From Department of Molecular Medicine and Surgery,
Karolinska Institutet, Stockholm, Sweden

IDENTIFICATION OF SUSCEPTIBILITY GENES IN TYPE 2 DIABETES

Sofia Nordman

Karolinska Institutet

Stockholm 2008

Supervisor: Associate Professor Harvest F. Gu
Department of Molecular Medicine and Surgery
Karolinska Institutet

Co-supervisor: Professor Suad Efendic
Department of Molecular Medicine and Surgery
Karolinska Institutet

Co-supervisor: Professor Claes-Göran Östenson
Department of Molecular Medicine and Surgery
Karolinska Institutet

Opponent: Professor Ann-Christine Syvänen
Department of Medical Sciences, Molecular Medicine
Uppsala University

Examination board:
Associate Professor Eva Ehrenborg
Department of Medicine, Karolinska University Hospital
Karolinska Institutet

Associate Professor Christina Bark
Department of Molecular Medicine and Surgery
Karolinska Institutet

Associate Professor Björn Eliasson
Department of Medicine, Sahlgrenska University Hospital
Göteborg University

Published by Karolinska Institutet. Printed by Universitets service US-AB.

ISBN 978-91-7357-492-1

In memory of my grandfather

ABSTRACT

Identification of the susceptibility genes will offer better understanding of molecular mechanisms underlying T2D pathogenesis, and subsequently may lead to development of novel therapeutic approaches. This thesis mainly concerns the genetic association studies of four candidate genes. They are selected from a region in chromosome 10q linked to T2D or based on involvement of the candidates in specific pathways related to T2D.

IDE plays a principal role in the proteolysis of several peptides in addition to insulin. The gene resides in a region of chromosome 10q linked to T2D. Fourteen SNPs in the IDE and IDE-HHEX regions were genotyped in 321 IGT and 403 NGT subjects selected from SDPP. The analyses of diplotypes (haplotypic genotypes) containing three tag SNPs provided evidence of the association with fasting plasma insulin levels, 2h plasma insulin levels, HOMA-IR and BMI in men, and suggested that the polymorphisms in/near the IDE gene contribute to variance in plasma insulin levels and correlated traits. The TCF7L2 gene is also located in the region of chromosome 10q. Five SNPs in the gene were genotyped in 243 T2D patients and 528 NGT subjects, and they were Swedish men. SNPs rs7901695, rs4506565, rs7903146 and rs12255372 are strongly associated with T2D. As to rs7903146, T2D patients carrying genotypes CT or TT had higher fasting plasma glucose levels, lower HOMA-β index and BMI compared to the patients carrying CC genotype. Furthermore, NGT subjects carrying the risk alleles of SNPs rs7901695 and rs4506565 demonstrated a more pronounced increase in fasting plasma glucose levels during the follow-up period. The study consistently indicated that TCF7L2 has a crucial contribution to impaired insulin secretion underlining the development of T2D. In order to evaluate whether Leu7Pro (T1128C) polymorphism in the NPY gene contributes to the development of T2D, genotyping of this SNP in 263 T2D patients, 309 IGT and 469 NGT subjects was performed. This SNP was significantly associated with IGT and T2D among Swedish men but not women. The carriers with TC and CC genotypes in male IGT subjects had significantly higher fasting plasma glucose in comparison with the TT carriers. A previous study using the Goto-Kakizaki rat implicated that AC3 may be a candidate for T2D. The variation screening in the putative promoter was performed and a novel variant -17A/T was identified. Genotyping of 14 SNPs covering the gene, including the novel variant, was performed in male subjects, including 243 T2D and 188 NGT. Interestingly, SNPs rs2033655 C/T and rs1968482 A/G were significantly associated with T2D patients with BMI $\geq$30 kg/m^2. Further genotyping and analysis in 199 male obese subjects with NGT demonstrated that these two polymorphisms were strongly associated with obesity. The present study thus provides the first evidence that AC3 polymorphisms confer the risk susceptibility to obesity in Swedish men with and without T2D.

In conclusion, this thesis has contributed with information on the role of the IDE, TCF7L2 and NPY genes in the development of T2D and/or control of insulin levels. AC3 may play a role in the development of obesity with and without T2D.

Key words: Type 2 diabetes, single nucleotide polymorphism, genetic association
ISBN 978-91-7357-492-1

LIST OF PUBLICATIONS

I. Gu HF, Efendic S, **Nordman S**, Ostenson CG, Brismar K, Brookes AJ, Prince JA. Quantitative trait loci near the insulin-degrading enzyme (IDE) gene contribute to variation in plasma insulin levels. *Diabetes. 2004 53(8):2137-42.*

II. **Nordman S**, Ostenson CG, Efendic S, Gu HF. Effects of the Transcription factor 7 like 2 (TCF7L2) genetic polymorphisms in type 2 diabetes among Swedish men *(Submitted manuscript).*

III. **Nordman S**, Ding B, Ostenson CG, Karvestedt L, Brismar K, Efendic S, Gu HF. Leu7Pro polymorphism in the neuropeptide Y (NPY) gene is associated with impaired glucose tolerance and type 2 diabetes in Swedish men. *Exp Clin Endocrinol Diabetes. 2005 113(5):282-7.*

IV. **Nordman S**, Abulaiti A, Hilding A, Langberg EC, Humphreys K, Ostenson CG, Efendic S, Gu HF. Genetic variation of the adenylyl cyclase 3 (AC3) locus and its influences to type 2 diabetes susceptibility in Swedish caucasians. *Int J Obes (Lond). 2007 Sept 25 [Epub ahead of print].*

OTHER PUBLICATIONS AND MANUSCRIPT BY THE SAME AUTHOR

I. **Nordman S**, Ståhl F, Abulaiti A. Ostensson CG, Efendic S, Gu HF. Department of Molecular Medicine, Karolinska Institutet, Stockholm, Sweden *Adenylate cyclase 8 (18 exons and 17 introns).* Ratmap database: http://ratmap.gen.gu.se/ResultSearchLocus.htm

II. Langberg EC, Gu HF, **Nordman S**, Efendic S, Östenson CG. Genetic variation in receptor protein tyrosine phosphatase σ is associated with type 2 diabetes in Swedish Caucasians. *Eur J Endocrinol. 2007 157(4):459-64.*

III. Ma J, **Nordman S**, Möllsten A, Falhammar H, Brismar K, Dahlquist G, Efendic S, Gu HF. Distribution of neuropeptide Y Leu7Pro polymorphism in patients with type 1 diabetes and diabetic nephropathy among Swedish and American populations. *Eur J Endocrinol. 2007 157(5):641-5.*

IV. **Nordman S**, Zhang DY, Ma J, Ostensson CG, Efendic S, Gu HF. Evaluation of over-expression of adenylyl cyclase 3 (AC3) in pancreatic islets of Goto-Kakizaki (GK) rat with effects of glucose and insulin. (*Manuscript in prep).*

CONTENTS

LIST OF ABBREVIATIONS

AA	Amino acid
AC	Adenylyl cyclase
AC3	Adenylyl cyclase 3
ALX4	Aristaless-like homeobox 4
APS	Adenosine 5’ phosphosulfate
ARC	Arcurate nucleus
ATP	Adenosine triphosphate
ATP	Adenosine triphosphate
BMI	Body mass index
bp	Base pair
cAMP	Cyclic adenosine monophosphate
CAPN10	Calpain 10
CCD	Charged couple device camera
CDKAL1	CDK5 regulatory subunit associated protein 1-like 1
CDKN2A/2B	Cyclin-dependent kinase inhibitor 2A
CDKN2B	Cyclin-dependent kinase inhibitor 2B
CEU	European Caucasians
CVD	Cardiovascular disease
DASH	Dynamic allele specific hybridization
DMH	Dorsomedial nucleus
DNA	Deoxyribonucleic acid
DZ	Dizygote
FHD	Family history of diabetes
FRET	Fluoroesence energy transfer
FTO	Fat mass and obesity associated
GAD2	Glutamate decarboxylase 2
GIP	Glucose-dependent insulinotropic polypeptide
GK	Goto-Kakizaki
GLP-1	Glucagon-like peptide 1
GWA	Genome-wide association
GWS	Genome-wide scan
HHEX	Hematopoietically expressed homeobox
HOMA	Homeostasis model assessment
HWE	Hardy Weinbergs equilibrium
IDE	Insulin degrading enzyme
IFG	Impaired fasting glucose
IGF2BP2	Insulin-like growth factor 2 mRNA binding protein 2
IGT	Impaired glucose tolerance
K_{ATP}	ATP sensitive potassium channel
KCNJ11	Potassium inwardly-rectifying channel, subfamily J, member 11
KIF11	Kinesin family member 11
LADA	Latent autoimmune diabetes of the adult
LD	Linkage disequilibrium
LDL	Low-density lipoprotein

LHA	Lateral hypothalamic area
MAF	Minor allele frequency
MGB	Minor grove binder
MODY	Maturity onset diabetes of the young
mtDNA	Mitochondrial DNA
MZ	Monozygote
NGT	Normal glucose tolerance
NPY	NeuropeptideY
OGTT	Oral glucose tolerance test
OLA	Oligonucleotide ligation assay
OR	Odds ratio
POMC	Pro-opiomelanocortin
PPARγ	Peroxisome proliferator-activated receptor γ
PPi	Pyrophosphate
PVN	Paravertricular nucleus
RFLP	Restriction fragment length polymorphism
SBE	Single-base extension
SDPP	Stockholm diabetes prevention study
SLC30A8	Solute carrier family 30 (zinc transporter), member 8
SNP	Single nucleotide polymorphism
T1D	Type 1 diabetes
T2D	Type 2 diabetes
TCF7L2	Transcription factor 7-like 2
TDT	Transcription disequilibrium test
Tm	Melting temperature
UTR	Untranslated region
VMH	Ventromedial nucleus
VNTR	Variable number tandem repeat
WHO	World health organization
WHR	Waist hip ratio

1 INTRODUCTION

1.1 GENETIC INTRODUCTION

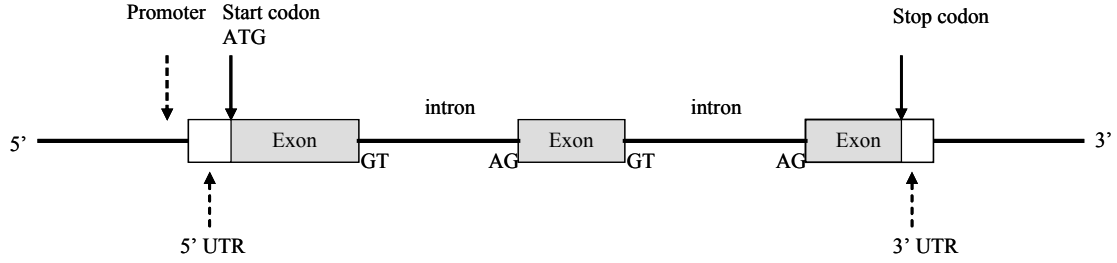
In the recent years, genetic research has been substantially developed and molecular genome projects have revealed the full sequence of human genome. A map of genetic variation, containing ~10 millions single nucleotide polymorphisms (SNPs), and a map of human haplotypes (HapMap) have been established. Genome wide association (GWA) study has recently become an important approach to identify the susceptibility genes in complex diseases(1).

1.1.1 DNA and Chromosomes

The structure of deoxyribonucleic acid (DNA) molecule was published by James Watson and Francis Crick in April 1953 and they shared Nobel prize in physiology or medicine with Maurice Wilkins in 1962. DNA is a double helical molecule carrying information of the genetic makeup in almost all living organisms. DNA sequences are composed with four bases: adenine (A), thymine (T), cytosine (C) and guanine (G). The double helix structure is built up by pairing of these bases, A with T and C with G. They pair up with two and three hydrogen bonds, respectively. Sets of three (triplets) base pairs code for an amino acid (AA). There are start and stop triplets. The size of human genome is ~2.9×10^9 base pairs (bp). The majority of DNA is situated in the cell nucleus, comprising ~25.000 genes. DNA is also found in the mitochondria. Mitochondrial DNA (mtDNA) is circular and consists of 37 genes. The mitochondrial DNA shows a maternal inheritance. The patients with type 2 diabetes (T2D) show a reduced mitochondrial capacity in skeletal muscle (2). There are genetic variations in mitochondrial DNA that are associated with T2D, deafness and/or other traits (3; 4).

In the nucleus, DNA molecule is stored in chromosomes. Human cell nucleus carries a total of 44 autosomes and 2 sex chromosomes. A chromosome consists of a constriction point called the centromere, which divides the short arm labeled petit (p) and long arm queue (q). Each arm is divided into regions delimited by specific landmarks, which are consistent and distinct morphological features. Regions are in turn divided into bands and sub-bands. All bands always count outwards from the centromere. For example, 10q25.3 means chromosome 10 q arm, region 2, band 5, and sub-band 3. The telomere is a structure in the ends of chromosomes.

Figure 1. A typical gene structure



A gene is the basic physical and functional unit of inheritance. A typical gene consists of promoter, exons, introns and 5'- or 3'-untranslated regions (UTRs) (Figure 1). Normally, the introns start with GT and end with AG.

1.1.2 Genetic Variation

Genetic variations comprise both chromosome aberrations (differences in number or structure) and differences in DNA sequences among individuals. Variations may confer the susceptibility to common diseases. There are different types of variations at genomic DNA level. Variable number tandem repeats (VNTR) mainly include satellites, minisatellites, microsatellites and telomeric sequences. If a variation is present with more than 1% in a population, it is defined as a polymorphism. SNPs are the most common variations in genome.

SNPs are the substitutions of nucleotides in genomic DNA. In general, small insertions/deletions are also included. SNPs located in the coding region of the gene are called as cSNPs, which include non-synonymous SNPs with AA changes and synonymous SNPs without AA changes. SNPs are also located in the uncoding regions of the gene, including promoter, intron and UTRs. SNPs in the promoter region may alter transcription binding site and thereby affect the transcriptional activity of the gene. In addition, a number of SNPs are found in inter-genic sequences.

Table 1. The codes of SNPs

	Bi-allelic SNPs						**Tri-allelic SNPs**			
Codes	A/C	A/G	A/T	C/G	C/T	G/T	A/C/T	A/C/G	A/G/T	C/G/T
SNPs	M	R	W	S	Y	K	H	V	D	B

All these SNPs are recorded in the public databases, including dbSNP, HGVbase, CGAP etc. Up to date, ~10 million SNPs are recorded in the databases. SNPs are commonly bi-allelic but rarely have three or four alleles. Table 1 summarizes the codes of SNPs, which are designed by American Association of Biochemistry. Y=C/T is the most common SNP in the genome.

A haplotype is a combination of alleles at multiple linked loci that are inherited together. Haplotype may refer to two or more loci. Information of haplotypes is useful for genetic association studies in complex diseases. The second generation of human haplotype map, including over 3.1 million SNPs (HapMap phase II)(5), has been developed by the international HapMap consortium (http://www.hapmap.org/).

1.1.3 Monogenic or Polygenic Disease Inheritance

An inherited disease is a condition that will pass on from parents to offspring, through the transmission of DNA. Inherited diseases are divided into monogenic and polygenic forms. Monogenic diseases result from modification(s) in a single gene, which are inherited according to Mendelian genetic law. Polygenic diseases are caused by variations in several genes and often influenced with environmental factors. Therefore, they are also referred as complex diseases.

1.1.4 Genetic Studies in Complex Diseases

Different strategies have been used for identification of the susceptibility genes in complex diseases. The principal destination lies between candidate gene approaches, where biological information plays a role in selecting genes to study, and positional cloning, where the genes of interest are localized by studies of pattern of co-segregation of variants between families and populations.

Genome wide scan (GWS) and linkage analyses have been used to reveal the susceptibility locus/loci in chromosomes for complex diseases such as T2D and obesity. In this approach, the aim is to find out location of the gene(s) relative to polymorphic genetic markers (often microsatellites) with known position and spaced along the entire genome. GWS requires family-based material, and this approach is time consuming.

Genetic linkage occurs when genetic loci are inherited together. Loci that are physically close tend to segregate together more often than genetic loci situated far away from each other during meioses. The closer two loci are, the higher is the chance that the loci are linked, since crossing over reactions is less likely to occur between them. Disease genes are mapped by measuring recombination against a panel of different markers spread over the entire genome. In most cases, the recombination will occur frequently, indicating that the disease gene and marker are far apart. Some markers, however, due to their proximity, will tend not to recombine with the disease gene, and these are considered to be linked. Ideally, close markers are identified that flank the disease gene and one can define a candidate region of the genome. The gene(s) participating in the disease is located somewhere in this chromosomal region.

Linkage analyses in humans can be performed by using the LOD (logarithm of odds) score method. This is a statistical test developed by Newton E. Morton. The method includes pedigree collection, estimates of recombination frequency and LOD score calculations for each estimate. The estimate with the highest LOD score will be considered to be the best estimate. The LOD score is calculated as follows:

$$LOD = \log \frac{probability\ of\ \ birth\ sequence\ with\ a\ given\ linkage\ value}{probability\ of\ \ birth\ sequence\ with\ no\ linkage}$$

A LOD score greater than 3.0 is considered evidence for linkage, whereas, a LOD score less than -2.0 is considered evidence to exclude linkage.

Genetic association studies are central for efforts to identify and characterize genomic variants underlying susceptibility to complex diseases. Genetic association studies may take one of two approaches depending on the material selection, either family-based, or population-based. The first approach uses the family materials for sib pair analysis or transmission disequilibrium test (TDT). The second is built on a case control approach.

The candidate gene approach has been widely used to study the genetic basis of complex diseases. The aim in a case-control genetic association study is to determine whether genetic variants occur more or less commonly in cases compared to controls(6). This approach includes five challenges: collection of appropriate patient resources for study, selection of candidate genes for study, assembling and use of tools for

identification of SNPs in the candidate genes, performing genotyping experiments with high-throughput SNP scorning techniques and data analyses.

The Hardy-Weinberg equilibrium (HWE) explains the frequency of genotypes at a given locus in a population under certain conditions. HWE formula is developed by Godfrey Hardy (English mathematician) and Wilhelm Weinberg (German physician), and can be used to discover the probable genotype frequencies in a population. The HWE equation is

$$p + q = 1$$

$$(p+q)^2 = 1 \Rightarrow p^2 + 2pq + q^2 = 1$$

Where
p=frequency of the dominant allele
q=frequency of the recessive allele
p^2*=predicted frequency of homozygote dominant subjects*
2pq=predicted frequency of heterozygous subjects
q^2= predicted frequency of homozygous recessive

This equation is valid for cases with two alleles. For a case controlled by a pair of alleles (A and a), p equals all of the alleles in individuals who are homozygous dominant (AA) and half of the individuals that are heterozygous (Aa) for this trait. In the same way q equals all of the alleles in individuals who are homozygous recessive (aa) and the other half of the individuals that are heterozygous (Aa). In mathematical terms this is: *p= AA+ ½ Aa* and *q=aa + ½ Aa.*

HWE equation is built on several assumptions, including diploid organism, sexual reproduction, equal allele frequencies in males and females, large population, random mating, no mutations, no migration, and no selection in order to be applicable on the sample sets.

In genetic association studies, SNP frequencies differ significantly between cases and controls if the studied SNP is involved in the disease. An SNP associated with the disease may reside in linkage disequilibrium (LD) with other SNPs. In genome, alleles at different loci are sometimes found to be inherited together more or less. Therefore, LD analysis is of importance in genetic association studies (7; 8).

There are several ways to measure LD values (9). The most common methods are r^2 and D', both of them are dependent on the measure of D. Most measures of LD quantify disequilibrium as the difference between the observed frequency of a two-locus haplotype and the expected frequency to show if the alleles are segregating at random. If assuming a two-locus, two-allele model, loci A and B with alleles A/a and B/b respectively in a gene on a specific chromosome, pA, pa, pB or pb represents the common allele and the rare allele frequency at locus respectively. Theoretically it is possible to make 4 different haplotypes from these alleles. These allele combinations can be denoted as AB, Ab, aB, ab and their corresponding frequencies are denoted as pAB, pAb, paB and pab. One of the earliest LD parameter was introduced by Robbins (1918) and named by Lewontin and Kojima (1960) and commonly denoted by a capital D. Following the model mentioned

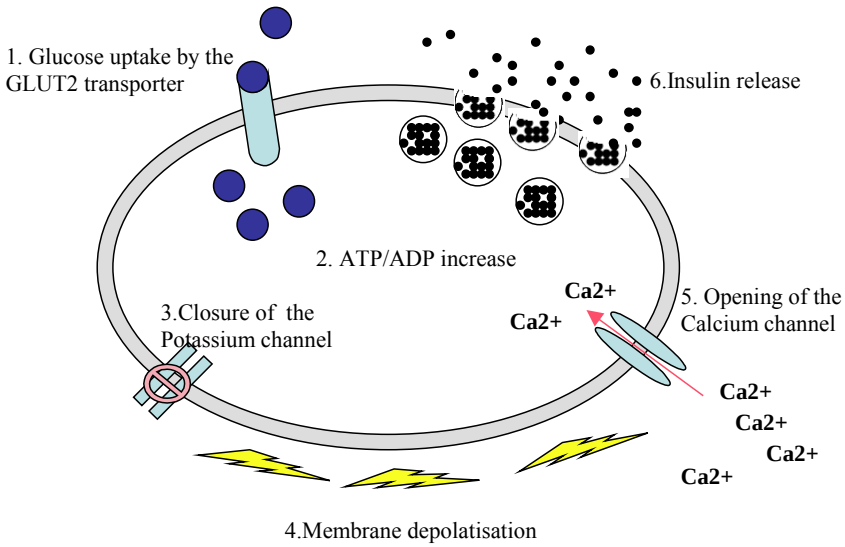
above, assuming the independent assortment of alleles at two loci, the expected halotype frequency is calculated as the product of the allele frequency of each of the two alleles, or $p_A \times p_B$. D is defined as $D=p_{AB}-p_A \times p_B$. Since *D* is dependent on the allele frequencies, no *D* is observed if any locus has an allele frequency 0 or 1. D′ is one of the most common measures for LD that attempts to avoid the allele frequency dependence by division of D with the maximum for the allele frequencies, which was first suggested by Lewontin (1964). D′=D/Dmax when D>0. D′=D/Dmin when D<0. D_{max} is given by the smaller of pApb and papB. D_{min} is given by the larger of -pApB and -papb. Because the sign is arbitrary, |D′| is often used. The case of D'=1 is known as complete LD. Values of D'<1 indicate that the complete ancestral LD has been disrupted. Another commonly used LD measure is called the correlation coefficient denoted by r^2 and is in some ways complementary to D' since it is not adjusted to the loci having different allele frequencies. The correlation coefficient is defined as $r^2=D^2/(pA \times Pa \times pBb)$. The values of r^2 also ranges from 0 (no disequilibrium) to 1 ('complete'disequilibrium).

1.2 DIABETES

Diabetes has been known since the antique time. The first known description of diabetes is from Egypt in 1500 BC. It was described as a rare disease that causes the patient to lose weight rapidly and urinate frequently. About 2000 years ago an ancient Greek doctor Aretaios first named the disease diabetes (passing through), as he described the body of a diabetic patient as a water-pipe – "the liquids do not stay in the body but just pass through". The word "mellitus" means honey sweet. In 1869, Paul Langerhans described islets in the pancreas that later were shown to be involved in the metabolism. In the experiments when the pancreas had been removed in dogs, the dogs developed diabetes. Canadians Frederick Banting and Charles Best achieved success to extract a substrate from the pancreas (insulin) that when injected into diabetic dogs, lowered blood sugar and prevented death. In 1922, Banting and Best tried their extract on Leonard Thompson, a 14-year old boy dying of diabetes, and saved his life. The insulin era had begun (10; 11). The division of diabetes into two categories "insulin sensitive" (today's Type I) and "insulin insensitive" (the modern Type 2) by Roger Himsworth was performed in 1935. Presently, diabetes mellitus, often referred to as diabetes, is recognized as a complex heterogeneous disorder characterized by hyperglycaemia.

The glucose homeostasis depends on the balance between glucose production by the liver and glucose uptake in the periphery (brain, muscles, liver, adipose tissue etc). In healthy subjects, the blood glucose levels are tightly controlled. Insulin is a hormone that lowers blood glucose levels. The islets of Langerhans produce hormones that are secreted into the blood. The islets consist of at least four types of hormone producing cells. Most common are ß-cells that produce insulin, and α-cells release glucagon, while delta cells produce somatostatin and PP cells secrete pancreatic polypeptide. The regulation of glucose - dependent insulin secretion in pancreatic ß-cells is linked to the expression and function of the ATP-sensitive potassium channel (K_{ATP}). Upon glucose metabolism, K_{ATP} cannels are closed in response to an increase in the ATP/ADP ratio, resulting in membrane depolarisation, Ca^{2+} influx through voltage dependent L-type Ca^{2+} channels, and subsequent insulin secretion (Figure 2).

Figure 2. Insulin release from a ß-cell



The classification of diabetes is based on a World Health Organization WHO-report with principles of diabetes definition (12). Two most important subtypes of diabetes are type 1 (T1D) and T2D. There are also other subtypes including maturity onset diabetes of the young (MODY) and latent autoimmune diabetes of the adults (LADA), gestational diabetes etc.

T1D develops on the basis of autoimmune destruction of pancreatic ß cells, which results in insulin deficiency. It mostly affects young people (<20 yrs), but is also occurs in adults. Several studies have indicated that auto-antibodies against specific islet cell antigens (ICA, GAD, IA-2 and insulin) are present in the majority of patients at the onset of T1D. A patient with T1D must rely on insulin treatment throughout life. Around 10% of all diabetic patients suffer from T1D.

T2D is the most common form of diabetes and it is increasing rapidly over the world. It accounts for approximately 90% of all diabetic patients. In T2D, hyperglycaemia results from a combination of impaired insulin secretion and insulin resistance (13). Insulin resistance decreases the ability of the body to respond to insulin. When the ß cells lose the ability to compensate for insulin resistance in skeletal muscle, liver and adipose tissue, hyperglycaemia becomes manifest. T2D develops slowly through stages of early disturbances of glucose metabolism, characterizing the prediabetic condition (14). The intermediate stages comprise impaired fasting glucose (IFG), impaired glucose tolerance (IGT) and a combination of both, i.e. combined glucose intolerance IFG+IGT. These are diagnosed on the basis of exaggerated increase in plasma glucose concentration after a standardized oral glucose tolerance test (OGTT), which was performed according to WHO. The onset of T2D unlike T1D mostly appears in elderly and middle-aged people, but is now also rapidly increasing in young people.

The prevalence of T2D is increasing in epidemic pattern in many countries and by 2010 the prevalence world-wide is expected to be 220 million and by the year 2025 will exceed 300 million (15). Another report including all age groups worldwide has estimated the

prevalence to be 4.4% in 2030 (16). In Sweden, the prevalence of T2D is about 3-4%, which corresponds to about 300,000 people (17). The overall incidence of diabetes has been rather stable in Sweden, even if the prevalence has tended to increase. Many risk factors have been identified which influence the prevalence or incidence of T2D. Factors of particular importance are a family history of diabetes, age, overweight, increased abdominal fat, hypertension, lack of physical exercise, and ethnic background.

It is estimated that globally diabetes is likely to be the fifth cause of death (18). Diabetes accounts for about 9-15% of total costs for healthcare system in the United States and other developed countries. This is mainly due to macrovascular and microvascular complications of diabetes. The chronic complications include injury at various organs e.g. retinopathy, neuropathy and nephropathy. Diabetes also increases the risk for cardiovascular disease (CVD) (myocardial infarction and stroke). About 75% of all diabetic patients die of CVD (19), and about, 40% of patients hospitalized for myocardial infarction have manifest T2D and 30% have IGT (20).

T2D is a serious, genetically influenced disease. The current available treatment does not cure the disease only milder the symptom. It includes diet and exercise, glucose lowering agents and insulin. Therefore, it is important to identify the susceptibility genes for the disease.

1.2.1 Genetics of T2D

It is believed that polygenic T2D results from inheritance of a set of susceptibility genes and that each exerts only a partial effect on the development of the disease. Only when the effect of these genes is added together in particular combinations and in the presence of certain risk factors, such as obesity, the disease is manifested. The predisposition to the disease could be determined by many different combinations of genetic variants (genotypes) and environmental factors. There is evidence from twin studies indicating that genetic determinants contribute to the development of T2D (21). The higher concordance rates are found in monozygotic (MZ) twins than in dizygotic (DZ) twins (22). In a population based cohort of twins in Finland, the concordance rate in MZ twins was 34% whereas in DZ twins it was 16% (23). In a Japanese study, these figures were 83% for MZ twins and 40% for DZ twins (24). The large variation in concordance rates between populations may be the result of bias or a different selection of the populations studied. It may also indicate differences in genetic susceptibility between these populations (25; 26).

In search for a better understanding of the pathogenesis of T2D, a genetic approach will help focusing on the underlying causes of the disease, and may provide new information for diagnostics, treatment and prevention. The progress in identification of specific genetic variants predisposing to T2D has been previously limited (27) but recently speeded up by genetic analyses including GWA (28-34)

The candidate gene studies are based on the selection of genes with the known or inferred biological functions which may predispose to disease or the observed phenotype. Studies examining specific candidates are mostly of the case-control association design. This is achieved by comparing a random sample of unrelated T2D patients with a matched

control group. So far, many candidate genes have been studied for their role in T2D (26). There are two major successes in genetic association studies in T2D. The common non-synonymous coding variant P12A in the proxisome proliferator-activated receptor γ (PPARγ) (35) and the variant E23K and potassium inwardly-rectifying channel, subfamily J, member 11 (KCNJ11) (36; 37) genes, respectively, have shown consistent evidence for association with T2D.

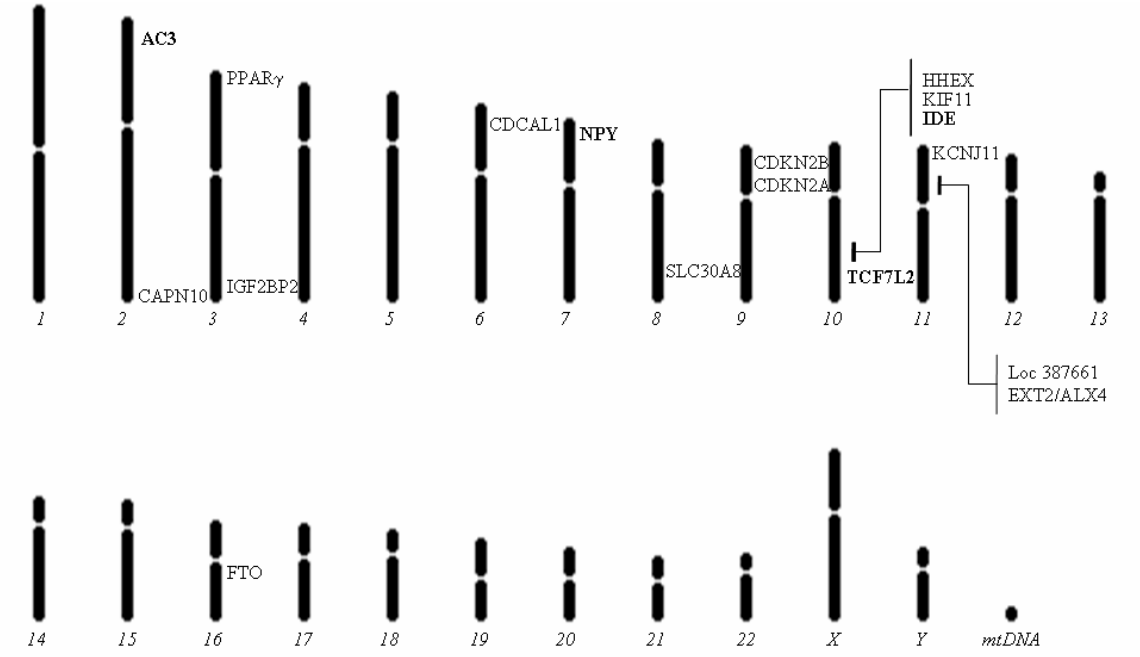
GWS and linkage analysis have been used for identification of chromosomal regions where the susceptibility loci for T2D reside. There are several chromosomal regions linked to T2D, for instance at 1q, 2q, 8p, 10q, 12q and 20q (38). Further investigation with positional cloning has indicated that the calpain 10 (CAPN10) genetic polymorphisms are associated with T2D (39). The CAPN10 gene is located on chromosome 2q37. In 2006, Grant et al. have found that the transcription factor 7-like 2 (TCF7L2) gene, which is located on chromosome 10q25 (40), may explain the linkage with T2D.

Until recently, linkage analysis has provided the only realistic means by which to undertake a comprehensive genetic survey of the entire genome. The availability of array-based reagents that allow massively parallel genotyping, combined with the development of SNP map and HapMap, provide the possibility of GWA study in T2D and other complex diseases. Companies such as Affymetrix and Illumina have developed chips that can capture information from more than two-thirds of the common variation in the human genome. Approximately 300,000–500,000 SNPs can be analysed using these chips.

The first GWA study in T2D has been published in Nature by Sladek et al. in 2007. This study in French population included a total of 1363 cases and controls, and demonstrated a strong associations with the TCF7L2 gene and the other four novel loci, including EXT2-ALX4 and LOC38776 on chromosome 11, HHEX-KIF11-IDE on chromosome 10, and SLC30A8 on chromosome 8 in T2D (32). Several GWA studies in different European populations have identified and/or confirmed additional susceptibility genes in T2D, including Insulin-like growth factor 2 mRNA binding protein 2 (IGF2BP2) on chromosome 3, CDK5 regulatory subunit associated protein 1-like 1 (CDKAL1) on chromosome 6, Cyclin-dependent kinase inhibitor 2A and 2B (CDKN2A/CDKN2B) on chromosome 9 (29-31; 33; 34).

The fat mass and obesity associated (FTO) gene, exerts its effect on T2D risk through an effect on adiposity (41), was found to be associated with obesity as well (42). In summary, we are all witnesses to a period of astonishing progress in identification of the susceptibility genes in T2D (43). With GWA approach, several susceptibility genes in T2D have been identified (Figure 3). Four candidate genes studied by using candidate gene approach in this thesis are also indicated in this figure but with bold letters. This type of genetics analyses provides novel insights into the pathogenesis of complex diseases.

Figure 3. Localization of the susceptibility genes predicted by GWA, linkage analysis and candidate gene approaches including genes covered by this thesis (bold)



1.2.2 Power of Genetic Association Study

Power to detect an association is dependent on several factors: the frequency of the predisposing allele, genotype, or haplotype; the accepted false-positive or Type 1 error rate (α); and the odds ratio (OR) or effect size of samples. Rarer alleles, genotypes, or haplotypes with small effects require larger sample sizes to attain the same power to detect an association, as compared to more frequent alleles or alleles with larger effects. Genetic association studies in large case-control populations may ultimately have the greatest power to detect alleles of small but significant effects on the susceptibility to common diseases such as T2D.

1.2.3 Diabetes and Environmental Factors

In addition to the genes, the environment is also responsible for the epidemic increase in the incidence of T2D. The environmental factors that may have a role in the development of the disease are obesity, physical inactivity, diet, toxic agents, viral infection, stress, and smoking (tobacco use). Physical activity increases insulin sensitivity and improves glucose tolerance. Obesity is implicated as a risk factor for T2D. The extent of intra-abdominal rather than subcutaneous fat is important in the development of T2D. The introduction of energy rich food contributes to the development of both obesity and T2D. The stress that is associated with today's life style may also be associated with glucose intolerance, hence increased risk of diabetes. The ethnicity is also an important factor since the prevalence is much higher in certain countries. It is amply documented that the development of T2D may be postponed or prevented by influencing lifestyle and other environmental factors (44).

1.2.4 Diabetes and Obesity

T2D is often associated with obesity. Thus, obesity is a major risk factor for T2D and cardiovascular disease (45). The prevalence of obesity diagnosed as BMI>30 has increased from 12% 1991 to 18% 1998 and was 20 % in 2000 (46). T2D and obesity are recognized as conditions of growing biomedical importance to societies worldwide. Overweight means increased weight and is a milder form of obesity. Obesity is commonly diagnosed by the body mass index (BMI) but other measures such as waist circumference, waist hip ratio (WHR) and body fat composition are also used. The current definitions of BMI commonly used are agreed in 1997 and published in 2000 (Table 2). BMI is defined by $Weight/height^2$ and expressed either in metric (kg/m^2) or US customary $lb * 703/in^2$. Waist measures (>102 cm in men and >88 cm in women) have become more important when considering different types of obesity, such as pear and apple types.

Table 2. Obesity classifications according to BMI

Classification	Under-weight	Normal-weight	Over-weight	Obese	Severely obese
BMI cut-of	<18,5	18.5-24.9	25.0-29.9	30.0-39.9	>40

Like T2D, obesity is partly determined by genetic factors, but an obesity-promoting environment is important for its phenotypic expression. The genetic factors predisposing obesity are poorly understood. On a population level, the thrifty gene hypothesis postulates that certain ethnic groups may be more prone to obesity than others. The ability to take advantage of rare periods of abundance to store energy efficiently may have been an evolutionary advantage in times when food was scarce (47). Individuals with greater adipose reserves were more likely to survive. This tendency to store fat is maladaptive in a society with stable food supplies.

2 THE PRESENT STUDY

2.1 AIMS

The overall aim of this thesis was to identify susceptibility genes in T2D.

The specific aims in each study included in this thesis are:

Paper I - IDE study
To investigate whether SNPs in IDE gene have a genetic influence on insulin levels in Swedish NGT and IGT subjects.

Paper II - TCF7L2 study
To confirm the genetic association between TCF7L2 polymorphisms and T2D in Swedish men and in addition explore the correlation between genotypes and phenotypes.

Paper III - NPY study
To evaluate whether the NPY Leu7Pro polymorphism contributes to the development of T2D in Swedish Caucasians.

Paper IV - AC3 study
To investigate the susceptibility of AC3 genetic variations in T2D in Swedish men, and to further analyze the association between the AC3 genetic polymorphisms and obesity.

2.2 SELECTION OF CANDIDATE GENES

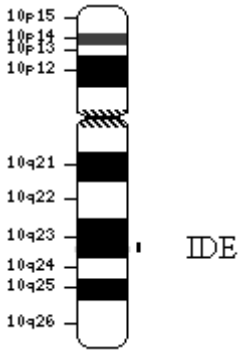
The candidate genes were selected by using two approaches:
First, from chromosomal regions linked to T2D. In recent years, several GWS reports have revealed a region in chromosome 10q, where the susceptibility genes for T2D may reside. From this chromosomal region, the candidate genes, including the insulin degrading enzyme (IDE) and transcription factor 7-like 2 (TCF7L2) genes were selected for our studies.

Second, based on biological functions and/or involvement in specific pathways related to T2D. Our knowledge of pathophysiology of T2D has increased, and we can select candidate genes from specific pathways related to T2D. By using this approach, the neuropeptide Y (NPY) and adenylate cyclase 3 (AC3) genes were selected.

2.2.1 Insulin Degrading Enzyme (IDE)

IDE is also called insulysin and it is a zink metallo endopeptidase. The human IDE protein consists of two domains, IDE-N and IDE-C, which are approximately equal in size. IDE-N contains the catalytic domain and IDE-C facilitates the substrate recognition and plays a role in oligomerization (48). The active site of IDE consists of the HEXXE (His-Glu-AA-AA-His) AAs (48; 49). Six human IDE transcrips are identified (50).

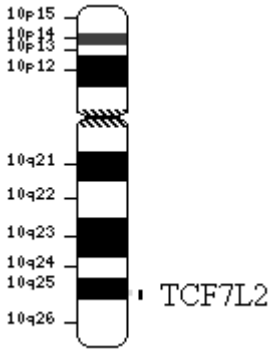
IDE is the major enzyme responsible for insulin proteolysis and shares structural and functional homology with bacterial protease III, which may function in the termination of the insulin response. IDE plays a principal role in the proteolysis of several peptides in addition to insulin, including amyloid ß, amylin glucagon, transforming growth factor α, ß-endorphin and atrial natriuretic protein. Kuo et al. have reported that over-expression of IDE in cell culture increases the rate of insulin degradation (51). A mouse model with homozygous deletion of the IDE gene (IDE-/-) shows impaired glucose tolerance and hyperinsulinemia (52). Studies on a transgenic mouse co-expressing human IDE show impaired glucose tolerance and lower serum insulin levels compared to wild type mouse (53). The transfer of an approximately 3.7 cM chromosomal region containing the IDE gene in Goto-Kakizaki (GK) rat to a normoglycemic rat recapitulated several features of the diabetic phenotype, including hyperinsulinemia and postprandial hyperglycemia. The IDE gene in GK rat was found to bear two AA change mutations (H18R and A890V). When they were transfected into COS-1 cells, it resulted in 31% less insulin degradation compared with cells transfected with wild type allele (54). IDE expression has been shown to be affected by aging, and IDE activity decreases significantly in liver and muscle of old animals compared to young (55).



The gene encoding IDE is located on chromosome 10q23-q24. This gene consists of 24 exons. The enzyme is highly conserved between species. Groves et al. initially carried out a variation screening of the IDE gene and association analysis in T2D patients among a British population. Although a borderline significance of association between the IDE genetic polymorphisms and T2D was found, no compelling association evidence was concluded (56). Moreover, Karamohamed et al. performed a haplotype analysis and found an

association between IDE genetic polymorphisms and levels of fasting plasma glucose, HbA1c and T2D in American populations of European descent (57). In order to explore the association between IDE genetic variation and T2D and to understand whether polymorphisms in this gene have a measurable influence upon insulin levels, we have carried out a genetic association study of the IDE gene in a Swedish population.

2.2.2 Transcription Factor 7-Like 2 (TCF7L2)

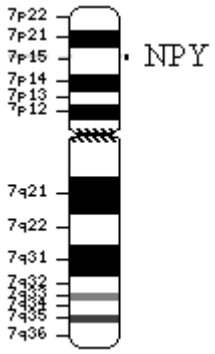


The TCF7L2 gene is located on chromosome 10q25, the same region as the gene coding for IDE. Both genes reside in the chromosomal region linked to T2D. The mRNA sequence (NM_030756) of the TCF7L2 gene spans 2439 bp. TCF7L2 is a high mobility box-containing transcription factor, expressed in many human tissues, including heart, placenta, lung, brain, liver, kidney, pancreas, adipocytes and omental adipose tissue (58). TCF7L2 has moved rapidly from a novel positional candidate gene to a reference gene for T2D susceptibility. Grant et al. have first reported that a microsatellite marker, DG10S478, within intron 3 of the TCF7L2 gene was strongly associated with T2D in Icelandic ($p=2.1*10^{-9}$), Danish ($p=4.8*10^{-3}$) and US ($P=3.3*10^{-9}$) cohorts. In the same study, five SNPs around the microsatellite marker were associated with T2D in the three studied cohorts (40). Since then, a number of genetic studies have provided evidence for the association between TCF7L2 genetic polymorphisms and T2D in different populations. In most ethnic groups (59; 60), except for Eastern Asians, the meta-analyses illustrate that the magnitude of the TCF7L2 effect is much higher than any other confirmed T2D candidate genes. The TCF7L2 gene expression was significantly increased in pancreatic islets from T2D patients with the CT/TT genotypes of SNP rs7903146. Furthermore, the incidence of hyperglycemia among carriers of the T allele of rs7903146 was increased in a French population (61). Over a three year period, subjects carrying two risk alleles of rs7901346 and rs12255372 were more likely to progress from IGT to T2D compared to subject not carrying the risk allele (62). TCF7L2 gene expression level in pancreatic islets of T2D patients, carrying the risk alleles of TCF7L2 polymorphisms, is increased by about 5-fold compared to the control subjects. In SNP rs7903146, the carriers with TT genotype have the highest expression levels of TCF7L2 mRNA in pancreatic islets (63). In the subcutaneous and omental fat from T2D patients with obesity, TCF7L2 expression is significantly decreased compared with obese normoglycemic individuals (58)

TCF7L2 has also an essential role in the developmental and growth regulatory mechanisms of intestinal epithelial cells, which secrete the glucagon-like peptide-1 (GLP-1), and TCF7L2-deficient mice lack an intestinal epithelial stem cell compartment. GLP-1 exerts a critical effect on blood glucose homeostasis by stimulating early insulin production from the pancreatic ß-cells and by increasing insulin secretion. TCF7L2 expression in the adipose tissue of T2D patients is decreased, which indicates that TCF7L2 plays a role in the regulation of adipogenesis by altering transcriptional regulation of the genes encoding CCAAT/enhancer-binding protein-α (CEBPA) and (PPARγ).

2.2.3 Neuropeptide Y (NPY)

NPY is a neuropeptide detected in the mammalian brain and is found throughout the central and peripheral nervous systems. The protein consists of 36 AAs. The NPY gene is located in chromosome 7p15.1. NPY has multiple functions but mainly plays a role in the regulation of satiety, ingestive behaviors, energy balance and expenditure. NPY also stimulates lipoprotein lipase activity in the adipose tissue. (64). NPY is also expressed in pancreatic islets and is implicated in the islet function NPY decreases glucose-stimulated insulin secretion from the islets (65).



Pancreatic islets from NPY-deficient mice have higher basal insulin secretion, glucose-stimulated insulin secretion and islet mass in comparison with wild-type mouse. The expression of NPY mRNA levels was decreased by 70% in the islets from mice with high-fat diet, compared with controls. Moreover, non-obese pre-menopausal women had significantly higher NPY serum levels than obese pre-menopausal and obese post-menopausal women (66). NPY has been also shown to be associated with insulin resistance.

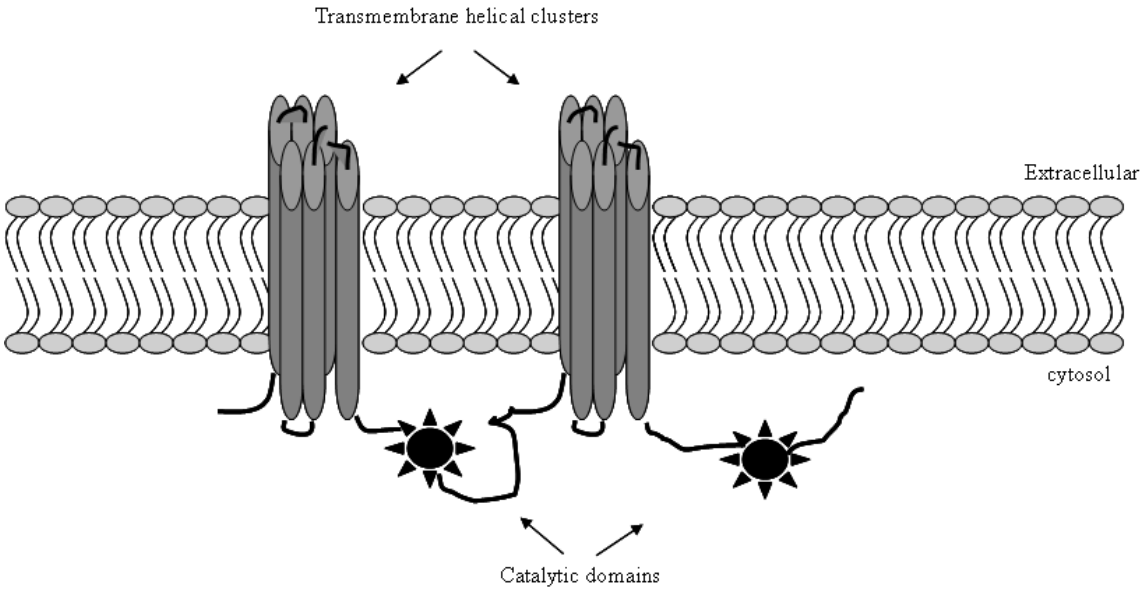
Leu7Pro (T1128C) is a non-synonymous SNP in exon 2 of the NPY gene (67). Several studies have demonstrated that Leu7Pro polymorphism is associated with increased levels of total cholesterol, LDL cholesterol and triglycerides in blood, accelerated development of atherosclerosis and alcohol dependence (68-71). Two genetic association studies have demonstrated that Leu7Pro polymorphism was linked to enhanced carotid atherosclerosis and retinopathy in patients with T2D (72; 73). Furthermore, Leu7Pro polymorphism was found to associate with increased BMI (74) and with an increased risk for T2D in middle aged subjects (75).

2.2.4 Adenylyl Cyclase 3 (AC3)

Mammalian adenylyl cyclases (ACs) are a family of diverse group of variously regulated signaling molecules. At least nine AC isoforms (AC 1-9) have been identified in mammals. AC isoforms can be classified into different families according to sequence homology and regulatory properties. One classification is based on the response of different AC isozymes to Ca^{2+} *in vitro*. The Ca^{2+}- stimulated ACs 1, 3 and 8, the Ca^{2+}-inhibited ACs 5 and 6 and the Ca^{2+}- unresponsive ACs 2, 4, 7 and 9. All family members are large polypeptides (1080–1248 amino acids). ACs is enzymes that catalyse the conversion of ATP to cAMP. The enzyme integrates signals that act through G protein-coupled cell-surface receptors with other extra-cellular stimuli to finely regulate intra-cellular levels of cAMP. The cAMP potentates glucose-stimulated insulin secretion through protein kinase A (PKA) activation. Mechanisms mediating cAMP action in cells include Ca^{2+} mobilization and Ca^{2+} influx.

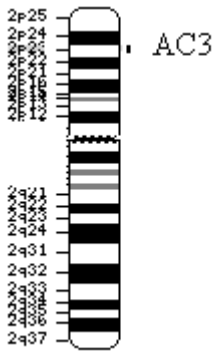
The ACs share the same structure conformation. They are trans-membrane-spanning protein helices. ACs consist of five domains, a cytoplasmic N-terminal region, a membrane anchoring hydrophobic domain (M1), a large cytoplasmic domain (C1), a second transmembrane helical cluster (M2) and a second cytoplasmic domain (C2) (76).

Figure 4. The structure of the mammalian AC isoforms



All nine isoforms contain at least one site predicted to undergo N-linked glycosylation in M2. M1 and M2 comprise cassettes with six transmembrane spanning domains. The transmembrane domains are not highly conserved among adenylyl cyclases. However, sections of two of the cytoplasmic domains (termed C1a and C2a) are highly conserved, when expressed separately they can combine to display basic catalytic activity. On the other hand, the N terminus C1b and C2b regions are poorly conserved and are the regions where type-specific regulatory features are speculated to reside.

ACs can couple with both stimulatory and inhibitory G-proteins. Interaction with Gs stimulates their activity and interaction with Gi inhibits its enzymatic activity. The G protein consists of three subunits, α, ß and γ. Ligand binding to the receptor changes the receptor conformation, allowing it to associate with a G-protein. This results in the activation of the specific G-protein via exchange of GTP for GDP bound to the α subunit of the G-protein. The G protein α-subunits bind GTP and adopt an active conformation of ACs. Subsequently, ACs are modulated until signalling terminated by the action of an intrinsic GTPase activity and reassociation with the G ß-γ complex.



AC3 is one of ACs, and has been named differently, such as adenylate cyclase 3 and adenylyl cyclase 3. The gene is located on chromosome 2p23.3. Previous studies have indicated that glucose-induced insulin release is markedly decreased in the GK rat pancreas. It has been shown that this defect is reversed by forskolin, which enhances cAMP generation in GK islets. These effects of forskolin were associated with an over-expression of AC3 mRNA in the ß-cells due to the presence of two functional point mutations in the promoter region of the AC 3 gene in GK rat(77). Using antibodies against ACs 1-8, the localisation of these AC isoforms in different endocrine cells types in both normal and diabetic GK rat pancreas demonstrated a clear immuno-reaction (IR) to AC1-4 and 6 in normal and GK islet ß-cells, while a smaller number of ACs were expressed in α- and

delta-cells. No AC-IR was observed in pancreatic polypeptide cells. Moreover, IR of Ca^{2+} stimulated AC l, AC 3 and AC 8 in diabetic ß- and α-cells was increased, compared with the corresponding IR in control pancreas (78). Additionally, liver adenylyl cyclase activity was increased in the membranes of male ob/ob mice in comparison to the lean control mice (79). These findings suggest a role for the AC3 gene in the pathogenesis of T2D and obesity. As for T2D patients and obesity subjects, however, there is no reported study of genetic association with the AC3 gene. In the present study, we investigated the association of AC3 genetic variation with T2D in Swedish men. We further analyze whether AC3 genetic variation is associated with obesity with NGT.

3 SUBJECTS

The subjects included in the present study are divided into several groups i.e. NGT, IGT, T2D and obese with NGT. All subjects were diagnosed according to the World Health Organization criteria (WHO) in 1985 (80) or 1998 (12), which are described in Tables 3a and b. In 1998, WHO defined new venous plasma glucose cut-off values for diagnosis of diabetes, and two more groups for pre-diabetes were included, i.e. impaired fasting glucose (IFG), and a combined group with IGT+IFG. The cut-off values measured by Oral Glucose Tolerance Test (OGTT) are given in the table below.

Table 3a. WHO (1985) venous plasma glucose cut-off values for diagnosis of T2D

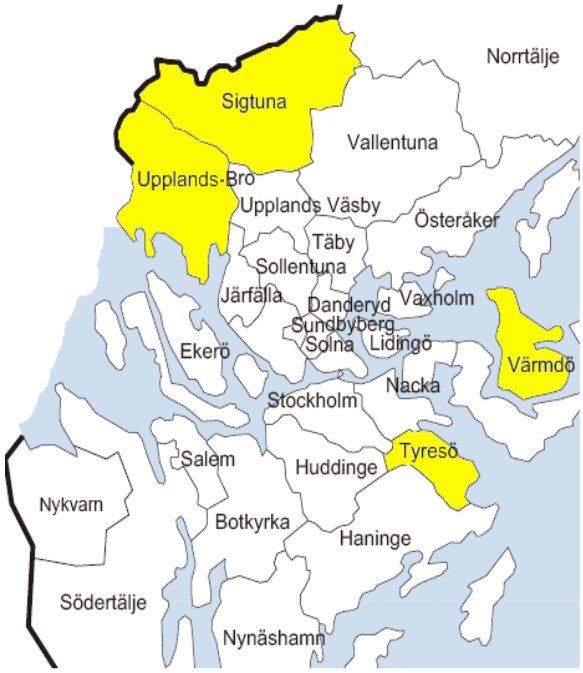
Glucose tolerance	OGTT 0h		OGTT 2h
NGT	<7.8		<7.8
IGT	<7.8		7.8≤ glucose <11.1
T2D	≥7.8	and/or	≥11.1

Table 3b. WHO (1998) venous plasma glucose cut-off values for diagnosis of T2D

Glucose tolerance	OGTT 0h		OGTT 2h
NGT	<6.1		<7.8
IFG	6.1≤ glucose <7.0		<7.8
IGT	<6.1		7.8≤ glucose <11.1
IGT+IFG	6.1≤ glucose <7.0		7.8≤ glucose <11.1
T2D	≥7.0	and/or	≥11.1

3.1 STOCKHOLM DIABETES PREVENTION PROGRAM (SDPP)

The subjects used in the studies are mainly recruited from the Stockholm diabetes prevention program (SDPP) (81-84). SDPP is a cohort study and comprises three stages: an initial baseline study in four municipalities: Värmdö, Upplands-Bro, Tyresö and Sigtuna situated in Stockholm suburbs, a 10 year follow-up study of the initial cohort and also a population based intervention program. The program comprises both men and women. The study in men was conducted 1992-1994 in totally 3128 participants. They were 35-55 years of age at the time of the study. The study in women was performed two years later but included one more municipality i.e. Upplands Väsby. The participants were selected in two steps. A short questionnaire concerning health and family history of diabetes (FHD), was sent by post to all men (nearly 13000 subjects). They were identified by the population registry with



the county council. The inclusion criteria were that each individual had to live in one of the four municipalities and having the appropriate age. FHD was defined as having at least one first-degree relative (parent, brother or sister) or at least two second-degree relatives (mother´s or father´s parents or sisters/brothers). Based on the results of this questionnaire, with a response rate of 79%, subjects with already known diabetes (2.5%), insufficient FHD, foreign origin as well as those giving incomplete response to the questionnaire were excluded from further studies. From the remaining material, the subjects with FHD (n=2106) and a group without known FHD was selected to age match with the FHD group (n=2424) and invited to a health examination. 70% of the subjects (n=3162) agreed to participate. The individuals were characterized by an OGTT, measurements of blood pressure, body weight, height and waist-hip ratio. They responded to an extensive questionnaire regarding lifestyle (food, exercise, tobacco-and alcohol habits, education, psychosocial and socio-economical factors etc). An additional 1% of the individuals were excluded due to insufficient FHD.

Finally, SDPP baseline study included 3128 male participants. The follow-up study was conducted ten years later (2002-2004). 87% (n=2383) subjects from 2746 invited (i.e. the subjects still living in the same area and that was not diagnosed with T2D at the baseline study) participated in the follow-up study. Development of T2D was assessed by an OGTT at baseline or at follow-up occasions or was self-reported by the patients diagnosed during the time period between baseline and follow-up (n=84). The OGTT demonstrated previously undiagnosed diabetes in 60 and 99 men at baseline and follow-up study respectively (WHO 1998). T2D patients had no medication when the data was collected. The same selection procedure was performed for women during 1996-1998 and resulted in a total of 4821 subjects in the baseline study. Of them, 3329 were participated in the follow-up study 8-10 years later. Blood samples were collected from both the baseline and follow-up studies. All subjects were Swedes. Genomic DNA was extracted from peripheral blood. Informed consent was received from all subjects. The study was approved by the local ethic committee.

3.2 KRONAN STUDY

An additional group of T2D patients selected from Sundbyberg, a municipality in the Stockholm region, was included in the NPY gene study (III). The subjects were born 1927-1957 and they were diagnosed with diabetes after 35 years of age. Patients with diabetes were acquired from three health care centers, Kronan, Hallonbergen and Rissne, within the municipality of Sundbyberg. 178 patients were included in the study and LADA patients were excluded. These T2D patients had anti-diabetic treatment, 24% were treated with diet alone, 46% with oral hypoglycaemic agents (OHA), 22% with insulin and 8% with a combination of insulin and OHA. The study was approved by the local ethic committee.

4 METHODS

4.1 DNA EXTRACTION

DNA extraction was performed from whole blood samples by using a Genomic DNA Purification Kit (Gentra). The kit relies on biological or environmental specimens as a source of genomic, mitochondrial or viral DNAs. The cells are lyzed to facilitate the separation from the white blood cells with an anionic detergent in the presence of a DNA stabiliser. Contaminating RNA is then removed by treatment with an RNA digesting enzyme. Genomic DNA is recovered by precipitation with alcohol and dissolved in a buffered solution containing DNA stabiliser.

4.2 DIRECT SEQUENCING

The sequencing analysis approach is based on the Sanger sequencing principle. With this dye terminator chemistry, each dideoxy nucleotide is labelled with a specific dye so that all four reactions can be performed in the same tube and run in one lane on the gel. The fluorescent-labelled sequencing products are detected using a laser beam. The laser beam stimulates fluorescence from each fragment with energy according to the terminator base added at the final position. The sequencing analysis protocol used in the present study is one line sequencing with four fluorescent dyes labelled ddNTPs, polymerase and buffer. The direct sequencing analysis using Big Dye terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystem, ABI, model 377 genetic analyzer, Perkin-Elmer, Foster City, USA) was performed. A computer program called Sequencher, which has the ability to compare several sequences with each other, is used for the analysis of sequencing data.

4.3 GENOTYPING METHODS

There are several genotyping techniques, including PCR-restriction fragment length polymorphism (RFLP), Pyrosequencing, Dynamic allele specific hybridisation (DASH), Mini-sequencing, TaqMan allelic discrimination, single-base extension (SBE), Oligo nucleotide ligation assay (OLA) and Direct sequencing etc. (85; 86). In this thesis, three high throughput SNP scoring methods, including DASH, pyrosequencing and Taqman allelic discrimination, were used. DASH and TaqMan allelic discrimination were mainly used for genotyping experiments, while Pyrosequencing was used for confirmation experiments in the AC3 study.

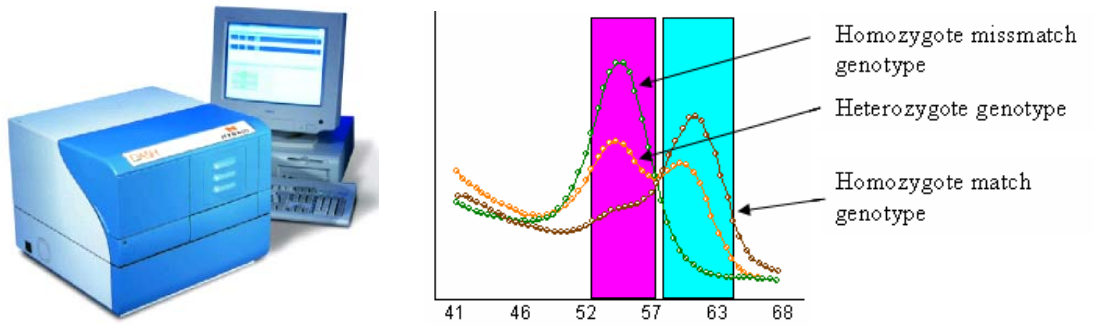
4.3.1 Dynamic Allele Specific Hybridisation (DASH)

DASH is a high throughput genotyping method, which is based on hybridization of an oligonucleotide probe to single stranded PCR product. It is used for scoring SNPs and detecting small insertions and deletions.

The procedure starts with assay design for primers used in PCR and probe. The amplicon is usually about 50 bp long and SNP of interest is located in or nearby the middle. One of the PCR primers is labeled with biotin. PCR product is then immobilized by transferring into a streptavidin-coated plate. The biotinylated primer will bind to Streptavidin on the well surface, whereas the non-biotinylated strand is removed by rinsing with a NaOH solution. The specific probe, complementary to one allele of SNP is added to the well

along with a hybridization buffer containing a fluorescent double-strand-specific dye, Sybr green. This dye will give a signal when it is bound to double strand DNA. The probe designed to match one allele, and thereby mis-match for other allele of interest will create a difference in the denaturating temperature during the detection with a DASH instrument. On the computer screen the loss of fluorescence is plotted as the negative derivative (slope of the fluorescence Vs temperature) the denaturation points are interpreted as peaks. The mismatch homozygous peak (pink) is observed at a relatively lower temperature and match homozygous peak at a higher temperature (blue). A heterozygous sample (containing both alleles) would undergo a two-phase denaturation and therefore produces two peaks in the negative first derivative (Figure 5) (87; 88).

Figure 5. DASH instrument and genotyping of SNP



The absolute Tm observed may vary depending on probe length and GC content, but the relative Tm difference between homozygous match and homozygous mismatch is normally 4-12°C. Probe with specific dye (Rox) can be used in order to improve genotyping peaks.

In a typical PCR-DASH assay design, there are two ~22 bp primers (one biotinylated) and one probe (~17 bp). To avoid the second structure of PCR probe for hybridization with the probe, it is recommended to use a folding analysis program named MFOLD (http://mfold2.wustl.edu/~mfold/dna/form1.cgi). The probe sequence is designed complementary to the biotinylated strand of PCR product.

4.3.2 Pyrosequencing

Pyrosequencing technology is based upon sequencing-by-synthesis, and uses an enzyme - based system to monitor DNA synthesis in real time. Essentially, the method allows sequencing of a single strand of DNA by synthesizing the complementary strand along it. It was developed by Mostafa Ronaghi and Pål Nyrén (89; 90).

The pyrosequencing procedure starts when a sequencing primer is hybridized to the single stranded DNA used as template for the sequencing, and incubated with the enzymes DNA polymerase, ATP sulfurylase, luciferase and apyrase, and with the substrates adenosine 5´ phosphosulfate (APS) and luciferin. The templates for pyrosequencing can be made both by solid phase template preparation (Streptavidin coated magnetic beads) and enzymatic template preparation (Apyrase + Exonuclease). The 5'-nuclease activity of DNA polymerase catalyzes the incorporation of deoxynucleotide into the DNA strand, if it is complementary to the base in the template strand. This incorporation releases

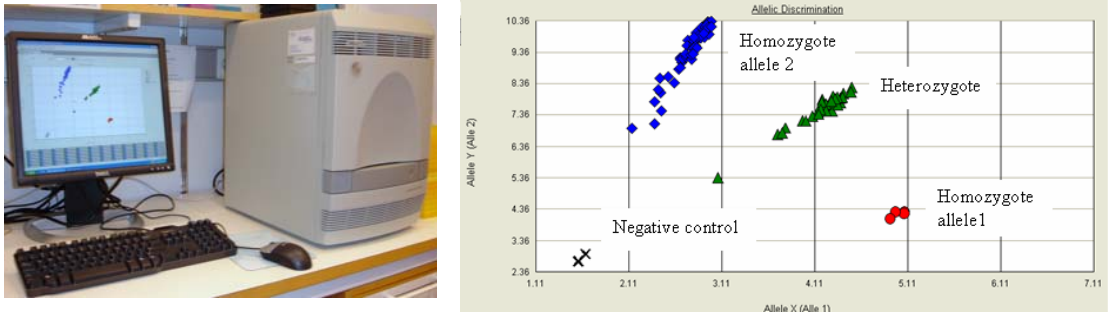
pyrophosphate (PPi) stoichiometrically. ATP sulfurylase quantitatively converts PPi to ATP in the presence of adenosine 5′- phosphosulfate. The produced ATP drives the luciferase-mediated conversion of luciferin to oxyluciferin, which generates visible light in amounts that are proportional to the amount of ATP. The light produced in the luciferase-catalyzed reaction is detected by a charge-coupled device (CCD) camera and is expressed as a peak in a pyrogram. Since the added nucleotide is known, the sequence of the template can be determined. The result can be analyzed in a program. Each light signal is proportional to the number of nucleotides incorporated. Apyrase, a nucleotide degrading enzyme, which continuously degrades ATP and unincorporated dNTPs and the reaction can restart with another nucleotide. The process continues, the complementary DNA strand is built up and the nucleotide sequence is determined from the signal peaks in the Pyrogram. This technique can be used for both sequencing and SNP genotyping experiments (91; 92).

4.3.3 TaqMan Allelic Discrimination

TaqMan technique has been used for quantification of mRNAs and also for SNP genotyping. It allows detection and measurement of products generated during each cycle of the PCR process. The technique is built on the 5'-exonuclease activity of the enzyme, Taq DNA polymerase, and it monitors degradation of fluorescently labeled probes. In this thesis, the method has been used for allelic discrimination of SNPs.

The procedure using the 5'-exonuclease activity of the enzyme Taq DNA polymerase is similar to conventional PCR, with the exception that a fluorescent probe is used and the result is detected in each cycle. In a TaqMan experiment, single stranded fluorogenic probe, complementary to the target sequence is added to the PCR reaction mixture. This probe is a dual labelled oligonucleotide with a reporter dye attached to the 5' end and a quencher dye attached to the 3' end. The probe is located between the two primers. Examples of reporter dyes are FAM, VIC and TET. The quencher dye is normally TAMRA. When the two fluorophours are attached to the probe proximity between them, only the length of the probe inhibits fluorescence from the fluorophore. This is called as fluorescent energy transfer (FRET). During PCR, the probe anneals specifically between the forward and reverse primer to an internal region of the PCR product. DNA polymerase then carries out the extension of the primer and replicates the template to which the primers and probes are bound. The 5'-exonuclease activity of the polymerase cleaves the probe, releasing the reporter molecule away from the close vicinity of the quencher. The fluorescence intensity of the reporter dye increases as a result. This process is repeated in every cycle and does not interfere with the accumulation of the PCR product. Hence, fluorescence detected in the real-time PCR thermal cycler is directly proportional to the fluorophore released and the amount of DNA template present in the PCR. To induce fluorescence during PCR, laser light is distributed to the sample wells via a multiplexed array of optical fibers. The resulting fluorescent emission returns via the fibers and is detected with a CCD camera (93; 94).

Figure 6. ABI 7300 instrument and TaqMan allelic discrimination



In the case, when this method is used for allelic discrimination, a minor groove binder (MGB) molecule is incorporated on the 3' end of the probes (95). The MGB binds to the minor groove of the DNA helix, improving hybridization by stabilizing the MGB-probe/template complex, thereby permitting the use of probes for improved mismatch discrimination and greater flexibility when designing assays. TaqMan probes can be designed to detect SNPs and small insertion/deletions (indels).

4.4 BIOINFORMATICS

Bioinformatics is a tool in which the computer is used to find out information from public databases, such as GenBank, Map Viewer, Blast Search, dbSNP, PubMed etc. The dbSNP is has served as a central, public repository for genetic variation, including SNPs, microsatellite repeats and small insertion/deletion polymorphisms. This database is established by the national center for biotechnology information (NCBI), USA (http://www.ncbi.nlm.nih.gov/sites/entrez?db=snp&cmd=search&term=). Several other SNP databases such as HGVbase, CGAP, GeneLynx etc. are also useful for searching information of gene sequences and genetic variations. The international HapMap project enables the study of LD in human populations online (http://www.hapmap.org). This has facilitated the selection of SNPs in genetic association studies.

Selection of the SNPs for study are based upon their locations (intronic, exonic or promoter), function and information from previous reports. All selected SNPs are blasted against the human genome to check for specificity of the sequences (http://www.ncbi.nlm.nih.gov/blast). The upstream and downstream sequences of the SNPs are examined by repeat masker because repeated sequences and duplicons may be deleterious for the genotype determination (http://repeatmasker.genome.washington.edu/cgi-bin/RepeatMasker). Tag SNPs are designated using an r^2 cut off ~0.8 and checked from the data in European Caucasians (CEU) population recorded in HapMap (release No. 22).

4.5 DATA ANALYSES

Data analyses are performed in both single and multiple marker (Haplotype) perspective. In single marker association analysis, comparison of allele and genotype frequencies between the cases and controls are conducted. If the difference of allele and/or genotype frequencies is significant, the analyses for association with phenotypes are followed. Further analyses for multiple marker association, including LD, haplotype and/or diplotypes, are performed. Programs used for these analyses in this thesis are Statistica,

version 7.1 (StatSoft, Tulsa, OK, USA), Statistical analysis system (SAS), version 8.2 or 9.1 (SAS Institute, Cary, NC, USA), StatView version 5.0 (Abacus Concepts, Piscataway, NJ, USA) and/or BMDP version1.1 (BMDP statistical software Inc, Los Angeles, CA, USA). P-values less than 0.05 are considered significant.

4.5.1 Single Marker Association

Using the χ^2-statistic, HWE in each SNP is assessed, and allele and genotype frequencies of the SNP are compared between the case and controls. Phenotype comparisons between different genotypes within the study groups are performed by appropriate statistical methods, such as analysis of variance (ANOVA) or co-variance analysis (adjusting for different traits where appropriate) or non-parametric methods. Traits important or interesting for T2D have been included in the quantitative trait analyses, and also as co-variates in several analyses. Non-normally distributed data are transformed by the natural logarithm before analyses to improve the normal distribution. Levene's test is performed in order to test for homogeneity of variances. Homeostasis model of assessment were used to assess insulin resistance (HOMA-IR) and β-cell function (HOMA-β). HOMA-IR and HOMA-β were calculated as [(fasting plasma glucose, mmol/l * fasting plasma insulin, mU/µl) /22,5] and [(fasting plasma insulin, mU/µl * 20) / (fasting plasma glucose, mmol/l -3,5)], respectively (96). Logistic regression analyses considering different modes of inheritance with and without inclusion of potential confounding factors are performed to study genotype distribution differences between the cases and controls.

4.5.2 Multiple Marker Association

Both r^2 and D' measurements of LD values are used. LD values and haplotype frequencies are estimated using either Haplotyper program (EH-plus) (ftp://linkage. rockefeller.edu/software/eh) or Arleqiun program version 2.0 (http://lgb.unige.ch/ arlequin/). Haplotype frequencies between the cases and controls are calculated using 2x2 contingency tables, and χ^2 test. ANOVA and/or co-variance analysis is used to test for differences in quantitative traits among diplotypes.

5 RESULTS

5.1 PAPER I – IDE STUDY

5.1.1 Single Marker Association

This study mainly addressed the fundamental question of whether the IDE genetic polymorphisms have measurable influence on insulin levels in Swedish population. A total of 403 (246 men/157 women) non-diabetic controls and 321 (165/156) IGT subjects were included in this study. All of them were Swedes and diagnosed according to the WHO 1985 criteria. Thirteen SNPs in the IDE gene and one SNP in the HHEX gene were genotyped. Genotype distribution of all studied SNPs were in accordance with HWE. In men but not in women, the common allele of SNP rs2251101 was associated with lower levels of fasting insulin, 2h insulin, HOMA-IS and BMI, while the rare allele of SNP rs2249960 was associated with lower BMI. The effect of SNP rs1887922 upon 2h insulin levels was however significant, albeit considerably attenuated in comparison with SNP rs2251101.

5.1.2 Multiple Marker Association

Three previously identified tag SNPs, i.e. rs2251101, rs832196 and rs1544210 were used for analyses of delineating common haplotypes in men. Six common haplotypes (>4.4%) were predicted using these markers, the sequences for which are shown in table 4.

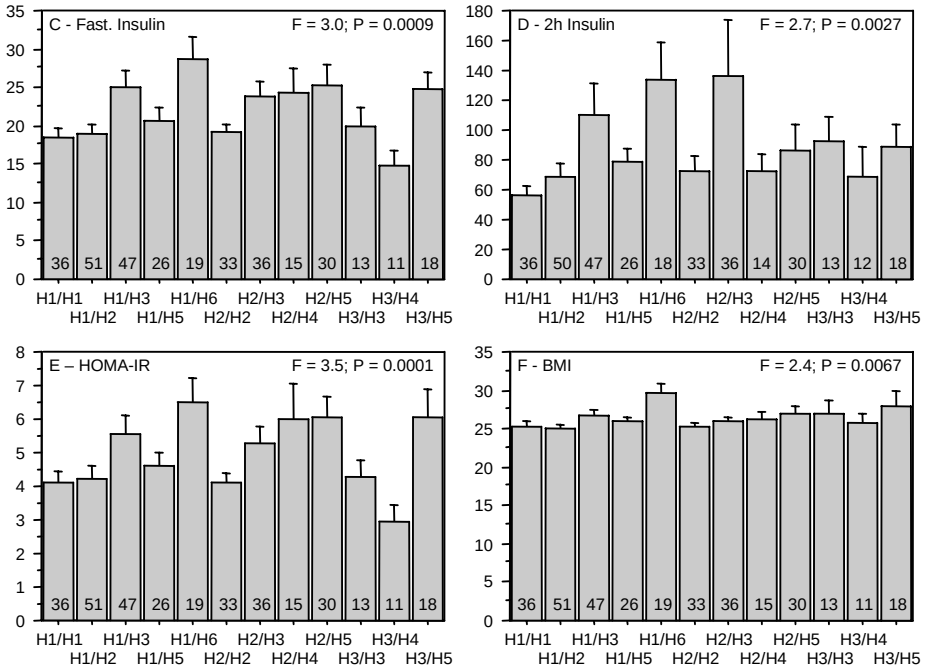
Table 4. Haplotypes inferring SNPs rs2251101, rs832196 and rs1544210

Haplotype	Frequency (%)	Sequence
H1	28.6	TCG
H2	20.5	TCA
H3	11.7	CCA
H4	6.2	TTA
H5	6.2	TTG
H6	4.4	CCG

Diplotypes were inferred haplotype-tagging markers and tested against the traits of fasting glucose, 2h glucose, fasting insulin, 2h insulin, HOMA-IR and BMI in men and women separately. For all analyses, IGT and NGT individuals were combined, and adjusted for phenotypic characteristics by including their group identity as a covariate in ANOVA models (Figure 7). The diplotype analyses including age and group as covariates revealed significant difference in the four traits between the diplotype groups in men.

Since throughout this study tests were conducted on the combined IGT and normal samples, it was considered important to confirm that an effect exists in both populations. Using the above model and focusing on the rs2251101 marker and the 2-h insulin trait, significant effects were observed for both the IGT ($p=0.0004$) and normal ($p=0.011$) groups.

Figure 7. Diplotype association with quantitative traits in men



5.2 PAPER II – TCF7L2 STUDY

5.2.1 Single Marker Association

In the present study, we have conducted an association study in Swedish men selected from SDPP. We also took advantage of the 10-year duration of the program and investigated the differences of quantitative traits with polymorphisms in the TCF7L2 gene. Five previously studied SNPs i.e. rs7901695, rs4506565, rs7903146, rs11196205 and rs12255372 were genotyped in 243 T2D patients and 528 NGT subjects selected from SDPP. All SNPs were kept in HWE. Four SNPs rs7901695, rs4506565, rs7903146 and rs12255372 were significantly associated with T2D in Swedish men. In SNP rs1196205, however, association with T2D was of borderline significance (Table 5).

Table 5. Allelic association and MAFof the TCF7L2 polymorphisms

SNP	NGT	T2D	p-value	OR (95% CI)
rs7901695	C: 0.235	C: 0.327	0.0001	1.585 (1.249 - 2.012)
rs4506565	T: 0.263	T: 0.329	0.0001	1.588 (1.251 - 2.015)
rs7903146	T: 0.217	T: 0.320	0.00001	1.700 (1.336 - 2.163)
rs11196205	C: 0.407	C: 0.456	0.0730	1.220 (0.981 - 1.517)
rs12255372	T: 0.223	T: 0.292	0.0036	1.437 (1.125 - 1.835)

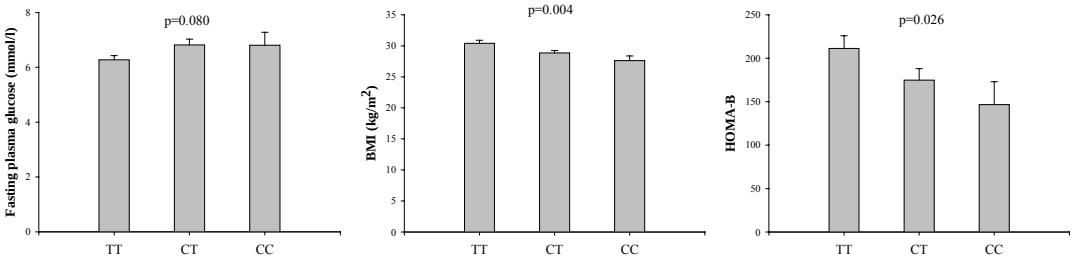
In multivariate logistic regression analysis applying the additive inheritance model and considering age, BMI and blood pressure as covariates, the association with T2D remained significant ($p<0.0008$) for these four SNPs rs7901695, rs4506565, rs7903146 and rs12255372.

Genotype-phenotype analyses demonstrated that HOMA-β index, BMI and 2h plasma insulin in T2D patients carrying the homozygous genotype of risk allele in SNPs rs7901695, rs4506565 and rs7903146 were significantly decreased. T2D patients carrying

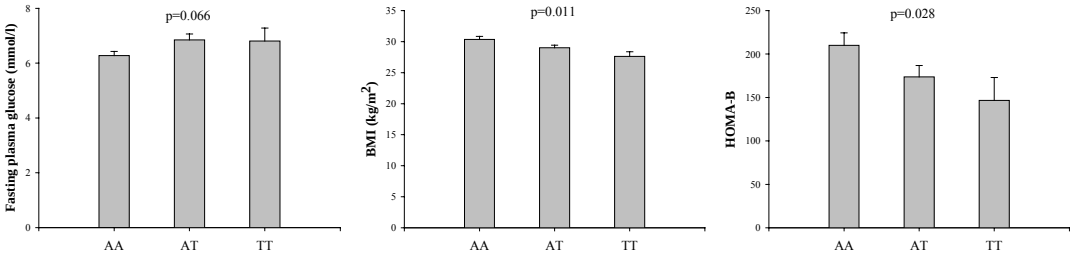
genotypes CT or TT in SNP rs7903146 had a more pronounced increase in fasting plasma glucose levels (p=0.042) and lower HOMA-β index (p=0.015) as well as BMI (p=0.015). Similarly, T2D patients carrying the genotypes with risk allele in SNPs rs7901695 and rs4506565 had decreased HOMA-β index (p=0.026 and 0.028) and BMI (p=0.004 and 0.011).

Figure 8. Quantitative traits in T2D patients

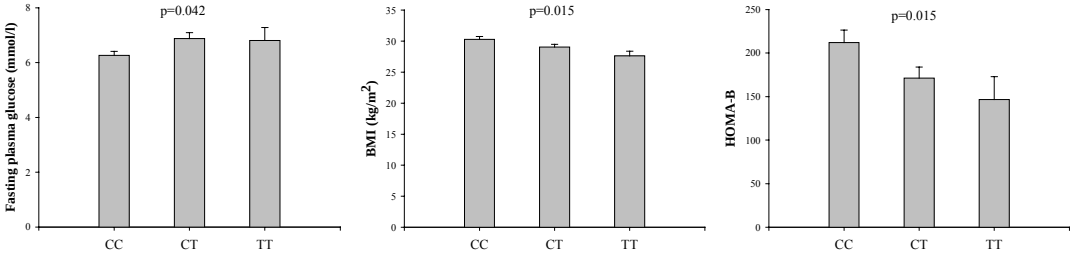
A. rs7901695



B. rs4506565



C. rs7903146



Both NGT subjects and T2D patients comprised a wide range of BMI from 18.3 to 45.6 kg/m^2. Analyses for allelic association of these 4 SNPs except rs11196205 in NGT subjects and T2D patients with BMI <30 kg/m^2 and >30 kg/m^2 were performed, respectively. SNPs rs7903146, rs4506565, rs7901695 and rs12255372 showed significant association with T2D with BMI <30 kg/m^2 but not with T2D with BMI >30 kg/m^2 (Table 6).

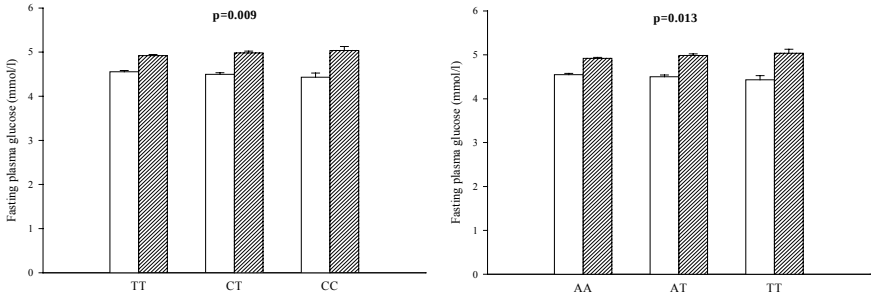
Table 6. Allelic association in T2D patients classified by BMI compared to lean NGT subjects

SNP	T2D with BMI<30	T2D with BMI>=30
rs7901695	0.00012:1.784 (1.313-2.237)	0.51366:1.131 (0.782-1.635)
rs4506565	0.00013:1.742 (1.308-2.319)	0.42964:1.159 (0.803-1.673)
rs7903146	0.00002:1,862 (1,395-2,486)	0.26526:1,234 (0,852-1,787)
rs11196205	0.2591:1.167 (0.892-1.527)	0.29568:1.190 (0.859-1.647)
rs12255372	0.0248:1.101 (1.043-1.883)	0.48317:1.141 (0.789-1.605)

Note: Data indicated p-value:OR (95% CI).

A prospective analysis was also carried out in order to predict genetic influence of the TCF7L2 polymorphisms in the development of T2D. Figure 9 demonstrates that NGT subjects from the baseline to follow-up studies, who carry the risk alleles of SNPs rs7901695 and rs4506565, had a more pronounced increase in fasting plasma glucose levels (p=0.009 and 0.013).

Figure 9. Fasting plasma glucose levels at baseline (white bars) and follow-up (dashed bars) among NGT subjects according to genotypes of rs7901695 and 450656



rs 7901695 rs4506565

Note: Data presented in this figure were calculated by the follow-up:baseline ratio.

5.2.2 Multiple Marker Association

LD and haplotype analyses for all studied five SNPs were carried out. Strong LD values from SNP rs7903146 to either rs4506565 or rs12255372 were found. Five common haplotypes and their frequencies in the cases and controls were analyzed. Haplotype H5 constructed with 5 risk alleles from all studied SNPs was strongly associated with T2D (p=0.002).

5.3 PAPER III – NPY STUDY

5.3.1 Single Marker Association

Genotyping for Leu7Pro (T1128C) polymorphism in the NPY gene was performed by DASH. A total of 309 IGT subjects and 469 non-diabetic individuals from SDPP were included, while 263 T2D patients were selected from both SDPP and Kronan study. The distribution of this polymorphism was in HWE and frequencies of the allele C in IGT subjects and T2D patients in Swedish men were 13 % (p=0.002) and 10 % (p=0.007) respectively, which were significantly higher than the frequency in non-diabetic controls (6 %). Further analyses indicated that a strong association for Leu7Pro polymorphism

with IGT and T2D was observed (p=0.002 and p=0.007, respectively) in men, when systolic and diastolic blood pressures, WHR, BMI and age were included as potential confounding factors(Table 7). However, no significant association of this polymorphism in female IGT subjects and T2D patients was found.

Table 7. Association of NPY with male IGT subjects and T2D patients

Leu7Pro Genotypes	IGT OR (95% CI)	T2D OR (95% CI)
TT	Reference	Reference
TC+CC	3.70 (1.65-8.29) p=0.002	4.08 (1.47-11.33) p=0.007

Note: Data were adjusted by age, BMI, WHR and blood pressures.

In this non-synomous SNP, the allele C is a risk allele, which predicts the susceptibility in the development of T2D. Therefore, analyses of the quantitative traits were performed between the carriers with CC and CT genotypes and the subjects carrying TT genotype. In male IGT subjects, the carriers with CT and CC genotypes had significantly higher fasting plasma glucose levels in comparison with the carriers with TT genotype (5.6±0.7 mmol/l vs 5.2±0.7 mmol/l, p=0.021). In male patients with T2D, the carriers with CT and CC genotypes had slightly increased BMI, waist circumference, fasting plasma glucose and fasting plasma insulin. There was no significant association with the studied traits and genotypes among Swedish women. Thus, this study provides the evidence that Leu7Pro (T1128 C) polymorphism in the NPY gene is associated with IGT and T2D in Swedish men.

5.4 PAPER IV – AC3 STUDY

5.4.1 Variation Screening in the Putative Promoter

It was previously found that the AC3 gene expression is increased in GK rats islets compared to Wistar rat due to two point mutations in the promoter (77). Variation screening of the putative promoter in human AC3 gene was then performed in a total of 48 samples (40 T2D patients and 8 NGT subjects). The sequence of 2236 bp upstream from the start codon in the AC3 gene was screened by using a direct sequencing protocol. A novel variant at -17A/T was identified.

5.4.2 Single Marker Association

A total of 14 SNPs in the AC3 gene, including the novel variant in the putative promoter region, were genotyped in the material of 630 Swedish men, including 188 control subjects (NGT, BMI ≤26 kg/m^2) and 243 T2D patients (BMI from 18.4 to 45.6 kg/m^2). Two SNPs rs2033655 and rs191968482 were significantly associated with T2D. Because T2D patients had a wide range in BMI, we divided them into two groups with BMI <30 kg/m^2 and >30 kg/m^2. Results demonstrated that SNPs rs2033655 C/T and rs1968482 A/G were found to be significantly associated with obesity when T2D patients had BMI ≥30 kg/m^2 (p=0.003 and 0.005). The significance was borderline in T2D patients with BMI <30 kg/m^2 (p=0.051 and 0.084) and disappeared in T2D patients with BMI ≤26 kg/m^2. This predicted that these two polymorphisms might be associated with obesity. Additional 199 obese subjects (BMI ≥30 kg/m^2) with NGT were therefore added into this study. Further genotyping and analyses with SNPs rs2033655 and rs191968482 indicated

that these two polymorphisms were associated with obesity in Swedish men (p=0.028 and 0.003) (Table 8).

Table 8. Allele frequencies of SNPs rs2033655 and rs1968482 of T2D and/or obese subjects compared with NGT lean subjects

Groups	rs2033655 p-value (OR, CI 95%)	rs1968482 p-value (OR, CI 95%)
All T2D	0.004 (1.509, 1.139-2.000)	0.008 (1.475, 1.106-1.965)
T2D (BMI ≤26 kg/m^2)	0.817	0.325
T2D (BMI <30 kg/m^2)	0.051 (1.370, 0.998-1.881)	0.084
T2D (BMI ≥30 kg/m^2)	0.003 (1.775, 1.215-2.592) /0.202*	0.005 (1.571, 1.186-2.585) /0.190*
Obese	0.028 (1.384, 1.036-1.850) /0.195¤	0.003 (1.586, 1.175-2.141) /0.622¤

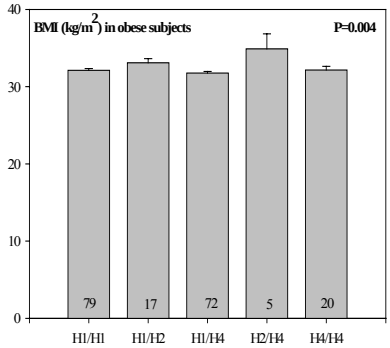
Note: *Comparison analysis for T2D (BMI ≥30 kg/m^2) vs T2D (BMI <30 kg/m^2). ¤Comparison analysis for obese vs T2D (BMI ≥30 kg/m^2).

The logistic regression analyses considering recessive inheritance model indicated a significant association with genotype distribution of these two SNPs in obese subjects (p=0.014, OR=1.714, CI 95% 1.114–2.635 and p=0.002, OR=1.906, CI 95% 1.264–2.876). The identified putative promoter variant had relatively low minor allele frequency and the association with T2D was not detected.

5.4.3 Multiple Marker Association

Four haplotypes constructed with SNPs rs2023655 and rs1968482 included H1:C-A, H2:T-A, H3:C-G, and H4:T-G. The frequency of H3 was low (1% or less) and it therefore was excluded from further analysis. Haplotype frequencies in NGT controls were significantly different from T2D patients and obese subjects. Diplotype (combined haplotypes) association analyses with quantitative traits were performed and data revealed a significant association with BMI (p=0.004) among obese subjects. The BMI was higher in the group of diplotype H2/H4 (Figure 10).

Figure 10. Diplotype association with BMI in obese subjects



6 DISCUSSION

SNP genotyping methods

Three high throughput genotyping methods, including DASH, Pyrosequencing and TaqMan allelic discrimination, have been used in the present study. All three methods require a PCR step. DASH and TaqMan are based upon the hybridization with specific probe(s), while Pyrosequencing is based upon the enzymatic reactions. Each method has its advantage and disadvantage. DASH technique is relatively cheap. TaqMan allelic discrimination does not require post PCR handling of the samples, and the instrument can be used for not only allelic discrimination but also for real time PCR or RT-PCR experiments. Pyrosequencing can be used for analysis of short sequences as well.

Genotyping quality

For genotyping quality control, the subjects were distributed randomly across plates with the similar numbers of cases and controls per plate. Negative controls (water blanks) were included on each plate. Successful genotype calls were ≥95% and plates were randomly genotyped twice for duplication accuracy, which was calculated to be 98%. HWE test for genotype distributions and allele frequencies is also used for prediction of the possible genotyping error. When it is necessary, direct sequencing protocol can be used for sequencing analysis of the PCR fragment, which contains the specific SNP.

Susceptibility of the IDE, TCF7L2, NPY and AC3 genes to T2D

The genetic association studies have become useful tools in identification of the susceptibility genes in T2D and other complex diseases. In the present study, the candidate gene approach has been used, and four genes, including IDE, TCF7L2, NPY and AC3 have been studied. The IDE gene is located in the region of chromosome 10q linked to T2D. We have mainly focused on genetic association study of the IDE gene with T2D but a polymorphism in 3'-strand of the HHEX gene is also included in the study. Three tag SNPs i.e. rs2251101 and rs1832196 in the IDE gene and rs1544210 in the HHEX gene were identified. Interestingly, diplotypes constructed with these three tag SNPs were found to be significantly associated with fasting plasma insulin levels, 2h insulin levels, HOMA-IR and BMI in NGT and IGT subjects. However, no statistically significant association of the IDE genetic polymorphisms with T2D is detected in the present study and also reported by Groves et al. (56) and Florez et al. (97). Sladek et al. have recently conducted a GWA study in T2D and identified that a LD block, covering IDE-KIF11-HHEX genes, confer the susceptibility to T2D (32). Therefore, IDE may not have the major risk susceptibility to the development of T2D but it may interact with the genes nearby such as HHEX and KIF11 in term of the genetic influence in T2D.

The TCF7L2 gene is also located in the same region of chromosome 10q as the IDE gene. In recent years, the TCF7L2 gene has been extensively studied and the polymorphisms in this gene are strongly associated with T2D in different populations. We have replicated the genetic association study of the TCF7L2 gene with T2D in Swedish Caucasians. Data indicate that the TCF7L2 polymorphisms are strongly associated with T2D in Swedish men. TCF7L2 may have crucial contribution to impaired insulin secretion underlining the development of T2D. However, there is no significant association between TCF7L2 polymorphisms and insulin resistance. These findings, one or another, have been reported by the previous studies in other populations, including, Finish, Italian, British,

Scandinavian, Polish and American. Hence so far obtained data are consistent with the hypothesis that TCF7L2 modulates glucose metabolism mainly by regulating insulin secretion. In the present study, the propensities of risk alleles in the TCF7L2 polymorphisms are more pronounced in the subjects with lower BMI. Furthermore, Humphries et al. have studied SNPs rs7903146 and rs12255372 and demonstrated that the subjects with lower BMI had an increased risk for T2D, whereas the subjects with higher BMI (>30 kg/m^2) did not show an increased risk for T2D (98). Additionally, several reports have indicated that the carriers with the genotypes of risk alleles for T2D among T2D patients and/or control subjects are associated with decreased BMI (99-102). Therefore, the TCF7L2 genetic polymorphisms associated with T2D unlikely contribute to the development of obesity, although T2D and obesity are known to be often associated with each other. Interestingly, we demonstrate that NGT subjects carrying the genotypes with the risk alleles of SNPs rs7903146 and rs12255372 had a more pronounced increase in fasting plasma glucose levels during a 10-year follow-up period in comparison with the subjects with the wild alleles. This is in agreement with the previous study in a French cohort revealing higher incidence of hyperglycemia in the subjects carrying risk allele of SNP rs7903146 (61). Similarly, Florez et al. have showed that IGT subjects, who carry the homozygous risk alleles of SNPs rs7903146 and rs12255372, have higher risk to develop T2D (62).

Schäfer et al. have demonstrated that GLP-1 induced insulin secretion in the subjects with the risk alleles of SNPs rs7903146 and rs12255372 is reduced, whereas concentrations of GLP-1 during OGTT is not influenced by the polymorphisms (103). Moreover, plasma concentrations of GIP and glucagon are not influenced with the TCF7L2 polymorphisms (63). These findings indicate that GLP-1 induced insulin secretion is impaired due to functional defects in the GLP-1 signaling in β-cells rather than a reduction in GLP-1 secretion.

Previously, several chromosome regions including chromosome 10q have been found to link with T2D by GWS and linkage analyses. Recently, several GWA reports and present study have confirmed that the TCF7L2, IDE, KIF11 and HHEX genes, which are located in the region of chromosome 10q, are associated with T2D. This is the first example suggesting that there are at least two loci under a linkage peak linked to T2D. However, further investigation of the interaction between TCF7L2 and other susceptibility genes in this chromosomal region by using multiplex gene analyses will provide useful information for better understanding of the impact of the interplay of this group of genes in the development of T2D.

NPY Leu7Pro polymorphism and its susceptibility to T2D

We have found evidence that the NPY Leu7Pro polymorphism is associated with both IGT and T2D in Swedish men and that the allele C confers the risk susceptibility to the development of T2D. The previous studies have demonstrated that this allele shows a decreasing frequency with the geographical distribution from north to south gradient. The highest C allele frequency was found in Finnish population and the second high frequency in Swedish population (104). Therefore, the risk allele C might be a factor facilitating the development of T2D under certain environmental conditions.

The hypothalamus consists of several nuclei involved in food intake, including the arcuate nucleus (ARC), the paraventricular nucleus (PVN), the lateral hypothalamic area (LHA), the ventromedial nucleus (VMH), and the dorsomedial nucleus (DMH). ARC neurons are located at the bottom of the hypothalamus around the third ventricle and are called ‘first

order neurons' because of their 'direct' contact with peripheral satiety factors like leptin and insulin. NPY has multiple functions, including regulation of satiety, energy balance and insulin release, while Pro-opiomelanocortin (POMC) works as a counterpart to NPY. NPY-producing neurons in the arcuate nucleus stimulate food intake, whereas arcuate nucleus neurons that release the POMC-derived peptide α-melanocyte-stimulating hormone (α-MSH) potently reduce food intake. Therefore, both NPY and POMC play a role in regulation of food intake and metabolism. It has been demonstrated that POMC neurons has a role in the overall physiological control of blood glucose (105). However, the interaction between NPY and POMC as well as their roles in the pathogenesis of obesity and T2D are still not fully understood.

Association of the AC3 genetic polymorphisms with obesity

We have conducted a genetic association study for the AC3 gene in T2D patients and obese subjects in Swedish men. The study revealed that two SNPs (rs2033655 and rs1968482) were associated with obese NGT subjects and obese T2D patients. Multiple marker association analyses of these SNPs predict a significant association with BMI in obese NGT subjects. There was no significant association of the SNPs with non-obese T2D patients, suggesting that this gene may play a role in the pathogenesis of obesity. However, we are unable to rule out the possibility that AC3 may have genetic influence in T2D because the number of non-obese T2D patients included in the present study is limited. Replication study with a large cohort of non-obese T2D is necessary to conclude whether AC3 genetic polymorphisms confer the susceptibility risk to the development of T2D.

The AC3 gene is located on chromosome 2p, and the POMC gene resides in the same chromosomal region. Evidence has indicated that POMC plays a role in the regulation of satiety (106). POMC locus is linked to leptin levels, a predictor of obesity, in American Caucasians, Mexican-American, and African-American families (107). POMC neuron is one of the neurons in the brain that is excited by glucose. This process is driven by the closure of KATP channels. The glucose sensing POMC neurons is impaired in obesity (105). Interestingly, variations in the POMC gene are found to be associated with obesity (108). The interaction between AC3 and POMC and their effects in obesity is unknown.

The human AC3 gene has ~94% sequence homology of the translated amino acids and ~80% sequence homology of the promoter regions compared to the gene in rat. In the present study, we found a novel variant in the promoter region of the human AC3 gene. However, no evidence was found that the polymorphisms in the promoter region of the AC3 gene in man are significantly associated with T2D, which differs from the study of variation in the promoter of the AC3 gene in the GK rat (77). Comparative genomic analysis using animal models may provide useful information for identifying the genetic and environmental factors responsible for complex diseases (109; 110). However, physiological distinction and/or similarity between man and animal models may not be explained by simple comparison of genomic identity (111).

In this thesis, we have studied the TCF7L2 and AC3 genes with male subjects while the IDE and NPY genes with both male and female subjects. The significant associations of the IDE and NPY genetic polymorphisms with insulin levels or T2D were found in male subjects only. In the recent years, several studies with SDPP have been carried out, and

the results may be interesting in this content (84). Kuhl et al. have found prevalence of early abnormalities of glucose metabolism was two to three times higher in male compared to female subjects (112). Thus, results from genetic association studies in this thesis are in agreement with the studies briefly described above in term of gender specificity. However, more evidence needs to be accumulated to conclude whether the gender specificity is related to the susceptibility of the genes in T2D.

T2D is a complex disease influenced by both genetic and environmental factors. Environmental factors such as food habits and living a sedentary lifestyle also play a role in the development of the disease. In fact the enormous increase in T2D prevalence over the past decades is thought to depend to a big extent on different environmental factors, since the genetic make-up of mammals from an evolutionary point of view can not change this fast. Factors in today's lifestyle are likely to pull the trigger for the development of the disease in people who have a certain genotypes in their genome. In this case the genes and genetic variants can be argued as a prerequisite for the environmental factors to cause the disease. Increased knowledge of the pathogenesis of the disease will increase our possibility of finding new therapeutic targets and thereby increase the life quality for T2D patients. Furthermore, identification of genetic variants associated with T2D or obesity open a possibility to identify peoples with a high-risk of developing the disease, and thereby an opportunity for them to change the lifestyle and likely postpone the breakthrough of the disease. Therefore, it is of great importance to find the susceptibility genes and environmental factors involved in the pathogenesis of the disease.

7 CONCLUSIONS

Paper I – IDE study
We provide evidence that sequence variation within or near the IDE gene contributes to variability in measures of insulin metabolism in Swedish men.

Paper II – TCF7L2 study
TCF7L2 polymorphisms are associated with T2D in Swedish men. TCF7L2 may have crucial contribution to impaired insulin secretion underlining the development of T2D.

Paper III – NPY study
Leu7Pro polymorphism in the NPY gene is associated with IGT and T2D in Swedish men. The allele C confers the risk susceptibility to the development of T2D.

Paper IV – AC3 study
The present study provides the first evidence that AC3 genetic polymorphisms are associated with obesity and obese T2D in Swedish men.

8 SAMMANFATTNING PÅ SVENSKA

Typ 2 diabetes (T2D) är en komplex sjukdom som utvecklas på basen av interaktionen mellan genetiska och miljömässiga faktorer. Denna avhandling omfattar associationsstudier av fyra kandidatgener för T2D. Dessa gener har valts ut från en region på kromosom 10q som är länkad till T2D, eller baserat på genens biologiska funktion och/eller medverkande i viktiga T2D-relaterade signalvägar.

Insulin degrading enzyme (IDE) har en viktig roll vid nedbrytningen av insulin och vissa andra peptider. Genen som kodar för IDE hittas på kromosom 10q i en region som är länkad till T2D. För att finna associationen mellan genetiska variationer och T2D, samt för att undersöka om polymorfismer i IDE-genen uppvisar mätbar påverkan på insulinnivåer, genotypades 14 single nucleotide polymorphisms (SNPs) bland 321 individer med nedsatt glukostolerans (IGT) och 403 individer med normal glukostolerans (NGT) utvalda från Stockholms Diabetespreventiva Program (SDPP). De studerade SNPs återfinns i ett område som sträcker sig över IDE- och IDE-HHEX-regionen. Analyser baserade på diplotyper (haplotypiska genotyper) konstruerade utifrån tre tag SNPs, visade association mellan denna kromosomregion och fasteinsulin ($p=0.0009$), 2h-plasmainsulin ($p=0.0027$), insulinresistens (mätt med HOMA-IR, $p=0.0001$) och BMI ($p=0.0067$). Resultaten fanns hos män men inte hos kvinnor. Således bidrar SNPs i eller nära IDE-genen väsentligen till varians i plasmainsulinnivåer och andra T2D-relaterade variabler.

Genen för transcription factor-7 like 2 (TCF7L2) återfinns också i den T2D-länkade regionen på kromosom 10q. Både en associations- och en 10 års prospektiv studie av genetiska variationer inom denna gen, utfördes bland manliga 243 T2D-patienter och 528 män med NGT. Fem genotypades. SNPs rs7901695, rs4506565, rs7903146 och rs12255372 är starkt associerade med T2D ($p<0.004$) hos svenska män utvalda från SDPP. Patienter med T2D som bär på CT- eller TT-genotypen av rs7903146 uppvisar en högre fasteglukosnivå ($p=0.042$), lägre HOMA-β ($p=0.015$) och BMI ($p=0.015$) jämfört med patienter med CC- genotypen. Liknande resultat återfanns bland T2D-patienter med olika genotyper i rs7901695, rs4506565 och rs12255372. Friska personer som bar på riskallel genotypen för rs7901695 och rs4506565 uppvisade en mer uttryckt ökning i fasteplasmaglukosnivå under uppföljningsperioden ($p=0.009$ respektive $p=0.013$). TCF7L2 polymorfismer är starkt associerade med T2D bland svenska män och kan ha en viktig roll vid den nedsatta insulinfrisättningen vid T2D.

Neuropeptide Y (NPY) reglerar mättnadskänsla, energibalans och insulinfrisättning. För att undersöka om Leu7Pro(1128C)-polymorfismen hos NPY-genen bidrar till utvecklingen av T2D, genotypades denna SNP bland 263 T2D-patienter, 309 personer med IGT och 469 med NGT. Resultaten visar att frekvensen av riskallelen C bland personer med IGT och T2D var 13% ($p=0.002$; OR=2.73; 95%CI 1.65-8.29) respektive 10% ($p=0.0070$; OR=4.08, 95%CI 1.47-11.33) bland svenska män, vilket var signifikant högre än C-allelfrekvensen hos individer med NGT (6%). Personerna med IGT som bär på TC- och CC-genotypen av denna SNP, har en signifikant högre fasteglukosnivå jämfört med TT-bärare (5.6 ± 0.7 mmol/l vs 5.2 ± 0.7 mmol/l $p=0.021$). Sålunda ger

denna studie bevis på att NPY Leu7Pro polymorfismen är associerad med IGT och T2D hos svenska män.

En tidigare studie utförd på Goto-Kakizakiråttor indikerade att adenylyl cyclase 3 (AC3) är en möjlig kandidatgen för T2D. Målet med vår studie var att undersöka om genetiska variationer inom denna gen uttrycker association med T2D hos män. Vid sökning efter genetiska variationer i promotorn hos den humana AC3-sekvensen identifierades en ny variant, -17A/T. Totalt 14 SNPs (inkluderande den nya varianten) genotypades bland 243 T2D-patienter och 188 personer med NGT valda från SDPP. Resultaten visade att SNPs rs2033655 C/T och rs1968482 A/G var associerade med T2D (p=0.003 and 0.005) hos obesa (BMI?30 kg/m^2) patienter. Genotypning av dessa två SNPs hos 199 obesa individer med NGT, visade starkta samband med fetma (p=0.028 respektive 0.003). Denna studie . tillhandahåller för första gången bevis på att genetiska variationer hos AC3-genen är associerade med obesitas hos svenska män med eller utan T2D

Sammanfattningsvis har denna avhandling bidragit med information angående rollen av IDE-, TCF7L2- och NPY-generna i utvecklingen av T2D. Vidare visar vi att AC3-genen kan ha en viktig roll för utvecklingen av fetma hos både personer med normal glukostolerans och T2D.

9 ACKNOWLEDGEMENTS

This thesis was carried out at the department of Molecular Medicine and Surgery at Karolinska Institutet, Solna. I would like to express my sincere gratitude to all the persons that have supported me during these years and made this thesis possible. Especially I whish to thank:

Docent **Harvest F. Gu,** my supervisor for great supervision, for introducing me to the field of genetics, for sharing your knowledge and working experience in molecular biology, for your never ending patience and enthusiasm, for your support, and for always being there for me.

Professor **Suad Efendic**, my co-supervisor for providing the excellent research environment, for your great scientific advice and guidance, fruitful discussions and for your never ending support. You are a great source of inspiration. I will always remember you, the nice man who introduced me to the exiting field of diabetic research. Thank you for letting me to take part in your excellent diabetes research.

Professor **Claes-Göran Östenson,** my co-supervisor for always being positive, for your support with material from the Stockholm diabetes prevention program, for your always valuable discussions, comments and advices.

From the bottom of my heart I would like to express my sincere gratitude to my family, especially, my boyfriend **Stefan Westlund** and my mother **Anja Nordman Palmgren** and of course also my father **Leif Palmgren** and my sister **Josefine Eurenius** and her husband **Daniel Eurenius** and their son **Elias Eurenius,** for coping with me. Without your love and support this would have been completely impossible. You all are the best!

Professor **Kerstin Brismar,** the head of Rolf Luft center for diabetic research and my co-author for nice collaboration and scientific support.

All nice peoples at the administration of the department of Molecular Medicine and Surgery, especially, **Helena Nässén, Christina Bremer, Kerstin Florell** and **Katarina Breitholtz**.

Lennart Helleday for excellent computer support, and for your efforts in bringing me sugar-free things.

Urban Rosenqvist for being my external mentor.

My co-authors **Lars Kärvestedt, Bo Ding, Anthony Brookes, Jonathan Prince, Keith Humphreys** and **Adil Abulati (former colleague), Ewa-Carin Långberg**, **Agneta Hilding** for collaboration and for valuable discussions and comments on manuscripts. **Ewa-Carin Långberg,** for being both a friend and a co-worker, for your encouragement and valuable discussions both in private and research. The lab wouldn't have been the

same without you. **Agneta Hilding,** for your never ending help and encouragement both with discussions and with statistical analyses, you taught me how to use Statistica.

My present and former colleagues: **Akhtar Kahn, Anita Maddock, Bee-Hoon Goh, Birgitta Isenberg, Dongying Zhang (Chris), Elisabeth Norén Krog, Fazliana Manzor, Galyna Bryzgalova, Hoa Nguyen Khanh, Huyen Vu Thi Than, Jie Chen, Kamal Yassin, Neil Portwood, Tina Wallin, Zuheng Ma** for creating a nice work environment. The lab should not have been the same without you all! Especially, I would like to thank **Yvonne Strömberg** for great laboratory assistance and many big laughs in the lab. **Anneli Björklund** for your kind concern and information about the dissertation procedure. **Anna-Karin Eriksson** for reading my thesis and always being supportive. **Jun Ma,** for your encouragement, understanding and support, for being a good listener, and for always spreading calmness around you.

My friend, **Lina Söderkvist** for always being there in good times and in bad times. Having a true friend like you is a privilege. **Per Sandgren** for valuable discussions. **Gunaratna Kuttuva** for being a good friend ever since we met as co-workers.

Erling Åkesson and **Cecilia Norlin** at Westnine Fitness for all help with keeping my body in a good physical condition.

Magnus Carlsson, without all the nice songs in your album "Live forever" I could never have finished writing my thesis. Your songs cheer me up and give me energy and faith in life.

I also want to thank all subjects for participating in the present studies.

The study was supported by grant from Novo Nordisk Foundation Consortium. The project was also supported by Swedish Research Council, Family Erling-Persson foundation, Novo Nordisk Scandinavia, Glaxo Smith Kline, Vetenskapligt arbete inom diabetologi foundation, Loo and Hans Osterman foundation and Swedish Diabetes Association.

10 REFERENCES

1. Amos C: Successful design and conduct of genome-wide association studies. *Hum Mol Genet* 16 Spec No. 2:R220-225, 2007

2. Toledo F, Menshikova E, Ritov V, Azuma K, Radikova Z, DeLany J, Kelley D: Effects of physical activity and weight loss on skeletal muscle mitochondria and relationship with glucose control in type 2 diabetes. *Diabetes* 56:2142-2147, 2007

3. Chen F, Liu Y, Song X, Hu H, Xu H, Zhang X, Shi J, Hu J, Shen Y, Lu B, Wang X, Hu R: A novel mitochondrial DNA missense mutation at G3421A in a family with maternally inherited diabetes and deafness. *Mutat Res* 602:26-33, 2006

4. Maassen J, 't Hart L, Janssen G, Reiling E, Romijn J, Lemkes H: Mitochondrial diabetes and its lessons for common Type 2 diabetes. *Biochem Soc Trans* 34:819-823, 2006

5. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449:851-861, 2007

6. Parikh H, Groop L: Candidate genes for type 2 diabetes. *Rev Endocr Metab Disord* 5:151-176, 2004

7. Goldstein D, Weale M: Population genomics: linkage disequilibrium holds the key. *Curr Biol* 11:R576-579, 2001

8. Pritchard J, Przeworski M: Linkage disequilibrium in humans: models and data. *Am J Hum Genet* 69:1-14, 2001

9. Wall J, Pritchard J: Haplotype blocks and linkage disequilibrium in the human genome. *Nat Rev Genet* 4:587-597, 2003

10. King K, Rubin G: A history of diabetes: from antiquity to discovering insulin. *Br J Nurs* 12:1091-1095, 2003

11. King K: A history of insulin: from discovery to modern alternatives. *Br J Nurs* 12:1137-1141, 2003

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

13. Fujimoto W: The importance of insulin resistance in the pathogenesis of type 2 diabetes mellitus. *Am J Med* 108 Suppl 6a:9S-14S, 2000

14. Leahy J: Pathogenesis of type 2 diabetes mellitus. *Arch Med Res* 36:197-209, 2005

15. King H, Aubert R, Herman W: Global burden of diabetes, 1995-2025: prevalence, numerical estimates, and projections. *Diabetes Care* 21:1414-1431, 1998

16. Wild S, Roglic G, Green A, Sicree R, King H: Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. *Diabetes Care* 27:1047-1053, 2004

17. Berger B, Stenström G, Chang Y, Sundkvist G: The prevalence of diabetes in a Swedish population of 280,411 inhabitants. A report from the Skaraborg Diabetes Registry. *Diabetes Care* 21:546-548, 1998

18. Roglic G, Unwin N, Bennett P, Mathers C, Tuomilehto J, Nag S, Connolly V, King H: The burden of mortality attributable to diabetes: realistic estimates for the year 2000. *Diabetes Care* 28:2130-2135, 2005

19. Haffner S, Lehto S, Rönnemaa T, Pyörälä K, Laakso M: Mortality from coronary heart disease in subjects with type 2 diabetes and in nondiabetic subjects with and without prior myocardial infarction. *N Engl J Med* 339:229-234, 1998

20. Norhammar A, Tenerz A, Nilsson G, Hamsten A, Efendíc S, Rydén L, Malmberg K: Glucose metabolism in patients with acute myocardial infarction and no previous diagnosis of diabetes mellitus: a prospective study. *Lancet* 359:2140-2144, 2002

21. Valsania P, Micossi P: Genetic epidemiology of non-insulin-dependent diabetes. *Diabetes Metab Rev* 10:385-405, 1994

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

23. Kaprio J, Tuomilehto J, Koskenvuo M, Romanov K, Reunanen A, Eriksson J, Stengård J, Kesäniemi Y: Concordance for type 1 (insulin-dependent) and type 2 (non-insulin-dependent) diabetes mellitus in a population-based cohort of twins in Finland. *Diabetologia* 35:1060-1067, 1992

24. Diabetes mellitus in twins: a cooperative study in Japan. Committee on Diabetic Twins, Japan Diabetes Society. *Diabetes Res Clin Pract* 5:271-280, 1988

25. MacGregor A, Snieder H, Schork N, Spector T: Twins. Novel uses to study complex traits and genetic diseases. *Trends Genet* 16:131-134, 2000

26. van Tilburg J, van Haeften T, Pearson P, Wijmenga C: Defining the genetic contribution of type 2 diabetes mellitus. *J Med Genet* 38:569-578, 2001

27. Malecki M: Genetics of type 2 diabetes mellitus. *Diabetes Res Clin Pract* 68 Suppl1:S10-21, 2005

28. Grant S, Hakonarson H: Recent development in pharmacogenomics: from candidate genes to genome-wide association studies. *Expert Rev Mol Diagn* 7:371-393, 2007

29. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 316:1331-1336, 2007

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

31. Scott L, Mohlke K, Bonnycastle L, Willer C, Li Y, Duren W, Erdos M, Stringham H, Chines P, Jackson A, Prokunina-Olsson L, Ding C, Swift A, Narisu N, Hu T, Pruim R, Xiao R, Li X, Conneely K, Riebow N, Sprau A, Tong M, White P, Hetrick K, Barnhart M, Bark C, Goldstein J, Watkins L, Xiang F, Saramies J, Buchanan T, Watanabe R, Valle T, Kinnunen L, Abecasis G, Pugh E, Doheny K, Bergman R, Tuomilehto J, Collins F, Boehnke M: A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 316:1341-1345, 2007

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

33. Steinthorsdottir V, Thorleifsson G, Reynisdottir I, Benediktsson R, Jonsdottir T, Walters G, Styrkarsdottir U, Gretarsdottir S, Emilsson V, Ghosh S, Baker A, Snorradottir

S, Bjarnason H, Ng M, Hansen T, Bagger Y, Wilensky R, Reilly M, Adeyemo A, Chen Y, Zhou J, Gudnason V, Chen G, Huang H, Lashley K, Doumatey A, So W, Ma R, Andersen G, Borch-Johnsen K, Jorgensen T, van Vliet-Ostaptchouk J, Hofker M, Wijmenga C, Christiansen C, Rader D, Rotimi C, Gurney M, Chan J, Pedersen O, Sigurdsson G, Gulcher J, Thorsteinsdottir U, Kong A, Stefansson K: A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat Genet* 39:770-775, 2007

34. Zeggini E, Weedon M, Lindgren C, Frayling T, Elliott K, Lango H, Timpson N, Perry J, Rayner N, Freathy R, Barrett J, Shields B, Morris A, Ellard S, Groves C, Harries L, Marchini J, Owen K, Knight B, Cardon L, Walker M, Hitman G, Morris A, Doney A: Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 316:1336-1341, 2007

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

36. Gloyn A, Weedon M, Owen K, Turner M, Knight B, Hitman G, Walker M, Levy J, Sampson M, Halford S, McCarthy M, Hattersley A, Frayling T: Large-scale association studies of variants in genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes. *Diabetes* 52:568-572, 2003

37. Nielsen E, Hansen L, Carstensen B, Echwald S, Drivsholm T, Glümer C, Thorsteinsson B, Borch-Johnsen K, Hansen T, Pedersen O: The E23K variant of Kir6.2 associates with impaired post-OGTT serum insulin response and increased risk of type 2 diabetes. *Diabetes* 52:573-577, 2003

38. Hansen L, Pedersen O: Genetics of type 2 diabetes mellitus: status and perspectives. *Diabetes Obes Metab* 7:122-135, 2005

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

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

41. Frayling T, Timpson N, Weedon M, Zeggini E, Freathy R, Lindgren C, Perry J, Elliott K, Lango H, Rayner N, Shields B, Harries L, Barrett J, Ellard S, Groves C, Knight B, Patch A, Ness A, Ebrahim S, Lawlor D, Ring S, Ben-Shlomo Y, Jarvelin M, Sovio U, Bennett A, Melzer D, Ferrucci L, Loos R, Barroso I, Wareham N, Karpe F, Owen K, Cardon L, Walker M, Hitman G, Palmer C, Doney A, Morris A, Smith G, Hattersley A, McCarthy M: A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* 316:889-894, 2007

42. Scuteri A, Sanna S, Chen W, Uda M, Albai G, Strait J, Najjar S, Nagaraja R, Orrú M, Usala G, Dei M, Lai S, Maschio A, Busonero F, Mulas A, Ehret G, Fink A, Weder A, Cooper R, Galan P, Chakravarti A, Schlessinger D, Cao A, Lakatta E, Abecasis G: Genome-Wide Association Scan Shows Genetic Variants in the FTO Gene Are Associated with Obesity-Related Traits. *PLoS Genet* 3:e115, 2007

43. Frayling T, McCarthy M: Genetic studies of diabetes following the advent of the genome-wide association study: where do we go from here? *Diabetologia* 50:2229-2233, 2007

44. Grarup N, Andersen G: Gene-environment interactions in the pathogenesis of type 2 diabetes and metabolism. *Curr Opin Clin Nutr Metab Care* 10:420-426, 2007

45. Keller U: From obesity to diabetes. *Int J Vitam Nutr Res* 76:172-177, 2006

46. Mokdad A, Bowman B, Ford E, Vinicor F, Marks J, Koplan J: The continuing epidemics of obesity and diabetes in the United States. *JAMA* 286:1195-1200, 2001

47. Hales C, Barker D: The thrifty phenotype hypothesis. *Br Med Bull* 60:5-20, 2001

48. Li P, Kuo W, Yousef M, Rosner M, Tang W: The C-terminal domain of human insulin degrading enzyme is required for dimerization and substrate recognition. *Biochem Biophys Res Commun* 343:1032-1037, 2006

49. Qiu W, Folstein M: Insulin, insulin-degrading enzyme and amyloid-beta peptide in Alzheimer's disease: review and hypothesis. *Neurobiol Aging* 27:190-198, 2006

50. Farris W, Leissring M, Hemming M, Chang A, Selkoe D: Alternative splicing of human insulin-degrading enzyme yields a novel isoform with a decreased ability to degrade insulin and amyloid beta-protein. *Biochemistry* 44:6513-6525, 2005

51. Kuo W, Gehm B, Rosner M, Li W, Keller G: Inducible expression and cellular localization of insulin-degrading enzyme in a stably transfected cell line. *J Biol Chem* 269:22599-22606, 1994

52. Farris W, Mansourian S, Chang Y, Lindsley L, Eckman E, Frosch M, Eckman C, Tanzi R, Selkoe D, Guenette S: Insulin-degrading enzyme regulates the levels of insulin, amyloid beta-protein, and the beta-amyloid precursor protein intracellular domain in vivo. *Proc Natl Acad Sci U S A* 100:4162-4167, 2003

53. Hwang D, Seo S, Kim Y, Kim C, Shim S, Jee S, Lee S, Sin J, Cho J, Kang B, Jang I: Significant change in insulin production, glucose tolerance and ER stress signaling in transgenic mice coexpressing insulin-siRNA and human IDE. *Int J Mol Med* 19:65-73, 2007

54. Fakhrai-Rad H, Nikoshkov A, Kamel A, Fernström M, Zierath J, Norgren S, Luthman H, Galli J: Insulin-degrading enzyme identified as a candidate diabetes susceptibility gene in GK rats. *Hum Mol Genet* 9:2149-2158, 2000

55. Runyan K, Duckworth W, Kitabchi A, Huff G: The effect of age on insulin-degrading activity in rat tissue. *Diabetes* 28:324-325, 1979

56. Groves C, Wiltshire S, Smedley D, Owen K, Frayling T, Walker M, Hitman G, Levy J, O'Rahilly S, Menzel S, Hattersley A, McCarthy M: Association and haplotype analysis of the insulin-degrading enzyme (IDE) gene, a strong positional and biological candidate for type 2 diabetes susceptibility. *Diabetes* 52:1300-1305, 2003

57. Karamohamed S, Demissie S, Volcjak J, Liu C, Heard-Costa N, Liu J, Shoemaker C, Panhuysen C, Meigs J, Wilson P, Atwood L, Cupples L, Herbert A: Polymorphisms in the insulin-degrading enzyme gene are associated with type 2 diabetes in men from the NHLBI Framingham Heart Study. *Diabetes* 52:1562-1567, 2003

58. Cauchi S, Meyre D, Dina C, Choquet H, Samson C, Gallina S, Balkau B, Charpentier G, Pattou F, Stetsyuk V, Scharfmann R, Staels B, Frühbeck G, Froguel P: Transcription factor TCF7L2 genetic study in the French population: expression in human beta-cells

and adipose tissue and strong association with type 2 diabetes. *Diabetes* 55:2903-2908, 2006

59. Cauchi S, El Achhab Y, Choquet H, Dina C, Krempler F, Weitgasser R, Nejjari C, Patsch W, Chikri M, Meyre D, Froguel P: TCF7L2 is reproducibly associated with type 2 diabetes in various ethnic groups: a global meta-analysis. *J Mol Med* 85:777-782, 2007

60. De Silva N, Steele A, Shields B, Knight B, Parnell K, Weedon M, Hattersley A, Frayling T: The transcription factor 7-like 2 (TCF7L2) gene is associated with Type 2 diabetes in UK community-based cases, but the risk allele frequency is reduced compared with UK cases selected for genetic studies. *Diabet Med*, 2007

61. Cauchi S, Meyre D, Choquet H, Dina C, Born C, Marre M, Balkau B, Froguel P: TCF7L2 variation predicts hyperglycemia incidence in a French general population: the data from an epidemiological study on the Insulin Resistance Syndrome (DESIR) study. *Diabetes* 55:3189-3192, 2006

62. Florez J, Jablonski K, Bayley N, Pollin T, de Bakker P, Shuldiner A, Knowler W, Nathan D, Altshuler D: TCF7L2 polymorphisms and progression to diabetes in the Diabetes Prevention Program. *N Engl J Med* 355:241-250, 2006

63. Lyssenko V, Lupi R, Marchetti P, Del Guerra S, Orho-Melander M, Almgren P, Sjögren M, Ling C, Eriksson K, Lethagen A, 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* 117:2155-2163, 2007

64. Billington C, Briggs J, Grace M, Levine A: Effects of intracerebroventricular injection of neuropeptide Y on energy metabolism. *Am J Physiol* 260:R321-327, 1991

65. Ahrén B, Larsson H: Plasma neuropeptide Y in impaired glucose tolerance. *Acta Diabetol* 33:295-297, 1996

66. Milewicz A, Bidzińska B, Mikulski E, Demissie M, Tworowska U: Influence of obesity and menopausal status on serum leptin, cholecystokinin, galanin and neuropeptide Y levels. *Gynecol Endocrinol* 14:196-203, 2000

67. Karvonen M, Pesonen U, Koulu M, Niskanen L, Laakso M, Rissanen A, Dekker J, Hart L, Valve R, Uusitupa M: Association of a leucine(7)-to-proline(7) polymorphism in the signal peptide of neuropeptide Y with high serum cholesterol and LDL cholesterol levels. *Nat Med* 4:1434-1437, 1998

68. Blumenthal J, Andersen R, Mitchell B, Seibert M, Yang H, Herzog H, Beamer B, Franckowiak S, Walston J: Novel neuropeptide Y1 and Y5 receptor gene variants: associations with serum triglyceride and high-density lipoprotein cholesterol levels. *Clin Genet* 62:196-202, 2002

69. Zhu G, Pollak L, Mottagui-Tabar S, Wahlestedt C, Taubman J, Virkkunen M, Goldman D, Heilig M: NPY Leu7Pro and alcohol dependence in Finnish and Swedish populations. *Alcohol Clin Exp Res* 27:19-24, 2003

70. Kallio J, Pesonen U, Jaakkola U, Karvonen M, Helenius H, Koulu M: Changes in diurnal sympathoadrenal balance and pituitary hormone secretion in subjects with Leu7Pro polymorphism in the prepro-neuropeptide Y. *J Clin Endocrinol Metab* 88:3278-3283, 2003

71. Pihlajamäki J, Karhapää P, Vauhkonen I, Kekäläinen P, Kareinen A, Viitanen L, Pesonen U, Kallio J, Uusitupa M, Laakso M: The Leu7Pro polymorphism of the neuropeptide Y gene regulates free fatty acid metabolism. *Metabolism* 52:643-646, 2003

72. Niskanen L, Karvonen M, Valve R, Koulu M, Pesonen U, Mercuri M, Rauramaa R, Töyry J, Laakso M, Uusitupa M: Leucine 7 to proline 7 polymorphism in the neuropeptide Y gene is associated with enhanced carotid atherosclerosis in elderly patients with type 2 diabetes and control subjects. *J Clin Endocrinol Metab* 85:2266-2269, 2000

73. Pettersson-Fernholm K, Karvonen M, Kallio J, Forsblom C, Koulu M, Pesonen U, Fagerudd J, Groop P: Leucine 7 to proline 7 polymorphism in the preproneuropeptide Y is associated with proteinuria, coronary heart disease, and glycemic control in type 1 diabetic patients. *Diabetes Care* 27:503-509, 2004

74. Ding B, Kull B, Liu Z, Mottagui-Tabar S, Thonberg H, Gu H, Brookes A, Grundemar L, Karlsson C, Hamsten A, Arner P, Ostenson C, Efendic S, Monné M, von Heijne G, Eriksson P, Wahlestedt C: Human neuropeptide Y signal peptide gain-of-function polymorphism is associated with increased body mass index: possible mode of function. *Regul Pept* 127:45-53, 2005

75. Ukkola O, Kesäniemi Y: Leu7Pro polymorphism of PreproNPY associated with an increased risk for type II diabetes in middle-aged subjects. *Eur J Clin Nutr* 61:1102-1105, 2007

76. Hanoune J, Defer N: Regulation and role of adenylyl cyclase isoforms. *Annu Rev Pharmacol Toxicol* 41:145-174, 2001

77. Abdel-Halim S, Guenifi A, He B, Yang B, Mustafa M, Höjeberg B, Hillert J, Bakhiet M, Efendić S: Mutations in the promoter of adenylyl cyclase (AC)-III gene, overexpression of AC-III mRNA, and enhanced cAMP generation in islets from the spontaneously diabetic GK rat model of type 2 diabetes. *Diabetes* 47:498-504, 1998

78. Guenifi A, Portela-Gomes G, Grimelius L, Efendić S, Abdel-Halim S: Adenylyl cyclase isoform expression in non-diabetic and diabetic Goto-Kakizaki (GK) rat pancreas. Evidence for distinct overexpression of type-8 adenylyl cyclase in diabetic GK rat islets. *Histochem Cell Biol* 113:81-89, 2000

79. Bégin-Heick N: Liver beta-adrenergic receptors, G proteins, and adenylyl cyclase activity in obesity-diabetes syndromes. *Am J Physiol* 266:C1664-1672, 1994

80. Diabetes mellitus. Report of a WHO Study Group. *World Health Organ Tech Rep Ser* 727:1-113, 1985

81. Carlsson S, Persson P, Alvarsson M, Efendic S, Norman A, Svanström L, Ostenson C, Grill V: Weight history, glucose intolerance, and insulin levels in middle-aged Swedish men. *Am J Epidemiol* 148:539-545, 1998

82. Hilding A, Eriksson A, Agardh E, Grill V, Ahlbom A, Efendic S, Ostenson C: The impact of family history of diabetes and lifestyle factors on abnormal glucose regulation in middle-aged Swedish men and women. *Diabetologia* 49:2589-2598, 2006

83. Andersson C, Bjärås G, Ostenson C: A stage model for assessing a community-based diabetes prevention program in Sweden. *Health Promot Int* 17:317-327, 2002

84. Agardh E, Ahlbom A, Andersson T, Efendic S, Grill V, Hallqvist J, Norman A, Ostenson C: Work stress and low sense of coherence is associated with type 2 diabetes in middle-aged Swedish women. *Diabetes Care* 26:719-724, 2003

85. Kwok P: Methods for genotyping single nucleotide polymorphisms. *Annu Rev Genomics Hum Genet* 2:235-258, 2001

86. Tsuchihashi Z, Dracopoli N: Progress in high throughput SNP genotyping methods. *Pharmacogenomics J* 2:103-110, 2002

87. Howell W, Jobs M, Gyllensten U, Brookes A: Dynamic allele-specific hybridization. A new method for scoring single nucleotide polymorphisms. *Nat Biotechnol* 17:87-88, 1999

88. Prince J, Feuk L, Howell W, Jobs M, Emahazion T, Blennow K, Brookes A: Robust and accurate single nucleotide polymorphism genotyping by dynamic allele-specific hybridization (DASH): design criteria and assay validation. *Genome Res* 11:152-162, 2001

89. Ronaghi M, Karamohamed S, Pettersson B, Uhlén M, Nyrén P: Real-time DNA sequencing using detection of pyrophosphate release. *Anal Biochem* 242:84-89, 1996

90. Ronaghi M, Uhlén M, Nyrén P: A sequencing method based on real-time pyrophosphate. *Science* 281:363, 365, 1998

91. Langaee T, Ronaghi M: Genetic variation analyses by Pyrosequencing. *Mutat Res* 573:96-102, 2005

92. Ahmadian A, Ehn M, Hober S: Pyrosequencing: history, biochemistry and future. *Clin Chim Acta* 363:83-94, 2006

93. Hampe J, Wollstein A, Lu T, Frevel H, Will M, Manaster C, Schreiber S: An integrated system for high throughput TaqMan based SNP genotyping. *Bioinformatics* 17:654-655, 2001

94. De la Vega F, Lazaruk K, Rhodes M, Wenz M: Assessment of two flexible and compatible SNP genotyping platforms: TaqMan SNP Genotyping Assays and the SNPlex Genotyping System. *Mutat Res* 573:111-135, 2005

95. Kutyavin I, Afonina I, Mills A, Gorn V, Lukhtanov E, Belousov E, Singer M, Walburger D, Lokhov S, Gall A, Dempcy R, Reed M, Meyer R, Hedgpeth J: 3'-minor groove binder-DNA probes increase sequence specificity at PCR extension temperatures. *Nucleic Acids Res* 28:655-661, 2000

96. Matthews D, Hosker J, Rudenski A, Naylor B, Treacher D, Turner R: Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* 28:412-419, 1985

97. Florez J, Wiltshire S, Agapakis C, Burtt N, de Bakker P, Almgren P, Bengtsson Boström K, Tuomi T, Gaudet D, Daly M, Hirschhorn J, McCarthy M, Altshuler D, Groop L: High-density haplotype structure and association testing of the insulin-degrading enzyme (IDE) gene with type 2 diabetes in 4,206 people. *Diabetes* 55:128-135, 2006

98. Humphries S, Gable D, Cooper J, Ireland H, Stephens J, Hurel S, Li K, Palmen J, Miller M, Cappuccio F, Elkeles R, Godsland I, Miller G, Talmud P: Common variants in the TCF7L2 gene and predisposition to type 2 diabetes in UK European Whites, Indian Asians and Afro-Caribbean men and women. *J Mol Med* 84:1005-1014, 2006

99. Wang J, Kuusisto J, Vänttinen M, Kuulasmaa T, Lindström J, Tuomilehto J, Uusitupa M, Laakso M: Variants of transcription factor 7-like 2 (TCF7L2) gene predict conversion to type 2 diabetes in the Finnish Diabetes Prevention Study and are associated with impaired glucose regulation and impaired insulin secretion. *Diabetologia* 50:1192-1200, 2007

100. Duan Q, Dubé M, Frasure-Smith N, Barhdadi A, Lesperance F, Théroux P, St-Onge J, Rouleau G, McCaffery J: Additive effects of obesity and TCF7L2 variants on risk for type 2 diabetes among cardiac patients. *Diabetes Care* 30:1621-1623, 2007

101. Kimber C, Doney A, Pearson E, McCarthy M, Hattersley A, Leese G, Morris A, Palmer C: TCF7L2 in the Go-DARTS study: evidence for a gene dose effect on both diabetes susceptibility and control of glucose levels. *Diabetologia* 50:1186-1191, 2007

102. Horikoshi M, Hara K, Ito C, Nagai R, Froguel P, Kadowaki T: A genetic variation of the transcription factor 7-like 2 gene is associated with risk of type 2 diabetes in the Japanese population. *Diabetologia* 50:747-751, 2007

103. Schäfer S, Tschritter O, Machicao F, Thamer C, Stefan N, Gallwitz B, Holst J, Dekker J, T'hart L, Nijpels G, van Haeften T, Häring H, Fritsche A: Impaired glucagon-like peptide-1-induced insulin secretion in carriers of transcription factor 7-like 2 (TCF7L2) gene polymorphisms. *Diabetologia*, 2007

104. Ding B: Distribution of the NPY 1128C allele frequency in different populations. *J Neural Transm* 110:1199-1204, 2003

105. Parton L, Ye C, Coppari R, Enriori P, Choi B, Zhang C, Xu C, Vianna C, Balthasar N, Lee C, Elmquist J, Cowley M, Lowell B: Glucose sensing by POMC neurons regulates glucose homeostasis and is impaired in obesity. *Nature* 449:228-232, 2007

106. Coll A: Effects of pro-opiomelanocortin (POMC) on food intake and body weight: mechanisms and therapeutic potential? *Clin Sci (Lond)* 113:171-182, 2007

107. Sutton B, Langefeld C, Williams A, Norris J, Saad M, Haffner S, Bowden D: Association of proopiomelanocortin gene polymorphisms with obesity in the IRAS family study. *Obes Res* 13:1491-1498, 2005

108. Farooqi I, Drop S, Clements A, Keogh J, Biernacka J, Lowenbein S, Challis B, O'Rahilly S: Heterozygosity for a POMC-null mutation and increased obesity risk in humans. *Diabetes* 55:2549-2553, 2006

109. Swanson K, Mazur M, Vashisht K, Rund L, Beever J, Counter C, Schook L: Genomics and clinical medicine: rationale for creating and effectively evaluating animal models. *Exp Biol Med (Maywood)* 229:866-875, 2004

110. Coleman R: Of mouse and man--what is the value of the mouse in predicting gene expression in humans? *Drug Discov Today* 8:233-235, 2003

111. Esmon C: Why do animal models (sometimes) fail to mimic human sepsis? *Crit Care Med* 32:S219-222, 2004

112. Kuhl J, Hilding A, Ostenson C, Grill V, Efendic S, Båvenholm P: Characterisation of subjects with early abnormalities of glucose tolerance in the Stockholm Diabetes Prevention Programme: the impact of sex and type 2 diabetes heredity. *Diabetologia* 48:35-40, 2005