmer CN, Frayling TM; DIAGRAM Consortium, Gloyn AL, McCarthy MI. **A role for coding functional variants in *HNF4A* in type 2 diabetes susceptibility.** Diabetologia. 2010 Sep 29. [Epub ahead of print]PMID: 20878384

2). **[Jafar-Mohammadi B*,** Reiling E**]**, van 't Riet E, Weedon MN, van Vliet-Ostaptchouk JV, Hansen T, Saxena R, van Haeften TW, Arp PA, Das S, Nijpels G, Groenewoud MJ, van Hove EC, Uitterlinden AG, Smit JW, Morris AD, Doney AS, Palmer CN, Guiducci C, Hattersley AT, Frayling TM, Pedersen O, Slagboom PE, Altshuler DM, Groop L, Romijn JA, Maassen JA, Hofker MH, Dekker JM, McCarthy MI, 't Hart LM. **Genetic association analysis of *LARS2* with type 2 diabetes**. Diabetologia. 2010 Jan;53(1):103-10. Epub 2009 Oct 22.PMID: 19847392

*** Joint First Author**

3). **Jafar-Mohammadi B,** Groves CJ, Owen KR, Frayling TM, Hattersley AT, McCarthy MI, Gloyn AL. **Low frequency variants in the exons only encoding isoform A of *HNF1A* do not contribute to susceptibility to type 2 diabetes**. PLoS One. 2009 Aug 12;4(8):e6615. PMID: 19672314.

4). **Jafar-Mohammadi B** & McCarthy MI**. Genetics of type 2 diabetes mellitus and obesity-a review**. Ann Med. 2008;40(1):2-10. PMID: 17952710

Glossary of common abbreviations

ADA	American Diabetes Association
ATP	Adenosine triphosphate
BMI	Body Mass Index
DNA	Deoxyribonucleic acid
DIAGRAM	Diabetes Genetics Replication And Meta-analysis Consortium
dNTPs	Deoxyribonucleotide triphosphate
GAD	Glutamic Acid Decarboxylase
GAII	Genome Analyser II
GWAS	Genome wide association study
HNF1A	Hepatocyte nuclear 1 alpha
HNF4A	Hepatocyte nuclear 4 alpha
KASPar	KBioscience allele specific PCR
KCNJ11	Potassium- inward rectifier
KCNJ15	Potassium inwardly rectifying channel, subfamily J, member 15
LARS2	Leucyl-tRNA synthetase 2, mitochondrial
LD	Linkage disequilibrium
LFIP	Low frequency intermediate penetrance
LOD	logarithm of the odds
MAF	Minor allele frequency
MODY	Maturity Onset Diabetes of the Young

NDM	Neonatal Diabetes Mellitus
NGS	Next generation sequencing
OR	Odds Ratio
PCR	Polymerase Chain Reaction
PDX1	Pancreatic and duodenal homeobox 1
SNP	Single nuclear polymorphism
TCF7L2	Transcription factor 7-like 2
T1D	Type 1 Diabetes
T2D	Type 2 Diabetes
UKT2DGC	United Kingdom Type 2 Diabetes Genetics Consortium
UTR	Untranslated region
WHO	World Health Organisation
WTCCC	Wellcome Trust Case Control Consortium
WTCHG	Wellcome Trust Centre for Human Genetics

Other abbreviations have been explained in the text, when first used.

Abstract

The evaluation of the contribution of low frequency, intermediate penetrance sequence variants to the pathogenesis of Type 2 Diabetes

Bahram Jafar-Mohammadi **St Edmund Hall**

Thesis submitted to the Faculty of Clinical Medicine for the Degree of Doctor of Philosophy

Trinity Term 2012

Genome wide association studies (GWAS) and their subsequent meta-analysis have identified a large number of susceptibility variants for Type 2 diabetes (T2D) risk. However, the familial aggregation seen in this disease is not yet fully explained. The sibling relative risk (λ_s) due to all known variants is ~1.104 which is well below the epidemiological estimates of λ_s of ~3.0. There has therefore been great interest in the potential role of variants that would have been largely invisible to the initial wave of GWAS and linkage approaches. Low frequency (minor allele frequency 1-5%), incompletely penetrant (odds ratio 2-4) variants (LFIP), are one such group of potential susceptibility variants.

The overall objective of this project (designed and implemented in 2007-2010) was to evaluate the contribution of LFIP variants to the inherited susceptibility to T2D. I tested the specific hypothesis that genes already-implicated in diabetes pathogenesis (due to an established role in monogenic or multifactorial disease) also harbour LFIP variants, and that those variants may contribute appreciably to the prediction of disease risk.

Mutations in exons only encoding isoform-A of *HNF1A* have been demonstrated to lead to a later age of diagnosis of HNF1A-MODY. This region was therefore felt to be auspicious for harbouring LFIP variants impacting on T2D risk. I have demonstrated that such variants impacting on T2D risk are unlikely to be present in this region by use of Sanger sequencing in a sample enriched for young onset, familial T2D.

The role in T2D risk of candidate LFIP variants across 5 genes (*HNF1A*, *HNF4A*, *PDX1*, *KCNJ15 and LARS2*), was evaluated by large scale association studies. For one variant, T130I of *HNF4A,* a modest association ($p=5x10^{-4}$) with T2D was seen in UK samples and the strength of association was marginally improved by incorporation of all previous studies of this variant in T2D in a meta-analysis ($p=2.1x10^{-5}$). This study demonstrated the difficulties encountered in confirming the association of low frequency variants to complex diseases, especially for those with modest effect sizes.

At the time of project design and inception “next-generation” sequencing platforms were in their infancy and the study design I planned (that of pooled, targeted sequencing) had not been widely applied. It was therefore necessary to design and optimise protocols for sample preparation for sequencing on this platform. I used the Genome Analyzer II platform to sequence ten genes previously implicated in T2D or monogenic diabetes pathogenesis in pooled DNA samples. This approach yielded in excess of 2900 variants, a large portion being novel. As part of this project I have highlighted heuristics that can be used in the follow-up of potential susceptibility variants discovered using high throughput sequencing. I have also established protocols and pathways for sample preparation that can be utilised across several next generation sequencing platforms for future studies in the host institution and beyond.

CHAPTER 1

Introduction

1.1 Introduction to Diabetes Mellitus

Diabetes Mellitus is a chronic heterogeneous disorder characterised by hyperglycaemia (American Diabetes Association 2007). Type 1 diabetes (T1D) is an autoimmune condition leading to pancreatic beta-cell destruction and accounts for approximately 5% of cases of diabetes in adults (http://diabetes.niddk.nih.gov). Type 2 diabetes (T2D) is the most common form of the condition accounting for 90-95% of cases of diabetes in the adult population (http://diabetes.niddk.nih.gov). Other subtypes of diabetes including, but not limited to, monogenic diabetes, diabetes secondary to underlying pancreatic disease or medications such as exogenous glucocorticoids, account for approximately 1-5% of cases of diabetes in adults (American Diabetes Association 2007). The thresholds for the diagnosis of diabetes mellitus have changed several times over the past 20-30 years, but are primarily based on the biochemical confirmation of hyperglycaemia. These thresholds have been based on glycaemic levels that are associated with a significantly increased risk of developing complications of diabetes mellitus (Sacks et al. 2011). The latest criteria for the diagnosis of this condition (American Diabetes Association 2007; World Health Organization 2011) have been summarised in Table1.1.

Table 1.1 Diagnostic criteria for the diagnosis of Diabetes Mellitus, impaired fasting glucose and impaired glucose tolerance based on the World Health Organisation and American Diabetes Association Guidelines 1

	Normal glucose tolerance	Impaired fasting glucose	Impaired glucose tolerance	Diabetes Mellitus
Fasting plasma glucose	**<6.1 mmol/L**	**6.1-6.9 mmol/L**	**-**	**>7.0 mmol/L**
OGTT*-2 hour plasma glucose in	**<7.8 mmol/L**	**-**	**7.8-11.0 mmol/L**	**>11.1 mmol/L**
HbA1c+	**-**	**-**	**-**	**6.5% (48 mmol/mol)**

*OGTT: Oral glucose tolerance test- plasma glucose 2 hours after the ingestion of 75g oral glucose load; +HbA1c: glycosylated Haemoglobin; [1](American Diabetes Association 2007; World Health Organization 2011)

1.2 Type 2 Diabetes Mellitus

Type 2 diabetes (T2D) represents a major cause of ill-health in societies worldwide (Zimmet et al. 2001) and is the primary focus of this thesis. The number of people being diagnosed annually is approaching epidemic proportions with great cost to health services as a result of the condition and its complications (Lam et al. 2012). It is estimated that there are currently in excess of 340 million people globally with T2D (Danaei et al. 2011). In the United Kingdom, diabetes management accounts for approximately 10% of the National Health Service budget annually (Sacks et al. 2011). The prevalence of T2D globally is reported to be approximately 6.3 % (World Health organisation) but this does vary from 3.4 % in parts of Africa to 10.2% in the Western Pacific (World Diabetes Foundation).

To allow for the development of effective healthcare strategies aimed at the treatment of T2D and the complications associated with it, an understanding of the pathophysiology of this condition is needed.

1.3 Pathogenesis of Type 2 Diabetes

T2D is a chronic progressive condition with pancreatic beta-cell dysfunction and insulin resistance central to its pathogenesis (Weyer et al. 1999; Butler et al. 2003; Kahn 2003; Hanley et al. 2010). Environmental, genetic and epigenetic factors have been implicated in the pathogenesis of T2D and these will each be discussed below. It is likely that a complex interplay of all or some of these factors are involved in the development of T2D.

1.3.1 Environmental factors in the pathogenesis of Type 2 Diabetes

Several environmental factors such as, obesity (Alberti et al. 2004), low socioeconomic status in developed nations (Connolly et al. 2000; Link et al. 2009), high socioeconomic status in developing nations (Ramachandran et al. 2002; Alberti et al. 2004), sedentary lifestyles (Alberti et al. 2004), ethnicity (Harris et al. 1999; Steinbrecher et al. 2011), smoking (Willi et al. 2007) and intrauterine environment (Dabelea et al. 2000) have been associated with an increased risk of developing T2D. However, association does not equate to causality, and an understanding of how these environmental factors contribute to the pathogenesis of disease is required. In the case of high socioeconomic status in developing nations, low socioeconomic status in developed nations and sedentary lifestyles, the observed increased risk of developing T2D is largely conferred through increased rates of obesity and insulin resistance as outlined in section 1.3.1.1 (Connolly et al. 2000; Ramachandran et al. 2002; Alberti et al. 2004; Link et al. 2009). Observational studies had initially highlighted the increased risk of developing T2D with a smoking history (Willi et al. 2007). This observation has reached nominal statistical significance ($p<0.05$) in a meta-analysis

incorporating 1.2 million participants across 25 prospective cohort studies (Willi et al. 2007). However, the mechanism underlying this association is not well understood and future studies are required to establish whether this association is causal. Such epidemiological associations are prone to spurious effects such as reverse causation and environmental confounders (Rockett 1994).

1.3.1.1 Obesity and the risk of developing Type 2 Diabetes

In the case of obesity, a parallel rise in the rates of this condition and T2D have been observed in many populations (Alberti et al. 2004). Increased adipose tissue and resultant insulin resistance with increasing weight have been thought to predispose to the development of T2D in obese individuals (body mass index, BMI, >30 Kg/m^2). However, not everyone classified as obese develops T2D (Hardy et al. 2012).

The mechanism leading to the development of T2D in obese individuals is not well understood. It is of interest that visceral adipose tissue and not subcutaneous fat have been associated with increased risk of development of insulin resistance (McLaughlin et al. 2011). The odds of developing insulin resistance have been reported to be increased by 80% with a single standard deviation increase in visceral adipose tissue mass (McLaughlin et al. 2011). The release of cytokines such as the pro-inflammatory interleukin (IL)-6 and tumour necrosis factor-α from mature adipocytes have been postulated to lead to insulin resistance by promoting mild inflammation (Hotamisligil 2003; Wellen et al. 2005; Hardy et al. 2012). Of note, the treatment of obese mice with “resolvins” that promote the resolution of inflammation have been noted to

improve glycaemic control (Hotamisligil 2003; Wellen et al. 2005; Hellmann et al. 2011; Hardy et al. 2012).

1.3.1.2 Intrauterine environment

Several epidemiological studies have identified a maternal role in the development of T2D (Martin et al. 1985; Pettitt et al. 1988; Aerts et al. 1990; Alcolado et al. 2002; Fetita et al. 2006). To evaluate the cause of this finding, a study in Pima native American families (Dabelea et al. 2000) that had at least one sibling born before and one after the mother was diagnosed with diabetes was carried out. This study has found that offspring born after the mother developed T2D had a 4 fold increased risk of developing diabetes as compared to those before the diagnosis of diabetes in the mother (Dabelea et al. 2000). A further study (Sobngwi et al. 2003), evaluating the effect on offspring of mothers having T1D during pregnancy found that in those non diabetic individuals in adulthood, that did not have immunological evidence of T1D (Anti- Glutamic decarboxylase antibodies [GAD] and Anti-islet cell Antibodies) did have 33% prevalence of impaired fasting glucose as compared to none in those with only fathers that had type 1 diabetes (Sobngwi et al. 2003). The findings from these studies would indicate that the intra-uterine exposure to hyperglycaemia could be a factor in the future development of hyperglycaemia and T2D in the offspring.

Intrauterine nutritional deficiency resulting in low birth weight has also been demonstrated to be associated with increased risk of developing T2D, obesity and cardiovascular disease (Hales et al. 1991; Barker et al. 1993; Barker 2004; Whincup et al. 2008). The cause for this finding is not yet fully understood. There are a number of theories to account for this finding. Firstly, the foetal

origins hypothesis (Hales et al. 1992; Barker 1995), postulates that an intrauterine stimulus, such as malnutrition, results in long term physiological "programming " (Barker 1995) and predispose the individual to later development of T2D. Another possible explanation for this finding is reduced beta cell mass in the infant, with subsequent impaired glucose tolerance and T2D in adulthood (Whincup et al. 2008). Finally, it is postulated that reduced low birth weight individuals have reduced insulin secretion which may predispose to later development of T2D (Jensen et al. 2002).

1.3.2 Epigenetic factors in the pathogenesis of Type 2 Diabetes

Epigenetics is broadly used to describe heritable alterations of gene function that are not attributable to changes in the nucleotide sequence of the gene (Bird 2007). It has been established that certain factors such as aging, obesity and reduced exercise increase the risk of developing T2D. It is postulated that these changes may be attributable, at least in part, to epigenetic factors (Ling et al. 2009). Epigenetic factors may also account, at least in part, to the effect of the intrauterine environment and later development of T2D described in Section 1.3.1.2. However, the contribution of epigenetic factors in the pathogenesis of complex diseases in man has not been fully deciphered and there are limited animal models described. Examples of epigenetic factors influencing the risk of developing complex diseases include DNA methylation and histone modifications that can interfere with gene transcription (Ling et al. 2009).

A study of the effects of methylation on the transcription of the *COX7A1* gene, an important factor in the respiratory chain (Ronn et al. 2008) illustrates how epigenetic factors may be involved in T2D pathogenesis. *COX7A1* gene is only expressed in cardiac and skeletal muscle cells

(Arnaudo et al. 1992). Reduced expression of this gene is seen in skeletal muscle from diabetic samples (Mootha et al. 2003; Ronn et al. 2008). Also, increased DNA methylation of the promoter region of the gene was demonstrated in human skeletal muscles with increasing age in diabetic individuals (Ronn et al. 2008). It has therefore been postulated that the increased DNA methylation of this gene may be associated with increased risk of developing T2D due to decreased glucose uptake and hence insulin resistance (Ling et al. 2009).

Furthermore, studies involving the Glucokinase gene (*GCK*), have demonstrated decreased activity of this gene with increasing age (Cockburn et al. 1997), which may, in part be due to epigenetic factors. As discussed in Section 1.4, Glucokinase is vital for hepatic glucose regulation. Studies in rat models have demonstrated increased methylation of the promoter regions of *Gck* and decreased activity of this gene in aged rats (Jiang et al. 2008). Subsequent in vitro studies of the hepatocytes of aged rats have demonstrated markedly increased expression of this gene when chemical demethylation of the promoter regions was conducted (Jiang et al. 2008). These studies demonstrate a possible explanation for observed increased risk of developing T2D with advancing age (Ling et al. 2009).

1.3.3 Genetic component in the pathogenesis of Type 2 Diabetes

Evidence for a genetic component in the pathogenesis of T2D exists from multiple sources. Firstly, evidence from twin studies in T2D have demonstrated a much higher concordance rate among monozygotic (range 0.3-1.0) as compared to dizygotic twins (range 0.1-0.4) (Barnett et al. 1981; Newman et al. 1987; Medici et al. 1999; Poulsen et al. 1999; Das et al. 2006)]. This finding would support a genetic component in the pathogenesis of T2D.

Secondly, epidemiological studies have demonstrated that the sibling relative risk (λ_s, the risk of developing the condition if one's sibling is affected) of the development of T2D is approximately 3 (Gottlieb 1980; Kobberling J 1982; Das et al. 2006). However, a recent report from Finland, focusing on non-twin siblings has indicated that this figure may be much higher and be of the order of 8 in the Botnia region of Finland (Almgren et al. 2011). Although, families share environmental factors, the familial aggregation of T2D is felt to be at least in part secondary to genetic factors (Das et al. 2006).

Finally, as discussed in Section 1.5, a portion of the heritability of T2D (up to 10 %) can now be attributed to genetic loci identified through genome wide association studies and their subsequent meta-analyses (Morris et al. 2012).

Specification of the genetic basis of this disease provides a powerful tool for improved understanding of pathogenesis and opportunities for clinical translation. Monogenic forms of diabetes provide clear examples of the value of this approach.

1.4 Monogenic Diabetes Mellitus

Monogenic forms of diabetes mellitus account for between 1-2% of cases of diabetes world-wide (Permutt et al. 2005; Kropff et al. 2011) and show a strong correlation between genotype and phenotype. The study of the genetic basis of these conditions has provided great insights into their molecular basis as well as providing clues about metabolic homeostasis which have broader

applications. The molecular classification of these conditions that is now possible due to a greater understanding of their genetic aetiology has important therapeutic implications (American Diabetes Association 2007).

1.4.1. Maturity onset Diabetes of the young

Maturity onset diabetes of the young (MODY) is a heterogeneous group of disorders arising from monogenic defects in the pancreatic beta cell (American Diabetes Association 2007). As a result of successes in gene identification, the original clinical definition (based around a non-autoimmune loss of beta-cell function, autosomal dominant inheritance and early age of diagnosis) has been largely replaced by a molecular classification (American Diabetes Association 2007).

This molecular classification has clear prognostic and therapeutic implications that increasingly see molecular diagnostic tools deployed as part of routine clinical care. Individuals with mutations in the glucokinase (GCK) gene (Froguel et al. 1992; Hattersley et al. 1992) account for approximately 32% of cases of MODY (Shields et al. 2010), but this form of the disease is characterised by mild, stable fasting hyperglycaemia, with little deterioration of beta-cell function over time. The majority of cases can be treated with dietary measures alone, and complications are few. A wide spectrum of mutations within the *GCK* gene has been described (Osbak et al. 2009). Monogenic diabetes resulting from mutations in beta-cell transcription factors (most commonly in *HNF1A*, encoding Hepatocyte nuclear factor-1A) (Yamagata et al. 1996) is, in contrast, more progressive with age, with substantial risk of diabetes complications if

treatment is not optimized (Isomaa et al. 1998; Lee et al. 1999; Steele et al. 2010). Those with *HNF1A* mutations are particularly sensitive to the hypoglycaemic effects of sulphonylureas (Pearson et al. 2003) and, in contrast to most patients with typical T2D, these, rather than metformin, represent optimal treatment. Obtaining a definitive diagnosis through molecular testing can thus lead to more rational treatment selection.

However, it has been reported that there is currently a mean delay of up to 13 years from the diagnosis of diabetes to the confirmation of the correct genetic cause of HNF1A-MODY (Shields et al. 2010) Although there may be multiple reasons for this finding, lack of awareness of the condition, expense of the genetic tests (£350 for a sequencing of the exonic regions of *HNF1α* in the Molecular Genetics Laboratory, Royal Devon Hospital, Exeter, UK; http://www.diabetesgenes.org) and genetic testing not being widely available, are likely contributors.

Even amongst the various subtypes of transcription-factor MODY [which can result from mutations in the genes encoding Hepatocyte nuclear factor-4A (HNF-4A) and Hepatocyte nuclear factor-1B (HNF-1B) as well as Hepatocyte nuclear factor-1A (HNF-1A)], there are subtle phenotypic differences. Although HNF-4A and HNF-1A are closely related functionally (Yamagata et al. 1996) and both lead to diabetes as a result of progressive beta-cell dysfunction, only mutations in *HNF4A* present with foetal and neonatal hyperinsulinaemia resulting in increased birth weight and hypoglycaemia (Pearson et al. 2007). Similarly, mutations in *HNF1B* (encoding HNF-1B) are associated with a spectrum of abnormalities in the renal and urogenital systems (Lindner et al. 1999). These, rather than the diabetes, tend to dominate presentation. In families segregating mutations in *HNF1*B, pancreatic atrophy (rather than beta-cell dysfunction) is

characteristic and many affected individuals are also insulin-resistant (Bellanne-Chantelot et al. 2004; Pearson et al. 2004). Once again, the capacity to allocate a specific molecular diagnosis has clear prognostic value for the index case and their families.

Other causes of MODY are extremely rare. Insulin promoter factor 1 (*PDX1*) is a homeodomain transcription factor required for pancreas development and for the transcriptional regulation of genes in the pancreatic beta cells. Mutations in the *PDX1* gene encoding PDX-1 underlie a rare form of the condition (Stoffers et al. 1997). Mutations in *NEUROD1* (Malecki et al. 1999) and *CEL* (Raeder et al. 2006) have also been implicated in a small number of families segregating early-onset non-autoimmune monogenic diabetes. Around 10-13% of families segregating MODY have, as yet, no defined genetic cause (Frayling et al. 2001; Murphy et al. 2008).

There has been a recent drive to find biomarkers that can aid in the diagnosis of MODY. Success in biomarkers in HNF1A-MODY has come in part due to the advent of genome wide association studies (Section 1.5.3). Variants in the region of *HNF1A* were identified as altering levels of C-reactive protein (CRP) (Reiner et al. 2008; Ridker et al. 2008). Subsequent studies of HNF1A-MODY patients have demonstrated robustly that these individuals have markedly reduced levels of highly sensitive CRP compared to those with T2D (Owen et al. 2010; McDonald et al. 2011; Thanabalasingham et al. 2011). This biomarker can be used in the clinical setting to identify patients that may have HNF1A-MODY. This represents one of the first clinical translations of GWA study data.

1.4.2 Neonatal Diabetes Mellitus

The most exciting developments in the field of monogenic diabetes in recent years have been in the area of neonatal diabetes (NDM). Until recently, classification of this condition (which is defined as diabetes with onset in the first six months of life) was based around the clinical course of disease; it was either permanent neonatal diabetes (PNDM) or transient neonatal diabetes (TNDM), the latter characterized by early remission, but frequent relapse in adulthood. It had been known for some time that many of these transient cases are attributable to abnormalities of an imprinted locus on chromosome 6q24 [involving the *ZAC* and *HYMA1* genes] (Gardner et al. 2000; Temple et al. 2000).

It is now clear that most cases of permanent NDM and many of the remaining cases of transient NDM are the result of activating mutations in the *KCNJ11* and *ABCC8* genes, each of which encodes one of the two subunits of the Adenosine Triphosphate (ATP)-sensitive potassium channel [K_{ATP}] (Gloyn et al. 2004; Sagen et al. 2004; Babenko et al. 2006; Proks et al. 2006). The most serious mutations in *KCNJ11* lead to the clinically-most severe syndromic form (DEND syndrome- Developmental Delay, Epilepsy and Neonatal Diabetes) (Gloyn et al. 2004) in which diabetes is accompanied by neurological features thought to reflect the extrapancreatic expression of the gene product.

In most cases, these mutations disrupt the capacity of the ATP signal generated by intracellular glucose metabolism to effect channel closure, depolarization of the plasma membrane and insulin release (Koster et al. 2000). Since sulphonylureas lead to membrane depolarization through an ATP-independent mechanism, it has been possible to transfer many NDM patients off

insulin therapy (not surprisingly, most of these individuals have been incorrectly labelled with type 1 diabetes) onto sulphonylureas, a therapeutic change which has generally resulted in improved glycaemic control and amelioration of some of the neurological sequelae (Gloyn et al. 2004; Pearson et al. 2006; Proks et al. 2006; Codner et al. 2007; Stanik et al. 2007).

Mutations in the preproinsulin (*INS*) gene have relatively recently been reported to be causative for approximately 20% of cases of NDM (Stoy et al. 2007).Interestingly dominant and recessive patterns of inheritance have been described for mutations in this gene leading to NDM (Stoy et al. 2007; Edghill et al. 2008; Garin et al. 2010; Stoy et al. 2010). Heterozygous dominant missense *INS* mutations are predicted to disrupt preproinsulin folding, leading to increased oxidative stress in the endoplasmic reticulum and beta cell apoptosis (Wang et al. 1999; Herbach et al. 2007; Liu et al. 2007). In contrast homozygous recessive mutations are predicted to decrease preproinsulin synthesis. Several different pathways have been reported to be important in modulating the effect of the recessive mutations. These include reduced mRNA stability, reduced transcription secondary to mutations in the promoter region and absent or altered translation due to a missense mutation or deletion (Garin et al. 2010; Stoy et al. 2010). Unsurprisingly, given the importance of insulin as a growth factor, patients with the more functionally severe homozygous recessive mutations, have a more marked intrauterine growth retardation and lower birth weight (median birth weight 1680g vs. 2530g) (Garin et al. 2010). These patients also require insulin therapy earlier (Stoy et al. 2010). Mutations in the non coding, promoter region of the *INS* gene have been associated with PNDM or TNDM, whereas mutations in the coding region or those affecting mRNA stability result in PNDM only (Garin et al. 2010). Dominant mutations have only reported to be causative for PNDM (Stoy et al. 2007), whereas recessive mutations can result in TNDM or PNDM (Garin et al. 2010).

More recently, mutations in the transcription factor *GATA6* (Allen et al. 2012) have been shown to account for approximately 56% of cases of pancreatic agenesis (defined as neonatal diabetes and pancreatic exocrine deficiency). The Genome Analyzer II platform was utilised for whole exome sequencing (Allen et al. 2012) in two individuals with pancreatic agenesis. A heterozygous non-synonymous mutation (p.Thr452Ala) and a frame-shift deletion (c.1448_1455del) in *GATA6* were identified (Allen et al. 2012). Subsequent Sanger sequencing of the exonic regions and intron exon boundaries of this gene in 24 other individuals identified a further 12 mutations causative for pancreatic agenesis. This is of particular interest in my project, as I have utilised the Genome Analyzer II platform in the search for low frequency sequence variants that may impact on T2D pathogenesis (Chapter 6).

1.5 Genetics of Type 2 Diabetes Mellitus

In contrast to monogenic diabetes, advances made in deciphering the genetic aetiology of multifactorial T2D, has until recently been far slower (McCarthy 2010). This is due to the fact that susceptibility to, and later development of, T2D likely involves the inheritance of multiple variants across multiple sites together with the presence of multiple environmental factors. In this setting, each susceptibility variant confers a relatively modest effect with a resultant reduced power for detection. This complex interplay of genetic and environmental factors was not fully considered in the early attempts at unravelling the genetic basis of this condition.

Two main approaches were historically utilised in the search for the genetic variants that may predispose to multifactorial diseases such as T2D: genome wide linkage and candidate gene association approaches.

1.5.1 Genome-wide linkage studies

Linkage analysis seeks evidence of co-segregation between anonymous genomic markers and the phenotype of interest. Typically, approximately 400 highly variable markers spanning the genome are utilised in looking for regions of DNA shared by affected relatives within families (Dawn Teare et al. 2005). This approach relies on high penetrance of the susceptibility variant to aid detection. The greatest disadvantage of the linkage approach, for complex traits at least, is one of power. Even for variants with substantial (but not extreme) effects on susceptibility at the population level, the numbers of families needed to offer reasonable power to detect linkage is prohibitive (Risch et al. 1996; Merikangas et al. 2003). It is therefore of little surprise that efforts to use linkage in T2D have failed to generate a clear view of genomic patterns of susceptibility. Loci reported in individual studies have usually proved difficult to replicate in others: chance, differences in the types and ethnic origins of families studied and the analytical methods deployed have also played their part (Guan et al. 2008).

In T2D, 18 different regions across the genome have been replicated at least 10 times in genome wide linkage studies (Lillioja et al. 2009). These regions are located on chromosomes 1q, 2q, 2p, 6q, 7q, 3q, 18p, 8p, 9q, 10q, 11q, 14q, 17q, 17p, 19q, 19p, 20q and 22q. These linkage findings played a part in the identification of T2D-associated variants in *TCF7L2* (see below) (Grant et al.

2006). Similarly, evidence of linkage of T2D to chromosome 2q in Mexican Americans led to identification of susceptibility variants within *CAPN10* (Horikawa et al. 2000), though the effect of these variants on T2D susceptibility in other populations remains controversial (Weedon et al. 2003).

1.5.2 Candidate gene studies

Given the low power of linkage approaches, there are good reasons to favour the application of methods based upon the use of association in the search for disease susceptibility loci (Risch et al. 1996). However, association mapping is reliant on linkage disequilibrium which, in human populations, typically extends only a few tens of kilobases. Until recently, this has restricted the scope of association-mapping efforts to the assessment of a limited number of variants selected through a candidacy-based approach. For T2D, many hundreds of such candidates have been selected for study, though very few genes have been tackled comprehensively or in data sets large enough to provide convincing inference. Many apparent associations have been reported, but very few have been replicated (Hattersley et al. 2005).

There are several factors that account for the relative failure of this approach, until recently at least. Firstly, these studies were largely underpowered due to small sample sizes as they often overestimated the likely effect size of the loci under investigation (McCarthy et al. 2009). Secondly, as the pathophysiology of the condition was relatively limited, the candidate gene selected was often a poor choice in retrospect. The thresholds for reporting of significant association were also usually set at $p<0.05$, in the early candidate gene studies. Considering the current thresholds for genome wide level of significance ($p<5x10^{-8}$), it can be seen that the initial

reported p values were not stringent enough and hence the reported signals not always replicated (McCarthy et al. 2009).

Association testing using candidate genes has had two major successes in T2D. Common variants in *PPARG* (Pro12Ala [P12A]) and *KCNJ11* (Glu23Lys [E23K]) have been shown to be associated with T2D in several studies, incorporating several thousand individuals (Altshuler et al. 2000; Gloyn et al. 2003). Though the effect size is modest in each case (OR ~1.20), the high prevalence of the disease associated variant in European populations (85% for P12A and 40% for E23K), equates to an appreciable effect size at the population level. Both variants had been selected for association testing based previous reported (albeit inconclusive) association with increasing T2D risk in smaller studies (Aguilar-Bryan et al. 1995; Hansen et al. 1997; Deeb et al. 1998; Hani et al. 1998; Clement et al. 2000; Gloyn et al. 2001). Both genes have also been implicated in the pathogenesis of monogenic forms of diabetes (Barroso et al. 1999; Hegele et al. 2002; Gloyn et al. 2004).

1.5.3 Genome Wide Association Studies

Prior to the current wave of genome-wide association studies (GWAS) and large meta-analyses (described below), variants in *TCF7L2* were confirmed as increasing T2D susceptibility risk (Grant et al. 2006). The identification of susceptibility variants within this gene came about as a result of fine mapping efforts aimed at the clarification of the linkage signal on chromosome 10 (Grant et al. 2006). However, the susceptibility variants identified thus far in this gene do not explain the linkage signal seen on this chromosome and therefore variants not captured by the GWA approach, such as low-frequency variants, may account for this linkage signal (if it exists).

The limited successes with the linkage and candidate gene approaches are in sharp contrast to the gains made through genome wide association scans over the past few years (Saxena et al. 2007; Scott et al. 2007; Zeggini et al. 2007; Zeggini et al. 2008; Shu et al. 2010; Tsai et al. 2010; Voight et al. 2010; Yamauchi et al. 2010; Kooner et al. 2011; Cho et al. 2012; Morris et al. 2012; Saxena et al. 2012). The principle rationale of GWAS is the common disease, common variant hypothesis, which stipulates that common disease will, at least in part, be attributable to variants that are found in 1-5% of the population (Collins et al. 1997; Collins et al. 1998; Reich et al. 2001). GWAS also rely on linkage disequilibrium (LD) at a population level (Wang et al. 2005). This is the non-random association between alleles at different loci. In general, the further apart that two loci reside, the weaker their LD (Hill et al. 1968). Linkage disequilibrium typically stretches over several kilobases (Reich et al. 2001). This phenomenon allows for a particular locus to be used as a marker for other variants that it is in LD. Hence 500,000 single nuclear polymorphisms (SNPs) can be used to capture the greater than 10,000,000 SNPs that are commonly variant in European samples (International HapMap Consortium 2005).

There are three main reasons for the gains made through the GWA approach. Firstly, efforts such as the HAPMAP project ("The International HapMap Project" 2003) have allowed better understanding of the patterns of sequence variation within and between populations. Secondly, as it has become recognised that variants of modest effect size are likely to be of great importance in the genetics of multifactorial traits, samples of adequate size have been collected which allow well powered association studies to be carried out. Finally, advances in genotyping technology which utilise array or bead technology, have allowed at least a million genotypes to be generated per assay (Fan et al. 2006).

GWA studies have successfully confirmed a large number of T2D susceptibility variants (Saxena et al. 2007; Scott et al. 2007; Zeggini et al. 2007; Zeggini et al. 2008; Dupuis et al. 2010; Qi et al. 2010; Shu et al. 2010; Tsai et al. 2010; Voight et al. 2010; Kooner et al. 2011; Cho et al. 2012; Morris et al. 2012; Saxena et al. 2012). There are currently (September 2012) 67 loci that have been implicated in T2D pathogenesis largely through testing of common variants in GWA studies (Morris et al. 2012)(Appendix, Table S1.1). The majority of these association signals have been robustly confirmed in multiple replication samples. The initial GWA studies were largely from samples of European descent (Saxena et al. 2007; Scott et al. 2007; Zeggini et al. 2007; Zeggini et al. 2008; Voight et al. 2010) , but more recently studies in non-European samples have also been completed (Shu et al. 2010; Kooner et al. 2011; Saxena et al. 2012).

However, despite the large number of common variants found to be associated with T2D, collectively they explain a relatively modest proportion of the heritability of T2D [λ_s=1.104 (Morris et al. 2012)]. Despite the relatively small proportion of the heritability of T2D being explained, these studies have demonstrated some insights into the pathophysiology of the condition.

For a large portion of these variants, the pathophysiological mechanisms leading to the development of T2D are not clear. For instance, some loci are not within the known boundaries of any gene. It is postulated that a variant in LD with the reported locus, is responsible for the observed association. Further fine-mapping efforts will be required to identify the causative locus.

For many loci clear pathophysiological tracts to the development of T2D have emerged. Although for some variants such as *KCNJ11* and *SLC30A8* clear biological mechanisms have been described, for the majority of variants the underlying biological mechanisms are not always clear. Efforts aimed at deciphering these mechanisms have included association analysis with associated traits such as insulin resistance (as measured by HOMO-IR) and continuous glycaemic measures (Voight et al. 2010; Morris et al. 2012). Expression analysis has also been used to highlight possible mechanism through which the identified T2D associated loci mediate their effect on disease state. This approach for instance has highlighted expression of *KLF14*A in adipose tissue hence indicating a possible mechanism for effect on T2D risk (Voight et al. 2010).However, the majority of signals appear to be involved in pancreatic beta-cell function and only a few are involved in insulin sensitivity (Morris et al. 2012). Broadly, the T2D associated loci can be divided into three categories based on how they are likely to manifest their effects on T2D risk (if known):

I. Loci manifesting effects though increased effect on obesity: *FTO* and *MC4R*
II. Loci manifesting their effect through insulin resistance without an effect on obesity: *PPARG*, *KLF14*, *ADAMTS9* , *GRB14 and AKNRD55*
III. Loci manifesting their effect through insulin secretion/ beta cell dysfunction: *TCF7L2, CDKN2A/B, IGF2BP2, HNF1B, THADA, CCNE2*, *KCNQ1*, *CAMK1D*, *MTNR1B*, *HHEX/IDE, JAZF1, CDKAL1, CENTD2, SLC30A8, and KCNJ11.*

1.5.4 Clinical implications of T2D signals identified to date.

Although the genetic loci implicated thus far in the pathogenesis of T2D can be used to calculate the genetic risk score for those that carry the greatest and least number of T2D susceptibility variants (Lango et al. 2008), this is not useful in the clinical setting. Clinical parameters, such as BMI, hypertension, hypertriglyceridaemia are far better predictors of the future development of T2D (McCarthy et al. 2009). The greatest impact of the variants thus far discovered is in highlighting pathophysiological pathways that may be targets of specific treatment modalities. It is of interest that variants within *KCNJ11* and *PPARG* have been associated with increased risk of T2D, as their protein products are targets of successful therapies for T2D. Further observations have been that patients that harbour susceptibility variants within *TCF7L2* have an altered response to sulphonylureas resulting in earlier treatment failure (Pearson et al. 2007; Holstein et al. 2011).These observations raise the possibility that, with improved understanding of the genetic susceptibility to T2D, personalised medicine may indeed be an achievable goal in the future.

1.6 A role for low frequency intermediate penetrance variants in Type 2 Diabetes pathogenesis

Genome wide association studies have generated valuable insights into pathogenesis but explain only a small proportion of the observed familial aggregation of T2D. The sibling relative risk (λs) attributable to all known T2D susceptibility variants combined is ~1.11, well-below the epidemiological estimate [~3] (Lango et al. 2008; Morris et al. 2012).

Several differing theories have been reported to account for this heritability "gap". Firstly, the possibility of the overestimation of the heritability of T2D needs to be considered (Manolio et al. 2009). Moreover, the proportion of this heritability that may be attributable to epigenetic factors or shared environments need to be taken into consideration (Ling et al. 2009). Although it would be difficult to quantify the contribution of these factors to the heritability of T2D, it would be reasonable to consider that these factors do account for at least part of the heritability seen for T2D (Ling et al. 2009; Nadeau 2009; Eichler et al. 2010).

Secondly, structural genetic defects such as copy number variations (insertions or deletions) or copy neutral variations (inversions and translocations) have been proposed as possible factors in the heritability of T2D. It is noted that an estimated 8% of the population carry rare (allele frequencies <0.05%), relatively large (greater than 500 kb) deletion or duplications (Eichler et al. 2010; Itsara et al. 2010). These structural defects have proved technically difficult to study due to a large number of artefacts and secondary false positive associations being observed (Craddock et al. 2010). However, the genome wide association study by Craddock et al, incorporating 16000 cases and 3000 control samples, did confirm 3 loci where CNV's were associated disease states, including one for T2D (*TSPAN8*). This locus had been previously identified by SNP GWAS (Zeggini et al. 2008). It is therefore postulated that CNV's that can be captured by current SNP platforms, are unlikely to contribute to complex disease pathogenesis (Craddock et al. 2010) as they would have been detected by the previous wave of GWAS. However, this finding does not out rule CNV's that are not captured by current GWAS platforms playing a part in complex disease pathogenesis.

Thirdly, gene-gene interactions (epistasis) have been proposed as another pathway that may explain a portion of the heritability if T2D (Bell et al. 2011). Epistatic effects can be complex to study and can be affected by single-locus association that can mask multi-locus findings (Bell et al. 2011). Although a weak association for epistatic effects was seen with T2D risk in one study (Bell et al. 2011), this was not at a genome wide level of significance. It is postulated that much larger studies, with increased power are required to evaluate fully the contribution of epistasis to T2D risk (Bell et al. 2011).

Another possible explanation for this heritability gap may be multiple common variants with small effect sizes (Park et al. 2010). It is postulated that further large scale GWAS with even larger number of samples than in current studies may implicate more common variants in the pathogenesis of T2D and other complex diseases (Park et al. 2010). However, it is unlikely that that these measures will fully explain the heritability of T2D (Manolio et al. 2009).

My thesis is based on the following premise: low-frequency (arbitrarily designated 1-5%) susceptibility-variants with intermediate penetrance (OR ~2-4), which at the time of project design and implementation, had remained largely-undetected by conventional approaches to susceptibility gene discovery, represent an important and neglected source of genome sequence variation.

To illustrate this statement, consider a model LFIP variant with a minor allele frequency (MAF) of 1% and an additive per allele relative risk of 3.0 (i.e. each copy of the risk allele increases the disease risk three fold). Using standard population genetics models it is possible to show that such a variant would have the following characteristics:

1) For a disease, such as T2D with prevalence of 5%, heterozygote carriers would have a penetrance of ~14% and rare homozygotes would have a penetrance of ~43% [calculated using CaTS Power Calculator (Skol et al. 2006)]. This trait does not conform to Mendelian patterns of segregation. Such a variant would tend to present within families with multiply-affected individuals, but would not meet criteria for inclusion in most collections of monogenic families. Even if such variants were captured, the causal variant would not be detected using traditional Mendelian linkage approaches because the low penetrance means that co-segregation would be hard to demonstrate;

2) The locus-specific sibling relative risk for this variant (λ_s) exceeds that of the best common variants known (1.038 vs. 1.025 for *TCF7L2;* calculated using CaTS Power Calculator) indicating the impact on familiality (and predictive power) that such variants could have. Fewer than 20 such variants genome-wide would generate a combined λ_s>2 and achieve respectable discriminative accuracy (area under receiver operative curve (ROC) ~72%). Thirty such variants would provide a discriminative accuracy of 77%, offering the prospect for useful clinical prediction and diagnostics;

3) The low MAF would mean that such a variant would have a very low chance of being detected through genome-wide association studies. These are largely targeted towards the identification of susceptibility variants with MAF > 5%. Low-frequency variants are poorly captured by this approach and many such studies exclude variants with MAF<1% from further analysis. Thus, whilst this model variant can readily be detected by association (see point 4 below), it would likely be missed by genome-wide association

efforts; However more recently, as a result of denser GWAS and imputation from sequence based reference panels, these variants are becoming more accessible (Huang et al. 2012; Magi et al. 2012; Voight et al. 2012).

4) Although current genome-wide approaches are not well-suited to the detection of a variant with these properties, the sample size required to demonstrate an association for such a variant is remarkably modest: only 2744 cases and 2744 control subjects would be required to achieve 90% power to detect such a variant for a very stringent α of 10^{-9} (in the setting of low frequency variants a more stringent α [type 1 error] than that for common variants [$\alpha < 5x10^{-8}$,(McCarthy et al. 2009)] in GWA is probably most appropriate].

There are several reasons to believe that variants with such characteristics are likely to be important in disease susceptibility.

1) Under population genetics theory (Crowe 1970), there is no reason to assume that variants must be either rare and highly penetrant or common and of low penetrance.

2) Secondly, low frequency variants have been shown to be contributing to the pathogenesis of other complex disease such as inflammatory bowel disease [*CARD15, CARD9, NOD2, CUL2 and IL18RAP*] (Hugot 2006; Rivas et al. 2011), sick sinus syndrome [MYH6} (Holm et al. 2011) hypertriglyceridaemia [*PCSK9* and *ANGPTL4*] (Cohen et al. 2006; Romeo et al. 2007) breast cancer [*CHEK2, ATM3, BRIP1, PALB*] (Xia et al. 2006), obesity [*MC4R*] (Larsen et al. 2005) and type 1 diabetes (T1D) [*IFIH1*] (Nejentsev et al.

2009). In the case of T1D, resequencing of genes implicated in the pathogenesis of this condition through common variants has identified four novel low frequency variants in *IFIH1* that showed association to T1D (Nejentsev et al. 2009). One of the variants (rs35667974), in exon 14 of *IFIH1*, achieved genome wide-association level of significance in subsequent large scale genotyping (Nejentsev et al. 2009). This is of particular interest for my project as I have also used a similar approach of screening genes implicated in T2D pathogenesis, in an effort to discover LFIP variants. Recently sequencing of the exonic regions of the *MTNR1B* gene that had previously shown association with T2D risk through common non coding variants (Section 1.3.1) has identified a large number of low frequency variants (Bonnefond et al. 2012). However, none of the coding variants identified in this gene achieved a genome wide association level of significance ($\alpha<5x10^{-8}$) with T2D risk (Bonnefond et al. 2012). Four variants with a MAF<0.1%, that were predicted to result in either complete or partial loss of function of the melatonin receptor did demonstrate a modest association with T2D risk [OR = 5.67 (2.17–14.82); P = 4.09×10^{-4} (Bonnefond et al. 2012)].

3) Thirdly, as discussed above, neither the GWA nor linkage approaches have been able to account fully for the familial aggregation seen in T2D. However, it is noted that recent statistical modelling using Bayesian inference analysis, has demonstrated that up to 49% of T2D disease risk may be amenable to discovery by GWAS SNPs (Stahl et al. 2012). It is therefore possible that other variants such as multiple LFIPs may account for the replicated T2D linkage signals seen on Chr 1, 10, 12, and 20, which have proved refractory to GWA scans. Whilst a single LFIP variant may not generate enough signal for

linkage, the co-localization of 2-3 within a single locus, might do so [such as that seen in *CARD15*, *NOD2* and *CARD9* referred to above(Hugot 2006; Rivas et al. 2011)].

In conclusion, there are several reasons to assume that LFIP variants exist and that they may explain a portion of the heritability gap seen for T2D. The characteristics of such variants have rendered them largely invisible to the initial wave of common variant GWAS. Denser GWAS and subsequent imputation may identify further low frequency variants. However, deep resequencing offers great promise for uncovering potential LFIP variants. The advent of high throughput sequencing approaches that have dramatically reduced the cost and time taken for deep resequencing means this approach can now be utilised.

1.7 High Throughput Sequencing

Over the last few years, great advances have been in the development of high throughput sequencing technology. There are several different platforms now available that differ in their biochemistry and read lengths achieved. The majority of these platforms require amplified segments of DNA for the sequencing reaction ("Second Generation" next generation high throughput sequencing ; NG-HTS), though the latest technologies are able to achieve de novo sequencing from a single molecule of DNA ("Third Generation" NG-HTS). This project was designed and completed between 2007 and 2009 and third generation NG-HTS platforms were not commercially available. I utilised the Genome Analyzer II platform (Second Generation – NG-HTS), due to local availability at the time of project commencement, for the studies outlined in Chapter 6.

1.8 Strategy employed

A logical place to initiate the search for LF variants influencing multifactorial T2D lies in exploring those genes already implicated in diabetes pathogenesis because they contain rare mutations causal for monogenic forms of diabetes, or variants associated with multifactorial T2D. In genes implicated due to their role in monogenic diabetes, there is already evidence that large-effect mutations are compatible with life, and that they result in a phenotype with substantial similarities (and clinical overlap) with T2D. It is likely therefore that variants with less dramatic effects on function and/or expression, where they exist, result in less extreme clinical phenotypes including multifactorial T2D. Since existing GWA study data effectively rule out the possibility of common variants of medium- to large-effect size, variants with such effect sizes are also likely to be rare or of low frequency. Low frequency variants have been implicated in the pathogenesis of other complex diseases, such as T1D (Nejentsev et al. 2009), although their contribution to T2D predisposition is as yet uncertain.

I utilised three approaches in the search for LFIP variants. Firstly, I used traditional Sanger resequencing in a targeted region of *HNF1A* to evaluate the presence of LFIP variants. Secondly, I carried out large scale genotyping to evaluate the role of candidate LFIP variants in T2D pathogenesis in two separate projects. Finally, high throughput sequencing was utilised in the search for LFIP variants by resequencing ten genes previously implicated in diabetes pathogenesis either through their common variants in T2D, or rare, penetrant variants in monogenic forms of diabetes.

1.9 Aims of thesis

The overall aim of this project was the investigation of the contribution of low frequency, intermediate penetrance variants in Type 2 Diabetes pathogenesis. More specifically, each study undertaken in this project has its specific aims as outlined below.

I. To evaluate the contribution of low frequency intermediate penetrance variants in exons only encoding isoform A of *HNF1A* to T2D risk (Chapter 3).

II. To evaluate the role in T2D pathogenesis of candidate LFIP variants in *HNF1A*, *HNF4A*, *PDX1* and *KCNJ15* that previous published works had highlighted as potentially impacting on T2D risk (Chapter 4).

III. To establish the role played in T2D pathogenesis by LFIP variants in *LARS2* by large scale genotyping (Chapter 5).

IV. To utilise "Next generation" targeted sequencing platform for discovery of novel low frequency variants that may potentially impact on T2D risk using exonic resequencing of ten genes previously implicated in T2D or monogenic diabetes pathogenesis (Chapter 6).

CHAPTER 2

Materials and Methods

2.1 Subjects included

Two sample groups were used in this project. **Group 1 samples** were comprised of samples from the UK that I utilised for the genotyping studies described in Chapters 4 and 5, the Sanger sequencing project described in Chapter 3 and the targeted sequencing project described in Chapter 6. **Group 2 samples** were those that were genotyped by collaborators in other centres and the genotyping data or the summary statistics were made available for inclusion in the studies described in Chapters 4 and 5.

2.1.1 Group 1 samples

UK sample 1 (UK1)

This sample was comprised of the United Kingdom Type 2 Diabetes Genetics Consortium case–control sample, comprising of a maximum of 5000 T2D cases and 6470 controls ascertained in Tayside, Scotland. All individuals were of white European ancestry. Control subjects were under 80 years old at recruitment and did not have a diagnosis of T2D. Control subjects with fasting blood glucose greater than 7.0 mmol/L or HbA1c value greater than 6.4% were excluded. All T2D subjects were diagnosed between the ages of 35 and 70. T2D subjects were defined as those that required hypoglycaemic agents at recruitment or those that were treated with diet alone but had hyperglycaemia in keeping with a diagnosis of T2D based on WHO criteria (World Health Organization 1999). Individuals with Type 1 diabetes (T1D) were excluded on the basis of previous diagnosis of T1D, Anti Glutamic Acid Decarboxylase (GAD) antibody positivity or the requirement of insulin therapy within 1 year of diagnosis. A subset of this sample has been

previously used as part of the genome wide association studies carried out in the UK (Wellcome Trust Case Control Consortium 2007; Zeggini et al. 2007). The enlarged sample used here represents continuing recruitment to this resource under precisely the same criteria. This sample was utilised for the genotyping of variants described in Chapters 4 and 5.

UK sample 2 (UK2)

UK sample 2 consisted of a maximum of 1853 T2D cases ascertained as part of the Diabetes UK Warren 2 collection and 10220 control samples. The latter represent the full British 1958 Birth Cohort (n=7133) and the United Kingdom Blood Services Collection of Common Controls (n=3087), a subset of which featured in the Wellcome Trust Case Control Consortium (WTCCC) genome-wide association scan (Wellcome Trust Case Control Consortium 2007; Zeggini et al. 2007). Both samples were collected throughout the UK. The British 1958 Birth Cohort collection represents a population based collection of individuals born in the same week in 1958. DNA sample collection was carried out when the individuals were 44-45 years of age. The United Kingdom Blood Services Collection of Common Controls represents samples ascertained from Blood Donors throughout the UK between September 2005 and February 2006.

The Diabetes UK Warren 2 collection was ascertained in 6 centres (Exeter, London, Oxford, Cambridge, Norwich and Newcastle) and all subjects were of Irish or British ancestry. The samples in this collection represent three related collections; namely (a) probands from the Warren 2 sibpair studies (at least one family member with T2D; age at diagnosis for probands 35-70 years), (b) cases from the Warren 2 case collection (age at diagnosis 35-65 years) and (c) probands from the Warren 2 trios collection (parent- offspring triad- DNA from both parents or

one parent and 2 siblings were available as well as the proband; age at diagnosis 25-65 years). All subjects were classified as having T2D based on treatment with oral hypoglycaemic agents or lack of treatment with insulin at diagnosis. Individuals being treated with diet alone were included, provided they had hyperglycaemia in keeping with T2D as per the WHO criteria (World Health Organization 1999). Individuals with T1D were excluded on the basis of previous diagnosis of T1D, Anti Glutamic Acid Decarboxylase (GAD) antibody positivity or the requirement of insulin within 1 year of diagnosis.

The "UK2" sample was utilised for the genotyping of variants described in Chapters 4 and 5. Five hundred subjects from the 1958 Birth Cohort collection that had been included as part of the WTCCC GWAS were selected at random for inclusion in the targeted sequencing project described in Chapter 6. One hundred and fifty probands from the Warren 2 sibpair studies, and 150 probands from the Warren 2 trios collection were included for the targeted sequencing project described in Chapter 6.

Young Diabetes in Oxford Cohort (UK3)

This sample includes subjects diagnosed with T2D between 18 and 45 years of age, recruited from Oxfordshire GP practices. Those with T1D were excluded by requiring that all subjects had no permanent requirement for insulin therapy within 12 months of diagnosis and no evidence of islet autoimmunity [Glutamic Acid Decarboxylase (GAD) antibody levels <14 WHO units/ml] (Bingley et al. 1997). The subjects classified as having T2D did not meet current clinical criteria for MODY diagnostic testing (Ellard et al. 2008) and all had demonstrable fasting C-peptide levels (mean 0.83 ± 0.44 nmol/L [non-diabetic reference range: 0.11-0.61 nmol/L]). Individuals from this

sample (n=86) were utilised in the sequencing of the latter 3 exons of *HNF1A* (Chapter 3) and 200 samples were included for the targeted sequencing project described in Chapter 6.

2.1.2 Group 2 samples

Netherlands sample 1 (NL1)

The first sample from the Netherlands was comprised of the Hoorn Study (NL1) samples (Mooy et al. 1995). This is a Dutch population-based study from the city of Hoorn, in the north-west of the Netherlands, from which 519 participants with normal glucose tolerance (NGT) and 480 T2D patients were included for the genotyping of the *LARS2* variants. Glucose tolerance was assessed using a fasting OGTT, according to 1999 WHO criteria(World Health Organization 1999). This sample was utilised for the genotyping efforts of the *LARS2* variants (Chapter 5).

Netherlands sample 2 (NL2)

The second Dutch sample (NL2) included 1517 controls and 821 T2D patients (Reiling et al. 2009; van 't Riet et al. 2010). The 1517 controls were randomly selected from the New Hoorn Study (NHS), which is an on-going, population-based study from the city of Hoorn which does not overlap with the NL1 sample. Of the T2D patients, 147 were from the NHS and the remainder (n=674) were recruited from the diabetes clinics of the Leiden University Medical Centre and the Vrije Universiteit Medical Centre, Amsterdam. All participants in this replication sample were Dutch whites. All NGT participants underwent an OGTT and were classified according to WHO

criteria (World Health Organization 1999). This sample was utilised for the genotyping efforts of the *LARS2* variants (Chapter 5).

Netherlands sample 3 (NL3)

The third sample from the Netherlands was ascertained from the Breda study (NL3) (van Tilburg et al. 2003; Monsuur et al. 2005). The case and control samples were recruited from the city of Breda, in the south of the Netherlands. The 920 controls were from the Dutch blood bank and self-reported a non-diabetic state. The 501 cases had T2D diagnosed on the basis of WHO criteria (World Health Organization 1999). This sample was utilised for the genotyping efforts of the *LARS2* variants (Chapter 5).

Netherlands sample 4 (NL4)

The fourth replication sample from the Netherlands was comprised of 5183 NGT participants and 1222 T2D patients from the population-based ERGO study (NL4) from Rotterdam in the south-west region of the Netherlands (Hofman et al. 1991). This sample was utilised for the genotyping efforts of the *LARS2* variants (Chapter 5).

Denmark Sample 1 (DK1)

This sample was a collection of samples from Inter99 (360 cases, 4520 controls), Steno Diabetes Center 1 (SDC1) (47 cases, 345 controls) and Steno Diabetes Center 2 (SDC2) (990 cases). Patients with diabetes due to known chronic pancreatitis, haemochromatosis, severe insulin resistance,

MODY, maternally inherited diabetes and deafness, family history of first-degree relatives with T1D, insulin requirement within the first year after diabetes diagnosis or a fasting serum C-peptide level ≤ 150 pmol/liter at time of recruitment were excluded from the category of clinically defined T2D. NGT participants underwent an OGTT according to WHO criteria (World Health Organization 1999). This sample was utilised for genotyping of the *HNF4A* and *KCNJ15* variants (Chapter 4).

Denmark sample 2 (DK2)

This sample consisted only of cases and consisted of a further 1020 cases from Steno Diabetes Centre 2 sample (SDC2) and 1626 from ADDITION (Anglo-Danish-Dutch Study of Intensive Treatment in People with Screen-Detected Diabetes in Primary Care: ClinicalTrials.gov ID-no: NCT00237548) (Lauritzen et al. 2000). For all samples, patients with diabetes due to known chronic pancreatitis, haemochromatosis, severe insulin resistance, MODY, maternally inherited diabetes and deafness, family history of first-degree relatives with T1D, insulin requirement within the first year after diabetes diagnosis or a fasting serum C-peptide level ≤ 150 pmol/liter at time of recruitment were excluded from the category of clinically defined T2D. This sample was utilised in the genotyping of the *HNF4A*, *LARS2* and *KCNJ15* variants (Chapters 4 and 5).

Denmark sample 3 (DK3)

This sample consisted of 514 NGT controls randomly selected from public registers at the Steno Diabetes Center and the Research Centre for Prevention and Health, Copenhagen, Denmark. The 706 T2D patients were recruited from the Steno Diabetes Center. NGT participants underwent an

OGTT according to WHO criteria (World Health Organization 1999). This sample was utilised for the genotyping efforts of the *LARS2* variants (Chapter 5).

Finland sample 1 (FL1)

Finland sample 1 was a case–control sample from the Botnia region of Finland, consisting of 353 controls and 402 T2D patients. This sample was utilised for the genotyping efforts of the *LARS2* variants (Chapter 5) and summary statistics form this variant were included in the meta-analysis of the A98V variant of *HNF1A* (Chapter 4).

Finland sample 2 (FL2)

Finland sample 2 was comprised of a set of trios originating from the Botnia region of Finland and consisted of 211 probands (multiple diabetic sibs) and 370 parents (Saxena et al. 2006; Florez et al. 2007). This sample was utilised for the genotyping efforts of the *LARS2* variants (Chapter 5).

Sweden sample 1 (SE1)

The sample from Sweden, SE1, was from a case–control study from Skara and Malmö, and consisted of 468 controls and 480 T2D patients. This sample was utilised for the genotyping efforts of the *LARS2* variants (Chapter 5).

FUSION samples

FUSION sample 1 ("FS1", n=1160 cases, 1173 controls) represents samples included as part of stage 1 genotyping of the DIAGRAM meta-analysis and was recruited in Finland (Scott et al. 2007; Zeggini et al. 2008). T2D samples were defined as those with a fasting glucose ≥ 7.0 mmol/L or 2 hour plasma glucose concentration ≥ 11.1 mmol/L, or use by use of diabetes medications (Scott et al. 2007). Normal glucose tolerance was defined as fasting glucose < 6.1 mmol/l and 2-h glucose < 7.8 mmol/L (Scott et al. 2007). This sample was genotyped using the Illumina HumanHap300 BeadChip (version 1.1) and was included in the FUSION GWAS for T2D (Scott et al. 2007). This sample was included in the analysis of the T130I variant of *HNF4A* (Chapter 4). Although T130I was included in this GWAS, the genotyping data for this variant was not specifically published.

FUSION sample 2 ("FS2", n=1211 cases, 1264 controls) represents samples included as part of stage 2 genotyping of the DIAGRAM meta-analysis (Zeggini et al. 2008) and was also recruited in Finland. T2D and normal glucose tolerance were defined in the same manner as for FUSION sample 1. This includes samples from the Dehko 2D, Health 2000, Finrisk 1987, Finrisk 2002, Savitaipale Diabetes, and Action LADA studies. Sequenom Homogeneous Mass Extended or iPlex Gold SBE assays were used for genotyping of this sample by collaborators in FUSION. T130I was genotyped in this sample as part of a lipid GWAS and subsequent replication (Kathiresan et al. 2009). This sample was also included in the analysis of the T130I variant of *HNF4A* (Chapter 4). However, the genotyping data for the T130I variant in relation to T2D in this sample have not been previously published.

METSIM sample

The METSIM (METabolic Syndrome In Men; "MS1") study is an ongoing population based collection aiming to investigate metabolic syndrome, T2D, cardiovascular disease, and cardiovascular risk factors, with the goal of collecting 10,000 men, randomly selected from the population of the town of Kuopio in Eastern Finland (population 95,000) (Stancakova et al. 2009). T2D and normal glucose tolerance were defined in the same manner as FUSION samples 1 and 2. This sample was included as part of a lipid GWAS follow up (Kathiresan et al. 2009). This sample was included as part of the analysis of the T130I variant of *HNF4A* (Chapter 4).

Broad samples

Broad sample 1 (BR1, "Canada sample" in original paper, n=127 cases, 127 controls), was recruited from the Saquenay Lac-St. Jean region in Quebec, Canada. Broad sample 2 (BR2, "Scandinavia sample", n=490 cases, 490 controls) were recruited in Finland and Sweden and Broad sample 3 (BR3, "Sweden sample", n=514 cases, 514 controls) was recruited in Sweden. All samples from the Broad Institute were individually matched for age, geographical location and BMI and were genotyped using the Sequenom platform (Winckler et al. 2005). T2D and normal glucose tolerance status were defined as per World Health Organization 1998 criteria for T2D (Winckler et al. 2005). These samples were included as part of the analysis of the T130I variant of *HNF4A* (Chapter 4). The genotyping of the T130I variant has been previously reported for these samples (Winckler et al. 2005).

All studies were approved by the appropriate medical ethical committees and were in accordance with the principles of the Declaration of Helsinki. All participants provided written, informed consent for this study. A table of the study subject characteristics has been included in the respective chapters that utilised the samples included above.

2.2 Methods

2.2.1 DNA and PCR product quantification

Quantification of DNA and Polymerase Chain Reaction (PCR) products (Section 2.5) in this project was carried out by use of Nanodrop 2000 (Thermo Scientific, Wilmington, DE) and Quant-iT PicoGreen double stranded DNA (dsDNA) assay kit (Invitrogen, Paisley, UK). Nanodrop 2000 determines the concentration of the DNA sample by measuring absorbance at 260nm and was utilised to obtain an initial estimate of the concentration of DNA samples and PCR products in this project. However, as the presence of single stranded DNA and RNA can lead to over-representation of the concentration obtained by this approach, the Quant-iT PicoGreen dsDNA assay kit was utilised to obtain a more accurate measure of the double stranded DNA (dsDNA) content of the samples being tested. The reagent for the latter assay is a fluorescent nucleic acid stain that can be utilised to quantitate the concentration of dsDNA in a solution.

A fresh working solution was prepared at the start of the experiment by a 200 fold dilution of the Picogreen reagent by mixture with 1X TE, having allowed the reagent to first reach room temperature. The resulting mixture was protected from sunlight as it can undergo photo-degradation. During this process the fluorescence of the samples being tested was compared to a

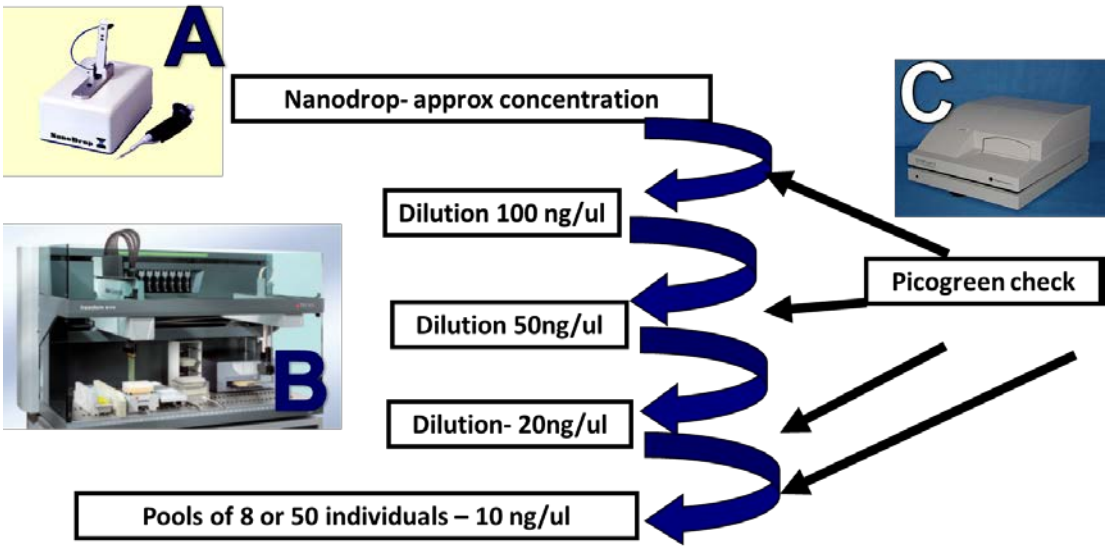
standard curve of fluorescence of known dsDNA concentrations. A stock solution of calf thymus dsDNA was prepared to construct the standard curves for use in this experiment. To limit sample loss, a 1:99 dilution of the DNA and PCR product to be quantified was used. Once the concentration of the 1:99 diluted samples was obtained, the correct concentration for the source samples was calculated. A 50µl aliquot of the 1:99 diluted samples was added to 50 µl of Picogreen working assay in duplicate in a Corning Costar 96-well polycarbonate plate (Corning Life sciences, Corning, NY) and allowed to incubate for 6 minutes and protected from sunlight. The fluorescence was subsequently read using a Cytofluor 4000 multi-well Plate Reader (Applied Biosystems, Foster City, CA), using fluorescein wavelength for excitation of 480 nm and 520 nm for emission. The Cytofluor software programme then generated an Excel (Microsoft Corporation, Seattle, USA) worksheet and an average of the concentrations for the samples that were prepared in duplicate was obtained.

2.2.2 DNA pooling

To allow a more efficient use of the Genome Analyzer II (GAII) capacity (Section 2.8) the DNA samples included in the targeted sequencing project (Chapter 6) were pooled. For the pilot target sequencing project (Section 6.3) the DNA samples were aggregated in pools of 8 individuals (24 case pools/ 6 control pools) and for the main targeted sequencing project (Section 6.4) the DNA samples were aggregated in groups of 50 (10 case and 10 control pools). Details of the samples included in each DNA pool have been included in Chapter 6. To ensure that each individual in the pool was proportionally represented, very accurate quantification of the DNA samples was carried out using a PicoGreen assay (Section 2.1). This was performed in a stepwise fashion with the DNA samples first diluted to 100 ng/µl, followed by dilution to 50ng/µl and 20ng/µl (Figure

2.1). The samples were then aggregated in equimolar fashion in pools of 10ng/µl. To expedite the pipetting and to minimise liquid handling errors, robotics (Freedom EVOware 2.1, Tecan Reading, UK) were utilised in pool generation and the serial dilutions of the DNA samples. A PicoGreen assay was used at each step to ensure accuracy of the dilution carried out.

Figure 2.1 Figure depicting the stepwise dilution of the DNA samples used in the targeted sequencing project and final pooling step (pools of 8 samples in the pilot study and pools of 50 samples in the main targeted sequencing study). Figure also demonstrates the Nanodrop 2000 machine (A), the Freedom EVOware 2.1, Tecan robot (B) and the Cytofluor 4000 multi-well Plate Reader (C).



2.2.3 Polymerase chain reaction amplification

All primers for this project were designed using primer3 software (http://frodo.wi.mit.edu/primer3/) and have been included in the corresponding chapters. Polymerase chain reaction (PCR) products were generated using either AmpliTaq Gold DNA polymerase (Applied Biosystems Warrington, UK; Sanger sequencing and pilot high throughput

sequencing, Chapters 3, 4 and 5) or Phusion high fidelity polymerase (New England Biolabs, Hitchin, Hertfordshire, UK; high throughput sequencing, Chapter 6). All PCR reactions were carried out using manufacturers' guidelines and at recommended primer annealing temperatures.

The master mix for the PCR reaction was generated by combining 2.5µl of 10X Buffer (Applied Biosystems), 1.5µl of 25mM Magnesium Chloride (Applied Biosystems), 1µl of 2.5mM Deoxynucleotide triphosphates (Applied Biosystems), 1µl each of the forward and reverse primer (5pmol), 0.1µl of AmpliTaq Gold DNA polymerase (5 units/µl) and 13.9µl of sterile water. For the amplification of the regions incorporated into the high throughput sequencing project described in Chapter 6, 0.2µl of Phusion high fidelity polymerase was used along with 13.8 µl of sterile water and all other reagents were the same as that described above. This master mix was then combined with 4µl of DNA at a concentration of 10ng/µl. The reaction was set up in 96 well non-skirted Thermo-fast PCR plates (Abgene, Epsom, UK) and the PCR reaction was carried out on a thermocycler (MJ Research Peltier Thermocycler- PTC 225). The following cycle conditions were used for the PCR reaction (unless otherwise stated in the corresponding chapter): 1 cycle at 95°C for 12 minutes; followed by 35 cycles of: (i) 94°C for 1 minute, (ii) 62°C for 1 minute, (iii) 72°C for 1 minute; followed by 1 cycle of 72°C for 10 minutes. Success of PCR reactions was assured by use of 2% gel electrophoresis.

2.2.4 Agarose gel electrophoresis

Agarose gel electrophoresis was used to evaluate the success of PCR reactions. Specifically, this method was used to assess the success of amplicons generated as part of: (i) the sequencing of

the latter 3 exons of *HNF1A* (Chapter 3) and (ii) the targeted sequencing using the Genome Analyzer 2 platform (Chapter 6).

The gel was prepared by the mixture of 400µl of 1X Trisborate ethylenediaminetetraacetic acid (EDTA) buffer (TBE), 8 µg of agarose (Sigma-Aldrich, Gillingham, UK) and 8µl of ethidium bromide (Sigma-Aldrich) (10mg/ml). Five microliters of each PCR product together with 5µl of an agarose loading buffer ("Orange G" in 30% glycerol solution) were dispensed into each well of the agarose gel. The gel was subsequently run in an electrophoresis tank containing a 1X TBE buffer together with 8µl of ethidium bromide at 120 volts for 30 minutes. The gel image was subsequently captured using a Bio Rad (Hercules, CA, USA) gel dock.

2.2.5 Capillary Sanger sequencing

Capillary Sanger sequencing was used to sequence the latter 3 exons of *HNF1A*, to evaluate whether any low frequency, incompletely penetrant (LFIP) variants may be present (Chapter 3). I utilised dye terminator capillary sequencing, which represents a modified version of the original chain-terminator sequencing described by Sanger (Sanger et al. 1977). Using this technique, dideoxynucleotides (ddNTPs) as well as deoxynucleotides (dNTPs) were used in the sequencing reaction. The incorporation of the ddNTPs, which lack a hydroxyl group on the 3' end, into the extending DNA fragment, prevents the addition of any further nucleotides and terminates the sequence that is being generated. This produces DNA fragments of differing lengths which can be further characterised by virtue of having a fluorescently labelled ddNTPs incorporated at the end of the sequence. Capillary electrophoresis can also be used to further separate these fragments

by their size based on their electric charge and length. The sequencing reaction was carried out by use of BigDye v1.1 cycle sequencing kit (Applied Biosystems Warrington, UK).

2.2.5.1 Clean-up of PCR products

The clean-up of the PCR products, which were all generated by use of M13 tailed primers (sequences provided in Chapter 3), was initially carried out by treating 1µl of PCR product with a 1µl solution of Exonuclease (New England Biolabs, Hitchin, Hertfordshire, UK) and Shrimp Alkaline Phosphatase (Promega, Southampton, Hampshire, UK). The Exonuclease removes the leftover primers whereas the Shrimp Alkaline Phospatase removes the leftover dNTPs. The solution was centrifuged at 2000 revolutions per minute (RPM) for 2 minutes. The sample was subsequently incubated for 30 minutes at 37°C on a thermocycler (MJ Research Peltier Thermocycler- PTC 225), followed by 15 minute incubation at 80°C to deactivate the enzyme.

2.2.5.2 Sequencing reaction

For a single sequencing reaction 0.25µl of BigDye terminator v1.1 (Applied Biosystems) was combined with 1.875µl of dilution buffer (Applied Biosystems), 2µl of sequencing primer (2.5pmol/µl) and 3.875µl of sterile water. This sequencing master mix was then added to each cleaned up sample. The sequencing reaction was then carried out on a thermocycler according to the following conditions: 25 cycles of (i) 96°C for10 seconds, (ii) 50°C for 5 seconds, (iii) 60°C for 4mins. The samples were again centrifuged briefly and 5µl of sterile water was added to each reaction.

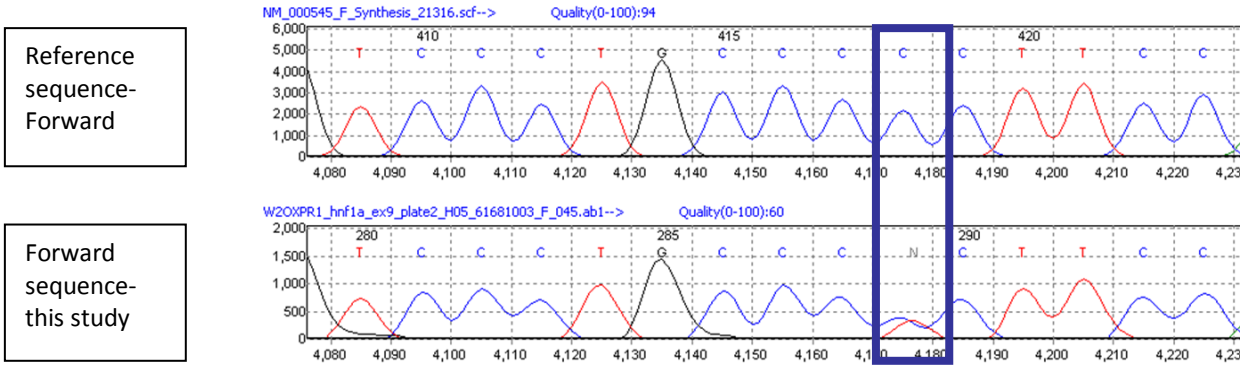
2.2.5.3 Clean-up of samples post sequencing reaction

The clean-up of samples post the sequencing reaction was carried out using the PCR Performa V3 plates (Edge Biosystems, Gaithersburg, MD, USA) using the manufacturers' guidelines. The samples were then dried in a 96 well plate by incubation at 70°C prior to sequencing on the ABI3700 sequencing platform in the Department of Zoology, University of Oxford.

2.2.5.4 Analysis

The resulting sequences were analysed using Mutation Surveyor v 3.2 (Soft Genetics, Cambridge UK) and compared to the appropriate reference sequence obtained from the National Center for Biotechnology Information (NCBI) [http://www.ncbi.nlm.nih.gov/RefSeq/]. Unidirectional analysis mode of Mutation surveyor V3.2 was utilised. This software package has been shown to have a sensitivity of >99% in the unidirectional analysis mode (Ellard S 2009). The accuracy of all calls was further checked by visual inspection of all electropherograms (Figure 2.2).

Figure 2.2 Electropherogram illustrating the heterozygous c.1768+44C>T variant of *HNF1A* (detected in sequencing of the latter 3 exons of *HNF1A*) highlighted in blue box.



2.2.6 Genotyping

In this project two genotyping methods were utilised. Taqman assays (T130I variant of *HNF4A*, I88I variant of *KCNJ15,* rs9825041 variant of *LARS2* and A98V variant of *HNF1A*) on the ABI 7900HT platform (Applied Biosystems, Warrington, Cheshire, UK) were used for the genotyping that I carried out in the Oxford Centre for Diabetes, Endocrinology and Metabolism (OCDEM). KBioscience allele specific PCR (KASPar) assay (V255M variant of *HNF4A,* H324Q variant of *LARS2* and D76N variant of *PDX1*) (KBioscience, Hoddesdon, U.K.) was used for genotyping that was carried out by KBiosciences (Kbiosciences, Hoddeston, UK).

Taqman genotyping (Figure 2.3) is dependent on a 5' nuclease PCR assay (Livak 2003). Each Taqman assay contains two primers to amplify the region of interest and two probes associated with fluorescent dyes (FAM and VIC) on the 5' end to detect the allele present at the site (Livak 2003). A nonfluorescent quencher is at the 3' end of the probe. Cleavage of the probe during PCR amplification releases the fluorescent dye which is then detected by the ABI 7900HT system

(Applied Biosystems, Warrington, UK) and the allele present called. The presence of the quencher dampens the fluorescence of the probe if it is not cleaved.

The DNA used (concentration 5ng/µl) was first dried down in Framestar 384 well PCR plates (4titude, Surrey, UK). A solution comprising of 1.5µl of Klearkall MasterMix (concentration 2X) (KBioscience, Hoddesdon, UK), 0.03µl of probe (concentration 40x) and 1.47µl of sterile water per reaction was made up and added to each well of the plate. The PCR reaction was then performed on a thermocycler (MJ Research Peltier Thermocycler- PTC 225) based on the following conditions: 10 minute cycle at 95°C, followed by 39 cycles of (i) 92°C for 15 seconds and (ii) 60°C for 1 minute. Each plate was subsequently analysed using allelic discrimination software (Figure 2.4) using the ABI7900HT system (Applied Biosystems). Details have been provided in the respective chapters.

Figure 2.3 Schematic Representation of the TAQMAN Genotyping Chemistry. Each assay contains two primers to amplify the region of interest and two probes associated with fluorescent dyes (FAM and VIC) on the 5' end to detect the allele present at the site and a nonfluorescent quencher is at the 3' end of the probe. Cleavage of the probe during the PCR amplification stage releases the fluorescent dye which is then detected by the ABI 7900HT system. (Figure adapted from www.Appliedbiosystems.com).

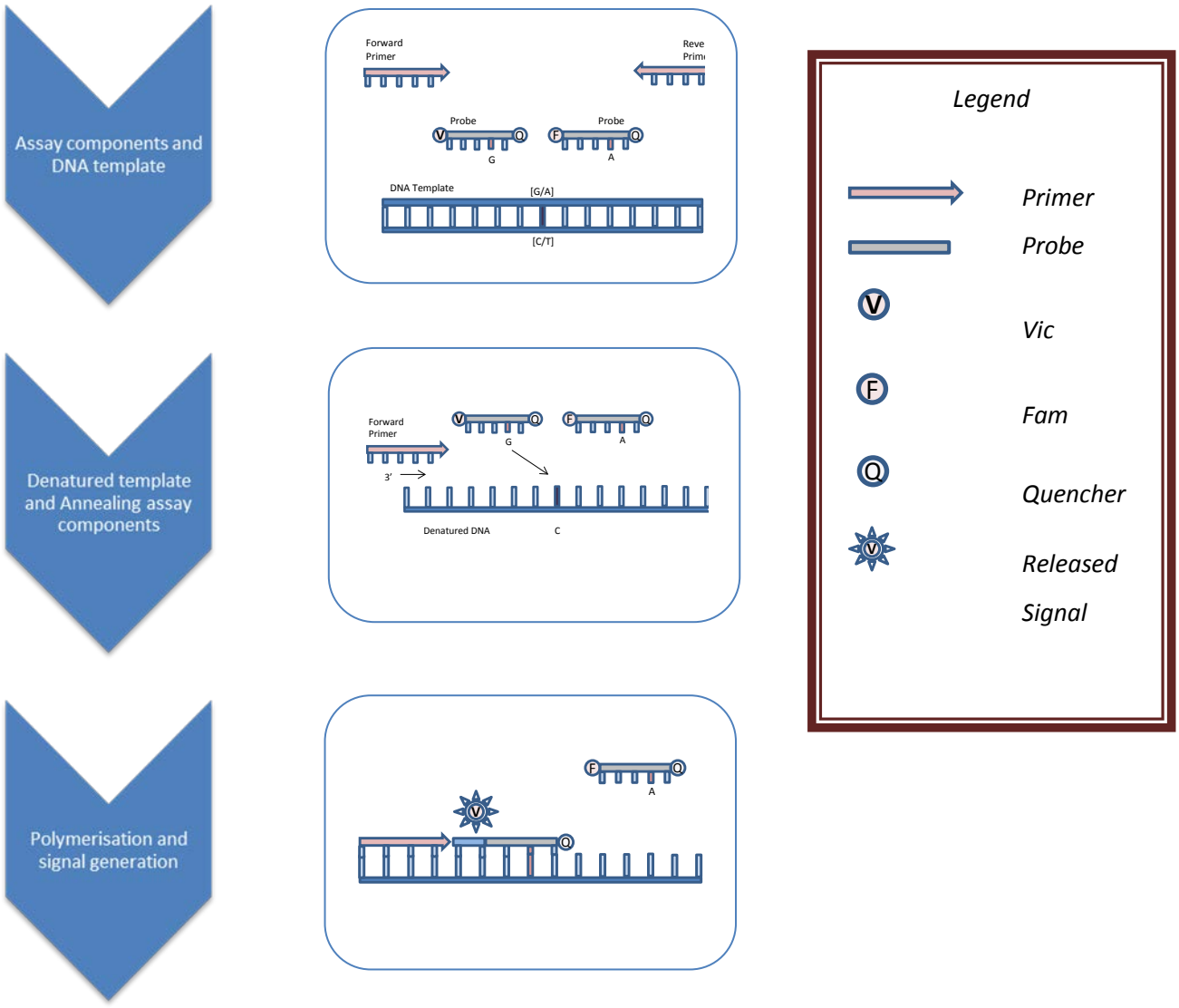
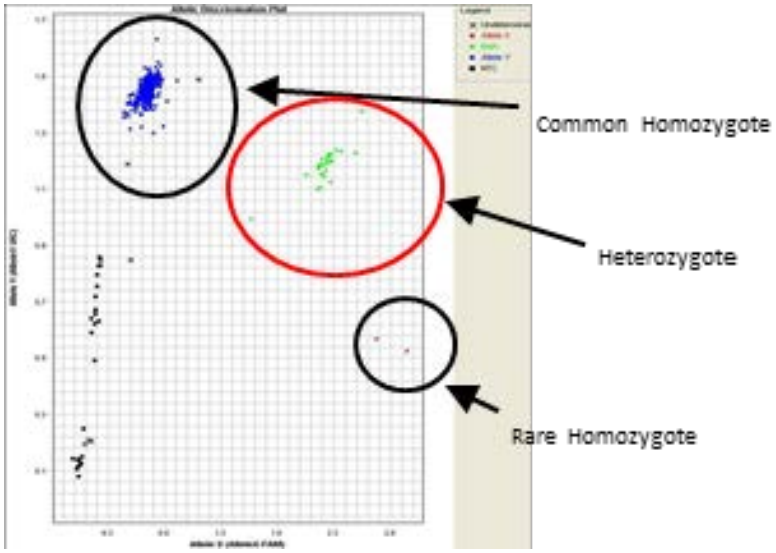
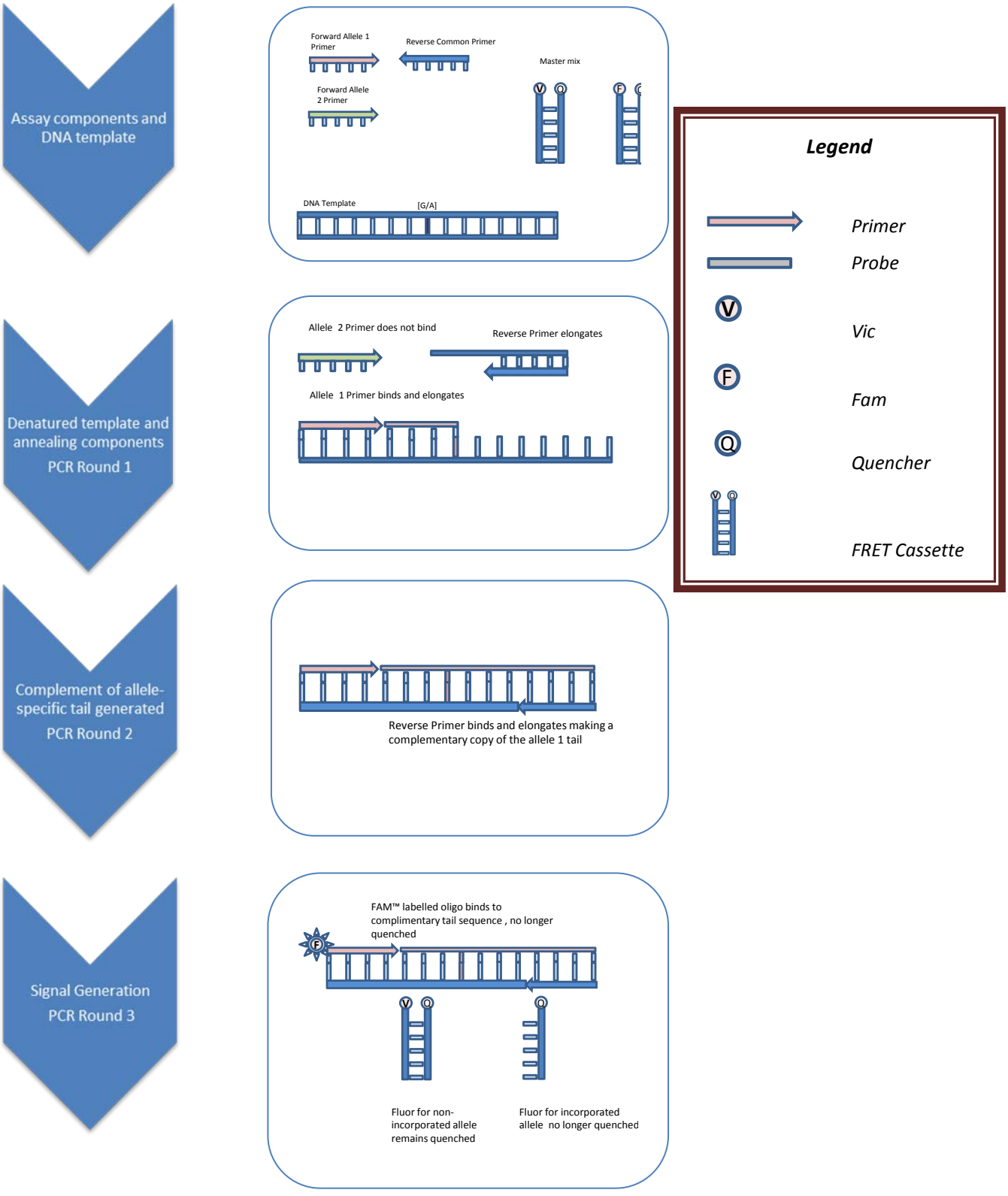


Figure 2.4 An example of the genotyping plot obtained using the ABI7900HT system in the analysis of the T130I variant of *HNF4A* (Chapter 4), demonstrating common homozygote, heterozygote and rare homozygote genotypes.



The KBiosciences Competitive Allele Specific PCR SNP genotyping system (KASPar) utilises 3 unlabelled primers (two allele specific primers and a common [reverse] primer) as well as a "reaction mix" (http://dna.uga.edu/docs/KASPar_SNP_Genotyping_Manual). The reaction mix contains a universal fluorescent reporter system. Fluorescence resonance energy transfer (FRET) is utilised for allele detection (Figure 2.5).

Figure 2.5 Schematic representation of the of the KASPAR genotyping chemistry (Figure adapted from www.KBioscience.com). Allelic discrimination is achieved through the competitive annealing of two allele specific forward primers each containing a unique tail sequence that corresponds with a distinctly labelled FRET cassette in the master mix.



The DNA (5 ng/µl) was first dried down prior to the addition of the primers and reaction mix. The following thermo-cycling conditions were used: 94°C for 15 minutes (hot-start activation) followed by 20 cycles of (i) 94°C for 10 seconds, (ii) 57°C for 5 seconds 20, (iii) 72°C for 10 seconds; followed by 18 cycles of: (i) 94°C for 10 seconds,(ii) 57°C for 20 seconds, (iii) 72°C for 40 seconds (http://dna.uga.edu/docs/KASPar_SNP_Genotyping_Manual). KASPAR uses FAM and VIC for fluorescence detection of alleles and also incorporates a passive reference dye (ROX). The latter allows for correction for background fluorescence and results in tighter clustering of the genotypes called. KBiosciences Klustercaller software (Kbiosciences, Hoddeston, UK) was utilised to plot the genotypes obtained for each SNP.

Probes for the genotyping carried out in this project were either developed through "assay on demand" or "assay by design" process (Applied Biosystems, Warrington, UK). For regions that that had no previous probes designed such as those for the *HNF1A* project (Chapter 3) an "assay by design" was utilised. For regions for which previous assays were available, such as those for the large scale genotyping of the low frequency variants of *HNF4A* and *LARS2* (Chapters 4 and 5), "assay by demand" was used.

The Taqman approach is robust and reliable, having been extensively tested and validated. The KASPaR assay has been developed more recently but has been shown to be as reliable as the Taqman technique and a concordance rate in excess of 99.5% has been reported with Taqman (http://www.kbioscience.co.uk/reagents/KB_AN_KASPar_chemistry). The Taqman assay was available locally and was utilised for "in-house" genotyping. To expedite the genotyping carried out as part of this experiment, the genotyping of the V255M variant of *HNF4A,* H324Q variant of

LARS2 and D76N variant of *PDX1* was outsourced to KBioscience, Hoddesdon, U.K., where the KASPaR assay was used.

I subsequently carried out the analysis of all data generated by both genotyping methods.

2.2.6.1 Analysis of genotyping data

Quality Control

All genotyping plots (Figure 2.3) were visually inspected to ensure accuracy of assigned genotyped. Allelic discrimination SDS2.3 software (Applied Biosystems) using the ABI7900HT system (Applied Biosystems) was used for the analysis of plates that utilised a Taqman assay. The KBiosciences Klustercaller software (Kbiosciences, Hoddeston, UK) was used to plot and analyse the genotypes obtained for each SNP genotyped using the KASPar assay.

Genotyping pass rate was calculated by evaluating the number of samples per plate that were assigned a genotype. Any plate with less than 95% success rate was deemed to have failed and the genotyping was repeated. As a further quality control measure, the estimated error rate based on completed duplicate pairs was calculated.

Finally, departure from Hardy Weinberg Equilibrium ($p < 0.05$) was assessed for each sample genotyped. This was evaluated using PEDSTATS software package; http://www.sph.umich.edu/csg/abecasis/PedStats (Wigginton et al. 2005).

Statistical analysis

Heterogeneity of genotype counts in cases and in controls was tested for each SNP using the Pearson chi-square test using StatXact (v6.0:Cytel software corps, Cambridge, MA, USA). Subsequent association testing was carried out using an exact implementation of the Cochran-Armitage trend test utilising StatXact. For the A98V variant of *HNF1A* and the T130I variant of *HNF4A* (Chapter 4) and the *LARS2* variants (Chapter 5), a meta-analysis incorporating summary statistics from other studies was carried out using an additive model, utilising GWAMA (Genome Wide Association Meta-Analysis) software package (http://www.well.ox.ac.uk/gwama). A forest plot of the meta-analysis findings was then generated using the Comprehensive Meta-Analysis software version 2.050 (Biostat Inc., NJ, USA).

2.2.7 Blunt-ended ligation

For the targeted sequencing element of this project (Section 2.7), the Genome Analyzer II system was used. This platform utilises amplified PCR products that undergo sonication prior to the sequencing run. It is reported that overrepresentation of the unfragmented end pieces of the amplicons will result if they are <1500bp in length (Druley et al. 2009). As the average amplicon generated in this experiment was of the order of 470 bp, it was felt that random concatenation of the amplicons should be performed prior to fragmentation with sonication. For this I utilised a protocol published in Nature Methods by Druley et al (Druley et al. 2009). Using this protocol, 1X of 10X T4 DNA Ligase buffer (New England Biolabs, Ipswitch, MA, USA), 2400 units of 400 units/µL T4 DNA Ligase (New England Biolabs, Ipswitch, MA, USA), 120 units of 10 units/µL T4 Polynucleotide Kinase (New England Biolabs, Ipswitch, MA, USA) and 15% (wt/vol) polyethylene

glycol 8000 MW (Sigma-Aldrich, St. Louis, U.S.A.) were combined with each PCR amplicon pool. The resulting mixture was then incubated for 17 hours at 22°C, followed by 20 minutes at 65°C. Concatenation was confirmed by agarose gel electrophoresis. Products in excess of 10kb were generated with this method.

2.2.8 Targeted sequencing- Genome Analyzer II

The Genome Analyzer II (GAII) platform was utilised for high throughput targeted sequencing as described in Chapter 6. This project was designed and carried out in 2008- 2009, when this technology was still relatively new and protocols were still being established for sample preparation at the time of project commencement.

Prior to the sequencing run, the amplified DNA sample of interest was sheared to appropriate size (average ~800bp). The Covaris S2 system (Applied Biosystems) was used for this process by the core genomics team at the Welcome Trust Centre for Human Genetics (WTCHG), Oxford, UK for the main targeted sequencing project (Section 6.4) and by Geneservice Ltd. (Nottingham, UK) for the pilot study (Section 6.2). The ends of the DNA were blunted, and two unique adapters were ligated to the fragments (5' P-GATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG; 5' ACACTCTTTCCCTACACGACGCTCTTCCGATCT). These adapters were complementary to two surface bound amplification primers located on flow cells.

This technology utilised a microfluidic cluster station in which samples were hybridized onto a flow cell and were amplified for later sequencing. The Hybridization process involves a heating and cooling step which is in turn followed by incubation with an isothermal polymerase which

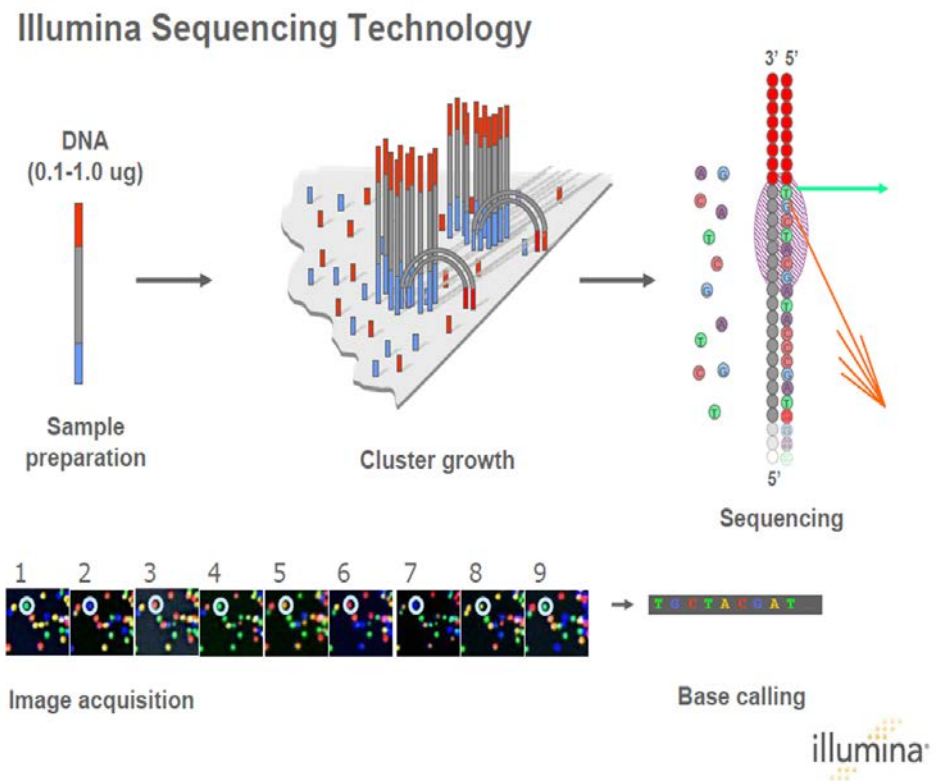
results in amplified fragments in discreet areas or “clusters” on the surface of the flowcell. An ultra-high density sequencing flow cell (with millions of clusters) was created by solid support bridge amplification (Figure 2.6). Each cluster contained approximately 1000 copies of the same template.

The GAII platform sequences clustered template DNA using a four colour DNA Sequencing-by-synthesis technology that employs reversible terminators with removable fluorophores. Laser excitation and total internal reflection optics are used to achieve highly sensitive fluorescence detection.

A high resolution image of the entire flow cell was taken, following the addition of a single fluorescent nucleotide. Repeating of the cycle, one base at a time generated a series of images each representing a single base extension at a specific cluster. These images represent the data collected for each base. Any signal above background identified the physical location of a cluster (or polony) and the fluorescent emission identifies which of the four bases was incorporated at that position. The emission colour over time is used to call the bases. The short sequence reads were subsequently aligned against a reference genome and genetic differences were called using a specially developed pipeline. All sequencing processes were carried out according to the manufacturer’s guidelines. As this experiment was aimed at SNP discovery in exonic regions of genes, single-end 50bp reads were targeted. At the time of designing the experiment the cost of paired end reads was prohibitively expensive for this project.

Figure 2.6 Schematic representation of the Genome Analyzer 2 platform including the cluster station where samples were hybridised to a flow cell. Image modified from http://www.Illumina.com

.



2.2.8.1 Data analysis

The GA pipeline software (version 1.0, Illumina software) was used to undertake image analysis and base calling at the WTCHG. The raw sequencing data were first aligned to a reference genome (the Human Build 36.1 finished human genome assembly [hg18, Mar. 2006]) using Eland (Illumina technologies) algorithm with the 'ELAND extended' option to enable better handling of >32 base reads. The output from the ELAND extended algorithm was then used in the MAQ [mapping and assembly with qualities (Li et al. 2008)] (http://maq.sourceforge.net/) software packages in the WTCHG to align to a reference sequence made up of the regions targeted in this study and for SNP calling. The Syzygy software package

(http://www.broadinstitute.org/software/syzygy/) was subsequently utilised for more accurate SNP calling in the WTCHG.

The data generated by the GAII run was primarily analysed using the Syzygy software package (release 0.9) developed by Manuel A. Rivas and Mark J. Daly (my supervisors' collaborators from the BROAD Institute, MA, USA). This software package had been newly released when I was performing my targeted sequencing project. Syzygy is a post processing analysis tool specifically targeted to the handling and analysis of complex sequencing data and allows rare variant SNP detection as well as pooled experiment statistical analysis (Calvo et al. 2010; Rivas et al. 2011).

CHAPTER 3

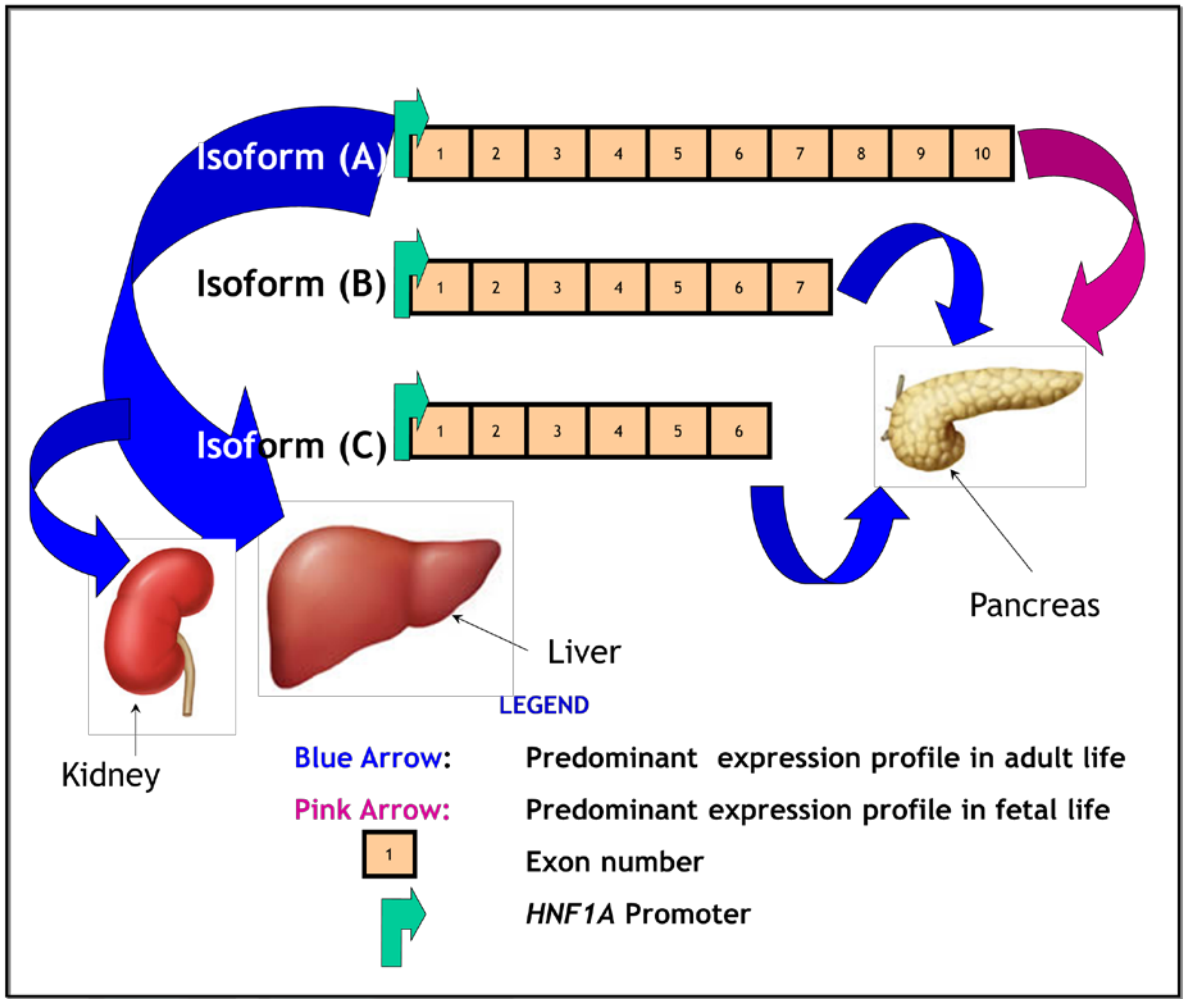
Evaluation of the contribution of low frequency variants in the exons only encoding isoform A of *HNF1A* to susceptibility to Type 2 Diabetes

3.1 Introduction

The majority (up to 65%) of cases of maturity onset diabetes of the young (MODY) in the United Kingdom can be attributed to rare, highly penetrant mutations in *HNF1A* (Frayling et al. 2001). This gene has also been implicated in the pathogenesis of T2D through candidate gene approaches (Weedon et al. 2005; Winckler et al. 2005; Winckler et al. 2007) and through a private mutation (G319S, located in exon 4) which is associated with T2D susceptibility in the Oji-Cree population (Triggs-Raine et al. 2002). More recently, a common (MAF 15%) variant of *HNF1A*, rs7957197, has been implicated in the pathogenesis of T2D through the genome wide association (GWA) approach (Voight et al. 2010).

HNF1A encodes three different isoforms, termed A (encoded by exons 1-10), B (encoded by exons 1-7) and C (encoded by exons 1-6) which arise by alternative splicing and polyadenylation (Bach et al. 1993) (Figure 3.1).

Figure 3.1 *HNF1A* isoform structure and expression profile



On average a later age of diagnosis of MODY is seen in individuals with mutations in exons 8-10 (present only in isoform A), in comparison to those individuals with mutations in exons 1-7 (18 vs. 25 years, p=0.0001 and 19 vs. 24 years, p=0.03) (Harries et al. 2006; Bellanne-Chantelot et al. 2008). The differing temporal and spatial expression profile of the isoforms is the most likely reason for this (isoform A predominates in the foetal pancreas, adult liver and kidney whilst isoforms B and C predominate in the adult pancreas (Harries et al. 2006). Thus, in contrast with mutations which affect the function of all isoforms, mutations in exons 8-10 display a lower penetrance and later age of onset leading to more clinical overlap with common forms of T2D.

These later 3 exons of *HNF1A* were therefore felt to be auspicious candidates in terms of harbouring low frequency incompletely penetrant (LFIP) variants involved in T2D pathogenesis.

The aim of this study was to perform deep resequencing of these exons in a large sample of patients enriched for a genetic aetiology for their diabetes, due to early onset and/or family history of T2D, as this patient population would be more likely to harbour LFIP variants. I tested the hypothesis that the later 3 exons of *HNF1A*, (i.e. the exons only present in isoform A), could contain LFIP variants which contribute to susceptibility to T2D by performing deep resequencing of these exons using traditional Sanger sequencing (Chapter 2, Section 2.2.5).

3.2 Subjects & Methods:

3.2.1 Subjects

Two samples of British ancestry were included (total n=591). Sample 1 (n=84) was enriched for early-adult onset T2D and sample 2 (n=507) was enriched for family history of T2D. This was to maximize the likelihood of identifying medium penetrance genetic variants influencing T2D-risk.

Sample 1 (the "Young Diabetes in Oxford Cohort", UK3, Chapter 2, Section 2.1.1) includes subjects diagnosed clinically with T2D between 18 - 45 years of age, recruited from Oxfordshire GP practices.

Sample 2 (UK2, Chapter 2, Section 2.1.1) consists of probands from Diabetes UK Warren 2 collection (Exeter, London, Oxford, Norwich and Newcastle). All individuals in this group were

diagnosed between 35 and 75 years (mean age of diagnosis 55.4 years) and had at least one affected sibling diagnosed with diabetes. These individuals were included due to their family history of T2D as they should be enriched for LFIP variants.

To establish the background allele frequency of variants identified in this study in a UK control population, I utilised a subset of the British Birth Cohort of 1958 (n=350 for the intronic variants identified, n=1050 for the exonic variant identified) (Wellcome Trust Case Control Consortium 2007; Zeggini et al. 2007). Full details of all samples have been included in chapter 2 (section 2.1).

Table 3.1 Characteristics of subjects with T2Dincluded in the sequencing of the latter 3 exons of *HNF1A*

	Sample 1	**Sample 2**
N	84	507
Male (%)	61.9	53.6
Mean age at diagnosis(years) ±SD	34.3±11.4	55.4±8.5
***Treatment (ins/OHA/diet) (%)**	26/66/8	14/69/17
First-degree relative with Diabetes (%)	48	100
Mean BMI(kg/m^2) ± SD	32.3±4.8	28.8±5.2

*Treatment at the time of ascertainment. ins, insulin; OHA, oral hypoglycaemic agent

3.2.2 Methods

Sequencing

Targeted capillary resequencing of exons 8-10 of *HNF1A* (including intron-exon boundaries) was performed on the ABI3700 platform based on Sanger sequencing (Sanger et al. 1977) (Chapter 2 section 2.2.5). All primers were designed using primer3 software (Table 3.2) (http://frodo.wi.mit.edu/primer3/).

Table 3.2 Primer sequences for exons 8, 9 and 10 of *HNF1A*

Target region	Forward primer sequence 5'-3'*	Reverse primer sequence 5'-3'#
Exon 8	AGGCTCCCCTGCATCAAG	GGGTGAGGGGCTCTGTCA
Exon 9	ACCAAGCAGGTAAGGTCCAG	GTGACGGACAGCAACAGAAG
Exon 10	GTACCCCTAGGGACAGGCAGG	ACCCCCCAAGCAGGCAGTACA

*All forward primers had the following M13 tail: TGTAAAACGACGGCCAGT
#All Reverse primers had the following M13 tail: CAGGAAACAGCTATGACC

Polymerase chain reaction (PCR) products were generated (Chapter 2, Section 2.2.3) using AmpliTaq Gold DNA polymerase (Applied Biosystems, Warrington, UK). All PCR reactions were carried out using manufacturers' guidelines and at recommended primer annealing temperatures (chapter 2, section 1.2). The PCR reaction was carried out on a thermocycler (MJ Research Peltier Thermocycler- PTC 225). The PCR program utilised was as follows:

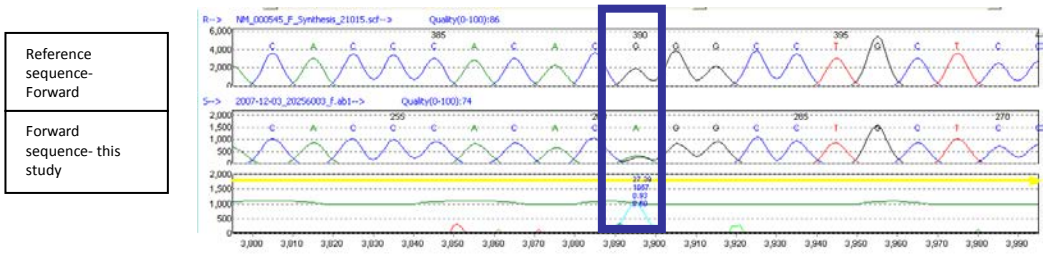
1 cycle at 95ºC for 12 minutes (initialization step), followed by 35 cycles of: (i) 94 ºC for 1 minute (denaturation step), (ii) 62ºC for 1 minute (Annealing step) and (iii) 72ºC for 1 minute (extension/elongation step). The programme was completed with a ten minute cycle at 72ºC

(final elongation step). Success of PCR reactions was determined by analysis on a 2% agarose gel (chapter 2, section 1.3).

The sequencing reaction was carried out by use of BigDye v1.1 (Applied Biosystems Warrington, UK), and M13 tailed sequencing primers (Eurofins MWG Operon, Acton, London, UK). The sequencing reaction was carried out on the ABI3700 sequencing platform (Chapter 2, section 2.2.5).

Sequences were analysed using Mutation Surveyor v 3.2 (Soft Genetics, Cambridge, UK). The sequences were compared to the reference sequence (NM_000545.3) using the unidirectional analysis mode of Mutation surveyor V3.2 (Soft Genetics, Cambridge, UK). This software package has been shown to have a sensitivity of >99% in the unidirectional analysis mode(Ellard S 2009). The accuracy of all calls was further checked by visual inspection of all electropherograms (Figure 3.2).

Figure 3.2 Electropherogram illustrating the heterozygous c.1545G>A; T515T variant of *HNF1A*



Electropherogram illustrating the heterozygous c.1545G>A; T515T variant highlighted in blue box.

3.3 Results

A total of 4 variants (three intronic and one exonic) were identified in resequencing the terminal 3 exons of *HNF1A* in 591 individuals (Table 3.3). However no novel coding variants (low frequency or otherwise) were identified.

Table 3.3 Variants identified using Sanger sequencing of exons 8-10 of *HNF1A*

Region	Approved cDNA level description	Description used in MODY literature	rs number	Sequence*
Intron 8	c.1623+29C>T	IVS8+29C>T	rs1169304	GCCCTCCCT**[T/C]**GGCCTGTGAC
Intron 9	c.1769-24T>C	IVS9-24T>C	rs735396	GTGGGTGGC**[T/C]**AGCAGCCTTG
Intron 9	c.1768+44C>T	IVS9+44C>T	N/A	TCCCTGCCC**[C/T]**CTTCCATGTT
Exon 8	c.1545G>A	T515T	rs61953349	CACCCACAC**[A/G]**GGCCTGCTCC

*sequence of the region spanning the variants identified. The base change is indicated in bold in parentheses. Within the parentheses the first letter relates to the wild type genotype and the second letter relates to the variant identified.

Genotyping (Chapter 2, Section 2.2.6) of a control population, to establish the background allele frequency in the UK, was performed on the Taqman ABI7900HT platform (Chapter 2, section 2.2.6). For regions that that had no previous probes designed (IVS9+44C>T) an "assay by design" was utilised. For regions, for which previous assays were available "assay by demand" was used (Chapter 2, section 2.2.6). The probes for the genotyping of the variants in this study have been included in Table 3.4. Each plate was subsequently analysed by use of allelic discrimination SDS2.3 software (Applied Biosystems) using the ABI7900HT system (Applied Biosystems). Quality of the genotyping was assured by assessing genotyping pass rate (> 98% for all variants) and analysing for departure ($p<0.05$) from Hardy Weinberg equilibrium (none detected).

Table 3.4: Probes used for genotyping of the variants detected in sequencing the latter 3 exons of *HNF1A*

	ABI Assay on Demand ID	**Assay by Design probe sequences**	**Assay by Design primer sequences**
IVS8+29C>T rs1169304	C___1263612_10	N/A	N/A
IVS9-24T>C rs735396	C___1263608_1_	N/A	N/A
IVS9+44C>T	N/A	**VIC*:** CCCTGCCCTCTTCCAT **FAM*:** CCTGCCCCCTTCCAT	**Forward:** CCACCCCCTCCCTTACTGT **Reverse:** AGTGACGGACAGCAACAGAA
T515T rs61953349	C__89163791_10	N/A	N/A

* VIC and FAM refer to the fluorescent dye employed in allelic discrimination.

N/A: Not available

The only coding variant identified (T515T, rs61953349) is a previously reported synonymous variant with a MAF of 14.6% in cases and of 18.3% in controls included in this study (p=0.056, using an exact implementation of the Cochran Armitage trend test-Table 3.5). For the 3 intronic variants, no statistically significant differences in genotype frequencies were noted between the cases and controls (calculated using an exact implementation of the Cochran Armitage trend test-Table 3.5).

Table 3.5. Allele frequencies of variants identified using Sanger sequencing of exons 8-10 of HNF1A

Approved cDNA level description	Description used in MODY literature	rs number	MAF current study (%)	MAF in population controls (%)	p-value
c.1623+29C>T	IVS8+29C>T	rs1169304	20.1	*17.9	P=0.28
c.1769-24T>C	IVS9-24T>C	rs735396	33.4	*34.3	P=0.63
c.1768+44C>T	IVS9+44C>T	N/A	4.1	*4.3	P=0.94
c.1545G>A	T515T	rs61953349	14.6	**18.3	p=0.05

N/A; not available. *Based on screening 350 population controls (part of the UK 1958 Birth Cohort), **Based on screening 1050 population controls (part of the UK 1958 Birth Cohort)

3.4 Discussion

According to one report, in MODY families showing classical Mendelian segregation, individuals with mutations in exons 8-10 of *HNF1A* have been diagnosed as late as 38 years, and the median age of diagnosis is ~27 years (Harries et al. 2006). It was this observation that led to the hypothesis that coding variants within the same exons might be playing a role in the pathogenesis of multifactorial T2D. Given that genome-wide association studies have already surveyed common variation (Saxena et al. 2007; Scott et al. 2007; Sladek et al. 2007; Zeggini et al. 2007; Zeggini et al. 2008; Dupuis et al. 2010; Voight et al. 2010; Cho et al. 2012; Morris et al. 2012), my particular focus was on variants of relatively low frequency (in the 0.5-5% range). If variants in this part of the allele frequency spectrum have effect sizes greater than those so far observed for common variants, they may be responsible for a substantial proportion of the genetic variance of T2D.

In this study, 60 (10%) individuals were diagnosed at 38.5 years or younger and the median age of diagnosis was 55 (compared to an average age of diagnosis in the UK of 59.6 years [based on 15653 T2D patients from Tayside Scotland. R. McAlpine & AD Morris, personal communication]). Almost all cases had close relatives with T2D, a feature which will have enriched for medium- and high-penetrance variants.

No novel variants (low frequency or otherwise) in the latter three exons of *HNF1A* were identified. However, the presence of additional low frequency variants in other regions of this gene impacting on T2D risk cannot be excluded based on these results. Recent advances in high-throughput sequencing allow for a more in-depth investigation of this gene. *HNF1A* has been included as one of the genes for the high throughput sequencing project discussed in chapter 6.

CHAPTER 4

Evaluation of the role of candidate coding, low frequency, incompletely penetrant variants in Type 2 Diabetes pathogenesis

4.1 Introduction

The familial aggregation (both genetic and non-genetic) seen in Type 2 diabetes (T2D) is as of yet not fully explained (Lango et al. 2008; Manolio et al. 2009) and low frequency variants may explain a portion of this heritability gap. A logical place to start the search for low frequency variants inferring increased T2D susceptibility risk is in genes that have been implicated in disease pathogenesis through previously discovered common or rare variants (Chapter 1, Section 1.2).

Five variants in four genes that previous published work had indicated may be candidate low frequency variants contributing to T2D pathogenesis were prioritised for further follow up. Previous T2D association reports regarding low frequency variants in key β cell transcription factors *HNF4A* (Ek et al. 2005; Winckler et al. 2007), *PDX1* (Hani et al. 1999; Macfarlane et al. 1999; Owen et al. 2004) and *HNF1A* (Holmkvist et al. 2006) as well as *KCNJ15* (Okamoto et al. 2010) that affects the β cell ion channel have been inconclusive. Therefore the aim of this project was to clarify the role played in T2D pathogenesis of these low frequency variants (T130I and V255M variants of *HNF4A,* D76N variant of *PDX11*, A98V variant of *HNF1A* and the I88I variant of *KCNJ15*) through large scale association analyses as outlined in sections 4.1.1-4.1.4.

4.1.1 A98V variant of *HNF1A*

Rare mutations in the transcription factor *HNF1A* are responsible for approximately 65% of cases of maturity onset diabetes of the young (MODY) (Ellard et al. 2008) while common variations in *HNF1A* have been implicated in the pathogenesis of T2D through the genome wide (Voight et al. 2010) and candidate gene approaches (Triggs-Raine et al. 2002; Weedon et al. 2005; Winckler et

al. 2005; Winckler et al. 2007). Linkage studies have also implicated Chromosome 12q as potentially impacting on T2D risk (Mahtani et al. 1996; Bowden et al. 1997; Shaw et al. 1998), though this was not seen in all studies (Frayling et al. 2000).

A low frequency (MAF 3.0%) variant of this gene, p.Ala98Val (A98V), rs1800574, that resides two amino acid residues upstream of the DNA binding domain of *HNF1A* has been highlighted as potentially increasing the susceptibility to T2D risk in a study of British subjects , n= 2010 cases, 1643 controls; OR 1.23 [95% CI 0.99-1.54], P = 0.07 (Weedon et al. 2005). This association was not seen in replication efforts in combined samples from the USA, Canada, Poland, Sweden and samples collected throughout Scandinavia; n= 5467 cases, 6152 controls; OR 1.07 (0.88–1.29), p = 0.49 (Winckler et al. 2005). However, a meta-analysis incorporating several published studies (n= 8545 cases, 9796 controls) has demonstrated a modest association of the A98V variant of *HNF1A* on T2D risk (OR 1.16 [1.03-1.31]; p= 0.02) (Holmkvist et al. 2006).

The biological candidacy of the A98V variant as potentially impacting on T2D pathogenesis has been strengthened by studies indicating that heterozygous carriers of this variant, in vivo, are associated with decreased glucose stimulated insulin secretion [p=0.000499 (Holmkvist et al.2006)]. This was quantified by calculating the I/G30 (insulin at 30 minutes –insulin at 0 minutes/glucose at 30 minutes- glucose at 0 minutes) during an oral glucose tolerance test incorporating 2293 samples from the Botnia prospective cohort recruited in the Botnia region of Finland (Holmkvist et al. 2006). This variant is predicted to be damaging by use of the SIFT (sorts intolerant from tolerant; http://sift.jcvi.org/) prediction software though this was not corroborated using the PolyPhen (Polymorphism Phenotyping; http://genetics.bwh.harvard.edu/pph/) prediction programme.

The aim of this study was to evaluate further the possible role of the A98V variant of *HNF1A* in T2D susceptibility through a large scale association study.

4.1.2 D76N variant of *PDX1*

The transcription factor pancreatic and duodenal homeobox 1 (*PDX1*) has been shown to be critical for pancreatic development and in the functioning of the adult pancreas (Weng et al. 2001). Very rare mutations in this gene have been shown to be causative for MODY in some pedigrees (Stoffers et al. 1997; Fajans et al. 2001). Pancreatic agenesis has been reported as a result of mutations in the *PDX1* gene (Stoffers et al. 1997; Fajans et al. 2010). In one pedigree, homozygous carriers of the Pro63fsX60 mutation of this gene were shown to have pancreatic agenesis and permanent neonatal diabetes whereas heterozygous carriers of this mutation had MODY (Fajans et al. 2010). A further low frequency (MAF 1.0%) nonsynonymous polymorphism in this gene, p.Asp76Asn (D76N), has been shown to be potentially implicated in the pathogenesis of multifactorial T2D (Hani et al. 1999; Macfarlane et al. 1999) though other studies dispute this finding (Hansen et al. 2000; Elbein et al. 2004; Owen et al. 2004).

The majority of the studies that have investigated the potential association of the D76N variant with T2D have been relatively small. The largest single study (n = 860 cases, 1275 controls) failed to demonstrate an association with T2D risk in a UK sample (OR 1.1 [95% CI 0.52-2.1], p=0.86) (Owen et al. 2004). However, a meta-analysis of all studies of this variant (n = 1767 cases, 1974 controls) did show a trend towards association of D76N with T2D disease state (OR 1.7 [1.0-2.8]; p = 0.07) (Owen et al. 2004). A further meta-analysis including only the subset of the subjects

with an age of onset of T2D of less than 45 years improved the strength of association (OR 2.5 [1.4-4.6]; p = 0.004) (Owen et al. 2004).

Recent sequencing of the exons and flanking intronic regions of *PDX1* in 910 T2D samples and 890 controls from the UK has identified 26 variants (6 novel) in this gene (Edghill et al. 2011). However, only one of these variants had a MAF >1.0% (D76N). This study did not identify an over accumulation of LF variants in T2D cases in comparison to non diabetic samples (p = 0.46) in *PDX1* (Edghill et al. 2011). However, the power to detect an association with T2D risk for each individual variant was limited to large effect sizes only (80% power to detect an OR of 3.2, for α = 0.001 for a variant with a MAF 1.0% by my calculations using the Quanto software package).

It is of interest that the D76N variant lies in the amino terminal region of *PDX1* that is required for the transcriptional activation of the insulin gene (Weng et al. 2001). The D76N variant has been reported to lead to decreased ability of endogenous *PDX1* to bind the insulin gene promoter compared to wild type *PDX1* (Weng et al. 2001).

Given the biological candidacy of the D76N variant and the discrepancy of the findings of the previous published works, this variant was prioritised for large scale genotyping to further clarify the role, if any, played by D76N in the pathogenesis of T2D.

4.1.3 The I88I variant of *KCNJ15*

Linkage studies in Japanese samples have highlighted the region of chromosome 21q, where the potassium inwardly rectifying channel, subfamily J, member 15 (*KCNJ15)* resides, as potentially

impacting on T2D pathogenesis (LOD score of 2.42) in lean Japanese samples (Iwasaki et al. 2003). Subsequent fine mapping efforts have highlighted a low frequency synonymous variant (MAF 3.1%), p.Ile88Ile (I88I), rs3746876, in exon 4 of *KCNJ15* as potentially impacting on T2D pathogenesis (OR of the T allele relative to the C allele = 1.76 [1.37-2.25]; $p=6.7x10^{-6}$) (Okamoto et al. 2010). This effect was more marked in lean individuals (OR 2.54 [1.76-3.67]; $p=2.5x10^{-7}$) (Okamoto et al. 2010). Subsequent genotyping efforts of this variant in Danish samples (n= 3531 cases, 4885 controls) demonstrated that this variant had a much lower allele frequency in Danish samples compared to Japanese samples (MAF 0.8% vs. 3.1%) and no association with T2D risk was demonstrated in the Danish sample for the I88I variant of *KCNJ15* (OR of the T allele relative to the C allele = 0.89 [0.62-1.28]; p=0.53) (Okamoto et al. 2010). However, given the lower allele frequency in the Danish samples, power to detect an association with T2D risk was limited [10% power to detect an OR of 0.89 for α=0.001, calculated using the Quanto software package (Gauderman et al. 2006) based on the figures quoted in the study]. It is of interest that *KCNJ15* has been shown to be highly expressed in the pancreas and in particular this expression is more marked in the cells of the islet of Langerhans as compared to cells of the exocrine pancreas in non-diabetic individuals (Okamoto et al. 2010). Characterisation of the effect of the I88I polymorphism has demonstrated that mRNA levels measured in peripheral blood samples were increased with carriage of the T allele (minor allele) as compared to the C (common) allele of the I88I variant, p= 0.0036 (Okamoto et al. 2010). In vitro functional characterisation of this variant has also shown that the degradation of the mRNA carrying the T allele was significantly reduced when compared to the mRNA carrying the C allele (Okamoto et al. 2010).

Recent work in relation to *KCNQ1* (Unoki et al. 2008; Yasuda et al. 2008) has demonstrated how a variant with differing allele frequencies in disparate populations may be identified as being

associated with a disease state first in samples with higher allele frequencies, with subsequent replication in samples with lower allele frequencies confirming the observed association. Variants within *KCNQ1* (rs2237897 [MAF 39%], rs2237895 [MAF 34%], rs2283228 [MAF 41%]) were initially highlighted as impacting on T2D pathogenesis in two separate Japanese genome wide association studies (Unoki et al. 2008; Yasuda et al. 2008). Subsequent replication efforts in European samples have demonstrated generally lower allele frequency in European Caucasians (MAF=3%, 40% and 6% respectively (Unoki et al. 2008)). An increased risk of T2D was associated with the rs2237895 variant of *KCNQ1* in samples from Denmark (Unoki et al. 2008). This association has been confirmed in subsequent replication efforts in other European samples (Jonsson et al. 2009).

Given the association of the I88I variant of *KCNJ15* with T2D risk in Japanese samples, this study aimed to clarify further the role, if any, played by this variant in the pathogenesis of T2D in UK samples in a large association study.

4.1.4 T130I and V255M variants of *HNF4A*

Rare mutations in the transcription factor *HNF4A* account for approximately 5% of cases of MODY (Ellard et al. 2008). Though *HNF4A* is expressed in multiple tissues, the expression in the pancreatic β-cells and liver is of particular interest. In pancreatic β-cells, *HNF4A* is required for glucose metabolism as well as normal insulin gene expression and secretion (Bartoov-Shifman et al. 2002). The protein encoded by this gene is required for hepatic gluconeogenesis (Rhee et al. 2003). Several studies have shown linkage between multifactorial T2D and the region of Chr 20q where *HNF4A* is located (Bowden et al. 1997; Ji et al. 1997; Zouali et al. 1997; Weedon et al.

2004). A Common (rs4812829, MAF 0.19 in European samples) variant in this gene has recently been associated ($p= 8.2 \times 10^{-12}$) with increased T2D in a Genome-Wide association study (GWAS) (Kooner et al. 2011). However, common variants in this region do not explain the findings of the linkage studies, it is possible that this region harbours more penetrant variants that might explain this observation (Lillioja et al. 2009).

HNF4A has been extensively resequenced not least as part of clinical diagnostic screening for MODY. These resequencing efforts have, *inter- alia*, identified two low frequency coding non-synonymous variants of particular interest: V255M (c.763G>A p.Val255Met) and T130I (c.389C>T p.Thr130Ile, rs1800961). V255M was first described following resequencing of Danish samples but no evidence of association with T2D risk was evident in analysis of 1434 cases and 4790 controls (Ek et al. 2005). The T130I variant, positioned in the DNA binding domain of *HNF4A*, did show a modest (p=0.04) association with T2D risk in the same sample, though subsequent efforts at replication failed to confirm this (Ek et al. 2005; Winckler et al. 2007). A recent meta-analysis of previous reports of the T130I variant and T2D risk in Japanese, Pima Indian and a combined North American/ Scandinavian sample (n=3517 cases, 3683 controls) has again demonstrated a very modest association (p=0.045) (Sookoian et al. 2009). Most recently, and of particular interest, given the relationship between lipids and T2D, the T130I variant has been shown to be significantly associated with HDL-cholesterol levels ($p=8x10^{-10}$) in a GWA meta-analysis incorporating 30714 individuals (Kathiresan et al. 2009). T130I and V255M are not in LD with the common variant (rs4812829) in this gene showing association with increased T2D risk.

Both variants have been shown to be functional based on studies of the transcriptional regulation of *HNF4A*-target genes in a range of cell lines and primary mouse hepatocytes (Moller

et al. 1997; Navas et al. 1999; Zhu et al. 2003; Ek et al. 2005). It was therefore reasoned that they remain interesting candidates for assessment in larger samples to establish more clearly their likely contribution to T2D susceptibility.

4.2 Methods

4.2.1 Subjects studied

Three sample categories were included. Category 1 consisted of samples specifically genotyped for this study. Category 2 was comprised of samples with previously reported genotyping information for the SNPs included in this study. Category 3 included samples for which only summary statistics were available from previous published reports. The samples included in each category have been summarised in Table 4.1. The overall subject characteristics are summarised in Table 4.2. Detailed descriptions of category 1 and 2 samples are included in chapter 2.

Category 1 samples

Category 1 samples were derived from three sources (two UK samples and one Danish sample). UK Sample 1 ("UK1", Chapter 2, Section 2.1; maximum of 5000 cases, 6470 controls) included the United Kingdom Type 2 Diabetes Genetics Consortium (UKT2DGC) collection recruited in Tayside, Scotland (Wellcome Trust Case Control Consortium 2007; Zeggini et al. 2007). For the study of the I88I variant of *KCNJ15*, further samples from this collection were available for genotyping. These extra samples represent on-going recruitment of UKT2DGC along the same criteria as the original samples. UK Sample 2 ("UK2", Chapter 2, Section 2.1) comprised of a

maximum of 1853 T2D cases ascertained from a subset of the Diabetes UK Warren 2 repository (Wiltshire et al. 2001). The controls for UK2 were taken from the population-based British 1958 Birth Cohort (n=7133), and the United Kingdom Blood Service collection (n=3087) (Wellcome Trust Case Control Consortium 2007).

Danish Sample 1 ("DK1", n=2646 cases, Chapter 2, Section 2.1) was also included in category 1 for the study of the T130I variant of *HNF4A*. This sample was also utilised for the study of the I88I variant of *KCNJ15* though the genotyping information was obtained from a previous publication (Okamoto et al. 2010) and hence this sample was included as category 2 for the study of this variant. DK1 represents samples collected in the Steno Diabetes centre and Danish samples from the Anglo-Danish-Dutch study of Intensive Treatment in People with Screen-Detected Diabetes in Primary Care (ADDITION) (Lauritzen et al. 2000; Ek et al. 2005). The new samples in DK1 were combined with the previously reported case and control data from DK2 (described below) to generate a combined DK analysis involving 4043 cases and 4865 controls.

Category 2 samples

Category 2 included samples from Denmark, Sweden, Finland and Canada. Danish Sample 2 ("DK2", n=1397 cases, 4865 controls) was previously genotyped for T130I and V255M (Ek et al. 2005) in addition to I88I variant of *KCNJ15* (Okamoto et al. 2010). Two samples from the Finland-United States Investigation of Non Insulin Dependent Diabetes Mellitus Genetics (FUSION) study were included for T130I (FUSION Sample 1, "FS1", [n=1160 cases, 1173 controls] and FUSION Sample 2, "FS2", [n=1211 cases, 1264 controls]) (Zeggini et al. 2008).

Table 4.1 Study samples included in each category for each candidate LFIP variant studied

Variant	Category 1 samples		Category 2 samples		Category 3 samples	
	Sample name short form	Cases; Controls	Sample name short form	Cases; Controls	Sample name short form	Cases; Controls
A98V variant of *HNF1A*	UK1	4124; 5126			BR1	127; 127
	UK2	1193; 10220			BR2	471; 471
					BR3	514; 514
					BR4	609; 580
					GCI US	1226; 1226
					GCI Poland	1009; 1009
					SI1	85; 328
					FL1	222; 295
					SC1	1511; 2225
					DK3	240; 245
D76N variant of PDX1	UK1	4124; 5126				
	UK2	1853; 10220				
I88I variant of *KCNJ15*	UK1	5000; 6470	DK1	2646; 0		
	UK2	1193; 10220	DK2	1397; 4865		
T130I variant of *HNF4A*	UK1	4124; 5126	DK2	1397; 4865	PI1	573; 464
	UK2	1193; 10220	FS1	1160; 1173	JP1	423; 354
	DK1	2646; 0	FS2	1211; 1264	MX1	100; 75
			MS1	774; 2957		
			BR1	127; 127		
			BR2	490; 490		
			BR3	514; 514		
V255M variant of *HNF4A*	UK1	4124; 5126	DK2	1397; 4865		
	UK2	1853; 10220				

UK1: UK sample 1; **UK2**: UK Sample 2; **DK1**: Denmark Sample 1; **DK2**: Denmark Sample 2; **DK3**: Denmark sample 3
FS1: FUSION Sample 1; **FS2**: FUSION Sample 2; **BR1**: Broad Sample 1; **BR2**: Broad Sample 2; **BR3**: Broad sample 3; **BR4**: Broad Sample 4
GCI US: Genomics Collaborative, Inc. USA Sample 1; **GCI Poland**: Genomics Collaborative, Inc. Poland sample
SI1: South India Sample 1; **FL1**: Finland Sample 1; **SC1**: Scandinavia Sample 1
PI1: Pima India Sample 1; **JP1**: Japan Sample 1; **MX1**: Mexico Sample 1

FS1 and FS2 represent the FUSION GWAS and replication samples respectively and have been included in a T2D (Scott et al. 2007) and a lipid GWAS (Kathiresan et al. 2009) and subsequent follow up of significant findings. Due to the withdrawal of some individuals and updated T2D status of others the numbers of individuals quoted for FS1 and FS2 vary slightly to those in the reference article.

Samples from The Metabolic Syndrome in Men (METSIM), ("MS1", n=774 cases, 2957 controls) recruited in Finland (Wang et al. 2007) were also included for the study of T130I. The genotype data for MS1 regarding T130I were included as part of the lipid GWAS (Kathiresan et al. 2009) though T2D data have not been published. Previously reported genotyping results for T130I from three samples from the Broad institute were also included (Winckler et al. 2005). These comprised of a Canadian sample (Broad Sample 1, "BR1", n=127 cases, 127 controls), a combined Swedish-Finnish sample (Broad Sample 2, "BR2", n=490 cases, 490 controls) and a Swedish sample (Broad sample 3, "BR3", n=514 cases, 514 controls).

Category 3 samples

For a more complete study of the T130I variant of *HNF4A* and the A98V variant of *HNF1A,* the results from these category 1 and 2 samples were included in a meta-analysis together with those from previously published studies. For these category 3 samples we had no access to genotype information. For the study of the T130I variant of *HNF4A* these included a Pima Indian sample ["PI1" n = 573 cases, 464 controls (Muller et al. 2005)] and a Japanese sample ["JP1" n = 423 cases, 354 controls (Zhu et al. 2003)] included in the meta-analysis by Sookoian et al (Sookoian et al. 2010) in addition to a Mexican sample ["MX1" n = 100 cases, 75 controls (Menjivar et al. 2008)].

For the further evaluation of the A98V variant of *HNF1A,* we included samples from Denmark ["DK3" n = 240 cases, 245 controls (Urhammer et al. 1997)], Finland ["FL1" n = 222 cases, 295 controls (Rissanen et al. 2000)], South India ["SI1" n =85 cases, 328 controls (Jackson et al. 2004)], Canada ["BR1" n = 127 cases, 127 controls (Winckler et al. 2005)], USA ["GCI US" n = 1226 cases,

1226 controls (Winckler et al. 2005)], Poland ["GCI Poland" n = 1009 cases, 1009 controls (Winckler et al. 2005)], discordant sib sample from North America [" BR4" n = 609 cases and 580 controls (Winckler et al. 2005)] as well as a combined Scandinavian sample [" SC1" n = 471 cases, 471 controls (Winckler et al. 2005)]. Samples from the UK included in a previous publication (Weedon et al. 2005) were excluded from the meta-analysis of the A98V variant as there was overlap with some of the samples included in the current study. A previously published sample from China (Rissanen et al. 2000) was also excluded from the study of the A98V variant as no heterozygote carriers were identified in the control population. In total, 14279 cases and 26835 controls were included in the meta-analysis of the T130I variant of *HNF4A.* For the meta-analysis of the A98V variant of *HNF1A,* 11410 cases and 20481 controls were included.

4.2.2 Genotyping and Quality Control

For regions that had no previous probes designed, an "assay by design" was utilised. For regions for which previous assays were available "assay by demand" was used (Chapter 2, section 2.2.6). I carried out the genotyping (Chapter 2, section 2.2.6) of the UK samples for the T130I variant of *HNF4A*, the I88I variant of *KCNJ15* and the A98V variant of *HNF1A* using a Taqman assay on the ABI 7900HT platform (Applied Biosystems, Warrington, Cheshire, UK). The genotyping of the V255M variant of *HNF4A* and the D76N variant of *PDX1* in the UK samples was carried out in KBioscience (Hoddesdon, U.K.) using the Bioscience allele specific PCR (KASPar) assay.

Table 4.2 Characteristics of study subject included in the genotyping of V255M and T130I variants of *HNF4A,* A98V variant of *HNF1A,* D76N variant of *PDX1* and I88I variant of *KCNJ15.*

Sample name	Short Form	Sample name in previous publications[A]	Subjects		% Male		Mean Age[B] (years) ± SD		Mean BMI (kg/m^2) ± SD	
			Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases
UK Sample 1	UK1	UKT2DGC Sample	5126	4124	51	55	60 ± 13	58 ± 12	27 ± 5	31 ± 6
UK Sample 2	UK2	Diabetes UK Warren 2 repository, British 1958 birth cohort, UK blood service collection of common controls	10220	1853	50	61	42 ± 7	52 ± 7	27 ± 6[C]	32 ± 7
Danish Sample 1	DK1	Inter99, SDC1, SDC2	-	2646	-	59	-	57 ± 10	-	31 ± 6
Danish Sample 2	DK2	SDC2 and ADDITION	4865	1397	47	60	46 ± 9	52 ± 10	26 ± 4	30 ± 5
FUSION sample 1	FS1	Stage 1 genotyping Sample for DIAGRAM meta-analysis	1173	1160	49	56	64 ± 7	54 ± 9	27 ± 4	30 ± 5
FUSION Sample 2	FS2	Stage 2 genotyping Sample for DIAGRAM meta-analysis	1264	1211	61	60	59 ± 8	55 ± 9	27 ± 4	31 ± 5
METSIM	MS1	METSIM	3043	801	100	100	59 ± 6	58 ± 7	26 ± 4	30 ± 5
Broad Sample 1	BR1	Canada Sample	127	127	55	55	52 ± 8	52 ± 8[D]	29 ± 4	29 ± 5
Broad Sample 2	BR2	Scandinavia Sample	490	490	52	52	60 ± 10	60 ± 10[D]	27 ± 4	28 ± 5
Broad Sample 3	BR3	Sweden Sample	514	514	52	52	66 ± 12	66 ± 12[D]	28 ± 4	28 ± 4

A: Full details available in chapter 2; B: Mean age refers to age at recruitment for controls and age at diagnosis for cases

C: Based on the British 1958 Birth Cohort (n=7133) and Panel 2 of the United Kingdom Blood Services Collection (n=1643)

D: Based on age at recruitment (age of diagnosis not available)

The probes for the genotyping of the variants in this study have been included in Table 4.3. For the study of the I88I variant of *KCNJ15,* three heterozygote samples were obtained from Denmark and were used as positive controls to aid the allelic discrimination of the UK samples.

I subsequently analysed each plate by use of allelic discrimination SDS2.3 software (Applied Biosystems) using the ABI7900HT system (Applied Biosystems) for plates that utilised a Taqman assay and utilised the KBiosciences Klustercaller software (Kbiosciences, Hoddeston, UK) to plot and analyse the genotypes obtained for each SNP genotyped using the KASPar assay. Genotyping methods and quality control measures for Danish sample 1 have been previously reported (Ek et al. 2005). The T130I variant was genotyped in Danish sample 2 by collaborators in Denmark using the KASPar assay (KBioscience, Hoddesdon, U.K.). Quality of the genotyping was assured by (a) assessing the genotyping pass rate (in excess of 96% globally), (b) evaluating the estimated error rate based on completed duplicate pairs (UK samples: 0.00% for the V255M variant of *HNF4A*, n=314 duplicate pairs, 0.18% for the T130I variant of *HNF4A*, 268 duplicate pairs; 0.00% for the D76N variant of *PDX1*, 302 duplicate pairs, 0.18% for the A98V variant of *HNF1A*, 276 duplicate pairs, 0.00% for the rs374876 variant of *KCNJ15*, 280 duplicate pairs; Danish sample 2: 0.20% based on 521 duplicate samples for T130I variant of *HNF4A*), and (c) assessing for departure ($p<0.05$) from Hardy Weinberg Equilibrium (none detected).

Table 4.3 Genotyping probes used in this study of V255M and T130I variants of *HNF4A*, A98V variant of *HNF1A*, D76N variant of *PDX1* and I88I variant of *KCNJ15*

Variant	ABI Assay on Demand ID	KBioscience Assay by Design Primer Allele specific Sequence	KBioscience Assay by Design Primer Common Sequence
rs1800961 (T130I) *HNF4A*	C___7591528_10	N/A	N/A
rs1800574 (A98V) *HNF1A*	C_____5478_10	N/A	N/A
V255M *HNF4A*	N/A	**Allele X sequence** CGTCAAGGATGCGTATGGACAC **Allele Y sequence** CGTCAAGGATGCGTATGGACAT	CAACCCTTCCAGGCAATGACTACAT
rs374876 (I88I) *KCNJ15*	C__27476178_10	N/A	N/A
[+]D76N *PDX1*	N/A	[+]probe 1- VIC* CTCGCCGACGACC, probe 2- FAM* CTCGCCGACAACC	[+]F primer CAGCCCCCCGGACATCT R primer CGGGAGGTGTGGTGAAGGT

[+] ABI assay by Design probe and primer sequences
* VIC and FAM refer to the fluorescent dye employed in allelic discrimination.
N/A: Not available

4.2.3 Statistical analysis

No heterogeneity of genotype counts was seen between the T2D cases in the category 1 samples for any SNP when tested by an exact implementation of the Pearson chi-square test using StatXact (v6.0:Cytel software corps, Cambridge, MA, USA). The same was true of controls. I subsequently carried out an association analysis of the Category 1 samples for each SNP with each sample as separate stratum using an exact implementation of the Cochran-Armitage trend test. A further association analysis was carried out for variants that had Category 2 samples included. For this analysis each sample from Category 1 and Category 2 was incorporated as a separate stratum. For the meta-analysis of OR's of the previous published work for the T130I variant of *HNF4A* and the A98V variant of *HNF1A*, an additive model using the GWAMA

(Genome Wide Association Meta-Analysis) software package was utilised. (http://www.well.ox.ac.uk/gwama). A forest plot of the meta-analysis findings was then generated using the Comprehensive Meta-Analysis software version 2.050 (Biostat Inc., NJ, USA). Heterogeneity in the meta-analysis was evaluated by the calculation of Cochran's Q statistic and evaluating the I^2% for the meta-analysis [quantifying the effect of heterogeneity on the meta-analysis (Higgins et al. 2003)]. A fixed effect model was used for the meta-analysis as it was felt that the degree of heterogeneity was low for both variants (Figures 4.1 and 4.2).

Power calculations were carried out at the start of the project using the Quanto software package (Gauderman et al. 2006), incorporating the previously reported OR for each variant (Table 4.4). The effect size (for each variant) that this study had 90% power to detect was also calculated (Table 4.5). These calculations using the previously reported OR for theT130I variant of *HNF4A* (Ek et al. 2005) and the samples incorporated in Category 1, indicated 94% power to detect an effect size of 1.3 (for α=0.001). Power to detect an effect by the V255M variant of *HNF4A* was limited to large effect sizes only (90% power to detect OR of 3.21 for alpha = 0.001). For the A98V variant of *HNF1A* power calculations based on the OR seen the previous UK study of the A98V (Weedon et al. 2005) indicated 78% power to detect an OR of 1.3 (for α=0.001). For the study of the I88I variant of *KCNJ15*, this study had 99% power to detect an OR of 1.75 (previously reported OR in Japanese samples (Okamoto et al. 2010)) for α=0.001. The power of this study to detect the previously reported OR of 1.1 for the D76N variant of *PDX1* (Owen et al. 2004) in UK samples and the MAF of 1% (Owen et al. 2004) was limited to 0.1% for α=0.001. However, there was 90% power to detect an OR of 1.58 for α=0.001

Table 4.4 Power calculations for association testing for V255M and T130I variants of *HNF4A*, A98V variant of *HNF1A*, D76N variant of *PDX1* and I88I variant of *KCNJ15* based on previously reported OR's

Gene	Variant	MAF	Alpha	Previously reported OR	Category 1 samples	Category 1+2 Samples	Category 1+2+ 3 samples
					Power (N= Cases, Controls)	Power (N= Cases, Controls)	Power (N= Cases, Controls)
HNF1A	A98V	2.75%	0.001	1.30	78% (5317, 15346)	NA	98% (11331, 22368)
PDX1	D76N	1.00%	0.001	1.10	0.1% (5977, 15336)	NA	NA
KCNJ15	I88I	0.89%	0.001	1.75	99% (6193, 16690)	99% (10236, 21555)	NA
HNF4A	T130I	3.00%	0.001	1.30	94% (7963, 15346)	99% (13636, 26736)	99% (14732, 27629)
HNF4A	V255M	0.10%	0.001	1.00	0.1% (5977, 15346)	0.1% (7374, 20211)	NA

OR: Odds Ratio; NA: Not Applicable

Table 4.5: Power calculations to evaluate the effect size which this project had 90% power to detect V255M and T130I variants of *HNF4A*, A98V variant of *HNF1A*, D76N variant of *PDX1* and I88I variant of *KCNJ15*

Gene	Variant	MAF	Alpha	Power	Category 1 samples	Category 1 +2	Category 1+2+3
					OR (N= Cases, Controls)	OR (N= Cases, Controls)	OR (N= Cases, Controls)
HNF1A	A98V	2.75%	0.001	90%	1.34 (5317, 15346)	NA	1.24 (11331, 22368)
PDX1	D76N	1.00%	0.001	90%	1.56 (5977, 15336)	NA	NA
KCNJ15	I88I	0.89%	0.001	90%	1.58 (6193, 16690)	1.47 (10236, 21555)	NA
HNF4A	T130I	3.00%	0.001	90%	1.28 (7963, 15346)	1.22 (13636, 26736)	1.21 (14732, 27629)
HNF4A	V255M	0.10%	0.001	90%	3.24 (5977, 15346)	2.92 (7374, 20211)	NA

OR: Odds Ratio; NA: Not Applicable MAF; Minor Allele Frequency

4.3 Results

4.3.1 Association analysis of the A98V variant of *HNF1A*

The A98V variant of *HNF1A* was successfully genotyped in 4714 cases and 14114 controls from the UK. The MAF of this variant was 2.75% in cases and 2.77% in controls. No association with T2D risk was seen in the analysis of the UK samples with an additive per allele OR of 1.05 (0.89-1.22); p = 0.55 (Table 4.4). However, the meta-analysis incorporating all previously published reports of the association analysis of the A98V variant of *HNF1A* (Figure 4.1) (n = 11410 cases, 20481 controls) did demonstrate a trend towards association with T2D risk, though failed to reach statistical significance (OR 1.11 [0.99-1.23]; p= 0.07).

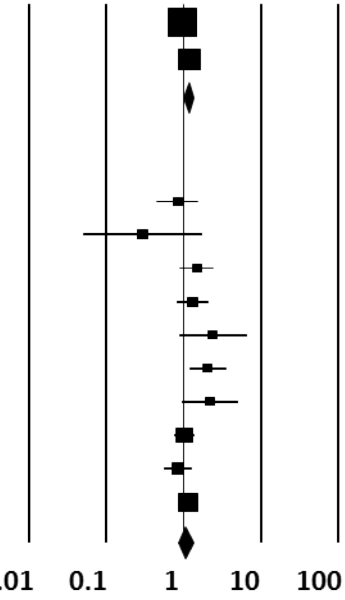
Table 4.6 Association testing of the A98V (rs1800574) variant of *HNF1A* – Category 1 samples

Study sample	**N**	**Cases: Genotype counts (Genotype frequencies)**			**Controls: Genotype counts (Genotype frequencies)**			**Statistical Analysis**	
	Cases/ Controls	CC	CT	TT	CC	CT	TT	OR (95% CI)	P_{Add}
UK sample 1	3598/ 4280	3418 (95.02%)	174 (4.84%)	5 (0.14%)	4057 (94.79%)	220 (5.14%)	3 (0.07%)	0.97 (0.79-1.18)	0.76
UK sample 2	1116/ 9864	1043 (93.46%)	71 (6.36%)	2 (0.17%)	9316 (99.84%)	539 (5.46%)	9 (0.09%)	1.20 (0.92-1.53)	0.16
UK sample 1 and 2	4714/ 14114	4461 (94.65%)	245 (5.20%)	7 (0.15%)	13373 (94.55%)	759 (5.37%)	12 (0.08%)	1.05 (0.89-1.22)	0.55

CC: Common Homozygote; CT: Heterozygote; TT: Rare Homozygote
N: Total number successfully genotyped
OR: Odds Ratio
P_{ADD}: P value for additive model

Figure 4.1 Meta-analysis of all published studies of the A98V variant of *HNF1A*

Study name	N (cases, controls)	Statistics for each study: Odds ratio	Lower limit	Upper limit	p-Value
Category 1 samples					
This study (UK1)	4280, 3597	0.97	0.79	1.18	0.75
This study (UK2)	1116, 9864	1.20	0.93	1.54	0.17
Subtotal (Category 1)	1116, 9864	1.05	0.89	1.22	0.55
(I^2 = 40%; Q statistic p=0.20)					
Category 3 Samples					
Urhammer et al (Denmark, DK3)	240, 245	0.84	0.44	1.60	0.60
Winckler et al (Canada, BR1)	127, 127	0.30	0.07	1.33	0.11
Winckler et al (Scandinavia, BR2)	471, 471	1.48	0.87	2.52	0.15
Winckler et al (Sweden, BR3)	514, 514	1.31	0.80	2.15	0.29
Winckler et al (Discordant sibs, BR4)	609, 580	2.40	0.87	6.61	0.09
Jackson et al (South India, SI1)	85, 328	2.05	1.16	3.62	0.01
Rissansen et al (Finland, FL1)	222, 295	2.20	0.94	5.13	0.07
Winckler et al (GCI US)	1226, 1226	1.03	0.75	1.42	0.86
Winckler et al (GCI Poland)	1009, 1009	0.83	0.54	1.27	0.39
Holmkvist et al (Scandinavia, SC1)	1511, 2225	1.13	0.86	1.49	0.38
Overall (Category 1 and 3)		1.11	0.99	1.23	0.07
(I^2 = 41%; Q statistic p=0.07)					



Plot generated using Comprehensive Meta-Analysis software version 2.2050, Biostat Inc., NJ, USA

4.3.2 Association analysis of the D76N Variant of *PDX1*

The D76N variant of *PDX1* was successfully genotyped in 5774 T2D cases and 15027 controls. The MAF in the T2D samples in this study was 1.66% and was 1.44% in controls. No association with T2D risk was seen in UK samples in this study, OR 0.96 (0.74-1.25); p = 0.79 (Table 4.7).

Table 4.7 Association analysis of the D76N variant of *PDX11*

Study sample	N	Cases :Genotype counts (Genotype frequencies)			Controls: Genotype counts (Genotype frequencies)			Statistical Analysis	
	Cases/ Controls	GG	GA	AA	GG	GA	AA	OR (95% CI)	P_{Add}
UK sample 1	3962/ 4900	3886 (98.08%)	76 (1.92%)	0 (0%)	4820 (98.37%)	80 (1.63%)	0 (0%)	1.18 (0.85-1.64)	0.33
UK sample 2	1812/ 10127	1792 (98.90%)	20 (1.10%)	0 (0%)	9993 (98.68%)	131 (1.29%)	3 (0.03%)	0.82 (0.49-1.31)	0.82
UK sample 1 and 2	5774/ 15027	5678 (98.34%)	96 (1.66%)	0 (0%)	14813 (98.58%)	211 (1.40%)	3 (0.02%)	0.96 (0.74-1.25)	0.79

GG: Common Homozygote; GA: Heterozygote; AA: Rare Homozygote
N: Total number successfully genotyped
OR Het: Odds Ratio for Heterozygote carriers compared to GG
P_{ADD}: P value for additive model

4.3.3 Association analysis of the I88I variant of *KCNJ15*

The I88I variant of *KCNJ15* was successfully genotyped in 6632 T2D samples and 16726 population controls from the UK. The MAF of this variant was 0.54% in cases and 0.78% in controls. It was noted that the MAF in control samples from UK sample 1 (recruited in Dundee) was 0.59%, though the MAF of the control samples from UK sample 2 (recruited throughout the UK) was 0.89%. To evaluate whether there may be a geographical difference in the distribution of the MAF across the UK, the samples for which there was geographical information were

further analysed. The samples were divided into three broad groups (1- the south of England, 2- Midlands and Wales, 3- Scotland). The Jonckeere-Terpestra test (using STATXACT v6.0) was utilised to evaluate whether there was evidence of stratification in the UK samples (Table 4.8). Although a trend was seen this did not reach statistical significance (p=0.16).

Table 4.8 Test for geographical stratification of the I88I variant of *KCNJ15* in the UK population

Study sample	**N**	**Cases Genotype counts (Genotype frequencies)**			**Statistical Analysis**	
		CC	**CT**	**TT**	**MAF**	**Jonckeere-Terpestra test P value**
Scotland	201	198 (98.51%)	3 (1.49%)	0 (0%)	0.74%	
Midlands	1044	1029 (98.56%)	15 (1.44%)	0 (0%)	0.72%	
South of England	992	969 (97.68%)	23 (2.32%)	0 (0%)	1.16%	
Overall	2237	2196 (98.17%)	41 (1.83%)	0 (0%)		0.16

CC: Common Homozygote; CT: Heterozygote; TT: Rare Homozygote
MAF: Minor Allele Frequency

Statistically significant association of the I88I variant of *KCNJ15* with T2D risk was not seen in the UK samples, OR 0.78 (0.58-1.04); p = 0.09 (Table 4.9). Incorporation of the Danish genotyping that has been previously published for this variant did not greatly alter this finding, OR 0.82 (0.65-1.03); p = 0.08 (Table 4.9).

Table 4.9 Association study of I88I variant *KCNJ15*

Study sample	N	Cases Genotype counts (Genotype frequencies)			Controls Genotype counts (Genotype frequencies)			Statistical Analysis	
	Cases/ Controls	CC	CT	TT	CC	CT	TT	OR (95% CI)	P_{Add}
UK sample 1	4996/ 6468	4945 (98.98%)	51 (1.02%)	0 (0%)	6391 (98.81%)	77 (1.19%)	0 (0%)	0.86 (0.58-1.24)	0.42
UK sample 2	1636/ 10258	1616 (98.78%)	20 (1.22%)	0 (0%)	10074 (98.21%)	184 (1.79%)	0 (0%)	0.68 (0.40-1.08)	0.10
Category 1 UK sample 1 and 2	6632/ 16726	6561 (98.93%)	71 (1.07%)	0 (0%)	16465 (98.44%)	261 (1.56%)	0 (0%)	0.78 (0.58-1.04)	0.09
Category 2* DK 1 and 2	3531/ 4885	3482 (98.61%)	49 (1.39%)	0 (0%)	4809 (98.44%)	76 (1.56%)	0 (0%)	0.89 (0.61-1.30)	0.58
Category 1+2 (UK1,UK2,DK1,DK2)	10163/ 21611							0.82 (0.65-1.03)	0.08

CC: Common Homozygote; **CT**: Heterozygote; **TT**: Rare Homozygote
N: Total number successfully genotyped
OR: Odds Ratio
P_{ADD}: P value for additive model
*data from previous publication (Okamoto et al. 2010)

4.3.5 Association analysis of the V255M and T130I variants of *HNF4A*

T130I was successfully genotyped in 7645 cases and 14756 controls in category 1. This variant had an MAF of 3.76% in cases and 3.00% in controls. In category 1 samples, a modest association T2D was conferred by the T allele of this variant (additive per allele OR 1.20 [1.08-1.33]; $p=5x10^{-4}$). The expanded association analysis incorporating the category 1 and 2 samples marginally increased the strength of this association (OR 1.17 [1.08-1.28]; $p=1.5x10^{-4}$) (Table 4.10). The meta-analysis (Figure 4.2) incorporating all available studies of T130I further increased the strength of this association (n=14279 cases, 26835 controls; OR 1.20 [1.10-1.30]; $p=2.1x10^{-5}$). There was no evidence for heterogeneity in our meta-analysis (I^2= 61%; Q statistic P value= 0.10).

The V255M variant was successfully genotyped in 5745 cases and 15044 controls in the UK study. The MAF for V255M was far lower than for T130I (cases 0.08%, controls 0.10%), and no T2D association was observed in either the UK sample (p=0.28) or the larger association analysis incorporating Danish genotyping data (p=0.40) (Table 4.11).

We did not find any evidence for linkage disequilibrium ($r2<0.005$) between either T130I or V255M and the common variants in *HNF4A* that had previously shown an association with T2D susceptibility.

Table 4.10 Association testing of the T130I variant of *HNF4*

Study sample	N	Cases: Genotype counts (Genotype frequencies)			Controls: Genotype counts (Genotype frequencies)			Statistical Analysis	
	Cases/Controls	CC	CT	TT	CC	CT	TT	OR (95% CI)	P_{Add}
UK1	4040/5012	3771 (93.34%)	266 (6.59%)	3 (0.07%)	4711 (94.00%)	291 (5.81%)	10 (0.19%)	1.09 (0.92-1.29)	0.333
UK2	1158[A]/9744	1069 (92.31%)	89 (7.69%)	0 (0.00%)	9175 (94.16%)	563 (5.78%)	6 (0.06%)	1.32 (1.04-1.66)	0.020
DK1	2447/0	2240 (91.54%)	200 (8.17%)	7 (0.29%)	-	-	-	-	-
DK2	1324/4727	1216 (91.84%)	106 (8.01%)	2 (0.15%)	4412 (93.33%)	305 (6.45%)	10 (0.21%)	1.21 (0.97-1.51)	0.090
DK1 and DK2	3771/4727	3456 (91.65%)	306 (8.11%)	9 (0.24%)	4412 (93.33%)	305 (6.45%)	10 (0.21%)	1.26 (1.07-1.47)	0.004
Category 1 (UK1, UK2, DK1, DK2)	7645/14756							1.20 (1.08-1.33)	$5x10^{-4}$
FS1	1160/1172	1057 (91.12%)	98 (8.44%)	5 (0.43%)	1087 (92.75%)	82 (7.00%)	3 (0.26%)	1.24 (0.93-1.66)	0.134
FS2	1181/1237	1069 (90.52%)	109 (9.23%)	3 (0.25%)	1142 (92.32%)	93 (7.52%)	2 (0.16%)	1.25 (0.94-1.67)	0.122
MS1	774/2957	720 (93.02%)	52 (6.72%)	2 (0.26%)	2760 (93.34%)	191 (6.46%)	6 (0.20%)	1.05 (0.77-1.43)	0.758
BR1	111/110	107 (96.40%)	4 (3.60%)	0 (0.00%)	106 (96.36%)	4 (3.64%)	0 (0.00%)	0.99 (0.18-5.46)	1.000
BR2	488/483	460 (94.26%)	26 (5.33%)	2 (0.41%)	451 (93.37%)	31 (6.42%)	1 (0.21%)	0.90 (0.54-1.51)	0.711
BR3	500/500	461 (92.20%)	36 (7.20%)	3 (0.60%)	459 (91.80%)	39 (7.80%)	2 (0.40%)	0.98 (0.63-1.52)	1.000
Category 1 and 2 (UK1, UK2, DK1, DK2, FS1, FS2, MS1, BR1, BR2, BR3)	13183/25942							1.17 (1.08-1.28)	$1.5x10^{-4}$
Category 1, 2 and 3 (UK1, UK2, DK1, DK2, FS1, FS2, MS1, BR1, BR2, BR3, PI1, JP1, MX1)	14279/26835							1.20 (1.10-1.30)	$2.1x10^{-5}$

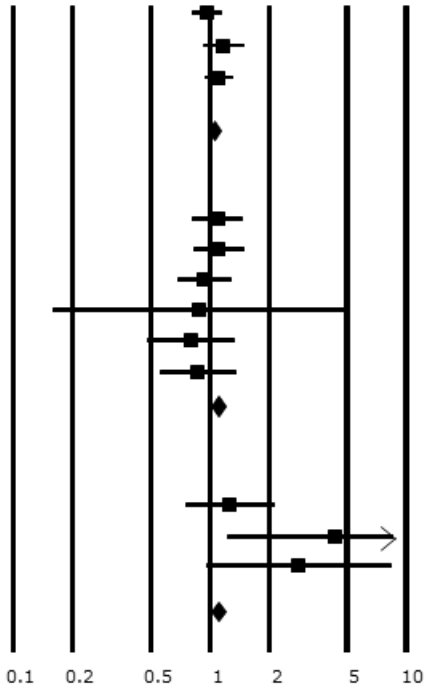
CC: Common Homozygote; CT: Heterozygote; TT: Rare Homozygote

N: Total number successfully genotyped

OR: Odds Ratio ; P_{ADD}: P value for additive model

Figure 4.2 Meta-analysis of the association studies of the T130I variant to Type 2 diabetes

Study name	N Cases/Controls	Risk Allele Frequency	Odds ratio	Lower limit	Upper limit
Category 1					
UK sample 1	4040/5012	0.03	1.09	0.92	1.29
UK sample 2	1158/9744	0.03	1.32	1.04	1.67
Danish samples	3771/4727	0.04	1.25	1.07	1.46
Subtotal; category 1 (I^2=7% , Q statistic p=0.34)	8969/19483		1.20	1.08	1.33
Category 2					
FUSION sample 1	1160/1172	0.04	1.24	0.93	1.66
FUSION sample 1	1181/1237	0.04	1.25	0.94	1.67
METSIM sample	774/2957	0.03	1.05	0.77	1.43
Broad sample 1	111/110	0.02	0.99	0.18	5.46
Broad sample 2	488/483	0.03	0.90	0.54	1.51
Broad sample 3	500/500	0.04	0.98	0.63	1.52
Subtotal; category 1,2 (I^2=0% , Q statistic p=0.77)	13183/25942		1.17	1.06	1.29
Category 3					
Pima sample	573/464	0.04	1.42	0.85	2.38
Japanese sample	423/354	0.02	5.05	1.44	17.65
Mexican sample	100/75	0.13	3.38	1.14	10.02
Overall category 1,2,3 (I^2=61% , Q statistic p=0.10)	14279/26835		1.20	1.10	1.30



[+] Plot generated using Comprehensive Meta-Analysis software version 2.2050, Biostat Inc., NJ, USA.

Table 4.11 Association testing of the V255M variant of *HNF4A*

Study sample	N	Cases Genotype counts (Genotype frequencies)			Controls Genotype counts (Genotype frequencies)			Statistical Analysis	
	Cases/Controls	GG	GA	AA	GG	GA	AA	OR (95% CI)	P_{Add}
UK1	3949/4925	3942 (99.82%)	7 (0.18%)	0 (0.00%)	4911 (99.72%)	14 (0.28%)	0 (0.00%)	0.62 (0.21-1.65)	0.38
UK2	1796/10119	1794 (99.89%)	2 (0.11%)	0 (0.00%)	10103 (99.84%)	16 (0.16%)	0 (0.00%)	0.76 (0.08-2.99)	0.76
Category 1 (UK1, UK2)	5745/15044							0.64 (0.26-1.44)	0.28
DK 2	1360/4791	1357 (99.78%)	3 (0.22%)	0 (0.00%)	4781 (99.79%)	10 (0.21%)	0 (0.00%)	1.05 (0.19-4.11)	1.00
Category 1 and 2 (UK1, UK2, DK2)	7105/19835							0.73 (0.34-1.47)	0.44

GG: Common Homozygote; GA: Heterozygote; AA: Rare Homozygote
N: Total number successfully genotyped
OR: Odds Ratio
P_{ADD}: P value for additive model

4.4 Discussion

Previous functional and association studies had highlighted coding LF variants within *HNF4A*, *HNF1A*, *PDX1* and *KCNJ15* as potentially influencing susceptibility to multifactorial T2D. I have carried out the largest association analysis for the V255M and T130I variants of *HNF4A*, the A98V variant of *HNF1A,* the D76N variant of *PDX1* and I88I of *KCNJ15* to better understand their role in T2D pathogenesis.

No association was found between the V255M variant of *HNF4A*, the A98V variant of *HNF1A*, the D76N variant of *PDX1* and I88I variant of *KCNJ15* and T2D risk in the UK samples or in our larger analysis. It is worth emphasising that the low MAF of the V255M variant of *HNF4A,* the I88I variant of *KCNJ15* and the D76N variant of *PDX1* means that the power to detect association was limited to larger effect sizes only (OR >1.5).

In contrast, evidence of association between the T130I variant of *HNF4A* and T2D risk was found in our association analysis. The strength of association was increased in the meta-analysis incorporating all previous studies of this variant. This association however, fails by some margin, to reach widely-accepted thresholds for genome-wide significance (α=5x10^{-8}), which in the context of LF variants should be even more stringent than those required for common variants (perhaps around α=5x10^{-9}) given the larger number of independent tests that are possible once lower-frequency variants are considered. We estimate that to achieve this level of significance for T130I (using the effect size [an OR of 1.20] observed in our expanded meta-analysis) would require almost 100,000 samples (in fact, 48697 cases and 48697 controls).

Recent evidence (achieving such levels of genome wide significance) that T130I is associated with altered HDL-cholesterol levels raises the prior odds that the T2D association we observe here is genuine. The previously reported association with lipid levels also raises an interesting question as to whether or not the T2D association is mediated by a direct influence of the variant (or a causal variant with which it is in linkage disequilibrium) on β-cell function, or a primary effect on lipid physiology. The latter question could in principle be answered by a suitably-scaled Mendelian randomisation experiment. The broad transcriptional effects of HNF-4α would be consistent with pleiotropic effects of the variant on multiple systems.

Although no association between the I88I variant of *KCNJ15* and T2D risk was seen in UK and Danish samples in this study, it is possible that the original association seen in the Japanese samples is due to another variant with which I88I is in linkage disequilibrium. Differing haplotype structures in the Japanese and European samples may explain the lack of association seen in the latter group. The 1000 genome project (http://www.1000genomes.org) achieved a low level of coverage around the I88I variant in Japanese samples. However, on review of the linkage disequilibrium data from the 1000 genome project, 6 SNPs (rs2836296, rs2836291, rs16996036, rs2836231, rs2836237 and rs2065306) achieved an $r^2 > 0.8$ with the I88I variant in CEPH samples (Utah residents with ancestry from northern and western Europe). In Japanese samples only 4 of these SNPs (rs2836296, rs2836291, rs16996036 and rs2066306) achieved an $r^2 > 0.8$. These data would indicate that there are differing haplotype structures around the I88I variant in Japanese and European samples. As large scale high throughput sequencing efforts are completed, a better understanding of the sequence variation and haplotype structure in disparate

samples will be achieved. This would allow a more thorough study of susceptibility variants in samples with differing haplotype structures.

Some of the variants prioritised in this study (V255M variant of *HNF4A* and D76N variant of *PDX1*) had been highlighted as potentially impacting on T2D risk largely based on functional studies. Such studies utilise various cell lines such as the COS7 (kidney cells from the African green monkey) and HepG2 cell lines (perpetual cell line from an individual with hepatocellular carcinoma). These are not the native cells that T2D disease state would impact. Although such studies are important in demonstrating the transcriptional regulation of target genes, their results need to viewed with caution as they may not necessarily reflect the situation *in vivo* in the disease state being studied.

As efforts are increasingly aimed at understanding the full allelic spectrum of variants involved in multifactorial disease pathogenesis, it is likely that many more LF variants will be identified. Large-scale genotyping will be required to clarify the role played by LF variants and as the results from this study exemplify, the numbers required for association testing of low frequency variants are substantial when the effect size is modest. It is clear that large collaborative efforts will be needed to maximise the samples available for any such studies

CHAPTER 5

Genetic association analysis of low frequency variants in *LARS2* with Type 2 Diabetes

5.1 Introduction

The *LARS2* gene encodes for the mitochondrial leucyl tRNA synthetase (EC 6.1.1.4), which catalyses the aminoacylation of both mitochondrial leucyl tRNAs with leucine and is therefore essential for mitochondrial protein synthesis. This is of particular interest as mitochondrial dysfunction is observed in patients with type 2 diabetes (T2D) and their first degree relatives (Gerbitz et al. 1996; Kelley et al. 2002; Bajpeyi et al. 2011). Previous studies have indicated that genes involved in oxidative phosphorylation are down regulated in the muscle cells of T2D patients (Mootha et al. 2003). Furthermore, the muscle mitochondria from T2D patients have an impaired bioenergetic capacity (Kelley et al. 2002). Mitochondria also play an important role in insulin secretion and sensitivity (Matschinsky et al. 1998; Maassen et al. 2007). Previous studies have shown that a variant in the mitochondrial DNA encoded tRNA-Leu (UUR) gene (M3243 A>G) is associated with maternally inherited diabetes and deafness (van den Ouweland et al. 1992).

Previous work in the Netherlands had highlighted a low frequency (MAF 3.0%) nonsynonymous variant of *LARS2*, p.His324Gln (H324Q), rs71645922, as potentially increasing T2D susceptibility risk in a meta-analysis of four independent cohorts from the Netherlands and Denmark (OR = 1.40 [95% CI 1.12-1.76], P = 0.004, n = 7836) (t Hart et al. 2005). The reported association for the H324Q variant (t Hart et al. 2005) did not achieve the now widely accepted genome wide association level of significance ($\alpha = 5 \times 10^{-8}$).

Fourteen low-frequency single nuclear polymorphisms (SNPs) (0.01>MAF<0.05) within *LARS2* (Figure 5.1) have been included in a Genome wide association study (GWAS) meta-analysis by the Diabetes Genetics Replication And Meta-analysis Consortium (DIAGRAM) (Zeggini et al. 2008). Ten of these SNPs are in high linkage disequilibrium with each other (r^2>0.95) and

were modestly associated with T2D (ORs 1.17–1.21; p 0.02–0.05). A proxy for this group (c.388-12390A>G, rs9825041; OR 1.20 [95% CI 1.03–1.39], p=0.02) was selected, by our group in Oxford and collaborators in the Netherlands, to further clarify their role T2D pathogenesis. The H324Q variant was not captured by this meta-analysis and is not in linkage disequilibrium with the group of low frequency variants noted above (r^2< 0.17).

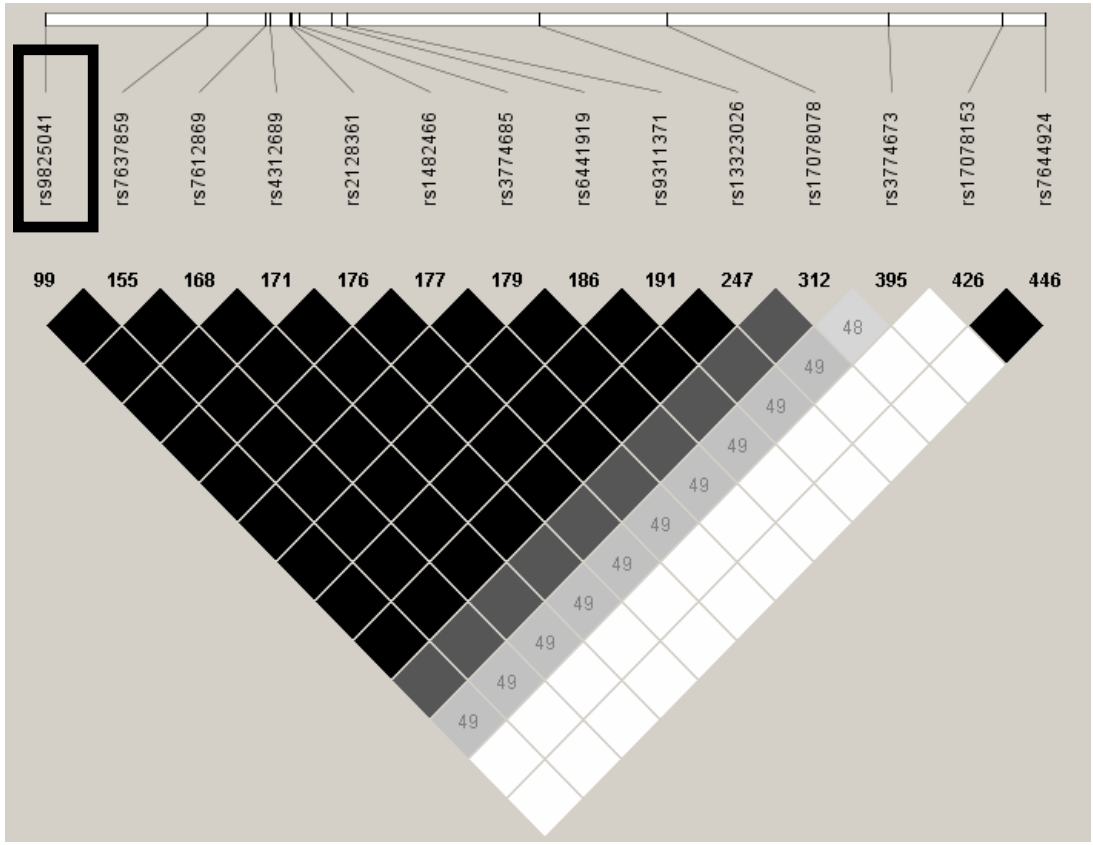
It was therefore the aim of this study to more accurately establish the contribution of both *LARS2* candidate low frequency incompletely penetrant variants (H324Q and rs9825041) to T2D pathogenesis through large scale association testing in samples from the UK. I subsequently carried out a meta-analysis incorporating the results from the UK samples and samples from Netherlands, Denmark, Finland and Sweden that were genotyped specifically for this study. An expanded meta-analysis was subsequently carried out incorporating all previous studies of the H324Q variant.

5.2 Materials and methods

5.2.1 Study samples

Three sample categories were included in the genotyping of the *LARS2* variants in this study. Category 1 samples included samples from the UK that I genotyped for this project. Category 2 samples included samples that were genotyped specifically for this study by collaborators and the summary statistics were made available for inclusion in this project.

Figure 5.1 Linkage disequilibrium plot of the 14 low-frequency SNPs in *LARS2* included in the DIAGRAM scan. rs9825041 was used as a proxy for the block and is shown in the box



For the study of the H324Q variant of *LARS2* only, category 3 samples were included. These were samples that had previously been genotyped for this variant and the results previously reported ('t Hart et al. 2005). The summary statistics from this latter study were then incorporated in an expanded meta-analysis.

Two samples from the UK were included in the genotyping of both *LARS2* variants, here designated UK sample 1 (UK1, maximum n=4124 cases, 5126 controls) and UK sample 2 (UK2, maximum n=1853 cases, 10220 controls). Following the association study in the UK samples, I carried out a meta-analysis incorporating the UK samples and samples genotyped in samples from the Netherlands (NL2, NL3 and NL4, total n=2027 cases, 7093 controls for the study of the H324Q variant and NL1, NL2, NL3 and NL4, total n = 3024 cases, 8139 controls for the study of rs9825041), Finland (FL1, n=402 cases, 353 controls, FL2, n=211 trios probands, 370 parents), one Danish sample (DK3, n=706 cases, 514 controls included for the study of rs9825041 only) and one Swedish sample (SE1, n=480 cases and 468 controls). The subject characteristics and collection details have been included in Chapter 2 (Section 2.1). All study samples are summarized in Table 5.1.

Category 1 samples

The first UK sample (UK1) was comprised of the UKT2DGC (United Kingdom Type 2 Diabetes Genetics Consortium) case-control sample including 4124 T2D cases and 5126 controls ascertained in Tayside, Scotland. Details of the ascertainment scheme and recruitment criteria for this sample have been described in Chapter 2 (Section 2.1). The second sample (UK2) consists of 1853 T2D cases ascertained as part of the BDA (British Diabetes Association- now Diabetes UK) Warren 2 collection and 10220 control samples. The latter represent the full British 1958 Birth Cohort (n=7133) and The United Kingdom Blood Services

Collection of Common Controls (UKBS) (n= 3087), a subset of which featured in the WTCCC genome wide association scan (Wellcome Trust Case Control Consortium 2007; Zeggini et al. 2007). Both UK samples were of British ancestry.

Category 2 and 3 samples

Netherlands sample 1 (NL1, Chapter 2, Section 2.1) was from the Hoorn Study (Mooy et al. 1995), a Dutch population-based study from the city of Hoorn (n= 480 T2D cases and 519 controls). This sample was used in the genotyping of rs9825041. As this sample was included in the original association study of the H324Q variant (t Hart et al. 2005), it was included as a category 3 sample for the study of this variant.

Netherlands sample 2 (NL2, Chapter 2, Section 2.1) included 821 T2D cases and 1517 controls (Reiling et al. 2009; van 't Riet et al. 2010). The 1517 controls were randomly selected from the New Hoorn Study (NHS) which is an on-going, population-based study from the city of Hoorn which does not overlap with the original NL1 sample. Of the T2D patients, 147 were from the NHS and the remainder (n = 674) were recruited from the diabetes clinics of the Leiden University Medical Centre and the Vrije Universiteit medical centre, Amsterdam.

Netherlands sample 3 (NL3, Chapter 2, Section 2.1) was ascertained from the Breda study (van Tilburg et al. 2003; Monsuur et al. 2005). This is a case–control study from the city of Breda, in the south of the Netherlands and included 501 T2D cases and 920 controls.

The fourth Dutch sample (Netherlands sample 4, NL4) included 1222 T2D cases and 5183 controls from the population-based ERGO study (NL4) from Rotterdam in the south-west

region of the Netherlands (Hofman et al. 1991). Participants in NL4 that had been included in the previous reported study of the H324Q variant (t Hart et al. 2005) and were included as category 3 samples only from the study of this variant (n=517 cases and 517 controls).

In total 2027 cases and 7093 controls from the Netherlands were included in the association study of the H324Q variant and 3024 cases and 8139 controls. All samples from the Netherlands were of Dutch white ancestry.

One sample from Denmark was included in the study of rs9825041 [Denmark sample 3; DK3 (Denmark sample 1 and 2 were included as part of the genotyping of variants highlighted in Chapter 4)] (Drivsholm et al. 2001). This sample consisted of 706 T2D cases and 514 controls (Chapter 2, Section 2.1. This sample was included in the original report of the association testing of the H324Q variant in T2D pathogenesis (t Hart et al. 2005).

Finally, samples from Finland and Sweden were included for the study of the H324Q variant. Finland sample 1 (FL1, Chapter 2, Section 2.1) was a case–control sample from the Botnia region of Finland, consisting of 402 T2D cases and 353 controls (Groop et al. 1996; Florez et al. 2007). A set of trios originating from the Botnia region of Finland was also included. This Finland Sample 2 (FL2, Chapter 2, Section 2.1) consisted of 211 probands (multiple diabetic siblings) and 370 parents (Groop et al. 1996; Florez et al. 2007). Sweden sample 1 (SE1, Chapter 2, Section 2.1) was from a case–control study from Skara and Malmö, and consisted of 480 T2D patients and 468 controls (Saxena et al. 2006).

In total, 10800 T2D patients and 25191 controls and were included for follow up of the *LARS2* variants. These samples were not included in the original association study of the H324Q variant with T2D risk.

5.2.2 Genotyping and quality control

Genotyping (Chapter 2, section 2.2.6) of the H324Q variant in the UK samples was carried out in KBioscience (Hoddesdon, U.K.) using the Bioscience allele specific PCR (KASPar) assay. I carried out the genotyping of the rs9825041 variant in the UK samples using a Taqman "assay on demand " (Chapter 2, Section 2.2.6) on the ABI 7900HT platform (Applied Biosystems, Warrington, Cheshire, UK). I subsequently analysed each plate by use of allelic discrimination SDS2.3 software (Applied Biosystems) using the ABI7900HT system (Applied Biosystems) for plates that utilised a Taqman assay and utilised the KBiosciences Klustercaller software (Kbiosciences, Hoddeston, UK) to plot and analyse the genotypes obtained for the H324Q variant, genotyped using the KASPar assay.

Table 5.1 Study Subject Characteristics included in the study of the *LARS2* variants.

Sample name	Short form	Sample name in previous publications	Subjects		%male		Mean Age at study entry (years) ± SD		Mean BMI (kg/m^2) ± SD	
			Cases	Controls	cases	controls	Cases	controls	cases	controls
UK sample 1	UK1	UKT2DGC Sample	4124	5126	55	51	66 ± 6	60 ± 13	31.2± 6.0	26.9 ± 4.6
UK sample 2	UK2	Diabetes UK Warren 2 repository, British 1958 birth cohort, UK blood service collection of common controls	1853	10220	61	50	57 ± 9	42 ± 7	31.8 ± 6.7	27.2± 6.4[a]
Netherlands sample 1	NL1	Hoorn Study	480	519	52	55	67 ± 8	65 ± 8	28.8 ± 4.6	26.4 ± 4.5
Netherlands sample 2	NL2	New Hoorn study	821	1517	50	44	61 ± 11	53 ± 7	29.0 ± 4.6	25.5 ± 3.6
Netherlands sample 3	NL3	Breda study	501	920	46	61	71 ± 10	48 ± 13	27.8 ± 4.1	n.a.
Netherlands sample 4	NL4	ERGO study	1222	5183	39	41	73 ± 9	69 ± 9	27.4 ± 4.0	26.0 ± 3.9
Danish sample 3	DK3	Steno Diabetes Centre Samples	706	514	46	48	59 ± 10	57 ± 10	n.a.	n.a.
Finland sample 1	FL1	Botnia case control sample	402	353	55	53	61 ± 10	60 ± 10	28.7 ± 4.5	26.1 ± 3.6
Finland sample 2	FL2	Botnia Trios	211[c]	371[b]	47[c]	47[b]	40 ± 9	n.a	n.a.	28.5 ± 5.5
Sweden sample 1	SE1	Skara and Malmö case-control study	480	468	53	52	67 ± 11	66 ± 12	27.9 ± 4.1	27.5 ± 4.1

n.a. not available

a. Based on the British 1958 Birth Cohort (7133) and Panel 2 of the United Kingdom Blood Services Collection of Common Controls (n=1643)

b. parents

c. proband

The genotyping of the H324Q and rs9825041 variants in all other samples was also carried out on the ABI 7900HT platform by collaborators in the Leiden University Medical Center, Netherlands (NL1, NL2, NL3 and NL4), Steno Diabetes Center, Denmark (DK3), and the Broad Institute, USA (FL1, FL2, and SE1). The genotyping probes used in this study have been included in Table 5.2.

Table 5.2 Genotyping probes used in the study of rs982541 and H324Q of *LARS2*

Variant	ABI Assay on Demand ID	KBioscience Assay by Design Primer Allele specific Sequence	KBioscience Assay by Design Primer Common Sequence
rs71645922 (H324Q) *LARS2*	C__98755403_10	**Allele X sequence** CAAGGCTTCCTTCAGAGAGCTG **Allele Y sequence** CAAGGCTTCCTTCAGAGAGCTT	GGCACCTCCCACGTGGCCAT
rs9825041 (c.388-12390A>G) *LARS2*	C__29969498_10	N/A	N/A

The overall genotyping success rate was in excess of 96% for all samples included in this study. No departure from Hardy Weinberg Equilibrium ($p < 0.05$) was noted for either variant. This was evaluated using PEDSTATS software package; http://www.sph.umich.edu/csg/abecasis/PedStats (Wigginton et al. 2005). Genotyping of duplicate samples (~5%) showed complete concordance.

5.2.3 Statistical analysis

No heterogeneity of genotype counts was seen between the T2D cases for either variant when tested by the Pearson chi-square test using StatXact (v6.0:Cytel software corps, Cambridge, MA, USA). The same was true of controls. An association analysis of the UK samples was subsequently carried out with UK sample 1 and UK sample 2 each as a separate stratum using an exact implementation of the Cochran Armitage Trend test using StatXact.

The differences in genotype distribution and allele frequencies in samples from the Netherlands, Denmark, Sweden and Finland were analysed by collaborators in the Netherlands using a chi-squared test and odds ratios were subsequently calculated using an additive model, which was the best fit for the data. I then carried out a meta-analysis incorporating the summary statistics from all samples using an additive model, utilising GWAMA (Genome Wide Association Meta-Analysis) software package (http://www.well.ox.ac.uk/gwama). A forest plot of the meta-analysis findings was then generated using the Comprehensive Meta-Analysis software version 2.050 (Biostat Inc., NJ, USA). Heterogeneity in the meta-analysis was evaluated by the calculation of Cochran's Q statistic and calculating the I^2% for the meta-analysis [quantifying the effect of heterogeneity on the meta-analysis (Higgins et al. 2003)]. A fixed effect model was used for the meta-analysis as it was assessed that the degree of heterogeneity was low for both variants (Figures 5.2 and 5.3). An expanded meta-analysis was subsequently carried out for the H324Q variant incorporating the previously published report for this variant (t Hart et al. 2005). The GWAMA software package and Comprehensive Meta-Analysis software version 2.050 were again utilised for the meta-analysis and subsequent forest plot generation using a fixed effect model (Figure 5.2).

For statistics involving the geographical distribution of the H324Q variant in the UK population (described below) StatXact v 6.0 (Cytel software corps) was utilised. Power calculations were performed using the Quanto software package (Gauderman 2002). This study, across all samples had 99% power to detect the previously reported OR for the H324Q variant (1.4) for $\alpha = 0.001$ and 90% power to detect an OR of 1.23 at the same α. For the study of rs9825041, this study had 90% power to detect an OR of 1.20 for $\alpha = 0.001$ for MAF =0.04 reported for CEU samples for this SNP on dbSNP [http://www.ncbi.nlm.nih.gov/projects/SNP].

5.3 Results

5.3.1 H324Q variant of *LARS2*

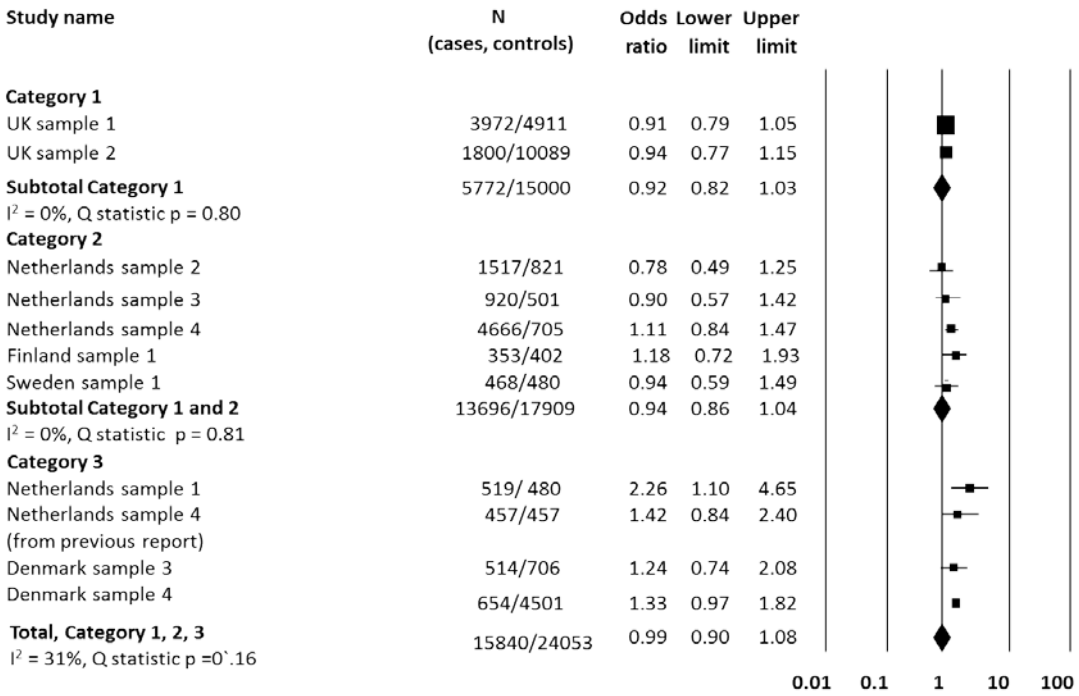
The H324Q variant was successfully genotyped in 3972 T2D cases and 4911 control samples from UK sample 1 and 1800 T2D cases and 10089 controls from UK sample 2 (Table 5.3). There was no evidence of association between the H324Q and increased T2D risk the UK population (OR = 0.92 [0.82-1.03]; p= 0.16). There was also no evidence between the H324Q variant and T2D risk in the meta-analysis incorporating samples from the Netherlands (NL2-4), UK (UK1, 2), Sweden (SE1) and Finland (FL1), OR = 0.94 (0.86-1.04); p= 0.24; (n = 13696 cases, 17909 controls). The incorporation of the previous data from the Netherlands (NL1 and the NL4 samples previously reported) and Denmark (DK3 and Denmark sample 4; DK4) (t Hart et al. 2005) did not alter this finding greatly with an overall OR of 0.99 (0.90 – 1.08), p = 0.87, (n = 15840 cases, 24053 controls) (Figure 5.2).

There was no significant excess of transmission of the risk allele in the Botnia trios (FL2, transmitted / untransmitted = 18 / 14, OR 1.29 (0.64 – 2.59), p = 0.48).

Table 5.3 Association Testing of the H32Q variant in the UK samples

Study sample	N	Cases: Genotype counts (Genotype frequencies)			Controls: Genotype counts (Genotype frequencies)			Statistical Analysis	
	Cases/ Controls	CC	CA	AA	CC	CA	AA	OR (95% CI)	P_{Add}
UK sample 1	3972/ 4911	3636 (91.54%)	326 (8.21%)	10 (0.25%)	4458 (94.79%)	437 (8.90%)	16 (0.33%)	0.91 (0.79-1.05)	0.19
UK sample 2	1800/ 10089	1680 (93.33%)	119 (6.61%)	1 (0.06%)	9383 (93.00%)	693 (6.87%)	13 (0.13%)	0.94 (0.77-1.15)	0.59
UK sample 1 and 2	5772/ 15000	5316 (92.10%)	445 (7.71%)	11 (0.19%)	13841 (92.28%)	1130 (7.53%)	29 (0.19%)	0.92 (0.82-1.03)	0.16

Figure 5.2 Meta-analysis of H324Q variant of *LARS2*



5.3.2. Geographical stratification of the H324Q variant

H324Q shows marked variation in MAF between the various European-descent samples examined (control MAF ranges from 1.9% to 4.8%). In the two large UK control samples, for

example, there was a highly significant ($p = 5 \times 10^{-7}$, using an exact implementation of the Cochran Armitage trend test) difference in allele frequencies between UK1 (recruited exclusively in Scotland) and UK2 (recruited throughout the UK) that raised the possibility that this variant was showing variation in allele frequency along the south-north cline previously described in the WTCCC study and others (Bersaglieri et al. 2004; Coelho et al. 2005; Sabeti et al. 2007; Wellcome Trust Case Control Consortium 2007).To test this, I made use of information on the region of ascertainment available for the UK 1958 Birth Cohort and UK Blood Service and analysed genotype frequencies based on 11 major categories and divided the UK into 4 regions, namely (1) Scotland, (2) Northern England (Yorkshire and the Humber, North East, North West), (3) UK Midlands (East Midlands, West Midlands, Wales, East of England) and (4) South of England (South East, Greater London, and South West). There was some evidence (Figure 5.3) for a North - South gradient across the UK. (MAF 4.66 %, 3.41%, 3.31%, and 3.28% respectively) with p= 0.038 calculated using the Jonckheere-Terpstra Test (StatXact v 6.0; Cytel software corporation). No such MAF gradient was observed in other European samples (Figure 5.4).

Figure 5.3 Geographic stratification in the UK for the H324Q variant (MAF %). The MAF of the H324Q variant based on the region of the ascertainment the samples.

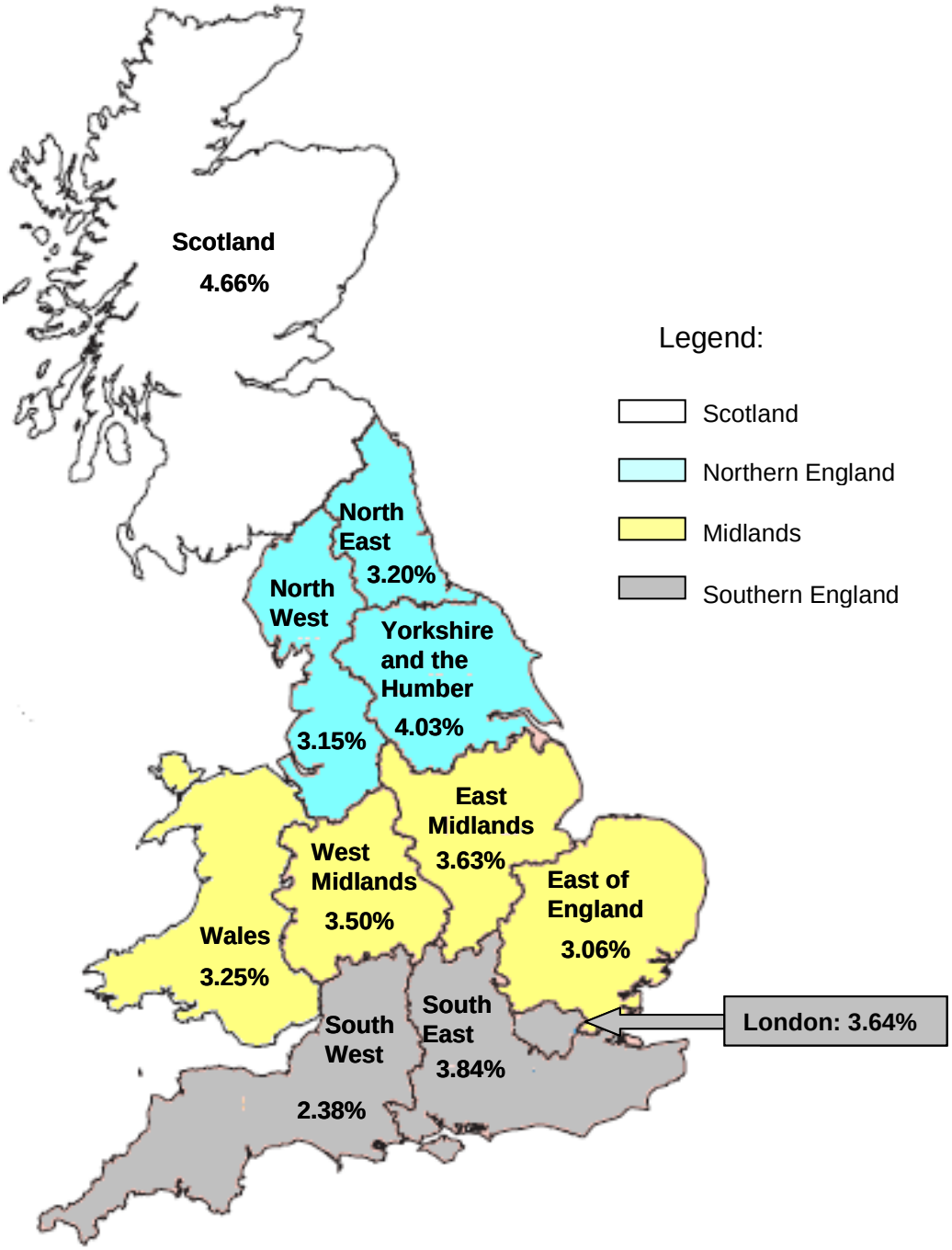
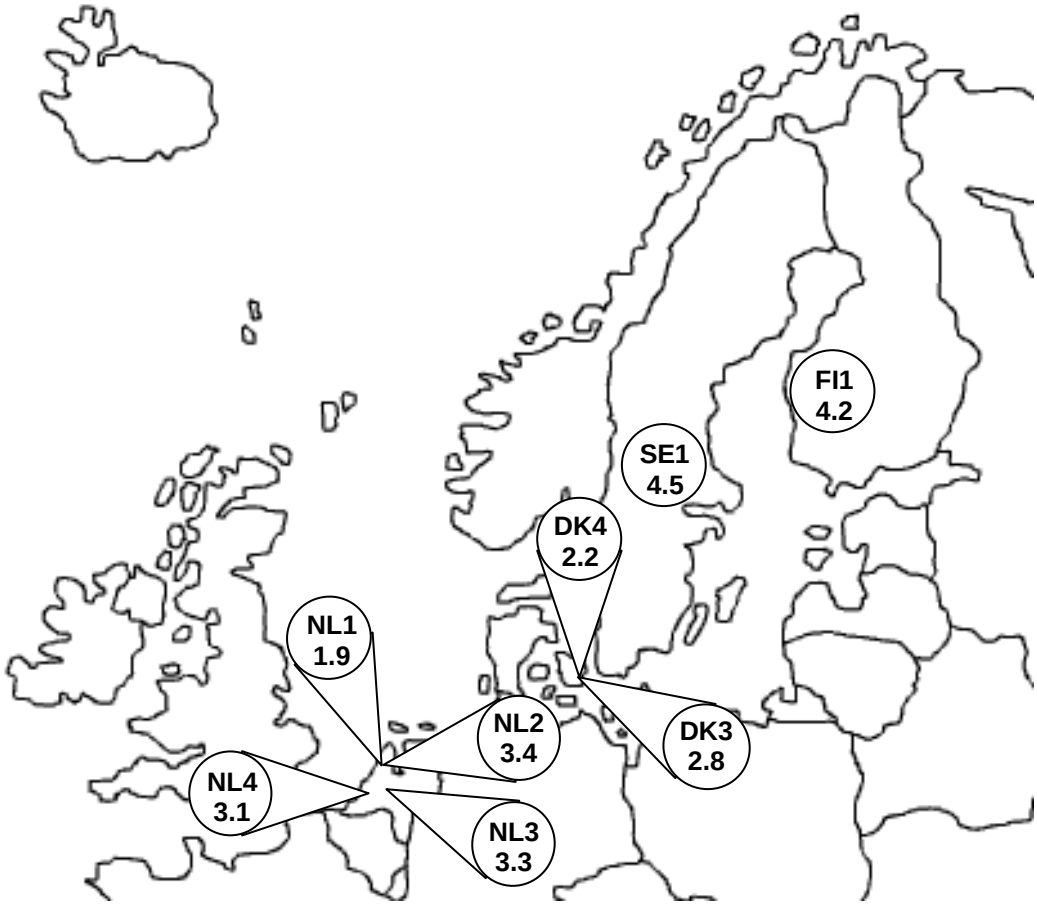


Figure 5.4. Geographic stratification in other European samples (MAF %). The MAF of the H324Q in each of the different regions in Europe (other than UK) that samples in this study were ascertained.



5.3.3 Association testing of rs9825041 with T2D risk

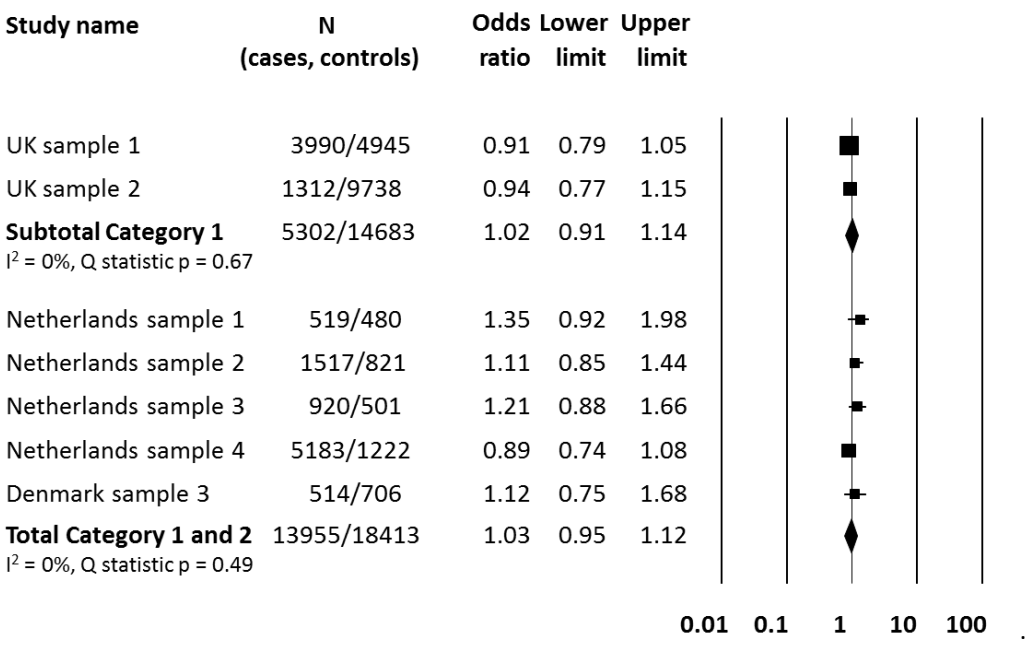
The rs9825041 variant was successfully genotyped in 3990 cases and 4945 controls in UK sample 1 and 1312 cases and 9738 controls in UK sample 2 (Table 5.4). The MAF in cases in UK samples was 5.0% and was 4.9 % in controls. There was no evidence of association with T2D risk in the UK samples in this study with an OR of 1.02 (0.91- 1.14); p = 0.77 (n= 5302 cases, 14683 controls). The meta-analysis incorporating the samples from the Netherlands

(NL1-4) and Denmark did not alter this finding greatly (OR = 1.03 [0.95- 1.12]; p = 0.49). This Meta-analysis included 13955 T2D cases and 18413 controls (Figure 5.5). There was no evidence of appreciable heterogeneity in the meta-analysis (I^2 = 0%, Q statistic p = 0.49).

Table 5.4. Association testing of the rs9825041 variant of *LARS2* in UK samples

Study sample	N	Cases: Genotype counts (Genotype frequencies)			Controls: Genotype counts (Genotype frequencies)			Statistical Analysis	
	Cases/ Controls	GG	GA	AA	GG	GA	AA	OR (95% CI)	P_{Add}
UK sample 1	3990/ 4945	3607 (90.40%)	372 (9.32%)	11 (0.28%)	4471 (91.26%)	459 (8.46%)	15 (0.28%)	1.00 (0.87-1.14)	1.00
UK sample 2	1312/ 9738	1187 (90.47%)	117 (8.92%)	8 (0.61%)	8820 (90.54%)	897 (9.27%)	21 (0.19%)	1.05 (0.88-1.27)	0.57
UK sample 1 and 2	5302/ 14683	4794 (90.42%)	489 (9.22%)	19 (0.36%)	13291 (90.52%)	1356 (9.24%)	36 (0.25%)	1.02 (0.91-1.14)	0.77

Figure 5.5 Meta-analysis of the rs9825041 variant of *LARS2*



5.4 Discussion

The previously observed association of the H324Q (rs71645922) variant in *LARS2* with T2D was not confirmed in this replication study (Table 5.3). Power calculations at the start of the project demonstrated at least 99% power to detect an effect size similar to the initial report for H324Q (rs71645922) (OR 1.4) and at least 80% power to detect an OR of 1.13 for α=0.001. Before excluding the previous association as false, the possibility of heterogeneity was considered. However, no evidence was found that age, age at diagnosis, BMI and gender could account for the previous findings.

Another possibility, raised by the evidence for clinal variation in H324Q MAF across the UK, is that the previous association in Dutch and Danish participants reflected the effects of hidden population structure. However, it seems unlikely that population stratification effects were responsible for the original reports of H324Q associations, as the cases and controls in that study were recruited from the same relatively narrow geographic regions within the Netherlands and Denmark (t Hart et al. 2005). Also in the additional Dutch replication samples there was no evidence of a MAF gradient across the country. However, there are differences in MAF between different countries in Europe (Figures 5.3 and 5.4). The difference in allele frequency for this variant in the UK population along the South- North cline raises the possibility that this finding may be reflecting migration patterns in the UK and therefore H324Q may be a potential migration marker. However, this needs to be demonstrated in other populations.

Several low-frequency SNPs, all in high linkage disequilibrium with each other (r^2>0.95), showed nominal evidence of association with T2D in the DIAGRAM meta-analysis. However, this

association was not confirmed in the large replication samples from the Netherlands, Denmark and the UK (n=31962). It is therefore likely that the nominal p values observed in the DIAGRAM meta-analysis are chance findings.

Three other low-frequency non-synonymous SNPs are present in the *LARS2* locus: K727N (rs36054230), E831D (rs9827689) and E868K (rs34965084). However, on review of dbSNP database, 1000 Genome project findings (http://www.1000genomes.org) and previous sequencing efforts (t Hart et al. 2005), these SNPs are only identified in the African population and not polymorphic in the European population. Therefore, these additional non-synonymous variants were not analysed in this study. As this study did not include a thorough resequencing of the complete *LARS2* locus, it cannot be fully excluded that other, as yet unknown low-frequency variants are present and associated with T2D. Large scale high through put sequencing efforts such as those described in the following chapter should allow for a deeper analysis of the contribution of low frequency variants in *LARS2* to T2D pathogenesis in the future.

This study is one of the largest (up to 35715 participants) replication studies of low-frequency variants in *LARS2* and conclude that the known low frequency variants in this gene do not play a significant role in T2D susceptibility.

CHAPTER 6

Targeted resequencing of exonic regions of genes implicated in Type 2 Diabetes pathogenesis

6.1 Introduction

Over the past five years there have been great advances made in the field of the genetics of type 2 diabetes (T2D) largely through genome wide association studies (GWAS) (Saxena et al. 2007; Scott et al. 2007; Zeggini et al. 2007) and their subsequent meta-analyses (Zeggini et al. 2008; Voight et al. 2010; Cho et al. 2012; Morris et al. 2012). However, it has become apparent that common genetic variants (with a minor allele frequency >5.0%) do not explain the full genetic aetiology of T2D (Lango et al. 2008; Morris et al. 2012) (Chapter 1, section 1.5). It has therefore been the hypothesis of this thesis that low frequency incompletely penetrant (LFIP) variants that would have been largely invisible to GWAS and linkage studies may explain a substantial proportion of the heritability of T2D (Chapter 1, Section 1.2).

This project was designed and completed in 2008-2009 using resequencing in pooled DNA samples (Chapter 2, Section 2.4) with the aim of identifying novel LFIP variants that may appreciably impact on T2D risk. For this project, due to local availability, the Genome Analyzer II (GAII) platform (Illumina Technologies, San Diego, CA, USA; Chapter 2 section 2.10) was utilised for high throughput sequencing.

Though current high throughput resequencing platforms have dramatically reduced the cost of deep resequencing, genome wide resequencing was prohibitively expensive at project inception. At the time of project commencement (2007), one sequencing run on the GAII platform produced ~1 Gigabase (Gb) of data, though by 2012 up to 1 terabase of data can be produced per sequencing run, equivalent to 1000x increase (http://www.illumina.com). The cost of whole genome sequencing in 2007 was approximately $1000000. However this cost has now

decreased dramatically and the current (2012) cost of whole genome sequencing is of the order of $4000 (http://www.illumina.com). This experiment was therefore concentrated on the exonic regions of genes and samples that were expected to increase the likelihood of detecting LFIP variants. To allow for higher throughput, this experiment was carried out in pooled DNA samples. This approach had not been widely reported at project inception.

A logical starting point for the deep resequencing approach is in a subset of genes that have already been implicated in diabetes pathogenesis (Table 6.1) by virtue of:

a) Common, low penetrance variants increasing T2D risk largely discovered through GWA studies (*TCF7L2, SLC30A8, CDKAL1, IGF2BP2, IDE, KCNJ11;* Chapter 1, Section1.1). At project design, the first round of genome wide association studies in T2D had just been published (Saxena et al. 2007; Scott et al. 2007; Zeggini et al. 2007). At that time, nine loci, that were close to or within the genetic boundaries of *TCF7L2*, *SLC30A8*, *CDKAL1*, *IG2BP2, IDE/HHEX, KCNJ11/ABCC8, FTO, CDKN2A/CDKN2B,* and *PPARG* had shown association with T2D risk at a genome wide association level of significance ($p \leq 5x10^{-8}$). Of these twelve genes, six (*TCF7L2*, *SLC30A8*, *CDKAL1*, *IG2BP2, IDE* and *KCNJ11*) were selected for follow up in this study and it was planned that the remaining genes (*FTO, CDKN2A, CDKN2B, HHEX*, *PPARG* and *ABCC8*) would be included in a second round of sequencing efforts.

b) Rare penetrant mutations, contributing to monogenic forms of Diabetes (*HNF1A, HNF4A, HNF1B* and *KCNJ11*). Rare, highly penetrants mutations in *HNF1A, HNF4A and HNF1B* account for 52%, 10% and 6% of cases of maturity onset diabetes of the young

(MODY) respectively (Shields et al. 2010). Rare mutations in *KCNJ11* can result in neonatal diabetes (permanent and transient) as well as the more severe Developmental Delay, Epilepsy and Neonatal Diabetes syndrome (Chapter 1, section 1.4.2). It was hypothesised that these genes may also harbour low frequency variants of intermediate penetrance that may appreciably increase susceptibility to T2D.

c) Genes within replicated regions of T2D-linkage (*LMNA, HNF4A*). Linkage studies had highlighted the regions on Chromosome 1q (Weyer et al. 2001) and 20q (Bowden et al. 1997), where *LMNA* and *HNF4A* reside, respectively, as potentially harbouring susceptibility variants contributing to T2D disease risk. Common variants in these two genes have not explained the Linkage signal identified. Both these genes have also been implicated in monogenic forms of Diabetes(Cao et al. 2000; Ellard et al. 2006) It was therefore envisaged that targeted sequencing of these genes may identify more penetrant, low frequency variants that would account for at least part of the linkage signal previously reported.

Strategy

There were no protocols for sample preparation for targeted sequencing in pooled samples available as a reference point at project commencement and hence I initially designed and completed a pilot project (Section 6.3) before carrying out the main project described in this chapter.

The following genes were selected to be taken forward for the initial pilot study (Section 6.3): *TCF7L2, SLC30A8, LMNA, HNF1A, HNF4A* and *KCNJ11* (85 amplicons, 33Kb). *TCF7L2, SLC30A8, LMNA, HNF1A, HNF4A, KCNJ11, CDKAL1, IGF2BP2, IDE* and *HNF1B* were included in the main high throughput sequencing project (162 amplicons – 60KB). The experimental design has been shown in Figure 6.1.

A study by Nejentsev et al (Nejentsev et al. 2009), published after the commencement of this project, attests to the viability of a similar approach to successfully identify novel LFIP variants. In that study the exonic regions of genes mapping to regions previously implicated in type 1 diabetes (T1D) pathogenesis through common variants in GWAS and genes causative for monogenic syndromes that may include T1D as a constituent element were resequenced in pooled DNA samples using the 454 platform (Margulies et al. 2005). Four novel low frequency variants in the *IFIH1* gene were identified that were associated with T1D independently of each other. For one variant (rs35667974) this association achieved GWA level of significance in subsequent genotyping in a large sample set. This variant is located in exon 14 of *IFIH1*, which is in a region that had previously been shown to have a common variant potentially contributing to T1D disease risk (Smyth et al. 2006).

Figure 6.1. Figure depicting the design of the pilot and main targeted sequencing projects

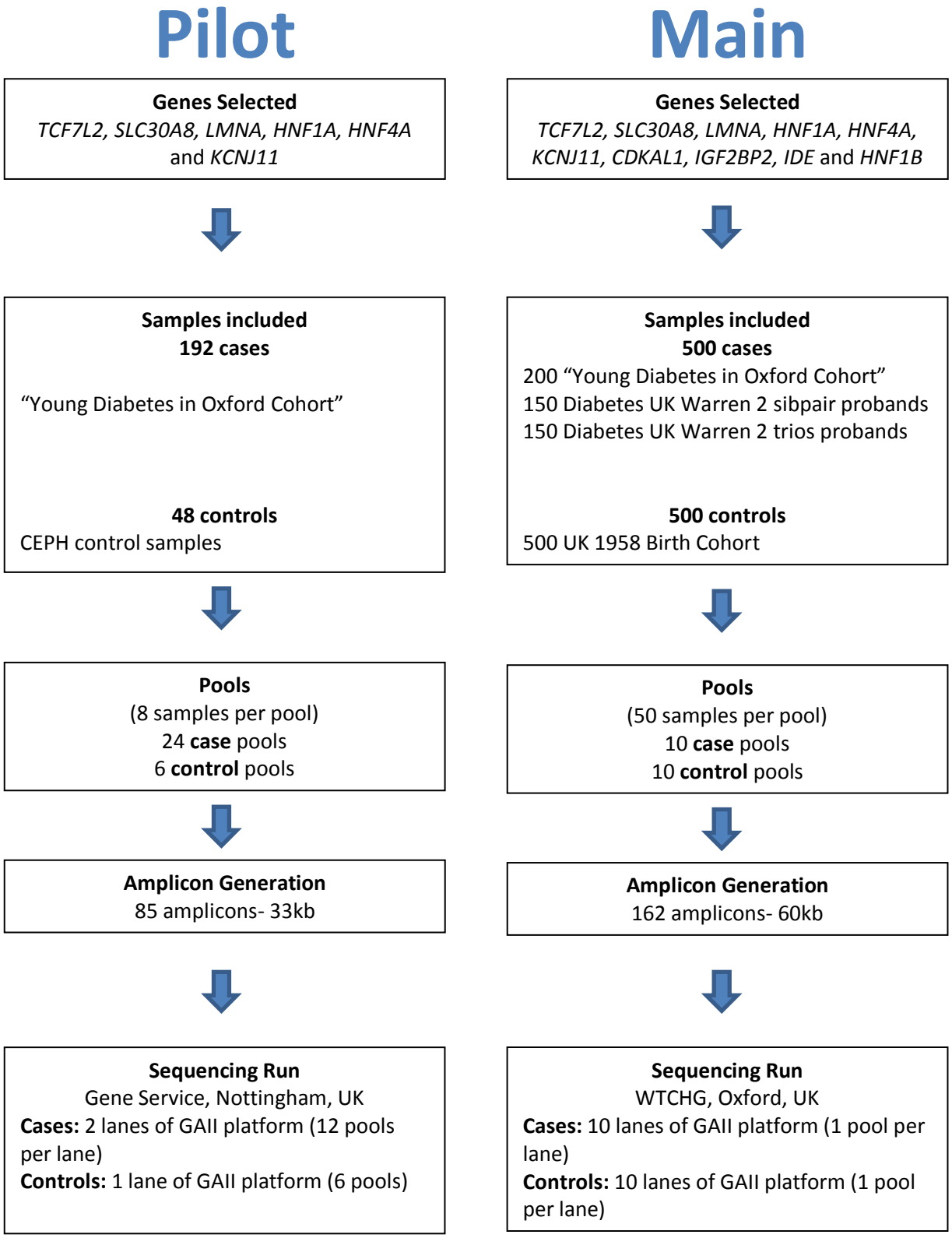


Table 6.1 Genes prioritised for targeted sequencing

Gene symbol	Gene name	Chr	location	No of Amplicons	Reason included	Reference
TCF7L2	Transcription factor 7-like 2	10	10q25.3	17	Common, low penetrance variant (T2D)	(Grant et al. 2006)
SLC30A8	Solute carrier family 30 (zinc transporter), member 8	8	8q24.11	20	Common, low penetrance variant (T2D)	(Saxena et al. 2007; Scott et al. 2007; Zeggini et al. 2007)
CDKAL1	CDK5 regulatory subunit associated protein 1-like 1	6	6p22.3	16	Common, low penetrance variant (T2D)	(Saxena et al. 2007; Scott et al. 2007; Zeggini et al. 2007)
IGF2BP2	Insulin-like growth factor 2 mRNA binding protein 2	3	3q27.2	21	Common, low penetrance variant (T2D)	(Saxena et al. 2007; Scott et al. 2007; Zeggini et al. 2007)
IDE	Insulin-degrading enzyme	10	10q23-q25	25	Common, low penetrance variant (T2D)	(Saxena et al. 2007; Scott et al. 2007; Zeggini et al. 2007)
KCNJ11	Potassium inwardly-rectifying channel, subfamily J, member 11	11	11p15.1	10	Common, low penetrance variant (T2D) Rare, high penetrance variant (monogenic diabetes)	(Gloyn et al. 2003)
HNF1A	Hepatocyte nuclear factor 1, alpha	12	12q24.2	12	Rare, high penetrance variant (monogenic diabetes) Common, low penetrance variant (T2D)	(Frayling et al. 1997; Hegele et al. 1999; Triggs-Raine et al. 2002; Voight et al. 2010)
HNF4A	Hepatocyte nuclear factor 4, alpha	20	20q12-q13.1	15	Rare, high penetrance variant monogenic diabetes/ Linkage (T2D)	(Winckler et al. 2007; Kooner et al. 2011)
HNF1B	Hepatocyte nuclear factor 1-beta	17	17cen-q21.3	10	Rare, high penetrance variant (monogenic diabetes Common, low penetrance variant (T2D))	(Horikawa et al. 1997; Gudmundsson et al. 2007; Winckler et al. 2007)
LMNA	Lamin A/C	1	1q21.2-q21.3	16	Linkage (T2D) Rare, high penetrance variant (monogenic diabetes)	(Shackleton et al. 2000; Weyer et al. 2001)

6.2 Samples included

In the search for the susceptibility variants that appreciably impact on multifactorial disease risk, previous reports have indicated that the study of individuals at the extremes of phenotype (such as younger age of onset) may identify more "loss of function" variants that would be more likely to impact on disease pathogenesis (Romeo et al. 2007). Therefore, in this study, to improve the likelihood of detecting LFIP variants that influence risk of T2D, individuals with early-onset, familial non-autoimmune diabetes were selected (Chapter 2, Section 2.1). These samples have been described in detail in Chapter 2, but a brief description is included below.

T2D samples

T2D cases were drawn from three samples:

1) The "Young Diabetes in Oxford Cohort" ("UK3", Chapter 2, Section 2.1.1) includes subjects of white British ethnicity diagnosed with T2D between 18 and 45 years of age, recruited from Oxfordshire GP practices (200 selected for this study from a total of 300).

2) Probands from the Diabetes UK Warren 2 sibpair collection (Wiltshire et al. 2001) (a subset of "UK sample 2"; Chapter 2, Section 2.1.1). One hundred and fifty samples were selected randomly for this study from a total sample number of 573. All individuals in this group were of white British ancestry and had at least one affected sibling diagnosed with diabetes. These individuals were included due to their family history of T2D as they should be enriched for LFIP variants (Chapter 1, Section 1.6). This sample has also been included in several genotyping projects (www-gene.cimr.cam.ac.uk/todd) and this information was used for quality control measures.

3) Probands from the Diabetes UK Warren 2 trios (Frayling et al. 2003) (parent-offspring triad, a subset of "UK sample 2", Chapter 2, Section 2.1.1) that had either both parents or one parent and at least two siblings available (150 selected randomly from a total of 350). All individuals were of British white ancestry. These samples were also used due to their family history of T2D and inclusion in previous genotyping projects.

Control samples

The control samples for the pilot study and the main targeted sequencing project were drawn from two separate samples.

For the pilot study (described in section 6.3), a subsection of the Centre d'Étude du Polymorphisme Humain (CEPH) (Utah residents with northern and western European ancestry [CEU]), DNA collection was used (n=48). Genotyping data from the HapMap project (International HapMap Consortium 2005) that incorporated these samples were used as a quality control measure in the pilot study.

The control samples for the main targeted sequencing project were taken from a random selection (n=500) of the UK 1958 Birth Cohort. The recruitment criteria for these collections are described in Chapter 2 (section 2.2). All individuals selected had been included as part of the Wellcome Trust Case Control Consortium (WTCCC) genome wide association scan (Wellcome Trust Case Control Consortium 2007).

The strategy in this project was primarily targeted at discovery of novel SNPs that may be functionally active and therefore impact on T2D pathogenesis and was designed prior to the availability of data from the 1000 Genome project (http://www.1000genomes.org/). However,

utilising 500 cases and 500 control samples also allowed for a comparison of minor allele frequency (MAF) differences in cases and controls and was used as an aid in identifying potential susceptibility variants.

6.3 Pilot targeted sequencing project

To inform and optimise protocols for the main study, a pilot study was undertaken. This study was concentrated on exon resequencing of 6 genes already implicated in T2D pathogenesis (*TCF7L2, SLC30A8, LMNA, HNF1A, HNF4A, and KCNJ11*). This encompassed 85 amplicons with an average size of 374bp. The study design involved resequencing 192 cases (from the "Young Diabetes in Oxford Cohort" described in section 6.2) and 48 CEPH control samples (section 6.2). The genotype frequency data for the CEPH sample, obtained as part of the HapMap project are publicly available (International HapMap Consortium 2005). These data were used as a quality control measure in this project to compare the genotype frequencies in the same 48 individuals in the current study and those obtained in the HapMap project.

The DNA samples were aggregated in pools of 8 individuals (24 case pools/ 6 control pools) as outlined in Chapter 2 (section 2.4) and Figure 6.1. Following amplification of all pools for 85 amplicons, two case PCR pools were generated by aggregating all amplicons generated in 12 case pools together and generated one control PCR aggregate by combining all amplicons generated in the control pools together.

Following clean-up of these pools, concatenation of the amplicons and subsequent shearing of the concatamers was carried out prior to the run on the GAII platform (Chapter 2, Section 2.2.8).

The reference genome used for sequence alignment was the Human Build 36.1 finished human genome assembly (hg18, Mar. 2006). The images from the Genome Analyzer (GA) were analysed with the GA pipeline software (version 1.0, Illumina software) to undertake image analysis, base calling and sequence alignment to the reference genome and was carried out by Gene Service, Nottingham, UK (Chapter 2, Section 2.2.8).

6.3.1 Sequence alignment and SNP calling

The output from the ELAND extended algorithm was then used in the MAQ software package (http://maq.sourceforge.net/maq-man.shtml) to align each of the 3 lanes against a 33kb reference made up of the regions targeted (85 Amplicons) and to identify SNPs. The Syzygy software package (Chapter 2, Section 2.2.8.1) has been designed for SNP identification in pooled DNA samples and was utilised for more accurate SNP calling. This was carried out by Manuel Rivas, the Broad Institute, Cambridge, Massachusetts, USA, who had designed the software package. The minimum depth of coverage to call a SNP was at least 3. The minimum phred like consensus quality for SNPs was at least 30. Phred (Ewing et al. 1998; Ewing et al. 1998) is a base-calling program for DNA sequence traces.

6.3.2 Results of pilot project

The total sequence yield from all 3 lanes of the GAII run was ~0.8Gb and read lengths of 36bp were achieved (both in line with expectation in December 2008 when the experiment was run). Approximately 70% of the reads obtained in the GAII run aligned to the reference genome (Human Build 36.1 finished human genome assembly [hg18, Mar. 2006]) and the alignments were then assembled into a consensus for SNP calling.

Using the thresholds for SNP calling noted above (Section 6.3.1), a total of 185 SNPs were called with high confidence, of which 19 were non-synonymous, 10 of which were novel at the time (Table 6.2). On comparison of the results with those of the 1000 Genome project (February 2012 release), seven of these variants remain novel. A further 20 synonymous variants were detected, 6 of which were novel at the time (Table 6.3). On comparison of the results with those of the 1000 Genome project (February 2012), 3 of these variants remained novel.

Table 6.2 Non-synonymous variants identified in pilot targeted sequencing project

Chr	Position (HG18)	Gene	Substitution	AA Change	SNP	Status	Control MAF (n=48)	Case MAF (n=192)
10	114915359	*TCF7L2*	C>**A**	Pro477Thr	rs77961654	Previously reported	0.026	0.042
10	114915396	*TCF7L2*	C>**G**	Pro489Arg	rs77673441	Previously reported	0.000	0.014
10	114915618	*TCF7L2*	C>**G**	Ser563Cys	N/A	**Not reported**	0.000	0.003
10	114915665	*TCF7L2*	T>**C**	Ser579Pro	1KG_10_114925675	Previously reported	0.024	0.003
11	17365061	*KCNJ11*	C>**G**	Ser385Cys	rs41282930	Previously reported	0.000	0.003
11	17365206	*KCNJ11*	G>**A**	Val337Ile	rs5215	Previously reported	0.362	0.322
11	17365407	*KCNJ11*	C>**G**	Leu270Val	rs1800467	Previously reported	0.029	0.055
11	17365730	*KCNJ11*	T>**C**	Ile162Thr	N/A	**Not reported**	0.010	0.003
11	17366148	*KCNJ11*	A>**G**	Lys23Glu	rs5219	Previously reported	0.419	0.334
12	119901033	*HNF1A*	A>**C**	Ile27Leu	rs1169288	Previously reported	0.231	0.262
12	119901096	*HNF1A*	G>**A**	Glu48Lys	N/A	**Not reported**	0.000	0.003
12	119901247	*HNF1A*	C>**T**	Ala98Val	rs1800574	Previously reported	0.056	0.027
12	119915736	*HNF1A*	T>**C**	Ile186Thr	N/A	**Not reported**	0.000	0.005
12	119919810	*HNF1A*	G>**A**	Ser487Asn	rs2464196	Previously reported	0.237	0.258
1	154374094	*LMNA*	G>**A**	Arg545His	N/A	**Not reported**	0.000	0.003
20	42475778	*HNF4A*	C>T	Thr130Ile	rs1800961	Previously reported	0.003	0.003
20	42475781	*HNF4A*	G>**A**	Arg140Gln	N/A	**Not reported**	0.000	0.003
20	42486372	*HNF4A*	C>**T**	Thr398Ile	N/A	**Not reported**	0.000	0.006
8	118253964	*SLC30A8*	C>**T**	Arg325Trp	rs13266634	Previously reported	0.178	0.217
8	118253965	*SLC30A8*	G>**A**	Arg325Gln	rs16889462	Previously reported	0.000	0.009

Chr: Chromosome number; **AA:** Amino Acid; **MAF**: Minor Allele Frequency; **N/A**: not available
Control: Ceph samples (48 individuals); **Cases**: T2D cases from Young onset in Oxford Cohort (192 individuals)

Table 6.3 Synonymous variants identified in pilot targeted sequencing project

Chr	Position (HG18)	Gene	Substitution	AA Change	SNP	Status	Control MAF (n=48)	Case MAF (n=192)
1	154367032	*LMNA*	C>**T**	Arg119Arg	rs41313880	Previously reported	0.000	0.003
1	154367113	*LMNA*	C>**T**	Ala146Ala	rs80356805	Previously reported	0.042	0.005
1	154370916	*LMNA*	G>**A**	Leu204Leu	rs12117552	Previously reported	0.000	0.003
1	154371652	*LMNA*	T>**C**	Ala287Ala	rs538089	Previously reported	0.069	0.123
1	154372809	*LMNA*	T>**C**	Asp446Asp	rs505058	Previously reported	0.136	0.145
1	154374158	*LMNA*	C>**T**	His566His	rs4641	Previously reported	0.207	0.233
10	114789863	*TCF7L2*	G>**A**	Pro157Pro	N/A	**Not Reported**	0.031	0.003
10	114915397	*TCF7L2*	C>**G**	Pro489Pro	1KG_10_114925407	Previously reported	0.000	0.008
10	114915700	*TCF7L2*	C>**T**	Leu590Leu	N/A	**Not Reported**	0.010	0.000
11	17365072	*KCNJ11*	G>**A**	Lys381Lys	rs8175351	Previously reported	0.000	0.011
11	17365126	*KCNJ11*	A>**G**	Ser363Ser	rs5214	Previously reported	0.000	0.006
11	17365414	*KCNJ11*	C>**G**	Leu267Leu	rs5216	Previously reported	0.030	0.022
11	17365645	*KCNJ11*	C>**T**	Ala190Ala	rs5218	Previously reported	0.226	0.312
11	17366107	*KCNJ11*	G>**A**	Val36Val	rs112070496	Previously reported	0.000	0.005
12	119901005	*HNF1A*	C>**G**	Leu17Leu	rs1169289	Previously reported	0.399	0.405
12	119916500	*HNF1A*	G>**C**	Gly288Gly	rs56348580	Previously reported	0.126	0.214
12	119919725	*HNF1A*	C>**T**	Leu459Leu	rs2259820	Previously reported	0.182	0.234
12	119919796	*HNF1A*	C>**T**	Ser482Ser	N/A	**Not Reported**	0.000	0.003
12	119921497	*HNF1A*	G>**A**	Thr515Thr	rs55834942	Previously reported	0.092	0.119
20	42490462	*HNF4A*	C>**T**	Asn401Asn	rs61737145	Previously reported	0.000	0.003

Chr: Chromosome number; AA: Amino Acid
MAF: Minor Allele Frequency
N/A: not available
Control: Ceph samples (48 individuals); Cases**:** T2D cases from Young onset in Oxford Cohort (192 individuals)

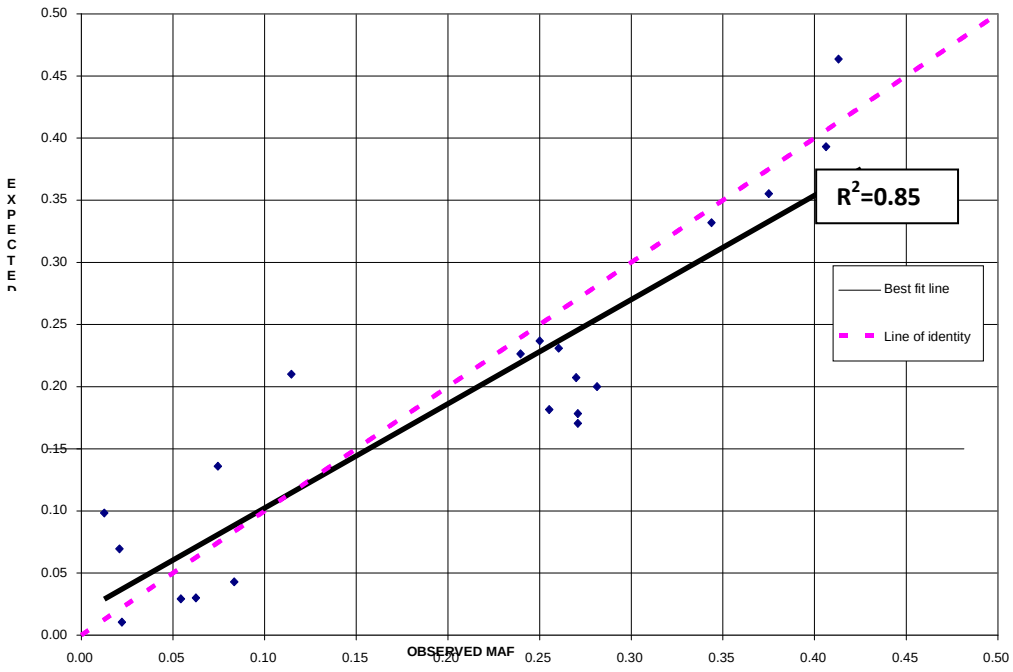
6.3.3 Validation

Thirty three SNPs that are not monomorphic in European samples (based on dbSNP in 2009) were in the regions covered by this experiment. All these SNPs were identified in this experiment. As a further quality control step, the MAF of the SNPs genotyped as part of the HAPMAP project in the same CEPH individuals was compared to the MAF obtained in this study. All 22 SNPs genotyped as part of the CEU HAPMAP project in the same 48 individuals were successfully identified and the concordance rate was calculated (R^2 = 0.85, Figure 6.2). The exact reason for this

underrepresentation of the MAFs is not known, though it was felt to be a reflection of error produced in this study rather than genotyping error in the HapMap data. Possible sequencing or analytical errors in this study were considered as potential causes. This may also be a reflection of the pool depth used in this study and the pooling strategy was therefore simplified in the next phase of the project.

As a further quality check, the ratio of the number of transitions (the substitution of a purine by a purine, or a pyrimidine by a pyrimidine) to the number of transversions (a change from a purine to a pyrimidine, or vice versa) for a pair of sequences was calculated. In regions with high quality reads this ratio was 3.32:1. This value is similar to that reported in other targeted sequencing projects in exomes [3.47-3.82(Marth et al. 2011)].

Figure 6.2 Comparison of the MAF of the 22 SNPs captured in this study (X axis) in the same 48 CEPH individuals that were also included in the HapMap project (Y-axis)



6.4 Protocol Development

As noted earlier, at project commencement, protocols for sample preparation and data analysis for the GAII platform were still in development. It was therefore necessary to overcome three key elements during the course of this project.

1. Ligation of PCR products

For the preparation of the samples for the GA platform, the DNA fragments need to be sheared to fragments of ~800bp (Chapter 2, Section 2.2.7). If the segment of DNA is much smaller than this to begin with the shearing step will result in a very poor sequencing yield. Concatenation of the PCR products was therefore necessary.

For the pilot study, this step was carried out by Gene Service ltd, Nottingham, UK. The protocol used by Gene Service ltd was proprietary and could not be shared. For the main targeted sequencing project, blunt ended ligation was attempted. However as the PCR products were generated using a Taq Gold Polymerase, "A" overhangs were produced. I attempted to blunt these ends by the incubation of the PCR products with T4 DNA Polymerase (Promega) prior to the ligation step but this approach yielded unreliably ligated products. I subsequently used Taq Phusion high fidelity DNA polymerase that did not leave "A" overhangs, but the ligation step again yielded unreliably ligated products. Subsequently a protocol using *Not1* digestion of the primer ends followed by ligation was obtained from Illumina Technologies (Mark Ross). This method produced unreliably ligated products and resulted in large amounts of sample loss. It was therefore deemed best to abandon this approach.

Fortunately, at this time a novel protocol was published in Nature Methods (Druley et al. 2009) which utilised blunt ended ligation (Chapter 2, Section 2.2.7). This protocol incorporated T4

ligase, T4 Polynucleotide kinase, Polyethyleneglycol as well as ligase buffer. I was able to use this protocol successfully to generate ligated product in excess of 10kb (confirmed by gel electrophoresis) which subsequently underwent a shearing step prior to the sequencing run.

2. Clean-up of ligated PCR products

To ensure the removal of unincorporated dNTPs, primers, salts or un-ligated products that may interfere with downstream sequencing, clean-up of the PCR products is required. Clean up columns can be used for the purification of products between 100bp-10kb (www.qiagen.com/Products/DnaCleanup/GelPcrSiCleanupSystems/QIAquickPCRPurificationKit). As the ligated products generated were >10Kb, it was not possible to use this method. It was therefore decided to use Agencourt AmPure clean up beads (New England Biolabs UK Ltd). The beads used solid phase paramagnetic bead technology (SPRI). Magnetic particles were used in an optimised buffer to bind amplicons 100 bp or larger. They were able to bind to fragments >10kb and hence were ideal for the clean-up of ligated products generated by this project. Initially sample loss was >70%, however with repeated washes of the beads >50% of each sample was recovered. This was enough for a run on GAII.

3. Software for analysis of GAII output

In excess of 1 terabyte of data were generated by the GAII platform. At the time that this study was conducted (2008-2009) there were no software programmes specifically designed to analyse sequencing data from pooled DNA samples. The raw sequencing data were first aligned to a reference genome using Eland (Illumina technologies) and MAQ [mapping and assembly with qualities (Li et al. 2008)] (http://maq.sourceforge.net/) software packages by the core bioinformatics group in the Wellcome Trust Centre for Human Genetics, Oxford. Both these packages were able to align short reads, such as those generated by GAII and were the main

software packages available for sequencing alignment at the time of running this experiment (2008-2009). These methods were designed for use with single samples (where SNPs would conform to three classes of genotype) but perform poorly at assessing SNPs in pools. The data generated from my samples proved to be complex with >1000 depth per base achieved (indeed depths up to 10000 per base achieved in certain areas). MAQ software for example was not able to accurately assess any region with >255X depth per base and in my data many SNPs were called when none existed. On inspection of the pileup files generated by the software, it was noted that the several SNPs that would be possible to call manually had been missed and SNPs had been called incorrectly in areas with low coverage. It was therefore decided not to use the SNP calling function of the MAQ software function as it had not been used for SNP calling in pooled samples previously.

Fortunately, my supervisors' collaborators in the BROAD institute (Manuel A. Rivas and Mark J. Daly) had developed software (Syzygy, http://broadinstitute.org/software/syzygy/) that was specifically designed for the analysis of variant discovery in pooled DNA samples and was made available for use in this project. The Syzygy software package incorporates error modelling per base by utilising the base context and coverage variability (Calvo et al. 2010). Evaluation of the error rate in pooled sequencing experiments is vital to differentiate sequencing error from true variants. The standard error rate of high throughput sequencing platforms at the time of performing this experiment was of the order of 1-5 % (Calvo et al. 2010). Error rates of this magnitude make the detection of variants in singleton cases in pools of 100 chromosomes (50 samples), such as that used in this experiment, difficult (Calvo et al. 2010) and specific software aimed at variant detection in pooled DNA samples is required.

The Syzygy software also provides allele frequency estimation, singleton and rare variant detection as well as single marker and group marker test for rare variants [C-alpha statistic

(Neale et al. 2011)]. Syzygy also provides the power of detection of a singleton variant at a particular base. Two tests of strand consistency are utilised in this software package to state "high confidence" in the identified polymorphisms. More details of this software package have been provided in Chapter 2.

6.5 Main targeted sequencing project

Following the successful completion of the pilot study, I proceeded to the main resequencing project. This represented an up-scaling of the pilot project to include 500 T2D samples and 500 control samples and to include ten genes implicated in T2D pathogenesis previously (Section 6.1). The total extent of this section of the project was ~60Kb.

6.5.1 DNA sample preparation

To allow a more efficient use of the GAII capacity, the DNA samples included were pooled in groups of 50 (10 case and 10 control pools). To ensure that each individual in the pool was proportionally represented, very accurate quantification of the DNA samples using a PicoGreen assay was carried out (Chapter 2, Section 2.2.1). The 50 samples were then aggregated in equimolar fashion in pools of 10ng/µl (Figure 6.3 and Chapter 2, Section 2.2.2).

6.5.2 Sample preparation for the Genome Analyser II platform

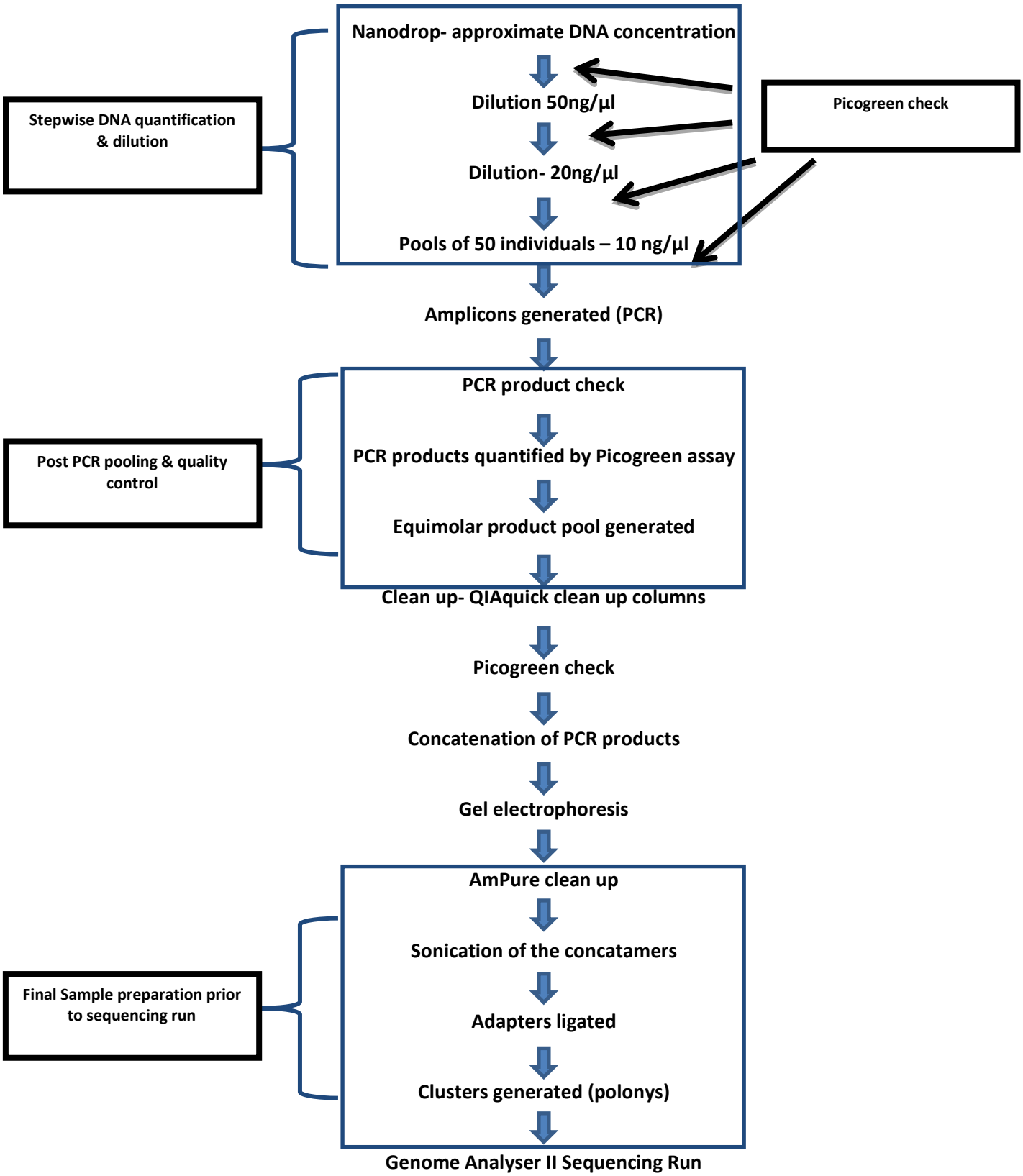
Following the accurate quantification of the DNA samples, PCR amplification of the regions of interest was carried out (Chapter 2, section 2.2.3). This equated to 162 amplicons for each pool. The primers used for the PCR amplification have been included in the Appendices (Supplementary Table S6.1). The PCR conditions have been included in Chapter 2, section 2.2.3.

The success of the PCR reaction was assessed by gel electrophoresis (Chapter 2, section 2.2.4). The PCR products were then quantified by PicoGreen assay (Chapter 2, section 2.2.1). Based on this quantification, the PCR products were aggregated for each pool ("PCR pool"), in an equimolar fashion. The PCR pools were cleaned up using QIAquick clean up columns (QIAGEN Ltd, West Sussex, UK).

Following this, blunt-ended ligation (Chapter 2, section 2.2.7; Chapter 6, section 6.4.) of the samples was carried out and the success of this step was assessed by gel electrophoresis. Between 3.2 and 4.0 µg of product was generated and was taken forward to clean up prior to being run on the GAII platform.

Following a shearing step (Chapter 2, Section 2.10), the samples subsequently underwent sequencing on the GAII platform, utilising sequencing by synthesis at the Wellcome Trust Centre for Human Genetics (WTCHG). Single end reads were targeted for this experiment and 50bp read lengths were achieved which was comparable to that typical at the time (2009). Sample preparation and the workings of this system are explained in more detail in Chapter 2 (section 2.2.8), but are summarised in Figure 6.3.

Figure 6.3 Sample preparation workflow for the targeted sequencing project



6.5.3 Analysis and evaluation

Analysis

The data generated by the GAII run were primarily analysed using the newly developed Syzygy (http://www.broadinstitute.org/software/syzygy/) software package (release 0.9) developed by Manuel A. Rivas and Mark J. Daly (Calvo et al. 2010; Rivas et al. 2011). The raw sequencing data were first aligned to a reference genome using Eland (Illumina technologies) and MAQ [mapping and assembly with qualities (Li et al. 2008)] (http://maq.sourceforge.net/) software packages as (Section 6.4) by the core bioinformatics group in the Wellcome Trust Centre for Human Genetics, Oxford.

The Syzygy software was utilised by Manuel Rivas in the Broad Institute for SNP calling in this experiment as high read depths in pooled samples were achieved. The output from the MAQ alignment was converted to SAM (Sequence Alignment/Map) format prior to being used in the Syzygy package and consequences (differences between the reference genome and the sequences obtained from GAII) called. The Syzygy software package was also able to estimate allele frequencies. I used these frequencies both as a quality control step (where genotyping information was available) and an estimate of association by comparison of the allele frequencies in cases vs. controls.

A series of heuristics were used to "vet" the low-MAF variants identified by this project, and to identify those of greatest potential interest. Variants deserving further study featured:

1. High-confidence base-call based on strand consistency (Section 6.4.)
2. implied functional relevance (e.g. coding change, alteration of canonical sequences). To evaluate whether an amino acid change was likely to affect

protein function, I utilised SIFT (Sorts Intolerant From Tolerant) prediction software (http://sift.jcvi.org/) and PolyPhen (Polymorphism Phenotyping) software package (http://genetics.bwh.harvard.edu/pph/) (Section 6.5.3.1; Table 6.6).

As intron-exon boundaries had been included in this experiment I utilised the predictive software packages, GeneSplicer (http://www.cbcb.umd.edu/software/GeneSplicer/) (Pertea et al. 2001) and NetGene2 (http://www.cbs.dtu.dk/services/NetGene2/) to evaluate whether splice sites may be affected (Section 6.5.3.1; Table 6.8) . The predictive software (Methylator, http://bio.dfci.harvard.edu/Methylator/) was used to assess whether methylation sites were likely to be affected by the variants discovered (Section 6.5.3.1; Table 6.7).

In addition to these "SNP-level" criteria, the C-Alpha statistic (Neale et al. 2011) was utilised to evaluate the distribution of rare alleles in cases vs. controls. Although there are now several different approaches available in evaluating the distribution of rare variants [such as CCRaVAT and QuTie (Lawrence et al. 2010)], the C-alpha statistic was the only approach available at the time (2009) and was used in this study as it is incorporated into the Syzygy software package. It is expected that risk alleles will be seen more in cases than in control samples (and protective alleles seen in more control samples) and are more likely to lead to "loss of function" of the gene (MacArthur et al. 2012). Loss of function variants have traditionally been regarded as being rare (MacArthur et al. 2012). The C-alpha statistic compares the burden of rare variants in cases vs. controls. Therefore, highlighting regions that harbour a large proportion of rare variants in cases or in controls by use of this statistic

would be expected to increase the likelihood of identifying variants appreciably increasing T2D risk (Neale et al. 2011).

6.5.4 Quality control

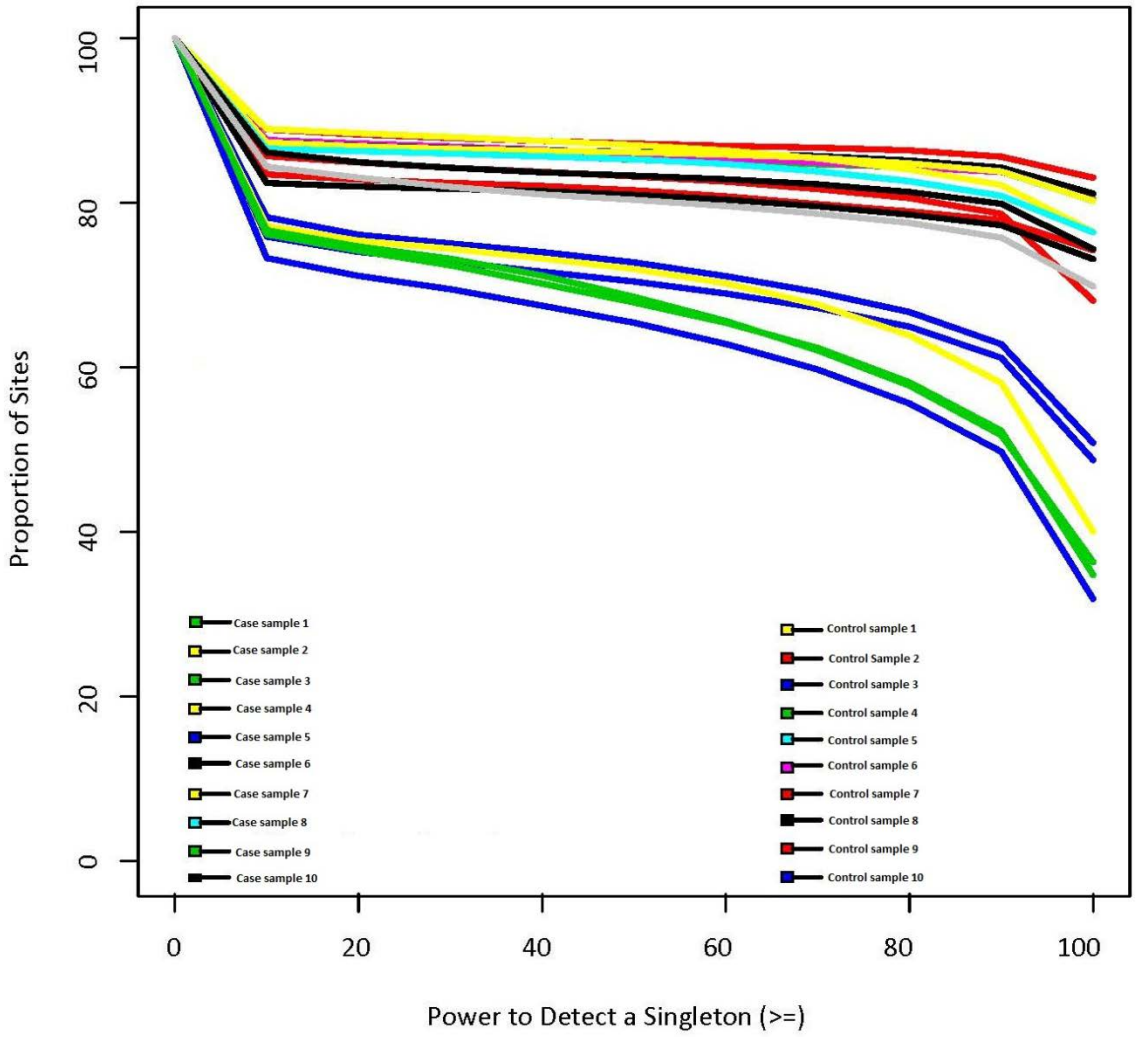
False negative rate

An estimate of the false negative rate for this study was calculated based on variants for which genotype frequencies in European samples were available (Hapmap2). A false negative rate of 6.4% was calculated based on the known variants that were not detected in this experiment (Table 6.4). This was calculated using the formula: [(number of variants missed)/ (total number of SNPs previously reported in the region covered by this experiment)] x100. Of note the only variants not detected in this study had an MAF <0.9% and no variants with an MAF >1.0% were missed. The power to detect a singleton at a locus was also evaluated (Figure 6.4) by use of the Syzygy software and was in excess of 60% in approximately 75% of all targeted regions (comparable to other studies performed at a similar time (Rivas et al. 2011).

Table 6.4 Table depicting the minor allele frequency bracket of variants previously reported in European samples and the number of samples called and missed in this experiment. A nominal false negative rate was calculated

MAF Bracket	SNPs called correctly	Missed SNPs	False negative rate (%)
>5%	43	0	0.00
>1, <5%	13	0	0.00
0.05-0.9%	6	3	33.33
<0.05%	11	2	15.38
Total	73	5	6.41

Figure 6.4 Figure depicting the power to detect a singleton across the 20 lanes of the GAII run. Power to detect a singleton was ≥ 60% in approximately 75% of all targeted regions. Each line represents a separate pool (and hence lane used on the GAII platform).



False positive rate

As part of the follow on studies generated from this project an estimated measure of the false positive rate (FPR) for this experiment has been attempted. Sanger sequencing has been used to validate the variants detected in singleton cases (Table 6.9) in this study by Ms Neelam Hassanali (DPhil Student with Professor Mark McCarthy and Dr Anna Gloyn). This work is on-going, though thus far 8 out of 10 singleton variants in the same DNA samples have been validated by Sanger sequencing (Table 6.9).

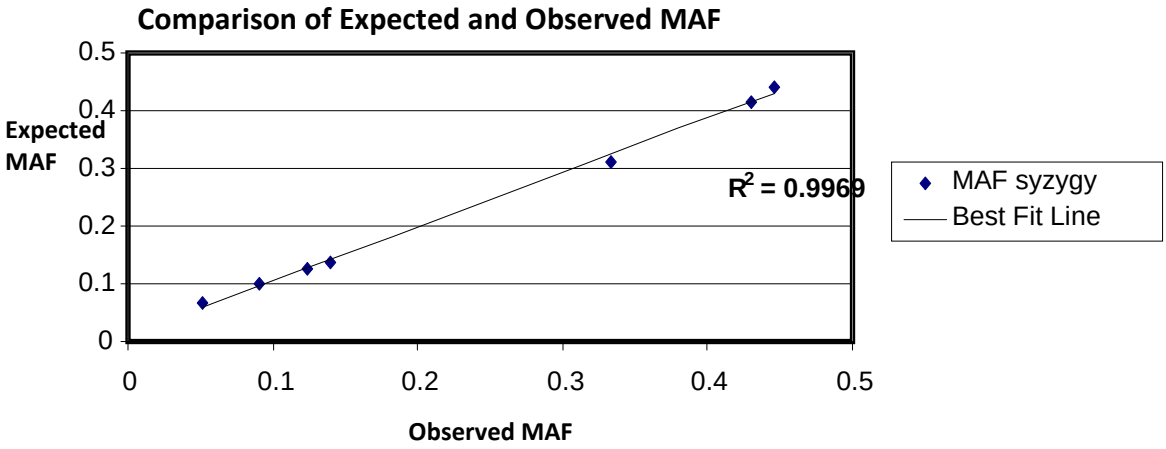
Transition: Transversion ratio

As a further quality check, the ratio of the number of transitions to the number of transversions for a pair of sequences was calculated. In regions with high quality reads this ratio was 2.5:1 which is comparable to that expected in coding regions as reported in some reports (2-2.5 (Halushka et al. 1999; Rivas et al. 2011)) but is lower than that achieved in the pilot study and in some targeted exonic sequencing studies (Marth et al. 2011).This may be reflection of differing amounts of noncoding regions being mapped and called in the main targeted sequencing experiment.

Tests of concordance

The control samples in this study had all been included as part of the WTCCC genotyping efforts. Seven SNPs (with MAF range between 5-31%) included as part of the WTCCC studies were covered by the regions included in this project (rs2819999, rs2820000, rs10946406, rs2820001, rs10282940, rs5215, rs6130615). The MAF obtained in pools to those from the WTCCC genotyping efforts for the same individual samples was compared as a further quality control measure (Figure 6.5). These results were highly comparable (r^2=0.9969) and exceeded the level of correlation reported in previous published works (Calvo et al. 2010) comparing the MAF of variants detected through high throughput sequencing to that achieved through genotyping.

Figure 6.5 Comparison of the MAF of the seven SNPs captured in this study (X axis) in the same individuals that were also included in the WTCCC study (Y-axis)



6.5.5 Results of the main high throughput resequencing project

In total, 20 lanes of the GAII (one per pool) were utilised for the sequencing runs and 50 bp single end reads were attained. Overall, 71.7% of the reads aligned to the reference sequence. An average per individual depth of coverage of 55X and average per base coverage of 2732X at pool level was achieved. These parameters compared favourably with other projects of a similar nature designed at a similar time (Gunnarsdottir et al. 2011).

A large number of variants were identified by this project. In total, 2913 variants of which 243 were non-synonymous (amino acid changing), 116 were synonymous (not amino acid changing), 765 were in the 3' UTR (un-translated region) and 1789 were intronic (Table 6.5) were identified in this study. The variants identified were subsequently further categorised based on the quality of the read and the overall power to detect a singleton at the locus (aimed for greater than 90% power). This reduced the number of variants to be taken forward for further analysis to 387 (all included in the appendices, Table S6.2). A total of 86 of these variants had previously been identified (dbSNP; v132). A further 50 of these variants have been identified subsequent to the completion of this project as part of the 1000

Genome project (February 2012). A total of 251 variants, with high quality reads, were therefore considered novel (Table S6.2).

Table 6.5 Overall summary of variants identified in the high throughput sequencing run and the C-Alpha statistic and C-Alpha P-Value for each gene (comparing the burden of rare variants in cases vs. controls) included in the main targeted sequencing project**.**

Gene	UTRNC	Syn	NS	Intronic	TOTAL	C-Alpha Statistic	C-Alpha P value
LMNA	85	14	30	190	319	3.499	**2.33×10^{-4}**
SLC30A8	154	3	17	61	235	18.34	**$<1.0 \times 10^{-16}$**
TCF7L2	76	13	31	141	261	21.53	**$<1.0 \times 10^{-16}$**
IGF2BP2	69	3	9	285	366	0.321	0.374
HNF1A	65	14	37	145	261	3.018	**0.001**
HNF4A	107	10	24	190	331	7.484	**3.62×10^{-14}**
HNF1B	38	1	24	161	224	0.475	0.317
IDE	36	7	18	326	387	-0.360	0.641
KCNJ11	107	13	43	9	172	-0.023	0.509
CDKAL1	28	9	10	271	318	3.621	**1.47×10^{-4}**
TOTAL	765	87	243	1779	2874	-	-

UTRNC: Untranslated region- non-coding
Syn: Synonymous
NS: Non-synonymous
Bold: C-Alpha P value <0.05

To further delineate regions that may be of interest for follow up, the C-Alpha statistic (Neale et al. 2011) was used. This test was included as part of the analysis carried out with the Syzygy software package. The C-Alpha statistic P-value for six genes achieved a p-value of <0.05 (Table 6.6), indicating that these genes were harbouring a large burden of rare variants in cases compared to controls.

A Chi-square test was used to compare the genotype counts of individual variants discovered in the T2D cases (500) to the controls samples (500). Although limited to larger effect sizes [80% power to detect an OR of 2.8, for a variant with a MAF 1% for alpha=0.05, calculated using Quanto software package (Gauderman 2002)], any variant with a p-value ≤0.05 was felt to be of interest for follow up in future studies provided the earlier prioritization criteria were met (Table 6.6, Table 6.7 and Table 6.8).

Table 6.6 Non-synonymous polymorphisms with MAF >0≤0.05 with SIFT and PolyPhen predictions with regard to the amino acid change affecting the protein function sites that were called with "high confidence" by the syzygy software package. A p-value of the chi-square testing comparing the 500 cases vs. 500 control samples in this study has also been included for each variant.

HG18 position with Chr number	gene	alleles	type	SIFT prediction	PolyPhen prediction	SNP*	Annotation	MAF Cases	MAF Controls	P-value Chi-square
chr1:154370876	*LMNA*	TG	ns	Damaging	Damaging	novel	Val191Gly	0.0099	0.0134	0.67
chr1:154373408	*LMNA*	CG	ns	Tolerated	Benign	novel	Pro485Ala	0.002	0.0009	1.00
chr1:154374094	*LMNA*	GA	ns	Tolerated	Damaging	1KG_1_156107470	Arg545His	0.001	0.001	1.00
chr1:154375134	*LMNA*	CT	ns	Damaging	Damaging	novel	Arg644Cys	0.005	0.008	0.58
chr3:186873042	*IGF2BP2*	GA	ns	Tolerated	Benign	novel	Ala394Val	0.0009	0.0019	1.00
chr3:186893227	*IGF2BP2*	TG	ns	Damaging	Damaging	novel	Thr120Pro	0.0165	0.0243	0.26
chr6:20654676	*CDKAL1*	GA	ns	Damaging	Benign	rs111739077	Arg39Gln	0.0119	0.011	1.00
chr6:21173428	*CDKAL1*	CT	ns	Damaging	Damaging	novel	Pro409Leu	0.0458	0.052	0.67
chr8:118252526	*SLC30A8*	AC	ns	Damaging	Damaging	novel	His301Pro	0.004	0.005	1.00
chr10:114915635	*TCF7L2*	CA	ns	Tolerated	Damaging	novel	Pro569Thr	0.0003	0.003	0.25
chr10:114915638	*TCF7L2*	TC	ns	Tolerated	Benign	novel	Ser570Pro	0.0016	0.0058	0.29
chr10:94204239	*IDE*	CA	ns	Tolerated	Benign	novel	Gly1001Val	0.002	0.0009	1.00
chr10:94254567	*IDE*	TG	ns	Tolerated	Damaging	novel	Glu465Ala	0.0001	0.002	0.50
chr10:94264674	*IDE*	AC	ns	Damaging	Damaging	novel	Val256Gly	0.0301	0.0259	0.68
chr10:94284426	*IDE*	TG	ns	Damaging	Damaging	novel	Tyr127Ser	0.0132	0.0239	0.13
chr11:17365407	*KCNJ11*	GC	ns	Tolerated	Benign	rs1800467	Leu183Val	0.0543	0.0451	0.47
chr12:119921457	*HNF1A*	TC	ns	Damaging	Damaging	novel	Leu502Pro	0.0127	0.0227	0.17
chr17:33139117	*HNF1B*	TG	ns	Damaging	Damaging	novel	His394Pro	0.0287	0.031	0.89
chr17:33144704	*HNF1B*	TG	ns	Damaging	Damaging	novel	Thr350Pro	0.0266	0.0188	0.28
chr20:42475778	*HNF4A*	CT	ns	Tolerated	Benign	rs1800961	Thr130Ile	0.0508	0.0389	0.28
chr20:42486153	*HNF4A*	AC	ns	Damaging	Damaging	novel	Asp325Ala	0.0119	0.01	0.83

ns: Non Synonymous; Chr: Chromosome

* Variants were deemed novel if not reported in dbSNP/100 Genome project (February 2012).

6.5.6 Functional evaluation

I also evaluated whether the variants that had been detected may have an implied functional relevance (e.g. coding change, alteration of canonical sequences, Section 6.4.2). Twenty one non-synonymous variants with a MAF <5% but greater than 0, excluding singleton cases, were identified (Table 6.5). Eleven of these variants were predicted to be damaging by both PolyPhen and SIFT software packages and were therefore considered for further evaluation (Table 6.6). However, none of these variants reached a P-Value of ≤0.05 in comparison of the 500 cases to 500 controls (Table 6.6).

For some genes (namely *TCF7L2, IDE, KCNJ11, LMNA* and *HNF4A)* parts of the promoter regions were also included. The Methylator (http://bio.dfci.harvard.edu/Methylator/) software package was used to assess whether methylation sites may be affected by variants in these promoter regions (Table 6.7). Three such variants, called with high confidence, were identified. One variant in *TCF7L2* (chr10:114700046), with a MAF in cases of 5.2% and in controls of 0.7%, achieved a p-value of $1.44X10^{-9}$ in comparison of 500 cases vs. 500 controls. A further variant in *LMNA* (chr1:154351035) with a MAF in cases of 1.5% and 1% in controls achieved a modest p value of 0.037 in comparison of 500 cases vs. 500 controls.

Table 6.7 Polymorphisms predicted to affect Methylation sites that were called with "high confidence" by the syzygy software package. A p-value of the chi-square testing comparing the 500 cases vs. 500 control samples in this study has also been included for each variant

Chr: position*	Gene	type	Base change	CPG site affected	MAF cases	MAF controls	Chi square test P Value
chr1:154351035	*LMNA*	nc	A/C	chr1_CGI_6107	0.015	0.010	0.037
chr1:154371306	*LMNA*	syn	G/A	chr1_CGI_6109	0.000	0.002	0.343
chr10:114700046	*TCF7L2*	utrsnc	C/T	chr10_CGI_3686	0.052	0.007	**$1.44X10^{-9}$**

syn: synonymous variant; Utrsnc Untranslated region- noncoding; nc Noncoding (Intronic); Chr: chromosome
*no rs numbers available for any of these variants.

With the use of the GeneSplicer (Pertea et al. 2001) and NetGene2 software packages, 13 variants were identified that were predicted to affect splice sites and were shortlisted for further evaluation (Table 6.8). Five of these variants achieved a p-value of <0.05 in comparison of 500 cases vs. 500 controls. It is of interest that one of these variants in *TCF7L2* (chr10:114700046) was also predicted to effect methylation as noted above. Another variant in *IGF2BP2* (chr3:186845492), with MAF in cases of 15.2% and in controls of 2.7% achieved a highly significant p-value of $1.23x10^{-23}$. Other variants with statistically significant p-values in comparison of 500 cases to 500 controls that were predicted to affect splice sites were variants in *LMNA* (chr1:154375620; MAF cases 3.6%, MAF controls 0.8%; p= $1.45x10^{-5}$), *SLC30A8* (chr8:118255973; MAF cases 2.0%, MAF controls 7.0%; p= $5.44x10^{-8}$) and *HNF4A* (chr20:42486355; MAF cases 0.9%, MAF controls 2.9%; p=$2.3x\ 10^{-3}$). All these results have been included in Table 6.8. These variants are not in linkage disequilibrium with the previously reported association signals for T2D. As part of .the follow-up work stemming from this project, these variants will undergo large scale association studies to establish if the observed association seen in this project is confirmed.

Table 6.8 Polymorphisms that are predicted to affect splice site that were called with "high confidence" by the syzygy software package. A p-value of the chi-square testing comparing the 500 cases vs. 500 control samples in this study has also been included for each variant

HG18 position	Gene	alleles	type	SNP	Annotation	MAF cases	MAF controls	P-Value(Chi-square)
chr1:154367032	*LMNA*	CT	syn	rs41313880	Arg119Arg	0.003	0.001	0.193
chr1:154374158	*LMNA*	CT	syn	rs4641	His566His	0.310	0.302	0.67
chr1:154375620	*LMNA*	GC	utrsnc	NA	NA	0.036	0.008	**1.45×10^{-5}**
chr3:186845492	*IGF2BP2*	AC	utrsnc	NA	NA	0.152	0.027	**1.23×10^{-23}**
chr8:118255973	*SLC30A8*	AC	utrsnc	NA	NA	0.020	0.070	**5.44×10^{-8}**
chr8:118255989	*SLC30A8*	CG	utrsnc	rs75043555	NA	0.067	0.051	0.12
chr8:118257040	*SLC30A8*	TC	utrsnc	NA	NA	0.030	0.020	0.24
chr10:114700046	*TCF7L2*	CT	utrsnc	NA	NA	0.052	0.007	**1.44×10^{-9}**
chr10:94213706	*IDE*	AG	syn	rs61736443	Asn841Asn	0.016	0.011	0.436
chr10:94224789	*IDE*	AC	nc	rs35223317	NA	0.031	0.028	0.78
chr10:94256325	*IDE*	TA	nc	rs34053974	NA	0.011	0.014	0.68
chr11:17364356	*KCNJ11*	TC	utrsnc	1KG_11_17407780	NA	0.007	0.013	0.26
chr20:42486355	*HNF4A*	AC	syn	NA	Pro392Pro	0.009	0.029	**2.3×10^{-3}**

syn: synonymous variant

utrsnc: Untranslated region- noncoding; nc: non-coding/ Intronic; NA: Not Applicable

Variants only seen in singleton cases

Variants only seen in singleton cases, that were predicted to be damaging by use of PolyPhen and SIFT software packages were also identified (Table 6.9). These variants were validated using Sanger sequencing (Chapter 2, section 2.2) of the specific region to confirm that these were true mutations and not sequencing errors. As described in section 6.5.5, this work was carried out by Ms Neelam Hassanali (DPhil student, University of Oxford) as part of the on-going work stemming from this project and thus far 8 out 10 variants have been validated (Table 6.9).

Two known MODY mutations were also identified [namely p.Gly326Val and p.Arg131Gln variants of *HNF4A* (Ellard et al. 2006)]. The particular samples have been identified by carrying out Sanger sequencing of the all samples that comprised the DNA pool in which the mutation was identified. These individuals have been invited to the monogenic diabetes clinic in the Oxford Centre for Diabetes and Endocrinology.

Eight additional variants only seen in singleton cases and not in control samples and resulting in an amino acid change were identified. These variants were predicted to be functionally damaging based on PolyPhen and SIFT prediction software (Table 6.9). These variants may be causative for monogenic diabetes and in one instance (His214Arg mutation in *HNF4A*) it has been established (as part of the work stemming from this project) that this mutation does indeed co-segregate with diabetic state in a family (involving parents and 1 sibling) and the proband has proved to be sensitive to sulphonylureas. This may be a novel mutation (not previously reported) causative for monogenic diabetes of the young (MODY).

For the remainder of these variants, the presence of the novel mutation has been validated by Sanger sequencing of all the samples in the DNA pool in which the mutation was discovered. This has identified the individual carrying the mutation. It is planned that sequencing of the parents +/- siblings (where DNA is available) will be carried out as part of the follow up work resulting from this project to establish if co-segregation with diabetic state occurs.

Table 6.9 Variants only seen in singleton cases, predicted to be damaging by PolyPhen and SIFT software packages and called with "high confidence" by the syzygy software package. The number of instances that each variant was detected has also been included**.**

Chr	Position (HG18)	Gene	Substitution	AA Change	Minor Allele	Number of instances
chr20	43042367	*HNF4A*	G>A	Arg131Gln	A	1
chr20	43042354	*HNF4A*	C>T	Arg127Trp^{+}	T	2
Chr20	43043184	*HNF4A*	G>A	Arg168Gln	A	1
Chr20	43047084	*HNF4A*	A>G	His214Arg	G	1
Chr20	43052769	*HNF4A*	G>T	Gly326Val	T	1
Chr20	43052756	*HNF4A*	C>T	Arg322Cys^{+}	T	1
Chr6	20546703	*CDKAL1*	G>A	Arg41Gln*	A	1
Chr6	21000551	*CDKAL1*	A>G	Arg335Gly	G	1
Chr12	121432125	*HNF1A*	C>A	Pro291Gln	A	4
Chr11	17409366	*KCNJ11*	C>T	Trp91Opl*	T	1

*Awaiting validation; += known mutation causative for MODY

6.5 Discussion

For many complex diseases such as T2D, the genetic variants identified to date only account for a relatively small proportion of the genetic disease component predicted from epidemiological studies. A heritability "gap" therefore exists that is to date not fully explained (Chapter, section 1.5). Advances in high throughput sequencing hold great promise in allowing for a more robust investigation of human allelic variation and may account for a proportion of this heritability gap.

This project was designed and implemented in 2007-2010 with the main aim of discovery of novel low frequency variants that may impact on the risk of T2D. To maximise the likelihood of detecting variants that may be functionally active, I resequenced the exons of ten genes that had been previously highlighted as potentially impacting on T2D risk. At project inception there were no guidelines or protocols widely available for sample preparation for high throughput targeted sequencing using this approach. It was therefore necessary to design and implement a pilot project to guide optimal experimental protocols. This pilot project was completed successfully and demonstrated that it was possible to identify novel low frequency variants in pooled DNA samples. Following this pilot study, I proceeded to upscale the project in much larger numbers and including larger target areas for sequencing. Protocols for sample preparation for targeted sequencing were not widely available at project inception and I have been able to help set these up in the host laboratory as a result of this project. Collaboration with colleagues in the Broad Institute was established for the analysis of targeted sequencing data and pipelines have been developed for future projects utilising targeted sequencing in pooled DNA samples.

It is clearly evident by the large number of variants detected in this project that strategies need to be in place to prioritise variants that may be of interest. Although the false negative rate in this project was relatively low (was in fact zero for variants with a MAF >1.0%), the false positive rate for low frequency variants can be high and strategies are needed to validate the results obtained. It is possible to utilise traditional Sanger sequencing for this but this can be both time consuming and costly if a large number of variants need to be validated. Strategies such as tagging of the individuals included in each DNA pool (Meyer et al. 2008) can greatly expedite this process by rapidly identifying the sample with the potential polymorphism or mutation. This approach was not possible at the inception of this project.

I have highlighted heuristics that can be implemented in prioritising variants that are detected through high throughput targeted sequencing. However, these strategies are guides only and to establish whether a variant is robustly associated with T2D risk, large scale genotyping will be required. Predictive software was used to evaluate whether variants detected were likely to be functionally active. However, these software packages are not always in agreement with each other. For instance PolyPhen and SIFT software packages did not always predict the same outcome from the same amino acid change (Table 6.7). Also, rs1800961 (T130I variant of *HNF4A*) is predicted to result in a tolerated amino acid change. However, as highlighted in Chapter 4, this variant has been shown to be functional based on studies of the transcriptional regulation of *HNF4A*-target genes in a range of cell lines and primary mouse hepatocytes (Moller et al. 1997; Navas et al. 1999; Zhu et al. 2003; Ek et al. 2005). It is therefore suggested that predictive software can be useful as a guide but functional studies, where available, should supersede the findings of these software packages (Bonnefond et al. 2012).

Several polymorphisms detected in this project have been highlighted as warranting follow up (Section 6.4.3). These are first undergo validation, if novel, to ensure that they have not resulted from sequencing error. This is being done by Sanger sequencing in the first instance as part of the work stemming from this project. Validated samples will be taken forward for genotyping in samples available locally (>10000 cases and >20000 controls) to assess association with T2D risk. However, projects such as the Exome Chip project (ftp://share.sph.umich.edu/exomeChip/), which are aimed at study of coding regions, variants which may affect splice sites and stop codons will be a great asset also in the follow up of some of the variants highlighted in this project. These arrays have incorporated the sequencing information from over 12,000 sequenced exomes and have included approximately 180000 nonsynonymous variants and approximately 300 splice variants. One

of the variants predicted to affect a splice site that has been highlighted in this study (chr1:154367032, *LMNA*, p=0.19 in the 500 cases vs. 500 controls) has been included in this array. Five nonsynonymous variants highlighted in this study (*HNF4A*, rs1800961; *KCNJ11*, rs1800467; CDKAL1, rs111739077 and the LMNA variants in positions chr1:154374094 and chr1:154375134) have also been included as part of this array. It is of interest that only one of these variants (*LMNA,* chr1:154375134) was predicted to be damaging by both the PolyPhen and SIFT software packages.

High throughput sequencing technologies have advanced greatly since this project was completed. For instance, read lengths of up to 150bp are now possible with the Genome Analyzer IIx machine (which has superseded the platform used in this project). However, this project , which was the first targeted sequencing project in the Oxford Centre for Diabetes and Endocrinology, has established protocols and pipelines for sample preparation and analysis that future studies will be able to utilize across several high throughput sequencing platforms.

CHAPTER 7

General Conclusions and Discussion

General Conclusions and Discussion

The ultimate aspiration of genetic studies in Type 2 diabetes (T2D) is improved care and outcome for patients diagnosed with the condition. It is hoped that with improved knowledge of the susceptibility variants appreciably increasing T2D risk, better biological models of the disease will be devised and further therapeutic targets identified. One aspect of this is a better understanding of genetic heritability of this condition.

Despite the significant advances in our understanding of the genetics of T2D driven largely through the success of GWAS approaches, only a small proportion (<10 %) of the heritability of this condition is currently explained (Morris et al. 2012). The overall aim of my thesis has been to try and close this heritability gap through the identification of low frequency, incompletely penetrant (LFIP) variants in genes previously implicated in T2D pathogenesis or monogenic diabetes that may appreciably impact on T2D risk. Sanger sequencing, large scale genotyping and high throughput targeted sequencing were utilised in the search for such variants in this thesis. This study was designed and completed between 2007 and 2009 when next generation technologies and sequencing platforms were in their infancy.

Firstly, as outlined in Chapter 3, Sanger sequencing was utilised to evaluate whether the latter 3 exons of *HNF1A* may harbour LFIP variants that may appreciably contribute to the T2D risk. This region of *HNF1A* was prioritised based on the previously published observation that mutations in the latter 3 exons of this gene resulted in a later age of diagnosis of HNF1A-MODY compared to mutations in exons 1-7 (Harries et al. 2006; Bellanne-Chantelot et al. 2008). I therefore hypothesised that this region of the gene could also harbour LFIP variants contributing to T2D pathogenesis. Traditional Sanger capillary sequencing was used to screen 591 samples of British ancestry for variants. No novel low frequency variants were identified

in this region of the gene, although four previously reported variants (minor allele frequency range 4.1%-33.4%) were successfully identified using this approach. This study was one of the first reporting the search for low frequency variants contributing to T2D risk. The approach was suitable for this targeted resequencing study but its application to whole genome sequencing is not practical both for cost and practicality reasons. However, sanger sequencing remains a powerful methodology for the validation of variants discovered using high throughput sequencing platforms.

In Chapter 4 of this thesis the role played in the pathogenesis of T2D by candidate LFIP variants was evaluated by large scale genotyping efforts. A literature search focused on variants identified in genes previously implicated in diabetes risk was performed to identify candidate LFIP variants for large-scale association analysis. Five variants, across four different genes (*HNF4A*, *HNF1A*, *PDX1* and *KCNJ15*) had been postulated to impact on T2D pathogenesis through previous functional or association studies (Stoffers et al. 1997; Hani et al. 1999; Zhu et al. 2003; Owen et al. 2004; Ek et al. 2005; Winckler et al. 2005; Okamoto et al. 2010). I carried out the largest association analysis for these variants in T2D samples.

For four of these variants (V255M; *HNF4A*, A98V; *HNF1A*, D76N; *PDX1* and I88I; *KCNJ15*), no statistically significant associations with T2D risk were found in this study. To increase statistical power I subsequently carried out a meta-analysis incorporating the results from my study and those of all previously reported studies. Despite the increased sample size and inclusion of positive association studies the meta-analysis did not alter my original results. Importantly, these studies demonstrate a number of important principles. First, due to the low allele frequency of these variants, the power to detect an association with increased T2D disease risk was limited to large effect sizes only, thus highlighting the difficulties in confirming or refuting the association of low frequency variants, especially when the effect

size is small. Second, for two of these variants (V255M; *HNF4A* and D76N;*PDX1*) previous studies had demonstrated decreased transcriptional regulation of various target genes in several different cell lines (Moller et al. 1997; Navas et al. 1999; Weng et al. 2001; Ek et al. 2005). Critically, these studies were performed in cell lines (such as COS7 and HEPG2) that were not derived from the primary tissue (pancreatic beta-cells) where these transcription factors mediate their effects on T2D risk. These results indicate that functional studies should be interpreted with caution when performed in cell lines derived from tissues in which the variant under investigation does not mediate its primary effect. Finally, the I88I variant of *KCNJ15* was selected for further evaluation as previous reports in a Japanese sample had indicated a possible association with increased T2D risk (Okamoto et al. 2010). However, this variant was noted to have a much lower MAF in European samples (0.89% in Danish samples (Okamoto et al. 2010), 0.79% in UK samples, Chapter 4, Section 4.3.2). Statistically significant association with T2D risk was not seen in the current study in European samples (UK and Denmark; Chapter 4, Section 4.3.2). The differing haplotype structures in the region of this variant in Japanese and European samples has been demonstrated in this study and may account for these disparate findings. The T2D signal in Japanese samples may therefore be secondary to another variant with which the I88I variant is in linkage disequilibrium.

With better knowledge of the genomic sequence variation in differing populations, such as that provided by the 1000 Genome project (The 1000 Genomes Project Consortium 2010), a better understanding of haplotype structures in these populations will be provided. This knowledge would aid in fine-mapping association signals detected in differing populations. Efforts aimed at understanding the allelic spectrum of disease causing variants in differing populations are already underway (The 1000 Genomes Project Consortium 2010; Morris 2011; Cho et al. 2012; Dastani et al. 2012; Huang et al. 2012). The majority of GWAS in T2D that have been completed to date have been carried out in samples of European white

ancestry (Saxena et al. 2007; Scott et al. 2007; Zeggini et al. 2007; Zeggini et al. 2008; Voight et al. 2010). With expanded studies in other ethnic groups, it is hoped that further susceptibility variants will be uncovered. Several examples of this have already been seen such as loci in *KCNQ1* (Unoki et al. 2008; Yasuda et al. 2008) that were highlighted as impacting on T2D pathogenesis in Japanese samples prior to confirmation in European samples, due to lower allele frequencies in the latter (Chapter 4, Section 4.1.4).

In contrast to the my findings with the above candidate variants, the data generated for the T130I variant of *HNF4A* suggested a role in T2D pathogenesis and demonstrated the challenges faced by the field in demonstrating a role for low frequency variants in disease pathogenesis. This variant was selected for further follow-up, based on previous work demonstrating decreased transcriptional regulation of *HNF4A* target genes secondary to this variant. Moreover, previous association studies had demonstrated a weak association of the T130I variant with T2D risk (Ek et al. 2005; Winckler et al. 2005; Sookoian et al. 2010). A modest association with T2D risk was found ($p=5x10^{-4}$) for this variant in UK samples in this study (Chapter 4, Section 4.3.5). The strength of this association was increased by the inclusion of all previous studies of this variant in T2D samples in a meta-analysis ($p=2.1x10^{-5}$). However, this association does not reach the widely accepted threshold for genome wide level of significance ($p=5x10^{-8}$). My study was one of the first published large scale genotyping projects aimed at evaluating the role of low frequency variants in T2D risk. As efforts are increasingly aimed at deciphering the allelic spectrum of T2D susceptibility variants, it is likely that further low frequency variants will be implicated in the pathogenesis of this condition. The data from this study exemplify the vast sample numbers required to confirm association to a complex disease state for low frequency variants, especially those with modest effect sizes.

In Chapter 5, the *LARS2* gene, encoding for the mitochondrial leucyl tRNA synthetase (EC 6.1.1.4), which is vital in mitochondrial protein synthesis was prioritised for further evaluation. This is of interest as reduced mitochondrial activity has been implicated in increased T2D risk (Gerbitz et al. 1996; Kelley et al. 2002; Bajpeyi et al. 2011). Furthermore, previous reports had implicated the low frequency (MAF 3.0%) H324Q variant of *LARS2* in T2D pathogenesis ('t Hart et al. 2005). However, the strength of this reported association [p=0.004 ('t Hart et al. 2005)] was far short of the now expected stringent genome wide levels of significance ($\alpha=5x10^{-8}$). This variant was therefore identified in the current study as a candidate LFIP variant that may impact on T2D pathogenesis and was taken forward for large scale genotyping to evaluate the role that it may play in T2D risk. Moreover, to fully study the possible role of known low frequency variants in *LARS2* in T2D pathogenesis, the results of GWAS meta-analysis by the DIAGRAM group (Zeggini et al. 2008) was scrutinised for any such variants. A group of low frequency variants that were in tight linkage disequilibrium were highlighted in this GWAS meta-analysis (Zeggini et al. 2008) as having a modest association with T2D (p=0.02–0.05). A proxy for this group was selected for further analysis.

This study did not support a role for the known low frequency variants in *LARS2* in T2D risk. This was one of the first published reports evaluating the potential role of all known low frequency variants in a single gene in T2D pathogenesis. However, the possibility that other, as yet undiscovered, low frequency variants in *LARS2* may impact on T2D pathogenesis cannot be out ruled based on this study. This point can be addressed by a better understanding of the sequence variation of this gene such as that outlined in Chapter 6 of this thesis.

The final of aim of this thesis (presented in Chapter 6) was the use of a “next-generation” high throughput sequencing platform to identify novel low frequency variants that may

impact on T2D pathogenesis. This study was designed and implemented between 2007 and 2009, when this technology was in its infancy. No protocols were available for sample preparation for this platform at project inception. It was therefore necessary to optimise pathways and establish protocols for the preparation of samples prior to sequencing on the Genome Analyzer II platform. At the time of project commencement, the cost of whole genome sequencing was prohibitively expensive. Therefore, this project was limited to genomic regions that were felt to increase the probability of detecting T2D susceptibility variants that may be functionally active. The exons of ten genes (total genomic extent 64 kb) that had previously been implicated in the pathogenesis of either T2D (through GWAS or linkage studies) or monogenic diabetes were prioritised for this project. To further increase the likelihood of detecting susceptibility variants that may impact on T2D pathogenesis, samples that were likely to be enriched for such variants, due to young onset or family history of T2D, were selected. This platform was used to successfully sequence the exonic regions of the selected genes in pooled DNA samples.

In excess of 2900 variants were identified using this approach, a sizeable portion of which were novel. It was therefore necessary to develop heuristics to prioritise variants that may be functionally active. Predictive software was utilised to identify variants that may be deleterious to gene function and those that may affect methylation or splice sites. This study has highlighted several susceptibility variants that may appreciably impact T2D disease state. As part of on-going work stemming from this project, these variants are undergoing validation and association to T2D will be further clarified by use of large scale genotyping, such as that highlighted in Chapters 4 and 5 of this thesis.

As a result of my study, protocols for preparation of samples for high throughput sequencing, which can be utilised across several platforms, have been established in the host institution.

In this study, I have also highlighted heuristics that can be implemented in prioritising the large number of variants that can be discovered using the high throughput sequencing approach.

This project was completed in 2009. Since that time, the development of high throughput sequencing platforms has substantially progressed, with lower costs, longer read lengths and shorter run times (Koboldt et al. 2012). This is allowing projects such as the 1000 Genome project (The 1000 Genomes Project Consortium 2010), the Exome sequencing project (http://evs.gs.washington.edu/EVS/) and the ENCODE project(Birney et al. 2007) to compile a better understanding of human genetic architecture. In turn, this will allow for a more in depth study of the genetic variants that may be causative for complex diseases such as T2D.

The known susceptibility variants implicated in the pathogenesis of T2D have shed some light on the biological mechanisms that may underlie this condition. However, for the majority of T2D susceptibility variants thus far implicated in T2D pathogenesis, the biological pathway affected is not yet known. A large portion of these loci lie in the intronic regions of genes or in the region between genes. Recently, it has been demonstrated that these loci may be biologically active, such as the intronic *TCF7L2* variant (rs7903146) that has been shown to be in a cis regulatory element (Gaulton et al. 2010). Intronic loci showing association with increased risk of T2D may also be in linkage disequilibrium with another, as yet undiscovered, causative variant and it will therefore be difficult to ascertain the biological mechanism affected. However, with improved understanding of the allelic spectrum of genetic variants causative for T2D, using methodologies highlighted in this thesis among others, it is hoped that a better understanding of the pathogenesis/biological pathways involved in disease susceptibility will be obtained. This in turn will aid efforts to develop novel therapeutics for T2D.

It is already known that variants in *TCF7L2* can reduce response to sulphonylureas (Pearson et al. 2007; Holstein et al. 2011) and variants in *ATM* improve response to metformin (Zhou et al. 2011). T2D susceptibility variants have also been discovered within *KCNJ11* and *PPARG*, two genes that encode known targets of sulphonylureas and thiazolidinediones respectively (Lehmann et al. 1995; Sagen et al. 2004). These examples demonstrate how genetic susceptibility may be associated with a therapeutic response. There is hope therefore, that with better understanding of the biological mechanisms underlying the T2D susceptibility loci, other therapeutic targets may be discovered as well as predicting which individuals may benefit from a particular therapeutic agent.

Efforts have been made to evaluate whether knowledge of the number of susceptibility variants that an individual harbours may be used to stratify their risk of developing T2D (Lango et al. 2008; Lyssenko et al. 2008; Lyssenko et al. 2012) . The loci thus far implicated in T2D pathogenesis, account for only approximately 5.7% of variance in disease susceptibility (Morris et al. 2012). At present, the incorporation of a genetic risk score, based on the susceptibility loci thus far implicated in T2D pathogenesis, adds little to the risk identifiable by clinical and biochemical criteria of increasing BMI, sex, family history, elevated systolic blood pressure and hyperlipidaemia (Lango et al. 2008; Imamura et al. 2011; Lyssenko et al. 2012).

However, it is hoped that in the future, with a fuller understanding of the genetic heritability of T2D, and genetic prediction of response to various therapeutic options as well as rapid whole genome sequencing, which is able to give results reliably and affordably, we may be able to deliver truly personalised medicine. Alas, we are far from this point at present, though with the current rate of advances in the genetics of T2D, the future appears bright.

In conclusion, through the course of this thesis, I have concentrated on strategies that can be employed to evaluate the potential role of low frequency, incompletely penetrant variants in the pathogenesis of T2D. This project has highlighted strategies and developed pipelines that may be utilised in deciphering a portion of the genetic heritability of T2D and have laid the ground work for future sequencing projects across several platforms in the Oxford Centre for Diabetes and Endocrinology and further afield. It is hoped that, these in turn will further expand our knowledge of the genetic heritability of T2D in the future.

Appendices

Supplementary Tables

Table S1.1. Loci reaching genome-wide significance (p<5x10-8) association with Type 2 Diabetes – arranged by chromosome number. The gene in closest proximity to the associated loci has also been quoted. The odds ratio quoted has been obtained from the referenced article in the last column**.**

SNP associated	Nearby gene symbol	Gene name	Chr	OR (95% CI)	Method of discovery	Reference
rs10923931	*NOTCH2*	Notch homolog 2 (Drosophila)	1	1.13 (1.08-1.17)	GWA	(Zeggini et al. 2008)
rs340874	*PROX1*	prospero homeobox 1	1	1.07 (1.05–1.09)	GWA	(Dupuis et al. 2010)
rs2943641	*IRS1*	Insulin receptor substrate 1	2	1.19 (1.13–1.25	LSA	(Rung et al. 2009)
rs7578597	*THADA*	Thyroid adenoma associated	2	1.15 (1.10-1.20)	GWA	(Zeggini et al. 2008)
rs243021	*BCL11A*	B-cell CLL/lymphoma 11A (zinc finger protein)	2	1.08 (1.06–1.10)	GWA	(Zeggini et al. 2008)
rs3923113	*GRB14*	Growth factor receptor-bound protein 14	2	1.08 (1.05-1.11)	GWA	(Kooner et al. 2011)
rs780094	*GCKR*	glucokinase (hexokinase 4) regulator	2	1.06 (1.04–1.08)	GWA	(Dupuis et al. 2010)
rs7593730	*RBMS1 /ITGB6*	RNA binding motif, single stranded interacting protein 1 / integrin, beta 6	2	0.90 (0.86-0.93)	GWA	(Qi et al. 2010)
rs831571	*PSMD6*	Proteasome (prosome, macropain) 26S subunit, non-ATPase, 6	3	1.09 (1.06–1.12)	GWA	Cho et al. 2012)
rs6780569	*UBE2E2*	ubiquitin-conjugating enzyme E2E 2	3	1.21 (1.14-1.30)	GWA	(Yamauchi et al. 2010)
rs4402960	*IGF2BP2*	Insulin-like growth factor 2 mrna binding protein 2	3	1.14 (1.11-1.18)	GWA	(Zeggini et al. 2007)
rs11708067	*ADCY5*	adenylate cyclase 5	3	1.12 (1.09–1.15)	GWA	(Dupuis et al. 2010)
rs4607103	*ADAMTS9*	ADAM metallopeptidase with thrombospondin type 1	3	1.09 (1.06-1.12)	GWA	(Zeggini et al. 2008)
rs16861329	*ST6GAL1*	ST6 beta-galactosamide alpha-2,6-sialyltranferase 1	3	1.08 (1.05-1.11)	GWA	(Kooner et al. 2011)
rs1801282	*PPARG*	Peroxisome proliferator-activated receptor gamma	3	0.78 (0.59-1.05)	CGA	(Altshuler et al. 2000)
rs10010131	*WFS1*	Wolframin	4	0.89 (0.86-0.92)	CGA	(Franks et al. 2008)
rs459193	*ANKRD55*	Ankyrin repeat domain 55	5	1.08 1.05–1.11)	GWA	(Morris et al. 2012)

rs1535500	*KCNK16*	potassium channel, subfamily K, member 16	5	1.08 (1.05–1.11)	GWA	Cho et al. 2012)
rs4457053	*ZBED3*	Zinc finger, BED-type containing 3	5	1.08 (1.06–1.11)	GWA	(Voight et al. 2010)
rs9470794	*ZFAND3*	Zinc finger, AN1-type domain 3	6	1.12 (1.08–1.16	GWA	Cho et al. 2012)
rs10946398	*CDKAL1*	CDK5 regulatory subunit associated protein 1-like 1	6	1.12 (1.08-1.16)	GWA	(Zeggini et al. 2007)
rs6467136	*GCC1*	GRIP and coiled-coil domain containing 1	7	1.11 (1.07–1.14)	GWA	Cho et al. 2012)
rs864745	*JAZF1*	JAZF zinc finger 1	7	1.10 (1.07-1.13)	GWA	(Zeggini et al. 2008)
rs972283	*KLF14*	Kruppel-like factor 14	7	1.07 (1.05–1.10)	GWA	(Voight et al. 2010)
rs4607597	*GCK*	glucokinase (hexokinase 4)	7	1.07 1.05-1.10	GWA	(Dupuis et al. 2010)
rs2191349	*DGKB-AGMO*	diacylglycerol kinase, beta 90kDa/ alkylglycerol monooxygenase	7	1.06 (1.04-1.08)	GWA	(Dupuis et al. 2010)
rs516946	*ANK1*	Ankyrin 1, erythrocytic	8	1.09 (1.06–1.12)	GWA	(Morris et al. 2012)
rs13266634	*SLC30A8*	Solute carrier family 30 (zinc transporter), member 8	8	1.12 (1.07-1.16)	GWA	(Zeggini et al. 2007)
rs896854	*TP53INP1*	Tumor protein p53 inducible nuclear protein 1	8	1.06 (1.04–1.09)	GWA	(Voight et al. 2010)
rs2796441	*TLE1*	Transducin-like enhancer of split 1 (E(sp1) homolog, Drosophila)	9	1.07 (1.05–1.10)	GWA	(Morris et al. 2012)
rs7041847	*GLIS3*	GLIS family zinc finger 3	9	1.10 1.07–1.13	GWA	(Cho et al. 2012)
rs17584499	*PTPRD*	Protein tyrosine phosphatase receptor type D	9	1.57 (1.36-1.82)	GWA	(Tsai et al. 2010)
rs10811661	*CDKN2A/2B*	Cyclin-dependent kinase inhibitor 2A/2B	9	1.20 (1.14-1.25)	GWA	(Zeggini et al. 2007)
rs13292136	*CHCHD9*	Coiled-coil-helix-coiled-coil-helix domain containing 2 pseudogene 9	9	1.11 (1.07–1.15)	GWA	(Voight et al. 2010)
rs7018475	*CDKN2B*	Cyclin-dependent kinase inhibitor 2B	9	0.74 (0.64–0.85)	GWA	(Huang et al. 2012)
rs12571751	*ZMIZ1*	Zinc finger, MIZ-type containing 1	10	1.08 (1.05–1.10)	GWA	(Morris et al. 2012)
rs7901695	*TCF7L2*	Transcription factor 7-like 2	10	1.56 (1.41-1.73)	LSA	(Grant et al. 2006)
rs1111875	*IDE/HHEX*	Insulin-degrading enzyme/ hematopoietically expressed homeobox	10	1.13 (1.08-1.17)	GWA	(Zeggini et al. 2007)
rs10906115	*CDC123*	cell division cycle 123 homolog (S. cerevisiae)	10	1.13 (1.08-1.18)	GWA	(Shu et al. 2010)
rs12779790	*CAMK1D*	Calcium/calmodulin-dependent protein kinase 1D	10	1.11 (1.07-1.14)	GWA	(Zeggini et al. 2008)

rs1802295	*VPS26A*	Vacuolar protein sorting 26 homolog A	10	1.07 (1.05-1.10	GWA	(Kooner et al. 2011)
rs2283228	*KCNQ1*	Potassium voltage-gated channel, KQT-like	11	1.26 (1.18-1.34)	GWA	(Unoki et al. 2008)
rs5215	*KCNJ11*	Potassium inwardly-rectifying channel, subfamily J, member 11	11	1.15 (1.09-1.21)	GWA	(Scott et al. 2007)
rs1552224	*ARAP1 (CENTD2)*	Arfgap with rhogap domain, ankyrin repeat and PH domain 1	11	1.14 (1.11–1.17)	GWA	(Voight et al. 2010)
rs1387153	*MTNR1B*	Melatonin receptor 1B	11	1.09 (1.06–1.11)	GWA	(Prokopenko et al. 2009)
rs10842994	*KLHDC5*	Kelch domain containing 5	12	1.10 (1.06–1.13)	GWA	(Morris et al. 2012)
rs11063069	**CCND2*	Cyclin D2	12	1.12 1.08–1.16)	GWA	(Morris et al. 2012)
rs1531343	*HMGA2*	High mobility group AT-hook 2	12	1.10 (1.07–1.14)	GWA	(Voight et al. 2010)
rs7961581	*TSPAN8*	Transmembrane 4 superfamily member 3	12	1.09 (1.06-1.12)	GWA	(Zeggini et al. 2008)
rs7957197	*HNF1A*	HNF1 homeobox A	12	1.07 (1.05–1.10)	GWA	(Voight et al. 2010)
rs1359790	*SPRY2*	sprouty homolog 2 (Drosophila)	13	1.15 (1.10-1.20)	GWA	(Shu et al. 2010)
rs7172432	*C2CD4A/B*	C2 calcium-dependent domain containing 4A/B	15	1.14 (1.09–1.20)	GWA	(Yamauchi et al. 2010)
rs7178572	*HMG20A*	High-mobility group 20A	15	1.08 (1.06-1.11)	GWA	(Kooner et al. 2011)
rs2028299	*AP3S2*	Adaptor-related protein complex 3, sigma 2 subunit	15	1.08 (1.06-1.11)	GWA	(Kooner et al. 2011)
rs8042680	*PRC1*	Protein regulator of cytokinesis 1	15	1.07 (1.05–1.09)	GWA	(Voight et al. 2010)
rs11634397	*ZFAND6*	Zinc finger, AN1-type domain 6	15	1.06 (1.04–1.08)	GWA	(Voight et al. 2010)
rs7202877	*BCAR1*	Breast cancer anti-estrogen resistance 1	16	1.12 1.07–1.16)	GWA	(Morris et al. 2012)
rs8050136	*FTO*	Fat mass and obesity associated	16	1.17 (1.12-1.22)	GWA	(Zeggini et al. 2007)
rs4523957	*SRR*	Serine racemase	17	1.28 (1.18-1.39)	GWA	(Tsai et al. 2010)
rs4430796	*HNF1B*	Hepatocyte nuclear factor 1-beta	17	0.91 (0.87–0.94)	CGA	(Gudmundsson et al. 2007)
rs12970134	*MC4R*	Melanocortin 4 receptor	18	1.08 (1.05–1.11)	GWA	(Morris et al. 2012)
rs12454712	*BCL2*	B-cell CLL/lymphoma 2	18	1.09 (1.05-1.11)	GWA	(Saxena et al. 2012)
rs10401969	*CILP2*	Cartilage intermediate layer protein 2	19	1.13	GWA	(Morris et

				(1.09–1.18)		al. 2012)
rs3786897	*PEPD*	Peptidase D	19	1.10 1.07–1.14)	GWA	Cho et al. 2012)
rs3794991	*GATAD2A*	GATA zinc finger domain containing 2A	19	1.12 (1.08–1.15)	GWA	(Saxena et al. 2012)
rs6017317	*FITM2*	Fat storage-inducing transmembrane protein 2	20	1.09 1.07–1.12)	GWA	Cho et al. 2012)
rs4812829	*HNF4A*	Hepatocyte nuclear factor 4, alpha	20	1.09 (1.06-1.12)	GWA	(Kooner et al. 2011)
rs5945326	*DUSP9*	Dual specificity phosphatase 9	X	1.27 (1.18–1.37)	GWA	(Voight et al. 2010)

Chr: Chromosome

OR: Odds Ratio

CI: Confidence Interval

GWA: Genome wide association study

CGA: Candidate gene approach

LSA: Large scale association

*: Association reaches GWA level of significance in men only

Table S6.1: Primers utilised in the amplification of the exonic regions of the Genes included in targeted high throughput sequencing

Gene Name	Amplicon Name	Forward Primer Sequence	Reverse Primer Sequence
CDKAL1	CDKAL1-01	CCAGAACTAGCGCTTTCAGC	TTGGCCATAAAAGGAGCTGT
CDKAL1	CDKAL1-02	TTTTTGGGTGCTTACTTTTCC	TGCTATGTAAAGCACAATGTTTCC
CDKAL1	CDKAL1-03	TTGAAGTTAGATCCAAAGGCTTG	CCAGTAGCGTCAGCTTCTCC
CDKAL1	CDKAL1-04	TTTAAAAATAACTCAATGAGGTTGG	TCACCTTAAAATCACTGTCATGC
CDKAL1	CDKAL1-05	TGTTTTTCTCTCCCCTACCG	TCTTTTGGTATCAATTCTTACTTTGA
CDKAL1	CDKAL1-06	CCAAGAGTATACCCATGAGCAA	GCTTACAAATAACCTACAAACAGAAAA
CDKAL1	CDKAL1-07	CCTAAAACAACTCCAGGCAGTTTA	GGCAAATTCATGAAGCCAAG
CDKAL1	CDKAL1-08	TGTTGAAAGCGTTAAGTTACAGTGA	TTTCCCAAGAAAATGTTCAGC
CDKAL1	CDKAL1-09	TGTTTTGAGGTATCCCCATGT	TTCAGAAAACTAAAGAAGTTGGTCTC
CDKAL1	CDKAL1-10	TGAGCAGCTGTATGATGAAGG	GCCACACTGTCTGGACTGG
CDKAL1	CDKAL1-11	TTGTGGTGTTGATGTGAGGAA	AAGTGCATTTTTGCCACAAG
CDKAL1	CDKAL1-12	CTATTTGCACGTGTGCTTTTT	GCAAGTTGTGTGAAGGTAGGC
CDKAL1	CDKAL1-13	GCCTCCACAAAGGAAGAATC	TCACCATATGTTGGTGAAAATTG
CDKAL1	CDKAL1-14	CCTGGGCTAGCCTTTATTTTT	CCCTTTAAATGCCCTTTGCT
CDKAL1	CDKAL1-15	GAAAACTGGGAGTTAGGTATACAGGA	GAAACATGGCCTGGAATCAC
CDKAL1	CDKAL1-16	GCACCTTCTGGAACCCATT	AAAATTGTCTGGGCTTGGTG
HNF1A	HNF1A-01	TGCAAGGAGTTTGGTTTGTG	GCCCCTCTAGGCTCTCCTG
HNF1A	HNF1A-02	GGGCTCCATAACTGCTTTCA	ATGAGTTAGGGGAGAGTCCTG
HNF1A	HNF1A-03	AAGAATCAAGGGCAAGGTCA	GGCAACTGGACAGCCTTTTA
HNF1A	HNF1A-04	CTCAGAACCCTCCCCTTCAT	GGACAGTCCTCCCCAACC
HNF1A	HNF1A-05	TTTGAAGTGCTGAGGGCTGT	CCAGAATCTCCCTGCCAAG
HNF1A	HNF1A-06	CTTTCCTTGGCAGGGAGATT	ATGAGTCCCAGTGGCTCTTC
HNF1A	HNF1A-07	GGCAGGGGTGGGATATAACT	GACAGCCAACCTCTATCATCATC
HNF1A	HNF1A-08	AGTCTTGAGGCCTGGGACTA	AAGGGGTGGGACCAACAT
HNF1A	HNF1A-09	TGAGTACCCCTAGGGACAGG	GCTTTGCTCCTAGCTCTCCA
HNF1A	HNF1A-10	ATCAGAAAGGGAGGGCTCTG	AGGCTCTGGCACAGAGTAGC
HNF1A	HNF1A-11	CTTCCAGCTAGTGACCCACA	CCTTGCTGCCTTGTGAGC
HNF1A	HNF1A-12	GCAGACCCTGCCCTTGTT	CCCCATCTGAATGATTTGTG
HNF1B	HNF1B-10	GACTTCTCTGGTGGGAAACG	CTTCTGCTCTCCAGGTCTGC
HNF1B	HNF1B-1	AGAAAATTCCAAGGAGACCAG	AACTGATGCGAAAACTTGAATC
HNF1B	HNF1B-2	GCCATTGGACAAATTTACTTCTTT	TGGGCATCATCTCCCTTAGA
HNF1B	HNF1B-3	CCACAACCTCTGCACATCC	AGATGGGAGCTATGGTGTGG
HNF1B	HNF1B-4	TTCAGACCCAGAGAGGGAAA	CGGTGACTGGGACATTGAG
HNF1B	HNF1B-5	GGAGACCTACAGTTATTTTTCACCA	CACATCGTGTTGGAAACTGC
HNF1B	HNF1B-6	CTGGGTCACCAGCACTACCT	AGAGGTGCCGAGTCATTGTT
HNF1B	HNF1B-7	TCCTCTGAAAACCCTTAAACCA	GTTTTTGGCCAAGCACCA
HNF1B	HNF1B-8	ATTGGGGTTCTGTGGAACAT	TCGTCCGTTGTCTGTCTGTC
HNF1B	HNF1B-9	CCTTGTCGATGAAAGGGAAA	GCACCCCCTAGAAAAAGAATG
HNF4A	HNF4A-01	ACCCCTTCACAGAGGCTAGG	CCAAAGATCTGCTCCTGGAC
HNF4A	HNF4A-02	GCTCCCTTAGATGCCTGACA	CCAAGTGTGCCCATTTCC
HNF4A	HNF4A-03	CAGGGCCCTAGTTCTGTCCT	ATGACTGTCGGGGAGCTGT
HNF4A	HNF4A-04	GCTCCCACTCCTCATCAGTC	GGTGAGCTCCTCTGTAGTGTGG
HNF4A	HNF4A-05	GAGGGCACCCACTATCCAG	CCACGGCTATATCCCAGGT
HNF4A	HNF4A-06	AGCCTTCCCAAGGGTACAGA	CACCATGTGAATCTCCTTGC
HNF4A	HNF4A-07	TTGCCAACTTAAAAGCCAAAA	GCAGAGCGTTCTGGAGAGAG
HNF4A	HNF4A-08	ACAAGTCAGGGGACATCTGG	TTGCCTTGTTCCCATTTTTC
HNF4A	HNF4A-09	GGAGACAGGCCTCACACAGT	CCTGTGGCCTCCCCTATCTA

HNF4A	HNF4A-10	AAGGCCTGAGGTCTGCATC	CACTTTCTCCTGCCTCCTGT
HNF4A	HNF4A-11	TACTCCCACAAAGGCTGGAA	AAGGCAGTGGCTTCAACAT
HNF4A	HNF4A-12	ACGCCCTGCTCTGGATAAC	CTTCTAGACCCTGCCTGGTG
HNF4A	HNF4A-13	TTCTCCAACACCACCTCTCC	GACCCTATGCTCTCCAGCAG
HNF4A	HNF4A-14	AAAGAACAGCCTGAGCCAAG	GGGAGGGAGGAGGAGAATAA
HNF4A	HNF4A-15	TCTGACATGCATCATCTCATTT	TTGACTTGGGGAGTGAAGAGA
IDE	IDE-01	GGACAAAATTGATGTAAGGTAAGTCTG	GAGGTATGCTCCCAGTGTCC
IDE	IDE-02	AAAAACCAAGCTGCTTCTTCTAAA	GCTGGTGCGGTGTTTTTAAT
IDE	IDE-03	AGCTATGACAAAGGCCAACC	AGTCTTCAGAGAACACCTCATTG
IDE	IDE-04	AAGTGATCCTTCCACCTTGG	CGTGGGTGACAGAGTGAAAC
IDE	IDE-05	CCCAACTCCCCAAATAGGTAA	TGCTCAGACAGGTTAGTTTTTGA
IDE	IDE-06	TGATTGCTTATCTTTTAGAAGTGGA	GTGGGCAGACATAGCATTGA
IDE	IDE-07	CCAAATTCAGCCCACCTGTA	GGGTGTCTGTTTCTTTCACCA
IDE	IDE-08	TCTTGAGAAAGACTAGAGGGAAAAA	AGGGAAAACTAGGTGAAGTGTTT
IDE	IDE-09	CCTGCCCACCACTGTTTTTA	GGGGAAAGAAGGGAATATCCTA
IDE	IDE-10	TCTTCTTTTCCTCCTTGCTTC	TCAAGTCGTTTATTTGGAAGTTG
IDE	IDE-11	AAACTCACACCCTCAATCAAAAA	AAATAGCCTCAAGAATTCAACCA
IDE	IDE-12	AACAGACAACTCCTGGAAACG	TTGCTTCTGTCCTCCTCCTC
IDE	IDE-13	ATTTACCCAGTCCCCCATGT	TTGCCCATATATTGAGCAGAGA
IDE	IDE-14	ACTCTGGGGCCTTAGTTGGT	TGGATATTGTTTATATTGTCCTGAGTC
IDE	IDE-15	AAAAGCATTTTATTCACAATTTCA	TGTGTTACAGATCTGCTTTAATAAACT
IDE	IDE-16	TGAAAATATATGCAAGATTGCTG	GCTTATAGTGCAAGCCAAACA
IDE	IDE-17	CTTTAAAAGTTTAACTGGGCCTTA	GGGACCATCTTCCACTTATCTG
IDE	IDE-18	AACAGATCCACAGCATGAAGAA	TCGCAAGAGGTTCATTCTGA
IDE	IDE-19	CAGGAGAAAATGAGAGCCAGA	TTGGGATTCTCCCATGAAAA
IDE	IDE-20	CCTCCAAAAATTTTCACCTGTT	CCTAAGATTTTAAGCACTCCCTGT
IDE	IDE-21	TTCTAACACTGTGAAAAAGCCAAC	AAAGTTCATCTTAAGCAATATCAGTG
IDE	IDE-22	CCTGAATTGTAAAATATGCCATGA	TTCCAAAAATCTTTCTAAAAGGCTA
IDE	IDE-23	TGAAGCCATCTCTGTGGAAA	TGCTAACAAGAACTACAAATCACCT
IDE	IDE-24	CATGGAACCGGTAACTATTGAG	GGGTTTACTTGACGTTATTGTTTG
IDE	IDE-25	GCGGCTTTAAGTGCTGAATC	CTGGGAAAGTCTGGGTGCT
IGF2BP2	IGF2BP2-01	GTCACACCTACCGCCTGTCT	GAAAACAGGAACCCACCAAA
IGF2BP2	IGF2BP2-02	TTTTGCTGAATATTGATGGCTTA	TCTGTGGCAAATCGTCTCTG
IGF2BP2	IGF2BP2-03	TTGATTAAAAATTGGAGTTAGTTGAG	CTGCCGTTTGTCCTCAGAA
IGF2BP2	IGF2BP2-04	CCACCAGCAGCTTAGCAGA	ACCTCTTGGCCCCTCACT
IGF2BP2	IGF2BP2-05	AGGGGAGACTGAGGGAACTC	CTTCGCAGCACTGATGAAGT
IGF2BP2	IGF2BP2-06	TTGAATGTGCCTCTGAGGTG	AAGACCTGGAACCTCCTGCT
IGF2BP2	IGF2BP2-07	TGGGACTGAGAGGGAACAAG	CTGGAGGAGAAAGCCTGGT
IGF2BP2	IGF2BP2-08	AACATTCACTGGCCACCAAG	GTGTCGCATGTTGGGAATCT
IGF2BP2	IGF2BP2-09	TTCCGTTCTGATTCCACACA	AAGACAAGAGAGGGAGGTTGA
IGF2BP2	IGF2BP2-10	AAGGCACTTCCTCATTAGCTG	CCCATCATCACTTGGTACCTG
IGF2BP2	IGF2BP2-11	TCTTGGGTTAAATGAGGGAGA	GCTGTACATGGGTTGGTGTG
IGF2BP2	IGF2BP2-12	GTGGCTGAGACCCACCAG	CGTTTTAGCCGGGATGGT
IGF2BP2	IGF2BP2-13	AGGTGGCACATATCCAGCTC	GCCAGTGTGATGAGGGTTTTA
IGF2BP2	IGF2BP2-14	TTTCTCCCTGGAGTTGACCT	AACTGTTTTTATGGAAGACCTGAGA
IGF2BP2	IGF2BP2-15	CCAGAAGCTTTGTGGGAAAC	ACGGGTTTAAGCCGGTTTTT
IGF2BP2	IGF2BP2-16	AACCACCAGGTCAAAGTTCA	AACGAGAAAGAGCTCCATGTG
IGF2BP2	IGF2BP2-17	TGATTTGCCTTTTGGTCTACAA	TCCTGATGCCCACTCTTTCT
IGF2BP2	IGF2BP2-18	TCTCTGACCAGGCTTCTTGC	TGCCATTCCTGTTCATTCTC
IGF2BP2	IGF2BP2-19	ATGTCCCAGGCTGCACTAAT	CTGTGGCCCATATGGTTTGT
IGF2BP2	IGF2BP2-20	GCCCCTTTAAAAACTACCAAGA	CTCGATCCGGGTAGGAGAC
IGF2BP2	IGF2BP2-21	GTCTGGTGCGTGGAAGTGA	GAGAGGGAGGCAGAGAGAGC

KCNJ11	KCNJ11-01	CCTGGCAAAGAAACATGACC	AACTCAGCCTTTGGGCACT
KCNJ11	KCNJ11-02	TATACCTCTCCCTGGGCACT	ACTGAGGTCCCTGAGGTGAA
KCNJ11	KCNJ11-03	GAGGGGGAGGCTTTATTGAC	CCAGTTTCAGGTCTGGAAGG
KCNJ11	KCNJ11-04	TGAGCGGGGCTCTACAAAG	TGGGTTATTTTTGTCCCTGCT
KCNJ11	KCNJ11-05	CAGTACCTCCCACAGCCTCT	GGAGGACGGACGTTACTCTG
KCNJ11	KCNJ11-06	GAGCTTCCAGTAGGCTGTGG	CCTCTGCTTCATGCTACGTG
KCNJ11	KCNJ11-07	GTCTTGCGTACCACCTGCAT	GCCACACACATTGCTCATCT
KCNJ11	KCNJ11-08	AAGGCGATGAGCCACCAG	GAAGAGTCTGGTGGGGAGTT
KCNJ11	KCNJ11-09	TAGGGCCTCACTGCAGAGTC	GTGATCTTGGCTCACTGCAA
KCNJ11	KCNJ11-10	CATGGTGAAACCCCATCTCT	TCCGGAGGTGTTAAGTTCTGA
LMNA	LMNA-01	CCCAATGGATCTGGGACTG	GGGTCGAAGGACAGAGACTG
LMNA	LMNA-02	CTACACCAGCCAACCCAGAT	CACCTCTTCAGACTCGGTGA
LMNA	LMNA-03	ACCTGCAGGAGCTCAATGAT	CCCTCTCACTCCCTTCCTG
LMNA	LMNA-04	CTGGCACTGTCTAGGCACAC	AGGACAGGTGAATGGCTCTG
LMNA	LMNA-05	TCCTTCCAGTTCTTGTGTTCTG	ACTAGGGCAAGGGACTCAGG
LMNA	LMNA-06	AGCACTCAGCTCCCAGGTTA	CTGATCCCCAGAAGGCATAG
LMNA	LMNA-07	ACCCCTCCCAGTCACCAC	CATCCGGCCCAGACTCTA
LMNA	LMNA-08	GTCCCTCCTTCCCCATACTT	CCAAGTGGGGGTCTAGTCAA
LMNA	LMNA-09	AGGTGCTGGCAGTGTCCTCT	CTCTGAGGGCAAGGATGTTC
LMNA	LMNA-10	GGGCCTTTGAGCAAGATACA	CAGCTGGCTCCGATGTTG
LMNA	LMNA-11	TCACTGGGGTAGACATGCTG	GAAAGTTCCCACTCCCTTCC
LMNA	LMNA-12	AGTGGTCAGTCCCAGACTCG	CCTACCCCTCGATGACCAG
LMNA	LMNA-13	GGGAGATGCTACCTCCCTTC	GTGGAGTTTCCTGGAAGCAG
LMNA	LMNA-14	CCAAAGAAAAATAACCCTTTGGT	GGAACCTCAGCCAGGTGTAG
LMNA	LMNA15	TCCGAGAGGGAGAGAGAGAG	TAGCGTTTGCCAAAACCACT
LMNA	LMNA-16	GGTCACTGGAAAGGGGAGAG	ATGCCCACGAGTCTGCAC
SLC30A8	SLC30A8-01	GCTCCTCTACCTCTTAGAAAGCA	AGGCCTACAATTCAGAAGCAA
SLC30A8	SLC30A8-02	TCAGTGAGTGGCTTGTTTGC	TGCAGAGTTTTGTTTGGTTGTT
SLC30A8	SLC30A8-03	TGTGAATCTAGCTGCATTCTGG	ACCAGCCTGCTGTTATTGCT
SLC30A8	SLC30A8-04	GGGGGTGTAGGTGTAGGTGA	AAAAGGTCGACTTTGTCCAGA
SLC30A8	SLC30A8-05	GCCAGCTTCTCCATTATGCT	GTTGGAGGGTCAGCTTCTGT
SLC30A8	SLC30A8-06	GACACAGGGAGGTTTACTTATTTT	TTCTAGGGAAGAGAAAACACTGA
SLC30A8	SLC30A8-07	TGGTGAGGACTTTGCAGCTT	TGACTTTACAGGACAAGGACTGA
SLC30A8	SLC30A8-08	GCGTGTGCAATCAGTGCTAA	CCAATTGATTGATGGATCTCAG
SLC30A8	SLC30A8-09	GGAACCAAAGGAAGAAATTCA	TGGCGTTAGCATACAGAGTCA
SLC30A8	SLC30A8-10	GAAATCAGTGGAAGGGACAAA	GTGGGGTCAACTCGAATGTT
SLC30A8	SLC30A8-11	ACCAGCCAAAGAATGTGGTT	GACTATGGCTCATCTTTAGAAGCA
SLC30A8	SLC30A8-12	CAGGCCAACATTCTGGAAAT	AAAATCACAGCGTGGAGAGG
SLC30A8	SLC30A8-13	GGCAAAAAGAAGACTAGGCACA	TTATTTCAGTGGTGCTGATTGA
SLC30A8	SLC30A8-14	GAGGCTATATCTGGGCCTTG	TTTTCAAAAAGAAGCAGAAGCA
SLC30A8	SLC30A8-15	TGCCAAATCACTTTTCCTGTC	TGGAGTTCTGACCACACCAG
SLC30A8	SLC30A8-16	GGGAGATTTTAAGGATTTTGAGA	TGGGTCCTAACATCACCACA
SLC30A8	SLC30A8-17	AAATCAGGTCAACTTTACCACCA	CTGTGCCCACTCATGTTTGT
SLC30A8	SLC30A8-18	AAATTAGGACACTTAAAAATGTTGGA	CATGACAGAGAAAAGCCAAGAA
SLC30A8	SLC30A8-19	AGCAGAATTGTCACTCCAAGG	AACAGATTGCCCTGTGATCC
SLC30A8	SLC30A8-20	TTGCAAAGAGACAAGCCTCA	AACATGTCCTGCCATGATTTT
TCF7L2	TCF7L2-01	CGGCATCTCATTAAGGCAGT	AGGGGAAAAGAGGGAGGAG
TCF7L2	TCF7L2-02	CGGATCCTTGGGAACGAG	CCAAGCCCAGGAAACAGATA
TCF7L2	TCF7L2-03	TCTTTTTCTCCCCCTTCTCC	CCCTTAAAGAGCCCTCCATC
TCF7L2	TCF7L2-04	AGACAAATCCCGGGAAAGTT	GCAAACAAATCCAGTCCACA
TCF7L2	TCF7L2-05	TGGGCTAGTTGTTCTGCATTT	TTTTCTTTTTCACCCCCTCA
TCF7L2	TCF7L2-06	AATCCAGGTGAGAGGCTGTG	ACACCCAGAAAAGCAACAGG

TCF7L2	TCF7L2-07	CATCCCCTAATGGATTGCTC	GCCCATGTAAGAAACAAAAAGG
TCF7L2	TCF7L2-08	CCCACCCTTGTTTCAAGTCT	ACACATGGCCTTGGATGAA
TCF7L2	TCF7L2-09	CATCCCTCCTCATTCATTCA	CGGCTACGAAGAAGGTGAGA
TCF7L2	TCF7L2-10	GTTGGACCACGACCTTGTTT	GACATTCTGCCACCTGACCT
TCF7L2	TCF7L2-11	GACCTTGGCGTAATGTGTGA	GACGCACAACACAGCTCTACA
TCF7L2	TCF7L2-12	CTTCCTCCTGCCTTCTCCTT	CTGGCCTTGAATTTCTCCTG
TCF7L2	TCF7L2-13	GGCCAGGCTTTTACAAACAA	ACCGTGTTGATGGTCATTTT
TCF7L2	TCF7L2-14	CTCCCTTTCTCCCACGTTCT	AAGGGGGAGAGGATCGAATA
TCF7L2	TCF7L2-15	TCTCTCCCTTGGCATCTGTG	AGAAGTCGACGGCTGTGC
TCF7L2	TCF7L2-16	CCCACCTGTCCATGATGC	TTGACCAATGAACTCGATAAACA
TCF7L2	TCF7L2-17	CCCACCTTGAAAGGTTTTGT	AAAATGATTGGTTCGCAAGC

Table S6.2: Variants detected with high confidence in the main targeted high throughput sequencing project. The P-value for the comparison of the 500 cases vs. 500 controls has been included or each value. The C-alpha statistic P-value for the burden of rare variants in cases vs. controls for the amplicon in which the variant is located is also included.

HG18 position with Chr number	gene	alleles	type	SNP	Base change	annotation	MAF Case	MAF Controls	C-alpha p-value for amplicon	P value-Chi square test
chr1:154351035	*LMNA*	AC	nc	N/A	A/C	NA	0.015	0.010	0.000	0.037
chr1:154366981	*LMNA*	CT	nc	N/A	C/T	NA	0.000	0.001	0.000	0.343
chr1:154367032	*LMNA*	CT	s	rs41313880	CGC119CGT	Arg119Arg	0.003	0.001	0.000	0.193
chr1:154367299	*LMNA*	CA	nc	N/A	A/C	NA	0.002	0.006	0.000	0.020
chr1:154370916	*LMNA*	GA	s	rs12117552	CTG204CTA	Leu204Leu	0.005	0.009	0.000	0.276
chr1:154370999	*LMNA*	GT	nc	rs11264442	G/T	NA	0.020	0.020	0.000	1.000
chr1:154371016	*LMNA*	CT	nc	rs11264443	C/T	NA	0.028	0.030	0.000	0.649
chr1:154371033	*LMNA*	CT	nc	N/A	C/T	NA	0.001	0.000	0.000	0.343
chr1:154371040	*LMNA*	CA	nc	N/A	A/C	NA	0.004	0.005	0.000	0.388
chr1:154371093	*LMNA*	CA	nc	N/A	A/C	NA	0.007	0.008	0.000	0.628
chr1:154371103	*LMNA*	GA	nc	N/A	A/G	NA	0.001	0.000	0.000	0.343
chr1:154371168	*LMNA*	CT	nc	rs41314033	C/T	NA	0.002	0.003	0.000	1.000
chr1:154371198	*LMNA*	GA	nc	N/A	A/G	NA	0.000	0.001	0.000	0.343
chr1:154371306	*LMNA*	GA	s	N/A	GCG242GCA	Ala242Ala	0.000	0.002	0.000	0.343
chr1:154371452	*LMNA*	GA	nc	N/A	A/G	NA	0.000	0.001	0.000	0.343
chr1:154371471	*LMNA*	GC	nc	N/A	C/G	NA	0.000	0.001	0.000	0.343
chr1:154371471	*LMNA*	GC	nc	N/A	C/G	NA	0.000	0.001	0.000	0.343
chr1:154371569	*LMNA*	AG	nc	N/A	A/G	NA	0.007	0.007	0.854	0.660
chr1:154371652	*LMNA*	TC	s	rs538089	GCT287GCC	Ala287Ala	0.053	0.047	0.854	0.596
chr1:154371769	*LMNA*	TC	nc	N/A	C/T	NA	0.007	0.007	0.854	0.628
chr1:154372552	*LMNA*	GA	nc	rs534807	A/G	NA	0.088	0.075	0.000	0.196
chr1:154372564	*LMNA*	CT	nc	N/A	C/T	NA	0.002	0.005	0.000	0.178
chr1:154372585	*LMNA*	CT	nc	1KG_1_156105961	C/T	NA	0.003	0.006	0.000	0.436
chr1:154372809	*LMNA*	TC	s	rs505058	GAT446GAC	Asp446Asp	0.059	0.051	0.000	0.299
chr1:154373272	*LMNA*	GT	nc	N/A	G/T	NA	0.003	0.006	0.097	0.081
chr1:154373279	*LMNA*	AT	nc	N/A	A/T	NA	0.000	0.004	0.097	0.037
chr1:154373408	*LMNA*	CG	ns	N/A	CCA485GCA	Pro485Ala	0.002	0.001	0.097	0.343
chr1:154373487	*LMNA*	CT	nc	rs553016	C/T	NA	0.084	0.076	0.097	0.401
chr1:154373507	*LMNA*	GA	nc	N/A	A/G	NA	0.000	0.001	0.097	0.000
chr1:154373701	*LMNA*	AG	nc	N/A	A/G	NA	0.000	0.001	0.097	0.343
chr1:154374094	*LMNA*	GA	ns	1KG_1_156107470	CGC545CAC	Arg545His	0.001	0.001	0.000	1.000

chr1:154374126	*LMNA*	GA	ns	N/A	GAG556AAG	Glu556Lys	0.000	0.001	0.000	0.343
chr1:154374158	*LMNA*	CT	s	rs4641	CAC566CAT	His566His	0.310	0.302	0.000	0.670
chr1:154374313	*LMNA*	CT	nc	N/A	C/T	NA	0.000	0.001	0.000	0.343
chr1:154374376	*LMNA*	AC	nc	N/A	A/C	NA	0.009	0.010	0.000	0.343
chr1:154374870	*LMNA*	GT	nc	N/A	G/T	NA	0.002	0.000	0.762	0.168
chr1:154374873	*LMNA*	GA	nc	N/A	A/G	NA	0.005	0.001	0.762	0.178
chr1:154374953	*LMNA*	GA	s	N/A	TCG583TCA	Ser583Ser	0.000	0.001	0.762	0.343
chr1:154374974	*LMNA*	CT	s	N/A	ACC590ACT	Thr590Thr	0.001	0.000	0.762	0.343
chr1:154375134	*LMNA*	CT	ns	N/A	CGC644TGC	Arg644Cys	0.005	0.008	0.762	0.424
chr1:154375145	*LMNA*	CT	s	N/A	CTC647CTT	Leu647Leu	0.002	0.000	0.762	0.343
chr1:154375600	*LMNA*	GC	utrsnc	rs7339	C/G	NA	0.072	0.063	0.000	0.518
chr1:154375620	*LMNA*	GC	utrsnc	N/A	C/G	NA	0.036	0.008	0.000	0.000
chr1:154375718	*LMNA*	CA	utrsnc	rs115645935	A/C	NA	0.010	0.011	0.000	0.343
chr1:154375850	*LMNA*	GD	utrsnc	N/A	G	NA	0.001	0.001	0.000	1.000
chr1:154375886	*LMNA*	CT	utrsnc	rs74116489	C/T	NA	0.023	0.021	0.000	0.777
chr1:154375905	*LMNA*	CT	utrsnc	N/A	C/T	NA	0.000	0.002	0.000	0.343
chr1:154376019	*LMNA*	AG	utrsnc	N/A	A/G	NA	0.007	0.005	0.000	0.388
chr1:154376072	*LMNA*	TC	utrsnc	N/A	C/T	NA	0.003	0.004	0.000	1.000
chr1:154376076	*LMNA*	TA	utrsnc	N/A	A/T	NA	0.004	0.004	0.000	1.000
chr1:154376335	*LMNA*	CT	utrsnc	N/A	C/T	NA	0.007	0.009	0.000	0.290
chr1:154376339	*LMNA*	AT	utrsnc	N/A	A/T	NA	0.002	0.005	0.000	0.178
chr1:154376529	*NA*	CT	nc	N/A	C/T	NA	0.001	0.000	0.000	0.343
chr10:114700046	*TCF7L2*	CT	utrsnc	N/A	C/T	NA	0.052	0.007	0.000	0.001
chr10:114700254	*TCF7L2*	AC	utrsnc	N/A	A/C	NA	0.012	0.013	0.000	0.777
chr10:114700274	*TCF7L2*	AG	utrsnc	N/A	A/G	NA	0.010	0.000	0.000	0.011
chr10:114893661	*TCF7L2*	TG	utrsnc	N/A	G/T	NA	0.002	0.000	0.548	0.343
chr10:114895869	*TCF7L2*	TG	nc	N/A	G/T	NA	0.004	0.000	0.693	0.037
chr10:114895876	*TCF7L2*	AC	nc	N/A	A/C	NA	0.006	0.002	0.693	0.178
chr10:114900883	*TCF7L2*	CT	utrsnc	N/A	C/T	NA	0.001	0.000	0.747	0.343
chr10:114901662	*TCF7L2*	GA	nc	rs3902056	A/G	NA	0.000	0.002	0.555	0.343
chr10:114901703	*TCF7L2*	GA	nc	N/A	A/G	NA	0.002	0.002	0.555	1.000
chr10:114902012	*TCF7L2*	AT	nc	N/A	A/T	NA	0.009	0.008	0.000	0.557
chr10:114902019	*TCF7L2*	AT	nc	N/A	A/T	NA	0.031	0.008	0.000	0.002
chr10:114902071	*TCF7L2*	GA	utrsnc	1KG_10_114912081	A/G	NA	0.001	0.000	0.000	0.343

chr10:114902087	*TCF7L2*	TC	s	N/A	CAT362CAC	His362His	0.001	0.000	0.000	0.343
chr10:114902201	*TCF7L2*	AG	utrsnc	N/A	A/G	NA	0.001	0.000	0.000	0.343
chr10:114907855	*TCF7L2*	AG	nc	N/A	A/G	NA	0.003	0.005	0.000	0.388
chr10:114908351	*TCF7L2*	AT	nc	N/A	A/T	NA	0.010	0.008	0.000	0.081
chr10:114908360	*TCF7L2*	CT	nc	N/A	C/T	NA	0.010	0.010	0.000	0.168
chr10:114908555	*TCF7L2*	GA	nc	N/A	A/G	NA	0.000	0.001	0.000	0.343
chr10:94203543	*IDE*	AT	utrsnc	N/A	A/T	NA	0.002	0.003	0.000	0.356
chr10:94203692	*IDE*	TC	utrsnc	N/A	C/T	NA	0.001	0.000	0.000	0.343
chr10:94204101	*IDE*	AG	utrsnc	N/A	A/G	NA	0.001	0.000	0.000	0.343
chr10:94204154	*IDE*	TC	utrsnc	N/A	C/T	NA	0.001	0.000	0.000	0.343
chr10:94204239	*IDE*	CA	ns	N/A	GGT1001GTT	Gly1001Val	0.002	0.001	0.000	0.557
chr10:94204339	*IDE*	TA	nc	rs4646958	A/T	NA	0.086	0.099	0.000	0.282
chr10:94205449	*IDE*	AI	nc	N/A	A	NA	0.011	0.012	0.000	0.660
chr10:94213381	*IDE*	CA	nc	N/A	A/C	NA	0.000	0.001	0.000	0.343
chr10:94213706	*IDE*	AG	s	rs61736443	AAT841AAC	Asn841Asn	0.016	0.011	0.000	0.427
chr10:94215640	*IDE*	CA	nc	N/A	A/C	NA	0.000	0.001	0.000	0.343
chr10:94218672	*IDE*	GA	ns	N/A	ACC755ATC	Thr755Ile	0.000	0.001	0.289	0.168
chr10:94218805	*IDE*	TC	nc	rs12243622	C/T	NA	0.001	0.001	0.289	1.000
chr10:94220188	*IDE*	TC	nc	N/A	C/T	NA	0.000	0.001	0.000	0.343
chr10:94224717	*IDE*	GA	nc	rs56035419	A/G	NA	0.001	0.000	0.720	0.343
chr10:94224735	*IDE*	TC	nc	rs12245118	C/T	NA	0.000	0.001	0.720	0.430
chr10:94224789	*IDE*	AC	nc	rs35223317	A/C	NA	0.031	0.028	0.720	1.000
chr10:94225562	*IDE*	CT	nc	N/A	C/T	NA	0.000	0.001	0.049	0.300
chr10:94225744	*IDE*	TA	utrsnc	rs72809183	A/T	NA	0.004	0.011	0.049	0.028
chr10:94229291	*IDE*	AT	nc	N/A	A/T	NA	0.005	0.000	0.000	0.015
chr10:94236972	*IDE*	CA	s	rs35426658	CCG539CCT	Pro539Pro	0.024	0.025	0.000	0.719
chr10:94240159	*IDE*	GT	nc	N/A	G/T	NA	0.004	0.010	0.000	0.005
chr10:94240165	*IDE*	TC	nc	N/A	C/T	NA	0.001	0.000	0.000	0.343
chr10:94254532	*IDE*	GA	ns	N/A	CGG477TGG	Arg477Trp	0.000	0.001	0.023	0.343
chr10:94254567	*IDE*	TG	ns	N/A	GAG465GCG	Glu465Ala	0.000	0.002	0.023	0.343
chr10:94256290	*IDE*	GA	nc	rs76303615	A/G	NA	0.001	0.001	0.000	1.000
chr10:94256325	*IDE*	TA	nc	rs34053974	A/T	NA	0.011	0.014	0.000	0.696
chr10:94257802	*IDE*	TC	nc	N/A	C/T	NA	0.000	0.001	0.841	0.168
chr10:94257828	*IDE*	AG	nc	N/A	A/G	NA	0.000	0.001	0.841	0.343

chr10:94257953	*IDE*	CG	utrsnc	N/A	C/G	NA	0.001	0.000	0.841	0.343
chr10:94257971	*IDE*	TC	nc	N/A	C/T	NA	0.001	0.001	0.841	1.000
chr10:94259949	*IDE*	CT	nc	rs61860640	C/T	NA	0.007	0.003	0.000	0.025
chr10:94259952	*IDE*	CA	nc	N/A	A/C	NA	0.004	0.009	0.000	0.020
chr10:94264606	*IDE*	CA	nc	N/A	A/C	NA	0.000	0.001	0.000	0.343
chr10:94264789	*IDE*	AG	utrsnc	rs17875327	A/G	NA	0.123	0.098	0.000	0.056
chr10:94281430	*IDE*	AG	nc	N/A	A/G	NA	0.001	0.000	0.000	0.343
chr10:94281659	*IDE*	GC	utrsnc	N/A	C/G	NA	0.000	0.001	0.000	0.343
chr10:94284268	*IDE*	GC	nc	rs56191416	C/G	NA	0.000	0.001	0.000	0.343
chr10:94284271	*IDE*	AG	nc	rs4646955	A/G	NA	0.276	0.242	0.000	0.072
chr10:94284464	*NA*	AD	NA	N/A	NA	NA	0.004	0.000	0.000	0.037
chr10:94284531	*IDE*	GA	utrsnc	N/A	A/G	NA	0.000	0.001	0.000	0.168
chr11:17363410	*KCNJ11*	TA	utrsnc	N/A	A/T	NA	0.000	0.001	0.147	0.343
chr11:17363448	*KCNJ11*	TC	utrsnc	N/A	C/T	NA	0.001	0.000	0.147	0.343
chr11:17363628	*KCNJ11*	GA	utrsnc	N/A	A/G	NA	0.001	0.000	0.147	0.343
chr11:17363670	*KCNJ11*	TG	utrsnc	N/A	G/T	NA	0.000	0.006	0.147	0.005
chr11:17363845	*KCNJ11*	CT	utrsnc	N/A	C/T	NA	0.003	0.001	0.000	0.412
chr11:17364194	*KCNJ11*	TC	utrsnc	rs5206	C/T	NA	0.001	0.001	0.000	1.000
chr11:17364276	*KCNJ11*	CT	utrsnc	1KG_11_17407700	C/T	NA	0.011	0.012	0.000	0.684
chr11:17364356	*KCNJ11*	TC	utrsnc	1KG_11_17407780	C/T	NA	0.007	0.013	0.000	0.372
chr11:17364480	*KCNJ11*	AC	utrsnc	N/A	A/C	NA	0.000	0.001	0.214	0.343
chr11:17364513	*KCNJ11*	CT	utrsnc	rs115826080	C/T	NA	0.008	0.014	0.214	0.478
chr11:17364555	*KCNJ11*	TC	utrsnc	rs79457490	C/T	NA	0.001	0.000	0.214	0.343
chr11:17364577	*KCNJ11*	GA	utrsnc	N/A	A/G	NA	0.000	0.001	0.214	0.343
chr11:17364601	*KCNJ11*	AG	utrsnc	rs2285676	A/G	NA	0.351	0.344	0.214	0.913
chr11:17364682	*KCNJ11*	CT	utrsnc	N/A	C/T	NA	0.000	0.002	0.214	0.343
chr11:17364827	*KCNJ11*	GA	utrsnc	rs5210	A/G	NA	0.384	0.392	0.214	0.862
chr11:17365098	*KCNJ11*	CT	ns	N/A	GTG286ATG	Val286Met	0.000	0.001	0.054	0.989
chr11:17365206	*KCNJ11*	CT	ns	rs5215	GTC250ATC	Val250Ile	0.647	0.691	0.054	0.034
chr11:17365253	*KCNJ11*	TC	ns	N/A	GAG234GGG	Glu234Gly	0.081	0.104	0.054	0.057
chr11:17365372	*KCNJ11*	GA	s	rs116392938	CTC194CTT	Leu194Leu	0.000	0.001	0.235	0.343
chr11:17365407	*KCNJ11*	GC	ns	rs1800467	CTG183GTG	Leu183Val	0.054	0.045	0.235	0.348
chr11:17365414	*KCNJ11*	GC	s	rs5216	CTC180CTG	Leu180Leu	0.027	0.024	0.235	1.000
chr11:17365537	*KCNJ11*	GA	s	rs74162102	CCC139CCT	Pro139Pro	0.000	0.001	0.235	0.343

chr11:17365568	*KCNJ11*	TC	ns	N/A	CAC129CGC	His129Arg	0.001	0.000	0.235	0.343
chr11:17365642	*KCNJ11*	CG	s	N/A	CTG104CTC	Leu104Leu	0.000	0.001	0.001	0.343
chr11:17365942	*KCNJ11*	CT	ns	N/A	TGG4TGA	Trp4Opl	0.002	0.000	0.001	0.343
chr11:17366148	*KCNJ11*	TC	ns	rs5219	AAG23GAG	Lys23Glu	0.550	0.564	0.427	0.674
chr12:119900846	*HNF1A*	CT	nc	N/A	C/T	NA	0.004	0.007	0.000	0.081
chr12:119901306	*HNF1A*	TC	nc	N/A	C/T	NA	0.001	0.001	0.000	1.000
chr12:119901309	*HNF1A*	CA	nc	N/A	A/C	NA	0.002	0.004	0.000	0.138
chr12:119910928	*HNF1A*	GA	nc	rs1169293	A/G	NA	0.920	0.896	0.000	0.042
chr12:119910977	*HNF1A*	GA	nc	rs1169294	A/G	NA	0.295	0.338	0.000	0.181
chr12:119915655	*HNF1A*	TA	nc	rs2071190	A/T	NA	0.364	0.377	0.002	0.532
chr12:119915683	*HNF1A*	CT	nc	rs1169301	C/T	NA	0.297	0.318	0.002	0.651
chr12:119915715	*HNF1A*	AG	ns	N/A	CAT179CGT	His179Arg	0.000	0.001	0.002	0.343
chr12:119915765	*HNF1A*	AG	ns	N/A	ACC196GCC	Thr196Ala	0.001	0.000	0.002	0.343
chr12:119916536	*HNF1A*	CT	s	N/A	CCC300CCT	Pro300Pro	0.000	0.002	0.003	0.343
chr12:119916638	*HNF1A*	TC	nc	N/A	C/T	NA	0.007	0.010	0.003	0.037
chr12:119918636	*HNF1A*	TC	nc	N/A	C/T	NA	0.024	0.011	0.000	0.005
chr12:119918685	*HNF1A*	GT	nc	rs3751156	G/T	NA	0.008	0.021	0.000	0.032
chr12:119918700	*HNF1A*	CT	nc	1KG_12_121434317	C/T	NA	0.003	0.005	0.000	0.280
chr12:119918712	*HNF1A*	GA	utrsnc	N/A	A/G	NA	0.000	0.001	0.000	0.340
chr12:119918723	*HNF1A*	GA	utrsnc	N/A	A/G	NA	0.001	0.000	0.000	0.343
chr12:119918778	*HNF1A*	CT	ns	N/A	CAC387TAC	His387Tyr	0.000	0.002	0.000	0.343
chr12:119919673	*HNF1A*	GC	s	CD052124	ACG441ACC	Thr441Thr	0.003	0.000	0.668	0.343
chr12:119919725	*HNF1A*	CT	s	rs2259820	CTG459TTG	Leu459Leu	0.284	0.321	0.668	0.229
chr12:119919810	*HNF1A*	GA	ns	rs2464196	AGC487AAC	Ser487Asn	0.294	0.341	0.668	0.063
chr12:119919858	*HNF1A*	GA	utrsnc	rs2464195	A/G	NA	0.329	0.382	0.668	0.243
chr12:119921497	*HNF1A*	GA	s	rs55834942	ACG515ACA	Thr515Thr	0.228	0.220	0.000	0.953
chr12:119921598	*HNF1A*	CT	nc	N/A	C/T	NA	0.001	0.002	0.000	0.343
chr12:119921604	*HNF1A*	TC	nc	rs1169304	C/T	NA	0.823	0.825	0.000	1.000
chr12:119923227	*HNF1A*	TC	nc	rs735396	C/T	NA	0.231	0.263	0.000	0.225
chr12:119923575	*HNF1A*	GT	utrsnc	rs1169309	G/T	NA	0.297	0.349	0.000	0.089
chr12:119923592	*HNF1A*	GA	utrsnc	1KG_12_121439209	A/G	NA	0.008	0.005	0.000	0.289
chr12:119923802	*HNF1A*	CT	utrsnc	N/A	C/T	NA	0.000	0.001	0.464	0.343
chr12:119923816	*HNF1A*	GA	utrsnc	rs1169310	A/G	NA	0.343	0.378	0.464	0.409
chr12:119923952	*HNF1A*	TA	utrsnc	N/A	A/T	NA	0.002	0.007	0.464	0.024

chr12:119923981	*HNF1A*	GA	utrsnc	rs11065390	A/G	NA	0.019	0.018	0.000	0.750
chr12:119924256	*HNF1A*	AG	utrsnc	N/A	A/G	NA	0.004	0.001	0.000	0.138
chr12:119924286	*HNF1A*	GA	utrsnc	1KG_12_121439903	A/G	NA	0.001	0.007	0.000	0.044
chr12:119924317	*HNF1A*	GA	utrsnc	N/A	A/G	NA	0.000	0.002	0.000	0.343
chr12:119924448	*HNF1A*	CT	utrsnc	N/A	C/T	NA	0.006	0.000	0.000	0.096
chr12:119924467	*HNF1A*	CA	utrsnc	N/A	A/C	NA	0.000	0.001	0.000	0.343
chr12:119924479	*HNF1A*	GT	utrsnc	N/A	G/T	NA	0.000	0.001	0.000	0.343
chr12:119924646	*HNF1A*	GA	utrsnc	rs41279096	A/G	NA	0.127	0.124	0.000	0.874
chr12:119924690	*HNF1A*	CT	utrsnc	rs56335342	C/T	NA	0.000	0.002	0.000	0.168
chr12:119924718	*HNF1A*	TC	nc	N/A	C/T	NA	0.007	0.008	0.000	1.000
chr12:119924724	*HNF1A*	AC	nc	N/A	A/C	NA	0.005	0.002	0.000	0.178
chr17:33120508	*HNF1B*	GC	nc	rs5206	C/G	NA	0.042	0.039	0.568	0.918
chr17:33120564	*HNF1B*	GC	utrsnc	rs10962	C/G	NA	0.219	0.211	0.568	0.376
chr17:33120650	*HNF1B*	GA	utrsnc	N/A	A/G	NA	0.000	0.001	0.568	0.343
chr17:33120654	*HNF1B*	GA	utrsnc	N/A	A/G	NA	0.001	0.000	0.568	0.343
chr17:33120787	*HNF1B*	CT	utrsnc	N/A	C/T	NA	0.002	0.000	0.568	0.343
chr17:33120797	*HNF1B*	TC	utrsnc	N/A	C/T	NA	0.000	0.004	0.568	0.015
chr17:33120828	*HNF1B*	AG	utrsnc	N/A	A/G	NA	0.001	0.000	0.568	0.343
chr17:33120863	*HNF1B*	GA	utrsnc	1KG_17_36046750	A/G	NA	0.000	0.003	0.568	0.037
chr17:33121000	*HNF1B*	GI	utrsnc	N/A	G	NA	0.000	0.001	0.568	0.343
chr17:33121044	*HNF1B*	GT	utrsnc	rs2688	G/T	NA	0.576	0.587	0.568	0.753
chr17:33121104	*HNF1B*	TC	utrsnc	rs1058166	C/T	NA	0.058	0.063	0.568	0.334
chr17:33121214	*HNF1B*	TA	utrsnc	rs2689	A/T	NA	0.437	0.464	0.001	0.259
chr17:33121530	*HNF1B*	AG	nc	rs3110641	A/G	NA	0.753	0.757	0.001	0.956
chr17:33121574	*HNF1B*	GA	nc	1KG_17_36047461	A/G	NA	0.000	0.002	0.001	0.081
chr17:33133135	*HNF1B*	AG	nc	N/A	A/G	NA	0.005	0.009	0.009	0.058
chr17:33133147	*HNF1B*	AI	nc	N/A	A	NA	0.075	0.040	0.009	0.015
chr17:33133166	*HNF1B*	CA	nc	N/A	A/C	NA	0.000	0.001	0.009	0.343
chr17:33133349	*HNF1B*	TC	nc	1KG_17_36059236	C/T	NA	0.001	0.000	0.009	0.343
chr17:33133359	*HNF1B*	AC	nc	rs2269842	A/C	NA	0.203	0.177	0.009	0.501
chr17:33133377	*HNF1B*	CT	nc	N/A	C/T	NA	0.000	0.002	0.009	0.343
chr17:33133395	*HNF1B*	CA	nc	N/A	A/C	NA	0.003	0.004	0.009	0.660
chr17:33135345	*HNF1B*	GT	nc	N/A	G/T	NA	0.000	0.001	0.000	0.979
chr17:33139010	*HNF1B*	AG	nc	rs2107133	A/G	NA	0.124	0.139	0.005	0.266

chr17:33139187	*HNF1B*	AG	nc	N/A	A/G	NA	0.001	0.000	0.005	0.343
chr17:33139220	*HNF1B*	GT	nc	N/A	G/T	NA	0.002	0.002	0.005	1.000
chr17:33144479	*HNF1B*	CT	nc	N/A	C/T	NA	0.000	0.001	0.000	0.343
chr17:33144521	*HNF1B*	GA	nc	N/A	A/G	NA	0.002	0.000	0.000	0.343
chr17:33167600	*HNF1B*	AG	nc	rs114017141	A/G	NA	0.000	0.001	0.000	0.343
chr17:33167814	*HNF1B*	CG	ns	N/A	GAT194CAT	Asp194His	0.000	0.001	0.000	0.343
chr17:33167903	*HNF1B*	GA	ns	N/A	TCT190TTT	Ser190Phe	0.000	0.001	0.000	0.343
chr17:33167994	*HNF1B*	AD	nc	N/A	A	NA	0.011	0.007	0.000	0.054
chr17:33173747	*HNF1B*	GA	utrsnc	N/A	A/G	NA	0.000	0.001	0.000	0.343
chr17:33173762	*HNF1B*	GA	nc	rs59527848	A/G	NA	0.000	0.001	0.000	0.343
chr20:42463574	*HNF4A*	GA	nc	rs113725562	A/G	NA	0.062	0.076	0.000	0.028
chr20:42463597	*HNF4A*	CA	nc	N/A	A/C	NA	0.005	0.004	0.000	0.356
chr20:42468057	*HNF4A*	AT	nc	N/A	A/T	NA	0.007	0.009	0.814	0.290
chr20:42468065	*HNF4A*	GT	nc	N/A	G/T	NA	0.009	0.008	0.814	0.557
chr20:42468074	*HNF4A*	TC	nc	rs736824	C/T	NA	0.430	0.436	0.814	0.414
chr20:42468107	*HNF4A*	CT	utrsnc	rs745975	C/T	NA	0.234	0.263	0.814	0.190
chr20:42468146	*HNF4A*	GA	s	rs41282026	GCG50GCA	Ala50Ala	0.000	0.003	0.814	0.193
chr20:42468197	*HNF4A*	CT	s	rs736823	GCC67GCT	Ala67Ala	0.092	0.103	0.814	0.436
chr20:42468200	*HNF4A*	GA	s	N/A	ACG68ACA	Thr68Thr	0.001	0.000	0.814	0.343
chr20:42468303	*HNF4A*	TI	nc	N/A	T	NA	0.002	0.002	0.814	1.000
chr20:42468313	*HNF4A*	GA	nc	N/A	A/G	NA	0.001	0.001	0.814	1.000
chr20:42469379	*HNF4A*	GT	nc	N/A	G/T	NA	0.004	0.007	0.010	0.196
chr20:42475768	*HNF4A*	CT	ns	CM971463	CGG136TGG	Arg136Trp	0.002	0.000	0.470	0.168
chr20:42475778	*HNF4A*	CT	ns	CM004479	ACT139ATT	Thr139Ile	0.051	0.039	0.470	0.150
chr20:42475781	*HNF4A*	GA	ns	N/A	CGA140CAA	Arg140Gln	0.001	0.000	0.470	0.343
chr20:42475860	*HNF4A*	GA	utrsnc	1KG_20_43042446	A/G	NA	0.003	0.001	0.470	0.290
chr20:42476465	*HNF4A*	CG	nc	rs3212194	C/G	NA	0.202	0.179	0.000	0.527
chr20:42476509	*HNF4A*	GA	nc	rs3212195	A/G	NA	0.268	0.239	0.000	0.260
chr20:42476512	*HNF4A*	CT	nc	N/A	C/T	NA	0.002	0.000	0.000	0.343
chr20:42476598	*HNF4A*	GA	ns	N/A	CGG177CAG	Arg177Gln	0.001	0.000	0.000	0.343
chr20:42480432	*HNF4A*	CT	nc	N/A	C/T	NA	0.010	0.012	0.820	0.168
chr20:42480498	*HNF4A*	AG	ns	N/A	CAT223CGT	His223Arg	0.001	0.000	0.820	0.343
chr20:42481754	*HNF4A*	CD	nc	N/A	C	NA	0.002	0.001	0.363	0.662
chr20:42481830	*HNF4A*	GA	s	N/A	GTG264GTA	Val264Val	0.001	0.003	0.363	0.557

chr20:42486001	*HNF4A*	CT	nc	N/A	C/T	NA	0.004	0.005	0.000	0.673
chr20:42486170	*HNF4A*	CT	ns	CM064051	CGT331TGT	Arg331Cys	0.001	0.000	0.000	0.343
chr20:42486183	*HNF4A*	GT	ns	N/A	GGA335GTA	Gly335Val	0.001	0.000	0.000	0.343
chr20:42486355	*HNF4A*	AC	s	N/A	CCA392CCC	Pro392Pro	0.009	0.029	0.000	0.000
chr20:42486589	*HNF4A*	GD	utrsnc	N/A	G	NA	0.003	0.002	0.000	0.628
chr20:42486671	*HNF4A*	TC	utrsnc	N/A	C/T	NA	0.000	0.001	0.000	0.168
chr20:42486682	*HNF4A*	CT	utrsnc	N/A	C/T	NA	0.000	0.001	0.000	0.343
chr20:42490462	*HNF4A*	CT	s	rs61737145	AAC401AAT	Asn401Asn	0.006	0.012	0.316	0.020
chr20:42490501	*HNF4A*	CT	s	N/A	AAC414AAT	Asn414Asn	0.001	0.001	0.316	1.000
chr20:42491631	*HNF4A*	AG	ns	N/A	TAT446TGT	Tyr446Cys	0.001	0.000	0.003	0.343
chr20:42491874	*HNF4A*	GA	utrsnc	rs41280258	A/G	NA	0.009	0.007	0.003	0.833
chr20:42491896	*HNF4A*	GC	utrsnc	N/A	C/G	NA	0.012	0.027	0.003	0.004
chr20:42491996	*HNF4A*	TC	utrsnc	1KG_20_43058582	C/T	NA	0.001	0.001	0.035	1.000
chr20:42492111	*HNF4A*	TG	utrsnc	rs11086926	G/T	NA	0.066	0.059	0.035	0.565
chr20:42492240	*HNF4A*	GA	utrsnc	rs111466682	A/G	NA	0.001	0.002	0.035	0.662
chr20:42492240	*HNF4A*	GA	utrsnc	rs111466682	A/G	NA	0.001	0.002	0.035	0.662
chr20:42492287	*HNF4A*	GA	utrsnc	1KG_20_43058873	A/G	NA	0.001	0.002	0.035	0.662
chr20:42492625	*HNF4A*	AC	utrsnc	rs3212210	A/C	NA	0.681	0.657	0.005	0.374
chr20:42492732	*HNF4A*	GA	utrsnc	N/A	A/G	NA	0.001	0.000	0.000	0.343
chr20:42492790	*HNF4A*	CG	utrsnc	N/A	C/G	NA	0.001	0.001	0.000	1.000
chr20:42492797	*HNF4A*	GA	utrsnc	N/A	A/G	NA	0.003	0.003	0.000	0.697
chr20:42492846	*HNF4A*	TC	utrsnc	N/A	C/T	NA	0.006	0.002	0.000	0.268
chr20:42492851	*HNF4A*	CT	utrsnc	rs6130615	C/T	NA	0.119	0.137	0.000	0.297
chr20:42492985	*HNF4A*	AC	utrsnc	N/A	A/C	NA	0.041	0.013	0.000	0.002
chr20:42493173	*HNF4A*	TA	utrsnc	N/A	A/T	NA	0.000	0.001	0.000	0.343
chr20:42493406	*HNF4A*	CT	utrsnc	N/A	C/T	NA	0.001	0.000	0.000	0.343
chr3:186844254	*IGF2BP2*	TG	utrsnc	N/A	G/T	NA	0.002	0.000	0.500	0.343
chr3:186844901	*IGF2BP2*	TC	utrsnc	N/A	C/T	NA	0.001	0.000	0.500	0.343
chr3:186845492	*IGF2BP2*	AC	utrsnc	N/A	A/C	NA	0.152	0.027	0.010	0.010
chr3:186845910	*IGF2BP2*	GA	utrsnc	N/A	A/G	NA	0.004	0.002	0.000	0.696
chr3:186847452	*IGF2BP2*	AT	nc	N/A	A/T	NA	0.005	0.008	0.000	0.196
chr3:186850024	*IGF2BP2*	AG	utrsnc	N/A	A/G	NA	0.000	0.001	0.000	0.343
chr3:186852509	*IGF2BP2*	AT	nc	N/A	A/T	NA	0.002	0.006	0.000	0.058
chr3:186872950	*IGF2BP2*	CT	nc	N/A	C/T	NA	0.005	0.005	0.730	0.178

chr3:186872956	*IGF2BP2*	GA	nc	N/A	A/G	NA	0.008	0.002	0.730	0.242
chr3:186873042	*IGF2BP2*	GA	ns	N/A	GCC394GTC	Ala394Val	0.001	0.002	0.730	0.343
chr3:186876064	*IGF2BP2*	TC	nc	rs41268601	C/T	NA	0.000	0.001	0.010	0.343
chr3:186876175	*IGF2BP2*	TG	nc	1KG_3_185393481	G/T	NA	0.000	0.002	0.010	0.168
chr3:186876436	*IGF2BP2*	TI	nc	N/A	T	NA	0.001	0.000	0.010	0.343
chr3:186876480	*IGF2BP2*	AC	nc	N/A	A/C	NA	0.000	0.002	0.010	0.168
chr3:186876487	*IGF2BP2*	GA	nc	N/A	A/G	NA	0.003	0.000	0.010	0.193
chr3:186893074	*IGF2BP2*	CT	nc	N/A	C/T	NA	0.001	0.000	0.000	0.343
chr3:186893120	*IGF2BP2*	CG	nc	1KG_3_185410426	C/G	NA	0.000	0.001	0.000	0.343
chr3:186893125	*IGF2BP2*	GC	nc	N/A	C/G	NA	0.001	0.000	0.000	0.343
chr3:186893135	*IGF2BP2*	AG	nc	1KG_3_185410441	A/G	NA	0.010	0.004	0.000	0.102
chr3:186893293	*IGF2BP2*	CG	nc	N/A	C/G	NA	0.001	0.000	0.000	0.343
chr3:186897246	*IGF2BP2*	GC	nc	rs56392101	C/G	NA	0.054	0.070	0.000	0.105
chr3:186897282	*IGF2BP2*	CA	nc	N/A	A/C	NA	0.004	0.010	0.000	0.005
chr3:186898734	*IGF2BP2*	AG	nc	N/A	A/G	NA	0.000	0.002	0.000	0.343
chr3:186913688	*IGF2BP2*	AT	nc	N/A	A/T	NA	0.002	0.003	0.309	0.628
chr3:186913756	*IGF2BP2*	TC	nc	N/A	C/T	NA	0.000	0.001	0.309	0.343
chr3:186913866	*IGF2BP2*	CT	nc	rs55701287	C/T	NA	0.098	0.101	0.309	0.870
chr3:186913891	*IGF2BP2*	CT	nc	rs80061571	C/T	NA	0.078	0.091	0.309	0.291
chr3:186913977	*IGF2BP2*	AG	nc	rs17826758	A/G	NA	0.193	0.168	0.309	0.189
chr3:186916901	*IGF2BP2*	CT	nc	N/A	C/T	NA	0.000	0.001	0.000	0.343
chr3:186917005	*IGF2BP2*	TA	nc	N/A	A/T	NA	0.002	0.000	0.000	0.343
chr3:186917137	*IGF2BP2*	GA	nc	rs77587335	A/G	NA	0.064	0.080	0.000	0.241
chr3:186917624	*IGF2BP2*	GT	nc	N/A	G/T	NA	0.002	0.000	0.000	0.343
chr3:186917877	*IGF2BP2*	TC	nc	N/A	C/T	NA	0.000	0.001	0.559	0.343
chr3:186918000	*IGF2BP2*	GC	nc	N/A	C/G	NA	0.000	0.002	0.559	0.343
chr3:186918113	*IGF2BP2*	GA	nc	N/A	A/G	NA	0.000	0.001	0.559	0.343
chr3:186918270	*IGF2BP2*	TC	nc	N/A	C/T	NA	0.037	0.041	0.559	0.440
chr3:186918270	*IGF2BP2*	TC	nc	N/A	C/T	NA	0.037	0.041	0.559	0.440
chr3:186918404	*IGF2BP2*	TC	nc	N/A	C/T	NA	0.037	0.041	0.559	0.556
chr3:186918404	*IGF2BP2*	TC	nc	N/A	C/T	NA	0.037	0.041	0.559	0.556
chr6:20642630	*CDKAL1*	AT	nc	N/A	A/T	NA	0.000	0.002	0.065	0.081
chr6:20642655	*CDKAL1*	CA	utrsnc	N/A	A/C	NA	0.000	0.001	0.065	0.343
chr6:20642706	*CDKAL1*	AC	utrsnc	N/A	A/C	NA	0.001	0.000	0.065	0.343

chr6:20642718	*CDKAL1*	TC	utrsnc	N/A	C/T	NA	0.001	0.002	0.065	0.290
chr6:20642765	*CDKAL1*	TA	utrsnc	1KG_6_20534786	A/T	NA	0.005	0.006	0.065	0.839
chr6:20642787	*CDKAL1*	AG	utrsnc	rs41271303	A/G	NA	0.010	0.000	0.065	0.008
chr6:20643689	*CDKAL1*	TG	nc	rs75615915	G/T	NA	0.027	0.017	0.000	0.208
chr6:20654506	*CDKAL1*	GA	nc	rs9465799	A/G	NA	0.000	0.001	0.851	0.343
chr6:20654676	*CDKAL1*	GA	ns	rs111739077	CGA39CAA	Arg39Gln	0.012	0.011	0.851	1.000
chr6:20654682	*CDKAL1*	GA	ns	N/A	CGA41CAA	Arg41Gln	0.001	0.000	0.851	0.343
chr6:20654777	*CDKAL1*	CT	nc	N/A	C/T	NA	0.008	0.009	0.851	0.557
chr6:20654867	*CDKAL1*	TG	nc	rs9465800	G/T	NA	0.106	0.086	0.851	0.033
chr6:20655063	*CDKAL1*	CT	nc	N/A	C/T	NA	0.002	0.000	0.851	0.343
chr6:20655101	*CDKAL1*	GA	nc	N/A	A/G	NA	0.001	0.001	0.851	1.000
chr6:20656443	*CDKAL1*	AT	nc	N/A	A/T	NA	0.005	0.000	0.000	0.015
chr6:20656466	*CDKAL1*	TC	nc	rs7773488	C/T	NA	0.026	0.011	0.000	0.055
chr6:20656531	*CDKAL1*	CT	nc	1KG_6_20548552	C/T	NA	0.001	0.001	0.000	1.000
chr6:20656742	*CDKAL1*	CA	nc	rs115391349	A/C	NA	0.012	0.010	0.000	0.591
chr6:20656989	*CDKAL1*	GA	nc	N/A	A/G	NA	0.000	0.001	0.000	0.343
chr6:20757531	*CDKAL1*	CT	s	rs9465854	CTC105CTT	Leu105Leu	0.000	0.001	0.653	0.343
chr6:20847710	*CDKAL1*	CT	nc	rs41272361	C/T	NA	0.051	0.068	nan	0.183
chr6:20847886	*CDKAL1*	CT	nc	rs4710942	C/T	NA	0.242	0.216	nan	0.231
chr6:20866598	*CDKAL1*	TC	nc	N/A	C/T	NA	0.000	0.001	0.139	0.343
chr6:20866612	*CDKAL1*	AG	nc	rs2819999	A/G	NA	0.942	0.934	0.139	0.551
chr6:20866627	*CDKAL1*	CT	nc	1KG_6_20758648	C/T	NA	0.000	0.001	0.139	0.343
chr6:20866650	*CDKAL1*	AG	nc	rs77156269	A/G	NA	0.018	0.011	0.139	0.601
chr6:20866683	*CDKAL1*	AG	nc	rs2820000	A/G	NA	0.446	0.442	0.139	1.000
chr6:20866739	*CDKAL1*	AG	nc	rs10946406	A/G	NA	0.407	0.415	0.139	0.769
chr6:20866742	*CDKAL1*	GT	nc	N/A	G/T	NA	0.000	0.001	0.139	0.343
chr6:20866922	*CDKAL1*	GT	nc	rs2820001	G/T	NA	0.875	0.875	0.139	0.674
chr6:20866986	*CDKAL1*	CA	nc	N/A	A/C	NA	0.002	0.004	0.139	0.628
chr6:20866993	*CDKAL1*	TC	nc	N/A	C/T	NA	0.005	0.010	0.139	0.015
chr6:20866996	*CDKAL1*	TA	nc	N/A	A/T	NA	0.001	0.004	0.139	0.290
chr6:20867002	*CDKAL1*	TC	nc	N/A	C/T	NA	0.002	0.005	0.139	0.178
chr6:20867002	*CDKAL1*	TC	nc	N/A	C/T	NA	0.002	0.005	0.139	0.178
chr6:20889473	*CDKAL1*	CT	s	N/A	ACC212ACT	Thr212Thr	0.000	0.002	0.000	0.343
chr6:21063842	*CDKAL1*	TC	nc	N/A	C/T	NA	0.002	0.007	0.000	0.024

chr6:21108530	*CDKAL1*	AG	ns	N/A	AGA335GGA	Arg335Gly	0.001	0.000	0.000	0.343
chr6:21173181	*CDKAL1*	GC	nc	N/A	C/G	NA	0.000	0.002	0.000	0.343
chr6:21173305	*CDKAL1*	GA	ns	N/A	GGA368GAA	Gly368Glu	0.000	0.001	0.000	0.343
chr6:21216482	*CDKAL1*	GT	nc	N/A	G/T	NA	0.004	0.007	0.000	0.196
chr6:21216528	*CDKAL1*	TC	nc	N/A	C/T	NA	0.002	0.002	0.000	1.000
chr6:21216597	*CDKAL1*	GT	utrsnc	N/A	G/T	NA	0.005	0.006	0.000	0.388
chr6:21216603	*CDKAL1*	TC	utrsnc	N/A	C/T	NA	0.001	0.003	0.000	0.290
chr6:21306180	*CDKAL1*	AG	nc	N/A	A/G	NA	0.000	0.001	0.000	0.343
chr6:21309244	*CDKAL1*	TC	nc	rs35662535	C/T	NA	0.012	0.029	0.000	0.013
chr6:21309472	*CDKAL1*	GA	s	rs9465994	TCG512TCA	Ser512Ser	0.489	0.472	0.000	0.243
chr8:118253964	*SLC30A8*	CT	ns	CM072050	CGG325TGG	Arg325Trp	0.329	0.317	0.013	0.177
chr8:118254156	*SLC30A8*	TC	utrsnc	N/A	C/T	NA	0.002	0.000	0.013	0.343
chr8:118254162	*SLC30A8*	GA	utrsnc	1KG_8_118184981	A/G	NA	0.005	0.004	0.013	0.723
chr8:118254206	*SLC30A8*	GA	utrsnc	rs3802177	A/G	NA	0.265	0.258	0.013	0.294
chr8:118254222	*SLC30A8*	CT	utrsnc	rs2466295	C/T	NA	0.624	0.633	0.013	0.808
chr8:118254244	*SLC30A8*	CG	utrsnc	rs2466294	C/G	NA	0.454	0.467	0.013	0.589
chr8:118254489	*SLC30A8*	AC	utrsnc	rs113040697	A/C	NA	0.023	0.022	0.203	0.877
chr8:118254547	*SLC30A8*	AG	utrsnc	N/A	A/G	NA	0.002	0.000	0.203	0.343
chr8:118254614	*SLC30A8*	TD	utrsnc	N/A	T	NA	0.009	0.004	0.632	0.005
chr8:118254751	*SLC30A8*	AT	utrsnc	N/A	A/T	NA	0.001	0.002	0.632	0.557
chr8:118254799	*SLC30A8*	CI	utrsnc	N/A	C	NA	0.012	0.019	0.632	0.526
chr8:118254877	*SLC30A8*	AG	utrsnc	N/A	A/G	NA	0.020	0.022	0.632	0.674
chr8:118254914	*SLC30A8*	AG	utrsnc	rs11558471	A/G	NA	0.303	0.291	0.608	0.237
chr8:118254953	*SLC30A8*	CT	utrsnc	N/A	C/T	NA	0.000	0.001	0.608	0.343
chr8:118255119	*SLC30A8*	AG	utrsnc	rs2466293	A/G	NA	0.413	0.419	0.608	0.654
chr8:118255333	*SLC30A8*	AG	utrsnc	N/A	A/G	NA	0.005	0.003	0.000	0.268
chr8:118255403	*SLC30A8*	GA	utrsnc	1KG_8_118186222	A/G	NA	0.001	0.000	0.000	0.343
chr8:118255426	*SLC30A8*	AC	utrsnc	N/A	A/C	NA	0.021	0.064	0.000	0.000
chr8:118255507	*SLC30A8*	AG	utrsnc	rs114965363	A/G	NA	0.001	0.000	0.000	0.343
chr8:118255561	*SLC30A8*	GA	utrsnc	N/A	A/G	NA	0.000	0.001	0.000	0.343
chr8:118255677	*SLC30A8*	TC	utrsnc	N/A	C/T	NA	0.010	0.008	0.000	0.168
chr8:118255694	*SLC30A8*	CG	utrsnc	N/A	C/G	NA	0.001	0.000	0.000	0.343
chr8:118255973	*SLC30A8*	AC	utrsnc	N/A	A/C	NA	0.020	0.070	0.000	0.166
chr8:118255989	*SLC30A8*	CG	utrsnc	rs75043555	C/G	NA	0.067	0.051	0.000	0.277

chr8:118256163	*SLC30A8*	TG	utrsnc	N/A	G/T	NA	0.000	0.001	0.000	0.343
chr8:118256213	*SLC30A8*	CT	utrsnc	N/A	C/T	NA	0.000	0.001	0.000	0.343
chr8:118256269	*SLC30A8*	AG	utrsnc	N/A	A/G	NA	0.000	0.002	0.000	0.343
chr8:118256317	*SLC30A8*	AC	utrsnc	N/A	A/C	NA	0.025	0.028	0.000	0.490
chr8:118256590	*SLC30A8*	TC	utrsnc	rs117009846	C/T	NA	0.001	0.001	0.000	1.000
chr8:118256700	*SLC30A8*	GC	utrsnc	N/A	C/G	NA	0.059	0.034	0.000	0.136
chr8:118256805	*SLC30A8*	TC	utrsnc	N/A	C/T	NA	0.079	0.076	0.000	0.839
chr8:118257007	*SLC30A8*	GA	utrsnc	rs10282940	A/G	NA	0.107	0.101	0.000	0.675
chr8:118257014	*SLC30A8*	CT	utrsnc	N/A	C/T	NA	0.001	0.000	0.000	0.343
chr8:118257040	*SLC30A8*	TC	utrsnc	N/A	C/T	NA	0.030	0.020	0.000	0.032
chr8:118257374	*SLC30A8*	GC	utrsnc	N/A	C/G	NA	0.019	0.041	0.000	0.000
chr8:118257584	*SLC30A8*	AD	utrsnc	N/A	A	NA	0.007	0.007	0.000	0.308
chr8:118257851	*SLC30A8*	TC	utrsnc	N/A	C/T	NA	0.090	0.124	0.000	0.056
chr8:118257921	*SLC30A8*	TG	utrsnc	N/A	G/T	NA	0.001	0.000	0.000	0.343
chr8:118257978	*SLC30A8*	GA	utrsnc	N/A	A/G	NA	0.000	0.001	0.000	0.343
chr8:118258106	*SLC30A8*	TG	utrsnc	N/A	G/T	NA	0.000	0.001	0.000	0.343

s: synonymous variant
utrsnc: Untranslated region- noncoding
nc: non-coding/ Intronic
NA: Not Applicable

References

't Hart, L. M., T. Hansen, I. Rietveld, J. M. Dekker, G. Nijpels, G. M. Janssen, P. A. Arp, A. G. Uitterlinden, T. Jorgensen, K. Borch-Johnsen, H. A. Pols, O. Pedersen, C. M. van Duijn, R. J. Heine and J. A. Maassen (2005). "Evidence that the mitochondrial leucyl tRNA synthetase (LARS2) gene represents a novel type 2 diabetes susceptibility gene." Diabetes **54**(6): 1892-1895.

Aerts, L., K. Holemans and F. A. Van Assche (1990). "Maternal diabetes during pregnancy: consequences for the offspring." Diabetes/metabolism reviews **6**(3): 147-167.

Aguilar-Bryan, L., C. G. Nichols, S. W. Weschsler, J. P. C. IV, A. E. B. III, G. Gonzalez, H. Herrera-Sosa, K. Nguy, J. Bryan and D. A. Nelson (1995). "Cloning of the beta cell high affinity sulphonylurea receptor: a regulator of insulin secretion." Science **268**(5209): 423-426.

Alberti, G., P. Zimmet, J. Shaw, Z. Bloomgarden, F. Kaufman and M. Silink (2004). "Type 2 diabetes in the young: the evolving epidemic: the international diabetes federation consensus workshop." Diabetes Care **27**(7): 1798-1811.

Alcolado, J. C., K. Laji and R. Gill-Randall (2002). "Maternal transmission of diabetes." Diabetic medicine : a journal of the British Diabetic Association **19**(2): 89-98.

Allen, H. L., S. E. Flanagan, C. Shaw-Smith, E. De Franco, I. Akerman, R. Caswell, J. Ferrer, A. T. Hattersley and S. Ellard (2012). "GATA6 haploinsufficiency causes pancreatic agenesis in humans." Nat Genet **44**(1): 20-22.

Almgren, P., M. Lehtovirta, B. Isomaa, L. Sarelin, M. R. Taskinen, V. Lyssenko, T. Tuomi and L. Groop (2011). "Heritability and familiality of type 2 diabetes and related quantitative traits in the Botnia Study." Diabetologia **54**(11): 2811-2819.

Altshuler, D., J. N. Hirschhorn, M. Klannemark, C. M. Lindgren, M.-C. Vohl, J. Nemesh, C. R. Lane, F. Schaffner, S. Bolk, C. Brewer, T. Tuomi, D. Gaudet, T. J. Hudson, M. Daly, L. Groop and E. S. Lander (2000). "The common PPARg Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes." Nature Genetics **26**: 76-80.

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**(1): 76-80.

American Diabetes Association (2007). "Diagnosis and classification of diabetes mellitus." Diabetes Care **30 Suppl 1**: S42-47.

Arnaudo, E., M. Hirano, R. S. Seelan, A. Milatovich, C. L. Hsieh, G. M. Fabrizi, L. I. Grossman, U. Francke and E. A. Schon (1992). "Tissue-specific expression and chromosome assignment of genes specifying two isoforms of subunit VIIa of human cytochrome c oxidase." Gene **119**(2): 299-305.

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." The New England journal of medicine **355**(5): 456-466.

Bach, I. and M. Yaniv (1993). "More potent transcriptional activators or a transdominant inhibitor of the HNF1 homeoprotein family are generated by alternative RNA processing." Embo J **12**(11): 4229-4242.

Bajpeyi, S., M. Pasarica, C. Moro, K. Conley, S. Jubrias, O. Sereda, D. H. Burk, Z. Zhang, A. Gupta, L. Kjems and S. R. Smith (2011). "Skeletal muscle mitochondrial capacity and insulin resistance in type 2 diabetes." The Journal of clinical endocrinology and metabolism **96**(4): 1160-1168.

Barker, D. J. (1995). "Fetal origins of coronary heart disease." Bmj **311**(6998): 171-174.

Barker, D. J. (2004). "The developmental origins of adult disease." Journal of the American College of Nutrition **23**(6 Suppl): 588S-595S.

Barker, D. J., C. N. Hales, C. H. Fall, C. Osmond, K. Phipps and P. M. Clark (1993). "Type 2 (non-insulin-dependent) diabetes mellitus, hypertension and hyperlipidaemia (syndrome X): relation to reduced fetal growth." Diabetologia **36**(1): 62-67.

Barnett, A. H., C. Eff, R. D. Leslie and D. A. Pyke (1981). "Diabetes in identical twins. A study of 200 pairs." Diabetologia **20**(2): 87-93.

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**(6764): 880-883.

Bartoov-Shifman, R., R. Hertz, H. Wang, C. B. Wollheim, J. Bar-Tana and M. D. Walker (2002). "Activation of the insulin gene promoter through a direct effect of hepatocyte nuclear factor 4 alpha." J Biol Chem **277**(29): 25914-25919.

Bell, J. T., N. J. Timpson, N. W. Rayner, E. Zeggini, T. M. Frayling, A. T. Hattersley, A. P. Morris and M. I. McCarthy (2011). "Genome-wide association scan allowing for epistasis in type 2 diabetes." Annals of human genetics **75**(1): 10-19.

Bellanne-Chantelot, C., C. Carette, J. P. Riveline, R. Valero, J. F. Gautier, E. Larger, Y. Reznik, P. H. Ducluzeau, A. Sola, A. Hartemann-Heurtier, P. Lecomte, L. Chaillous, M. Laloi-Michelin, J. M. Wilhem, P. Cuny, F. Duron, B. Guerci, N. Jeandidier, H. Mosnier-Pudar, M. Assayag, D. Dubois-Laforgue, G. Velho and J. Timsit (2008). "The type and the position of HNF1A mutation modulate age at diagnosis of diabetes in patients with maturity-onset diabetes of the young (MODY)-3." Diabetes **57**(2): 503-508.

Bellanne-Chantelot, C., D. Chauveau, J. F. Gautier, D. Dubois-Laforgue, S. Clauin, S. Beaufils, J. M. Wilhelm, C. Boitard, L. H. Noel, G. Velho and J. Timsit (2004). "Clinical spectrum associated with hepatocyte nuclear factor-1beta mutations." Annals of internal medicine **140**(7): 510-517.

Bersaglieri, T., P. C. Sabeti, N. Patterson, T. Vanderploeg, S. F. Schaffner, J. A. Drake, M. Rhodes, D. E. Reich and J. N. Hirschhorn (2004). "Genetic signatures of strong recent positive selection at the lactase gene." Am J Hum Genet **74**(6): 1111-1120.

Bingley, P. J., E. Bonifacio, A. J. Williams, S. Genovese, G. F. Bottazzo and E. A. Gale (1997). "Prediction of IDDM in the general population: strategies based on combinations of autoantibody markers." Diabetes **46**(11): 1701-1710.

Bird, A. (2007). "Perceptions of epigenetics." Nature **447**(7143): 396-398.

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. 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, M. Xu, J. N. Haidar, Y. Yu, 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**(7146): 799-816.

Bonnefond, A., N. Clement, K. Fawcett, L. Yengo, E. Vaillant, J. L. Guillaume, A. Dechaume, F. Payne, R. Roussel, S. Czernichow, S. Hercberg, S. Hadjadj, B. Balkau, M. Marre, O. Lantieri, C. Langenberg, N. Bouatia-Naji, G. Charpentier, M. Vaxillaire, G. Rocheleau, N. J. Wareham, R. Sladek, M. I. McCarthy, C. Dina, I. Barroso, R. Jockers and P. Froguel (2012). "Rare MTNR1B variants impairing melatonin receptor 1B function contribute to type 2 diabetes." Nature Genetics.

Bowden, D. W., M. Sale, T. D. Howard, A. Qadri, B. J. Spray, C. B. Rothschild, G. Akots, S. S. Rich and B. I. Freedman (1997). "Linkage of genetic markers on human chromosomes 20 and 12 to NIDDM in Caucasian sib pairs with a history of diabetic nephropathy." Diabetes **46**(5): 882-886.

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**(1): 102-110.

Calvo, S. E., E. J. Tucker, A. G. Compton, D. M. Kirby, G. Crawford, N. P. Burtt, M. Rivas, C. Guiducci, D. L. Bruno, O. A. Goldberger, M. C. Redman, E. Wiltshire, C. J. Wilson, D. Altshuler, S. B. Gabriel, M. J. Daly, D. R. Thorburn and V. K. Mootha (2010). "High-throughput, pooled sequencing identifies mutations in NUBPL and FOXRED1 in human complex I deficiency." Nature genetics **42**(10): 851-858.

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**(1): 109-112.

Cho, Y. S., C. H. Chen, C. Hu, J. Long, R. T. Ong, X. Sim, F. Takeuchi, Y. Wu, M. J. Go, T. Yamauchi, Y. C. Chang, S. H. Kwak, R. C. Ma, K. Yamamoto, L. S. Adair, T. Aung, Q. Cai, L. C. Chang, Y. T. Chen, Y. Gao, F. B. Hu, H. L. Kim, S. Kim, Y. J. Kim, J. J. Lee, N. R. Lee, Y. Li, J. J. Liu, W. Lu, J. Nakamura, E. Nakashima, D. P. Ng, W. T. Tay, F. J. Tsai, T. Y. Wong, M. Yokota, W. Zheng, R. Zhang, C. Wang, W. Y. So, K. Ohnaka, H. Ikegami, K. Hara, Y. M. Cho, N. H. Cho, T. J. Chang, Y. Bao, A. K. Hedman, A. P. Morris, M. I. McCarthy, R. Takayanagi, K. S. Park, W. Jia, L. M. Chuang, J. C. Chan, S. Maeda, T. Kadowaki, J. Y. Lee, J. Y. Wu, Y. Y. Teo, E. S. Tai, X. O. Shu, K. L. Mohlke, N. Kato, B. G. Han and M. Seielstad (2012). "Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians." Nature Genetics **44**(1): 67-72.

Clement, K., S. Hercberg, B. Passinge, P. Galan, M. Varroud-Vial, A. R. Shuldiner, B. A. Beamer, G. Charpentier, B. Guy-Grand, P. Froguel and C. Vaisse (2000). "The Pro115Gln and Pro12Ala PPAR gamma gene mutations in obesity and type 2

diabetes." International journal of obesity and related metabolic disorders : journal of the International Association for the Study of Obesity **24**(3): 391-393.
Cockburn, B. N., D. M. Ostrega, J. Sturis, C. Kubstrup, K. S. Polonsky and G. I. Bell (1997). "Changes in pancreatic islet glucokinase and hexokinase activities with increasing age, obesity, and the onset of diabetes." Diabetes **46**(9): 1434-1439.
Codner, E., S. E. Flanagan, F. Ugarte, H. Garcia, T. Vidal, S. Ellard and A. T. Hattersley (2007). "Sulfonylurea treatment in young children with neonatal diabetes: dealing with hyperglycemia, hypoglycemia, and sick days." Diabetes care **30**(5): e28-29.
Coelho, M., D. Luiselli, G. Bertorelle, A. I. Lopes, S. Seixas, G. Destro-Bisol and J. Rocha (2005). "Microsatellite variation and evolution of human lactase persistence." Hum Genet **117**(4): 329-339.
Cohen, J. C., E. Boerwinkle, T. H. Mosley, Jr. and H. H. Hobbs (2006). "Sequence variations in PCSK9, low LDL, and protection against coronary heart disease." N Engl J Med **354**(12): 1264-1272.
Collins, F. S., M. S. Guyer and A. Charkravarti (1997). "Variations on a theme: cataloging human DNA sequence variation." Science **278**(5343): 1580-1581.
Collins, F. S., A. Patrinos, E. Jordan, A. Chakravarti, R. Gesteland and L. Walters (1998). "New goals for the U.S. Human Genome Project: 1998-2003." Science **282**(5389): 682-689.
Connolly, V., N. Unwin, P. Sherriff, R. Bilous and W. Kelly (2000). "Diabetes prevalence and socioeconomic status: a population based study showing increased prevalence of type 2 diabetes mellitus in deprived areas." Journal of epidemiology and community health **54**(3): 173-177.
Craddock, N., M. E. Hurles, N. Cardin, R. D. Pearson, V. Plagnol, S. Robson, D. Vukcevic, C. Barnes, D. F. Conrad, E. Giannoulatou, C. Holmes, J. L. Marchini, K. Stirrups, M. D. Tobin, L. V. Wain, C. Yau, J. Aerts, T. Ahmad, T. D. Andrews, H. Arbury, A. Attwood, A. Auton, S. G. Ball, A. J. Balmforth, J. C. Barrett, I. Barroso, A. Barton, A. J. Bennett, S. Bhaskar, K. Blaszczyk, J. Bowes, O. J. Brand, P. S. Braund, F. Bredin, G. Breen, M. J. Brown, I. N. Bruce, J. Bull, O. S. Burren, J. Burton, J. Byrnes, S. Caesar, C. M. Clee, A. J. Coffey, J. M. Connell, J. D. Cooper, A. F. Dominiczak, K. Downes, H. E. Drummond, D. Dudakia, A. Dunham, B. Ebbs, D. Eccles, S. Edkins, C. Edwards, A. Elliot, P. Emery, D. M. Evans, G. Evans, S. Eyre, A. Farmer, I. N. Ferrier, L. Feuk, T. Fitzgerald, E. Flynn, A. Forbes, L. Forty, J. A. Franklyn, R. M. Freathy, P. Gibbs, P. Gilbert, O. Gokumen, K. Gordon-Smith, E. Gray, E. Green, C. J. Groves, D. Grozeva, R. Gwilliam, A. Hall, N. Hammond, M. Hardy, P. Harrison, N. Hassanali, H. Hebaishi, S. Hines, A. Hinks, G. A. Hitman, L. Hocking, E. Howard, P. Howard, J. M. Howson, D. Hughes, S. Hunt, J. D. Isaacs, M. Jain, D. P. Jewell, T. Johnson, J. D. Jolley, I. R. Jones, L. A. Jones, G. Kirov, C. F. Langford, H. Lango-Allen, G. M. Lathrop, J. Lee, K. L. Lee, C. Lees, K. Lewis, C. M. Lindgren, M. Maisuria-Armer, J. Maller, J. Mansfield, P. Martin, D. C. Massey, W. L. McArdle, P. McGuffin, K. E. McLay, A. Mentzer, M. L. Mimmack, A. E. Morgan, A. P. Morris, C. Mowat, S. Myers, W. Newman, E. R. Nimmo, M. C. O'Donovan, A. Onipinla, I. Onyiah, N. R. Ovington, M. J. Owen, K. Palin, K. Parnell, D. Pernet, J. R.

Perry, A. Phillips, D. Pinto, N. J. Prescott, I. Prokopenko, M. A. Quail, S. Rafelt, N. W. Rayner, R. Redon, D. M. Reid, Renwick, S. M. Ring, N. Robertson, E. Russell, D. St Clair, J. G. Sambrook, J. D. Sanderson, H. Schuilenburg, C. E. Scott, R. Scott, S. Seal, S. Shaw-Hawkins, B. M. Shields, M. J. Simmonds, D. J. Smyth, E. Somaskantharajah, K. Spanova, S. Steer, J. Stephens, H. E. Stevens, M. A. Stone, Z. Su, D. P. Symmons, J. R. Thompson, W. Thomson, M. E. Travers, C. Turnbull, A. Valsesia, M. Walker, N. M. Walker, C. Wallace, M. Warren-Perry, N. A. Watkins, J. Webster, M. N. Weedon, A. G. Wilson, M. Woodburn, B. P. Wordsworth, A. H. Young, E. Zeggini, N. P. Carter, T. M. Frayling, C. Lee, G. McVean, P. B. Munroe, A. Palotie, S. J. Sawcer, S. W. Scherer, D. P. Strachan, C. Tyler-Smith, M. A. Brown, P. R. Burton, M. J. Caulfield, A. Compston, M. Farrall, S. C. Gough, A. S. Hall, A. T. Hattersley, A. V. Hill, C. G. Mathew, M. Pembrey, J. Satsangi, M. R. Stratton, J. Worthington, P. Deloukas, A. Duncanson, D. P. Kwiatkowski, M. I. McCarthy, W. Ouwehand, M. Parkes, N. Rahman, J. A. Todd, N. J. Samani and P. Donnelly (2010). "Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls." Nature **464**(7289): 713-720.

Crowe, J. F., Kimura M (1970). "An introduction to population genetics theory." An introduction to population genetics theory.

Dabelea, D., R. L. Hanson, R. S. Lindsay, D. J. Pettitt, G. Imperatore, M. M. Gabir, J. Roumain, P. H. Bennett and W. C. Knowler (2000). "Intrauterine exposure to diabetes conveys risks for type 2 diabetes and obesity: a study of discordant sibships." Diabetes **49**(12): 2208-2211.

Danaei, G., M. M. Finucane, Y. Lu, G. M. Singh, M. J. Cowan, C. J. Paciorek, J. K. Lin, F. Farzadfar, Y. H. Khang, G. A. Stevens, M. Rao, M. K. Ali, L. M. Riley, C. A. Robinson and M. Ezzati (2011). "National, regional, and global trends in fasting plasma glucose and diabetes prevalence since 1980: systematic analysis of health examination surveys and epidemiological studies with 370 country-years and 2.7 million participants." Lancet **378**(9785): 31-40.

Das, S. K. and S. C. Elbein (2006). "The Genetic Basis of Type 2 Diabetes." Cellscience **2**(4): 100-131.

Dastani, Z., M. F. Hivert, N. Timpson, J. R. Perry, X. Yuan, R. A. Scott, P. Henneman, I. M. Heid, J. R. Kizer, L. P. Lyytikainen, C. Fuchsberger, T. Tanaka, A. P. Morris, K. Small, A. Isaacs, M. Beekman, S. Coassin, K. Lohman, L. Qi, S. Kanoni, J. S. Pankow, H. W. Uh, Y. Wu, A. Bidulescu, L. J. Rasmussen-Torvik, C. M. Greenwood, M. Ladouceur, J. Grimsby, A. K. Manning, C. T. Liu, J. Kooner, V. E. Mooser, P. Vollenweider, K. A. Kapur, J. Chambers, N. J. Wareham, C. Langenberg, R. Frants, K. Willems-Vandijk, B. A. Oostra, S. M. Willems, C. Lamina, T. W. Winkler, B. M. Psaty, R. P. Tracy, J. Brody, I. Chen, J. Viikari, M. Kahonen, P. P. Pramstaller, D. M. Evans, B. St Pourcain, N. Sattar, A. R. Wood, S. Bandinelli, O. D. Carlson, J. M. Egan, S. Bohringer, D. van Heemst, L. Kedenko, K. Kristiansson, M. L. Nuotio, B. M. Loo, T. Harris, M. Garcia, A. Kanaya, M. Haun, N. Klopp, H. E. Wichmann, P. Deloukas, E. Katsareli, D. J. Couper, B. B. Duncan, M. Kloppenburg, L. S. Adair, J. B. Borja, J. G. Wilson, S. Musani, X. Guo, T. Johnson, R. Semple, T. M. Teslovich, M. A. Allison, S. Redline, S. G. Buxbaum, K. L. Mohlke, I. Meulenbelt, C. M.

Ballantyne, G. V. Dedoussis, F. B. Hu, Y. Liu, B. Paulweber, T. D. Spector, P. E. Slagboom, L. Ferrucci, A. Jula, M. Perola, O. Raitakari, J. C. Florez, V. Salomaa, J. G. Eriksson, T. M. Frayling, A. A. Hicks, T. Lehtimaki, G. D. Smith, D. S. Siscovick, F. Kronenberg, C. van Duijn, R. J. Loos, D. M. Waterworth, J. B. Meigs, J. Dupuis, J. B. Richards, B. F. Voight, L. J. Scott, V. Steinthorsdottir, C. Dina, R. P. Welch, E. Zeggini, C. Huth, Y. S. Aulchenko, G. Thorleifsson, L. J. McCulloch, T. Ferreira, H. Grallert, N. Amin, G. Wu, C. J. Willer, S. Raychaudhuri, S. A. McCarroll, O. M. Hofmann, A. V. Segre, M. van Hoek, P. Navarro, K. Ardlie, B. Balkau, R. Benediktsson, A. J. Bennett, R. Blagieva, E. Boerwinkle, L. L. Bonnycastle, K. B. Bostrom, B. Bravenboer, S. Bumpstead, N. P. Burtt, G. Charpentier, P. S. Chines, M. Cornelis, G. Crawford, A. S. Doney, K. S. Elliott, A. L. Elliott, M. R. Erdos, C. S. Fox, C. S. Franklin, M. Ganser, C. Gieger, N. Grarup, T. Green, S. Griffin, C. J. Groves, C. Guiducci, S. Hadjadj, N. Hassanali, C. Herder, B. Isomaa, A. U. Jackson, P. R. Johnson, T. Jorgensen, W. H. Kao, A. Kong, P. Kraft, J. Kuusisto, T. Lauritzen, M. Li, A. Lieverse, C. M. Lindgren, V. Lyssenko, M. Marre, T. Meitinger, K. Midthjell, M. A. Morken, N. Narisu, P. Nilsson, K. R. Owen, F. Payne, A. K. Petersen, C. Platou, C. Proenca, I. Prokopenko, W. Rathmann, N. W. Rayner, N. R. Robertson, G. Rocheleau, M. Roden, M. J. Sampson, R. Saxena, B. M. Shields, P. Shrader, G. Sigurdsson, T. Sparso, K. Strassburger, H. M. Stringham, Q. Sun, A. J. Swift, B. Thorand, J. Tichet, T. Tuomi, R. M. van Dam, T. W. van Haeften, T. van Herpt, J. V. van Vliet-Ostaptchouk, G. B. Walters, M. N. Weedon, C. Wijmenga, J. Witteman, R. N. Bergman, S. Cauchi, F. S. Collins, A. L. Gloyn, U. Gyllensten, T. Hansen, W. A. Hide, G. A. Hitman, A. Hofman, D. J. Hunter, K. Hveem, M. Laakso, A. D. Morris, C. N. Palmer, I. Rudan, E. Sijbrands, L. D. Stein, J. Tuomilehto, A. Uitterlinden, M. Walker, R. M. Watanabe, G. R. Abecasis, B. O. Boehm, H. Campbell, M. J. Daly, A. T. Hattersley, O. Pedersen, I. Barroso, L. Groop, R. Sladek, U. Thorsteinsdottir, J. F. Wilson, T. Illig, P. Froguel, C. M. van Duijn, K. Stefansson, D. Altshuler, M. Boehnke, M. I. McCarthy, N. Soranzo, E. Wheeler, N. L. Glazer, N. Bouatia-Naji, R. Magi, J. Randall, P. Elliott, D. Rybin, A. Dehghan, J. J. Hottenga, K. Song, A. Goel, T. Lajunen, A. Doney, C. Cavalcanti-Proenca, M. Kumari, N. J. Timpson, C. Zabena, E. Ingelsson, P. An, J. O'Connell, J. Luan, A. Elliott, R. M. Roccasecca, F. Pattou, P. Sethupathy, Y. Ariyurek, P. Barter, J. P. Beilby, Y. Ben-Shlomo, S. Bergmann, M. Bochud, A. Bonnefond, K. Borch-Johnsen, Y. Bottcher, E. Brunner, S. J. Bumpstead, Y. D. Chen, P. Chines, R. Clarke, L. J. Coin, M. N. Cooper, L. Crisponi, I. N. Day, E. J. de Geus, J. Delplanque, A. C. Fedson, A. Fischer-Rosinsky, N. G. Forouhi, M. G. Franzosi, P. Galan, M. O. Goodarzi, J. Graessler, S. Grundy, R. Gwilliam, G. Hallmans, N. Hammond, X. Han, A. L. Hartikainen, C. Hayward, S. C. Heath, S. Hercberg, D. R. Hillman, A. D. Hingorani, J. Hui, J. Hung, M. Kaakinen, J. Kaprio, Y. A. Kesaniemi, M. Kivimaki, B. Knight, S. Koskinen, P. Kovacs, K. O. Kyvik, G. M. Lathrop, D. A. Lawlor, O. Le Bacquer, C. Lecoeur, Y. Li, R. Mahley, M. Mangino, M. T. Martinez-Larrad, J. B. McAteer, R. McPherson, C. Meisinger, D. Melzer, D. Meyre, B. D. Mitchell, S. Mukherjee, S. Naitza, M. J. Neville, M. Orru, R. Pakyz, G. Paolisso, C. Pattaro, D. Pearson, J. F. Peden, N. L. Pedersen, A. F. Pfeiffer, I. Pichler, O. Polasek, D.

Posthuma, S. C. Potter, A. Pouta, M. A. Province, K. Rice, S. Ripatti, F. Rivadeneira, O. Rolandsson, A. Sandbaek, M. Sandhu, S. Sanna, A. A. Sayer, P. Scheet, U. Seedorf, S. J. Sharp, B. Shields, G. Sigurethsson, E. J. Sijbrands, A. Silveira, L. Simpson, A. Singleton, N. L. Smith, U. Sovio, A. Swift, H. Syddall, A. C. Syvanen, A. Tonjes, A. G. Uitterlinden, K. W. van Dijk, D. Varma, S. Visvikis-Siest, V. Vitart, N. Vogelzangs, G. Waeber, P. J. Wagner, A. Walley, K. L. Ward, H. Watkins, S. H. Wild, G. Willemsen, J. C. Witteman, J. W. Yarnell, D. Zelenika, B. Zethelius, G. Zhai, J. H. Zhao, M. C. Zillikens, I. B. Borecki, P. Meneton, P. K. Magnusson, D. M. Nathan, G. H. Williams, K. Silander, S. R. Bornstein, P. Schwarz, J. Spranger, F. Karpe, A. R. Shuldiner, C. Cooper, M. Serrano-Rios, L. Lind, L. J. Palmer, F. B. s. Hu, P. W. Franks, S. Ebrahim, M. Marmot, A. F. Wright, M. Stumvoll, A. Hamsten, T. A. Buchanan, T. T. Valle, J. I. Rotter, B. W. Penninx, D. I. Boomsma, A. Cao, A. Scuteri, D. Schlessinger, M. Uda, A. Ruokonen, M. R. Jarvelin, L. Peltonen, V. Mooser, K. Musunuru, A. V. Smith, A. C. Edmondson, I. M. Stylianou, M. Koseki, J. P. Pirruccello, D. I. Chasman, C. T. Johansen, S. W. Fouchier, G. M. Peloso, M. Barbalic, S. L. Ricketts, J. C. Bis, M. F. Feitosa, M. Orho-Melander, O. Melander, X. Li, Y. S. Cho, M. J. Go, Y. J. Kim, J. Y. Lee, T. Park, K. Kim, X. Sim, R. T. Ong, D. C. Croteau-Chonka, L. A. Lange, J. D. Smith, A. Ziegler, W. Zhang, R. Y. Zee, J. B. Whitfield, J. R. Thompson, I. Surakka, J. H. Smit, J. Sinisalo, J. Scott, J. Saharinen, C. Sabatti, L. M. Rose, R. Roberts, M. Rieder, A. N. Parker, G. Pare, C. J. O'Donnell, M. S. Nieminen, D. A. Nickerson, G. W. Montgomery, W. McArdle, D. Masson, N. G. Martin, F. Marroni, G. Lucas, R. Luben, M. L. Lokki, G. Lettre, L. J. Launer, E. G. Lakatta, R. Laaksonen, I. R. Konig, K. T. Khaw, L. M. Kaplan, A. Johansson, A. C. Janssens, W. Igl, G. K. Hovingh, C. Hengstenberg, A. S. Havulinna, N. D. Hastie, T. B. Harris, T. Haritunians, A. S. Hall, L. C. Groop, E. Gonzalez, N. B. Freimer, J. Erdmann, K. G. Ejebe, A. Doring, A. F. Dominiczak, S. Demissie, U. de Faire, M. J. Caulfield, S. M. Boekholdt, T. L. Assimes, T. Quertermous, M. Seielstad, T. Y. Wong, E. S. Tai, A. B. Feranil, C. W. Kuzawa, H. A. Taylor, Jr., S. B. Gabriel, H. Holm, V. Gudnason, R. M. Krauss, J. M. Ordovas, P. B. Munroe, J. S. Kooner, A. R. Tall, R. A. Hegele, J. J. Kastelein, E. E. Schadt, D. P. Strachan, M. P. Reilly, N. J. Samani, H. Schunkert, L. A. Cupples, M. S. Sandhu, P. M. Ridker, D. J. Rader and S. Kathiresan (2012). "Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: a multi-ethnic meta-analysis of 45,891 individuals." PLoS genetics **8**(3): e1002607.

Dawn Teare, M. and J. H. Barrett (2005). "Genetic linkage studies." Lancet **366**(9490): 1036-1044.

Deeb, S. S., L. Fajas, M. Nemoto, J. Pihlajamaki, L. Mykkanen, J. Kuusisto, M. Laakso, W. Fujimoto and J. Auwerx (1998). "A Pro12Ala substitution in PPARgamma2 associated with decreased receptor activity, lower body mass index and improved insulin sensitivity." Nat Genet **20**(3): 284-287.

Drivsholm, T., H. Ibsen, M. Schroll, M. Davidsen and K. Borch-Johnsen (2001). "Increasing prevalence of diabetes mellitus and impaired glucose tolerance among 60-year-old Danes." Diabetic medicine : a journal of the British Diabetic Association **18**(2): 126-132.

Druley, T. E., F. L. Vallania, D. J. Wegner, K. E. Varley, O. L. Knowles, J. A. Bonds, S. W. Robison, S. W. Doniger, A. Hamvas, F. S. Cole, J. C. Fay and R. D. Mitra (2009). "Quantification of rare allelic variants from pooled genomic DNA." Nat Methods **6**(4): 263-265.

Druley, T. E., F. L. Vallania, D. J. Wegner, K. E. Varley, O. L. Knowles, J. A. Bonds, S. W. Robison, S. W. Doniger, A. Hamvas, F. S. Cole, J. C. Fay and R. D. Mitra (2009). "Quantification of rare allelic variants from pooled genomic DNA." Nature methods **6**(4): 263-265.

Dupuis, J., C. Langenberg, I. Prokopenko, R. Saxena, N. Soranzo, A. U. Jackson, E. Wheeler, N. L. Glazer, N. Bouatia-Naji, A. L. Gloyn, C. M. Lindgren, R. Magi, A. P. Morris, J. Randall, T. Johnson, P. Elliott, D. Rybin, G. Thorleifsson, V. Steinthorsdottir, P. Henneman, H. Grallert, A. Dehghan, J. J. Hottenga, C. S. Franklin, P. Navarro, K. Song, A. Goel, J. R. Perry, J. M. Egan, T. Lajunen, N. Grarup, T. Sparso, A. Doney, B. F. Voight, H. M. Stringham, M. Li, S. Kanoni, P. Shrader, C. Cavalcanti-Proenca, M. Kumari, L. Qi, N. J. Timpson, C. Gieger, C. Zabena, G. Rocheleau, E. Ingelsson, P. An, J. O'Connell, J. Luan, A. Elliott, S. A. McCarroll, F. Payne, R. M. Roccasecca, F. Pattou, P. Sethupathy, K. Ardlie, Y. Ariyurek, B. Balkau, P. Barter, J. P. Beilby, Y. Ben-Shlomo, R. Benediktsson, A. J. Bennett, S. Bergmann, M. Bochud, E. Boerwinkle, A. Bonnefond, L. L. Bonnycastle, K. Borch-Johnsen, Y. Bottcher, E. Brunner, S. J. Bumpstead, G. Charpentier, Y. D. Chen, P. Chines, R. Clarke, L. J. Coin, M. N. Cooper, M. Cornelis, G. Crawford, L. Crisponi, I. N. Day, E. J. de Geus, J. Delplanque, C. Dina, M. R. Erdos, A. C. Fedson, A. Fischer-Rosinsky, N. G. Forouhi, C. S. Fox, R. Frants, M. G. Franzosi, P. Galan, M. O. Goodarzi, J. Graessler, C. J. Groves, S. Grundy, R. Gwilliam, U. Gyllensten, S. Hadjadj, G. Hallmans, N. Hammond, X. Han, A. L. Hartikainen, N. Hassanali, C. Hayward, S. C. Heath, S. Hercberg, C. Herder, A. A. Hicks, D. R. Hillman, A. D. Hingorani, A. Hofman, J. Hui, J. Hung, B. Isomaa, P. R. Johnson, T. Jorgensen, A. Jula, M. Kaakinen, J. Kaprio, Y. A. Kesaniemi, M. Kivimaki, B. Knight, S. Koskinen, P. Kovacs, K. O. Kyvik, G. M. Lathrop, D. A. Lawlor, O. Le Bacquer, C. Lecoeur, Y. Li, V. Lyssenko, R. Mahley, M. Mangino, A. K. Manning, M. T. Martinez-Larrad, J. B. McAteer, L. J. McCulloch, R. McPherson, C. Meisinger, D. Melzer, D. Meyre, B. D. Mitchell, M. A. Morken, S. Mukherjee, S. Naitza, N. Narisu, M. J. Neville, B. A. Oostra, M. Orru, R. Pakyz, C. N. Palmer, G. Paolisso, C. Pattaro, D. Pearson, J. F. Peden, N. L. Pedersen, M. Perola, A. F. Pfeiffer, I. Pichler, O. Polasek, D. Posthuma, S. C. Potter, A. Pouta, M. A. Province, B. M. Psaty, W. Rathmann, N. W. Rayner, K. Rice, S. Ripatti, F. Rivadeneira, M. Roden, O. Rolandsson, A. Sandbaek, M. Sandhu, S. Sanna, A. A. Sayer, P. Scheet, L. J. Scott, U. Seedorf, S. J. Sharp, B. Shields, G. Sigurethsson, E. J. Sijbrands, A. Silveira, L. Simpson, A. Singleton, N. L. Smith, U. Sovio, A. Swift, H. Syddall, A. C. Syvanen, T. Tanaka, B. Thorand, J. Tichet, A. Tonjes, T. Tuomi, A. G. Uitterlinden, K. W. van Dijk, M. van Hoek, D. Varma, S. Visvikis-Siest, V. Vitart, N. Vogelzangs, G. Waeber, P. J. Wagner, A. Walley, G. B. Walters, K. L. Ward, H. Watkins, M. N. Weedon, S. H. Wild, G. Willemsen, J. C. Witteman, J. W. Yarnell, E. Zeggini, D. Zelenika, B. Zethelius, G. Zhai, J. H. Zhao, M. C. Zillikens, I. B. Borecki, R. J. Loos, P.

Meneton, P. K. Magnusson, D. M. Nathan, G. H. Williams, A. T. Hattersley, K. Silander, V. Salomaa, G. D. Smith, S. R. Bornstein, P. Schwarz, J. Spranger, F. Karpe, A. R. Shuldiner, C. Cooper, G. V. Dedoussis, M. Serrano-Rios, A. D. Morris, L. Lind, L. J. Palmer, F. B. Hu, P. W. Franks, S. Ebrahim, M. Marmot, W. H. Kao, J. S. Pankow, M. J. Sampson, J. Kuusisto, M. Laakso, T. Hansen, O. Pedersen, P. P. Pramstaller, H. E. Wichmann, T. Illig, I. Rudan, A. F. Wright, M. Stumvoll, H. Campbell, J. F. Wilson, R. N. Bergman, T. A. Buchanan, F. S. Collins, K. L. Mohlke, J. Tuomilehto, T. T. Valle, D. Altshuler, J. I. Rotter, D. S. Siscovick, B. W. Penninx, D. I. Boomsma, P. Deloukas, T. D. Spector, T. M. Frayling, L. Ferrucci, A. Kong, U. Thorsteinsdottir, K. Stefansson, C. M. van Duijn, Y. S. Aulchenko, A. Cao, A. Scuteri, D. Schlessinger, M. Uda, A. Ruokonen, M. R. Jarvelin, D. M. Waterworth, P. Vollenweider, L. Peltonen, V. Mooser, G. R. Abecasis, N. J. Wareham, R. Sladek, P. Froguel, R. M. Watanabe, J. B. Meigs, L. Groop, M. Boehnke, M. I. McCarthy, J. C. Florez and I. Barroso (2010). "New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk." Nature Genetics **42**(2): 105-116.

Edghill, E. L., S. E. Flanagan, A. M. Patch, C. Boustred, A. Parrish, B. Shields, M. H. Shepherd, K. Hussain, R. R. Kapoor, M. Malecki, M. J. MacDonald, J. Stoy, D. F. Steiner, L. H. Philipson, G. I. Bell, A. T. Hattersley and S. Ellard (2008). "Insulin mutation screening in 1,044 patients with diabetes: mutations in the INS gene are a common cause of neonatal diabetes but a rare cause of diabetes diagnosed in childhood or adulthood." Diabetes **57**(4): 1034-1042.

Edghill, E. L., A. Khamis, M. N. Weedon, M. Walker, G. A. Hitman, M. I. McCarthy, K. R. Owen, S. Ellard, T. H. A and T. M. Frayling (2011). "Sequencing PDX1 (insulin promoter factor 1) in 1788 UK individuals found 5% had a low frequency coding variant, but these variants are not associated with Type 2 diabetes." Diabetic medicine : a journal of the British Diabetic Association **28**(6): 681-684.

Eichler, E. E., J. Flint, G. Gibson, A. Kong, S. M. Leal, J. H. Moore and J. H. Nadeau (2010). "Missing heritability and strategies for finding the underlying causes of complex disease." Nature reviews. Genetics **11**(6): 446-450.

Ek, J., C. S. Rose, D. P. Jensen, C. Glumer, K. Borch-Johnsen, T. Jorgensen, O. Pedersen and T. Hansen (2005). "The functional Thr130Ile and Val255Met polymorphisms of the hepatocyte nuclear factor-4alpha (HNF4A): gene associations with type 2 diabetes or altered beta-cell function among Danes." J Clin Endocrinol Metab **90**(5): 3054-3059.

Elbein, S. C. and M. A. Karim (2004). "Does the aspartic acid to asparagine substitution at position 76 in the pancreas duodenum homeobox gene (PDX1) cause late-onset type 2 diabetes?" Diabetes care **27**(8): 1968-1973.

Ellard, S., C. Bellanne-Chantelot and A. T. Hattersley (2008). "Best practice guidelines for the molecular genetic diagnosis of maturity-onset diabetes of the young." Diabetologia **51**(4): 546-553.

Ellard, S. and K. Colclough (2006). "Mutations in the genes encoding the transcription factors hepatocyte nuclear factor 1 alpha (HNF1A) and 4 alpha (HNF4A) in maturity-onset diabetes of the young." Hum Mutat **27**(9): 854-869.

Ellard S, S. B., Tysoe C, Treacy R,Yau S, Mattocks C, Wallace A (2009). "Semi-Automated Unidirectional Sequence Analysis for Mutation Detection in a Clinical Diagnostic Setting." Genetic Testing and Molecular Biomarkers: In press.

Ewing, B. and P. Green (1998). "Base-calling of automated sequencer traces using phred. II. Error probabilities." Genome Res **8**(3): 186-194.

Ewing, B., L. Hillier, M. C. Wendl and P. Green (1998). "Base-calling of automated sequencer traces using phred. I. Accuracy assessment." Genome Res **8**(3): 175-185.

Fajans, S. S., G. I. Bell, V. P. Paz, J. E. Below, N. J. Cox, C. Martin, I. H. Thomas and M. Chen (2010). "Obesity and hyperinsulinemia in a family with pancreatic agenesis and MODY caused by the IPF1 mutation Pro63fsX60." Translational research : the journal of laboratory and clinical medicine **156**(1): 7-14.

Fajans, S. S., G. I. Bell and K. S. Polonsky (2001). "Molecular mechanisms and clinical pathophysiology of maturity-onset diabetes of the young." The New England journal of medicine **345**(13): 971-980.

Fan, J. B., M. S. Chee and K. L. Gunderson (2006). "Highly parallel genomic assays." Nat Rev Genet **7**(8): 632-644.

Fetita, L. S., E. Sobngwi, P. Serradas, F. Calvo and J. F. Gautier (2006). "Consequences of fetal exposure to maternal diabetes in offspring." The Journal of clinical endocrinology and metabolism **91**(10): 3718-3724.

Florez, J. C., M. Sjogren, C. M. Agapakis, N. P. Burtt, P. Almgren, U. Lindblad, G. Berglund, T. Tuomi, D. Gaudet, M. J. Daly, K. G. Ardlie, J. N. Hirschhorn, D. Altshuler and L. Groop (2007). "Association testing of common variants in the insulin receptor substrate-1 gene (IRS1) with type 2 diabetes." Diabetologia **50**(6): 1209-1217.

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**(3): 458-463.

Frayling, T. M., M. P. Bulamn, S. Ellard, M. Appleton, M. J. Dronsfield, A. D. Mackie, J. D. Baird, P. J. Kaisaki, K. Yamagata, G. I. Bell, S. C. Bain and A. T. Hattersley (1997). "Mutations in the hepatocyte nuclear factor-1alpha gene are a common cause of maturity-onset diabetes of the young in the U.K." Diabetes **46**(4): 720-725.

Frayling, T. M., J. C. Evans, M. P. Bulman, E. Pearson, L. Allen, K. Owen, C. Bingham, M. Hannemann, M. Shepherd, S. Ellard and A. T. Hattersley (2001). "beta-cell genes and diabetes: molecular and clinical characterization of mutations in transcription factors." Diabetes **50 Suppl 1**: S94-100.

Frayling, T. M., M. I. McCarthy, M. Walker, J. C. Levy, S. O'Rahilly, G. A. Hitman, P. V. Rao, A. J. Bennett, E. C. Jones, S. Menzel, S. Ellard and A. T. Hattersley (2000). "No evidence for linkage at candidate type 2 diabetes susceptibility loci on chromosomes 12 and 20 in United Kingdom Caucasians." J Clin Endocrinol Metab **85**(2): 853-857.

Frayling, T. M., S. Wiltshire, G. A. Hitman, M. Walker, J. C. Levy, M. Sampson, C. J. Groves, S. Menzel, M. I. McCarthy and A. T. Hattersley (2003). "Young-onset type

2 diabetes families are the major contributors to genetic loci in the Diabetes UK Warren 2 genome scan and identify putative novel loci on chromosomes 8q21, 21q22, and 22q11." Diabetes **52**(7): 1857-1863.

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**(6365): 162-164.

Gardner, R. J., D. J. Mackay, A. J. Mungall, C. Polychronakos, R. Siebert, J. P. Shield, I. K. Temple and D. O. Robinson (2000). "An imprinted locus associated with transient neonatal diabetes mellitus." Hum Mol Genet **9**(4): 589-596.

Garin, I., E. L. Edghill, I. Akerman, O. Rubio-Cabezas, I. Rica, J. M. Locke, M. A. Maestro, A. Alshaikh, R. Bundak, G. del Castillo, A. Deeb, D. Deiss, J. M. Fernandez, K. Godbole, K. Hussain, M. O'Connell, T. Klupa, S. Kolouskova, F. Mohsin, K. Perlman, Z. Sumnik, J. M. Rial, E. Ugarte, T. Vasanthi, K. Johnstone, S. E. Flanagan, R. Martinez, C. Castano, A. M. Patch, E. Fernandez-Rebollo, K. Raile, N. Morgan, L. W. Harries, L. Castano, S. Ellard, J. Ferrer, G. Perez de Nanclares and A. T. Hattersley (2010). "Recessive mutations in the INS gene result in neonatal diabetes through reduced insulin biosynthesis." Proceedings of the National Academy of Sciences of the United States of America **107**(7): 3105-3110.

Gauderman, W. and J. Morrison (2006). QUANTO 1.1: A computer program for power and sample size calculations for genetic-epidemiology studies, http://hydra.usc.edu/gxe.

Gauderman, W. J. (2002). "Sample size requirements for association studies of gene-gene interaction." American journal of epidemiology **155**(5): 478-484.

Gaulton, K. J., T. Nammo, L. Pasquali, J. M. Simon, P. G. Giresi, M. P. Fogarty, T. M. Panhuis, P. Mieczkowski, A. Secchi, D. Bosco, T. Berney, E. Montanya, K. L. Mohlke, J. D. Lieb and J. Ferrer (2010). "A map of open chromatin in human pancreatic islets." Nature Genetics **42**(3): 255-259.

Gerbitz, K. D., K. Gempel and D. Brdiczka (1996). "Mitochondria and diabetes. Genetic, biochemical, and clinical implications of the cellular energy circuit." Diabetes **45**(2): 113-126.

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**(3): 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**(18): 1838-1849.

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**(2): 568-572.

Gottlieb, M. S. (1980). "Diabetes in offspring and siblings of juvenile- and maturity-onset-type diabetics." Journal of chronic diseases **33**(6): 331-339.

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**(3): 320-323.

Groop, L., C. Forsblom, M. Lehtovirta, T. Tuomi, S. Karanko, M. Nissen, B. O. Ehrnstrom, B. Forsen, B. Isomaa, B. Snickars and M. R. Taskinen (1996). "Metabolic consequences of a family history of NIDDM (the Botnia study): evidence for sex-specific parental effects." Diabetes **45**(11): 1585-1593.

Guan, W., A. Pluzhnikov, N. J. Cox and M. Boehnke (2008). "Meta-analysis of 23 type 2 diabetes linkage studies from the International Type 2 Diabetes Linkage Analysis Consortium." Human heredity **66**(1): 35-49.

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." Nature genetics **39**(8): 977-983.

Gunnarsdottir, E. D., M. Li, M. Bauchet, K. Finstermeier and M. Stoneking (2011). "High-throughput sequencing of complete human mtDNA genomes from the Philippines." Genome research **21**(1): 1-11.

Hales, C. N. and D. J. Barker (1992). "Type 2 (non-insulin-dependent) diabetes mellitus: the thrifty phenotype hypothesis." Diabetologia **35**(7): 595-601.

Hales, C. N., D. J. Barker, P. M. Clark, L. J. Cox, C. Fall, C. Osmond and P. D. Winter (1991). "Fetal and infant growth and impaired glucose tolerance at age 64." Bmj **303**(6809): 1019-1022.

Halushka, M. K., J. B. Fan, K. Bentley, L. Hsie, N. Shen, A. Weder, R. Cooper, R. Lipshutz and A. Chakravarti (1999). "Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis." Nat Genet **22**(3): 239-247.

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**(12): 1511-1515.

Hani, E. H., D. A. Stoffers, J. C. Chevre, E. Durand, V. Stanojevic, C. Dina, J. F. Habener and P. Froguel (1999). "Defective mutations in the insulin promoter factor-1 (IPF-1) gene in late-onset type 2 diabetes mellitus." J Clin Invest **104**(9): R41-48.

Hanley, S. C., E. Austin, B. Assouline-Thomas, J. Kapeluto, J. Blaichman, M. Moosavi, M. Petropavlovskaia and L. Rosenberg (2010). "{beta}-Cell mass dynamics and islet cell plasticity in human type 2 diabetes." Endocrinology **151**(4): 1462-1472.

Hansen, L., S. M. Echwald, T. Hansen, S. A. Urhammer, J. O. Clausen and O. Pedersen (1997). "Amino acid polymorphisms in the ATP-regulatable inward rectifier Kir6.2 and their relationships to glucose- and tolbutamide-induced insulin secretion, the insulin sensitivity index, and NIDDM." Diabetes **46**(3): 508-512.

Hansen, L., S. Urioste, H. V. Petersen, J. N. Jensen, H. Eiberg, F. Barbetti, P. Serup, T. Hansen and O. Pedersen (2000). "Missense mutations in the human insulin promoter factor-1 gene and their relation to maturity-onset diabetes of the young and late-onset type 2 diabetes mellitus in caucasians." The Journal of clinical endocrinology and metabolism **85**(3): 1323-1326.

Hardy, O. T., M. P. Czech and S. Corvera (2012). "What causes the insulin resistance underlying obesity?" Current opinion in endocrinology, diabetes, and obesity **19**(2): 81-87.

Harries, L. W., S. Ellard, A. Stride, N. G. Morgan and A. T. Hattersley (2006). "Isomers of the TCF1 gene encoding hepatocyte nuclear factor-1 alpha show differential expression in the pancreas and define the relationship between mutation position and clinical phenotype in monogenic diabetes." Hum Mol Genet **15**(14): 2216-2224.

Harris, M. I., R. C. Eastman, C. C. Cowie, K. M. Flegal and M. S. Eberhardt (1999). "Racial and ethnic differences in glycemic control of adults with type 2 diabetes." Diabetes Care **22**(3): 403-408.

Hattersley, A. T. and M. I. McCarthy (2005). "What makes a good genetic association study?" Lancet **366**(9493): 1315-1323.

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**(8805): 1307-1310.

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**(12): 3586-3590.

Hegele, R. A., H. Cao, S. B. Harris, A. J. Hanley and B. Zinman (1999). "The hepatic nuclear factor-1alpha G319S variant is associated with early-onset type 2 diabetes in Canadian Oji-Cree." J Clin Endocrinol Metab **84**(3): 1077-1082.

Hellmann, J., Y. Tang, M. Kosuri, A. Bhatnagar and M. Spite (2011). "Resolvin D1 decreases adipose tissue macrophage accumulation and improves insulin sensitivity in obese-diabetic mice." FASEB journal : official publication of the Federation of American Societies for Experimental Biology **25**(7): 2399-2407.

Herbach, N., B. Rathkolb, E. Kemter, L. Pichl, M. Klaften, M. H. de Angelis, P. A. Halban, E. Wolf, B. Aigner and R. Wanke (2007). "Dominant-negative effects of a novel mutated Ins2 allele causes early-onset diabetes and severe beta-cell loss in Munich Ins2C95S mutant mice." Diabetes **56**(5): 1268-1276.

Higgins, J. P., S. G. Thompson, J. J. Deeks and D. G. Altman (2003). "Measuring inconsistency in meta-analyses." Bmj **327**(7414): 557-560.

Hill, W. G. and A. Robertson (1968). "The effects of inbreeding at loci with heterozygote advantage." Genetics **60**(3): 615-628.

Hofman, A., D. E. Grobbee, P. T. de Jong and F. A. van den Ouweland (1991). "Determinants of disease and disability in the elderly: the Rotterdam Elderly Study." European journal of epidemiology **7**(4): 403-422.

Holm, H., D. F. Gudbjartsson, P. Sulem, G. Masson, H. T. Helgadottir, C. Zanon, O. T. Magnusson, A. Helgason, J. Saemundsdottir, A. Gylfason, H. Stefansdottir, S. Gretarsdottir, S. E. Matthiasson, G. M. Thorgeirsson, A. Jonasdottir, A. Sigurdsson, H. Stefansson, T. Werge, T. Rafnar, L. A. Kiemeney, B. Parvez, R. Muhammad, D. M. Roden, D. Darbar, G. Thorleifsson, G. B. Walters, A. Kong, U. Thorsteinsdottir, D. O. Arnar and K. Stefansson (2011). "A rare variant in MYH6 is associated with high risk of sick sinus syndrome." Nature Genetics **43**(4): 316-320.

Holmkvist, J., C. Cervin, V. Lyssenko, W. Winckler, D. Anevski, C. Cilio, P. Almgren, G. Berglund, P. Nilsson, T. Tuomi, C. M. Lindgren, D. Altshuler and L. Groop (2006). "Common variants in HNF-1 alpha and risk of type 2 diabetes." Diabetologia **49**(12): 2882-2891.

Holstein, A., M. Hahn, A. Korner, M. Stumvoll and P. Kovacs (2011). "TCF7L2 and therapeutic response to sulfonylureas in patients with type 2 diabetes." BMC medical genetics **12**: 30.

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**(4): 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. Horikawa, 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**(2): 163-175.

Hotamisligil, G. S. (2003). "Inflammatory pathways and insulin action." International journal of obesity and related metabolic disorders : journal of the International Association for the Study of Obesity **27 Suppl 3**: S53-55.

Huang, J., D. Ellinghaus, A. Franke, B. Howie and Y. Li (2012). "1000 Genomes-based imputation identifies novel and refined associations for the Wellcome Trust Case Control Consortium phase 1 Data." European journal of human genetics : EJHG.

Hugot, J. P. (2006). "CARD15/NOD2 mutations in Crohn's disease." Ann N Y Acad Sci **1072**: 9-18.

Imamura, M. and S. Maeda (2011). "Genetics of type 2 diabetes: the GWAS era and future perspectives [Review]." Endocrine journal **58**(9): 723-739.

International HapMap Consortium (2005). "A haplotype map of the human genome." Nature **437**(7063): 1299-1320.

"The International HapMap Project." (2003). Nature **426**(6968): 789-796.

Isomaa, B., M. Henricsson, M. Lehto, C. Forsblom, S. Karanko, L. Sarelin, M. Haggblom and L. Groop (1998). "Chronic diabetic complications in patients with MODY3 diabetes." Diabetologia **41**(4): 467-473.

Itsara, A., H. Wu, J. D. Smith, D. A. Nickerson, I. Romieu, S. J. London and E. E. Eichler (2010). "De novo rates and selection of large copy number variation." Genome research **20**(11): 1469-1481.

Iwasaki, N., N. J. Cox, Y. Q. Wang, P. E. Schwarz, G. I. Bell, M. Honda, M. Imura, M. Ogata, M. Saito, N. Kamatani and Y. Iwamoto (2003). "Mapping genes influencing type 2 diabetes risk and BMI in Japanese subjects." Diabetes **52**(1): 209-213.

Jackson, A. E., P. G. Cassell, B. V. North, S. Vijayaraghavan, S. V. Gelding, A. Ramachandran, C. Snehalatha and G. A. Hitman (2004). "Polymorphic variations in the neurogenic differentiation-1, neurogenin-3, and hepatocyte nuclear factor-1alpha genes contribute to glucose intolerance in a South Indian population." Diabetes **53**(8): 2122-2125.

Jensen, C. B., H. Storgaard, F. Dela, J. J. Holst, S. Madsbad and A. A. Vaag (2002). "Early differential defects of insulin secretion and action in 19-year-old caucasian men who had low birth weight." Diabetes **51**(4): 1271-1280.

Ji, L., M. Malecki, J. H. Warram, Y. Yang, S. S. Rich and A. S. Krolewski (1997). "New susceptibility locus for NIDDM is localized to human chromosome 20q." Diabetes **46**(5): 876-881.

Jiang, M. H., J. Fei, M. S. Lan, Z. P. Lu, M. Liu, W. W. Fan, X. Gao and D. R. Lu (2008). "Hypermethylation of hepatic Gck promoter in ageing rats contributes to diabetogenic potential." Diabetologia **51**(8): 1525-1533.

Jonsson, A., B. Isomaa, T. Tuomi, J. Taneera, A. Salehi, P. Nilsson, L. Groop and V. Lyssenko (2009). "A variant in the KCNQ1 gene predicts future type 2 diabetes and mediates impaired insulin secretion." Diabetes **58**(10): 2409-2413.

Kahn, S. E. (2003). "The relative contributions of insulin resistance and beta-cell dysfunction to the pathophysiology of Type 2 diabetes." Diabetologia **46**(1): 3-19.

Kathiresan, S., C. J. Willer, G. M. Peloso, S. Demissie, K. Musunuru, E. E. Schadt, L. Kaplan, D. Bennett, Y. Li, T. Tanaka, B. F. Voight, L. L. Bonnycastle, A. U. Jackson, G. Crawford, A. Surti, C. Guiducci, N. P. Burtt, S. Parish, R. Clarke, D. Zelenika, K. A. Kubalanza, M. A. Morken, L. J. Scott, H. M. Stringham, P. Galan, A. J. Swift, J. Kuusisto, R. N. Bergman, J. Sundvall, M. Laakso, L. Ferrucci, P. Scheet, S. Sanna,

M. Uda, Q. Yang, K. L. Lunetta, J. Dupuis, P. I. de Bakker, C. J. O'Donnell, J. C. Chambers, J. S. Kooner, S. Hercberg, P. Meneton, E. G. Lakatta, A. Scuteri, D. Schlessinger, J. Tuomilehto, F. S. Collins, L. Groop, D. Altshuler, R. Collins, G. M. Lathrop, O. Melander, V. Salomaa, L. Peltonen, M. Orho-Melander, J. M. Ordovas, M. Boehnke, G. R. Abecasis, K. L. Mohlke and L. A. Cupples (2009). "Common variants at 30 loci contribute to polygenic dyslipidemia." Nat Genet **41**(1): 56-65.

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**(10): 2944-2950.

Kobberling J, T. H. (1982). Empirical risk figures for first degree relatives of non-insulin dependent diabetics. London, Academic Press.

Koboldt, D. C., D. E. Larson, K. Chen, L. Ding and R. K. Wilson (2012). "Massively parallel sequencing approaches for characterization of structural variation." Methods in molecular biology **838**: 369-384.

Kooner, J. S., D. Saleheen, X. Sim, J. Sehmi, W. Zhang, P. Frossard, L. F. Been, K. S. Chia, A. S. Dimas, N. Hassanali, T. Jafar, J. B. Jowett, X. Li, V. Radha, S. D. Rees, F. Takeuchi, R. Young, T. Aung, A. Basit, M. Chidambaram, D. Das, E. Grunberg, A. K. Hedman, Z. I. Hydrie, M. Islam, C. C. Khor, S. Kowlessur, M. M. Kristensen, S. Liju, W. Y. Lim, D. R. Matthews, J. Liu, A. P. Morris, A. C. Nica, J. M. Pinidiyapathirage, I. Prokopenko, A. Rasheed, M. Samuel, N. Shah, A. S. Shera, K. S. Small, C. Suo, A. R. Wickremasinghe, T. Y. Wong, M. Yang, F. Zhang, G. R. Abecasis, A. H. Barnett, M. Caulfield, P. Deloukas, T. M. Frayling, P. Froguel, N. Kato, P. Katulanda, M. A. Kelly, J. Liang, V. Mohan, D. K. Sanghera, J. Scott, M. Seielstad, P. Z. Zimmet, P. Elliott, Y. Y. Teo, M. I. McCarthy, J. Danesh, E. S. Tai and J. C. Chambers (2011). "Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci." Nature genetics.

Koster, J. C., B. A. Marshall, N. Ensor, J. A. Corbett and C. G. Nichols (2000). "Targeted overactivity of beta cell K(ATP) channels induces profound neonatal diabetes." Cell **100**(6): 645-654.

Kropff, J., M. P. Selwood, M. I. McCarthy, A. J. Farmer and K. R. Owen (2011). "Prevalence of monogenic diabetes in young adults: a community-based, cross-sectional study in Oxfordshire, UK." Diabetologia **54**(5): 1261-1263.

Lam, D. W. and D. LeRoith (2012). "The worldwide diabetes epidemic." Current opinion in endocrinology, diabetes, and obesity **19**(2): 93-96.

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.

Larsen, L. H., S. M. Echwald, T. I. Sorensen, T. Andersen, B. S. Wulff and O. Pedersen (2005). "Prevalence of mutations and functional analyses of melanocortin 4 receptor variants identified among 750 men with juvenile-onset obesity." J Clin Endocrinol Metab **90**(1): 219-224.

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.

Lawrence, R., A. G. Day-Williams, K. S. Elliott, A. P. Morris and E. Zeggini (2010). "CCRaVAT and QuTie-enabling analysis of rare variants in large-scale case control and quantitative trait association studies." BMC bioinformatics **11**: 527.

Lee, B. C., M. Appleton, A. C. Shore, J. E. Tooke and A. T. Hattersley (1999). "Impaired maximum microvascular hyperaemia in patients with MODY 3 (hepatocyte nuclear factor-1alpha gene mutations)." Diabetic medicine : a journal of the British Diabetic Association **16**(9): 731-735.

Lehmann, J. M., L. B. Moore, T. A. Smith-Oliver, W. O. Wilkison, T. M. Willson and S. A. Kliewer (1995). "An antidiabetic thiazolidinedione is a high affinity ligand for peroxisome proliferator-activated receptor gamma (PPAR gamma)." The Journal of biological chemistry **270**(22): 12953-12956.

Li, H., J. Ruan and R. Durbin (2008). "Mapping short DNA sequencing reads and calling variants using mapping quality scores." Genome research **18**(11): 1851-1858.

Lillioja, S. and A. Wilton (2009). "Agreement among type 2 diabetes linkage studies but a poor correlation with results from genome-wide association studies." Diabetologia **52**(6): 1061-1074.

Lindner, T. H., P. R. Njolstad, Y. Horikawa, L. Bostad, G. I. Bell and O. Sovik (1999). "A novel syndrome of diabetes mellitus, renal dysfunction and genital malformation associated with a partial deletion of the pseudo-POU domain of hepatocyte nuclear factor-1beta." Hum Mol Genet **8**(11): 2001-2008.

Ling, C. and L. Groop (2009). "Epigenetics: a molecular link between environmental factors and type 2 diabetes." Diabetes **58**(12): 2718-2725.

Link, C. L. and J. B. McKinlay (2009). "Disparities in the prevalence of diabetes: is it race/ethnicity or socioeconomic status? Results from the Boston Area Community Health (BACH) survey." Ethnicity & disease **19**(3): 288-292.

Liu, M., I. Hodish, C. J. Rhodes and P. Arvan (2007). "Proinsulin maturation, misfolding, and proteotoxicity." Proceedings of the National Academy of Sciences of the United States of America **104**(40): 15841-15846.

Livak, K. J. (2003). "SNP genotyping by the 5'-nuclease reaction." Methods in molecular biology **212**: 129-147.

Lyssenko, V., A. Jonsson, P. Almgren, N. Pulizzi, B. Isomaa, T. Tuomi, G. Berglund, D. Altshuler, P. Nilsson and L. Groop (2008). "Clinical risk factors, DNA variants, and the development of type 2 diabetes." The New England journal of medicine **359**(21): 2220-2232.

Lyssenko, V., T. Jorgensen, R. W. Gerwien, T. Hansen, M. W. Rowe, M. P. McKenna, J. Kolberg, O. Pedersen, K. Borch-Johnsen and L. Groop (2012). "Validation of a multi-marker model for the prediction of incident type 2 diabetes mellitus: combined results of the Inter99 and Botnia studies." Diabetes & vascular disease

research : official journal of the International Society of Diabetes and Vascular Disease **9**(1): 59-67.

Maassen, J. A., J. A. Romijn and R. J. Heine (2007). "Fatty acid-induced mitochondrial uncoupling in adipocytes as a key protective factor against insulin resistance and beta cell dysfunction: a new concept in the pathogenesis of obesity-associated type 2 diabetes mellitus." Diabetologia **50**(10): 2036-2041.

MacArthur, D. G., S. Balasubramanian, A. Frankish, N. Huang, J. Morris, K. Walter, L. Jostins, L. Habegger, J. K. Pickrell, S. B. Montgomery, C. A. Albers, Z. D. Zhang, D. F. Conrad, G. Lunter, H. Zheng, Q. Ayub, M. A. DePristo, E. Banks, M. Hu, R. E. Handsaker, J. A. Rosenfeld, M. Fromer, M. Jin, X. J. Mu, E. Khurana, K. Ye, M. Kay, G. I. Saunders, M. M. Suner, T. Hunt, I. H. Barnes, C. Amid, D. R. Carvalho-Silva, A. H. Bignell, C. Snow, B. Yngvadottir, S. Bumpstead, D. N. Cooper, Y. Xue, I. G. Romero, J. Wang, Y. Li, R. A. Gibbs, S. A. McCarroll, E. T. Dermitzakis, J. K. Pritchard, J. C. Barrett, J. Harrow, M. E. Hurles, M. B. Gerstein and C. Tyler-Smith (2012). "A systematic survey of loss-of-function variants in human protein-coding genes." Science **335**(6070): 823-828.

Macfarlane, W. M., T. M. Frayling, S. Ellard, J. C. Evans, L. I. Allen, M. P. Bulman, S. Ayres, M. Shepherd, P. Clark, A. Millward, A. Demaine, T. Wilkin, K. Docherty and A. T. Hattersley (1999). "Missense mutations in the insulin promoter factor-1 gene predispose to type 2 diabetes." J Clin Invest **104**(9): R33-39.

Magi, R., J. L. Asimit, A. G. Day-Williams, E. Zeggini and A. P. Morris (2012). "Genome-Wide Association Analysis of Imputed Rare Variants: Application to Seven Common Complex Diseases." Genetic epidemiology.

Mahtani, M. M., E. Widen, M. Lehto, J. Thomas, M. McCarthy, J. Brayer, B. Bryant, G. Chan, M. Daly, C. Forsblom, T. Kanninen, A. Kirby, L. Kruglyak, K. Munnelly, M. Parkkonen, M. P. Reeve-Daly, A. Weaver, T. Brettin, G. Duyk, E. S. Lander and L. C. Groop (1996). "Mapping of a gene for type 2 diabetes associated with an insulin secretion defect by a genome scan in Finnish families." Nat Genet **14**(1): 90-94.

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**(3): 323-328.

Manolio, T. A., F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff, D. J. Hunter, M. I. McCarthy, E. M. Ramos, L. R. Cardon, A. Chakravarti, J. H. Cho, A. E. Guttmacher, A. Kong, L. Kruglyak, E. Mardis, C. N. Rotimi, M. Slatkin, D. Valle, A. S. Whittemore, M. Boehnke, A. G. Clark, E. E. Eichler, G. Gibson, J. L. Haines, T. F. Mackay, S. A. McCarroll and P. M. Visscher (2009). "Finding the missing heritability of complex diseases." Nature **461**(7265): 747-753.

Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y. J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. Alenquer, T. P. Jarvie, K. B. Jirage, J. B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P.

McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. Yu, R. F. Begley and J. M. Rothberg (2005). "Genome sequencing in microfabricated high-density picolitre reactors." Nature **437**(7057): 376-380.

Marth, G. T., F. Yu, A. R. Indap, K. Garimella, S. Gravel, W. F. Leong, C. Tyler-Smith, M. Bainbridge, T. Blackwell, X. Zheng-Bradley, Y. Chen, D. Challis, L. Clarke, E. V. Ball, K. Cibulskis, D. N. Cooper, B. Fulton, C. Hartl, D. Koboldt, D. Muzny, R. Smith, C. Sougnez, C. Stewart, A. Ward, J. Yu, Y. Xue, D. Altshuler, C. D. Bustamante, A. G. Clark, M. Daly, M. Depristo, P. Flicek, S. Gabriel, E. Mardis, A. Palotie and R. Gibbs (2011). "The functional spectrum of low-frequency coding variation." Genome biology **12**(9): R84.

Martin, A. O., J. L. Simpson, C. Ober and N. Freinkel (1985). "Frequency of diabetes mellitus in mothers of probands with gestational diabetes: possible maternal influence on the predisposition to gestational diabetes." Am J Obstet Gynecol **151**(4): 471-475.

Matschinsky, F. M., B. Glaser and M. A. Magnuson (1998). "Pancreatic beta-cell glucokinase: closing the gap between theoretical concepts and experimental realities." Diabetes **47**(3): 307-315.

McCarthy, M. I. (2010). "Genomics, type 2 diabetes, and obesity." The New England journal of medicine **363**(24): 2339-2350.

McCarthy, M. I. and E. Zeggini (2009). "Genome-wide association studies in type 2 diabetes." Current diabetes reports **9**(2): 164-171.

McDonald, T. J., B. M. Shields, J. Lawry, K. R. Owen, A. L. Gloyn, S. Ellard and A. T. Hattersley (2011). "High-sensitivity CRP discriminates HNF1A-MODY from other subtypes of diabetes." Diabetes Care **34**(8): 1860-1862.

McLaughlin, T., C. Lamendola, A. Liu and F. Abbasi (2011). "Preferential fat deposition in subcutaneous versus visceral depots is associated with insulin sensitivity." The Journal of clinical endocrinology and metabolism **96**(11): E1756-1760.

Medici, F., M. Hawa, A. Ianari, D. A. Pyke and R. D. Leslie (1999). "Concordance rate for type II diabetes mellitus in monozygotic twins: actuarial analysis." Diabetologia **42**(2): 146-150.

Menjivar, M., M. A. Granados-Silvestre, I. Montufar-Robles, M. Herrera, M. T. Tusie-Luna, S. Canizales-Quinteros, C. A. Aguilar-Salinas and M. G. Ortiz-Lopez (2008). "High frequency of T130I mutation of HNF4A gene in Mexican patients with early-onset type 2 diabetes." Clin Genet **73**(2): 185-187.

Merikangas, K. R. and N. Risch (2003). "Genomic priorities and public health." Science **302**(5645): 599-601.

Meyer, M., U. Stenzel and M. Hofreiter (2008). "Parallel tagged sequencing on the 454 platform." Nature protocols **3**(2): 267-278.

Moller, A. M., S. A. Urhammer, L. T. Dalgaard, R. Reneland, L. Berglund, T. Hansen, J. O. Clausen, H. Lithell and O. Pedersen (1997). "Studies of the genetic variability of the coding region of the hepatocyte nuclear factor-4alpha in Caucasians with maturity onset NIDDM." Diabetologia **40**(8): 980-983.

Monsuur, A. J., P. I. de Bakker, B. Z. Alizadeh, A. Zhernakova, M. R. Bevova, E. Strengman, L. Franke, R. van't Slot, M. J. van Belzen, I. C. Lavrijsen, B. Diosdado, M. J. Daly, C. J. Mulder, M. L. Mearin, J. W. Meijer, G. A. Meijer, E. van Oort, M. C. Wapenaar, B. P. Koeleman and C. Wijmenga (2005). "Myosin IXB variant increases the risk of celiac disease and points toward a primary intestinal barrier defect." Nature genetics **37**(12): 1341-1344.

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**(3): 267-273.

Mooy, J. M., P. A. Grootenhuis, H. de Vries, H. A. Valkenburg, L. M. Bouter, P. J. Kostense and R. J. Heine (1995). "Prevalence and determinants of glucose intolerance in a Dutch caucasian population. The Hoorn Study." Diabetes care **18**(9): 1270-1273.

Morris, A. P. (2011). "Transethnic meta-analysis of genomewide association studies." Genetic epidemiology **35**(8): 809-822.

Morris, A. P., B. F. Voight, T. M. Teslovich, T. Ferreira, A. V. Segre, V. Steinthorsdottir, R. J. Strawbridge, H. Khan, H. Grallert, A. Mahajan, I. Prokopenko, H. M. Kang, C. Dina, T. Esko, R. M. Fraser, S. Kanoni, A. Kumar, V. Lagou, C. Langenberg, J. Luan, C. M. Lindgren, M. Muller-Nurasyid, S. Pechlivanis, N. W. Rayner, L. J. Scott, S. Wiltshire, L. Yengo, L. Kinnunen, E. J. Rossin, S. Raychaudhuri, A. D. Johnson, A. S. Dimas, R. J. Loos, S. Vedantam, H. Chen, J. C. Florez, C. Fox, C. T. Liu, D. Rybin, D. J. Couper, W. H. Kao, M. Li, M. C. Cornelis, P. Kraft, Q. Sun, R. M. van Dam, H. M. Stringham, P. S. Chines, K. Fischer, P. Fontanillas, O. L. Holmen, S. E. Hunt, A. U. Jackson, A. Kong, R. Lawrence, J. Meyer, J. R. Perry, C. G. Platou, S. Potter, E. Rehnberg, N. Robertson, S. Sivapalaratnam, A. Stancakova, K. Stirrups, G. Thorleifsson, E. Tikkanen, A. R. Wood, P. Almgren, M. Atalay, R. Benediktsson, L. L. Bonnycastle, N. Burtt, J. Carey, G. Charpentier, A. T. Crenshaw, A. S. Doney, M. Dorkhan, S. Edkins, V. Emilsson, E. Eury, T. Forsen, K. Gertow, B. Gigante, G. B. Grant, C. J. Groves, C. Guiducci, C. Herder, A. B. Hreidarsson, J. Hui, A. James, A. Jonsson, W. Rathmann, N. Klopp, J. Kravic, K. Krjutskov, C. Langford, K. Leander, E. Lindholm, S. Lobbens, S. Mannisto, G. Mirza, T. W. Muhleisen, B. Musk, M. Parkin, L. Rallidis, J. Saramies, B. Sennblad, S. Shah, G. Sigurethsson, A. Silveira, G. Steinbach, B. Thorand, J. Trakalo, F. Veglia, R. Wennauer, W. Winckler, D. Zabaneh, H. Campbell, C. van Duijn, A. G. Uitterlinden, A. Hofman, E. Sijbrands, G. R. Abecasis, K. R. Owen, E. Zeggini, M. D. Trip, N. G. Forouhi, A. C. Syvanen, J. G. Eriksson, L. Peltonen, M. M. Nothen, B. Balkau, C. N. Palmer, V. Lyssenko, T. Tuomi, B. Isomaa, D. J. Hunter, L. Qi, A. R. Shuldiner, M. Roden, I. Barroso, T. Wilsgaard, J. Beilby, K. Hovingh, J. F. Price, J. F. Wilson, R. Rauramaa, T. A. Lakka, L. Lind, G. Dedoussis, I. Njolstad, N. L. Pedersen, K. T. Khaw, N. J. Wareham, S. M. Keinanen-Kiukaanniemi, T. E. Saaristo, E. Korpi-Hyovalti, J. Saltevo, M. Laakso, J. Kuusisto, A. Metspalu, F. S. Collins, K. L. Mohlke, R. N. Bergman, J. Tuomilehto, B. O. Boehm, C. Gieger, K. Hveem, S. Cauchi, P. Froguel, D. Baldassarre, E. Tremoli,

S. E. Humphries, D. Saleheen, J. Danesh, E. Ingelsson, S. Ripatti, V. Salomaa, R. Erbel, K. H. Jockel, S. Moebus, A. Peters, T. Illig, U. de Faire, A. Hamsten, A. D. Morris, P. J. Donnelly, T. M. Frayling, A. T. Hattersley, E. Boerwinkle, O. Melander, S. Kathiresan, P. M. Nilsson, P. Deloukas, U. Thorsteinsdottir, L. C. Groop, K. Stefansson, F. Hu, J. S. Pankow, J. Dupuis, J. B. Meigs, D. Altshuler, M. Boehnke and M. I. McCarthy (2012). "Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes." Nature Genetics **44**(9): 981-990.

Muller, Y. L., A. M. Infante, R. L. Hanson, L. Love-Gregory, W. Knowler, C. Bogardus and L. J. Baier (2005). "Variants in hepatocyte nuclear factor 4alpha are modestly associated with type 2 diabetes in Pima Indians." Diabetes **54**(10): 3035-3039.

Murphy, R., S. Ellard and A. T. Hattersley (2008). "Clinical implications of a molecular genetic classification of monogenic beta-cell diabetes." Nature clinical practice. Endocrinology & metabolism **4**(4): 200-213.

Nadeau, J. H. (2009). "Transgenerational genetic effects on phenotypic variation and disease risk." Human molecular genetics **18**(R2): R202-210.

Navas, M. A., E. J. Munoz-Elias, J. Kim, D. Shih and M. Stoffel (1999). "Functional characterization of the MODY1 gene mutations HNF4(R127W), HNF4(V255M), and HNF4(E276Q)." Diabetes **48**(7): 1459-1465.

Neale, B. M., M. A. Rivas, B. F. Voight, D. Altshuler, B. Devlin, M. Orho-Melander, S. Kathiresan, S. M. Purcell, K. Roeder and M. J. Daly (2011). "Testing for an unusual distribution of rare variants." PLoS genetics **7**(3): e1001322.

Nejentsev, S., N. Walker, D. Riches, M. Egholm and J. A. Todd (2009). "Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes." Science **324**(5925): 387-389.

Newman, B., J. V. Selby, M. C. King, C. Slemenda, R. Fabsitz and G. D. Friedman (1987). "Concordance for type 2 (non-insulin-dependent) diabetes mellitus in male twins." Diabetologia **30**(10): 763-768.

Okamoto, K., N. Iwasaki, C. Nishimura, K. Doi, E. Noiri, S. Nakamura, M. Takizawa, M. Ogata, R. Fujimaki, N. Grarup, C. Pisinger, K. Borch-Johnsen, T. Lauritzen, A. Sandbaek, T. Hansen, K. Yasuda, H. Osawa, K. Nanjo, T. Kadowaki, M. Kasuga, O. Pedersen, T. Fujita, N. Kamatani, Y. Iwamoto and K. Tokunaga (2010). "Identification of KCNJ15 as a susceptibility gene in Asian patients with type 2 diabetes mellitus." American journal of human genetics **86**(1): 54-64.

Osbak, K. K., K. Colclough, C. Saint-Martin, N. L. Beer, C. Bellanne-Chantelot, S. Ellard and A. L. Gloyn (2009). "Update on mutations in glucokinase (GCK), which cause maturity-onset diabetes of the young, permanent neonatal diabetes, and hyperinsulinemic hypoglycemia." Human mutation **30**(11): 1512-1526.

Owen, K. R., J. C. Evans, T. M. Frayling, A. T. Hattersley, M. I. McCarthy, M. Walker and G. A. Hitman (2004). "Role of the D76N polymorphism of insulin promoter factor-1 in predisposing to Type 2 diabetes." Diabetologia **47**(5): 957-958.

Owen, K. R., G. Thanabalasingham, T. J. James, F. Karpe, A. J. Farmer, M. I. McCarthy and A. L. Gloyn (2010). "Assessment of high-sensitivity C-reactive protein levels as

diagnostic discriminator of maturity-onset diabetes of the young due to HNF1A mutations." Diabetes Care **33**(9): 1919-1924.

Park, J. H., S. Wacholder, M. H. Gail, U. Peters, K. B. Jacobs, S. J. Chanock and N. Chatterjee (2010). "Estimation of effect size distribution from genome-wide association studies and implications for future discoveries." Nature Genetics **42**(7): 570-575.

Pearson, E. R., M. K. Badman, C. R. Lockwood, P. M. Clark, S. Ellard, C. Bingham and A. T. Hattersley (2004). "Contrasting diabetes phenotypes associated with hepatocyte nuclear factor-1alpha and -1beta mutations." Diabetes care **27**(5): 1102-1107.

Pearson, E. R., S. F. Boj, A. M. Steele, T. Barrett, K. Stals, J. P. Shield, S. Ellard, J. Ferrer and A. T. Hattersley (2007). "Macrosomia and hyperinsulinaemic hypoglycaemia in patients with heterozygous mutations in the HNF4A gene." PLoS medicine **4**(4): e118.

Pearson, E. R., L. A. Donnelly, C. Kimber, A. Whitley, A. S. Doney, M. I. McCarthy, A. T. Hattersley, A. D. Morris and C. N. Palmer (2007). "Variation in TCF7L2 influences therapeutic response to sulfonylureas: a GoDARTs study." Diabetes **56**(8): 2178-2182.

Pearson, E. R., I. Flechtner, P. R. Njolstad, M. T. Malecki, S. E. Flanagan, B. Larkin, F. M. Ashcroft, I. Klimes, E. Codner, V. Iotova, A. S. Slingerland, J. Shield, J. J. Robert, J. J. Holst, P. M. Clark, S. Ellard, O. Sovik, M. Polak and A. T. Hattersley (2006). "Switching from insulin to oral sulfonylureas in patients with diabetes due to Kir6.2 mutations." The New England journal of medicine **355**(5): 467-477.

Pearson, E. R., B. J. Starkey, R. J. Powell, F. M. Gribble, P. M. Clark and A. T. Hattersley (2003). "Genetic cause of hyperglycaemia and response to treatment in diabetes." Lancet **362**(9392): 1275-1281.

Permutt, M. A., J. Wasson and N. Cox (2005). "Genetic epidemiology of diabetes." J Clin Invest **115**(6): 1431-1439.

Pertea, M., X. Lin and S. L. Salzberg (2001). "GeneSplicer: a new computational method for splice site prediction." Nucleic Acids Res **29**(5): 1185-1190.

Pettitt, D. J., K. A. Aleck, H. R. Baird, M. J. Carraher, P. H. Bennett and W. C. Knowler (1988). "Congenital susceptibility to NIDDM. Role of intrauterine environment." Diabetes **37**(5): 622-628.

Poulsen, P., K. O. Kyvik, A. Vaag and H. Beck-Nielsen (1999). "Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance--a population-based twin study." Diabetologia **42**(2): 139-145.

Prokopenko, I., C. Langenberg, J. C. Florez, R. Saxena, N. Soranzo, G. Thorleifsson, R. J. Loos, A. K. Manning, A. U. Jackson, Y. Aulchenko, S. C. Potter, M. R. Erdos, S. Sanna, J. J. Hottenga, E. Wheeler, M. Kaakinen, V. Lyssenko, W. M. Chen, K. Ahmadi, J. S. Beckmann, R. N. Bergman, M. Bochud, L. L. Bonnycastle, T. A. Buchanan, A. Cao, A. Cervino, L. Coin, F. S. Collins, L. Crisponi, E. J. de Geus, A. Dehghan, P. Deloukas, A. S. Doney, P. Elliott, N. Freimer, V. Gateva, C. Herder, A. Hofman, T. E. Hughes, S. Hunt, T. Illig, M. Inouye, B. Isomaa, T. Johnson, A. Kong, M. Krestyaninova, J. Kuusisto, M. Laakso, N. Lim, U. Lindblad, C. M. Lindgren, O. T. McCann, K. L. Mohlke, A. D. Morris, S. Naitza, M. Orru, C. N. Palmer, A. Pouta,

J. Randall, W. Rathmann, J. Saramies, P. Scheet, L. J. Scott, A. Scuteri, S. Sharp, E. Sijbrands, J. H. Smit, K. Song, V. Steinthorsdottir, H. M. Stringham, T. Tuomi, J. Tuomilehto, A. G. Uitterlinden, B. F. Voight, D. Waterworth, H. E. Wichmann, G. Willemsen, J. C. Witteman, X. Yuan, J. H. Zhao, E. Zeggini, D. Schlessinger, M. Sandhu, D. I. Boomsma, M. Uda, T. D. Spector, B. W. Penninx, D. Altshuler, P. Vollenweider, M. R. Jarvelin, E. Lakatta, G. Waeber, C. S. Fox, L. Peltonen, L. C. Groop, V. Mooser, L. A. Cupples, U. Thorsteinsdottir, M. Boehnke, I. Barroso, C. Van Duijn, J. Dupuis, R. M. Watanabe, K. Stefansson, M. I. McCarthy, N. J. Wareham, J. B. Meigs and G. R. Abecasis (2009). "Variants in MTNR1B influence fasting glucose levels." Nature genetics **41**(1): 77-81.

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." Human molecular genetics **15**(11): 1793-1800.

Proks, P., C. Girard, H. Baevre, P. R. Njolstad and F. M. Ashcroft (2006). "Functional effects of mutations at F35 in the NH2-terminus of Kir6.2 (KCNJ11), causing neonatal diabetes, and response to sulfonylurea therapy." Diabetes **55**(6): 1731-1737.

Qi, L., M. C. Cornelis, P. Kraft, K. J. Stanya, W. H. Linda Kao, J. S. Pankow, J. Dupuis, J. C. Florez, C. S. Fox, G. Pare, Q. Sun, C. J. Girman, C. C. Laurie, D. B. Mirel, T. A. Manolio, D. I. Chasman, E. Boerwinkle, P. M. Ridker, D. J. Hunter, J. B. Meigs, C. H. Lee, F. B. Hu and R. M. van Dam (2010). "Genetic variants at 2q24 are associated with susceptibility to type 2 diabetes." Human molecular genetics **19**(13): 2706-2715.

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**(1): 54-62.

Ramachandran, A., C. Snehalatha, R. Vinitha, M. Thayyil, C. K. Kumar, L. Sheeba, S. Joseph and V. Vijay (2002). "Prevalence of overweight in urban Indian adolescent school children." Diabetes research and clinical practice **57**(3): 185-190.

Reich, D. E., M. Cargill, S. Bolk, J. Ireland, P. C. Sabeti, D. J. Richter, T. Lavery, R. Kouyoumjian, S. F. Farhadian, R. Ward and E. S. Lander (2001). "Linkage disequilibrium in the human genome." Nature **411**(6834): 199-204.

Reich, D. E. and E. S. Lander (2001). "On the allelic spectrum of human disease." Trends Genet **17**(9): 502-510.

Reiling, E., J. V. van Vliet-Ostaptchouk, E. van 't Riet, T. W. van Haeften, P. A. Arp, T. Hansen, D. Kremer, M. J. Groenewoud, E. C. van Hove, J. A. Romijn, J. W. Smit, G. Nijpels, R. J. Heine, A. G. Uitterlinden, O. Pedersen, P. E. Slagboom, J. A. Maassen, M. H. Hofker, L. M. t Hart and J. M. Dekker (2009). "Genetic association analysis of 13 nuclear-encoded mitochondrial candidate genes with type II diabetes mellitus: the DAMAGE study." European journal of human genetics : EJHG **17**(8): 1056-1062.

Reiner, A. P., M. J. Barber, Y. Guan, P. M. Ridker, L. A. Lange, D. I. Chasman, J. D. Walston, G. M. Cooper, N. S. Jenny, M. J. Rieder, J. P. Durda, J. D. Smith, J. Novembre, R. P. Tracy, J. I. Rotter, M. Stephens, D. A. Nickerson and R. M. Krauss (2008). "Polymorphisms of the HNF1A gene encoding hepatocyte nuclear factor-1 alpha are associated with C-reactive protein." American journal of human genetics **82**(5): 1193-1201.

Rhee, J., Y. Inoue, J. C. Yoon, P. Puigserver, M. Fan, F. J. Gonzalez and B. M. Spiegelman (2003). "Regulation of hepatic fasting response by PPARgamma coactivator-1alpha (PGC-1): requirement for hepatocyte nuclear factor 4alpha in gluconeogenesis." Proc Natl Acad Sci U S A **100**(7): 4012-4017.

Ridker, P. M., G. Pare, A. Parker, R. Y. Zee, J. S. Danik, J. E. Buring, D. Kwiatkowski, N. R. Cook, J. P. Miletich and D. I. Chasman (2008). "Loci related to metabolic-syndrome pathways including LEPR,HNF1A, IL6R, and GCKR associate with plasma C-reactive protein: the Women's Genome Health Study." American journal of human genetics **82**(5): 1185-1192.

Risch, N. and K. Merikangas (1996). "The future of genetic studies of complex human diseases." Science **273**(5281): 1516-1517.

Rissanen, J., H. Wang, R. Miettinen, P. Karkkainen, P. Kekalainen, L. Mykkanen, J. Kuusisto, P. Karhapaa, L. Niskanen, M. Uusitupa and M. Laakso (2000). "Variants in the hepatocyte nuclear factor-1alpha and -4alpha genes in Finnish and Chinese subjects with late-onset type 2 diabetes." Diabetes care **23**(10): 1533-1538.

Rivas, M. A., M. Beaudoin, A. Gardet, C. Stevens, Y. Sharma, C. K. Zhang, G. Boucher, S. Ripke, D. Ellinghaus, N. Burtt, T. Fennell, A. Kirby, A. Latiano, P. Goyette, T. Green, J. Halfvarson, T. Haritunians, J. M. Korn, F. Kuruvilla, C. Lagace, B. Neale, K. S. Lo, P. Schumm, L. Torkvist, M. C. Dubinsky, S. R. Brant, M. S. Silverberg, R. H. Duerr, D. Altshuler, S. Gabriel, G. Lettre, A. Franke, M. D'Amato, D. P. McGovern, J. H. Cho, J. D. Rioux, R. J. Xavier and M. J. Daly (2011). "Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease." Nature genetics **43**(11): 1066-1073.

Rockett, I. (1994). "Population and Health: An introduction to epidimiology." Population Bull **49**(3): 11.

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**(4): 513-516.

Ronn, T., P. Poulsen, O. Hansson, J. Holmkvist, P. Almgren, P. Nilsson, T. Tuomi, B. Isomaa, L. Groop, A. Vaag and C. Ling (2008). "Age influences DNA methylation and gene expression of COX7A1 in human skeletal muscle." Diabetologia **51**(7): 1159-1168.

Rung, J., S. Cauchi, A. Albrechtsen, L. Shen, G. Rocheleau, C. Cavalcanti-Proenca, F. Bacot, B. Balkau, A. Belisle, K. Borch-Johnsen, G. Charpentier, C. Dina, E. Durand, P. Elliott, S. Hadjadj, M. R. Jarvelin, J. Laitinen, T. Lauritzen, M. Marre, A. Mazur, D. Meyre, A. Montpetit, C. Pisinger, B. Posner, P. Poulsen, A. Pouta, M. Prentki, R. Ribel-Madsen, A. Ruokonen, A. Sandbaek, D. Serre, J. Tichet, M. Vaxillaire, J. F.

Wojtaszewski, A. Vaag, T. Hansen, C. Polychronakos, O. Pedersen, P. Froguel and R. Sladek (2009). "Genetic variant near IRS1 is associated with type 2 diabetes, insulin resistance and hyperinsulinemia." Nature genetics **41**(10): 1110-1115.

Sabeti, P. C., P. Varilly, B. Fry, J. Lohmueller, E. Hostetter, C. Cotsapas, X. Xie, E. H. Byrne, S. A. McCarroll, R. Gaudet, S. F. Schaffner, E. S. Lander, K. A. Frazer, 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, 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, W. Sun, H. 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, P. C. Sham, 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. A. Johnson, J. C. Mullikin, S. T. Sherry, M. Feolo, A. Skol, H. Zhang, 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, D. Muzny, L. Nazareth, E. Sodergren, G. M. Weinstock, I. Yakub, B. W. Birren, 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, 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). "Genome-wide detection and characterization of positive selection in human populations." Nature **449**(7164): 913-918.

Sacks, D. B., M. Arnold, G. L. Bakris, D. E. Bruns, A. R. Horvath, M. S. Kirkman, A. Lernmark, B. E. Metzger and D. M. Nathan (2011). "Guidelines and

recommendations for laboratory analysis in the diagnosis and management of diabetes mellitus." Diabetes Care **34**(6): e61-99.

Sagen, J. V., H. Raeder, E. Hathout, N. Shehadeh, K. Gudmundsson, H. Baevre, D. Abuelo, C. Phornphutkul, J. Molnes, G. I. Bell, A. L. Gloyn, A. T. Hattersley, A. Molven, O. Sovik and P. R. Njolstad (2004). "Permanent neonatal diabetes due to mutations in KCNJ11 encoding Kir6.2: patient characteristics and initial response to sulfonylurea therapy." Diabetes **53**(10): 2713-2718.

Sanger, F., S. Nicklen and A. R. Coulson (1977). "DNA sequencing with chain-terminating inhibitors." Proc Natl Acad Sci U S A **74**(12): 5463-5467.

Sanger, F., S. Nicklen and A. R. Coulson (1977). "DNA sequencing with chain-terminating inhibitors." Proceedings of the National Academy of Sciences of the United States of America **74**(12): 5463-5467.

Saxena, R., Clara C. Elbers, Y. Guo, I. Peter, Tom R. Gaunt, Jessica L. Mega, Matthew B. Lanktree, A. Tare, Berta A. Castillo, Yun R. Li, T. Johnson, M. Bruinenberg, D. Gilbert-Diamond, R. Rajagopalan, Benjamin F. Voight, A. Balasubramanyam, J. Barnard, F. Bauer, J. Baumert, T. Bhangale, Bernhard O. Böhm, Peter S. Braund, Paul R. Burton, Hareesh R. Chandrupatla, R. Clarke, Rhonda M. Cooper-DeHoff, Errol D. Crook, G. Davey-Smith, Ian N. Day, A. de Boer, Mark C. H. de Groot, F. Drenos, J. Ferguson, Caroline S. Fox, Clement E. Furlong, Q. Gibson, C. Gieger, Lisa A. Gilhuijs-Pederson, Joseph T. Glessner, A. Goel, Y. Gong, Struan F. A. Grant, Diederick E. Grobbee, C. Hastie, Steve E. Humphries, Cecilia E. Kim, M. Kivimaki, M. Kleber, C. Meisinger, M. Kumari, Taimour Y. Langaee, Debbie A. Lawlor, M. Li, Maximilian T. Lobmeyer, A.-H. Maitland-van der Zee, Matthijs F. L. Meijs, Cliona M. Molony, David A. Morrow, G. Murugesan, Solomon K. Musani, Christopher P. Nelson, Stephen J. Newhouse, Jeffery R. O'Connell, S. Padmanabhan, J. Palmen, Sanjey R. Patel, Carl J. Pepine, M. Pettinger, Thomas S. Price, S. Rafelt, J. Ranchalis, A. Rasheed, E. Rosenthal, I. Ruczinski, S. Shah, H. Shen, G. Silbernagel, Erin N. Smith, Annemieke W. M. Spijkerman, A. Stanton, Michael W. Steffes, B. Thorand, M. Trip, P. van der Harst, Daphne L. van der A, Erik P. A. van Iperen, J. van Setten, Jana V. van Vliet-Ostaptchouk, N. Verweij, Bruce H. R. Wolffenbuttel, T. Young, M. H. Zafarmand, Joseph M. Zmuda, M. Boehnke, D. Altshuler, M. McCarthy, W. H. L. Kao, James S. Pankow, Thomas P. Cappola, M. Caulfield, A. Dominiczak, Denis C. Shields, D. Bhatt, L. Zhang, Sean P. Curtis, J. Danesh, Juan P. Casas, Yvonne T. van der Schouw, N. C. Onland-Moret, Pieter A. Doevendans, Gerald W. Dorn Ii, M. Farrall, Garret A. FitzGerald, A. Hamsten, R. Hegele, Aroon D. Hingorani, Marten H. Hofker, Gordon S. Huggins, T. Illig, Gail P. Jarvik, Julie A. Johnson, Olaf H. Klungel, William C. Knowler, W. Koenig, W. März, James B. Meigs, O. Melander, Patricia B. Munroe, Braxton D. Mitchell, Susan J. Bielinski, Daniel J. Rader, Muredach P. Reilly, Stephen S. Rich, Jerome I. Rotter, D. Saleheen, Nilesh J. Samani, Eric E. Schadt, Alan R. Shuldiner, R. Silverstein, K. Kottke-Marchant, Philippa J. Talmud, H. Watkins, F. Asselbergs, Paul I. W. de Bakker, J. McCaffery, C. Wijmenga, Marc S. Sabatine, James G. Wilson, A. Reiner, Donald W. Bowden, H. Hakonarson, David S. Siscovick and Brendan J. Keating (2012). "Large-Scale Gene-Centric Meta-Analysis across 39

studies Identifies Type 2 Diabetes Loci." The American Journal of Human Genetics(0).

Saxena, R., L. Gianniny, N. P. Burtt, V. Lyssenko, C. Giuducci, M. Sjogren, J. C. Florez, P. Almgren, B. Isomaa, M. Orho-Melander, U. Lindblad, M. J. Daly, T. Tuomi, J. N. Hirschhorn, K. G. Ardlie, L. C. Groop and D. Altshuler (2006). "Common single nucleotide polymorphisms in TCF7L2 are reproducibly associated with type 2 diabetes and reduce the insulin response to glucose in nondiabetic individuals." Diabetes **55**(10): 2890-2895.

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**(5829): 1331-1336.

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**(5829): 1341-1345.

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**(2): 153-156.

Shaw, J. T., P. K. Lovelock, J. B. Kesting, J. Cardinal, D. Duffy, B. Wainwright and D. P. Cameron (1998). "Novel susceptibility gene for late-onset NIDDM is localized to human chromosome 12q." Diabetes **47**(11): 1793-1796.

Shields, B. M., S. Hicks, M. H. Shepherd, K. Colclough, A. T. Hattersley and S. Ellard (2010). "Maturity-onset diabetes of the young (MODY): how many cases are we missing?" Diabetologia **53**(12): 2504-2508.

Shu, X. O., J. Long, Q. Cai, L. Qi, Y.-B. Xiang, Y. S. Cho, E. S. Tai, X. Li, X. Lin, W.-H. Chow, M. J. Go, M. Seielstad, W. Bao, H. Li, M. C. Cornelis, K. Yu, W. Wen, J. Shi, B.-G. Han, X. L. Sim, L. Liu, Q. Qi, H.-L. Kim, D. P. K. Ng, J.-Y. Lee, Y. J. Kim, C. Li, Y.-T. Gao, W. Zheng and F. B. Hu (2010). "Identification of New Genetic Risk Variants for Type 2 Diabetes." PLoS genetics **6**(9): e1001127.

Skol, A. D., L. J. Scott, G. R. Abecasis and M. Boehnke (2006). "Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies." Nature genetics **38**(2): 209-213.

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**(7130): 881-885.

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**(6): 617-619.

Sobngwi, E., P. Boudou, F. Mauvais-Jarvis, H. Leblanc, G. Velho, P. Vexiau, R. Porcher, S. Hadjadj, R. Pratley, P. A. Tataranni, F. Calvo and J. F. Gautier (2003). "Effect of a diabetic environment in utero on predisposition to type 2 diabetes." Lancet **361**(9372): 1861-1865.

Sookoian, S., C. Gemma and C. J. Pirola (2009). "Influence of hepatocyte nuclear factor 4alpha (HNF4alpha) gene variants on the risk of type 2 diabetes: A meta-analysis in 49,577 individuals." Mol Genet Metab.

Sookoian, S., C. Gemma and C. J. Pirola (2010). "Influence of hepatocyte nuclear factor 4alpha (HNF4alpha) gene variants on the risk of type 2 diabetes: a meta-analysis in 49,577 individuals." Mol Genet Metab **99**(1): 80-89.

Stahl, E. A., D. Wegmann, G. Trynka, J. Gutierrez-Achury, R. Do, B. F. Voight, P. Kraft, R. Chen, H. J. Kallberg, F. A. Kurreeman, S. Kathiresan, C. Wijmenga, P. K. Gregersen, L. Alfredsson, K. A. Siminovitch, J. Worthington, P. I. de Bakker, S. Raychaudhuri and R. M. Plenge (2012). "Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis." Nature Genetics **44**(5): 483-489.

Stancakova, A., M. Javorsky, T. Kuulasmaa, S. M. Haffner, J. Kuusisto and M. Laakso (2009). "Changes in insulin sensitivity and insulin release in relation to glycemia and glucose tolerance in 6,414 Finnish men." Diabetes **58**(5): 1212-1221.

Stanik, J., D. Gasperikova, M. Paskova, L. Barak, J. Javorkova, E. Jancova, M. Ciljakova, P. Hlava, J. Michalek, S. E. Flanagan, E. Pearson, A. T. Hattersley, S. Ellard and I. Klimes (2007). "Prevalence of permanent neonatal diabetes in Slovakia and successful replacement of insulin with sulfonylurea therapy in KCNJ11 and ABCC8 mutation carriers." The Journal of clinical endocrinology and metabolism **92**(4): 1276-1282.

Steele, A. M., B. M. Shields, M. Shepherd, S. Ellard, A. T. Hattersley and E. R. Pearson (2010). "Increased all-cause and cardiovascular mortality in monogenic diabetes as a result of mutations in the HNF1A gene." Diabetic medicine : a journal of the British Diabetic Association **27**(2): 157-161.

Steinbrecher, A., Y. Morimoto, S. Heak, N. J. Ollberding, K. S. Geller, A. Grandinetti, L. N. Kolonel and G. Maskarinec (2011). "The preventable proportion of type 2

diabetes by ethnicity: the multiethnic cohort." Annals of epidemiology **21**(7): 526-535.
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**(2): 138-139.
Stoffers, D. A., N. T. Zinkin, V. Stanojevic, W. L. Clarke and J. F. Habener (1997). "Pancreatic agenesis attributable to a single nucleotide deletion in the human IPF1 gene coding sequence." Nature genetics **15**(1): 106-110.
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." Proceedings of the National Academy of Sciences of the United States of America **104**(38): 15040-15044.
Stoy, J., D. F. Steiner, S. Y. Park, H. Ye, L. H. Philipson and G. I. Bell (2010). "Clinical and molecular genetics of neonatal diabetes due to mutations in the insulin gene." Reviews in endocrine & metabolic disorders **11**(3): 205-215.
t Hart, L. M., T. Hansen, I. Rietveld, J. M. Dekker, G. Nijpels, G. M. Janssen, P. A. Arp, A. G. Uitterlinden, T. Jorgensen, K. Borch-Johnsen, H. A. Pols, O. Pedersen, C. M. van Duijn, R. J. Heine and J. A. Maassen (2005). "Evidence that the mitochondrial leucyl tRNA synthetase (LARS2) gene represents a novel type 2 diabetes susceptibility gene." Diabetes **54**(6): 1892-1895.
Temple, I. K., R. J. Gardner, D. J. Mackay, J. C. Barber, D. O. Robinson and J. P. Shield (2000). "Transient neonatal diabetes: widening the understanding of the etiopathogenesis of diabetes." Diabetes **49**(8): 1359-1366.
Thanabalasingham, G., N. Shah, M. Vaxillaire, T. Hansen, T. Tuomi, D. Gasperikova, M. Szopa, E. Tjora, T. J. James, P. Kokko, F. Loiseleur, E. Andersson, S. Gaget, B. Isomaa, N. Nowak, H. Raeder, J. Stanik, P. R. Njolstad, M. T. Malecki, I. Klimes, L. Groop, O. Pedersen, P. Froguel, M. I. McCarthy, A. L. Gloyn and K. R. Owen (2011). "A large multi-centre European study validates high-sensitivity C-reactive protein (hsCRP) as a clinical biomarker for the diagnosis of diabetes subtypes." Diabetologia **54**(11): 2801-2810.
The 1000 Genomes Project Consortium (2010). "A map of human genome variation from population-scale sequencing." Nature **467**(7319): 1061-1073.
Triggs-Raine, B. L., R. D. Kirkpatrick, S. L. Kelly, L. D. Norquay, P. A. Cattini, K. Yamagata, A. J. Hanley, B. Zinman, S. B. Harris, P. H. Barrett and R. A. Hegele (2002). "HNF-1alpha G319S, a transactivation-deficient mutant, is associated with altered dynamics of diabetes onset in an Oji-Cree community." Proc Natl Acad Sci U S A **99**(7): 4614-4619.
Tsai, F. J., C. F. Yang, C. C. Chen, L. M. Chuang, C. H. Lu, C. T. Chang, T. Y. Wang, R. H. Chen, C. F. Shiu, Y. M. Liu, C. C. Chang, P. Chen, C. H. Chen, C. S. Fann, Y. T. Chen and J. Y. Wu (2010). "A genome-wide association study identifies susceptibility variants for type 2 diabetes in Han Chinese." PLoS genetics **6**(2): e1000847.
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." Nature genetics **40**(9): 1098-1102.

Urhammer, S. A., S. K. Rasmussen, P. J. Kaisaki, N. Oda, K. Yamagata, A. M. Moller, M. Fridberg, L. Hansen, T. Hansen, G. I. Bell and O. Pedersen (1997). "Genetic variation in the hepatocyte nuclear factor-1 alpha gene in Danish Caucasians with late-onset NIDDM." Diabetologia **40**(4): 473-475.

van 't Riet, E., M. Alssema, J. M. Rijkelijkhuizen, P. J. Kostense, G. Nijpels and J. M. Dekker (2010). "Relationship between A1C and glucose levels in the general Dutch population: the new Hoorn study." Diabetes care **33**(1): 61-66.

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**(5): 368-371.

van Tilburg, J. H., L. A. Sandkuijl, E. Strengman, H. van Someren, C. A. Rigters-Aris, P. L. Pearson, T. W. van Haeften and C. Wijmenga (2003). "A genome-wide scan in type 2 diabetes mellitus provides independent replication of a susceptibility locus on 18p11 and suggests the existence of novel Loci on 2q12 and 19q13." The Journal of clinical endocrinology and metabolism **88**(5): 2223-2230.

Voight, B. F., H. M. Kang, J. Ding, C. D. Palmer, C. Sidore, P. S. Chines, N. P. Burtt, C. Fuchsberger, Y. Li, J. Erdmann, T. M. Frayling, I. M. Heid, A. U. Jackson, T. Johnson, T. O. Kilpelainen, C. M. Lindgren, A. P. Morris, I. Prokopenko, J. C. Randall, R. Saxena, N. Soranzo, E. K. Speliotes, T. M. Teslovich, E. Wheeler, J. Maguire, M. Parkin, S. Potter, N. W. Rayner, N. Robertson, K. Stirrups, W. Winckler, S. Sanna, A. Mulas, R. Nagaraja, F. Cucca, I. Barroso, P. Deloukas, R. J. Loos, S. Kathiresan, P. B. Munroe, C. Newton-Cheh, A. Pfeufer, N. J. Samani, H. Schunkert, J. N. Hirschhorn, D. Altshuler, M. I. McCarthy, G. R. Abecasis and M. Boehnke (2012). "The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits." PLoS genetics **8**(8): e1002793.

Voight, B. F., L. J. Scott, V. Steinthorsdottir, A. P. Morris, C. Dina, R. P. Welch, E. Zeggini, C. Huth, Y. S. Aulchenko, G. Thorleifsson, L. J. McCulloch, T. Ferreira, H. Grallert, N. Amin, G. Wu, C. J. Willer, S. Raychaudhuri, S. A. McCarroll, C. Langenberg, O. M. Hofmann, J. Dupuis, L. Qi, A. V. Segre, M. van Hoek, P. Navarro, K. Ardlie, B. Balkau, R. Benediktsson, A. J. Bennett, R. Blagieva, E. Boerwinkle, L. L. Bonnycastle, K. Bengtsson Bostrom, B. Bravenboer, S. Bumpstead, N. P. Burtt, G. Charpentier, P. S. Chines, M. Cornelis, D. J. Couper, G. Crawford, A. S. Doney, K. S. Elliott, A. L. Elliott, M. R. Erdos, C. S. Fox, C. S. Franklin, M. Ganser, C. Gieger, N. Grarup, T. Green, S. Griffin, C. J. Groves, C. Guiducci, S. Hadjadj, N. Hassanali, C. Herder, B. Isomaa, A. U. Jackson, P. R. Johnson, T. Jorgensen, W. H. Kao, N. Klopp, A. Kong, P. Kraft, J. Kuusisto, T. Lauritzen, M. Li, A. Lieverse, C. M. Lindgren, V. Lyssenko, M. Marre, T. Meitinger, K. Midthjell, M. A. Morken, N.

Narisu, P. Nilsson, K. R. Owen, F. Payne, J. R. Perry, A. K. Petersen, C. Platou, C. Proenca, I. Prokopenko, W. Rathmann, N. W. Rayner, N. R. Robertson, G. Rocheleau, M. Roden, M. J. Sampson, R. Saxena, B. M. Shields, P. Shrader, G. Sigurdsson, T. Sparso, K. Strassburger, H. M. Stringham, Q. Sun, A. J. Swift, B. Thorand, J. Tichet, T. Tuomi, R. M. van Dam, T. W. van Haeften, T. van Herpt, J. V. van Vliet-Ostaptchouk, G. B. Walters, M. N. Weedon, C. Wijmenga, J. Witteman, R. N. Bergman, S. Cauchi, F. S. Collins, A. L. Gloyn, U. Gyllensten, T. Hansen, W. A. Hide, G. A. Hitman, A. Hofman, D. J. Hunter, K. Hveem, M. Laakso, K. L. Mohlke, A. D. Morris, C. N. Palmer, P. P. Pramstaller, I. Rudan, E. Sijbrands, L. D. Stein, J. Tuomilehto, A. Uitterlinden, M. Walker, N. J. Wareham, R. M. Watanabe, G. R. Abecasis, B. O. Boehm, H. Campbell, M. J. Daly, A. T. Hattersley, F. B. Hu, J. B. Meigs, J. S. Pankow, O. Pedersen, H. E. Wichmann, I. Barroso, J. C. Florez, T. M. Frayling, L. Groop, R. Sladek, U. Thorsteinsdottir, J. F. Wilson, T. Illig, P. Froguel, C. M. van Duijn, K. Stefansson, D. Altshuler, M. Boehnke and M. I. McCarthy (2010). "Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis." Nature genetics **42**(7): 579-589.

Wang, J., J. Kuusisto, M. Vanttinen, T. Kuulasmaa, J. Lindstrom, J. Tuomilehto, M. Uusitupa and M. Laakso (2007). "Variants of transcription factor 7-like 2 (TCF7L2) gene predict conversion to type 2 diabetes in the Finnish Diabetes Prevention Study and are associated with impaired glucose regulation and impaired insulin secretion." Diabetologia **50**(6): 1192-1200.

Wang, J., T. Takeuchi, S. Tanaka, S. K. Kubo, T. Kayo, D. Lu, K. Takata, A. Koizumi and T. Izumi (1999). "A mutation in the insulin 2 gene induces diabetes with severe pancreatic beta-cell dysfunction in the Mody mouse." J Clin Invest **103**(1): 27-37.

Wang, W. Y., B. J. Barratt, D. G. Clayton and J. A. Todd (2005). "Genome-wide association studies: theoretical and practical concerns." Nat Rev Genet **6**(2): 109-118.

Weedon, M. N., K. R. Owen, B. Shields, G. Hitman, M. Walker, M. I. McCarthy, A. T. Hattersley and T. M. Frayling (2005). "A large-scale association analysis of common variation of the HNF1alpha gene with type 2 diabetes in the U.K. Caucasian population." Diabetes **54**(8): 2487-2491.

Weedon, M. N., K. R. Owen, B. Shields, G. Hitman, M. Walker, M. I. McCarthy, L. D. Love-Gregory, M. A. Permutt, A. T. Hattersley and T. M. Frayling (2004). "Common variants of the hepatocyte nuclear factor-4alpha P2 promoter are associated with type 2 diabetes in the U.K. population." Diabetes **53**(11): 3002-3006.

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**(5): 1208-1212.

Wellcome Trust Case Control Consortium (2007). "Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls." Nature **447**(7145): 661-678.

Wellen, K. E. and G. S. Hotamisligil (2005). "Inflammation, stress, and diabetes." The Journal of clinical investigation **115**(5): 1111-1119.

Weng, J., W. M. Macfarlane, M. Lehto, H. F. Gu, L. M. Shepherd, S. A. Ivarsson, L. Wibell, T. Smith and L. C. Groop (2001). "Functional consequences of mutations in the MODY4 gene (IPF1) and coexistence with MODY3 mutations." Diabetologia **44**(2): 249-258.

Weyer, C., C. Bogardus, D. M. Mott and R. E. Pratley (1999). "The natural history of insulin secretory dysfunction and insulin resistance in the pathogenesis of type 2 diabetes mellitus." J Clin Invest **104**(6): 787-794.

Weyer, C., J. K. Wolford, R. L. Hanson, J. E. Foley, P. A. Tataranni, C. Bogardus and R. E. Pratley (2001). "Subcutaneous abdominal adipocyte size, a predictor of type 2 diabetes, is linked to chromosome 1q21--q23 and is associated with a common polymorphism in LMNA in Pima Indians." Mol Genet Metab **72**(3): 231-238.

Whincup, P. H., S. J. Kaye, C. G. Owen, R. Huxley, D. G. Cook, S. Anazawa, E. Barrett-Connor, S. K. Bhargava, B. E. Birgisdottir, S. Carlsson, S. R. de Rooij, R. F. Dyck, J. G. Eriksson, B. Falkner, C. Fall, T. Forsen, V. Grill, V. Gudnason, S. Hulman, E. Hypponen, M. Jeffreys, D. A. Lawlor, D. A. Leon, J. Minami, G. Mishra, C. Osmond, C. Power, J. W. Rich-Edwards, T. J. Roseboom, H. S. Sachdev, H. Syddall, I. Thorsdottir, M. Vanhala, M. Wadsworth and D. E. Yarbrough (2008). "Birth weight and risk of type 2 diabetes: a systematic review." JAMA : the journal of the American Medical Association **300**(24): 2886-2897.

Wigginton, J. E. and G. R. Abecasis (2005). "PEDSTATS: descriptive statistics, graphics and quality assessment for gene mapping data." Bioinformatics **21**(16): 3445-3447.

Willi, C., P. Bodenmann, W. A. Ghali, P. D. Faris and J. Cornuz (2007). "Active smoking and the risk of type 2 diabetes: a systematic review and meta-analysis." JAMA : the journal of the American Medical Association **298**(22): 2654-2664.

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, G. F. Bottazzo, R. Foxon, 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**(3): 553-569.

Winckler, W., N. P. Burtt, J. Holmkvist, C. Cervin, P. I. de Bakker, M. Sun, P. Almgren, T. Tuomi, D. Gaudet, T. J. Hudson, K. G. Ardlie, M. J. Daly, J. N. Hirschhorn, D. Altshuler and L. Groop (2005). "Association of common variation in the HNF1alpha gene region with risk of type 2 diabetes." Diabetes **54**(8): 2336-2342.

Winckler, W., R. R. Graham, P. I. de Bakker, M. Sun, P. Almgren, T. Tuomi, D. Gaudet, T. J. Hudson, K. G. Ardlie, M. J. Daly, J. N. Hirschhorn, L. Groop and D. Altshuler (2005). "Association testing of variants in the hepatocyte nuclear factor 4alpha gene with risk of type 2 diabetes in 7,883 people." Diabetes **54**(3): 886-892.

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**(3): 685-693.

World Health Organization (1999). "Definition, diagnosis and classification of diabetes mellitus, report of a WHO consultation, part 1: diagnosis and classification of diabetes mellitus." WHO, Geneva.

World Health Organization (2011). "Use of glycated haemoglobin (HbA1c) in the diagnosis of diabetes mellitus." WHO, Geneva.

Xia, B., Q. Sheng, K. Nakanishi, A. Ohashi, J. Wu, N. Christ, X. Liu, M. Jasin, F. J. Couch and D. M. Livingston (2006). "Control of BRCA2 cellular and clinical functions by a nuclear partner, PALB2." Mol Cell **22**(6): 719-729.

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 (1996). "Mutations in the hepatocyte nuclear factor-4alpha gene in maturity-onset diabetes of the young (MODY1)." Nature **384**(6608): 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. (1996). "Mutations in the hepatocyte nuclear factor-1alpha gene in maturity-onset diabetes of the young (MODY3)." Nature **384**(6608): 455-458.

Yamauchi, T., K. Hara, S. Maeda, K. Yasuda, A. Takahashi, M. Horikoshi, M. Nakamura, H. Fujita, N. Grarup, S. Cauchi, D. P. Ng, R. C. Ma, T. Tsunoda, M. Kubo, H. Watada, H. Maegawa, M. Okada-Iwabu, M. Iwabu, N. Shojima, H. D. Shin, G. Andersen, D. R. Witte, T. Jorgensen, T. Lauritzen, A. Sandbaek, T. Hansen, T. Ohshige, S. Omori, I. Saito, K. Kaku, H. Hirose, W. Y. So, D. Beury, J. C. Chan, K. S. Park, E. S. Tai, C. Ito, Y. Tanaka, A. Kashiwagi, R. Kawamori, M. Kasuga, P. Froguel, O. Pedersen, N. Kamatani, Y. Nakamura and T. Kadowaki (2010). "A genome-wide association study in the Japanese population identifies susceptibility loci for type 2 diabetes at UBE2E2 and C2CD4A-C2CD4B." Nature Genetics **42**(10): 864-868.

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." Nature genetics **40**(9): 1092-1097.

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**(5): 638-645.

Zeggini, E., M. N. Weedon, C. M. Lindgren, T. M. Frayling, K. S. Elliott, H. Lango, N. J. Timpson, J. R. Perry, N. W. Rayner, R. M. Freathy, J. C. Barrett, B. Shields, A. P. Morris, S. Ellard, C. J. Groves, L. W. Harries, J. L. Marchini, K. R. Owen, B. Knight, L. R. Cardon, M. Walker, G. A. Hitman, A. D. Morris, A. S. Doney, M. I. McCarthy and A. T. Hattersley (2007). "Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes." Science **316**(5829): 1336-1341.

Zhou, K., C. Bellenguez, C. C. Spencer, A. J. Bennett, R. L. Coleman, R. Tavendale, S. A. Hawley, L. A. Donnelly, C. Schofield, C. J. Groves, L. Burch, F. Carr, A. Strange, C. Freeman, J. M. Blackwell, E. Bramon, M. A. Brown, J. P. Casas, A. Corvin, N. Craddock, P. Deloukas, S. Dronov, A. Duncanson, S. Edkins, E. Gray, S. Hunt, J. Jankowski, C. Langford, H. S. Markus, C. G. Mathew, R. Plomin, A. Rautanen, S. J. Sawcer, N. J. Samani, R. Trembath, A. C. Viswanathan, N. W. Wood, L. W. Harries, A. T. Hattersley, A. S. Doney, H. Colhoun, A. D. Morris, C. Sutherland, D. G. Hardie, L. Peltonen, M. I. McCarthy, R. R. Holman, C. N. Palmer, P. Donnelly and E. R. Pearson (2011). "Common variants near ATM are associated with glycemic response to metformin in type 2 diabetes." Nature Genetics **43**(2): 117-120.

Zhu, Q., K. Yamagata, A. Miura, N. Shihara, Y. Horikawa, J. Takeda, J. Miyagawa and Y. Matsuzawa (2003). "T130I mutation in HNF-4alpha gene is a loss-of-function mutation in hepatocytes and is associated with late-onset Type 2 diabetes mellitus in Japanese subjects." Diabetologia **46**(4): 567-573.

Zimmet, P., K. G. Alberti and J. Shaw (2001). "Global and societal implications of the diabetes epidemic." Nature **414**(6865): 782-787.

Zouali, H., E. H. Hani, A. Philippi, N. Vionnet, J. S. Beckmann, F. Demenais and P. Froguel (1997). "A susceptibility locus for early-onset non-insulin dependent (type 2) diabetes mellitus maps to chromosome 20q, proximal to the phosphoenolpyruvate carboxykinase gene." Hum Mol Genet **6**(9): 1401-1408.