Genetics of Type 2 Diabetes and Metabolic Syndrome: From Genome Wide Linkage Scan and Candidate Genes to Genome Wide Association Studies

Sjögren, Marketa

2008

Genetics of Type 2 Diabetes and Metabolic Syndrome: From Genome Wide Linkage Scans and Candidate Genes to Genome Wide Association Studies

Academic Dissertation

Marketa Sjögren

Lund University
Department of Clinical Sciences Malmö, Diabetes & Endocrinology
Clinical Research Centre, Malmö University Hospital

LUND
UNIVERSITY

With the permission of the Medical Faculty of Lund University, to be presented for public examination in the Grand Hall at the Medical Research Centre, Entrance 59, Malmö University Hospital, on December 5th, 2008, at 01.15 p.m.

Faculty Opponent
Professor Ann-Christine Syvänen
Molecular Medicine
Department of Medical Sciences
Uppsala University
Sweden

ISSN 1652-8220
ISBN 978-91-86059-78-1
Lund University, Faculty of Medicine Doctoral Dissertation Series 2008:125

Printed by Media-Tryck, Lund University, Lund, Sweden

Cover: Chromosomes created by Zuzana Kolkova, 2008.

” Every flower of tomorrow is in the seeds of today”
Chinese proverb

Organization
LUND UNIVERSITY

Department of Clinical Sciences, Malmö
Diabetes & Endocrinology
CRC, Malmö University Hospital
Sweden

Document name
DOCTORAL DISSERTATION

Date of issue
December 5th, 2008

Sponsoring organization

Author(s)
Marketa Sjögren

Title and subtitle

Genetics of Type 2 Diabetes and Metabolic syndrome

Abstract

Type 2 diabetes (T2D) and metabolic syndrome (MetS) are major health problems associated with cardiovascular disease. Both diseases are influenced by a combination of genetic and environmental factors. The aim of this thesis was to identify genetic risk factors for T2D, particularly T2D associated with obesity, and for MetS. To achieve this goal, we 1) followed-up a region linked to obese T2D in an earlier genome wide linkage study; 2) studied IRS1 as a candidate gene for T2D and MetS and 3) studied the role of 28 candidate genes for T2D, earlier identified in candidate gene and genome wide association studies (GWAS), in MetS.
In study I, extensive fine mapping by genotyping 501 haplotype tag SNPs in 27 genes within the 15 Mb region linked to obese T2D on chromosome 18p was undertaken in Botnia 2 family material. Nominal association ($p < 0.05$) between T2D, obese T2D and/or BMI in T2D patients and 74 SNPs was identified. Five of the strongest associated SNPs were selected for replication in a large Swedish T2D case-control material. In addition, six SNPs indicating strongest association in three GWAS (DGI, WTCCC and Mexican Americans) with T2D, obese T2D, BMI in T2D patients or fat mass as well as a SNP in the lipin 2 gene showing association with T2D in a Dutch population (rs3745012, $p = 0.03$) were pursued further in the same cohort. However, no significant association was observed between the selected SNPs in the case-control material. These findings argue against common variation contributing to the observed linkage on chromosome 18p.
In studies II and III, the role of polymorphisms in IRS1 in T2D risk was investigated in > 9 000 individuals. We found no association between T2D and G972R nor any of the 20 tag SNPs capturing 85% of the common variation in IRS1. These findings argue against any major involvement of common polymorphisms in IRS1 in the development of T2D.
In studies IV and V, we investigated the role of candidate or susceptibility genes for T2D or components of MetS in the development of MetS. In the family based Botnia prospective study (study IV) we found that polymorphisms in PPARG and ADRB1 predicted development of MetS, indicating a role of altered free fatty acid metabolism in the pathogenesis of MetS. In study V, polymorphisms in candidate genes for T2D (TCF7L2, WFS1, IGF2BP2) and obesity (FTO) predicted development of MetS and the risk to develop MetS seemed to be driven by associations with the previously reported phenotypes. These data do not support the view that the different components of MetS share a common genetic background.

Key words: Type 2 Diabetes, Metabolic Syndrome, Obesity, SNP, Genetics, Genome Wide Linkage Study, Association Study, Prospective Study, IRS1, Candidate Gene, GWAS, Polymorphism, tag SNP

Classification system and/or index termes (if any):

Supplementary bibliographical information:

Language
English

ISSN and key title:
1652-8220

ISBN
978-91-86059-78-1

Recipient's notes

Number of pages
122

Price

Security classification

TABLE OF CONTENTS

LIST OF PAPERS INCLUDED IN THE THESIS

I **Sjögren M,** Almgren P, Svensson M, Perola M, Peltonen L, Groop L and Orho-Melander M: Genetic dissection of chromosome 18p11, a region linked to obese T2D and HDL-cholesterol. *Manuscript*

II Florez JC, **Sjögren M,** Burtt N, Orho-Melander M, Schayer S, Sun M, Almgren p, Lindblad U, Tuomi T, Gaudet D, Hudson TJ, Daly MJ, Ardlie KG, Hirschorn JN, Altshuler D and Groop L: Association testing in 9000 people fails to confirm the association of the insulin receptor substrate-1 G972R polymorphism with type 2 diabetes. *Diabetes* 53: 3313-3318, 2004

III Florez JC, **Sjögren M**, Agapakis CM, Burt NP, Almgren P, Lindblad U, Berglund G, Tuomi T, Gaudet D, Daly MJ, Ardlie KG, Hirschhorn JN, Altshuler D, Groop L: Association testing of common variants in the insulin receptor substrate-1 gene (IRS1) with type 2 diabetes. *Diabetologia* 50:1209-17, 2007

IV Lyssenko V, **Sjögren M,** Almgren P, Anevski D, Orho-Melander M, Tuomi T, Taskinen M-R and Groop L for the Botnia Study Group: Genetic prediction of the metabolic syndrome. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews, in press, Available online 15 August 2008*

V **Sjögren M**, Lyssenko V, Jonsson A, Berglund G, Nilsson P, Groop L, Orho-Melander M: The search for putative unifying genetic factors for components of the metabolic syndrome.
Diabetologia, in press, Available online 14 October 2008

LIST OF PAPERS NOT INCLUDED IN THE THESIS

Demenais F, Kanninen T, Lindgren CM, Wiltshire S, Gaget S, Dandrieux C, Almgren P, **Sjögren M**, Hattersley A, Dina C, Tuomi T, McCarthy MI, Froguel P, Groop LC: A meta-analysis of four European genome screens (GIFT Consortium) shows evidence for a novel region on chromosome 17p11.2-q22 linked to type 2 diabetes. *Hum Mol Genet*. 12:1865-1873, 2003

Lyssenko V, Almgren P, Anevski D, Orho-Melander M, **Sjögren M**, Saloranta C, Tuomi T, Groop L, the Botnia Study Group. Genetic prediction of future type 2 diabetes. *PloS Med* 12: 1299-1308, 2005

Lindholm E, Bakhtadze E, **Sjögren M**, Cilio CM, Agardh E, Groop L, Arardh C-D. The –374 T/A polymorphism in the gene encoding RAGE is associated with diabetic nephropathy and retinopathy in type 1 diabetic patients. *Diabetologia* 49: 2745-2755, 2006

Saxena R, Gianniny L, Burtt NP, Lyssenko V, Giuducci C, **Sjögren M**, Florez JC, Almgren P, Isomaa B, Orho-Melander M, Lindblad U, Daly M, Tuomi T, Hirschhorn J, Groop L and Altshuler D. Common SNPs in *TCF7L2* are reproducibly associated with type 2 diabetes and reduce the insulin response to glucose in non-diabetic individuals. *Diabetes* 55: 2890-2895, 2006

Fredriksson J, Anevski D, Almgren P, **Sjögren M**, Lyssenko V, Carlson J, Isomaa B, Taskinen M-R, Groop L, Orho-Melander M for the Botnia Study Group. Polymorphisms in the apolipoprotein E and glycogen synthase genes on chromosome 19q13 predict cardiovascular mortality. *PLoS ONE* 14;2:e285, 2007

Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research, Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ, Hughes TE, Groop L, Altshuler D, Almgren P, Florez JC, Meyer J, Ardlie K, Bengtsson Boström K, Isomaa B, Lettre G, Lindblad U, Lyon HN, Melander O, Newton-Cheh C, Nilsson P, Orho-Melander M, Råstam L, Speliotes EK, Taskinen MR, Tuomi T, Guiducci C, Berglund A, Carlson J, Gianniny L, Hackett R, Hall L, Holmkvist J, Laurila E, **Sjögren M**, Sterner M, Surti A, Svensson M, Svensson M, Tewhey R, Blumenstiel B, Parkin M, Defelice M, Barry R, Brodeur W, Camarata J, Chia N, Fava M, Gibbons J, Handsaker B, Healy C, Nguyen K, Gates C, Sougnez C, Gage D, Nizzari M, Gabriel SB, Chirn GW, Ma Q, Parikh H, Richardson D, Ricke D, Purcell S. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 316:1331-6, 2007

Lyssenko V, Lupi R, , Marchetti P, Del Guerra S, Orho-Melander M**,** Almgren P**, Sjögren M**, Ling C, Eriksson K-E, Lethagen Å, Mancarella R, Berglund G, Tuomi T,

Nilsson P, Del Prato S, Groop L, Mechanisms by which common variants in the *TCF7L2* gene increase risk of type 2 diabetes. *J Clin Invest*. 117:2155-63, 2007

Nilsson E, Poulsen P**, Sjögren M**, Ling C, Ridderstråle M, Groop L, Vaag A. Regulation of skeletal muscle PPARδ mRNA expression in twins, *J Physiol*. 584: 1011-1017, 2007

Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, de Bakker PIW, Abecasis GR, Almgren P, Andersen G, Ardlie K, Bengtsson Boström K, Bergman RN, Bonnycastle LL, Borch-Johnsen K, Burtt NP, Chen H, Chines PS, Daly MJ, Deodhar P, Ding C-J, Doney ASF, 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, Jørgensen 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 CNA, Payne F, Perry JRB, Pettersen E, Platou C, Prokopenko I, Qi L, Qin L, Rayner NW, Rees M, Roix JJ, Sandbæk A, Shields B, **Sjögren 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, Wellcome Trust Case Control Consortium, 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. *Nature Genetics* 40: 638-45, 2008

Brorsson C, Bergholdt R, **Sjögren M**, Eising S, Meden Sørensen K, Hougaard DM, Orho-Melander M, Groop L and Pociot F, A non-synonymous variant in *SLC30A8* is not associated with type 1 diabetes in the Danish population. *Mol Gen Metab*. 94: 386-8, 2008

ABBREVIATIONS

ADAMTS9	A Disintegrin-like and Metalloproteinase with Thrombospondin Type 1 Motif
ADRB1	Adrenergic β-receptor 1
ADRB2	Adrenergic β-receptor 2
ADRB3	Adrenergic β-receptor 3
ARMS	Amplification Refractory Mutation System
AUC	Area Under the Curve
BMI	Body Mass Index
bp	base pairs
CAMK1D	Calcium/Calmodulin-dependent Protein Kinase 1-delta
CAPN10	Calpain 10
CDKAL1	CDK5 Regulatory Subunit-Associated Protein 1-like 1
CDKN2A	Cyclin-dependent Kinase Inhibitor 2A
CI	Confidence Interval
cM	Centimorgan
CNV	Copy Number Variation
CVD	Cardiovascular Disease
ddNTP	Dideoxynucleotide Triphosphate
DM	Diabetes Mellitus
DNA	Deoxyribonucleic Acid
dNTP	Deoxynucleotide Triphosphate
FPG	Fasting Plasma Glucose
FRET	Fluorescence Resonance Energy Transfer
FTO	Fat Mass- and Obesity-Associated Gene
GCI	Genomics Collaborative, Inc
GCKR	Glucokinase Regulatory Protein
GEE	Generalised Estimating Equation
HDL	High Density Lipoprotein
HGP	Human Genome Project
HHEX	Hematopoetically Expressed Homeobox
HOMA-β	Homeostasis Model Assessment of β-cell function
HOMA-IR	Homeostasis Model Assessment of Insulin Resistance
HR	Hazard Ratio
IBD	Identical by Descent
IBS	Identical by State
IDF	International Diabetes Federation
IFG	Impaired Fasting Glucose
IGF2BP2	Insulin-like Growth Factor 2 mRNA-Binding Protein 2
IGT	Impaired Glucose Tolerance
IRS1	Insulin Receptor Substrate 1
JAZF1	Juxtaposted with Another Zinc Finger Gene
KCNJ11	Potassium Channel, Inwardly Rectifying, Subfamily J, Member 11

LADA	Latent Autoimmune Diabetes in Adults
LCR	Low Copy Repeat
LD	Linkage Disequilibrium
LGR5	Leucine-rich Repeat-containing G Protein-coupled Receptor 5
LINE	Long Interspersed Nuclear Elements
LOD	Logarithm of odds
LPIN2	Lipin 2
MAF	Minor Allele Frequency
MALDI-TOF MS	Matrix-Assisted Laser Desorption Ionisation-Time of Flight Mass Spectrometry
MetS	Metabolic Syndrome
MIDD	Maternally Inherited Diabetes and Deafness
MODY	Maturity Onset Diabetes of the Young
MPP	Malmö Preventive Project
NCEP-ATP III	National Cholesterol Education Program Adult Treatment Panel III
NGT	Normal Glucose Tolerance
NOTCH2	Notch Homolog 2
OGTT	Oral Glucose Tolerance Test
OR	Odds Ratio
PCR	Polymerase Chain Reaction
PPARG	Peroxisome Proliferator-activated Receptor Gamma
SBE	Single Base Extension
SINE	Short Interspersed Nuclear Elements
SLC30A8	Solute Carrier Family 30 (zinc transporter) Member 8
SNP	Single Nucleotide Polymorphism
STR	Short Tandem Repeats
SUR1	Sulfonylurea Receptor 1
T1D	Type 1 Diabetes
T2D	Type 2 Diabetes
TCF7L2	Transcription Factor 7-like 2
THADA	Thyroid Adenoma Associated
TSPAN8	Tetraspanin 8
TZD	Thiazolidinedione
WFS1	Wolframin
WHO	World Health Organization
WHR	Waist-to-Hip Ratio

INTRODUCTION

Genetics

Genetics studies heredity, or the transmission of traits between generations. The first groundbreaking study was conducted in 1865 by Gregor Mendel, who studied garden peas. The result of his work lead to the postulation of Mendel's laws of inheritance, stating that traits (phenotypes) are transmitted independently and in equal frequencies from one generation to the next. The molecular basis for the transmission of traits is now known and the discovery of the molecule encoding the genetic information, the deoxyribonucleic acid (DNA), and characterization of the double helix structure of the DNA, are considered as major milestones in the field of molecular genetics.

Human Genome

The human genome consists of 3.1 billion base pairs (bp) organized in 22 pairs of autosomal and two sex specific chromosomes. The first draft of the sequenced genome was published in 2001 [1,2] and the complete sequence, covering 99%, was published by the Human Genome Project (HGP) in 2004 [3]. It has been estimated that the human genome contains 20 000 – 25 000 protein coding genes [3].

Genetic variation

Single Nucleotide Polymorphisms

Approximately 0.1% of the genome differs between any two unrelated individuals. A position where two, or in rare circumstances more than two, alternative bases are present in different individuals in one nucleotide position of the genome is called a single nucleotide polymorphism (SNP) and represent the most abundant variation in human genome [4]. The number of SNPs is estimated to be around 10 million, i.e. on average one SNP every 300 bp [5,6]. Depending on the location the SNP may or may not have functional consequences. If a polymorphism is located within a coding region of a protein, it can either alter the amino acid sequence (a non-synonymous SNP) which can in some cases change the structure or function of the protein; or the variant can be silent (synonymous SNP), i.e. it does not change the amino acid sequence of the protein. Most of the SNPs are located outside the coding regions of genes and are thus referred to as non-coding SNPs. These variants may be functional if they are located in functional elements, such as promoters, silencer or enhancers, in which case they may alter expression of a gene or several genes.

SNPs which are abundant in the genome and have high frequency in the population are commonly used as bi-allelic markers in association studies as they can relatively easily be detected using different genotyping methods (see page 16).

Short Tandem Repeats

Short tandem repeats (STR), also known as microsatellites, are 2-4 nucleotide sequences with variable number of repeats, the most common being a various number of CA repeats $(CA)_n$, ranging from below ten to over hundred repeats. Microsatellites have traditionally been used in genome-wide linkage studies in families because they are highly polymorphic and relatively stable.

Structural variants

Repetitive elements have been shown to compose > 50% of the human genome. Such elements can be divided into different subclasses. A majority of them are derived from transposable elements. The most common classes include long interspersed nuclear elements (LINE), short interspersed nuclear elements (SINE) and segmental duplications or low copy repeats (LCR). LINEs make up about 20% of the genome and are capable of autonomous mobilization via an RNA intermediate. The transposones are 6 kb long but many of the SINEs are truncated and the average length is 900 kb. SINEs are non-autonomous sequences of 100 – 400 bp and make up about 13% of the genome [1,7]. Segmental duplications are defined as DNA sequences of 1 kb and upward with at least 90% homology and present in the genome in at least two copies. These variation comprise ~5% of the genome [8].

Copy number variation (CNV) is defined as a DNA segment 1 kb or larger that is present at a variable number of copies in the genome [9].

Linkage Disequilibrium

Genetic markers located close to each other are often inherited together more frequently than would be expected by chance i.e. if they segregated independently in a population. This phenomenon is described as linkage disequilibrium (LD) or allelic association. Two statistical measures are typically used to describe LD. D´ estimates the number of recombination events and ranges from zero to one [10]. Two markers are considered to be in complete LD if D´ equals one, indicating that no recombination occurred in the studied population. Another measure, r^2, denotes statistical correlation between the markers and is related to allele frequency and can range between zero and one. When two alleles of two SNPs are always observed together, the r^2 equals one and the SNPs are in perfect LD. The measures of LD, specifically r^2, are used to identify haplotype tagging SNPs (tag SNPs), that is SNPs that capture the genetic information at a nearby locus and therefore can reduce the amount of genotyping needed to identify the genetic variation.

Haplotypes

A haplotype is a linear order of alleles linked on a chromosome and inherited together. Haplotypes are defined by the LD structure of a specific DNA segment and can be utilized for genetic studies. Large portions of the genome are organized in haplotype blocks, i.e. regions of high LD interspersed by recombination hot-spots. The size of haplotype blocks varies between populations, the average size being 11 kb in populations with African ancestry and 22 kb in European and Asian populations [11].

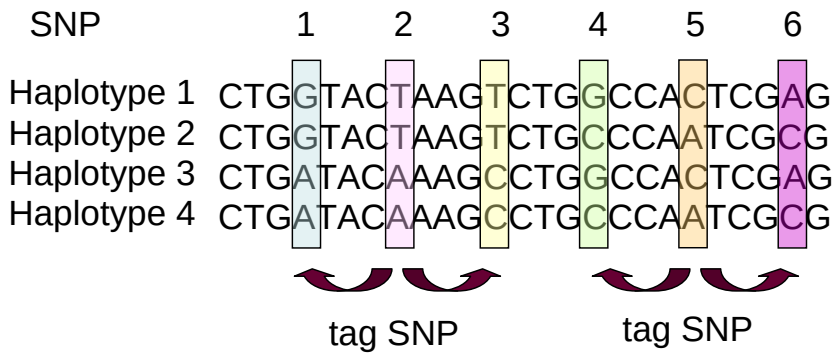


Figure 1. Schematic representation of haplotypes. SNPs are highlighted in colour. Groups of three SNPs (SNPs 1-3 and SNPs 4-6) show high allelic association and each SNP in the group can therefore serve as tag SNP for the other two SNPs. Addapted from [12].

SNP genotyping technologies

SNP genotyping methods have undergone rapid development in the last decades, increasing both in multiplexing capacity and accuracy. In the following section principles of the most commonly used methods are described; methods used in this thesis are described in more detail.

Hybridization methods

Hybridization methods are based on the strength of binding between short complementary strands of DNA and the ability of allele specific oligonucleotides to detect single base mismatches. Hybridization methods are used in large variety of genotyping methods including, homogenous hybridization methods, e.g. TaqMan™ (Applied Biosystems, Foster City, CA, USA) as well as high density arrays, e.g. Affymetrix and Illumina. The chip based methods are used in genome wide association studies (GWAS) because their high throughput and possibility of high

multiplexing enable relatively easy genotyping of large number of SNPs in large sample sizes at reasonable costs.

Taqman

The TaqMan™ (Applied Biosystems, Foster City, CA, USA) allelic discrimination assay is based on 5'-exonuclease assay. The allele specific probes are labelled with two different fluorophores (FAM™ and VIC™) at the 3'-ends and with a quencher molecule at the 5'-ends. In the intact probe the close proximity of the quencher interacts with the reporter dye and reduces the fluorescence signal. The TaqMan™ probes have been modified by minor groove binder (MGB) molecule in order to improve their ability for allelic discrimination. During polymerase chain reaction (PCR) the perfect match probe hybridizes to the PCR products and during the extension phase the 5'-3' exonuclease activity of the TaqDNA polymerase degrades the perfectly matched probes, separating the fluorophore from the quencher leading to increased fluorescence, taking advantage of the fluorescence resonance energy transfer (FRET). The mismatched probe is thus displaced without cleavage. As the probes can be distinguished by the different fluorophores, emission of only one of the signals indicates homozygosity while emission of both signals indicates heterozygosity [13,14]. The TaqMan™ allelic discrimination assays are detected on an ABI 7900HT sequence detection system (Applied Biosystems, Foster City, CA, USA).

Allele specific PCR

PCR primers complementary at the 3'-end to either nucleotide combined with a common reverse primer allow alleles to be selectively amplified in a technique called amplification refractory mutation system (ARMS) [15]. A commonly used method to detect the allele specific PCR products is by addition of fluorescent dye into the reaction mixture.

Kaspar assay

The KASPar® genotyping system (KBioscience, Hoddesdon, UK) utilizes competitive allele specific PCR with two allele specific forward primers and one common reverse primer different from conventional ARMS (patent pending) and taking advantage of FRET [16]. The emitted fluorescence can be detected on ABI 7900HT sequence detection system (Applied Biosystems, Foster City, CA, USA).

Single Base Extension

The single base extension (SBE) is a DNA polymerase assisted method. A primer that anneals adjacent to the SNP with its 3' end is elongated by a single labelled deoxynucleoside triphosphate (dNTP) or dideoxynucleoside triphosphate (ddNTP) by a DNA polymerase [17]. This method was originally called minisequencing [18] and can be coupled with several detection steps. The SBE method is today implemented in many assays enabling multiplexing, including iPLEX™ and SNaPshot assays described in more detail below.

iPLEX™ assay

The iPLEX™ assay, based on SBE, is terminated after incorporation of only one base by ddNTPs. The primer extension products are separated by mass by matrix-assisted laser desorption ionisation-time of flight mass spectrometry (MALDI-TOF MS) on the MassARRAY platform (Sequenom, San Diego, CA, USA) [19-21]. To increase the mass differences between alleles, specific mass-modified terminators are used. The assay can be multiplexed by adding non-template nucleotides to the 5'-end of the extension primer and up to 36 SNPs can be pooled.

SNaPshot assay

Another assay based on SBE is the SNaPshot assay (Applied Biosystems, Foster City, CA, USA), which utilizes four differentially fluorescently labelled ddNTPs terminating the extension reaction. The primer extension products are detected on DNA-sequencing instrument (in our case ABI3100, Applied Biosystems, Foster City, CA, USA), where the electrophoretic size separating step enables multiplexing. By addition of nucleotide tails of varying length to the 5'-end of the extension primer up to ten SNPs can be analyzed together.

Genetic analysis of multifactorial diseases

A combination of several genetic (i.e. polygenic) and environmental factors predispose to multifactorial, complex diseases. Multifactorial diseases lack clear inheritance patterns as many genes with often small effects may contribute to the development of the disease and environmental factors may modify the risk. Uncovering genes predisposing to complex diseases has proven difficult. Mainly two strategies have been used to search for genetic variation increasing the risk for complex diseases, linkage and association studies.

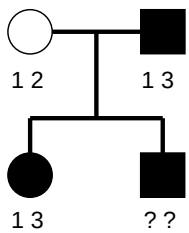
Heritability

The proportion of phenotypic variance that is explained by a genetic component is demarcated as heritability (h^2) for a specific trait. Monogenic traits have usually high heritability (often 100%) while only a smaller fraction of complex diseases can be explained by genetic factors.

Linkage studies

Linkage is used to analyze the co-segregation of two or more loci (genes or traits) in a family to verify whether loci segregate independently or tend to be inherited together more often than expected by chance. Two loci are considered linked if they are transmitted from a parent to offspring more frequently than expected, i.e. if recombination occurs between them with the probability of significantly less than 50% [22]. Two main analysis methods are used, parametric and non-parametric linkage analysis [23]. Parametric methods are used to identify genetic loci co-segregating in

pedigrees and these methods rely on the recombination fraction θ, defined as probability of recombination between two loci during meiosis. Linkage is usually reported as a logarithm of odds (LOD) score, which is a function of θ and the disease locus or map position measured in centimorgans (cM) [24]. A LOD score of 3 is considered as significant evidence for linkage while LOD score of -2 or less indicate evidence against linkage. Parametric linkage analysis requires prior knowledge about the inheritance model and has been successful in identifying genes predisposing to monogenic traits with Mendelian mode of inheritance. Non-parametric methods are used for complex traits and make no assumption about the inheritance model. These methods are based on the expectation that affected individuals share copies of chromosomal loci (haplotypes) identical by descent (IBD) in a region containing genetic variation predisposing to the disease more often than expected by chance, irrespective of the mode of inheritance. Assuming no linkage, two siblings have a 0.25 probability to not share any allele IBD, 0.5 probability to share one allele and 0.25 probability to share both alleles IBD at a certain locus. Linkage is indicated when affected siblings or other family members share significantly more alleles IBD than expected [23]. Genome wide linkage studies have been performed usually genotyping approximately 400 microsatellite markers distributed throughout the genome, ca 10 cM apart, in families to track recombination and identify putative disease predisposing loci. The threshold for significance for non-parametric genome wide linkage studies has been proposed at LOD score > 3.6 (p-value = 0.00002) to encounter false positive results at no more than 5% rate [25]. In contrast to monogenic disease, linkage studies for complex diseases have not been particularly successful. This might be due to the modest contribution of each gene and the lack of power to detect such small effects in the small sample sizes used [26].



Sibling 1	Sibling 2	IBD	IBS
1 3	1 1	1	1
	1 3	2	2
	2 1	0	1
	2 3	1	1

Figure 2. Allele sharing in sibling pairs. The table indicates number of alleles shared IBD or IBS depending on the genotype of sibling two.

Association studies

Association analysis can be used to detect relationship between alleles and disease by comparing frequencies of alleles between affected individuals and matched healthy controls. Association can arise for several reasons. First, the investigated allele may be causative of the disease, leading to direct association. Second, the allele may be in LD with the causative allele and show indirect association. The third cause for association is population stratification where the frequency of the investigated allele is different in the affected group compared to the control group not because of the investigated trait but because they originate from different populations. Although association studies represent the most powerful tool to investigate genetic make-up of complex diseases, many inconsistent results have been published.

Candidate gene studies

Until recently, association studies were used to investigate genetic markers located in genes coding for proteins or other gene products with probable or known function for development of the disease. Candidate gene studies require at least some prior knowledge about pathogenesis of the disease and have been quite successful for some traits like lipids but less successful for T2D, obesity and hypertension.

Genome wide association studies

Genome wide association studies (GWAS) investigate SNPs distributed across the whole genome to uncover association in an unbiased way and require therefore no knowledge about the pathophysiology of the disease. It has been shown to be a powerful tool to discover new pathways and mechanisms involved in the pathogenesis of complex diseases and might provide new insights into possible future treatments. Several requirements had to be met before GWAS became plausible; sequencing of the human genome, the knowledge of existing common genetic variation in the human genome, collection of large well powered cohorts and the technological platforms to genotype large set of markers in these large cohorts at reasonable costs as well as development of genetic statistical tools for analysis of the data. In recent years these prerequisites have been met with public SNP databases and chip-based genotyping strategies. In addition, collaborative efforts between several groups that have published GWAS have been crucial for lowering costs for replication studies as GWAS produce a vast number of false positive signals. Several GWAS have now been published and several novel susceptibility genes for complex diseases have already been discovered (Figure 3).

Typically, several hundred thousands to one million markers are genotyped, leading to many false positive results caused by chance. Therefore, a p-value of 5×10^{-8}, equivalent to p-value of 0.05 after a Bonferroni correction for 1 million independent tests, has been proposed and is commonly used to define a genome-wide significant result [27].

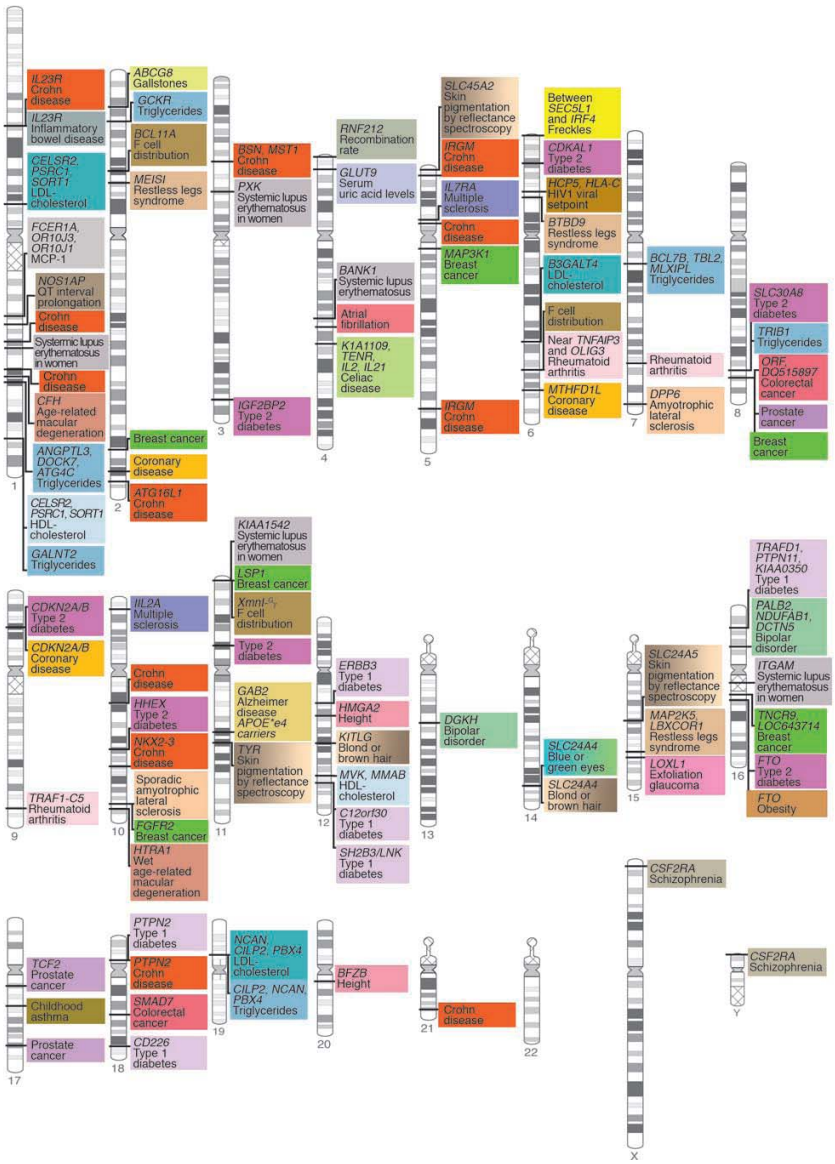


Figure 3. **Novel genes for complex diseases.** Associations between SNPs and traits detected by GWAS with p-values < 9.9×10^{-7} are shown according to chromosomal location. Colour coded boxes indicate similar diseases or traits. JOURNAL OF CLINICAL INVESTIGATION [12]. ONLINE by Teri A. Manolio. Copyright 2008 by American Society for Clinical Investigation. Reproduced with permission of American Society for Clinical Investigation in the format Dissertation via Copyright Clearance Center.

Diabetes mellitus

Diabetes mellitus (DM) describes a group of metabolic disorders with different aetiologies characterized by chronic hyperglycaemia [28]. The definition and diagnosis, as defined by World Health Organization (WHO) is fasting plasma glucose (FPG) ≥ 7.0 mmol/l and/or 2-hour plasma glucose levels ≥ 11.1 mmol/l during an oral glucose tolerance test (OGTT). DM is caused by defects in insulin secretion and insulin action (insulin sensitivity). Several sub-types of DM are recognized; Type 1 Diabetes (T1D), Latent Autoimmune Diabetes in Adults (LADA), Maturity Onset Diabetes of the Young (MODY), Maternally Inherited Diabetes and Deafness (MIDD) and Type 2 Diabetes (T2D). T1D is characterized by early onset, typically in childhood and autoimmune destruction of insulin secreting β-cells. T1D accounts for about 10% of all diabetes cases, although the prevalence varies greatly between different populations [28,29]. Autoimmunity is also present in LADA which demonstrates slowly progressing destruction of β-cells, and is often pathophysiologically considered as a combination of T1D and T2D [30,31]. MODY, which consists of several monogenic forms of DM accounts for ~ 5% of all diabetes. It manifests early in life with an age of onset typically below 25 years. Characteristical for MODY is β-cell dysfunction and a dominant mode of inheritance. It is caused by mutations in single genes and so far mutations in six different genes have been identified, each leading to a specific phenotype and a specific subtype of MODY, MODY 1-6 [32-37]. MIDD is caused by a mutation in mitochondrial DNA and this type of DM is often accompanied by hearing loss or other neurological symptoms [38].

T2D

T2D is the most prevalent type of DM, constituting 85 – 95% of all diabetes. It is usually characterized by late onset, typically above 45 years, and is caused by defects in both insulin secretion and insulin action. T2D patients do not require exogenous insulin for survival and are often treated with diet and glucose lowering oral antidiabetic agents or insulin to obtain metabolic control. The sibling relative risk for T2D is estimated to about 3 [39], indicating a clear genetic component for T2D.

Genetics of T2D

T2D is a polygenic, multifactorial disease and the risk for T2D is thought to be influenced by many genes with minor effects together with environmental factors, such as physical inactivity and energy rich food. Until recently, genetics of T2D had limited success with only a few truly verified genes. Until 2007, the dissection of genetic predisposition to T2D was performed by linkage using microsatellites and candidate gene association studies. With GWAS and large meta-analyses several novel T2D genes have finally been identified and verified, the function of which still remains to be elucidated.

Linkage studies for T2D

Several genome-wide linkage studies searching for region conferring susceptibility to T2D have been published. However, no single region has been widely replicated and only a few have shown significant linkage with LOD score > 3.6 or been replicated in independent studies. Regions showing evidence for linkage in more than one study are located on chromosomes 1q25.3, 2q37.3, 3p24.1, 3q28, 10q26.13, 12q24.31 and 18p11 [40-53]. Although attempts to identify the genetic variation underlying these linkage peaks have been made, the success has been limited. So far, two genes located in the linked regions have been associated with T2D, the calpain 10 gene (*CAPN10*) on chromosome 2q [54] and the transcription factor 7-like 2 gene (*TCF7L2*) on chromosome 10q [55].

One of the most widely replicated regions linked to T2D is located on chromosome 18p. It was first identified in Swedish and Finnish sibling pairs in the 20% most obese individuals and fine-mapping using microsatellites could, however, not further narrow the 15 Mb region [47]. The linkage was not seen until the analysis was restricted to the quintile of patients with highest BMI indicating that different genes may contribute to susceptibility to obese T2D as compared to all cases of T2D. Subsequently, three other scans, two in Dutch and one in an Icelandic population replicated the linkage between obese T2D and chromosome 18p [51-53]. Several related phenotypes show linkage to this region including fasting glucose levels [56-58], BMI [59,60], T2D [43,61], obesity and insulin resistance [62], abdominal obesity [63], dyslipidemia [64], total cholesterol [65] and HDL cholesterol [66]. Further investigations of genetic variation in this region in a Dutch population previously showing linkage to this 15 Mb region was undertaken [51]. Altogether 26 evenly distributed SNPs were genotyped and nominal association for T2D was reported to one SNP (rs3745012) in the lipin 2 (*LPIN2*) gene ($p = 0.03$) [67].

Candidate genes for T2D

Many genes have been proposed as candidate genes for T2D but only few susceptibility genes have been convincingly associated in several studies, including *PPARG, KCNJ11, TCF7L2* and *WFS1*. Many other genes have been associated with T2D in some studies but not in others, including *IRS1, CAPN10, ADBR3, PPARGC1A, ENPP1* and others [68-70]. Many of these genes have been investigated as candidate genes because of their biological function. Common polymorphisms in MODY genes, such as *TCF2, GCK, HNF1α* and *HNF4α* have also been implicated in the development of T2D [71-75].

IRS1

Insulin receptor substrate 1 (*IRS1*) is one of the proteins involved in signal transduction of the activated insulin receptor, when phosphorylated on tyrosine residues [76]. A common polymorphism G972R (rs1801278), located close to the tyrosine phosphorylation motifs, has been associated with T2D in some but not in all studies. A meta-analysis of 27 studies shows modest association with T2D [68], but this association could not been confirmed by recent large study or GWAS for T2D [77]. The

same polymorphism has also been associated with insulin resistance in obese but not in lean individuals [78]. Also the expression of *IRS1* mRNA has been reported to be lower in skeletal muscle from insulin resistant non-diabetic subjects [79]. Mice lacking the *IRS1* display mild insulin resistance and hyperinsulinemia but do not develop diabetes [80]. Tissue specific knock-out experiments indicate a role for IRS1 in insulin signalling in skeletal muscle, adipose tissue and pancreatic β-cells [81].

PPARG

The peroxisome proliferator-activated receptor gamma gene (*PPARG*) encodes a transcription factor highly expressed in adipose tissue and involved in adipocyte differentiation. *PPARG* is also a target of the anti-diabetic, insulin sensitizing drugs thiazolidinediones (TZD) [82]. The minor Ala-allele in in *PPARG*, (Pro12Ala, rs1801282) has been associated with lower BMI, increase in insulin sensitivity and reduced risk of T2D [83,84]. This association has been confirmed by several recent GWAS for T2D [85-89].

KCNJ11

A common, non-synonymous SNP (E23K, rs5219) in the potassium channel, inwardly rectifying, subfamily J, member 11 (*KCNJ11*) has been associated with T2D in several large studies [90-95] and this finding was confirmed by several recent GWAS [85-89]. The gene encodes the Kir6.2 subunit of the inwardly rectifying K_{ATP}-channel. The K_{ATP}-channels in pancreatic β-cells consist of four pore forming Kir6.2 subunits surrounded by four regulatory sulfonylurea receptor 1 (SUR1) subunits and regulate insulin secretion by coupling metabolism to electrical activity [96]. The E23K polymorphism affects ATP sensitivity of the K_{ATP}-channel; homozygous KK-genotype carriers show two fold reduced sensitivity to ATP [97]. The K_{ATP}-channels are targets for sulfonylureas, drugs used to treat diabetes by closing the channel and triggering insulin secretion. Activating mutations in *KCNJ11* are known to cause neonatal diabetes and inactivating mutations neonatal hypoglycemia [98].

TCF7L2

The association between polymorphisms in *TCF7L2* and T2D was first discovered in 2006 by Grant and colleagues when investigating a region linked to T2D on chromosome 10q25 [55]. This association did, however, not explain the linkage to this region. The association has thereafter been confirmed by several studies, including all GWAS and *TCF7L2* is therefore considered the strongest susceptibility gene for T2D [85-89]. *TCF7L2* encodes a transcription factor involved in the Wnt signalling pathway but the mechanism by which it contributes to the pathogenesis of T2D is poorly understood. Several studies have shown that an intronic SNP in *TCF7L2* is associated with impaired insulin secretion and β-cell function but not insulin action [99-104]. It has been suggested that the impaired insulin secretion can be mediated by an impaired incretin effect.

WSF1

WFS1 was identified as a candidate gene for T2D in a study of 83 candidate genes for β-cell function and T2D [105]. Meta analysis of 11 studies confirmed rs10010131 as a susceptibility variant for T2D with genome-wide significant p-value ($p = 5.4\times10^{-11}$) [106]. *WFS1* encodes wolframin, a membrane glycoprotein regulating calcium homeostasis in the endoplasmic reticulum. Mutations in *WSF1* cause the Wolfram syndrome, characterised by diabetes insipidus, DM, optic atrophy and deafness [107,108].

GWAS for T2D

During 2007 six new gene regions involved in T2D were identified by GWAS *(HHEX, SLC30A8, CDKAL1, CDKN2A, IGF2BP2, FTO*) [85-89]. None of these regions contains previously known obvious candidate genes, thereby showing the ability of GWAS to uncover new pathophysiological pathways. Meta-analysis of the three large GWAS for T2D revealed additional six loci (*JAZF1, CDC123/CAMK1D, TSPAN8/LGR5, THADA, ADAMTS9, NOTCH2*) with genome wide significant association [109]. Most of the identified signals have shown association with defects of insulin secretion and β-cell function [88,110-116].

One of the T2D susceptibility genes identified by the GWAS, the Fat mass- and obesity-associated gene (*FTO*), is associated with increase in BMI but also with increased risk of T2D, most likely by its effect on weight gain [117]. Although the mechanisms by which *FTO* predisposes to obesity are mainly unknown, they could involve appetite regulation as *FTO* is strongly expressed in hypothalamus [118,119]

Metabolic syndrome

Metabolic syndrome (MetS) is a cluster of phenotypes, such as obesity, hypertension, hyperglycaemia and dyslipidaemia, increasing the risk of cardiovascular disease (CVD) and T2D. The MetS was described in 1988 by Reaven [120], and many different definitions of the syndrome exist. The most widely used definitions are WHO [121], the National Cholesterol Education Program Adult Treatment Panel III (NCEP-ATP III) [122] and International Diabetes Federation (IDF) [123] (Table 1). Prevalence of MetS varies between ethnic groups and between the different definitions, ranging from 8 – 46% [124]. MetS clusters in families, indicating a genetic component and the heritability has been estimated to ~ 30%. Heritability of components of MetS ranges between ~ 16% for systolic blood pressure to ~ 60% for HDL cholesterol levels [125-129].

The debate is still ongoing whether the components of MetS have a common pathogenic basis; therefore identification of genetic factors contributing to MetS could help to identify this common pathogenesis, if present.

Table 1. Different classifications of the metabolic syndrome

WHO (1999)	NCEP (2004)	IDF (2005)
Diabetes, IGT, IFG or insulin resistance and at least two of:	Three or more of :	Central obesity (waist circumference ≥ 94/80 cm in Europid men/women) and at least two of:
• High WHR (≥ 0.90/0.85 in men/women) • Hypertriglyceridaemia • (≥ 1.7 mmol/l) and/or low HDL-C (≤ 0.9/1.0 mmol/l, in men/women) • Elevated blood pressure (≥ 140/90 mmHg) • Microalbuminuria	• High waist circumference (≥ 102/88 cm in men/women) • Hypertriglyceridaemia • (≥ 1.7 mmol/l) • Low HDL-C (≤ 1.0/1.3 mmol/l in men/women) • Elevated blood pressure (≥ 130/85 mmHg) • Elevated FPG (≥ 6.1 mmol/l)	• Elevated blood pressure (≥ 130/85 mmHg) or treatment • Hypertriglyceridaemia • (≥ 1.7 mmol/l) and/or treatment and/or low HDL-C (≤ 0.9/1.1 mmol/l in men/women) or treatment • Elevated FPG (≥ 5.6 mmol/l)

IGT − impaired glucose tolerance, IFG − impaired fasting glucose, WHR − Waist-to-hip ratio, HDL-C − High density lipoprotein cholesterol, FPG – fasting plasma glucose.

Genetics of MetS

Genetics of MetS has been difficult to dissect because of the complexity of the phenotype, differences in definitions and the relatively low power of studies to detect the subtle effects of genetic variation. The search for genetic variation predisposing to MetS has been conducted by both genome wide linkage studies and investigation of candidate genes. The proposed candidate genes for MetS are involved in energy storage and often support the thrifty gene hypothesis [130]. Individuals living in an environment with unstable food supply, often present during evolution, may have had higher probability of survival if they could store surplus energy when the food was plentiful and use it in times of famine. Genetic predisposition to energy storing would therefore be exposed to positive selection during evolution and beneficial for the individual. When the energy storing genetic variants are exposed to the westernized environment with abundance of high caloric food and physical inactivity (low energy expenditure), they may have become detrimental and cause the phenotype with metabolic disturbances observed in MetS, including obesity and glucose intolerance.

To date, no unifying genetic factors predisposing to MetS have been identified. Several genes have however been associated with at least two features of MetS and are therefore considered as the most promising candidate genes. The adrenergic β-receptors (*ADRB1, ADRB2* and *ADRB3*) have been associated with obesity, hypertension and glucose intolerance and can therefore be considered as good candidates for predisposing to development of MetS [131-136]. Other genes, implicated in T2D and related disorders could also be considered as candidates for MetS, including *APM1*, *CAPN10, PPARGC1A, ENPP1, UCP2, IRS1, PTPN1* and *PPARG* [69,70,129,137-141].

Several genome wide linkage studies for MetS or its components have been performed yielding few chromosomal regions displaying linkage to MetS or its components, but no MetS susceptibility genes in these regions have been identified [60,62,82,142-147].

AIMS

The overall aim of the thesis was to identify genetic risk factors for T2D, particularly T2D associated with obesity and MetS. To achieve this goal, we followed-up a region linked to obese T2D in an earlier genome wide linkage study (I), studied *IRS1* as a candidate gene for T2D and MetS (II, III and V) and studied the role of candidate or susceptibility genes for T2D, identified in candidate gene studies (IV) and GWAS (V), in MetS.

The specific aims were:

Study I To identify genetic variation explaining linkage between obese T2D (diabesity) and a region on chromosome 18p. To achieve this, we performed tag SNP mapping of the region and took advantage of publicly available results from GWAS for related phenotypes to choose SNPs for replication in a large Swedish case-control material.

Study II To investigate whether previously established [68] association between the G972R polymorphism in the *IRS1* gene could be replicated in a well powered study of 9 000 individuals.

Study III To investigate whether other common variants apart from the G972R polymorphism in the *IRS1* gene would increase risk of T2D by investigating the haplotype structure of *IRS1* and genotyping tag SNPs capturing common variation in the gene in several large cohorts of northern European ancestry.

Study IV To investigate whether common polymorphisms in seven genes previously associated with T2D or features of the MetS (*PPARG, APM1, ADRB1, ADRB2, ADRB3* and *PPARGC1A*) would predispose to MetS in non-diabetic individuals from the Botnia prospective study.

Study V To investigate whether 27 polymorphisms in genes previously associated with T2D or features of the MetS would predict development of MetS or components thereof in a large prospective study including 16 000 Swedish individuals followed-up for 24 years.

SUBJECTS

Study I

Botnia 2 families

The Botnia study is a large family based study initiated in 1990 in the Botnia region in the Western part of Finland to investigate genetic determinants and metabolic consequences of T2D [148]. Later on, the Botnia study was extended to include families from other parts of Finland and Sweden. Today the Botnia study comprises over 11 000 individuals from more than 1 400 families. Two genome wide linkage studies using microsatellites have been performed within the Botnia study a) Botnia 1 in 58 families from the original Botnia region [44,149] and b) Botnia 2 in 480 sibling pairs from Finalnd and Sweden. The families included in the Botnia 2 linkage study had at least two family members affected by T2D (mainly two affected siblings) [47,150]. Altogether 3 064 individuals from 555 families agreed to participate in the Botnia 2 scan [47].

In study I, 752 individuals from 164 families with a family mean BMI > 30 kg/m^2 were included (males/females 322/420, age 58 ± 13 years, BMI 28.8 ± 4.8 kg/m^2, 60% with T2D and 44% of T2D patients with BMI > 30 kg/m^2).

Swedish case-control replication sample

A Swedish case-control cohort consisting of 2 830 T2D patients and 3 550 normoglycemic control subjects was used to test for association with T2D. T2D patients were chosen from the Malmö diabetes registry (Diabetes 2000) comprising of more than 7 000 patients with T1D and T2D [151]. Patients of Scandinavian origin with an age of onset of diabetes > 35 years, C-peptide ≥ 0.3 nmol/l and without GAD antibodies were selected as T2D cases. Controls were selected from the Malmö Diet and Cancer study (MDC), which includes 30 000 individuals [152], including ~ 6 000 for whom a more extensive metabolic data concerning cardiovascular risk factors have been collected (MDC-cardiovascular cohort; MDC-CC). The control individuals were of Scandinavian origin, had no known 1st degree family history of diabetes and had fasting blood glucose < 5.5 mmol/l (Table 2).

As our primary interest was in T2D in combination with obesity (diabesity), a subgroup of obese T2D patients with BMI > 30 kg/m^2 (n = 1 098) and lean nondiabetic controls with BMI < 25 kg/m^2 (n = 1 927) were used in study I to identify genetic variants associated with diabesity (Table 2).

Table 2. Clinical characteristics of the Swedish T2D case control cohort, as well as of the subgroup of obese T2D patients and lean controls

	T2D		Controls	
	All	Obese	All	Lean
N (m/f)	2 830 (1 667/1 163)	1 098 (595/503)	3 550 (1 340/2 210)	1 927 (633/1 294)
Age (years)	63 ± 11	60 ± 11	57 ± 6	57 ± 6
BMI (kg/m^2)	29.6 ± 5.5	34.7 ± 4.2	25.1 ± 3.6	22.5 ± 1.7
FPG (mmol/l)	11.9 ± 4.3	12.0 ± 4.4	5.4 ± 0.41	5.3 ± 0.41
HDL cholesterol (mmol/l)	1.14 ± 0.33	1.09 ± 0.29	1.42 ± 0.37	1.51 ± 0.38

Obesity was defined as BMI > 30 kg/m^2 and leanness as had BMI < 25 kg/m^2.

Study II and III

In studies II and III, altogether nine different cohorts were included to analyze association between genetic variation in *IRS1* and T2D. In study II, cases were defined as individuals with T2D or severe impaired glucose tolerance (IGT). In study III, only individuals with manifest T2D from the same cohorts were included as cases. A detailed description of the different cohorts is given in Table 3.

The Scandinavian subjects underwent an OGTT and insulinogenic index (insulin at 30 min – insulin at 0 min)/glucose at 30 min [153], insulin sensitivity index (ISI) (10 000/√[fasting glucose x fasting insulin] x [mean glucose x mean insulin during OGTT]) [154] and insulin area under the curve (AUC) (insulin at 30 min x 30) + (insulin at 60 min x 45) + (insulin at 120 min x 30) – (insulin at 0 min x 105) were calculated. Disposition index, i.e. a measure of β-cell function corrected for insulin resistance, was calculated both as (insulinogenic index/HOMA-IR) and (insulinogenic index x ISI/100). Percent homeostasis model assessment of β-cell function (HOMA-β) was estimated as (20 x fasting serum insulin/[fasting plasma glucose – 3.5]) and estimate of insulin resistance was derived by HOMA of insulin resistance (HOMA-IR), calculated as (fasting serum insulin x fasting plasma glucose/22.5) [155].

Table 3. Clinical characteristics of samples cohorts analyzed in studies II and III

Sample set	Study II N (m/f)	Study III N (m/f)	Age[1] (years)	BMI[1] (kg/m^2)	FPG[1] (mmol/l)
Sibships					
Cases[2]	280/329	212/247	65 ± 10	29 ± 5	9.3 ± 3.3
Controls	275/305	167/235	62 ± 10	26 ± 3	5.4 ± 0.4
Scandinavia CC					
Cases[2]	252/219	220/182	60 ± 10	28 ± 5	9.8 ± 3.4
Controls	254/217	187/166	60 ± 10	27 ± 4	6.2 ± 1.8
Sweden CC					
Cases[2]	267/247	252/228	66 ± 12	28 ± 4	8.5 ± 2.5
Controls	267/247	244/224	66 ± 12	28 ± 4	4.8 ± 0.6
Botnia CC					
Cases[2]	425/507	NI	68 ± 12	27 ± 8	9.1 ± 3.2
Controls	60/125	NI	48 ± 11	24 ± 5	5.3 ± 0.4
Canada CC					
Cases	70/57	70/57	53 ± 8	29 ± 5	6.4 ± 1.8
Controls	70/57	70/57	52 ± 8	29 ± 4	5.1 ± 0.6
USA CC					
Cases	644/582	644/582	63 ± 11	33 ± 7	9.8 ± 3.0
Controls	644/582	644/582	61 ± 10	27 ± 5	5.1 ± 0.9
Poland CC					
Cases	422/587	422/587	62 ± 10	30 ± 5	8.9 ± 4.0
Controls	422/587	422/587	59 ± 7	26 ± 4	4.8 ± 1.2
Botnia trios					
Probands	NI	99/112	40 ± 9	28 ± 5	7.5 ± 3.0
Sweden replication CC					
Cases	NI	1 008/991	61 ± 11	30 ± 6	12.5 ± 4.1
Controls	NI	1 009/1251	59 ± 5	25 ± 4	4.8 ± 0.4

Data are presented as mean ± SD; CC – case control material; NI – not included [1]The clinical characteristics are given for the larger sample included in study II, except for Botnia trios and Sweden replication CC; [2] Cases include subjects with T2D and severe IGT in study II but only subjects with T2D in study III.

Study IV

Botnia prospective study

Within the Botnia study, non-diabetic subjects, either family members of T2D patients or spouses without known first or second degree family history of T2D aged 18 -70 years were invited to prospective visits with OGTT every 2 -3 years [148]. In total 2293 individuals participated in at least two OGTTs with a median follow-up time of 6 years (range 2-12 years) [156,157]. Carriers of known mutations causing MODY (n = 20) were excluded from the study.

MetS was defined according to NCEP-ATP III criteria [122] and 1 937 individuals (M/F 873/1064) were free of MetS at baseline (Table 4).

Table 4. Clinical characteristics of subjects without MetS according to NCEP-ATPIII definition at baseline from the Botnia prospective study.

	Characteristics at baseline	Data Available (N)
N (m/f)	1937 (873/1064)	1937
Age (years)	43.7 ± 13.8	1937
BMI (kg/m^2)	24.9 ± 3.6	1926
Waist (cm)	84.9 ± 10.8	1881
Systolic blood pressure (mmHg)	126 ± 17	1888
Diastolic blood pressure (mmHg)	77 ± 10	1887
Fasting glucose (mmol/l)	5.5 ± 0.5	1937
Triglycerides (mmol/l)	1.1 ± 0.6	1743
HDL–cholesterol (mmol/l)	1.4 ± 0.3	1740
Smoking (%)	37.7	1877
Obesity (%)	11.9	1881
Low HDL (%)	17.6	1739
Hypertriglyceridemia (%)	8.7	1742
Hypertension (%)	39.7	1888
Hyperglycemia (%)	12.9	1937

Study V

Malmö Preventive Project

The Malmö Preventive Project (MPP) is a large population based prospective study from Malmö, Sweden, initiated in 1974 as a population based health screening [158,159]. In MPP, 33 346 subjects (22 444 men and 10 902 women) with mean age of 49 years underwent a physical examination and measurements of fasting blood glucose and triglyceride concentrations were performed. In addition, 18 900 consecutive individuals underwent OGTT. Information on lifestyle factors and medical history was obtained from a questionnaire. Of individuals participating in the initial screening, 4 931 had died and 551 were lost from the follow-up for other reasons. 25 000 of the eligible individuals were invited to a re-screening visit during 2002-2006, including a physical examination and fasting blood samples for measurements of glucose, triglycerides and HDL-cholesterol concentrations. Of the invited individuals, 17 284 participated in the re-screening and of them 1 141 were excluded from the present study because of lack of DNA or crucial clinical information, or due to T2D at baseline. Diagnosis of diabetes was confirmed from patient records or was based upon a fasting plasma glucose concentration greater than 7.0 mmol/l. Thereby, 16 143 non-diabetic subjects, 2 063 of whom developed T2D during the mean follow-up time of 24 years, were included in the study (Table 5).

Table 5. Clinical characteristics of subjects in MPP at baseline and follow-up

	Characteristics at baseline	Data Available (N)	Characteristics at follow-up	Data Available (N)
Males/Females	10 455/5 688	16 143	10 455/5 688	16 143
Age (years)	45.5 ± 6.9	16 143	68.4 ± 6.0	16 143
BMI (kg/m^2)	24.3 ± 3.3	16 136	27.1 ± 4.1	16 055
Waist (cm)	78.6 ± 9.8	827	94.8 ± 12.2	16 051
SBP (mmHg)	127.2 ± 14.3	16 140	144.9 ± 20.0	16 099
DBP (mmHg)	83.9 ± 8.9	16 140	83.6 ± 10.54	16 099
FPG (mmol/l)	5.4 ± 0.56	16 083	5.8 ± 1.30	16 111
Triglycerides (mmol/l)	1.26 ± 0.77	16 104	1.26 ± 0.80	16 118
HDL (mmol/l)	1.54 ± 0.37	289	1.40 ± 0.42	16 112
Anti-hypertensive medication (%)	4.2	16 119	37.7	16 141
Lipid lowering medication (%)	NA	0	20.0	16 141
T2D (%)	0	16 143	13.2	16 122
Obesity[1] (%)	5.1	16 135	21.2	16 042
Hypertriglyceridaemia (%)	18.2	16 104	19.7	16 118
Hypertension[2] (%)	34.2	16 124	84.5	16 099
Hyperglycaemia (%)	39.9	16 143	47.2	16 117

SBP = systolic blood pressure; DBP = diastolic blood pressure; FPG = fasting plasma glucose; NA = not available; [1] Obesity is defined as BMI ≥ 30 kg/m^2; [2] hypertension is defined as untreated blood pressure of 140/90 mmHg or treatment for hypertension.

In study V, a person with MetS was defined as an individual having at least three of the following features: (1) obesity (BMI ≥ 30 kg/m^2); (2) dyslipidaemia (triacylglycerides ≥ 1.7 mmol/l and/or lipid-lowering treatment); (3) hypertension (blood pressure ≥ 140/90 mmHg and/or antihypertensive treatment); and (4) hyperglycaemia (fasting plasma glucose ≥ 5.6 mmol/l and/or overt diabetes). The lack of data for waist and HDL cholesterol at baseline prevented the definition of MetS according to the established definitions (NCEP-ATP III, WHO, IDF) [121-123].

METHODS

Genotyping

The TaqMan™ allelic discrimination assays on the ABI 7900HT sequence detection system (Applied Biosystems, Foster City, CA, USA) were performed in all five studies.

The iPLEX™ genotyping by matrix-assisted laser desorption ionisation-time of flight mass spectrometry (MALDI-TOF MS) on the MassARRAY platform (Sequenom, San Diego, CA, USA) was used in studies I, II, III and V.

The SNaPshot assay (Applied Biosystems, Foster City, CA, USA) genotyping using ABI3100 (Applied Biosystems, Foster City, CA, USA) was performed in study IV.

The KASPar® genotyping assay was used to genotype rs13266634 in study V.

Statistical analysis

The association with dichotomized phenotypes in family based studies was tested using the transmission disequilibrium test (TDT) [160] (studies I and III) or discordant allele test in the sib-pairs [161] (studies II and III).

Association between the SNPs and BMI was tested by using the method for family based association analysis implemented in the computer program Merlin. All analyses in study I were adjusted for age and sex whereas all T2D association analyses were adjusted for age, gender and BMI.

Association to dichotomous phenotypes in population based studies was analysed by logistic regression (studies I and V) or χ^2 test (studies II and III). Risk was described by odds ratio (OR) which gives an estimate of the increased risk per risk allele carried. In study I, all analyses were adjusted for age and gender (T2D and obese T2D). In study V, the analyses of risk of MetS or components thereof were adjusted for gender, age at inclusion to the study and follow-up time.

Results from the different cohorts in studies II and III were combined by Mantel-Haenszel meta-analysis of the ORs [162]. Homogeneity among study samples was tested by a Pearson χ^2 goodness of fit test in study II or by asymptotic Breslow-Day statistics in study III [162,163].
Power calculations were performed with the genetic power calculator at http://pngu.mgh.harvard.edu/~purcell/gpc/cc2.html [164] (studies II, III and V)

Survival analyses were used to estimate the effect of genetic variants (risk and non-risk genotypes) on the risk of developing MetS by defining age at onset (the proportion of individuals developing MetS at certain age) in the Botnia prospective study. The

risk of developing MetS was expressed as a hazard function using an age-adjusted Cox proportional hazard regression model [165]. The relative effect was presented as the ratio between the hazard functions (hazard ratio [HR]) of the two groups (carriers versus non-carriers). All survival analyses were stratified for gender and performed with a robust variance estimate to adjust for within family dependence extended to large pedigrees. In using a robust variance estimate we treated each pedigree (instead of each individual) as an independent entity for calculating the variance of the estimates (study IV).

A generalised estimating equation (GEE) procedure was used to test differences in individual MetS components over time between risk and non-risk genotype carriers, treating the components as continuous variables. The analyses were adjusted for age, sex and smoking. GEE corrects for the potential problem that one person with more visits contributed more to the change over time than a person with fewer visits. In the GEE the standard errors were adjusted for repeated measurements in the same person with a robust variance estimator [165,166] (study IV).

Multivariate logistic regression analyses were performed adjusting for age at baseline, follow-up time and sex and including 17 polymorphisms identified by GWAS in the model. In addition, we conducted backward elimination of SNPs with a retention threshold of $p<0.05$ (study V).

RESULTS

Study I: Genetic dissection of chromosome 18p11, a region linked to obese T2D and HDL-cholesterol

Several strategies were used to investigate genetic variation with potential contribution to the linkage on chromosome 18p. First, tag SNPs were identified by Tagger program implemented in Haploview (MAF > 0.05, $r^2 \geq 0.8$) in 11 positional and functional candidate genes located in the 15 Mb region identified by earlier studies performed in our laboratory [47] (*LPIN2*, *EPB41L3, LAMA1, NDUFV2, ANKRD12, VAPA, NAPG, CIDEA, PTPN2* and *MC2R/MC5R*). Due to the more recent results of a collaborative study between our and Leena Peltonens laboratories showing linkage to HDL cholesterol in a meta analysis of four genome wide linkage studies [66], tag SNPs were selected in all known and hypothetical genes in the 1 Mb region centred around 5.3 Mb and thus fully overlapping with the T2D linked region (*LOC388458, LOC284215, PPIAP14, LOC642597, LOC339290, ZFP161, TTMA-LOC388459, LOC645355, TTMA, LOC729309, L3MBTL4* and *LOC645387*). In addition, we also identified 58 tag SNPs in four genes (*PTPRM, APCDD1, FAM38B* and *C18orf58)* located around 10 Mbp based on association results for T2D and obese T2D from DGI GWAS (Figure 4A). Altogether 501 tag SNPs were genotyped in Botnia 2 families in this step.

Second, we utilized the publicly available GWAS data from DGI and other published GWAS [89,167]. In the 12 Mb region on chromosome 18p between 2 – 14 Mb linked to obese T2D, 2017 SNPs were successfully genotyped in the DGI GWAS. We plotted all nominally significant p-values for association with T2D, obese T2D (T2D and BMI $\geq$ 30 kg/m^2 compared with all controls) and BMI in T2D patients (Figure 4A). Three SNPs with the most significant results from DGI GWAS on chromosome 18 were selected for replication in the Botnia 2 famillies and the Swedish case control material (Figure 4A). First, the most significant SNPs for T2D and obese T2D (rs7237226 in *C18orf58*, $p = 1.1 \times 10^{-5}$) and second, two most significant SNPs for BMI in T2D patients (rs644399 and rs4121619 in *PTPRM*, $p = 2.0 \times 10^{-5}$ and $p = 0.00028$, respectively) were selected for replication. We also evaluated SNPs associated with fatmass in DGI and the most significant result (rs20659 in *APCDD1*, $p = 5.6 \times 10^{-5}$) was further investigated. In addition, SNPs showing the most significant association on chromosome 18p in the WTCCC (rs543759 in *PTPRM*, $p = 2.6 \times 10^{-5}$) and in Mexican Americans (rs1941011 in *EPB41L3*, $p = 5.0 \times 10^{-5}$) were selected for replication in both the Botnia 2 families and the Swedish case-control material [67,89,167]. Also, a SNP in the *LPIN2* gene showing association with T2D in a Dutch population (rs3745012, p = 0.03) was pursued further in the same cohorts.

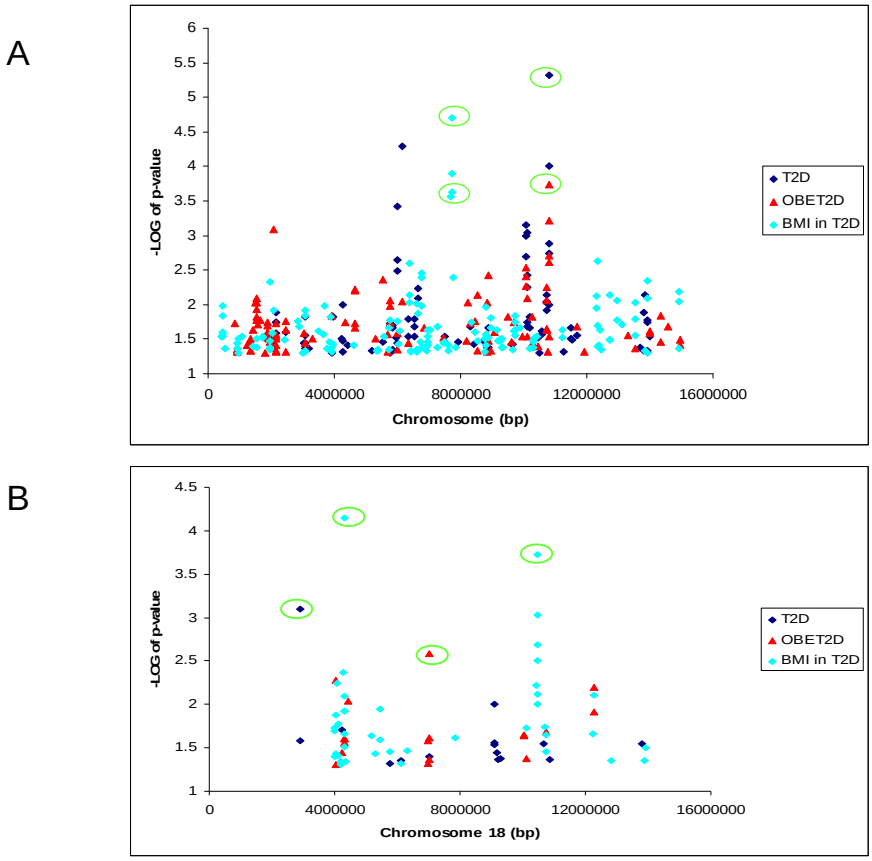


Figure 4. Association signals on chromosome 18p. A) Significant association results from DGI GWAS for T2D, obese T2D and BMI in T2D patients. The –log of p-values for all three traits are plotted in the graph. SNPs marked with green circle were chosen for replication in the Swedish T2D case-control material. Several association signals in intron one of the *PTPRM* gene were associated with BMI in T2D patients in the DGI GWAS; therefore rs4121619 (p = 0.00028) was chosen as the best proxy as it was highly correlated with the other significantly associated SNPs in this region. B) Significant association results from the Botnia 2 families for T2D, obese T2D and BMI in T2D patients. The –log of p-values for all three traits are plotted in the graph. SNPs selected for replication in the Swedish case-control material are marked with green circles.

In total 508 SNPs were successfully genotyped in 752 individuals from the Botnia 2 families and analysed for three phenotypes; obese T2D, T2D and BMI in T2D patients and the results are summarized in Figure 4B. As an exploratory analysis, BMI in all individuals from Botnia 2 families was also investigated. The strongest association with obese T2D was observed for rs1537422 in the *LAMA1* gene ($p = 0.0026$) and the same SNP was also associated with T2D ($p = 0.040$). For T2D the strongest associated SNP was rs1985 in the *LPIN2* gene ($p = 0.0008$). The rs9303948 in *LOC388458* showed the strongest association to BMI in T2D patients ($p = 7.2 \times 10^{-5}$) and the same SNPs was also associated with BMI in all individuals ($p = 0.0032$). The second strongest association with BMI in T2D patients was observed for rs564991 in the *APCDD1* gene ($p = 0.00064$); the same SNP was also associated to BMI in T2D patients ($p = 0.00019$). All of these most significant results in the Botnia 2 families (rs1537422, rs1985, rs9303948 and rs564991) were selected for replication in the Swedish case-control material (Figure 4B).

Altogether twelve SNPs were selected for replication (Table 6). Based on the results from three GWAS, including DGI, WTCCC and Mexican American population [85,89,167], six SNPs were selected. In addition, rs3745012 in *LPIN2*, previously nominally associated with T2D the Dutch population ($p = 0.03$) [67], was included in the replication study. Finally, five SNPs most significantly associated with T2D, obese T2D and BMI in T2D patients in Botnia 2 families were included in the replication. All twelve SNPs were analyzed for association with obese T2D, T2D and BMI in T2D patients. However, no significant association was observed between the selected SNPs and any of the investigated phenotypes in the case control material (Table 6).

Our results from extensive fine mapping of the chromosome 18p region by tag SNPs and replication of association signals from this fine mapping, published GWAS and rs3745012 in *LPIN2* show no association with T2D, obese T2D or BMI in T2D patients in a large Swedish case control material. These findings argue against common variation contributing to observed linkage on chromosome 18p.

Table 6. Association results in Swedish T2D case-control material.

SNP	Gene	T2D		OBET2D		BMI in T2D	Previous association		
		OR [95% CI]	p-value	OR [95% CI]	p-value	p-value	Study	Trait	p-value
rs7237226	C18orf58	1.01 [0.94-1.09]	0.76	1.06 [0.89-1.27]	0.51	0.62	DGI	OBET2D + T2D	0.00019/1.1 x 10^{-5}
rs644399	PTPRM	1.00 [0.93-1.08]	0.93	1.04 [0.93-1.16]	0.50	0.74	DGI	BMI in T2D patients	1.97 x 10^{-5}
rs4121619	PTPRM	0.97 [0.90-1.05]	0.46	1.01 [0.90-1.13]	0.85	0.89	DGI	BMI in T2D patients	0.00028
rs206539	APCDD1	1.06 [0.96-1.16]	0.28	1.10 [0.96-1.27]	0.18	0.43	DGI	Fatmass	5.6 x 10^{-5}
rs543759	PTPRM	0.94 [0.86-1.03]	0.17	0.94 [0.83-1.08]	0.38	0.28	WTCCC	T2D	2.6 x 10^{-5}
rs1941011	EPB41L3	0.98 [0.90-1.07]	0.68	0.96 [0.85-1.09]	0.55	0.20	Mexican Americans		0.000050
rs3745012	LPIN2	1.05 [0.96-1.15]	0.24	0.97 [0.85-1.10]	0.63	0.94	Aulchenko et al		0.03
rs1537422	LAMA1	1.02 [0.94-1.10]	0.66	0.95 [0.85-1.07]	0.42	0.57	Botnia 2 families	OBET2D	0.0026
rs11081094	LOC388458	1.03 [0.95-1.12]	0.45	1.04 [0.92-1.18]	0.50	0.68	Botnia 2 families	BMI	0.0077
rs9303948	LOC388458	1.01 [0.94-1.10]	0.73	1.11 [0.99-1.25]	0.071	0.30	Botnia 2 families	BMI in T2D	7.2 x 10^{-5}
rs564991	APCDD1	0.96 [0.88-1.03]	0.26	0.92 [0.82-1.03]	0.15	0.91	Botnia 2 families	BMI	0.00064
rs1985	LPIN2	1.04 [0.96-1.13]	0.40	1.09 [0.96-1.23]	0.18	0.45	Botnia 2 families	T2D	0.0008

Study II: Association testing in 9000 people fails to confirm the association of the insulin receptor substrate-1 G972R polymorphism with type 2 diabetes

The aim of study II was to investigate whether the results from a meta-analysis by Jellema *et al*. [68] indicating that G972R (rs1801278) in *IRS1* is associated with T2D could be replicated in an adequately powered collection of several family based and case-control cohorts of northern European ancestry.

Our study had 98.9% power to reject the null hypothesis of no association between G972R in *IRS1* and T2D at $p < 0.05$ under a dominant model of inheritance, assuming MAF of 10% and OR of 1.25 (as reported by Jellema *et al*.) and prevalence of T2D of 10%.

No association between G972R in *IRS1* and T2D was observed in any of the cohorts nor in the combined meta-analysis. In the Scandinavian/Canadian samples (2 044 T2D cases, 1 297 controls and 1 189 siblings discordant for T2D) and the Polish sample (1 009 case control pairs) no association was observed (OR = 1.01, p = 0.91 and 1.13, p = 0.39, respectively). In the North American Caucasians (1 226 case control pairs) the association trended in the opposite direction (OR = 0.81, p = 0.07). When combining our data together with diabetic trios reported by Altshuler *et al*. [83] and all the studies in Jellema *et al*. [68] the association between G972R in *IRS1* was not significant (OR = 1.07 [95% CI 0.97 – 1.18], p = 0.17).

Neither did we observe association to age of onset, measures of insulin secretion (insulinogenic index, percent HOMA-β and disposition index) nor to measures of insulin resistance (HOMA-IR and ISI) in the Scandinavian samples.

Study III: Association testing of common variants in the insulin receptor substrate-1 gene (IRS1) with type 2 diabetes

To follow-up study II, where no association between G972R and T2D was observed, study III was designed to investigate the haplotype structure of *IRS1* gene region and to assess whether common variation in *IRS1*, other than G972R, could contribute to increased risk of T2D in several family based and case control cohorts of northern European ancestry (same as in study II).

The haplotype structure of *IRS1* was obtained from the CEU samples using phase 1 data of the HapMap project [168]. Using this information we identified a region spanning 183 kb consisting of 130 polymorphic SNPs and ranging from 105 kb upstream and 13 kb downstream of the *IRS1* gene. A decay in LD was observed at 105 kb upstream and 13 kb downstream and defined the ends of the haplotype blocks. Using the Tagger

program, 20 tag SNP were selected, capturing 85% of all common variation (MAF ≥ 0.05) with $r^2 \geq 0.8$ in that region.

All 20 tag SNPs were genotyped in samples originating from USA and Poland. After combined meta-analysis of these two samples only one SNP, rs934167 showed nominally significant association to T2D (OR 1.25 [1.03 – 1.51], p = 0.03). This SNP was then genotyped in the Scandinavian and Canadian sample sets. The combined analysis of all samples resulted in an OR of 1.20 [1.05 – 1.37], p = 0.008. After adjustment for multiple testing by 10 000 permutations, these results did not remain significant (p = 0.086). No heterogeneity was observed among the sub-samples (p = 0.29).

When investigating quantitative phenotypes that may influence the risk of T2D, no significant differences in fasting and 2-h insulin levels, insulinogenic index, insulin AUC, HOMA-IR or ISI were observed between the different genotype carriers of rs934167.

From these results we conclude that common variation in the *IRS1* gene region does not have a major impact on the development of T2D. It is, however, possible that yet uncaptured variation in *IRS1* might increase risk of T2D.

Study IV: Genetic prediction of the metabolic syndrome

The aim of study IV was to investigate whether common variants in genes previously associated with T2D or features of MetS (*PPARG, ADRB1, ADRB2, ADRB3* and *APM1*) could predict development of MetS in non-diabetic individuals from the Botnia prospective study. MetS was defined according to the NCEP-ATP III definition [122]. In addition, the G487S polymorphism in the *PPARGC1A* gene was included as this gene is considered a master regulator in energy, fat and glucose metabolism [169,170], thereby making it a good candidate for MetS.

Out of 1937 individuals free of MetS at baseline 267 (13.8%) developed MetS during a mean follow-up period of 6 years. At baseline, 729 (37.6%) individuals did not have any of the components of MetS while 735 (38.0%) individuals had one and 473 (24.4%) had two components of MetS.

The risk to develop MetS was increased by polymorphisms in *PPARG* (P12A (rs1801282), HR = 1.49 [1.10 – 3.01], p = 0.011) and *ADRB1* (G398R (rs1801253), HR = 1.43 [1.14 – 1.80], p = 0.0022) in the univarite analysis as well as the multivariate analysis (p = 0.040 and 0.0020 respectively).

Additionally, we investigated if changes in components of the MetS over time of were associated with the studied genetic variation, using a GEE model. Carriers of the PP genotype in *PPARG* P12A showed increase in triglyceride concentration over time

(coefficient = 0.016 [0.0002 – 0.0317], p = 0.047). The GG/GR genotype carriers of the *ADRB1* G398R polymorphism showed increase in both triglyceride (0.0162 [0.0017 – 0.0307], p = 0.028) and fasting plasma glucose (0.013 [0.0009 – 0.0255], p = 0.036) concentration over time.

There are two other commonly used definitions for MetS in addition to NCEP-ATP III, the WHO [121] and IDF [123] definitions. We investigated whether the same polymorphism would predict development of MetS if defined by WHO and IDF criteria. Out of 1759 individuals without MetS according to WHO criteria ($MetS_{WHO}$) at baseline 361 (20.5%) developed MetS during the follow-up. Polymorphisms in *ADRB1* (G398R (rs18012539), HR = 1.32 [1.08 – 1.62], p = 0.008) and *APM1* (I2019D (rs59618226) HR = 1.36 [1.02 – 1.82], p = 0.035) genes increased the risk for $MetS_{WHO}$. Totally, 1609 individuals were free from MetS according to IDF criteria ($MetS_{IDF}$) at baseline. Out of these 380 (23.6%) developed $MetS_{IDF}$ during the 6-year follow-up and none of the studied polymorphisms significantly increased the risk of $MetS_{IDF}$. However, if the IDF criteria were modified and the FPG cut-off was increased to 6.1 mmol/l (instead of 5.6 mmol/l), the polymorphism in *ADRB1* (G398R (rs18012539), HR = 1.76 [1.20 – 2.58], p = 0.004) increased the risk to develop $MetS_{IDF}$.

In conclusion, polymorphisms in *PPARG* and *ADRB1* predicted development of MetS according to NCEP-ATP III criteria. The G398R variant in *ADRB1* also predicted development of MetS according to all three criteria, indicating a role of altered free fatty acid metabolism in the pathogenesis of MetS.

Study V: The search for putative unifying genetic factors for components of the metabolic syndrome

The aim of this study was to investigate whether polymorphisms in genes previously associated with T2D (*TCF7L2, PPARG, KCNJ11, WFS1, SLC30A8, HHEX, IGF2BP2, CDKAL1, CDKN2A, FTO, JAZF1, CDC123/CAMK1D, TSPAN8/LGR5, THADA, ADAMTS9* and *NOTCH2*) or at least two features of the MetS (*ADRB1, ADRB2, ADRB3, CAPN10, IRS1, UCP2, PPARGC1A, PTPN1, ENPP1* and *GCKR*) would predict MetS or it's components in a large prospective study of 16 000 individuals from MPP.

Polymorphisms in *TCF7L2* (rs7903146, OR 1.10 [1.04-1.17], p=0.00097), *FTO* (rs9939609, OR 1.08 [1.02-1.14], p=0.0065), *WFS1* (rs10010131, OR 1.07 [1.02-1.13], p=0 .0078) and *IGF2BP2* (rs4402960, OR 1.07 [1.01-1.13], p=0.021) predicted MetS in MPP using univariate analysis adjusted for age at baseline, sex and follow-up time (Table 7). After correction for multiple testing (Bonferroni correction for analysis of 27 SNPs included in the study) only the association with *TCF7L2* SNP remained significant (p_c=0.026). Multivariate analysis was conducted and included the 16 variants previously associated with T2D and *GCKR*. In multivariate analysis rs7903146 in *TCF7L2* (p=0.0012), rs111875 in *HHEX* (p=0.0052) and rs9939609 in

FTO (p=0.018) were retained as significant predictors of MetS in backward elimination model.

Table 7. The risk of developing MetS in MPP

Gene	SNP	OR [95% CI]	p-value
TCF7L2	**rs7903146**	**1.11 [1.04-1.17]**	**0.00097**
FTO	**rs9939609**	**1.08 [1.02-1.14]**	**0.0065**
WSF1	**rs10010131**	**0.93 [0.89-0.98]**	**0.010**
IGF2BP2	**rs4402960**	**1.07 [1.01-1.13]**	**0.023**
PPARg	**rs1801282**	**0.97 [0.90-1.05]**	**0.50**
CDKN2A	**rs10811661**	**0.93 [0.87-1.00]**	**0.060**
HHEX	**rs111875**	**0.95 [0.90-1.00]**	**0.068**
SLC30A8	**rs13266634**	**0.97 [0.91-1.02]**	**0.22**
GCKR	**rs1260326**	**1.03 [0.97-1.09]**	**0.33**
THADA	**rs7578597**	**1.05 [0.96-1.14]**	**0.33**
CDKAL1	**rs7754840**	**0.97 [0.91-1.03]**	**0.34**
NOTCH2	**rs10923931**	**0.96 [0.88-1.06]**	**0.45**
TSPAN8_LGR5	**rs7961581**	**0.98 [0.93-1.04]**	**0.57**
KCNJ11	**rs5219**	**1.01 [0.96-1.07]**	**0.66**
ADAMTS9	**rs4607103**	**1.00 [0.94-1.06]**	**0.87**
JAZF1	**rs864745**	**1.00 [0.95-1.06]**	**0.88**
CAMK1D	**rs12779790**	**1.00 [0.93-1.07]**	**0.89**
CAPN 10	rs3792267	1.03 [0.97-1.09]	0.39
CAPN 10	rs2975760	0.99 [0.92-1.06]	0.77
UCP2	rs659366	1.02 [0.97-1.08]	0.41
IRS1	rs1801278	0.98 [0.87-1.10]	0.68
PPARGC1A	rs8192678	1.00 [0.95-1.06]	0.92
ADRB1	rs1801253	1.04 [0.98-1.10]	0.24
ADRB2	rs1042714	1.01 [0.95-1.06]	0.83
ADRB3	rs4994	0.97 [0.87-1.07]	0.51
PTPN1	rs3787348	1.05 [0.99-1.11]	0.087
ENPP1	rs1044498	1.00 [0.93-1.08]	0.99

All analyses are adjusted for age at baseline, follow-up time and gender. The p-values given are 2-sided, uncorrected for multiple testing. Polymorphisms highlighted in bold were included in the composite genetic score.

When studying the specific components of MetS, polymorphisms in *TCF7L2* (rs7903146, OR 1.17 [1.09-1.25], p<0.00001), *IGF2BP2* (rs4402960, OR 1.10 [1.03-1.18], p=0.0033), *SLC30A8* (rs13266634, OR 1.12 [1.05-1.19], p=0 .00061), *CDKAL1* (rs7754840, OR 1.12 [1.04-1.20], p=0 .0017), *WFS1* (rs10010131, OR 1.07 [1.01-1.14], p=0 .027) and *ENPP1* (rs1044498, OR 1.10 [1.01–1.19], p=0.037) nominally predicted hyperglycaemia in univariate analysis. The same genes predicted hyperglycaemia also in multivariate analysis (rs7903146, p<0.00001; rs4402960, p=0.025; rs7754840, p=0.0019; and rs13266634, p=0.0037).

Development of obesity was predicted by polymorphisms in *FTO* (rs9939609, OR 1.09 [1.03-1.16], p=0.004) and *ADRB3* (rs4994, OR 1.13 [1.01–1.26], p=0.034); none of the SNPs remained significant after correction after multiple testing and none of the polymorphisms were retained as significant predictors of obesity in multivariate analysis.

The risk to develop dyslipidaemia was increased in carriers of T allele in *GCKR* both in univariate (OR 1.15 [1.09-1.22], p<0.00001, p_c=0.000053) and multivariate analysis (p<0.00001 for *GCKR*).

A composite genetic risk score was constructed calculating the number of risk alleles carried by each individual in the 17 SNPs identified and/or verified in GWAS (*TCF7L2, PPARG, KCNJ11, WFS1, SLC30A8, HHEX, IGF2BP2, CDKAL1, CDKN2A, FTO, JAZF1, CDC123/CAMK1D, TSPAN8/LGR5, THADA, ADAMTS9, NOTCH2* and *GCKR*). Complete data for all 17 SNPs were available for 9740 individuals. Carriers of ≥19 risk alleles (9.7% of individuals) had 51% increased risk of developing at least three components of MetS (OR 1.51 [1.26–1.82], p=0.000012) compared with individuals carrying ≤12 risk alleles (15% of individuals) (Figure 5), translating into 4% increased risk per carried risk allele (OR 1.04 [1.02–1.06], p<0.00001).

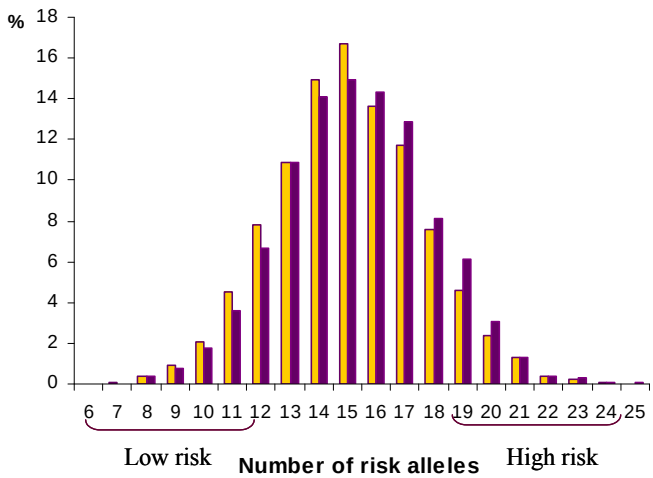


Figure 5. Composite genetic score of 17 SNPs identified by GWAS. Individuals that did not develop MetS during the follow-up time of 24 years are indicated in yellow and individuals that developed MetS are indicated in purple.

DISCUSSION

In this thesis, several aspects of genetics of T2D, obese T2D and MetS have been studied including follow-up investigations of a microsatellite genome wide linkage scan for T2D and candidate gene studies for T2D and MetS. Also, novel candidate genes identified in GWAS for T2D have been explored for their role in MetS.

Identification of genetic variants associated with increased risk of T2D has been cumbersome during the last two decades. The attempts to uncover T2D susceptibility genes have involved candidate gene studies, genome wide linkage studies and fine mapping of the regions that provided linkage signals. Although a number of regions have indicated linkage to T2D in several studies, identification of genes contributing to these signals has been very limited.

As chromosome 18p is the most widely replicated region linked to obese T2D and related phenotypes, identification of susceptibility gene or genes in this region might provide new insights into pathogenesis of T2D and obesity. In our Botnia 2 families from Finland and Sweden, linkage was observed only in the sample stratified for obesity, indicating that genes, partially different from those increasing the risk of T2D, might predispose to obese T2D and that the pathogenesis of obese T2D could differ from non-obese T2D. For this purpose we carried out extensive fine mapping of the region on chromosome 18p11. In a previous attempt in Dutch families, 26 evenly distributed SNPs across the 15 Mb linked region on chromosome 18p11 were studied and a weak association between a polymorphism in *LPIN2* gene and T2D was identified (rs3745012, $p = 0.03$) [67]. We could, however, not replicate this association in our large Swedish case control material ($p = 0.24$ for T2D and $p = 0.63$ for obese T2D). In our study, we genotyped a total of 508 common SNPs (MAF > 0.05) and found nominal association in the family based material for 74 SNPs (Botnia 2 families), but none of the best association signals (n = 5) could be replicated. In addition, more than 2 000 SNPs in the region were analyzed for association with T2D, obese T2D, BMI in T2D patients and fat mass in the DGI GWAS data. From these analyses, four SNPs with best results were selected for replication analysis. However, none of these results could be replicated either. This would argue against common variation contributing to the observed linkage signal.

Results of this extensive study would indicate that common variation in the chromosome 18p11 region may not contribute to a larger extent to the observed linkage result, which supports the current hypothesis that rare variants aggregating in families may contribute to linkage but might be difficult to follow-up using association studies of common variants and difficult to capture by common (MAF > 5%) SNPs. Resequencing of candidate regions may therefore identify novel genes contributing to complex phenotypes. Even though rare variants may be more specific and may contribute to disease in only smaller subgroups of patients, they can provide new insights and mechanism for common diseases. Whether rare variants in any of the

genes on chromosome 18p will explain the observed linkage remains to be answered in future studies.

In studies II and III, we investigated the role of polymorphisms in *IRS1* in T2D. *IRS1* is a good biological candidate gene for T2D as it is a central component in the signal transduction from insulin receptor. Mice with disrupted IRS1 gene develop insulin resistance and hyperinsulinemia but do not become diabetic. This could indicate a compensatory effect from other molecules, i.e. the other IRS molecules (*IRS2-4*). As knocking out *IRS1* in mice causes insulin resistance, *IRS1* makes an obvious candidate gene even for MetS. The G972R polymorphism in *IRS1* has been associated with T2D in some but not all studies and the evidence for it's involvement in the pathogenesis of T2D has been unclear. In our studies we did not observe any association between the G972R in *IRS1* and T2D comprising more than 9 000 individuals or MetS. To further investigate the putative association between *IRS1* and T2D 20 tag SNP, capturing 85% of the common variation in *IRS1*, were genotyped in several family and population based materials. However, as no convincing association could be observed, our studies indicate that common genetic variation in *IRS1* does not have a major role in the pathogenesis of T2D. In very recent and still unpublished studies, an intergenic SNP identified by GWAS and located around 400 kb upstream of *IRS1* (rs2943641) was found associated with T2D ($p = 9.3 \times 10^{-12}$), increased HOMA-IR, elevated triglyceride levels as well as fasting insulin levels [171]. This would make *IRS1* the first gene identified by GWAS predisposing to T2D via effects on insulin resistance. If this SNP would be located in a regulatory element, it could be affecting expression of *IRS1* or other genes in the vicinity. Even though this SNP is located relatively far from the original association of G972R, the inconsistent association observed between G972R and T2D could theoretically depend on population specific allelic association between these two SNPs. It is also possible that genetic variation in *IRS1* predisposes to insulin resistance but not to T2D as such, proposed by animal models [80]. Also, not even all individuals with severe insulin resistance due to mutations in the insulin receptor gene develop diabetes, indicating that insulin resistance alone without some accompanying defect in insulin secretion is not sufficient to cause diabetes. We did not identify association between MetS and G972R in *IRS1* in MPP. In MPP, no measure of insulin resistance is available for testing the hypothesis that *IRS1* associates with insulin resistance. In MPP, measures of abdominal obesity, often associated with insulin resistance, were not available at baseline and were therefore not used in the definition of MetS. In addition, the rs2943641 has not yet been studied in MPP and the association signal from this polymorphism is most likely distinct from G972R.

The role of candidate genes or confirmed susceptibility genes for T2D or components of MetS in the development of MetS were investigated in studies IV and V. In the family based Botnia prospective study, polymorphisms in genes earlier identified as candidate genes for insulin resistance, *PPARG* and *ADRB1*, were found to be associated with increased risk to develop MetS. On the other hand, in the population based larger MPP study with longer follow-up time, mainly polymorphisms in genes affecting hyperglycaemia predicted development of MetS.

Our aim was to search for a unifying genetic factors predisposing to the development of the MetS by investigating polymorphism in genes previously associated with T2D or features of the MetS. Although some of these SNPs were associated with increased risk of the MetS, the risk to develop MetS seemed to be driven by their association with the previously reported phenotype. The results argue against the view that any of these genes would be a common unifying genetic factor predisposing to MetS.

Several genes were investigated in both studies (*PPARG, ADRB1, ADRB2*, *ADRB3* and *PPARGC1A*) and the results for some of these differed between the two studies. In the Botnia prospective study, polymorphisms in *PPARG* and *ADRB1* were associated with development of MetS. These polymorphisms were not associated with prediction of MetS in MPP. There are some differences between the studies that could contribute to the observed differences. First, Botnia prospective study is a family based study enriched for T2D. It is well established that insulin resistance is common among relatives of T2D patients, increasing the probability to identify genes acting through insulin resistance to influence development of MetS. Second, the definitions of MetS in Botnia prospective study and MPP differed. While the NCEP-ATP III definition could be used in the Botnia prospective study, measures of abdominal obesity (waist) and HDL cholesterol levels were not available for a majority of subjects in MPP at baseline and therefore a modified definition of MetS was used (see the methods section for study V). The inconsistencies between the results in the two studies highlight the complexity of the phenotype, the importance of how the phenotypes are defined in association studies as well as the putative power problems even in larger cohorts.

Several definitions of MetS exist and are widely used, emphasizing different aspects of the syndrome. While the NCEP-ATP III definition assigns equal weight to all components, the WHO definition uses insulin resistance and IDF definition uses abdominal obesity as common denominators. This leads to different individuals being diagnosed as having the MetS with the different definitions and also the prevalence varies depending on which definition is used. Thus, genetic factors influencing the risk to develop MetS syndrome could differ depending on how MetS is defined. The discussion whether common pathogenic pathways are involved in MetS is ongoing and the question has been difficult to answer. Identification of gene or genes predisposing to MetS or several of it's components could provide arguments for the common pathogenesis. So far no such genes have been identified. Therefore, an unbiased approach to MetS, such as GWAS could provide novel genes indicating new pathways involved in the pathogenesis of MetS. To date, no GWAS for MetS have been conducted leaving the question of common pathogenic pathway unanswered.

The genetic studies of MetS have focused on candidate gene and linkage studies. Very few, if any genes have been convincingly associated with MetS, although genes associating with more than one of the subcomponents have been reported. Genetic studies of MetS have to deal with heterogeneity in the definition of the phenotype in addition to the other pitfalls of genetic association studies. The genes so far reported as candidate genes for MetS have often been found associated with one or two

components of the syndrome but not to associate with more than two or all of the components. This does not exclude the possibility of a unifying genetic factor for MetS but rather highlights the fact that the unknown pathophysiology of the syndrome makes prediction about a unifying candidate gene difficult. Liver fat content is correlated with all components of MetS, indicating a central role for non-alcoholic fatty liver in the pathogenesis of MetS [172]. Recently, a polymorphism in the adiponutrin gene (*PNPLA3*) has been associated with hepatic fat content [173], making adiponutrin a novel candidate gene for MetS for future studies.

T2D is defined by chronic hyperglycaemia and relatively late onset (typically above 45 years). The prevalence of T2D increases with age but T2D appears with increasing frequencies even in younger age groups and children. T2D thus covers a broad spectrum of patients, with large span in age of onset, duration and severity of the disease. While most patients with long duration of the disease develop micro- and/or macrovascular complications there are patients with long duration of the disease but without any complications. Obesity is a common feature of T2D but far from all patients suffering from T2D are obese. Neither are all obese individuals insulin resistant nor develop T2D. It has therefore been discussed whether T2D indeed can be defined as a single disease or rather a cluster of diseases presenting with chronic hyperglycaemia. The heterogeneity of the disease is a potential problem in genetic association studies, as accurate phenotype definition increases the power of the studies. On the other hand, information about genetic factors influencing the risk of the disease can be used to differentiate between the potentially different T2D subgroups and can lead not only to better diagnostic criteria and better treatment for the patients but also improved possibilities for prediction and prevention of the disease.

Recent progress and future of genetics of complex diseases

The recent progress in genetics of complex diseases has been greatly aided by the development of high-throughput genotyping technologies as well as the rapidly increasing knowledge about the variation in the human genome aided HGP and HapMap projects. Many publicly available databases of human genetic variation exist today, augmenting genetic research. Validation of the information in the public databases and information about the population specific variations has improved dramatically in the last ten years.

The discovery of TCF7L2 demonstrated that identification of new genes with functions previously not implicated in the pathogenesis of T2D was possible and lead to increased optimism in the field. The hope that GWAS could identify new genes and provide new insights into the aetiology of T2D proved true in 2007 when several GWAS for T2D were published. To date, up to 20 new genes have been associated with T2D. Most of the identified genes have not been previously implied in T2D and studies trying to identify mechanisms by which these genes could affect the pathophysiological processes are ongoing. So far, most of the identified genes were found associated with measures of insulin secretion rather than those of insulin action.

To date (September 2008), over 180 GWASs for complex phenotypes have been published [174], including cancer, psychiatric diseases and metabolic traits.
Many novel genes influencing complex traits, like T2D, lipid traits, myocardial infarction and blood pressure have been identified. The GWAS for lipid traits have identified several new genes and confirmed previously known susceptibility genes and all to all 30 loci influencing lipid traits have now been identified.

The recent success in genetics of polygenic diseases still leaves many questions unanswered. Many of the identified loci contain many genes in LD with each other and a more thorough investigation of association signals together with functional studies will be needed to identify the susceptibility genes. The first attempts for genetic prediction show promising, although yet modest results. As new genes and pathways are discovered it is probable that genetic prediction measures will be improved which can lead to possibilities to identify individuals at high risk to develop diseases such as T2D, MetS, dyslipideamia and CVD in addition to teaching us new insights about biological processes operating in these common diseases. Although many novel genes for complex diseases have been identified, known variants explain only a small part of the familial aggregation [175,176]. The future challenges to identify genetic factors contributing to complex diseases will most certainly include investigation of the role of rare variants by sequencing large segments of the genome; enabled by the new emerging sequencing technologies. Also the much less understood role of CNV in the development of complex diseases has lately come into focus.

SUMMARY & CONCLUSIONS

Study I Investigation of common genetic variation on chromosome 18p, a region linked to obese T2D, by tag SNP mapping in our T2D family material showed several nominal association signals. These results and publicly available results from GWAS for related phenotypes could not be replicated in a large Swedish case-control material. These results would argue against common variation contributing to observed linkage on chromosome 18p.

Study II The well powered study of 9 000 individuals could not confirm association between G972R in the *IRS1* gene and T2D. This would indicate that G972R does not have a major role in development of T2D.

Study III No association between common variation in *IRS1*, captured by 20 tag SNPs, and T2D was observed in patients of northern European ancestry. Based on our findings, we conclude that common variation in the IRS1 gene region does not have a major impact on the development of T2D. It is, however, possible that yet uncaptured variation in IRS1 might increase risk of T2D.

Study IV Genetic variation in *PPARG* and *ADRB1* predicted development of MetS using the NCEP-ATP III definition and the G389R polymorphism in *ADRB1* predicted MetS using all three definitions (NCEP-ATP III, WHO and IDF) in the Botnia prospective study. These findings emphasize the role of free fatty acid metabolism in the pathogenesis of MetS.

Study V In MPP, polymorphisms in candidate genes for T2D (*TCF7L2, WFS1, IGF2BP2*) and obesity (*FTO)* predicted development of MetS by increasing risk of development of the component of the syndrome to which the original association was reported. These data do not support the view that the different components of MetS share a common genetic background.

POPULÄRVETENSKAPLIG SAMMANFATTNING

Diabetes är en folksjukdom som drabbar ungefär tio procent av befolkning. Den vanligaste formen är typ 2 diabetes, också kallad åldersdiabetes eftersom de flesta människor insjuknar senare i livet. Typ 2 diabetes kännetecknas av höga blodsockerhalter, eller glukoshalter, och behandlas med diet, blodsockersänkande tabletter och i vissa fall insulin. Diabetes är också sammankopplat med flera följdsjukdomar, så som hjärt- och kärlsjukdomar. Typ 2 diabetes utvecklas när produktionen av insulin i bukspottskörteln inte längre klarar av att kompensera för ett nedsatt glukosupptag i celler, vilket resulterar i kroniskt förhöjda halter av glukos i blodet. Vad som orsakar det försämrade glukosupptag eller insulinutsöndringen är inte känt idag och det är därför viktigt att leta efter de underliggande mekanismerna för att bättre kunna behandla och förebygga uppkomsten av typ 2 diabetes.

Metabolt syndrom definieras som en grupp av ämnesomsättningsstörningar som tillsammans ökar risken för både typ 2 diabetes och kardiovaskulära sjukdomar. Dessa störningar inkluderar fetma, högt blodtryck, dåliga blodfetter och höga blodsockerhalter. Dessa störningar förekommer ofta tillsammans och man har därför konstaterat att det skulle kunna finnas en gemensam sjukdomsalstrande mekanism bakom dessa, även om orsaken är okänd.

Både typ 2 diabetes och metabolt syndrom är s.k. komplexa sjukdomar med en ärftlig komponent. Dessa sjukdomar orsakas av många gener där var och en bidrar endast lite till sjukdomsbilden och vars effekt modifieras av miljöfaktorer så som fysisk inaktivitet och högt näringsintag. Detta är bidragande faktorer till varför identifiering av gener som medverkar i utvecklingen av komplexa sjukdomar har varit svår. Först under senare år har man kunnat identifiera ett flertal gener av betydelse för komplexa sjukdomar genom studier där ett stort antal genetiska markörer fördelade över hela genomet undersökts. Dessa studier har bl. a. bidragit till identifiering av 12 tidigare okända typ 2 diabetes gener. Många av dessa gener har ingen känd funktion inom typ 2 diabetes och identifierar nya möjliga mekanismer för sjukdomsutvecklingen.

Ett sätt att söka efter orsaker till varför typ 2 diabetes och metabolt syndrom utvecklas är att leta efter genetiska varianter som ökar risken för typ 2 diabetes. Målet med denna avhandling var att undersöka genetiska varianter och dess roll i utvecklingen av typ 2 diabetes och metabolt syndrom och flera olika strategier har använts i sökandet.

Den korta armen på kromosom 18 visar en koppling till typ 2 diabetes och fetma i flera oberoende studier och är därför en region som skulle kunna innehålla en eller flera gener som ökar risken för att utveckla typ 2 diabetes kombinerat med fetma. I den första studien undersökte vi 508 olika varianter lokaliserade i och kring gener i den intressanta regionen på kromosom 18 i familjematerial från Sverige och Finland. Vi fann ett samband mellan flera av varianterna och typ 2 diabetes samt typ 2 diabetes tillsammans med fetma. De varianter med starkast samband till sjukdomen försökte vi bekräfta i en studie där vi jämförde frekvensen av varianten mellan sjuka och friska

individer, en s.k. associationsstudie. Då vi inte kunde bekräfta en koppling mellan dessa varianter och typ 2 diabetes tillsammans med fetma, drar vi slutsatsen att vanlig genetisk variation på kromosom 18 inte bidrar avsevärt till den noterade kopplingen.

I studie II och III undersöktes den genetiska variationen i *IRS1* genen och dess typ 2 diabetes. I studie II undersöktes mer än 9 000 individer med hänsyn till en variant som sedan tidigare är associerad med typ 2 diabetes. I vår studie kunde vi inte se någon skillnad i frekvensen av denna variant mellan friska och sjuka individer och vi drog därmed slutsatsen att denna variant inte har någon roll i utvecklande av typ 2 diabetes i den studerade populationen population.

Eftersom *IRS1* genen kodar för ett protein som överför signaler från insulinreceptorn in i cellen och därför är en bra kandidatgen för typ 2 diabetes, ville vi i studie III undersöka om någon annan genetisk variation kan bidra till ökad risk av typ 2 diabetes. Vi valde därför 20 varianter på ett sätt så att de skulle täcka mer än 85% av den vanliga variationen och jämförde frekvenserna av dessa varianter mellan friska och sjuka individer. Då vi inte såg någon skillnad i frekvensen av de studerade varianterna drog vi slutsatsen att vanlig genetisk variation i *IRS1* genen inte bidrar nämnvärt till risken att utveckla typ 2 diabetes.

I studie IV och V undersöktes sambandet mellan varianter i olika kandidatgener för det metabola syndromet. Detta gjordes i s.k. prospektiva studier, där friska individer följts under en tid och frekvensen av varianter jämförs sedan mellan individer som utvecklat sjukdomen, i vårt fall metabola syndromet, och individer som förblivit friska. I studie IV undersöktes åtta varianter i sex gener som tidigare visat på ett samband med de ämnesomsättningsstörningar som utgör en del av metabola syndromet. Vi såg ett samband mellan varianter i *PPARG* och *ADRB1* gener och metabola syndromet. Dessa gener har en funktion i omsättningen av fetter och vi såg att dessa varianter också hade en påverkan på hur halterna av blodfetter förändrades under den tid individerna följdes.

I studie V undersöktes 27 varianter i gener som antingen visat ett väletablerat samband till typ 2 diabetes eller en mindre etablerad samband med åtminstone två ämnesomsättningsstörningar som utgör en del av metabola syndromet, hos 16 000 individer som följts i 24 år. Av dessa gener kunde man se ett samband mellan metabola syndromet och *TCF7L2*, *FTO* och *WFS1* generna. Dessa gener bidrog till ökad risk av metabola syndromet genom att påverka specifika ämnesomsättningsstörningar; *FTO* medförde en ökad risk för fetma och *TCF7L2* och *WFS1* medförde ökad risk för höga blodsockerhalter. Ingen av de undersökta varianterna visade samband till mer än två av de ämnesomsättningsstörningar som ingår i metabola syndromet. Genom att kombinera riskvarianter av de 17 väletablerade generna i varje individ, undersökte vi om kombinationen av dessa varianter ökar risken att utveckla metabola syndromet. Vi noterade att individer med ett högt antal riskvarianter hade 50% högre risk att utveckla metabola syndromet jämfört med individer med ett lågt antal riskvarianter. Sammanfattningsvis tyder våra resultat från studierna IV och V på att det inte finns en sammanlänkade gen som ökar risken för att utveckla metabola syndromet. Eftersom vi endast undersökt gener med känd funktion

eller med ett känt samband i utvecklingen av en eller flera av de ämnesomsättningsstörningar som är en del av metabola syndromet, kan vi inte utesluta att andra gener med ännu ej känd funktion i ämnesomsättningsstörningar kan vara sammanlänkade för metabola syndromet.

De framsteg som nyligen gjorts inom genetiken av komplexa sjukdomar har lett till identifiering av gener med okänd funktion i sjukdomsbilden och kommer sannolikt leda till upptäckten av nya mekanismer för sjukdomens utvecklande, vilket i sin tur kan leda till förbättrad behandlig och förebyggande av dessa sjukdomar. Vidare undersökning av dessa gener, så som vi har gjort i dessa studier, bidrar till ökad kunskap om eventuella mekanismer.

ACKNOWLEDGEMENTS

There are number of people who greatly contributed to this work in various ways and to whom I would like thank:

During my thesis I had two people guiding me through the field of science in an excellent way, my supervisors **Marju Orho-Melander** and **Lief Groop** and who I wish to thank for all their support during my years in the lab. To **Marju**, for your enthusiasm and devotion to science, for your ability to identify the core of the problem and find the solutions, I have learned so much from you! To **Leif,** for giving me the opportunity to work in your lab, your truly inspiring enthusiasm and faith in me and for giving me the chance to try my wings under your supervision!

My warmest thanks to **Malin S**, a great mentor in all things practical and a friend, who taught me everything about lab work and who I can always count on to help me "think" when solving genotyping (and other) problems! Without your expertise there would be very little thesis to write…

To all my co-authors, especially **Jose Florez** and **Valeriya Lyssenko** for all the great work on the manuscripts included in the thesis, it's really been a pleasure working with you.

To all my fellow PhD students, present and past, for making these past years memorable. During our travels around the world (or at least Europe), parties, working out the mysteries in the lab and sharing each others failures and successes you have all become important to me and I am proud to call you my friends. It has been a pleasure working with you! Thank you all for the great times we hade, both at and outside work: **Jenny** my partner in crime in so many ways… there are things you understand like nobody else and you are hard to match at partying although I keep trying☺, **Johan** for being the best personal shopper in the world and for never complaining about us girls and for being my greatest support during our days in the lab, **Maria C** thank you for being a great friend and room mate during these years and for all the laughs, the laughing cat made my day more then once ☺, **Fredrik**, **Camilla**, **Emma** & **Charlotte G** for sharing laughters, rooms (be it offices or hotel rooms during our travels) and knowledge, **Charlotta** & **Anna J** for being such a great students (my first!) when you came to the lab, I had fun teaching and learned a lot from you, **Eka** for our cigarette breaks and your spontaneous dinner parties, **Hemang** for taking me out for dinners during late nights in the lab, **Tina & Lovisa** for always being first and foremost my friends! I'm thankful for all the nights of fun either partying or organizing the parties and for always finding you online when I needed support or just a chat, sometimes late into the night… you have made these last month so much easier!, **Anders O**, you are not to bad at shopping either ☺, **Lina & Maja** expanding our group and adding new perspectives on life to our discussions.
I'm looking forward to all the new adventures ahead of us!

To all co-workers in the lab: **Margareta** for being a great mentor in the lab and for teaching us all the routines so needed for good lab work! We all learned from your experience! **Anna B** for taking care of the lab and other important issues like Friday breakfasts, **Esa** for always sharing all you know and always finding the answers to what you don't know, **Lena** for sharing the lab, the responsibilities, the laughs, **June & Mona** for all the laughs at coffee table, **Peter A** for always taking your time to explain the statistical issues and listen to my frustrations, **Johan Hu** for taking care of all the computer and server problems and for always writing clever scripts to help us in our everyday work, **Ulrika B-N** for help with all kinds of administrative things, like getting hold of signatures, and for organizing great work parties, **Barbro** for helping me tackle the ABI-people in the very beginning and all the long talks **Ola, Peter O, Marloes, Nael, Mikael , Charlotte L, Martin, Olle, Cristiano, Phillipe, Tord, Jacqueline, Jasmina, Mattias, Jalal, Lisbeth, Malin N**, **Nicolo, Kristina, Damon**, for being such a nice people to be around both at and off work.

To the all DPLU people: **Hindrik, Erik, Corrado, Olga** & **Albert** for believing in our PhD student activities and especially to our PhD student organizing group (we never did find any catchy name for us ;)!): **Kalle, Helena, Anders R, Ulrika A, Cecilia N** and **Karin** for all the fun during our parties and planning dinners!

To my colleagues at RSKC in Malmö, especially **Ann-Kristin** for taking me in when staring the new group, **Maria S & Liselotte** for teaching me everything I know about Sequenom and **Holger** for always being ready to engage in any kind of discussion, about science or life in general.

To other co-workers**: Avinash, Hee-Bok, Lisa, Anita, Bodil, Marie N, Elisabeth, Carl-David, Maj-Lis, Caroline, Eero, Martins, Jeanette, Britt, Mozghan, Targ, Ylva, Ida, Rasmus, Thea, Sabina, Tomas, Åke** for creating such a nice environment at work and always bringing interesting topics to the coffee table!

I have been fortunate to be blessed with many true friends and a fantastic family and I wish to thank them all for being part of my life! I would especially like to thank:

Marketa, my very oldest and dearest friend in the whole world! We have always belonged together and there are stories to prove that! No matter what will happen in life you will always be a part of it and I hate that I spend a lot of time missing you! Luckily, means of communication have improved considerably, making our separation a bit easier… Thank you for always listening! Pořád jsem to já, pořád jsme to my…

Anna Å & Anna H-U, for being my friends for so long I can hardly remember… The summer of 92' is unforgettable, who said "hybris" was bad? ;)

Karro, **Maja** & **Petra** (appearing in alphabetical order! ;)), thank you for being an enormous support when life was tough and an enormous source of joy during all other times! Our evenings of Pictionary can still make me laugh… Thank you for always knowing when I need chocolate, dinner invitation, walk in the park or just a phone call…

To the Czech women's mafia for being such a good friends, **Tereza**, who doubled the number of Czech girls in the lab, thank you for showing me that a two language event actually can work and making all of us to cook Czech food, **Marcela** for letting me practice my teaching skills on you, for your travel company and for our long talks about everything and nothing & **Zuzana** for being a great room mate, for the spontaneous parties and loooong talks in the sofa, for the cover picture and thank you (and **Edouard**) for distracting me during the time of writing this thesis.

To my large, noisy, extended and beloved family, especially my aunts and cousins, for always being my safe port, even when life was rough, for always giving me a sense of belonging even if I was hundreds of miles away. For all the great family reunions, for always taking time to see me when I'm swooping through town and for all the things yet to come…To my **Grandfather**, for always having faith in me, for understanding me better then I realized and for building our summer house, the place for all of us to meet and where my roots always will be! It has always been the place where I go to recharge my batteries and a place I always call home!

To my parents, Eva and Pavel, for always giving me all the love and support I needed but also giving me the right amount of space to grow. My mum, for showing me that you can achieve whatever you set your mind to and for always asking if I didn't forget to eat (even if you are thousands miles away). Your generosity and love give me strength every day. To Pavel, for always knowing when to stop mum when she starts to get a little overprotective… and for taking care of her when I can't. I love you both and I always miss you when you travel around the world (which is most of the time)!

REFERENCES

1. Lander, E.S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
2. Venter, J.C. et al. The sequence of the human genome. *Science* **291**, 1304-51 (2001).
3. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931-45 (2004).
4. The International HapMap Project. *Nature* **426**, 789-96 (2003).
5. Kruglyak, L. & Nickerson, D.A. Variation is the spice of life. *Nat Genet* **27**, 234-6 (2001).
6. Reich, D.E., Gabriel, S.B. & Altshuler, D. Quality and completeness of SNP databases. *Nat Genet* **33**, 457-8 (2003).
7. Prak, E.T. & Kazazian, H.H., Jr. Mobile elements and the human genome. *Nat Rev Genet* **1**, 134-44 (2000).
8. Bailey, J.A., Yavor, A.M., Massa, H.F., Trask, B.J. & Eichler, E.E. Segmental duplications: organization and impact within the current human genome project assembly. *Genome Res* **11**, 1005-17 (2001).
9. Feuk, L., Carson, A.R. & Scherer, S.W. Structural variation in the human genome. *Nat Rev Genet* **7**, 85-97 (2006).
10. Lewontin, R.C. The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. *Genetics* **49**, 49-67 (1964).
11. Gabriel, S.B. et al. The structure of haplotype blocks in the human genome. *Science* **296**, 2225-9 (2002).
12. Manolio, T.A., Brooks, L.D. & Collins, F.S. A HapMap harvest of insights into the genetics of common disease. *J Clin Invest* **118**, 1590-605 (2008).
13. Livak, K.J. SNP genotyping by the 5'-nuclease reaction. *Methods Mol Biol* **212**, 129-47 (2003).
14. Syvanen, A.C. Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nat Rev Genet* **2**, 930-42 (2001).
15. Newton, C.R. et al. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res* **17**, 2503-16 (1989).
16. http://www.kbioscience.co.uk/chemistry/chemistry_application_note.html.
17. Syvanen, A.C., Aalto-Setala, K., Harju, L., Kontula, K. & Soderlund, H. A primer-guided nucleotide incorporation assay in the genotyping of apolipoprotein E. *Genomics* **8**, 684-92 (1990).
18. Syvanen, A.C. From gels to chips: "minisequencing" primer extension for analysis of point mutations and single nucleotide polymorphisms. *Hum Mutat* **13**, 1-10 (1999).
19. Braun, A., Little, D.P. & Koster, H. Detecting CFTR gene mutations by using primer oligo base extension and mass spectrometry. *Clin Chem* **43**, 1151-8 (1997).
20. Ross, P., Hall, L., Smirnov, I. & Haff, L. High level multiplex genotyping by MALDI-TOF mass spectrometry. *Nat Biotechnol* **16**, 1347-51 (1998).
21. Oeth, P., Beaulieu, M., Park, C., Kosman, D., del Mistro, G., van den Boom, v.d.D., Jurinke, C. iPLEX™ Assay: Increased Plexing Efficiency and Flexibility for MassARRAY System Through Single Base Primer Extension with Mass-Modified Terminators. *SEQUENOM Application Note* **Doc. No. 8876-006, R01**, 1-12 (2005).

22. Burton, P.R., Tobin, M.D. & Hopper, J.L. Key concepts in genetic epidemiology. *Lancet* **366**, 941-51 (2005).
23. Dawn Teare, M. & Barrett, J.H. Genetic linkage studies. *Lancet* **366**, 1036-44 (2005).
24. Morton, N.E. Sequential tests for the detection of linkage. *Am J Hum Genet* **7**, 277-318 (1955).
25. Lander, E. & Kruglyak, L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* **11**, 241-7 (1995).
26. Risch, N.J. Searching for genetic determinants in the new millennium. *Nature* **405**, 847-56 (2000).
27. Risch, N. & Merikangas, K. The future of genetic studies of complex human diseases. *Science* **273**, 1516-7 (1996).
28. Alberti, K.G. & Zimmet, P.Z. 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**, 539-53 (1998).
29. Karvonen, M. et al. Incidence of childhood type 1 diabetes worldwide. Diabetes Mondiale (DiaMond) Project Group. *Diabetes Care* **23**, 1516-26 (2000).
30. Tuomi, T. et al. Clinical and genetic characteristics of type 2 diabetes with and without GAD antibodies. *Diabetes* **48**, 150-7 (1999).
31. Tuomi, T. et al. Antibodies to glutamic acid decarboxylase reveal latent autoimmune diabetes mellitus in adults with a non-insulin-dependent onset of disease. *Diabetes* **42**, 359-62 (1993).
32. Froguel, P. et al. Close linkage of glucokinase locus on chromosome 7p to early-onset non-insulin-dependent diabetes mellitus. *Nature* **356**, 162-4 (1992).
33. Horikawa, Y. et al. Mutation in hepatocyte nuclear factor-1 beta gene (TCF2) associated with MODY. *Nat Genet* **17**, 384-5 (1997).
34. Malecki, M.T. et al. Mutations in NEUROD1 are associated with the development of type 2 diabetes mellitus. *Nat Genet* **23**, 323-8 (1999).
35. Vaxillaire, M. et al. A gene for maturity onset diabetes of the young (MODY) maps to chromosome 12q. *Nat Genet* **9**, 418-23 (1995).
36. Yamagata, K. et al. Mutations in the hepatocyte nuclear factor-4alpha gene in maturity-onset diabetes of the young (MODY1). *Nature* **384**, 458-60 (1996).
37. Yamagata, K. et al. Mutations in the hepatocyte nuclear factor-1alpha gene in maturity-onset diabetes of the young (MODY3). *Nature* **384**, 455-8 (1996).
38. van den Ouweland, J.M., Lemkes, H.H., Gerbitz, K.D. & Maassen, J.A. Maternally inherited diabetes and deafness (MIDD): a distinct subtype of diabetes associated with a mitochondrial tRNA(Leu)(UUR) gene point mutation. *Muscle Nerve* **3**, S124-30 (1995).
39. Frayling, T.M. Genome-wide association studies provide new insights into type 2 diabetes aetiology. *Nat Rev Genet* **8**, 657-62 (2007).
40. Bowden, D.W. et al. Linkage of genetic markers on human chromosomes 20 and 12 to NIDDM in Caucasian sib pairs with a history of diabetic nephropathy. *Diabetes* **46**, 882-6 (1997).
41. Busfield, F. et al. A genomewide search for type 2 diabetes-susceptibility genes in indigenous Australians. *Am J Hum Genet* **70**, 349-57 (2002).
42. Duggirala, R. et al. Linkage of type 2 diabetes mellitus and of age at onset to a genetic location on chromosome 10q in Mexican Americans. *Am J Hum Genet* **64**, 1127-40 (1999).

43. Elbein, S.C., Hoffman, M.D., Teng, K., Leppert, M.F. & Hasstedt, S.J. A genome-wide search for type 2 diabetes susceptibility genes in Utah Caucasians. *Diabetes* **48**, 1175-82 (1999).
44. Lindgren, C.M. et al. Genomewide search for type 2 diabetes mellitus susceptibility loci in Finnish families: the Botnia study. *Am J Hum Genet* **70**, 509-16 (2002).
45. Luo, T.H. et al. A genome-wide search for type II diabetes susceptibility genes in Chinese Hans. *Diabetologia* **44**, 501-6 (2001).
46. Mori, Y. et al. Genome-wide search for type 2 diabetes in Japanese affected sib-pairs confirms susceptibility genes on 3q, 15q, and 20q and identifies two new candidate Loci on 7p and 11p. *Diabetes* **51**, 1247-55 (2002).
47. Parker, A. et al. A gene conferring susceptibility to type 2 diabetes in conjunction with obesity is located on chromosome 18p11. *Diabetes* **50**, 675-80 (2001).
48. Shaw, J.T. et al. Novel susceptibility gene for late-onset NIDDM is localized to human chromosome 12q. *Diabetes* **47**, 1793-6 (1998).
49. Wiltshire, S. et al. 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**, 553-69 (2001).
50. Vionnet, N. et al. Genomewide search for type 2 diabetes-susceptibility genes in French whites: evidence for a novel susceptibility locus for early-onset diabetes on chromosome 3q27-qter and independent replication of a type 2-diabetes locus on chromosome 1q21-q24. *Am J Hum Genet* **67**, 1470-80 (2000).
51. Aulchenko, Y.S. et al. A genome-wide search for genes involved in type 2 diabetes in a recently genetically isolated population from the Netherlands. *Diabetes* **52**, 3001-4 (2003).
52. Reynisdottir, I. et al. Localization of a susceptibility gene for type 2 diabetes to chromosome 5q34-q35.2. *Am J Hum Genet* **73**, 323-35 (2003).
53. van Tilburg, J.H. et al. A genome-wide scan in type 2 diabetes mellitus provides independent replication of a susceptibility locus on 18p11 and suggests the existence of novel Loci on 2q12 and 19q13. *J Clin Endocrinol Metab* **88**, 2223-30 (2003).
54. Horikawa, Y. et al. Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat Genet* **26**, 163-75 (2000).
55. Grant, S.F. et al. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet* **38**, 320-3 (2006).
56. Chen, C.H., Chang, C.J., Yang, W.S., Chen, C.L. & Fann, C.S. A genome-wide scan using tree-based association analysis for candidate loci related to fasting plasma glucose levels. *BMC Genet* **4 Suppl 1**, S65 (2003).
57. Hsueh, W.C. et al. Genome-wide and fine-mapping linkage studies of type 2 diabetes and glucose traits in the Old Order Amish: evidence for a new diabetes locus on chromosome 14q11 and confirmation of a locus on chromosome 1q21-q24. *Diabetes* **52**, 550-7 (2003).
58. Jun, G., Song, Y., Stein, C.M. & Iyengar, S.K. An autosome-wide search using longitudinal data for loci linked to type 2 diabetes progression. *BMC Genet* **4 Suppl 1**, S8 (2003).
59. Atwood, L.D., Heard-Costa, N.L., Fox, C.S., Jaquish, C.E. & Cupples, L.A. Sex and age specific effects of chromosomal regions linked to body mass index in the Framingham Study. *BMC Genet* **7**, 7 (2006).

60. Shmulewitz, D. et al. Linkage analysis of quantitative traits for obesity, diabetes, hypertension, and dyslipidemia on the island of Kosrae, Federated States of Micronesia. *Proc Natl Acad Sci U S A* **103**, 3502-9 (2006).
61. Ng, M.C. et al. Genome-wide scan for type 2 diabetes loci in Hong Kong Chinese and confirmation of a susceptibility locus on chromosome 1q21-q25. *Diabetes* **53**, 1609-13 (2004).
62. Kraja, A.T. et al. Quantitative trait loci for metabolic syndrome in the Hypertension Genetic Epidemiology Network study. *Obes Res* **13**, 1885-90 (2005).
63. Perusse, L. et al. A genome-wide scan for abdominal fat assessed by computed tomography in the Quebec Family Study. *Diabetes* **50**, 614-21 (2001).
64. Yip, A.G. et al. Search for genetic factors predisposing to atherogenic dyslipidemia. *BMC Genet* **4 Suppl 1**, S100 (2003).
65. Liu, X.Q., Hanley, A.J. & Paterson, A.D. Genetic analysis of common factors underlying cardiovascular disease-related traits. *BMC Genet* **4 Suppl 1**, S56 (2003).
66. Sammalisto, S. European Human Genetics Conference (http://www.eshg.org/eshg2006/index1.htm). (2006).
67. Aulchenko, Y.S. et al. LPIN2 is associated with type 2 diabetes, glucose metabolism, and body composition. *Diabetes* **56**, 3020-6 (2007).
68. Jellema, A., Zeegers, M.P., Feskens, E.J., Dagnelie, P.C. & Mensink, R.P. Gly972Arg variant in the insulin receptor substrate-1 gene and association with Type 2 diabetes: a meta-analysis of 27 studies. *Diabetologia* **46**, 990-5 (2003).
69. McAteer, J.B. et al. The ENPP1 K121Q polymorphism is associated with type 2 diabetes in European populations: evidence from an updated meta-analysis in 42,042 subjects. *Diabetes* **57**, 1125-30 (2008).
70. Weedon, M.N. et al. Meta-analysis and a large association study confirm a role for calpain-10 variation in type 2 diabetes susceptibility. *Am J Hum Genet* **73**, 1208-12 (2003).
71. Gudmundsson, J. et al. Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat Genet* **39**, 977-83 (2007).
72. Winckler, W. et al. Evaluation of common variants in the six known maturity-onset diabetes of the young (MODY) genes for association with type 2 diabetes. *Diabetes* **56**, 685-93 (2007).
73. Weedon, M.N. et al. Genetic regulation of birth weight and fasting glucose by a common polymorphism in the islet cell promoter of the glucokinase gene. *Diabetes* **54**, 576-81 (2005).
74. Weedon, M.N. et al. A large-scale association analysis of common variation of the HNF1alpha gene with type 2 diabetes in the U.K. Caucasian population. *Diabetes* **54**, 2487-91 (2005).
75. Holmkvist, J. et al. Common variants in maturity-onset diabetes of the young genes and future risk of type 2 diabetes. *Diabetes* **57**, 1738-44 (2008).
76. Sun, X.J. et al. Structure of the insulin receptor substrate IRS-1 defines a unique signal transduction protein. *Nature* **352**, 73-7 (1991).
77. Zeggini, E. et al. Association studies of insulin receptor substrate 1 gene (IRS1) variants in type 2 diabetes samples enriched for family history and early age of onset. *Diabetes* **53**, 3319-22 (2004).
78. Clausen, J.O. et al. Insulin resistance: interactions between obesity and a common variant of insulin receptor substrate-1. *Lancet* **346**, 397-402 (1995).

79. Kovacs, P. et al. The role of insulin receptor substrate-1 gene (IRS1) in type 2 diabetes in Pima Indians. *Diabetes* **52**, 3005-9 (2003).
80. Araki, E. et al. Alternative pathway of insulin signalling in mice with targeted disruption of the IRS-1 gene. *Nature* **372**, 186-90 (1994).
81. Nandi, A., Kitamura, Y., Kahn, C.R. & Accili, D. Mouse models of insulin resistance. *Physiol Rev* **84**, 623-47 (2004).
82. Lehmann, J.M. et al. An antidiabetic thiazolidinedione is a high affinity ligand for peroxisome proliferator-activated receptor gamma (PPAR gamma). *J Biol Chem* **270**, 12953-6 (1995).
83. Altshuler, D. et al. The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat Genet* **26**, 76-80 (2000).
84. Deeb, S.S. et al. A Pro12Ala substitution in PPARgamma2 associated with decreased receptor activity, lower body mass index and improved insulin sensitivity. *Nat Genet* **20**, 284-7 (1998).
85. Saxena, R. et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316**, 1331-6 (2007).
86. Scott, L.J. et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316**, 1341-5 (2007).
87. Sladek, R. et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **445**, 881-5 (2007).
88. Steinthorsdottir, V. et al. A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat Genet* **39**, 770-5 (2007).
89. Zeggini, E. et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* **316**, 1336-41 (2007).
90. Barroso, I. et al. Candidate gene association study in type 2 diabetes indicates a role for genes involved in beta-cell function as well as insulin action. *PLoS Biol* **1**, E20 (2003).
91. Florez, J.C. et al. Haplotype structure and genotype-phenotype correlations of the sulfonylurea receptor and the islet ATP-sensitive potassium channel gene region. *Diabetes* **53**, 1360-8 (2004).
92. Gloyn, A.L. et al. Association studies of variants in promoter and coding regions of beta-cell ATP-sensitive K-channel genes SUR1 and Kir6.2 with Type 2 diabetes mellitus (UKPDS 53). *Diabet Med* **18**, 206-12 (2001).
93. Gloyn, A.L. et al. 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**, 568-72 (2003).
94. Hani, E.H. et al. Missense mutations in the pancreatic islet beta cell inwardly rectifying K+ channel gene (KIR6.2/BIR): a meta-analysis suggests a role in the polygenic basis of Type II diabetes mellitus in Caucasians. *Diabetologia* **41**, 1511-5 (1998).
95. Nielsen, E.M. et al. The E23K variant of Kir6.2 associates with impaired post-OGTT serum insulin response and increased risk of type 2 diabetes. *Diabetes* **52**, 573-7 (2003).
96. Koster, J.C., Marshall, B.A., Ensor, N., Corbett, J.A. & Nichols, C.G. Targeted overactivity of beta cell K(ATP) channels induces profound neonatal diabetes. *Cell* **100**, 645-54 (2000).
97. Schwanstecher, C., Meyer, U. & Schwanstecher, M. K(IR)6.2 polymorphism predisposes to type 2 diabetes by inducing overactivity of pancreatic beta-cell ATP-sensitive K(+) channels. *Diabetes* **51**, 875-9 (2002).

98. Gloyn, A.L. et al. Activating mutations in the gene encoding the ATP-sensitive potassium-channel subunit Kir6.2 and permanent neonatal diabetes. *N Engl J Med* **350**, 1838-49 (2004).
99. Lyssenko, V. et al. Mechanisms by which common variants in the TCF7L2 gene increase risk of type 2 diabetes. *J Clin Invest* **117**, 2155-63 (2007).
100. Cauchi, S. et al. TCF7L2 variation predicts hyperglycemia incidence in a French general population: the data from an epidemiological study on the Insulin Resistance Syndrome (DESIR) study. *Diabetes* **55**, 3189-92 (2006).
101. Dahlgren, A., Zethelius, B., Jensevik, K., Syvanen, A.C. & Berne, C. Variants of the TCF7L2 gene are associated with beta cell dysfunction and confer an increased risk of type 2 diabetes mellitus in the ULSAM cohort of Swedish elderly men. *Diabetologia* **50**, 1852-7 (2007).
102. Damcott, C.M. et al. Polymorphisms in the transcription factor 7-like 2 (TCF7L2) gene are associated with type 2 diabetes in the Amish: replication and evidence for a role in both insulin secretion and insulin resistance. *Diabetes* **55**, 2654-9 (2006).
103. Loos, R.J. et al. TCF7L2 polymorphisms modulate proinsulin levels and beta-cell function in a British Europid population. *Diabetes* **56**, 1943-7 (2007).
104. Marzi, C. et al. Variants of the transcription factor 7-like 2 gene (TCF7L2) are strongly associated with type 2 diabetes but not with the metabolic syndrome in the MONICA/KORA surveys. *Horm Metab Res* **39**, 46-52 (2007).
105. Sandhu, M.S. et al. Common variants in WFS1 confer risk of type 2 diabetes. *Nat Genet* **39**, 951-3 (2007).
106. Franks, P.W. et al. Replication of the association between variants in WFS1 and risk of type 2 diabetes in European populations. *Diabetologia* **51**, 458-63 (2008).
107. Inoue, H. et al. A gene encoding a transmembrane protein is mutated in patients with diabetes mellitus and optic atrophy (Wolfram syndrome). *Nat Genet* **20**, 143-8 (1998).
108. Strom, T.M. et al. Diabetes insipidus, diabetes mellitus, optic atrophy and deafness (DIDMOAD) caused by mutations in a novel gene (wolframin) coding for a predicted transmembrane protein. *Hum Mol Genet* **7**, 2021-8 (1998).
109. Zeggini, E. et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet* **40**, 638-45 (2008).
110. Grarup, N. et al. Studies of association of variants near the HHEX, CDKN2A/B, and IGF2BP2 genes with type 2 diabetes and impaired insulin release in 10,705 Danish subjects: validation and extension of genome-wide association studies. *Diabetes* **56**, 3105-11 (2007).
111. Groenewoud, M.J. et al. Variants of CDKAL1 and IGF2BP2 affect first-phase insulin secretion during hyperglycaemic clamps. *Diabetologia* **51**, 1659-63 (2008).
112. Pascoe, L. et al. Beta cell glucose sensitivity is decreased by 39% in non-diabetic individuals carrying multiple diabetes-risk alleles compared with those with no risk alleles. *Diabetologia* **51**, 1989-1992 (2008).
113. Pascoe, L. et al. Common variants of the novel type 2 diabetes genes CDKAL1 and HHEX/IDE are associated with decreased pancreatic beta-cell function. *Diabetes* **56**, 3101-4 (2007).
114. Staiger, H. et al. Polymorphisms within novel risk loci for type 2 diabetes determine beta-cell function. *PLoS ONE* **2**, e832 (2007).

115. Staiger, H. et al. A candidate type 2 diabetes polymorphism near the HHEX locus affects acute glucose-stimulated insulin release in European populations: results from the EUGENE2 study. *Diabetes* **57**, 514-7 (2008).
116. Lyssenko, V. et al. Clinical and genetic factors predicting future type 2 diabetes. *N Engl J Med in press* (2008).
117. Frayling, T.M. et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* **316**, 889-94 (2007).
118. Fredriksson, R. et al. The obesity gene, FTO, is of ancient origin, up-regulated during food deprivation and expressed in neurons of feeding-related nuclei of the brain. *Endocrinology* **149**, 2062-71 (2008).
119. Gerken, T. et al. The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science* **318**, 1469-72 (2007).
120. Reaven, G.M. Banting lecture 1988. Role of insulin resistance in human disease. *Diabetes* **37**, 1595-607 (1988).
121. World Health Organisation. Definition, Diagnosis, and Classification of Diabetes Mellitus and its Complications. Report of a WHO consultation. Part 1: Diagnosis and Classification of Diabetes Mellitus. Geneva: World Health Organisation. (1999).
122. National Cholesterol Education Program: Executive Summary of the Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol In Adults (Adult Treatment Panel III). *JAMA* **285**, 2486 –2497 (2001).
123. Alberti, K.G., Zimmet, P. & Shaw, J. The metabolic syndrome--a new worldwide definition. *Lancet* **366**, 1059-62 (2005).
124. Cameron, A.J., Shaw, J.E. & Zimmet, P.Z. The metabolic syndrome: prevalence in worldwide populations. *Endocrinol Metab Clin North Am* **33**, 351-75, table of contents (2004).
125. Bosy-Westphal, A. et al. Common familial influences on clustering of metabolic syndrome traits with central obesity and insulin resistance: the Kiel obesity prevention study. *Int J Obes (Lond)* **31**, 784-90 (2007).
126. Freeman, M.S., Mansfield, M.W., Barrett, J.H. & Grant, P.J. Heritability of features of the insulin resistance syndrome in a community-based study of healthy families. *Diabet Med* **19**, 994-9 (2002).
127. Lin, H.F. et al. Heritabilities of the metabolic syndrome and its components in the Northern Manhattan Family Study. *Diabetologia* **48**, 2006-12 (2005).
128. Poulsen, P., Vaag, A., Kyvik, K. & Beck-Nielsen, H. Genetic versus environmental aetiology of the metabolic syndrome among male and female twins. *Diabetologia* **44**, 537-43 (2001).
129. Teran-Garcia, M. & Bouchard, C. Genetics of the metabolic syndrome. *Appl Physiol Nutr Metab* **32**, 89-114 (2007).
130. Neel, J.V. Diabetes mellitus: a "thrifty" genotype rendered detrimental by "progress"? *Am J Hum Genet* **14**, 353-62 (1962).
131. Bengtsson, K. et al. Polymorphism in the beta(1)-adrenergic receptor gene and hypertension. *Circulation* **104**, 187-90 (2001).
132. Dionne, I.J. et al. Association between obesity and a polymorphism in the beta(1)-adrenoceptor gene (Gly389Arg ADRB1) in Caucasian women. *Int J Obes Relat Metab Disord* **26**, 633-9 (2002).

133. Fujisawa, T., Ikegami, H., Kawaguchi, Y. & Ogihara, T. Meta-analysis of the association of Trp64Arg polymorphism of beta 3-adrenergic receptor gene with body mass index. *J Clin Endocrinol Metab* **83**, 2441-4 (1998).
134. Lange, L.A. et al. Association of adipose tissue deposition and beta-2 adrenergic receptor variants: the IRAS family study. *Int J Obes (Lond)* **29**, 449-57 (2005).
135. Large, V. et al. Human beta-2 adrenoceptor gene polymorphisms are highly frequent in obesity and associate with altered adipocyte beta-2 adrenoceptor function. *J Clin Invest* **100**, 3005-13 (1997).
136. Widen, E. et al. Association of a polymorphism in the beta 3-adrenergic-receptor gene with features of the insulin resistance syndrome in Finns. *N Engl J Med* **333**, 348-51 (1995).
137. Meyre, D. et al. Variants of ENPP1 are associated with childhood and adult obesity and increase the risk of glucose intolerance and type 2 diabetes. *Nat Genet* **37**, 863-7 (2005).
138. Olivier, M. et al. Single nucleotide polymorphisms in protein tyrosine phosphatase 1beta (PTPN1) are associated with essential hypertension and obesity. *Hum Mol Genet* **13**, 1885-92 (2004).
139. Pihlajamaki, J. et al. Common polymorphisms of calpain-10 are associated with abdominal obesity in subjects at high risk of type 2 diabetes. *Diabetologia* **49**, 1560-6 (2006).
140. Wang, H. et al. Uncoupling protein-2 polymorphisms in type 2 diabetes, obesity, and insulin secretion. *Am J Physiol Endocrinol Metab* **286**, E1-7 (2004).
141. Vimaleswaran, K.S. et al. The Gly482Ser genotype at the PPARGC1A gene and elevated blood pressure: a meta-analysis of 13,949 individuals. *J Appl Physiol* (2008).
142. Kissebah, A.H. et al. Quantitative trait loci on chromosomes 3 and 17 influence phenotypes of the metabolic syndrome. *Proc Natl Acad Sci U S A* **97**, 14478-83 (2000).
143. Arya, R. et al. Factors of insulin resistance syndrome--related phenotypes are linked to genetic locations on chromosomes 6 and 7 in nondiabetic mexican-americans. *Diabetes* **51**, 841-7 (2002).
144. Cai, G. et al. Principal component for metabolic syndrome risk maps to chromosome 4p in Mexican Americans: the San Antonio Family Heart Study. *Hum Biol* **76**, 651-65 (2004).
145. Kraja, A.T. et al. Two major QTLs and several others relate to factors of metabolic syndrome in the family blood pressure program. *Hypertension* **46**, 751-7 (2005).
146. Loos, R.J. et al. Genome-wide linkage scan for the metabolic syndrome in the HERITAGE Family Study. *J Clin Endocrinol Metab* **88**, 5935-43 (2003).
147. McQueen, M.B., Bertram, L., Rimm, E.B., Blacker, D. & Santangelo, S.L. A QTL genome scan of the metabolic syndrome and its component traits. *BMC Genet* **4 Suppl 1**, S96 (2003).
148. Groop, L. et al. Metabolic consequences of a family history of NIDDM (the Botnia study): evidence for sex-specific parental effects. *Diabetes* **45**, 1585-93 (1996).
149. Mahtani, M.M. et al. Mapping of a gene for type 2 diabetes associated with an insulin secretion defect by a genome scan in Finnish families. *Nat Genet* **14**, 90-4 (1996).

150. Demenais, F. et al. A meta-analysis of four European genome screens (GIFT Consortium) shows evidence for a novel region on chromosome 17p11.2-q22 linked to type 2 diabetes. *Hum Mol Genet* **12**, 1865-73 (2003).
151. Lindholm, E., Agardh, E., Tuomi, T., Groop, L. & Agardh, C.D. Classifying diabetes according to the new WHO clinical stages. *Eur J Epidemiol* **17**, 983-9 (2001).
152. Berglund, G., Elmstahl, S., Janzon, L. & Larsson, S.A. The Malmo Diet and Cancer Study. Design and feasibility. *J Intern Med* **233**, 45-51 (1993).
153. Bergman, R.N., Phillips, L.S. & Cobelli, C. Physiologic evaluation of factors controlling glucose tolerance in man: measurement of insulin sensitivity and beta-cell glucose sensitivity from the response to intravenous glucose. *J Clin Invest* **68**, 1456-67 (1981).
154. Matsuda, M. & DeFronzo, R.A. Insulin sensitivity indices obtained from oral glucose tolerance testing: comparison with the euglycemic insulin clamp. *Diabetes Care* **22**, 1462-70 (1999).
155. Matthews, D.R. et al. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* **28**, 412-9 (1985).
156. Lyssenko, V. et al. Genetic prediction of future type 2 diabetes. *PLoS Med* **2**, e345 (2005).
157. Lyssenko, V. et al. Predictors of and longitudinal changes in insulin sensitivity and secretion preceding onset of type 2 diabetes. *Diabetes* **54**, 166-74 (2005).
158. Berglund, G. et al. Long-term outcome of the Malmo preventive project: mortality and cardiovascular morbidity. *J Intern Med* **247**, 19-29 (2000).
159. Nilsson, P. & Berglund, G. Prevention of cardiovascular disease and diabetes: lessons from the Malmo Preventive Project. *J Intern Med* **248**, 455-62 (2000).
160. Spielman, R.S., McGinnis, R.E. & Ewens, W.J. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* **52**, 506-16 (1993).
161. Boehnke, M. & Langefeld, C.D. Genetic association mapping based on discordant sib pairs: the discordant-alleles test. *Am J Hum Genet* **62**, 950-61 (1998).
162. Lohmueller, K.E., Pearce, C.L., Pike, M., Lander, E.S. & Hirschhorn, J.N. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* **33**, 177-82 (2003).
163. Reis, I.M., Hirji, K.F. & Afifi, A.A. Exact and asymptotic tests for homogeneity in several 2 x 2 tables. *Stat Med* **18**, 893-906 (1999).
164. Purcell, S., Cherny, S.S. & Sham, P.C. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* **19**, 149-50 (2003).
165. Klein, J.P. & Moeschberger, M.L. *Survival analysis: Techniques for censored and truncated data*, 536p (Springer, 2003).
166. Levy, J.C., Matthews, D.R. & Hermans, M.P. Correct homeostasis model assessment (HOMA) evaluation uses the computer program. *Diabetes Care* **21**, 2191-2 (1998).
167. Hayes, M.G. et al. Identification of type 2 diabetes genes in Mexican Americans through genome-wide association studies. *Diabetes* **56**, 3033-44 (2007).
168. A haplotype map of the human genome. *Nature* **437**, 1299-320 (2005).

169. Mootha, V.K. et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* **34**, 267-73 (2003).
170. Patti, M.E. et al. Coordinated reduction of genes of oxidative metabolism in humans with insulin resistance and diabetes: Potential role of PGC1 and NRF1. *Proc Natl Acad Sci U S A* **100**, 8466-71 (2003).
171. Froguel, P. The French/Canadian angle. in *EASD* (http://conference2web.com/easd2008_webcast/Sessions/LatebreakingGWAS.aspx, Rome, 2008).
172. Kotronen, A. & Yki-Jarvinen, H. Fatty liver: a novel component of the metabolic syndrome. *Arterioscler Thromb Vasc Biol* **28**, 27-38 (2008).
173. Romeo, S. et al. Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. *Nat Genet* (2008).
174. http://www.genome.gov/26525384.
175. McCarthy, M.I. et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* **9**, 356-69 (2008).
176. Altshuler, D. & Daly, M. Guilt beyond a reasonable doubt. *Nat Genet* **39**, 813-5 (2007).