Chapter 2

Materials and Methods

2.1 Definition of cohorts

2.1.1 Severe Insulin Resistance (SIR) cohort

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

2.1.2 Control panels

2.1.2.1 CEPH

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

2.1.2.2 HGDP-CEPH Human Genome Diversity Cell Line Panel

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

2.1.2.3 European-Indian control panel

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

2.1.3 Case-control populations

2.1.3.1 Cambridgeshire case-control

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

2.1.3.2 EPIC- Norfolk case-control study

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

2.1.3.3 Exeter case-control study

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

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

2.1.3.4 ADDITION case-control study

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

2.1.4 Population-based populations

2.1.4.1 MRC Ely study

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

2.1.4.2 Hertfordshire cohort study

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

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

2.2 Reagents

2.2.1 DNA preparation

20 X TE

200 mM Tris-HCL pH7.5

20 mM EDTA

2.2.2 Polymerase Chain Reaction (PCR) and sequencing

Loading buffer

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

Reaction buffer

200 mM Tris-HCL pH8.0

100 mM $MgCl_2$

Dilution buffer

50 mM Tris-HCL pH8.0

Sequencing mix

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

Precipitation mix

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

2.2.3 Sequenom reagents

10 X PCR Buffer

10 mM Tris-HCl

500 mM KCl

15 mM $MgCl_2$

0.01 % (w/v) gelatine

10 X Thermosequenase buffer

260 mM Tris-HCl pH: 9.5

65 mM $MgCl_2$

2.3 Protocols

2.3.1 DNA preparation

2.3.1.1 Whole-genome amplification

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

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

2.3.1.2 Quantification of DNA using Pico-Green

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

2.3.2 PCR

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

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

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

2.3.3 Gel electrophoresis

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

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

2.3.4 DNA purification

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

2.3.5 Sequencing

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

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

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

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

2.3.6 Sequence analysis

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

2.3.7 Genotyping

2.3.7.1 Sequenom

2.3.7.1.1 MassArray Homogeneous MassEXTEND (hME) assay

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

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

6) 72°C for 3 minutes

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

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

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

2.3.7.1.2 MassArray iPLEX assay

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

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

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

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

2.3.7.2 Taqman

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

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

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

2.3.8 Statistical analysis

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