View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Queen Mary Research Online

Genes and type 2 diabetes: polymorphisms of the EIF2AK3 gene and its relationship to type 2 diabetes mellitus

Sudagani, Jaidev

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author

For additional information about this publication click this link.
https://qmro.qmul.ac.uk/jspui/handle/123456789/411

Information about this research object was correct at the time of download; we occasionally make corrections to records, please therefore check the published record when citing. For more information contact scholarlycommunications@qmul.ac.uk

GENES AND TYPE 2 DIABETES: POLYMORPHISMS OF THE *EIF2AK3* GENE AND ITS RELATIONSHIP TO TYPE 2 DIABETES MELLITUS

A dissertation submitted to the Queen Mary University of London for the degree of Doctor of Medicine (M.D.)

Dr. Jaidev Sudagani. MBBS, MRCP

Barts and The London School of Medicine and Dentistry,
Queen Mary University of London

DECLARATION OF STATEMENT OF CONJOINT WORK

The work on type 2 diabetes for the candidate gene association study (*EIF2AK3* gene) was done by me together with Dr. Rebecca Allotey. I had a major role in recruiting South Asian subjects from London for the purpose of the study. I spent considerable amount of time questioning the basis of the study hypothesis and study design under the guidance of Professor Hitman. The majority of sequencing work was done by Dr. Rebecca Allotey and only a few runs by me. Analysis of the sequencing was done jointly with Dr. Allotey. Genotyping work at Genome centre was done by me. The majority of genotyping work was out sourced to KBiosciences as this worked out more economical leaving me time to carry out the major workload of the data handling, analysis and interpretation for the candidate gene association study. I spent considerable amount of time and effort for the purpose of quality control checks which became a major issue in the study. I had significant input and support from Professor Graham Hitman in the major work that was needed after the genotyping stage of the research. I am grateful to Professor Hitman for his guidance in the completion of the M.D. thesis.

Signature:

Dr. Jaidev Sudagani

ABSTRACT

Aims/ Hypothesis: Wolcott- Rallison syndrome (WRS) is a rare autosomal recessively inherited Mendelian disorder. It is characterised by a short trunk compared to arm span, multiple epiphyseal dysplasia, multiple fractures, hepatosplenomegaly and renal insufficiency in addition to insulin dependent diabetes. The onset of diabetes in WRS families is mainly below the age of 6 months and is characterised by permanent severe non-autoimmune insulin deficiency. Mutations of the gene encoding eukaryotic translational initiation factor 2 - alpha kinase 3 (*EIF2AK3*) were found to account for diabetes in WRS. The aim of our study was to determine whether common polymorphisms in the *EIF2AK3* gene (Candidate gene association study) could be associated with type 2 diabetes.

Methods: Direct sequencing was performed on all 17 exons/coding regions and intron/exon boundaries of *EIF2AK3* gene in 48 diabetes and control subjects. Single Nucleotide Polymorphisms (SNPs) tagging the common haplotypes (tag SNPs) were identified and 11 SNPs were genotyped initially in 2,835 subjects with type 2 diabetes, 3,538 control subjects in the British Irish, Bangladeshi and South Indian Populations and 522 families (n= 1,722) in the British Irish and South Indian Populations.

Results: We identified 19 SNPs by direct sequencing. There was no association (all $p>0.05$) between the SNPs and type 2 diabetes in the case–control study and in the family study. In the one marker, rs7605713, that showed a nominal significance in Warren 2 European samples, further replication studies in the Dundee samples (3,334 diabetes cases and 3,456 controls) proved to be negative thereby avoiding a false positive result. The results also showed several of the SNPs had different minor allele frequencies between the British/Irish Caucasians as compared to the South Asians.

Conclusions/interpretation: Common variations in the *EIF2AK3* gene were not associated with type 2 diabetes in the British Irish and the South Asian population.

INDEX

Title page 1

Statement of conjoint work 2

Abstract 3

Table of Contents **4**

List of Tables 7

List of Figures 9

Abbreviations 10

TABLE OF CONTENTS

CHAPTER ONE:

INTRODUCTION TO DIABETES **12**

1.1 THE CLASSIFICATION OF DIABETES MELLITUS 12

1.2 AETIOLOGY AND EPIDEMIOLOGY OF DIABETES 15

1.2.1 TYPE 1 DIABETES 15

1.2.2 TYPE 2 DIABETES 24

1.2.3 MATURITY ONSET DIABETES OF THE YOUNG 40

THE WRS GENE- *EIF2AK3* GENE **43**

1.3 THE WOLCOTT-RALLISON SYNDROME-*EIF2AK3* GENE 43

1.4 THE *EIF2AK3* GENE REGION AND TYPE 1 DIABETES IN SUBJECTS FROM SOUTH INDIA 47

1.5 STRUCTURE OF THE *EIF2AK3* GENE AND CHROMOSOMAL LOCATION 50

PANCREATIC β CELL, APOPTOSIS AND ER STRESS **51**

1.6 ROLE OF APOPTOSIS IN PANCREATIC β CELL DEATH 51

1.7 *EIF2AK3* GENE AND ENDOPLASMIC RETICULUM STRESS/ ROLE OF APOPTOSIS IN β CELL DEATH 57

1.8 **HYPOTHESIS AND AIMS OF THE STUDY** **60**

CHAPTER TWO: SUBJECTS **64**

2.1 SUBJECT DEFINITIONS 64

2.2 INDIVIDUAL ETHNIC GROUPS

2.2.1 BRITISH IRISH RESOURCES 66

2.2.2 SOUTH ASIAN RESOURCES 69

2.2.3 DUNDEE REPLICATION RESOURCES 71

CHAPTER THREE: MATERIALS AND METHODS **72**

3.1 COLLECTION OF BLOOD SAMPLES AND DNA PREPARATION 72

3.2 DNA SEQUENCING 74

3.3 TAQMAN HIGH THROUGHPUT GENOTYPING 78

3.4 AMPLIFLUOR HIGH THROUGHPUT GENOTYPING 83

3.5 GENOTYPE ERRORS- QUALITY CONTROL (QC) 86

CHAPTER FOUR: STATISTICS **87**

4.1 HARDY- WEINBERG EQUILIBRIUM TESTS 87

4.2 UNPHASED GENOTYPING ANALYSIS 88

4.3 LINKAGE DISEQUILIBRIUM STATISTICS 91

CHAPTER FIVE: SEQUENCING RESULTS **92**

5.1 SEQUENCING RESULTS 92

CHAPTER SIX: QUALITY CONTROL **98**

6.1 QUALITY CONTROL 98

6.2 QUALITY CONTROL ISSUES FOR TAQMAN GENOTYPING 105

6.3 QUALITY CONTROL AMPLIFUOR AND KASpar GENOTYPING 110

6.4 MINOR ALLELE FREQUENCIES 111

6.5 HARDY WEINBERG EQUILIBRIUM 113

CHAPTER SEVEN: GENOTYPING RESULTS **116**
7.1 DISEASE ASSOCIATION STATISTICS-GENOTYPING RESULTS 116
7.2 LINKAGE DISEQUILIBRIUM STATISTICS 131

CHAPTER EIGHT: DISCUSSION **134**
8.1 PARADIGM SHIFT IN THE STUDY OF GENETICS OF COMMON DISEASES 134
8.2 COMMON DISEASE COMMON VARIANT HYPOTHESIS-GWAS 136
8.3 COMMON DISEASE RARE VARIANT HYPOTHESIS 139
8.4 MONOGENIC DISEASE- COMMON DISEASE 140
8.5 *EIF2AK3* GENE- POLYGENIC TYPE 2 DIABETES 142
8.6 GENOTYPING QUALITY CONTROL ISSUES 145
8.7 SEQUENCING 146
8.8 MINOR ALLELE FREQUENCIES- POPULATION CHARACTERISTICS 146
8.9 HARDY WEINBERG EQUILIBRIUM RESULTS 147
8.10 GENOTYPING RESULTS ON *EIF2AK3* GENE 148
8.11 STUDY DESIGN AS OF 2009-2010 152
8.12 THE FUTURE 156
8.13 **CONCLUSIONS** **162**

REFERENCE LIST **163**

ACKNOWLEDGEMENTS **184**

LIST OF TABLES

CHAPTER ONE: INTRODUCTION

Table 1.a Aetiologic classification of diabetes mellitus 12

Table 1.b Incidence of type 1 diabetes mellitus in Europe 16

Table 1.c Extremes of prevalence rates of type 2 diabetes mellitus in different ethnic groups 24

Table 1.d Gene discovery in type 2 diabetes 34

Table 1.e Maturity onset diabetes of the young 41

Table 1.f Potential inducers of β cell apoptosis 52

CHAPTER TWO: SUBJECTS

Table 2.a Subjects for type 2 diabetes association study 65

CHAPTER THREE: MATERIALS AND METHODS

Table 3.a Selection of samples for sequencing 77

Table 3.b PCR reaction mix for Taqman 81

Table 3.c Temperature settings for PCR 81

CHAPTER FIVE: SEQUENCING RESULTS

Table 5.a Tagging SNPs as per HAPMAP for *EIF2AK3* gene 97

Table 5.b Polymorphisms genotyped in *EIF2AK3* gene 97

CHAPTER SIX: QUALITY CONTROL

Table 6.a Quality control checks 105

Table 6.b W 2 New trios plate 1 genotyping failures on taqman and gel run 107

Table 6.c W 2 New trios plate 2 genotyping failures on taqman and gel run 108

Table 6.d Minor allele frequencies 111

Table 6.e Minor allele frequency- Population characteristics 112

Table 6.f Hardy-Weinberg equilibrium results for 3 populations 113

Table 6.g-1, 2, 3, 4, 5 and 6 HWE Failures 114

CHAPTER SEVEN: GENOTYPING RESULTS

Table 7.a	Genotyping results – p-values all markers, 3 groups	119
Table 7.b	Genotyping results – p-values all markers, British Irish Cohort	120
Table 7.c	Genotyping results – p-values all markers, British Irish case control	121
Table 7.d	Genotyping results – p-values all markers, Warren 2 trios	122
Table 7.e	Genotyping results – p-values all markers, Bangladeshi Cohort	123
Table 7.f	Genotyping results – p-values all markers, Young Bangladeshi case control	124
Table 7.g	Genotyping results – p-values all markers, Bangladeshi case control	125
Table 7.h	Genotyping results – p-values all markers, SI Cohort	126
Table 7.i	Genotyping results- p-values all markers, SI case control	127
Table 7.j	Genotyping results – p-values all markers, SI Urban survey	128
Table 7.k	Genotyping results – p-values all markers, SI families	129
Table 7.l	Genotyping results – Dundee replication study	130
Table 7.m	Pairwise linkage disequilibrium for *EIF2AK3* markers	132

CHAPTER EIGHT: DISCUSSION

Table 8.a	GWAS and gene variants of T2DM	137
Table 8.b	Power calculations for gene association study	153

LIST OF FIGURES

CHAPTER ONE: INTRODUCTION

Figure 1.a The aetiology of type 1 diabetes mellitus 17

Figure 1.b Pathway for posititional cloning and candidate gene testing 30

THE WRS GENE- *EIF2AK3* GENE

Figure 1.c Mutations within *EIF2AK3* and WRS 45

Figure 1.d Structure of the *EIF2AK3* gene 50

HYPOTHESIS AND AIMS OF THE STUDY

Figure 1.e Relative reduction in beta cell mass 51

Figure 1.f β cell replication, apoptosis and β cell volume 53

Figure 1.g A model of β-cell apoptosis in type 2 diabetes 55

CHAPTER THREE: MATERIAL AND METHODS

Figure 3.a Taqman SNP genotyping 79

Figure 3.b Amplifluor SNP Genotyping 84

CHAPTER FIVE: SEQUENCING RESULTS

Figure 5.a Sequencing strategy 92

Figure 5.b Sequencing chromatogram 94

CHAPTER SIX: QUALITY CONTROL

Figure 6.a ECACC 1 (rs1805164) Allelic discrimination plot 100

Figure 6.b ECACC 1 (rs1805165) Allelic discrimination plot 101

Figure 6.c ECACC Raw data plots 102

Figure 6.d W2 new trios plate 1 (rs867529) Allelic discrimination plot 103

Figure 6.e W2 new trios plate 2 (rs867529) Allelic discrimination plot 104

CHAPTER SEVEN: RESULTS

Figure 7.a LD plot 131

CHAPTER EIGHT: DISCUSSION

Figure 8.a Common variants- common disease/
Rare variants- common disease 140

ABBREVIATIONS

β cells	Pancreatic beta cells (β Islets of Langerhans)
BMI	Body mass index
Bp	Base Pairs
CAPN10	Calpain 10 gene
CTLA4	Cytotoxic T lymphocyte antigen 4 gene
cDNA	Complementary Deoxyribonucleic acid
χ^2	Chi Square test
DNA	Deoxyribonucleic acid
ECACC	European Collection of Cell Cultures
EIF2AK3	Eukaryotic translational initiation factor 2 - alpha kinase 3 gene
ER	Endoplasmic reticulum
FBG	Fasting blood glucose
GAD antibody	Glutamic acid decarboxylase antibody
GC	Genome centre
GCK	Glucokinase
GWAS	Genome wide association studies
HLA	Human Leucocyte antigen
HNF1α	Hepatic transcription factor 1 *α* gene
HNF1β	Hepatic transcription factor 1 *β* gene
HWE	Hardy–Weinberg equilibrium
IGT	Impaired glucose tolerance
INDEL	Insertion Deletion product
INS	Insulin gene
GCK	Glucokinase gene
HLA	Human leucocyte antigen
Kb	Kilo base
KCNJ11	Kir6.2 component of the pancreatic beta-cell KATP channel

LADA	Latent autoimmune diabetes in adults
LD	Linkage disequilibrium
LOD scores	Logarithm of odds ratio/ likelihood ratio
MAF	Minor allele frequency
MHC	Major Histocompatibility complex
MODY	Maturity onset diabetes of the young
MRC 1958 birth cohort	Medical research Council 1958 birth cohort
mRNA	Messenger RNA
NOD mice	Non obese diabetic mice
OGTT	Oral glucose tolerance test
OR	Odds ratio
PCR	Polymerase chain reaction
PEK /PERK	pancreatic eIF2-alpha kinase
PPARγ	Peroxisomal proliferative activated receptor gamma
PTPN11	Protein tyrosine phosphatase, non-receptor type 11
PTPN22	Protein Tyrosine Phosphatase N22 gene
QC	Quality control
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
SI	South Indian DNA collection
Sib-pair	Sibling pair
SNP	Single nucleotide polymorphism
T1DM	Type 1 diabetes
T2DM	Type 2 diabetes
TCF7L2	Transcription factor 7-like 2
tRNA	Transfer RNA
VDR	Vitamin D receptor
W2	Warren 2 DNA collection
WHO	World Health Organisation
WRS	Wolcott-Rallison Syndrome
WTCC	Wellcome Trust Case-control Consortium

CHAPTER ONE: INTRODUCTION TO DIABETES

1.1 THE CLASSIFICATION OF DIABETES MELLITUS

Diabetes mellitus is aetiologically and clinically heterogeneous group of disorders that share hyperglycaemia in common. The National Diabetes Data Group (NDDG) and World Health Organisation (WHO) Expert committee on Diabetes have come up with the classification for diabetes in 1979 **(1)** and in 1997, an international committee of Diabetologists recommended several changes in the classification that have been endorsed by the American Diabetes Association and WHO. The terms insulin dependent diabetes (IDDM) and non-insulin dependent diabetes (NIDDM) were eliminated as they were based upon pharmacologic rather than aetiological considerations. The two major forms of diabetes are type 1 diabetes and type 2 diabetes. Type 1 diabetes, an autoimmune disease, is due to pancreatic beta cell destruction resulting in insulin deficiency. Type 2 diabetes is a heterogeneous condition characterised by insulin resistance and also associated with impairment in compensatory insulin secretion. The current expert committee classification **(Table 1.a)** has proposed the above changes.

Table 1.a: Aetiologic classification of diabetes mellitus[1]

I. Type 1 diabetes (β-cell destruction, usually leading to absolute insulin deficiency)
   A. Immune mediated
   B. Idiopathic

II. Type 2 diabetes (may range from predominantly insulin resistance with relative insulin deficiency to a predominantly secretory defect with insulin resistance)

III. Other specific types
   A. Genetic defects of β-cell function
      1. Chromosome 12, HNF-1α (MODY3)
      2. Chromosome 7, glucokinase (MODY2)
      3. Chromosome 20, HNF-4α (MODY1)
      4. Chromosome 13, insulin promoter factor-1 (IPF-1; MODY4)
      5. Chromosome 17, HNF-1β (MODY5)
      6. Chromosome 2, *NeuroD1* (MODY6)
      7. Mitochondrial DNA
      8. Others

B. Genetic defects in insulin action
1. Type A insulin resistance
2. Leprechaunism
3. Rabson-Mendenhall syndrome
4. Lipoatrophic diabetes
5. Others

C. Diseases of the exocrine pancreas
1. Pancreatitis
2. Trauma/pancreatectomy
3. Neoplasia
4. Cystic fibrosis
5. Hemochromatosis
6. Fibrocalculous pancreatopathy
7. Others

D. Endocrinopathies
1. Acromegaly
2. Cushing's syndrome
3. Glucagonoma
4. Pheochromocytoma
5. Hyperthyroidism
6. Somatostatinoma
7. Aldosteronoma
8. Others

E. Drug- or chemical-induced
1. Vacor
2. Pentamidine
3. Nicotinic acid
4. Glucocorticoids
5. Thyroid hormone
6. Diazoxide
7. β-adrenergic agonists
8. Thiazides
9. Dilantin
10. α-Interferon
11. Others

F. Infections
   1. Congenital rubella
   2. Cytomegalovirus
   3. Others

G. Uncommon forms of immune-mediated diabetes
   1. "Stiff-man" syndrome
   2. Anti–insulin receptor antibodies
   3. Others

H. Other genetic syndromes sometimes associated with diabetes
   1. Down's syndrome
   2. Klinefelter's syndrome
   3. Turner's syndrome
   4. Wolfram's syndrome
   5. Friedreich's ataxia
   6. Huntington's chorea
   7. Laurence-Moon-Biedl syndrome
   8. Myotonic dystrophy
   9. Porphyria
   10. Prader-Willi syndrome
   11. Others

IV. Gestational diabetes mellitus (GDM)

---

[1]Modified from American Diabetes Association: Diabetes Care 2008 31:S55-S60, 2008

1.2 **Aetiology and Epidemiology of Diabetes**

Introduction

Diabetes mellitus is a chronic metabolic disorder characterised by disturbance in glucose metabolism leading to a state of hyperglycaemia and is associated with microvascular and macrovascular complications in the long term. Diabetes is the leading cause of chronic diseases worldwide and has reached epidemic proportions in certain parts of the world and in certain ethnic groups. This has widespread implications for health resources. The two main types of diabetes, type 1 diabetes mellitus (T1DM) and type 2 diabetes mellitus (T2DM) are quite distinct from each other in their aetiology and epidemiology. T2DM is the most common form of diabetes world-wide accounting for 90% of cases globally (2) and affecting approximately 4% of the worlds adult population. Type 1 diabetes is an autoimmune disease that results in insulin deficiency.

1.2.1 Type 1 Diabetes

Worldwide prevalence

In 1997 there were 11.5 million people with T1DM in the world; the figure has been estimated to rise to 23.7 million in the year 2010. These increasing figures will have most impact in Asia, where there are 4.5 million people with T1DM, estimated to rise to 12 million by the year 2010. One of the best incidence studies has come from Europe as part of a European collaboration. The highest incidence of T1DM is found in Finland; the lowest rates in Europe are in Romania **(Table 1.b)**. The incidence of T1DM follows a north-south gradient, with the notable exception of Sardinia. The figures from countries such as India are less precise, although one study in Chennai (3) suggested an incidence equivalent to that found in Southern European countries. These different rates of T1DM are likely to reflect both the genetic background of individual countries and differences in

exposure to environmental agents. In recent years in several different countries the incidence of T1DM is increasing. These changes must reflect environmental influences.

Table 1.b: Incidence of type 1 diabetes mellitus in Europe

	Rate (a)
Finland	42.9
Sardinia	30.2
Denmark	21.5
Norway	20.8
UK	16.4
Luxembourg	12.4
Netherlands	11.0
Sicily	10.1
Belgium	9.8
France	7.8
Italy	6.8
Israel	5.5
Poland	5.5
Romania	5.1

a Standardized incidence rates (per 100 000 per year), age 0-14 years, in a selection of countries in Europe taken from EURODIAB study

Aetiology of Type 1 Diabetes

Type 1 diabetes is due to autoimmune destruction of insulin secreting pancreatic β cells. T1DM typically occurs in young individuals with an age of onset in childhood or early adult life. The autoimmune reaction is likely to be triggered by an environmental agent in utero or in very early life **(Figure 1.a)**. The earliest markers of β cell destruction are the appearance of autoantibodies to glutamic acid decarboxylase (GAD), islet cells and

insulin. Autoantibodies have been detected 10-15 years before the onset of disease and, furthermore, have been known to disappear without T1DM occurring in a few individuals. One to two years before onset of the disease, evidence of β cell impairment can be detected, initially evidenced by a reduction in the first phase of insulin response to intravenous glucose and in the later stages by an abnormal oral glucose tolerance. In contrast to the slow β cell destruction, the onset of T1DM is acute and is usually measured in weeks. At this stage in the aetiological process, it is likely that 70% of β cells have been destroyed and those remaining are inhibited by the action of cytokines.

Figure 1.a: The aetiology of type 1 diabetes mellitus

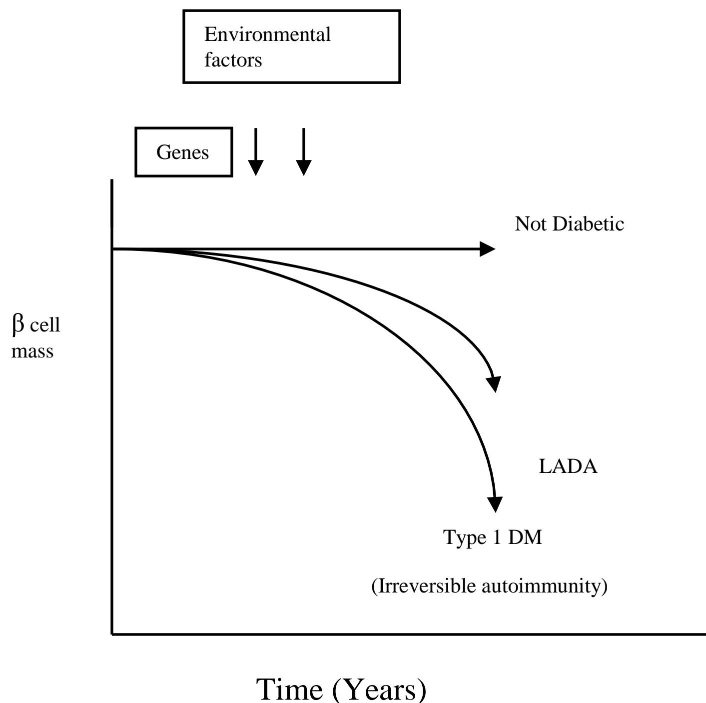


Figure 1.a: Gene- Environment interaction and decline in β cell mass in auto-immune diabetes

There is a subgroup of patients who develop diabetes in adult life (> 25 years of age) and they do not require insulin during the first few years after diagnosis but progress rapidly to insulin dependence, and have an autoimmune component to their disease with positive GAD and islet cell antibodies (4). This condition is named Latent autoimmune diabetes in adults (LADA). There are several common features between T1DM and LADA,

including T cell mediated insulitis, islet and GAD antibody positivity and high rates of HLA DR3 and DR4 genes. The prevalence of LADA in newly diagnosed diabetics has been shown to range from 2.8% to 22.3% in different studies depending on the markers used and characteristics of the patients. Although these patients present as type 2 diabetes, they have been shown to progress quickly to insulin dependency especially if the diabetes is diagnosed at a younger age and the patient is not overweight (5). Therefore measuring GAD antibody and C- peptide levels, in newly diagnosed young, non-obese “type 2 diabetes patients” and no family history of T2DM, to identify the LADA sub group should be considered in routine clinical practice.

Genetics of type 1 Diabetes

Type 1 diabetes is a multi-factorial disease with both genetic and environmental components. The largest genetic contribution to T1DM is determined by genes in the major histocompatibility complex (MHC) located to the short arm of chromosome 6 (IDDM1-HLA, 6p21). Initial associations between T1DM and the MHC were described for the HLA class I antigens A1-B8 and B15. With advent of HLA class II serology, closer associations were found with HLA –DR with an increased frequency of DR3 and DR4 alleles and a decreased frequency of DR2 allele in T1DM subjects. At the population level the strongest genetic association with T1DM is with HLA-DQ alleles (6). This is best defined by DNA typing of HLA- DQ1, DQB1 and DRB1 (7). However, due to the strong linkage disequilibrium between these loci it has been very difficult to study the effect of individual HLA DQ or HLA DR Genes separately. For the individual, susceptibility is best defined by allelic combinations of MHC genes located to all three major regions (classes I, II, and III), called HLA haplotypes. Haplotypes occur because of strong linkage disequilibrium observed in the MHC whereby the combinations of alleles are seen more frequently than would be expected by their individual gene frequencies. An example of a haplotype would be A2, Cw1, B56, TNFa6, DRB1*401, DQA1*0301, DQB1*0302. In a recent study there has been a re-focus on the MHC class I genes HLA-B and HLA-A- principally involving HLA-B*39, contributing to the aetiology of type 1 diabetes.(8)

The MHC accounts for approximately 40% of the genetic component to T1DM. Evidence from genome scans and candidate gene studies indicates the existence of a large number of putative non-MHC genes contributing to the aetiology of T1DM, although all of comparatively small effect compared to the MHC (9). The most reproducible T1DM associations have been found with the insulin gene and Cytotoxic T lymphocyte antigen 4 gene (*CTLA4*). An association between the insulin gene [located on chromosome 11p15.5; *INS*), and T1DM was described in the 1980s and subsequently confirmed by linkage studies. The *INS* locus on chromosome 11p15.5 contains a major polymorphism 5' to the transcription site, which is a variable number of tandem repeats (VNTR) region (10). One functional hypothesis to explain the association between the insulin gene and T1DM is that 'hypersecretors' of insulin determined by the disease- associated polymorphism might induce thymic intolerance to insulin, thus increasing a risk of autoimmune reaction. Studies have showed association of the T- cell regulatory gene *CTLA4* with susceptibility to autoimmune disease, including type 1 diabetes (11). *CTLA4*, gene located on chromosome 2q33, plays an important role in the counter regulation of CD28 T cell antigen receptor activation of T cells. In the mouse model of T1DM, susceptibility was also associated with variation in *CTLA4* gene splicing with reduced production of a splice form encoding a molecule lacking the CD80/CD86 ligand-binding domain.

A single nucleotide polymorphism in the gene Protein Tyrosine Phosphatase N22 *(PTPN22)* encoding the lymphoid protein tyrosine phosphatase (LYP), a suppressor of T-cell activation, has been shown to be associated with T1DM (12). The variants encoded by the two alleles, 1858C (Arginine) and 1858T (Tryptophan), differ in a crucial amino acid residue involved in association of LYP with the negative regulatory kinase Csk. Unlike the variant encoded by the more common allele 1858C, the variant associated with T1DM does not bind Csk. The *PTPN22* allele 1858T was more frequent in individuals with T1DM than in healthy individuals: 30.6% of individuals with T1DM were heterozygous with respect to 1858T, compared with only 21.3% of the healthy controls and 3.7% of individuals with T1DM were homozygous with respect to 1858T,

compared with only 1.0% of the healthy controls. These results suggest that the *PTPN22* allele 1858T predisposes an individual to developing T1DM.

A joint Genome wide association study (GWAS) in the British population has examined 2000 individuals for each of 7 major diseases and a shared set of 3000 controls, identified association signals at 7 regions in type 1 diabetes (13). There are six genes/regions for which there is strong pre-existing statistical support for a role in T1DM-susceptibility: these are the major histocompatibility complex (MHC), the genes encoding insulin, *CTLA-4* (cytotoxic T-lymphocyte associated 4) and *PTPN22* (protein tyrosine phosphatase, non-receptor type 22), and the regions around the interleukin 2 receptor alpha (*IL2RA/CD25*) and interferon-induced helicase 1 genes (*IFIH1/MDA5*). However, these signals can explain only part of the familial aggregation of T1DM. Five of these previously identified associations were detected in this scan. In this study, single-point analyses revealed three novel regions (on chromosomes 12q13, 12q24 and 16p13) showing strong evidence of association. Four further regions attained similar levels of significance either through multi-locus analyses (chromosomes 4q27 and 12p13), or through the combined analysis of autoimmune cases (chromosomes 18p11 and the 10p15 CD25 region). The associations with T1DM for chromosomes 12q13, 12q24, 16p13 and 18p11 have been confirmed in independent and multiple populations. The two signals on chromosome 12 (at 12q13 and 12q24) map to regions of extensive linkage disequilibrium covering more than ten genes. Several of these represent functional candidates because of their presumed roles in immune signalling, considered to be a major feature of T1DM susceptibility. These include *ERBB3* (receptor tyrosine-protein kinase erbB-3 precursor) at 12q13 and *SH2B3/LNK* (SH2B adaptor protein 3), *TRAFD1* (TRAF-type zinc finger domain containing 1) and *PTPN11* (protein tyrosine phosphatase, non-receptor type 11) at 12q24. For these signal regions in particular, extensive resequencing, further genotyping and targeted functional studies will be essential steps in identifying which gene, or genes, are causal. Of those listed, *PTPN11* is a particularly attractive candidate given a major role in insulin and immune signalling. It is also a member of the same family of regulatory phosphatases as *PTPN22*, already established as an important

susceptibility gene for T1DM and other autoimmune diseases. A further GWA study confirmed unequivocally the associations at 12q24, 12q13, 16p13 and 18p11 (14).
Results from other candidate gene studies and GWAS have identified other chromosomal regions, which may be involved in disease pathogenesis. The finding that so many genes are involved in T1DM, raises the possibility that there are several disease processes that might lead to β cell destruction. The identification of the variants, genes and pathways involved in disease through these approaches offers 1) a potential route to new therapies, 2) improved diagnosis and 3) better disease prevention.

Environmental factors- Type 1 Diabetes

Environmental factors play a significant part in the aetiology of T1DM and have been implicated in both initiation and progression of β cell damage. The majority of evidence points to the effects of viruses and/or dietary factors as aetiological agents.

Many viruses have been implicated in the pathogenesis of T1DM (15); they may have a direct effect on β cells by infection and cell lysis or alternatively they may act as triggers to the autoimmune process. Amongst the viruses that have been implicated in humans are coxsackie A, coxsackie B, rubella, cytomegalovirus, mumps and Epstein- Barr viruses. The Enteroviruses (coxsackie A, coxsackie B and echovirus) are the most commonly associated viruses with diabetes and serve as a major trigger for T1DM in the young possibly by induction of islet cell antibodies. The evidence for viral involvement in type 1 diabetes came from several sources, including anecdotal case reports, epidemiological studies, seasonal incidence studies and animal models. There is data to support that enterovirus infection either accompanies or precedes the development of T1DM in young people in many instances (16).

Coxsackie B was first implicated in the early 1970s by Gamble, who found an increased titre of coxsackie B antibodies in newly diagnosed T1DM patients. Coxsackie virus has been identified in very young onset type 1 diabetes (in patients under 5 years of age) using the polymerase chain reaction (17). Furthermore, when the coxsackie virus was

sequenced, although it had extensive homology to coxsackie virus B4, there was some unique sequence variation indicating an T1DM variant (18). There have also been many anecdotal reports of coxsackie B virus causing T1DM, presumably by a direct cytolytic effect on β cells. A previously fit child died in diabetic ketoacidosis three days after flu-like illness. At, necropsy, there was an extensive lymphocytic infiltration into the β cells of pancreas and coxsackie B4 was found in the child's serum. This virus was extracted from the pancreas and, when used to infect mice, led to diabetes.

There is a high incidence of T1DM among patients with the congenital rubella syndrome (19). Clearly, this results from an in utero infection, but the diabetes that ensues is indistinguishable from type 1 diabetes; the disease presents in the second decade of life, the onset is preceded by islet cell antibodies and the genetic predisposition is defined by the same HLA association as T1DM. This is likely to be a good example of virus triggering the immune process.

Dietary factors have also been implicated in the development of T1DM. Amongst the dietary factors indirectly linked to either susceptibility or protection to T1DM are cow's milk protein (including bovine serum albumin and β - lactoglobulin), β cell toxic drugs (alloxan, streptozotocin, rodenticides), dietary toxins (in particular nitroso-containing compounds) and other such as coffee and sugar. There is an interesting interplay between vitamin D, vitamin D receptor (VDR) and association with T1DM. The contribution of vitamin D as a potent modulator of the immune system is well recognised. The main sources of vitamin D are ergocalciferol and cholecalciferol found in dietary sources and cholecalciferol produced in the skin by ultraviolet radiation of 7 dehydrocholesterol. Vitamin D deficiency in infancy and VDR polymorphisms may be risk factors for T1DM. In non-obese diabetic (NOD) mice, long term treatment with high doses of vitamin D3 reduced the incidence of diabetes by changing the cytokine balance at the local pancreatic lesion (20).

The Future

The working out of the complete genetic basis of type 1 diabetes will lead to a better understanding of disease pathogenesis and will be further aided by studies of genetic and environmental interaction. T1DM may be the first multi-factorial disease to benefit from primary prevention of both insulitis and of T1DM (once autoantibodies have been detected). Diapep277 immune modulator- Immune modulation has been attempted to attenuate the process of autoimmune destruction of insulin-producing pancreatic beta cells. The 60 kDa heat-shock protein (hsp60) is one of the many different self antigens involved in this process. Treatment with DiaPep277, a peptide derived from human HSP60, has been shown to save residual beta-cell function in NOD mice (21). In humans, C-peptide concentrations were maintained in the DiaPep277 treated group as opposed to a fall in the placebo group, insulin requirements were lower in the treatment group compared to placebo (22). T-cell reactivity to hsp60 and p277 in the DiaPep277 group showed an enhanced T-helper-2 cytokine phenotype.

Anti-CD3 monoclonal antibodies can affect immune responses and possess the unique capacity to induce immunological tolerance. Short-term treatment with CD3 antibody has been shown to preserve residual beta-cell function for at least 18 months in patients with recent-onset type 1 diabetes (23). In the Diabetes Prevention Trial, low dose insulin was administered to persons with high risk of T1DM as ascertained by family history, islet antibodies and HLA typing. It concluded that low dose insulin does not delay or prevent the onset of T1DM. In a European study (24) (ENDIT) high dose nicotinamide was being used to protect the β cell in high risk individuals for type 1 diabetes; unfortunately this trial also failed to show benefit on active treatment (25). However, 90% of type 1 diabetes patients do not have a family history of type 1 diabetes. The approach in this latter group might be to identify the genetically susceptible by the use of genetic markers and then test for autoantibodies. If the latter subjects are autoantibody positive then intervention may be considered in the future. The sensitivity and specificity that would be required of such testing would depend on how safe and effective the proposed intervention would be.

1.2.2 Type 2 diabetes

Worldwide prevalence

Type 2 diabetes is one of the most common non-communicable diseases in the world with an estimated 147.2 million people suffering from this disorder; by 2010 this figure is estimated to reach 212.9 million people (26). Furthermore, it has been predicted that by the year 2010 over half the people with T2DM will be living in Asia. This trend is likely to be due to increasing urbanization and industrialization (27). According to WHO estimates the figure is likely to double by the year 2025. The prevalence of diabetes in a selection of studies is given in **Table 1.c**.

Table 1.c: Extremes of prevalence rates of type 2 diabetes mellitus in different ethnic groups

Higher	Rate (a)	Lower	Rate (a)
Pima Indians	50.3	Hispanic Central Mexico	5.6
Nauruans	41.3	Micronesian	4.3
Australian Aborigines	22.5	European (Poland)	3.5
Fijian Indians	22.0	Rural Asian Indians	2.7
US Mexican (Hispanic)	14.1	Rural Melanesian (Fiji)	1.9
Mauritian Chinese	13.1	Rural Chinese	1.6
US African	10.3	Rural African (Tanzania)	1.2
Southern Italy	10.2		

a Age –specific prevalence rates in percentages estimated for 1994 (28)

As with T1DM, the incidence of diabetes in different countries is likely to reflect the different genetic architecture as well as the differing environment. A good example is afforded by the population of Nauru; in full- blooded Nauruans over the age of 60 years the prevalence of T2DM is 83%, whereas in those with genetic admixture as adduced by the HLA typing the prevalence of diabetes is 17%; this clearly reflects the genetic component. However, rapid increase of T2DM in the world in the last few decades can only be ascribed to environmental factors. This illustrates the multi-factorial nature of T2DM, with strong genetic and environmental contributions.

Aetiology of Type 2 Diabetes

Type 2 diabetes is a multi-factorial disease with genetic and environmental factors playing a key role in its pathogenesis. Central to the aetiology is a defect in insulin action, increased hepatic glucose output and impaired insulin secretion. Although insulin resistance is frequently the first detectable abnormality in the progression of T2DM, insulin resistance by itself does not cause the disease, and is only manifested when there is a coexisting insulin secretory defect. T2DM typically occurs in middle aged and elderly people but there is increasing trend of T2DM occurring in young individuals. The main question yet to be answered is whether T2DM is one disorder, or a group of disorders with hyperglycaemia as the end point in disease pathogenesis. Insulin resistance is common to several other disorders, including ischaemic heart disease, hypertension, dyslipidaemia, central obesity and coagulation defects; the clustering of these disorders is known as metabolic syndrome or the insulin resistance syndrome or Syndrome X. The interface of T2DM with obesity is complex one, highlighted by the discovery of leptin and adiponectin. The cause of obesity and T2DM in the ob mouse is a mutation of the ob gene. With administration of the ob gene protein (leptin) the ob mouse decreases its food consumption and increases exercise, leading to a dramatic weight loss; if given earlier enough it will also prevent diabetes. Common human obesity, by contrast is associated with increased leptin levels, and leptin levels have been found to correlate with hyperinsulinaemia suggesting a degree of leptin resistance. The protein adiponectin signals adipose tissue mass and reduced levels are found in obese subjects and there is an

important interplay between adiponectin, insulin resistance, T2DM and atherosclerosis (29). Ghrelin is a gut hormone that is a signal of satiety and therefore has a direct effect on obesity (30). In the obese diabetic subject, therefore, there may be interplay between leptin, adiponectin, ghrelin and insulin, contributing to insulin resistance and the metabolic syndrome.

The Genetic contribution to Type 2 Diabetes mellitus

Type 2 Diabetes is a classic example of multi-factorial disease with the phenotype of an individual being influenced by an interaction of genes and environment. Some of the most common chronic diseases in the modern world, including heart disease, cancer, diabetes, arthritis and osteoporosis, are a result of environmental/lifestyle factors, including diet and nutrition, activity levels and smoking, superimposed on genetic predisposition. Such environmental factors have an inter-play with genetic predisposition, from foetal development to childhood and adult life, resulting in disease manifestation. The global increase in T2DM over the last century is a salient example of how interaction of lifestyle changes with genotype can dramatically impact on the health of entire populations.

Evidence for a genetic basis for T2DM and the metabolic syndrome has been derived from studies of families, twins and populations with genetic admixture. In some genetic diseases, the basis is mutation of a single gene and this is a feature of monogenic diabetes, which will be discussed later. More commonly however the genetic basis of T2DM and other chronic diseases is polygenic, i.e. many genes are involved, each with relatively modest effect on the phenotype. Genetically complex diseases such as type 2 diabetes may exhibit the following properties: incomplete penetrance - not all susceptible individuals with at risk genes are affected, involvement of several disease-predisposing loci, some of which may have a major effect, but many of which may have a relatively minor effect i.e. polygenic, interaction between genes and the environmental factors,

heterogeneity so that different loci and/or alleles cause disease in different ethnic groups(31).

Family aggregation studies have shown that siblings of family members with T2DM are four times more likely to develop T2DM in their lifetime than the general population (λsib ~4) (32). Evaluation of familial aggregation of metabolic syndrome components (Framingham Heart Study data) suggest some components have genetic determinants: lipid levels and BMI were found to correlate between siblings and between parents and offspring (33). Heritability studies have also provided support for the role of genetic factors in susceptibility to T2DM. For example, high heritability estimates for fasting blood glucose, β-cell function and components of the metabolic syndrome has been found in non-diabetic relatives from families with T2DM of Northern European extraction (34), while in a comparable study of families from the same background, but healthy families not characterized by a high degree of insulin resistance, the heritability estimate for fasting glucose was significantly lower (0.77 vs 0.21, $P < 0.001$) (35).

Further evidence for the role of genetic factors in predisposition to T2DM comes from studies of populations with genetic admixture and twins. There is a higher prevalence of T2DM in full-blooded Nauruans aged 60 years or older than in those with ancestral foreign admixture (83 vs. 17%) (36). Twin studies have shown a much higher concordance rate for T2DM among monozygotic (28–34%) versus dizygotic (14–16%) twins. In addition, long-term follow-up of monozygotic twins has shown that if one has T2DM, the concordance rate for type 2 diabetes/ IGT in monozygotic twin is very high, even in twins initially ascertained discordant for diabetes. (37).

Study of genes involved in complex disease like Type 2 Diabetes

There has been a significant change in approaching the study of genes that confer susceptibility to a complex disease like type 2 diabetes and other such chronic diseases over the past few years.

There is the

1) Traditional approach by **Linkage analyses** (Positional cloning)
2) **Candidate gene approach**
3) Recently popular methodology of **Genome wide association studies** (GWAS).

Linkage analysis (positional cloning) involves investigating the entire genome to search for genetic causes of disease (38). This kind of random gene search assumes no knowledge of the underlying defects. Instead, positional cloning aims at localizing the disease gene on the basis of its position in the genome. The tools utilized for linkage are microsatellite markers. Linkage analysis studies are performed in family pedigrees, affected sib- pairs. Once a chromosomal region has been linked to the disease, the next step would be the search for attractive candidate genes in the region or narrowing the region by linkage disequilibrium mapping. Linkage analysis tests for the segregation of a marker and disease phenotype in a pedigree and has been particularly useful in detecting genes involved in monogenic diabetes (Hepatocyte nuclear factor-1 alpha -*HNF1α* mutations in MODY).

Many groups worldwide have completed the genome scans for genes predisposing to T2DM, although the results are generally disappointing. In the UK Warren 2 sib pair repository, an autosomal genome scan was carried out on 743 UK sib pairs with T2DM (39). This was followed by nonparametric linkage analysis of the entire dataset, which identified seven regions showing linkage located on chromosomes 1q24.2, 5q13, 5q32, 7p15.3, 8p21–22, 8q24.2 and 10q23.3 (LOD scores ≥1.18). Five of the regions were found to coincide with loci previously identified in other genome scans of French, Finnish, American (European descent), Mexican American and Native American (Pima Indians) pedigrees. It was hoped this data should aid positional cloning of genes

associated with susceptibility to T2DM. Only two genome scans have lead to gene discovery with the identification of the *calpain10* gene (located to chromosome 2q37) as a major T2DM susceptibility gene in Mexican-Americans (40) and *TCF7L2* (encoding transcription factor 7-like 2) in Icelandic population. However, it is pertinent to note that *calpain10* gene was not picked out in recent Genome wide association studies whereas, *TCF7L2* was identified in GWAS and has been recognized as the most important T2DM-candidate gene to date.

The **candidate gene** approach aims at the identification of genes based upon information of their biological function(41). Association studies are based on candidate gene approach, frequently using single nucleotide polymorphisms (SNPs) and tag SNPs. Tagging SNPs exploit the extensive linkage disequilibrium that is present in the genome. A candidate gene association study tests for significant differences in the genotype/ allele frequencies of the gene between cases and controls. When choosing a candidate gene for association studies in polygenic disease like T2DM **(see Figure 1.b)** the following information helps to arrive at a candidacy for the gene (a priori evidence) - 1) Biological significance 2) Animal models 3) Human models (monogenic syndromes) 4) Linkage evidence (positional cloning). Therefore for the study of T2DM, choosing a gene involved in insulin secretion and pancreatic beta cell function or a gene involved in insulin action on the muscle, adipose tissue and liver would be an appropriate candidate gene. An animal model with the gene defect expressing the T2DM phenotype would give further credence to the candidate gene. Extrapolation by studying genes contributing to monogenic diabetes in T2DM has proved fruitful in some candidate gene association studies. Positional cloning with positive LOD scores of >+3 indicates that two gene loci are close to each other on a chromosome and provide evidence for linkage. The more of the above criteria that a gene satisfies the higher is the chance of finding a successful candidate gene. The Hapmap project has catalogued majority of the common variants (single nucleotide polymorphisms-SNPs and tag SNPs) in the human genome(42). This publicly available database has been a huge resource for candidate gene testing.

Figure 1.b: Pathway for posititional cloning and candidate gene testing

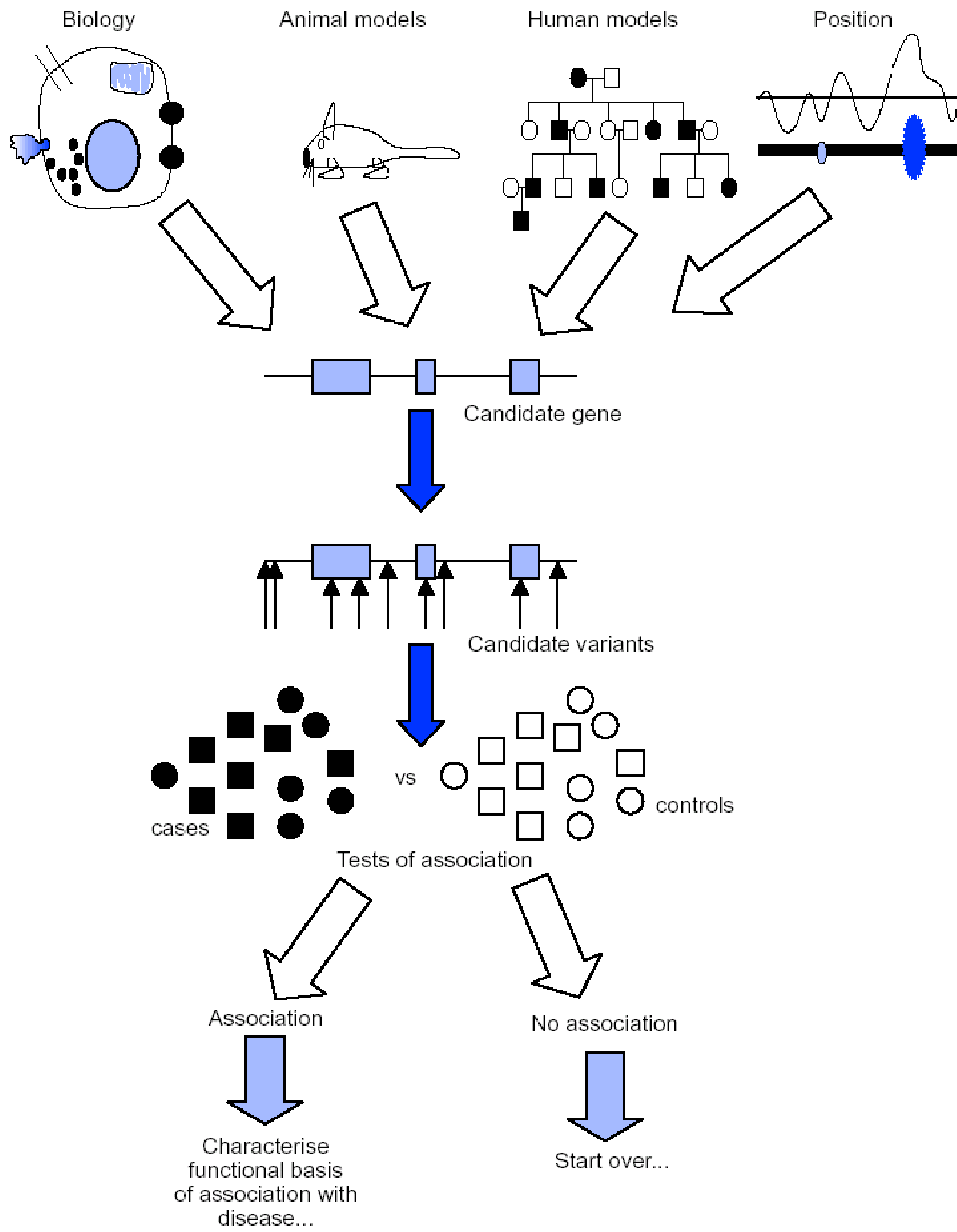


Figure 1.b: The selection of a candidate gene based on putative biological action of the gene product, animal model and human model with genetic defect leading to disease manifestation, linkage evidence of the gene for the condition of interest, in our case T2DM. Tests are carried out between cases and controls looking into association of the candidate gene variants with the disease. A positive candidate gene association study needs further replication in independent datasets to establish its candidacy. Further studies into the functional basis of the gene with relevance to its effects on T2DM would shed light on possible mechanistic pathways(**43).**

Genome wide association studies (GWAS)

Genome wide association studies are hugely popular at present (44). Three recent advances mean that Genome Wide Association studies that are powered to detect plausible effect sizes are now possible. 1) The International Hapmap resource, which documents patterns of genome-wide variation and linkage disequilibrium in four population samples, greatly facilitating both the design and analysis of association studies. 2) The availability of dense genotyping chips, (up to a million of single nucleotide polymorphisms (SNPs) that provide good coverage of much of the human genome, means that for the first time GWA studies for thousands of cases and controls are technically and financially feasible. 3) Appropriately large and well characterized clinical samples have been assembled for many common diseases like diabetes.

A GWA study is defined by the National Institutes of Health as a study of common genetic variation across the entire human genome designed to identify genetic associations with observable traits. The typical GWA study has 4 parts: (1) selection of a large number of individuals with the disease or trait of interest and a suitable comparison group; (2) DNA isolation, genotyping, and data review to ensure high genotyping quality; (3) statistical tests for associations between the SNPs passing quality thresholds and the disease/ trait; and (4) replication of identified associations in an independent population sample or examination of functional implications experimentally. The most frequently used GWA study design has been the case-control design, in which allele frequencies in

patients with the disease of interest are compared to those in a disease-free comparison group.

The approach in a GWA study involves rapidly scanning markers across the complete sets of DNA, or genomes, of many people to find genetic variations associated with a particular disease. Once new genetic associations are identified, researchers use the information to conduct further replication studies for robust associations in various datasets, and to develop better strategies to detect, treat and prevent the disease. Such studies are particularly useful in finding genetic variations that contribute to common, complex diseases, such as type 2 diabetes. The tools include computerized databases that contain the reference human genome sequence, a map of human genetic variation and a set of new technologies (Affymetrix Gene Chip, illumina, perlegen) that can quickly and accurately analyze whole-genome samples for genetic variations that contribute to the onset of a disease. DNA is placed on tiny chips and scanned on automated laboratory machines. The machines quickly survey each individual's genome for strategically selected markers of genetic variation, single nucleotide polymorphisms, or SNPs.

If certain genetic variations are found to be significantly more frequent in people with the disease compared to people without disease, the variations are said to be "associated" with the disease. The associated genetic variations can serve as powerful pointers to the region of the human genome where the disease-causing problem resides. However, the associated variants themselves may not directly cause the disease. They may just be in linkage disequilibrium with the actual causal variants. The large number of tests in a Genome wide association studies, up to 400,000, mean that p-values of ~5 X 10^{-7} are necessary to provide a study-wide p-value of 0.05.

Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls by The Wellcome Trust Case-control Consortium describe a joint GWA study (using the Affymetrix Gene Chip 500K Mapping Array Set) undertaken in the British population (45).

The most widely replicated associations for T2DM in various populations are limited to variants in three genes: *PPARG* (encoding the peroxisomal proliferative activated receptor gamma; (P12A), *KCNJ11* (the inwardly rectifying Kir6.2 component of the pancreatic beta-cell KATP channel; E23K and *TCF7L2* (transcription factor 7-like 2; rs7903146). All three of these signals were detected with effect-sizes in this GWA study consistent with previous reports. A cluster of SNPs on chromosome 10q, within *TCF7L2,* represented by rs4506565 (trend test, OR 1.36, $P = 5.7 \times 10^{-13}$) generated the strongest association signal for T2DM. rs4506565 is in tight linkage disequilibrium (r^2 of 0.92 in the CEU component of HapMap) with rs7903146, the variant with the strongest aetiological claims. The signals associated with variants within the other established T2DM-susceptibility genes, *KCNJ11* (rs5215, r^2 of 0.9 with rs5219, E23K) and *PPARG* (rs17036328, r^2 of 1 with rs1801282, P12A) are less dramatic (trend test, OR 1.15 and 1.23 respectively, both $P = \sim 0.001$). Apart from *TCF7L2,* the scan revealed two signals for T2DM with p-values less than 5×10^{-7}. The first of these maps within the *FTO* (fat-mass and obesity-associated) gene on chromosome 16q. Several adjacent SNPs (including rs9939609, rs7193144 and rs8050136) generate signals characterized by a per-allele OR for T2DM of ~1.25 and a risk-allele frequency of ~ 40% in controls.

Genes and Type 2 Diabetes

Genome scanning, candidate gene studies and GWA studies have facilitated the identification of a number of loci for T2DM and the metabolic syndrome, including *TCF7L2, KCNJ11, CAPN10*, and *PPARγ* and many other genes. **(Table 1.d)**

Table 1.d: Gene discovery in type 2 diabetes

Linkage Studies	Candidate wide association studies	Genome wide association studies (Incomplete list)
TCF7L2	*TCF7L2*	*TCF7L2*
Capn10	*KCNJ11 (E23K variant of Kir 6.2)*	*HHEX–IDE*
	PPARG (P12A)	*SLC30A8*
	TCF2	*CDKAL1*
	WFS1	*FTO*
		PPARG (P12A)
		IGF2BP2
		KCNJ11
		CDKN2A–2B

TCF7L2

TCF7L2 was identified by Decode Genetics group. This group was following up a type 2 diabetes linkage peak on chromosome 10q in Icelandic individuals, when they found very strong association of type 2 diabetes with variants of the *TCF7L2* gene. There was reported suggestive linkage of type 2 diabetes mellitus to chromosome 10q in earlier studies. The investigators then genotyped 228 microsatellite markers in Icelandic individuals with type 2 diabetes and controls throughout a 10.5-Mb interval on 10q. A microsatellite, DG10S478, within intron 3 of the transcription factor 7-like 2 gene

(*TCF7L2*) was associated with type 2 diabetes (P = 2.1 x 10^{-9}) (46). The variants were all strongly correlated with each other and represent one risk allele that is present in ~28% of control subjects and ~36% of individuals with type 2 diabetes. The risk allele confers an estimated RR of 1.50 per copy carried compared to the risk of non-carriers. These were followed up by replication studies in other cohorts/ ethnic groups and there were similar findings confirming *TCF7L2* (encoding transcription factor 7-like 2) as the most important T2DM susceptibility gene to date. It has the largest allelic odds ratio (OR) of any established common variant for T2D and is ~1.35.

In the Diabetes Prevention Program (47) participants with the risk conferring TT genotype at rs7903146 were more likely to have progression from impaired glucose tolerance to diabetes than were CC homozygotes (hazard ratio- 1.55). The TT genotype was associated with decreased insulin secretion but not increased insulin resistance at baseline. Similar results were obtained for rs12255372.

A case-control study and family based association study in UK using four *TCF7L2* single nucleotide polymorphisms (SNPs: rs7903146, rs12255372, rs4506565, rs12243326) was performed (48). All SNPs were significantly associated with T2DM. In the family-based analyses strongest signal was seen for rs4506565. Similar positive replication studies in other populations clearly established *TCF7L2* as a type 2 diabetes susceptibility gene.

The *TCF7L2* gene product is thought to act through regulation of proglucagon gene expression in enteroendocrine cells via the Wnt signalling pathway. The proposed mechanism by which Variants of *TCF7L2* contribute to T2DM suggests impaired GLP-1-induced insulin secretion. This seems to be rather the result of a functional defect in the GLP-1 signalling in beta cells than a reduction in GLP-1 secretion. This defect might explain the impaired insulin secretion in carriers of the risk alleles and confers the increased risk of type 2 diabetes (49).

KCNJ11

Inwardly rectifying potassium channels (Kir channels) control cell membrane K (+) fluxes and electrical signalling in diverse cell types, like pancreatic β cells. *KCNJ11* encodes *Kir6.2,* which is an essential subunit of the pancreatic β cell potassium ATP (K_{ATP}) channel controls insulin secretion. Rare mutations of this locus lead to the monogenic syndrome of familial hyperinsulinaemia, confirming the important role of *KCNJ11* in insulin secretion. For some, mutations lead to permanent neonatal diabetes mellitus, and accompanied by neurological sequelae- developmental delay, muscle weakness, and epilepsy. Polymorphisms in the gene *KCNJ11* E23K (glutamate is exchanged by lysine at codon 23) have been shown to be associated with type 2 diabetes in several populations. The K allele was associated with diabetes in a large case-control study (odds ratio [OR] 1.18 [95% CI 1.04–1.34], $P = 0.01$) suggesting E23K variant of *KCNJ11* increases the risk of type 2 diabetes (50).

PPARγ

Variants of the *PPARγ* gene are strong candidates for conferring susceptibility to T2DM and obesity, as PPARγ regulates adipocyte differentiation, and lipid and glucose metabolism. Two *PPARγ* isoforms, *PPARγ1* and *PPARγ2*, have been characterised, and are encoded by a single PPAR gamma gene. The *PPARγ* gene has been mapped to human chromosome 3p25 by somatic cell hybridization and linkage analysis (51). The peroxisome proliferator-activated receptor γ (*PPARγ*) gene is mainly expressed in adipose tissue and is the target of the thiazolidinedione class of drugs used to treat T2DM by improving insulin action and secretion.

One of the first examples of a meta-analysis in complex disease demonstrated that the Pro12Ala variant of *PPARγ2* (proline to alanine change at position 12) is associated with predisposition to T2DM. They found a modest (1.25-fold) but significant (P=0.002) increase in diabetes risk associated with the more common proline allele (~85% frequency) (52). The less frequent alanine allele (16%) was associated with decreased diabetes risk (odds ratio 0.78, P<0.045).

CAPN10

Calpains (CAPNs) are ubiquitously expressed cysteine proteases that regulate a variety of cellular functions. *CAPN10* is expressed in β cells where evidence suggests it mediates apoptosis and insulin exocytosis (53,54); in fat and muscle *CAPN10* modifies insulin-mediated glucose transport. It also appears to be involved in myoblast/preadipocyte differentiation (55).

Calpain 10 was the first T2DM gene identified by genome-wide scan of Mexican-American families. Initially linkage was found on chromosome 2 (LOD 4.03) (56), an interaction of this locus with genes on chromosome 15 was then found and used to narrow the chromosome 2 interval (57). The critical interval was then sequenced and informative SNPs identified and the gene was identified as *calpain 10*. Three intron variants account for most of the haplotype diversity and accounted for 14% of the population attributable risk of T2DM in Mexican-Americans (58). Meta-analyses of association studies assessing *CAPN10* and T2DM risk have confirmed a role for *CAPN10* polymorphisms in T2DM susceptibility (59) increasing risk by ~ 14 % for T2DM. In one of the analyses, SNP44, a rare *CAPN10* allele, was shown to be over-transmitted from heterozygous parents to their affected offspring with T2DM. However it is pertinent to note *Calpain 10* was not identified in GWA studies casting possible doubts on the previous observations.

In contrast to the inconsistent SNP data, exciting new data is emerging on the functional significance of Calpain in insulin signalling and secretion. In Pima Indians, a *CAPN10* polymorphism (SNP43) correlates with impaired insulin action and reduced expression of *CAPN10* in the skeletal muscle of prediabetic subjects (60).

CAPN10 polymorphisms have also been associated with insulin secretion, adipocyte biology and microvascular function. Data from our lab (Prof. Hitman's lab) has demonstrated a role for *calpain 10* isoform in both phase 1 and phase 2 insulin secretion (61). Particularly *calpain 10* interacts with snare proteins leading to vesicle fusion and insulin release. Secondly we identified an important role of *Calpain 10* on actin,

proposing that calpain plays a role in facilitating the actin reorganization required for glucose-stimulated insulin secretion in INS-1 cells (62). Additionally another group has found *calpain 10* to be an important molecule for fuel sensing in beta cell of mouse islets (63). Lastly, mirroring the beta cell work it has been shown *calpain 10* facilitates GLUT4 translocation and actin reorganization of the adipocytes (64).

Environmental factors- Type 2 Diabetes

Evidence of a strong environmental element to T2DM has come from the studies of Barker and Hales (65). In a number of separate studies, a strong relationship of the development of glucose intolerance and other associated factors of the insulin resistance syndrome with low birth weight or thinness at birth has been demonstrated. Furthermore, these associations are not confined to those with growth retardation in utero but extend to through the whole range of birth weight. As a consequence of these epidemiologic studies, the 'thrifty phenotype' hypothesis has been proposed, whereby nutritional deficiencies in utero lead to poor foetal and infant growth and the subsequent development of T2DM in later life, especially when combined with obesity due to excess food intake and lack of physical activity. These changes are recognised to be due to insulin resistance, which is favourable for survival in the immediate post natal period but plays a significant role in the progression to T2DM and metabolic syndrome, and to a certain extent insulin secretion (66). While there is much discussion regarding this hypothesis, it illustrates the importance of environmental factors in early life, which might prime the foetus for T2DM in later life.

Dietary factors and physical inactivity undoubtedly affect the progression of abnormal glucose tolerance to diabetes in a genetically predisposed individual. The best way to lower the risk of diabetes is to lead a healthy lifestyle by eating a healthy balanced diet, engaging in regular physical activity and balancing the energy intake with energy expenditure. Indeed evidence would suggest that the adoption of a healthy lifestyle in high risk subjects can decrease the risk of developing T2DM by ~ 60% (67,68). There is

a close relationship between diabetes and obesity, especially when the later has central distribution (69). Apart from obesity, several other nutritional factors affect glucose metabolism and the risk of T2DM. Evidence suggests association between different types of fats and carbohydrates and insulin resistance and T2DM (70). Diets rich in saturated fats are associated with insulin resistance; a multi-centre study in a group of healthy individuals showed that a diet high in saturated fat decreased insulin sensitivity compared with a diet high in monounsaturated fat with the same total fat content (71). Prospective and cross- sectional studies suggest a role of specific types of fat rather than the total fat content in the development of T2DM, where high intake of vegetable oils, oils consisting primarily of polyunsaturated fat was associated with reduced risk of developing diabetes (72) and a positive association between saturated fat and hyperglycaemia or glucose intolerance. In a 20-years follow-up of the Finnish and Dutch cohorts of the Seven countries study, it was found that a high intake of fat (73) (in particular saturated fatty acids) contributed to the risk of glucose intolerance and T2DM.

Dietary carbohydrates are classified into simple or complex carbohydrates depending on their chemical structure. The traditional view is that simple carbohydrates be avoided and substituted with complex (starchy) carbohydrates to reduce postprandial glucose response, but this has been challenged by various studies that recognised starchy foods such as baked potatoes and white bread produce even higher glycaemic responses than simple sugars. Glycaemic index (GI) was developed to quantify the different glycaemic responses induced by different carbohydrate foods. A low GI diet with a greater amount of fibre and minimally processed whole grain products seem to improve glycaemic and insulin responses and lower the risk of T2DM. This goes to show that dietary recommendations to prevent and manage diabetes should focus more on the quality of fat and carbohydrate than the quantity alone.

A number of environmental toxins have also been shown to cause diabetes in humans, including nitrosated compounds, as well as streptozotocin, the rat poison Vacor and foods such as smoked mutton; depending on the amount consumed could either lead to T1DM or T2DM, presumably dependent on the amount of direct β cell destruction. It has also

been proposed that vitamin D might modulate the diabetic process. Vitamin D deficiency has been shown to reduce insulin secretion. In a UK study of Bangladeshi subjects (74) living in east London who were particularly prone to vitamin D deficiency, vitamin D levels were found to be low in those most at risk of diabetes. Furthermore, there was a correlation between vitamin D levels and 30 min oral glucose tolerance test, blood glucose, insulin and C-peptide levels.

1.2.3 Maturity onset diabetes of the young (MODY)

The most common forms of monogenic syndromes causing diabetes are maturity onset diabetes of the young (MODY) and maternal inheritance of diabetes and deafness (MIDD) accounting for 2–5% of 'T2DM' cases. Maturity onset diabetes of the young, are a group of monogenic disorders inherited in an autosomal dominant pattern (75). MODY is characterised by early onset of 'type 2 diabetes' (usually before the age of 25 years), β cell dysfunction and a family history (at least two generations) of early onset diabetes. The defect is in insulin secretion due to mutations in the *glucokinase* and β cell transcription factor genes **(Table 1.e)** (76). *HNF1α* also known as hepatic transcription factor 1 *(TCF1), HNF1β* (also known as *TCF2*) *and 4α,* insulin promoter factor *(IPF1)* and neurogenic differentiation-1 *(NEUROD1)* have an important role in the normal development and function of the β cell of pancreas.

In the UK mutations in *HNF1α* is the commonest cause of MODY accounting for 63% of cases, followed by mutations in the glucokinase gene (20% of the cases) (77). The clinical presentation and progression of diabetes is different between patients with mutations of *glucokinase*, *HNF1α* and *HNF1β*. Subjects with *glucokinase* mutations are frequently asymptomatic or detected with gestational diabetes and when diagnosed associated with a milder form of diabetes, which is frequently treated with diet alone and not associated with the complications of diabetes. In contrast those subjects with *HNF1α* mutations are more like lean T2DM with susceptibility to microvascular complications

and progressive loss of β cell function made worse by increasing body mass index. In comparison to patients with T2DM, subjects with *HNF1α* mutations are very sensitive to sulphonylurea treatment as might be predicted from the genetic defect (78). Patients with *HNF1β* in addition to T2DM have renal cysts that may lead to renal failure and hence such patients are more frequently found in the renal clinic. Due to its monogenic aetiology, MODY is a good example of a disease for which genetic diagnosis is clinically useful.

Table 1.e: Maturity onset diabetes of the young

Gene	Chromosome	MODY frequency
HNF 4 α (MODY1)	20q	5%
GCK (MODY2)	7p	10-15%
HNF1 α (MODY3)	12q	20-75%
IPF1 (MODY4)	13q	rare
HNF1 β (MODY5)	17q	rare
NEUROD1 (MODY6)	2q	rare

GCK

Genes known to be involved in β-cell function are obvious candidates for conferring predisposition to diabetes. *GCK* (Glucokinase) is one such gene, acts as a pancreatic glucose sensor by catalysing the rate-limiting step of glucose to glucose-6-phosphate in β cells. Investigation of *GCK* as a candidate gene in large families and nucleotide sequencing demonstrated it as a cause of MODY in some families (79). Mutations in the *GCK* gene are associated with glucose-sensing defects in β cells. These mutations tend to be associated with 'mild', asymptomatic or gestational diabetes, or IGT and patients are free of diabetes related complications.

HNF1α

Hepatocyte nuclear factor-1 alpha (*HNF1α)* also known as hepatic transcription factor 1 (*TCF1*), functions as a regulator of multiple genes, including insulin like growth factor and insulin in rodents. The *HNF1α* gene is located on chromosome 12q and was identified from positional cloning and genome scans in 12 families with MODY (80,81)The clinical features associated with *HNF1α* variants are markedly different to those seen in patients with *GCK* defects and are more typical of T2DM. *HNF1α* associated diabetes is characterised by insulin secretory defects and hyperglycaemia and weight gain worsens the metabolic control. Unlike *GCK* mutations, patients with *HNF1α* mutations are at risk of microvascular complications. Although approximately 40% of patients are treated with insulin, diabetic patients with *HNF1α* mutations are sensitive to sulphonylureas and treatment with sulphonylurea is effective in a subset of patients (82).

THE WRS GENE- *EIF2AK3* GENE

1.3 THE WOLCOTT- RALLISON SYNDROME-*EIF2AK3*

Wolcott- Rallison syndrome (WRS) is a rare autosomal recessively inherited disorder. The syndrome is characterised by a short trunk compared to arm span, multiple epiphyseal dysplasia, bone demineralisation, multiple fractures, tooth discolouration, abnormal skin, hepatosplenomegaly, renal insufficiency, exocrine pancreatic dysfunction in addition to endocrine pancreatic dysfunction leading to insulin dependent diabetes (83). The onset of diabetes in WRS families is mainly below the age of 6 months and is characterised by permanent severe insulin deficiency. In those families examined, islet cell antibodies were not detected in the diabetic subjects excluding autoimmunity as a cause and in a case of WRS that came to autopsy the pancreas was found to be hypoplastic (84), with interstitial fibrosis and poor staining for insulin. It is therefore possible that the diabetes of WRS is due to beta cell dysfunction and represents a non-autoimmune cause of insulin requiring diabetes. Mutations of the gene encoding eukaryotic translational initiation factor 2 - alpha kinase 3 (*EIF2AK3*) were found to account for WRS in 2 families (85). Initial investigators studied 2 consanguineous families and carried out genome -wide linkage study which mapped WRS gene to a region of less than 3 cM on chromosome 2p12, with maximum evidence of linkage at 4 microsatellite markers (D2S113, D2S1786, D2S2181 and D2S2222). The gene encoding eukaryotic translational initiation factor 2- alpha kinase 3 (*EIF2AK3*) resides in this interval.

Markers D2S1994/WI-6863 was assigned to the segment between D2S1786 and D2S2181 and all 3 markers are located in a single YAC clone from the Whitehead contig WC2.7. D2S1994/WI-6863 was identified as *EIF2AK3* encoding a serine/threonine kinase (also known as pancreatic eukaryotic initiation factor 2-alpha kinase, *PEK*). Previous mapping of *EIF2AK3* to chromosome 2p12 was carried out by Hayes et al (86). *EIF2AK3*, also designated as pancreatic eIF2-alpha kinase *PEK/ PERK* by various authors, was considered as a possible candidate gene for WRS as it was highly expressed

in pancreatic islets and has a role to play in endoplasmic reticulum stress (to be discussed in detail in next chapter) and apoptotic pancreatic beta cell death. They identified distinct mutations in both index cases in homozygous state in both the families. The other affected siblings were homozygous for the mutation and the unaffected siblings and parents were heterozygous. In WRS1 (family one), the *EIF2AK3* mutation consisted of an insertion (T) at position 1103 (1103insT) creating a frameshift at position 345 and premature termination. This probably results in complete loss of function of the protein as it is devoid of part of the regulatory domain (aa 1-576) and the entire catalytic domain (aa 577-1,115). In WRS2 (family two) the mutation is a 1832 G to A transition resulting in a change of glutamine for arginine at position 587 (R587Q) within the catalytic domain (87). This amino acid is completely conserved in the gene of different organisms (humans, mouse and rat). These results showed that mutations in *EIF2AK3* are responsible for WRS in the two consanguineous families studied. **(Figure 1.c)**

Following the initial studies, direct sequencing of the WRS gene in a girl with clinical features of WRS revealed the presence of a homozygous T to C mutation in exon 13 leading to a missense S877P (Serine to Proline) mutation in the catalytic domain of the *EIF2AK3* protein (88). Serine 877 is conserved in human, mouse and rat proteins. Functional studies of the particular mutation showed the P877 (Proline) mutated protein is less efficient than the wild type kinase in phosphorylating itself and is completely unable to phosphorylate eIF2alpha.

Case studies on WRS (89) looked into the genetic composition for the *EIF2AK3* gene and reported 2 further mutations. In the 1st case the offspring of a consanguineous Saudi parents presented with diabetic ketoacidosis and was treated with insulin, had skeletal dysplasia and direct sequencing of *EIF2AK3* revealed a homozygous deletion of four nucleotides (1563delGAAA) at the site of a GAAA 4 base pair direct repeat in exon 9 which results in an immediate premature termination codon at amino acid 523. Heterozygous deletions were seen in the parents and unaffected sibling. In the 2nd case with similar clinical presentation and family background sequencing of *EIF2Ak3* gene showed a homozygous G to A substitution in Intron 14 resulting in splicing

abnormalities. Both of these mutations were novel and result in truncated proteins that lack the critical kinase domain.

Figure 1.c: Mutations within *EIF2AK3* in Walcott-Rallison Syndrome families

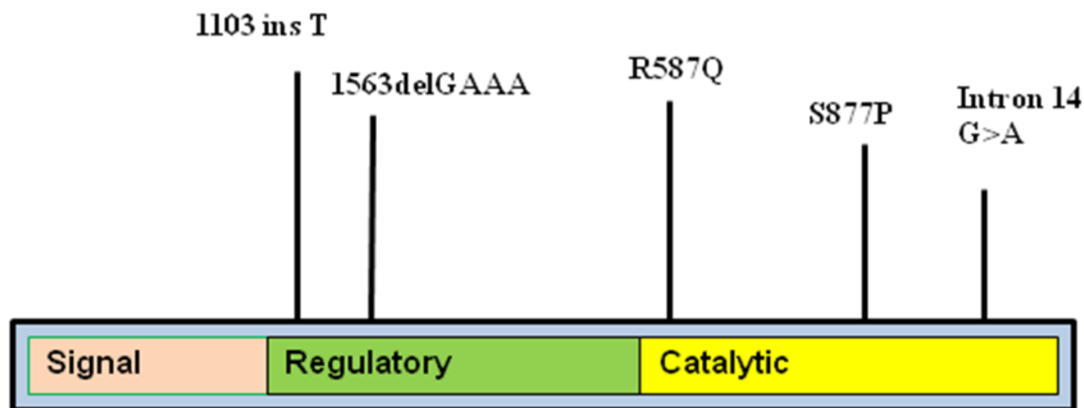


Figure 1.c: 110insT- Insertion (T) at position 1103, 1563delGAAA- 4 base pair direct repeat in Exon 9 leading to premature termination codon and loss of function. R587Q- Glutamate for Arginine at position 587, S877P Proline for Serine at position 877 in the catalytic domain leading to loss of function. G to A substitution in Intron 14 leading to splicing abnormalities and truncated protein.

In a more recent study of 12 families (18 patients) with WRS (90) the age at onset was generally very young: the mean age at onset was 3 months. There was multiple epiphyseal dysplasia common to all patients and various degrees of osteopenia, with multiple fractures, slight to severe mental retardation or developmental delay in the majority of the patients. Four patients showed signs of exocrine pancreas dysfunction, with pancreatic hypotrophy observed in one patient and fibrotic infiltrations in pancreas in another. Sequencing of the coding regions of *EIF2AK3* was performed in all the WRS cases and in available parents and siblings. *EIF2AK3* mutations were identified in 11 of the 12 families, in the homozygous state in the patients of 10 of them, and as a compound heterozygote in one individual, whose parents were not consanguineous. None of these

12 mutations were found in a control population of 95 Caucasian individuals. The gene defects- nonsense mutations, frameshift mutations, splice-site mutations, missense mutation, resulted in truncated proteins. All of the *EIF2AK3* mutations that resulted in truncated proteins are missing all or part of the kinase domain, which would be expected to lead to a complete loss of function. The functional consequence of the missense mutations was evaluated by measuring the *EIF2AK3* activity using the yeast translational assay. There was defective autophosphorylation of EIF2AK3 protein and of eIF2α phosphorylation activity in the mutant variants of *EIF2AK3* as compared to normal function in the wild type indicating loss of *EIF2AK3* function.

1.4 THE *EIF2AK3* GENE REGION AND TYPE 1 DIABETES IN SUBJECTS FROM SOUTH INDIA

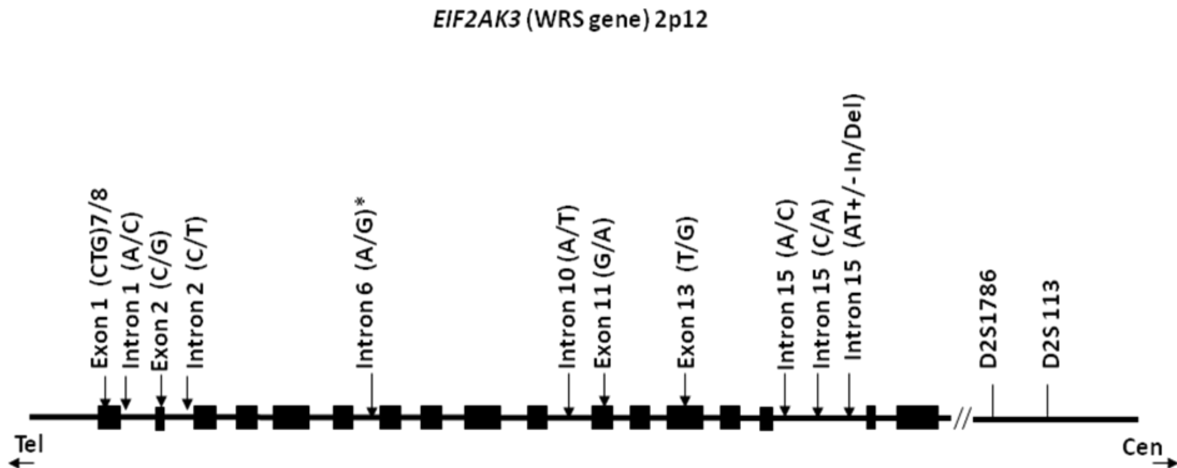
Researchers in our lab, Dr. Rebecca Allotey et al, were involved in investigating whether there is an association between the *EIF2AK3* gene and T1DM (91). Using a family based approach (234 South Indian families) they studied markers from the 3' flanking region of the locus: D2S113 (the linked marker in the T1DM Scandinavian genome scan), D2S1786 and an AT insertion/deletion at intron 15 (15INDEL). The families had been ascertained through a proband with T1DM, with both parents also available for study. The D2S1786 marker was found to be significantly transmitted to the affected offspring (overall allele-wise ETDT; P=0.0098 uncorrected; P=0.03 corrected); this was partly attributable to allele 4 (215 bp) being transmitted 142 times compared to 102 non-transmissions (P=0.010) and decreased transmission of allele 5 (219 bp) (46 transmissions vs. 71 non-transmissions; P=0.021). Excess transmission was also found of the common allele of 15INDEL. Allele 1 was transmitted 99 times and not transmitted 69 times (allele wise ETDT; uncorrected P=0.02; P=0.06 corrected); this could be largely explained by a paternal effect (P=0.005; allele 1, 33 transmissions vs. 11 non-transmissions). No association was found between T1DM and D2S113 (13 alleles) in the South Indian families. D2S113 is located on a separate YAC (WC 2.8) to the disease-associated marker D2S1786 (WC2.7 that also includes *EIF2AK3*). TRANSMIT was used to analyse the potential effects on diabetes risk of the D2S1786 and 15INDEL marker combinations. Of 14 possible combinations, only four were found to have a frequency greater than 5%. There was evidence for an association between T1DM and particular *EIF2AK3* marker combinations (overall global P<0.001), with excess transmission of both the 15INDEL allele 1/D2S1786 allele 4 and 15INDEL allele 1/D2S1786 allele 3 combination (both P<0.0001), and decreased transmission of 15INDEL allele 1/D2S1786 allele 5 combination (P<0.001). Since there were a large number of D2S113 alleles, it was not possible to perform an accurate TRANSMIT analysis of all the three markers (15INDEL two alleles, D2S1786 five alleles and D2S113 13 alleles), because of the large number of haplotypes generated. Since there was evidence for an association between markers from the *EIF2AK3* region and T1DM, the researcher's proceeded to compare

allele frequencies of the disease associated 15INDEL and D2S1786 markers in ethnically matched unaffected subjects, with normal glucose tolerance and the previously typed probands with T1DM. The frequencies of the 15INDEL genotypes (1/1, 1/2 and 2/2) in 286 subjects without diabetes were 40.9, 47.9 and 11.2%, respectively, and the corresponding figures in the 225 probands with T1DM were 68.9, 30.2 and 0.9% (Fisher's exact two-sided $P<0.0001$; odds ratio for 1/1 genotype 3.3, 95% CI 2.2–4.7). Genotype frequencies in the subjects were in Hardy–Weinberg equilibrium. Supporting the family data, allele 4 of D2S1786 was present in 82.4% of the probands compared to 60.9% of subjects without diabetes ($P<0.0001$). Similarly, there were fewer subjects with allele 5 of D2S1786 in the probands (22.5%) compared to subjects without diabetes (35.4%; $P=0.001$). The allele frequencies in the control subjects therefore support an association between T1DM susceptibility and *EIF2AK3* in this ethnic group. The allele frequency of the 15INDEL in South Asians is more frequent to that previously reported in Europeans. In order to validate the genotyping, sequencing was carried out in 48 subjects (30% possessing the uncommon allele); in all cases, the sequencing results were concordant with the PCR-RFLP assay.

Based on this evidence of a disease association with the two markers closest to the *EIF2AK3* gene, 24 subjects were selected for comprehensive sequencing studies of all 17 exons and exon/intron boundaries of the gene to identify any novel mutations and define haplotype blocks within the gene. These subjects comprised 10 South Indian T1DM probands with the disease-associated combination (allele 1 of 15INDEL and allele 4 of D2S1786, subsequently referred to as the 1:4 combination) that had been transmitted from a parent, 12 South Indian subjects (three parents, two probands and six controls), who did not possess the 1:4 combination, and two European controls (untyped). The previously reported mutations leading to WRS were not found in this study. Researchers would not have anticipated detecting the mutations in our subjects given that they did not have any features of WRS and the small sample size of the population studied has insufficient power to detect rare variants. However, single nucleotide polymorphisms at 11 different loci of the *EIF2AK3* gene were detected, corresponding to nine previously reported polymorphisms and two novel SNPs found in exon 1 (position 242 G/T;

Arg57Leu) and intron 6 (position 2661 A/G), respectively. Inspection of the sequences in these subjects suggests linkage disequilibrium (LD) of all SNPs apart from the 15INDEL, and that five possible haplotypes were defined by the four SNPs 2, 3, 10 and 11. Researchers therefore genotyped the families for the four relevant SNPs by a PCR RFLP assay. LD between all markers used in this study was assessed using the PM (permutation and model-free analysis) and PM plus programmes. These programs allow for both a permutational approach, as well as analysis between two loci and between haplotype blocks, and they confirmed the results from the 24 sequenced subjects. SNPs 2, 3, 10 and 11 were all in tight LD and formed a specific haplotype block. Of 16 possible haplotypes, only 14 were observed, and five of these were present at a frequency of greater than 5%. The haplotype block composed of SNPs 2, 3, 10 and 11 was in linkage equilibrium with all other surrounding markers. No individual marker or haplotype was transmitted in excess to the diabetic offspring (TRANSMIT global P= 0.18) from within the block. The two disease-associated markers in the study (D2S1786 and 15INDEL) were in LD with D2S113 (the marker linked to T1DM in the Scandinavian population), thereby placing disease susceptibility in the 3' flanking region of the *EIF2AK3* locus. To confirm the preliminary findings that disease susceptibility lies 3' of the gene, Dr.R. Allotey used a PCR-RFLP assay to investigate two other reported SNPs, SNP C (rs2364564) and SNP E (rs1606803) (located approximately 4 and 6 kb, respectively, 5' of *EIF2AK3*) in the study population. No association was found between SNP C and SNP E and T1DM susceptibility. Analysis using the PM plus programmes revealed that SNP C and E are in LD with each other and also with markers located to the central haplotype block (SNPs 2, 3, 10 and 11). The two SNPs located 5' of the gene were not in linkage with the disease-associated markers, 15INDEL and D2S1786, which are 3' position, further supporting the hypothesis of disease susceptibility located either 3' or downstream of the *EIF2AK3* locus in this population.

1.5 STRUCTURE OF THE *EIF2AK3* GENE AND CHROMOSOMAL LOCATION (Figure 1.d)



EIF2AK3 gene showing results of nucleotide sequencing, SNPs genotyping and microsatellite markers.

EIF2AK3 gene is found on Chromosome 2 at location 88,637,376-88,708,209 (OMIM *604032). It is also known by other names- *PEK/ PERK*. Researchers identified a rat eIF2-alpha kinase that they designated *PEK* (pancreatic eIF2-alpha kinase) (92). The *EIF2AK3* gene was mapped to chromosome 2p12 (93). A targeted mutation of the mouse *EIF2AK3* gene, which they called Perk, abolished the phosphorylation of eIF2-alpha in response to accumulation of malfolded proteins in the ER, resulting in abnormally elevated protein synthesis and higher levels of ER stress (94). Mutations (nonsense mutations, frameshift mutations, splice-site mutations, missense mutation) in the gene contributing to Wolcott- Rallison syndrome (monogenic Mendelian disease) in humans have been reported (95) (96). *EIF2AK3* gene size is 7.8 kb and it has 17 exons.

PANCREATIC β CELL, APOPTOSIS AND ER STRESS

1.6 ROLE OF APOPTOSIS IN PANCREATIC β CELL DEATH IN DIABETES

Type 2 diabetes is the result of two major defects – insulin resistance and pancreatic β cell dysfunction leading to chronic hyperglycaemia. Insulin resistance in the muscle and liver is thought to be the primary defect in individuals with type 2 diabetes. However, pancreatic β cell dysfunction plays a critical role in manifestation of T2D. The disease progression from IGT to type 2 diabetes involves not only peripheral insulin resistance but also a defect in β cell insulin secretory capability due to relative reduction in β cell mass **(Figure 1.e).** There is compensatory hyperinsulinaemia in insulin resistant individuals in the early stages of the disease. This increased insulin secretion maintains normal blood glucose levels in the initial stages but as time progresses there is progressive decline in β cell function leading to impaired glucose tolerance and finally to the manifestation of T2D. Studies in non-diabetic relatives of type 2 diabetes patients show evidence of defects in insulin secretory pattern. There are animal and human studies, which show conclusive evidence that apoptosis may be involved in β cell failure in type 2 diabetes. Pancreatic β cells are sensitive to a number of proapoptotic stimuli. The critical role of endoplasmic reticulum (ER) stress in the process of apoptosis and the specific role of *EIF2AK3* in regulation of protein translation and ER stress will be discussed later in the chapter.

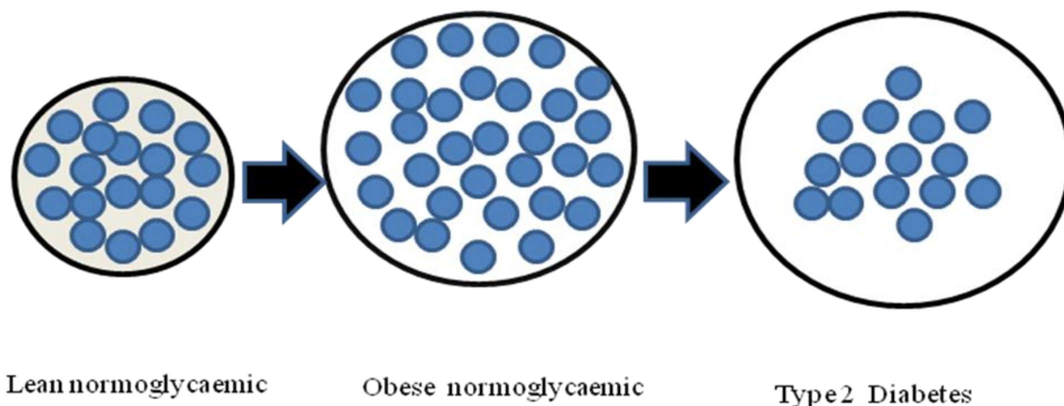


Figure 1.e: Relative reduction in beta cell mass, although not reduced compared to lean normoglycaemic individuals β cell mass is reduced in type 2 DM patients when compared to obese normoglycaemic individuals.

Apoptosis or programmed cell death occurs during normal cellular development and is an energy dependent process characterised by cell shrinkage, condensation of nuclear chromatin, membrane blebbing and the formation of apoptotic bodies that undergo phagocytosis (97). Activation of apoptosis in β cells is elicited by various stimuli- death receptor activation by proteases, oxidative stress, Ca dependent processes, endoplasmic reticulum stress, mitochondrial dysfunction etc **(Table 1.f).**

Table 1.f: Potential inducers of β cell apoptosis

Potential inducers
Death receptors
Fas ligand/Fas receptor
Perforin
Cytokines (IL-1*β*, IFN-*γ*, TNF-*α*)
Membrane permeable mediators
Reactive oxygen species (hydrogen peroxide, hydroxyl radicals)
Reactive nitrogen species (nitric oxide, peroxynitrite)
Alkylating agents (streptozotocin, MMS)
Ceramid
Inadequate growth factors
IGF-1
EGF
FGF
Perturbation of the metabolic and signal pathways
Increased
glucose
Free fatty acids
Amylin
Calcium
Decreased
Cell cycle regulator (cyclin dependent kinase4)
Transcription factor (BETA/NeuroD,HNF-1*α*)
Mitochondria function (mitochondrial transcription factor A)
ER stress transducer (PERK/EIF2AK3)

In a study that was done to determine whether β cell apoptosis increased in type 2 diabetes in humans, pancreatic tissue from post-mortem samples in lean and obese type 2 diabetes patients, and lean and obese controls was looked into for relative β cell volume, frequency of β cell apoptosis and replication, and new islet formation from exocrine pancreatic tissue; neogenesis (98). Pancreatic tissue from 124 autopsies was examined:

91 obese cases; 41 with type 2 diabetes, 15 with impaired fasting glucose, and 35 non diabetic subjects and 33 lean cases; 16 type 2 diabetic and 17 non diabetic subjects. They measured relative β-cell volume, frequency of β cell apoptosis and replication, and new islet formation from exocrine ducts. Relative β cell volume was increased in obese versus lean non-diabetic cases (P = 0.05) through the mechanism of increased neogenesis (P < 0.05). **(Figure 1.f)**

Figure 1.f: β cell replication, apoptosis and β cell volume

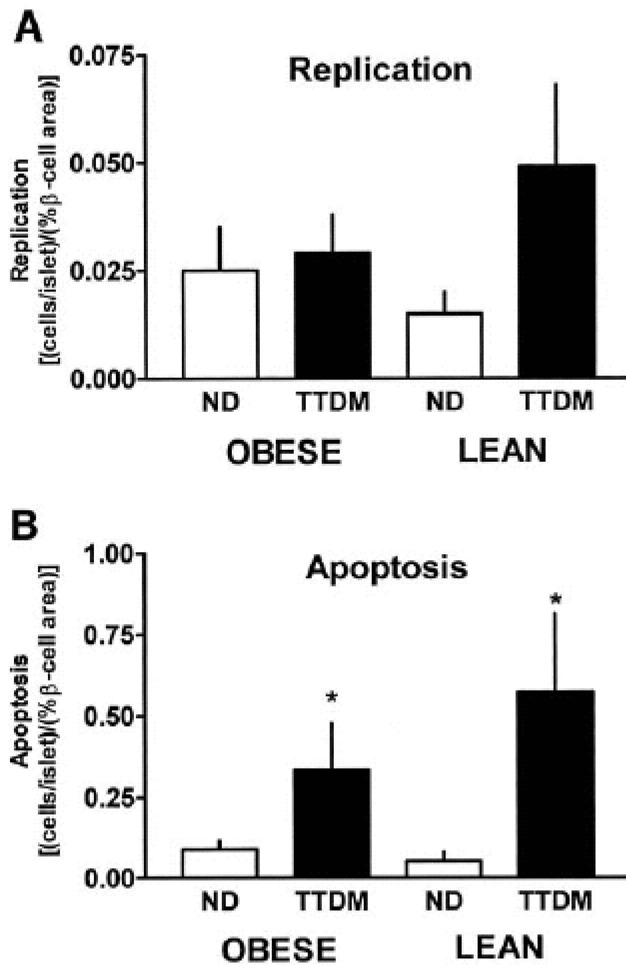


Figure 1.f: The frequency of β cell replication (*A*) and β cell apoptosis (*B*) normalized to relative β cell volume in each case. ND, nondiabetic; TTDM, type 2 diabetes. (99)

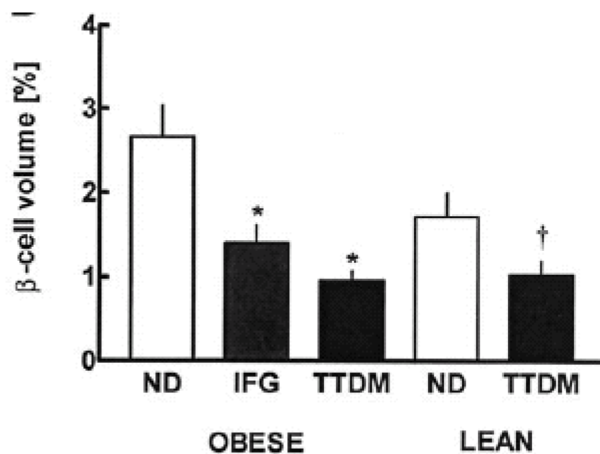


Figure 1.f: The mean β cell volume in obese (nondiabetic [ND], IFG, and diabetic subjects [TTDM]) and lean cases (nondiabetic and type 2 diabetic subjects). (100)

Obese humans with IFG and type 2 diabetes had a 40% ($P < 0.05$) and 63% ($P < 0.01$) deficit and lean cases of type 2 diabetes had a 41% deficit ($P < 0.05$) in relative β cell volume compared with non diabetic obese and lean cases, respectively. The frequency of β cell replication was very low in all cases and no different among groups. Neogenesis, while increased with obesity, was comparable in obese type 2 diabetic, IFG, or non-diabetic subjects and in lean type 2 diabetic or non-diabetic subjects. However, the frequency of β cell apoptosis was increased 10-fold in lean and 3-fold in obese cases of type 2 diabetes compared with their respective non-diabetic control group ($P < 0.05$). This study shows that β cell mass is decreased in type 2 diabetes and that the mechanism underlying this is increased beta-cell apoptosis.

There could be several mechanisms contributing to β cell apoptosis in type 2 diabetes. Hyperglycaemia in T2DM individuals can potentially induce apoptosis in several types of cells. The processes involve high glucose activated NFκB, mitochondrial cytochrome C-mediated caspase 3 activation and formation of reactive oxygen species (ROS). Similar pathways have been suggested to be activated by glucose in pancreatic β cells (101).

The toxicity in T2DM may in part indirectly be caused by "lipo-toxicity" implicating alterations in β cell malonyl-CoA, peroxisome proliferator activated receptors α and γ and steroid regulatory element binding protein expression (102). Free fatty acid induced β cell

apoptosis was suggested to involve formation of ceramide, increased production of nitric oxide and mitochondrial pathways (103). **(Figure 1.g)**

Figure 1.g: A model of β-cell apoptosis in type 2 diabetes

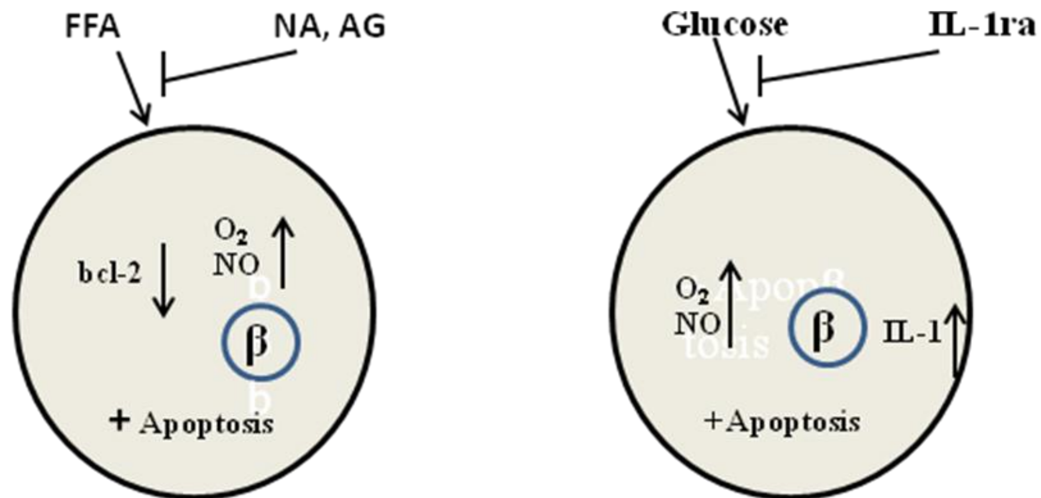


Figure 1.g: Free fatty acids (FFA) cause β cell apoptosis by induction of free oxygen radicals (O2), nitric oxide (NO) synthesis as well as down-regulation of anti-apoptotic proteins such as bcl-2 (B-cell lymphoma). The toxic effect of free fatty acids can be blocked by nicotinamide (NA) or aminoguanidine (AG) both blockers of nitric oxide synthase. Glucose-induced β cell apoptosis involves the induction of free oxygen radicals and β-cell synthesis of IL- 1 (Interleukin-1) and activation of IL-1 signalling leading to β cell apoptosis. The pro-apoptotic effect of glucose on human islets can be blocked by interleukin-1 receptor antagonist (IL-1Ra) **(104)**.

Hyperglycaemia up-regulates the Fas-receptor in human islets (105) which might cause induction of apoptosis in neighbouring β cells constitutively expressing Fas-ligand. The mechanism underlying glucose induced β cell Fas-expression was explained by glucose-induced production of IL-1-β (106). Thus β cells exposed to high glucose expressed IL-1-β thereby inducing apoptosis. These events were associated with NFκB activation, DNA fragmentation and impaired β cell function. The apoptotic effect of glucose could be blocked by the natural antagonist of IL-1 action interleukin-1 receptor antagonist (IL-1Ra). A recombinant interleukin-1 receptor antagonist has been used in clinical trials for

the treatment of type 2 diabetes and showed improved glycaemic control and beta-cell secretory function and reduced markers of systemic inflammation (107).

Thus, inflammatory mediators expressed as part of the autoimmune response in type 1 diabetes or inflammatory mediators expressed in β cells in response to high glucose in type 2 diabetes may be a common denominator for β cell apoptosis in these two diseases(108).

1.7 ENDOPLASMIC RETICULUM STRESS AND ROLE OF *EIF2AK3* IN ER STRESS

There is a highly developed endoplasmic reticulum (ER) in pancreatic β cells to facilitate insulin secretion. Pancreatic β-cells utilise nearly 70% of its protein synthesis for insulin production. The ER is a highly dynamic organelle and plays a central role in lipid and protein biosynthesis (109). The ER produces the transmembrane proteins and lipids for most cell organelles and is responsible for the synthesis of almost all secreted proteins. The ER also has an important role in calcium storage and signalling. Due to its ability to store and secrete Ca^{2+}, the ER controls a wide range of cellular processes such as organogenesis, transcriptional activity, stress responses, and apoptosis. The ER with its several important functions- synthesis, folding and assembly of proteins, post-translational modifications, lipid synthesis and a cellular calcium store, and disruption of any of these processes leads to ER stress. The ER stress (110) inducers include hyperglycaemia, high free fatty acids, inhibition of protein glycosylation or disulphide bond formation, calcium depletion from the ER lumen, expression of mal-folded proteins and impairment of protein transport from the ER to the Golgi apparatus. Therefore it could be argued ER stress is of particular relevance in type 2 diabetes. Chronic hyperglycaemia leading to pancreatic β cell stimulation and increased insulin synthesis could potentially activate ER stress pathways. A high fat diet and obesity may contribute to the development of type 2 diabetes by causing β cell lipotoxicity and insulin resistance. FFAs activate an ER stress response in β cells, with palmitate being more potent than oleate. Glucotoxicity and lipotoxicity have been shown to contribute to the β cell insulin secretory failure and potentially precipitate β cell apoptosis and failure and play a role in the pathophysiology of T2DM (111-113).

Animal studies have shown that interfering with normal ER function can result in apoptosis (114). To survive under ER stress conditions, cells have a self- protective response, termed the ER stress response or also known as unfolded protein response (115). ER stress and unfolded protein response can be protective to the β cells under euglycaemic and normal physiological conditions. Moreover, the same response has the

potential for triggering β cells dysfunction through apoptosis under the conditions which potentiate ER stress.

The four distinctive responses in unfolded protein response under ER stress include

1. Upregulation of genes encoding ER chaperone proteins to increase protein folding activity and to prevent protein aggregation.
2. Translational attenuation.
3. ER associated degradation of malfolded proteins.
4. Apoptosis when ER function is severely impaired.

Translational attenuation- Role of *EIF2AK3* on endoplasmic reticulum stress

In response to various environmental stresses, eukaryotic cells downregulate protein synthesis by phosphorylation of the alpha subunit of eukaryotic translation initiation factor-2 (eIF2-alpha). The *EIF2AK3* gene has an essential role in protecting the cell from endoplasmic reticulum stress. mRNA translation is attenuated as part of ER stress response due to malfolded proteins. This occurs at the level of translational initiation via eIF2α phosphorylation. There are four protein kinases known to phosphorylate eIF2α- GCN2 (general control non-derepressible-2), heme-regulated inhibitor (HRI), double-stranded RNA dependent protein kinase (PKR), and *EIF2AK3*. *EIF2AK3* (protein 1115 amino acid) enriched in pancreatic cells, is uniquely located to the endoplasmic reticulum and is activated by the accumulation of unfolded proteins in the ER lumen. *EIF2AK3* is the kinase responsible for this phosphorylation during ER stress. *EIF2AK3* phosphorylates eukaryotic initiation factor- 2 alpha (eIF2 α) which then inhibits eIF2 β, leading to cellular events that inhibit mRNA translation (116). This is one of the fundamental responses to ER stress and prevents further cell damage by apoptosis. Conversely, prolonged activation of *EIF2AK3* as part of unfolded protein response can induce apotosis of β cells and potentially implicating this response in the pathophysiology of T2D suggesting a dual role of ER unfolded protein response(117). The ER stress pro- apoptotic downstream signals include CHOP, IRE 1, JNK, CFOS/ JUN, IL- 1β(118).

As was discussed in the previous section, increased β cell apoptosis is known to occur in the diabetes state in both obese and non obese individuals and this contributes to their diabetes phenotype. Type 2 diabetes is characterized by decreased insulin secretion in the presence of increased insulin resistance. The susceptibility to endoplasmic reticulum stress and thereby apoptosis differs from cell to cell and β cells of pancreas may be one of the most susceptible cells given their extensive role in protein synthesis and secretion.

1.8 HYPOTHESIS AND AIMS OF THE STUDY

We propose that **common polymorphisms in the *EIF2AK3* gene (Candidate gene association study)** could be associated with **type 2 diabetes** based on the following observations:

Monogenic syndrome

Mutations in *EIF2AK3* have been discovered in humans and the resultant syndrome, Wolcott-Rallison syndrome results in severe insulin deficiency due to a non-autoimmune process in infancy and a particular phenotype as was discussed in the earlier chapter, clearly establishing the role of the *EIF2AK3* mutations in WRS. Other clinical features of WRS include mental retardation, hepatic and kidney dysfunction, cardiac abnormalities, exocrine pancreatic dysfunction, and neutropenia.

Genomic and cDNA clones encoding human *PEK*/ *EIF2AK3* have been isolated. And it was shown *EIF2AK3* is specifically expressed in pancreatic islets. The *EIF2AK3* gene is expressed ubiquitously and the highest expression of *EIF2AK3* was shown to be in the delta cells of the pancreas, which is the main site of secretion of somatostatin hormone (119). Subsequent studies in humans have shown expression of *EIF2AK3* in the β cells of pancreas. Study of fixed sections of human adult pancreas by immunohistochemistry with a polyclonal antibody to *EIF2AK3* showed *EIF2AK3* to be expressed extensively in the islet, with a predominance in β cells (120).

The islets of Langerhans in the pancreas play a critical role in maintaining normoglycemia in humans. They secrete different hormones in a highly regulated fashion in response to changes in environmental conditions. In the pancreatic islets- α cells secrete glucagon, β cells secrete insulin, δ cells secrete somatostatin, and PP cells secrete pancreatic polypeptide. The changes in blood glucose levels are an important factor that affects the synthesis and secretion of various islet hormones. The action of several hormones including insulin and glucagon helps to maintain the blood glucose in a narrow range under physiologic conditions. Somatostatin also plays an important role in

maintaining normal blood glucose levels by modulating secretion of both insulin and glucagon. Therefore any defects in the islet formation during embyrogenesis as happens with transcription factor mutations in MODY or pancreatic damage related to apoptosis as occurring with *EIF2AK3* mutations in WRS leads to manifestation of diabetes.

Animal Studies

The *EIF2AK3* gene has an essential role in protecting the cell from endoplasmic reticulum stress as discussed in the previous section. Targeted mutation of the mouse *EIF2AK3/PERK* gene abolishes phosphorylation of eIF2 α, and the resultant phenotype in mice homozygous for the mutation is characterised by insulin deficiency, hyperglycaemia, growth retardation, steatorrhoea and high mortality between 2 and 4 weeks of life (121). In contrast, heterozygous mice are phenotypically similar to wild type mice, although they are glucose intolerant following intra-peritoneal injection of glucose. Pancreatic islets at postnatal day 12 are reduced in size, with decrease in insulin containing cells and an increase in glucagon secreting cells. *PERK –/ –* mice show rapid and progressive decline in both the endocrine and exocrine pancreatic function. Furthermore, there is evidence of increased programmed cell death/ apoptosis of the beta cells.

Biology of *EIF2AK3*

EIF2AK3/ *PERK/ PEK* is an endoplasmic reticulum transmembrane resident protein and is activated by ER stress. *EIF2AK3* by its actions of phosphorylating eIF2α leading to attenuating effect on protein synthesis in the β cell of pancreas helps protect the cell from ER stress. Targeted mutation of the *PERK* gene in the mouse embryonic stem cells was undertaken to highlight the effects of such experimentation. The *PERK –/ –* mice cells were unable to phosphorylate eIF2 α and thereby unable to lead to translational repression in response to ER stress (122). When the cells were exposed to pharmacological agents that induce ER stress, the wild type mice cells exhibited a significant attenuation in protein synthesis rates. However the *PERK –/ –* mice cells did not reduce protein synthesis. Furthermore, the *PERK* mutation lead to increased parallel

ER stress pathways leading to IRE1 phosphorylation and caspase 12 processing thereby leading to increase in pro-apoptotic activity by these proteins. In *PERK –/ –* mice studies histological analysis and immunochemistry demonstrate the normal development of pancreas but rapid deterioration in the pancreatic function indicating the tissue dysfunction being acquired post natally. The progressive loss of β cell function is well recognised in type 2 diabetes. Therefore it could be plausible that variations in *EIF2AK3* could contribute to T2DM by way of chronic endoplasmic reticulum stress pathways.

Linkage evidence (T1DM)/ Preliminary data on type 2 diabetes- *EIF2AK3*

At the time of initiating the hypothesis LOD scores of > 2 were observed on chromosome 2, at D2S113 near *EIF2AK3* (LOD score 2.1) in a Genome wide Scan for type 1 diabetes Susceptibility in Scandinavian Families (123). The evidence of linkage at this locus was increased in the HLA DR3/4$^+$ sib pairs (multilocus LOD score 2.6), compared with that observed in the complete data set. This suggests that *EIF2AK3*—or another gene located in this region—may contribute to T1DM susceptibility. Since endoplasmic reticulum stress is common to type 1 diabetes and type 2 diabetes it might be hypothesised that common variants of the *EIF2AK* gene could be associated with type 2 diabetes.

As a preliminary study, we examined the disease associated INDEL15 in 96 South Indian families, selected for a proband with T2DM. Excess transmission was found of the T1DM associated allele 1 of INDEL15 (p=0.03). However, similar results were not obtained in a separate case-control study of 72 South Indian subjects with T2DM. As part of the Diabetes UK Warren 2 study, we (in collaboration with Prof. Mark McCarthy's group) have also studied 150 British/Irish trios, ascertained through a proband with T2DM. Five polymorphisms- SNP 2, 10, 11 and 13 and INDEL15 were genotyped through this cohort. Although no single marker including INDEL15 was statistically significant of itself, marked deviation from expected transmission was found for haplotypes with a frequency of >5% (p=0.002). These are unpublished data. The two datasets in different ethnic groups would therefore suggest an association between *EIF2AK3* gene and T2DM, but again the exact aetiological variants that might account for the association have not been identified. There appeared to be a positive signal albeit

a minor signal for association of *EIF2AK3* gene variants with T2D. In the preliminary study only a few SNPs were studied thereby lacking adequate coverage of the gene. Furthermore only a small cohort of individuals was studied, thereby lacking power. This has the potential of leading to false positive or false negative associations. A comprehensive study of gene variants in the exonic and intronic location along with study of SNPs 5' and 3' of *EIF2AK3* gene was planned for T2DM and control groups. Given the biology of the *EIF2AK3* gene and the links with pancreatic β cell dysfunction noted with gene variants, albeit in monogenic WRS, we felt it was worthwhile to pursue the gene in polygenic Type 2 diabetes.

As we have a

1) Human model of monogenic disease with *EIF2AK3* mutation, the WRS syndrome
2) Animal model of PEK deficient mice showing insulin secretory defects
3) Linkage evidence for *EIF2AK3* though in T1DM and preliminary evidence of association of the gene with T2DM in South Indian and Warren 2 trios
4) Biological significance of this protein kinase in regulating ER stress

It could therefore be hypothesised that in some subjects with type 2 diabetes, dysregulation of the *EIF2AK3* gene may contribute to beta cell apoptosis and cause relative insulin deficiency leading to diabetes.

Therefore the hypothesis that polymorphic variants in the *EIF2AK3* gene could be associated with T2DM in different population groups is the basis of our investigation.

Aims of the study

1) Compiling a SNP inventory by sequencing and reference to Hapmap projects
2) To determine whether common polymorphisms in the *EIF2AK3* gene (Candidate gene association study) could be associated with type 2 diabetes in the British/Irish, the Bangladeshi and the South Indian cohorts comprising of 8,095 individuals
3) Replication study in SNPs of *EIF2AK3* gene showing positive association with type 2 diabetes.

CHAPTER TWO: SUBJECTS

The study included genotype analysis of subjects from the British/Irish and the South Asian populations for the association of polymorphisms of the *EIF2AK3* gene and type 2 diabetes. For the purpose of case- control study we studied diabetes cases from the British Irish population- Warren 2 cases, W2 Sib pair probands and population based controls from the 1958 MRC birth cohort, ECACC samples. Furthermore, we also conducted a case-control study in the South Asian population- the Bangladeshi and the South Indian. For the family based association studies, we studied the British Irish resources- Warren 2 trios and the South Indian families. **(Table 2.a)**

The common Single nucleotide polymorphisms and the tagged SNPs were genotyped (11 SNPs) in 2,835 subjects with type 2 diabetes and in 3,538 control subjects and 522 families. Finally, we utilised the Dundee replication resources for a case-control study for the analysis of one marker (rs7605713) after the initial studies (n=6790).

2.1 SUBJECT DEFINITIONS

Controls - A person aged 25 years and above with no history of type 2 diabetes with a fasting blood glucose of less than 7 mmol/l and random blood glucose of less than 7.8 mmol/l and no history of diabetes and diabetes related symptoms.

Cases/Type 2 DM- A person aged 25 years and above with a history of type 2 diabetes and is on diet control, oral hypoglycaemic medication or on insulin treatment (Not started on insulin for at least 1 year after diagnosis)/ fasting glucose above 7 mmol/l and/or random glucose above 11 mmol/l or contemporary laboratory evidence of hyperglycemia satisfying the diagnosis of T2D at the time of recruitment (as defined by World Health Organization).

Table 2.a: Subjects for type 2 diabetes association study (n=14,885)

	Cases	**Controls**	**Family (Trios)**
British/Irish subjects (n= 12,902)			
UK Warren 2 n=6112	**2206** (W2 Cases, OCDEM_OXGN= 1713 + Sib pair probands= 493)	**2652** (MRC Birth cohort= 2088 + ECACC= 564)	**418 trios** (1254 subjects)
Dundee Replication resources n=6790	**3334**	**3456**	
South Asian subjects (n=1983)			
Bangladeshi n=985	**426** Young n=330 T2DM n=96	**559** Young n=374 Non DM n=155 IGT n=30 Total= 155+30	
South Indian n=998	**203** Urban Survey n=85 Newham n=118	**327** Urban Survey n=215, (IGT n=45) 215+45 Newham n=67	**104 families** (468 subjects)

2.2 INDIVIDUAL ETHNIC GROUPS

We studied the Caucasian British Irish population, and the South Asian populations from South India and Bangladesh for the type 2 diabetes candidate gene association study. The purpose of studying the *EIF2AK3* gene common variants in a candidate gene association study in different ethnic groups is due to the realisation differences in genetic architecture might exist among the different populations- i.e. minor allele frequencies might be different in different ethnic populations. Moreover, there could also be population differences in gene variants associated with the disease, with SNP variations showing positive associations in one ethnic group which cannot be replicated when studied in another ethnic group and vice-versa. We had access to genetic resources for the different ethnic groups, and were closely working with collaborators from the Warren 2 group in UK, Chennai diabetes research group in India and Bangladeshi diabetes research group, for conducting a comprehensive candidate gene association study. The pros and cons of choosing subjects from different populations, especially in relation to power of the study will be addressed later in the thesis.

2.2.1 British Irish resources

The case-control resources: This group includes the 1713 W2 cases and 493 W2 sib-pair probands totalling 2206 T2DM cases. As a control resource we have utilised the ECACC (n= 564) and MRC 1958 birth cohort (n=2088) samples totalling 2652 controls.

Warren 2 Cases

All type 2 diabetic subjects were unrelated U.K. Caucasians who had diabetes defined by either World Health Organization criteria at the time of diagnosis or being treated with medication for diabetes. Clinical criteria and/or genetic testing excluded known subtypes such as maturity-onset diabetes of the young or mitochondrial-inherited diabetes and deafness. The type 2 diabetic case group was a collection of type 2 diabetic subjects from the Warren 2 repository with an age of diagnosis between 35 and 65 years but not selected for having a family history. The cases tested positive for GAD autoantibodies had been excluded.

Warren 2 Sib Pair probands

The Warren 2 Sib Pair Repository currently comprises sibship pedigrees ascertained through six U.K. research centers (two in London and one each in Exeter, Oxford, Cambridge/Norwich, and Newcastle), according to a unified ascertainment protocol. All ascertained families include, at minimum, a sib pair with T2DM, together with parents and additional siblings when available. Validation of the diagnosis of diabetes in the index sib pair was based on either current prescribed treatment with sulphonylureas, biguanides, and/or insulin or, in the case of individuals treated with diet alone, historical or contemporary laboratory evidence of hyperglycemia (as defined by World Health Organization guidelines in place at the time of recruitment). Age at diagnosis of both members of the index sib pair, was initially restricted to the age range 35−75 years and subsequently was narrowed to 35−70 years, with 97.6% families meeting the latter criterion. Other forms of diabetes (e.g., maturity-onset diabetes of the young, mitochondrial diabetes, and type 1 diabetes) were excluded by standard clinical criteria based personal and family history, including an absence of first-degree relatives with type 1 diabetes and an interval of ⩾1 year between diagnosis and institution of regular insulin therapy. In addition evidence for autoimmunity to islet antigens was sought by measurement of titres of antibodies to glutamic acid decarboxylase (anti-GAD). All sibships were of European descent, with all four grandparents having exclusively British and/or Irish origin, both by self-reported ethnicity and by place of birth. Finally, pedigrees either reporting bilineal inheritance (both parents diabetic) or having a high proportion of affected individuals within large sibships were excluded from collection.

Controls

ECACC (European collection of cell cultures)

These are nationally recruited random UK population control samples from the European Collection of Cell Cultures (ECACC) without a diagnosis of type 2 diabetes, providing a control population totalling 564 samples. Trait information is not available for these individuals from the ECACC collection. (Website: www.hpacultures.org.uk)

1958 MRC birth cohort Controls

The 1958 British birth cohort study includes all children born in Scotland, Wales, and England between 3 and 9 March 1958. A proportion of these samples were released by MRC, Bristol, for Genetic studies and provided a control population totalling 2088. These were control individuals without a diagnosis of type 2 diabetes. However trait information is not available for this population. (Website: www.mrc.ac.uk)

The British/Irish resources for the family studies

The Warren T2DM trios collection for family based association studies

The families consist of a proband with T2DM (who are GAD negative) and both parents (n=418 families, 1254 subjects). The DNA and clinical characteristics were obtained by establishing a nationwide collection run from 6 UK centres. Recruitment of patients was principally from hospital diabetes clinics, general practitioners, and the questionnaire survey. Subjects were considered suitable for the trios collection if the proband had type 2 diabetes and had two living parents and four grandparents of European descent (>98% were of UK/ Irish origin). To avoid inadvertently recruiting subjects with type 1 diabetes, recruitment was limited to trios for whom the proband was diagnosed after the age of 25 years, had been treated without insulin for at least 1 year after diagnosis, and did not have a family history of type 1 diabetes. Potential maturity-onset diabetes of the young (MODY) families were excluded if there were two generations with an autosomal dominant history of diabetes and at least one family member diagnosed before the age of 25 years. Maternally inherited diabetes and deafness was excluded if there was a maternal transmission of diabetes and deafness.

2.2.2 South Asian Resources

The Case-control resources: The 2 main resources of subjects are from Bangladesh and South India.

Bangladeshi subjects for a case-control study

The first group (n = 330) consisted of Bangladeshi subjects presenting with diabetes before the age of 30 years (subsequently referred to as "under-30 diabetes" [mean age at onset 18.7 ± 6.2 years; mean BMI 18.3 ± 5.1 kg/m^2]) at the Bangladesh Institute of Research and Rehabilitation in Diabetes, Endocrine and Metabolic Disorders (BIRDEM) clinic. The leaner, under 30 age group subjects likely to be the more insulin dependent group closely matches the hypothesis we were testing of gene variants of *EIF2AK3* contributing to pancreatic insufficiency and T2DM. The controls were recruited from Bangladesh totaling 374 subjects with a mean age of 23 ± 5 years at recruitment and mean BMI 19.9 ± 3.3 kg/$m^{2.}$

The second cohort we studied were Bangladeshi subjects from Sylhet in Bangladesh, and those ascertained either from a diabetes clinic at the Royal London Hospital, London- diabetes subjects (n=96), non diabetic subjects (n=155), impaired glucose tolerance subjects (n=30). The subjects with T2D had a mean age at onset of diabetes of 44.7 ± 10.0 years and a mean BMI 26.5 ± 3.4 kg/m^2. The ethnically matched control individuals had a mean age of 41.5 ± 10.4 years and a mean BMI 26.9 ± 9.7 kg/m^2.

Subjects from Sri Lankan Tamil and South Indian Dravidian population for a case-control study

The subjects (cases n= 118 and controls n= 67) were recruited from East London (Ethical committee approval from East London health authority) and were Sri Lankan Tamil and South Indian Dravidian population. They were identified through the hospital diabetes clinics and the local community centres. The inclusion criteria for cases included those who had type 2 diabetes for more than 1 year, aged above 25 and 70 years old, not started on insulin for at least 1 year after diagnosis. The diagnosis was based on the WHO classification of diabetes (as of 1997 classification) and information of their glycaemic

status was obtained from the diabetes database (Diamond). The inclusion criteria for controls included those who were aged between 25 and 70 years of age, family members (Parents and 1st degree relatives) were not taking part in the study, had no complaints of osmotic symptoms and had random blood glucose in the normal range (< 7.8 mmol/l). The subjects with T2D had a mean age of 59.8 ± 11 years and a mean BMI 27 ± 3.8 kg/m^2. The ethnically matched control individuals had a mean age of 50 ± 14 years and a mean BMI 26.8 ± 4.5 kg/m^2.

South Indian cross-sectional urban survey

DNA samples were available from 345 South Indian subjects initially recruited as part of an urban population-based survey of the prevalence of type 2 diabetes and associated risk factors, using a cluster analysis design across all socio-economic groups. Of these subjects, 48.6% were female, with mean age of 42 years (range 34–53), a mean BMI of 23.1 ± 4.6 kg/m^2. The other 51.4% were male, with a mean age of 47 years (40–55), a mean BMI of 22.0 ± 3.7 kg/m^2. In the urban survey, glucose tolerance was defined by the most recent WHO criteria (124). Male type 2 diabetic subjects were recruited at a mean had a mean BMI of 23.4 ± 3.3 kg m^2; the females had a mean BMI of 24.1 ± 4.1. In the total study population, 215 subjects had normal glucose tolerance, 45 subjects had impaired glucose tolerance, and 85 had type 2 diabetes.

South Indian families

A total of 104 South Indian families (468 subjects) were recruited from a diabetic clinic in Chennai, India. Ascertainment was via an offspring with type 2 diabetes, as defined by World Health Organization (WHO) criteria. Of the probands, 63.5% were male, with a mean age of onset for type 2 diabetes of 34 years (range 29–38), a mean BMI of 26.5 ± 4.5 kg/m^2. Clinical details for female probands were: a mean age of onset of type 2 diabetes of 32 years (range 23–37), a mean BMI of 27.8 ± 4.6 kg/m^2. Maturity onset diabetes of the young was excluded in families if the proband had an age of onset of diabetes before 25 years and an autosomal-dominant history of diabetes in two generations. Type 1 diabetes was excluded on clinical grounds. None of the diabetic

subjects in the family had a history of ketoacidosis or ketonuria, an acute onset of symptoms, or weight loss before diagnosis, nor was there a family history of type 1 diabetes. Furthermore, if the proband was insulin treated, then this did not commence within a year of diagnosis.

2.2.3 Dundee Wellcome Trust UK T2DM Genetic consortium Case-control Collection for replication study

The cases (n=3334) were European White descent and validation of the diagnosis of type 2 diabetes was based on either current prescribed treatment with sulphonylureas, biguanides, and/or insulin or, in the case of individuals treated with diet alone, historical or contemporary laboratory evidence of hyperglycemia. (As defined by World Health Organization) No clinical or laboratory diagnosis of other forms of diabetes (e.g., maturity-onset diabetes of the young, mitochondrial diabetes, and type 1 diabetes). At least 1 year between diagnosis and institution of regular insulin therapy. Antibodies to glutamic acid decarboxylase (anti-GAD) were measured and were negative. The age range of individuals was 35-70 years inclusive. The controls (n=3456) were of European White descent not diagnosed with diabetes and not likely to be diabetic on biochemical testing of fasting glucose < 7mmol/l, HbA1c< 6.2%, aged 35-80 years.

CHAPTER THREE: MATERIALS AND METHODS

3.1 Collection of blood samples and DNA preparation

I was involved in the collection of blood and DNA preparation for the Newham South Indian/ Sri Lanka Tamil case-control resource (T2DM subjects n = 118 and Non DM n = 67) recruited from the boroughs of Newham and Tower Hamlets in East London and a small proportion of Bangladeshi Whitechapel case-control resource. The samples were collected after appropriate ethics approval by the East London health authority.

Majority of the DNA samples used for our study were previously collected and DNA was plated in a 96 well format. The UK Warren 2 British Irish resources were collected as part of a combined effort through six UK centers (two in London and one each in Exeter, Oxford, Cambridge/Norwich and Newcastle). We obtained the samples from Prof. Hattersley's laboratory in Exeter and stored the DNA in our lab at - 20° Celsius. KBioSciences were also in receipt of Warren 2 resources which were utilized for our experiments. The control samples ECACC were obtained from Prof. Hattersley's laboratory in Exeter. 1958 birth cohort samples were provided by Prof. David Strachan at St.George's Hospital medical school, University of London after appropriate approval.

The rest of the South Asian resources- South Indian and Bangladeshi samples were previously collected in those countries by our collaborators and were already available in our laboratory and they were stored at - 20° Celsius in the lab. The South Asian resources have been utilized in previous genetic association studies conducted by the researchers in our unit.

The Dundee replication set (Wellcome Trust UK Type 2 Diabetes Genetics Consortium Case- control samples collected in Tayside region of Scotland by Dr. Palmer's group and were provided to KBioSciences) was utilized for the replication study.

Identification of suitable candidates for South Asian resource

The Diabetes cases for the Newham and Tower Hamlets East London resources were identified from the local Diabetes Register, Diamond by me. The controls were included from the same population by seeking the help of the local Tamil community forums and district nurses. The Bangladeshi Whitechapel case-control resource was identified from the diabetes centre at Royal London hospital with the help of Prof. Hitman.

We collected demographic data and data for weight, height, BMI, waist hip ratio, blood pressure for each individual. And blood samples were collected in 2 X 10 ml EDTA tube for DNA extraction, 1 X 7 ml clotted serum for Cholesterol estimation, 1 X 4 ml sodium fluoride tube for glucose estimation, were taken from each individual. The blood samples were stored at - 20° Celsius in the lab prior to DNA extraction. I did majority of this work and created a database for the South Asian resource.

DNA preparation

Protocol: DNA Purification from Blood (blood samples collected in EDTA tube) (Spin Protocol) (QIAGEN DNA extraction protocol) (Website: www.qiagen.com)

The method used for DNA extraction for the South Asian resource was spin protocol for purification of total genomic DNA from whole blood using a microcentrifuge. First step involves pipetting 20 µl QIAGEN Protease into a 1.5 ml microcentrifuge tube. Add 200 µl blood sample and 200 µl Buffer AL to the microcentrifuge tube. Mix by pulse-vortexing for 15 seconds and incubate the solution for 10 minutes at 56°C. Centrifuge the tube to remove drops from the inside of the lid. Then add 200 µl of ethanol (96–100%) to the sample, and vortex and centrifuge the tube to remove drops from the inside of the lid.

Take Qiamp spin column and transfer the whole mixture without wetting the rim. Centrifuge at 8000 rpm for 1 minute. Discard the tube containing the filtrate and keep spin column in a 2 ml collection tube. Add 500 µl Buffer AW1 to the QIAamp Mini spin

column without wetting the rim and centrifuge at 8000 rpm for 1 minute. Then add 500 µl Buffer AW2 to the QIAamp Mini spin column without wetting the rim and centrifuge at full speed (14,000 rpm) for 3 minutes.

Place the QIAamp Mini spin column in a new 2 ml collection tube and discard the old collection tube with the filtrate. Centrifuge at full speed for 1 min. Then place the QIAamp Mini spin column in a clean 1.5 ml microcentrifuge tube, and discard the collection tube containing the filtrate. Add 200 µl Buffer AE to the QIAamp Mini spin column. Incubate at room temperature (15–25°C) for 10- 30 minutes, and then centrifuge at 8000 rpm for 1 minute. Run the DNA (5 µl) on a gel with 2 µl of loading buffer.
A 200 µl sample of whole blood typically yields 6 µg of DNA in 200 µl water (30 ng/µl).

3.2 DNA SEQUENCING

Nucleotide sequencing reactions

Samples were sequenced on the ABI 3100 capillary sequencer. The direct automated sequencing of the double stranded PCR generated templates for this system was performed using cycle sequencing with fluorescent ddNTPs and electrophoresis *(reference- Applied Biosystems Sequencing protocol)* (125). ABI PRISM TM Dye terminator Cycle Sequencing Ready Reaction Kit containing the ABI Big Dye version 3 sequencing kit (Perkin Elmer) for the ABI 3100 sequencer was used.

ABI Dye-Deoxy Terminator Cycle Sequencing

ABI 3100 reaction

Fluorescence-labelled cycle sequencing reactions were performed in 10µl volumes using the ABI Big Dye version 3 sequencing kit (Perkin Elmer). Each reaction contained 100ng of PCR products template DNA, 8µl of Terminator Ready Reaction Mix (containing AmpliTaq DNA polymerase, ddNTPs, magnesium, fluorescent dichlirorhodamine dye Terminators, Tris-HCl) (PE Biosystems) and 3.2pmol primer. The mixture was overlaid

with one drop of paraffin oil, and the reaction was performed in a PTC-100 Programmable Thermal Controller (MJ Research, Inc.). Reaction profiles included 25 cycles of 30 seconds denaturation at 96°C for 15 seconds, annealing at 50°C for 1 minute, and 4 minutes of extension at 60°C. After PCR, extension products were purified from unincorporated dye-labelled terminators by ethanol/sodium acetate precipitation. All 20µl of the reaction contents were transferred to a 1.5ml Eppendorf tube that contained 2µl of 3M sodium acetate (pH 4.6) and 50µl of 95% ethanol. The tube was vortexed and kept on ice for 30 minutes. Then the tubes were placed in a centrifuge and spun at 13000 rpm for 30 minutes. Supernatant was carefully removed and the pellet rinsed with 125µl of 70% ethanol. After re-centrifugation at 13000 rpm for 5 minutes, the supernatant was again removed and the tubes left on bench uncapped for 10-15 minutes for DNA pellet to air dry. The dried DNA sample was resuspended by adding 5µl of auto-sequencing gel loading buffer and completely dissolved by vortexing. Then, the DNA sample was denatured at 95°C for 5 minutes, and placed on ice until all of the sample solution was loaded in a sequencing gel

Automated ABI 3100 analysis

The calibration was done according to manufacturer instructions and all functions were controlled by PC. The samples were electrophoresed through a 16 x 47 cm sequencing capillaries using a POP6 polymer (PE Biosystems) with recommended electrophoresis buffer (PE Biosystems). Raw data was converted to Macintosh sequence chromatograms and analysed as described in following section.

Sequence Analysis using Computer Software

Following visual inspection of the analysed digitised raw sequence data and the attainment of correct and optimal base spacing (between 9-12 scans), sample files were imported into the Sequence Navigator Software (ABI) version. The wild type sequence corresponding to the target template was retrieved from Genbank (GDB), imported into the Sequence Navigator programme and designated as a reference sequence. The multiple alignment comparisons using the Cluster algorithm were performed between the reference sequence and imported sequences. Forward and reverse (reversed and

complementary) sequences were included for all template comparisons. The mismatch option of the programme identified mismatched/mis-aligned sequence compared with the reference sequence within the sequence text file. Mismatched regions were examined on the electrophoregrams to establish whether they were genuine mutations or sequence artefacts. All potential mutations had to be confirmed by being present in both forward and reverse sequence data for each sample. Potential mutations/polymorphisms in individual samples were further confirmed by repeating the sample PCR-sequencing reactions. This was necessary to eliminate false positives due to random non-template incorporation of dNTPs by *Taq* polymerase during the PCR reaction.

EIF2AK3 **5' AND 3' SEQUENCING**

To detect all possible disease associated variants at a given gene it would be necessary to examine, in large samples, every base at which variation might alter gene function or expression. Only then could we be confident that an association had not been missed just because the wrong markers had been typed. This is difficult to achieve, in terms of costs and procedural constraints, if not impossible given the technology available at the present time. However this was definitely not the case in 2002 when we were working on the *EIF2AK3* project. We therefore required sequencing data for the candidate gene- *EIF2AK3* to identify SNPs to use in a case-control study thereby having a substantial effect on the power and quality of the study.

EIF2AK3 gene together with 1800bp 5' prime and 3000bp 3' prime were sequenced on 48 individuals from different ethnic backgrounds (Table 3.a). The 48 individuals included all the 3 ethnic groups and both T2D cases and controls. At the time the study was designed in the year 2002, sequencing of 24- 48 individuals was deemed more than adequate to gather adequate information on SNP variants. The sequencing work would enable us to identify common polymorphisms with minor allele frequencies of around 30% in the gene of interest, in our study the sequencing of *EIF2AK3* gene. This was in keeping with our plans to study common disease common polymorphism association. If we were to

hypothesise common disease rare polymorphisms we would have needed to sequence a lot more individuals to capture the rare alleles. However, we were not pursuing rare polymorphisms to be associated with T2D.

Table 3.a: Selection of samples for sequencing (n= 48)

Subjects	**Number of individuals**
WRS family,Chennai (WRS proband, affected sibling, 4 non affected)	6
Caucasian controls	5
British/Irish T2DM (Warren 2 samples)	12
SI T2DM (Newham samples)	8
SI T2DM (Indian)	9
Bangladeshi controls	8

Direct sequencing of all 17 exons/coding region and intron/exon boundaries was performed using the Big Dye Terminator chemistry on the ABI 3100 sequencer. Single nucleotide polymorphisms, already in the public databases and novel SNPs, were identified in this preliminary group. Further investigations of the SNPs in a larger group revealed which of them were truly polymorphic and the ones that were monomorphic.

3.3 TAQMAN HIGH THROUGHPUT GENOTYPING

Taqman is a powerful high throughput assay, which can be used for genotyping large number of samples in a relatively short period of time *(reference- Applied Biosystems real time PCR protocol)*. The technology uses an ABI Prism 7900 instrument (Taqman) to detect accumulation of PCR products continuously during the PCR process, real time quantitative PCR. The polymerase chain reaction (PCR) is a common technique used to rapidly amplify several copies of a specific region of DNA exponentially (126). It requires a thermally stable DNA Taq polymerase, each of the four nucleotide dNTPs in equimolar concentration, and a source of template DNA, usually a double stranded genomic DNA but may also be from a cDNA. The technique also requires synthetic oligonucleotide primers, designed to complement DNA sequences flanking the region of interest. The taqman technology is based on the polymerase chain reaction to amplify and simultaneously quantify a targeted DNA sequence- real time PCR (127).

The system integrates 4 major elements:

1. Fluorogenic chemistry for target –specific oligonucleotide probes.
2. Exploitation of the polymerisation dependant 5' nuclease activity of the DNA polymerase.
3. Instrumentation to measure fluorescence signal within a closed PCR reaction tube.
4. Software to process and analyse data.

Steps in real time quantitative PCR

A taqman probe is designed to anneal at the target sequence between the traditional forward and reverse primers **(Figure 3.a)**. Two fluorescent dyes, a reporter (R) and a quencher (Q), are attached to the probes used with the TaqMan PCR Reagent Kit. The probe is labelled at the 5' end with a reporter fluorochrome and a quencher fluorochrome at the 3' end. The 3' end of the probe is blocked, so it is not extended during the PCR reaction.

Figure 3.a: Taqman SNP genotyping

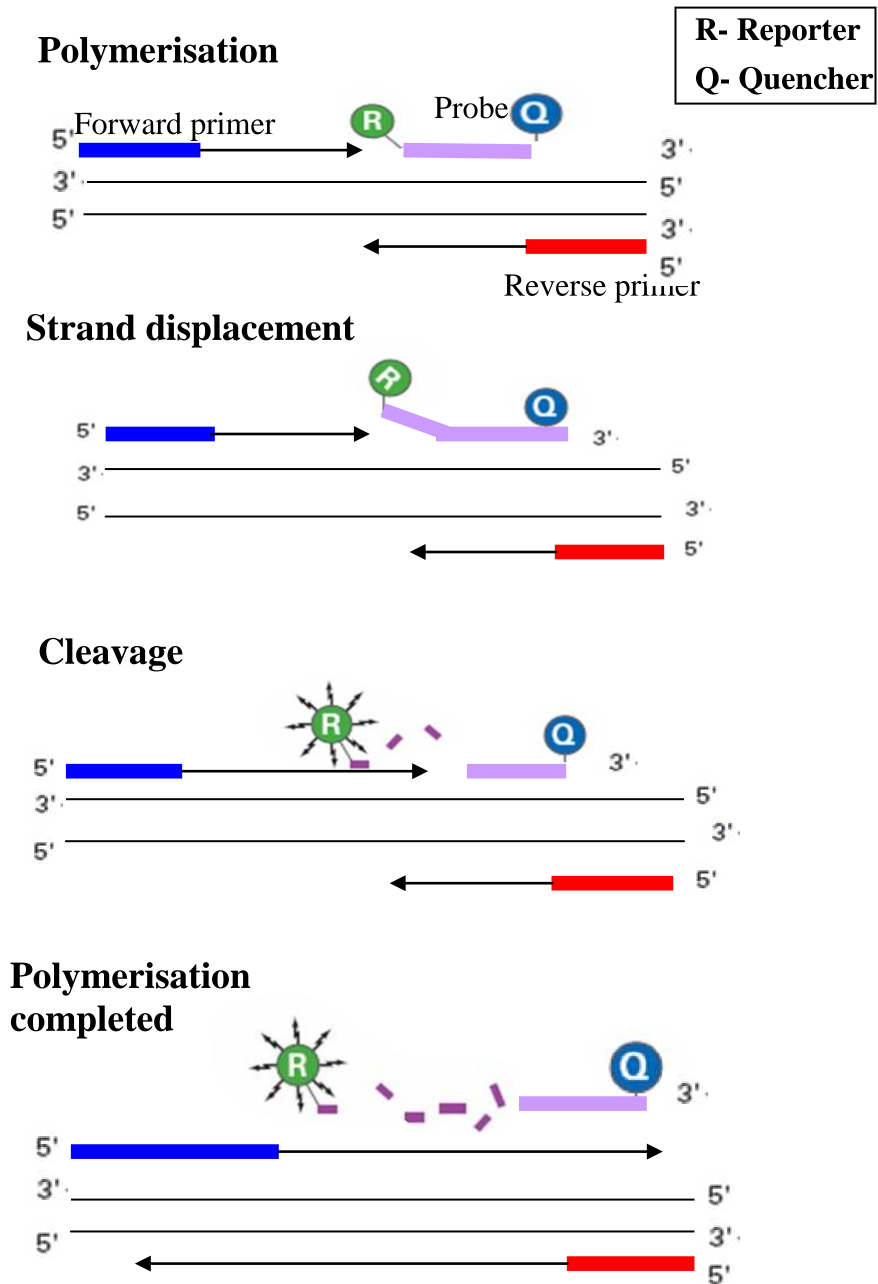


Figure 3.a: Taqman probe with its attached reporter and quencher fluorescent probes anneals between forward and reverse primer onto the target sequence. TaqDNA polymerase cleaves reporter dye from the probe, whereby emitting characteristic fluorescence. Depending on the characteristic fluorescent signal the common allele, rare allele or a heterozygote are identified.

When both dyes are attached to the probe, reporter dye emission is quenched due to fluorescence energy transfer from the reporter dye to the quencher dye. During each extension cycle, the probe is displaced at the 5' end by the DNA polymerase. Taq DNA polymerase then cleaves the reporter dye from the probe via its 5'-3' exonuclease activity. Once separated from the quencher, the reporter dye emits its characteristic fluorescence, which can then be measured by the 7900. The amount of fluorescence measured is proportional to the amount of PCR product made.

The design of TaqMan probes and primers is critical to the success of the experiment. The target sequences with the necessary SNP information were sent to Applied Biosystems (Commercial enterprise) and the Taqman primers and probes were designed by Applied Biosystems. (Website: www.appliedbiosystems.com). Optimisation of PCR reaction was performed at the Genome centre.

WRS Taqman primers and probes

- **EXON 2 SNP C/G rs867529**

GTAAAAAGCAGTGGGATTTGGATGT **[C/G]** CATCCAGCCTTAGCAAACCAGAG

- **Intron 10 SNP A/T rs6750998**

CCAATGACAGTAGCTGGAATGACA [**A/T**] caaccatctgaatcaaagaagaaatgac

- **Exon 11 SNP A/G rs1805164**

CCACAAAATTTAAAACTGTTGCAAACCTTTAT **[A/G]** GACGTGGTGGCTTTGGAGT

- **Exon 13 SNP T/G rs1805165**

GCACCATCAGTTAAAATACGCAGAA **[T/G]** AATCATAGCTCCTTCACCACAAAGA

In my work Taqman chemistry was used for allelic discrimination of SNPs 2, 10, 11 and 13 of the *EIF2AK3* gene. The genotyping was performed on various data sets using the ABI 7900 taqman at the Genome centre, Queen Mary University of London. Optimisation of the PCR reaction is required for each primer and probe set. The taqman reaction requires two primers for amplification of the sequence of interest (Primer information given below) and the two allele specific probes labelled with the reporter dyes VIC and FAM at the 5' end and a non fluorescent quencher bound to the 3' end of allele specific probe. VIC label detects the common or the wild type allele and FAM detects the rare or the mutant allele.

Table 3.b: PCR reaction mix for Taqman

DNA	2 ul
Taqman Universal master mix 2 X	2.5 ul
40 X primers & probes	0.125 ul
Water	0.375 ul
Total Volume of reaction	5 ul

Table 3.c: Temperature settings for PCR

Step 1	50° C	2 minutes
Step 2	95° C	10 minutes
Step 3	95° C	15 seconds
Step 4	60° C	60 seconds
Step 5	Repeat step 3-4 for 40 cycles	

The primers, taqman mix and water were mixed (PCR reaction mix) in a 1.5 ml eppendorf **(Table 3.b)** and the eppendorf is vortexed to mix the components and centrifuged. 2 µl DNA was pipetted (robotic) into each well of an ABI PRISM™ 384-Well Clear Optical Reaction Plate. Subsequently 3 µl of PCR reaction mix was added in each well. PCR amplification was performed on DNA Engine Tetrad (MJ Research) as per the temperature settings in **Table 3.c**.

Steps in PCR- Thermocycling

1. DNA denaturation by heating - The first step involves heat denaturation at 95° C, of DNA template and separation of the 2 strands of DNA
2. Annealing of primers to DNA strands-The forward and reverse primer anneal to the DNA strands and involves primers binding to their complementary sequences in the DNA template and the annealing temperature is at 95° C.
3. Extension of DNA strands in both 5' and 3' directions- The DNA strands extend in both directions and as the PCR process continues the extension is limited to the sequence of interest between the 2 primers.
4. Amplification of DNA segments by manifold- the cycle is repeated up to 40 times and this result in a manifold increase in the target sequence.

Optimisation of the PCR reaction is required for each primer and probe set. This was performed with the temperature settings for PCR reactions as mentioned in Table 3.c. The temperature settings were set constant for the denaturation of the DNA, annealing of the primers and the extension of DNA of the PCR process for all the experiments. The amplification step was repeated for 40 cycles in one set of experiments and another set of experiments with 60 cycles. Both the settings generated consistently good results. Therefore, the majority of taqman assays were performed with the settings as mentioned above with the amplification cycle repeated for 40 cycles. After PCR amplification an endpoint plate read was performed using SDS software on ABI Prism 7900 instrument (Taqman), which calculates the fluorescence measurements made during the plate read and plots fluorescent values based on the signals from each well. Thus, the different alleles (homozygous wild allele, heterozygote, homozygous rare allele) can be determined.

3.4 AMPLIFLUOR HIGH THROUGHPUT GENOTYPING

The Amplifluor SNPs Genotyping System is based on competitive allele-specific PCR, allowing the simultaneous amplification and detection of DNA within a closed reaction vessel (128). This homogenous assay utilizes two fluorescently labelled Amplifluor SNP Primers and three unlabelled standard oligonucleotides – two Z-tailed, allele-specific forward primers and a common reverse primer – in a one-step, single-tube reaction with standard PCR reaction components (Reference- Flowgen Bioscience Amplifluor protocol*)*.

First, the two Z - tailed, unlabelled, allele-specific primers with the common reverse primer initiate a competitive allele-specific PCR reaction. During early rounds of PCR, the allele-specific primer with its unique "Z" tail is incorporated into the amplicon. The corresponding Amplifluor SNP Primer recognizes the complement of the "Z" tail sequence and is able to prime off of that sequence. Incorporation of the Amplifluor Primer into an allele-specific amplicon melts its hairpin structure, thus separating the fluorophore from the quencher and generating a fluorescent signal. Depending on which base is present in the target SNP, either a green or red signal is generated. Heterozygotes produce a yellow signal combined for both fluorophores. The fluorescent signal can then be measured on a real-time PCR instrument or fluorescent plate reader. Technology similar to Amplifluor is KASpar (KBiosciences allele specific PCR) technology.

Figure 3.b: Amplifluor SNP Genotyping

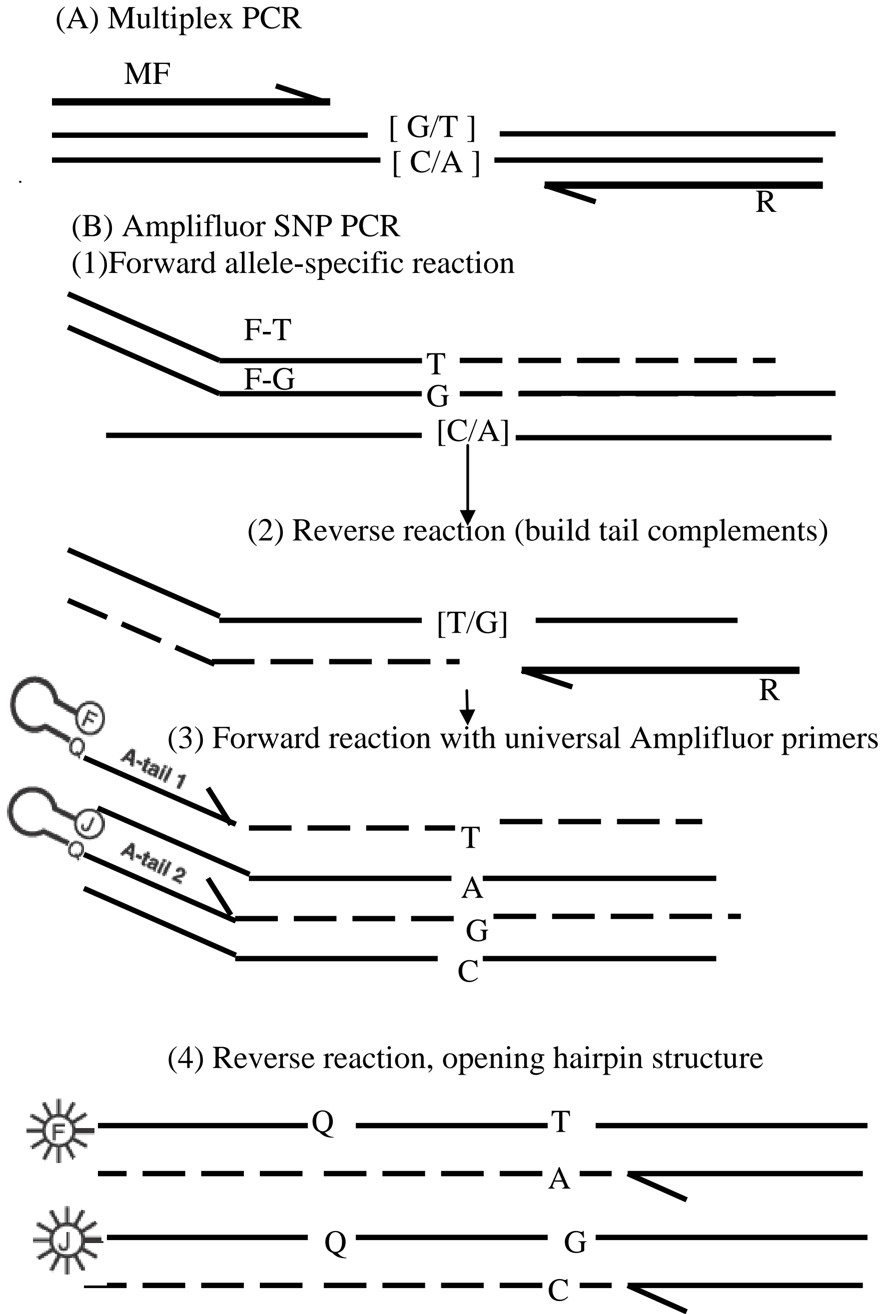


Figure 3.b: Amplifluor SNP genotyping process for a G/T SNP is being shown. Two fluorescently labelled Amplifluor SNP Primers and three unlabelled standard oligonucleotides – two Z-tailed, allele-specific forward primers and a common reverse primer – in a one-step, single-tube reaction with standard PCR reaction components. Primers are: multiplex forward (MF), common reverse (R), allele-specific forward primers for the T allele (F-T) and G allele (F–G), and Amplifluor universal primers. The tail sequences on the 5' ends of the allele-specific forward primers and at the 3' end of the Amplifluor universal primers are labelled A-tail 1 and A-tail 2. Dabsyl quencher is represented by Q. Fluorophores FAM and JOE are represented by F and J. Incorporation of the Amplifluor Primer into an allele-specific amplicon melts its hairpin structure, thus separating the fluorophore from the quencher and generating a fluorescent signal.

In my work Amplifluor chemistry and KASpar (KBiosciences allele specific PCR) genotyping with underlying principle very similar to amplifluor, were used for allelic discrimination of all 11 SNPs of the *EIF2AK3* gene. Genotyping was outsourced to KBiosciences. The Amplifluor and KASpar technology was utilised for genotyping performed on all the data sets at KBiosciences.

3.5 GENOTYPE ERRORS – QUALITY CONTROL (QC)

Quality control checks of the genotyping forms an important step to generate reliable data prior to carrying out the statistical analysis for disease association. There are set rules to be followed for QC checks whilst doing candidate gene association studies(129). We performed a variety of QC checks in our study to this effect.

Steps for passing QC checks:

- Checks on genotyping quality should include careful visual inspection of genotype plots.
- The proportion of samples for which a SNP can be measured (the SNP call rate, typically >95%)
- Concordance rates in duplicate samples (typically >99.5%).
- The minor allele frequency often >1%, as rarer SNPs are difficult to measure reliably
- Severe violations of Hardy-Weinberg equilibrium (HWE) to be excluded

CHAPTER FOUR: STATISTICS

The statistical methodologies utilised in our study include Hardy-Weinberg equilibrium tests for quality control, UNPHASED-Output log- likelihood ratio test for the purpose of candidate gene association study in case- control study and family study, Likelihood ratio tests as implemented in LDPAIRS program for the linkage disequilibrium statistics. Power calculations were not performed prior to our candidate gene association study. However, prior power calculations would have been ideal to investigate the candidate gene in the three ethnic groups, thereby providing a robust study design. In our defence, when the study was envisaged in the year 2002, published candidate gene association studies included hundreds of cases and controls unlike the thousands of individuals we were studying(130). Furthermore, we did not have the minor allele frequencies for the South Asian population to enable us to do appropriate power calculations. Further detailed discussion of power of the study and the drawback to our study is discussed in the final chapter.

4.1 HARDY- WEINBERG EQUILIBRIUM TESTS

Hardy- Weinberg equilibrium tests were conducted for the purpose of quality control issues to detect genotyping errors in our candidate gene association study. Gene frequencies and genotype ratios in a randomly breeding population remain constant from generation to generation. This is the basis for the Hardy-Weinberg principle (131). The total number of genes in a population is its gene pool and allele/genotype frequencies remain constant through the generations if certain conditions are met. HWE is an important tool for understanding population structure and can be tested by mathematical formula(132). If p represents the frequency of one allele, for example a, and q the frequency of the other allele, for example b, then $p + q = 1$ in the total population.

- p^2 = the fraction of the population homozygous for a
- q^2 = the fraction homozygous for b
- $2pq$ = the fraction of heterozygotes

Then HW equilibrium states that $(p+q)^2 = p^2+2pq+q^2= 1$.

However, there are times when there is deviation from HW equilibrium in the studied population-

- Non-random mating in the population increases the chances of homozygosity in the genes. Selective mating leads to changes in gene frequencies.
- Gene flow-migration is whereby members of one population group may breed with immigrants from another population group. This has the propensity to introduce new genes or alter existing gene frequencies.
- New mutations in the genes within the population.
- Genetic drift is whereby a random change in the genotype frequencies occurs in a small population size. There could be chance elimination and changes in allele frequency due to a sampling effect in a small population.

HWE testing was carried out by the χ^2 test using an excel spreadsheet supplied by Prof. Curtis. Any significant deviation from HWE was taken into account as a data quality check and those samples which were not in HWE were excluded from the final statistical analysis.

4.2 GENOTYPING ANALYSIS- UNPHASED-*Output* FOR CASE- CONTROL STUDIES AND FAMILY STUDIES

The statistical methodology used in our study was based on the principle of calculating log likelihood ratios as implemented in the program UNPHASED (133). The overall test of association is a likelihood ratio test comparing the log likelihoods for the null hypothesis that a marker or allele is not associated with the disease and alternative hypotheses that the marker or allele is associated with the disease. We tested whether case/ control status is associated with disease i.e. is there a difference in the distribution of alleles between cases and controls.

The log-likelihoods for the null and alternative hypotheses are displayed in the analysis section of UNPHASED- Output. The results generated show the relevant information

pertaining to the **Allele**, the estimated count of the alleles in cases, the estimated count of the alleles in controls. **Case frequency** is the marginal frequency of the alleles in cases. **Control frequency** is the marginal frequency of the alleles in controls. Odds ratio are generated for the allele and diseases association and are shown relative to one allele, termed the *reference* allele.

The **likelihood ratio** statistic is minus twice the difference in log likelihoods, and is asymptotically distributed as χ^2 with degrees of freedom (**df**) equal to the difference in number of free parameters between the two hypotheses. The **p-value** is the probability of observing a likelihood ratio statistic at least as large as this one, if the null hypothesis were true.

A candidate gene association study tests for significant differences in the – genotype/allele frequencies of the markers within a gene between cases and controls, the aim being to detect association with a trait of interest, which in our study is type 2 diabetes (134).

A brief discussion of alternative methodologies to log- likelihood ratio tests is given below. The classic case-control study design compares allele frequencies or genotypes in a sample of unrelated affected individuals for T2D and a sample of matched controls. This can be done by testing the null hypothesis of no association between rows and columns of the 2 X 3 table that contains the counts of the three genotypes (the wild type homozygote, rare type homozygote and the heterozygote) among cases and controls using a Pearson test (2 df) or a Fisher exact test. For complex traits like T2D, it is thought that contributions to disease risk from individual SNPs will be additive. That is, the heterozygote risk will be intermediate between the two homozygote risks. One way to improve power to detect additive risks is to count alleles (i.e. major allele and minor allele in a 2 X 2 table) rather than genotypes so that each individual contributes twice to a 2 X 2 table and a Pearson 1-df test can be applied.

The case-control design has frequently been criticized because of the potential for spurious associations due to population stratification. Population stratification implies the existence of genetically different groups in the population under study. It occurs, for example, when cases and controls are not well matched ethnically or when people in the population under study have not mated randomly for at least several generations. This problem can be overcome by careful selection of the subjects- by matching for ethnicity, age and a random selection from the population for control subjects. However, in a complex disease like T2DM selecting young cases of T2DM and having a group of 'hyper normal controls' (i.e. older individuals who have not got T2DM) has the potential to detect gene variants associated with the disease.

One of the approaches to overcome the problem of population stratification is doing a family based association study utilising the transmission disequilibrium test (TDT) (135). The rationale behind the TDT is that, in the absence of both linkage and association between marker and disease loci, marker alleles will be transmitted randomly from parents to offspring. The TDT compares the transmission vs. nontransmission of marker alleles to affected offspring by means of a simple chi-square statistic. The TDT is widely known and used as a test for association in the presence of linkage, that is, as a tool for fine-mapping of disease genes. The undisputed strength of the TDT is that it eliminates population stratification effects completely. The original formulation of the TDT was based on a biallelic, one-marker model. Extension for multiallelic markers is done by extended TDT (ETDT), based on a logistic regression procedure (136). All these TDT-like tests share the method of comparing the transmission and nontransmission of marker alleles to affected offspring. While family-based association study designs are generally more robust to population stratification, they are less powerful than case-control designs. The numbers required for a family based association study, say for example 1000 trios will have the same power as 1000 case vs. 1000 control subjects but this would mean that there are nearly 3000 samples to be genotyped in the trios.

The reasons why our statisticians Anna Vine and Prof. Curtis chose the log likelihood ratio test as in the program UNPHASED- Output for our study- 1) implements maximum-likelihood inference on genotype effects while allowing for missing data such as uncertain phase and missing genotypes. Data imputation consists of replacing missing genotypes with predicted values that are based on the observed genotypes at neighbouring SNPs. Data imputation can be reliable for tightly linked markers and it allows better use of the observed data. 2) The UNPHASED program can analyse both case-control studies and the family association study which is an added advantage.

4.3 LINKAGE DISEQUILIBRIUM (LD) STATISTICS

In order to measure marker-marker association we calculated a likelihood ratio test comparing the likelihood for the genotypes assuming LD to that assuming no LD, as implemented in the LDPAIRS program (137). The p-values measure the statistical significance of the test for linkage disequilibrium between the markers. LD plot was obtained using the program Haploview. (Website: www.broadinstitute.org/haploview)

GENECOUNTING is a program utilising the estimation- maximization algorithm to estimate haplotype frequencies in a group of unrelated subjects. Moreover the program allows the estimation of haplotype frequencies even if some subjects are not genotyped at some markers. GENECOUNTING inputs a set of multi- locus genotypes along with the count of how often each one occurs, and outputs the log likelihood and estimated haplotype frequencies under the null hypothesis of no linkage disequilibrium between markers, and under the alternative hypothesis of LD being present between all markers according to maximum likelihood haplotype frequencies. There are a range of programs which enhance the usability of GENECOUNTING. LDPAIRS is one such program enhancing the usability of GENECOUNTING. LDPAIRS facilitates obtaining measures of linkage disequilibrium (LD) between markers. The LDPAIRS program measures LD between all pairs of markers in the dataset and produces the results in a tabular format. The p-values measure the statistical significance of the test for linkage disequilibrium between the markers. The absolute value of D' as measured between the commonest allele at each of the two markers is also output.

CHAPTER FIVE: SEQUENCING RESULTS

Compiling a complete SNP inventory of the *EIF2AK3* gene

For the type 1 diabetes work we had previously sequenced the coding region and immediate flanking sequences. In this study the *EIF2AK3* gene together with 1800bp 5' and 3000bp 3' were sequenced in 48 subjects which included 31 T2DM cases and 17 controls **(Figure 5.a).**

Figure 5.a: Sequencing of the 5' and 3' regions of the *EIF2AK3* gene [in 48 selected individuals]

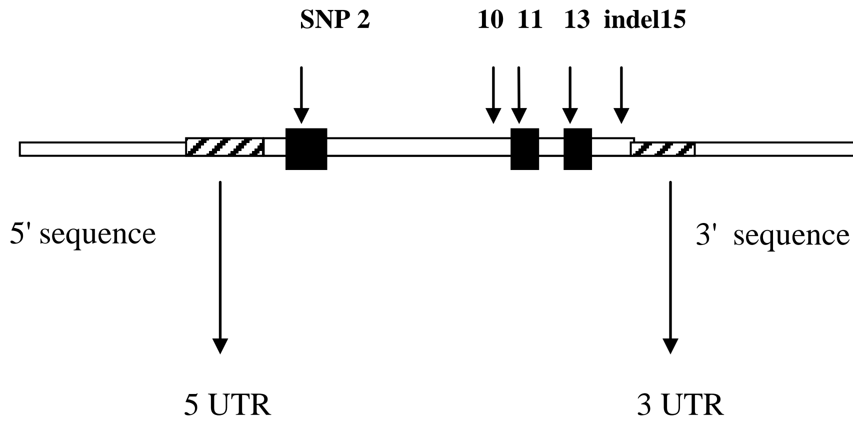


Figure 5.a: The *EIF2AK3* gene was sequenced together with 1800bp 5' upstream and 3000bp 3' downstream were sequenced. UTR refers to untranslated regions. A few SNPs are shown highlighting INDEL15 to be present near the 3' region of the gene. (SNP 2 - rs867529, SNP 10- rs6750998, SNP 11- rs1805164, SNP 13- rs1805165, INDEL15- rs1805185)

SNPs identified (all by sequencing unless otherwise stated)

- rs1606803
- rs2364564
- rs867529
- rs13045
- rs6750998
- rs1805164
- rs1805165
- rs1805185
- rs1554098
- rs13425096
- rs7579242
- rs7605713
- SNP5F (C/A) subsequently monomorphic
- SNP3B (T/C) subsequently monomorphic
- SNP3C (A/G) subsequently monomorphic
- SNP3D (C/G) subsequently monomorphic
- rs10189080
- rs6740205
- rs4449134
- rs867014 (HapMap2)
- rs17689440 (HapMap2)

DNA sequencing resulting in SNPs identified for further genotyping

The path that was followed which lead us to the 11 SNPs that were finally genotyped in the datasets began with the sequencing data that was generated on type 1 diabetes. Further sequencing of *EIF2AK3* gene together with 1800bp 5' and 3000bp 3' of the gene in type 2 diabetes individuals and controls (n=48) was subsequently undertaken. 19 SNPs were identified by direct sequencing and by comparison with HapMap2, 5 of the SNPs were novel. 2 further tagging SNPs were identified from HapMap2 – these 21 SNPs formed the core of the analysis. Four of these SNPs were monomorphic, four SNPs failed genotyping due to assay problems (genotyping performed on amplifuor and KASpar at Kbiosicences) leaving 11 SNPs to be studied in all datasets. A sample sequence chromatogram for INDEL15 of *EIF2AK3 gene* is being shown in **Figure5.b**.

Figure 5.b: INDEL15 $AT^{+}/^{-}$ Sequence chromatogram

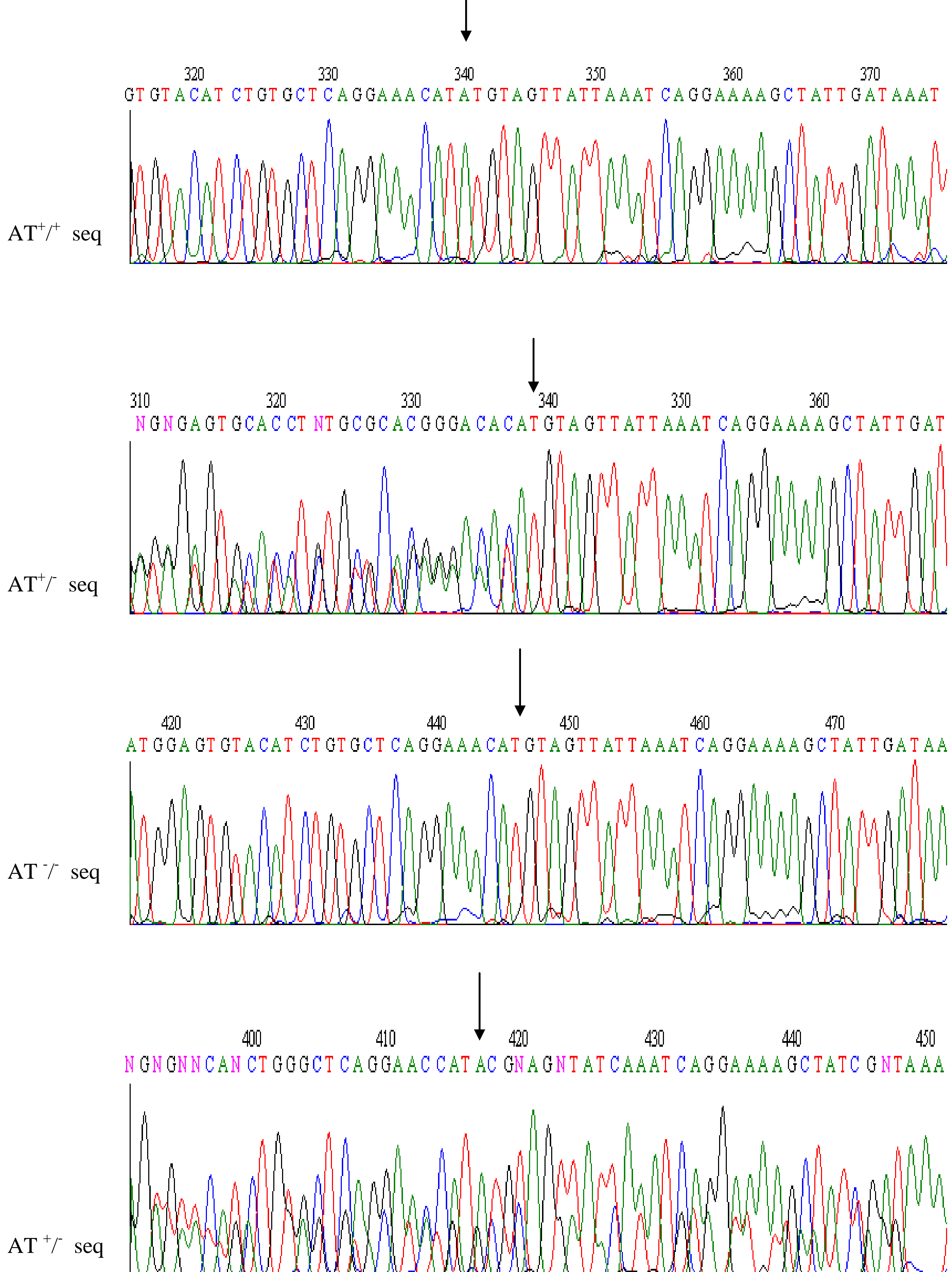


Figure5.b: A sample sequence chromatogram of INDEL 15 of *EIF2AK3* gene. This is the insertion deletion of AT sequence of *EIF2AK3* gene. Figure showing the wild type AT$^{+}/^{+}$ sequence at the top, the rare type AT $^{-}/^{-}$ sequence in the 3rd row and the heterozygote AT$^{+}/^{-}$ sequence in the 2nd and 4th row.

***EIF2AK3* SNPs: Nucleotide sequences identified**

1. WRS/SNP 2 [C/G] rs867529
GTCATTAGTAATTATCAGCACTTTAGATGGGAGAATTGCTGCCTTGGATCCTG
AAAATCATGGTAAAAAGCAGTGGGATTTGGATGTGGGATCCGGTT**[C/G]**CTTG
GTGTCATCCAGCCTTAGCAAACCAGAG

2. WRS/SNP 10 [T/A] rs6750998
TCAAACTGAAAATAAATATGATTCTGTAAGTGGTGAAGCCAATGACAGTAGC
TGGAATGACATAAAAAACTCTGGATATATATCACGGTAAG**[T/A]**GTCTTATAA
AATACAACCATCTGAATCAAAGAAGAAATGACCTAAGATCTTGTTTAACTTTT
TTTTTAATGTGTGGATATCTAGAAA

3. WRS/SNP 11 [A/G] rs1805164
ATATCTAACTGATTTTGAGCCAATTCA**[A/G]**TGCCTGGGACGTGGTGGCTTTG
GAGTTGTTTTTGAAGCTAAAAACAAAGTAGATGACTGCAATTATGCTATCAA
GAGGATCCGTCTCCCCAATAG

4. WRS/SNP 13 [G/T] rs1805165
CACAGACTGGCCACTCAGCTCTCCTAGCCCAATGGATGCACCATCAGTTAAA
ATACGCAGAATGGATCCTTTC**[G/T]**CTACAAAAGAACATATTGAAATCATAG
CTCCTTCACCACAAAGAAGCAGGTCTTTTTCAGTAGGGATTTCCTGTGACCA
GACAAGTTCATCTGAGAGCCAGTTCTCACCACTGGAATTCTCAGGAATGGAC
CATGA

5. WRS/INDEL 15 [AT+/-AT] rs1805185
TTTTTAAATTTTTTAAATATTTTCATTCTGTGGTTGGTTGAATCCTTGGATGTG
GAATCTGTGGATGTGGAAGACCAGCTGTATATTTTGAGGATATTTGATGGAG
TGTACATCTGTGCTCAGGAAACAT**[AT$^{+/-}$]**GTAGTTATTAAATCAGGAAAAGCT
ATTGATAAATTTTCCATTTGATAGATGTACAACCTCTTAGTCATTTTGTTAGA
GTATCAAAAAATATTTTCATGTTGTATGTCA

6. SNP5C [T/C] rs7579242
TATTTGGAGATAAGATCTTTAAAGCAGTGATTAAGTTAAAATGAGGCTGTTA
AAGGTAGCCCCACACTCCAATCTGACTGGTGTTAGAAGAATAGGAAGAGAC
ATCAGAGGGAAGAGGAAGTAAGAGGGCTG[T/C]CATCGGCAAGCCAAGGAG
AGAGGCCTCAGAATCCCACCTTGATCTTGGACTTCTAGCCTCCAGAACTTCG
AGAAAATAAACTAGTTTTGT

7. SNP5D [C/A] rs7605713
TCAGGCCACCTGGTCTGTGGTATTTGTTAGGGCAACTCTAGCAAACTCATAT
ACCTGGGAAGTTTCCTAGGCAGGCAAAACAACAAAGGAAAACCCCC[C/A]A
GGTGGGTCTTGATTGGCCCAGTGGAGTGAGCAAAGACAGTATGAAATGAGA
TCAGAAAAGCCATAGGAACCAGATGCTGTAGCACCCTGTAGTCTATTTTAA
GGACTTGACTTTTA

8. SNP3E [A/G] rs 10189080
CTGTGATTTGTTCAAGGCCACAAAGACAGCAAGAGCAGAACCGGGATCAAA
AGGGATTGGTAAATTCTTACATGATGT[A/G]ATCAGGTTTTCTTTCATTTTAA
CCTGGGTTTGCCTTTGCTGACCACCAGCAGGAAAGCAGGCATAGGGTTGGG
GAAGTTAAGAGCCTGGGCCAGCAACACTTAAGTCCGT

9. SNP3G [T/C] rs 4449134
TGCAATGGACTGTGGTGAAGTTTAGATTTATTTCAGCCCTTCTCCTGTACTGG
CCCTACCCTGACTACTTCTGGGATAGGCTGGCACTAACAAGATAAAATGCAC
CTGGGCCA[T/C]GGTTCTTTCTACTAACATCTCCCAAGCAATTCCTAAGAGGA
AGGGACTATTCTAGAAGGGACCTATAACCATTCCCCCAGATCCCTAAGCTCT
GGATGGACAAATGCAGGCCGA

SNPs identified for genotyping via HAPMAP

1. WRS Intron 1 [A/C] rs867014 (HAPMAP SNP)

CTGATAAACTTCCAAGGTTCCCTGTGCCCACTGAAAAAGAAAGGTCACACTG
CTTGGTGTGGCACAT[A/C]AAGCCTTTCACAATTTGGTTTCGGTGTACTTTCCT
AATCTCATTTTTCTGCCTCTTGTACCAGCCCAACCCTCTCTCCCCTCTCCTTCC
CAGCTCCACACTCCCCCCTTTCCATGGGGTG

2. WRS Intron 2 [C/T] rs17689440 (HAPMAP SNP)

GTGCATTATTTATTCTAAGACACACAAACATGCACCTTAACGGTTTTAGCTTG
GTTATGTATCCACACTCTTTTCAGGAGATAGTGGCAAATGAGATGTG[C/T]GT
CCCAAGGTACACAATTCATACAATTCATTCTAAATGTAATTTAAAACATCTGT
TCTTAGTGGAACTGTCAATTTCCCTTATCTTCCAGGA GCTTATTT

The non-random association of alleles at nearby loci (Linkage disequilibrium (LD), which is an important and widespread feature of the genome, facilitates survey in an association study a significant proportion of the common variation of a large number of SNPs that occur in regions of high LD by genotyping a few TagSNPs. Tagging SNPs exploit the extensive linkage disequilibrium that is present in the genome. We found that rs6750998 (SNP10) is a tagSNP and has allele capture for SNPs -rs4972060, rs12478762, rs4972059, rs7563643. There were four other SNPs that were tagging at least one other marker SNP with r^2 of 1.0 **(Table 5.a).** By this method we could genotype a few variants with extensive linkage disequilibrium resulting in extensive coverage across the whole gene- *EIF2AK3* and also have significant cost savings. The SNPs genotyped in our candidate gene association study are shown in table 5.b.

Table 5.a: Tagging SNPs as per HAPMAP for *EIF2AK3* gene

rs numbers	Tagging SNPs	r^2
rs1805165	rs4972221	1.0
rs1805164	rs17037621	1.0
rs867014	rs17037621	1.0
rs6750998	rs7563643	1.0

Table 5.b: Polymorphisms genotyped in *EIF2AK3* gene

Intron/exon *EIF2AK3*	**cDNA position/SNP**	**Minor Allele Frequency according to HAPMAP (CEU)**
Exon2	C/G rs867529	0.26
Intron 10	A/T rs6750998	0.24
Exon 11	A/G rs1805164	0.35
Exon 13	T/G rs1805165	0.35
Intron 15	$AT^{+/-}$ rs1805185	—**
SNP5C - 5'	T/C rs7579242	—*
SNP5D- 5'	C/A rs7605713	—*
SNP3E- 3'	A/G rs10189080	—*
SNP3G- 3'	T/C rs4449134	—*
Intron 1	A/C rs867014	0.35
Intron 2	C/T rs17689440	0.008

****- INDEL15 $AT^{+/-}$**
***- Novel SNPs : no MAF quoted in the public database/ OMIM.**

CHAPTER SIX: QUALITY CONTROL

6.1 GENOTYPE ERRORS – QUALITY CONTROL (QC)

Quality control checks were carried out for generating reliable data prior to carrying out the statistical analysis for disease association.

Steps for passing QC checks in our study:

- Checks on genotyping quality included careful visual inspection of genotype plots
- The proportion of samples with a SNP call rate typically >95%
- Concordance rates in duplicate samples (typically >99.5%)
- The minor allele frequency often >1%, as rarer SNPs are difficult to measure reliably
- Severe violations of Hardy-Weinberg equilibrium (HWE) were excluded

The process of internal quality control for the genotyping included identifying the duplicates and comparing for concordance for the genotypes among the duplicates. The second method involved use of blanks with no DNA in the plates and to check that they were showing negative signals. External quality control included genotyping and comparing the concordance for the same SNP analysis on 2 different platforms, i.e. Taqman and Amplifluor technology for different datasets. We performed SNP analysis for SNP 2, 10, 11 and 13 for the 1958 birth cohort and Bangladeshi case-control study on Taqman methodology at the Genome centre and Amplifluor technology at the KBiosciences.

In the final analyses, we performed statistical analysis only on those SNPs that were successfully genotyped in > 95% of the samples and discarded those datasets that failed QC. Those datasets/ genotypes that showed excessive discordant results for the duplicates and wrong calls on the blanks, not in HWE were also excluded from the final statistical analysis.

QC failure for rs1805185/ INDEL15

In the type 1 diabetes study this marker had been typed manually using a PCR-RFLP method. The number of samples to be genotyped and the amount of DNA that would be required prohibited the use of this assay in our study, despite in Rebecca Allotey's hand having a good QC.
Samples were genotyped by taqman but we only had successful genotyping in less than 95% of the sample set and failed the duplicate checks for QC. This marker therefore was outsourced to KBiosciences using their in house assay. However, it did not pass their initial QC steps in 2 separate assays (Amplifluor and KASpar). Since INDEL15 failed on 3 assays we did not further pursue this marker nor use any of the generated data. The exact reasons for failure of INDEL15 by high throughput genotyping on three different platforms were not identified. The probable reasons which accounted for INDEL15 failure on high throughput genotyping were likely to be - 1) INDEL15 being a 2 base pair as opposed to a single base pair polymorphism 2) Selected primers for INDEL15 attaching to numerous sites in the genome due to base pairs being identical 3) Sequence of interest being in a highly methylated region.

Quality control for SNPs by Taqman genotyping by visual check

Taqman genotyping for SNPs 2,10,11 and 13 in British Irish and South Asian cohorts generated allelic discrimination plots. Visual check for the allelic discrimination plots was undertaken and the quality of calling the wild type allele, the rare allele and the heterozygote was an important factor that formed part of the QC measures. The allelic discrimination plots for SNPs 2, 11 and 13 are shown as an example of the sample in Figures 6.a-6.e.

Figure 6.a: ECACC 1 (rs1805164) Allelic discrimination plot

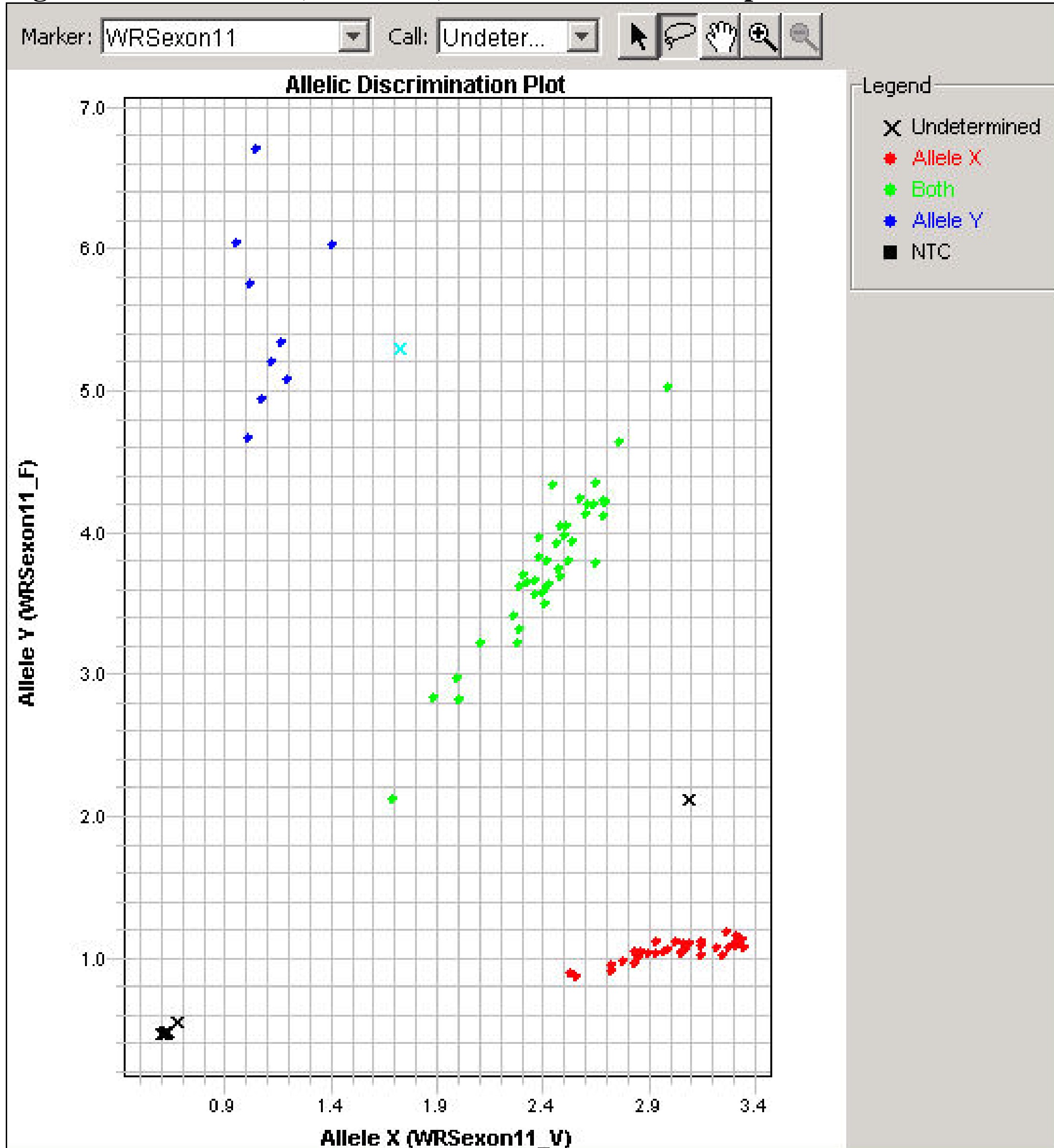


Figure 6.a: The quality of allelic discrimination plot was excellent in ECACC controls with clusters being clear and distinct. The wild type allele shown as red and is represented by Allele X, i.e. Allele A of rs1805164. The rare allele represented shown as blue and is represented by Allele Y, i.e. Allele G of rs1805164. The heterozygote alleles AG are shown as green and is represented as both. There were very few undetermined samples and they were the negative controls (blanks).

Figure 6.b: ECACC 1 (rs1805165) Allelic discrimination plot

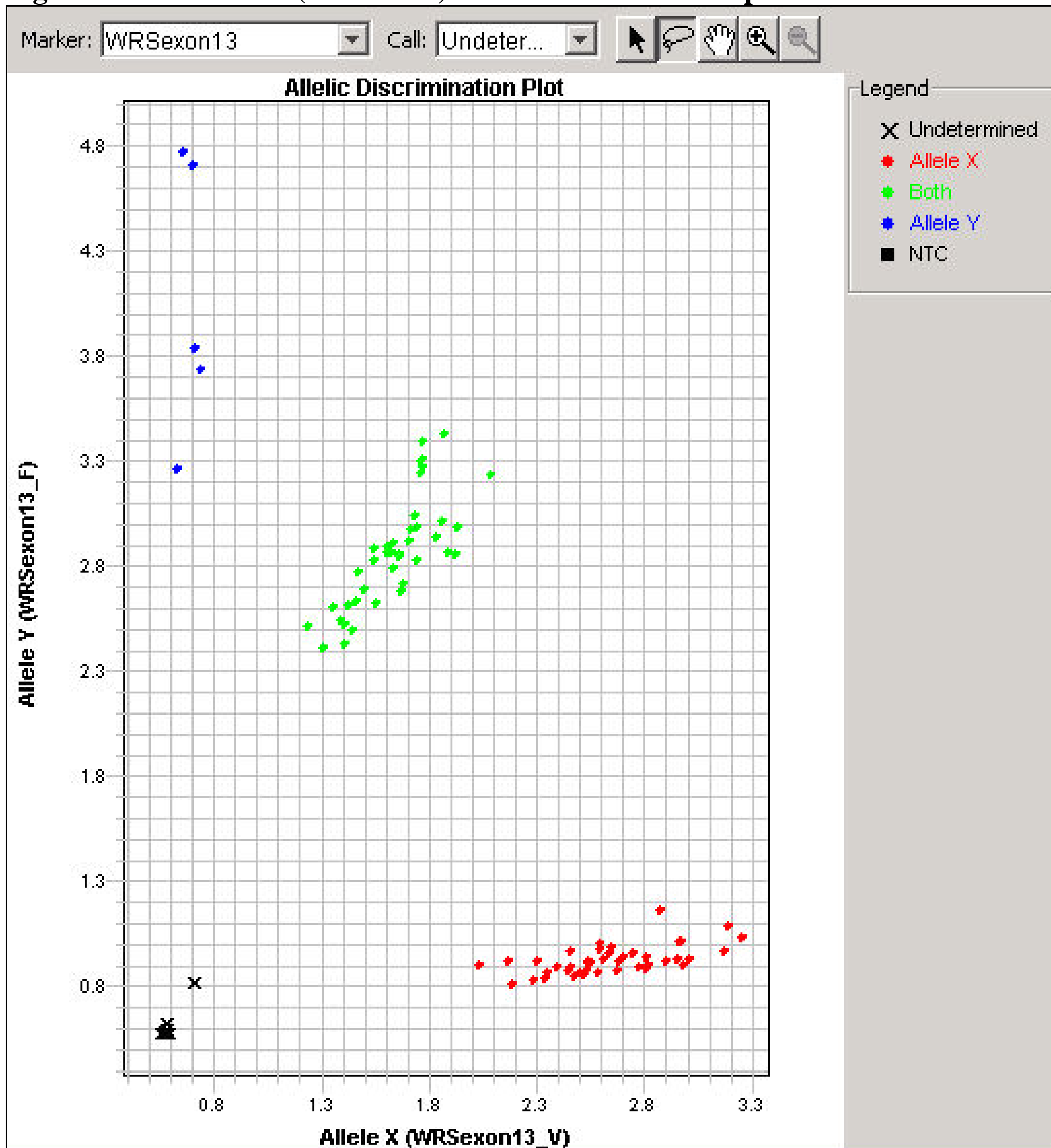


Figure 6.b: The quality of allelic discrimination plot was excellent in ECACC controls with clusters being clear and distinct. The wild type allele shown as red and is represented by Allele X, i.e. Allele T of rs1805165. The rare allele represented shown as blue and is represented by Allele Y, i.e. Allele G of rs1805165. The heterozygote alleles TG are shown as green and is represented as both. There were very few undetermined samples and they were the negative controls (blanks).

Figure 6.c: ECACC Raw data plots

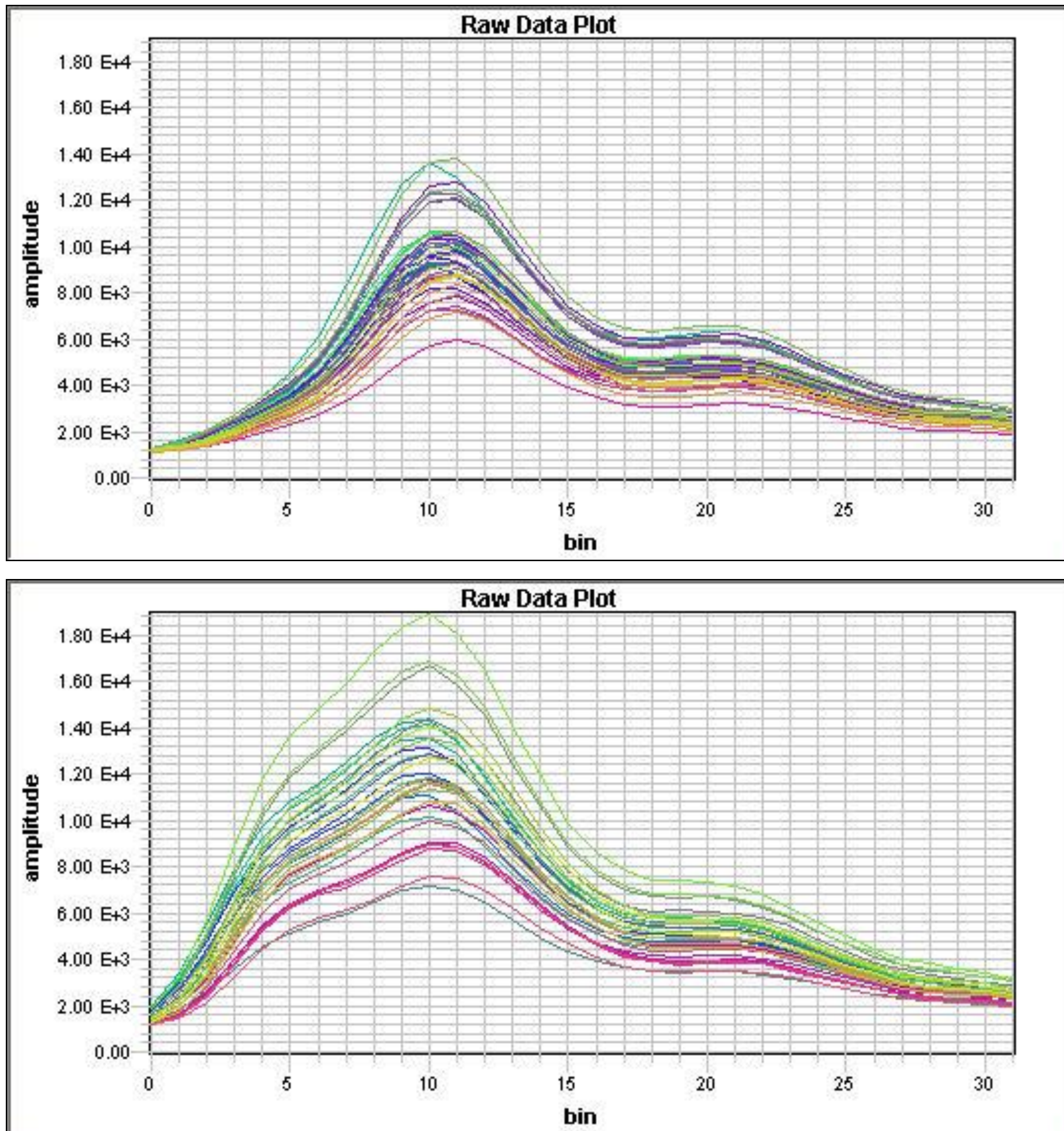


Figure 6.c: Raw data plots are shown above demonstrating the genotyping quality of individual DNA samples for the marker studied. Each of the coloured lines represents the individual DNA specimen showing the amplitude of the signal. The raw data plot in the top panel with a smoother amplitude indicating homozygote wild type/ rare alleles for the marker and the raw data plot in the bottom panel with a notch in the signal indicating heterozygote alleles.

Figure 6.d: W2 new trios plate 1 (rs867529) Allelic discrimination plot

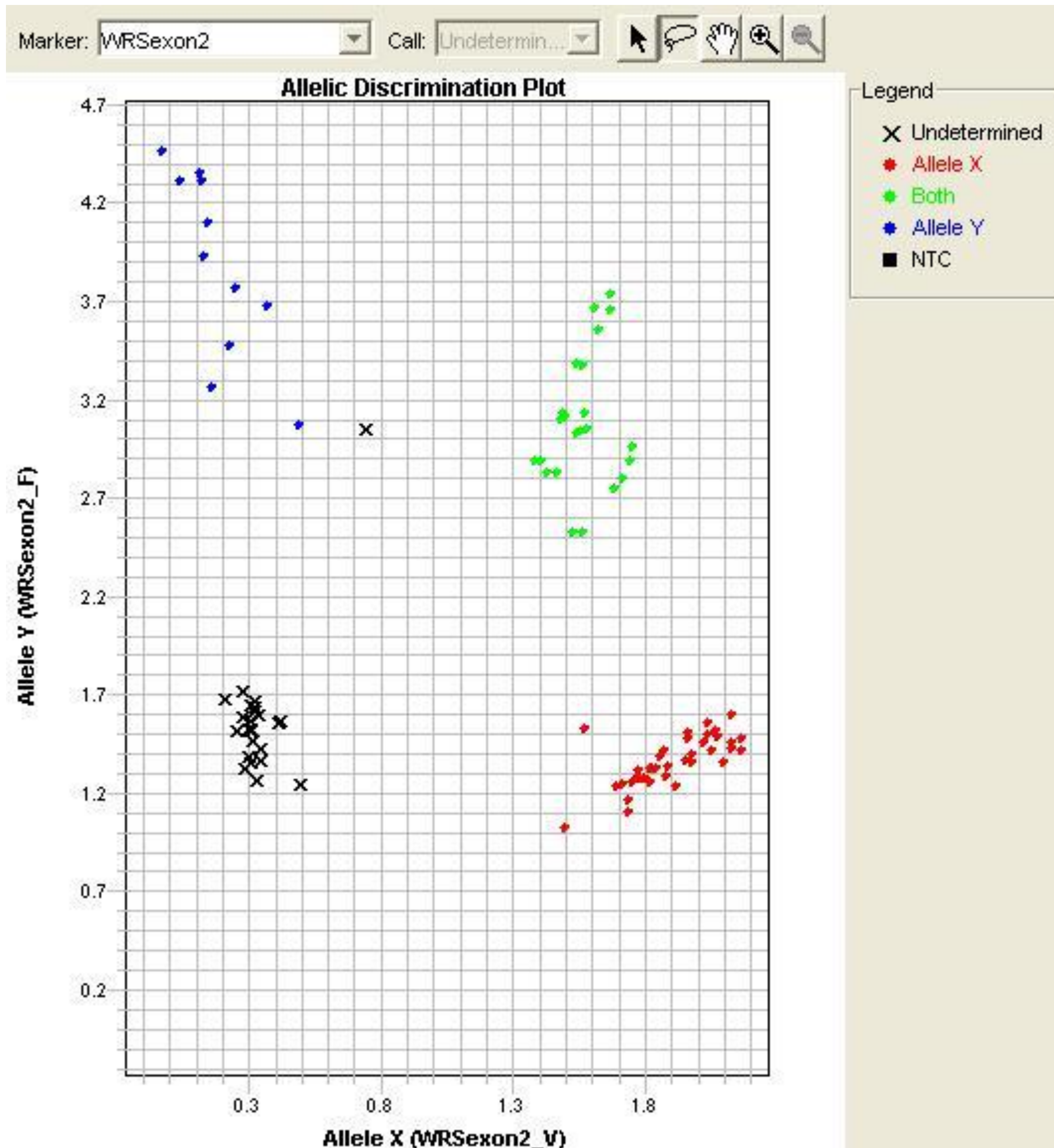


Figure 6.d: The quality of allelic discrimination was good in W2 trios plate 1 for rs867529 with clusters being clear and distinct for the wild type allele, the rare allele and the heterozygote. But there was a high failure of genotyping being shown as undetermined samples shown in black and marked as × in the above figure.

Figure 6.e: W2 new trios plate 2 (rs867529) Allelic discrimination plot

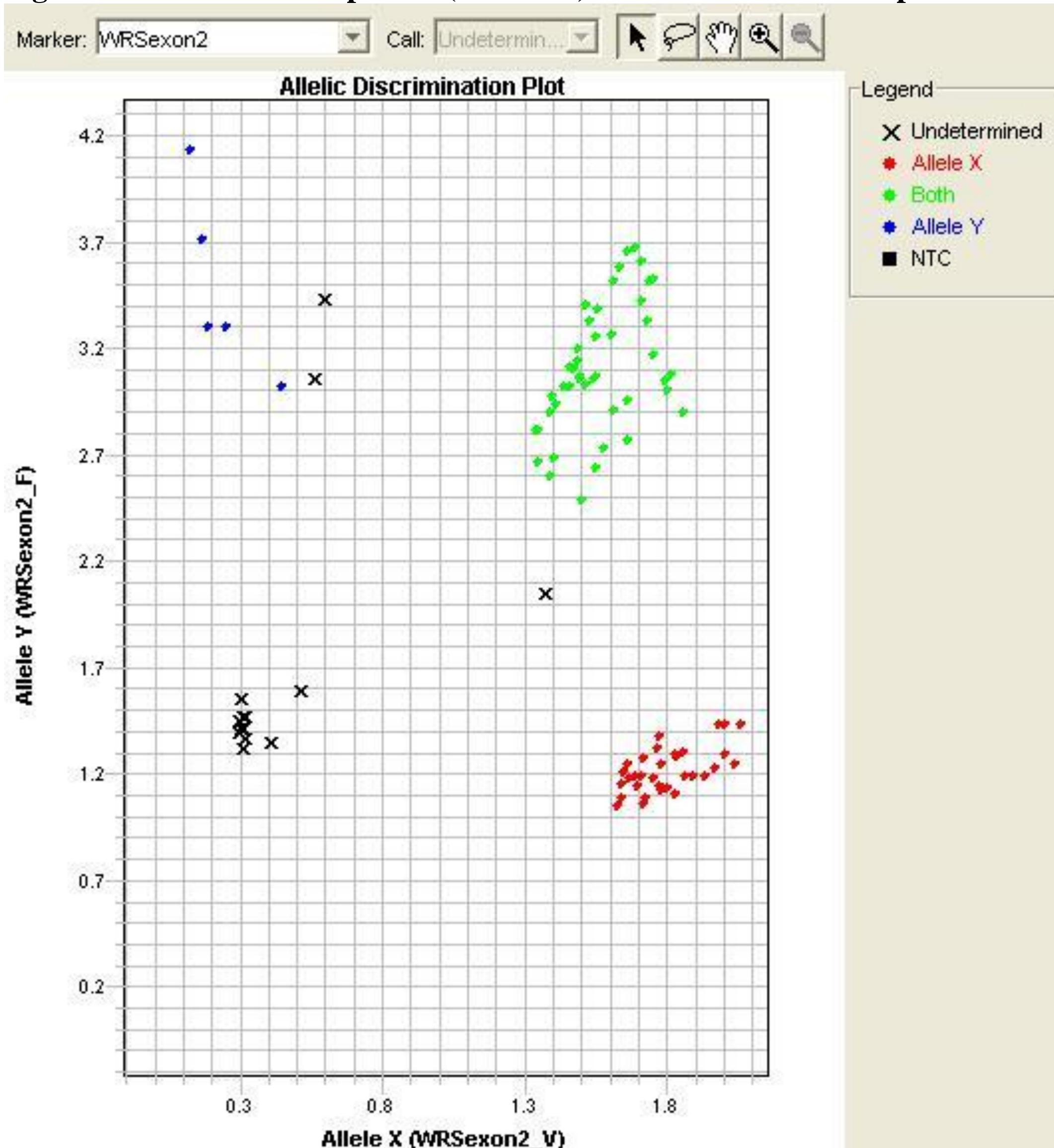


Figure 6.e: The quality of allelic discrimination was good in W2 trios plate 2 for rs867529 with clusters being clear and distinct for the wild type allele, the rare allele and the heterozygote. But there was a high failure of genotyping being shown as undetermined samples in the above figure.

Quality control for SNPs by Taqman Genotyping

Quality control checks for genotyping performed by Taqman methodology at the Genome centre are shown below. (Table 6.a) The genotyping error rates- i.e. percentage of failed genotyping/drop outs for each subset. Those markers with < 95 % genotyping rate were discarded and not utilised for the final analysis.

Table 6.a: Quality control checks

	W2 trios	**W2 Sib Pair Probands**	**CCS BAN**	**MRC 1958 BC**	**ECACC**
SNP2 = Failed Genotype	17%	1.7%	4.8%	2%	<1%
SNP 10 = Failed Genotype	8%	4.6%	8.4%	3%	<1%
SNP11 = Failed Genotype	17%	4.6%	9.2%	4%	<1%
SNP 13 = Failed Genotype	7%	10.4%	4.3%	1%	<1%

Combined SNP genotype failure for W2 trios was 12.3%. Therefore, the genotyping data for these markers was not analysed further. Combined SNP genotype failure for Warren 2 sib pair Probands was 5.3%. This dataset just passed QC. Combined SNP genotype failure for Bangladeshi CCS was 6.7% but genotyping data for final analysis was supplied by KBiosciences which passed the criteria for > 95% successful genotyping. Combined SNP genotype failure for MRC 1958 Birth cohort was 1.9%. Combined SNP genotype failure for ECACC was 0.46%. Concordance rates for duplicate samples was >99 % in W2 Sib Pair Probands, Bangladeshi CCS, ECACC and MRC 1958 BC cohorts. Concordance rates for W2 trios was <90 % and was unsuitable for further analysis.

QUALITY CONTROL ISSUES FOR TAQMAN GENOTYPING

The Warren 2 family trios, ECACC, 1958 BC Control samples and the South Asian resources were genotyped using taqman technology for SNPs 2, 10, 11 and 13 of the *EIF2AK3* gene at the Barts and The London Genome Centre by myself and supervised by Rebecca Allotey and Charles Mein. The genotyping data needed thorough quality control checks, especially for the Warren 2 trios data set.

High throughput genotyping was performed by taqman on ABI 7900. Initial results of the genotyping on the Warren 2 new trios plates 1-3 showed a much higher than expected failed genotyping (> 5% of the samples), i.e. the results were undetermined for either of the alleles. In the 1st plate there were 20 % genotyping failures. In the 2nd there were 8 % and in the 3rd there were 24 % genotyping failures (see **table 6.b and 6.c** below). The assay was set up in such a way that the 4 markers, SNPs 2, 10, 11 and 13 were being genotyped on a single plate of DNA in each run, i.e. 4 × 96= 384 plate. The results showed a clear pattern, in that the same DNA sample in the 384 well plate turned out to be undetermined for each of the SNP marker and this would not have been anticipated to be the case if the problem were the primers or taqman probe for one of the 4 SNPs accounting for the excess failure rate. This indicated a problem with the DNA rather than the taqman assay being the source of the dropouts. I went through the details of the process of DNA aliquoting performed by R. Allotey and myself at Prof. A. Hattersley's lab at Exeter. The original DNA samples were in eppendorf's in a concentration of 200 ng/ul. The datasets that were in eppendorf were the W2 old trios, new trios. Whereas W2 sib pair probands and ECACC samples were in working dilution of 20 ng/ul in a 96 well plate. The W2 sib pair probands and the ECACC samples in the 96 well plate format provided good results for the same SNPs using the taqman assay with a high genotyping call rate (> 95%). The samples in eppendorf were vortexed, centrifuged and then collected in a 96 well plate. We took 20 ul of the original DNA and diluted with 180 ul of water to give a final concentration of 20 ng/ul. The samples were brought to our lab in London and stored in – 20 $^{\circ}$C freezer on the same evening.

The original samples were already quantified for DNA by picogreen at the lab in Exeter. Therefore, we planned to do a gel run with the DNA. This experiment was performed for all the data sets collected in Exeter. In the experiment I took 2 ul of DNA and 2 ul of loading dye and running it on a 3% agarose gel. The product was visualised under UV light. The results of this experiment showed us the lack of a DNA band in some wells from the W2 trios data sets with the exception of ECACC control samples. In our initial collection from Exeter it is pertinent to note that the ECACC collection was in a 96 well format from which we took the aliquots and this could explain the good results we observed on the gel run with these samples.

Table 6.b: W2 New trios plate 1-genotyping failures on taqman and gel run

	1	2	3	4	5	6	7	8	9	10	11	12
A	X (Plate ID)	X (Plate ID)	X (Plate ID)	X (17/75/1)								
B					X (13/4/2)	X (13/4/3)		X (13/12/2)		X (13/36/1)	X (13/36/2)	X (13/36/3)
C				X (18/1/1)	X (18/1/2)	X (18/1/3)						
D					X (13/6/2)							
E							X (13/23/1)					
F								X (12/78/2)	X (12/78/3)			
G							X (water)					
H		X (12/98/2)	X (12/98/3)					X (13/30/2)	X (13/30/3)	X (Blank)	X (Blank)	X (Blank)

The next step involved checking if the initial failed genotypes on taqman assays and the empty DNA wells (Agarose gel run) tallied. I found in the three W2 new trios plates there was a clear pattern of this being the case (Represented by X for the empty DNA and the sample ID number for the failed genotype as shown in **Table 6.b & 6.c**). This suggested the problem was DNA quality rather than a problem with taqman assay.

Table 6.c: Warren 2 New trios plate 2 genotyping failures on taqman and gel run

	1	2	3	4	5	6	7	8	9	10	11	12
A												
B	X (Plate ID)	X (Plate ID)	X (Plate ID)	X 18/25/1	X 18/25/2	X 18/25/3				X (Water)	X (Water)	X (Water)
C							X 13/87/1					
D												
E												
F												
G												
H							X 13/53/1	X 13/53/2				

To confirm the problem lay with DNA being of poor quality I genotyped SNP 44 of *Calpain 10* by taqman on a South Asian data set with very small genotype failure rate (4 %), and typed the same marker on W2 New trios plate and the genotype failure rate was much higher in the W2 New trios plate 1 (20%) and the failed DNA wells were identical to those obtained with *EIF2AK3* markers.

This suggested the problem was the lack of DNA in the some of the wells rather than the taqman assay; either the plates dried up or less likely that they were incorrectly aliquoted in the first place. We therefore decided to genotype the rest of the data sets for SNPs 2, 10, 11 and 13 by taqman as we were confident of the taqman assay not being the problem.

Warren 2 trios failed the pre-set QC targets likely to be due to varying DNA concentrations in the plates. Therefore results from these plates were discarded. The Sib Pair probands, South Indian and Bangladeshi cohorts and the controls- ECACC, MRC birth cohort all had good genotyping data generated by taqman methodology for the 4 markers, except INDEL15, which coincidentally failed at KBiosciences as well.

This process highlights the importance of 1) good quality DNA to be available for genotyping and 2) robust quality control measures in genotyping to give reliable data.

The DNA resources that were available at KBiosciences were of better quality and all the 10 SNPs that were analysed by Amplifuor technology passed QC according to their internal quality control measures, except INDEL15, which also failed by taqman previously. Indeed the majority that was successfully genotyped were those DNA that were available at KBiosciences from the GWA and replication studies.

6.3 QUALITY CONTROL AMPLIFUOR AND KASpar GENOTYPING

The majority of the genotyping was outsourced to KBiosciences for cost effectiveness. After initial set up and the exclusion of either monomorphic markers or those not able to be typed by the KBioscience in house assays 11 SNPs were studied in British Irish, Bangladeshi and South Indian cohorts. After a further QC step 10 SNPs were analysed excluding INDEL15. The genotyping performed at KBiosciences passed their internal QC measures.

Quality control criteria for KBiosciences

KBiosciences criteria for routine quality control measures on all its Genotyping are shown below:

- Water controls must be negative
- Inter plate duplicate testing of a known DNA
- Intra plate testing of a known DNA
- Percentage success rate of calls must be greater than 95%
- Clusters are clear and distinct

The results released by KBiosciences were further checked as a QC measure by comparing with genotyping performed on Taqman. The datasets for Bangladeshi case control study, 1958 MRC birth cohort were re-genotyped (Initial genotyping was performed on taqman) for the SNPs 2, 10, 11, 13 at KBiosciences as an external quality step; there was 99% concordance of genotyping calls between both methodologies.

6.4 MINOR ALLELE FREQUENCIES

The minor allele frequencies (MAFs) in the genotyped SNPs varied from 2% to 32% in the British Irish resources, 1% to 41% in the Bangladeshi resources and from 1% to 45% in the South Indian resources. We compared the MAFs that were obtained in the British Irish population with the MAFs quoted in the NCBI/HAPMAP public databases and were very similar. There are no publicly available data on either the Bangladeshi or the South Indian populations for MAF. **(Table 6.d)**

Table 6.d: Minor allele frequencies

Marker	**Minor Allele Frequency HAPMAP (CEU)**	**British Irish**	**Bangladeshi**	**South Indian**
rs number		**N=6112**	**N=985**	**N=998**
		MAF	**MAF**	**MAF**
1805165	0.35	0.28	0.39	0.45
867529	0.26	0.28	0.41	0.44
6750998	0.24	0.27	0.25	0.30
10189080	—*	0.08	0.02	0.01
7579242	—*	0.08	0.03	0.02
1805164	0.35	0.30	0.19	0.09
4449134	—*	0.32	0.22	0.13
7605713	—*	0.27	0.25	0.28
17689440	0.008	0.02	0.01	0.01
867014	0.35	0.30	0.20	0.11

Hapmap CEU-30 mother-father-child trios from the CEPH collection (Utah residents with ancestry from northern and western Europe), representing one of the populations studied in the International HapMap project

—* Novel SNPs

MINOR ALLELE FREQUENCIES- POPULATION CHARACTERISTICS

Given the possibility of differences in MAF in the three different cohorts we proceeded to compare the MAFs among the cohorts utilising a chi squared test (χ^2); British Irish with South Indian (SI), British Irish with Bangladeshi (Bang), South Indian with Bangladeshi. The results (p-values) are shown in **table 6.e**. Comparing the minor allele frequencies (MAFs) in the 3 ethnic groups showed that 9 of the 10 minor alleles had frequencies that differed between UK whites and the South Indians and all 10 minor alleles differed between UK whites and the Bangladeshi, whereas in the South Asian sub-groups MAFs were very similar.

Table 6.e: Minor allele frequency- Population characteristics

rs number	British Irish vs. SI	British Irish vs. Bang	SI vs. Bang
1805165	1.3×10^{-32}	1.7×10^{-23}	0.16
867529	9.3×10^{-32}	1.2×10^{-22}	0.14
6750998	0.0360394	1.1×10^{-05}	0.83
10189080	5.6×10^{-12}	0.000791481	0.78
7579242	5.9×10^{-8}	0.0081	0.07
1805164	1.1×10^{-110}	3.0×10^{-25}	0.14
4449134	3.3×10^{-13}	2.4×10^{-32}	0.37
7605713	0.18	1.2×10^{-05}	0.98
17689440	0.0009	0.00089	0.70
867014	1.5×10^{-31}	5.8×10^{-08}	0.30

6.5 HARDY-WEINBERG EQUILIBRIUM

The results showed majority of the SNPs were in HWE in all three cohorts. **(Table 6.f)** The SNPs that failed HWE are highlighted in red. SNP rs4449134 was not in HWE in all three population sub groups. SNPs rs10189080 and rs4449134 were not in HWE in the Bangladeshi and South Indian subjects. SNP rs17689440 was not in HWE in South Indian population.

Table 6.f: Hardy-Weinberg equilibrium results for 3 populations

Marker	**British Irish**	**Bangladeshi**	**South Indian**
rs number	HWE PValue	HWE PValue	HWE PValue
1805165	0.71	0.88	0.41
867529	0.90	0.82	0.41
6750998	0.67	0.86	0.17
10189080	0.99	6.24×10^{-49} *	0.005*
7579242	0.84	0.53	0.15
1805164	0.93	0.29	0.10
4449134	0.02*	9.3×10^{-08} *	0.001*
7605713	0.64	0.90	0.18
17689440	0.54	0.81	2.24×10^{-13} *
867014	0.52	0.62	0.06

HARDY-WEINBERG EQUILIBRIUM

The details of the observed and expected genotype counts for the markers not in HWE are shown in the following tables- **(Table 6.g-1, 2, 3, 4, 5 and 6)**

Table 6.g-1, 2, 3, 4, 5 and 6: HWE Failures

1. **rs 4449134 (British Irish)**

Observed	**TT**	**CT**	**CC**	**NUM**	
	2184	1925	494	4603	
Expected	**TT**	**CT**	**CC**		
	2150.872	1991.257	460.8717	4603	
	0.510252	2.204612	2.381329	**5.096193**	**chi=sq**
				0.023978	**p-value**

2. **rs 4449134 (Bangladeshi)**

Observed	**TT**	**CT**	**CC**	**NUM**	
	652	**344**	**101**	1097	
Expected	**TT**	**CT**	**CC**		
	618.9389	410.1222	67.93892	1097	
	1.765982	10.66058	16.08849	**28.51505**	**chi=sq**
				9.3×10^{-08}	**p-value**

3. **rs 4449134 (South Indian)**

Observed	**TT**	**CT**	**CC**	**NUM**	
	335	**128**	**28**	491	
Expected	**TT**	**CT**	**CC**		
	324.2383	149.5234	17.23829	491	
	0.357189	3.098228	6.71844	**10.17386**	**chi=sq**
				0.001424	**p-value**

4. **rs10189080 (Bangladeshi)**

Observed	**AA**	**GA**	**GG**	**NUM**	
	996	38	17	1051	
Expected	**AA**	**GA**	**GG**		
	980.2331	69.53378	1.233111	1051	
	0.253608	14.30066	201.5996	**216.1539**	**chi=sq**
				6.24×10^{-49}	**p-value**

5. **rs10189080 (South Indian)**

Observed	**AA**	**GA**	**GG**	**NUM**	
	474	23	2	499	
Expected	**AA**	**GA**	**GG**		
	472.3652	26.26954	0.36523	499	
	0.005658	0.406931	7.317219	**7.729808**	**chi=sq**
				0.005432	**p-value**

6. **rs17689440 (South Indian)**

Observed	**TT**	**TC**	**CC**	**NUM**	
	491	4	1	496	
Expected	**TT**	**TC**	**CC**		
	490.0181	5.96371	0.018145	496	
	0.001967	0.646604	53.12926	**53.77783**	**chi=sq**
				2.2×10^{-13}	**p-value**

SNP rs4449134 was not in HWE in all three population sub groups and was not analysed for association study. SNPs rs10189080 and rs4449134 were not in HWE in the Bangladeshi and the South Indian subjects, SNP rs17689440 was not in HWE in the South Indian population and were not analysed for association study.

CHAPTER SEVEN: GENOTYPING RESULTS

7.1 DISEASE ASSOCIATION STATISTICS- GENOTYPING RESULTS

Genotyping analysis for candidate gene association for the 10 single nucleotide polymorphisms (SNPs) of the *EIF2AK3* gene was performed using the program - UNPHASED 3.0.6 log likelihood ratio tests. We generated clean data with the quality control measures aforementioned.

There were three ways in which the statistical analysis was approached (hierarchical approach to statistical analysis of the three populations). The reasoning behind our approach to statistical analysis of all the subjects for a combined analysis was to look for a trend or positive signal for any of the markers. Following on after the combined analysis our effort was to delve into the ethnic specific sub-groups and further sub-group analysis within the ethnic groups for all three populations studied. The rationale to take this approach was to avoid false positive results in the candidate gene association study.

- All three groups were combined which included geographical region as a confounder. The way we performed the statistical analysis with the help of program UNPHASED using log likelihood ratios facilitated the analysis of all three groups combined, i.e. British Irish cohort, South Indian cohort and Bangladeshi cohort. This analysis includes the Warren 2 trios, Warren 2 diabetes cases and Warren 2 controls from the British Irish cohort, the Bangladeshi young diabetes cases and controls, the Bangladeshi diabetes case-control study cohort, the South Indian urban survey diabetes cases and controls, the South Indian Newham diabetes cases and controls and the South Indian family trios. Geographical confounders may have values that could influence the population frequency of minor allele frequency and UNPHASED allowed for modelling of interactions between haplotype and geographical covariates.

- Within each of the ethnic group the various datasets were analysed as a combined group. The British Irish combined analysis includes the Warren 2 T2DM cases

and controls, Warren 2 trios. The Bangladeshi combined analysis includes the Young diabetes cases and controls, Bangladeshi case-control study. The South Indian combined analysis includes urban survey diabetes cases and controls, the South Indian Newham diabetes cases and controls and the South Indian family trios. Age was applied as a confounder when testing as a combined data set within the particular ethnic group as different data sets had differences in age. For example, the young Bangladeshi cohort had a mean age of 18 years and the rest of the Bangladeshi cohort had a mean age of 36 years.

- We studied all three ethnic groups in detail and analysed each subset in a particular ethnic group separately. For example, the young Bangladeshi cohort was analysed on its own and the Bangladeshi Whitechapel and Sylhet cohort was analysed separately.

SUBJECTS STUDIED -

a) All 3 groups combined (British Irish, Bangladeshi, South Indian)
b) Warren2 British Irish population
c) Bangladeshi population
d) South Indian population
e) Replication set from Dundee

- SNP rs1805185 (previously referred to as 15INDEL) failed QC by 2 separate assays and therefore was not included in the final analysis.
- Ten SNPs were included in the final analysis except those SNPs that failed HWE
- Four SNPs failed QC by Taqman in the W2trios plates.

Genotyping results / UNPHASED Output

Combined analysis (British Irish, Bangladeshi, South Indian)

The following table **(Table 7.a)** summarizes the results and the p-values generated as a measure of significance of the association of the marker, i.e. the SNP, with being a case (Type 2 diabetes) for all samples in the combined analysis of the British Irish, the Bangladeshi and the South Indian subjects. The p-values were generated for each SNP using the log likelihood ratio for all samples- combined group of British Irish, Bangladeshi and South Indian resources. Since there was a significant difference in allele frequencies between regions we included geographical region as a confounder (which is a type of covariate) with discrete levels. (i.e. treated as a "factor" by UNPHASED). The results are detailed in the following table (Table 7.a) that has all samples combined for all our data sets.

Disease association statistics- Genotyping results- p-values

Comparing the alleles between cases and controls and allelic transmission in the family studies showed that in 8 of 9 SNPs analysed for *EIF2AK3* gene in the UK white cohort there was no statistically significant difference; in the 7 SNPs analysed in the South Indian cohort there was no statistically significant difference. In the novel SNP in the 5' UTR (rs7605713) an equivocal p-value was obtained that was individually significant in the UK whites and a subset of the Bangladeshi cohort and further replication study was carried out for this marker in the white population. An equivocal p-value was also seen for the marker rs6750998 in the Bangladeshi cohort alone. None of the 9 SNPs in the combined analysis for all three populations showed any statistically significant difference between type 2 diabetes cases and the controls (Table 7.a). The results for the combined analysis within the British Irish, Bangladeshi and South Indian cohorts and their summarized information for these analyses are given in the future tables.

Table 7.a: Genotyping results – p-values all SNPs, 3 groups

Marker	All samples	British Irish	Bangladeshi	South Indian
	p-value	p- value	p- value	p- value
rs number	N=	N=6112	N= 985	N=998
1805165	0.82	0.40	0.91	0.26
867529	0.45	0.12	0.78	0.24
6750998	0.63	0.48	0.04*	0.37
10189080	0.78	0.78	—	—
7579242	0.40	0.43	0.58	0.83
1805164	0.15	0.22	0.19	0.66
4449134	—	—	—	—
7605713	0.44	0.05***	0.05**	0.36
17689440	0.97	0.88	0.56	—
867014	0.14	0.31	0.11	0.92

—rs4449134 failed HWE in all 3 cohorts

—rs10189080 failed HWE in South Asian cohorts

—rs17689440 failed HWE in South Indian cohort

*** p=0.05 in British Irish combined dataset/ W2 trios

** p=0.007 in the young Bangladeshi dataset

* p=0.005 in the young Bangladeshi dataset

All p- values were generated using UNPHASED 3.0.6 (p-value is a measure of significance of the association of the SNP with being a case)

British Irish cohort combined analysis

The following table **(Table 7.b)** summarizes the results and the p-values generated as a measure of significance of the association of the marker, i.e. the SNP, with being a case (Type 2 diabetes) in the combined analysis of the British Irish subjects which include the W2 T2D cases, Sib pair probands and OCDEM_OXGN Cases (Total cases=2206). It also includes the analysis of 418 W2 trios (n=1254). The control population includes the MRC Birth cohort and the ECACC (Total control group= 2652).

Table 7.b: Genotyping results – p-values all SNPs, British Irish Cohort

British Irish combined	rs numbers	Allele	p-values	Case Freq	Control Freq
1	rs1805165	G	0.4	0.286	0.277
		T		0.714	0.723
2	rs867529	C	0.12	0.707	0.722
		G		0.293	0.278
3	rs6750998	A	0.48	0.733	0.726
		T		0.267	0.274
4	rs10189080	A	0.78	0.918	0.917
		G		0.082	0.083
5	rs7579242	C	0.43	0.082	0.082
		T		0.918	0.918
6	rs1805164	A	0.22	0.704	0.693
		G		0.296	0.307
7	rs7605713	A	0.05	0.734	0.720
		C		0.266	0.280
8	rs17689440	C	0.88	0.019	0.019
		T		0.981	0.981
9	rs867014	A	0.31	0.702	0.694
		C		0.298	0.306

The table 7.b shows the reference SNPs in 2nd column, the alleles studied in the 3rd column, and the p- values generated using log likelihood ratios. The results also show the frequency of the alleles in the cases and controls respectively. rs449134 failed HWE. SNP rs7605713 (novel 5' of *EIF2AK3* SNP) in the British Irish combined analysis showed a nominally significant p-value of 0.05.

British Irish case-control cohort

The following table (Table 7.c) summarizes the results and the p-values generated as a measure of significance of the association of the marker, i.e. the SNP, with being a case (Type 2 diabetes) in the analysis of the British Irish W2 T2D cases, Sib pair probands and OCDEM_OXGN Cases (Total cases=2206). The control population includes the MRC Birth cohort and the ECACC (Total control group= 2652).

Table 7.c: Genotyping results – p-values all SNPs, British Irish case-control

British Irish case-control	rs numbers	Allele	p-values	Case Freq	Control Freq
1	rs1805165	G	0.39	0.286	0.277
		T		0.714	0.723
2	rs867529	C	0.12	0.707	0.722
		G		0.293	0.278
3	rs6750998	A	0.47	0.733	0.726
		T		0.267	0.274
4	rs10189080	A	0.92	0.917	0.916
		G		0.083	0.084
5	rs7579242	C	0.91	0.085	0.084
		T		0.915	0.916
6	rs1805164	A	0.21	0.704	0.693
		G		0.296	0.307
7	rs7605713	A	0.27	0.734	0.724
		C		0.266	0.276
8	rs17689440	C	0.83	0.020	0.020
		T		0.980	0.980
9	rs867014	A	0.34	0.703	0.694
		C		0.297	0.306

The table 7.c shows the reference SNPs in 2nd column, the alleles studied in the 3rd column, and the p- values generated using log likelihood ratios. The results also show the frequency of the alleles in the cases and controls respectively. rs449134 failed HWE. None of the markers showed any statistically significant difference between cases and controls in the British Irish case-control cohort.

Warren 2 trios (family study)

The results in table 7.d show the analysis of the Warren 2 family study subjects analysed separately using the log likelihood ratio test. rs449134 failed HWE and four markers failed QC measures (rs867529, rs6750998, rs1805164 and rs1805165) leaving 5 markers which could be analysed in the family study.

Table 7.d: Genotyping results – p-values all SNPs, Warren 2 trios

British Irish trios	**rs numbers**	**p-values**
1	rs10189080	0.91
2	rs7579242	0.81
3	rs7605713	0.05
4	rs17689440	0.36
5	rs867014	0.89

The table 7.d shows the reference SNPs in 2nd column, and the p- values generated using log likelihood ratios. rs7605713 showed a nominally significant p-value of 0.05 in the Warren 2 trios family study.

Bangladeshi cohort combined analysis

The following table (Table 7.e) summarizes the results and the p-values generated as a measure of significance of the association of the marker, i.e. the SNP, with being a case (Type 2 diabetes) in the combined analysis of the Bangladeshi subjects which include the T2D cases (n= 426) and controls (n= 559). This group analysis includes the young Bangladeshi cohort, Whitechapel and Sylhet Bangladeshi case-control cohort.

Table 7.e: Genotyping results – p-values all SNPs, Bangladeshi Cohort

Bangladeshi combined	**rs numbers**	**Allele**	**p-values**	**Case Freq**	**Control Freq**
1	rs1805165	G	0.91	0.396	0.398
		T		0.604	0.602
2	rs867529	C	0.78	0.605	0.599
		G		0.395	0.401
3	rs6750998	A	0.04	0.684	0.727
		T		0.316	0.273
4	rs7579242	C	0.58	0.038	0.033
		T		0.962	0.967
5	rs1805164	A	0.19	0.830	0.806
		G		0.170	0.194
6	rs7605713	A	0.05	0.689	0.729
		C		0.311	0.271
7	rs17689440	C	0.56	0.006	0.008
		T		0.994	0.992
8	rs867014	A	0.11	0.832	0.804
		C		0.168	0.196

The table 7.e shows the reference SNPs in 2nd column, the alleles studied in the 3rd column, and the p- values generated using log likelihood ratios. The results also show the frequency of the alleles in the cases and controls respectively. rs4449134 and rs10189080 failed HWE. SNPs rs6750998 and rs7605713 showed p-value of 0.04 and 0.05 respectively, a nominally statistically significant difference between cases and controls in the combined analysis of the Bangladeshi cohort.

Bangladeshi case-control cohort (Young diabetes cohort)

The following table (Table 7.f) summarizes the results and the p-values generated as a measure of significance of the association of the marker, i.e. the SNP, with being a case (Type 2 diabetes) in the the analysis of the young Bangladeshi subjects which include the T2D cases (n= 330) and controls (n= 374).

Table 7.f: Genotyping results – p-values all SNPs, young Bangladeshi case-control

Bangladeshi CCS	**rs numbers**	**Allele**	**p-values**	**Case Freq**	**Control Freq**
1	rs1805165	G	0.61	0.392	0.406
		T		0.608	0.594
2	rs867529	C	0.50	0.607	0.589
		G		0.393	0.411
3	rs6750998	A	0.005	0.672	0.741
		T		0.328	0.259
4	rs7579242	C	0.54	0.037	0.031
		T		0.963	0.969
5	rs1805164	A	0.17	0.832	0.803
		G		0.168	0.197
6	rs7605713	A	0.007	0.677	0.743
		C		0.323	0.257
7	rs17689440	C	0.33	0.006	0.011
		T		0.994	0.989
8	rs867014	A	0.06	0.838	0.800
		C		0.162	0.200

The table 7.f shows the reference SNPs in 2nd column, the alleles studied in the 3rd column, and the p- values generated using log likelihood ratios. The results also show the frequency of the alleles in the cases and controls respectively. The results are for the 8 SNPs that passed QC and HWE. SNP rs7605713 was showing a significant p- value of 0.007. SNP rs6750998 was showing a significant p- value of 0.005 in this sub set alone.

Bangladeshi Whitechapel, Sylhet case-control cohort

The following table (Table 7.g) summarizes the results and the p-values generated as a measure of significance of the association of the marker, i.e. the SNP, with being a case (Type 2 diabetes) in the analysis of Bangladeshi subjects (Whitechapel and Sylhet cohort) which include the T2D cases (n= 96) and controls (n= 185).

Table 7.g: Genotyping results – p-values all SNPs, Bangladeshi case-control Whitechapel and Sylhet cohort

Bangladeshi Whitechapel/ Sylhet	**rs numbers**	**Allele**	**p-values**	**Case Freq**	**Control Freq**
1	rs1805165	G	0.58	0.407	0.382
		T		0.593	0.618
2	rs867529	C	0.65	0.60	0.620
		G		0.40	0.380
3	rs6750998	A	0.48	0.725	0.697
		T		0.275	0.303
4	rs7579242	C	0.89	0.039	0.036
		T		0.961	0.964
5	rs1805164	A	0.81	0.822	0.814
		G		0.178	0.186
6	rs7605713	A	0.47	0.731	0.701
		C		0.269	0.299
7	rs17689440	C	0.63	0.006	0.002
		T		0.994	0.998
8	rs867014	A	0.99	0.813	0.813
		C		0.187	0.187

The table 7.g shows the reference SNPs in 2nd column, the alleles studied in the 3rd column, and the p- values generated using log likelihood ratios. The results also show the frequency of the alleles in the cases and controls respectively. The results are for the 8 SNPs that passed QC/ HWE. SNPs rs7605713 and rs6750998 which were showing a significant p-value in young Bangladeshi sub set were not statistically significant in this cohort.

South Indian cohort combined analysis

The following table (Table 7.h) summarizes the results and the p-values generated as a measure of significance of the association of the marker, i.e. the SNP, with being a case (Type 2 diabetes) in the combined analysis of the South Indian subjects which include the T2D cases (n= 203) and controls (n= 322) from the Urban survey and Newham collection. It also includes the analysis of the 104 families (n=468).

Table 7.h: Genotyping results – p-values all SNPs, SI Cohort

3.South Indian combined	**rs numbers**	**Allele**	**p-values**	**Case Freq**	**Con Freq**
1	rs1805165	G	0.26	0.408	0.430
		T		0.592	0.570
2	rs867529	C	0.24	0.595	0.572
		G		0.405	0.428
3	rs6750998	A	0.37	0.69	0.712
		T		0.31	0.288
4	rs7579242	C	0.83	0.040	0.049
		T		0.960	0.951
5	rs1805164	A	0.66	0.83	0.837
		G		0.17	0.163
6	rs7605713	A	0.36	0.695	0.719
		C		0.305	0.281
7	rs867014	A	0.92	0.829	0.829
		C		0.171	0.171

The table 7.h shows the reference SNPs in 2nd column, the alleles studied in the 3rd column, and the p- values generated using log likelihood ratios. The results also show the frequency of the alleles in the cases and controls respectively. Three SNPs failed HWE (rs4449134, rs10189080, rs17689440). None of the SNPs showed any statistically significant difference between cases and controls in the combined analysis of the South Indian cohort.

South Indian Newham case-control cohort

The following table (Table 7.i) summarizes the results and the p-values generated as a measure of significance of the association of the marker, i.e. the SNP, with being a case (Type 2 diabetes) in the analysis of South Indian Newham subjects which include the T2D cases (n= 118) and controls (n= 67).

Table 7.i: Genotyping results- p-values all SNPs Newham SI case-control cohort

South Indian Newham	rs numbers	Allele	p-values	Case Freq	Con Freq
1	rs1805165	G	0.42	0.410	0.453
		T		0.590	0.547
2	rs867529	C	0.55	0.587	0.555
		G		0.413	0.445
3	rs6750998	A	0.13	0.616	0.695
		T		0.384	0.305
4	rs7579242	C	0.12	0.057	0.023
		T		0.943	0.977
5	rs1805164	A	0.74	0.898	0.91
		G		0.102	0.09
6	rs7605713	A	0.11	0.633	0.714
		C		0.377	0.286
7	rs867014	A	0.69	0.896	0.883
		C		0.104	0.117

The table 7.i shows the reference SNPs in 2nd column, the alleles studied in the 3rd column, and the p- values generated using log likelihood ratios. The results also show the frequency of the alleles in the cases and controls respectively. The results are for the 7 markers that passed QC/ HWE. None of the SNPs showed any statistically significant difference between cases and controls in South Indian Newham case-control cohort.

South Indian Urban survey

The following table (Table 7.j) summarizes the results and the p-values generated as a measure of significance of the association of the marker, i.e. the SNP, with being a case (Type 2 diabetes) in the analysis of South Indian Urban survey subjects which include the T2D cases (n= 85) and controls (n= 260).

Table 7.j: Genotyping results- p-values all SNPs South Indian Urban survey

South Indian Urban survey	**rs numbers**	**Allele**	**p-values**	**Case Freq**	**Con Freq**
1	rs1805165	G	0.25	0.388	0.446
		T		0.612	0.554
2	rs867529	C	0.16	0.625	0.556
		G		0.375	0.444
3	rs6750998	A	0.90	0.694	0.688
		T		0.306	0.312
4	rs7579242	C	0.12	0.016	0.042
		T		0.984	0.958
5	rs1805164	A	0.01	0.771	0.863
		G		0.229	0.137
6	rs7605713	A	0.82	0.683	0.694
		C		0.317	0.306
7	rs867014	A	0.04	0.779	0.856
		C		0.221	0.144

The table 7.j shows the reference SNPs in 2nd column, the alleles studied in the 3rd column, and the p- values generated using log likelihood ratios. The results also show the frequency of the alleles in the cases and controls respectively. The results are for the 7 markers that passed QC/ HWE. rs1805164 and rs867014 showed a statistically significant difference between cases and controls in the South Indian Urban survey cohort but the result should be interpreted with caution for this sub-group analysis given the small number of cases (n= 85) and disproportionately larger number of controls (n=260) whereas there was no difference noted in the combined SI analysis.

South Indian families

The results in table 7.k show the analysis of the South Indian family study (n=104 families, 468 subjects) analysed separately using the log likelihood ratio test. Three markers failed HWE leaving 7 markers which could be analysed. And the p- values are shown in the following **table 7.k**. None of the markers showed any statistically significant association.

Table 7.k: Genotyping results – p-values all SNPs, SI families

South Indian families	rs numbers	p- values
1	rs1805165	0.59
2	rs867529	0.54
3	rs6750998	0.26
4	rs7579242	0.43
5	rs1805164	0.71
6	rs7605713	0.26
7	rs867014	0.67

The table 7.k shows the reference SNPs in 2nd column, and the p- values generated using log likelihood ratios. None of the SNPs showed any statistically significant association in South Indian family study.

Disease association statistics- Genotyping results- p-values

In the novel SNP in the 5' of *EIF2AK3* gene (rs7605713) an equivocal p-value was obtained that was individually significant in the British Irish subjects in the combined analysis and Warren 2 trios and also in the Bangladeshi cohort in the combined analysis. In the Bangladeshi young diabetes cohort there was a statistically significant p-value for SNPs rs7605713 and rs6750998 (Intron 10 of *EIF2AK3* gene). SNP rs1805164 (Exon 11 of *EIF2AK3* gene) and SNP rs867014 (Intron 1 of *EIF2AK3* gene) showed a nominally significant p-value in the South Indian urban survey. Further replication study was carried out for rs7605713 in the British population.

Dundee Replication Study

The following table (Table 7.l) summarizes the results and the p-values generated as a measure of significance of the association of the marker, i.e. the SNP, with being a case (Type 2 diabetes) using the log likelihood ratio test. SNP rs7605713 which gave an equivocal p-value in the Warren 2 trios family study, combined British Irish cohort further replication study was undertaken in 3334 cases and 3456 controls for rs7605713 in the DARTS cohort (Table 7.l) which are all Scottish and therefore similar to the Warren 2 cohorts.

Table 7.l: Genotyping results – p-values, Dundee Replication study

Allele (rs7605713)	p- values	Case Frequency	Control Frequency
A	0.56	0.709	0.705
C		0.291	0.295

There was no statistically significant difference between cases and controls for the SNP rs7605713 in the Dundee replication study.

7.2 LINKAGE DISEQUILIBRIUM STATISTICS

Linkage disequilibrium statistics measuring marker- marker association for the *EIF2AK3* gene calculated by a likelihood ratio test are shown in the LD plot in **Figure 7.a.** and results are presented in the **Table 7.m.**

Figure 7.a: LD plot for 10 SNPs within and flanking the *EIF2AK3* gene

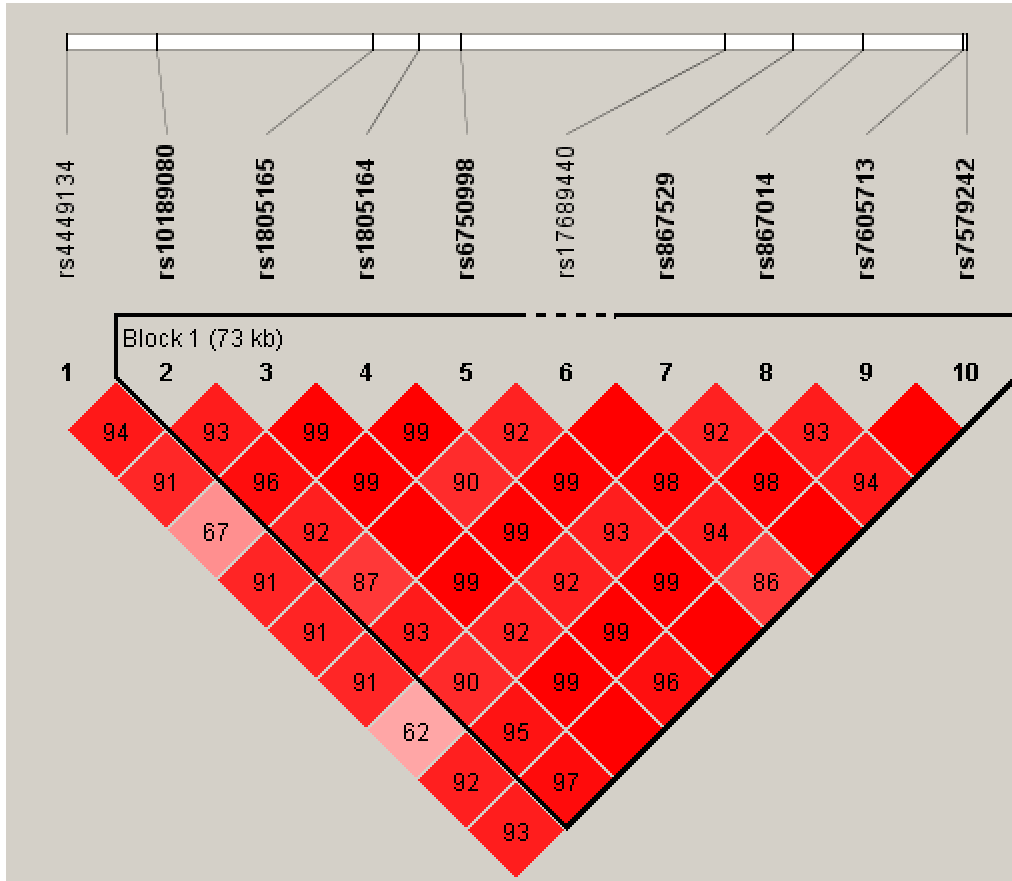


Figure 7.a: LD plot for linkage disequilibrium for the 10 SNPs of *EIF2AK3* gene are being shown. The LD plot was obtained using the program Haploview(138).The colour schemata of the plots highlighted in red colour with D' values of >0.8 indicate SNPs in strong LD with each other. The 5 empty red plots are SNPs in perfect LD with each other with D' values of ~1. The plots in pink colour with D' values of <0.8 indicate SNPs not in strong LD with each other.

Table 7.m: The table in the previous page displays the LD for all 10 SNPs within and flaking the *EIF2AK3* gene. D' calculated by LDPAIRS below diagonal. P-values calculated by LDPAIRS above diagonal.

In order to measure marker-marker association we calculated a likelihood ratio test comparing the likelihood for the genotypes assuming LD to that assuming no LD, as implemented in the LDPAIRS program. The p values measure the statistical significance of the test for linkage disequilibrium between the markers.

D' of 1 equals complete LD. D' of >0.8 indicates strong LD. D' of 0.2- 0.8 indicates incomplete LD. D' of <0.2 indicates negligible LD(139). The results show rs7579242 to be in complete LD with rs1805165, rs867529 and rs6750998. rs17689440 is in complete LD with rs867529. rs7605713 is in complete LD with rs7579242. rs4449134 and rs1805164 are in incomplete LD. rs4449134 and rs867014 are in incomplete LD. All of the other SNPs, except rs4449134, rs1805164 and rs867014, genotyped in our study are in strong LD with each other with a D' of >0.8.

CHAPTER EIGHT: DISCUSSION

Type 2 Diabetes is a quintessentially complex disease with significant gaps in our knowledge in understanding the pathogenesis of the condition. It is well recognised that insulin resistance and insulin secretory failure occur in T2DM, but what is not clear is the underlying molecular mechanisms that lead to these defects. The incidence of T2DM reaching epidemic proportions in recent times does support the underlying concept that calorie excess and sedentary lifestyle have had a major impact on the disease. Successful discovery of genes that contribute to T2DM based on candidate gene testing have been very few and the risk that is associated with these gene variants low (odds ratio, OR ~1.3). There has been remarkable progress made in recent years in genetic studies of polygenic diseases that has led to the discovery of new loci/ genes associated with T2DM using a genome wide approach but again the risk associated with these loci are low (OR 1.1-1.7). Indeed most of the familial risk of T2DM still remains unexplained, highlighting the need for ongoing efforts to detect novel disease associated variants.

8.1 Paradigm shift in the study of genetics of common diseases

Since starting work in the laboratory in 2002 to the present day there has been a major shift in approaches to identifying the susceptibility genes for type 2 diabetes. It spans the era of:

- Investigating **candidate genes** in small subsets (Studying a few hundred subjects in a case-control design and family based candidate gene association study) (140) of individuals which resulted in false positive associations that failed future replication studies (141,142). The problem of failed association studies, either positive or negative signals and the failed replication studies could be attributable to low power to detect small effect, i.e. low odds ratios. There was also the problem of having low threshold for suggesting a significant p-value, leading to wrong interpretation of results and false positive associations. Finally, there was

the problem of investigating only a few variants across the gene, i.e. SNPs, which led to inadequate capture of gene variation(143).

- **Low throughput genotyping** which was labour intensive and a relatively slow process involving carrying out PCR reaction, post- PCR sample handling and running the DNA product on agarose gel, as compared (144) to **high throughput genotyping** (145) facilitating study of large number of samples for association studies in a quick and accurate manner and not requiring post- PCR sample handling thereby avoiding potential PCR product contamination leading to wrong genotype data generation.

- Testing **single candidate gene variants** with a few SNPs to the now popular **Genome wide association studies** with ~ 0.5 – 1 million SNPs spanning the whole genome with the capability to detect novel genes associated with disease.

- Studying **single SNPs** to compiling SNP inventories by direct sequencing and the use of publicly available in silico databases as a consequence of the International Hapmap project and the Human Genome project(146,147).

In the **year 2002** facilities existed for doing a large scale candidate gene association study but we did not have the tools to investigate the whole genome for association study. We tried to address some of the issues pertaining to power to detect modest effects. However, this was only possible in the British Irish cohort. We did a robust large scale association study of the *EIF2AK3* gene, with thousands of individuals studied, resulting in more reliable results dependent on sufficient power to detect the relevant effect. We utilised **TagSNP** information from the International Hapmap project to capture the majority of variations across the gene.

8.2 Common disease common variants – Genome wide association studies

In the era of Genome wide association studies and advances in high throughput genotyping platforms identifying disease susceptibility genes, would mean that new genes incriminated in type 2 diabetes are being discovered at a rapid pace. Replication studies would further support or refute such candidate genes being associated with type 2 Diabetes. The power or the numbers needed to study to detect gene variants associated with a complex disease such as T2D, in GWAS depends on the effect size and the risk allele frequency. The GWAS design of gene association studies in polygenic disease like T2DM has been to identify common variants with minor allele Frequency > 5% that could be associated with disease, i.e. common disease common variant hypothesis(148). The common disease-common variant (CDCV) (**C3 in the figure 8.a**) hypothesis infers that common variants, interacting disease alleles are associated with most common diseases. The hypothesis has been the scientific paradigm for genome-wide association (GWA) studies that have been conducted on many common diseases. As an example in GWAS the detection of the vast majority of the effect sizes of risk alleles are small, typically OR around 1.1 to 1.5, which are the limits of detection given the experimental sample sizes employed to date. GWAS SNP panels are limited to polymorphisms with MAF >0.01. Rare causal variants are likely to exist but will not be represented by the genotyped SNPs in the current GWAS design.

During the process of writing up my thesis there have been GWAS and also GWA meta analysis (149) which combined the results across a total of 10,128 samples (4,549 cases and 5,579 controls (see **table 8.a** for the gene variants associated with T2DM) but still *EIF2AK3* did not come up in the Genome Wide Association studies. However, as will be discussed later in this chapter, *EIF2AK3* was a potentially good candidate gene to be studied in Type 2 diabetes and our study was hypothesized pre-GWAS. **GWAS** have identified several T2DM susceptibility loci and accumulating evidence suggests that these loci confer the diabetes risk through their **effects on beta cell function.**

Table 8.a: GWAS and genetic variants associated with T2DM

SNP	Nearest Gene	Odds ratio	p- value
rs864745	*JAZF1*	1.10	5.0×10^{-14}
rs12779790	*CDC123, CAMK1D*	1.11	1.2×10^{-12}
rs7961581	*TSPAN8, LGR5*	1.09	1.1×10^{-09}
rs7578597	*THADA*	1.15	1.1×10^{-09}
rs4607103	*ADAMTS9*	1.09	1.2×10^{-08}
rs10923931	*NOTCH2*	1.13	4.1×10^{-08}
rs1153188	*DCD*	1.08	1.8×10^{-07}
rs17036101	*SYN2, PPARG*	1.15	2.0×10^{-07}
rs2641348	*ADAM30*	1.10	4.0×10^{-07}
rs9472138	*VEGFA*	1.06	4.0×10^{-06}
rs10490072	*BCL11A*	1.05	1.0×10^{-04}
rs13266634	*SLC30A8*	1.12	5.3×10^{-08}
rs1111875	*HHEX*	1.13	5.7×10^{-10}
rs8050136	*FTO*	1.17	1×10^{-12}
rs7903146	*TCF7L2*	1.37	1×10^{-48}
rs1801282	*PPARG*	1.19	1.5×10^{-7}
rs10010131	*WFS1*	1.15	4.5×10^{-5}

The GWAS findings are consistent with pathophysiology of T2DM occurring due to beta cell insulin secretory failure(150,151). The majority of the genes identified at various loci were previously unknown to be associated with T2D and they were not even considered as potential candidate genes. The finding of these novel genes for T2DM susceptibility raises the possibility of exploring the biological pathways by which disease may manifest. For example, common variants in SLC30A8 and HHEX were discovered as novel T2DM association genes(152). The *SCL30A8* gene encodes a zinc transporter and is expressed in pancreatic beta cells(153). The *HHEX* gene encodes a transcription factor involved in early pancreatic development(154). The candidacy was confirmed for *TCF7L2* gene in GWAS which was already identified as strong T2DM candidate gene (Linkage studies and Candidate gene association study).

The results from GWAS do not however mean that the **insulin resistance genes** are not there to be found. The two genes identified which are conferring diabetes risk through insulin action are the *PPARG and FTO* gene. In a GWAS in U.K population *FTO* gene has been identified associated with BMI and obesity from childhood to old age that predisposed to T2D through an effect on BMI. The adults who have the homozygous risk allele (Allele A of rs9939609) of *FTO* gene region on chromosome16, weighed 3 kilograms more and had 1.67 fold increased odds as compared to those who did not have the risk allele and OR of 1.27 for T2D(155). In the Finnish GWAS rs8050136 of *FTO* gene giving an odds ratio of 1.17 per at risk allele was identified(156). The reason for fewer number of new variants of insulin resistance genes found in GWAS as opposed to insulin secretion genes could be potentially multi-factorial- design of the study where subjects are matched for BMI or have lean T2D cases, modest affect sizes of at risk alleles leading to loss of power, potentially fewer variants that affect insulin resistance and potential for strong environmental influences in obesity and metabolic syndrome phenotypes(157). This would imply the critical importance of design of the GWAS takes BMI/ Obese individuals into account to improve the chances of detecting insulin resistance genes.

Candidate gene association study vs. GWAS

There are several possible explanations why candidate gene studies may identify genes missed by GWA studies. Most importantly, depending on the genotyping platform (affymetrix, illumina, perlegen) and SNP density (500k-1 million) there may be poor coverage of the relevant candidate gene. This is illustrated by the fact that *SLC30A3* was not identified as a result of WTCC GWA but was identified once tag SNPs were added to identify the gene in another GWA scan on a different population (158). Candidate gene association studies use more densely spaced SNPs to target the gene in question. Recent candidate gene association studies have included more participants and are therefore more likely to uncover lower risk alleles. Finally, in a GWA study the results are corrected for the large number of tests performed, and the threshold p-value for positive associations is lower, leaving room for more type 2 error. Furthermore, Genome wide association studies cannot take into account the possibility of rare variants contributing to the disease risk- common disease rare variants hypothesis.

8.3 Common disease rare variant hypothesis

One emerging hypothesis is that low frequency variants with intermediate penetrance effects (**B2 in figure 8.a**) could be associated with common disease(159). Most associated variants that have been detected to date are common (minor allele frequency, MAF > 5%), although this is mostly a reflection of study design in terms of SNPs selected for genotyping and power of the sample size(160). If susceptibility alleles have minor allele frequencies (MAFs) of less than 0.1 i.e. rare, then large sample sizes of greater than 10,000 cases and 10,000 controls (or 10,000 families) would be required to achieve convincing statistical support for a disease association(161). The exact size of the study cohorts would depend on power calculations but the expectation is that these low frequency variants would be associated with a high risk predisposition compared to those recently identified by GWA. Thereby to achieve adequate power at even lower MAFs with the same effect sizes we are talking about sample sizes in excess of 20,000 individuals.

Figure 8.a: Common variants- common disease / rare variants- common disease

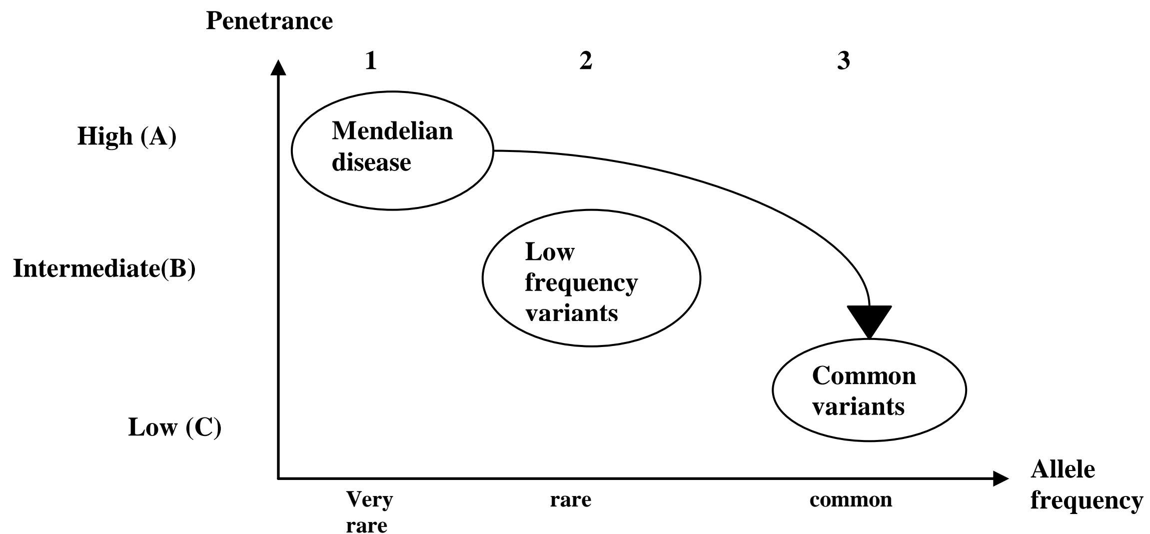


Figure 8.a: Common disease common variants (3C), Common disease low frequency variants (2B), and Mendelian disease very rare variants (1A)

8.4 Monogenic disease – common disease

The alternative approach is the study of Mendelian disease, such as Wolcott- Rallison Syndrome, and once a causal gene has been identified **(A1 in figure 8.a)** to follow through these findings to see if common variants of the gene is associated with the polygenic disease - T2DM, the common disease common variant hypothesis **(C3 in figure 8.a)**. The study of monogenic disease to identify polygenic disease has had its success, for example- *HNF 1a/TCF-1, SUR 1/Kir6.2 E23K, Wolfram syndrome/WFS 1, MC4R* receptor genes and its variants and their association with common type 2 diabetes. Rare mutations of these genes lead to monogenic forms of diabetes.

Mutations of *transcription factor 1 (TCF1, or hepatocyte nuclear factor-1α)* affect β-cell function. Common variants in the *HNF-1α* gene influences transcriptional activity and insulin secretion in vivo. These variants have been shown to be associated with a modestly increased risk of type 2 diabetes in elderly overweight individuals (162).
In pancreatic *β*-cells ATP-sensitive potassium (K^+ATP) channels control insulin secretion by coupling metabolism to changes in electrical activity. The K^+ATP channels are composed of two subunits, the sulfonylurea receptor *(SUR1 or ABCC8)* and the inwardly rectifying potassium channel *Kir 6.2*. Mutations in the *SUR1* gene cause decreased K^+ ATP channel activity and congenital hyperinsulinemia. E23K variant in the *Kir 6.2* gene has been shown in several studies in various ethnic groups to be associated with type 2 diabetes (163).

Mutations in the *WFS1* gene cause β-cell death, resulting in a monogenic form of diabetes known as Wolfram syndrome. The role of variation in *WFS1* in type 2 diabetes susceptibility was investigated in various ethnic populations and data provide strong evidence that variation in *WFS1*, a gene with an essential role in the endoplasmic reticulum stress response in insulin-producing pancreatic β- cells, contributes to risk of common type 2 diabetes (164).

Mutations in the *melanocortin 4 receptor gene (MC4R)* are a common cause of monogenic human obesity. The common SNP rs17782313 near *MC4R* gene was significantly associated with higher intakes of total energy and dietary fat and was shown to be related to greater long-term weight change and increased risk of type 2 diabetes in women. (165)

The studies into monogenic syndromes and the gene mutations and their positive associations seen with the common type 2 diabetes underlies the fact that such studies could bear fruitful results and could potentially invoke a new aetiopathogenesis for type 2 diabetes. This hypothesis would argue that common and rare variants of genes that lead to monogenic disease should be studied as plausible candidate genes in the common form of T2DM.

8.5 *EIF2AK3* gene- polygenic Type 2 Diabetes

When the study was envisaged in the yea2002, the best approach for performing candidate gene association study was to identify a suitable candidate gene which had strong a priori credentials to be associated with type 2 diabetes. The candidate gene was selected on the basis of prior linkage evidence, biological plausibility and mechanistic model of action of the gene product- protein, animal model with underlying gene defect and a monogenic model. However, as explained in the previous chapters "a priori", even in the era of GWAS, there are good reasons to investigate the candidacy of *EIF2AK3* as a cause of non-auto immune diabetes. In our study we selected the WRS gene- *EIF2AK3,* which satisfied most of the criteria and appeared to be a promising gene to investigate for candidacy for type 2 diabetes. *EIF2AK3* was a very attractive candidate gene to test given its role in Walcott- Rallison syndrome (monogenic form of insulin dependent diabetes), its role in attenuating endoplasmic reticulum stress in the pancreas, animal model of diabetes in *EIF2AK3/PEK* deficient mice and preliminary association data on the South Indian and UK Warren 2 type 2 diabetes patients; although it is acknowledged that the preliminary study was grossly underpowered.

EIF2AK3 gene- Wolcott- Rallison syndrome

Wolcott- Rallison syndrome (WRS) is a rare autosomal recessively inherited disorder. The syndrome, as previously explained, is characterised by a short trunk compared to arm span, multiple epiphyseal dysplasia, bone demineralisation, multiple fractures, tooth discolouration, abnormal skin, hepatosplenomegaly and renal insufficiency in addition to pancreatic exocrine insufficiency and non autoimmune insulin dependent diabetes.

Aetio-pathogenesis of type2 diabetes and the role of ER stress

Type 2 Diabetes is characterized by the insulin resistance in the liver, muscle and adipose tissue. In the initial stages of the disease there is relative hyperinsulinaemia to compensate for the insulin resistance. But, there is progressive β cell loss necessitating escalation of diabetes treatment and frequently needing insulin treatment in type 2 diabetes to achieve normoglycaemia. This process highlights the β cell failure in the

development and progression of Type 2 diabetes(166). The decreased β cell mass is due to increased β cell apoptosis and decreased β cell replication and regeneration. This was shown in the post mortem studies of pancreatic specimens of T2D patients(167). Glucotoxicity and lipotoxicity leading to oxidative stress causes pancreatic β cell dysfunction. This glucolipotoxicity is a feature of both obese and lean T2D phenotype individuals. However, glucotoxicity is proposed as a prerequisite for β cell lipotoxicity (168). ER stress pathways have been proposed as potential route involved in pancreatic β cell apoptosis. ER stress response also known as unfolded protein response (UPR) can be trigger of pancreatic β cell dysfunction and apoptosis under situation of chronic stress (High glucose and high free fatty acid exposure). ER is a dynamic organelle involved in protein (insulin) biosynthesis in the pancreatic islets. ER stress due to lipotoxicity and its role in pancreatic dysfunction has been shown in in-vitro experiment. Palmitate and oleate caused apoptosis in the well differentiated insulin producing INS-1E cells by triggering ER stress response(169). ER has an important role in calcium storage and signalling. Any disruption of the ER function potentially has a deleterious effect on pancreatic insulin secretory function. In the PERK knock out mice due to the ER stress there is disruption of ca^{2+} signalling complexes and ca^{2+} regulation in pancreatic cells(170).

The ER in pancreatic β cells has a major function of synthesis of insulin, folding and processing of the protein. There are many ER stress signals as discussed previously. The response to ER stress is called the unfolded protein response to protect the β cells. *PERK*/ *EIF2AK3* has a central role to play in UPR. *EIF2AK3* enriched in pancreatic cells, is uniquely located to the endoplasmic reticulum and is activated by the accumulation of unfolded proteins in the ER lumen. *EIF2AK3* phosphorylates eukaryotic initiation factor-2 alpha (eIF2 α) which then inhibits eIF2 β, leading to cellular events that inhibit mRNA translation. Downregulation of protein synthesis occurs by phosphorylation of the alpha subunit of eukaryotic translation initiation factor-2 (eIF2-alpha). This is one of the fundamental responses to ER stress and prevents further β cell damage by apoptosis(171). Targeted mutation of the mouse *EIF2AK3* gene abolishes phosphorylation of eIF2 α, and the resultant phenotype in mice homozygous for the mutation is characterised by non-

autoimmune insulin deficiency, hyperglycaemia, growth retardation, steatorrhoea and high mortality between 2 and 4 weeks of life and heterozygous mice have impaired glucose tolerance.

The candidate gene association study was carried out with a plan to have a large sample size (total study size = 8095 subjects) to have adequate power to either implicate or refute *EIF2AK3* as a candidate gene for T2DM, as it is accepted in genetic research into polygenic diseases that small sample sizes looking into SNP polymorphisms and associations with disease can lead to mostly false positive results. We did a robust association study for T2DM looking into a plausible gene *EIF2AK3* and had a good admixture of Caucasian and South Asian resources. The Caucasian resources were the largest cohort with 6112 subjects.

In T1DM study there was evidence for an association between T1DM and particular *EIF2AK3* marker combinations (overall global $P<0.001$), with excess transmission of both the 15INDEL allele 1/D2S1786 allele 4 and 15INDEL allele 1/D2S1786 allele 3 combination (both $P<0.0001$), and decreased transmission of 15INDEL allele 1/D2S1786 allele 5 combination ($P<0.001$) in South Indian families. Preliminary T2DM studies in UK Warren 2 trios (n = 150) showed marked deviation from expected transmission for haplotypes for SNPs 2, 10, 11 and 13 of the *EIF2AK3* gene. These preliminary results however were not adequately powered and only a few SNPs across the *EIF2AK3* gene were investigated.

We carried out a case-control study as well as a family based association study. A detailed SNP inventory of *EIF2AK3* was compiled in the British Irish and the South Asian subjects and 19 SNPs were identified. Four SNPs were not found in the public databases. Four markers were subsequently not found to be polymorphic. Our final analysis included three exonic and three intronic SNPs within the *EIF2AK3* gene and four SNPs in the 5' and 3' of the *EIF2AK3* gene.

8.6 Genotyping Quality control issues

Quality control is critical to avoid pitfalls leading to false interpretation of the genotyping results. In our study we had to undertake a large amount of work for quality control purposes. The reason for the QC checks is to obtain reliable and clean data and to eliminate the likelihood of generating false positive or false negative associations. When processing large number of samples these problems would be magnified and is therefore especially important to pay attention to Quality control. The stringent criteria set for QC pass enabled us to identify the data sets that could be identified for further analysis as discussed in chapter six. In the year 2002, less attention was focussed on QC issues. This led to false positive associations which failed future replication studies(172). There have been standards recommended by leading experts in the complex disease genetics field to address QC issues(173). These standards have been followed in my work for quality control checks.

The DNA samples from ECACC and 1958 MRC birth cohort were of good quality and easily passed QC- the genotyping was performed at Genome Centre by taqman methodology. However, there were a lot of QC problems with the results generated from Warren 2 trios by taqman methodology as was thoroughly discussed in chapter six. The conclusion from the experiments carried out was that the wells were lacking DNA-W2trios DNA samples were either dried up or were inadequately aliquoted in the first place. The DNA resources that were available at KBiosciences were of better quality and all the 10 SNPs that were analysed by Amplifuor technology passed QC according to their internal quality control measures, except INDEL15, which also failed by taqman previously. Indeed the majority that was successfully genotyped were those DNA that were available at KBiosciences.

The process of QC checks and the results obtained highlights the importance of 1) good quality DNA to be available for genotyping and 2) robust quality control measures in genotyping and data analysis phase to give reliable data.

8.7 Sequencing

Sequencing of the *EIF2AK3* gene together with 1800bp 5' and 3000 base pairs 3' was performed on ABI 3100 capillary sequencer utilising Big Dye terminator chemistry. Direct sequencing of all 17 exons and intron/ exon boundaries was performed on 48 individuals (29 T2DM cases, 2 possible WRS cases and 17 controls- subjects were chosen from Chennai WRS family, British/ Irish T2DM and controls, South Asian T2DM and controls). 19 SNPs were identified by direct sequencing. Four of these SNPs were monomorhic SNP (i.e. only one form of allele present in the population of interest). Previously identified SNPs (174) detected in *EIF2AK3* gene were also detected in our study. No mutations were detected that would be predicted to contribute to T2DM. Four of these SNPs detected were novel and were further studied in the candidate gene association study in both the British Irish and the South Asian cohorts. The novel SNPs rs7579242 and rs7605713 were located 5' of *EIF2AK3* gene and the novel SNPs rs10189080 and rs4449134 were located 3' of *EIF2AK3* gene. We also utilised two TagSNPs within the *EIF2AK3* gene -rs867014 and rs17689440- from the HapMap2 project for the purpose of *EIF2AK3* gene candidate gene study to give extensive coverage of SNPs across the whole gene.

8.8 Minor allele frequency- Population characteristics

Genetic heterogeneity is a feature which has particular relevance when it comes to doing candidate gene association study in different ethnic groups. There could be differences in the minor allele frequencies among different ethnic groups. This difference in MAFs will have a bearing on the power of the study to detect association with the disease. The analysis showed several of the SNPs had different minor allele frequencies between the British/Irish Caucasians as compared to the South Asians (see **table 6.e** in results section on MAFs differences). However, comparisons between Bangladeshi and South Indians showed very similar minor allele frequencies despite the genetic heterogeneity among these groups. Bangladeshis are of Aryan origin with genetic structure more closely related to Caucasians whereas the South Indians are of Dravidian origin with genetic

structure more closely related to the Africans. This might suggest that for genetic studies of these markers for the WRS gene in the South Asian populations it may not be necessary to study the gene in different ethnic groups within the Indian sub-continent, although it should be borne in mind that culturally there are differences between South Indians and Bangladeshis. Therefore the environmental impact on a disease like T2DM may be different in these groups. Furthermore, in a recent extensive population study in India differences in allele frequency have been found at the genetic level in people from different regions within India(175).

The minor allele frequencies obtained in the British Irish cohort in our study compared very favourably with the MAFs quoted in the public databases (HAPMAP CEU). This comparison adds further support to the quality control measures that were undertaken in our study. Since the MAFs for some of the SNPs were different in the South Asian population- it underlies the need for caution in using MAFs in HAPMAP for the South Asian population unless they are independently studied and quoted in HAPMAP.

8.9 Hardy- Weinberg equilibrium results

The principle of Hardy- Weinberg equilibrium is that in a randomly mating population the genotype frequencies remain constant from generation to generation. HWE testing was carried out on all datasets after the genotyping phase. Three reference SNPs failed HWE among the three populations studied (British Irish, Bangladeshi, and South Indians). The SNP rs4449134 was not in HWE in all three populations and therefore genotyping analysis was not undertaken for this marker. The SNP rs10189080 was not in HWE in the Bangladeshi and South Indian population. Finally, rs17689440 was not in HWE in the South Indian population. The reason for the deviation from HWE was an excess of homozygote rare alleles as compared to the expected number (see **table 6.f** in HWE failures) in the studied subjects.

8.10 Genotyping results on *EIF2AK3* gene

Single nucleotide polymorphisms within the *EIF2AK3* gene have not been shown to be associated with type 2 diabetes previously. Our effort was to investigate the *EIF2AK3* gene for a candidate gene association study in T2DM in a case-control and family based association study. Genotyping was performed at two different sites- one at Genome centre in London and the other at KBioSciences. The technology used was the best available at the time the study was conducted for a large scale candidate gene association study. Taqman technology was utilised at Genome centre and Amplifluor/ KASPAR technology was utilised at KBioSciences. These technologies facilitated analysis of large sets of samples and provided us with robust data for the genotyping analysis. Genotyping analysis was performed in-house in ICMS London using the programme UNPHASED Log likelihood ratio tests. Statistical analysis was supported by Dr. Anna Vine under supervision from Professor David Curtis.
(Website address: www.wolfson.qmul.ac.uk/psychiatry/staff/curtis_d.html)

In order to measure marker-marker association of the *EIF2AK3* gene we calculated a likelihood ratio test comparing the likelihood for the genotypes assuming LD to that assuming no LD, as implemented in the LDPAIRS program. The results showed the novel SNP rs7579242 to be in complete LD with rs1805165, rs867529 and rs6750998. rs17689440 is in complete LD with rs867529. rs7605713 is in complete LD with rs7579242. SNPs rs4449134 and rs1805164 are in incomplete LD. SNPs rs4449134 and rs867014 are in incomplete LD. All of the other SNPs, except SNPs rs4449134, rs1805164 and rs867014, genotyped in our study are in strong LD with each other with a D' of >0.8. These results highlight the fact that we could have reached similar conclusions by genotyping a few of the markers without genotyping all the markers that were in strong and complete LD. The aforementioned results were not anticipated prior to the LD analysis phase. There were four SNPs which were novel and we did not have LD structure in Hapmap. Furthermore, there is no LD information for the South Asian subjects and the information to this date is restricted to Guajarati Indians living in America. Therefore, we took the decision to genotype all the ten SNPs in the *EIF2AK3*

gene candidate gene association study in all the subjects. In retrospect it could be argued at least in the British Irish population we should have restricted the markers studied, since LD relationships in six of the markers is found in Hapmap.

Allele frequencies and the p-values generated using UNPHASED 3.0.6 for all SNP markers of the *EIF2AK3* gene are shown in Tables 7.a. to 7.k.

The statistical results generated from our study for SNP rs7605713 (novel 5' of *EIF2AK3* SNP) in the British Irish combined analysis and in the Warren 2 trios showed a nominally significant p-value. A significant increase in the major allele of the SNP rs7605713, allele A, was seen in subjects with type 2 diabetes in the Warren 2 trios and British Irish combined genotyping analysis (P= 0.05). The rest of the SNPs showed no association between any of the reference SNPs/ markers of the *EIF2AK3* gene and type 2 diabetes in the British Irish cohort.

The statistical results generated from our study for SNP rs7605713 (novel 5' of *EIF2AK3* SNP) in the Bangladeshi combined analysis and young Bangladeshi case-control study showed a nominally significant p-value. A significant increase in the minor allele of the SNP rs7605713, allele C, was seen in subjects with type 2 diabetes in the Bangladeshi combined analysis (P= 0.05) and subjects with type 2 diabetes in the young Bangladeshi case-control study (P= 0.007).

The statistical results generated from our study for SNP rs6750998 (Intron 10 SNP of *EIF2AK3*) in the Bangladeshi combined analysis and the young Bangladeshi case-control study showed a nominally significant p-value. A significant increase in the minor allele of the SNP rs6750998, allele T, was seen in subjects with type 2 diabetes in the Bangladeshi combined analysis (P= 0.04) and subjects with type 2 diabetes in the young Bangladeshi case-control study (P= 0.005).

It is pertinent to note that the young Bangladeshi cohort was small (330 cases and 374 controls). The patients with diabetes were lean with a mean BMI ~ 18. This is not the

typical phenotype of T2DM. Whether the SNPs- rs6750998 and rs7605713 are really showing a positive signal needs further study in a larger cohort with the same phenotype. Our study did not have adequate power for the young Bangladeshi cohort to arrive at a firm conclusion. To avoid false positives large numbers of subjects need to be studied (at least nearly 4,500 individuals) in candidate gene studies. Moreover it is important to have an independent replication cohort.

The statistical results generated from our study showed no association between any of the reference SNPs/ markers and type 2 diabetes in the South Indian cohorts, even though SNP rs867014 (Intron 1 SNP of *EIF2AK3*) and SNP rs1805164 (Exon 11 SNP of *EIF2AK3*) showed a nominally significant p-value in the South Indian urban survey. A significant increase in the minor allele of rs867014, allele C, was seen in subjects with type 2 diabetes in the South Indian urban survey (P= 0.04). A significant increase in the minor allele of rs1805164, allele G, was seen in subjects with type 2 diabetes in the South Indian urban survey (P= 0.01). The results should be interpreted with caution for this sub-group analysis given the small number of cases (n= 85) and disproportionately larger number of controls (n=260) whereas there was no difference noted in the combined South Indian analysis.

In the one marker, rs7605713, that showed a nominal significance in the British Irish combined analysis and Warren 2 trios family study, further replication studies in a similar population group, the Dundee Wellcome Trust UK T2DM Genetic consortium Case-control cohort (3334 diabetes cases and 3456 controls) proved to be negative (P= 0.5) thereby avoiding a false positive association. This replication work is an important step in a robust candidate gene association study.

We therefore concluded in our candidate gene association study that no association exists between common variants of *EIF2AK3* gene and type 2 diabetes.

The following SNPs that were studied in our group- rs867529, rs1805164, rs1805165, rs867014- were also independently studied by Geneticists in Cambridge, U.K (176) in a case-control study (9533 cases and 11389 controls) and they found no association with for these SNPs with type 2 diabetes. All three SNPs are in tight LD (D' >0.9) with rs7605713. This would further support the likelihood the association between rs 7605713 and T2DM in the original British Irish cohort in our study was a false positive association.

Rare mutations of *EIF2AK3* and association with type 2 diabetes have not been excluded in our case-control and family study. However no coding variants that would lead to an amino acid change or regulatory region variants were identified in sequencing work of 48 subjects in our study including a South Indian family with phenotype for WRS. Our findings are further supported by a French study of young type 2 diabetes individuals (30 T2DM cases) (177) in whom direct sequencing of 17 exons and intron- exon boundaries of *EIF2AK3* did not detect any mutations in the cases. The results of these two datasets and experiments taken as a whole make it very unlikely that the common disease rare variant hypothesis would apply to the *EIF2AK3* gene. But this can by no means be ruled out. In order to exclude the common disease rare variant hypothesis, at least 1000 people would need sequencing(178). This approach will be made substantially easier once the 1000 Genomes project is completed. 1000 Genomes project (web link: www.1000genomes.org) is an international research effort to establish the most detailed catalogue of human genetic variation by sequencing approximately 1200 individuals across the globe over the next few years.

It is likely that the association we described previously between T1DM and markers D2S1786 and INDEL15/rs1805181 of *EIF2AK3* gene (179) is a false positive result (Type 1 error) likely related to the small sample size thereby once again reinforcing the need to conduct a study on a large scale with large number of individuals. Furthermore, if the Bonferroni correction set the threshold of significance at p= 0.05/n to avoid errors of association, where n is the number of markers studied, statistical significance would not have been observed.

Despite having very good coverage of markers throughout the *EIF2AK3* gene which were in tight LD with each other none of our markers tagged INDEL15/rs1805185. Ideally we would have wanted to identify a good tag SNP for rs1805185 to complete the analysis but searching the Hapmap database (Web link: www.hapmap.org) no appropriate marker exists.

In view of the overwhelming negative *EIF2AK3* candidate gene association study in T2DM there seemed little point pursuing any functional aspects of *EIF2AK3* gene that were originally envisaged.

8.11 Study design as of 2009-2010

There have been major advances in technology- automation and high throughput techniques, the International Hapmap project with readily available SNP inventory and Genome wide scanning approach, that if I were to do the study in the present time I would approach it in a different way

- Find evidence of a preliminary association by genotyping of Tag SNPs identified in silico and establishing minor allele frequencies in all the ethnic groups.
- Adequately powered data sets within each ethnic group.
- Technology advances for single gene SNP analysis
- Study design to account for common disease- common polymorphisms (current GWAS design) vs. common disease- rare polymorphisms (Future study designs).

Tag SNPs

Tagging SNPs exploit the extensive linkage disequilibrium that is present in the genome. Genotyping a few variants with extensive linkage disequilibrium resulting in extensive coverage across the whole gene- *EIF2AK3* is a practical approach. We carried out this process for the identification of SNPs to be studied. Even though we have SNP information for the European cohort, tag SNP data is not currently available in the Hapmap project for the Bangladeshi and South Indian population; indeed we found different MAFs in these ethnic groups for some of the markers.

For cost effectiveness, we could do a candidate gene association study by testing the tag SNPs in the first phase. And if they were positive, we could proceed using a denser set of markers/SNPs to extensively cover the gene, especially when there is strong a priori evidence for the gene to be associated with the disease. One could argue that the tagging option is not warranted if the a priori evidence for a gene is strong and proceed with the classical design for a candidate gene association study. We had four Tag SNPs which we genotyped in our candidate gene association study.

Power

Ideally we should have done power calculations before we embarked on the candidate gene association study. Furthermore, at the time it would not have been predicted that majority of T2DM candidate genes would only have an odds ratio between 1.1- 1.3 as has been revealed by recent GWAS.

Table 8.b: Power calculations for gene association study
(Assuming 80% power and p value of 0.05)

Relative risk	Minor allele frequency	Case-control
1.1	0.1	18,493
1.2	0.1	4,888
1.3	0.1	2291
1.1	0.2	10,503
1.2	0.2	2,801
1.3	0.2	1,324
1.1	0.3	8,080
1.2	0.3	2,174
1.3	0.3	1,037

Nonetheless, the British Irish cohort for the case-control candidate gene association study comprised of 2206 cases and 2652 controls. Thereby the study had adequate power to detect a relative risk of 1.2 with 80% power and a p- value 0.05, assuming a MAF of 0.3. However, we had smaller numbers in the Bangladeshi and the South Indian populations in our study. Ideally we would have liked to have a bigger sample size for the purpose of the study. There were logistical constraints to recruit more individuals of South Asian origin. For instance, we would have needed a sample size of at least 4,348 individuals from the different ethnic groups to give us the power to identify a relative risk of 1.2 with 80% power and a p- value 0.05, assuming a MAF of 0.3 (Website- http://stat.ubc.ca/~rollin/stats/ssize/caco.html) as indicated in the **table 8.b**. As would have been predicted all the South Asian datasets in our study are therefore underpowered for an ethnic specific analysis.

Technology advances for single gene SNP analysis

Technological advances for conducting a single candidate gene association study have also taken place since the time of our study. Taqman and Amplifuor technologies would still be considered adequate at present for a study of similar size to our study (Analysis of 10-11 SNPs in a single candidate gene). However, technology exists for simultaneous genotyping of multiple SNPs (From 10 up to 48 SNPs can be multiplexed). SNaPshot Multiplex system enables multiplexing up to 10 SNPs for genotyping (180). The Snapshot chemistry is based on dideoxy single base extension of an unlabeled oligonucleotide primer. Each of the primers bind to a complementary template in the presence of fluorescently labeled ddNTPs and AmpliTaq DNA polymerase. The DNA polymerase extends the primer by one nucleotide, adding a single ddNTP to the 3' end. The generated products are analysed with GeneScan analysis software. SNPlex genotyping system enables simultaneous genotyping of up to 48 SNPs (181). SNPlex assay utilizes an oligonucleotide ligation assay followed by PCR. It uses capillary electrophoresis instrument to separate selectively amplified gene regions and identifies the eluted probes with the aid of GeneMapper software. The technologies are available from Applied Biosystems. (Web address: www.appliedbiosystems.com)

Common disease- common polymorphisms (current GWAS design) vs. common disease- rare polymorphisms (Future study designs)

The GWAS design of gene association studies in T2DM has been to identify common variants with minor allele Frequency > 5% that could be associated with disease, i.e. common disease common variant hypothesis. If as suspected susceptibility alleles have minor allele frequencies (MAFs) of less than 0.1 i.e. rare, then large sample sizes of greater than 20,00 would be required to achieve convincing statistical support for a disease association studies. The study design will therefore depend on the power calculations based on minor allele frequencies in various ethnic cohorts.

In an **'ideal world'** scenario adequately powered cross sectional cohort ethnic study design would be appropriate. Furthermore, we would have done power calculations for each ethnic population prior to genotyping to have a robust study design. We would have had access to DNA from more subjects in the South Asian cohorts to boost the power for the study. There could also be an argument made for analysing each ethnic group as one sampling frame rather than having sub group analysis as we have done in our study. We studied *EIF2AK3* as a T2DM candidate gene in isolation in the various cohorts. We could have strengthened our study to include genes that were the main regulators and effectors of the *EIF2AK3* gene. Therefore studying common variants of eukaryotic initiation factor- 2 alpha *(eIF2 α) eIF2 β,* and the ER stress pro- apoptotic downstream signaling genes which include *CHOP, IRE 1, JNK, CFOS/ JUN, IL- 1β*, would have enabled us to enquire about the complete biological system of endoplasmic reticulum stress and type 2 diabetes rather than testing an individual component in the *EIF2AK3* gene. Furthermore, if we were to argue that the gene variants could be linked to pancreatic insufficiency in subjects with T2DM with lean body habitus and low BMI; having detailed phenotypic data on all the cohorts would have been ideal to pursue this line of investigation. Moreover, comparing allelic frequencies for the *EIF2AK3* gene in type 2 diabetes in young individuals below the age of 30 years vs. controls would further strengthen the study design for establishing or refuting candidacy.

8.12 The future

Detecting heritability in complex disease such as type 2 diabetes is a hard and arduous task. We have seen progression from family based linkage studies, case-control candidate gene association studies and family association study to the currently popular genome wide association studies. We are realising the scale and magnitude of understanding heritability in complex disease such as type 2 diabetes. The signals from the GWAS are identifying novel susceptibility variants and mostly pointing towards pancreatic β cell. However, the realisation there is missing heritability especially in insulin resistance genes due to lack of adequate power of GWAS or inappropriate marker selection needs to be addressed in future studies. Thus far the variants discovered in T2D gene association studies account for only 5-10% of inherited predisposition. Obviously there is scope for ongoing efforts to detect the missing heritability. There is a trend to progress from common disease common variant hypothesis testing in the current GWAS to common disease rare variant hypothesis testing. There is lot more work needed to be done in the genetics of complex disease, involving various consortia across the globe working closely. The need of the hour is to obtain large data sets numbering in tens to hundreds of thousands, to improve power and identify the missing heritability. Some of these factors have led to the first GWAS meta- analysis (135). The DIAGRAM consortium (Diabetes Genetics, Replication And Meta-analysis), a collaborative effort by the investigators from 3 major GWAS groups world wide, boosted the power to detect common susceptibility loci with modest effects, although the power to detect smaller effect remained low. The way forward in the history of gene discovery in T2D needs to be focused on even more such collaborative efforts and also to utilize the fast emerging technology to study the genome in much further detail. Indeed a further comprehensive meta-analysis of all GWAS study for T2DM has recently been completed (DIAGRAM plus consortium); unpublished data suggest that there are now 35 genes identified for T2DM. However, the identified gene variations still only accounts for ~10% genetic component of T2DM. The other area of effort in the field is comparative genomics with GWAS underway in diverse ethnic groups where the LD structure is likely to be different to the Caucasians already studied(182).

A complimentary approach to **GWAS SNP** based analysis, would be to focus on **copy number variations (CNV).** CNV are the number of variations in DNA sequence (1 kb to several megabases) in the genomes of different individuals with some individuals having more copies of specific genes than others. CNVs can be caused by deletions, duplications, inversions and translocations leading to significant differences among the genomes of different individuals. A study of eight individuals with diverse geographic ancestry revealed novel insertion sequences many of which were CNV between individuals(183). This study was a high resolution sequence map of human structural variation employing a clone based sequencing methodology. Another study involving 270 individuals from four different population groups discovered that copy number variants covered 12% of the human genome(184). These observations have ignited interest in researchers that CNV might underlie disease susceptibility in complex diseases. The study of CNV to detect association in complex diseases (eg- Crohn's disease, Schizophrenia, Autism) have been undertaken and CNV are likely to be involved in disease susceptibility; results in T2DM is still awaited. The current commercial GWAS platforms include "CNV aware" chips in addition to SNP genotyping(185). The data generated from such GWAS analysis can enable integrated analysis of SNPs and CNVs. The same constraints faced by SNP analysis could also be relevant to CNV, i.e. common variants vs. rare variants. Such enquiries of common disease rare variant hypothesis will eventually lead to efforts in **large scale sequencing studies** to detect the missing heritability in complex diseases.

The search for rare variants in common disease will be enabled by **large scale sequencing studies** with technology expanding at a rapid pace. Our study concluded that common variants of WRS gene do not contribute to development of type 2 diabetes. But rare variants of WRS gene have not been excluded in contributing to T2DM, and clonal sequencing would be the most robust approach to investigate rare variants in the *EIF2AK3* gene and possible association with T2DM.The alternative approach to our study will be **whole genome sequencing** to the whole dataset, i.e. clonal sequencing using Solexa. Nonetheless, the future will be non hypothesis driven genome wide sequencing. The principles of clonal sequencing are exemplified by Solexa Sequencing

technology(186). The Sequencing is performed by synthesis on arrays. Single DNA molecules are clonally amplified in spatially separate locations in a highly parallel array. The use of clonal single-DNA molecule templates is very useful for detecting SNP alleles in the DNA. Because each sequencing template is clonal in origin, high quality base calls are generated separately for each variant. Sequencing templates are immobilized on a proprietary flow cell surface designed to present the DNA in a manner that facilitates access to enzymes while ensuring high stability of surface-bound template and low non-specific binding of fluorescently labelled nucleotides. Single molecules are covalently attached to a planar surface and amplified in situ. Solid phase amplification is employed to create up to 1,000 identical copies of each single molecule in close proximity (diameter of one micron or less). Sequencing by synthesis is carried out by adding a mixture of four fluorescently labelled reversible chain terminators and DNA polymerase to the template. This results in addition of a single reversible terminator to each template. The fluorescent signal is detected for each template, and the flourophore and the reversible block are removed. The terminator–enzyme mix is then added to start the next cycle, and the process is reiterated until the end of the run. Solexa sequencing technology can achieve densities of up to ten million single molecule clusters per square centimeter. Solexa sequencing uses four proprietary fluorescently labelled modified nucleotides to sequence the millions of clusters present on the flow cell surface. Given that all four are present in the reaction, the risk of misincorporation is minimised, increasing sequencing accuracy. The Illumina Genome Analyser has the capability to generate over a billion bases of DNA sequence per run. The 454 system is one of the other next generation sequencing technology and it uses the pyrosequencing technology (187).

At the same time as efforts are directed to the study of genome, the critical study of gene-environment interaction and its role in disease pathogenesis needs to be addressed(188). An important interface between gene and environment is the epigenome. **Epigenetics** studies the heritable changes occurring in gene function during cellular replication, in the absence of changes in the DNA sequence. Epigenetics are often involved in switching on or switching off the gene without any changes in the base sequence leading to different phenotypes. Epigenetic mechanisms, such as DNA methylation, histone acetylation,

RNA interference, and their effects on gene activation and inactivation, are opening new insights into gene regulation, RNA transcription, and protein homeostasis. Even though all human cells contain the same genome, the gene function is different in different cells and organelles. The cells contain different epigenomes based on cell type, function, developmental age and stage of the cell. The study of epigenetics is of significant importance in type 2 diabetes for obvious reasons. The realisation is that epigenetic alteration by DNA methylation, histone acetylation and deacetylation can introduce epigenetic changes during the life time. Such changes may influence age related changes in gene-expression and thereby contribute to age-related diseases like T2DM.

Research is being directed at the gene-environment interaction, as this will indicate the appropriate population strategies to combat the increasing incidence of T2DM. Methods to investigate epigenetics are rapidly advancing and this can now be studied at the whole genome level(189). Regulation of gene expression is a complex process. The influence that epigenetic phenomena may have on disease onset and outcomes is also becoming clearer, paving the way for the development of new diagnostic and therapeutic strategies. Intra uterine growth retardation has been linked to development of T2DM in adult life. Foetal programming and the effect of abnormal intra uterine environment modifies the gene expression of cells like the pancreatic cells. Altered gene expression persists after birth and epigenetic mechanisms are thought to be involved. This links up nicely with the Barker and Hale thrifty phenotype theory of developmental origins of metabolic disorders. The current terminology "Developmental Origins of Health and Disease" DOHaD model hypothesises the phases of IUGR, foetal programming, rapid childhood growth increases the susceptibility to type 2 diabetes in later life. Research has highlighted the possible role of methyl donors like vitamin B12 and folate playing a major role of epigenetics in foetal programming. Low maternal B12 levels along with normal to high folate levels have been shown to predict adiposity and insulin resistance in Indian children(190).

The Human Epigenome Project (HEP) aims to identify genome wide DNA methylation patterns of all human genes in all major tissues(191). The goals of HEP are to generate

methylation reference maps, to identify methylation variable positions (MVPs) in the genome and to correlate methylation with gene activity, and thus with a specific phenotype or disease. Hence DNA methylation leading to change in gene function under environmental influence constitutes the main and so far missing link between genetics, disease and the environment that is widely thought to play a decisive role in the aetiology of many human pathologies. In the human genome methylation occurs naturally on cytosine bases at CpG sequences (Cytosine-phosphate- Guanine sequence in DNA) and is involved in controlling the correct expression of genes(192). Differentially methylated cytosines give rise to distinct patterns specific for tissue type and disease state. Such methylation variable positions (MVPs) are studied as the common epigenetic markers given their stability in the genome. Similar to the study of single nucleotide polymorphisms (SNPs), the study of MVPs promise to significantly advance our ability to understand chronic disease like Type 2 Diabetes.

Technologies for DNA methylation analyses

PCR based (193) technology after bisulphite conversion of unmethylated cytosines to uracil. At present, bisulphite conversion of DNA followed by PCR and sequencing remains the gold standard of methylation analysis. Real time PCR approaches enable to quantitatively detect methylation specific amplicons. Pyrosequencing assays that enable high-throughput analyses of up to 10 CpGs in an amplicon size of 300 bp have been developed(194).

Restriction enzyme based methylation analysis followed by PCR- For example the utilisation of restriction enzyme HpaII's differential ability to recognise and cleave methylated and unmethylated regions of CpG DNA sites(195). The restriction enzyme will only cut if the cytosine is unmethylated and is unable to cut the DNA if the cytosine is methylated.

Immunoprecipitation of methylated DNA targeted towards a genome wide assessment of methylation (MeDIP)(196). Microarray technology and comparative genomic

hybridization have further opened the field for high-throughput methylation analyses MeDIP- chip and Illumina 27K are technologies available to generate genome wide methylation data. MeDIP-seq combines MeDIP with the next generation sequencing technologies such as Solexa, SoLiD and 454(197).

Modifications of the methodology include using matrix-assisted laser desorption/ ionization mass spectrometry (MALDI-MS) for quantitative detection of methylation after primer extension for discrimination between methylated and unmethylated CpGs on bisulphite treated DNA(198).

Epigenetics could be used to study either a selected candidate gene such as *EIF2AK3* gene or an unbiased approach of studying the whole methylome. For instance if a study of *EIF2AK3* gene was proposed, one approach would be to do bisulphite conversion of unmethylated cytosines to uracil followed by pyrosequencing. This methodology is also used to validate top hits of a whole methylome analysis. It is unclear how useful it would be to study leukocyte DNA for *EIF2AK3* gene study, since this would depend on an early life environmental event affecting all three germ cell layers. As in relationship to insulin secretion the epigenetic mark may be tissue specific, thereby requiring access to human islets to carry out the study.

Thus whole genome studies in T2DM with integrated approach of genome wide association studies, study of rare variants and common disease and methylation studies would be a comprehensive investigative approach for type 2 diabetes genetic link.

8.13 CONCLUSIONS

Genome wide association studies are throwing up new genes associated with T2DM and other chronic diseases at a rapid pace. They promise to greatly enhance our understanding of the genetic basis of common and complex diseases. Thus far the results generated in type 2 diabetes from GWA studies highlight the genes involved in pancreatic β cell development and function. The ideal approach would be to follow through on this initial success and investigate these genes further by way of replication studies in various ethnic groups to confirm their candidacy. The results from GWAS do not however mean that the insulin resistance genes are not there to be found in T2DM. Missing heritability is likely to be identified by integration of common disease rare variant hypothesis, genome wide epigenetic and genome wide expression studies with genome wide association studies in a range of different ethnic groups. Furthermore, functional studies of the gene and its protein to elucidate the aetio-pathogenetic basis of type 2 diabetes should be conducted. Proteomics (199) (large-scale study of proteins) and metabolomics (200) (study of metabolites linked to metabolic dysfunction and type 2 diabetes and the study of metabolic responses to drugs, environmental changes and diseases), utilising tools such as nuclear magnetic resonance and mass spectrometry, will eventually lead to insights in the pathogenesis of this common condition and the ideal drug treatments for type 2 diabetes. With the knowledge of the precise biochemical variants involved in disease pathogenesis, we may be in a better position to classify the disease, identify novel drugs and design more rational therapeutic manoeuvres to prevent and ameliorate this condition. Nonetheless, as of 2008-9 lifestyle interventions (healthy diet and exercise) (201) have been demonstrated to show significant reduction in onset of diabetes in high-risk individuals for type 2 diabetes (202,203). The future might see genetic profiling of individuals to identify who might respond to such strategies including targeted pharmacological treatment.

Reference List

1. Classification and diagnosis of diabetes mellitus and other categories of glucose intolerance. National Diabetes Data Group. *Diabetes* 28:1039-1057, 1979

2. Zimmet P, Alberti KG, Shaw J: Global and societal implications of the diabetes epidemic. *Nature* 414:782-787, 2001

3. Ramachandran A, Snehalatha C, Krishnaswamy CV: Incidence of IDDM in children in urban population in southern India. Madras IDDM Registry Group Madras, South India. *Diabetes Res Clin Pract* 34:79-82, 1996

4. Pozzilli P, Di Mario U: Autoimmune diabetes not requiring insulin at diagnosis (latent autoimmune diabetes of the adult): definition, characterization, and potential prevention. *Diabetes Care* 24:1460-1467, 2001

5. Turner R, Stratton I, Horton V, Manley S, Zimmet P, Mackay IR, Shattock M, Bottazzo GF, Holman R: UKPDS 25: autoantibodies to islet-cell cytoplasm and glutamic acid decarboxylase for prediction of insulin requirement in type 2 diabetes. UK Prospective Diabetes Study Group. *Lancet* 350:1288-1293, 1997

6. She JX: Susceptibility to type I diabetes: HLA-DQ and DR revisited. *Immunol Today* 17:323-329, 1996

7. Pociot F, McDermott MF: Genetics of type 1 diabetes mellitus. *Genes Immun* 3:235-249, 2002

8. Nejentsev S, Howson JM, Walker NM, Szeszko J, Field SF, Stevens HE, Reynolds P, Hardy M, King E, Masters J, Hulme J, Maier LM, Smyth D, Bailey R, Cooper JD, Ribas G, Campbell RD, Clayton DG, Todd JA, Burton PR, Clayton DG, Cardon LR, Craddock N, Deloukas P, Duncanson A, Kwiatkowski DP, McCarthy MI, Ouwehand WH, Samani NJ, Todd JA, Donnelly P, Barrett JC, Burton PR, Davison D, Donnelly P, Easton D, Evans D, Leung HT, Marchini JL, Morris AP, Spencer CC, Tobin MD, Cardon LR, Clayton DG, Attwood AP, Boorman JP, Cant B, Everson U, Hussey JM, Jolley JD, Knight AS, Koch K, Meech E, Nutland S, Prowse CV, Stevens HE, Taylor NC, Walters GR, Walker NM, Watkins NA, Winzer T, Todd JA, Ouwehand WH, Jones RW, McArdle WL, Ring SM, Strachan DP, Pembrey M, Breen G, St CD, Caesar S, Gordon-Smith K, Jones L, Fraser C, Green EK, Grozeva D, Hamshere ML, Holmans PA, Jones IR, Kirov G, Moskvina V, Nikolov I, O'Donovan MC, Owen MJ, Craddock N, Collier DA, Elkin A, Farmer A, Williamson R, McGuffin P, Young AH, Nicol F, I, Ball SG, Balmforth AJ, Barrett JH, Bishop DT, Iles MM, Maqbool A, Yuldasheva N, Hall AS, Braund PS, Burton PR, Dixon RJ, Mangino M, Stevens S, Tobin MD, Thompson JR, Samani NJ, Bredin F, Tremelling M, Parkes M, Drummond H, Lees CW, Nimmo ER, Satsangi J, Fisher SA, Forbes A, Lewis CM, Onnie CM, Prescott NJ, Sanderson J, Mathew CG, Barbour J, Khalid MM, Todhunter CE, Mansfield JC, Ahmad T, Cummings FR, Jewell DP, Webster J,

Brown MJ, Clayton DG, Lathrop GM, Connell J, Dominiczak A, Samani NJ, Braga CA, Burke B, Dobson R, Gungadoo J, Lee KL, Munroe PB, Newhouse SJ, Onipinla A, Wallace C, Xue M, Caulfield M, Farrall M, Barton A, Bruce IN, Donovan H, Eyre S, Gilbert PD, Hider SL, Hinks AM, John SL, Potter C, Silman AJ, Symmons DP, Thomson W, Worthington J, Clayton DG, Dunger DB, Nutland S, Stevens HE, Walker NM, Widmer B, Todd JA, Frayling TM, Freathy RM, Lango H, Perry JR, Shields BM, Weedon MN, Hattersley AT, Hitman GA, Walker M, Elliott KS, Groves CJ, Lindgren CM, Rayner NW, Timpson NJ, Zeggini E, McCarthy MI, Newport M, Sirugo G, Lyons E, Vannberg F, Hill AV, Bradbury LA, Farrar C, Pointon JJ, Wordsworth P, Brown MA, Franklyn JA, Heward JM, Simmonds MJ, Gough SC, Seal S, Stratton MR, Rahman N, Ban M, Goris A, Sawcer SJ, Compston A, Conway D, Jallow M, Newport M, Sirugo G, Rockett KA, Kwiatkowski DP, Bryan C, Bumpstead SJ, Chaney A, Downes K, Ghori J, Gwilliam R, Hunt SE, Inouye M, Keniry A, King E, McGinnis R, Potter S, Ravindrarajah R, Whittaker P, Withers D, Deloukas P, Leung HT, Nutland S, Stevens HE, Walker NM, Todd JA, Easton D, Clayton DG, Burton PR, Tobin MD, Barrett JC, Evans D, Morris AP, Cardon LR, Cardin NJ, Davison D, Ferreira T, Pereira-Gale J, Hallgrimsdottir IB, Howie BN, Marchini JL, Spencer CC, Su Z, Ying TY, Vukcevic D, Donnelly P, Bentley D, Brown MA, Cardon LR: Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A. *Nature* 450:887-892, 2007

9. Pociot F, McDermott MF: Genetics of type 1 diabetes mellitus. *Genes Immun* 3:235-249, 2002

10. Bennett ST, Lucassen AM, Gough SC, Powell EE, Undlien DE, Pritchard LE, Merriman ME, Kawaguchi Y, Dronsfield MJ, Pociot F, .: Susceptibility to human type 1 diabetes at IDDM2 is determined by tandem repeat variation at the insulin gene minisatellite locus. *Nat Genet* 9:284-292, 1995

11. Ueda H, Howson JM, Esposito L, Heward J, Snook H, Chamberlain G, Rainbow DB, Hunter KM, Smith AN, Di Genova G, Herr MH, Dahlman I, Payne F, Smyth D, Lowe C, Twells RC, Howlett S, Healy B, Nutland S, Rance HE, Everett V, Smink LJ, Lam AC, Cordell HJ, Walker NM, Bordin C, Hulme J, Motzo C, Cucca F, Hess JF, Metzker ML, Rogers J, Gregory S, Allahabadia A, Nithiyananthan R, Tuomilehto-Wolf E, Tuomilehto J, Bingley P, Gillespie KM, Undlien DE, Ronningen KS, Guja C, Ionescu-Tirgoviste C, Savage DA, Maxwell AP, Carson DJ, Patterson CC, Franklyn JA, Clayton DG, Peterson LB, Wicker LS, Todd JA, Gough SC: Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature* 423:506-511, 2003

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

13. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447:661-678, 2007

14. Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, Plagnol V, Bailey R, Nejentsev S, Field SF, Payne F, Lowe CE, Szeszko JS, Hafler JP, Zeitels L, Yang JH, Vella A, Nutland S, Stevens HE, Schuilenburg H, Coleman G, Maisuria M, Meadows W, Smink LJ, Healy B, Burren OS, Lam AA, Ovington NR, Allen J, Adlem E, Leung HT, Wallace C, Howson JM, Guja C, Ionescu-Tirgoviste C, Simmonds MJ, Heward JM, Gough SC, Dunger DB, Wicker LS, Clayton DG: Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat Genet* 39:857-864, 2007

15. Hyoty H, Taylor KW: The role of viruses in human diabetes. *Diabetologia* 45:1353-1361, 2002

16. Clements GB, Galbraith DN, Taylor KW: Coxsackie B virus infection and onset of childhood diabetes. *Lancet* 346:221-223, 1995

17. Michalkova D, Petrovicova A, Kolar J, Jancova E, Silesova J: [The role of coxsackie virus infection in the development of type 1 diabetes mellitus and its effect on the postinitial course of the disease in childhood]. *Cesk Pediatr* 44:257-262, 1989

18. Horwitz MS, Bradley LM, Harbertson J, Krahl T, Lee J, Sarvetnick N: Diabetes induced by Coxsackie virus: initiation by bystander damage and not molecular mimicry. *Nat Med* 4:781-785, 1998

19. Forrest JM, Menser MA, Harley JD: Diabetes mellitus and congenital rubella. *Pediatrics* 44:445-447, 1969

20. Mathieu C, Laureys J, Sobis H, Vandeputte M, Waer M, Bouillon R: 1,25-Dihydroxyvitamin D3 prevents insulitis in NOD mice. *Diabetes* 41:1491-1495, 1992

21. Elias D, Cohen IR: Treatment of autoimmune diabetes and insulitis in NOD mice with heat shock protein 60 peptide p277. *Diabetes* 44:1132-1138, 1995

22. Raz I, Elias D, Avron A, Tamir M, Metzger M, Cohen IR: Beta-cell function in new-onset type 1 diabetes and immunomodulation with a heat-shock protein peptide (DiaPep277): a randomised, double-blind, phase II trial. *Lancet* 358:1749-1753, 2001

23. Keymeulen B, Vandemeulebroucke E, Ziegler AG, Mathieu C, Kaufman L, Hale G, Gorus F, Goldman M, Walter M, Candon S, Schandene L, Crenier L, De BC, Seigneurin JM, De PP, Pierard D, Weets I, Rebello P, Bird P, Berrie E, Frewin M, Waldmann H, Bach JF, Pipeleers D, Chatenoud L: Insulin needs after CD3-antibody therapy in new-onset type 1 diabetes. *N Engl J Med* 352:2598-2608, 2005

24. Gale EA: Intervening before the onset of Type 1 diabetes: baseline data from the European Nicotinamide Diabetes Intervention Trial (ENDIT). *Diabetologia* 46:339-346, 2003

25. Couzin J: Clinical trials. Diabetes' brave new world. *Science* 300:1862-1865, 2003

26. Amos AF, McCarty DJ, Zimmet P: The rising global burden of diabetes and its complications: estimates and projections to the year 2010. *Diabet Med* 14 Suppl 5:S1-85, 1997

27. Ramachandran A, Snehalatha C, Kapur A, Vijay V, Mohan V, Das AK, Rao PV, Yajnik CS, Prasanna Kumar KM, Nair JD: High prevalence of diabetes and impaired glucose tolerance in India: National Urban Diabetes Survey. *Diabetologia* 44:1094-1101, 2001

28. Zimmet PZ, McCarty DJ, de Court: The global epidemiology of non-insulin-dependent diabetes mellitus and the metabolic syndrome. *J Diabetes Complications* 11:60-68, 1997

29. Goldfine AB, Kahn CR: Adiponectin: linking the fat cell to insulin sensitivity. *Lancet* 362:1431-1432, 2003

30. Eisenstein J, Greenberg A: Ghrelin: update 2003. *Nutr Rev* 61:101-104, 2003

31. Hitman GA, Sudagani J: Searching for genes in diabetes and the metabolic syndrome. *Int J Clin Pract Suppl*3-8, 2004

32. Rich SS: Mapping genes in diabetes. Genetic epidemiological perspective. *Diabetes* 39:1315-1319, 1990

33. Lee KE, Klein BE, Klein R: Familial aggregation of components of the multiple metabolic syndrome in the Framingham Heart and Offspring Cohorts: Genetic Analysis Workshop Problem 1. *BMC Genet* 4 Suppl 1:S94, 2003

34. Mills GW, Avery PJ, McCarthy MI, Hattersley AT, Levy JC, Hitman GA, Sampson M, Walker M: Heritability estimates for beta cell function and features of the insulin resistance syndrome in UK families with an increased susceptibility to Type 2 diabetes. *Diabetologia* 2004

35. Freeman MS, Mansfield MW, Barrett JH, Grant PJ: Heritability of features of the insulin resistance syndrome in a community-based study of healthy families. *Diabet Med* 19:994-999, 2002

36. Serjeantson SW, Owerbach D, Zimmet P, Nerup J, Thoma K: Genetics of diabetes in Nauru: effects of foreign admixture, HLA antigens and the insulin-gene-linked polymorphism. *Diabetologia* 25:13-17, 1983

37. Medici F, Hawa M, Ianari A, Pyke DA, Leslie RD: Concordance rate for type II diabetes mellitus in monozygotic twins: actuarial analysis. *Diabetologia* 42:146-150, 1999

38. Lander ES, Schork NJ: Genetic dissection of complex traits. *Science* 265:2037-2048, 1994

39. Wiltshire S, Hattersley AT, Hitman GA, Walker M, Levy JC, Sampson M, O'Rahilly S, Frayling TM, Bell JI, Lathrop GM, Bennett A, Dhillon R, Fletcher C, Groves CJ, Jones E, Prestwich P, Simecek N, Rao PV, Wishart M, Bottazzo GF, Foxon R, Howell S, Smedley D, Cardon LR, Menzel S, McCarthy MI: A genomewide scan for loci predisposing to type 2 diabetes in a U.K. population (the Diabetes UK Warren 2 Repository): analysis of 573 pedigrees provides independent replication of a susceptibility locus on chromosome 1q. *Am J Hum Genet* 69:553-569, 2001

40. Horikawa Y, Oda N, Cox NJ, Li X, Orho-Melander M, Hara M, Hinokio Y, Lindner TH, Mashima H, Schwarz PE, Bosque-Plata L, Horikawa Y, Oda Y, Yoshiuchi I, Colilla S, Polonsky KS, Wei S, Concannon P, Iwasaki N, Schulze J, Baier LJ, Bogardus C, Groop L, Boerwinkle E, Hanis CL, Bell GI: Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat Genet* 26:163-175, 2000

41. Florez JC, Hirschhorn J, Altshuler D: The inherited basis of diabetes mellitus: implications for the genetic analysis of complex traits. *Annu Rev Genomics Hum Genet* 4:257-291, 2003

42. The International HapMap Project. *Nature* 426:789-796, 2003

43. Gloyn AL, McCarthy MI: The genetics of type 2 diabetes. *Best Pract Res Clin Endocrinol Metab* 15:293-308, 2001

44. Altshuler D, Daly MJ, Lander ES: Genetic mapping in human disease. *Science* 322:881-888, 2008

45. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447:661-678, 2007

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

47. Florez JC, Jablonski KA, Bayley N, Pollin TI, de Bakker PI, Shuldiner AR, Knowler WC, Nathan DM, Altshuler D: TCF7L2 polymorphisms and progression to diabetes in the Diabetes Prevention Program. *N Engl J Med* 355:241-250, 2006

48. Groves CJ, Zeggini E, Minton J, Frayling TM, Weedon MN, Rayner NW, Hitman GA, Walker M, Wiltshire S, Hattersley AT, McCarthy MI: Association analysis of 6,736 U.K. subjects provides replication and confirms TCF7L2 as a type 2 diabetes susceptibility gene with a substantial effect on individual risk. *Diabetes* 55:2640-2644, 2006

49. Schafer SA, Tschritter O, Machicao F, Thamer C, Stefan N, Gallwitz B, Holst JJ, Dekker JM, T'hart LM, Nijpels G, van Haeften TW, Haring HU, Fritsche A: Impaired glucagon-like peptide-1-induced insulin secretion in carriers of transcription factor 7-like 2 (TCF7L2) gene polymorphisms. *Diabetologia* 2007

50. Gloyn AL, Weedon MN, Owen KR, Turner MJ, Knight BA, Hitman G, Walker M, Levy JC, Sampson M, Halford S, McCarthy MI, Hattersley AT, Frayling TM: Large-scale association studies of variants in genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes. *Diabetes* 52:568-572, 2003

51. Greene ME, Blumberg B, McBride OW, Yi HF, Kronquist K, Kwan K, Hsieh L, Greene G, Nimer SD: Isolation of the human peroxisome proliferator activated receptor gamma cDNA: expression in hematopoietic cells and chromosomal mapping. *Gene Expr* 4:281-299, 1995

52. Altshuler D, Hirschhorn JN, Klannemark M, Lindgren CM, Vohl MC, Nemesh J, Lane CR, Schaffner SF, Bolk S, Brewer C, Tuomi T, Gaudet D, Hudson TJ, Daly M, Groop L, Lander ES: The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat Genet* 26:76-80, 2000

53. Johnson JD, Han Z, Otani K, Ye H, Zhang Y, Wu H, Horikawa Y, Misler S, Bell GI, Polonsky KS: RyR2 and calpain-10 delineate a novel apoptosis pathway in pancreatic islets. *J Biol Chem* 279:24794-24802, 2004

54. Sreenan SK, Zhou YP, Otani K, Hansen PA, Currie KP, Pan CY, Lee JP, Ostrega DM, Pugh W, Horikawa Y, Cox NJ, Hanis CL, Burant CF, Fox AP, Bell GI, Polonsky KS: Calpains play a role in insulin secretion and action. *Diabetes* 50:2013-2020, 2001

55. Patel YM, Lane MD: Role of calpain in adipocyte differentiation. *Proc Natl Acad Sci U S A* 96:1279-1284, 1999

56. Hanis CL, Boerwinkle E, Chakraborty R, Ellsworth DL, Concannon P, Stirling B, Morrison VA, Wapelhorst B, Spielman RS, Gogolin-Ewens KJ, Shepard JM, Williams SR, Risch N, Hinds D, Iwasaki N, Ogata M, Omori Y, Petzold C, Rietzch H, Schroder HE, Schulze J, Cox NJ, Menzel S, Boriraj VV, Chen X, .: A

genome-wide search for human non-insulin-dependent (type 2) diabetes genes reveals a major susceptibility locus on chromosome 2. *Nat Genet* 13:161-166, 1996

57. Cox NJ, Frigge M, Nicolae DL, Concannon P, Hanis CL, Bell GI, Kong A: Loci on chromosomes 2 (NIDDM1) and 15 interact to increase susceptibility to diabetes in Mexican Americans. *Nat Genet* 21:213-215, 1999

58. Horikawa Y, Oda N, Cox NJ, Li X, Orho-Melander M, Hara M, Hinokio Y, Lindner TH, Mashima H, Schwarz PE, Bosque-Plata L, Horikawa Y, Oda Y, Yoshiuchi I, Colilla S, Polonsky KS, Wei S, Concannon P, Iwasaki N, Schulze J, Baier LJ, Bogardus C, Groop L, Boerwinkle E, Hanis CL, Bell GI: Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat Genet* 26:163-175, 2000

59. Weedon MN, Schwarz PE, Horikawa Y, Iwasaki N, Illig T, Holle R, Rathmann W, Selisko T, Schulze J, Owen KR, Evans J, Del Bosque-Plata L, Hitman G, Walker M, Levy JC, Sampson M, Bell GI, McCarthy MI, Hattersley AT, Frayling TM: Meta-analysis and a large association study confirm a role for calpain-10 variation in type 2 diabetes susceptibility. *Am J Hum Genet* 73:1208-1212, 2003

60. Baier LJ, Permana PA, Yang X, Pratley RE, Hanson RL, Shen GQ, Mott D, Knowler WC, Cox NJ, Horikawa Y, Oda N, Bell GI, Bogardus C: A calpain-10 gene polymorphism is associated with reduced muscle mRNA levels and insulin resistance. *J Clin Invest* 106:R69-R73, 2000

61. Marshall C, Hitman GA, Partridge CJ, Clark A, Ma H, Shearer TR, Turner MD: Evidence that an isoform of calpain-10 is a regulator of exocytosis in pancreatic beta-cells. *Mol Endocrinol* 19:213-224, 2005

62. Turner MD, Fulcher FK, Jones CV, Smith BT, Aganna E, Partridge CJ, Hitman GA, Clark A, Patel YM: Calpain facilitates actin reorganization during glucose-stimulated insulin secretion. *Biochem Biophys Res Commun* 352:650-655, 2007

63. Zhou YP, Sreenan S, Pan CY, Currie KP, Bindokas VP, Horikawa Y, Lee JP, Ostrega D, Ahmed N, Baldwin AC, Cox NJ, Fox AP, Miller RJ, Bell GI, Polonsky KS: A 48-hour exposure of pancreatic islets to calpain inhibitors impairs mitochondrial fuel metabolism and the exocytosis of insulin. *Metabolism* 52:528-534, 2003

64. Paul DS, Harmon AW, Winston CP, Patel YM: Calpain facilitates GLUT4 vesicle translocation during insulin-stimulated glucose uptake in adipocytes. *Biochem J* 376:625-632, 2003

65. Hales CN, Barker DJ: Type 2 (non-insulin-dependent) diabetes mellitus: the thrifty phenotype hypothesis. *Diabetologia* 35:595-601, 1992

66. Hales CN, Barker DJ: The thrifty phenotype hypothesis. *Br Med Bull* 60:5-20, 2001

67. Lindstrom J, Louheranta A, Mannelin M, Rastas M, Salminen V, Eriksson J, Uusitupa M, Tuomilehto J: The Finnish Diabetes Prevention Study (DPS): Lifestyle intervention and 3-year results on diet and physical activity. *Diabetes Care* 26:3230-3236, 2003

68. Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, Nathan DM: Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med* 346:393-403, 2002

69. Bray GA, Jablonski KA, Fujimoto WY, Barrett-Connor E, Haffner S, Hanson RL, Hill JO, Hubbard V, Kriska A, Stamm E, Pi-Sunyer FX: Relation of central adiposity and body mass index to the development of diabetes in the Diabetes Prevention Program. *Am J Clin Nutr* 87:1212-1218, 2008

70. Hu FB, van Dam RM, Liu S: Diet and risk of Type II diabetes: the role of types of fat and carbohydrate. *Diabetologia* 44:805-817, 2001

71. Vessby B, Unsitupa M, Hermansen K, Riccardi G, Rivellese AA, Tapsell LC, Nalsen C, Berglund L, Louheranta A, Rasmussen BM, Calvert GD, Maffetone A, Pedersen E, Gustafsson IB, Storlien LH: Substituting dietary saturated for monounsaturated fat impairs insulin sensitivity in healthy men and women: The KANWU Study. *Diabetologia* 44:312-319, 2001

72. Salmeron J, Hu FB, Manson JE, Stampfer MJ, Colditz GA, Rimm EB, Willett WC: Dietary fat intake and risk of type 2 diabetes in women. *Am J Clin Nutr* 73:1019-1026, 2001

73. Feskens EJ, Virtanen SM, Rasanen L, Tuomilehto J, Stengard J, Pekkanen J, Nissinen A, Kromhout D: Dietary factors determining diabetes and impaired glucose tolerance. A 20-year follow-up of the Finnish and Dutch cohorts of the Seven Countries Study. *Diabetes Care* 18:1104-1112, 1995

74. Boucher BJ, Mannan N, Noonan K, Hales CN, Evans SJ: Glucose intolerance and impairment of insulin secretion in relation to vitamin D deficiency in east London Asians. *Diabetologia* 38:1239-1245, 1995

75. Hattersley AT: Diagnosis of maturity-onset diabetes of the young in the pediatric diabetes clinic. *J Pediatr Endocrinol Metab* 13 Suppl 6:1411-1417, 2000

76. Stride A, Hattersley AT: Different genes, different diabetes: lessons from maturity-onset diabetes of the young. *Ann Med* 34:207-216, 2002

77. Frayling TM, Evans JC, Bulman MP, Pearson E, Allen L, Owen K, Bingham C, Hannemann M, Shepherd M, Ellard S, Hattersley AT: beta-cell genes and

diabetes: molecular and clinical characterization of mutations in transcription factors. *Diabetes* 50 Suppl 1:S94-100, 2001

78. Pearson ER, Starkey BJ, Powell RJ, Gribble FM, Clark PM, Hattersley AT: Genetic cause of hyperglycaemia and response to treatment in diabetes. *Lancet* 362:1275-1281, 2003

79. Hattersley AT, Turner RC, Permutt MA, Patel P, Tanizawa Y, Chiu KC, O'Rahilly S, Watkins PJ, Wainscoat JS: Linkage of type 2 diabetes to the glucokinase gene. *Lancet* 339:1307-1310, 1992

80. Vaxillaire M, Boccio V, Philippi A, Vigouroux C, Terwilliger J, Passa P, Beckmann JS, Velho G, Lathrop GM, Froguel P: A gene for maturity onset diabetes of the young (MODY) maps to chromosome 12q. *Nat Genet* 9:418-423, 1995

81. Yamagata K, Oda N, Kaisaki PJ, Menzel S, Furuta H, Vaxillaire M, Southam L, Cox RD, Lathrop GM, Boriraj VV, Chen X, Cox NJ, Oda Y, Yano H, Le Beau MM, Yamada S, Nishigori H, Takeda J, Fajans SS, Hattersley AT, Iwasaki N, Hansen T, Pedersen O, Polonsky KS, Bell GI, .: Mutations in the hepatocyte nuclear factor-1alpha gene in maturity-onset diabetes of the young (MODY3). *Nature* 384:455-458, 1996

82. Pearson ER, Starkey BJ, Powell RJ, Gribble FM, Clark PM, Hattersley AT: Genetic cause of hyperglycaemia and response to treatment in diabetes. *Lancet* 362:1275-1281, 2003

83. Wolcott CD, Rallison ML: Infancy-onset diabetes mellitus and multiple epiphyseal dysplasia. *J Pediatr* 80:292-297, 1972

84. Thornton CM, Carson DJ, Stewart FJ: Autopsy findings in the Wolcott-Rallison syndrome. *Pediatr Pathol Lab Med* 17:487-496, 1997

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

86. Hayes SE, Conner LJ, Stramm LE, Shi Y: Assignment of pancreatic eIF-2alpha kinase (EIF2AK3) to human chromosome band 2p12 by radiation hybrid mapping and in situ hybridization. *Cytogenet Cell Genet* 86:327-328, 1999

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

88. Biason-Lauber A, Lang-Muritano M, Vaccaro T, Schoenle EJ: Loss of kinase activity in a patient with Wolcott-Rallison syndrome caused by a novel mutation in the EIF2AK3 gene. *Diabetes* 51:2301-2305, 2002

89. Brickwood S, Bonthron DT, Al-Gazali LI, Piper K, Hearn T, Wilson DI, Hanley NA: Wolcott-Rallison syndrome: pathogenic insights into neonatal diabetes from new mutation and expression studies of EIF2AK3. *J Med Genet* 40:685-689, 2003

90. Senee V, Vattem KM, Delepine M, Rainbow LA, Haton C, Lecoq A, Shaw NJ, Robert JJ, Rooman R, atloff-Zito C, Michaud JL, Bin-Abbas B, Taha D, Zabel B, Franceschini P, Topaloglu AK, Lathrop GM, Barrett TG, Nicolino M, Wek RC, Julier C: Wolcott-Rallison Syndrome: clinical, genetic, and functional study of EIF2AK3 mutations and suggestion of genetic heterogeneity. *Diabetes* 53:1876-1883, 2004

91. Allotey RA, Mohan V, McDermott MF, Deepa R, Premalatha G, Hassan Z, Cassell PG, North BV, Vaxillaire M, Mein CA, Swan DC, O'Grady E, Ramachandran A, Snehalatha C, Sinnot PJ, Hemmatpour SK, Froguel P, Hitman GA: The EIF2AK3 gene region and type I diabetes in subjects from South India. *Genes Immun* 5:648-652, 2004

92. Shi Y, Vattem KM, Sood R, An J, Liang J, Stramm L, Wek RC: Identification and characterization of pancreatic eukaryotic initiation factor 2 alpha-subunit kinase, PEK, involved in translational control. *Mol Cell Biol* 18:7499-7509, 1998

93. Hayes SE, Conner LJ, Stramm LE, Shi Y: Assignment of pancreatic eIF-2alpha kinase (EIF2AK3) to human chromosome band 2p12 by radiation hybrid mapping and in situ hybridization. *Cytogenet Cell Genet* 86:327-328, 1999

94. Harding HP, Zhang Y, Bertolotti A, Zeng H, Ron D: Perk is essential for translational regulation and cell survival during the unfolded protein response. *Mol Cell* 5:897-904, 2000

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

96. Brickwood S, Bonthron DT, Al-Gazali LI, Piper K, Hearn T, Wilson DI, Hanley NA: Wolcott-Rallison syndrome: pathogenic insights into neonatal diabetes from new mutation and expression studies of EIF2AK3. *J Med Genet* 40:685-689, 2003

97. Kerr JF, Wyllie AH, Currie AR: Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. *Br J Cancer* 26:239-257, 1972

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

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

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

101. Evans JL, Goldfine ID, Maddux BA, Grodsky GM: Are oxidative stress-activated signaling pathways mediators of insulin resistance and beta-cell dysfunction? *Diabetes* 52:1-8, 2003

102. Prentki M, Joly E, El Assaad W, Roduit R: Malonyl-CoA signaling, lipid partitioning, and glucolipotoxicity: role in beta-cell adaptation and failure in the etiology of diabetes. *Diabetes* 51 Suppl 3:S405-S413, 2002

103. Lupi R, Dotta F, Marselli L, Del Guerra S, Masini M, Santangelo C, Patane G, Boggi U, Piro S, Anello M, Bergamini E, Mosca F, Di Mario U, Del Prato S, Marchetti P: Prolonged exposure to free fatty acids has cytostatic and pro-apoptotic effects on human pancreatic islets: evidence that beta-cell death is caspase mediated, partially dependent on ceramide pathway, and Bcl-2 regulated. *Diabetes* 51:1437-1442, 2002

104. Mandrup-Poulsen T: Apoptotic signal transduction pathways in diabetes. *Biochem Pharmacol* 66:1433-1440, 2003

105. Maedler K, Spinas GA, Lehmann R, Sergeev P, Weber M, Fontana A, Kaiser N, Donath MY: Glucose induces beta-cell apoptosis via upregulation of the Fas receptor in human islets. *Diabetes* 50:1683-1690, 2001

106. Maedler K, Sergeev P, Ris F, Oberholzer J, Joller-Jemelka HI, Spinas GA, Kaiser N, Halban PA, Donath MY: Glucose-induced beta cell production of IL-1beta contributes to glucotoxicity in human pancreatic islets. *J Clin Invest* 110:851-860, 2002

107. Donath MY, Mandrup-Poulsen T: The use of interleukin-1-receptor antagonists in the treatment of diabetes mellitus. *Nat Clin Pract Endocrinol Metab* 4:240-241, 2008

108. Donath MY, Storling J, Maedler K, Mandrup-Poulsen T: Inflammatory mediators and islet beta-cell failure: a link between type 1 and type 2 diabetes. *J Mol Med* 81:455-470, 2003

109. Csala M, Banhegyi G, Benedetti A: Endoplasmic reticulum: a metabolic compartment. *FEBS Lett* 580:2160-2165, 2006

110. Scheuner D, Kaufman RJ: The unfolded protein response: a pathway that links insulin demand with beta-cell failure and diabetes. *Endocr Rev* 29:317-333, 2008

111. Unger RH, Grundy S: Hyperglycaemia as an inducer as well as a consequence of impaired islet cell function and insulin resistance: implications for the management of diabetes. *Diabetologia* 28:119-121, 1985

112. Poitout V, Robertson RP: Glucolipotoxicity: fuel excess and beta-cell dysfunction. *Endocr Rev* 29:351-366, 2008

113. Unger RH: Lipotoxicity in the pathogenesis of obesity-dependent NIDDM. Genetic and clinical implications. *Diabetes* 44:863-870, 1995

114. Oyadomari S, Koizumi A, Takeda K, Gotoh T, Akira S, Araki E, Mori M: Targeted disruption of the Chop gene delays endoplasmic reticulum stress-mediated diabetes. *J Clin Invest* 109:525-532, 2002

115. Schroder M: Endoplasmic reticulum stress responses. *Cell Mol Life Sci* 65:862-894, 2008

116. Shi Y, Vattem KM, Sood R, An J, Liang J, Stramm L, Wek RC: Identification and characterization of pancreatic eukaryotic initiation factor 2 alpha-subunit kinase, PEK, involved in translational control. *Mol Cell Biol* 18:7499-7509, 1998

117. Schroder M, Kaufman RJ: The mammalian unfolded protein response. *Annu Rev Biochem* 74:739-789, 2005

118. Oyadomari S, Araki E, Mori M: Endoplasmic reticulum stress-mediated apoptosis in pancreatic beta-cells. *Apoptosis* 7:335-345, 2002

119. Shi Y, An J, Liang J, Hayes SE, Sandusky GE, Stramm LE, Yang NN: Characterization of a mutant pancreatic eIF-2alpha kinase, PEK, and co-localization with somatostatin in islet delta cells. *J Biol Chem* 274:5723-5730, 1999

120. Brickwood S, Bonthron DT, Al-Gazali LI, Piper K, Hearn T, Wilson DI, Hanley NA: Wolcott-Rallison syndrome: pathogenic insights into neonatal diabetes from new mutation and expression studies of EIF2AK3. *J Med Genet* 40:685-689, 2003

121. Harding HP, Zeng H, Zhang Y, Jungries R, Chung P, Plesken H, Sabatini DD, Ron D: Diabetes mellitus and exocrine pancreatic dysfunction in perk-/- mice reveals a role for translational control in secretory cell survival. *Mol Cell* 7:1153-1163, 2001

122. Harding HP, Zhang Y, Bertolotti A, Zeng H, Ron D: Perk is essential for translational regulation and cell survival during the unfolded protein response. *Mol Cell* 5:897-904, 2000

123. Nerup J, Pociot F: A genomewide scan for type 1-diabetes susceptibility in Scandinavian families: identification of new loci with evidence of interactions. *Am J Hum Genet* 69:1301-1313, 2001

124. Alberti KG, Zimmet PZ: Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. *Diabet Med* 15:539-553, 1998

125. Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, Heiner C, Kent SB, Hood LE: Fluorescence detection in automated DNA sequence analysis. *Nature* 321:674-679, 1986

126. Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA: Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239:487-491, 1988

127. Heid CA, Stevens J, Livak KJ, Williams PM: Real time quantitative PCR. *Genome Res* 6:986-994, 1996

128. Nazarenko IA, Bhatnagar SK, Hohman RJ: A closed tube format for amplification and detection of DNA based on energy transfer. *Nucleic Acids Res* 25:2516-2521, 1997

129. Hattersley AT, McCarthy MI: What makes a good genetic association study? *Lancet* 366:1315-1323, 2005

130. Hara K, Tobe K, Okada T, Kadowaki H, Akanuma Y, Ito C, Kimura S, Kadowaki T: A genetic variation in the PGC-1 gene could confer insulin resistance and susceptibility to Type II diabetes. *Diabetologia* 45:740-743, 2002

131. Edwards AW: G. H. Hardy (1908) and Hardy-Weinberg equilibrium. *Genetics* 179:1143-1150, 2008

132. Ryckman K, Williams SM: Calculation and use of the Hardy-Weinberg model in association studies. *Curr Protoc Hum Genet* Chapter 1:Unit, 2008

133. Dudbridge F: Likelihood-based association analysis for nuclear families and unrelated subjects with missing genotype data. *Hum Hered* 66:87-98, 2008

134. Cordell HJ, Clayton DG: Genetic association studies. *Lancet* 366:1121-1131, 2005

135. Spielman RS, McGinnis RE, Ewens WJ: Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 52:506-516, 1993

136. Sham PC, Curtis D: An extended transmission/disequilibrium test (TDT) for multi-allele marker loci. *Ann Hum Genet* 59:323-336, 1995

137. Curtis D, Knight J, Sham PC: Program report: GENECOUNTING support programs. *Ann Hum Genet* 70:277-279, 2006

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

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

140. Minton JA, Hattersley AT, Owen K, McCarthy MI, Walker M, Latif F, Barrett T, Frayling TM: Association studies of genetic variation in the WFS1 gene and type 2 diabetes in U.K. populations. *Diabetes* 51:1287-1290, 2002

141. Cassell PG, Jackson AE, North BV, Evans JC, Syndercombe-Court, Phillips C, Ramachandran A, Snehalatha C, Gelding SV, Vijayaravaghan S, Curtis D, Hitman GA: Haplotype combinations of calpain 10 gene polymorphisms associate with increased risk of impaired glucose tolerance and type 2 diabetes in South Indians. *Diabetes* 51:1622-1628, 2002

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

143. van Dam RM, Hoebee B, Seidell JC, Schaap MM, Blaak EE, Feskens EJ: The insulin receptor substrate-1 Gly972Arg polymorphism is not associated with Type 2 diabetes mellitus in two population-based studies. *Diabet Med* 21:752-758, 2004

144. Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA: Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239:487-491, 1988

145. Heid CA, Stevens J, Livak KJ, Williams PM: Real time quantitative PCR. *Genome Res* 6:986-994, 1996

146. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H,

Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la BM, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de JP, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ: Initial sequencing and analysis of the human genome. *Nature* 409:860-921, 2001

147. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, Hunt SE, Cole CG, Coggill PC, Rice CM, Ning Z, Rogers J, Bentley DR, Kwok PY, Mardis ER, Yeh RT, Schultz B, Cook L, Davenport R, Dante M, Fulton L, Hillier L, Waterston RH, McPherson JD, Gilman B, Schaffner S, Van Etten WJ, Reich D, Higgins J, Daly MJ, Blumenstiel B, Baldwin J, Stange-Thomann N, Zody MC, Linton L, Lander ES, Altshuler D: A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409:928-933, 2001

148. Zeggini E, Rayner W, Morris AP, Hattersley AT, Walker M, Hitman GA, Deloukas P, Cardon LR, McCarthy MI: An evaluation of HapMap sample size and tagging SNP performance in large-scale empirical and simulated data sets. *Nat Genet* 37:1320-1322, 2005

149. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, de Bakker PI, Abecasis GR, Almgren P, Andersen G, Ardlie K, Bostrom KB, Bergman RN, Bonnycastle LL, Borch-Johnsen K, Burtt NP, Chen H, Chines PS, Daly MJ, Deodhar P, Ding CJ, Doney AS, Duren WL, Elliott KS, Erdos MR, Frayling TM, Freathy RM, Gianniny L, Grallert H, Grarup N, Groves CJ, Guiducci C, Hansen T, Herder C, Hitman GA, Hughes TE, Isomaa B, Jackson AU, Jorgensen T, Kong A, Kubalanza K, Kuruvilla FG, Kuusisto J, Langenberg C, Lango H, Lauritzen T, Li Y, Lindgren CM, Lyssenko V, Marvelle AF, Meisinger C, Midthjell K, Mohlke KL, Morken MA, Morris AD, Narisu N, Nilsson P, Owen KR, Palmer CN, Payne F, Perry JR, Pettersen E, Platou C, Prokopenko I, Qi L, Qin L, Rayner NW, Rees M, Roix JJ, Sandbaek A, Shields B, Sjogren M, Steinthorsdottir V,

Stringham HM, Swift AJ, Thorleifsson G, Thorsteinsdottir U, Timpson NJ, Tuomi T, Tuomilehto J, Walker M, Watanabe RM, Weedon MN, Willer CJ, Illig T, Hveem K, Hu FB, Laakso M, Stefansson K, Pedersen O, Wareham NJ, Barroso I, Hattersley AT, Collins FS, Groop L, McCarthy MI, Boehnke M, Altshuler D: Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet* 40:638-645, 2008

150. McCarthy MI, Zeggini E: Genome-wide association studies in type 2 diabetes. *Curr Diab Rep* 9:164-171, 2009

151. Florez JC: Newly identified loci highlight beta cell dysfunction as a key cause of type 2 diabetes: where are the insulin resistance genes? *Diabetologia* 51:1100-1110, 2008

152. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S, Balkau B, Heude B, Charpentier G, Hudson TJ, Montpetit A, Pshezhetsky AV, Prentki M, Posner BI, Balding DJ, Meyre D, Polychronakos C, Froguel P: A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 445:881-885, 2007

153. Chimienti F, Devergnas S, Favier A, Seve M: Identification and cloning of a beta-cell-specific zinc transporter, ZnT-8, localized into insulin secretory granules. *Diabetes* 53:2330-2337, 2004

154. Bort R, Martinez-Barbera JP, Beddington RS, Zaret KS: Hex homeobox gene-dependent tissue positioning is required for organogenesis of the ventral pancreas. *Development* 131:797-806, 2004

155. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, Perry JR, Elliott KS, Lango H, Rayner NW, Shields B, Harries LW, Barrett JC, Ellard S, Groves CJ, Knight B, Patch AM, Ness AR, Ebrahim S, Lawlor DA, Ring SM, Ben-Shlomo Y, Jarvelin MR, Sovio U, Bennett AJ, Melzer D, Ferrucci L, Loos RJ, Barroso I, Wareham NJ, Karpe F, Owen KR, Cardon LR, Walker M, Hitman GA, Palmer CN, Doney AS, Morris AD, Smith GD, Hattersley AT, McCarthy MI: A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* 316:889-894, 2007

156. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding CJ, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li XY, Conneely KN, Riebow NL, Sprau AG, Tong M, White PP, Hetrick KN, Barnhart MW, Bark CW, Goldstein JL, Watkins L, Xiang F, Saramies J, Buchanan TA, Watanabe RM, Valle TT, Kinnunen L, Abecasis GR, Pugh EW, Doheny KF, Bergman RN, Tuomilehto J, Collins FS, Boehnke M: A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 316:1341-1345, 2007

157. Florez JC: Newly identified loci highlight beta cell dysfunction as a key cause of type 2 diabetes: where are the insulin resistance genes? *Diabetologia* 51:1100-1110, 2008

158. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S, Balkau B, Heude B, Charpentier G, Hudson TJ, Montpetit A, Pshezhetsky AV, Prentki M, Posner BI, Balding DJ, Meyre D, Polychronakos C, Froguel P: A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 445:881-885, 2007

159. Pritchard JK: Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet* 69:124-137, 2001

160. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN: Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 9:356-369, 2008

161. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM: Finding the missing heritability of complex diseases. *Nature* 461:747-753, 2009

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

163. Gloyn AL, Weedon MN, Owen KR, Turner MJ, Knight BA, Hitman G, Walker M, Levy JC, Sampson M, Halford S, McCarthy MI, Hattersley AT, Frayling TM: Large-scale association studies of variants in genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes. *Diabetes* 52:568-572, 2003

164. Sandhu MS, Weedon MN, Fawcett KA, Wasson J, Debenham SL, Daly A, Lango H, Frayling TM, Neumann RJ, Sherva R, Blech I, Pharoah PD, Palmer CN, Kimber C, Tavendale R, Morris AD, McCarthy MI, Walker M, Hitman G, Glaser B, Permutt MA, Hattersley AT, Wareham NJ, Barroso I: Common variants in WFS1 confer risk of type 2 diabetes. *Nat Genet* 39:951-953, 2007

165. Qi L, Kraft P, Hunter DJ, Hu FB: The Common Obesity Variant near MC4R Gene Is Associated with Higher Intakes of Total Energy and Dietary Fat, Weight Change, and Diabetes Risk in Women. *Hum Mol Genet* 2008

166. Kahn SE: Clinical review 135: The importance of beta-cell failure in the development and progression of type 2 diabetes. *J Clin Endocrinol Metab* 86:4047-4058, 2001

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

168. Poitout V, Robertson RP: Glucolipotoxicity: fuel excess and beta-cell dysfunction. *Endocr Rev* 29:351-366, 2008

169. Kharroubi I, Ladriere L, Cardozo AK, Dogusan Z, Cnop M, Eizirik DL: Free fatty acids and cytokines induce pancreatic beta-cell apoptosis by different mechanisms: role of nuclear factor-kappaB and endoplasmic reticulum stress. *Endocrinology* 145:5087-5096, 2004

170. Huang G, Yao J, Zeng W, Mizuno Y, Kamm KE, Stull JT, Harding HP, Ron D, Muallem S: ER stress disrupts Ca2+-signaling complexes and Ca2+ regulation in secretory and muscle cells from PERK-knockout mice. *J Cell Sci* 119:153-161, 2006

171. Herbert TP: PERK in the life and death of the pancreatic beta-cell. *Biochem Soc Trans* 35:1205-1207, 2007

172. Freely associating. *Nat Genet* 22:1-2, 1999

173. Cardon LR, Bell JI: Association study designs for complex diseases. *Nat Rev Genet* 2:91-99, 2001

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

175. Reich D, Thangaraj K, Patterson N, Price AL, Singh L: Reconstructing Indian population history. *Nature* 461:489-494, 2009

176. Sandhu MS, Weedon MN, Fawcett KA, Wasson J, Debenham SL, Daly A, Lango H, Frayling TM, Neumann RJ, Sherva R, Blech I, Pharoah PD, Palmer CN, Kimber C, Tavendale R, Morris AD, McCarthy MI, Walker M, Hitman G, Glaser B, Permutt MA, Hattersley AT, Wareham NJ, Barroso I: Common variants in WFS1 confer risk of type 2 diabetes. *Nat Genet* 39:951-953, 2007

177. Vaxillaire M, Benmezroua Y, Durand E, Vasseur F, Froguel P: No evidence for diabetes-associated mutations of PEK/EIF2AK3 gene in French patients with early-onset type II diabetes. *Diabetologia* 44:786, 2001

178. Li B, Leal SM: Discovery of rare variants via sequencing: implications for the design of complex trait association studies. *PLoS Genet* 5:e1000481, 2009

179. Allotey RA, Mohan V, McDermott MF, Deepa R, Premalatha G, Hassan Z, Cassell PG, North BV, Vaxillaire M, Mein CA, Swan DC, O'Grady E, Ramachandran A, Snehalatha C, Sinnot PJ, Hemmatpour SK, Froguel P, Hitman GA: The EIF2AK3 gene region and type I diabetes in subjects from South India. *Genes Immun* 5:648-652, 2004

180. Hurst CD, Zuiverloon TC, Hafner C, Zwarthoff EC, Knowles MA: A SNaPshot assay for the rapid and simple detection of four common hotspot codon mutations in the PIK3CA gene. *BMC Res Notes* 2:66, 2009

181. Tobler AR, Short S, Andersen MR, Paner TM, Briggs JC, Lambert SM, Wu PP, Wang Y, Spoonde AY, Koehler RT, Peyret N, Chen C, Broomer AJ, Ridzon DA, Zhou H, Hoo BS, Hayashibara KC, Leong LN, Ma CN, Rosenblum BB, Day JP, Ziegle JS, De L, V, Rhodes MD, Hennessy KM, Wenz HM: The SNPlex genotyping system: a flexible and scalable platform for SNP genotyping. *J Biomol Tech* 16:398-406, 2005

182. Tan JT, Ng DP, Nurbaya S, Ye S, Lim XL, Leong H, Seet LT, Siew WF, Kon W, Wong TY, Saw SM, Aung T, Chia KS, Lee J, Chew SK, Seielstad M, Tai ES: Polymorphisms Identified through Genome-Wide Association Studies and Their Associations with Type 2 Diabetes in Chinese, Malays, and Asian-Indians in Singapore. *J Clin Endocrinol Metab* 2009

183. Kidd JM, Cooper GM, Donahue WF, Hayden HS, Sampas N, Graves T, Hansen N, Teague B, Alkan C, Antonacci F, Haugen E, Zerr T, Yamada NA, Tsang P, Newman TL, Tuzun E, Cheng Z, Ebling HM, Tusneem N, David R, Gillett W, Phelps KA, Weaver M, Saranga D, Brand A, Tao W, Gustafson E, McKernan K, Chen L, Malig M, Smith JD, Korn JM, McCarroll SA, Altshuler DA, Peiffer DA, Dorschner M, Stamatoyannopoulos J, Schwartz D, Nickerson DA, Mullikin JC, Wilson RK, Bruhn L, Olson MV, Kaul R, Smith DR, Eichler EE: Mapping and sequencing of structural variation from eight human genomes. *Nature* 453:56-64, 2008

184. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, Gonzalez JR, Gratacos M, Huang J, Kalaitzopoulos D, Komura D, Macdonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, Zhang J, Zerjal T, Zhang J, Armengol L, Conrad DF, Estivill X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurles ME: Global variation in copy number in the human genome. *Nature* 444:444-454, 2006

185. Fan JB, Chee MS, Gunderson KL: Highly parallel genomic assays. *Nat Rev Genet* 7:632-644, 2006

186. Porreca GJ, Zhang K, Li JB, Xie B, Austin D, Vassallo SL, LeProust EM, Peck BJ, Emig CJ, Dahl F, Gao Y, Church GM, Shendure J: Multiplex amplification of large sets of human exons. *Nat Methods* 4:931-936, 2007

187. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM: Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376-380, 2005

188. Rakyan VK, Beck S: Epigenetic variation and inheritance in mammals. *Curr Opin Genet Dev* 16:573-577, 2006

189. Rakyan VK, Down TA, Thorne NP, Flicek P, Kulesha E, Graf S, Tomazou EM, Backdahl L, Johnson N, Herberth M, Howe KL, Jackson DK, Miretti MM, Fiegler H, Marioni JC, Birney E, Hubbard TJ, Carter NP, Tavare S, Beck S: An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs). *Genome Res* 18:1518-1529, 2008

190. Yajnik CS, Deshmukh US: Maternal nutrition, intrauterine programming and consequential risks in the offspring. *Rev Endocr Metab Disord* 9:203-211, 2008

191. Moving AHEAD with an international human epigenome project. *Nature* 454:711-715, 2008

192. Klose RJ, Bird AP: Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci* 31:89-97, 2006

193. Liu ZJ, Maekawa M: Polymerase chain reaction-based methods of DNA methylation analysis. *Anal Biochem* 317:259-265, 2003

194. Tost J, Gut IG: DNA methylation analysis by pyrosequencing. *Nat Protoc* 2:2265-2275, 2007

195. Khulan B, Thompson RF, Ye K, Fazzari MJ, Suzuki M, Stasiek E, Figueroa ME, Glass JL, Chen Q, Montagna C, Hatchwell E, Selzer RR, Richmond TA, Green RD, Melnick A, Greally JM: Comparative isoschizomer profiling of cytosine methylation: the HELP assay. *Genome Res* 16:1046-1055, 2006

196. Vucic EA, Wilson IM, Campbell JM, Lam WL: Methylation analysis by DNA immunoprecipitation (MeDIP). *Methods Mol Biol* 556:141-153, 2009

197. Pomraning KR, Smith KM, Freitag M: Genome-wide high throughput analysis of DNA methylation in eukaryotes. *Methods* 47:142-150, 2009

198. Tost J, Schatz P, Schuster M, Berlin K, Gut IG: Analysis and accurate quantification of CpG methylation by MALDI mass spectrometry. *Nucleic Acids Res* 31:e50, 2003

199. Chapal N, Molina L, Molina F, Laplanche M, Pau B, Petit P: Pharmacoproteomic approach to the study of drug mode of action, toxicity, and resistance: applications in diabetes and cancer. *Fundam Clin Pharmacol* 18:413-422, 2004

200. Sebedio JL, Pujos-Guillot E, Ferrara M: Metabolomics in evaluation of glucose disorders. *Curr Opin Clin Nutr Metab Care* 12:412-418, 2009

201. Li G, Zhang P, Wang J, Gregg EW, Yang W, Gong Q, Li H, Li H, Jiang Y, An Y, Shuai Y, Zhang B, Zhang J, Thompson TJ, Gerzoff RB, Roglic G, Hu Y, Bennett PH: The long-term effect of lifestyle interventions to prevent diabetes in the China Da Qing Diabetes Prevention Study: a 20-year follow-up study. *Lancet* 371:1783-1789, 2008

202. Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, Nathan DM: Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med* 346:393-403, 2002

203. Tuomilehto J, Lindstrom J: The major diabetes prevention trials. *Curr Diab Rep* 3:115-122, 2003

ACKNOWLEDGEMENTS:

This work was carried out whilst I was the Diabetes Research Fellow at the Centre for Diabetes and Metabolic Medicine, Barts and The London School of Medicine and Dentistry and a Specialist Registrar in Diabetes and Endocrinology in North Western Deanery, Manchester (2002- 2008) under the direction of Professor Hitman. I am indebted for his continued support and encouragement. I carried out all the experimental work and analysis together with Dr. Rebecca Allotey at the Centre for Diabetes and Metabolic Medicine unless stated elsewhere below. I would also like to thank the input and support from Dr. Paul Cassell and Mrs Irene Smith (Diabetes centre administrator) from our research unit.

The London based clinical collections were conducted within the catchment areas of the Royal London Hospital and Newham General Hospital, with the support and supervision of Professor Graham Hitman, Dr. Shanti Vijayaragahavan and Professor Susan Gelding. I would also like to thank Professor A. Hattersley and his research group at Exeter for their help in our work (DNA samples, Quality control). I would like to thank the Warren 2 UK type 2 diabetes consortium (Professor A. Hattersley, Dr. T. Frayling, Dr. M. Sampson, Dr. S. O. Rahilly, Professor M. McCarthy, Professor M. Walker, Dr. A. Morris, Dr. C. Palmer) for their support. Most of sequencing was done by Dr. Rebecca Allotey and only a few runs by me. Analysis of the sequencing was done jointly with Dr. Allotey. The majority of genotyping work was out sourced to KBiosciences apart from the work done at the Genome centre by me. I received good support and technical advice from Dr. Charles Mein and his colleagues at the Genome centre in Charterhouse Square. I did the majority of data cleaning and data basing prior to the analysis. Advice and help on statistics was provided by Dr. Anna Vine and Professor David Curtis (Joint Academic Department of Psychological Medicine). Thanks to Diabetes UK who helped fund the project and to all the patients with diabetes, families and healthy volunteers without whose kind contribution this body of research would not have been possible.

I would like to thank my family for their support and being a source of inspiration.