Integrating bioinformatics and physiology to describe genetic effects in complex polygenic diseases

Parikh, Hemang

2009

Link to publication

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Integrating bioinformatics and physiology to describe genetic effects in complex polygenic diseases

ACADEMIC DISSERTATION

Hemang Parikh

Lund University
Department of Clinical Sciences
Diabetes and Endocrinology
Malmö University Hospital

LUND UNIVERSITY
Faculty of Medicine

With the permission of the Medical Faculty of Lund University, to be presented for public examination in the Grand Hall at the Medical Research Center, Entrance 59, Malmö University Hospital, on January 23rd, 2009, at 9:00 a.m.

Faculty Opponent
Professor Björn Carlsson
Department of Molecular and Clinical Medicine
Sahlgrenska University Hospital
University of Gothenburg
Sweden
Type 2 diabetes mellitus (T2DM) results from interaction between genetic and environmental factors. This thesis has focused on dissecting the genetic contribution in T2DM using large-scale genomic approaches with a particular emphasis on analysis of gene transcripts in different tissues, predominantly muscle.

In paper I, we identified TXNIP as a gene whose expression is powerfully suppressed by insulin yet stimulated by glucose. In healthy individuals, its expression was inversely correlated to total body measures of glucose uptake. TXNIP regulates both insulin-dependent and insulin-independent pathways of glucose uptake in human skeletal muscle.

In paper II, we investigated molecular mechanisms associated with insulin sensitivity in skeletal muscle by relating global skeletal muscle gene expression to physiological measurements of insulin sensitivity. We identified 70 genes positively and 110 genes inversely correlated with insulin sensitivity in human skeletal muscle. More specifically, expression of CPT1B was positively correlated with insulin sensitivity. These data suggest that a high capacity of fat oxidation in mitochondria is reflected by a high expression of CPT1B which is a marker of insulin sensitivity.

In paper III, we investigated molecular mechanisms associated with maximal oxygen uptake (VO2max) and type 1 fiber in human skeletal muscle. We identified 66 genes positively and 83 genes inversely correlated with VO2max and 171 genes positively and 217 genes inversely correlated with percentage of type 1 fibers in human skeletal muscle. Genes involved in oxidative phosphorylation (OXPHOS) showed high expression in individuals with high VO2max. These findings indicate that VO2max closely reflects expression of OXPHOS genes.

In paper IV, we combined results from genome-wide association (GWA) studies with genome-wide expression profiling in pancreas, adipose tissue, liver, and skeletal muscle in patients with or without T2DM or animal models thereof to identify novel T2DM susceptibility loci. Most notably, we were able to identify four genes namely IGF2BP2, CDKL1, TSPAN8, and NOTCH2 for which SNPs located in vicinity of these genes have shown association with T2DM in different populations. In addition, we identified a SNP (rs27582) in the CAST gene which was associated with future risk of T2DM. In conclusion, using gene expression in different tissues from T2DM patients is a powerful tool for prioritizing SNPs from GWA studies for replication studies.

Key words: Type 2 diabetes mellitus, Insulin sensitivity, Gene expression profiling, TXNIP, Maximal oxygen uptake, CPT1B, NDUFB5, ATP5C1, AHI1, CAST, Genome-wide association studies

Supplementary bibliographical information:

<table>
<thead>
<tr>
<th>Language</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISSN and key title:</td>
<td>1652-8220</td>
</tr>
<tr>
<td>Recipient’s notes</td>
<td>Number of pages</td>
</tr>
<tr>
<td>Security classification</td>
<td></td>
</tr>
</tbody>
</table>

Distribution by (name and address)

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature  

Date 08/12/18
Integrating bioinformatics and physiology to describe genetic effects in complex polygenic diseases

ACADEMIC DISSERTATION

Hemang Parikh

Lund University
Department of Clinical Sciences
Diabetes and Endocrinology
Malmö University Hospital
“The only thing that interferes with my learning is my education.”

Albert Einstein

To my mother
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of papers</td>
<td>9</td>
</tr>
<tr>
<td>List of papers not included in the thesis</td>
<td>11</td>
</tr>
<tr>
<td>Abbreviations</td>
<td>13</td>
</tr>
<tr>
<td>Abstract</td>
<td>17</td>
</tr>
<tr>
<td>Introduction</td>
<td>19</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>19</td>
</tr>
<tr>
<td>Epidemiology of T2DM</td>
<td>19</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>20</td>
</tr>
<tr>
<td>Risk factors for T2DM</td>
<td>20</td>
</tr>
<tr>
<td>Pathophysiology of T2DM</td>
<td>21</td>
</tr>
<tr>
<td>Skeletal muscle</td>
<td>22</td>
</tr>
<tr>
<td>Muscle fiber types</td>
<td>23</td>
</tr>
<tr>
<td>Pathophysiology of insulin resistance in skeletal muscle</td>
<td>23</td>
</tr>
<tr>
<td>Impaired muscle glycogen synthesis</td>
<td>23</td>
</tr>
<tr>
<td>Free fatty acid induced insulin resistance</td>
<td>24</td>
</tr>
<tr>
<td>Oxidative stress and mitochondrial dysfunction</td>
<td>26</td>
</tr>
<tr>
<td>Physical activity</td>
<td>27</td>
</tr>
<tr>
<td>Maximal oxygen uptake (VO$_2$\text{max})</td>
<td>28</td>
</tr>
<tr>
<td>Methods for evaluating the genetic component of T2DM</td>
<td>29</td>
</tr>
<tr>
<td>Association studies</td>
<td>29</td>
</tr>
<tr>
<td>Gene expression studies</td>
<td>31</td>
</tr>
<tr>
<td>Global gene expression profiling</td>
<td>31</td>
</tr>
<tr>
<td>Affymetrix oligonucleotide arrays</td>
<td>31</td>
</tr>
<tr>
<td>Agilent oligonucleotide arrays</td>
<td>32</td>
</tr>
<tr>
<td>Global gene expression profiling for T2DM</td>
<td>33</td>
</tr>
<tr>
<td>Analysis of microarray data</td>
<td>33</td>
</tr>
<tr>
<td>Pre-processing of microarray data</td>
<td>33</td>
</tr>
<tr>
<td>Analysis of differentially expressed genes</td>
<td>34</td>
</tr>
<tr>
<td>Pathway analysis of microarray data</td>
<td>35</td>
</tr>
<tr>
<td>Genome-wide association studies</td>
<td>35</td>
</tr>
<tr>
<td>Aims</td>
<td>36</td>
</tr>
<tr>
<td>Study participants</td>
<td>37</td>
</tr>
<tr>
<td>Paper I</td>
<td>37</td>
</tr>
<tr>
<td>Paper II</td>
<td>38</td>
</tr>
<tr>
<td>Paper III</td>
<td>38</td>
</tr>
<tr>
<td>Paper IV</td>
<td>38</td>
</tr>
</tbody>
</table>
Methods

Phenotypic characterization 41
Oral glucose tolerance test (OGTT) 41
Laboratory methods 41
Measurements 41
Hyperinsulinemic euglycemic clamp 42
Indirect calorimetry 42
Maximal oxygen uptake (VO_{2\text{max}}) measurement 42
Human adipocyte cell culture and treatment 43
RNA extraction and hybridization 43
Real-time polymerase chain reaction (real-time PCR) 43
RNA interference, viral transduction, and glucose uptake assays 45
Sequencing 45
Genotyping using TaqMan allelic discrimination 45
Genotyping using MassARRAY technology 46
Statistical analysis 46

Results

Paper I: TXNIP regulates peripheral glucose metabolism in humans 49
Paper II: Relationship between insulin sensitivity and gene expression in human skeletal muscle 51
Paper III: Molecular correlates for maximal oxygen uptake and type 1 fibers 52
Paper IV: Prioritizing genes for follow-up from genome wide association studies using information on gene expression in tissues relevant for type 2 diabetes mellitus 55

Discussion

Methodology 59
TXNIP 61
CPT1B 63
NDUFB5 64
ATP5C1 64
AHNAK 64
CAST 65
Future challenges 65

Summary 67

Conclusions 69

References 77
List of papers


Paper I and III are reproduced with the permission from the publishers.
List of papers not included in the thesis


Ridderstrale M, Parikh H, Groop LC: Calpain 10 and type 2 diabetes: are we getting closer to an explanation? Curr Opin Clin Nutr Metab Care 8:361-366, 2005


Shaat N, Lernmark A, Karlsson E, Ivarsson S, Parikh H, Berntorp K, Groop LC: A variant in the transcription factor 7-like 2 (TCF7L2) gene is associated with an

Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>acetyl-CoA</td>
<td>Acetyl coenzyme A</td>
</tr>
<tr>
<td>AHNAK</td>
<td>AHNAK nucleoprotein</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of variance</td>
</tr>
<tr>
<td>ASK1</td>
<td>Apoptosis signal–regulating kinase 1</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>ATP5C1</td>
<td>ATP synthase, H+ transporting, mitochondrial F1 complex, gamma polypeptide 1</td>
</tr>
<tr>
<td>BMI</td>
<td>Body mass index</td>
</tr>
<tr>
<td>BPS</td>
<td>Botnia prospective study</td>
</tr>
<tr>
<td>C/EBP</td>
<td>CCAAT/enhancer binding protein</td>
</tr>
<tr>
<td>CAST</td>
<td>Calpastatin</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary deoxyribonucleic acid</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence interval</td>
</tr>
<tr>
<td>CIR</td>
<td>Corrected insulin response</td>
</tr>
<tr>
<td>CPT1B</td>
<td>Carnitine palmitoyltransferase 1B</td>
</tr>
<tr>
<td>ddNTP</td>
<td>Dideoxynucleotide triphosphate</td>
</tr>
<tr>
<td>DEXA</td>
<td>Dual energy X-ray absorptiometry</td>
</tr>
<tr>
<td>DI</td>
<td>Disposition index</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>ECM</td>
<td>Extracellular matrix</td>
</tr>
<tr>
<td>ES</td>
<td>Enrichment score</td>
</tr>
<tr>
<td>EST</td>
<td>Expressed sequence tag</td>
</tr>
<tr>
<td>FDR</td>
<td>False discovery rate</td>
</tr>
<tr>
<td>FFA</td>
<td>Free-fatty acid</td>
</tr>
<tr>
<td>FH-</td>
<td>Without family history of T2DM</td>
</tr>
<tr>
<td>FH+</td>
<td>With family history of T2DM</td>
</tr>
<tr>
<td>FPG</td>
<td>Fasting plasma glucose concentration</td>
</tr>
<tr>
<td>FRET</td>
<td>Fluorescence resonance energy transfer</td>
</tr>
<tr>
<td>FWER</td>
<td>Family-wise error rate</td>
</tr>
<tr>
<td>G6P</td>
<td>Glucose-6-phosphate</td>
</tr>
<tr>
<td>GC-RMA</td>
<td>GC-content robust multi-array average</td>
</tr>
<tr>
<td>GEE</td>
<td>Generalized estimating equations</td>
</tr>
<tr>
<td>GLUT4</td>
<td>Glucose transporter 4</td>
</tr>
<tr>
<td>GO</td>
<td>Gene ontology</td>
</tr>
<tr>
<td>GRAIL</td>
<td>Gene relationships across implicated loci</td>
</tr>
<tr>
<td>GSEA</td>
<td>Gene set enrichment analysis</td>
</tr>
<tr>
<td>GWA</td>
<td>Genome-wide association</td>
</tr>
</tbody>
</table>
HK  Hexokinase II
HOMA  Homeostasis model assessment
HOMA-IR  Homeostasis model assessment of insulin resistance
HRR  Haplotype relative risk
IFG  Impaired fasting glucose
IGT  Impaired glucose tolerance
IRS-1  Insulin receptor substrate 1
IRS-2  Insulin receptor substrate 2
ISI  Insulin sensitivity index
IVGTT  Intravenous glucose tolerance test
JNK  c-jun N-terminal kinase
KEGG  Kyoto encyclopedia of genes and genomes
KLF5  Krüppel-like transcription factor 5
MALDI-TOF  Matrix-assisted laser desorption/ionization-time of flight
MAS5.0  Microarray suite version 5.0
MBEI  Model based expression index
MIAME  Minimum information about a microarray experiment
MM  Mismatch
MPP  Malmö preventive project
mRNA  messenger ribonucleic acid
mTOR  Mammalian target-of-rapamycin
NDUFB5  NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 5
NGT  Normal glucose tolerance
NMR  Nuclear magnetic resonance
OGTT  Oral glucose tolerance test
OR  Odds ratio
OXPHOS  Oxidative phosphorylation
OXPHOS-CR  A subset of genes involved in oxidative phosphorylation
PCR  Polymerase chain reaction
PDH  Pyruvate dehydrogenase
PFK  Phosphofructokinase
PI3K  Phosphatidylinositol-3 kinase
PKC  Protein kinase C
PLIER  Probe logarithmic intensity error
PM  Perfect match
PPAR  Peroxisome proliferator-activated receptor
PPP  Prevalence, prediction, and prevention of diabetes
RMA  Robust multi-array average
RNA  Ribonucleic acid
RNAi  RNA interference
ROS  Reactive oxygen species
RT-PCR  Real-time polymerase chain reaction
SAM  Significance analysis of microarrays
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>SEM</td>
<td>Standard error of the mean</td>
</tr>
<tr>
<td>siRNA</td>
<td>Short interfering RNA</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td>T2DM</td>
<td>Type 2 Diabetes Mellitus</td>
</tr>
<tr>
<td>TDT</td>
<td>Transmission disequilibrium test</td>
</tr>
<tr>
<td>Tg</td>
<td>Triglyceride</td>
</tr>
<tr>
<td>TNF</td>
<td>Tumor necrosis factor</td>
</tr>
<tr>
<td>TXN</td>
<td>Thioredoxin</td>
</tr>
<tr>
<td>TXNIP</td>
<td>Thioredoxin interacting protein</td>
</tr>
<tr>
<td>UDP</td>
<td>Uridine 5'-diphosphate</td>
</tr>
<tr>
<td>VCAM1</td>
<td>Vascular cell adhesion molecule-1</td>
</tr>
<tr>
<td>VO_{2max}</td>
<td>Maximal oxygen uptake</td>
</tr>
<tr>
<td>WebGestalt</td>
<td>Web-based gene set analysis</td>
</tr>
<tr>
<td>WHO</td>
<td>World health organization</td>
</tr>
<tr>
<td>ZDF</td>
<td>Zucker diabetic fatty</td>
</tr>
<tr>
<td>ZLC</td>
<td>Zucker lean control</td>
</tr>
</tbody>
</table>
Abstract

Type 2 diabetes mellitus (T2DM) results from interaction between genetic and environmental factors. The worldwide prevalence of T2DM is increasing rapidly due to reduction in physical activity, increase in dietary intake, and the aging of the population. This thesis has focused on dissecting the genetic contribution in T2DM using large-scale genomic approaches with a particular emphasis on analysis of gene transcripts in different tissues, predominantly muscle.

In paper I, we identified TXNIP as a gene whose expression is powerfully suppressed by insulin yet stimulated by glucose. In healthy individuals, its expression was inversely correlated to total body measures of glucose uptake. Forced expression of TXNIP in cultured adipocytes significantly reduced glucose uptake, while silencing with RNA interference in adipocytes and in skeletal muscle enhanced glucose uptake, confirming that the gene product is also a regulator of glucose uptake. TXNIP expression is consistently elevated in the muscle of pre-diabetics and diabetics, although in a panel of 4,450 Scandinavian individuals, we found no evidence for association between common genetic variation in the TXNIP gene and T2DM. TXNIP regulates both insulin-dependent and insulin-independent pathways of glucose uptake in human skeletal muscle. Combined with recent studies that have implicated TXNIP in pancreatic β-cell glucose toxicity, our data suggest that TXNIP might play a key role in defective glucose homeostasis preceding overt T2DM.

In paper II, we investigated molecular mechanisms associated with insulin sensitivity in skeletal muscle by relating global skeletal muscle gene expression to physiological measures of the insulin sensitivity. We identified 70 genes positively and 110 genes inversely correlated with insulin sensitivity in human skeletal muscle. Most notably, genes involved in a mammalian target-of-rapamycin signaling pathway were positively whereas genes encoding extracellular matrix structural constituent such as extracellular matrix-receptor, cell communication, and focal adhesion pathways were inversely correlated with insulin sensitivity. More specifically, expression of CPT1B was positively and that of LEO1 inversely correlated with insulin sensitivity, a finding which was replicated in an independent study of 9 non-diabetic men. These data suggest that a high capacity of fat oxidation in mitochondria is reflected by a high expression of CPT1B which is a marker of insulin sensitivity.

In paper III, we investigated molecular mechanisms associated with maximal oxygen uptake (VO₂max) and type 1 fibers in human skeletal muscle. We identified 66 genes positively and 83 genes inversely correlated with VO₂max and 171 genes positively and
217 genes inversely correlated with percentage of type 1 fibers in human skeletal muscle. Genes involved in oxidative phosphorylation (OXPHOS) showed high expression in individuals with high VO\textsubscript{2}\text{max}, whereas the opposite was not the case in individuals with low VO\textsubscript{2}\text{max}. Instead, genes such as AHNAK and BCL6 were associated with low VO\textsubscript{2}\text{max}. Also, expression of the OXPHOS genes, NDUFB5 and ATP5C1, increased with exercise training and decreased with aging. In contrast, expression of AHNAK in skeletal muscle decreased with exercise training and increased with aging. These findings indicate that VO\textsubscript{2}\text{max} closely reflects expression of OXPHOS genes, particularly that of NDUFB5 and ATP5C1 in skeletal muscle and high expression of these genes suggest good muscle fitness. In contrast, a high expression of AHNAK was associated with a low VO\textsubscript{2}\text{max} and poor muscle fitness.

In paper IV, we combined results from the Diabetes Genetics Initiative (DGI) and the Wellcome Trust Case Control Consortium (WTCCC) genome-wide association (GWA) studies with genome-wide expression profiling in pancreas, adipose tissue, liver, and skeletal muscle in patients with or without T2DM or animal models thereof to identify novel T2DM susceptibility loci. We identified 453 single nucleotide polymorphisms (SNPs) associated with T2DM with \( P < 0.01 \) in at least one of the GWA studies and 150 genes that were located in vicinity of these SNPs. Out of these 150 genes, we identified 41 genes differentially expressed using publicly available gene expression profiling data. Most notably, we were able to identify four genes namely IGF2BP2, CDKAL1, TSPAN8, and NOTCH2 for which SNPs located in vicinity of these genes have shown association with T2DM in different populations. In addition, we identified a SNP (rs27582) in the CAST gene which was associated with future risk of T2DM (odds ratio (OR) = 1.10, 95% CI: 1.00-1.20, \( P < 0.05 \)) in a prospective study of 16,061 Swedish individuals followed for more than 25 years; this association was stronger in lean individuals (OR = 1.19, 95% CI: 1.03-1.36, \( P = 0.024 \)). Moreover in the Botnia Prospective Study (BPS) involving 2,770 individuals followed for more than 7 years, carriers of the A-allele were more insulin resistant than carriers of the G-allele as indicated by higher fasting insulin concentrations (regression coefficient (\( \beta \)) = 0.048, \( P = 0.017 \)) and higher HOMA-IR index (\( \beta = 0.044, P = 0.025 \)) as well as lower insulin sensitivity index during OGTT (\( \beta = -0.039, P = 0.039 \)) at follow-up. In conclusion, using gene expression in different tissues from patients with T2DM and animal models is a powerful tool for prioritizing SNPs from GWA studies for replication studies. We thereby identified association of a variant (rs27582) in the CAST gene with T2DM and insulin resistance.

**Keywords:** Type 2 diabetes mellitus, Insulin sensitivity, Skeletal muscle, Gene expression profiling, TXNIP, Maximal oxygen uptake, Type 1 fibers, CPT1B, NDUFB5, ATP5C1, AHNAK, CAST, Genome-wide association studies.
Introduction

About 250 million people suffer from type 2 diabetes mellitus (T2DM) in the world today and prevalence of the disease is increasing worldwide due to reduction in physical activity, increase in dietary intake, and the aging of the population. Several large-scale genomic approaches have been employed to identify genes predisposing to T2DM such as deoxyribonucleic acid (DNA) microarrays and genome-wide association (GWA) studies. DNA microarrays can be used to measure levels of messenger ribonucleic acid (mRNA) in biological samples for more than 20,000 gene transcripts. Moreover, GWA studies offer an unbiased approach to identify genetic variants that influence susceptibility to disease.

This thesis describes the use of these genomic approaches for identifying different sets of genes regulated by insulin, correlated with insulin sensitivity, and associated with maximal oxygen uptake (VO2max) and type 1 fibers in human skeletal muscle. In addition, it presents an approach for prioritizing single nucleotide polymorphisms (SNPs) from GWA studies for further replication in other samples and examination of their pathophysiological role in T2DM etiology.

Diabetes mellitus

Diabetes mellitus is a pathological state in which the blood glucose concentration is chronically raised. It was already described as a disease in the Ebers Papyrus dating back to 1500 BC1. Diabetes mellitus is caused by defects in insulin secretion, insulin action, or both resulting from a complex interaction of genes, and environmental factors2. Diabetes mellitus is broadly classified into four different etiological categories designated as type 1 diabetes mellitus, T2DM, gestational diabetes mellitus, and other specific types3. T2DM accounts for 90-95% of all cases of diabetes mellitus. T2DM is a key risk factor for vascular complications, stroke, kidney failure, neuropathy, blindness, and amputation4.

Epidemiology of T2DM

T2DM is a major health concern today. Prevalence of T2DM is increasing rapidly worldwide and it has been predicted that about 366 million people will be affected by T2DM in the year 20305. There are large ethnic and geographic variations in the prevalence of T2DM5,6. Asians, Hispanics, African-Americans, and Pima Indians appear
to have much higher incidence of T2DM as compared to European populations. Moreover, the “westernized lifestyle” is strongly associated with an increase in the prevalence of T2DM across all ethnic groups. According to Centers for Disease Control and Prevention, incidence of T2DM in the United States has increased by 90% during the last 10 years. Also, it is no longer a disease of the elderly as the age at onset has decreased. It occurs among children and adolescents as well.

**Diagnosis**

The World Health Organization (WHO) diagnostic criteria for diabetes mellitus are based on either measurements of the fasting plasma glucose concentration or a 2-hour plasma glucose concentration after drinking a solution of 75g glucose; oral glucose tolerance test (OGTT) (Table 1). In a symptomless patient the measurements need to be repeated to confirm the diagnosis. Impaired fasting glucose (IFG) and impaired glucose tolerance (IGT) represent early abnormalities of glucose metabolism and are considered as pre-diabetic conditions. Individuals with IFG or IGT have 40% higher risk of developing T2DM.

**Table 1: Criteria for the diagnosis of diabetes mellitus and other categories of hyperglycaemia**

<table>
<thead>
<tr>
<th>Glucose (mmol/l)</th>
<th>Impaired fasting glucose</th>
<th>Impaired glucose tolerance</th>
<th>Diabetes mellitus</th>
</tr>
</thead>
<tbody>
<tr>
<td>FPG</td>
<td>≥ 6.1 to &lt; 7.0</td>
<td>-</td>
<td>≥ 7.0</td>
</tr>
<tr>
<td>2 hours after OGTT</td>
<td>-</td>
<td>≥ 7.8 to &lt; 11.1</td>
<td>≥ 11.1</td>
</tr>
</tbody>
</table>

Abbreviations: FPG, fasting plasma glucose concentration; OGTT, oral glucose tolerance test.

**Risk factors for T2DM**

Risk factors for development of T2DM comprise of genetic (family history of the disease) as well as non-genetic factors such as low birth weight, obesity, excess food intake, physical inactivity, and aging. Hence, T2DM is considered as a multifactorial polygenic disease where common variations in several genes interact (epistasis) to cause the disease especially when exposed to the affluent environment of too much food and too little exercise. Evidence for genetic susceptibility to T2DM comes from studies demonstrating difference in incidence of T2DM between different populations as well as twins studies. Moreover, the lifetime risk for developing T2DM in offspring of one T2DM parent is about 40% and the ratio of the risk for a sibling of a T2DM individual to that of the general population ($\lambda_s$) is about 3.5.
Pathophysiology of T2DM

T2DM is characterized by impaired insulin secretion and action, the latter manifested as peripheral and hepatic insulin resistance (poor suppression of excessive glucose production from the liver by insulin)\textsuperscript{18}. Insulin resistance impairs glucose uptake by insulin sensitive tissues (e.g., skeletal muscle, liver, and adipose tissue) and increases hepatic glucose production; these effects contribute to elevation in glucose concentration and increases ambient plasma free-fatty acids (FFAs) concentration\textsuperscript{19}. The FFAs are the main form in which energy is transferred from stores in adipose tissue to other sites in the body for metabolic utilization. However, elevated levels of FFAs can impair glucose utilization in skeletal muscle, promote glucose production by the liver, and impair pancreatic β-cell function\textsuperscript{20}. The pancreatic β-cell is initially able to compensate by secreting more insulin to maintain normal glucose levels. As insulin resistance and compensatory hyperinsulinemia progress, the pancreatic β-cell compensation fails. A decline in insulin secretion and an increase in hepatic glucose production lead to overt T2DM (Figure 1)\textsuperscript{21}.

**Figure 1: Pathophysiology of T2DM.**

Overeating and physical inactivity along with pre-disposed genetic risk and aging can leads to insulin resistance. In insulin resistant state hepatic glucose production along with storage of triglycerides (Tg) are increased while glucose uptake is decreased in skeletal muscle and adipose
tissue which leads to mild hyperglycemia and higher plasma free-fatty acids (FFAs) concentration; which results in hyperinsulinemia because small increases in glucose concentration stimulate $\beta$-cells to overproduce insulin to restore normal glycemic levels but eventually impaired insulin secretion from pancreatic $\beta$-cells and hyperglycemia leads to T2DM.

Skeletal muscle

About 40% of the body's mass is composed of skeletal muscle. Muscle is made up of fascicule which are grouped together and surrounded by a connective tissue called epimysium. The fasciculus contains muscle fibers which are made up of many myofibrils that are highly organized bundles of long polymers of myosin and actin proteins (Figure 2).

Figure 2: Basic structure of skeletal muscle.

Each end of the muscle is connected to a tendon, which connects to the bone mediating the body's lever system. The muscle is made up of fascicule containing the muscle fibers, which are made up of myofibrils and ultimately the myofilaments actin and myosin. This figure is adapted from Kraemer WJ et al.
Muscle fiber types

Muscle fibers are classified based upon their functional capabilities and enzymatic profiles. Muscle fibers are broadly classified into two distinct fiber types: (1) Type 1 (slow-twitch) muscle fiber and (2) Type 2 (fast-twitch) muscle fibers; the latter further subdivided into type 2A and type 2B. Type 1 fibers contain many relatively large size mitochondria and generate energy for adenosine triphosphate (ATP) synthesis by aerobic energy transfer. Type 2 fibers contain fewer mitochondria and possess rapid contraction speed and high capacity for anaerobic ATP production in glycolysis. Type 2A fibers are considered as fast-twitch oxidative glycolytic fibers while type 2B fibers are considered as anaerobic glycolytic fibers which possess the greatest potential for anaerobic energy transfer. A low proportion of type 1 fibers and high proportion of type 2B fibers is associated with decreased mitochondrial function/numbers in the insulin-resistant offspring of T2DM patients and predicts T2DM.

Pathophysiology of insulin resistance in skeletal muscle

Insulin resistance is a key feature of the pre-diabetic state and a predictor of T2DM. Insulin resistance can be observed 10–20 years before the onset of overt T2DM. It is also seen in healthy glucose tolerant first degree relatives of patients with T2DM. Insulin resistance in skeletal muscle has been attributed to different pathological conditions such as impaired glycogen synthesis, accumulation of lipids with subsequent impairment of insulin signaling, and mitochondrial dysfunction and oxidative stress. However, the underlying molecular mechanisms associated with insulin resistance are still poorly understood. In an attempt to elucidate the underlying pathophysiology, we have investigated the molecular mechanisms associated with insulin sensitivity in skeletal muscle in paper II. Moreover, we have studied novel mechanisms of skeletal muscle insulin resistance and have identified mediators of glucose homeostasis in paper I.

Impaired muscle glycogen synthesis

Skeletal muscle is the principal site of glucose uptake in the postprandial state. After entering the cell, glucose is phosphorylated by hexokinase and either stored as glycogen via the activation of glycogen synthase, or catabolized to generate ATP via glycolysis and oxidative phosphorylation (OXPHOS) in the mitochondria. Using nuclear magnetic resonance (NMR) spectroscopy, which measures the concentration of intracellular metabolites noninvasively, it has been shown that impaired muscle glycogen synthesis plays a major role in causing insulin resistance in T2DM patients. Impaired glucose transport and/or hexokinase activity may account for the impairment of muscle glycogen synthesis observed in insulin-resistant subjects including obese individuals, T2DM patients, and their offspring.
Figure 3: The pathway of muscle glycogen synthesis.

Uptake of glucose into the muscle cell occurs via glucose transporter 4 (GLUT4), whereupon it is phosphorylated by hexokinase to glucose-6-phosphate (G6P). After isomerization to G1P and activation to uridine 5’-diphosphate (UDP)-glucose, the final step is polymerization into glycogen by glycogen synthase. This figure is adapted from Petersen KF et al.37

Free fatty acid induced insulin resistance

Increased plasma FFAs concentration is strongly associated with insulin resistance. In a classic series of studies, Randle and colleagues demonstrated that FFAs compete with glucose for substrate oxidation which leads to decreased glucose oxidation when FFAs are elevated40. They proposed that an increase in FFAs causes elevation of the intramitochondrial acetyl-coenzyme A (acetyl-CoA)/CoA and NADH/NAD⁺ ratios with subsequent inactivation of pyruvate dehydrogenase. This event in turn would cause an increase in intracellular citrate concentrations, which leads to inhibition of phosphofructokinase and subsequent accumulation of glucose-6-phosphate. The resulting increased concentration of glucose-6-phosphate would inhibit hexokinase II activity and increase intracellular glucose concentration and decrease glucose uptake (Figure 4).
An increase in fatty acid concentration results in an elevation of the intramitochondrial acetyl-CoA/CoA and NADH/NAD⁺ ratios, with subsequent inactivation of pyruvate dehydrogenase (PDH). This in turn causes citrate concentrations to increase, leading to inhibition of phosphofructokinase (PFK). Subsequent increases in intracellular glucose-6-phosphate concentration would inhibit hexokinase II (HK) activity, which would result in an increase in intracellular glucose concentration and a decrease in muscle glucose uptake. This figure is adapted from Shulman GI et al.41

However, Shulman and colleagues have challenged this hypothesis by a series of studies which employed in-vivo tracer and NMR. They demonstrated that elevated plasma FFAs concentrations cause a reduction in intracellular glucose concentration. After entering the muscle cell, FFAs promote the intramyocellular accumulation of fatty acid-derived metabolites such as fatty acyl-CoA and diacylglycerol which leads to impaired insulin signaling via the activation of protein kinase C (PKC) isoforms (Figure 5). This event in turn promotes a serine/threonine phosphorylation cascade and increases serine phosphorylation of the insulin receptor substrate 1 and 2 (IRS-1 and IRS-2), which in turn leads to decreased tyrosine phosphorylation of IRS-1, decreased activity of phosphatidylinositol-3 kinase (PI3K), and ultimately decreased translocation of glucose transporter 4 (GLUT4) and glucose transport. Also, FFAs produce low-grade inflammation by activating nuclear factor-kappa beta which leads to release of several proinflammatory and proatherogenic cytokines.

Figure 4: Mechanism of fatty acid–induced insulin resistance in skeletal muscle as proposed by Randle et al.40
Figure 5: Proposed alternative mechanism for fatty acid–induced insulin resistance in human skeletal muscle.

An increase in delivery of fatty acids to muscle or a decrease in intracellular metabolism of fatty acids leads to an increase in intracellular fatty acid metabolites such as diacylglycerol, fatty acyl-CoA, and ceramides. These metabolites activate a serine/threonine kinase cascade (possibly initiated by protein kinase C (PKC) isoforms) leading to phosphorylation of serine/threonine sites on insulin receptor substrates (IRS-1 and IRS-2), which in turn reduces the ability of the insulin receptor substrates to activate phosphatidylinositol-3 kinase (PI3K). As a consequence, glucose transport activity and other events downstream of insulin receptor signaling are diminished. This figure is adapted from Shulman GI et al.41

Oxidative stress and mitochondrial dysfunction

Oxidative stress is defined as a persistent imbalance between the production of highly reactive molecular species (such as superoxide and hydrogen peroxide) and antioxidant defenses48. Mitochondria play a key role in energy homeostasis by metabolizing nutrients and producing ATP and heat and thereby supply more than 95% of the total energy required for the cell. Imbalance between energy intake and expenditure leads to mitochondrial dysfunction, characterized by a reduced mitochondrial oxidative capacity49. Inefficient nutrient oxidation and a low ratio of ATP production/oxygen consumption result in an increased formation of reactive oxygen species (ROS) and ultimately lead to oxidative stress34. Mitochondrial dysfunction and oxidative stress are associated with insulin resistance34,50-55 (Figure 6). Also, the quantity and morphology of mitochondria are correlated with mitochondrial oxidative capacity56. Fewer and
smaller-sized mitochondria are observed in skeletal muscle of T2DM patients\textsuperscript{57,58}. Mitochondrial dysfunction predisposes to intramyocellular lipid accumulation, which provides a link to insulin resistance\textsuperscript{33} through the mechanisms described earlier in Figure 5.

In addition, several defects in insulin signaling are associated with insulin resistance\textsuperscript{36}.

**Figure 6:** Mechanism of mitochondrial dysfunction.

Excess food intake, including overloaded free-fatty acids (FFAs) or hyperglycemia conditions, increases reactive oxygen species (ROS) production and reduces mitochondrial biogenesis, causing mitochondrial dysfunction. Mitochondrial dysfunction leads to decreased \(\beta\)-oxidation and adenosine triphosphate (ATP) production and increased ROS production, resulting in insulin resistance and T2DM. This figure is adapted from Kim JA \emph{et al.}\textsuperscript{34}

**Physical activity**

Physical activity is defined as body movement produced by muscle action which substantially contributes to energy expenditure\textsuperscript{59}. Physical inactivity is an independent risk factor for T2DM\textsuperscript{60,61}. Also, several prospective studies have concluded that physically active individuals have a 33-50\% lower risk of developing T2DM\textsuperscript{62-64}. Physical activity has been shown to increase glycogen synthase activity, muscle mass, and mitochondrial biogenesis, leading to improvement in insulin sensitivity (Figure 7)\textsuperscript{34,65-67}. In addition, physical activity reduces the risk for a wide spectrum of other diseases including cardiovascular disease, colon cancer, breast cancer, prostate cancer, osteoporosis, and depression\textsuperscript{68}. 

27
Figure 7: Improvement of mitochondrial function by exercise can improve insulin sensitivity, which leads to normal metabolism. This figure is adapted from Kim JA et al.34

Maximal oxygen uptake (VO$_{2\text{max}}$)

VO$_{2\text{max}}$ is defined as the highest oxygen uptake achievable by an individual for a given exercise profile and is widely used as a measure of physical fitness. Moreover, the VO$_{2\text{max}}$ provides useful information about long-term energy system capacity. Many factors influence VO$_{2\text{max}}$, e.g., physical activity, heredity, gender, body composition, and age. Several studies have reported heritability of VO$_{2\text{max}}$ ranging from 40 to 70%69-72. Untrained individuals with T2DM have been shown to have lower VO$_{2\text{max}}$ compared to untrained healthy control individuals. Also, a low VO$_{2\text{max}}$ predicts T2DM73,74. There are several potential mechanisms by which a low VO$_{2\text{max}}$ could increase risk for T2DM such as mitochondrial dysfunction75, a change in muscle fiber type25, and insulin resistance76. In addition, there is a relatively strong positive correlation between insulin sensitivity and VO$_{2\text{max}}$77, and exercise training can increase both VO$_{2\text{max}}$ and insulin sensitivity78. Intriguingly, one study suggested that after exercise, VO$_{2\text{max}}$ correlated with insulin sensitivity only in individuals without but not in individuals with a family history of T2DM, suggesting that in people genetically predisposed to T2DM, an increase in VO$_{2\text{max}}$ does not translate into improved insulin sensitivity79. Moreover, a high VO$_{2\text{max}}$ is associated with a high proportion of type 1 fibers80. We have also investigated molecular mechanisms associated with VO$_{2\text{max}}$ and type 1 fibers in skeletal muscle in paper III.
Methods for evaluating the genetic component of T2DM

The human genome comprises about 3 billion base pairs and only 0.1% of it contains polymorphic changes, which includes SNP, insertions, deletions, and repetitive elements of variable copy number. The most frequent type of these changes is a SNP which is a difference between chromosomes in the base present at a particular site in the DNA sequence. Identifying the inherited contribution to human disease involves linking genomic variation to clinical phenotype. However, the identification of this relationship is challenging for complex diseases such as T2DM because of interactions between genes and between genes and environment, heterogeneity, and limited statistical power due to insufficient number of samples. Several approaches are being used in the search for genes predisposing to T2DM. The candidate gene approach aims at the identification of genes based upon information of their biochemical function. The random gene search assumes no knowledge of the underlying defects. Positional cloning aims at localizing the disease gene on the basis of its position in the genome.

Association studies

Association studies are performed in case-control samples to evaluate if the specific allele is more frequent in a group of cases than in ethnically matched control subjects. To avoid the population-stratification problem in case-control studies, association studies can be performed in family-based samples by estimating haplotype relative risk (HRR) or using transmission disequilibrium test (TDT). Collecting a larger number of family-based samples is challenging for late onset disease such as T2DM as many subjects' parents might not be alive.
A Case-control association studies

Unrelated affecteds         Matched controls

B Family-based association studies

C Haplotype relative risk method (HRR)

D Transmission disequilibrium test (TDT)

Figure 8: Association studies.

(A) Case–control association studies compare allele frequencies between a patient cohort and an ethnically matched control cohort. (B) Family-based association studies overcome the problem of ethnic matching by ascertaining patients as well as their parents. Non-transmitted alleles from parents can be used as controls. (C) In the haplotype relative risk method (HRR), the allele from each parent that is not present in the patient is used as a control. (D) In the transmission disequilibrium test (TDT), each allele, in this example allele A, is tested separately. Only parents heterozygous for allele A are relevant. A heterozygous parent is expected to give allele A to 50% of his/her offspring; however, if allele A is a predisposing factor, it is transmitted more often than expected to affected offspring (4/4 times = 100%, in this small example). This figure is adapted from Burmeister M et al.88
**Gene expression studies**

Gene expression is regulated at multiple levels including transcription, ribonucleic acid (RNA) processing, mRNA transport, mRNA translation, and mRNA degradation. The amount of a protein that a cell expresses depends on many different parameters such as tissue, the developmental state of the organism, and the metabolic state of the cell. Gene expression studies can be used for gene discovery, detecting gene function or modifications, drug discovery, and identification of potentially important genes involved in the pathogenesis of a disease. Several techniques are available to measure expression of individual gene (i.e., real-time polymerase chain reaction (RT-PCR) and northern blot) or thousands of genes simultaneously (i.e., serial analysis of gene expression and DNA microarray).

**Global gene expression profiling**

Global gene expression profiling provides a powerful and unbiased approach for examining the expression of tens of thousands of gene transcripts simultaneously. DNA microarrays constitute a miniaturized, ordered arrangement of nucleic acid fragments derived from individual genes located at defined positions on a solid surface which enables the analysis of thousands of genes in parallel by specific hybridization. There are mainly two kinds of DNA microarrays: (1) complementary-DNA (cDNA) arrays and (2) oligonucleotide arrays. In the cDNA arrays, about 100-5000 base long probes (a labeled single-stranded DNA molecule of specific base sequence) are spotted on a glass slide by microrobots called DNA arrayers. In oligonucleotide arrays, in situ synthesis produces the short 25-mers probes by photolithography (Affymetrix oligonucleotide arrays) or the long 60-mers probes by inkjet technology (Agilent oligonucleotide arrays). The key difference between the two microarray formats is the length of the probe as in oligonucleotide arrays the probes are of constant length while in cDNA arrays the lengths of the probes differ between spots. Hence, in oligonucleotide arrays fluorescence intensities between different spots are comparable, whereas in cDNA arrays the fluorescence intensity of each spot from the samples always requires to be compared to controls. There are several platforms available for global gene expression profiling. We have used Affymetrix and Agilent oligonucleotide arrays in our studies.

**Affymetrix oligonucleotide arrays**

Affymetrix oligonucleotide arrays are manufactured based on two techniques: photolithography and solid-phase DNA synthesis. The probes are selected from a region 600 bases proximal to the polyadenylation site (3' end) of gene, cDNA, or expressed sequence tags (ESTs) (Figure 9). The probes are designed based on a set of empirically derived parameters and selected to hybridize with high affinity and specificity. Typically 11 probe pairs (consisting a perfect match (PM) probe and a corresponding mismatch (MM) probe) defined as a probeset, are selected from the gene region (Figure 9). The MM probe has a single base mismatch in the middle of the
sequence unlike the PM probe sequence. The MM is designed to provide measures of background and non-specific hybridization and aids in determining if hybridization to the PM is specific.

![Fluorescence intensity image](image)

**Figure 9: Expression probe and array design.**

Oligonucleotide probes are chosen based on uniqueness criteria and composition design rules. For eukaryotic organisms, probes are chosen typically from the 3’ end of the gene or transcript (nearer to the poly(A) tail) to reduce problems that may arise from the use of partially degraded mRNA. The use of the perfect match (PM) minus mismatch (MM) differences averaged across a set of probes greatly reduces the contribution of background and cross-hybridization and increases the quantitative accuracy and reproducibility of the measurements. This figure is adapted from Lipshutz RJ et al.95

**Agilent oligonucleotide arrays**

The Agilent oligonucleotide arrays are manufactured using an in situ-synthetic scheme based on ink-jet process99-101. The long 60-mer probes are designed in silico using empirical measurements that increases hybridization sensitivity100,102. The 60-mers provide certain advantages as only one 60-mers probe per gene or transcript is required compared with 11 perfect match and mismatch probe pairs for each probe set used by Affymetrix. This makes the Agilent array somewhat more tolerant to sequence mismatches and increases sensitivity partly due to larger area available for hybridization91. Nevertheless, there is a strong concordance between these two platforms103.
Global gene expression profiling for T2DM

Several global gene expression profiling studies have been performed in the field of T2DM research. Using this approach, expression of genes regulating OXPHOS pathway, including their master regulator, PPARGC1A, has been shown to be downregulated in skeletal muscle of patients with T2DM\textsuperscript{104} and their non-diabetic first-degree relatives\textsuperscript{105}. Also, expression of genes from this pathway in particular from the electron transport chain, was downregulated in visceral adipose tissue of patient with T2DM independently of obesity\textsuperscript{106}. Gene expression has also been studied in pancreatic islets from T2DM patients although the numbers for obvious reasons (islets from islet transplantation programs) have been small. One study compared global gene expression profiling in pancreatic islets isolated from five humans with T2DM to seven normal glucose-tolerant islet donors and identified decreases in expression of transcription factor ARNT, HNF4alpha, IRS2, Akt2, and genes involved in glucose metabolism\textsuperscript{107}.

Analysis of microarray data

Microarray data analysis can be sub-divided into three categories: (1) Pre-processing of microarray data, (2) Analysis of differentially expressed genes, and (3) Pathway analysis of microarray data.

Pre-processing of microarray data

Pre-processing of microarray data involves several processes namely image analysis, background adjustment, and normalization. Image analysis is used to summarize the information in each spot and to estimate the level of expression of each probe from the pixel intensities in the scanned images\textsuperscript{108}. Background adjustment is employed to adjust observed probe intensities for non-specific hybridization and noise in the optical detection system\textsuperscript{109}. Normalization compensates for non-biological variations between experiments due to different efficiencies of reverse transcription, labeling, or hybridization reactions, physical problems with the arrays, reagent batch effects, and laboratory conditions. For Affymetrix array data, summarization is also needed to obtain an overall measure of RNA transcripts by combining the background adjusted and normalized intensities of multiple probes. A substantial number of the Affymetrix array’s original probe groupings and mappings are shown to be inaccurate\textsuperscript{110}. Several approaches such as Affyproberminer\textsuperscript{110} or the Entrez Gene database based can be applied to regroup the individual probes into more consistent probesets and remap these to the correct sets of genes.

Several statistical methods are available for pre-processing of Affymetrix arrays\textsuperscript{111}. Affymetrix’s microarray suite version 5.0 (MAS5.0) performs ideal mismatch based background adjustment or adjusts for background noise level using estimates of the
distribution of probe intensities, scaling based normalization, and one-step Tukey Biweight based summarization\textsuperscript{112}. The model based expression index (MBEI) uses the invariant set normalization method to normalize arrays and implements a model to estimate the background subtraction\textsuperscript{113,114}. The robust multi-array average (RMA) method estimates the background adjustment by convoluting the signal and noise distributions from the PM probe intensities and performs quantile normalization and summarization based on a multi-array model fit using the median polish algorithm\textsuperscript{109,115}. The GC-content robust multi-array average (GC-RMA) is a modified RMA procedure and includes the additional background adjustment using sequence information to estimate probe affinity for non-specific binding\textsuperscript{116}. The probe logarithmic intensity error (PLIER) method (Affymetrix) estimates the measure of RNA transcript for the entire probe set by utilizing probe affinities, empirical probe performance, and by handling error appropriately across low and high concentrations and implements quantile based normalization. The variance stabilizing normalization\textsuperscript{117}, position-dependent nearest neighbor method\textsuperscript{118}, factor analysis for robust microarray summarization\textsuperscript{119}, and several other methods\textsuperscript{111} can also be employed for pre-processing of Affymetrix microarray data. The model-based algorithms which integrate information from multiple arrays to calculate the expression of a gene are expected to provide more reproducible results\textsuperscript{120}. However, methods used for pre-processing can have a major impact on the results\textsuperscript{121} and one study has reported only 27 to 36\% overlap between different methods\textsuperscript{120}. Therefore, the best approach might be to compare results from several of these methods to obtain a reliable list of differentially expressed genes\textsuperscript{122}.

For the two-color Agilent arrays, simple loess within array normalization without background subtraction method has shown to produce consistent results\textsuperscript{123}. Moreover, data filtering should be conducted to eliminate data that are not reliably detected over all samples\textsuperscript{124,125}.

**Analysis of differentially expressed genes**

Several statistical methods can be applied to identify differentially expressed genes. For interclass two group comparison with an adequate number of replicate samples, the Student’s \textit{t}-test or the Wilcoxon rank-sum test or setting a simple fold change cut-off (typically between 1.5 to 3.0) can be used. However, global gene expression profiling studies involve the simultaneous testing of thousands of statistical hypotheses. Hence, approaches based on false discovery rate (FDR) (which is defined as the expected proportion of false positives among the tests where the null hypothesis was rejected) are adapted to provide correction for multiple hypothesis testing\textsuperscript{126-128}. The significance analysis of microarrays (SAM) algorithm is based on a modified \textit{t}-test and estimates the FDR using permutations under the assumption that all null hypotheses are true\textsuperscript{129}. The patterns from gene expression\textsuperscript{130}, linear models and empirical Bayes\textsuperscript{131}, weighted analysis of general microarray experiments\textsuperscript{132}, rank product method\textsuperscript{133}, data-adaptive test statistics\textsuperscript{134}, integrating differing statistics via a distance synthesis\textsuperscript{135}, and several other methods\textsuperscript{136} can also be employed to detect differentially expressed genes. The selection of differentially expressed genes heavily depends on the choice of the methods used and
performance of these methods is affected by sample size, distributional assumption, and variance structure\textsuperscript{137}. Hence, it is advisable to explore the characteristic of the data first and then apply the most appropriate testing method under the given situation\textsuperscript{137}. Spearman's or Pearson's correlation coefficients and corresponding hypothesis testing procedures can be used to examine the association of each gene's expression with the quantitative phenotype. Spearman partial correlation or Pearson partial correlation or regression analysis can be used to explore the association after adjustment for covariates.

Pathway analysis of microarray data

Pathway analysis of microarray data evaluates gene expression profiles of \textit{a priori} defined biological pathways in association with disease phenotype. The gene set enrichment analysis (GSEA) method calculates a score (called an enrichment score (ES)) for a given gene set based on rank of genes and infers statistical significance of each ES against ES background distribution calculated by permuting the class labels, which preserves gene-gene correlations and provides a more accurate null model\textsuperscript{138}. Moreover, web-based gene set analysis (WebGestalt) identifies whether a group of differentially expressed genes is enriched in Kyoto encyclopedia of genes and genomes (KEGG) pathway or gene ontology (GO) database (\textit{i.e.}, biological process, molecular function, and cellular component) by using overlap statistics such as the cumulative hypergeometric distribution\textsuperscript{139}. Several other methods are also available for inferring GO categories from a group of differentially expressed genes\textsuperscript{140-142}. However, there are a number of methodological issues to consider in the interpretation of the results based on pathway analysis\textsuperscript{143}.

In addition, global gene expression profiling data can also be use for class prediction and class discovery\textsuperscript{144}.

Genome-wide association studies

GWA studies offer a great potential to identify genetic polymorphisms contributing to complex human diseases by examining association of more than 500,000 SNPs with disease\textsuperscript{145}. Several platforms are available for GWA studies including the Affymetrix Genechip\textsuperscript{®} and Illumina BeadChip platform\textsuperscript{146}. Several statistical issues need to be considered in the analysis of such large-scale genetic studies\textsuperscript{137,148}. A number of GWA studies have identified novel genomic regions influencing risk for T2DM\textsuperscript{149-153}. However, the challenge remains to prioritize SNPs from GWA studies for further replication in other samples\textsuperscript{146}; even if one would try to replicate only the top 1% with strongest genetic evidence at least 5,000 SNPs need to be genotyped in replication studies for GWA studies including 500,000 SNPs. Genetic variation can influence gene expression by altering transcript stability or rate of transcription or splicing\textsuperscript{154}. Hence, we have presented an approach for prioritizing SNPs from GWA studies by combining the results from these studies with genome-wide expression profiling of target tissues (\textit{i.e.} pancreas, adipose tissue, liver, and skeletal muscle) of T2DM in paper IV.
Aims

The aim of this thesis was to try to dissect the genetics of T2DM using large-scale genomic approaches with a particular focus on analysis of gene transcripts in different tissues, predominantly muscle. The specific aims were:

I. To identify mediators of glucose homeostasis and to investigate whether these mediators are causally involved in the development of T2DM.

II. To investigate molecular mechanisms associated with insulin sensitivity in skeletal muscle by relating global skeletal muscle gene expression to physiological measures of insulin sensitivity.

III. To investigate molecular mechanisms associated with VO\textsubscript{2max} and type 1 fibers in human skeletal muscle.

IV. To combine GWA studies with genome-wide expression profiling of target tissues (i.e. pancreas, adipose tissue, liver, and muscle) of T2DM in order to identify novel T2DM susceptibility loci and to understand their pathophysiological role in T2DM etiology.
Study participants

Paper I

To identify genes whose expression is responsive to the action of insulin in human skeletal muscle, participants from three different clinical studies (paper I-studies A, B, and C) were included. Paper I-study A consisted of 6 non-diabetic men (age 55 ± 2 years, body mass index (BMI) 32 ± 1 kg/m²), paper I-study B consisted of 6 non-diabetic volunteers (4 men and 2 women, age 66 ± 3 years, BMI 28 ± 2 kg/m²), and paper I-study C consisted of 96 young non-diabetic twins (28 monozygotic and 20 dizygotic pairs, 56 men and 40 women, age 28 [26–29] years, BMI 24 [22–25] kg/m²).

In paper I-studies A and C, skeletal muscle biopsies from the vastus lateralis muscle were obtained before and after a 2-hour hyperinsulinemic euglycemic clamp (Figure 10). Indirect calorimetry was performed during the basal and insulin-stimulated steady state periods to estimate rates of glucose and lipid oxidation. Also, an intravenous glucose tolerance test (IVGTT) was performed prior to the clamp and indirect calorimetry was performed in the basal state as well as at the end of the clamp.

![Figure 10: Design of the hyperinsulinemic, euglycemic clamp protocol.](image)

Muscle biopsies were obtained before and after a hyperinsulinemic euglycemic clamp. An intravenous glucose tolerance test (IVGTT) was administered prior to the clamp and indirect calorimetry was performed in the basal state as well as at the end of the clamp.
For mutation screening of the *thioredoxin interacting protein* (*TXNIP*), we included 48 individuals, 23 of whom showed expression of *TXNIP* in muscle in the highest quartile, while 25 had *TXNIP* expression in lowest quartile from our previous studies\(^\text{104,155}\).

To study the association between *TXNIP* variants and T2DM, we included 4,450 individuals. The sample population consisted of 333 Scandinavian parent–offspring trios, 1,189 discordant Scandinavian sib-pairs, and 1,094 case-control pairs (969 from Scandinavia and 125 from Canada). All case-control pairs were individually matched for sex, age, region of origin, and BMI as previously described\(^\text{156}\).

**Paper II**

To identify genes which were correlated with insulin sensitivity in participants from two different studies (paper II-studies A, and B) were included. Paper II-study A consisted of 39 non-diabetic men (age 38 [35–41] years, BMI 28 [27–30] kg/m\(^2\)) from Malmö, Sweden. To replicate the findings from paper II-study A, we studied an additional 10 young non-diabetic men without any family history of diabetes in paper II-study B (age 24–27 years, BMI 25 [23–26] kg/m\(^2\)). Two participants (one from each paper II-studies A and B) with extreme values of insulin sensitivity (more than 1.5 * interquartile range) were excluded from the analyses.

**Paper III**

To identify genes correlated with VO\(_{2\text{max}}\) and type 1 fibers, participants from two different studies (paper III-studies A, and B) were included. Paper III-study A consisted of 43 age-matched men (age 66 [65–67] years, BMI 25 [23–28] kg/m\(^2\)) from our previously published studies\(^\text{104,157}\). To replicate the findings from paper III-study A and to examine the influence of aging on gene expression, we included the young twins from paper I-study C as well as 78 elderly non-diabetic twins in paper III-study B (18 monozygotic and 21 dizygotic pairs, 34 men and 44 women, age 63 [60–64] years, BMI 26 [24–29] kg/m\(^2\))\(^\text{158}\).

The muscle biopsies were obtained from the vastus lateralis muscle under local anesthesia from the participants using a modified Bergström needle in paper I, II, and III\(^\text{159}\).

**Paper IV**

To prioritize SNPs from GWA studies for further replication, we analyzed whether genes in their neighborhood showed differences in their expression in published gene expression profiling studies in key tissues (*i.e.* pancreas, adipose tissue, liver, and skeletal
muscle). In addition, we analyzed the expression patterns in whole pancreas of 3-weeks old NIDDM1I (which is a well-characterized insulin-deficient congenic strain derived from the Goto-Kakizaki rat, an animal model for T2DM) and control F344 rats.

To replicate the key findings, we have included participants from several studies namely the Malmö case-control study, two prospective cohorts (Malmö Preventive Project (MPP) and Botnia Prospective Study (BPS)). The Malmö case-control study consisted of 2,830 T2DM patients from the Malmö Diabetes Registry and 3,550 controls from the Malmö Diet and Cancer study. All T2DM patients were of Scandinavian origin, had age at onset > 35 years, connecting peptide ≥ 0.3 nmol/l and no glutamic acid decarboxylase antibody. The MPP study included 16,061 participants, 2,263 of whom developed T2DM during a median follow-up period of 24.4 years. The BPS included 2,770 participants, 138 of whom developed T2DM during a median follow-up period of 7.6 years. 2,328 non-diabetic individuals in the BPS with available longitudinal measurements of insulin secretion and action were included in the analyses.

All studies were approved by local ethics committees and all participants gave their informed consent for participation. All studies were conducted in accordance with the principles of the Helsinki Declaration.
Methods

Phenotypic characterization

The participants’ body weight and height were measured in light clothing and BMI was calculated as weight (kg) divided by height-squared (m²). Body composition (lean body mass and fat mass) was determined by dual energy X-ray absorptiometry (DEXA) scanning using a Norland XR-36 scanner or electrical bioimpedance. Glucose tolerance was determined by a 75g OGTT and classified in accordance with WHO criteria

Oral glucose tolerance test (OGTT)

The OGTT is a method to determine how quickly glucose is cleared from the blood. It is a widely used technique for diagnosis of diabetes mellitus. After an overnight fast, the subject is given a solution of 75g glucose to drink. Blood samples are then collected at certain time-points for analysis of glucose and insulin.

Laboratory methods

All laboratory specimens were obtained after overnight fasting. Plasma glucose concentrations were analyzed using an automated glucose oxidation method (Glucose Analyzer 2, Beckman instruments, Fullerton, CA). Plasma insulin concentrations were measured using an immunoassay (Delfia, Turku, Finland).

Measurements

Insulin sensitivity index (ISI) from the OGTT was calculated as 10,000/√(fasting plasma-glucose x fasting plasma-insulin x mean OGTT glucose x mean OGTT insulin). The basal insulin resistance index (HOMA-IR) was calculated from fasting insulin and glucose concentrations. β-cell function was assessed as corrected insulin response during OGTT (CIR = 100 x insulin at 30 min or 40 min in MPP/[glucose at 30 min or 40 min in MPP -3.89])) or as disposition index, i.e. insulin secretion adjusted for insulin sensitivity (DI = CIRxISI).
Hyperinsulinemic euglycemic clamp

The hyperinsulinemic euglycemic clamp is a gold standard method for investigating and quantifying insulin resistance. This technique measures the amount of glucose needed to compensate for an increased insulin level without causing hypoglycemia. After an overnight fast, a specific dose of insulin (40 mU/m²/min) was infused intravenously at a constant rate over a 2 to 3 hour period. In order to compensate for the insulin infusion, glucose (180 g/l) was infused to maintain blood glucose levels between 5 and 5.5 mmol/l. The rate of glucose infusion was determined by measuring the plasma glucose concentrations every 5-10 min. The rate of glucose infusion during the last 30 min of the test was used to determine insulin sensitivity and was expressed as mg/kg-lean body mass/min. In paper II, we used the homeostasis model assessment (1/HOMA = 22.5 / (fasting plasma insulin (μU/ml) x fasting plasma glucose (mmol/l))) as a surrogate measure of insulin sensitivity.

Indirect calorimetry

Indirect calorimetry is a method to quantify energy expenditure and respiratory quotient by measuring oxygen consumption and carbon dioxide production. Indirect calorimetry was performed during basal and insulin-stimulated steady state periods of hyperinsulinemic euglycemic clamp using a computerized flow-through canopy gas analyzer system (Deltatrac, Datex, Helsinki, Finland). Inhaled and exhaled air flow was analyzed for oxygen content using a paramagnetic differential oxygen sensor and for carbon dioxide tension using an infrared carbon dioxide sensor.

Maximal oxygen uptake (VO₂max) measurement

In paper III, VO₂max was measured using an incremental work-conducted upright exercise test with a bicycle ergometer (Monark Varberg, Halland, Sweden) combined with continuous analysis of expiratory gases and minute ventilation. The exercise was started at a workload varying from 30 to 100 W depending on the previous history of endurance training or exercise habits and was then increased by 20–50 W every 3 min until a perceived exhaustion or a respiratory quotient of 1.0 was reached. VO₂max was defined as the oxygen uptake measured during the last 30 seconds of exercise and was expressed per kilogram of total body weight. Fiber-type composition was determined as previously described. We quantified and calculated the fibers using the COMFAS image analysis system (Scan Beam).
Human adipocyte cell culture and treatment

A human preadipocyte cell line derived from subcutaneous white adipose tissue from a patient with Simpson Golabi Behmel syndrome was allowed to differentiate. These cells exhibit a high capacity for adipose differentiation, which results in mature fat cells, which are both biochemically and functionally similar to primary human adipocytes. The mature adipocytes were treated for four hours with glucose (1, 5, or 25 mM) in the presence or absence of the physiological concentration of insulin (1 nM).

RNA extraction and hybridization

RNA from adipocyte cell cultures was extracted using the RNeasy Mini Kit (Qiagen, Germany). RNA from skeletal muscle biopsies was extracted using the guanidinium thiocyanate method or TRI reagent (Sigma-Aldrich, St. Louis, MO). cDNA was synthesized from total RNA using random hexamer primers (Life Technologies, MD) and Superscript II RNase H Reverse Transcriptase (Life Technologies, MD). In paper I, RNA from human skeletal muscle biopsies was hybridized to the Affymetrix HG-U133A arrays in paper I-study A and Affymetrix Hu6800 arrays in paper I-study B. In paper II, we have used the Affymetrix Custom Array NuGO-Hs1a520180 arrays in paper II-study A and the one-color (Cy3, green) Agilent Whole Human Genome Oligo Microarray (G4112A) in paper II-study B. In paper IV, RNA from whole pancreases of 3-week old NIDDM11 and F344 rats was hybridized to the Affymetrix Rat Genome 230 2.0 arrays.

Real-time polymerase chain reaction (real-time PCR)

The TaqMan real-time PCR is a technique to quantify gene expression. It is based on the 5' to 3' exonuclease activity of the Taq polymerase. The probe which has a reporter fluorescent dye attached to its 5' end and a quencher attached to its 3' end, hybridizes to the target gene (Figure 11). During PCR amplification, the quencher is cleaved by the 5' nuclease activity of Taq polymerase resulting in the accumulation of reporter fluorescence. The release of the fluorescent dye during amplification allows for rapid detection and quantification of cDNA. The real-time PCR was performed using the ABI PRISM 7900 sequence detection systems (Applied Biosystems, Foster City, CA). Primers and probes for different genes namely, TXNIP, G0S2, BCL6, NDUFB5, ATP5C1, and AHNAK mRNA, were purchased as a ready-to-use mix of primers and FAM-labeled probes (Applied Biosystems, Foster City, CA). Cyclophilin A was used as an endogenous control to standardize the amount of cDNA added to the reactions using a ready-to-use mix of primers and a VIC-labeled probe (Applied Biosystems, Foster City, CA). All samples were run in duplicate, and data were calculated using the standard curve method and expressed as a ratio to the cyclophilin A reference.
Figure 11: Principle of real-time PCR.

This technique exploits the fluorescence resonance energy transfer (FRET) principle. According to this principle the fluorescent dye, when excited, transfers the energy to the quencher dye when they are in close proximity. The fluorescent labeled probe anneals between the forward and reverse primer sites and is cleaved by the Taq DNA polymerase in each PCR cycle. This cleavage results in loss of FRET and fluorescence can be detected, which is proportional to the initial level of cDNA. R = reporter (FAM, TET or VIC), Q = quencher (TAMRA or non-fluorescent). This figure is adapted from ABI Prism 7900 sequence detection systems (Applied Biosystems, Foster City, CA, USA) user manual.
RNA interference, viral transduction, and glucose uptake assays

RNA interference (RNAi) is a mechanism by which short RNA oligonucleotides can mediate repression of complementary genes. The RNAi works through inhibition of mRNA translation, destruction of mRNA, or silencing of the promoter which governs expression of the mRNA. In paper I, human myoblast cells were transfected with short interfering RNA (siRNA) directed against TXNIP (Dharmacon, Chicago, IL) as described previously. In addition, 3T3-L1 mouse fibroblast cells were transfected with mouse Txnip-specific siRNA species (individual siRNA sequences, Ambion, Austin, TX).

Viral transduction is a method to transduce genes into mammalian cells and cause over-expression of proteins. Human TXNIP was subcloned into the pCDH1-MCS1-EF1-Puro vector (System Biosciences, Mountain View, CA). A vector containing TXNIP or an empty vector was transfected into 293TN cells (ATCC) with Fugene 6 (Roche, Basel, Switzerland) to generate pseudoviral particles.

Glucose uptake assays on virally infected adipocytes, siRNA gene-silenced adipocytes, and siRNA gene-silenced human myoblast cells were performed with insulin stimulation and subsequent measurement of lysate radioactivity using scintillation counting and the data were expressed in pmol/(mg lysate protein-min). These experiments were performed by the groups of Richard Lee, Brigham and Women’s Hospital, Cambridge, MA, and Juleen Zierath, Karolinska Institute, Stockholm, Sweden.

Sequencing

Sequencing of the TXNIP gene was performed using Cycle Sequencing kit (Applied Biosystems, Foster City, CA) from genomic DNA with an ABI3730 DNA Analyzer. Each base was called using the PHRED software package, and quality scores were assigned. Sequence assembly and analysis were performed using the Staden software package.

Genotyping using TaqMan allelic discrimination

The TaqMan allelic discrimination method is a sequence-specific method for SNP genotyping. The allelic discrimination assay contains two TaqMan probes, one probe for each allele of the SNP. Each probe consists of an oligonucleotide with a reporter fluorescent dye at 5’ end and a quencher at 3’ end. During the amplification phase of PCR, the TaqMan probes hybridize only to the perfectly matching DNA target and not to those with a one base mismatch. Cleavage of the hybridized probe separates
the quencher from the reporter, which results in fluorescence of the reporter dye. The SNP genotype is inferred by monitoring the fluorescence of the reaction mixture\(^{187}\). As the probes are distinguished by different fluorophores, emission of only one of the signals corresponds to homozygosity while emission of both signals corresponds to heterozygosity\(^{188,189}\). Allelic discrimination was performed using the ABI PRISM 7900 sequence detection system (Applied Biosystems, Foster City, CA).

**Genotyping using MassARRAY technology**

The principle of the MassARRAY system is that an oligonucleotide probe is extended over a SNP site in a PCR product with the use of dideoxynucleotide triphosphates (ddNTPs) terminators which produce different size products for each allele of a SNP\(^{190}\). The extended products are analyzed using SEQUENOM matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry which allows precise determination of the size of products generated, which is used to determine SNP genotype\(^{191}\). A large number of SNPs can be genotyped simultaneously using this technology (Sequenom, San Diego, CA).

**Statistical analysis**

Data are presented as mean ± standard deviation (SD) or mean ± standard error of the mean (SEM) or median with an interquartile range were appropriate.

Mostly, non-parametric statistics were used due to small sample sizes and the fact that variables were not normally distributed. For two-group comparisons we have used Mann-Whitney U test for unpaired comparisons and Wilcoxon’s signed rank sum test for paired comparisons. The association between two variables was assessed by Spearman rank correlation analysis. Normally distributed variables were analyzed using Student’s \(t\)-test with equal variance for unpaired comparisons, and paired Student’s \(t\)-test for paired comparisons between two groups or two-way analysis of variance (ANOVA) for comparing three or more groups. Two-tailed \(P < 0.05\) was considered statistically significant.

Allele and genotype frequencies were compared by \(\chi^2\) statistics and/or Fisher’s exact test and by McNemar’s statistics in the matched case-control study. The Hardy-Weinberg equilibrium was assessed by \(\chi^2\) test. Multiple linear regression analysis was used to determine genotype and quantitative phenotype correlations after adjusting for potential confounding effects of other covariates such as age, sex, BMI, and family dependence. Multiple logistic regression analysis was performed to assess association between a genotype and occurrence of T2DM after adjusting for different covariates. In paper I, the Mantel-Haenszel meta-analysis was performed to combine results from different sample populations.
Generalized estimating equations (GEE) were used to fit general linear models for the twin data in papers I and III. The GEE model was used to adjust for interdependence between twins. In this model the correlation was allowed to be different for monozygotic and dizygotic twins. Variables included in the model were selected using a backward selection regression.

A global scaling normalization was performed to normalize the gene expression profiling data in paper I. To eliminate data that were not reliably detected over all samples, the MAS5.0 present/absent calls based filtering was performed which classified each gene as expressed above background (present call) or not (absent or marginal call) for Affymetrix arrays in all papers. In paper I, the SAM method with the fold change values $>1.5$, delta $=0.37$, and median false discovery rate $=0\%$ was used for paired data to identify insulin-regulated genes from study A. Normalized gene expression data from previously published microarray studies were used for analyzing expression of individual genes. In paper II-study A, three different procedures for normalization and summarization were used namely GC-RMA, PLIER, and RMA for the Affymetrix arrays. For one-color Agilent arrays in paper II-study B, a quantile-based normalization between arrays without background subtraction was performed. To identify genes whose expression correlated consistently with VO$_{2\text{max}}$ and percentage of type 1 fibers, we have used normalized gene expression data from our previous microarray study in paper III. Spearman partial correlation analysis was performed to determine the individual effects of each gene expression on different quantitative phenotypes (e.g., a surrogate measure of insulin sensitivity ($1/HOMA$) in paper II, VO$_{2\text{max}}$ and percentage of type 1 fibers in paper III) after adjusting for different sets of covariates (such as BMI, age, and family history of T2DM in paper II-study A, BMI and age in paper II-study B, and BMI and T2DM in paper III-study A). In paper IV, normalization of raw data from different published gene expression profiling studies was performed using four different methods, namely MAS5.0, GC-RMA, PLIER, and RMA. Due to the hypothesis generating nature of this study, we accepted genes that were significantly altered in diabetes or associated phenotypes with a $P<0.05$ in at least one normalization method.

The GSEA of metabolic pathways was performed between two groups divided by mean values of VO$_{2\text{max}}$ as well as percentage of type 1 fibers in paper III. Also, The KEGG pathway analysis of genes of interest was performed using WebGestalt which implements a hypergeometric test in papers II and III.

All statistical analyses were performed using Number Cruncher Statistical Systems software (NCSS, Kaysville, UT), STATA (StataCorp, College Station, TX), SPSS (SPSS, Chicago, IL), MATLAB (The MathWorks, Natick, MA), GSEA, PLINK, and R statistical software.
Results

Paper I: TXNIP regulates peripheral glucose metabolism in humans

The aim of this paper was to identify mediators of glucose homeostasis and to investigate whether these mediators are causally involved in the development of T2DM.

We obtained skeletal muscle biopsies from non-diabetic participants before and after a euglycemic hyperinsulinemic clamp using two different clinical protocols (paper I-studies A and B) each of which examined six individuals and then profiled gene expression in these biopsies using Affymetrix arrays. Expression of G0S2 was induced by insulin and expression of TXNIP and BCL6 was repressed by insulin in both studies. These findings were independently validated by real-time PCR in a third clamp protocol (paper I-study C) that involved a panel of 96 young non-diabetic individuals (Figure 12).

![Graph showing TXNIP expression in human muscle in vivo](image)

**Figure 12:** Effects of insulin on TXNIP expression in human muscle *in vivo.*

TXNIP mRNA levels were measured in skeletal muscle biopsies of healthy individuals before and after the hyperinsulinemic euglycemic clamp (N = 96, P < 1 x 10^{-14}). Levels were measured using real-time PCR and normalized to cyclophilin A mRNA. Inset shows the distribution of expression changes over individuals.
We analyzed two previously published microarray datasets from muscle of T2DM patients and controls\textsuperscript{104,105} and found that \textit{TXNIP} expression was elevated in IGT (1.95-fold, \( P = 0.02 \)) or T2DM patients (1.60-fold, \( P = 0.01 \)) but not in healthy individuals with a family history of T2DM (1.05-fold, \( P = 0.67 \)). Moreover, \textit{Txnip} gene was highly expressed (1.50-fold, \( P < 0.05 \)) in the muscle of mice treated with streptozotocin, an agent that is selectively toxic to pancreatic \( \beta \)-cells and results in insulin deficiency and hyperglycemia\textsuperscript{194}. \textit{Txnip} expression levels reverted (0.53-fold, \( P < 0.05 \)) to baseline in the muscle of these mice following treatment with insulin, which normalizes circulating glucose levels. However, in mice with knock-out of the insulin-receptor in skeletal muscle\textsuperscript{161}, insulin treatment of streptozotocin-diabetic mice failed to suppress \textit{Txnip} expression (1.33-fold increase with insulin, \( P = 0.08 \)), despite the fact that glucose was well controlled in these mice. These findings indicate that the suppression of \textit{Txnip} by insulin is not simply secondary to a reduction in circulating glucose concentrations, but rather or in addition, requires intact insulin receptor signaling.

We performed \textit{in vitro} assays of \textit{TXNIP} expression in a human adipocyte cell line that is known to respond to insulin by stimulating glucose uptake. \textit{TXNIP} expression increased following elevations in glucose concentration, and was suppressed by insulin. Together, these cell culture studies confirm that \textit{TXNIP} expression is reciprocally regulated by insulin and glucose in human muscle and fat tissues.

Using gene expression data from skeletal muscle of normal glucose tolerance (NGT), IGT, and T2DM patients from our previously published study\textsuperscript{104}, we found that \textit{TXNIP} expression is inversely correlated with the total body rate of insulin-stimulated glucose. This relationship is robust in NGT (\( r = -0.57, P = 0.02 \)) or IGT (\( r = -0.86, P < 0.01 \)), but not in patients with T2DM (\( r = -0.21, P = 0.41 \)). We also related the expression of skeletal muscle \textit{TXNIP} to glucose uptake in 96 healthy, young individuals. Using a GEE model, we modeled insulin-stimulated glucose uptake as a function of age, sex, basal and insulin-stimulated \textit{TXNIP} expression, zygosity, birth weight, percentage of body fat, BMI, and \( VO_{2\text{max}} \). Basal (\( P = 0.02 \)) and insulin-stimulated (\( P < 0.01 \)) \textit{TXNIP} expression as well as BMI (\( P < 0.01 \)) were inversely correlated, while \( VO_{2\text{max}} \) (\( P < 0.01 \)) was positively correlated with insulin-stimulated glucose uptake (Table 2). These data are consistent with the notion that \textit{TXNIP} is an independent determinant of insulin-stimulated glucose uptake in humans.

### Table 2: Factors influencing insulin-stimulated glucose uptake in humans

<table>
<thead>
<tr>
<th>Factors</th>
<th>Regression coefficients</th>
<th>95 % CI</th>
<th>( P )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>\textit{TXNIP} before clamp</td>
<td>-4.18</td>
<td>-7.76</td>
<td>-0.61</td>
</tr>
<tr>
<td>\textit{TXNIP} after clamp</td>
<td>-20.29</td>
<td>-35.20</td>
<td>-5.39</td>
</tr>
<tr>
<td>BMI</td>
<td>-0.35</td>
<td>-0.53</td>
<td>-0.16</td>
</tr>
<tr>
<td>( VO_{2\text{max}} )</td>
<td>0.11</td>
<td>0.03</td>
<td>0.18</td>
</tr>
</tbody>
</table>

The GEE model was used to determine whether basal and insulin-stimulated \textit{TXNIP} expression, age, sex, birth weight, zygosity, percentage of body fat, BMI, or \( VO_{2\text{max}} \) influence insulin-stimulated glucose uptake.
To directly test if TXNIP may regulate glucose uptake, we used genetic manipulations to alter the gene expression in mouse and human cell lines. We performed *in vitro* studies on an insulin-sensitive cell line, the adipocyte differentiated 3T3-L1. A 2-fold forced overexpression of TXNIP using lentiviral transduction resulted in both diminished basal and insulin-stimulated glucose uptake (59% and 28% reduction, respectively). In agreement with this result, forced reduction of TXNIP expression by siRNA gene silencing achieved the opposite effect and enhanced both basal and insulin-stimulated glucose uptake (157% and 61%, respectively). In addition, we examined the effects of TXNIP siRNA gene silencing in primary human skeletal muscle myocytes. Again, we observed a significant increase in both basal and insulin-stimulated glucose uptake. Together, these *in vitro* studies indicate that changes in TXNIP expression can directly influence glucose uptake in insulin-responsive cells.

To determine whether genetic variation in TXNIP may be associated with measures of insulin resistance or T2DM in humans, we assessed the pattern of common genetic variation encompassing the TXNIP gene using data from the HapMap project and by re-sequencing the gene. We then performed a genetic association study of nine “tag” SNPs that capture the most of the common variation observed within the vicinity of this gene. In 4,450 Scandinavian individuals, no significant association was detected between SNPs in TXNIP and T2DM or between the genotypes and four quantitative phenotypes: fasting glucose, fasting insulin, HOMA-IR as a measure of insulin resistance, and HOMA-β as a measure of insulin secretion.

In conclusion, TXNIP regulates both insulin-dependent and insulin-independent pathways of glucose uptake in human skeletal muscle. Combined with recent studies that have implicated TXNIP in pancreatic β-cell glucose toxicity, our data suggest that TXNIP might play a key role in defective glucose homeostasis preceding overt T2DM.

**Paper II: Relationship between insulin sensitivity and gene expression in human skeletal muscle**

The aim of this paper was to investigate molecular mechanisms associated with insulin sensitivity in skeletal muscle by relating global skeletal muscle gene expression to physiological measures of insulin sensitivity.

We obtained serial muscle biopsies from 38 non-diabetic participants in paper II-study A and then profiled gene expression in biopsies using the Affymetrix oligonucleotide microarrays. We used the 1/HOMA as a surrogate measure of insulin sensitivity. We identified 70 genes positively and 110 genes inversely correlated with insulin sensitivity in human skeletal muscle. Among genes positively correlated with insulin sensitivity we found two, TSC2 and ULK1, involved in the mammalian target-of-rapamycin (mTOR)
signaling pathway using WebGestalt ($P < 0.01$). Other positively correlated genes had various functions, e.g., \textit{CAMK2A} ($r = 0.37$, $P = 0.03$) from the Wnt signaling pathway, \textit{UCP2} ($r = 0.56$, $P < 0.01$) from a family of inner mitochondrial membrane proteins, and \textit{CPT1B} ($r = 0.47$, $P < 0.01$) from peroxisome proliferator-activated receptor (PPAR) signaling pathway. We found six genes (\textit{COL1A1}, \textit{COL1A2}, \textit{COL3A1}, \textit{LAMA4}, \textit{MYL6}, and \textit{RAP1A}) involved in focal adhesion ($P < 1 \times 10^{-4}$), cell communication ($P < 1 \times 10^{-3}$), and extracellular matrix (ECM)-receptor interaction ($P < 1 \times 10^{-3}$) pathways among the genes inversely correlated with insulin sensitivity.

To replicate the finding from paper II-study A, we included 9 non-diabetic participants in paper II-study B. We again used 1/HOMA as a surrogate measure of insulin sensitivity. We performed skeletal muscle gene expression profiling from these participants using the Agilent oligonucleotide microarrays. \textit{SIRT2} ($r = 0.95$, $P < 0.01$), \textit{FBXW5} ($r = 0.95$, $P < 0.01$), \textit{RAB11FIP5} ($r = 0.93$, $P < 0.01$), \textit{CPT1B} ($r = 0.85$, $P = 0.02$), \textit{CI6orf86} ($r = 0.79$, $P = 0.04$), \textit{UCKL1} ($r = 0.77$, $P = 0.04$), and \textit{ARFGAP2} ($r = 0.76$, $P < 0.05$) were positively correlated while \textit{ZNF613} ($r = -0.81$, $P = 0.03$), \textit{UTP6} ($r = -0.81$, $P = 0.03$), and \textit{LEO1} ($r = -0.78$, $P = 0.04$) were inversely correlated with insulin sensitivity in this study. Moreover, a recent study has shown that the skeletal muscle expression of \textit{CPT1B} was increased in response to treatment with a PPARδ agonist\textsuperscript{196} and this agonist is shown to increase mitochondrial biogenesis in muscle by enhancing fatty acid catabolism\textsuperscript{197}. Taken together, these findings point at the possibility that beneficial effect of PPARδ agonist is in part regulated by CPT1B.

We conclude that a high capacity of fat oxidation in mitochondria is reflected by a high expression of \textit{CPT1B} which is a marker of insulin sensitivity.

\textbf{Paper III: Molecular correlates for maximal oxygen uptake and type 1 fibers}

The aim of this paper was to investigate molecular mechanisms associated with VO\textsubscript{2max} and type 1 fibers in human skeletal muscle.

To identify genes whose expression was consistently correlated with VO\textsubscript{2max} and type 1 fibers, we reanalyzed our previously published gene expression data from human skeletal muscle in paper III-study A\textsuperscript{104}. We observed a modest positive correlation between VO\textsubscript{2max} and percentage of type 1 fibers ($r = 0.37$, $P = 0.03$) in these 43 individuals. We identified 66 genes positively correlated and 83 genes inversely correlated with VO\textsubscript{2max}. Among positively correlated genes, we found 13 genes involved in OXPHOS using WebGestalt ($P < 1 \times 10^{-19}$), i.e., \textit{NDUFS3}, \textit{NDUFS8}, \textit{NDUFB4}, \textit{NDUFB5}, and \textit{NDUFB11} from complex I, \textit{UQCRB} and \textit{UQCRC2} from complex III, \textit{COX5A}, \textit{COX5B}, and \textit{COX6A1} from complex IV, and \textit{ATP5C1}, \textit{ATP5G3}, and \textit{ATP6V0C} from complex V (Figure 13A). There were no OXPHOS genes among genes inversely correlated with VO\textsubscript{2max}. Instead, this list comprised genes with varying function, including \textit{AHNAK}.
(r = -0.50, P < 0.01), NACA (r = -0.46, P < 0.01), and BCL6 (r = -0.42, P < 0.01). We performed a GSEA of metabolic pathways by comparing groups with VO\textsubscript{2\text{max}} below or above the mean of VO\textsubscript{2\text{max}} in all individuals. We hereby identified three gene sets (OXPHOS-CR, the electron transport chain, and oxidative phosphorylation) to be significantly enriched in the high-VO\textsubscript{2\text{max}} group but none in the low-VO\textsubscript{2\text{max}} group (Table 3). Among the list of genes correlated with VO\textsubscript{2\text{max}}, we also identified 6 genes (ATP5C1, NDUFB5, PTP4A1, MTCH2, TSG101, and GOLGA7) whose expression was consistently increased whereas the expression of 10 genes (PSMB1, AHNAK, CIRBP, ABCA8, ZBTB20, FOXO1, HNRPA3P1, TTC3, CYLD, and ZNF611) was decreased after 3 months of endurance training.

Table 3 Gene sets enriched in high VO\textsubscript{2\text{max}} group

<table>
<thead>
<tr>
<th>Gene Set</th>
<th>Size</th>
<th>ES</th>
<th>NES</th>
<th>P</th>
<th>FDR, q-value</th>
<th>FWER, P</th>
</tr>
</thead>
<tbody>
<tr>
<td>OXPHOS-CR</td>
<td>77</td>
<td>0.66</td>
<td>1.91</td>
<td>0.002</td>
<td>0.049</td>
<td>0.097</td>
</tr>
<tr>
<td>Electron transport chain</td>
<td>80</td>
<td>0.65</td>
<td>1.89</td>
<td>0.006</td>
<td>0.031</td>
<td>0.115</td>
</tr>
<tr>
<td>MAP00190 oxidative phosphorylation</td>
<td>42</td>
<td>0.56</td>
<td>1.84</td>
<td>0.004</td>
<td>0.039</td>
<td>0.191</td>
</tr>
</tbody>
</table>

OXPHOS-CR, a subset of genes involved in oxidative phosphorylation; ES, enrichment score; NES, normalized enrichment score; FDR, False discovery rate; FWER, family-wise error rate. Size refers to numbers of genes in the gene set.

We identified 171 probe sets positively correlated and 217 probe sets inversely correlated with percentage of type 1 fibers. There were 15 OXPHOS genes whose expression correlated positively with percentage of type 1 fibers using WebGestalt (P < 1 x 10\textsuperscript{-16}, i.e., NDUFB4, NDUFC2, UQCRB, UQRC2, UQCRQ, COX5A, COX7A2, COX7B, ATP5C1, ATP5J2, ATP5G3, ATP5L, ATP5O, ATP6V1D, and ATP5F1 (Figure 13B).
Figure 13: List of genes from the oxidative phosphorylation pathway positively correlated with (A) Maximal oxygen uptake (VO$_{2\text{max}}$), and (B) Percentage of type 1 fibers.

For replication of some of the above findings by applying real-time PCR to archived RNA samples from skeletal muscle of 154 non-diabetic individuals in paper III-study B, we selected four genes namely NDUF5, ATP5C1, AHNAK, and BCL6, for the following reasons: (1) NDUF5 because its expression was positively correlated with VO$_{2\text{max}}$ and increased after training, and also because a study has shown that expression of NDUF5 decreases after a high-fat diet; (2) ATP5C1 because its expression was positively correlated with VO$_{2\text{max}}$ and type 1 fibers and increased after training; (3) AHNAK because its expression was strongly inversely correlated with VO$_{2\text{max}}$ and decreased after training; and (4) BCL6 expression because we previously have shown that its expression is influenced by insulin in paper I.

We used a GEE to model expression of these genes individually as a function of age, sex, BMI, and zygosity. Interestingly, the expression of NDUF5 (6.33 ± 0.25 vs. 7.70 ± 0.26, $P < 0.01$) and ATP5C1 (1.13 ± 0.06 vs. 1.59 ± 0.08, $P < 0.01$) was reduced, whereas expression of AHNAK (2.01 ± 0.12 vs. 1.36 ± 0.09, $P < 0.01$) was increased in elderly compared with young individuals. We also modeled VO$_{2\text{max}}$ as a function of age, sex, BMI, zygosity, and expression of individual genes (NDUF5, ATP5C1, AHNAK, and BCL6). Expression of NDUF5 ($P = 0.03$; Table 4) and ATP5C1 ($P = 0.02$; Table
4) but not of AHNAK (regression coefficient $\beta = -0.10$, 95% confidence interval (CI) = $-1.34$–$-1.14$, $P = 0.87$) or BCL6 ($\beta = 3.13$, 95% CI = $-2.65$–$8.91$, $P = 0.29$) was significantly correlated with VO$_{2\text{max}}$.

### Table 4 Genetic and non-genetic factors influencing VO$_{2\text{max}}$ in humans

<table>
<thead>
<tr>
<th>Factors</th>
<th>Regression coefficients</th>
<th>95 % CI</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td><strong>NDUFB5</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expression</td>
<td>0.44</td>
<td>0.04</td>
<td>0.84</td>
</tr>
<tr>
<td>BMI</td>
<td>-0.69</td>
<td>-0.96</td>
<td>-0.42</td>
</tr>
<tr>
<td>Age</td>
<td>-0.29</td>
<td>-0.36</td>
<td>-0.22</td>
</tr>
<tr>
<td>Sex</td>
<td>-6.35</td>
<td>-8.69</td>
<td>-3.99</td>
</tr>
<tr>
<td><strong>ATP5C1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expression</td>
<td>0.95</td>
<td>0.15</td>
<td>1.74</td>
</tr>
<tr>
<td>BMI</td>
<td>-0.85</td>
<td>-1.02</td>
<td>-0.67</td>
</tr>
<tr>
<td>Age</td>
<td>-0.28</td>
<td>-0.35</td>
<td>-0.21</td>
</tr>
<tr>
<td>Sex</td>
<td>-7.44</td>
<td>-9.93</td>
<td>-4.95</td>
</tr>
</tbody>
</table>

The generalized estimating equation model was used to determine whether BMI, age, sex, zygosity, and expression of individual genes (NDUFB5 and ATP5C1) influence VO$_{2\text{max}}$.

**Abbreviations**: CI, confidence interval.

In conclusion, VO$_{2\text{max}}$ closely reflects expression of OXPHOS genes, particularly that of NDUF5 and ATP5C1, in skeletal muscle. The expression of these genes was decreased with aging and increased with exercise training. Increased expression of these genes thereby seems to reflect good muscle fitness. In contrast, expression of AHNAK was associated with low VO$_{2\text{max}}$, increased with aging, and decreased with exercise training. Increased expression of AHNAK thus seems to be a marker of poor muscle fitness.

**Paper IV: Prioritizing genes for follow-up from genome wide association studies using information on gene expression in tissues relevant for type 2 diabetes mellitus**

The aim of this paper was to combine GWA studies with genome-wide expression profiling of target tissues (i.e. pancreas, adipose tissue, liver, and skeletal muscle) of T2DM in order to identify novel T2DM susceptibility loci and to understand their pathophysiological role in T2DM etiology.

To identify novel T2DM susceptibility loci, we compiled results from two previously reported GWA studies which used the Affymetrix GeneChip® Human Mapping 500K Array Set containing 500,000 SNPs$^{149,153}$. For first stage of SNPs selection, we identified 453 SNPs associated with T2DM with $P < 0.01$ in at least one of the GWA study (Figure
and 150 genes were located in the vicinity of these SNPs. To determine whether any of these genes exhibit altered expression in diabetes or associated phenotypes in humans or rodents, we analyzed gene expression profiling data from pancreas, adipose tissue, liver, and skeletal muscle\cite{104,107,160,161}. Out of these 150 genes, we identified 41 genes differentially expressed in these studies. 11 genes were differentially expressed in at least two studies (Table 5).

**Figure 14:** Schematic diagram of prioritizing SNPs from GWA studies.

453 SNPs were associated with T2DM with $P < 0.01$ in at least one of the GWA study\cite{149,153} and 150 genes were located in vicinity of these SNPs. Out of these 150 genes, 41 were differentially expressed in different gene expression profiling studies\cite{104,107,160,161}. 16 SNPs from 7 genes were selected based upon the results from the GWA studies for replication in well powered Malmö case-control study.
Table 5: Genes differentially expressed in at least two different studies

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Entrez GeneID</th>
<th>Dataset</th>
<th>Fold-change</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERAP1</td>
<td>51752</td>
<td>T2DM vs. NGT human pancreatic islets</td>
<td>2.18</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Diabetic vs. healthy rat pancreas</td>
<td>0.83</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NIDDM1I vs. F344 rat pancreas</td>
<td>1.26</td>
<td>0.025</td>
</tr>
<tr>
<td>HNRNPC</td>
<td>3183</td>
<td>T2DM vs. NGT human pancreatic islets</td>
<td>0.55</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NIDDM1I vs. F344 rat pancreans</td>
<td>0.92</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T2DM vs. NGT human skeletal muscle</td>
<td>1.36</td>
<td>0.029</td>
</tr>
<tr>
<td>NCOR1</td>
<td>9611</td>
<td>T2DM vs. NGT human pancreatic islets</td>
<td>1.28</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Diabetic vs. healthy mice skeletal muscle</td>
<td>1.47</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>CAST</td>
<td>831</td>
<td>T2DM vs. NGT human pancreatic islets</td>
<td>1.31</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Liver of 12 weeks old ZDF vs. ZLC</td>
<td>-0.68&lt;sup&gt;5&lt;/sup&gt;</td>
<td>-</td>
</tr>
<tr>
<td>PQLC1</td>
<td>80148</td>
<td>T2DM vs. NGT human pancreatic islets</td>
<td>1.44</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Diabetic vs. healthy rat pancrean</td>
<td>0.81</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>CADPS</td>
<td>8618</td>
<td>NIDDM1I vs. F344 rat pancreans</td>
<td>1.19</td>
<td>0.043</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Diabetic vs. healthy rat pancrean</td>
<td>0.60</td>
<td>0.011</td>
</tr>
<tr>
<td>TGFB3</td>
<td>7043</td>
<td>NIDDM1I vs. F344 rat pancreans</td>
<td>1.16</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Diabetic vs. healthy rat pancrean</td>
<td>0.77</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>KCNS3</td>
<td>3790</td>
<td>T2DM vs. NGT human skeletal muscle</td>
<td>0.83</td>
<td>0.029</td>
</tr>
<tr>
<td>SAMD4A</td>
<td>23034</td>
<td>T2DM vs. NGT human skeletal muscle</td>
<td>0.59</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NIDDM1I vs. F344 rat pancreans</td>
<td>0.63</td>
<td>0.019</td>
</tr>
<tr>
<td>FAM45A</td>
<td>55855</td>
<td>NIDDM1I vs. F344 rat pancreans</td>
<td>0.88</td>
<td>0.023</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Muscle of 12 weeks old ZDF vs. ZLC</td>
<td>-0.62&lt;sup&gt;5&lt;/sup&gt;</td>
<td>-</td>
</tr>
<tr>
<td>ACTN1</td>
<td>87</td>
<td>NIDDM1I vs. F344 rat pancreans</td>
<td>1.18</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Muscle of 12 weeks old ZDF vs. ZLC</td>
<td>0.62&lt;sup&gt;5&lt;/sup&gt;</td>
<td>-</td>
</tr>
</tbody>
</table>

*Expression ratios are shown as log2 scale of normalized mean signal intensities of ZDF to ZLC rats.

We selected 5 genes namely - HNRNPC, NCOR1, CAST, KCNS3, and FAM45A as these genes were differentially expressed in two different tissues for further investigation (Table 5). Furthermore, ZBTB16 was selected because it has been shown to interact with NCOR1 and act as a transcriptional repressor. Similarly, protein-protein interactions have been reported between MSRA and TXNIP which involves TRX. Since TXNIP is a regulator of insulin-dependent and insulin-independent pathways of glucose uptake in human skeletal muscle and is involved in pancreatic β-cell apoptosis, we decided to investigate MSRA further. We selected 16 SNPs from 7 genes for replication in a well powered Malmö case-control study consisting of 2,830 T2DM patients and 3,550 controls. However, we were not able to replicate association of these 16 SNPs with T2DM in the Malmö case-control study. Though we further genotyped a SNP (rs27582) in the calpastatin (CAST) gene in two prospective cohorts from the MPP study and the BPS, for the following reasons: (1) the SNP was strongly associated with T2DM in the DGI (P = 0.007) and the WTCCC (P = 0.008) studies, (2) CAST gene
expression was increased in human pancreatic islets from T2DM patients compared to controls (1.31-fold, \( P < 0.05 \); Table 5), (3) Cast gene expression was decreased in liver of Zucker diabetic fatty (ZDF) rats compared to and Zucker lean control (ZLC) rats (1.60-fold; Table 5), and (4) CAST is an endogenous inhibitor of calpains and a member of this family, \( CAPNI10 \), has been associated with T2DM\(^{199} \). In the MPP study, carriers of rs27582 AA/AG genotypes had a higher risk of future T2DM than GG genotype carriers (odds ratio (OR) = 1.10, 95% CI: 1.00-1.20, \( P < 0.05 \) adjusted for BMI at follow-up; OR = 1.09, 95% CI: 1.00-1.19, \( P = 0.069 \) adjusted for BMI at baseline), this effect was particularly strong in individuals with BMI below the median (below median at follow-up; OR = 1.19, 95% CI: 1.03-1.36, \( P = 0.024 \); below median of BMI at baseline; OR = 1.25, 95% CI: 1.08-1.43, \( P = 0.005 \)).

We also investigated whether a SNP (rs27582) in the CAST gene would influence insulin secretion and action during the 7.6 year follow-up in the BPS. Carriers of the A-allele were more insulin resistant than carrier of the G-allele as indicated by higher fasting insulin concentrations (regression coefficient (\( \beta \)) = 0.048, \( P = 0.017 \)) and higher HOMA-IR index (\( \beta = 0.044, \ P = 0.025 \)) as well as lower insulin sensitivity index during OGTT (\( \beta = -0.039, \ P = 0.039 \)) at follow-up.

In conclusion, using gene expression in different tissues from patients with T2DM and animal models is a powerful tool for prioritizing SNPs from GWA study for replication studies. We thereby identified association of a variant (rs27582) in the CAST gene with T2DM and insulin resistance.
Discussion

The search for susceptibility genes for T2DM is a challenging task due to polygenic nature of the disease and the interaction between genetic and non-genetic factors (e.g., diet, physical activity, and aging) contributing to the disease. Identifying individuals at high risk of developing the disease is a prerequisite for disease prevention. Also, development of efficacious anti-diabetic treatments depends heavily on a deeper understanding of the molecular events that lead to T2DM. Recent advances in high-throughput technologies related to biomedical research have provided tremendous opportunities for understanding the pathogenesis of T2DM. This thesis has focused on the identification of genes regulated by insulin, correlated with insulin sensitivity, and associated with VO₂ max and type 1 fibers in human skeletal muscle as well as described an approach for prioritizing SNPs from GWA studies.

Methodology

Gene expression profiling studies, which enable the simultaneous investigation of more than 20,000 genes, allow a more comprehensive understanding of the genes involved in a wide variety of biological processes. However, after the initial euphoria over the successful development of gene expression profiling technology; there are several caveats with data analysis and interpretation associated with this technology. The complex nature of a microarray experiment introduces many sources of variations which includes sample extraction, sample quality, array design (i.e., probe length and sequence homology to other sequences and replicates), labeling protocol, hybridization conditions, wash conditions, choice of scanning instrument, image processing, data normalization and analysis, data quality assessment, and interpretation of results. Several standardization projects have been pursued by the Microarray Gene Expression Data Society to facilitate the sharing of data generated using the microarray technologies. The best known among these standardization projects is the Minimum Information About a Microarray Experiment (MIAME), which describes the minimum information required to ensure that microarray data can be easily interpreted and that results derived from its analysis can be independently verified. Although, this MIAME standard is appropriate for reporting methodological and design aspects of microarray experiments; it does not address data analysis and validation. Moreover, gene expression profiles are strongly influenced by gender and age, therefore, these confounding factors need to be considered while designing of microarray experiments. Also, the site from which a biopsy sample is obtained can have an effect on the gene expression profiles due to large functional and morphologic heterogeneity of muscle tissue.
Several statistical algorithms have been developed for background correction, normalization, and summarization of Affymetrix microarray data. The ideal probe set algorithm should compare information about probe characteristics, based upon the performance of each probe across arrays, and use this to give a better estimate of transcript abundance. The model-based algorithms, e.g., dchip, RMA, PLIER, and GC-RMA, have been developed based upon these principles. Moreover, the statistical analysis of microarray data is challenging due to the fact that a large numbers of comparisons are made in parallel\(^{204}\). This hampers many studies employing only a small number of replicates with a considerable risk of false positive findings. The required sample size for microarray experiments predominantly depends on the distribution of the truly differentially expressed genes i.e. when the fold-change differences are small, a large sample size is needed to control for the FDR\(^{205}\). Moreover, differences in gene expression are usually modest and would often not reach significant when correction for multiple hypothesis testing. To address this problem we exploited several strategies. In paper II-study A, we used three different model-based algorithms to calculate gene expression from Affymetrix array data. We required that all three methods namely GC-RMA, PLIER, and RMA should show consistent results. Only in paper IV due to hypothesis generating nature of the study, we used a less stringent cut-off with \(P < 0.05\) in at least one of these methods. In addition, we have integrated the results of our microarray analysis with previously published microarray data by re-analyzing these data\(^{104,105,107,160,161,193,194}\). This approach was successfully implemented by Coletta and colleagues who showed that the transcription factor \(HES1\) was regulated by insulin in human skeletal muscle\(^{206}\). However, further validation using alternative methods (e.g. real-time PCR) is still necessary to extrapolate results for broad use of the data. Hence, we have validated our key findings in independent clinical studies using different quantitation platforms (e.g. real-time PCR) which have provided biological and technical replication of our microarray experiments.

Prioritizing SNPs from GWA studies for further replication in other samples is challenging task due to the fact that effect sizes of individual SNPs are usually very small and would often not reach significance after correction for multiple hypothesis testing. To address this issue several strategies have been developed including meta-analysis of GWA studies and Gene Relationships Across Implicated Loci (GRAIL)\(^{207}\). The meta-analysis of several GWA studies from different populations is very powerful approach for identifying variants with modest effects. Six novel loci associated with T2DM have been identified using this approach\(^{152}\). However, there are several difficulties in integrating GWA studies from different platforms and also this strategy requires large-scale GWA studies in several populations. The GRAIL is a web-tool to examine relationships between genes in different disease associated loci. It is based upon similarities in the published scientific text among the associated genes. We used the GRAIL for 453 SNPs associated with T2DM from paper IV (Figure 14), but could only identify one gene (\(HHEX\)) for which neighboring SNPs have been associated with T2DM\(^{151}\). Also, a more general approach e.g. GeneMiner (which is a meta-analysis
approach that integrates data of heterogeneous origin e.g. DNA microarrays and complementing qualitative data covering several human and mouse tissues related to T2DM) has identified several functional T2DM candidate genes\textsuperscript{208}. However using this approach, we were not able to identify any of the 22 genes for which there is genome-wide significant evidence for association with T2DM\textsuperscript{209}. In paper IV, we tested a novel approach for prioritizing SNPs from GWA studies by combining the results from these studies with publicly available genome-wide expression profiling data of target tissues (i.e. pancreas, adipose tissue, liver, and skeletal muscle) of T2DM. Notably, we were able to identify four genes namely \textit{IGF2BP2}, \textit{CDKAL1}, \textit{TSPAN8}, and \textit{NOTCH2} for which neighboring SNPs have consistently shown to be associated with T2DM in several populations\textsuperscript{149,152,153,209}.

**TXNIP**

In paper I, we have combined human physiology, genome-wide expression profiling, and cellular studies to identify TXNIP as a physiologic regulator of peripheral glucose uptake in humans. A study has shown that TXNIP is glucose inducible in pancreatic β-cells and mediates β-cell death through apoptosis\textsuperscript{210}. Together, these findings suggest that TXNIP may be involved in glucose toxicity both in β-cells and in the periphery, helping to reconcile the dynamic relationship between insulin deficiency and impaired glucose uptake that is observed in the pre-diabetic state (Figure 15). Although in a panel of 4,450 individuals, we found no evidence for association between common genetic variation in the \textit{TXNIP} gene and T2DM; another study has shown that a genetic variation within the \textit{TXNIP} gene is associated with hypertriglyceridemia and increased diastolic blood pressure in patients with T2DM\textsuperscript{211}.

![Figure 15: Role of TXNIP in glucose toxicity in the β-cell and in impaired glucose uptake in the periphery.](image-url)
Insulin deficiency or hyperglycemia can increase TXNIP levels in muscle, resulting in impaired peripheral glucose uptake. The pancreatic β-cell is initially able to compensate by secreting more insulin, but eventually the β-cell compensation fails. The resulting hyperglycemia may then elevate pancreatic β-cell TXNIP expression, which can induce apoptosis. The loss of β-cells, in turn, results in decreased insulin production that further exacerbates peripheral IGT. The vicious cycle would eventually spiral to T2DM.

Thioredoxin is a multifunctional protein and acts as an intracellular oxidoreductase that associates with thioredoxin reductase and thioredoxin peroxidase to reduce oxidized proteins and scavenge free radicals. TXNIP interacts directly with the reduced form of thioredoxin and thereby negatively regulates the expression and function of thioredoxin. In addition, TXNIP has multifunctional roles in a variety of cellular functions including growth inhibition, transcriptional regulation, natural killer cells development, and lipid metabolism. A recent study has shown that TXNIP is a key regulator of PPARα expression and signaling, and coordinated regulation of PPARα and insulin secretion by TXNIP is crucial in the feeding–fasting nutritional transition.

TXNIP expression is induced by glucocorticoids and by glucose, and TXNIP elevations can stimulate ROS production and suppress thioredoxin-mediated antioxidant function. Hence, TXNIP represents a candidate intermediate linking diabetogenic stimuli to ROS production. In addition, we have observed an inverse correlation in the expression of TXNIP and the genes regulating mitochondrial OXPHOS, and previous studies have shown that TXNIP can serve as a transcriptional repressor. Taken together, these data raise the possibility that TXNIP may contribute to the elevations in ROS and mitochondrial dysfunction that accompany insulin resistance.

Over-expression of TXNIP is associated with chronic diabetic neuropathy, and diabetic nephropathy which is related to collagen synthesis. Moreover, elevated levels of TXNIP can enhance the atherosclerotic process by increasing vascular inflammation through inhibition of thioredoxin–ASK1 interaction effect on the negative regulation of TNF-mediated stimulation of JNK, p38, and VCAM1 expression. While expression of TXNIP is downregulated in various types of human cancer cells derived from breast cancer, colon cancer, prostate cancer, bladder cancer, gastrointestinal cancer, malignant pheochromocytomas, and high-grade B cell lymphoma. Taken together, TXNIP plays crucial role in cancer and metabolic diseases.
Figure 16: Schematic diagram showing the role of TXNIP in glucose and its related diseases.

Hyperglycemia-induced reactive oxygen species (ROS) in T2DM up-regulates TXNIP and suppresses thioredoxin (TRX) mediated antioxidant function\textsuperscript{222}. Over-expression of TXNIP enhances β-cell death and impairs insulin secretion\textsuperscript{229,230}, which may result in further aggravation of T2DM. In T2DM patients, increased TXNIP is associated with chronic diabetic neuropathy\textsuperscript{224} and diabetic nephropathy\textsuperscript{225}, which may be related to collagen synthesis. TXNIP also enhances the atherosclerotic process by increasing vascular inflammation through inhibition of TRX–apoptosis signal–regulating kinase 1 (ASK1) interaction activity on the negative regulation of tumor necrosis factor (TNF)-mediated stimulation of c-jun N-terminal kinase (JNK), p38, and vascular cell adhesion molecule-1 (VCAM1) expression\textsuperscript{226,227}. This figure is adapted from Kaimul AM \textit{et al.}\textsuperscript{228}

CPT1B

In paper II, we have identified that expression of \textit{carnitine palmitoyltransferase 1B} (\textit{CPT1B}) is positively correlated with insulin sensitivity in human skeletal muscle. CPT1B regulates transport of long-chain fatty acyl-CoAs from the cytoplasm into the mitochondria and often considered a regulatory step in lipid oxidation. The Krüppel-like transcription factor 5 (KLF5) together with C/EBP-β regulate expression of \textit{CPT1B} as well as \textit{UCP2} in skeletal muscle as promoter regions of these genes contain binding motifs for the transcription factor CCAAT/enhancer binding protein (C/EBP).\textsuperscript{231} Moreover, expression of these genes was up-regulated in the \textit{Klf5}-knockout heterozygous
mice (Klf5\textsuperscript{+/−}) which is resistant to high fat-induced obesity, hypercholesterolemia, and glucose intolerance\textsuperscript{233}. The expression of CPT1B in skeletal muscle was increased in response to treatment with a PPARδ agonist\textsuperscript{196}. This PPARδ agonist is shown to increase mitochondrial biogenesis in muscle by enhancing fatty acid catabolism\textsuperscript{197}. These findings point at the possibility that beneficial effect of PPARδ agonist is in part regulated by CPT1B.

**NDUFB5**

In paper III, we have identified that expression of NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 5 (NDUFB5) is positively correlated with VO\textsubscript{2max} in human skeletal muscle. NDUFB5 is located on the inner mitochondrial membrane. It transfers electrons from NADH to the respiratory chain. Intriguingly, expression of this gene was decreased in human skeletal muscle after a 3-day isoenergetic high-fat diet as well as in mouse muscle after 21 days of high-fat feeding\textsuperscript{198}. No mutations in the gene, which is located on chromosome 3 have been linked with human pathology. However, increased expression of this gene in skeletal muscle seems to reflect fitness of muscle, as the expression increased with endurance training and decreased with aging.

**ATP5C1**

In paper III, we have found that expression of ATP synthase, H+ transporting, mitochondrial F1 complex, gamma polypeptide 1 (ATP5C1) is positively correlated with VO\textsubscript{2max} and percentage of type 1 fibers, increased with endurance training, and decreased with aging in human skeletal muscle. The ATP5C1 gene on chromosome 10p encodes for a subunit of the soluble catalytic portion of mitochondrial ATP synthase, which catalyzes ATP synthesis during oxidative phosphorylation, thereby utilizing an electrochemical gradient of protons across the inner membrane. No obvious pathology has been associated with ATP5C1, and we could not observe any association between common variants in this gene and T2DM or related metabolic traits in the DGI scan\textsuperscript{149}.

**AHNAK**

In paper III, we have shown that enhanced expression of AHNAK nucleoprotein (AHNAK) reflects poor muscle fitness; as the expression level was associated with low VO\textsubscript{2max} and increased with aging but decreased with exercise training in human skeletal muscle. AHNAK encodes for a very large protein, desmoyokin (700 kDa), the carboxyterminal domain of which has been ascribed a stabilizing effect on muscle contractility via interaction with actin\textsuperscript{232}. AHNAK also seems to mediate activation of
phospholipase C and release of arachidonic acid through protein kinase C. This opens up the interesting possibility that \textit{AHNAK} could represent a link to inflammation. In fact, AHNAK was recently suggested to mediate the repressing effect of arachidonic acid on glucose transporter GLUT4 expression in muscle and interfere with phosphatidylinositol 3-kinase-Akt signaling. Moreover, AHNAK has been suggested to link actin with L-type Ca\textsuperscript{2+} channels in cardiomyocytes.

\section*{CAST}

In paper IV, we have identified a SNP (rs27582) in the \textit{CAST} gene that was associated with T2DM in 16,061 Swedish individuals, particularly in lean individuals. Furthermore, this SNP was also associated with more severe insulin resistance. \textit{CAST} is an inhibitor of calpains which bind to \textit{CAST} in a calcium-dependent, reversible manner. Calpains play a key role in the regulation of insulin secretion and insulin action. Taken together, these data raise the possibility that the variation in the \textit{CAST} gene might affect the activity of the calpastatin-calpain system and thereby alters insulin action and eventually secretion.

\section*{Future challenges}

A major future challenge for dissecting the pathogenesis of complex diseases is the need for integration of data from the various high-throughput platforms such as whole-genome sequencing, GWA studies, studies of genetic differences in terms of copy number variations (comparative genomic hybridization arrays), whole-genome methylation profiling, transcriptome profiling (gene expression arrays, exon arrays, and massive-scale RNA sequencing), protein-DNA binding (chromatin immunoprecipitations (ChIP)-chip), micro-RNA arrays, proteomics (protein arrays), and metabolomics. However, cross-platform integration is an intricate task as different data types are interrelated in a complex manner through a highly sophisticated biological network. In general, understanding the combined effects of genetics and environmental factors contributing to T2DM through a systems biology framework will enable the advancement of personalized medicine.
Summary

The main findings of the studies included in this thesis are:

- TXNIP gene expression is reciprocally regulated by insulin and by glucose.
- TXNIP expression levels are consistently elevated in humans with T2DM and pre-diabetes.
- Elevation in TXNIP expression can inhibit glucose uptake.
- In a panel of 4,450 Scandinavian individuals, we found no evidence for association between common genetic variation in the TXNIP gene and T2DM.
- Genes from various biochemical pathways were related to insulin sensitivity. Notably, genes from the mTOR signaling pathway correlated positively while genes involved in extracellular matrix structural constituent such as focal adhesion, cell communication, and ECM-receptor pathways were inversely correlated with insulin sensitivity.
- Expression of CPT1B was positively and that of LEO1 inversely correlated with insulin sensitivity, a finding which was replicated in an independent study of 9 non-diabetic men.
- VO$_{2\text{max}}$ closely reflects expression of OXPHOS genes, particularly that of NDUFB5 and ATP5C1 in skeletal muscle.
- The expression of NDUFB5 and ATP5C1 was increased with exercise training and decreased with aging. High expression of these genes reflects good muscle fitness.
- In contrast, expression of AHNAK was associated with low VO$_{2\text{max}}$, decreased with exercise training, increased with aging. High expression of AHNAK reflects poor muscle fitness.
- A variant (rs27582) in CAST gene is associated with T2DM and insulin resistance.
Conclusions

- TXNIP regulates both insulin-dependent and insulin-independent pathways of glucose uptake in human skeletal muscle. Combined with recent studies that have implicated TXNIP in pancreatic beta-cell glucotoxicity, our data suggest that TXNIP might play a key role in defective glucose homeostasis preceding overt T2DM.
- A high capacity of fat oxidation in mitochondria is reflected by a high expression of CPT1B which is a marker of insulin sensitivity.
- VO$_{2\text{max}}$ closely reflects expression of OXPHOS genes while expression of AHNAK was associated with low VO$_{2\text{max}}$.
- A variant (rs27582) in CAST gene is associated with T2DM and insulin resistance. More generally, we proposed a novel approach for prioritizing SNPs from GWA studies.
Populärvetenskaplig sammanfattning

Omkring 250 miljoner människor lider av diabetes mellitus typ 2 (T2DM) i världen i dag och förekomsten av sjukdomen ökar i hela världen på grund av minskad fysisk aktivitet, ökat energintag och den åldrande befolkningen. T2DM är ett sjukdomstillstånd där blodglukoskoncentrationen är kroniskt förhöjd. T2DM kännetecknas av nedsatt insulinsekretion och verkan, den senare manifesteras som perifer och hepatisk insulinresistens (dålig dämpning av glukosproduktionen från levern av insulin). Riskfaktorer för utveckling av T2DM omfattar genetiska (familjehistoria av sjukdomen) och icke-genetiska faktorer såsom låg födelsevikt, fetma, högt födointag, fysisk inaktivitet, och åldrande. Därför betraktas T2DM som en av flera polygeniska sjukdomar där gemensamma förändringar i flera gener samverkar för att orsaka sjukdom när de utsätts för en miljö av för mycket mat och för lite motion (epistasis). Flera storskaliga genetiska metoder har använts för att identifiera gener som predisponerar för T2DM. DNA microarrays används för att mäta nivåerna av mRNA i biologiska prover för mer än 20,000 gentranskript. Helgenomskannar (GWA) erbjuder en objektiv metod för att identifiera genetiska varianter som påverkar känsligheten för att få sjukdomen. Denna avhandling har fokuserat på att analysera det genetiska bidraget till T2DM genom att använda storskaliga genetiska metoder med en särskilt fokus på analys av gentranskript i olika vävnader, huvudsakligen muskel.

I paper I har vi identifierat TXNIP som en gen vars uttryck är kraftfullt minskat av insulin, men stimulerat av glukos. Hos friska individer, var uttrycket omvänt korrelerat till kroppens totala glukosupptag. Tvingat uttryck av TXNIP i odlade adipocyter minskar avsevärt glukosupptaget, medan dämpning med RNA-interferens i adipocyter och i skelettmuskulatur förbättrar glukosupptaget, vilket bekräftar att genprodukten också är en regulator av glukosupptaget. TXNIP uttrycket är genomgående förhöjt i muskeln i prediabetiker och diabetiker, men i en panel bestående av 4,450 skandinaviska individer, kunde vi inte hitta några bevis för ett samband mellan vanliga genetiska variationer i TXNIP genen och T2DM. Slutsatsen var att TXNIP reglerar både insulin-beroende och insulin-oberoende vägar av glukosupptag i människans skelettmuskulatur. I kombination med nyligen genomförda studier som involverade TXNIP i glukostoxicitet av bukspottkörtelns β-cell, tyder våra data på att TXNIP kan spela en nyckelroll i defekt glukoshomeostas föregående fastställt T2DM.

I paper II undersökte vi molekylära mekanismer i samband med känsligheten för insulin i skelettmuskulatur genom att jämföra skelettmuskulaturens genuttryck med fysiologiska mätningar av insulinkänsligheten. Vi har identifierat 70 gener positivt och 110 gener omvänt korrelerade med insulinkänsligheten i människans skelettmuskulatur.
Anmärkningsvärt är att gener som är inblandade i en signalväg hos däggdjur target-of-rapamycin var positivt medan gener som kodar för extracellulär matrix strukturella komponenter såsom extracellulär matrix-receptor, cellkommunikation, och fokal vidhäftningssignalvägar var omvänt korrelerad med insulin känslighet. Mer specifikt var uttrycket för CPT1B positivt och det för LEO1 omvänt korrelerat med känsligheten för insulin, ett konstaterande som reproducerades i en oberoende undersökning av 9 icke-diabetiska män. Dessa data tyder på att en hög kapacitet av fettoxidation i mitokondrierna återspeglas i ett högt uttryck av CPT1B som är en markör för insulinkänsligheten.

I paper III undersökte vi molekylära mekanismer i samband med maximal syreupptagning (VO₂max) och typ 1 fiber i människans skelettmuskulatur. Vi har identifierat 66 gener positivt och 83 gener omvänt korrelerade med VO₂max och 171 gener positivt och 217 gener omvänt korrelerade med andel av typ 1 fiber i människans skelettmuskulatur. Gener som är inblandade i oxidativ fosforylering (OXPHOS) visade högt uttryck hos individer med hög VO₂max, medan det motsatta inte var fallet hos individer med låg VO₂max. I stället var gener som AHNAK och BCL6 förenade med låg VO₂max. Uttrycket av OXPHOS generna NDUFB5 och ATP5C1 ökade med träning och minskade med åldrande. Uttryck av AHNAK i skelettmuskel minskade med träning och ökade med åldrande. Dessa resultat visar att VO₂max nära återspeglar uttryck av OXPHOS gener särskilt av NDUFB5 och ATP5C1 i skelettmusklar och högt uttryck av dessa gener tyder på god muskelkondition. I motsats var ett högt uttryck av AHNAK förenat med en låg VO₂max och dålig muskelfitness.

I paper IV kombinerade vi GWA studier för T2DM med profilering av genuttrycket i målvävnaderna (buksportkörtel, fettvävnad, lever och skelettmuskulatur) för att hitta nya riskgener för T2DM. Genom att använda denna metod identifierade vi en enbaspolymorf (SNP) (rs27582) i CAST genen som associerades med T2DM i 16,061 svenska personer och denna association är tydligast i magra individer. Vi har även sett en association mellan denna SNP och insulinresistens. Mer allmänt har vi föreslagit en ny strategi för prioritering av SNPs från GWA studier.
I would like to acknowledge and extend my sincere thanks to several people who greatly contributed to this work in various ways. Their expertise, experience, and support have made this work more valuable.

Leif Groop, my supervisor for your support, enthusiasm, advice, generosity, and patience. I am really grateful for your instrumental guidance throughout the years. I have always learnt from you and truly admire your brilliant knowledge and memory as well as your art of scientific writing. You have inspired me in a lot of ways, above and beyond the scope of our work. I would also like to thank Birgitta Groop for your kindness and exceptional hospitality.

Vamsi Mootha, my co-supervisor, thank you for your essential suggestions, discussions, critiques, and for sharing your great knowledge in several different scientific fields. I absolutely admire your analytical and management skills. I am indebted to you for all your help and guidance.

Thanks Allan Vaag for your help, for discussing, critically reviewing manuscripts and providing your invaluable feedback. Thanks for allowing me to have access to your well-characterized muscle biopsy materials.

Thanks to all my co-authors, Martin Ridderstråle, Charlotte Ling, Targ Elgzyri, Juleen Zierath, David Altshuler, Richard Lee, Charlotte Granhall, Henrietta Nittby, Richa Saxena, Christine Ladd, Christian Schulze, Michael Mazzini, Anna Krook, Marie Björnholm, Hans Tornqvist, Bo Isomaa, Tiinamaija Tuomi, Göran Berglund, Peter Nilsson, and Natalie Hiscock, for ongoing outstanding help and discussions.

I especially would like to thank, Emma Nilsson and William A. Chutkow, for your great work and sharing your expertise in wet labs.

Thanks to my colleagues at Steno Diabetes Center in Denmark: Pernille Poulsen, Christine Bjørn Jensen, Jette Bork-Jensen, Heidi Storgaard, and Amra Alibegovic for sharing your knowledge and materials.

I would also like to express my sincere thanks to my parents, Mahendra and Surbhi, for your endless love, patience, encouragement, and support. My dearest sister, Avani, thanks for all your insights and motivation every step of the way. Thanks to my brother-in-law, Rakesh, for lots of fun and laughter. And to my sweet niece, Tulsi, thanks for always making me feel so special and significant.

I thank Avinash Abhyankar for your great friendship and sharing your knowledge in medicine. Thanks for providing me warm homely atmosphere. I knew you were there to support me during crisis of Indian cricket team. I would also like to thank his wife, Amruta, for cooking delicious Indian food for me and their cutest son, Yash, for playing with me.
Thank you, Ekaterine Bakhtadze for your invaluable friendship and your generosity. You are incredible. Thanks for cooking delicious Georgian food and always providing me boundless support and fun. I sincerely appreciate your great help with many things throughout these years.

Thanks Damon Toijar for being a great friend and providing me lots of support. I would also like to thank your family members for welcoming me in your family and considering me a part of it.

I would like to thank Peter Almgren for your help and sharing your statistics knowledge. I very much appreciate our scientific and non-scientific discussions. I have learnt many things from you. Thanks to your family for exceptional cordiality.

Anna Jonsson, I appreciate your friendship and help. Thanks for cooking luscious food. I will definitely miss your answer “No” to all my questions.

I thank Peter Osmark for proof-reading my manuscripts and thesis. Thanks for giving me a historical perspective of science and sharing your in-depth knowledge of the stock market. I really appreciate your support and enthusiasm to my exercise related activities including power Yoga and Tai Chi.

Thanks, Ola Hansson for sharing your in-depth knowledge of molecular biology. I have learnt many things from you. I greatly appreciate our discussions and your generous nature.

Nicolo Pulizzi, I thank you for your friendship, for being a great roommate, and for sharing your expertise in medicine, and lots of fun.

Thanks, Hee-Bok Park for being a great friend, and sharing your expertise in statistics, and always giving me right advice all these years. Also, thanks to your wife for cooking delicious spicy Korean food.

Thanks Valeriya Lyssenko for your friendship, joyful nature, a lots of fun and contributing your knowledge.

Lovisa Johansson, I thank you for your friendliness and help with experiments. I will definitely miss the chocolates.

Thanks Nael Shaat for your company, advice and our interesting conversations.

Jasmina Kravic, thanks for your friendship and help with translating all Swedish bureaucracy letters.

I appreciate Marketa Sjögren for solving many problems with genotyping and joining me for late evening dinners.

Ulrika Blom-Nilsson, thanks for helping me with many different things.

Beatrice Yang and Yuedan Zhou, my roommates, I thank you for your friendship and all the fun.

I thank my previous roommates Malin Eliasson, Johana, and Sara for being wonderful people that they are.
Thanks to Andreas Andersson, Inga-Britt Jönsson, Anne Vahaniemi, Jacqueline Postma, June Ljungberg, Alice McKenney, and Susan Henneberg for helping me with administrative work.

Johan Hultman and Mattias Borell, thanks for solving computer and server problems.

To other colleagues: Camilla Cervin, Johan Holmkvist, Anastasia Katsarou, Kristian Lynch, Lena Rosberg, Esa Laurila, Tereza Planck, Margareta Svensson, Tina Rönn, Mozghan Dorkhan, Mikael Lantz, Marju Orho-Melander, Ylva Wessman, Philippe Burri, Maria, Corrado Cilio, Albert Salehi, Marloes Dekker Nitert, Tord Ajanki, Anders Olsson, Jalal Taneera, Lina Johansson, Sabina Lindehammer, Caroline Bolmeson, Martin Kalis, and Maja, for being such nice people to be around both at and off work.

Thanks to several people from Karolinska Institutet: Per-Erik Jansson and Lena Lewin for helping me with administrative things and Erik Fredlund, Jens Nilsson, and Michael Green for Bioinformatics discussions.

I truly thank Anders Bloomberg for allowing me to be a part of the Research School in Genomics and Bioinformatics and for helping me make such good friends across Sweden to share and discuss my research with a different perspective.

Thanks to Alex Evilevitch for teaching me Biochemistry.

Several friends from Lund University: Yingchun Liu, Ahmed Hadad, Ewa Ohlsson, Helena Christianson, Mia Landin, Hanna Iderberg, and Clare O’Sullivan thanks for your help with several courses.

To my Georgian friends: Alexander and David thanks for fun and parties.

I would like to acknowledge Jatin Desai for our long 24 years’ friendship and for all your help, support, and guidance though different stages of my life.

Thanks Anjali Haryana for teaching me Yoga. Thanks Dino for your friendship. I would also like to thank Anjali’s mother, Geeta Haryana, for her outstanding warmth and cooking exquisite Punjabi food.

Thanks to Victoria Persson for your guidance with physical activity.

Thanks Mario Bitsanis for renting out your apartment and your friendship.

Thanks to my friends during my stay in the UK: Egmont, Shantanu, Karishma, Kunal, Rohit, Sotiris, Sunil, Darshak, Deepan, Noemi, Saloni, Keyur, Behher, and Ravi for your friendship. Many special thanks to Anjali Samani and Nirzari Parikh for introducing me to Bioinformatics.

Thanks to my friends during my Engineering training: Pratik, Rajan, Samir, Manan, Charmy, Ankita, Lopa, Bhavin, Dipali, Kedar, Karan, Animesh, Nilay, Neel, Karan, and Nirali for your friendship and great fun memories.

This thesis was supported by grants from the Swedish Knowledge Foundation through the Industrial PhD program in Medical Bioinformatics at the Center for Medical Innovations (CMI) at the Karolinska Institute, the Diabetes Programme at Lund University, Diabetesföreningen in Malmö, and the Medical Faculty at Lund University.
References


168. The Homeostasis Model Assessment (HOMA)


234. Hohaus, A. et al. The carboxyl-terminal region of ahnak provides a link between cardiac L-type Ca2+ channels and the actin-based cytoskeleton. *FASEB J* 16, 1205-16 (2002).