LUND UNIVERSITY

The role of genetic variation and DNA methylation in human glucose metabolism and type 2 diabetes

Rönn, Tina

2010

From the Department of Clinical Sciences, Malmö, Diabetes and Endocrinology
Clinical Research Centre, Malmö University Hospital, Lund University

Academic Dissertation

The role of genetic variation and DNA methylation in human glucose metabolism and type 2 diabetes

Tina Rönn

LUND UNIVERSITY
Faculty of Medicine

With the permission of the Medical Faculty of Lund University, to be presented for public examination in the CRC aula at the Clinical Research Centre, entrance 72, Malmö University Hospital, on May 27th 2010, at 9.00

Faculty opponent
Professor Karen Temple
University of Southampton
United Kingdom

Organization LUND UNIVERSITY	Document name DOCTORAL DISSERTATION
	Date of issue 12 April 2010
	Sponsoring organization
Author(s) Tina Rönn	

Title and subtitle: The role of genetic variation and DNA methylation in human glucose metabolism and type 2 diabetes

Abstract

The aim of this thesis was to explore the role of DNA methylation and common genetic variation on glucose metabolism and the pathogenesis of type 2 diabetes.

Reduced oxidative capacity of the mitochondria in skeletal muscle has been suggested to play a role in insulin resistance and type 2 diabetes. In studies I and II, we investigated the regulation of *COX7A1* and *ATP5O*, which encode two subunits of the mitochondrial respiratory chain. We found that genetic variation and age were associated with skeletal muscle mRNA expression in both studies. mRNA levels were also positively correlated with the expression of the transcriptional co-activator *PPARGC1A* and insulin-stimulated glucose uptake, *i.e.*, elderly individuals had reduced mRNA expression levels and reduced *in vivo* glucose uptake. Additionally, DNA methylation of the *COX7A1* promoter was increased in elderly individuals concordant with the decrease in *COX7A1* mRNA expression, suggesting a role for genetic, epigenetic and non-genetic factors in gene regulation.

In study III, we investigated a common genetic variant in *MTNR1B* that has previously been found to be associated with increased risk of type 2 diabetes, increased fasting plasma glucose and impaired insulin secretion in populations of European ancestry. We aimed to replicate these findings in a type 2 diabetes case-control cohort of Han Chinese ancestry. We confirmed the association between rs10830963 and both the risk of type 2 diabetes and increased fasting plasma glucose levels, suggesting a relatively ancient origin for this variant.

In study IV, common genetic variants that introduce or remove potential DNA methylation sites were selected based on their association with the risk of type 2 diabetes and changes in gene expression in blood. These genetic variants were analysed together with the level of DNA methylation and gene expression in human skeletal muscle, adipose tissue, blood and pancreatic islets. We found that 18 of the 19 sites that we analysed were associated with a difference in DNA methylation related to genotype, and for 11 of these sites this finding was consistent in all four tissues. Additionally, our data suggested a tissue-specific pattern of DNA methylation. Our results confirm an interaction between genetic and epigenetic mechanisms, which introduces a new level of complexity to our knowledge of gene regulation in type 2 diabetes.

Key words: type 2 diabetes, epigenetics, DNA methylation, oxidative phosphorylation, SNP, GWAS, gene expression

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

Supplementary bibliographical information:		Language
ISSN and key title: 1652-8220		ISBN 978-91-86443-73-3
Recipient's notes	Number of pages	Price
	Security classification	

The role of genetic variation and DNA methylation in human glucose metabolism and type 2 diabetes

Tina Rönn

LUND UNIVERSITY
Faculty of Medicine

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

Lund University, Faculty of Medicine Doctoral Dissertation Series 2010:58
ISSN 1652-8220
ISBN 978-91-86443-73-3

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

“Ju mer man tänker, desto mer inser man att det inte finns något enkelt svar.”
Nalle Puh

TABLE OF CONTENTS

ABSTRACT 9

POPULÄRVETENSKAPLIG SAMMANFATTNING 10

LIST OF PUBLICATIONS 12
- Scientific papers contributing to this thesis 12
- Publications not included in this thesis 13

ABBREVIATIONS 14

INTRODUCTION 15
- Diabetes mellitus 15
- Type 2 diabetes 16
  - *Insulin resistance 16*
- Oxidative phosphorylation 18
- Genetics of type 2 diabetes 19
  - *Genome-wide association studies 20*
- Epigenetics in type 2 diabetes 20
  - *DNA methylation 21*

AIMS OF THIS THESIS 23

STUDY PARTICIPANTS 24
- Twin studies 24
- Type 2 diabetes case-control cohorts 24
  - *Botnia 24*
  - *Malmö 25*
  - *Han Chinese 25*
- Intervention studies 25
  - *Malmö intervention study 25*
  - *Botnia PPP intervention study 26*
- Human islets 26

METHODOLOGY 27
- Phenotype characterisation 27
- Selection of single nucleotide polymorphisms (SNPs) 27
- Genotyping 28
  - *iPLEX 28*
  - *TaqMan 28*

DNA methylation analysis 28
Bisulfite sequencing *28*
EpiTYPER *30*
mRNA expression analysis 30
TaqMan *31*
Microarray *31*
Statistical analysis 31
Association testing *32*
Linear regression *32*
Generalised estimating equations *32*
Biometric modelling *32*
DNA methylation *32*
Hierarchical clustering *33*
mQTL analysis *33*
RESULTS 34
Study I 34
Age influences DNA methylation and gene expression of COX7A1 in human skeletal muscle *34*
Study II 36
Genetic variation in ATP5O is associated with skeletal muscle ATP5O mRNA expression and glucose uptake in young twins *36*
Study III 37
A common variant in MTNR1B, encoding melatonin receptor 1B, is associated with type 2 diabetes and fasting plasma glucose in Han Chinese individuals *37*
Study IV 38
Polymorphisms associated with type 2 diabetes and gene expression influence DNA methylation in human skeletal muscle, adipose tissue, pancreatic islets and blood *38*
DISCUSSION 41
SUMMARY AND GENERAL CONCLUSION 45
ACKNOWLEDGEMENTS 46
LIST OF REFERENCES 48

ABSTRACT

The incidence of diabetes is increasing worldwide, with the most prevalent form being type 2 diabetes. Two fundamental processes contribute to the development of type 2 diabetes: insulin resistance in target organs and insufficient insulin secretion from the pancreatic beta-cells. The aim of this thesis was to explore the role of DNA methylation and common genetic variation on glucose metabolism and the pathogenesis of type 2 diabetes.

Reduced oxidative capacity of the mitochondria in skeletal muscle has been suggested to play a role in insulin resistance and type 2 diabetes. In studies I and II, we investigated the regulation of *COX7A1* and *ATP5O*, which encode two subunits of the mitochondrial respiratory chain. We found that genetic variation and age were associated with skeletal muscle mRNA expression in both studies. mRNA levels were also positively correlated with the expression of the transcriptional co-activator *PPARGC1A* and insulin-stimulated glucose uptake, *i.e.*, elderly individuals had reduced mRNA expression levels and reduced *in vivo* glucose uptake. Additionally, DNA methylation of the *COX7A1* promoter was increased in elderly individuals concordant with the decrease in *COX7A1* mRNA expression, suggesting a role for genetic, epigenetic and non-genetic factors in gene regulation.

In study III, we investigated a common genetic variant in *MTNR1B* that has previously been found to be associated with increased risk of type 2 diabetes, increased fasting plasma glucose and impaired insulin secretion in populations of European ancestry. We aimed to replicate these findings in a type 2 diabetes case-control cohort of Han Chinese ancestry. We confirmed the association between rs10830963 and both the risk of type 2 diabetes and increased fasting plasma glucose levels, suggesting a relatively ancient origin for this variant.

In study IV, common genetic variants that introduce or remove potential DNA methylation sites were selected based on their association with the risk of type 2 diabetes and changes in gene expression in blood. These genetic variants were analysed together with the level of DNA methylation and gene expression in human skeletal muscle, adipose tissue, blood and pancreatic islets. We found that 18 of the 19 sites that we analysed were associated with a difference in DNA methylation related to genotype, and for 11 of these sites this finding was consistent in all four tissues. Additionally, our data suggested a tissue-specific pattern of DNA methylation. Our results confirm an interaction between genetic and epigenetic mechanisms, which introduces a new level of complexity to our knowledge of gene regulation in type 2 diabetes.

POPULÄRVETENSKAPLIG SAMMANFATTNING

Diabetes är en grupp sjukdomar som alla kännetecknas av en förhöjd nivå av sockerarten glukos i blodet. Antalet individer som drabbas av diabetes ökar hela tiden, en utveckling som ses över hela världen. Bara i Sverige har drygt 5% av befolkningen diagnosen diabetes, motsvarande 484.400 personer i åldern 20-79 år.

Typ 2 diabetes är den vanligaste formen av diabetes, och kallades tidigare också åldersdiabetes. Denna form av diabetes är förknippad med övervikt och för lite motion. För att kroppen ska fungera optimalt måste halten av glukos i blodet hela tiden hållas på en jämn nivå, oavsett om man precis ätit en stor måltid eller om man sover. Stigande glukosnivåer i blodet regleras med hjälp av hormonet insulin, som får kroppens celler, framför allt i muskel och lever, att ta upp och lagra glukos. Insulin tillverkas och utsöndras från betaceller i bukspottskörteln.

Orsaken till typ 2 diabetes är ännu inte helt känd, men man vet att både miljöfaktorer och flera olika ärftliga faktorer spelar roll. Sjukdomen föregås ofta av en längre tid med insulin-resistens, ett tillstånd där kroppens celler inte längre svarar normalt på insulinets verkan, vilket försämrar upptaget av glukos från blodet. Till en början kan kroppen kompensera genom att öka utsöndringen av insulin, men när betacellerna i bukspottskörteln inte längre klarar av att upprätthålla detta stiger glukoshalterna och resulterar i typ 2 diabetes. Kroniskt höga halter av glukos i blodet kan ge upphov till en rad komplikationer, tex hjärt- och kärlsjukdomar som också är en av de största dödorsakerna i typ 2 diabetes.

Glukos som tas upp av kroppens celler kan lagras för senare bruk eller användas för att producera energi. Mitokondrien är den enhet i kroppens alla celler som står för produktionen av energi, i form av en molekyl som kallas ATP. Mitokondrien är uppbyggd av flera mindre enheter, och delar av dessa har visat sig vara mindre uttryckta i muskel hos patienter med typ 2 diabetes än i friska individer. Detta kan leda till att energiproduktion och omsättning av glukos i cellen inte fungerar optimalt, och därmed försämras också upptag av glukos från blodet till cellerna.

Vi har studerat två delar i mitokondriens mindre enheter, *COX7A1* och *ATP5O*, för att ta reda på vad som reglerar deras uttryck och på så sätt kunna förstå varför nivåerna av dem är sänkta i muskelceller från typ 2 diabetiker. Vi fann att både *COX7A1* och *ATP5O* påverkas negativt av ökad ålder och av ärftliga faktorer, och vi upptäckte också att en minskning av dem bidrog till sämre glukosupptag från blodet. *COX7A1* visade sig också vara negativt påverkad av epigenetiska faktorer, i form av ökad DNA-metylering.

Epigenetik är ett dynamiskt fenomen som kan kontrollera när, var och hur ett arvsanlag ska komma till uttryck. Epigenetiska förändringar inträffar under livets gång och skiljer sig åt mellan olika celler och olika organ. Detta är i stor kontrast till genetiken, som omfattar vårt arvsanlag ordnat i en lång kedja av DNA som är konstant över tid och i alla kroppens celler. Ordningen på beståndsdelarna i DNA-kedjan utgör den genetiska informationen, vårt arv. Epigenetiken påverkar inte arvsmassan direkt, men kan påverka hur arvsmassan tolkas och kommer till uttryck i kroppen, och fungerar som cellens minne.

DNA-metylering är ett exempel på epigenetisk reglering, där en metylgrupp ($-CH_3$) binder till DNA-kedjan och påverkar hur arvsanlaget kommer att tolkas. Detta är en reversibel process, dvs metylgrupperna kan tas bort igen, och beroende på i vilka celler och organ, samt när DNA-metyleringen sker, så kommer vårt arvsanlag att uttryckas på olika sätt.

Vanligt förekommande genetisk variation kan användas för att identifiera regioner i vår arvsmassa som ökar risken för att drabbas av sjukdomar, som tex typ 2 diabetes. Vi har studerat genetisk variation i en region av arvsmassan som innehåller genen för *MTNR1B*. Uttryck av denna gen ger produktion av en enhet där melatonin kan binda in, ett hormon som påverkas av ljuset och hjälper kroppen att reglera dygnsrytmen. Genetisk variation i *MTNR1B* har tidigare visat sig ha ett samband med risk att drabbas av typ 2 diabetes. Arvsanlaget skiljer sig åt mellan olika individer och det är därför vi alla är olika och har olika risk för att drabbas av sjukdomar. Skillnaderna i arvsanlag skiljer sig dessutom åt mellan personer med olika etniskt och geografiskt ursprung. Vårt syfte var att ta reda på om den genetiska variationen i *MTNR1B* som ökar risken för typ 2 diabetes i européer, också har samma effekt hos personer från Shanghai i Kina. Resultaten visade att samma genetiska variation var förknippad med ökad risk för typ 2 diabetes och med högre blodsockernivåer också i de studerade individerna från Kina. Effekten var av samma storlek som visats i européer i de tidigare studierna, men riskvarianten är vanligare i individerna med kinesiskt ursprung och har därför påverkan på en större andel av befolkningen.

Slutligen ville vi kombinera genetisk variation med DNA metylering för att få djupare förståelse av hur regleringen av vårt arvsanlag går till. Vi valde genetiska varianter som tidigare visat sig ha en koppling till typ 2 diabetes och som också påverkar hur arvsanlagen i närheten uttrycks i blod. Dessutom inkluderades bara genetiska varianter som kan påverka DNA-metylering, dvs som antingen ger upphov till eller tar bort en sekvens i DNAt där en metylgrupp kan binda in. Vi fann att majoriteten av de varianterna vi undersökt förändrade graden av DNA-metylering i det specifika området, och att detta också i vissa fall påverkade uttrycket av arvsanlaget.

Identifiering av nya gener och nya mekanismer för hur dessa kan påverka kroppens funktion leder till ökad förståelse för sjukdomens utveckling. I slutändan kan detta användas för förbättrad behandling eller förebyggande av typ 2 diabetes.

LIST OF PUBLICATIONS

Scientific papers contributing to this thesis

I. Age influences DNA methylation and gene expression of *COX7A1* in human skeletal muscle.
Rönn T, Poulsen P, Hansson O, Holmkvist J, Almgren P, Nilsson P, Tuomi T, Isomaa B, Groop L, Vaag A, Ling C.
Diabetologia. 2008 Jul;51(7):1159-68.

II. Genetic variation in *ATP5O* is associated with skeletal muscle *ATP50* mRNA expression and glucose uptake in young twins.
Rönn T, Poulsen P, Tuomi T, Isomaa B, Groop L, Vaag A, Ling C.
PLoS One. 2009;4(3):e4793.

III. A common variant in *MTNR1B*, encoding melatonin receptor 1B, is associated with type 2 diabetes and fasting plasma glucose in Han Chinese individuals.
Rönn T, Wen J, Yang Z, Lu B, Du Y, Groop L, Hu R, Ling C.
Diabetologia. 2009 May;52(5):830-3.

IV. Polymorphisms associated with type 2 diabetes and gene expression influence DNA methylation in human skeletal muscle, adipose tissue, pancreatic islets and blood.
Rönn T, Randall J, Dekker Nitert M, Herrera B, Elgzyri T, Hansson O, Östman B, Söderström J, Eriksson K-F, McCarthy M, Isomaa B, Groop L, Ling C, Lindgren C.
Manuscript

Publications not included in this thesis

Genetic and epigenetic factors are associated with expression of respiratory chain component *NDUFB6* in human skeletal muscle.
Ling C, Poulsen P, Simonson S, Rönn T, Holmkvist J, Almgren P, Hagert P, Nilsson E, Mabey A, Nilsson P, Vaag A, Groop L.
The Journal of Clinical Investigation. 2007 Nov;117(11):3427-35.

Epigenetic regulation of *PPARGC1A* in human type 2 diabetic islets and effect on insulin secretion.
Ling C, Del Guerra S, Lupi R, Rönn T, Granhall C, Luthman H, Masiello P, Marchetti P, Groop L, Del Prato S.
Diabetologia. 2008 Apr;51(4):615-22.

Unique splicing pattern of the *TCF7L2* gene in human pancreatic islets.
Osmark P, Hansson O, Jonsson A, Rönn T, Groop L, Renström E.
Diabetologia. 2009 May;52(5):850-4.

Investigation of type 2 diabetes risk alleles support *CDKN2A/B*, *CDKAL1* and *TCF7L2* as susceptibility genes in a Han Chinese cohort.
Wen J, Rönn T, Olsson A, Yang Z, Lu B, Du Y, Groop L, Ling C, Hu R.
PLoS One. 2010 Feb 10;5(2):e9153.

DNA methylation and gene expression of *PPARGC1A* in human muscle is influenced by high-fat overfeeding in a birth weight dependent manner.
Brøns C, Jacobsen S, Nilsson E, Rönn T, Jensen C, Storgaard H, Poulsen P, Groop L, Ling C, Astrup A, Vaag A.
Journal of Clinical Endocrinology & Metabolism. Accepted on March 15th, 2010.

ABBREVIATIONS

BMI	body mass index
CI	confidence interval
CpG	cytosine guanine dinucleotide
DNMT	DNA methyl transferase
DZ	dizygotic
eQTL	expression quantitative trait locus
FPG	fasting plasma glucose
GWAS	genome-wide association study
HDAC	histone deacetylase
LBM	lean body mass
MAF	minor allele frequency
mQTL	methylation quantitative trait locus
MZ	monozygotic
OGTT	oral glucose tolerance test
OR	odds ratio
RMA	robust multi-array average
SAP	shrimp alkaline phosphatase
SD	standard deviation
SNP	single nucleotide polymorphism
T2D	type 2 diabetes
VO_2max	maximal aerobic capacity
WHO	World Health Organization

INTRODUCTION

Many diseases of concern today are multifactorial in nature, making them difficult to predict and cure. Type 2 diabetes is one example of this type of disease and has been subjected to intense study over a long period of time. There are well established factors involved in the pathogenesis of the disease, but none of these factors constitutes a single and inevitable cause. The many risk factors involved in type 2 diabetes operate at different levels, the societal level, at the level of individual behaviour and at the biological level, resulting in a complex network of interactions.

Diabetes mellitus

Diabetes mellitus describes a group of disorders with different aetiologies, with the common denominator being a chronic increase in blood glucose levels [1]. Glycaemia is normally maintained within a narrow range by the opposing actions of the key hormones insulin and glucagon, both of which are produced in the pancreatic islets of Langerhans. Insulin is an anabolic hormone that promotes glucose uptake into liver, muscle and adipose tissue after food intake, thereby storing energy for later use. During fasting, the catabolic action of glucagon promotes the mobilisation of stored energy, mainly through hepatic glucose production. In diabetes mellitus, the interplay between these two hormones is disturbed, causing deterioration of fundamental body functions.

The definition of chronic hyperglycaemia that is characteristic of diabetes mellitus, according to the World Health Organization (WHO) is: a fasting plasma glucose concentration ≥ 7.0 mmol/l or plasma glucose ≥ 11.1 mmol/l two hours after a 75 g oral glucose load measured on two occasions in asymptomatic patients [1]. The most well-characterised sub-types of diabetes include type 1 diabetes, type 2 diabetes, latent autoimmune diabetes in adults (LADA), maturity onset diabetes of the young (MODY) and maternally inherited diabetes and deafness (MIDD), of which type 2 diabetes is by far the most prevalent (> 90% of all diabetic patients).

For 2010, the estimated global prevalence of diabetes is 6.4%, representing 285 million individuals of the world's adult population. In Sweden, 484,400 people are known to have diabetes, *i.e.*, 5.2% of the Swedish population aged 20-79 [2]. While the prevalence of diabetes is increasing worldwide, the largest increase will occur in developing countries, particularly in Asia, including China and India, which already contain the largest number of individuals with diabetes [3].

Type 2 diabetes

Type 2 diabetes is a slowly progressing, chronic disease that is characterised by hyperglycaemia resulting from impaired pancreatic beta-cell function and insulin resistance in the liver and peripheral target tissues (Figure 1). Pancreatic insulin secretion is determined both by the total beta-cell mass and the function of each individual beta-cell. In the early stages of insulin resistance, prior to the development of type 2 diabetes, there appears to be an increase in islet mass [4]. However, in patients with overt type 2 diabetes, beta-cell mass is often reduced, partially due to increased apoptosis induced by high concentrations of glucose (glucotoxicity) and free fatty acids (lipotoxicity) [4].

The first choice for treatment of type 2 diabetes is life-style interventions followed by oral antidiabetic agents that act by increasing endogenous insulin secretion or improving insulin sensitivity in the liver and/or peripheral tissues. In later stages of the disease, insulin injections often become necessary to maintain metabolic control, due to the decrease in beta-cell mass and deterioration of beta-cell function. High glucose levels can have long-term damaging effects on various organs, including the cardiovascular system, eyes, nerves and kidneys [5, 6].

Though there is a high degree of ethnic and geographic variation in the prevalence of type 2 diabetes, its prevalence is rapidly increasing worldwide, mainly due to aging populations, urbanisation and the increasing prevalence of obesity and physical inactivity [2]. Our knowledge of how these factors increase susceptibility to the disease still remains limited, and molecular mechanisms linking environmental factors to type 2 diabetes currently require further attention.

Insulin resistance

Insulin resistance is a condition that occurs when cells in the body become unable to respond to normal amounts of insulin [7]. The major target organs for insulin are skeletal muscle and the liver because it is the sites where the majority of glucose uptake occurs after a meal [8]. Insulin is required to activate GLUT4, the transporter that is responsible for glucose uptake in both skeletal muscle and fat cells. Insulin also stimulates glycogen synthase, which catalyses the process of converting glucose post-prandially to glycogen [8]. Insulin resistance leads to reduced glucose clearance from blood and, subsequently, a decrease in glucose oxidation and glycogen storage. Though adipose tissue only accounts for a small portion of glucose clearance, it is important in total glucose homeostasis, as a reduced effect of insulin in fat cells results in increased hydrolysis of triglycerides, which may further increase insulin resistance and hyperinsulinaemia by oxidative substrate competition [8, 9]. Insulin resistance in hepatocytes results in failure to suppress glucose production [8]. The decrease in glucose clearance from the blood and the increase in hepatic glucose production contribute to hyperglycaemia, forcing the pancreatic beta-cells to compensate by

producing more insulin. Enhanced insulin release from the pancreatic beta-cells maintains normoglycaemia in the early stages of insulin resistance, resulting in high plasma levels of insulin. When the beta-cells fail to adapt to the increasing demands for insulin production, maintenance of this balance is lost and hyperglycaemia and, eventually, type 2 diabetes occur (Figure 1).

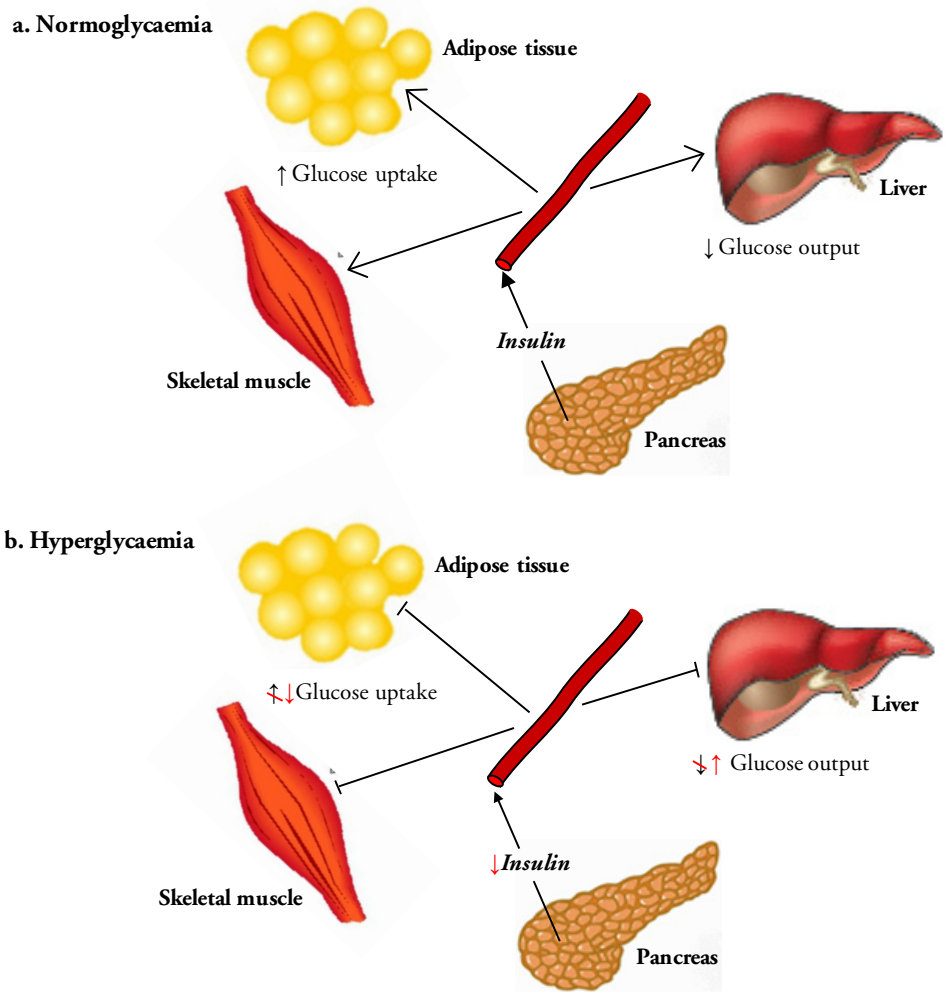

Figure 1. Insulin action and pathophysiology of hyperglycaemia in type 2 diabetes

a. In normoglycaemic individuals, insulin suppresses hepatic glucose production and stimulates glucose uptake into skeletal muscle and adipose tissue.
b. Type 2 diabetic patients display defects of both pancreatic insulin secretion and insulin action in target tissues, leading to a decrease in glucose clearance from the blood and an increase in glucose output from the liver, resulting in hyperglycaemia.

Insulin resistance is common in people with visceral adiposity, hypertension, hyperglycaemia and dyslipidaemia. Adipose cells produce significant amounts of adipokines, including leptin, adiponectin, cytokines and proteins involved in coagulation and vascular control [10]. The primary role for leptin is to signal energy deficiency to the brain, but it also suppresses hepatic glucose production and stimulates lipid catabolism [10, 11]. Several studies have suggested that proinflammatory cytokines disrupt normal insulin action in fat and muscle cells and may, thus, be a major factor causing whole-body insulin resistance in obese patients [11]. Additionally, visceral adiposity is related to an accumulation of fat in the liver, leading to an excessive release of free fatty acids into the bloodstream and increased hepatic glucose production that, in turn, exacerbates peripheral insulin resistance and increases the likelihood of type 2 diabetes. Insulin resistance has also been found to be associated with infections, which are mediated primarily by cytokines, and in association with the use of glucocorticoids.

Oxidative phosphorylation

Oxidative phosphorylation takes place in the inner membrane of the mitochondria and is the main source of energy production in human cells. Mitochondrial oxidation of carbohydrates, proteins and fat requires proper cellular uptake of nutrients and results in the production of H_2O, CO_2 and ATP [12]. During oxidative phosphorylation, a set of redox reactions occurs in which electrons are transferred from electron donors to electron acceptors that release the energy that is used to form ATP. These redox reactions are carried out by sets of enzymes in the electron transport chain, which contains five protein complexes (Figure 2) [12]. In pancreatic beta-cells, the production of ATP is crucial for normal insulin secretion to occur. In other tissues, such as skeletal muscle and fat, the production of ATP is also essential for energy release and to maintain proper whole-body metabolism. In the muscle and liver of insulin-resistant individuals, there is an increase in lipid accumulation that is associated with impaired mitochondrial oxidative activity [13]. Oxidative phosphorylation is thought to be of great interest in understanding the pathogenesis of type 2 diabetes because impaired ATP production due to reduced oxidative phosphorylation is considered a common denominator in pancreatic islets and skeletal muscle, both of which are target tissues in type 2 diabetes [14]. However, whether mitochondrial dysfunction is a cause or a consequence of metabolic disorders is not clear [12].

Oxidative capacity and mitochondrial function in skeletal muscle decline with age, as well as in insulin-resistant and type 2 diabetic patients [13, 15]. Mitochondrial function and mitochondrial biogenesis require the expression of genes that are encoded by both nuclear and mitochondrial DNA. Both mitochondrial- and a number of nuclear-encoded genes that are involved in oxidative phosphorylation are down-regulated in skeletal muscle from patients with type 2 diabetes [16-19] and in

elderly individuals [20, 21], which is most likely a result of both inherited and environmental factors [20, 21].

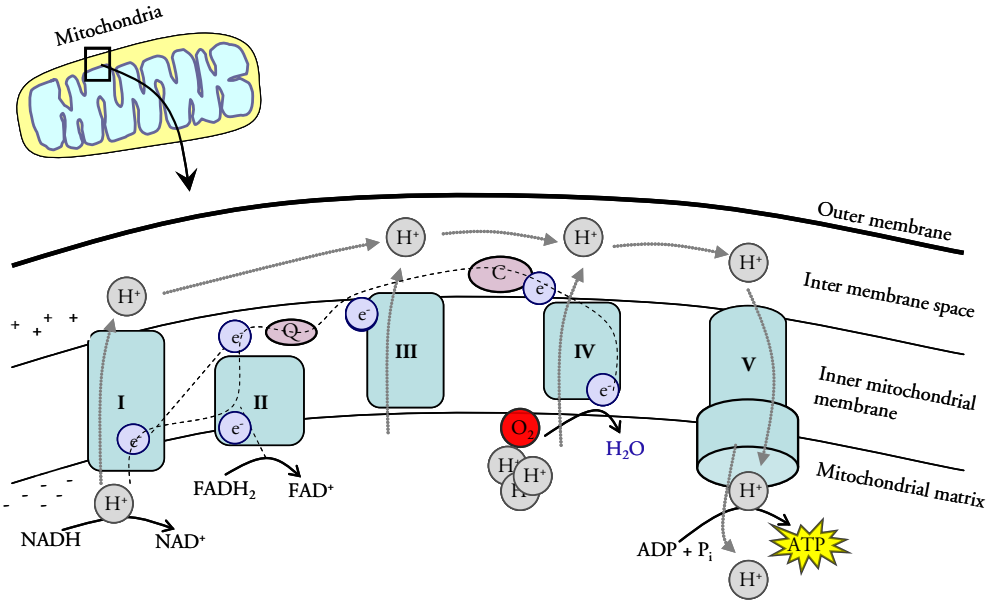

Figure 2. The electron transport chain and oxidative phosphorylation

The respiratory chain contains five multiprotein enzyme complexes and two electron carriers. Complex I, NADH dehydrogenase; Complex II, succinate dehydrogenase; Complex III, ubiquinol-cytochrome c; Complex IV, cytochrome c oxidase; Complex V, ATP synthase; Q, coenzyme Q ubiquinone; C, cytochrome c.

Genetics of type 2 diabetes

The human genome comprises approximately three billion base pairs, and the information that it stores is determined by the order of the four different nucleotides within DNA sequences. The total number of protein-coding genes, not including non-coding RNAs, is estimated to be from 20,000 – 25,000 [22]. Recently, it has been shown that most of the human genome is actually transcribed, though there are many transcripts to which a biological role cannot yet be assigned [23]. Nucleotide diversity between humans is approximately 0.1%, and the most abundant genetic variation is in the form of single nucleotide polymorphisms (SNPs). It is estimated that the genome contains approximately 10 million common SNPs, *i.e.*, SNPs having a minor allele frequency (MAF) ≥ 5% [24].

Type 2 diabetes has long been attributed to a complex interaction between an individual's genetic background and multiple environmental factors. The genetic

contribution has been confirmed by twin, family and population studies. For example, the type 2 diabetes concordance rate is higher in monozygotic twins compared with dizygotic twins [25, 26], and individuals who have one parent with type 2 diabetes have a 40% lifetime risk for developing the disease, and if both parents are affected the risk is higher [27]. This finding clearly demonstrates that the closer the genetic relationship between two individuals, the more likely they are to have the same glucose tolerance status. Many genetic variants have been convincingly and repeatedly found to associate with the disease, each of which confers only a small increase in risk, making causality difficult to prove and also limiting the prognostic and diagnostic potential of these individual variants [28].

Genome-wide association studies

Although the genetics of type 2 diabetes have been subjected to intense study for a long period of time, until recently the progress of these investigations has been limited. The identification of genetic variation in *TCF7L2* in 2006 followed by the completion of a number of genome-wide association studies (GWASs) was a breakthrough in the genetics of type 2 diabetes [29-33]. GWASs and subsequent large meta-analysis identified several novel type 2 diabetes risk variants. However, for many of these variants the function still remains to be elucidated [28, 34, 35]. GWAS are performed using large case-control cohorts without a prior hypothesis, making them ideal for finding new risk genes and pathophysiological pathways. Most of the established type 2 diabetes risk loci have been found to have a predominant effect on insulin secretion, including SNPs in or near *KCNJ11, TCF7L2, CDKAL1, CDKN2A/B, IGF2BP2, HNF1B, HHEX/IDE, JAZF1* and *SLC30A8*, whereas a *PPARG* SNP influences insulin action. Other SNPs influence the risk of type 2 diabetes through influencing related traits: for example *FTO* has a primary association with body mass index (BMI) [36], and *MTNR1B* was first identified to associate with fasting glucose levels [37-39].

Epigenetics in type 2 diabetes

Epigenetics has been described as the study of the heritable changes in gene function that occur without a change in the DNA sequence. There are two distinct epigenetic processes that are known to occur: changes in chromatin-associated proteins, such as histone modifications, and DNA methylation. By regulating chromatin structure and DNA accessibility, these chemical modifications influence genome structure during different developmental stages and diseases. Whereas every cell in the human body has the same genome, the epigenome is different in different cells and tissue types and may also change over time. Epigenetic modifications can also be more easily manipulated than genomic mutations and, thus, present the potential for pharmacological interventions.

Epigenetic mechanisms such as DNA methylation and histone modifications are increasingly considered to be important in phenotype transmission and the development of different diseases. In 1992, the thrifty phenotype hypothesis was introduced by Hales and Barker [40], who proposed that the environment in early life might influence later risk of developing type 2 diabetes. In particular, low birth weight and poor nutrition in early life have been shown to have a connection to adult type 2 diabetes and, accordingly, to impaired insulin secretion and insulin resistance [41]. Inadequate nutrition may lead to chronic alterations in the body's ability to maintain metabolism, hormone levels and the cell number of important organs [42]. One hypothesis to explain how events in early life can give rise to disease first decades later is that developmental plasticity allows the early human embryo to adapt to the intrauterine condition of malnutrition, and, when the environmental situation changes later in life, the benefit of the improved ability to make use of nutrients becomes a disadvantage. Because the genome itself cannot change, this developmental programming could potentially be explained by epigenetic regulation. However, there is currently limited knowledge about whether epigenetic factors influence the pathogenesis of type 2 diabetes.

Some evidence for the involvement of epigenetics in the pathogenesis of type 2 diabetes comes from a data-mining analysis of more than 12 million Medline records. Methylation and chromatin were both found among the top hits having a relation to type 2 diabetes [43]. There are also several examples of epigenetic involvement in the regulation of insulin gene expression, both at the levels of DNA methylation and chromatin structure [44]. Additionally, type 2 diabetic patients have been shown to have a decrease in S-adenosylmethionine, the main physiological donor of methyl groups. This decrease was also associated with the progression of the disease [45]. Our lifestyles and environment can, therefore, affect the way that our genes are expressed, leading even identical twins to become distinct as they age.

DNA methylation

In differentiated mammalian cells, the addition of methyl groups to DNA occurs on cytosine residues, and these modifications are mostly established in the context of cytosine guanine dinucleotides (CpGs). The enzymes that are responsible for the process of DNA methylation in humans are DNA methyl transferases (DNMTs), which act in concert with different regulatory factors. DNMT1 is the most abundant of these and is considered the key maintenance methyl transferase, acting primarily on hemimethylated DNA, *i.e.*, DNA where only one of the two strands is methylated. DNMT3A and 3B are required for *de novo* methylation, *i.e.*, to establish DNA methylation patterns early in development, but they also have a role in maintaining the global pattern of DNA methylation. DNMT3L has no catalytic activity but is nevertheless essential for methylation imprinting and appropriate gene expression, as suggested by its ability to stimulate and direct the other *de novo* DNMTs [46].

DNA methylation is essential for normal development to occur and is responsible for X-chromosome inactivation in females. It also remains an important process for the survival of differentiated cells. Methylation of CpGs can regulate gene expression in different ways (Figure 3) and is generally associated with gene repression [47]: the methyl group attached to the DNA could directly prevent transcription factors or other proteins from binding to the target DNA sequence (Figure 3b), or transcriptional repressors with a methyl-CpG binding domain may bind and inhibit transcription by either preventing interactions with activators or by recruiting histone deacetylases and additional co-repressors and thereby modifying the surrounding chromatin structure (Figure 3c) [47]. Recently, it was shown that the nucleosomal confirmation can be transformed directly by methylation into a more compact, rigid and more tightly wrapped nucleosome structure [48].

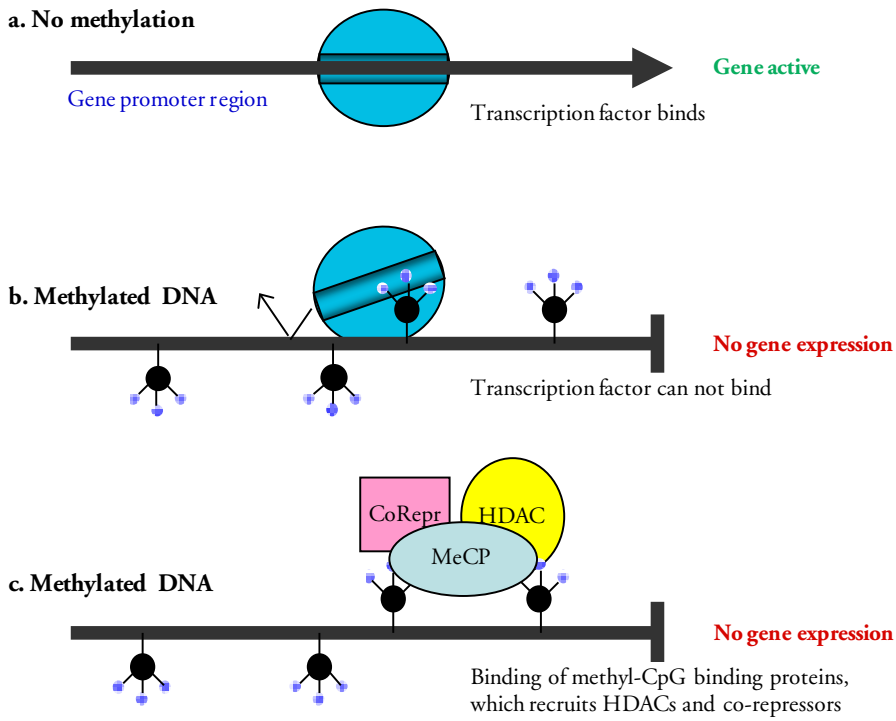

Figure 3. CpG DNA methylation and regulation of gene expression

A simplified picture of how DNA methylation interferes with the binding of activator or repressor molecules, thereby having the potential to influence gene expression. The black lines represent the gene promoter region; the blue ovals represent transcription factors; and the black balls with protruding blue dots represent methyl groups. **a.** No methyl group is attached and a transcription factor has access to and binds the DNA and activates gene expression. **b.** The transcription factor cannot recognise and access its binding site when it contains a methylated CpG and no activation of gene expression occurs. **c.** Methyl-CpG binding proteins (MeCP) are attracted to methylated CpGs and act to repress gene expression by additional recruitment of histone deacetylases (HDACs) and different co-repressors (CoRepr).

AIMS OF THIS THESIS

The overall aim of this thesis was to explore the role of DNA methylation and common genetic variation in the pathogenesis of type 2 diabetes. Environmental factors were also considered and the studies that were carried out include both candidate gene and genome-wide approaches.

In studies I and II, genetic and epigenetic factors were analysed in candidate genes that are involved in oxidative phosphorylation and then related to their expression and to glucose metabolism and risk of type 2 diabetes.

Study III aimed to replicate in a Chinese cohort an association that was previously reported in Europeans of a genetic variant in *MTNR1B* with fasting plasma glucose and type 2 diabetes.

The aim of study IV was to select common genetic variants (SNPs) that introduce or delete possible DNA methylation sites (CG-dinucleotides) and that are associated with type 2 diabetes and gene expression and to determine whether these SNPs are associated with DNA methylation status and gene expression in human skeletal muscle, adipose tissue, blood cells and pancreatic islets.

STUDY PARTICIPANTS

All studies were approved by the regional ethics committees and conducted according to the Declaration of Helsinki.

Twin studies

Subjects included in the twin cohort in studies I and II were selected from the Danish Twin Register. A total number of 98 non-diabetic monozygotic (MZ) and dizygotic (DZ) same sex twin pairs were selected; 55 were young (22-31 years), and 43 were elderly (57-66 years). Blood samples and biopsies from the *vastus lateralis* muscle were obtained before and after a two-hour hyperinsulinaemic euglycaemic clamp. To estimate rates of glucose and lipid oxidation, indirect calorimetry was performed in the basal state and at the end of the clamp. Body composition was determined by DEXA scanning [49].

Table 1. Clinical characteristics of participants in the twin study

Characteristics	Twins	
	Young	Elderly
n (male/female)	110 (60/50)	86 (38/48)
n (monozygotic/dizygotic)	110 (66/44)	86 (42/44)
Age (years)	28.0 ± 1.9	62.4 ± 2.0
BMI (kg/m^2)	24.1 ± 3.1	26.1 ± 4.4
VO_2max (maximal aerobic capacity, ml/kg LBM/min)	39.6 ± 7.8	26.3 ± 6.9
Insulin-stimulated glucose uptake (mg/kg LBM/min)	11.7 ± 3.2	9.9 ± 3.3

Data are expressed as the means ± SD. LBM, lean body mass.

Type 2 diabetes case-control cohorts

In all studies, type 2 diabetes was diagnosed according to 1999 WHO criteria [1].

Botnia

The Botnia Study cohort, which was used in papers I and II, is a family-based study that was established in 1990 with the aim of identifying type 2 diabetes susceptibility genes [50]. 1466 individuals from the Botnia Study were included in our studies: 751 type 2 diabetic patients, diagnosed after the age of 35, and 715 non-diabetic controls, without any first-degree relatives with type 2 diabetes.

Malmö

A second type 2 diabetes case-control cohort from Malmö that included 2830 Scandinavian type 2 diabetes cases from a local diabetes register with the age of onset above 35 years was used in study I [51]. 3550 unrelated, ethnically matched control individuals were selected from the population-based Malmö Diet and Cancer Study [52]. The controls had fasting plasma glucose levels below 5.6 mmol/l and no family history of type 2 diabetes.

Han Chinese

In study III, a Han Chinese cohort from Shanghai was used, including 1165 patients diagnosed with type 2 diabetes after the age of 27 and 1105 normoglycaemic controls. All individuals were unrelated and those with known subtypes of diabetes were excluded based on antibody measurements and pattern of inheritance. Controls were older than 50 years, had no family history of diabetes, and normoglycaemia was verified by an oral glucose tolerance test (OGTT).

Table 2. Clinical characteristics of study participants in case-control cohorts

Characteristics	Botnia cohort		Malmö cohort		Chinese cohort	
	Cases	Controls	Cases	Controls	Cases	Controls
N	751	715	2830	3550	1165	1105
male/female	399/352	345/370	1667/1163	1340/2210	455/710	349/756
Age (years)	54.5 ± 9.4	53.7 ± 11.4	57.9 ± 11.5	57.5 ± 6.0	60.3 ± 10.9	59.4 ± 7.7
BMI (kg/m^2)	28.9 ± 4.8	25.8 ± 3.7	29.6 ± 5.5	25.1 ± 3.6	25.2 ± 3.4	24.1 ± 3.0
FPG (mmol/l)	9.1 ± 3.2	5.3 ± 0.5	11.9 ± 4.3	5.4 ± 0.4	8.4 ± 3.0	5.2 ± 0.4
2 h PG (mmol/l)	13.8 ± 5.8	5.2 ± 1.2			15.1 ± 5.3	6.0 ± 1.0

Data are expressed as the means ± SD. FPG, fasting plasma glucose; PG, plasma glucose.

Intervention studies

In study IV, two cohorts from two different exercise intervention studies were included for a total of 114 individuals. All subjects underwent a physical examination, an oral glucose tolerance test and a test for aerobic capacity: either a submaximal exercise test (Malmö) or a 10-min walking test (Botnia). Muscle biopsies from the *vastus lateralis* were obtained during the fasting state simultaneously with subcutaneous fat biopsies from the thigh and venous blood samples. Both DNA and RNA were extracted from all tissues including the blood samples.

Malmö intervention study

The Malmö cohort included 50 sedentary but healthy young men, with a mean age of 37.6 years. They were selected to have a first-degree family history of type 2 diabetes (n=24) or not (n=26), and the two groups were matched for age, BMI and VO_2max.

Botnia PPP intervention study

The Botnia PPP (prevalence, prediction and prevention of diabetes) study is a population-based study from the Botnia region in Finland that was initiated in 2004 [53]. From the Botnia PPP, 64 normoglycaemic men and women of different ages were included in the study: 32 of them with and 32 without a family history of type 2 diabetes.

Table 3. Clinical characteristics of participants in exercise intervention studies

Characteristics	Malmö intervention		Botnia intervention	
	FH+	FH-	FH+	FH-
n (male/female)	24 (24/0)	26 (26/0)	32 (17/15)	32 (18/14)
Age (years)	38.0 ± 3.5	37.4 ± 4.9	49.5 ± 11.4	51.0 ± 12.5
BMI (kg/m^2)	28.2 ± 3.0	27.8 ± 3.3	28.4 ± 4.8	27.1 ± 3.1
FPG (mmol/l)	4.2 ± 0.4	4.3 ± 0.6	5.4 ± 0.6	5.3 ± 0.5
VO_2max (ml/kg LBM/min)	41.8 ± 4.6	40.0 ± 5.4	39.1 ± 6.8	38.9 ± 6.5

Data are expressed as the means ± SD. FH+, first degree family history of type 2 diabetes; FH-, no family history of type 2 diabetes; FPG, fasting plasma glucose; VO_2max, maximal aerobic capacity.

Human islets

24 DNA and RNA samples of human pancreatic islets from the human tissue laboratory at the Lund University Diabetes Centre (LUDC) were used for these studies. Human islets and clinical information about the donors were provided by Professor Olle Korsgren at the Nordic network for clinical islet transplantation, Uppsala University, Sweden. Both men and women with a mean age of 55.4 (range 26 to 73 years) and a mean BMI of 25.2 kg/m^2 (range 17.6 to 29.4 kg/m^2) were included.

Table 4. Clinical characteristics of donors of human pancreatic islets

Characteristics	Human pancreatic islets
n (male/female)	24 (14/10)
Age (years)	55.4 ± 13.1
BMI (kg/m^2)	25.2 ± 3.0

Data are expressed as the means ± SD.

METHODOLOGY

Phenotype characterisation

An oral glucose tolerance test (OGTT) can be used to diagnose diabetes mellitus [54]. A dose of 75 g of glucose is given to a fasting patient, and blood glucose levels are then measured over the following two hours. After two hours, a glucose level < 7.8 mmol/l is considered normal; between 7.8 and 11.0 mmol/l is classified as impaired glucose tolerance; and a glucose levels ≥ 11.1 mmol/l confirms a diagnosis of diabetes [1].

The hyperinsulinaemic euglycaemic clamp is a method for investigating and quantifying insulin resistance [55]. A primed-constant infusion of insulin (40 $mU/m^2/min$) is given for two hours, and the amount of glucose that must be infused to maintain normoglycaemia is used as a measure of insulin-stimulated glucose metabolism. Data from the last 30 minutes of the test are used because the glucose infusion rate is assumed to have achieved steady state during this period. A glucose infusion rate of > 7.5 mg/min indicates that the patient is insulin-sensitive, and rates < 4.0 mg/min suggest insulin resistance. Adding tritiated-labelled glucose to the study enables the calculation of endogenous glucose production and whole-body insulin-stimulated glucose metabolism as the rate of appearance (Ra) and the rate of disappearance (Rd) of glucose.

Selection of single nucleotide polymorphisms (SNPs)

The HapMap project is an international collaboration with the aim to catalogue common genetic variants in the human genome in populations of different ethnicities [24, 56]. Genetic markers or SNPs that are located close to each other are inherited together more frequently than would occur by chance. This phenomenon is described as linkage disequilibrium (LD), where a high LD indicates that no or low recombination has occurred between the markers. This inheritance pattern of SNPs can be used to create haplotypes, a combination of alleles at multiple loci that are transmitted together on the same chromosome. Tag SNPs are a reduced number of SNPs that are used to describe a particular haplotype [57]. In studies I and II, tag SNPs in two candidate gene regions were selected for genotyping based on HapMap data (Release #23 phase II, on NCBI B36 assembly and dbSNP 126). In study IV we also used HapMap data for SNP selection and design of the study.

Genotyping

Two different methods were used for SNP genotyping: TaqMan genotyping to obtain data on individual genotypes, and iPLEX genotyping to analyse several SNPs simultaneously. For the design of studies III and IV, we used already existing genotype data from GWAS [30, 31, 33, 35].

iPLEX

iPLEX genotyping by matrix-assisted laser desorption ionisation-time of flight mass spectrometry (MALDI-TOF MS) on the MassARRAY platform (Sequenom, San Diego, CA, USA) was used in all four studies. The iPLEX assay is based on single base extension and uses a DNA polymerase to incorporate a single labelled dideoxynucleoside triphosphate (ddNTP) next to the SNP. Primer extension products, representing the SNP alleles, are then separated based on mass. The assay can be multiplexed by adding non-template nucleotides to the 5' end of the extension primer, thereby increasing the mass difference between alleles [58, 59].

TaqMan

TaqMan allelic discrimination assays, run on the ABI7900HT sequence detection system (Applied Biosystems, Foster City, CA, USA), were used in studies I and II in cases where iPLEX failed and for the purpose of technical replication. The allelic discrimination assay contains two different fluorescently labelled probes, one for each allele of the SNP. Each probe consists of an oligonucleotide with a fluorescent reporter dye at the 5' end and a quencher at the 3' end. During PCR amplification, the TaqMan probes hybridise only to perfectly matching DNA, and Taq polymerase with 5' to 3' exonuclease activity cleaves the hybridised probe. This cleavage separates the quencher from the reporter, allowing the fluorescence of the reporter dye to be detected. The measured emission represents the genotype of each sample [60, 61].

DNA methylation analysis

Methods for DNA methylation analysis can be either gene-specific or global; in the current studies we have focused on gene-specific methods. For gene-specific methylation analysis, a number of different techniques have been developed. Today, the bisulfite conversion method is the most commonly used DNA methylation technique for identifying specific methylation patterns.

Bisulfite sequencing

Treatment of DNA with sodium bisulfite converts cytosine residues to uracil, but does not affect 5-methylcytosine residues, *i.e.*, methylated DNA. In the resulting

genomic sequence that is produced after PCR amplification, a cytosine residue denotes a methylated site whereas unmethylated DNA is seen as a thymine (Figure 4).

The first bisulfite sequencing method that we used in our studies (I and II) included bisulfite treatment using the EZ DNA methylation kit (Zymo Research, Orange, CA, USA) followed by nested PCR with primers designed using MethPrimer [62]. The resulting PCR products were cloned into plasmid vectors (pCR 4-TOPO; Invitrogen, Carlsbad, CA, USA) and transformed and amplified using *Escherichia coli*. DNA from ten colonies of each sample was isolated and sequenced, and the level of DNA methylation was determined using BiQ Analyzer [63].

Secondly, following the bisulfite conversion and PCR amplification steps, we directly sequenced t samples. To be able to obtain a quantitative result, the sequencing trace files were analysed using ESME (epigenetic sequencing methylation analysis software; Epigenomics, Berlin, Germany) [64].

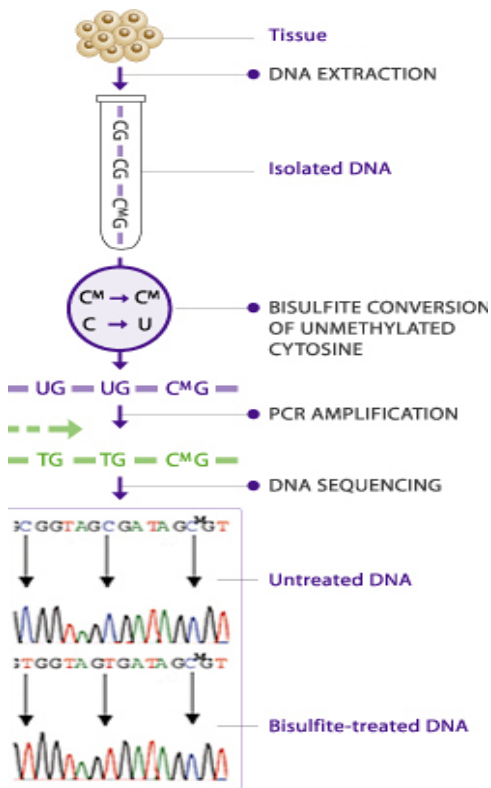

Figure 4. Analysis of DNA methylation using bisulfite conversion followed by PCR amplification and sequencing

A simplified picture of the bisulfite conversion method and the following sequencing trace file. The figure is adapted from http://www.acgtinc.com/specialty_dna_sequencing.htm.

EpiTYPER

EpiTYPER (Sequenom) is a tool for discovery and quantification of DNA methylation using the MassARRAY system, and was used in study IV. This method starts with bisulfite treatment of genomic DNA, followed by PCR amplification of selected regions in which a T7-promoter tag is introduced. These first steps are followed by *in vitro* RNA transcription on the reverse strand, and in the next step RNase A is used for base-specific cleavage. The cleavage products are analysed using MALDI-TOF MS, which discriminates between methylated and non-methylated DNA due to a shift in mass. By comparing the signal intensity between the mass signal of non-methylated and methylated template, the relative amount of methylation can be calculated [65, 66].

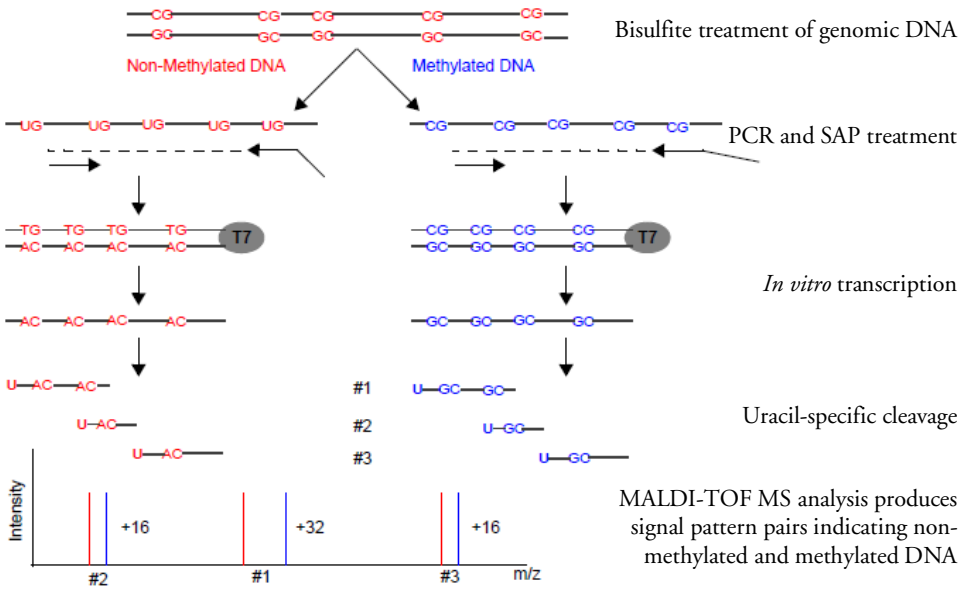

Figure 5. Overview of EpiTYPER process

Picture adapted from Sequenom [65]. SAP, shrimp alkaline phosphatase.

mRNA expression analysis

In studies I and II, mRNA levels were quantified by TaqMan real-time PCR with an ABI7900HT detection system (Applied Biosystems) to identify changes in gene expression. In contrast to real-time PCR, the microarray method allows for quantification of most transcripts present in a sample, and this was applied in study IV. Both of these methods require that mRNA is converted into cDNA, using reverse transcription and random primers.

TaqMan

TaqMan real-time PCR uses continuous measurements of fluorescence throughout the PCR reaction to quantify differences in cDNA levels between samples. The technique is based on the 5' to 3' exonuclease activity of Taq polymerase. In the real-time PCR reaction, a combination of primers and a probe are used, where the probe is marked with a 5' fluorescent reporter dye and a 3' quencher. When the probe is free in solution, the quencher prevents the reporter dye from emitting a fluorescent signal. However, when the probe hybridises to the DNA and gets incorporated into a new DNA strand, both the quencher and the reporter dye are cleaved and released. The reporter dye then emits the fluorescent signal, which is measured each cycle by a detector. The cycle number where the fluorescence exceeds a threshold value is then recorded as the Ct-value, and these values can be compared between different samples [67]. Transcript quantity of the studied genes was normalised to the mRNA level of *cyclophilin A*, an endogenous control known to have consistent expression levels in skeletal muscle. Data were calculated using the standard curve method, *i.e.*, a standard curve was generated for each primer/probe set confirmed to increase linearly with increasing amounts of cDNA [68].

Microarray

The Affymetrix oligonucleotide microarray method (Affymetrix, Santa Clara, CA, USA) is based on slides with a solid surface that are covered with thousands of short DNA molecules called oligomere probes. The DNA sequence for each probe is specific for a certain transcript. The samples to be analysed are first labelled with a fluorescent dye and then hybridised to the arrays, and a laser is used to visualise the hybridised transcripts. The light emission is detected as a measure of the transcript level and is used for analysis of differences in gene expression between groups of samples. On the Human Gene 1.0 ST Array (Affymetrix), all well-annotated genes are represented by approximately 26 probes each [69]. Basic Affymetrix chip and experimental quality analyses were performed using the Expression Console Software package, and the robust multi-array average (RMA) method was used for probe summarisation and data normalisation [70].

Statistical analysis

Phenotype characteristics are described as the means ± standard deviation (SD). Non-normally distributed data were logarithmically transformed. Statistical calculations in studies I-III were performed using Number Cruncher Statistical Systems 2004 release (NCSS; Kaysville, UT, USA).

Association testing

Association of individual genetic variants and type 2 diabetes was calculated using logistic regression, adjusted for age, sex and BMI, assuming an additive genetic model. The results are presented as odds ratio (OR) with 95% confidence intervals (CI). Power calculations were performed using the genetic power calculator (http://pngu.mgh.harvard.edu/~purcell/gpc/), assuming a type 2 diabetes prevalence of 6% in the population, a genotype relative risk of 1.3 at alpha = 0.05 and a minor allele frequency (MAF) > 5%.

Linear regression

In studies I and II, we used backward-elimination multivariate regression analysis to identify factors independently associated with the response variable with $p > 0.05$ set as the level for exclusion of model terms. In study III, a multivariate linear regression analysis including age, sex and BMI as covariates was used to test genetic association with fasting plasma glucose.

Generalised estimating equations

A basic assumption of regression analysis is that all observations are statistically independent. To correct for the strong intrapair correlation of twin data, generalised estimating equations (GEE) methodology was used in studies I and II [71, 72]. All observations were used for estimation of the beta-coefficient, whereas the variance was calculated using each twin pair as one cluster.

Biometric modelling

The intraclass correlation gives an estimate of the similarity within a monozygotic or dizygotic twin pair, respectively, and is used to calculate heritability. This analysis was applied in studies I and II, together with biometric modelling, which is a method used to estimate the degree of genetic and environmental influence on a phenotypic variable [73]. The models tested include the following parameters: genetic variance due to additive genetic effects or dominant genetic effects, and environmental variance due to an individual environment not shared with co-twin or a common environment shared among co-twins.

DNA methylation

Power to detect differences in DNA methylation in studies I and II was calculated using DSS research statistical power calculator (www.dssresearch.com; Fort Worth, TX, USA). In study I, a robust rank-order test was used to test whether the total amount of DNA methylation was different between the two age groups [74].

Hierarchical clustering

In study IV, a hierarchical clustering analysis was performed in R using the 'hclust' and 'dist' methods of the Core Stats package [75]. Each observation was initially assigned to its own cluster, and the distance between each cluster was calculated using the N-dimensional Euclidean distance between the two vectors of the observed methylation ratio. The two clusters with the nearest distance were merged, and the distances between all clusters were recalculated, and the process was then repeated until there was only one remaining cluster containing all observations. Bootstrap resampling was used to generate p-values, to demonstrate with confidence that each cluster was different than the others.

mQTL analysis

The experimental variables analysed in the mQTL analysis were the percentage of methylation detected at CpG sites in selected amplicons and the genotype of a SNP in the same region. Both quality control and analysis were carried out separately in each of the four tissues in a stratified manner. We additionally controlled for several other covariates including age, sex, BMI, and plate ID (where methylation data within a single tissue was collected across more than one reaction plate).

We performed an inverse-normal transformation on the methylation data to yield a normally distributed phenotype that could be used in a standard parametric test. For each CpG site, the following procedure was performed:
The age, sex, and BMI covariates were taken in their raw form, while plate ID was converted to a contrast (a set of variables that only take on the values of 0 and 1). A linear model was then fitted using the percentage methylation as the response variable and covariates as the linear predictor variables to calculate residuals of percent methylation after regressing out the effect of the covariates. The resulting residuals were then inverse-normally transformed to yield a normal distribution by ranking the data from lowest to highest and then fitting the ranks to a normal distribution with mean of 0 and standard deviation of 1. A linear model was then fitted using the normalised methylation data as the response and the genotype of the nearby SNP (using the additive genetic model) as the predictor variable. The significance of the genotype-methylation association was then tested using a two-sided t-test. P-values were adjusted for multiple testing using Bonferroni correction to correct for a total of 1180 tests (295 CpG sites times 4 tissues).

RESULTS

Study I

Age influences DNA methylation and gene expression of COX7A1 in human skeletal muscle

The aim of this study was to investigate the regulation of *COX7A1* gene expression in human skeletal muscle and its role in glucose metabolism and type 2 diabetes. Genetic (SNPs), epigenetic (DNA methylation) and non-genetic (age, insulin and body composition) factors were studied. *COX7A1* resides at a locus that exhibits tissue-specific DNA methylation [76] and encodes a subunit of complex IV in the mitochondrial respiratory chain that is down-regulated in skeletal muscle of type 2 diabetic patients [17].

The *COX7A1* mRNA expression was decreased in skeletal muscle from elderly compared to young twins, both in the basal (1.00 ± 0.05 versus 1.68 ± 0.06) and insulin-stimulated (1.04 ± 0.05 versus 1.71 ± 0.06) states (Figure 6a). The hyperinsulinaemic euglycaemic clamp did not influence the mRNA level of *COX7A1*, suggesting that this gene is not regulated by insulin (Figure 6a). An effect of age was also seen on the level of DNA methylation in the selected 5' region of *COX7A1*, with higher methylation in elderly (19.9 ± 8.3%) than in young (1.8 ± 2.7%) twins (Figure 6b).

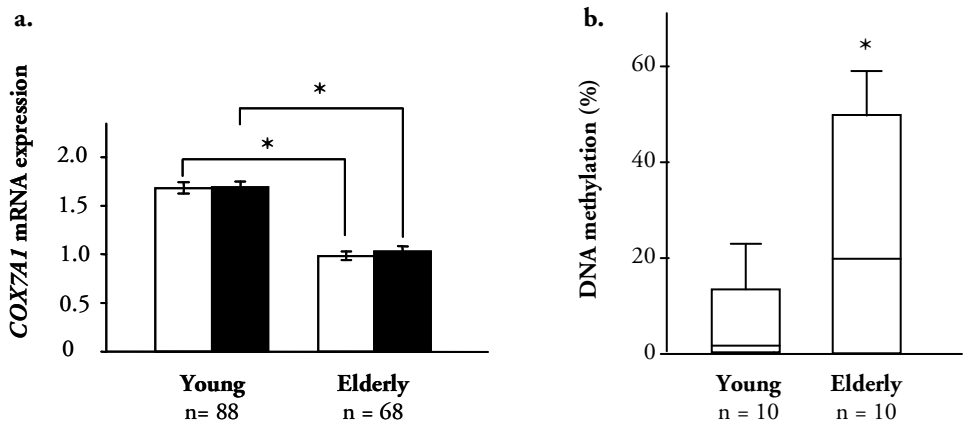

Figure 6. Age influences *COX7A1* mRNA expression and the level of DNA methylation

In **a**, the effects of age and insulin on human skeletal muscle *COX7A1* mRNA expression are shown. White bars represent the basal and black bars the insulin-stimulated state. The results are presented as the mean ± SEM. The level of DNA methylation is shown in **b**, where boxes represent the 25th to 75th percentiles with median value, and the upper whiskers represent the maximum values. * $p < 0.05$.

A multivariate regression analysis showed that the level of *COX7A1* was positively associated with *PPARGC1A* mRNA expression (regression coefficient = 0.27, p = $3x10^{-3}$). Furthermore, one genetic variant located in the 5' untranslated region of *COX7A1* (rs753420) was shown to influence *COX7A1* gene expression in skeletal muscle, but this effect was only seen in young twins. The minor allele (G) was associated with an increase in *COX7A1* expression (T/T: 1.53 ± 0.08, T/G: 1.71 ± 0.08 and G/G 2.3 ± 0.19), p = $1x10^{-4}$ based on a recessive model.

Next, we investigated whether common variants in or near the *COX7A1* gene are associated with an increased risk of type 2 diabetes. Three SNPs were genotyped in two independent cohorts. In the Botnia case-control cohort, we could not detect an association with disease for any of these three SNPs. However, in the Malmö cohort, one SNP (rs7255180) showed a nominal association with risk of type 2 diabetes (OR: 0.80 [0.66 – 0.97], p = 0.027).

To test the effect of *COX7A1* mRNA expression in skeletal muscle on *in vivo* metabolism, multivariate regression analyses including age, sex and BMI as covariates were performed. *COX7A1* was positively correlated with both insulin-stimulated glucose uptake (p = 0.01) and maximal aerobic capacity (VO_2max; p = $1x10^{-3}$).

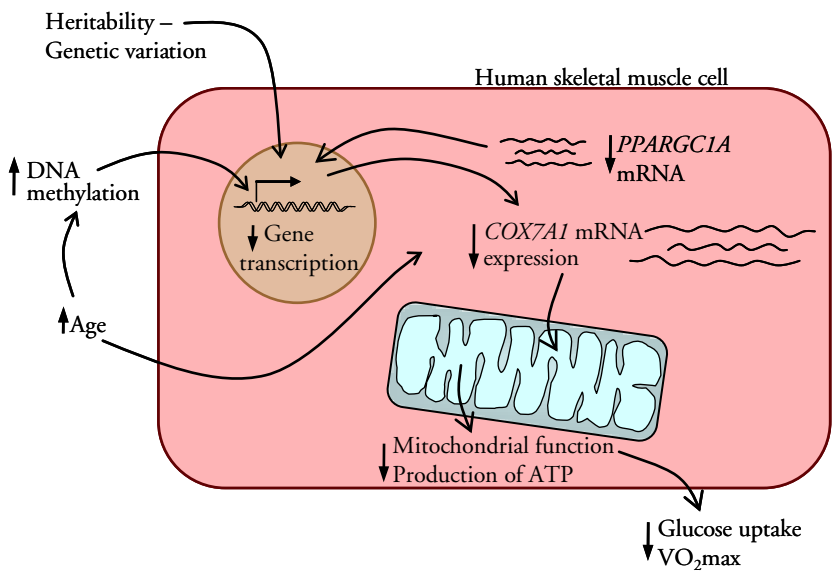

Figure 7. Proposed mechanisms for the regulation of *COX7A1* in human skeletal muscle and its effect on *in vivo* metabolism

According to our findings, age increases DNA methylation and decreases *COX7A1* mRNA expression. Furthermore, *COX7A1* mRNA expression is associated with genetic variation (SNP rs753420) and *PPARGC1A* mRNA expression. As a result, these expression changes could affect *in vivo* metabolism because *COX7A1* mRNA expression was positively associated with both insulin-stimulated glucose uptake and VO_2max.

Study II

Genetic variation in ATP5O is associated with skeletal muscle ATP5O mRNA expression and glucose uptake in young twins

The purpose of this study was similar to that of study I, to explore mechanisms regulating gene expression of a component of the respiratory chain, *ATP5O*, in human skeletal muscle and its association with glucose metabolism and type 2 diabetes. A previous study showed that *ATP5O* was the most significantly reduced oxidative phosphorylation gene in skeletal muscle from type 2 diabetics [17]. The product of this gene is part of the ATP synthase complex, which is the final step in the respiratory chain and the complex that produces ATP.

We found that the mRNA level of *ATP5O* was reduced in skeletal muscle of elderly compared to young twins, both in the basal (0.19 ± 0.01 versus 0.28 ± 0.01; $p < 5x10^{-4}$) and in the insulin-stimulated (0.21 ± 0.01 versus 0.30 ± 0.01; $p < 5x10^{-4}$) state (Figure 8). Again, insulin did not influence mRNA expression (Figure 8).

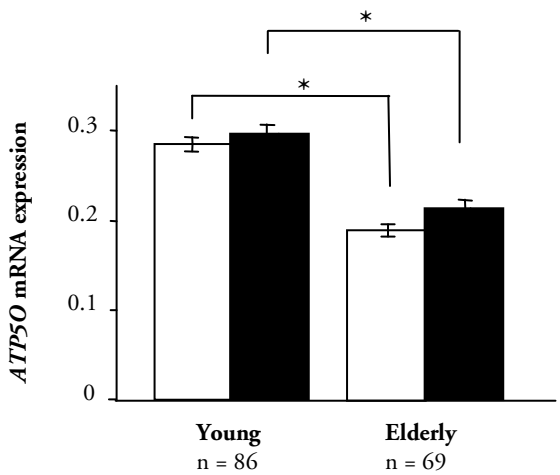

Figure 8. Age, but not insulin, associates with *ATP5O* mRNA expression in human skeletal muscle

Data are expressed as the mean ± SEM. White bars representing the basal and black bars the insulin-stimulated state. * $p < 0.05$.

Eleven tag SNPs in the *ATP5O* gene region were genotyped and related to mRNA expression. We found that two polymorphisms influenced the expression of *ATP5O* in skeletal muscle of young twins, rs12482697 (T/T 0.30 ± 0.010 and T/G 0.25 ± 0.012; $p = 0.02$) and rs11088262 (A/A 0.31 ± 0.012 and A/G 0.25 ± 0.011; $p = 4x10^{-3}$). These two SNPs are in strong LD ($r^2 = 0.96$) and are likely to represent the

same association. The 11 tag SNPs were additionally genotyped in a case-control cohort, but no association with type 2 diabetes was observed.

We also analysed DNA methylation in a region close to the *ATP5O* transcription start site located in a CpG island. The level of methylation was very low in both young (0.14 ± 0.07%) and elderly (0.73 ± 0.33%) twins, without a significant difference. We also could not observe any significant correlation between DNA methylation and *ATP5O* mRNA expression.

Finally, we found that the *ATP5O* mRNA expression in human skeletal muscle was positively correlated with insulin-stimulated glucose uptake (p = 0.02) in young and elderly twins. Additionally, the two SNPs shown to influence *ATP5O* expression were also associated with glucose uptake in the young twins (rs12482697: T/T 12.1 ± 0.36 versus T/G 9.8 ± 0.75; p = $5x10^{-3}$ and rs11088262: A/A 12.2 ± 0.4 versus A/G 9.8 ± 0.5; p = $2x10^{-3}$).

Study III

A common variant in MTNR1B, encoding melatonin receptor 1B, is associated with type 2 diabetes and fasting plasma glucose in Han Chinese individuals

Genetic variation in *MTNR1B* was recently shown to be associated with an increased risk of type 2 diabetes, increased fasting plasma glucose and impaired insulin secretion [30, 37-39]. Following these results, we aimed to replicate this finding in a Han Chinese cohort to elucidate whether this effect is seen across different populations.

The genetic variant rs10830963 is located within intron 1 of the *MTNR1B* gene. We found this SNP to be associated with increased risk of type 2 diabetes in a Han Chinese cohort including 1165 cases and 1105 controls, resulting in an OR of 1.16 (95% CI 1.03-1.31, p = 0.015). The risk variant was also associated with increased fasting plasma glucose in normoglycaemic individuals, showing an increase of 0.068 mmol/l (95% CI 0.036 – 0.100, p = $4x10^{-5}$) per risk allele.

Table 5. Association between rs10830963 and type 2 diabetes and fasting plasma glucose level

rs10830963	Genotype frequency		Association with T2D		Fasting plasma glucose (mmol/l)		
	T2D cases	Controls	OR_{add} (95% CI)[a]	p value	Controls	Per-allele effect[b]	p value
CC	0.318	0.338			5.15		
CG	0.475	0.505	1.16 (1.03-1.31)	0.015	5.25	0.068	$4*10^{-5}$
GG	0.207	0.156			5.27		

[a] calculated using logistic regression, assuming an additive model and adjusted for age, sex and BMI.
[b] calculated using multiple regression and adjusted for age, sex and BMI.
T2D, type 2 diabetes.

Study IV

Polymorphisms associated with type 2 diabetes and gene expression influence DNA methylation in human skeletal muscle, adipose tissue, pancreatic islets and blood

We initially selected genome-wide SNPs that introduce or delete possible DNA methylation sites, *i.e.*, CG dinucleotides (CpGs). Our analysis was restricted to regions 10 kb upstream (5') of the transcription start site of annotated genes and included a total of 92,766 CpG-SNPs or 28.3% of all SNPs in these regions. The CpG-SNPs were then sorted based on their association with changes in mRNA transcription levels in blood cells ($p < 1x10^{-3}$) [77] and their association with increased risk of type 2 diabetes ($p < 0.05$) [35]. This analysis resulted in a list of 657 CpG-SNPs with a MAF ≥ 0.30. Of these, 33 were selected for further analysis, where we related genotype to DNA methylation and gene expression. DNA methylation was analysed using EpiTYPER and assay design was successful for 29 CpG-SNPs and a total number of 340 CpG units, and only these units were considered in further analyses.

A hierarchical clustering analysis of the methylation data was performed, where samples with closer methylation patterns were more closely clustered. Our results clearly suggest that there is a tissue-specific pattern of DNA methylation that is different between all four tissues.

We next examined whether the CpG-SNPs affect the degree of methylation in the corresponding regions as a methylation quantitative trait locus (mQTL). DNA methylation results were obtained for 19 of the 29 analysed CpG-SNP sites. At 11 of these sites, a significant difference in DNA methylation with regard to genotype was found in all four tissues, and these differences were consistent for both the forward and reverse DNA strands (where applicable). Additionally, a significant difference in DNA methylation was seen in another seven CpG-SNPs in at least one tissue (Table 6). Only one CpG-SNP (rs922957) was not associated with the level of methylation in any of the analysed tissues, and it is of note that the DNA methylation in this case was very low in the majority of the samples. The mQTL in blood of one CpG-SNP, rs9422541 (*ZNF239*), is visualised in Figure 9.

In addition to the CpG-SNP sites, a number of ordinary CpG sites showed a difference in DNA methylation with regard to the genotype of the lead CpG-SNP in that amplicon, suggesting that genetic variation has the potential to affect the degree of methylation in the nearby region, as exemplified by rs2329573 (*CSTB*) in Figure 10. We did not obtain methylation data for the lead CpG-SNP in this amplicon, and hence, these data are missing from both Table 6 and Figure 10.

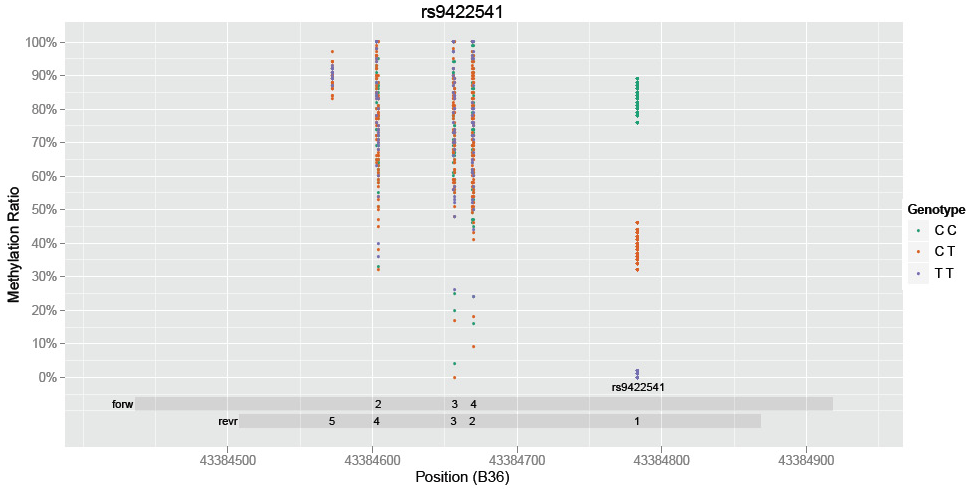

Figure 9. mQTL analysis of rs9422541 (*ZNF239*) in blood

Percentage methylation is shown on the y axis, and the x axis displays chromosomal position and CpG coverage of the forward (forw) and reverse (revr) amplicons, respectively. The methylation ratios for each blood sample in five different CpG sites are shown. Each sample is coloured based on the genotype in the CpG-SNP site, *i.e.*, site 1 on the reverse amplicon (rs9422541, C/T polymorphism).

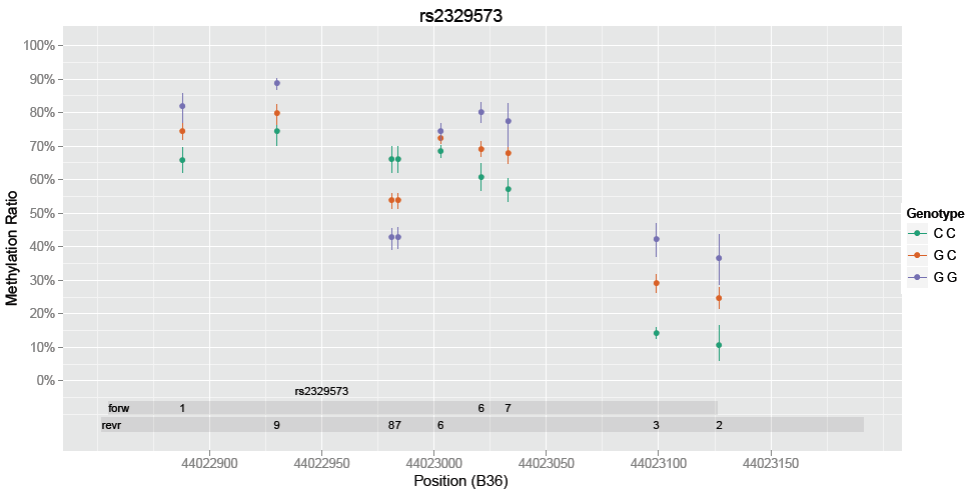

Figure 10. mQTL analysis of rs2329573 (*CSTB*) in blood

Percentage methylation is shown on the y axis, and the x axis displays the chromosomal position and CpG coverage of the forward (forw) and reverse (revr) amplicons, respectively. The methylation is shown as the mean and 95% confidence interval for individuals grouped by genotype at each CpG site. The genotype is based on the lead CpG-SNP site (rs2329573, C/G polymorphism), which did not yield methylation data itself in this example.

Table 6. mQTL results of lead CpG-SNPs

CpG-SNP	Gene	Blood		Muscle		Fat		Islets	
		Effect	*P* corr	Effect	*P* corr	Effect	*P* corr	Effect	*P* corr
rs1041456_R13	*RRP1*	-0,72	$<1x10^{-270}$	-0,71	$<1x10^{-270}$	-0,74	$1x10^{-230}$	-0,37	$2x10^{-7}$
rs1041456_F2		-0,9	$<1x10^{-270}$	-0,76	$<1x10^{-270}$	-1	$<1x10^{-270}$	-0,75	$2x10^{-38}$
rs12355908_R7	*ANK3*	1,28	$<1x10^{-270}$	0,92	$<1x10^{-270}$	0,59	$8x10^{-93}$	1,03	$5x10^{-36}$
rs12889309_R14	*NEK9*	-0,23	$3x10^{-69}$	-0,51	$3x10^{-235}$	-0,28	$6x10^{-26}$	-0,16	ns
rs12889309_F2		0,23	$3x10^{-70}$	-0,14	$7x10^{-15}$	-0,05	Ns	0,38	$1x10^{-04}$
rs1531798_F4	*USP36*	0,14	$3x10^{-19}$	0,06	ns	0,22	$6x10^{-15}$	0,43	$2x10^{-02}$
rs1872164_F2	*PTPN23*	0,04	ns	0,26	$9x10^{-44}$	0,2	$4x10^{-08}$	0,14	ns
rs2665795_R2	*SMARCD2*	0,67	$<1x10^{-270}$	0,45	$8x10^{-184}$	0,3	$2x10^{-29}$	0,63	$8x10^{-09}$
rs2665795_F1		1,26	$<1x10^{-270}$	1,12	$<1x10^{-270}$	1,18	$<1x10^{-270}$	1,5	$2x10^{-82}$
rs2953802_R5	*THEX1*	0,73	$<1x10^{-270}$	0,56	$5x10^{-249}$	-		1,12	$2x10^{-89}$
rs2953802_F3		-0,38	$2x10^{-174}$	-0,43	$1x10^{-155}$	0,15	$7x10^{-05}$	-0,52	$1x10^{-7}$
rs4660880_R2	*GPBP1L1*	0,07	$1x10^{-3}$	-		-		-	
rs6785251_F2	*WWTR1*	-0,06	$4x10^{-3}$	0,18	$1x10^{-22}$	-		0,28	ns
rs7395920_F2	*MRPL23*	-1,42	$<1x10^{-270}$	-1,41	$<1x10^{-270}$	-1,42	$<1x10^{-270}$	-0,94	$1x10^{-47}$
rs757110_R9	*KCNJ11*	-0,23	$9x10^{-51}$	-0,21	$2x10^{-28}$	-0,12	$3x10^{-2}$	-0,03	ns
rs757110_F8		-0,61	$<1x10^{-270}$	-0,34	$6x10^{-74}$	-0,21	$9x10^{-11}$	-0,59	$5x10^{-22}$
rs764129_R1	*IGF2BP2*	0,87	$<1x10^{-270}$	0,52	$2x10^{-265}$	0,86	$1x10^{-269}$	0,72	$5x10^{-16}$
rs7766189_F2	*HBS1L*	-0,37	$8x10^{-140}$	-0,7	$<1x10^{-270}$	-0,48	$1x10^{-62}$	-0,92	$4x10^{-15}$
rs799905_F11	*BRCA1*	-0,1	$5x10^{-11}$	0,04	ns	0,24	$5x10^{-18}$	-0,12	ns
rs799912_R1	*BRCA1*	1,27	$<1x10^{-270}$	1,07	$<1x10^{-270}$	0,93	$<1x10^{-270}$	0,72	$6x10^{-13}$
rs799912_F3		1,16	$<1x10^{-270}$	1,03	$<1x10^{-270}$	0,52	$2x10^{-72}$	0,93	$7x10^{-26}$
rs898893_R1	*AIM1*	-0,04	ns	-0,08	$8 x10^{-03}$	-0,06	Ns	-0,18	ns
rs922957_R12	*PTPN23*	0,42	ns	0,19	ns	-0,14	Ns	0,52	ns
rs922957_F13		0,02	ns	0,18	$8x10^{-23}$	0,25	$2x10^{-18}$	0,43	$8x10^{-04}$
rs9422541_R1	*ZNF239*	-1,3	$<1x10^{-270}$	-1,26	$<1x10^{-270}$	-1,15	$<1x10^{-270}$	-0,83	$5x10^{-30}$
rs9422545_F4	*ZNF239*	-1,22	$<1x10^{-270}$	-0,84	$<1x10^{-270}$	-0,98	$<1x10^{-270}$	-1,07	$6x10^{-91}$

Data are presented for the forward (F) and reverse (R) amplicon separately. All data are adjusted for age, sex, BMI and plate effects. CpG-SNP, lead CpG-SNP for each amplicon; Gene, nearest downstream gene of selected CpG-SNP; Effect, effect size (beta coefficient); *P* corr, p-value of differences in DNA methylation with regard to genotype, assuming an additive genetic model applied to the effect allele and corrected for multiple testing (number of analysed CpG units and tissues); -, missing data.

DISCUSSION

Type 2 diabetes has long been known to be a multifactorial disease with a complex aetiology, including both genetic and non-genetic factors. However, when the research presented in this thesis was initiated, a relatively small number of studies had attempted to integrate this complexity; most studies focussed only on either the genetic contribution or environmental factors when estimating the risk of type 2 diabetes. Additionally, the role of DNA methylation in the pathogenesis of the disease was previously unknown. In the studies we present here, we investigated not only genetic but also non-genetic factors and DNA methylation in the pathogenesis of type 2 diabetes and glucose metabolism.

Today, many genes are known to be associated with type 2 diabetes, but the effect size of each gene is generally low. Investigations based on candidate genes have mostly chosen candidates on the basis of biological function, but from many plausible candidates, only a few susceptibility genes have been convincingly replicated in several studies. In more recent years, breakthrough GWASs have resulted in the discovery of numerous genes using an unbiased means of investigating the association between a large number of common polymorphisms in the genome and type 2 diabetes or related traits. For example, polymorphisms in *MTNR1B* have been found to be associated with fasting plasma glucose, glucose-stimulated insulin secretion and type 2 diabetes [37-39]. These studies were, however, all performed in populations of European ancestry, so the aim of study III was to investigate if this association was also seen in an ethnically different population. We replicated both the association with fasting plasma glucose and the risk of type 2 diabetes in a Han Chinese cohort, suggesting that the origin of rs10830963 could be relatively ancient. The effect size of this variant was similar independent of ancestry; however, the risk allele frequency was higher in Chinese than European individuals.

Studies I and II investigated genes that are involved in oxidative phosphorylation, as impairment of oxidative capacity in skeletal muscle has been suggested to contribute to insulin resistance [15]. A number of studies have reported reduced expression of genes involved in oxidative phosphorylation and their regulators in skeletal muscle from type 2 diabetic patients [16-19] and elderly [20, 21]. *COX7A1* and *ATP5O* were both identified in a group of *PPARGC1A*-responsive genes involved in oxidative phosphorylation, using a pathway analysis designed to detect modest but coordinated changes in the expression of groups of functionally related genes [17]. As expected, we confirmed a positive association between the expression of the genes we investigated and the transcriptional co-activator *PPARGC1A* in our studies. For both *COX7A1* and *ATP5O*, we found an age-related decrease in skeletal muscle mRNA expression,

in line with earlier studies [13, 21, 78, 79]. This decrease was related to reduced insulin-stimulated glucose uptake, suggesting a role for *COX7A1* and *ATP5O* in the regulation of *in vivo* metabolism. Furthermore, genetic variation in these genes was found to be associated with skeletal muscle mRNA expression for both *COX7A1* and *ATP5O*, respectively, but not with increased risk of type 2 diabetes.

A novel finding from study I was that there is an age-dependent increase in DNA methylation of the *COX7A1* 5' region, which clearly coincides with the decreased *COX7A1* mRNA expression, though this correlation was not statistically significant. Together with previous data, this result suggests that ageing has the ability to alter the level of DNA methylation [21, 80], which could then influence gene expression and, subsequently, human metabolism. The increase in DNA methylation with age seems not to be a general phenomenon, as we did not detect increased methylation around the *ATP5O* transcription start site in the elderly individuals. A similar result was seen in a previous study of genes involved in oxidative phosphorylation, where some, but not all genes, showed an age-related increase in DNA methylation [21]. A change in DNA methylation with increasing age has also been proposed in hepatic glucokinase (*GCK*), which encodes a key enzyme in glucose utilisation that is associated with insulin resistance and type 2 diabetes. In rat hepatocytes, *gck* expression and activity declined with age, concurrent with an increase in DNA methylation. Additionally, de-methylation using 5-aza-deoxycytidine in cultured hepatocytes from elderly rats restored the *gck* expression [81].

The role of DNA methylation in the study of complex diseases such as type 2 diabetes is still not clear. The prevalence of type 2 diabetes increases with increasing age as does epigenetic dysregulation [82]. It seems likely that there is a connection between these phenomena, but it remains to be seen whether DNA methylation is the cause or a consequence of the disease. In support of a role for DNA methylation in diabetes, a recent study suggested that insulin gene expression is partially regulated by DNA methylation. Several tissues were investigated in this study, and the *INS* promoter was found to be demethylated specifically in insulin-producing beta-cells. Additionally, methylation of these CpG sites was shown experimentally to suppress insulin gene expression [83]. In another study in human pancreatic islets, we showed that there was an increased level of DNA methylation in a region upstream of *PPARGC1A* in islets from type 2 diabetic donors compared with islets from non-diabetic human donors. This increase in DNA methylation was related to a decrease in *PPARGC1A* mRNA expression, which showed a positive correlation with glucose-stimulated insulin secretion [84]. An animal model of intrauterine growth retardation, leading to diabetes in adulthood, has been used to investigate the transcription factor *Pdx1*, which is important for pancreas development and beta-cell differentiation. A progressive decrease is *Pdx1* mRNA expression levels was observed concurrent with the appearance of marks that indicate epigenetic silencing, including histone modifications and increased DNA methylation [85].

DNA methylation is a dynamic, naturally occurring process, and changes in the epigenetic profile are more common than changes in genetic variation. Differentially methylated cytosines give rise to distinct patterns that are specific for tissue type or disease state. These methylation variable positions can be seen as common epigenetic markers that will hopefully improve our understanding of the pathogenesis of common human diseases in a way that is similar to what has been found from the study of common genetic variants [86].

Our understanding of DNA methylation and its connection to transcriptional control is growing but far from complete. We previously reported that a SNP that introduces a CpG site was associated with DNA methylation, gene expression and metabolism in human skeletal muscle [21]. Additionally, the aberrant DNA methylation seen in transient neonatal diabetes has been associated with mutations in *ZFP57* [87]. In study IV, we further demonstrated that common genetic variation in CpG sites introduces changes in DNA methylation, adding a new level of complexity to our understanding of the regulation of gene expression. This relationship between genetics and epigenetics was also supported by another recent study, which showed that although genetic variation was not strongly associated with a phenotype, it could still affect the variability of that phenotype through epigenetic mechanisms [88].

In study IV, CpG-SNPs were selected based on previously discovered associations with type 2 diabetes ($p < 0.05$) [35] and gene expression in blood ($p < 1x10^{-3}$) [77], an attempt to increase the likelihood that we are analysing functional CpG-SNPs. We also restricted our search to a region 10 kb upstream of annotated genes because epigenetic influence on gene activity has been shown to mainly act in a *cis*-regulatory fashion [89]. Furthermore, it has been shown that 17% of human genes are differentially methylated in their 5' untranslated region and that one third of these cases are also correlated inversely with the transcription level of the affected gene [90]. We found a difference in DNA methylation with regard to genotype in 11 genomic regions in all analysed tissues, corresponding to 10 unique genes. Both *KCNJ11* and *IGF2BP2* are type 2 diabetes candidate genes that have been convincingly replicated to associate with the disease [30, 31, 33, 91]. *ANK3* encodes a neuronal adaptor that regulates the assembly of voltage-gated sodium channels and is suggested to have a role in bipolar disorder [92]. *HBS1L* is located in a region known to influence blood cell counts and hemoglobin content and is associated with b-Thalassemia/HbE [93]. A number of these genes are also proposed to be involved in gene regulation: *SMARCD2* is a member of the SWI/SNF family, which alters chromatin structure [94]; *THEX1* interacts with histone mRNA [95]; *RRP1* acts in the processing of rRNA [96]; the tumour suppressor gene *BRCA1* plays a role in maintenance and regulation of genome stability [97] and *ZNF239* has the ability to bind both DNA and RNA to act on both transcriptional and posttranscriptional regulation of specific genes [98]. *MRPL23* is located within a region of imprinted genes and encodes a

mitochondrial ribosomal protein that is up-regulated in chronic fatigue syndrome [99, 100].

Our current knowledge about the human epigenome is very limited, and there have been only a few studies that have investigated DNA methylation patterns in multiple tissues. In study IV, we collected DNA and RNA from four tissues that are relevant to type 2 diabetes: blood, muscle and adipose tissue from the same individuals and pancreatic islets from a separate cohort. We observed a clear tissue-specific pattern of DNA methylation, which suggests an important role for epigenetics in tissue-specific regulation of gene activity. This result also emphasises the importance of choosing the correct tissue when designing a study.

SUMMARY AND GENERAL CONCLUSION

The time as a PhD student is a constant learning process. Each time that you get the feeling that you actually understand what you are doing, it is time to move on. In the course of my PhD studies, I have experienced the methodological development to move forward in a similar way, both in the genetic and, in particular, in the epigenetic field. Techniques that seemed hardly possible when I entered the science world are well-established methods today. For instance, mapping the sequence of the human genome and the HapMap project made GWASs possible, which led to the discovery of multiple type 2 diabetes genes.

In studies I and II, we found that the mRNA expression of *ATP5O* and *COX7A1* was associated with age, genetic variation and *PPARGC1A* expression. We also found that there was a positive correlation with *in vivo* glucose uptake, suggesting that genes involved in oxidative phosphorylation may have a role in insulin sensitivity. In the case of *COX7A1*, we also found an increase in DNA methylation in elderly individuals concurrent with a decrease in mRNA expression, which demonstrates how genetic, epigenetic and non-genetic factors act in concert to guide gene expression.

Study III replicated an association between a genetic variant in *MTNR1B* and fasting plasma glucose and type 2 diabetes, showing that the Han Chinese individuals that we studied share the same risk variant as has previously been reported in Europeans.

In study IV, we showed that common genetic variants interfere with possible sites for DNA methylation, thereby causing a change in the methylation profile. This finding serves as an interesting model for a possible mechanism of how the epigenetic pattern may be inherited.

There is much work left to do until we have a clear understanding of the process of DNA methylation and its role in the pathogenesis of type 2 diabetes. It will also be interesting to determine whether DNA methylation and epigenetics might have a role in the interplay between genetics, environment and disease.

ACKNOWLEDGEMENTS

During my time as a PhD student, I have met a lot of wonderful people, both in the house and all around the world. You have brightened my time here by friendship and laughter, both at the lab and in leisure, and have given me support and encouragement whenever needed. Thank all of you who have generously shared your knowledge, provided thoughtful remarks and inspired me.

I want to express my sincere gratitude to my supervisor Charlotte Ling and my co-supervisor Leif Groop, who guided me through the elegant art of doing science. Charlotte has been a supportive, inspiring and encouraging supervisor, as well as an excellent collaborator. I really appreciate having been part of your group throughout this journey, which became more fascinating than I could ever have anticipated. Thank you Leif for always finding good solutions to problems that presented themselves; you are an everlasting source of knowledge.

Marketa & Lovisa – you have been by my side since my first day in the lab. Thanks for all of your valuable advice in the lab and for how to survive as a PhD student. Most important, thank you for a great friendship!

Thanks to the Ling group, Anders, Marloes, Beatrice, Siri, Tasnim and Elin, for being such nice and bright people to work with. Additionally, thanks to all of the past and present PhD students of the Groop group and from the time at the Wallenberg lab, who have been part of creating a pleasurable work environment and with whom I have shared a lot of fun in addition to science: Anna J, Charlotta, Maja, Tereza, Lina, Camilla, Jenny, Johan, Maria, Hemang, Yuedan, Nael, Emma N, Fredrik, Yang, Thomas, Martins, Caroline, Kristian, Eka, Hee-Bok, Avinash, Targ, Eero, Mozhgan, Charlotte G, Gull and Ivana. We have shared each others failures and successes, and we have travelled to conferences, making sure to enjoy ourselves in as many ways as possible. Marketa and Anders have been my best travelling room-mates of course, and adding Lovisa makes up a strong team for sight-seeing, shopping, dinners etc. And partying all night long…

I also appreciated the opportunity to be part of a large research network, including all of the members of DPLU and LUDC. A special thank to the DPLU PhD student group for all of the good times we have had organising seminars and dinners. You have been a great team for creating events that are always appreciated: Lovisa, Marketa, Kalle, Anders, Helena, Cecilia, Ulrika, Karin, Emma, Siri and Hedvig.

There are certain colleagues who have made my daily work easier. Thanks to June, Ulrika and Mona for assisting with practical and administrative tasks; to the technicians Malin S, Lena, Margareta, Anna B and Malin N for teaching and

problem solving in the lab; to Peter A and Claes for statistical skills and willingness to help; to Ola for good scientific discussions; to Peter O for helping to take my focus away from science with a smile; to Johan H and Mattias B, without whom my research would still be struggling without a working computer and software; and finally, to Esa for being a good room-mate - I really like your small tree. I also would like to thank Marju, Olle, Petter, Stefan, Jasmina, Valeri, Jacqueline, Gabriella, Tord, Emma A, Jalal, Emily, Emilia, Damon, Amitabh and Tarun in addition to all the people mentioned above for creating a nice environment.

I would like to thank all of the people who participated in the clinical studies and Ylva for taking good care of these people. Without you these projects would not be possible. Thanks to all of my co-authors for fruitful collaborations and engagement in reading of manuscripts for this thesis; Ola Hansson, Johan Holmkvist, Peter Almgren, Peter Nilsson, Allan Vaag, Pernille Poulsen, Bo Isomaa, Tiinamaija Tuomi, Jie Wen and Renming Hu. A special thanks to the team in Oxford, Cecilia Lindgren, Joshua Randall and Blanca Herrera, for true inspiration and sharing of knowledge in our great, long-lasting project.

Life wouldn't be complete without good friends outside of research. Thanks to the always growing "dinner group" for lovely dinners, cakes and time to talk about both science and life; Anneli & Calle, Sidinh & John, Susanne & Pontus, Katti & Thomas, Maria & Peter, Åsa & Andreas, Calle & Maja, Matilda & Anton, Marta and babies….

I would like to thank my beloved family for invaluable support and for always having faith in me: Mum, Dad, Marie and Magnus, and my favourite nephews Bastian and Jasper. Your generosity and love give me strength every day. Stefan, my dearest, thanks for giving me strength to believe in myself, never letting my work become more important than training, and keeping me down to earth this last year.

Finally I am grateful for all of you, those that I mentioned above as well as the ones I forgot, who have been part of making my time as a PhD student the best possible!

LIST OF REFERENCES

1. Alberti, K.G. and P.Z. Zimmet, *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, 1998. 15(7): p. 539-53.
2. The International Diabetes Federation (IDF), *IDF Diabetes Atlas*. 4th edition, 2009. URL: http://www.diabetesatlas.org/.
3. Yoon, K.H., et al., *Epidemic obesity and type 2 diabetes in Asia.* Lancet, 2006. 368(9548): p. 1681-8.
4. Del Prato, S., et al., *Beta-cell mass plasticity in type 2 diabetes.* Diabetes Obes Metab, 2004. 6(5): p. 319-31.
5. UK Prospective Diabetes Study (UKPDS) Group, *Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33).* Lancet, 1998. 352(9131): p. 837-53.
6. Boyle, P.J., *Diabetes mellitus and macrovascular disease: mechanisms and mediators.* Am J Med, 2007. 120(9 Suppl 2): p. S12-7.
7. Kahn, S.E., R.L. Hull, and K.M. Utzschneider, *Mechanisms linking obesity to insulin resistance and type 2 diabetes.* Nature, 2006. 444(7121): p. 840-6.
8. Bugianesi, E., A.J. McCullough, and G. Marchesini, *Insulin resistance: a metabolic pathway to chronic liver disease.* Hepatology, 2005. 42(5): p. 987-1000.
9. Groop, L.C., et al., *The role of free fatty acid metabolism in the pathogenesis of insulin resistance in obesity and noninsulin-dependent diabetes mellitus.* J Clin Endocrinol Metab, 1991. 72(1): p. 96-107.
10. Ahima, R.S., *Metabolic actions of adipocyte hormones: focus on adiponectin.* Obesity (Silver Spring), 2006. 14 Suppl 1: p. 9S-15S.
11. Bastard, J.P., et al., *Recent advances in the relationship between obesity, inflammation, and insulin resistance.* Eur Cytokine Netw, 2006. 17(1): p. 4-12.
12. Johannsen, D.L. and E. Ravussin, *The role of mitochondria in health and disease.* Curr Opin Pharmacol, 2009. 9(6): p. 780-6.
13. Petersen, K.F., et al., *Mitochondrial dysfunction in the elderly: possible role in insulin resistance.* Science, 2003. 300(5622): p. 1140-2.
14. Stark, R. and M. Roden, *ESCI Award 2006. Mitochondrial function and endocrine diseases.* Eur J Clin Invest, 2007. 37(4): p. 236-48.
15. Kelley, D.E., et al., *Dysfunction of mitochondria in human skeletal muscle in type 2 diabetes.* Diabetes, 2002. 51(10): p. 2944-50.
16. Huang, X., et al., *Insulin-regulated mitochondrial gene expression is associated with glucose flux in human skeletal muscle.* Diabetes, 1999. 48(8): p. 1508-14.

17. Mootha, V.K., et al., *PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes.* Nat Genet, 2003. 34(3): p. 267-73.
18. 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, 2003. 100(14): p. 8466-71.
19. Sreekumar, R., et al., *Gene expression profile in skeletal muscle of type 2 diabetes and the effect of insulin treatment.* Diabetes, 2002. 51(6): p. 1913-20.
20. Ling, C., et al., *Multiple environmental and genetic factors influence skeletal muscle PGC-1alpha and PGC-1beta gene expression in twins.* J Clin Invest, 2004. 114(10): p. 1518-26.
21. Ling, C., et al., *Genetic and epigenetic factors are associated with expression of respiratory chain component NDUFB6 in human skeletal muscle.* J Clin Invest, 2007. 117(11): p. 3427-35.
22. International Human Genome Sequencing Consortium, *Finishing the euchromatic sequence of the human genome.* Nature, 2004. 431(7011): p. 931-45.
23. Birney, E., et al., *Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project.* Nature, 2007. 447(7146): p. 799-816.
24. Frazer, K.A., et al., *A second generation human haplotype map of over 3.1 million SNPs.* Nature, 2007. 449(7164): p. 851-61.
25. Kaprio, J., et al., *Concordance for type 1 (insulin-dependent) and type 2 (non-insulin-dependent) diabetes mellitus in a population-based cohort of twins in Finland.* Diabetologia, 1992. 35(11): p. 1060-7.
26. Newman, B., et al., *Concordance for type 2 (non-insulin-dependent) diabetes mellitus in male twins.* Diabetologia, 1987. 30(10): p. 763-8.
27. Köbberling, J. and H. Tillil, *Empirical risk figures for first-degree relatives of non-insulin dependent diabetics*, in *The genetics of diabetes mellitus*. 1982. p. 201-209.
28. McCarthy, M.I. and E. Zeggini, *Genome-wide association studies in type 2 diabetes.* Curr Diab Rep, 2009. 9(2): p. 164-71.
29. Grant, S.F., et al., *Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes.* Nat Genet, 2006. 38(3): p. 320-3.
30. Saxena, R., et al., *Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels.* Science, 2007. 316(5829): p. 1331-6.
31. Scott, L.J., et al., *A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants.* Science, 2007. 316(5829): p. 1341-5.
32. Sladek, R., et al., *A genome-wide association study identifies novel risk loci for type 2 diabetes.* Nature, 2007. 445(7130): p. 881-5.
33. Zeggini, E., et al., *Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes.* Science, 2007. 316(5829): p. 1336-41.
34. Dupuis, J., et al., *New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk.* Nat Genet, 2010. 42(2): p. 105-16.

35. 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, 2008. 40(5): p. 638-45.
36. Frayling, T.M., et al., *A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity.* Science, 2007. 316(5826): p. 889-94.
37. Bouatia-Naji, N., et al., *A variant near MTNR1B is associated with increased fasting plasma glucose levels and type 2 diabetes risk.* Nat Genet, 2008.
38. Lyssenko, V., et al., *Common variant in MTNR1B associated with increased risk of type 2 diabetes and impaired early insulin secretion.* Nat Genet, 2008.
39. Prokopenko, I., et al., *Variants in MTNR1B influence fasting glucose levels.* Nat Genet, 2008.
40. Hales, C.N. and D.J. Barker, *Type 2 (non-insulin-dependent) diabetes mellitus: the thrifty phenotype hypothesis.* Diabetologia, 1992. 35(7): p. 595-601.
41. Lindsay, R.S. and P.H. Bennett, *Type 2 diabetes, the thrifty phenotype - an overview.* Br Med Bull, 2001. 60: p. 21-32.
42. Barker, D.J., *The developmental origins of insulin resistance.* Horm Res, 2005. 64 Suppl 3: p. 2-7.
43. Wren, J.D. and H.R. Garner, *Data-mining analysis suggests an epigenetic pathogenesis for type 2 diabetes.* J Biomed Biotechnol, 2005. 2005(2): p. 104-12.
44. Ling, C. and L. Groop, *Epigenetics: a molecular link between environmental factors and type 2 diabetes.* Diabetes, 2009. 58(12): p. 2718-25.
45. Poirier, L.A., et al., *Blood S-adenosylmethionine concentrations and lymphocyte methylenetetrahydrofolate reductase activity in diabetes mellitus and diabetic nephropathy.* Metabolism, 2001. 50(9): p. 1014-8.
46. Ooi, S.K., A.H. O'Donnell, and T.H. Bestor, *Mammalian cytosine methylation at a glance.* J Cell Sci, 2009. 122(Pt 16): p. 2787-91.
47. Bird, A., *DNA methylation patterns and epigenetic memory.* Genes Dev, 2002. 16(1): p. 6-21.
48. Choy, J.S., et al., *DNA Methylation Increases Nucleosome Compaction and Rigidity.* J Am Chem Soc, 2010.
49. Poulsen, P., et al., *Age-dependent impact of zygosity and birth weight on insulin secretion and insulin action in twins.* Diabetologia, 2002. 45(12): p. 1649-57.
50. Groop, L., et al., *Metabolic consequences of a family history of NIDDM (the Botnia study): evidence for sex-specific parental effects.* Diabetes, 1996. 45(11): p. 1585-93.
51. Lindholm, E., et al., *Classifying diabetes according to the new WHO clinical stages.* Eur J Epidemiol, 2001. 17(11): p. 983-9.
52. Leosdottir, M., et al., *The association between total energy intake and early mortality: data from the Malmo Diet and Cancer Study.* J Intern Med, 2004. 256(6): p. 499-509.
53. Pyykkonen, A.J., et al., *Stressful life events and the metabolic syndrome: the prevalence, prediction and prevention of diabetes (PPP)-Botnia Study.* Diabetes Care. 33(2): p. 378-84.

54. Phillips, D.I., et al., *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, 1994. 11(3): p. 286-92.
55. DeFronzo, R.A., J.D. Tobin, and R. Andres, *Glucose clamp technique: a method for quantifying insulin secretion and resistance.* Am J Physiol, 1979. 237(3): p. E214-23.
56. The International HapMap Consortium, *The International HapMap Project.* Nature, 2003. 426(6968): p. 789-96.
57. de Bakker, P.I., et al., *Efficiency and power in genetic association studies.* Nat Genet, 2005. 37(11): p. 1217-23.
58. Jurinke, C., P. Oeth, and D. van den Boom, *MALDI-TOF mass spectrometry: a versatile tool for high-performance DNA analysis.* Mol Biotechnol, 2004. 26(2): p. 147-64.
59. Oeth, P., et al., *iPLEX Assay: Increased Plexing Efficiency and Flexibility for MassARRAY System Through Single Base Primer Extension with Mass-Modified Terminators.* SEQUENOM Application note, 2007. Doc. No. 8876-006, R05.
60. Applied Biosystems, *Allelic Discrimination Assay Getting Started Guide for the 7900HT Fast System.* Applied Biosystems, 2007. Part Number 4364015 Rev. B.
61. Livak, K.J., *SNP genotyping by the 5'-nuclease reaction.* Methods Mol Biol, 2003. 212: p. 129-47.
62. Li, L.C. and R. Dahiya, *MethPrimer: designing primers for methylation PCRs.* Bioinformatics, 2002. 18(11): p. 1427-31.
63. Bock, C., et al., *BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing.* Bioinformatics, 2005. 21(21): p. 4067-8.
64. Lewin, J., et al., *Quantitative DNA methylation analysis based on four-dye trace data from direct sequencing of PCR amplificates.* Bioinformatics, 2004. 20(17): p. 3005-12.
65. Ehrich, M., D. Correll, and D. van den Boom, *Introduction to EpiTYPER for quantitative DNA methylation analysis using the MassARRAY system.* SEQUENOM Application note, 2006. Doc. No. 8876-007, R02.
66. Ehrich, M., et al., *Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry.* Proc Natl Acad Sci U S A, 2005. 102(44): p. 15785-90.
67. Heid, C.A., et al., *Real time quantitative PCR.* Genome Res, 1996. 6(10): p. 986-94.
68. Applied Biosystems, *Absolute Quantitation Using Standard Curve Getting Started Guide.* Applied Biosystems, 2007. Part Number 4364014 Rev. C.
69. Affymetrix, *Whole-transcript Expression Analysis.* Affymetrix Application note, 2007. P/N 702503-2.
70. Irizarry, R.A., et al., *Exploration, normalization, and summaries of high density oligonucleotide array probe level data.* Biostatistics, 2003. 4(2): p. 249-64.

71. Liang, K.Y. and S.L. Zeger, *Regression analysis for correlated data.* Annu Rev Public Health, 1993. 14: p. 43-68.
72. Zeger, S.L. and K.Y. Liang, *Longitudinal data analysis for discrete and continuous outcomes.* Biometrics, 1986. 42(1): p. 121-30.
73. Storgaard, H., et al., *Genetic and nongenetic determinants of skeletal muscle glucose transporter 4 messenger ribonucleic acid levels and insulin action in twins.* J Clin Endocrinol Metab, 2006. 91(2): p. 702-8.
74. Siegel, S. and N.J.J. Castellan, *Nonparametric statistics for the behavioral sciences*. 2nd edition, 1988: McGraw-Hill International Editions. 399 pages.
75. R Development Core Team, *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. URL: http://www.R-project.org.
76. Chalaya, T.V., et al., *Tissue specificity of methylation of cytosines in regulatory regions of four genes located in the locus FXYD5-COX7A1 of human chromosome 19: correlation with their expression level.* Biochemistry (Mosc), 2006. 71(3): p. 294-9.
77. Dixon, A.L., et al., *A genome-wide association study of global gene expression.* Nat Genet, 2007. 39(10): p. 1202-7.
78. Trounce, I., E. Byrne, and S. Marzuki, *Decline in skeletal muscle mitochondrial respiratory chain function: possible factor in ageing.* Lancet, 1989. 1(8639): p. 637-9.
79. Zahn, J.M., et al., *Transcriptional profiling of aging in human muscle reveals a common aging signature.* PLoS Genet, 2006. 2(7): p. e115.
80. Fraga, M.F., et al., *Epigenetic differences arise during the lifetime of monozygotic twins.* Proc Natl Acad Sci U S A, 2005. 102(30): p. 10604-9.
81. Jiang, M.H., et al., *Hypermethylation of hepatic Gck promoter in ageing rats contributes to diabetogenic potential.* Diabetologia, 2008. 51(8): p. 1525-33.
82. Fraga, M.F., *Genetic and epigenetic regulation of aging.* Curr Opin Immunol, 2009. 21(4): p. 446-53.
83. Kuroda, A., et al., *Insulin gene expression is regulated by DNA methylation.* PLoS One, 2009. 4(9): p. e6953.
84. Ling, C., et al., *Epigenetic regulation of PPARGC1A in human type 2 diabetic islets and effect on insulin secretion.* Diabetologia, 2008. 51(4): p. 615-22.
85. Park, J.H., et al., *Development of type 2 diabetes following intrauterine growth retardation in rats is associated with progressive epigenetic silencing of Pdx1.* J Clin Invest, 2008. 118(6): p. 2316-24.
86. Eckhardt, F., et al., *Future potential of the Human Epigenome Project.* Expert Rev Mol Diagn, 2004. 4(5): p. 609-18.
87. Mackay, D.J., et al., *Hypomethylation of multiple imprinted loci in individuals with transient neonatal diabetes is associated with mutations in ZFP57.* Nat Genet, 2008. 40(8): p. 949-51.
88. Feinberg, A.P. and R.A. Irizarry, *Evolution in Health and Medicine Sackler Colloquium: Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease.* Proc Natl Acad Sci U S A, 2010.

89. Schalkwyk, L.C., et al., *Allelic skewing of DNA methylation is widespread across the genome.* Am J Hum Genet. 86(2): p. 196-212.
90. Eckhardt, F., et al., *DNA methylation profiling of human chromosomes 6, 20 and 22.* Nat Genet, 2006. 38(12): p. 1378-85.
91. Flanagan, S.E., et al., *Update of mutations in the genes encoding the pancreatic beta-cell K(ATP) channel subunits Kir6.2 (KCNJ11) and sulfonylurea receptor 1 (ABCC8) in diabetes mellitus and hyperinsulinism.* Hum Mutat, 2009. 30(2): p. 170-80.
92. Barnett, J.H. and J.W. Smoller, *The genetics of bipolar disorder.* Neuroscience, 2009. 164(1): p. 331-43.
93. Nuinoon, M., et al., *A genome-wide association identified the common genetic variants influence disease severity in beta(0)-thalassemia/hemoglobin E.* Hum Genet, 2009.
94. Weissman, B. and K.E. Knudsen, *Hijacking the chromatin remodeling machinery: impact of SWI/SNF perturbations in cancer.* Cancer Res, 2009. 69(21): p. 8223-30.
95. Yang, X.C., et al., *Characterization of 3'hExo, a 3' exonuclease specifically interacting with the 3' end of histone mRNA.* J Biol Chem, 2006. 281(41): p. 30447-54.
96. Savino, T.M., et al., *The nucleolar antigen Nop52, the human homologue of the yeast ribosomal RNA processing RRP1, is recruited at late stages of nucleologenesis.* J Cell Sci, 1999. 112 (Pt 12): p. 1889-900.
97. Huen, M.S., S.M. Sy, and J. Chen, *BRCA1 and its toolbox for the maintenance of genome integrity.* Nat Rev Mol Cell Biol. 11(2): p. 138-48.
98. Harper, M., et al., *Phosphorylation-dependent binding of human transcription factor MOK2 to lamin A/C.* FEBS J, 2009. 276(11): p. 3137-47.
99. Ishihara, K. and H. Sasaki, *An evolutionarily conserved putative insulator element near the 3' boundary of the imprinted Igf2/H19 domain.* Hum Mol Genet, 2002. 11(14): p. 1627-36.
100. Kaushik, N., et al., *Gene expression in peripheral blood mononuclear cells from patients with chronic fatigue syndrome.* J Clin Pathol, 2005. 58(8): p. 826-32.