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

Parikh, Hemang

2009

Link to publication

Total number of authors:
1

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

DOKUMENTDATABLAD enl SIS 61 41 21

Organization LUND UNIVERSITY Department of Clinical Sciences Diabetes and Endocrinology Malmö University Hospital	Document name DOCTORAL DISSERTATION
	Date of issue January 23rd, 2009
	Sponsoring organization
Author(s) Hemang Parikh	

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

Abstract

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 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 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, 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. 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, AHNAK, CAST, Genome-wide association studies

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

Supplementary bibliographical information:	Language English
ISSN and key title: 1652-8220	ISBN 978-91-86059-96-5
Recipient's notes	Number of pages	Price
	Security classification	

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

"The only thing that interferes with my learning is my education."
Albert Einstein

To my mother

Contents

List of papers

I.   **Parikh H**, Carlsson E, Chutkow WA, Johansson LE, Storgaard H, Poulsen P, Saxena R, Ladd C, Schulze PC, Mazzini MJ, Jensen CB, Krook A, Bjornholm M, Tornqvist H, Zierath JR, Ridderstrale M, Altshuler D, Lee RT, Vaag A, Groop LC, Mootha VK: TXNIP regulates peripheral glucose metabolism in humans. *PLoS Med* 4:e158, 2007

II.   **Parikh H**, Elgzyri T, Hansson O, Alibegovic A, Hiscock N, Eriksson KF, Vaag A, Groop LC: Relationship between insulin sensitivity and gene expression in human skeletal muscle. 2008 [manuscript]

III.   **Parikh H**, Nilsson E, Ling C, Poulsen P, Almgren P, Nittby H, Eriksson KF, Vaag A, Groop LC: Molecular correlates for maximal oxygen uptake and type 1 fibers. *Am J Physiol Endocrinol Metab* 294:E1152-E1159, 2008

IV.   **Parikh H**, Lyssenko V, Granhall C, Isomaa B, Tuomi T, Berglund G, Nilsson P, Luthman H, Groop LC: Prioritizing genes for follow-up from genome wide association studies using information on gene expression in tissues relevant for type 2 diabetes mellitus. 2008 [manuscript]

Paper I and III are reproduced with the permission from the publishers.

List of papers not included in the thesis

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

Holmkvist J, Almgren P, **Parikh H**, Zucchelli M, Kere J, Groop LC, Lindgren CM: Haplotype construction of the FRDA gene and evaluation of its role in type II diabetes. *Eur J Hum Genet* 13:849-855, 2005

Lantz M, Vondrichova T, **Parikh H**, Frenander C, Ridderstrale M, Asman P, Aberg M, Groop LC, Hallengren B: Overexpression of immediate early genes in active Graves' ophthalmopathy. *J Clin Endocrinol Metab* 90:4784-4791, 2005

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

Johansson LE, Hoffstedt J, **Parikh H**, Carlsson E, Wabitsch M, Bondeson AG, Hedenbro J, Tornqvist H, Groop LC, Ridderstrale M: Variation in the adiponutrin gene influences its expression and associates with obesity. *Diabetes* 55:826-833, 2006

Shaat N, Karlsson E, Lernmark A, Ivarsson S, Lynch K, **Parikh H**, Almgren P, Berntorp K, Groop L: Common variants in MODY genes increase the risk of gestational diabetes mellitus. *Diabetologia* 49:1545-1551, 2006

Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ, Hughes TE, Groop LC, Altshuler D, Almgren P, Florez JC, Meyer J, Ardlie K, Bengtsson Bostrom K, Isomaa B, Lettre G, Lindblad U, Lyon HN, Melander O, Newton-Cheh C, Nilsson P, Orho-Melander M, Rastam L, Speliotes EK, Taskinen MR, Tuomi T, Guiducci C, Berglund A, Carlson J, Gianniny L, Hackett R, Hall L, Holmkvist J, Laurila E, Sjogren M, Sterner M, Surti A, Svensson M, Svensson M, Tewhey R, Blumenstiel B, Parkin M, Defelice M, Barry R, Brodeur W, Camarata J, Chia N, Fava M, Gibbons J, Handsaker B, Healy C, Nguyen K, Gates C, Sougnez C, Gage D, Nizzari M, Gabriel SB, Chirn GW, Ma Q, **Parikh H**, Richardson D, Ricke D, Purcell S: Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 316:1331-1336, 2007

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

increased risk of gestational diabetes mellitus. *Diabetologia* 50:972-979, 2007

Vondrichova T, de Capretz A, **Parikh H**, Frenander C, Asman P, Aberg M, Groop LC, Hallengren B, Lantz M: COX-2 and SCD, markers of inflammation and adipogenesis, are related to disease activity in Graves' ophthalmopathy. *Thyroid* 17:511-517, 2007

Abbreviations

acetyl-CoA	Acetyl coenzyme A
AHNAK	AHNAK nucleoprotein
ANOVA	Analysis of variance
ASK1	Apoptosis signal–regulating kinase 1
ATP	Adenosine triphosphate
ATP5C1	ATP synthase, H+ transporting, mitochondrial F1 complex, gamma polypeptide 1
BMI	Body mass index
BPS	Botnia prospective study
C/EBP	CCAAT/enhancer binding protein
CAST	Calpastatin
cDNA	Complementary deoxyribonucleic acid
CI	Confidence interval
CIR	Corrected insulin response
CPT1B	Carnitine palmitoyltransferase 1B
ddNTP	dideoxynucleotide triphosphate
DEXA	Dual energy X-ray absorptiometry
DI	Disposition index
DNA	Deoxyribonucleic acid
ECM	Extracellular matrix
ES	Enrichment score
EST	Expressed sequence tag
FDR	False discovery rate
FFA	Free-fatty acid
FH-	Without family history of T2DM
FH+	With family history of T2DM
FPG	Fasting plasma glucose concentration
FRET	Fluorescence resonance energy transfer
FWER	Family-wise error rate
G6P	Glucose-6-phosphate
GC-RMA	GC-content robust multi-array average
GEE	Generalized estimating equations
GLUT4	Glucose transporter 4
GO	Gene ontology
GRAIL	Gene relationships across implicated loci
GSEA	Gene set enrichment analysis
GWA	Genome-wide association

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

SD	Standard deviation
SEM	Standard error of the mean
siRNA	Short interfering RNA
SNP	Single nucleotide polymorphism
T2DM	Type 2 Diabetes Mellitus
TDT	Transmission disequilibrium test
Tg	Triglyceride
TNF	Tumor necrosis factor
TXN	Thioredoxin
TXNIP	Thioredoxin interacting protein
UDP	Uridine 5'-diphosphate
VCAM1	Vascular cell adhesion molecule-1
VO_{2max}	Maximal oxygen uptake
WebGestalt	Web-based gene set analysis
WHO	World health organization
ZDF	Zucker diabetic fatty
ZLC	Zucker lean control

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 (VO_{2max}) and type 1 fibers in human skeletal muscle. We identified 66 genes positively and 83 genes inversely correlated with VO_{2max} and 171 genes positively and

17

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_{2max}, whereas the opposite was not the case in individuals with low VO_{2max}. Instead, genes such as *AHNAK* and *BCL6* were associated with low VO_{2max}. 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_{2max} 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_{2max} 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 (β) = 0.048, $P = 0.017$) and higher HOMA-IR index (β = 0.044, $P = 0.025$) as well as lower insulin sensitivity index during OGTT (β = -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.

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 (VO_{2max}) 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 BC[1]. Diabetes mellitus is caused by defects in insulin secretion, insulin action, or both resulting from a complex interaction of genes, and environmental factors[2]. Diabetes mellitus is broadly classified into four different etiological categories designated as type 1 diabetes mellitus, T2DM, gestational diabetes mellitus, and other specific types[2]. 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 amputation[3].

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 2030[4]. There are large ethnic and geographic variations in the prevalence of T2DM[5,6]. Asians, Hispanics, African-Americans, and Pima Indians appear

to have much higher incidence of T2DM as compared to European populations[7]. Moreover, the "westernized lifestyle" is strongly associated with an increase in the prevalence of T2DM across all ethnic groups[7]. According to Centers for Disease Control and Prevention, incidence of T2DM in the United States has increased by 90% during the last 10 years[8]. 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[9].

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)[10] (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[11].

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

Glucose (mmol/l)	Impaired fasting glucose	Impaired glucose tolerance	Diabetes mellitus
FPG	≥ 6.1 to < 7.0	-	≥ 7.0
2 hours after OGTT	-	≥ 7.8 to < 11.1	≥ 11.1

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[12]. 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[5,6] as well as twins studies[13,14]. Moreover, the lifetime risk for developing T2DM in offspring of one T2DM parent is about 40%[15] and the ratio of the risk for a sibling of a T2DM individual to that of the general population (λ_s) is about 3.5-4[16,17].

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)[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[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[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)[21].
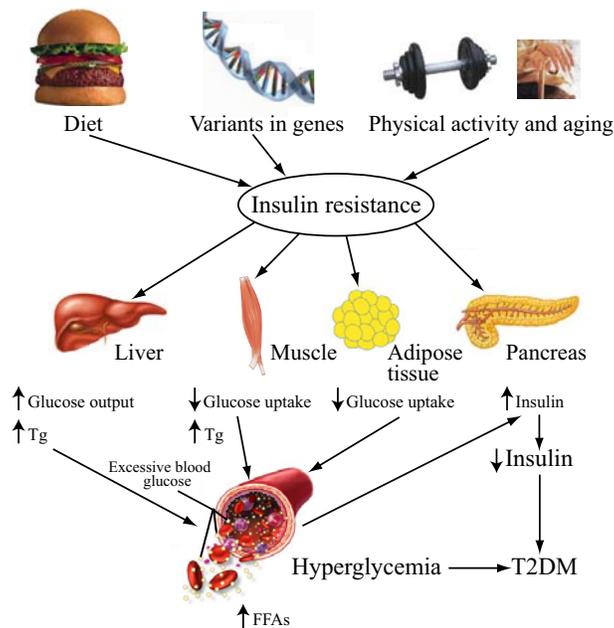
Figure 1: Pathophysiology of T2DM.

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 β-cells to overproduce insulin to restore normal glycemic levels but eventually impaired insulin secretion from pancreatic β-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)[22].
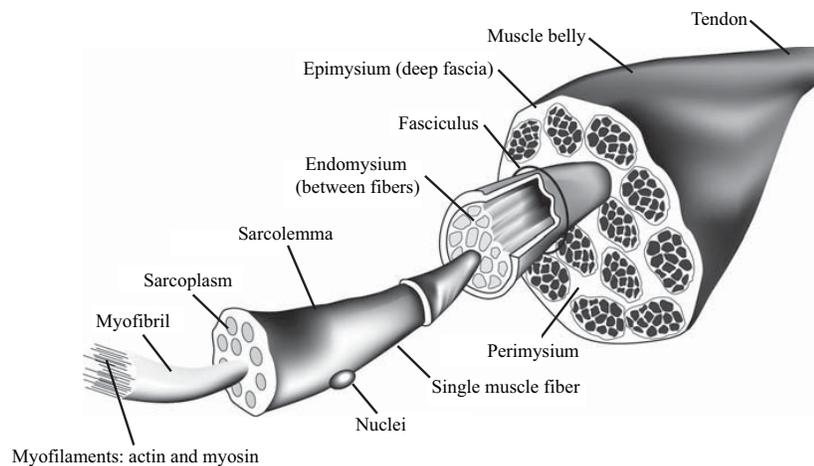
Figure 2: Basic structure of skeletal muscle.

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.*[22]

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[22,23]. 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[24] and predicts T2DM[25].

Pathophysiology of insulin resistance in skeletal muscle

Insulin resistance is a key feature of the pre-diabetic state and a predictor of T2DM[26,27]. Insulin resistance can be observed 10–20 years before the onset of overt T2DM[28,29]. It is also seen in healthy glucose tolerant first degree relatives of patients with T2DM[30]. Insulin resistance in skeletal muscle has been attributed to different pathological conditions such as impaired glycogen synthesis[31,32], accumulation of lipids with subsequent impairment of insulin signaling[33], and mitochondrial dysfunction and oxidative stress[34]. 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[35]. 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 mitochondira[36]. 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[26]. 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[37-39].
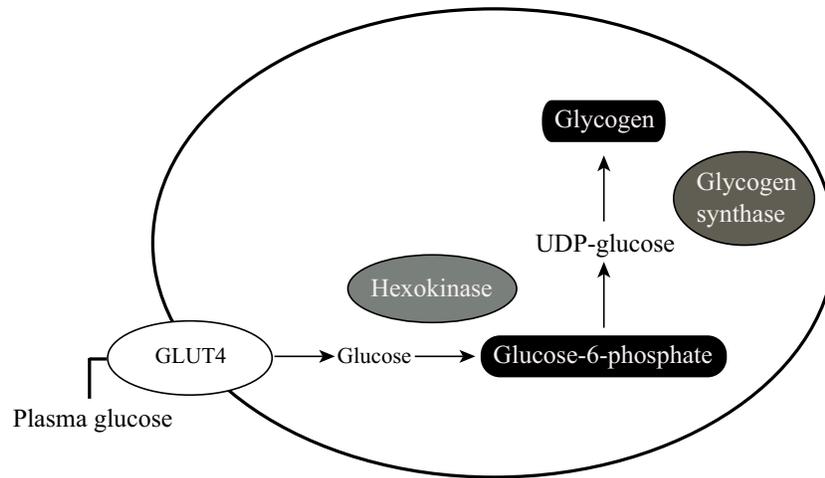
Figure 3: The pathway of muscle glycogen synthesis.

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 elevated[40]. 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-phophate would inhibit hexokinase II activity and increase intracellular glucose concentration and decrease glucose uptake (Figure 4).
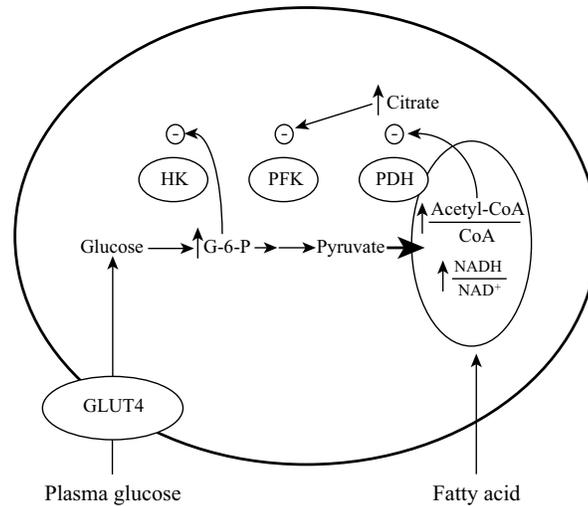
Figure 4: Mechanism of fatty acid–induced insulin resistance in skeletal muscle as proposed by Randle et al.[40]

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[42,43]. They demonstrated that elevated plasma FFAs concentrations cause a reduction in intracellular glucose concentration[42]. 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[44-46]. Also, FFAs produce low-grade inflammation by activating nuclear factor-kappa beta which leads to release of several proinflammatory and proatherogenic cytokines[47].
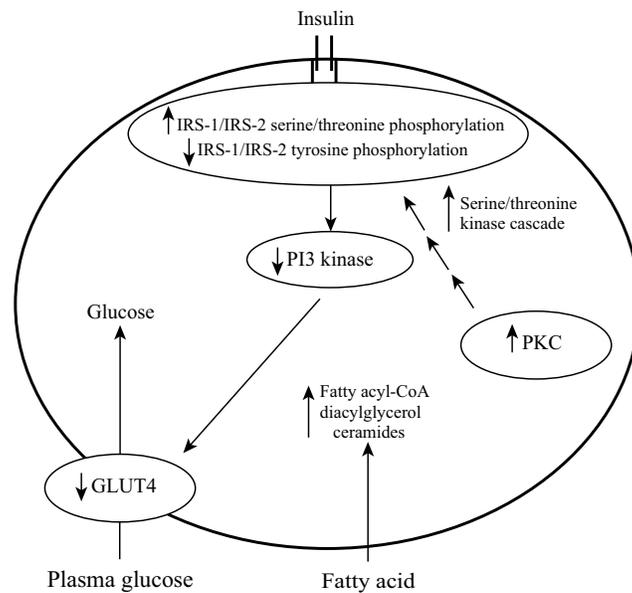
Figure 5: Proposed alternative mechanism for fatty acid–induced insulin resistance in human skeletal muscle.

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 defenses[48]. 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 capacity[49]. 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 stress[34]. Mitochondrial dysfunction and oxidative stress are associated with insulin resistance[34,50-55] (Figure 6). Also, the quantity and morphology of mitochondria are correlated with mitochondrial oxidative capacity[56]. Fewer and

smaller-sized mitochondria are observed in skeletal muscle of T2DM patients[57,58]. Mitochondrial dysfunction predisposes to intramyocellular lipid accumulation, which provides a link to insulin resistance[33] through the mechanisms described earlier in Figure 5.

In addition, several defects in insulin signaling are associated with insulin resistance[36].
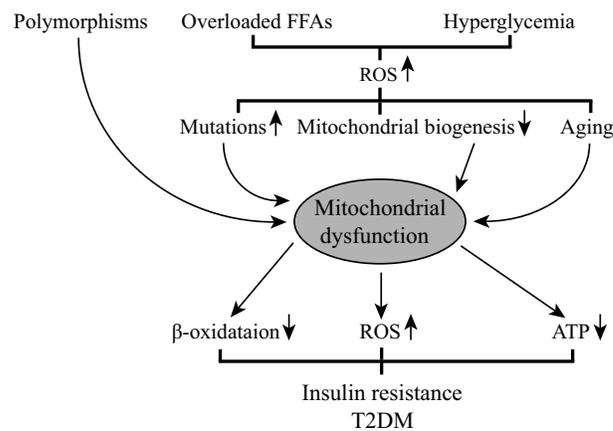
Figure 6: Mechanism of mitochondrial dysfunction.

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 β-oxidation and adenosine triphosphate (ATP) production and increased ROS production, resulting in insulin resistance and T2DM. This figure is adapted from Kim JA *et al.*[34]

Physical activity

Physical activity is defined as body movement produced by muscle action which substantially contributes to energy expenditure[59]. Physical inactivity is an independent risk factor for T2DM[60,61]. Also, several prospective studies have concluded that physically active individuals have a 33-50% lower risk of developing T2DM[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)[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[68].
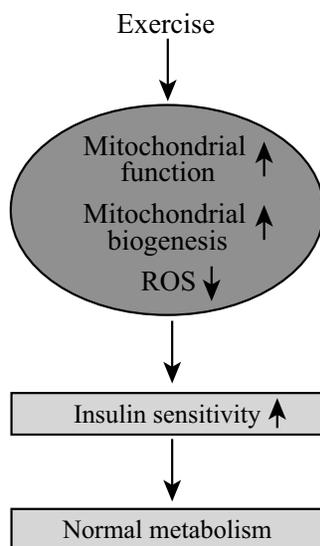
Figure 7: Improvement of mitochondrial function by exercise can improve insulin sensitivity, which leads to normal metabolism. This figure is adapted from Kim JA *et al.*[34]

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

VO$_{2max}$ 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$_{2max}$ provides useful information about long-term energy system capacity. Many factors influence VO$_{2max}$ *e.g.*, physical activity, heredity, gender, body composition, and age. Several studies have reported heritability of VO$_{2max}$ ranging from 40 to 70%[69-72]. Untrained individuals with T2DM have been shown to have lower VO$_{2max}$ compared to untrained healthy control individuals. Also, a low VO$_{2max}$ predicts T2DM[73,74]. There are several potential mechanisms by which a low VO$_{2max}$ could increase risk for T2DM such as mitochondrial dysfunction[75], a change in muscle fiber type[25], and insulin resistance[76]. In addition, there is a relatively strong positive correlation between insulin sensitivity and VO$_{2max}$[77], and exercise training can increase both VO$_{2max}$ and insulin sensitivity[78]. Intriguingly, one study suggested that after exercise, VO$_{2max}$ 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$_{2max}$ does not translate into improved insulin sensitivity[79]. Moreover, a high VO$_{2max}$ is associated with a high proportion of type 1 fibers[80]. We have also investigated molecular mechanisms associated with VO$_{2max}$ 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[81-83]. 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[84,85]. Identifying the inherited contribution to human disease involves linking genomic variation to clinical phenotype[86]. 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[87]. 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 (Figure 8A). 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)[88] (Figure 8B-D). However, 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**
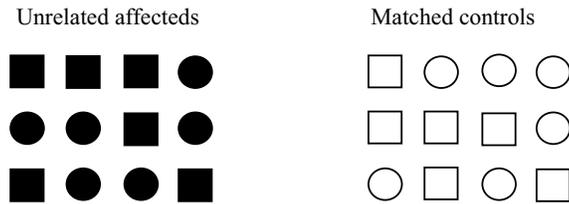
Ascertain affecteds and their parents
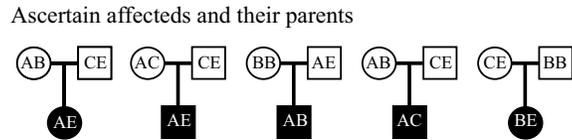
C   **Haplotype relative risk method (HRR)**

Alleles in affecteds   Control allels

AE, AE, AB, AC, BE   BC, CC, BE, BE, BC

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[89,90]. 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 florescence intensity of each spot from the samples always requires to be compared to controls. There are several platforms available for global gene expression profiling[91-93]. 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[94,95]. 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[96-98] 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.
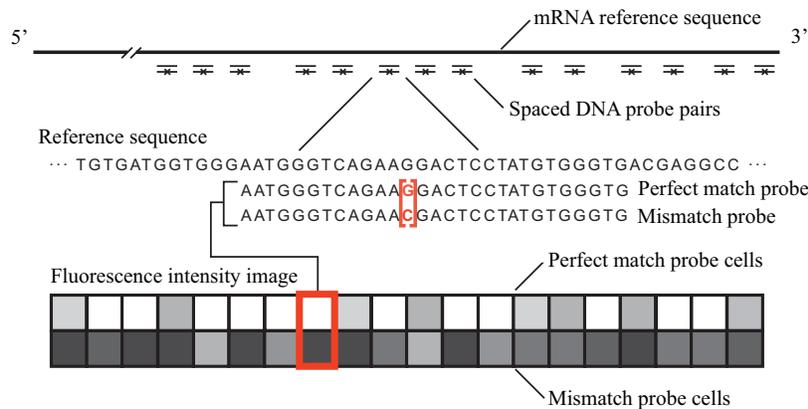
Figure 9: Expression probe and array design.

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 process[99-101]. The long 60-mer probes are designed *in silico* using empirical measurements that increases hybridization sensitivity[100,102]. 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 hybridization[91]. Nevertheless, there is a strong concordance between these two platforms[103].

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[104] and their non-diabetic first-degree relatives[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[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[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[108]. Background adjustment is employed to adjust observed probe intensities for non-specific hybridization and noise in the optical detection system[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[110]. Several approaches such as Affyproberminer[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[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[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[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[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[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[117], position-dependent nearest neighbor method[118], factor analysis for robust microarray summarization[119], and several other methods[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[120]. However, methods used for pre-processing can have a major impact on the results[121] and one study has reported only 27 to 36% overlap between different methods[120]. Therefore, the best approach might be to compare results from several of these methods to obtain a reliable list of differentially expressed genes[122].

For the two-color Agilent arrays, simple loess within array normalization without background subtraction method has shown to produce consistent results[123].
Moreover, data filtering should be conducted to eliminate data that are not reliably detected over all samples[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 *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[126-128]. The significance analysis of microarrays (SAM) algorithm is based on a modified *t*-test and estimates the FDR using permutations under the assumption that all null hypotheses are true[129]. The patterns from gene expression[130], linear models and empirical Bayes[131], weighted analysis of general microarray experiments[132], rank product method[133], data-adaptive test statistics[134], integrating differing statistics via a distance synthesis[135], and several other methods[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[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[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 use to explore the association after adjustment for covariates.

Pathway analysis of microarray data

Pathway analysis of microarray data evaluates gene expression profiles of *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[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 (*i.e.*, biological process, molecular function, and cellular component) by using overlap statistics such as the cumulative hypergeometric distribution[139]. Several other methods are also available for inferring GO categories from a group of differentially expressed genes[140-142]. However, there are a number of methodological issues to consider in the interpretation of the results based on pathway analysis[143].

In addition, global gene expression profiling data can also be use for class prediction and class discovery[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[145]. Several platforms are available for GWA studies including the Affymetrix Genechip® and Illumina BeadChip platform[146]. Several statistical issues need to be considered in the analysis of such large-scale genetic studies[147,148]. A number of GWA studies have identified novel genomic regions influencing risk for T2DM[149-153]. However, the challenge remains to prioritize SNPs from GWA studies for further replication in other samples[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[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 (*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_{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/m^2), paper I-study B consisted of 6 non-diabetic volunteers (4 men and 2 women, age 66 ± 3 years, BMI 28 ± 2 kg/m^2), 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/m^2). 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 to characterize the first-phase insulin response to glucose. In paper I-study B, skeletal muscle biopsies were obtained before and after a 3-hour hyperinsulinemic euglycemic clamp.
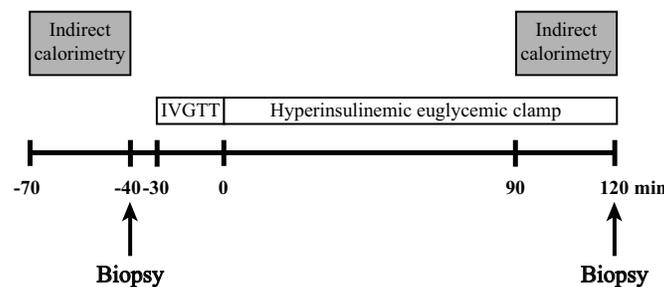

Figure 10: Design of the hyperinsulinemic, euglycemic clamp protocol.

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[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[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$_{2max}$ 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[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)[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[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)[104,107,160,161]. 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)[162] 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 2,830 T2DM patients from the Malmö Diabetes Registry[163] and 3,550 controls from the Malmö Diet and Cancer study[164]. All T2DM patients were of Scandinavian origin, had age at onset > 35 years, connecting peptide $\geq$ 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[165]. The BPS included 2,770 participants, 138 of whom developed T2DM during a median follow-up period of 7.6 years[17]. 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 (m^2). 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[10].

Oral glucose tolerance test (OGTT)

The OGTT is a method to determine how quickly glucose is cleared from the blood[166]. 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 $10,000/\sqrt{(}$fasting plasma-glucose x fasting plasma-insulin x mean $OGTT_{glucose}$ x mean $OGTT_{insulin}$)[167]. The basal insulin resistance index (HOMA-IR) was calculated from fasting insulin and glucose concentrations[168]. 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 x (glucose 30 min or 40 min in MPP -3.89)]) or as disposition index, *i.e.* insulin secretion adjusted for insulin sensitivity (DI = CIRxISI)[169].

Hyperinsulinemic euglycemic clamp

The hyperinsulinemic euglycemic clamp is a gold standard method for investigating and quantifying insulin resistance[170]. 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[171,172].

Indirect calorimetry

Indirect calorimetry is a method to quantify energy expenditure and respiratory quotient by measuring oxygen consumption and carbon dioxide production[173]. 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[174,175].

Maximal oxygen uptake (VO_{2max}) measurement

In paper III, VO_{2max} 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. VO_{2max} 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[176]. 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[177]. 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[178] 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 Transciptase (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 NIDDM1I 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[179]. 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[180]. 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.

Polymerization

Forward primer   R   Probe   Q

5' ⟶ 3'
3' ————————————————— 5'
5' ————————————————— 3'
5' ⟵
Reverse primer

Strand displacement

R

Q

5' ⟶
3' ————————————————— 5'
5' ————————————————— 3'
⟵ 5'

Clevage

R

Q

5' ⟶
3' ————————————————— 5'
5' ————————————————— 3'
⟵ 5'

Polymerization completed

R   Q

5' ⟶
3' ————————————————— 5'
5' ————————————————— 3'
⟵ 5'
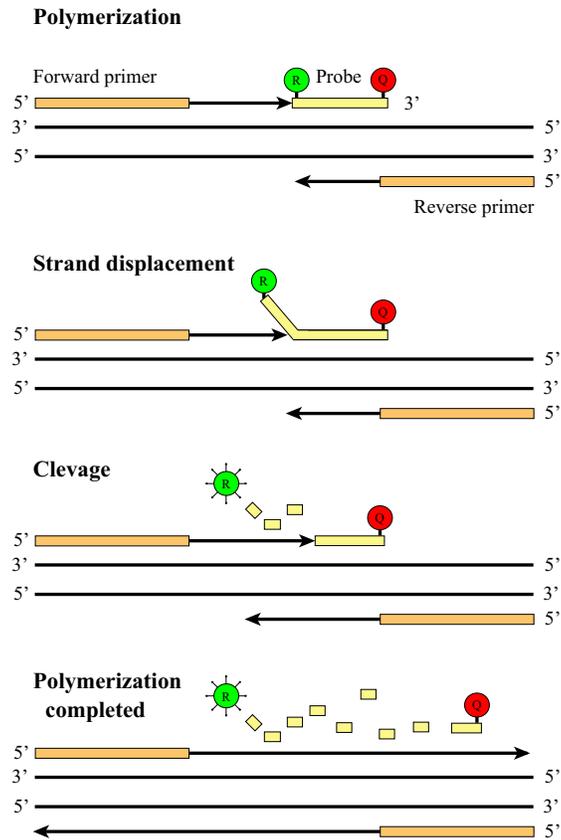
Figure 11: Principle of real-time PCR.

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 fluorescence 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[181]. 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[182,183], and quality scores were assigned. Sequence assembly and analysis were performed using the Staden software package[184].

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[185]. 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[186]. 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 χ^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 χ^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[155,192]. 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[125]. In paper I, the SAM method[129] 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[104,105,161,193,194] 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[195]. To identify genes whose expression correlated consistently with VO_{2max} and percentage of type 1 fibers, we have used normalized gene expression data from our previous microarray study in paper III[104]. 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_{2max} 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[104,107,160,161] 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[138] of metabolic pathways was performed between two groups divided by mean values of VO_{2max} as well as percentage of type 1 fibers in paper III. Also, The KEGG pathway analysis of genes of interest was performed using WebGestalt[139] 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).
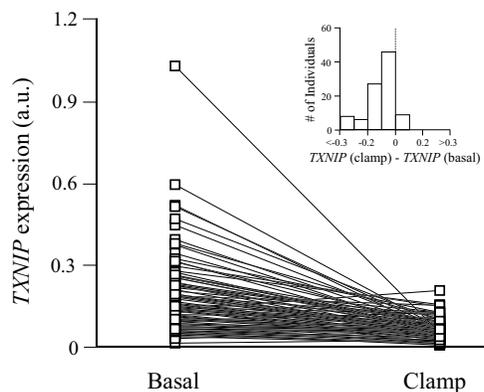
Figure 12: Effects of insulin on TXNIP expression in human muscle *in vivo.*

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[104,105] and found that *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, *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 β-cells and results in insulin deficiency and hyperglycemia[194]. *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[161], insulin treatment of streptozotocin-diabetic mice failed to suppress *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 *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 *in vitro* assays of *TXNIP* expression in a human adipocyte cell line that is known to respond to insulin by stimulating glucose uptake. *TXNIP* expression increased following elevations in glucose concentration, and was suppressed by insulin. Together, these cell culture studies confirm that *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[104], we found that *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 *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 *TXNIP* expression, zygosity, birth weight, percentage of body fat, BMI, and VO_{2max}. Basal (P = 0.02) and insulin-stimulated (P < 0.01) *TXNIP* expression as well as BMI (P < 0.01) were inversely correlated, while VO_{2max} (P < 0.01) was positively correlated with insulin-stimulated glucose uptake (Table 2). These data are consistent with the notion that *TXNIP* is an independent determinant of insulin-stimulated glucose uptake in humans.

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

Factors	Regression coefficients	95 % CI		P
		Low	High	
TXNIP before clamp	-4.18	-7.76	-0.61	0.02
TXNIP after clamp	-20.29	-35.20	-5.39	< 0.01
BMI	-0.35	-0.53	-0.16	< 0.01
VO_{2max}	0.11	0.03	0.18	< 0.01

The GEE model was used to determine whether basal and insulin-stimulated *TXNIP* expression, age, sex, birth weight, zygosity, percentage of body fat, BMI, or VO_{2max} influence insulin-stimulated glucose uptake.

50

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.*, *CAMK2A* (r = 0.37, $P = 0.03$) from the Wnt signaling pathway, *UCP2* (r = 0.56, $P < 0.01$) from a family of inner mitochondrial membrane proteins, and *CPT1B* (r = 0.47, $P < 0.01$) from peroxisome proliferator-activated receptor (PPAR) signaling pathway. We found six genes (*COL1A1*, *COL1A2*, *COL3A1*, *LAMA4*, *MYL6*, and *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. *SIRT2* (r = 0.95, $P < 0.01$), *FBXW5* (r = 0.95, $P < 0.01$), *RAB11FIP5* (r = 0.93, $P < 0.01$), *CPT1B* (r = 0.85, $P = 0.02$), *C16orf86* (r = 0.79, $P = 0.04$), *UCKL1* (r = 0.77, $P = 0.04$), and *ARFGAP2* (r = 0.76, $P < 0.05$) were positively correlated while *ZNF613* (r = -0.81, $P = 0.03$), *UTP6* (r = -0.81, $P = 0.03$), and *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 *CPT1B* was increased in response to treatment with a PPARδ agonist[196] and this agonist is shown to increase mitochondrial biogenesis in muscle by enhancing fatty acid catabolism[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 *CPT1B* which is a marker of insulin sensitivity.

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_{2max} and type 1 fibers in human skeletal muscle.

To identify genes whose expression was consistently correlated with VO_{2max} and type 1 fibers, we reanalyzed our previously published gene expression data from human skeletal muscle in paper III-study A[104]. We observed a modest positive correlation between VO_{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_{2max}. Among positively correlated genes, we found 13 genes involved in OXPHOS using WebGestalt ($P < 1 \times 10^{-19}$), *i.e.*, *NDUFS3*, *NDUFS8*, *NDUFB4*, *NDUFB5*, and *NDUFB11* from complex I, *UQCRB* and *UQCRC2* from complex III, *COX5A*, *COX5B*, and *COX6A1* from complex IV, and *ATP5C1*, *ATP5G3*, and *ATP6V0C* from complex V (Figure 13A). There were no OXPHOS genes among genes inversely correlated with VO_{2max}. Instead, this list comprised genes with varying function, including *AHNAK*

52

(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[139] of metabolic pathways by comparing groups with VO_{2max} below or above the mean of VO_{2max} 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_{2max} group but none in the low-VO_{2max} group (Table 3). Among the list of genes correlated with VO_{2max}, 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[193].

Table 3 Gene sets enriched in high VO_{2max} group

Gene Set	Size	ES	NES	*P*	FDR, q-value	FWER, *P*
OXPHOS-CR	77	0.66	1.91	0.002	0.049	0.097
Electron transport chain	80	0.65	1.89	0.006	0.031	0.115
MAP00190 oxidative phosphorylation	42	0.56	1.84	0.004	0.039	0.191

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 ($P < 1 \times 10^{-16}$), *i.e.*, *NDUFB4*, *NDUFC2*, *UQCRB*, *UQCRC2*, *UQCRQ*, *COX5A*, *COX7A2*, *COX7B*, *ATP5C1*, *ATP5J2*, *ATP5G3*, *ATP5L*, *ATP5O*, *ATP6V1D*, and *ATP5F1* (Figure 13B).

| | Complex I | Complex II | Complex III | Complex IV | Complex V |

(a) VO2max (ml/kg/min)

Complex I	Complex III	Complex IV	Complex V
NDUFS3	UQCRB	COX5A	ATP5C1
NDUFS8	UQCRC2	COX5B	ATP5G3
NDUFB4		COX6A1	ATP6V0C
NDUFB5			
NDUFB11			

(b) Type 1 fibers (%)

Complex I	Complex III	Complex IV	Complex V
NDUFB4	UQCRB	COX5A	ATP5C1
NDUFC2	UQCRC2	COX7A2	ATP5J2
	UQCRQ	COX7B	ATP5G3
			ATP5L
			ATP5O
			ATP6V1D
			ATP5F1

Figure 13: List of genes from the oxidative phosphorylation pathway positively correlated with (A) Maximal oxygen uptake (VO_{2max}). 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 *NDUFB5*, *ATP5C1*, *AHNAK*, and *BCL6*, for the following reasons: (1) *NDUFB5* because its expression was positively correlated with VO_{2max} and increased after training, and also because a study has shown that expression of *NDUFB5* decreases after a high-fat diet[198]; (2) *ATP5C1* because its expression was positively correlated with VO_{2max} and type 1 fibers and increased after training; (3) *AHNAK* because its expression was strongly inversely correlated with VO_{2max} 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 *NDUFB5* (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_{2max} as a function of age, sex, BMI, zygosity, and expression of individual genes (*NDUFB5*, *ATP5C1*, *AHNAK*, and *BCL6*). Expression of *NDUFB5* ($P = 0.03$; Table 4) and *ATP5C1* ($P = 0.02$; Table

4) but not of *AHNAK* (regression coefficient β = -0.10, 95% confidence interval (CI) = -1.34 –1.14, *P* = 0.87) or *BCL6* (β = 3.13, 95% CI = -2.65– 8.91, *P* = 0.29) was significantly correlated with VO_{2max}.

Table 4 Genetic and non-genetic factors influencing VO_{2max} in humans

Factors	Regression coefficients	95 % CI		*P*
		Low	High	
NDUFB5				
Expression	0.44	0.04	0.84	0.03
BMI	-0.69	-0.96	-0.42	< 0.01
Age	-0.29	-0.36	-0.22	< 0.01
Sex	-6.35	-8.69	-3.99	< 0.01
ATP5C1				
Expression	0.95	0.15	1.74	0.02
BMI	-0.85	-1.02	-0.67	< 0.01
Age	-0.28	-0.35	-0.21	< 0.01
Sex	-7.44	-9.93	-4.95	< 0.01

The generalized estimating equation model was used to determine whether BMI, age, sex, zygosity, and expression of individual genes (*NDUFB5* and *ATP5C1*) influence VO_{2max}.
Abbreviations: CI, confidence interval.

In conclusion, VO_{2max} closely reflects expression of OXPHOS genes, particularly that of *NDUFB5* 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_{2max}, 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

14) 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[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).
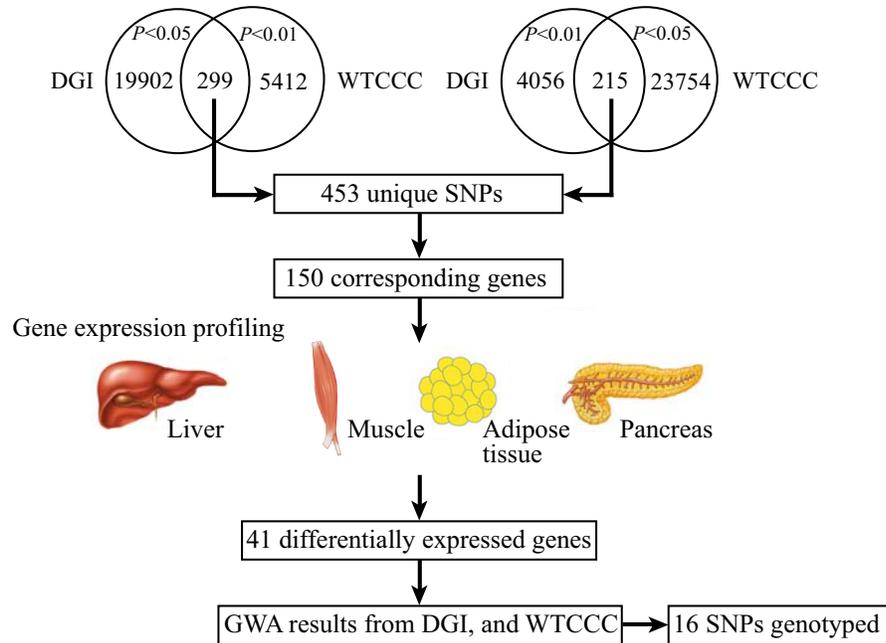

Figure 14: Schematic diagram of prioritizing SNPs from GWA studies.

453 SNPs were associated with T2DM with $P < 0.01$ in at least one of the GWA study[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[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

Gene Symbol	Entrez GeneID	Dataset	Fold-change	*P*
ERAP1	51752	T2DM vs. NGT human pancreatic islets	2.18	0.018
		Diabetic vs. healthy rat pancreas	0.83	0.006
		NIDDM1I vs. F344 rat pancreas	1.26	0.025
HNRNPC	3183	T2DM vs. NGT human pancreatic islets	0.55	0.007
		NIDDM1I vs. F344 rat pancreas	0.92	0.011
		T2DM vs. NGT human skeletal muscle	1.36	0.029
NCOR1	9611	T2DM vs. NGT human pancreatic islets	1.28	0.021
		Diabetic vs. healthy mice skeletal muscle	1.47	< 0.001
CAST	831	T2DM vs. NGT human pancreatic islets	1.31	0.048
		Liver of 12 weeks old ZDF vs. ZLC	-0.68[§]	-
PQLC1	80148	T2DM vs. NGT human pancreatic islets	1.44	0.004
		Diabetic vs. healthy rat pancreas	0.81	< 0.001
CADPS	8618	NIDDM1I vs. F344 rat pancreas	1.19	0.043
		Diabetic vs. healthy rat pancreas	0.60	0.011
TGFB3	7043	NIDDM1I vs. F344 rat pancreas	1.16	0.009
		Diabetic vs. healthy rat pancreas	0.77	< 0.001
KCNS3	3790	Diabetic vs. healthy rat pancreas	0.66	0.039
		T2DM vs. NGT human skeletal muscle	0.83	0.029
SAMD4A	23034	T2DM vs. NGT human skeletal muscle	0.59	0.024
		NIDDM1I vs. F344 rat pancreas	0.63	0.019
FAM45A	55855	NIDDM1I vs. F344 rat pancreas	0.88	0.023
		Muscle of 12 weeks old ZDF vs. ZLC	-0.62[§]	-
ACTN1	87	NIDDM1I vs. F344 rat pancreas	1.18	0.015
		Muscle of 12 weeks old ZDF vs. ZLC	0.62[§]	-

[§]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, *CAPN10*, 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 (β) = 0.048, $P = 0.017$) and higher HOMA-IR index (β = 0.044, $P = 0.025$) as well as lower insulin sensitivity index during OGTT (β = -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 VO_{2max} 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[200]. 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[201]. 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[202]. 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[203].

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

60

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[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[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 *IGF2BP2*, *CDKAL1*, *TSPAN8*, and *NOTCH2* for which neighboring SNPs have consistently shown to be associated with T2DM in several populations[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[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 *TXNIP* gene and T2DM; another study has shown that a genetic variation within the *TXNIP* gene is associated with hypertriglyceridemia and increased diastolic blood pressure in patients with T2DM[211].
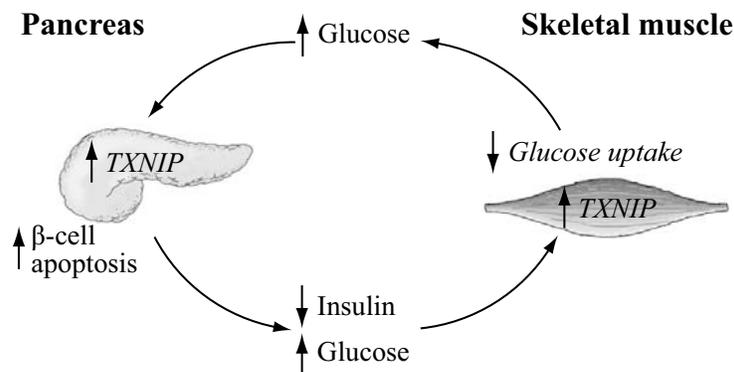
Figure 15: Role of TXNIP in glucose toxicity in the β-cell and in impaired glucose uptake in the periphery.

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[210]. 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 oxidoreducatase that associates with thioredoxin reductase and thioredoxin peroxidase to reduce oxidized proteins and scavenge free radicals[212-214]. TXNIP interacts directly with the reduced form of thioredoxin and thereby negatively regulates the expression and function of thioredoxin[215-218]. In addition, TXNIP has multifunctional roles in a variety of cellular functions including growth inhibition, transcriptional regulation, natural killer cells development, and lipid metabolism[219]. 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[220].

TXNIP expression is induced by glucocorticoids[221] and by glucose, and TXNIP elevations can stimulate ROS production[222] and suppresses 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[223]. 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[224], and diabetic nephropathy[225] 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[226,227] (Figure 16). 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[228]. Taken together, TXNIP plays crucial role in cancer and metabolic diseases.
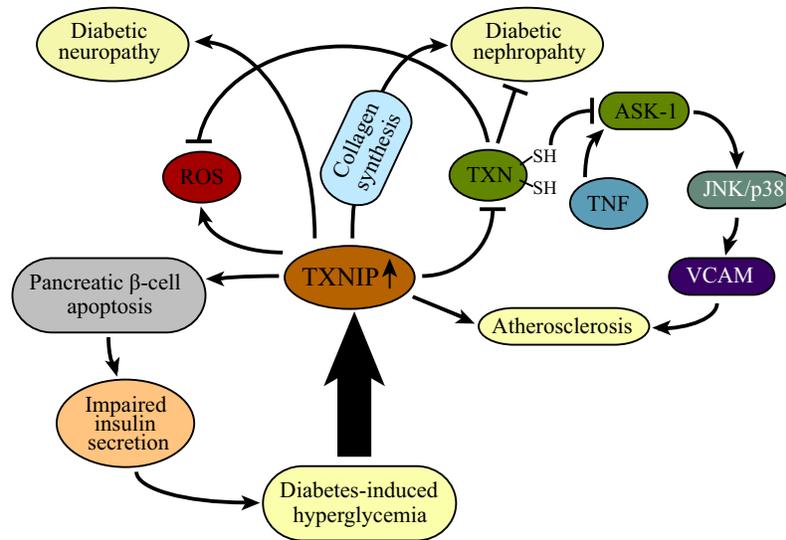
Figure 16: Schematic diagram showing the role of TXNIP in glucose and its related diseases.

Hyperglycemia-induced reactive oxygen species (ROS) in T2DM up-regulates TXNIP and suppresses thioredoxin (TRX) mediated antioxidant function[222]. Over-expression of TXNIP enhances β-cell death and impairs insulin secretion[229,230], which may result in further aggravation of T2DM. In T2DM patients, increased TXNIP is associated with chronic diabetic neuropathy[224] and diabetic nephropathy[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[226,227]. This figure is adapted from Kaimul AM *et al.*[228]

CPT1B

In paper II, we have identified that expression of *carnitine palmitoyltransferase 1B* (*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 *CPT1B* as well as *UCP2* in skeletal muscle as promoter regions of these genes contain binding motifs for the transcription factor CCAAT/enhancer binding protein (C/EBP)[231]. Moreover, expression of these genes was up-regulated in the *Klf5*-knockout heterozygous

mice ($Klf5^{+/-}$) which is resistant to high fat-induced obesity, hypercholesterolemia, and glucose intolerance[231]. The expression of *CPT1B* in skeletal muscle was increased in response to treatment with a PPARδ agonist[196]. This PPARδ agonist is shown to increase mitochondrial biogenesis in muscle by enhancing fatty acid catabolism[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_{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[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_{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/and related metabolic traits in the DGI scan[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_{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 carboxylterminal domain of which has been ascribed a stabilizing effect on muscle contractility via interaction with actin[232]. *AHNAK* also seems to mediate activation of

phospholipase C and release of arachidonic acid through protein kinase C[233]. This opens up the interesting possibility that *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^{2+} channels in cardiomyocytes[234].

CAST

In paper IV, we have identified a SNP (rs27582) in the *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. CAST is an inhibitor of calpains which bind to CAST in a calcium-dependent, reversible manner[235]. Calpains play a key role in the regulation of insulin secretion and insulin action[236]. Taken together, these data raise the possibility that the variation in the *CAST* gene might affect the activity of the calpastatin-calpain system and thereby alters insulin action and eventually secretion.

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[237].

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_{2max} 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_{2max}, 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_{2max} closely reflects expression of OXPHOS genes while expression of *AHNAK* was associated with low VO_{2max}.
- 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 energiintag, 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 används 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 glukosupptag. *TXNIP* uttrycket är genomgående förhöjt i muskeln i pre-diabetiker 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-celler, tyder våra data på att *TXNIP* kan spela en nyckelroll i defekt glukoshomeostas föregående fastställd 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 (VO_{2max}) och typ 1 fibrer i människans skelettmuskulatur. Vi har identifierat 66 gener positivt och 83 gener omvänt korrelerade med VO_{2max} och 171 gener positivt och 217 gener omvänt korrelerade med andel av typ 1 fibrer i människans skelettmuskulatur. Gener som är inblandade i oxidativ fosforylering (OXPHOS) visade högt uttryck hos individer med hög VO_{2max}, medan det motsatta inte var fallet hos individer med låg VO_{2max}. I stället var gener som *AHNAK* och *BCL6* förenade med låg VO_{2max}. 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 VO_{2max} nära återspeglar uttryck av OXPHOS gener särskilt av *NDUFB5* och *ATP5C1* i skelettmuskler och högt uttryck av dessa gener tyder på god muskelkondition. I motsats var ett högt uttryck av *AHNAK* förenat med en låg VO_{2max} och dålig muskelfitness.

I paper IV kombinerade vi GWA studier för T2DM med profilering av genuttrycket i målvävnaderna (bukspottkörtel, fettvävnad, lever och skelettmuskulatur) för att hitta nya riskgener för T2DM. Genom att använda denna metod identifierade vi en enbaspolymorfi (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.

Acknowledgements

73

References

1. King, K.M. & Rubin, G. A history of diabetes: from antiquity to discovering insulin. *Br J Nurs* **12**, 1091-5 (2003).
2. Diagnosis and classification of diabetes mellitus. *Diabetes Care* **31 Suppl 1**, S55-60 (2008).
3. Zimmet, P. Globalization, coca-colonization and the chronic disease epidemic: can the Doomsday scenario be averted? *J Intern Med* **247**, 301-10 (2000).
4. Wild, S., Roglic, G., Green, A., Sicree, R. & King, H. Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. *Diabetes Care* **27**, 1047-53 (2004).
5. Knowler, W.C., Pettitt, D.J., Saad, M.F. & Bennett, P.H. Diabetes mellitus in the Pima Indians: incidence, risk factors and pathogenesis. *Diabetes Metab Rev* **6**, 1-27 (1990).
6. Zimmet, P., Dowse, G., Finch, C., Serjeantson, S. & King, H. The epidemiology and natural history of NIDDM--lessons from the South Pacific. *Diabetes Metab Rev* **6**, 91-124 (1990).
7. Abate, N. & Chandalia, M. The impact of ethnicity on type 2 diabetes. *J Diabetes Complications* **17**, 39-58 (2003).
8. State-specific incidence of diabetes among adults--participating states, 1995-1997 and 2005-2007. *MMWR Morb Mortal Wkly Rep* **57**, 1169-73 (2008).
9. Shaw, J. Epidemiology of childhood type 2 diabetes and obesity. *Pediatr Diabetes* **8 Suppl 9**, 7-15 (2007).
10. Alberti, K.G. & Zimmet, P.Z. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. *Diabet Med* **15**, 539-53 (1998).
11. Unwin, N., Shaw, J., Zimmet, P. & Alberti, K.G. Impaired glucose tolerance and impaired fasting glycaemia: the current status on definition and intervention. *Diabet Med* **19**, 708-23 (2002).
12. Patti, M.E. Gene expression in the pathophysiology of type 2 diabetes mellitus. *Curr Diab Rep* **4**, 176-81 (2004).
13. Newman, B. et al. Concordance for type 2 (non-insulin-dependent) diabetes mellitus in male twins. *Diabetologia* **30**, 763-8 (1987).
14. Pyke, D.A., Theophanides, C.G. & Tattersall, R.B. Genetic origin of diabetes: Re-evaluation of twin data. *Lancet* **2**, 464 (1976).
15. Martin, B.C. et al. Role of glucose and insulin resistance in development of type 2 diabetes mellitus: results of a 25-year follow-up study. *Lancet* **340**, 925-9 (1992).

16. Elbein, S.C. Perspective: the search for genes for type 2 diabetes in the post-genome era. *Endocrinology* **143**, 2012-8 (2002).

17. Lyssenko, V. et al. Predictors of and longitudinal changes in insulin sensitivity and secretion preceding onset of type 2 diabetes. *Diabetes* **54**, 166-74 (2005).

18. Kahn, C.R. Banting Lecture. Insulin action, diabetogenes, and the cause of type II diabetes. *Diabetes* **43**, 1066-84 (1994).

19. Reaven, G.M. Banting lecture 1988. Role of insulin resistance in human disease. *Diabetes* **37**, 1595-607 (1988).

20. Wilding, J.P. The importance of free fatty acids in the development of Type 2 diabetes. *Diabet Med* **24**, 934-45 (2007).

21. DeFronzo, R.A., Bonadonna, R.C. & Ferrannini, E. Pathogenesis of NIDDM. A balanced overview. *Diabetes Care* **15**, 318-68 (1992).

22. Kraemer, W.J. & Spiering, B.A. Skeletal muscle physiology: Plasticity and responses to exercise. in *9th KIGS/KIMS Expert Meeting on Growth Hormone and Growth Disorders* 2-16 (Karger, Stresa, ITALY, 2006).

23. Kraus, W.E., Torgan, C.E. & Taylor, D.A. Skeletal muscle adaptation to chronic low-frequency motor nerve stimulation. *Exerc Sport Sci Rev* **22**, 313-60 (1994).

24. Petersen, K.F., Dufour, S., Befroy, D., Garcia, R. & Shulman, G.I. Impaired mitochondrial activity in the insulin-resistant offspring of patients with type 2 diabetes. *N Engl J Med* **350**, 664-71 (2004).

25. Lillioja, S. et al. Skeletal muscle capillary density and fiber type are possible determinants of in vivo insulin resistance in man. *J Clin Invest* **80**, 415-24 (1987).

26. Eriksson, J. et al. Early metabolic defects in persons at increased risk for non-insulin-dependent diabetes mellitus. *N Engl J Med* **321**, 337-43 (1989).

27. Lillioja, S. et al. Insulin resistance and insulin secretory dysfunction as precursors of non-insulin-dependent diabetes mellitus. Prospective studies of Pima Indians. *N Engl J Med* **329**, 1988-92 (1993).

28. Lillioja, S. et al. Impaired glucose tolerance as a disorder of insulin action. Longitudinal and cross-sectional studies in Pima Indians. *N Engl J Med* **318**, 1217-25 (1988).

29. Warram, J.H., Martin, B.C., Krolewski, A.S., Soeldner, J.S. & Kahn, C.R. Slow glucose removal rate and hyperinsulinemia precede the development of type II diabetes in the offspring of diabetic parents. *Ann Intern Med* **113**, 909-15 (1990).

30. Volk, A. et al. Insulin action and secretion in healthy, glucose tolerant first degree relatives of patients with type 2 diabetes mellitus. Influence of body weight. *Exp Clin Endocrinol Diabetes* **107**, 140-7 (1999).

31. Groop, L.C. et al. Association between polymorphism of the glycogen synthase gene and non-insulin-dependent diabetes mellitus. *N Engl J Med* **328**, 10-4 (1993).

32. Schalin-Jantti, C., Harkonen, M. & Groop, L.C. Impaired activation of glycogen synthase in people at increased risk for developing NIDDM. *Diabetes* **41**, 598-604 (1992).

33. Petersen, K.F. & Shulman, G.I. Etiology of insulin resistance. *Am J Med* **119**, S10-6 (2006).

34. Kim, J.A., Wei, Y. & Sowers, J.R. Role of mitochondrial dysfunction in insulin resistance. *Circ Res* **102**, 401-14 (2008).

35. DeFronzo, R.A. Lilly lecture 1987. The triumvirate: beta-cell, muscle, liver. A collusion responsible for NIDDM. *Diabetes* **37**, 667-87 (1988).

36. Sesti, G. Pathophysiology of insulin resistance. *Best Pract Res Clin Endocrinol Metab* **20**, 665-79 (2006).

37. Petersen, K.F. & Shulman, G.I. Pathogenesis of skeletal muscle insulin resistance in type 2 diabetes mellitus. *Am J Cardiol* **90**, 11G-18G (2002).

38. Rothman, D.L. et al. Decreased muscle glucose transport/phosphorylation is an early defect in the pathogenesis of non-insulin-dependent diabetes mellitus. *Proc Natl Acad Sci U S A* **92**, 983-7 (1995).

39. Shulman, G.I. et al. Quantitation of muscle glycogen synthesis in normal subjects and subjects with non-insulin-dependent diabetes by 13C nuclear magnetic resonance spectroscopy. *N Engl J Med* **322**, 223-8 (1990).

40. Randle, P.J., Garland, P.B., Hales, C.N. & Newsholme, E.A. The glucose fatty-acid cycle. Its role in insulin sensitivity and the metabolic disturbances of diabetes mellitus. *Lancet* **1**, 785-9 (1963).

41. Shulman, G.I. Cellular mechanisms of insulin resistance. *J Clin Invest* **106**, 171-6 (2000).

42. Dresner, A. et al. Effects of free fatty acids on glucose transport and IRS-1-associated phosphatidylinositol 3-kinase activity. *J Clin Invest* **103**, 253-9 (1999).

43. Roden, M. et al. Mechanism of free fatty acid-induced insulin resistance in humans. *J Clin Invest* **97**, 2859-65 (1996).

44. Griffin, M.E. et al. Free fatty acid-induced insulin resistance is associated with activation of protein kinase C theta and alterations in the insulin signaling cascade. *Diabetes* **48**, 1270-4 (1999).

45. Itani, S.I., Ruderman, N.B., Schmieder, F. & Boden, G. Lipid-induced insulin resistance in human muscle is associated with changes in diacylglycerol, protein kinase C, and IkappaB-alpha. *Diabetes* **51**, 2005-11 (2002).

46. Yu, C. et al. Mechanism by which fatty acids inhibit insulin activation of insulin receptor substrate-1 (IRS-1)-associated phosphatidylinositol 3-kinase activity in muscle. *J Biol Chem* **277**, 50230-6 (2002).

47. Boden, G. Fatty acid-induced inflammation and insulin resistance in skeletal muscle and liver. *Curr Diab Rep* **6**, 177-81 (2006).

48. Evans, J.L., Maddux, B.A. & Goldfine, I.D. The molecular basis for oxidative stress-induced insulin resistance. *Antioxid Redox Signal* **7**, 1040-52 (2005).

49. Brehm, A. et al. Increased lipid availability impairs insulin-stimulated ATP synthesis in human skeletal muscle. *Diabetes* **55**, 136-40 (2006).

50. Befroy, D.E. et al. Impaired mitochondrial substrate oxidation in muscle of insulin-resistant offspring of type 2 diabetic patients. *Diabetes* **56**, 1376-81 (2007).

51. Bruce, C.R., Carey, A.L., Hawley, J.A. & Febbraio, M.A. Intramuscular heat shock protein 72 and heme oxygenase-1 mRNA are reduced in patients with type 2 diabetes: evidence that insulin resistance is associated with a disturbed antioxidant defense mechanism. *Diabetes* **52**, 2338-45 (2003).

52. Houstis, N., Rosen, E.D. & Lander, E.S. Reactive oxygen species have a causal role in multiple forms of insulin resistance. *Nature* **440**, 944-8 (2006).

53. Mogensen, M. et al. Mitochondrial respiration is decreased in skeletal muscle of patients with type 2 diabetes. *Diabetes* **56**, 1592-9 (2007).

54. Paolisso, G. et al. Evidence for a relationship between oxidative stress and insulin action in non-insulin-dependent (type II) diabetic patients. *Metabolism* **43**, 1426-9 (1994).

55. West, I.C. Radicals and oxidative stress in diabetes. *Diabet Med* **17**, 171-80 (2000).

56. Ritz, P. & Berrut, G. Mitochondrial function, energy expenditure, aging and insulin resistance. *Diabetes Metab* **31 Spec No 2**, 5S67-5S73 (2005).

57. Kelley, D.E., He, J., Menshikova, E.V. & Ritov, V.B. Dysfunction of mitochondria in human skeletal muscle in type 2 diabetes. *Diabetes* **51**, 2944-50 (2002).

58. Ritov, V.B. et al. Deficiency of subsarcolemmal mitochondria in obesity and type 2 diabetes. *Diabetes* **54**, 8-14 (2005).

59. Sigal, R.J., Kenny, G.P., Wasserman, D.H. & Castaneda-Sceppa, C. Physical activity/exercise and type 2 diabetes. *Diabetes Care* **27**, 2518-39 (2004).

60. Hu, F.B. et al. Walking compared with vigorous physical activity and risk of type 2 diabetes in women: a prospective study. *JAMA* **282**, 1433-9 (1999).

61. Knowler, W.C. et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med* **346**, 393-403 (2002).

62. Folsom, A.R., Kushi, L.H. & Hong, C.P. Physical activity and incident diabetes mellitus in postmenopausal women. *Am J Public Health* **90**, 134-8 (2000).

63. Helmrich, S.P., Ragland, D.R., Leung, R.W. & Paffenbarger, R.S., Jr. Physical activity and reduced occurrence of non-insulin-dependent diabetes mellitus. *N Engl J Med* **325**, 147-52 (1991).

64. Helmrich, S.P., Ragland, D.R. & Paffenbarger, R.S., Jr. Prevention of non-insulin-dependent diabetes mellitus with physical activity. *Med Sci Sports Exerc* **26**, 824-30 (1994).

65. Goodyear, L.J. & Kahn, B.B. Exercise, glucose transport, and insulin sensitivity. *Annu Rev Med* **49**, 235-61 (1998).

66. Menshikova, E.V. et al. Effects of weight loss and physical activity on skeletal muscle mitochondrial function in obesity. *Am J Physiol Endocrinol Metab* **288**, E818-25 (2005).

80

67.   Perseghin, G. et al. Increased glucose transport-phosphorylation and muscle glycogen synthesis after exercise training in insulin-resistant subjects. *N Engl J Med* **335**, 1357-62 (1996).

68.   Miles, L. Physical activity and health. *Nutrition Bulletin* **32**, 314–63 (2007).

69.   Bouchard, C., Boulay, M.R., Simoneau, J.A., Lortie, G. & Perusse, L. Heredity and trainability of aerobic and anaerobic performances. An update. *Sports Med* **5**, 69-73 (1988).

70.   Bouchard, C. et al. Familial resemblance for VO2max in the sedentary state: the HERITAGE family study. *Med Sci Sports Exerc* **30**, 252-8 (1998).

71.   Fagard, R., Bielen, E. & Amery, A. Heritability of aerobic power and anaerobic energy generation during exercise. *J Appl Physiol* **70**, 357-62 (1991).

72.   Sundet, J.M., Eilertsen, D.E., Tambs, K. & Magnus, P. No differential heritability of intelligence test scores across ability levels in Norway. *Behav Genet* **24**, 337-9 (1994).

73.   Eriksson, K.F. & Lindgarde, F. Poor physical fitness, and impaired early insulin response but late hyperinsulinaemia, as predictors of NIDDM in middle-aged Swedish men. *Diabetologia* **39**, 573-9 (1996).

74.   Wei, M. et al. The association between cardiorespiratory fitness and impaired fasting glucose and type 2 diabetes mellitus in men. *Ann Intern Med* **130**, 89-96 (1999).

75.   Hagen, J.L. et al. Skeletal muscle aging in F344BN F1-hybrid rats: I. Mitochondrial dysfunction contributes to the age-associated reduction in VO2max. *J Gerontol A Biol Sci Med Sci* **59**, 1099-110 (2004).

76.   Ruderman, N., Chisholm, D., Pi-Sunyer, X. & Schneider, S. The metabolically obese, normal-weight individual revisited. *Diabetes* **47**, 699-713 (1998).

77.   Dela, F., Mikines, K.J., von Linstow, M., Secher, N.H. & Galbo, H. Effect of training on insulin-mediated glucose uptake in human muscle. *Am J Physiol* **263**, E1134-43 (1992).

78.   Bruce, C.R., Kriketos, A.D., Cooney, G.J. & Hawley, J.A. Disassociation of muscle triglyceride content and insulin sensitivity after exercise training in patients with Type 2 diabetes. *Diabetologia* **47**, 23-30 (2004).

79.   Ostergard, T. et al. Impact of exercise training on insulin sensitivity, physical fitness, and muscle oxidative capacity in first-degree relatives of type 2 diabetic patients. *Am J Physiol Endocrinol Metab* **290**, E998-1005 (2006).

80.   Matolin, S. et al. Histochemical and functional parameters in Nordic combination athletes. *Physiol Res* **43**, 243-51 (1994).

81.   Forton, J.T. & Kwiatkowski, D.P. Searching for the regulators of human gene expression. *Bioessays* **28**, 968-72 (2006).

82.   Lander, E.S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).

83.   Snyder, M. & Gerstein, M. Genomics. Defining genes in the genomics era. *Science* **300**, 258-60 (2003).

84.   The International HapMap Project. *Nature* **426**, 789-96 (2003).

85.	Kruglyak, L. & Nickerson, D.A. Variation is the spice of life. *Nat Genet* **27**, 234-6 (2001).

86.	Giallourakis, C., Henson, C., Reich, M., Xie, X. & Mootha, V.K. Disease gene discovery through integrative genomics. *Annu Rev Genomics Hum Genet* **6**, 381-406 (2005).

87.	Parikh, H. & Groop, L. Candidate genes for type 2 diabetes. *Rev Endocr Metab Disord* **5**, 151-76 (2004).

88.	Burmeister, M. Basic concepts in the study of diseases with complex genetics. *Biol Psychiatry* **45**, 522-32 (1999).

89.	Arcellana-Panlilio, M. & Robbins, S.M. Cutting-edge technology. I. Global gene expression profiling using DNA microarrays. *Am J Physiol Gastrointest Liver Physiol* **282**, G397-402 (2002).

90.	Schena, M., Shalon, D., Davis, R.W. & Brown, P.O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467-70 (1995).

91.	Hardiman, G. Microarray platforms--comparisons and contrasts. *Pharmacogenomics* **5**, 487-502 (2004).

92.	Lockhart, D.J. & Winzeler, E.A. Genomics, gene expression and DNA arrays. *Nature* **405**, 827-36 (2000).

93.	Shippy, R. et al. Using RNA sample titrations to assess microarray platform performance and normalization techniques. *Nat Biotechnol* **24**, 1123-31 (2006).

94.	Dalma-Weiszhausz, D.D., Warrington, J., Tanimoto, E.Y. & Miyada, C.G. The affymetrix GeneChip platform: an overview. *Methods Enzymol* **410**, 3-28 (2006).

95.	Lipshutz, R.J., Fodor, S.P., Gingeras, T.R. & Lockhart, D.J. High density synthetic oligonucleotide arrays. *Nat Genet* **21**, 20-4 (1999).

96.	Lockhart, D.J. et al. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* **14**, 1675-80 (1996).

97.	Mei, R. et al. Probe selection for high-density oligonucleotide arrays. *Proc Natl Acad Sci U S A* **100**, 11237-42 (2003).

98.	Wodicka, L., Dong, H., Mittmann, M., Ho, M.H. & Lockhart, D.J. Genome-wide expression monitoring in Saccharomyces cerevisiae. *Nat Biotechnol* **15**, 1359-67 (1997).

99.	Caruthers, M.H., Beaton, G., Wu, J.V. & Wiesler, W. Chemical synthesis of deoxyoligonucleotides and deoxyoligonucleotide analogs. *Methods Enzymol* **211**, 3-20 (1992).

100.	Hughes, T.R. et al. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat Biotechnol* **19**, 342-7 (2001).

101.	Kronick, M.N. Creation of the whole human genome microarray. *Expert Rev Proteomics* **1**, 19-28 (2004).

102.	Wolber, P.K., Collins, P.J., Lucas, A.B., De Witte, A. & Shannon, K.W. The Agilent in situ-synthesized microarray platform. *Methods Enzymol* **410**, 28-57 (2006).

103. Yauk, C.L., Berndt, M.L., Williams, A. & Douglas, G.R. Comprehensive comparison of six microarray technologies. *Nucleic Acids Res* **32**, e124 (2004).

104. Mootha, V.K. et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* **34**, 267-73 (2003).

105. Patti, M.E. et al. Coordinated reduction of genes of oxidative metabolism in humans with insulin resistance and diabetes: Potential role of PGC1 and NRF1. *Proc Natl Acad Sci U S A* **100**, 8466-71 (2003).

106. Dahlman, I. et al. Downregulation of electron transport chain genes in visceral adipose tissue in type 2 diabetes independent of obesity and possibly involving tumor necrosis factor-alpha. *Diabetes* **55**, 1792-9 (2006).

107. Gunton, J.E. et al. Loss of ARNT/HIF1beta mediates altered gene expression and pancreatic-islet dysfunction in human type 2 diabetes. *Cell* **122**, 337-49 (2005).

108. Glasbey, C.A. & Ghazal, P. Combinatorial image analysis of DNA microarray features. *Bioinformatics* **19**, 194-203 (2003).

109. Irizarry, R.A. et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249-64 (2003).

110. Liu, H. et al. AffyProbeMiner: a web resource for computing or retrieving accurately redefined Affymetrix probe sets. *Bioinformatics* **23**, 2385-90 (2007).

111. Irizarry, R.A., Wu, Z. & Jaffee, H.A. Comparison of Affymetrix GeneChip expression measures. *Bioinformatics* **22**, 789-94 (2006).

112. Gautier, L., Cope, L., Bolstad, B.M. & Irizarry, R.A. affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**, 307-15 (2004).

113. Li, C. & Hung Wong, W. Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. *Genome Biol* **2**, RESEARCH0032 (2001).

114. Li, C. & Wong, W.H. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc Natl Acad Sci U S A* **98**, 31-6 (2001).

115. Irizarry, R.A. et al. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* **31**, e15 (2003).

116. Wu, Z. & Irizarry, R.A. Stochastic models inspired by hybridization theory for short oligonucleotide arrays. *J Comput Biol* **12**, 882-93 (2005).

117. Huber, W., von Heydebreck, A., Sultmann, H., Poustka, A. & Vingron, M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* **18 Suppl 1**, S96-104 (2002).

118. Zhang, L., Miles, M.F. & Aldape, K.D. A model of molecular interactions on short oligonucleotide microarrays. *Nat Biotechnol* **21**, 818-21 (2003).

119. Hochreiter, S., Clevert, D.A. & Obermayer, K. A new summarization method for Affymetrix probe level data. *Bioinformatics* **22**, 943-9 (2006).

120. Millenaar, F.F. et al. How to decide? Different methods of calculating gene expression from short oligonucleotide array data will give different results. *BMC Bioinformatics* 7, 137 (2006).

121. Seo, J. & Hoffman, E.P. Probe set algorithms: is there a rational best bet? *BMC Bioinformatics* 7, 395 (2006).

122. Kibriya, M.G. et al. Changes in gene expression profiles in response to selenium supplementation among individuals with arsenic-induced pre-malignant skin lesions. *Toxicol Lett* **169**, 162-76 (2007).

123. Zahurak, M. et al. Pre-processing Agilent microarray data. *BMC Bioinformatics* **8**, 142 (2007).

124. Gandemer, V. et al. Five distinct biological processes and 14 differentially expressed genes characterize TEL/AML1-positive leukemia. *BMC Genomics* **8**, 385 (2007).

125. McClintick, J.N. & Edenberg, H.J. Effects of filtering by Present call on analysis of microarray experiments. *BMC Bioinformatics* 7, 49 (2006).

126. Allison, D.B. et al. A mixture model approach for the analysis of microarray gene expression data. *Computational Statistics & Data Analysis* **39**, 1-20 (2002).

127. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate - a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological* **57**, 289-300 (1995).

128. Genovese, C. & Wasserman, L. Operating characteristics and extensions of the false discovery rate procedure. *Journal of the Royal Statistical Society Series B-Statistical Methodology* **64**, 499-517 (2002).

129. Tusher, V.G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* **98**, 5116-21 (2001).

130. Grant, G.R., Liu, J. & Stoeckert, C.J., Jr. A practical false discovery rate approach to identifying patterns of differential expression in microarray data. *Bioinformatics* **21**, 2684-90 (2005).

131. Smyth, G.K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**, Article3 (2004).

132. Sjogren, A., Kristiansson, E., Rudemo, M. & Nerman, O. Weighted analysis of general microarray experiments. *BMC Bioinformatics* **8**, 387 (2007).

133. Breitling, R., Armengaud, P., Amtmann, A. & Herzyk, P. Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS Lett* **573**, 83-92 (2004).

134. Mukherjee, S., Roberts, S.J. & van der Laan, M.J. Data-adaptive test statistics for microarray data. *Bioinformatics* **21 Suppl 2**, ii108-14 (2005).

135. Yang, Y.H., Xiao, Y. & Segal, M.R. Identifying differentially expressed genes from microarray experiments via statistic synthesis. *Bioinformatics* **21**, 1084-93 (2005).

136.   Jeffery, I.B., Higgins, D.G. & Culhane, A.C. Comparison and evaluation of methods for generating differentially expressed gene lists from microarray data. *BMC Bioinformatics* 7, 359 (2006).

137.   Kim, S.Y., Lee, J.W. & Sohn, I.S. Comparison of various statistical methods for identifying differential gene expression in replicated microarray data. *Stat Methods Med Res* 15, 3-20 (2006).

138.   Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102, 15545-50 (2005).

139.   Zhang, B., Kirov, S. & Snoddy, J. WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res* 33, W741-8 (2005).

140.   Barry, W.T., Nobel, A.B. & Wright, F.A. Significance analysis of functional categories in gene expression studies: a structured permutation approach. *Bioinformatics* 21, 1943-9 (2005).

141.   Nam, D. et al. ADGO: analysis of differentially expressed gene sets using composite GO annotation. *Bioinformatics* 22, 2249-53 (2006).

142.   Volinia, S. et al. GOAL: automated Gene Ontology analysis of expression profiles. *Nucleic Acids Res* 32, W492-9 (2004).

143.   Goeman, J.J. & Buhlmann, P. Analyzing gene expression data in terms of gene sets: methodological issues. *Bioinformatics* 23, 980-7 (2007).

144.   Golub, T.R. et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286, 531-7 (1999).

145.   Hirschhorn, J.N. & Daly, M.J. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 6, 95-108 (2005).

146.   Maresso, K. & Broeckel, U. Genotyping platforms for mass-throughput genotyping with SNPs, including human genome-wide scans. *Adv Genet* 60, 107-39 (2008).

147.   Balding, D.J. A tutorial on statistical methods for population association studies. *Nat Rev Genet* 7, 781-91 (2006).

148.   Teo, Y.Y. Common statistical issues in genome-wide association studies: a review on power, data quality control, genotype calling and population structure. *Curr Opin Lipidol* 19, 133-43 (2008).

149.   Saxena, R. et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 316, 1331-6 (2007).

150.   Scott, L.J. et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 316, 1341-5 (2007).

151.   Sladek, R. et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 445, 881-5 (2007).

152.   Zeggini, E. et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet* 40, 638-45 (2008).

153.   Zeggini, E. et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 316, 1336-41 (2007).

154. Schadt, E.E. et al. Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422**, 297-302 (2003).

155. Ling, C. et al. Multiple environmental and genetic factors influence skeletal muscle PGC-1alpha and PGC-1beta gene expression in twins. *J Clin Invest* **114**, 1518-26 (2004).

156. Florez, J.C. et al. Association testing in 9,000 people fails to confirm the association of the insulin receptor substrate-1 G972R polymorphism with type 2 diabetes. *Diabetes* **53**, 3313-8 (2004).

157. Eriksson, K.F. & Lindgarde, F. Impaired glucose tolerance in a middle-aged male urban population: a new approach for identifying high-risk cases. *Diabetologia* **33**, 526-31 (1990).

158. Poulsen, P. et al. Heritability of insulin secretion, peripheral and hepatic insulin action, and intracellular glucose partitioning in young and old Danish twins. *Diabetes* **54**, 275-83 (2005).

159. Eriksson, K.F., Saltin, B. & Lindgarde, F. Increased skeletal muscle capillary density precedes diabetes development in men with impaired glucose tolerance. A 15-year follow-up. *Diabetes* **43**, 805-8 (1994).

160. Suh, Y.H. et al. Analysis of gene expression profiles in insulin-sensitive tissues from pre-diabetic and diabetic Zucker diabetic fatty rats. *J Mol Endocrinol* **34**, 299-315 (2005).

161. Yechoor, V.K. et al. Distinct pathways of insulin-regulated versus diabetes-regulated gene expression: an in vivo analysis in MIRKO mice. *Proc Natl Acad Sci U S A* **101**, 16525-30 (2004).

162. Granhall, C., Rosengren, A.H., Renstrom, E. & Luthman, H. Separately inherited defects in insulin exocytosis and beta-cell glucose metabolism contribute to type 2 diabetes. *Diabetes* **55**, 3494-500 (2006).

163. Lindholm, E., Agardh, E., Tuomi, T., Groop, L. & Agardh, C.D. Classifying diabetes according to the new WHO clinical stages. *Eur J Epidemiol* **17**, 983-9 (2001).

164. Berglund, G., Elmstahl, S., Janzon, L. & Larsson, S.A. The Malmo Diet and Cancer Study. Design and feasibility. *J Intern Med* **233**, 45-51 (1993).

165. Lyssenko, V. et al. Clinical risk factors, DNA variants, and the development of type 2 diabetes. *N Engl J Med* **359**, 2220-32 (2008).

166. Phillips, D.I., Clark, P.M., Hales, C.N. & Osmond, C. Understanding oral glucose tolerance: comparison of glucose or insulin measurements during the oral glucose tolerance test with specific measurements of insulin resistance and insulin secretion. *Diabet Med* **11**, 286-92 (1994).

167. Matsuda, M. & DeFronzo, R.A. Insulin sensitivity indices obtained from oral glucose tolerance testing: comparison with the euglycemic insulin clamp. *Diabetes Care* **22**, 1462-70 (1999).

168. The Homeostasis Model Assessment (HOMA) http://www.dtu.ox.ac.uk/. (2008).

169. Hanson, R.L. et al. Evaluation of simple indices of insulin sensitivity and insulin secretion for use in epidemiologic studies. *Am J Epidemiol* **151**, 190-8 (2000).

170. DeFronzo, R.A., Tobin, J.D. & Andres, R. Glucose clamp technique: a method for quantifying insulin secretion and resistance. *Am J Physiol* **237**, E214-23 (1979).

171. Matthews, D.R. et al. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* **28**, 412-9 (1985).

172. Muniyappa, R., Lee, S., Chen, H. & Quon, M.J. Current approaches for assessing insulin sensitivity and resistance in vivo: advantages, limitations, and appropriate usage. *Am J Physiol Endocrinol Metab* **294**, E15-26 (2008).

173. Matarese, L.E. Indirect calorimetry: technical aspects. *J Am Diet Assoc* **97**, S154-60 (1997).

174. Ferrannini, E. The theoretical bases of indirect calorimetry: a review. *Metabolism* **37**, 287-301 (1988).

175. Frayn, K.N. Calculation of substrate oxidation rates in vivo from gaseous exchange. *J Appl Physiol* **55**, 628-34 (1983).

176. Schalin-Jantti, C., Laurila, E., Lofman, M. & Groop, L.C. Determinants of insulin-stimulated skeletal muscle glycogen metabolism in man. *Eur J Clin Invest* **25**, 693-8 (1995).

177. Wabitsch, M. et al. Characterization of a human preadipocyte cell strain with high capacity for adipose differentiation. *Int J Obes Relat Metab Disord* **25**, 8-15 (2001).

178. Chomczynski, P. & Sacchi, N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* **162**, 156-9 (1987).

179. Holland, P.M., Abramson, R.D., Watson, R. & Gelfand, D.H. Detection of specific polymerase chain reaction product by utilizing the 5'----3' exonuclease activity of Thermus aquaticus DNA polymerase. *Proc Natl Acad Sci U S A* **88**, 7276-80 (1991).

180. Heid, C.A., Stevens, J., Livak, K.J. & Williams, P.M. Real time quantitative PCR. *Genome Res* **6**, 986-94 (1996).

181. Al-Khalili, L., Cartee, G.D. & Krook, A. RNA interference-mediated reduction in GLUT1 inhibits serum-induced glucose transport in primary human skeletal muscle cells. *Biochem Biophys Res Commun* **307**, 127-32 (2003).

182. Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* **8**, 186-94 (1998).

183. Ewing, B., Hillier, L., Wendl, M.C. & Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* **8**, 175-85 (1998).

184. Staden, R. The Staden sequence analysis package. *Mol Biotechnol* **5**, 233-41 (1996).

185.	Livak, K.J. Allelic discrimination using fluorogenic probes and the 5' nuclease assay. *Genet Anal* **14**, 143-9 (1999).

186.	Kwok, P.Y. Methods for genotyping single nucleotide polymorphisms. *Annu Rev Genomics Hum Genet* **2**, 235-58 (2001).

187.	Latif, S., Bauer-Sardina, I., Ranade, K., Livak, K.J. & Kwok, P.Y. Fluorescence polarization in homogeneous nucleic acid analysis II: 5'-nuclease assay. *Genome Res* **11**, 436-40 (2001).

188.	Livak, K.J. SNP genotyping by the 5'-nuclease reaction. *Methods Mol Biol* **212**, 129-47 (2003).

189.	Syvanen, A.C. Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nat Rev Genet* **2**, 930-42 (2001).

190.	Jurinke, C., van den Boom, D., Cantor, C.R. & Koster, H. The use of MassARRAY technology for high throughput genotyping. *Adv Biochem Eng Biotechnol* **77**, 57-74 (2002).

191.	van den Boom, D. et al. MALDI-TOF MS: a platform technology for genetic discovery. *International Journal of Mass Spectrometry* **238**, 173-188 (2004).

192.	Zeger, S.L. & Liang, K.Y. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* **42**, 121-30 (1986).

193.	Radom-Aizik, S. et al. Effects of aerobic training on gene expression in skeletal muscle of elderly men. *Med Sci Sports Exerc* **37**, 1680-96 (2005).

194.	Yechoor, V.K., Patti, M.E., Saccone, R. & Kahn, C.R. Coordinated patterns of gene expression for substrate and energy metabolism in skeletal muscle of diabetic mice. *Proc Natl Acad Sci U S A* **99**, 10587-92 (2002).

195.	Smyth, G.K. & Speed, T. Normalization of cDNA microarray data. *Methods* **31**, 265-73 (2003).

196.	Riserus, U. et al. Activation of peroxisome proliferator-activated receptor (PPAR)delta promotes reversal of multiple metabolic abnormalities, reduces oxidative stress, and increases fatty acid oxidation in moderately obese men. *Diabetes* **57**, 332-9 (2008).

197.	Reilly, S.M. & Lee, C.H. PPAR delta as a therapeutic target in metabolic disease. *FEBS Lett* **582**, 26-31 (2008).

198.	Sparks, L.M. et al. A high-fat diet coordinately downregulates genes required for mitochondrial oxidative phosphorylation in skeletal muscle. *Diabetes* **54**, 1926-33 (2005).

199.	Horikawa, Y. et al. Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat Genet* **26**, 163-75 (2000).

200.	Nadler, S.T. & Attie, A.D. Please pass the chips: genomic insights into obesity and diabetes. *J Nutr* **131**, 2078-81 (2001).

201.	Wilkes, T., Laux, H. & Foy, C.A. Microarray data quality - review of current developments. *OMICS* **11**, 1-13 (2007).

202.	Brazma, A. et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet* **29**, 365-71 (2001).

203. Sun, G. Application of DNA microarrays in the study of human obesity and type 2 diabetes. *OMICS* **11**, 25-40 (2007).

204. Rensink, W.A. & Hazen, S.P. Statistical issues in microarray data analysis. *Methods Mol Biol* **323**, 359-66 (2006).

205. Pawitan, Y., Michiels, S., Koscielny, S., Gusnanto, A. & Ploner, A. False discovery rate, sensitivity and sample size for microarray studies. *Bioinformatics* **21**, 3017-24 (2005).

206. Coletta, D.K. et al. Effect of acute physiological hyperinsulinemia on gene expression in human skeletal muscle in vivo. *Am J Physiol Endocrinol Metab* **294**, E910-7 (2008).

207. The gene relationships across implicated loci (GRAIL) http://www.broad.mit.edu/mpg/grail/. (2008).

208. Rasche, A., Al-Hasani, H. & Herwig, R. Meta-analysis approach identifies candidate genes and associated molecular networks for type-2 diabetes mellitus. *BMC Genomics* **9**, 310 (2008).

209. Prokopenko, I., McCarthy, M.I. & Lindgren, C.M. Type 2 diabetes: new genes, new understanding. *Trends Genet* **24**, 613-21 (2008).

210. Minn, A.H., Hafele, C. & Shalev, A. Thioredoxin-interacting protein is stimulated by glucose through a carbohydrate response element and induces beta-cell apoptosis. *Endocrinology* **146**, 2397-405 (2005).

211. van Greevenbroek, M.M. et al. Genetic variation in thioredoxin interacting protein (TXNIP) is associated with hypertriglyceridaemia and blood pressure in diabetes mellitus. *Diabet Med* **24**, 498-504 (2007).

212. Haendeler, J. et al. Redox regulatory and anti-apoptotic functions of thioredoxin depend on S-nitrosylation at cysteine 69. *Nat Cell Biol* **4**, 743-9 (2002).

213. Holmgren, A. Thioredoxin. *Annu Rev Biochem* **54**, 237-71 (1985).

214. Muoio, D.M. TXNIP links redox circuitry to glucose control. *Cell Metab* **5**, 412-4 (2007).

215. Junn, E. et al. Vitamin D3 up-regulated protein 1 mediates oxidative stress via suppressing the thioredoxin function. *J Immunol* **164**, 6287-95 (2000).

216. Nishiyama, A. et al. Identification of thioredoxin-binding protein-2/vitamin D(3) up-regulated protein 1 as a negative regulator of thioredoxin function and expression. *J Biol Chem* **274**, 21645-50 (1999).

217. Patwari, P., Higgins, L.J., Chutkow, W.A., Yoshioka, J. & Lee, R.T. The interaction of thioredoxin with Txnip. Evidence for formation of a mixed disulfide by disulfide exchange. *J Biol Chem* **281**, 21884-91 (2006).

218. Yamanaka, H. et al. A possible interaction of thioredoxin with VDUP1 in HeLa cells detected in a yeast two-hybrid system. *Biochem Biophys Res Commun* **271**, 796-800 (2000).

219. Kim, S.Y., Suh, H.W., Chung, J.W., Yoon, S.R. & Choi, I. Diverse functions of VDUP1 in cell proliferation, differentiation, and diseases. *Cell Mol Immunol* **4**, 345-51 (2007).

220. Oka, S.I. et al. Thioredoxin Binding Protein-2 (TBP-2)/Txnip is a Critical Regulator of Insulin Secretion and PPAR Function. *Endocrinology* (2008).

221. Wang, Z. et al. Thioredoxin-interacting protein (txnip) is a glucocorticoid-regulated primary response gene involved in mediating glucocorticoid-induced apoptosis. *Oncogene* **25**, 1903-13 (2006).

222. Schulze, P.C. et al. Hyperglycemia promotes oxidative stress through inhibition of thioredoxin function by thioredoxin-interacting protein. *J Biol Chem* **279**, 30369-74 (2004).

223. Han, S.H. et al. VDUP1 upregulated by TGF-beta1 and 1,25-dihydorxyvitamin D3 inhibits tumor cell growth by blocking cell-cycle progression. *Oncogene* **22**, 4035-46 (2003).

224. Price, S.A. et al. Thioredoxin interacting protein is increased in sensory neurons in experimental diabetes. *Brain Res* **1116**, 206-14 (2006).

225. Kobayashi, T., Uehara, S., Ikeda, T., Itadani, H. & Kotani, H. Vitamin D3 up-regulated protein-1 regulates collagen expression in mesangial cells. *Kidney Int* **64**, 1632-42 (2003).

226. World, C.J., Yamawaki, H. & Berk, B.C. Thioredoxin in the cardiovascular system. *J Mol Med* **84**, 997-1003 (2006).

227. Yamawaki, H., Pan, S., Lee, R.T. & Berk, B.C. Fluid shear stress inhibits vascular inflammation by decreasing thioredoxin-interacting protein in endothelial cells. *J Clin Invest* **115**, 733-8 (2005).

228. Kaimul, A.M., Nakamura, H., Masutani, H. & Yodoi, J. Thioredoxin and thioredoxin-binding protein-2 in cancer and metabolic syndrome. *Free Radic Biol Med* **43**, 861-8 (2007).

229. Chen, J., Couto, F.M., Minn, A.H. & Shalev, A. Exenatide inhibits beta-cell apoptosis by decreasing thioredoxin-interacting protein. *Biochem Biophys Res Commun* **346**, 1067-74 (2006).

230. Minn, A.H. et al. Gene expression profiling in INS-1 cells overexpressing thioredoxin-interacting protein. *Biochem Biophys Res Commun* **336**, 770-8 (2005).

231. Oishi, Y. et al. SUMOylation of Kruppel-like transcription factor 5 acts as a molecular switch in transcriptional programs of lipid metabolism involving PPAR-delta. *Nat Med* **14**, 656-66 (2008).

232. Haase, H. et al. The carboxyl-terminal ahnak domain induces actin bundling and stabilizes muscle contraction. *FASEB J* **18**, 839-41 (2004).

233. Lee, I.H. et al. AHNAK-mediated activation of phospholipase C-gamma1 through protein kinase C. *J Biol Chem* **279**, 26645-53 (2004).

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).

235. Bertipaglia, I. & Carafoli, E. Calpains and human disease. *Subcell Biochem* **45**, 29-53 (2007).

236.   Sreenan, S.K. et al. Calpains play a role in insulin secretion and action. *Diabetes* **50**, 2013-20 (2001).

237.   Schnackenberg, L.K. Global metabolic profiling and its role in systems biology to advance personalized medicine in the 21st century. *Expert Rev Mol Diagn* 7, 247-59 (2007).