THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Systems Biology of Type 2 Diabetes in Skeletal Muscle

LEIF VÄREMO

Department of Biology and Biological Engineering

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2016

Systems Biology of Type 2 Diabetes in Skeletal Muscle
Leif Väremo
Gothenburg, Sweden 2016
ISBN 978-91-7597-322-7

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 4003
ISSN 0346-718X

Department of Biology and Biological Engineering
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone + 46 (0)31-772 1000

Cover: An illustration summarizing the topic of the thesis. Skeletal myocyte transcriptome and proteome data was used to reconstruct the myocyte metabolic network. Transcriptome data from controls and subjects with type 2 diabetes were analyzed using gene-set analysis and integrated with the metabolic network, shown in the background, to identify metabolite hubs and subnetworks implicated in the disease.

Printed by Chalmers Reproservice
Gothenburg, Sweden 2016

Systems Biology of Type 2 Diabetes in Skeletal Muscle

LEIF VÄREMO

Department of Biology and Biological Engineering

CHALMERS UNIVERSITY OF TECHNOLOGY

Abstract

Type 2 diabetes (T2D) is a heterogeneous and complex disease that currently affects more than 350 million people worldwide. A wide range of risk factors influence the pathogenesis of T2D, including genetic and epigenetic components, as well as controllable factors such as diet, obesity, and sedentary lifestyle. T2D is characterized by abnormally high blood glucose levels as a consequence of the development of insulin resistance in multiple tissues (primarily skeletal muscle, liver, and adipose tissue) in combination with impaired insulin secretion in the pancreas. Skeletal muscle accounts for around 75-80% of the insulin-stimulated glucose uptake from the blood. Consequently, deficiency in glucose uptake mediated by insulin resistance in skeletal myocytes is an important factor for the disrupted glucose homeostasis associated with T2D. In fact, skeletal muscle insulin resistance can appear long before the onset of the disease itself, making it one of the primary defects preceding the development of T2D. The pathophysiology of T2D and the mechanisms underlying the development of insulin resistance in skeletal muscle are not yet fully understood. In light of the multifactorial complexity of T2D we have adopted a systems biology approach to study skeletal muscle in response to this disease, using network modeling of metabolism and analysis of genome-wide data from human subjects.

We developed three tools for analyzing gene expression data and facilitating its interpretation. The R package piano enables functional characterization and interpretation of gene expression profiles (and other omics data), through so called gene-set analysis (GSA). The skeletal myocyte genome-scale metabolic model (GEM), that we reconstructed based on transcriptome and proteome data, constitutes a comprehensive map of the myocyte metabolic network that can be used for simulation and integration of genome-wide data. The Python tool Kiwi visualizes the output from GSA using metabolite gene-sets and the topology of a GEM so that significant metabolite subnetworks affected by gene expression changes can be identified.

Leveraged by these tools, we performed two studies of T2D. In the first study, we carried out a meta-analysis of muscle tissue transcriptome data from 6 published datasets, providing a holistic insight into the metabolic state of T2D muscle. In particular, we identified a metabolic signature that has the power to predict T2D in individual subjects, highlighting its potential use for biomarkers or drug targets. In the second study, we analyzed transcriptome data from primary differentiated myocytes to explore inherent properties associated with T2D and obesity. We found a remarkable similarity between the transcriptional profiles in response to T2D and obesity, independent of each other, and identified a possible epigenetic mechanism behind these patterns. We performed a systematic characterization of the individual intrinsic effects of T2D and obesity, which are hardwired in the myocytes rather than attributable to a diabetic or obese extracellular environment. In summary, this thesis provides novel methods for analysis of genome-wide data and contributes to disentangling the complexity of T2D.

Keywords: skeletal muscle, myocyte, type 2 diabetes, obesity, metabolism, transcriptomics, gene expression, gene-set analysis, network analysis, genome-scale metabolic model

List of publications

This thesis is based on the work contained in the following publications:

I. **Väremo**, Nielsen and Nookaew. (2013) Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. *Nucleic Acids Research* 41:8

II. **Väremo**, Nookaew and Nielsen. (2013) Novel insights into obesity and diabetes through genome-scale metabolic modeling. *Frontiers in Physiology* 4:92

III. **Väremo**, Scheele, Broholm, Mardinoglu, Kampf, Asplund, Nookaew, Uhlén, Pedersen and Nielsen. (2015) Proteome- and transcriptome-driven reconstruction of the human myocyte metabolic network and its use for identification of markers for diabetes. *Cell Reports* 11:6

IV. **Väremo***, Gatto* and Nielsen. (2014) Kiwi: a tool for integration and visualization of network topology and gene-set analysis. *BMC Bioinformatics* 15:408

V. **Väremo** and Nielsen. (2015) Networking in metabolism and human disease. *Oncotarget* 6:18

VI. **Väremo**, Scheele, Broholm, Pedersen, Uhlén, Pedersen and Nielsen. (2016) Type 2 diabetes and obesity are independently associated with a similar inherent transcriptional profile in human myocytes. *Manuscript*

Additional publications not included in this thesis:

VII. Garcia-Albornoz*, Thankaswamy-Kosalai*, Nilsson, **Väremo**, Nookaew and Nielsen. (2014) BioMet Toolbox 2.0: Genome-wide analysis of metabolism and omics data. *Nucleic Acids Research* 42:W1

VIII. Aspuria, Lunt*, **Väremo***, Vergnes, Gozo, Beach, Salumbides, Reue, Wiedemeyer, Nielsen, Karlan, Orsulic and Lunt. (2014) Succinate dehydrogenase inhibition leads to epithelial-mesenchymal transition and reprogrammed carbon metabolism. *Cancer and Metabolism* 2:21

IX. Lindskog, Linné, Fagerberg, Hallström, Sundberg, Lindholm, Huss, Kampf, Choi, Liem, Ping, **Väremo**, Mardinoglu, Nielsen, Larsson, Pontén and Uhlén. (2015) The human cardiac and skeletal muscle proteomes defined by transcriptomics and antibody-based profiling. *BMC Genomics* 16:1

X. Svahn*, **Väremo***, Gabrielsson, Peris, Nookaew, Grahnemo, Sandberg, Wernstedt Asterholm, Jansson, Nielsen and Johansson. (2016) The impact of dietary fatty acids composition on the transcriptomes of six tissues reveals specific regulation of immune related genes. *Submitted manuscript*

*Authors contributed equally to this work.

Contribution summary

I. Designed the study, performed the analysis, designed and created the software, and wrote the paper.
II. Carried out the literature review, and wrote the paper.
III. Contributed to the design of the study, performed the statistical and bioinformatics analysis, participated in the reconstruction of the model, and wrote the paper.
IV. Participated in the design and creation of the software, and wrote the paper.
V. Wrote the paper.
VI. Contributed to the design of the study, performed the statistical and bioinformatics analysis, and wrote the paper.
VII. Contributed to the design of the study, and read and edited the paper.
VIII. Performed bioinformatics analysis, and contributed to the writing and editing of the paper.
IX. Contributed to the writing and editing of the paper.
X. Performed the bioinformatics analysis, and contributed to the writing and editing of the paper.

Preface

This dissertation is submitted for the partial fulfillment of the degree of doctor of philosophy. It is based on work carried out between September 2011 and February 2016 in the Systems and Synthetic Biology group, Department of Biology and Biological Engineering, Chalmers University of Technology, under the supervision of Professor Jens Nielsen. The research was funded by the Knut and Alice Wallenberg Foundation, the Chalmers Foundation, and the Bill & Melinda Gates Foundation.

Leif Väremo

January 2016

Table of contents

Abbreviations

AGE	Advanced glycation end-products
AUC	Area under the ROC curve
BCAA	Branched-chain amino acid
BMI	Body mass index
ChIP	Chromatin immunoprecipitation
CVD	Cardiovascular disease
ECM	Extracellular matrix
eQTL	Expressed quantitative trait loci
ER	Endoplasmic reticulum
FPKM	Fragments per kilobase of transcript per million mapped reads
GEM	Genome-scale metabolic model
GLUT	Glucose transporter
GO	Gene Ontology
GSA	Gene-set analysis
GWAS	Genome-wide association study
HMR	Human Metabolic Reaction database
HOMA	Homeostasis model assessment
HPA	Human protein atlas
IFG	Impaired fasting glucose
IGT	Impaired glucose tolerance/tolerant
IL	Interleukin
IRS	Insulin receptor substrate
KEGG	Kyoto Encyclopedia of Genes and Genomes
NGS	Next-generation sequencing
NGT	Normal glucose tolerance/tolerant
OGTT	Oral glucose tolerance test
ORA	Overrepresentation analysis
PCA	Principle component analysis
PI3K	Phosphatidylinositol 3-kinase
PKC	Protein kinase C
ROC	Receiver operator characteristic
SNP	Single nucleotide polymorphism
SPL	Shortest path length
T2D	Type 2 diabetes/diabetic
TCA	Tricarboxylic acid
TNF	Tumor necrosis factor

To Isa,
for always being by my side
throughout this journey

Introduction

The human genome contains just above 20,000 coding genes, according to current statistics from Ensembl. In the cell, these genes are transcribed and translated into proteins, under the control of advanced regulatory mechanisms. The proteins have a vast range of functions serving many of the cell's needs. The protein landscape of a given cell at a given time will depend on numerous factors, including cell type, signals and interactions from nearby cells and hormones, availability of various nutrients and molecules, and other environmental factors. Genetic and epigenetic factors, such as mutations, DNA methylation, and chromatin modifications, also influence gene expression (Lee and Young, 2013). Knowledge about what genes a cell is expressing at a given time, and at what levels, can give a clue of the current state of the cell. With established methods, such as microarrays, and the rapid development of next-generation sequencing technology, researchers are now able to simultaneously quantify the transcription of virtually all genes. This field of research is referred to as transcriptomics.

Transcriptomics has enabled a holistic and unbiased way to study human disease, in terms of finding disease-associated gene expression differences, without having to specify beforehand what genes to study. Unfortunately, many diseases are complex and do not exhibit changes in only a few genes. Instead, there often is a combination of distinct and subtle changes across many genes that act together either as a cause or a consequence of the disease (Lee and Young, 2013). Detecting and quantifying these gene expression changes is relatively simple. However, to figure out the impact of these changes on the function of the cell, or what these changes imply about the cell state, is not trivial. This challenge, of how to interpret the meaning of numerous differentially expressed genes, has intrigued me since I first got introduced to systems biology.

During my doctoral studies I have pursued this challenge by using and advancing a methodology called gene-set analysis (GSA). We have applied GSA as a basis to interpret transcriptomic data with a particular interest in understanding how type 2 diabetes (T2D) is affecting skeletal muscle. T2D is a metabolic disorder and, as it turns out, almost one fifth of the 20,000 coding genes are associated with metabolism, according the Human Metabolic Reaction (HMR) database (Mardinoglu et al., 2014). Metabolism constitutes a network of reactions, metabolites and enzymes, and can as such be used as a scaffold for integrating gene expression data. In our goal to gain insight into the function of skeletal muscle under the influence of T2D, we have reconstructed a metabolic network model for muscle cells and used this in connection with transcriptome data and GSA to dissect the meaning of gene expression changes, with a particular focus on metabolism.

In particular, the aim with this thesis was:

- To develop and provide software tools that improve the contextual and functional interpretation and characterization of omics data, in general, and transcriptomic data, in particular.
- To reconstruct the skeletal myocyte metabolic reaction network in form of a comprehensive genome-scale metabolic model, to provide a scaffold for omics data integration in the context of muscle metabolism.
- To use systems biology and bioinformatics analysis to gain new insight into the function and dysfunction of muscle cells, with a focus on metabolism, in response to obesity and T2D.

This thesis summarizes the main part of the scientific research that I have contributed to during my doctoral studies, represented by four original research papers, one review paper, and one editorial (all listed on page V and included in full-text in the end of the thesis). **Paper I** describes a software tool, piano, for GSA. **Paper II** gives a review of the use of genome-scale metabolic models to study obesity and diabetes. **Paper III** describes a genome-scale metabolic model for skeletal muscle cells and its application to study T2D. **Paper IV** describes a software tool, Kiwi, which visualizes the output of GSA using genome-scale metabolic models. **Paper V** is a summary and outlook of **Paper III** in form of an editorial. **Paper VI** describes a study of inherent properties of muscle cells in association with obesity and T2D.

Structure of the thesis

Following this introduction there is a background section covering T2D and presenting the holistic framework that we have used to study this disease, in the field of systems biology. The main body of this thesis is divided into two parts and more specific background relevant to these sections will be given there. Part I (based on results from **Papers I**-**IV**) introduces GSA and genome-scale metabolic models, and describes and discusses two software tools and a network model of metabolism that address the difficulties of condensing high-throughput omics data (in particular transcriptomics) into interpretable results. Part II (based on results from **Papers III**, **V** and **VI**) turns to the specific topic of studying the impact of T2D on skeletal muscle and describes how we applied the tools presented in Part I, along with other methods, to add insight into the effects that T2D has on skeletal muscle cells. Lastly, there are some concluding remarks and future perspectives.

Background

An overview of type 2 diabetes

A very brief history of diabetes

The oldest known reference to diabetes comes from an Egyptian medical papyrus (Ebers Papyrus) that dates back to around 3,500 years from now, where a condition is mentioned and described as *"too great emptying of the urine"* (Zajac et al., 2010). Meanwhile, in India they gave this condition the name *"honey urine"*, noticing that ants were attracted by its sweetness. Later, around 2,000 years ago, the Greek coined the term diabetes, meaning *"to pass through"*. In 1776 it was discovered that the sweet substance was sugar (Dobson and Fothergill, 1776) and in 1815 it was shown to be glucose (Chevreul, 1815). In the early 1920's Banting, Best, Collip and colleagues discovered the hormone insulin (which we now know is responsible for, among other things, the clearance of glucose from the blood), which they extracted and purified from the pancreas. This extract was used on diabetic dogs and, in 1922, the first human patient was successfully treated (Banting et al., 1922). After an agreement with the University of Toronto, Eli Lilly and Company received manufacturing rights which led to the first commercial large-scale production of insulin (Zajac et al., 2010). In 1922, the Danish researcher August Krogh, who the year before was awarded the Nobel Prize in Physiology or Medicine for his work on capillaries, was traveling overseas together with his wife Marie, who coincidently had just been diagnosed with diabetes. During her husband's lectures in the United States, Marie Krogh was informed about Banting's discoveries and they decided to visit the group. Before their return to Denmark, Krogh was given a license for production of insulin. This resulted in the establishment of Nordisk Insulinlaboratorium in Denmark, which has evolved to the current company Novo Nordisk, the world's largest insulin producer. In 1923, Banting and laboratory director Macleod were awarded the Nobel Prize in Physiology or Medicine for their findings. Controversially however, there have been claims that Paulescu, who during the same time was treating dogs with insulin extract, should have been included in the prize (Murray, 1971). After these discoveries diabetes has continued to be immensely studied, which has resulted in a deeper understanding of the molecular basis of the disease as well as in improvements of treatment strategies. However, there is still much more to learn before we fully understand this disease. In the next few sections I will give a brief overview of the state of diabetes today and our current understanding of its pathophysiology.

Different types of diabetes

Diabetes (formally diabetes mellitus, to be distinguished from diabetes insipidus) refers to a group of syndromes characterized by hyperglycemia, i.e. abnormally high blood glucose levels. There are several less common forms of diabetes (e.g. gestational diabetes) but the most common variants are type 1 and type 2, where the latter stands for the vast majority (roughly 90%) of the cases (Scully, 2012). Type 1 diabetes is primarily an autoimmune disease and results from the pancreas losing the ability to produce insulin (thus leading to high glucose concentration in the blood stream). Individuals with type 1 diabetes rely on treatment with external insulin. On the other hand, T2D, which is the form of the disease that we are focusing on in our research, involves reduced sensitivity to insulin in different tissues, and eventually compromised insulin production in the pancreas. T2D is associated with obesity and a sedentary lifestyle. Individuals with T2D do not necessarily rely on external insulin supply, depending on the extent of insulin resistance and pancreas dysfunction.

Type 2 diabetes epidemiology

T2D has been described as a global epidemic (Zimmet et al., 2001). The numbers certainly support this statement. It has been estimated that around 350 million people worldwide are suffering from diabetes (Scully, 2012) and, according to the World Health Organization (WHO), 9% of adults had the disease in 2014 (WHO, 2014b). These numbers are expected to rise, and it is projected that in 2030 440-550 million people will have diabetes, equivalent to at least a 50% increase from 2010 (Shaw et al., 2010; Whiting et al., 2011).

The traditional view of T2D being considered a disease with onset in middle-aged adults is now challenged by the serious facts of increasing prevalence among young adults, adolescents and children (Pinhas-Hamiel and Zeitler, 2005; Rosenbloom et al., 1999). In the 1990's T2D represented only 3% of diabetes cases among children and adolescents, whereas in 2005 a striking increase to 45% was reported (Pinhas-Hamiel and Zeitler, 2005). The incidence varies by ethnicity however, and the highest frequencies are seen in minority populations, e.g. Native Americans, Pacific Islanders, Australian Indigenous populations, and African Americans (Chen et al., 2012; Craig et al., 2007; Dabelea et al., 2007). The risk of increasing onset of T2D among youth and a higher prevalence of people with chronic complications in the near future, may have a huge impact on the economy and public health, in terms of negative effects on work capacity and premature morbidity and mortality (Chen et al., 2012; Dowse et al., 1991; Zimmet et al., 2001).

The estimated number of deaths caused by diabetes varies with reports from 1.5 million worldwide (WHO, 2014a) to as many as 3.5 million in middle income countries only (Scully, 2012). In 2030, diabetes is predicted to be the 7th leading cause of death in the world and, ranked even higher, the 4th in high-income countries (Mathers and Loncar, 2006). Still, around 80% of deaths caused by diabetes occur in low- and middle-income countries (WHO, 2014a) and the mortality rate is more than double compared to high-income countries (Scully, 2012).

The diabetes epidemic is taking place all over the globe. Africa, in particular driven by countries in the northern parts, is predicted to have the fastest increase of diabetes cases in the world (Scully, 2012). In 2010 14.1 million people in Africa had the disease (Zimmet et al., 2001). Similar patterns are seen in South America and in the Middle East, with some extreme examples being Qatar, Saudi Arabia, and the United Arab Emirates, all with more than 20% of the population being diabetic (Scully, 2012; Whiting et al., 2011). Asia is currently considered the epicenter of diabetes (Hu, 2011). In China 92 million adults have diabetes (Yang et al., 2010), representing the largest national population of diabetics in the world, followed by India (Scully, 2012). In India, an increase in diabetes cases has been reported both in urbanized (13.9%-18.2% in 2000-2006) and rural (6.4%-9.2% in 2000-2006) areas (Ramachandran et al., 2008).

Given the high global incidence of diabetes, largely represented by the epidemic increase in T2D (Zimmet, 1999), and the predicted escalation of the number of diabetic people in the coming years, it is no big surprise that a fair amount of money is spent on health care costs related to this disease. It is estimated that 12% of the global health expenditure, equivalent to 376 billion US dollars, is used for treatment of diabetes (Zhang et al., 2010). There is however a huge difference in the amount spent per patient between countries where e.g. USA and Australia spend around 8,000 US dollars per patient, whereas China and India, with by far the highest absolute number of deaths per year caused by diabetes, only spend a couple of hundred US dollars per patient (Scully, 2012).

Diagnosis and classification

Different criteria have been used over the years to diagnose and classify diabetes. In 2006, the WHO and the International Diabetes Foundation (IDF) published updated guidelines for the definition and diagnosis of diabetes (WHO and IDF, 2006). They recommend that venous blood plasma glucose concentration measurements should be used for classification, at fasting state and two hours after an oral glucose tolerance test (OGTT). During an OGTT, the subject is given 75 g of glucose dissolved in water, and plasma glucose levels are measured at baseline and after two hours.

Table 1 summarizes the criteria for diabetes, impaired glucose tolerance (IGT), and impaired fasting glucose (IFG), based on these kinds of measurements. In brief, a subject with too high baseline glucose levels or with reduced ability to clear glucose from the blood is classified as diabetic.

This classification is used for both type 1 and type 2 diabetes. If the type is not clear from the circumstances, additional tests can be performed, e.g. using

Table 1. Diagnostic criteria for diabetes, IGT, and IFG, recommended by WHO and IDF (2006).

	Fasting plasma glucose		2 hour plasma glucose
Diabetes	≥ 7.0 mmol/L	or	≥ 11.1 mmol/L
Impaired glucose tolerance	< 7.0 mmol/L	and	≥ 7.8 and < 11.1 mmol/L
Impaired fasting glucose	≥ 6.1 and < 7.0 mmol/L	and	< 7.8 mmol/L

autoantibodies to detect destroyed and lost pancreatic islet cell typical for type 1 diabetes (Sacks et al., 2011) or measuring C-peptide levels to assess insulin secretion capability (Jones and Hattersley, 2013). C-peptide and insulin are produced in equimolar amounts but, unlike insulin, C-peptide is not metabolized by the liver. Further on, C-peptide enables assessment of pancreatic insulin secretion even in subjects on exogenous insulin.

An alternative measure that has recently started to be used for diabetes classification and diagnosis is glycated hemoglobin (HbA_{1c}) which reflects the average glycemia (blood glucose levels) during the previous 2-3 months (Nathan et al., 2007). The WHO recommends a cut point at 6.5% HbA_{1c} for diagnosing diabetes, but states that lower levels does not necessarily exclude diabetes (WHO, 2011). The choice of method used to classify diabetes has a considerable effect on prevalence estimations and will be important to take into account when considering longitudinal studies (Chen et al., 2012).

A technique which is often used in diabetes research is the glucose clamp (which does not describe a device but refers to clamping the levels of glucose or insulin at a fixed predetermined value). Two common versions are the hyperglycemic clamp and the euglycemic clamp (DeFronzo et al., 1979). During the hyperglycemic clamp the blood plasma glucose concentration is raised to a fixed high level and thereafter maintained by intravenous glucose infusion. The infusion rate needed to maintain the high glucose concentration will reflect how fast the body can clear high glucose levels from the blood (e.g. by insulin secretion and signaling). During the euglycemic clamp the blood plasma insulin concentration is raised and maintained at a high constant value. Meanwhile, glucose infusion is used to keep the blood plasma glucose concentrations at a baseline level. The glucose infusion rate reflects how fast the body clears glucose from the blood at a given high insulin concentration. These tests quantify how sensitive the pancreatic beta cells are to glucose and how sensitive tissues are to insulin (DeFronzo et al., 1979).

Another common measure used in research to assess beta cell function and insulin resistance is the homeostasis model assessment (HOMA) (Wallace et al., 2004). Unlike the glucose clamp technique, HOMA is based only on basal plasma glucose and insulin or C-peptide concentrations, thus more feasible for large cohort studies (Levy et al., 1998; Matthews et al., 1985). A nonlinear computer model, based on a simulation of factors influencing blood insulin and glucose concentrations, is used to estimate beta cell function and insulin sensitivity as percentages of normal young adults, given input values within clinically realistic ranges. There also exists a linear approximation to the first version of the computer model. Several studies have shown good (but far from perfect) correlation of insulin resistance and beta cell function estimates between HOMA and euglycemic or hyperglycemic clamps (Wallace et al., 2004).

Pathophysiology

As described in the previous section, the diagnosis of T2D is based on the notion of elevated plasma glucose levels and reduced capability to maintain glucose homeostasis after the acute increases associated with e.g. a meal. This effect is

tightly connected to impairments in both insulin secretion, i.e. pancreatic beta cell dysfunction, and insulin action in insulin sensitive tissues, i.e. insulin resistance. Normally, after a meal, the increase in circulating glucose stimulates the beta cells to secrete insulin into the blood stream. First, stored insulin is released, followed by a prolonged release of synthesized insulin. Circulating insulin is quickly degraded, resulting in a half-life of less than ten minutes (Tomasi et al., 1967). Insulin binds to the insulin receptor on cells in insulin sensitive tissues (primarily skeletal muscle, liver, and adipose) and induces a signaling cascade. The signaling cascade involves activation of the insulin receptor which leads to tyrosine phosphorylation of several different substrates, including insulin receptor substrates (IRSs). IRS activates the phosphatidylinositol 3-kinase (PI3K) pathway, involving stimulation of the protein kinases Akt and protein kinase C (PKC). The insulins signaling pathway eventually results in uptake of glucose, fatty acids, and amino acids, and expression of genes promoting glycogen, lipid, and protein synthesis (Saltiel and Kahn, 2001). Insulin signaling triggers the translocation of glucose transporter 4 (GLUT4) to the cell membrane, which leads to a rapid uptake of glucose by skeletal myocytes and adipocytes. In myocytes, excess glucose is stored as glycogen. In adipocytes insulin inhibits lipolysis and promotes an increased lipogenesis, so that glucose is converted into and stored as triglycerides. In liver, insulin stimulates glycogenesis and inhibits gluconeogenesis.

There is a feedback loop of unknown mechanism that enables crosstalk between the tissues responding to insulin action and the beta cells in the pancreas, so that insulin release can be continuously adjusted to serve the needs of the cells (Kahn et al., 2014). This means that the beta cells can compensate for impaired insulin action by increasing insulin secretion, thus maintaining glucose homeostasis even in the presence of insulin resistance. During the pathogenesis of T2D, blood glucose levels increase, indicating beta cell dysfunction and failure to fully compensate for the increased need of insulin secretion. Beta cell dysfunction is progressive during the development of T2D, ranging from a disability to clear acute postprandial blood glucose increases to resulting in permanently elevated glucose levels (Kohei, 2010). Beta cell function can be impaired already in non-diabetic subjects at high risk of developing T2D (Cnop et al., 2007; Dunaif and Finegood, 1996). There is also a reduction in the number of beta cells, through apoptosis, in T2D, but this does not on its own explain the reduced insulin secretion capacity (Butler et al., 2003; Kahn et al., 2014). The mechanism behind beta cell loss is unknown but may involve toxic effects from the high levels of glucose and free fatty acids that are seen in connection to T2D, damage related to the increased secretory demand, or negative effects from an increased inflammatory response (Meier and Bonadonna, 2013). There is likely a combination of loss of function and reduced beta cell mass in the impaired insulin secretion seen in T2D subjects.

Insulin resistance develops before and during T2D and blunts the normal effect of insulin, thus leading to reduced glucose uptake in muscle and adipose tissue, increased lipolysis in adipose, and increased gluconeogenesis in liver (Eckel et al., 2005; Kohei, 2010). Most single nucleotide polymorphisms (SNPs) related to T2D are targeting beta cell function, but several genetic variants have nevertheless been detected to have an association with insulin resistance and affect proteins in the

insulin signaling pathway (Billings and Florez, 2010; Kohei, 2010). There are several hypotheses about the mechanisms underlying the development of insulin resistance, including metabolic overload, endoplasmic reticulum (ER) stress, and inflammation. Associated with obesity and high-fat diet, metabolic overload involves the elevation of dietary nutrients and accumulation of lipids in non-adipose tissues. This, in combination with decreased fatty acid metabolism, can lead to increased intracellular levels, in liver and muscle, of diacylglycerol, fatty acyl-CoA, and ceramides that could cause mitochondrial stress and interfere with insulin signaling (Kahn et al., 2006; Muoio and Newgard, 2008; Shulman, 2000). Metabolic overload could also lead to an ER stress response that suppresses the insulin signaling pathway (Özcan et al., 2004). Adipocytes seem to play an important role in the development of insulin resistance by affecting other tissues through the secretion of several adipokines, including tumor necrosis factor alpha (TNF-alpha), resistin, interleukin 6 (IL-6), retinol binding protein 4, and free fatty acids (Kahn et al., 2006; Lin and Sun, 2010; Yang et al., 2005). On the other hand, the adipokines adiponectin and leptin are considered beneficial for T2D subjects (Eckel et al., 2005; Muoio and Newgard, 2008). Obesity is associated with an increased adipose infiltration of macrophages, which can also release proinflammatory cytokines like TNF-alpha and IL-6. These can act locally by reducing insulin signaling, but also lead to increased insulin resistance in muscle and liver (Kahn et al., 2014; Kahn et al., 2006). Specific knockout or inactivation of GLUT4 in adipose tissue has been shown to result in impaired insulin action in muscle and liver, providing evidence for tissue crosstalk in the development of insulin resistance (Abel et al., 2001).

Risk factors

The development of T2D is connected to a complex interaction of several different environmental and genetic risk factors. Examples of these are e.g. diet, smoking, excessive alcohol intake, aging, gender, ethnicity, and intrauterine environment (Chen et al., 2012; Doria et al., 2008; Hu, 2011). Other indicators of an increased risk of T2D include impaired glucose tolerance, abnormal blood lipid levels, hypertension, inflammation, and history of diseases such as gestational diabetes, polycystic ovary syndrome, or nonalcoholic fatty liver disease (Chen et al., 2012). However, one of the absolute main driving forces behind the increased global spread of T2D is overweight and obesity, which are the most important predictors of the development of the disease (Hu et al., 2001). In 2005 it was estimated that 23% of the world's adult population was overweight (BMI of 25-30) and 9.8% was obese (BMI ≥ 30), and considering the current trend, these numbers are predicted to drastically increase in the coming years (Kelly et al., 2008). Obesity is tightly connected to excessive caloric intake, diet quality, sedentary lifestyle, and decreased physical activity. As an example, a two hour per day increase in television watching was associated with a 14% increase of the risk of T2D, whereas if the same amount of time was spent on standing or walking around at home the risk decreased with 12% (Hu et al., 2003). Adding to the complexity is the fact that T2D can appear in non-obese subjects. This is in particular the case for many Asian populations, one example being India with a very low rate of obesity but a high

incidence of T2D (Yoon et al., 2006). It has been suggested that these discrepancies may be due to a higher percentage of body fat, in particular intra-abdominal fat, among Asians, thus enabling a so-called metabolically obese phenotype at lower BMI levels (Brunetti, 2007; Misra, 2003). The BMI of T2D patients is on average lower for subjects diagnosed at older ages (Hillier and Pedula, 2001), whereas intra-abdominal fat increases with age (Utzschneider et al., 2004). It is thus possible that the metabolically obese phenotype plays a role in older T2D patients with lower BMI levels (Brunetti, 2007; Goodpaster et al., 2003). It has also been suggested that different mechanisms affect the development of T2D in obese compared to non-obese subjects (Arner et al., 1991).

A low birth weight and poor nutrition during fetal and infant development can increase the risk of developing T2D later in life (Hales and Barker, 2001; Whincup et al., 2008). It is suggested that undernutrition promote changes that are beneficial for surviving starvation but are detrimental during exposure to normal food intake in adult life (Chen et al., 2012; Hu, 2011).

Genetic differences do also contribute to the risk of developing T2D. A history of T2D in a first-degree family was associated with a doubled risk of T2D in a Scandinavian cohort (Lyssenko et al., 2008) and a person with diabetic parents have a substantially higher risk of developing T2D compared to the general population (Leslie et al., 1986). Genome-wide association studies have enabled the association between genetic variants and susceptibility to T2D. Variants in around 50 genetic loci have been established to contribute to T2D, but do not improve the prediction of the disease compared to other common risk factors (Cho et al., 2012; Hu, 2011; Morris et al., 2012). It is likely that genetic susceptibility enhance the risk of T2D in combination with the presence of additional environmental risk factors (Hu, 2011). Epigenetic factors, including DNA methylation and histone modifications may also have an important impact on the development of T2D, but is still poorly understood (Ling and Groop, 2009).

There is also evidence for that alterations in the function and composition of the gut microbiome are associated with T2D and can be used for prediction and classification of the disease (Karlsson et al., 2013; Qin et al., 2012). It can however be challenging to isolate these effects from the effects on the microbiome associated with drug treatment (Forslund et al., 2015).

Complications

The negative effects of hyperglycemia on the vascular system represent the primary source of morbidity and mortality associated with diabetes (Fowler, 2008). Microvascular damage from high glucose levels can lead to chronic complications including retinopathy (eye damage, which can lead to blindness), nephropathy (impaired kidney function and failure), and neuropathy (which can lead to loss of sensation in the feet, with increased risk of ulcers and infections, and can ultimately require amputation). These complications are caused by damage to cells where glucose uptake is independent of insulin, including capillary endothelial cells, mesangial cells (lining capillaries in the kidney), and peripheral neurons and Schwann cells (supporting neurons) (Brownlee, 2005; Stolar, 2010).

The mechanisms underlying the development of these complications are not fully understood, but several theories exist, some of which will be briefly described here. High glucose levels can lead to formation of advanced glycation end-products (AGEs), e.g. non-enzymatic glycation of intracellular proteins involved in transcriptional regulation, modification of the extracellular matrix, and glycation of circulating proteins which can trigger an inflammatory response (Brownlee, 2005). High glucose levels can overload the polyol pathway resulting in increased sorbitol and fructose production through aldolase reductase, at the same time depleting the NADPH pool, which in turn can lead to decreased levels of reduced glutathione, thus making the cell more vulnerable to oxidative stress (Brownlee, 2005; Chung et al., 2003). In addition, intracellular accumulation of sorbitol and fructose can induce osmotic stress (Chung et al., 2003; Fowler, 2008). Oxidative stress can arise from increased flux through the tricarboxylic acid (TCA) cycle leading to electron buildup in the electron transport chain and superoxide production (Du et al., 2001; Nishikawa et al., 2000). Hyperglycemia can also lead to activation of protein kinase C (PKC), which can regulate gene expression and may indirectly be pro-inflammatory and have a deleterious effect on vascular function (Brownlee, 2001, 2005). Finally, increased flux through the hexosamine pathway, leading to the addition of *N*-acetylglucosamine to serine and threonine residues of transcription factors, may be linked to increased transcription associated with hyperglycemia (Brownlee, 2001). There is unifying evidence that oxidative stress, specifically through the increased production of superoxide, may be the underlying cause to PKC activation, AGE formation, and increased flux through aldolase reductase and the hexosamine pathway (Brownlee, 2001; Nishikawa et al., 2000).

Diabetes and hyperglycemia can also cause macrovascular damage by affecting larger blood vessels and the formation of atherosclerotic plaque (Fowler, 2008; Stolar, 2010). Around 50% of diabetic people die from cardiovascular disease (CVD), making it the primary cause of death (Laing et al., 2003; Morrish et al., 2001). It has been reported that the risk of myocardial infarction (MI) in diabetic people (without history of earlier MI) is comparable to that of non-diabetic people with a history of earlier MI (Haffner et al., 1998). The risk of having a stroke is more than doubled in people with T2D, independent of other known risk factors for CVD (Almdal et al., 2004). However, in contrast to this, there have been recent reports that intensive glycemic control does not have any beneficial effect on CVD, challenging the role of hyperglycemia in macrovascular damage (Duckworth et al., 2009; Gerstein et al., 2008; Patel et al., 2008).

Apart from the chronic complications mentioned earlier, diabetes can also cause a number of acute health problems. This includes hyperosmolar hyperglycemic non-ketotic syndrome, where high glucose concentrations can lead to severe dehydration through osmosis (Pasquel and Umpierrez, 2014). Hypoglycemia, too low glucose levels, may also occur, as a consequence of anti-diabetic medication (Yanai et al., 2015). Both of these conditions can lead to diabetic coma, with high mortality rate (Ben-Ami et al., 1999; Gill and Alberti, 1985; Pasquel and Umpierrez, 2014).

Prevention, management, and treatment

Prevention of T2D can be achieved by reducing controllable risk factors and adopting a healthy lifestyle. This includes maintaining a normal body weight, healthy diet, daily physical activity, avoiding tobacco, and moderate alcohol intake (Hu, 2011; Mozaffarian et al., 2009). Lifestyle interventions in high-risk subjects have been shown to prevent or delay T2D by 50%, being as effective as pharmacological treatment (Gillies et al., 2007).

Management of T2D revolves around reducing the risk of hyperglycemic complications by controlling blood glucose concentrations (Inzucchi et al., 2012; Kahn et al., 2014). This can be achieved by weight-loss through diet control and exercise, without or in combination with medication. There is no universal therapeutic strategy and it is important to individualize treatment (Inzucchi et al., 2012). A long list of drugs have been developed and are used to treat T2D, targeting classic organs like the pancreas, adipose tissue, muscle tissue, and liver, but also kidneys, brain, and the gastrointestinal tract (Kahn et al., 2014). Table 2 gives an overview of the different cellular mechanisms of the common classes of drugs used in diabetes treatment (although the exact mechanism is unknown for many drugs). Metformin is generally used as initial monotherapy, but can subsequently be combined with one or several other drugs and insulin (Inzucchi et al., 2012; Kahn et al., 2014; Qaseem et al., 2012).

Table 2. Different classes of drugs used for the treatment of T2D.

Class	Mechanism
Biguanides (metformin)	Lowers plasma glucose levels by reducing hepatic gluconeogenesis (glucose production) and opposing the action of glucagon (glucagon is a hormone with opposite effect of insulin). Metformin inhibits complex I in the mitochondrial respiratory chain, which reduces energy availability (increased ratios of ADP/ATP and AMP/ATP) eventually leading to reduced gluconeogenesis and lipid/cholesterol synthesis, partly through AMP-mediated signaling (Rena et al., 2013).
Sulfonylureas	Improves insulin secretion by binding to ATP-sensitive potassium channels in the cell membrane of pancreatic beta cells. This causes a membrane depolarization which opens voltage-gated calcium channels, increasing intracellular calcium concentrations which stimulates insulin secretion.
Meglitinides/glinides	Has a similar mode of action as sulfonylureas but with a weaker binding affinity to the potassium channels.
Thiazolidinediones	Activates the nuclear receptor peroxisome proliferator-activated receptor γ (PPAR-γ) which regulates transcription of several genes (involved in glucose and lipid metabolism), indirectly improving insulin sensitivity in adipose, skeletal muscle, and liver (Hauner, 2002).
α-Glucosidase inhibitors	Inhibits α-glucosidase in the small intestine, which slows down carbohydrate absorption, thus decreasing blood glucose concentrations.
DPP4 inhibitors	Dipeptidyl peptidase 4 (DPP4) normally inactivates the gut hormones glucagon-like peptide-1 (GPL-1) and gastric inhibitory polypeptide (GIP). DPP4 inhibitors thus promote GPL-1 and GIP action, which includes increased insulin and reduced glucagon secretion.
GLP-1 receptor agonists	Mimics the effect of GPL-1 (but has longer half-life) by binding to the GLP-1 receptor and stimulating insulin secretion and reducing glucagon release.
Dopamine-2 agonists	Activates dopamine receptors and thereby influencing central regulation of metabolism by the hypothalamus.
Bile acid binding resins	Lowers glucose through a poorly understood mechanism.
Amylin analogues	The hormone amylin is normally secreted together with insulin and decreases glucagon secretion, delays gastric emptying, and increases satiety (Schmitz et al., 2004). Amylin analogues activate amylin receptors and may improve glycemic control.
SGLT2 inhibitors	Sodium–glucose co-transporter 2 (SGLT2) acts in the kidney by reabsorbing glucose from the urine. SGLT2 inhibitors thus decrease this absorption, leading to lower levels of glucose in the blood, but higher in the urine.
Insulins	Various modified insulins are available with varying pharmacokinetics. For instance, long-acting insulins provide a prolonged maintenance of basal levels, whereas short-acting insulins can be used if required after meals for more rapid responses.

Adopted and modified from Inzucchi et al. (2012) and Kahn et al. (2014).

In the combination with obesity, the treatment of T2D is challenging since many drugs have the risk of resulting in weight gain (Inzucchi et al., 2012; Mingrone et al., 2012). Bariatric surgery is used as a weight reducing therapy but has also shown positive effects on T2D, possibly through mechanisms unrelated to the weight loss (Guidone et al., 2006; Sjöström et al., 2004). In subjects with T2D and severe obesity, bariatric surgery has been reported to result in better glucose control compared to medical treatment (Mingrone et al., 2012).

Even though there exists a range of treatment strategies for T2D there is a need for research that can result in novel drug targets and improved medications that are efficient and have reduced side effects. In fact, a large portion of patients that are treated with insulin or other antidiabetic drugs, still retain a poor glycemic control (Liebl et al., 2002).

A complex multi-organ disease

It is apparent that T2D is a complex heterogenic disorder and its pathophysiology and underlying molecular causes are to this date not fully understood. The research efforts that remain to be undertaken to unravel these mechanisms are hampered by the fact that the development of T2D involves complex interactions of multiple environmental and genetic factors, cross-tissue communication, and variations in its pathogenesis between different subjects. Consequently, it is sensible to use a holistic approach to study T2D. In our research we have therefore used methods from systems biology and bioinformatics, relying heavily on so-called omics data and network modeling. A brief introduction to this field is given in the next section. Nevertheless, I strongly believe that both a systems and reductionist biology approach, in collective symbiosis, is needed to uncover the factors that are crucial to reach a final comprehension and efficient treatment of T2D.

Systems biology and bioinformatics

The system is more than the sum of its parts

In essence, science has always been driven by the desire to understand and explain the unknown and unexplored. In biology, few entities act in a complete autonomous manner, but are in contrast connected in one way or another to form a system of components that interact and affect each other. Biological research undertakes to study the function and properties of these components, as well as the systems they constitute. Biological systems can be defined on a hierarchy of different levels. For instance, ecology studies the interactions between organisms as well as their environment. The human body is a complex system built up by different tissues and organs that are connected and controlled by the nervous and circulatory systems. Organs themselves, can also be described as systems. Thus what is seen as a component of one system can also be a system on its own. In my research the system that we study is the cell. Efforts have been made to computationally model the whole cell (Karr et al., 2012). However, the exact definition of the system depends on what one wants to study. One focus that we

have is metabolism, the complete connected network of chemical reactions that are catalyzed by enzymes and enable the conversion of molecules within the cell, leading to the production of energy, amino acids, lipids, carbohydrates, nucleotides, and a range of other important biomolecules. Using so called genome-scale metabolic models (GEMs) allows us to holistically model and simulate cellular metabolism. As the focus is on metabolism, these models do intentionally not take into account other perspectives of the cell, like e.g. cell signal transduction pathways or gene regulatory networks. GEMs have been a central component to many of my research projects and they are therefore further introduced in their own section, in Part I.

Systems biology is an interdisciplinary research field that makes use of mathematical and computational modeling to study the interactions of biological components. A key motivation to employ a systems biology approach is the concept of emergent properties. These are properties of a complex system that cannot be deduced from the isolated individual components, but requires, in order to emerge, that the components and their interactions are collectively considered in the system they constitute.

Networks and Big Data

Complex perturbations to molecular networks often stand at the origin of human disease, emphasizing the importance of network science in medical research (Barabasi et al., 2011). The components of metabolic networks are metabolites, reactions, and genes, and as such they provide a bridge between metabolism and the genome. An important part of systems biology is the integration of genome-wide data, networks and models. Today we are flooded with big data owing to the advancement of several high-throughput technologies, in particular next-generation sequencing (NGS). A recent report estimated that between 100 million to 2 billion human genomes will be sequenced by 2025 (Stephens et al., 2015). Apart from sequencing genomes, NGS technology can also be used for e.g. quantifying mRNA and non-coding RNA, identifying and quantifying DNA-protein interactions, and sequencing heterogeneous microbial populations. These type of data are commonly referred to as omics data, in the sense of being complete or total. Sequencing can generate e.g. genomic and transcriptomic data, whereas other high-throughput technologies are useful in areas such as proteomics, lipidomics, and metabolomics. Consequently, it is not uncommon that systems biology research is data-driven, exploratory, and hypothesis generating. The field of bioinformatics is tightly connected to systems biology and provides an extensive pool of methods and software for analyzing and understanding such data. In my research, both in tool development and in bioinformatic analysis of T2D, the primary measured quantity has been gene expression, in other words transcriptomic data. Therefore, the last part of this background section will be devoted to a few words on transcriptomics.

Transcriptomics

Transcriptomics is the study of the complete collection of transcribed RNA molecules in a single cell or a collection of cells (e.g. from a tissue or microbiota). A common aim is to assess the expression levels of all genes, or at least a fair portion of all genes, in contrast to studies focusing on expression of a predefined selection of interesting genes. In that sense, the analysis of transcriptome data provides an unbiased way to study gene expression. A gene expression program reflects a specific cell state, and its misregulation is implicated in numerous diseases (Lee and Young, 2013). Quantification of gene transcription can be seen as a proxy for protein expression, even though several post-transcriptional and post-translational events play a role in determining the final level and activity of a protein. Protein abundances and mRNA levels have been shown to be positively correlated (Lundberg et al., 2010; Nagaraj et al., 2011; Schwanhausser et al., 2011).

One of the most common applications of gene expression profiling is to search for differential expression, i.e. compare two or more conditions and identify the genes that show statistical differences in their transcript levels. These results can then be integrated with models, networks, and other data, to deduce what functions and properties are affected in the cells, when comparing different conditions. A popular method that is used in this context is gene-set analysis, which will be introduced in Part I.

The most common platforms used for transcriptome profiling are DNA microarrays and RNA-sequencing (RNA-seq). I will not go into any deep technical details about these methods here, as they are both well-established and standard approaches, but just briefly mention a few points. Microarrays were developed during the 1990s and 2000s and have continuously been improved since then (Bumgarner, 2001; DeRisi et al., 1996; Fodor et al., 1991). Microarrays are based on probes, which are short DNA sequences collectively representing e.g. the full transcriptome of an organism. A sample containing mRNA (or actually cDNA) can then be hybridized to these probes and detected using fluorescence. Microarrays are used for relative quantification, where the change in fluorescence intensity between the same spot on different arrays can be related to the concentration change of the corresponding mRNA molecule. The raw signal data is processed, normalized, quality controlled, and statistically analyzed using one of several available bioinformatics pipelines. RNA-seq, on the other hand, is based on NGS technology. Here, a sample of cDNA is sequenced, producing short sequence reads around 100 base pairs (this technology is however developing rapidly and the ability to sequence considerably longer stretches of DNA at high-throughput rates is evolving). The sequence reads are optionally quality trimmed, and thereon aligned to an available genome or transcriptome, or assembled *de novo*. Quantification is done by counting the number of reads that have aligned to a given genomic location. Differential expression analysis is performed by one of several available software tools that handle the discrete count-based nature of RNA-seq data, using various statistical models. Some advantages of RNA-seq, compared to microarrays, includes the increased ability to detect alternative

splicing isoforms and allelic expression, a virtually unlimited dynamic range, and the possibility to identify novel transcripts (Wang et al., 2009).

Part I: The toolbox

This part of my thesis will describe the tools that were central to my research and publications, and which I took part in developing. Gene-set analysis (GSA), as mentioned in the introduction is a method to facilitate the interpretation of (primarily) transcriptomic data by exploring overlap with annotated sets of genes representing specific functions or properties. Here this concept is presented in more detail, and our contribution, in form of a software package piano, is described. Further on, to study myocyte metabolism, which is central to T2D, and to bridge it to transcriptomics, we reconstructed a myocyte-specific genome-scale metabolic model (GEM). Finally, our software tool Kiwi is presented, which improves the visualization and interpretation of the results from a GSA, based on the topology of e.g. a GEM. Although I have mainly been applying these tools to study T2D and obesity in skeletal muscle, all tools described here are general, in the sense that they can be applied to a wide range of data and biological research topics.

Gene-set analysis (Paper I)

The difficulty of interpreting gene expression profiles

In biological research it is very common to compare two or more groups to each other, with the aim of identifying statistically significant differences. This can be e.g. to assess the effect of a perturbation or stimulation, or to define distinguishing characteristics between different populations, stages, or conditions. The properties that are measured and compared are decided by the researcher, in consideration of the scientific question and the experimental design. In transcriptomic research the variables of interest are the genes and the measured values are their expression. Typically, researchers will perform an appropriate statistical test to assess the extent of differential expression between the different groups, resulting in a list of *p*-values that should be adjusted for multiple testing. At this point it is easy to identify the list of genes that seem to display changes in expression levels (e.g. by selecting genes with a low probability that their expression changes between the groups occurred by chance, i.e. with low *p*-values). What remains is the difficult task of interpreting what these collective changes mean, and what implication they have on cell function. Of course, if the number of significantly differentially expressed genes is low, it may be possible to manually go through the list and evaluate the function of each individual gene. In many cases however, a large number of genes are significant, making the task of interpretation too complex to be efficiently carried out by a human individual, justifying the need for computational methods to aid in this step.

Gene-sets – less is more

For well-studied organisms, including humans, the genome has been extensively studied and therefore widely annotated, in terms of gene function and interaction. An enormous amount of data is available through a wide range of databases. A good example of this is the Gene Ontology (GO) project which through collaborative efforts aims to provide a consistent annotation of genes (or to be correct, gene products) by associating them to defined terms (ranging from broad, e.g. metabolic process, to specific, e.g. oxidoreductase activity). These terms ultimately belong to one of three main ontologies: biological processes, cellular components, and molecular functions (Ashburner et al., 2000). Each GO-term can be traced back, through a number of more broadly defined parent GO-terms, and eventually to one of the three main ontologies. A given gene can thus be described by a number of associated GO-terms. What is maybe more interesting from a systems biology perspective is the reverse, that a given GO-term is associated to a number of genes. This is an example of a gene-set. Other examples of categories of gene-sets could be transcription factors (genes sharing a specific transcription factor binding site), or pathways (genes participating in specific metabolic or signaling pathways). Gene-set collections can be compiled from available databases, but there also exists dedicated databases providing various gene-set collections, e.g. the Molecular Signatures Database (Liberzon et al., 2011; Subramanian et al., 2005) and Enrichr (Chen et al., 2013).

Returning to the problem of interpreting long lists of significant genes, it has become popular to exploit the collective information provided by gene-sets. Moving from the gene level to the gene-set level has two major advantages. First, gene-sets typically have a descriptive name, focusing on the function or property of its genes, which is often easier to understand than the set of individual gene names. Second, even though a gene can belong to multiple gene-sets, the total number of gene-sets in a gene-set collection is smaller than the total number of genes. This means that, moving from the gene to the gene-set level, the researcher typically has a smaller list of gene-sets to go through than the initial list of significant genes.

Overrepresentation analysis

A common approach to interpret gene expression changes on the level of gene-sets has been to select genes using a binary cutoff (e.g. all genes with p-value<0.01) and determine whether these genes are significantly overrepresented in any gene-set. That is, if any gene-set contains more genes from the selected list than expected by random chance. Overrepresentation analysis (ORA) can be done e.g. by using Fisher's exact test, chi-square test, or a binomial test (Khatri and Drăghici, 2005). The drawback of ORA is that the use of an arbitrary binary cutoff omits a lot of information since genes falling outside the cutoff are discarded and the individual statistical values for the remaining genes are neglected. An advantage is that ORA is a fast and computationally light method, and has therefore been employed by online web-based user interfaces like DAVID (Hosack et al., 2003), and Enrichr (Chen et al., 2013).

Gene-set analysis – using all gene-level statistics

In 2003 Mootha et al. introduced a method and tool, termed gene-set enrichment analysis (GSEA), that has become widely popular (Mootha et al., 2003; Subramanian et al., 2005). A primary difference to ORA is the use of all gene-level statistics to calculate a score for each gene-set. I will use the term gene-set analysis (GSA) to refer to methods like GSEA. Formally, where a gene-set score or statistic can be calculated by a function $f(G, S)$, where G is a vector of all gene-level statistics and S is a vector giving the positions of the gene-set genes in G. To assess the statistical significance of the gene-sets, p-values can be calculated for the gene-set statistics, either from a theoretical null distribution or by permuting the genes or samples. The advantage with GSA is that no specific cutoff is required. Rather, information for all genes is used, even enabling the detection of small but collectively coordinated responses that converge on a specific function or property, represented by the gene-set. There are several methods that fall in the definition of GSA, but I will not attempt to list them here as they have already been reviewed by us (**Paper I**) and others (Ackermann and Strimmer, 2009; Hung et al., 2012; Maciejewski, 2013; Tarca et al., 2013).

A refined framework for gene-set analysis

In **Paper I** we focused our attention on some of the current limitations of GSA. We identified four issues that needed to be addressed:

1. The lack of a gold standard for GSA makes it difficult to evaluate different methods and decide which one to use.
2. It is unclear what approach to use when calculating the p-values of the gene-set statistics.
3. Available methods are implemented in various tools, on different platforms and in different programming languages, making it difficult for users to test and compare methods in a straightforward manner.
4. It is difficult, but useful, to assess the directionality of a gene-set, in terms of the different fold changes of its member genes.

To deal with these issues we developed a new analysis tool called piano (Platform for Integrative Analysis of Omics data), fully documented and freely available as an R/Bioconductor package (Huber et al., 2015; R Core Team, 2005). In piano, we implemented 11 methods for calculating gene-set statistics and 3 ways to calculate gene-set p-values, enabling the comparison and evaluation of different GSA methods, thus addressing point 1-3. With regards to point 4, we introduced the concept of directionality classes, where a gene-set is assigned different scores, each capturing different characteristics of the collective pattern of up- and down-regulation of the genes in the set. These concepts will be briefly elaborated in the following two sections and an overview of the workflow is given in Figure 1.

Gene-set statistics and significance estimation

At the time of developing piano, we identified 11 methods to calculate gene-set statistics that fitted our criteria:

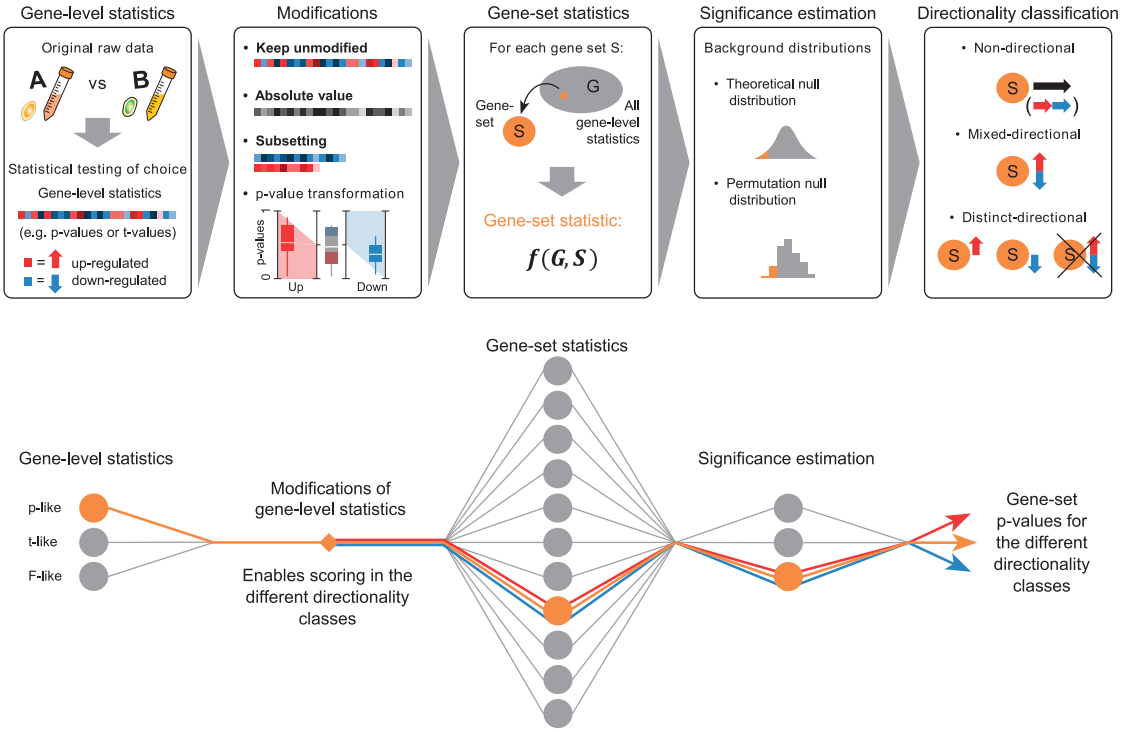

Figure 1. Overview of the GSA workflow in piano. The input gene-level statistics are used to calculate gene-set statistics according to one of the 11 available methods (orange path). In a similar manner, modifications of the gene-level statistics (e.g. absolute vales, or subsets of up- and down-regulated genes) are used in parallel (red and blue paths). This enables the calculation of gene-set statistics and corresponding gene-set *p*-values in different directionality classes (non-directional, mixed-directional, and/or distinct-directional).

- Fisher's combined probability test (Fisher, 1932)
- Stouffer's method (Stouffer et al., 1949)
- Reporter features (Oliveira et al., 2008; Patil and Nielsen, 2005)
- Parametric analysis of gene-set enrichment, PAGE (Kim and Volsky, 2005)
- Tail strength (Taylor and Tibshirani, 2006)
- Wilcoxon rank-sum test
- Gene-set enrichment analysis, GSEA (Subramanian et al., 2005)
- Mean
- Median
- Sum
- Maxmean (Efron and Tibshirani, 2007)

The idea was to create a unified framework with consistent input and output, for all methods. This was implemented using a single function where the user can input gene-level statistics and simply select the desired GSA method and significance estimation procedure to be used. This provided a simple system to run and compare different GSAs based on the same input data, as opposed to having to move between different platforms and tools. In practice, there is one limitation however, in that not all types of gene-level statistics can be used as input to all GSA methods. As an example, Fisher's test, Stouffer's method, and Tail strength, are all designed to work specifically on *p*-values.

After the gene-set statistics are calculated, using one of the listed methods (the calculation for each method is described in detail in **Paper I** and in its supplementary), gene-set *p*-values are estimated using one of three methods. In all cases, either gene sampling or sample permutation can be used. In these procedures, either the gene-level statistics are permuted or the sample labels are permuted and new random gene-level statistics are calculated. These random gene-level statistics are then used to recalculate the gene-set statistics, a procedure that is repeated e.g. 10,000 times, thus generating a background distribution of gene-set statistics that can be used for significance estimation. Alternatively, a few of the methods are defined so that a theoretical null distribution can be used to calculate *p*-values from the gene-set statistics. As an example, for Stouffer's method *p*-values can be estimated from the gene-set statistics using the normal cumulative distribution.

Directionality classes

When comparing transcriptional profiles between two different conditions, it is clear that a significantly differentially expressed gene is either up- or down-regulated, and thus its functional activity is either increased or decreased, respectively (assuming of course that a change in transcript level is reflected in protein level and activity). For a gene-set, the situation becomes slightly more complex, as the fold changes of multiple genes have to be taken into account to determine if the gene-set is up- or down-regulated (if it is even meaningful to talk about up- and down-regulation of a gene-set). The question of directionality can be handled in different ways.

For example, for the reporter features method, Oliveira et al. (2008) first use the gene-level *p*-values to calculate gene-set scores. This procedure does not take into account directionality, and significant gene-sets will thus be those containing genes with low p-values, regardless of the sign of their fold changes. To capture the direction of regulation the authors propose to subset the input data (gene-level *p*-values) into up-regulated and down-regulated genes. An effect of this is that a gene-set with only a minority of down-regulated genes will still be deemed significantly down-regulated, given that the few down-regulated genes have low enough *p*-values. Further on, a gene-set can simultaneously be significantly up- and down-regulated if both subsets of the gene-set contain genes with low enough *p*-values. (The authors do point out that the subset analyses should be interpreted in connection to the initial analysis.)

Another example is the PAGE method (Kim and Volsky, 2005). Here, the gene-set statistic is a function of the difference between the average fold change of the gene-set and the average fold change of the whole dataset. This means that the sign of the gene-set statistic will reflect whether the gene-set is considered up- or down-regulated. An effect of this approach is that gene-sets containing a mix of up- and down-regulated genes will not be significant, even if all the genes in the set are significant, as their fold changes will cancel out.

Obviously, the meaning of an up- or down-regulated gene-set is different in these two examples. To address this, we introduced three different directionality classes,

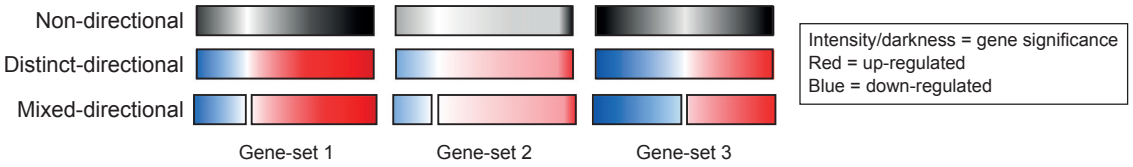

Figure 2. Illustration of the directionality classes. Gene-set analysis identifies significant gene-sets. However, the criteria of being significant differs depending on the choice of method. To address this we defined three directionality classes in which piano scores the gene-sets. The non-directional class identifies gene-sets with significant genes, disregarding the direction of change (gene-set 1 and 3). The distinct-directional class finds gene-sets that display a significant change in one distinct direction, either up or down (gene-set 1 and 2, but not 3). The mixed-directional class separates the subset of up-regulated genes and down-regulated genes in a gene-set, and checks if they are significant, independently of each other. Gene-set 1 and 3 are both significantly affected by up-regulation and down-regulation in the mixed-directional class. Gene-set 2 was found to be distinct-directional-up, but is not significant in the other classes. This means that the significance is mainly due to the majority of up-regulated genes (a coordinated small up-regulation), rather than the highly significant genes.

in an attempt to unify the interpretation of the results from the different GSA methods (Figure 2). As an example, the gene-set *p*-values calculated by the standard reporter features method are classified as non-directional. A gene-set that is significant in the non-directional class can be interpreted as containing more "high-scoring" (e.g. low p-values, or high absolute fold changes) genes than expected by random chance, discarding the gene-level directionality. On the other hand, the gene-set *p*-values calculated in the subset analysis approach of the reporter features method are classified as mixed-directional. A gene-set that is significant in the mixed-directional class when the subset (small or large) of genes regulated in the same direction and more "high-scoring" than expected by random chance. A gene-set can be significantly up- and down-regulated simultaneously, hence the name mixed-directional. Finally, as in the example of PAGE, these gene-set *p*-values are classified as distinct-directional. Only gene-sets containing "high-scoring" genes with consistent fold changes will be significant, and they will be either up- or down-regulated, not both, hence the name distinct-directional.

By knowing what directionality class the gene-set *p*-values belong to, makes it easier for the user to interpret the meaning of directionality on the gene-set level and compare different GSA methods. In piano we took this one step further, by automatically calculating gene-set *p*-values in all possible directionality classes. This was enabled by modifying the original gene-level statistics (e.g. absolute vales, or subsetting up- and down-regulated genes) and running parallel GSA analyses. The possible directionality classes depend on the specific combination of gene-level statistics and gene-set statistic calculation method used. Having access to information in several directionality classes enables a more comprehensive interpretation of the gene-level changes underlying a significant gene-set.

Comparison of gene-set analysis methods

With the established GSA framework, unifying the workflow and enabling the use of different gene-level statistics, different methods to calculate gene-set statistics, different methods to estimate gene-set significance, and structuring the output in different directionality classes, it was possible to start to compare different GSAs. One GSA can be seen as a unique path through the graph shown in Figure 1. We

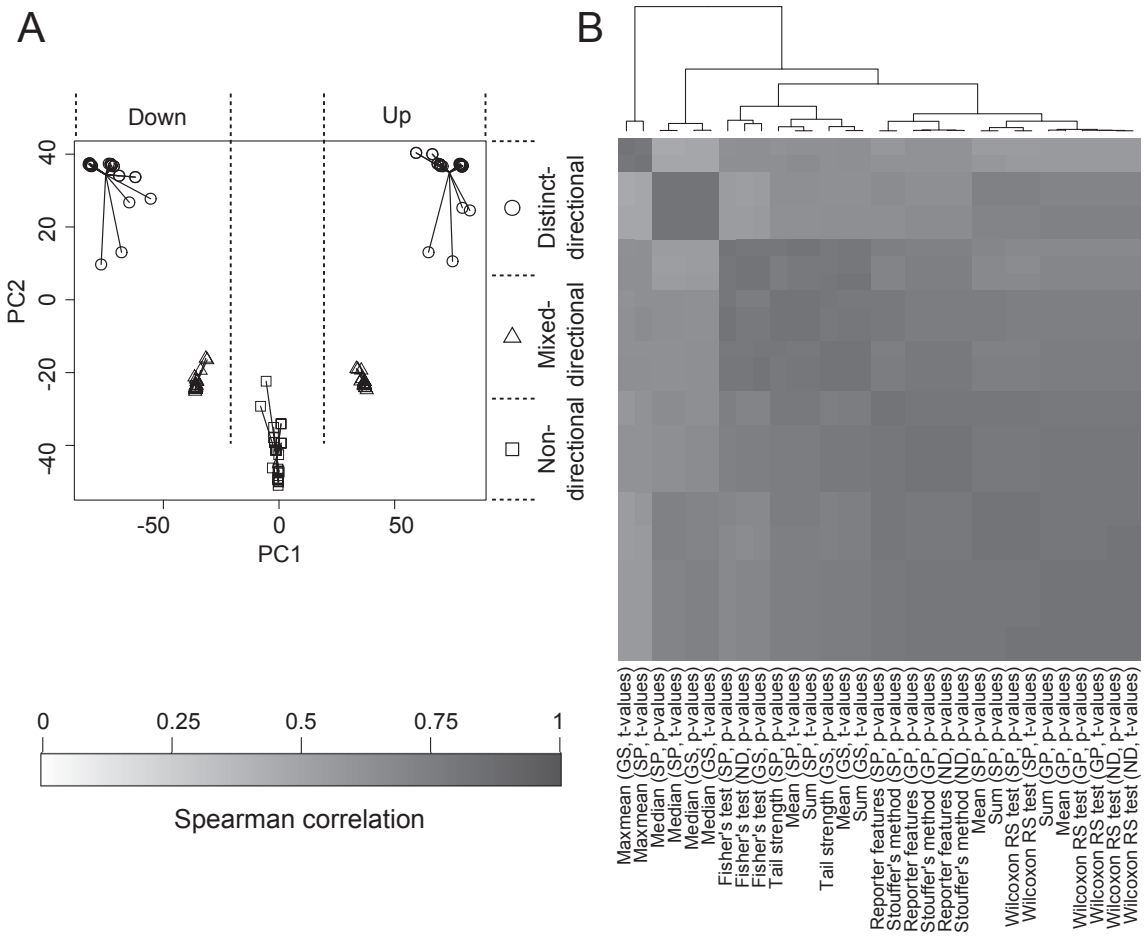

Figure 3. Comparison of different GSA runs. A) PCA plot of 127 gene-set *p*-value vectors resulting from different GSA runs (varying gene-level statistic type, gene-set statistic calculation method, and significance estimation method). The results separate into our defined directionality classes, supporting our approach. B) Pairwise Spearman correlation between gene-set *p*-value vectors (of the non-directional class) for the different GSA runs, showing a high consistency of the results from the different runs. The results are similar for the other directionality classes. GS: gene sampling, SP: sample permutation, ND: null distribution.

used a published microarray dataset (Mootha et al., 2003) and performed a differential expression analysis between NGT and T2D subjects, generating gene-level *t*-values and *p*-values. Using a collection of GO-term gene-sets, these gene-level statistics were used as input to run GSA for all possible combinations of gene-level statistic type, gene-set statistic calculation method, and significance estimation method. Counting also the separate runs for the different directionality classes, this resulted in a total of 127 GSA runs. The 127 gene-set *p*-value vectors were then compared using principle component analysis (PCA) and Spearman correlation (Figure 3). The primary separation of *p*-value vectors (representing the different GSA runs) in the PCA plot depended on the directionality classification. This supports our classification scheme and implies that the way to interpret a GSA result can be based on the corresponding directionality class, as earlier described, without having to consider the chosen GSA method. The high correlation between different GSA runs is good from a user perspective, meaning that different GSA methods will not yield drastically different results. It has been suggested before to run several GSA tools and combine the results (Huang et al., 2009; Naeem et al., 2012). Through our implemented framework in piano we have laid the groundwork

for performing such a task and thus proposed, in **Paper I**, a consensus scoring approach.

Consensus gene-set analysis

The consensus GSA is carried out by running a multiple of parallel GSAs (varying the input, gene-set statistic calculation method, and/or significance estimation method). For each directionality class and each GSA run, the gene-sets are ranked according to their *p*-values. This can be seen as a ranked voting problem, where each GSA run is a voter that ranks the candidates (gene-sets). A rank aggregation method can then be used to select the winning gene-set (or actually to condense the multiple rank-lists into a consensus ranking of all gene-sets). In piano this can be done using the average or median rank, or by one of the classical methods proposed by de Borda (1781) and Copeland (1951). Ranks rather than actual gene-set *p*-values are used to treat conservative and less conservative methods equally. When using the consensus GSA approach, I would recommend to investigate the gene-set rank (and *p*-value) distribution for a given gene-set, across GSA runs, in connection to drawing biological conclusions. Piano offers visualization functions to view the consensus rank as well as individual ranks from each GSA run in order to evaluate if the results are consistent across runs or not. An alternative to varying the methods between different GSA runs, is to instead vary the input. As an example, on could perform one GSA based on gene-level *p*-values and another on fold changes, and use the consensus GSA approach to detect significant gene-sets based on both statistical significance and biological change of the member genes.

Other types of methods

There are other kinds of methods that do not fit the piano workflow. I will not attempt to give a complete overview here but rather list some examples to give an idea of the range of methods. See e.g. reviews by Ackermann and Strimmer (2009), Huang et al. (2009), Maciejewski (2013), or Nam and Kim (2008) for more information. For instance, some methods start from the raw expression data by using multivariate and global tests (Goeman et al., 2004; Hummel et al., 2008; Kong et al., 2006; Mansmann and Meister, 2005; Tsai and Chen, 2009). Other methods incorporate gene (or gene product) interaction networks into the analysis of gene-sets (Alexeyenko et al., 2012; Glaab et al., 2012) or take into account gene overlap between gene-sets (Tarca et al., 2012). There are also methods and tools specifically designed for e.g. RNA-seq data (Lee et al., 2015) or genome-wide association studies (de Leeuw et al., 2015; Nam et al., 2010; Segrè et al., 2010).

Considerations when performing gene-set analysis

As powerful as GSA can seem, it has its pitfalls, and it is important to remember that it will never be better than the quality of the input data, i.e. the gene-level statistics and the gene-set collection. I will here go over some points that I have found useful to consider when running GSA.

Gene-sets are typically acquired from various dedicated online databases with different levels of criteria and validity for gene and gene-set associations, as well

as being biased to more heavily researched biological processes. Therefore, when the GSA result points towards e.g. significant regulation of glycolysis, the researcher has to trust that the genes ascribed to the glycolysis gene-set are correct. Further on, the gene-set name may in some cases also be misleading. As an example, of the 11 genes in the GO-term gene-set oocyte development, 3 are also in the GO-term male sex differentiation. The former gene-set might therefore, counterintuitively, be significant largely due to genes associated to male-specific gene expression. Consequently, it is important to revisit the gene-level data when interpreting the GSA results. In addition to this, it is important to consider the gene-set size, as a gene-set containing 3 genes compared to one containing 100 genes may have different impact on the biological interpretation.

Another issue to take into account is that genes that are both positively or negatively associated with a specific biological process (e.g. inhibiting or activating) may be part of the same gene-set. This makes it very difficult to correctly interpret directionality on the gene-set level. Further on, there is no perfect rule that connects the level of gene expression change, or its statistical significance, to the level of influence it has on the biological process represented by the gene-set. For instance, the same change in gene expression of two genes in a gene-set may in reality affect the represented biological process to different extent. Similarly, a single gene may be associated with several gene-sets and thus contribute to them with the same score, but influence them differently biologically.

Finally, when using the permutation-based approaches for gene-set significance estimation it is important to consider the gene-level dataset properties. As an example, take a dataset where no genes are deemed significant. It could still be possible that some gene-sets contain genes with extreme enough values (in relation to the dataset) to generate a significant gene-set. In other words, a gene-set may become significant since it contains the top differentially expressed genes (hence the most extreme genes in the dataset), even if none of these genes are consider statistically significant. Because of this, and the other reasons mentioned above, it is important to revisit the gene-level information when interpreting the GSA results.

The piano package – three years later

The purpose of a tool is to be used to solve the task it was designed for. In contrast to published biological results, it is therefore relevant to evaluate the actual usage of a tool after its publication. Through the publication peer review process and acceptance of piano into Bioconductor, some quality control was assured. However, its true impact is judged by the usage statistics, a number that, to be blunt, probably to a larger extent reflects the tool's user friendliness, documentation quality, and ease of installation, rather than the scientific quality of the software. Table 3 shows the number of publications using piano (and total

Table 3. Usage and total citations for piano.

	2013	2014	2015
Using piano (number of publications)	9	21	24
Total number of citations	12	32	31

number of citations) in the last three years, covering a range of biological topics. As a developer it is reassuring to see that the tool is used to assist in impactful research, justifying the countless hours spent on coding. In addition, piano is available through BioMet Toolbox, an online web-based user interface (Garcia-Albornoz et al., 2014), and has been included to perform analyses in an RNA-seq analysis pipeline (Fonseca et al., 2014), and in the online Expression Atlas from EBI (Petryszak et al., 2015).

Genome-scale metabolic models (Papers II & III)

The structure of genome-scale metabolic models

In **Paper II** we described genome-scale metabolic models and their application to research on diabetes and obesity. A GEM is a comprehensive list of metabolic reactions aiming at covering the complete metabolism of a cell. Each reaction is stoichiometrically defined and consists of the participating metabolites and is linked to its associated enzymes (additional annotation can also be included). It is common to compartmentalize the GEM, so that the same reaction taking place in both the cytosol and mitochondria can be properly modeled autonomously. This is solved by introducing compartment specific metabolite names, e.g. glucose[c] for cytosolic glucose, and transport reactions for connecting metabolites in different compartments, e.g. glucose[s] $\leftrightarrow$ glucose[c] for transport of glucose between the extracellular space and the cytosol. For further structure, groups of reactions can also be assigned to specific pathways or metabolic subsystems. Figure 4A shows the relation between reactions, metabolites, and enzymes, and this topology enables the construction of various networks (e.g. metabolite-metabolite, reaction-reaction, gene-metabolite, or gene-pathway) that can be used for network-dependent analysis and data integration.

A GEM can also be formulated as a stoichiometric matrix, as shown in Figure 4B, to enable simulation and prediction of metabolic phenotypes (Bordbar and Palsson, 2012). By assuming steady state, i.e. that the metabolite concentrations remain constant during the simulated condition, and consequently requiring that the production and consumption rates of each metabolite are equal, it is possible to calculate fluxes for the reactions in the system. Typically an indefinite number of metabolic flux distributions will fulfill the steady-state assumption (e.g. setting all fluxes to zero would be an example of a trivial solution). By introducing constraints on specific fluxes, the solution space can be decreased. Finally, by including an optimization (e.g. maximize the sum of fluxes contributing to synthesis of macromolecules essential for growth) the problem reduces to a unique flux distribution, the optimal solution (unless alternative optima are present due to e.g. futile cycles). This kind of procedure is referred to as flux balance analysis (Orth et al., 2010).

In my research I have not focused on simulation-based analysis of GEMs, but rather to exploit the metabolic network topology to contextualize and interpret

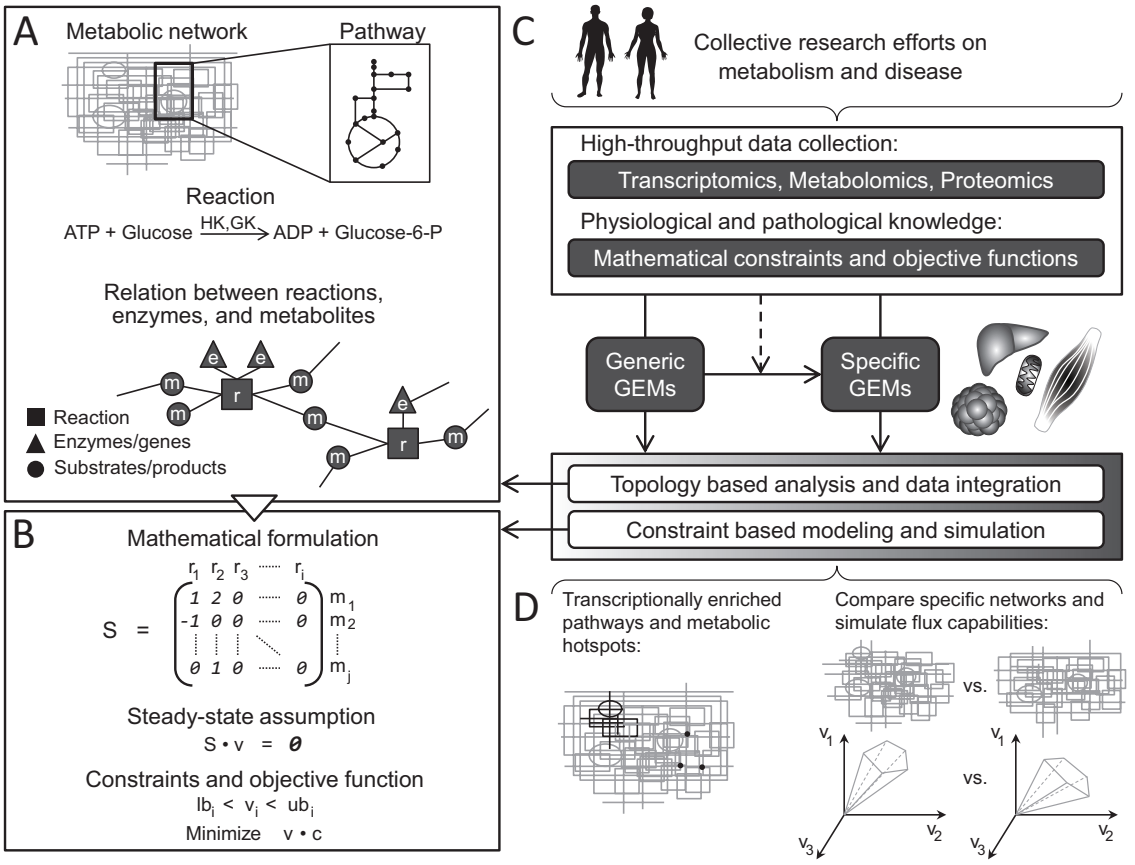

Figure 4. Overview of the structure and application of GEMs. A and B) A GEM can be represented both as a network (A) and in mathematical terms (B). C) High throughput data and knowledge from molecular biology can be used for topology based analysis as well as constraint based modeling and simulation. D) This can be used to identify transcriptionally enriched pathways or metabolites, or to find differences in network structure or simulation capabilities.

gene expression data, primarily in connection to T2D as presented in Part II, but also to explore the metabolism of ovarian cancer (Aspuria et al., 2014).

Human genome-scale metabolic models

Initially GEMs were developed and used to study microbial metabolism (Oberhardt et al., 2009). In 2007, the first GEMs of human metabolism, Recon 1 (Duarte et al., 2007) and EHMN (Ma et al., 2007), were created and anticipated to provide a systematic and holistic approach to study the complexity of human metabolism and its dysfunction in connection to disease. A few years later, information from Recon1, an updated version of EHMN (Hao et al., 2010), HumanCyc (Romero et al., 2004), and the Kyoto Encyclopedia of Genes and Genomes (KEGG) database, was used to construct the Human Metabolic Reaction database (HMR) (Agren et al., 2012; Mardinoglu et al., 2013). These efforts of mapping out the complete human metabolism have continuously progressed. In 2013 Recon2 (Thiele et al., 2013) was published, and in 2014 HMR was updated to HMR2 with a more comprehensive coverage of lipid metabolism (Mardinoglu et al., 2014). Just to give an idea of the scope of these models: HMR2 contains around 8,000 reactions, 3,000 unique metabolites, and 3,700 genes.

The human body consists of a range of different cell types with different phenotypes. The generic human GEMs do not directly account for the differences in metabolism across these different cell types, but are great resources for constructing cell type-specific GEMs. Typically this involves creating a subset of the generic metabolic network (representing the active metabolism in a given tissue or cell type) that fulfils some network and simulation specific requirements, e.g. fully connected network with no dead-end metabolites, and all reactions being able to carry flux. The selection of reactions, metabolites, and genes to be included is based on evidence from omics data available for the specific cell type in question. A handful of algorithms with different approaches have been developed for creating cell type- or context-specific GEMs (Agren et al., 2012; Agren et al., 2014; Becker and Palsson, 2008; Jerby et al., 2010; Robaina Estévez and Nikoloski, 2015; Schmidt et al., 2013; Shlomi et al., 2008; Wang et al., 2012; Vlassis et al., 2014; Yizhak et al., 2014). In a recent review the range of different available cell type-specific GEMs was summarized, which included models for brain cells, heart cells, liver and hepatocytes, kidney, and adipocytes (Ryu et al., 2015). As there was not a comprehensive GEM for skeletal myocytes available at time, we set out to reconstruct a myocyte metabolic network in order to have a framework to study muscle metabolism in connection to T2D and obesity.

A myocyte genome-scale metabolic model

In **Paper III** we describe the procedure of reconstructing the myocyte GEM. The general workflow for this is shown in Figure 5A. High throughput data at the transcript and protein level was used to score the reactions in HMR2 to evaluate whether they should be included in the final myocyte model. To do this in a systematic way we started by comparing the transcriptome and proteome data. The transcriptome data was generated using deep RNA-seq of human primary myocyte cell cultures from three males and three females and transcript levels were estimated by calculating FPKM-values (fragments per kilobase of transcript per million mapped reads). This data is myocyte-specific, in contrast to RNA isolated from skeletal muscle biopsies which can be contaminated by other cell types. The proteome data comes from the Human Protein Atlas (HPA) (Uhlén et al., 2015). Briefly, the structure of the HPA data is such that a protein abundance score is associated to each protein, based on immunohistochemistry assays. This score has four levels (not detected, low, medium, high). Further on, the evidence-based reliability of the abundance score for a given protein is classified as either supportive or uncertain. At the time of reconstructing the myocyte GEM, the majority of the abundance scores were classified as uncertain. For the proteins with supportive abundance scores there was a clear difference in FPKM-value distributions between genes detected on the protein level (low, medium, or high) and genes not detected on the protein level (Figure 5B). A gene with a high FPKM-value would more likely also be expressed on the protein level. This pattern was however not seen for the genes with uncertain protein abundance scores. A density plot of all FPKM-values revealed a bimodal distribution of lowly expressed (LE) and highly expressed (HE) genes (Figure 5C). The LE genes have been shown to likely have non-functional transcripts whereas the HE genes are translated into

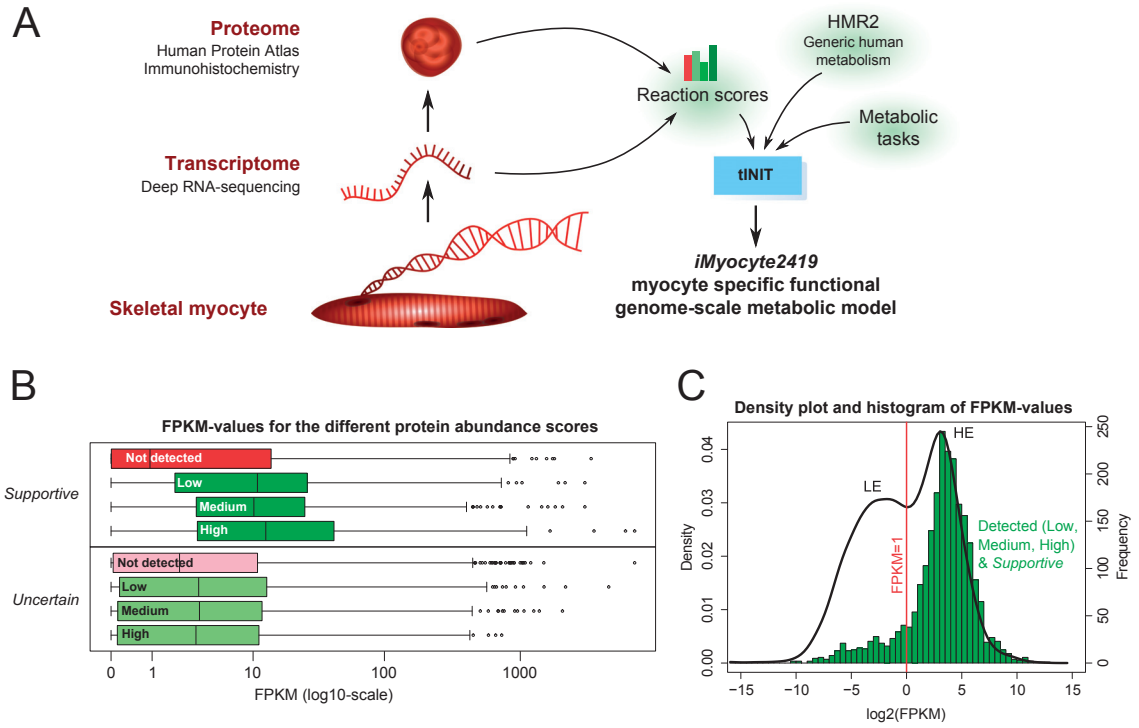

Figure 5. Reconstruction of the myocyte metabolic network. A) An overview of the model reconstruction workflow. B) Relation between myocyte protein and transcript levels. Higher FPKM-values (transcript levels) were observed for expressed proteins. This information was used to correct the uncertain protein abundances. C) A bimodal distribution of transcript levels was observed. The histogram of FPKM-values, corresponding to the subset of proteins expressed with high certainty, is shown in green. This information was used to determine a FPKM-based cutoff to predict protein presence.

functional proteins (Hebenstreit et al., 2011). The FPKM-values of the subset of genes that were detected on the protein level (low, medium, or high) with supportive reliability coincide with the HE part of the density plot. An FPKM-value of 1 separates the LE and HE distributions and we used this as a rough cutoff to predict, from the RNA-seq data, whether or not the corresponding protein was present. In practice, we considered only the genes that were either not measured on the protein level, or that were, with uncertain reliability, not detected on the protein level (pink box in Figure 5B). The abundance levels of these genes were then corrected using the RNA-seq data, so that if their FPKM>1 their abundance level was set to detected (low). For the remaining genes, we used the available protein abundance scores. We believe that this procedure is a fair tradeoff between using only the proteome data (which is biologically closer to the presence of a reaction, but with a large extent of uncertain and unmeasured abundance scores) and using only the transcriptome data (which is biologically farther away from the reactions, but has higher gene coverage and is more robust across genes in terms of reliable estimates of transcript abundances).

The combined transcriptome and proteome abundance scores were then used as input to the tINIT (task-driven Integrative Network Inference for Tissues) algorithm, which has been developed to reconstruct cell type-specific GEMs from a draft generic human metabolic network (Agren et al., 2012; Agren et al., 2014). Briefly, tINIT starts from a reference model (HMR2 in our case) and a list of

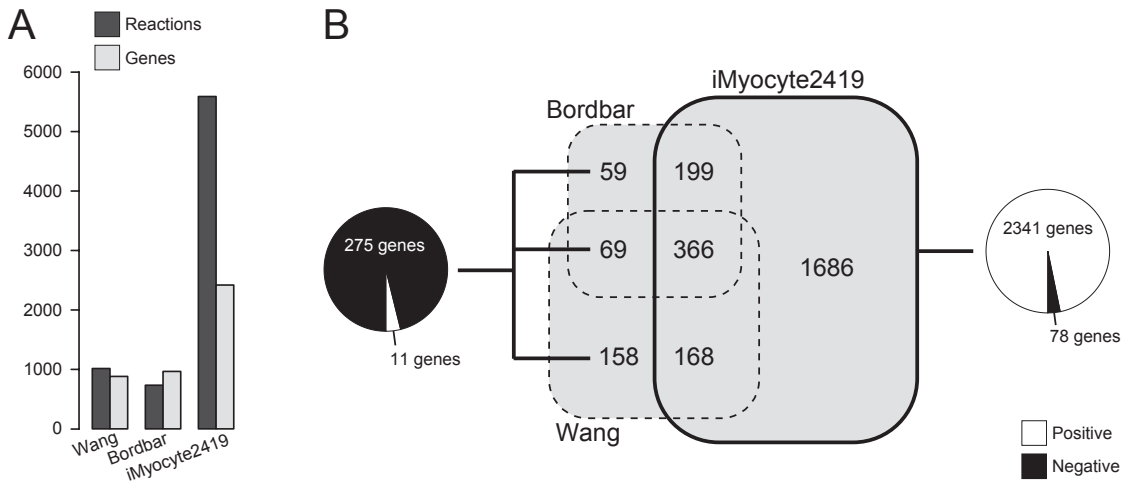

Figure 6. Comparison of myocyte GEMs. A) Number of reactions and genes for the three GEMs. B) Venn diagram showing the overlap in gene content. The genes excluded from iMyocyte2419 are mainly negatively scored (not detected in myocytes) whereas the included genes are mainly positively scored (detected in myocytes).

metabolic tasks that the final model should be able to perform (e.g. produce ATP from glucose). The abundance levels of the reaction-associated genes are used to score the reactions, where a non-detected gene gives a negative score, whereas a detected gene gives a positive score. Through a multistep optimization procedure, tINIT returns a model favoring the inclusion of positively scored reactions while ensuring a fully connected network where all reactions can carry flux and all metabolic tasks can be carried out. The 247 metabolic tasks that were used led to an addition of only 22 reactions on top of the ones included from the reaction scoring and connectivity requirements. This meant that most of the metabolic tasks could be simulated already from a model reconstruction based on only the abundance data. The final GEM, iMyocyte2419, contains 5,590 reactions, 2,396 metabolites, and 2,419 genes.

We compared iMyocyte2419 to two previously published skeletal muscle GEMs (Bordbar et al., 2011; Wang et al., 2012). Both of these models are smaller in scope (Figure 6A) but it was still of interest to see if there was any overlap between the GEMs. In terms of genes, 82% and 70% of the Bordbar and Wang GEMs, respectively, are encompassed by iMyocyte2419 (Figure 6B). Of the genes not included in iMyocyte2419, the vast majority were negatively scored, i.e. not present in myocytes based on the transcriptome and proteome data. On the other hand, only 78 out of the 2419 genes included in iMyocyte2419 had negative scores. These genes correspond to 5.2% of the reactions in iMyocyte2419 and had to be included to ensure connectivity or functionality of the model. In fact, 14 of these genes have received updated abundance levels in later HPA versions and are now classified as detected in myocytes on the protein level.

In Part II I will summarize how we have exploited the topology of the myocyte GEM and applied it to study T2D and obesity. However, the myocyte GEM also remains available as a resource for other kinds of studies relating to muscle metabolism.

Combining GEMs and GSA (Paper IV)

The network structure of a GEM, connecting reactions, genes, metabolites, and pathways, can be used to construct gene-sets for the specific purpose of analyzing genome-wide gene-level data (typically transcriptomics) in the context of metabolism.

Pathway gene-sets

Metabolic pathway gene-sets can be extracted from GEMs if the reactions are annotated to belong to a specific pathway or metabolic subsystem. This is not an exclusive feature of GEMs, as pathway gene-sets also can be acquired from databases such as KEGG (Kanehisa et al., 2012), BioCyc (Caspi et al., 2014), or Reactome (Croft et al., 2014). One advantage of using GEM-derived pathways could be that gene-sets are filtered to represent a specific context (like a certain cell type).

Metabolite gene-sets – reporter metabolites

Another way to exploit the topology of a GEM can be done by extracting the metabolite-gene network. Each metabolite is then connected to all genes associated to reactions the metabolite is involved in. These genes thus have the possibility to influence the production and consumption rate of that metabolite. High throughput transcriptome data is readily available and relatively easy to generate, and can be integrated with a GEM to reveal metabolites surrounded by significant transcriptional regulation, thus translating transcriptome information into the context of metabolism. By using metabolites as gene-sets, defined by the metabolite-gene network, GSA can be used to pinpoint metabolite nodes in the huge metabolic network that are connected to differential gene expression. These metabolites were termed reporter metabolites by Patil and Nielsen (2005).

Gene-set interaction networks

Metabolite gene-sets are special in the sense that they themselves make up a network. The metabolite-metabolite network connects metabolites that participate in the same reaction, and an edge in this network thus represents one reaction step. This property is not limited to GEM-derived metabolite gene-sets. Other types of gene-sets may represent biological entities that connect to or interact with each other. In **Paper IV** we termed this the gene-set interaction network. Other examples of such networks are transcription factor gene-sets connected in a gene regulatory network (Oliveira et al., 2008), or GO-terms connected in the directed acyclic graph that is defined by the GO-term relationships in the GO hierarchy. An important distinction that we make for gene-set interaction networks is that the gene-set connections are not simply based on gene overlap between gene-sets. In other words, the gene-set interaction network should add an additional layer of information that could not simply be extracted from the gene members.

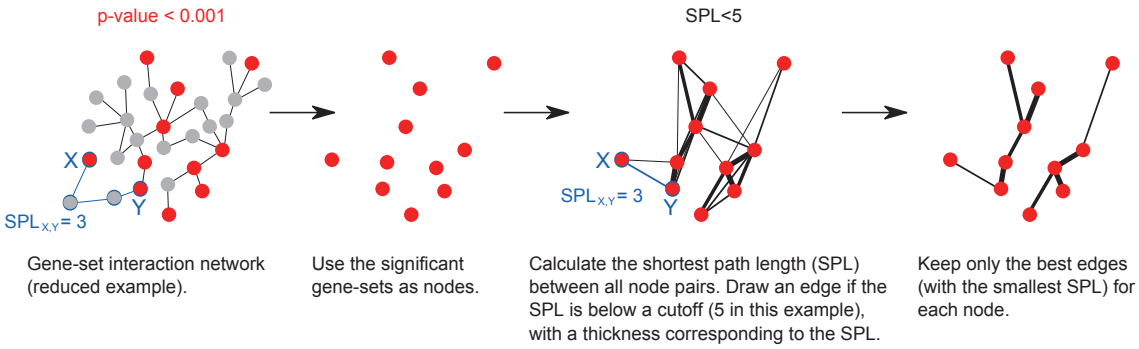

Figure 7. The Kiwi workflow. A toy example of a gene-set interaction network, heavily reduced in size, is used to illustrate the different steps of the visualization algorithm.

Some problems with visualization

GSA results are often presented as tables or heatmaps, which do not take into account the possible connections between gene-sets, represented by the gene-set interaction network. In the case of metabolite gene-sets, as an example, it may not always be obvious to know which of the significant gene-sets are closely connected in the metabolic network. This piece of information would be valuable during the interpretation of the GSA results, as a connected group of metabolites may point to regulation of a specific part of metabolism, something that would not immediately be identified if the information from the metabolite gene-set interaction network was omitted. At the time, there was no optimal tool for visualizing GSA results in the context of the gene-set interaction network, which we addressed by developing the tool Kiwi.

Kiwi – integrating GSA and gene-set interaction networks

In **Paper IV** we present the network-based visualization tool Kiwi, implemented in Python and available through a web-based user interface in BioMet Toolbox. In particular, Kiwi addresses the problem of huge gene-set interaction networks (which would be impossible to visualize in their complete form in any meaningful way) by reducing the network so that the significant gene-sets and their interactions become apparent.

The inputs to Kiwi are the results from a GSA (for instance, but not limited to, the output from piano) and a gene-set interaction network (e.g. a metabolite-metabolite network). The visualization algorithm is outlined in Figure 7. First, significant gene-sets are selected, based on the results from the GSA. Next, the shortest path length (SPL) is calculated between all pairs of these gene-sets, based on the topology of the gene-set interaction network. Edges are then drawn between significant gene-sets, if they are close enough, i.e. if the SPL is smaller than an arbitrary cutoff. The edge thickness is set in relation to the SPL, so that gene-sets that are close to each other will have a thick edge. The choice of a sensible cutoff can depend on network properties, like the average SPL, or the meaning of the gene-set interaction network. For example, for a highly connected metabolite-metabolite network it would not make much biological sense to connect metabolites that are maybe three or four steps apart, since these numbers of steps

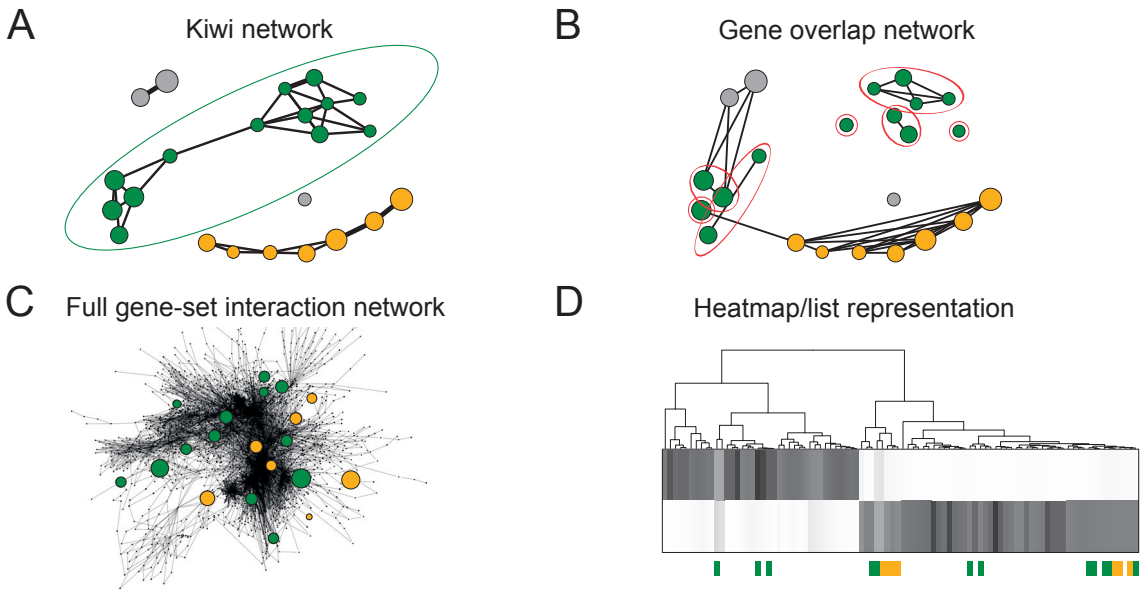

Figure 8. Visualization of GSA results. A) The Kiwi network identifies groups of metabolites with close proximity in the metabolic network (marked green and orange in this example). B) A similar network, but with edges based on gene overlap, does not capture the connection between the green metabolites. C) The full gene-set interaction network is too big to be able to highlight the relevant connections. D) Metabolite gene-sets shown as a list or a heatmap will also not convey the metabolite interactions.

can connect even distant parts of metabolism. On the other hand, for a GO-term network, it could be useful to have a higher cutoff, since GO-terms are connected in a more hierarchy-like structure, allowing for connections between parent and child GO-terms that are several steps away from each other, but biologically related. As a last step (although this is optional) only the best edge or edges are kept for each gene-set node, i.e. the thickest edge (representing the shortest connections), or in case of a tie, the thickest edges. This results in a network where each gene-set is only connected to its closest neighbor. Finally, a force-based layout is employed and nodes are scaled and colored in relation to their significance and general direction of change, respectively.

In **Paper IV** we carried out two case studies that emphasized the advantages of Kiwi. I will not reiterate those analyses here, but use Figure 8 (see figure legend) to conceptually highlight the purpose of Kiwi compared to other visualization approaches (using real data from case study 1 in **Paper IV**). A group of significant reporter metabolites that are connected in the metabolic network were identified using Kiwi, but are not necessarily connected based on gene overlap. Nevertheless, a gene-overlap network can also be useful for the purpose of detecting gene-sets that are driven to be significant based on a similar set of genes. This type of network can be obtained by using piano.

This concludes Part I that described tools that we have developed and that are available for the research community to apply to a vast range of different projects. I have mainly used these tools to enable advanced analysis of skeletal muscle gene transcription in connection to T2D and obesity. This work will be described in Part II.

Part II: Type 2 diabetes

Skeletal muscle and T2D

An overview and introduction to T2D was given in the background section. Here I will briefly elaborate on the specific role of skeletal muscle in the development of the disease. Skeletal muscle is responsible for the majority (around 75-80%) of the required uptake of glucose from the circulation, induced by increased insulin secretion following e.g. a meal (Björnholm and Zierath, 2005; Stump et al., 2006). A majority of this glucose is converted to glycogen (DeFronzo and Tripathy, 2009). Deficiency in skeletal muscle glucose uptake consequently plays an important role for the hyperglycemia connected to T2D, even though this also relies on an incapacity of the beta cells to compensate for muscle insulin resistance. Muscle insulin resistance has been suggested to be one of the primary defects in the development of T2D, present long before the disease itself (DeFronzo and Tripathy, 2009). This conclusion is based on observations of muscle insulin resistance in normal glucose tolerant (NGT) subjects with family history of T2D (Ferrannini et al., 2003; Kashyap et al., 2004; Tripathy et al., 2003; Vaag et al., 1992). Initially, insulin secretion is increased, so that these persons can remain NGT. Eventually however, as insulin resistance gets worse and is accompanied by beta cell deficiency, T2D will evolve. Increased insulin secretion as a response to early development of muscle insulin resistance can also have negative effects. Increased and sustained high levels of insulin may induce insulin resistance as shown by a reduction of glucose uptake and a decreased activity of glycogen synthase with a concordant decrease in glycogen synthesis in skeletal muscle (Iozzo et al., 2001). These observations point to the importance of skeletal muscle insulin resistance and impaired glucose uptake and glycogen production in the progression of T2D. The underlying mechanisms behind insulin resistance in muscle is not yet fully understood.

Insulin resistance in muscle

Studies have shown that insulin resistance is related to defects in the insulin signaling pathway in T2D skeletal muscle, specifically decreased tyrosine phosphorylation of the insulin receptor and insulin receptor substrate 1 (IRS1) and reduced phosphoinositide 3-kinase (PI3K) activity (Bouzakri et al., 2003; Cusi et al., 2000; Krook et al., 2000). Similar effects have been observed in skeletal muscle of non-obese NGT subjects with family history of T2D (Pratipanawatr et al., 2001). The metabolic overload associated with T2D has been implicated in contributing to impaired insulin signaling. A raise in circulating free fatty acid levels has been shown to decrease IRS1-associated PI3K activity and IRS1 tyrosine phosphorylation, and increase activation of PKC theta activity (Dresner et al.,

1999; Griffin et al., 1999). PKC theta has been suggested to be an important mediator of fatty acid-induced insulin resistance in muscle (Kim et al., 2004). Elevated levels of diacylglycerol, acyl-CoA, and ceramide, resulting from incomplete fatty acid oxidation and lipid overload, could activate PKC and lead to serine phosphorylation of IRS1 and consequently impaired insulin signaling (Dresner et al., 1999; Itani et al., 2002; Szendroedi et al., 2012; Yu et al., 2002).

Mitochondrial dysfunction has been implicated in insulin resistance and T2D of skeletal muscle (Szendroedi et al., 2012). Reduced ability to switch between carbohydrate and fatty acid oxidation under insulin stimulation, so called metabolic inflexibility, has been reported in T2D skeletal muscle (Kelley and Mandarino, 2000). Furthermore, both decreased oxidative phosphorylation, connected to lower expression of the transcriptional coactivator PGC1 (Mootha et al., 2003; Patti et al., 2003), and reduced number of mitochondria is associated with T2D muscle (Chomentowski et al., 2010; Morino et al., 2005). It is however unclear if, in the development of insulin resistance, lipid overload leads to mitochondrial dysfunction or if mitochondrial incapacity to oxidize fatty acids precedes the accumulation of intramyocellular fatty acid metabolites, (DeFronzo and Tripathy, 2009).

Apart from disrupted lipid metabolism, branched-chain amino acids may also play a role in the pathogenesis of T2D (Muoio and Newgard, 2008). These metabolites have been found to be elevated in subjects with T2D and could impair insulin signaling by activating mTORC1 and S6K1, leading to serine phosphorylation of IRS1 (Lynch and Adams, 2014).

Genome-scale metabolic models and skeletal muscle T2D

In **Paper II** we reviewed the application of GEMs to study metabolism related to obesity and diabetes. The generic human GEMs Recon 1 and EHMN have been used to identify metabolic signatures from T2D muscle gene expression data (Zelezniak et al., 2010). This analysis highlighted metabolites in the TCA cycle, oxidative phosphorylation, and lipid metabolism, as well as NAD+/NADH and ATP/ADP. Furthermore, the authors report several transcription factors that are members of the CREB, NRF1, and PPAR families, to potentially regulate the differentially expressed genes. Recon 1 has also been used to construct a multi-tissue GEM covering myocytes, adipocytes, hepatocytes, and blood (Bordbar et al., 2011). The model was used as a basis for constructing context-specific GEMs from gene expression data from normal obese and T2D obese subjects. This analysis indicated reduced activity of lactate dehydrogenase and catalase in the myocyte model of T2D subjects.

Recently, a small-scale metabolic model (388 reactions) was used to identify network perturbations that reproduced metabolic properties of insulin resistant muscle (Nogiec et al., 2015). From their simulations the authors report that a dual knockdown of pyruvate dehydrogenase (PDH) and electron-transferrin-flavoprotein dehydrogenase (ETFDH) resulted in a reduction in ATP synthesis, TCA cycle flux, and metabolic flexibility. Experimental validation in form of dual

PDH/ETFDH knockdowns in cultured myocytes, and human transcriptomic and metabolomics data analysis, supported their simulation results.

In **Papers III**, **V**, and **VI**, we have exploited our myocyte-specific metabolic network iMycoyte2419, by using piano and Kiwi to analyze both published and newly generated transcriptome data on muscle T2D, to add pieces to the unsolved puzzle of the underlying mechanisms of insulin resistance and T2D in skeletal muscle. These projects are summarized in the remaining pages of Part II.

A metabolic signature of T2D muscle (Papers III & V)

Transcriptional meta-analysis of skeletal muscle

In **Paper III** (and subsequently discussed and summarized in **Paper V**) our aim was to characterize the effects of T2D on skeletal muscle, in particular focusing on metabolism. We addressed this by analyzing and connecting the results from multiple published datasets of muscle gene expression by performing a transcriptional meta-analysis. Meta-analysis is a general statistical approach that combines the results from similar studies to, through an increased sample size and gain in statistical power, get a better estimate of the effect measured by the individual studies. Through database searches we identified microarray datasets related to T2D and skeletal muscle, which we narrowed down to six consistent studies comparing T2D vs NGT in skeletal muscle at baseline, shown in Figure 9A (Chibalin et al., 2008; Jin et al., 2011; Patti et al., 2003; Pihlajamäki et al., 2011; Sears et al., 2009; van Tienen et al., 2012). We then used a meta-analysis method, designed for microarray data, proposed by Choi et al. (2003) and recommended in a recent review (Ramasamy et al., 2008). This method builds on the work of Hedges and Olkin (1985). Briefly, an effect-size (which is related to the *t*-statistic) and variance is estimated for each gene in each study. The effect-sizes and

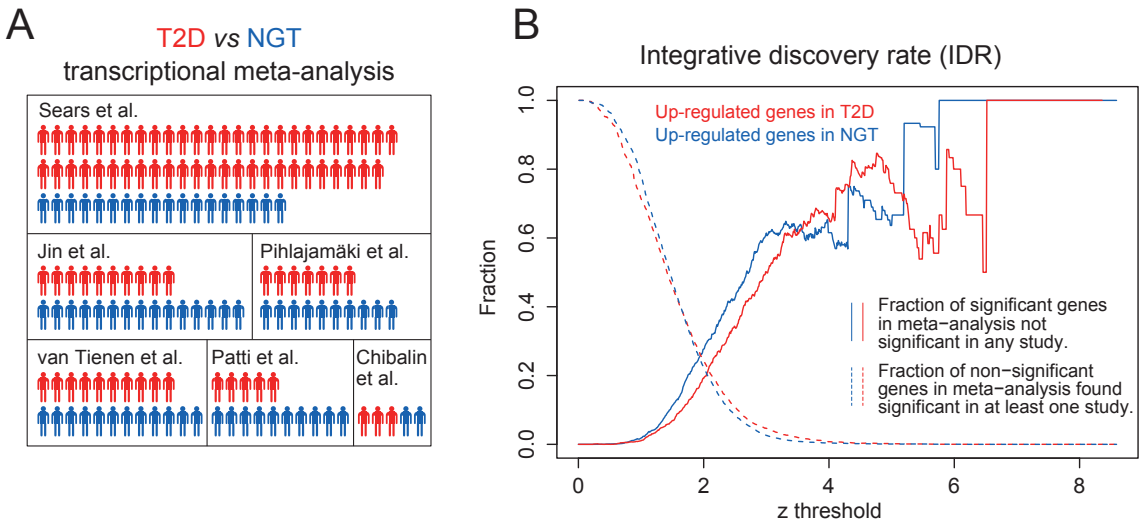

Figure 9. Meta-analysis of T2D vs NGT gene expression. A) The number of T2D and NGT subjects in each of the six studies used in the meta-analysis. B) The integrative discovery rate (at different *z*-score cutoffs), i.e. the fraction of significant genes in the meta-analysis that were not significant in any of the studies.

variances are then pooled across all studies using a random effects model. Finally, a *z*-score is calculated for each gene so that genes with altered expression between T2D and NGT can be identified by having an absolute *z*-score larger than a given threshold. By combining multiple datasets, the statistical power is improved, which can enable the identification of differentially expressed genes that could not be detected in any individual study. Figure 9B shows the integrative discovery rate (IDR) (Choi et al., 2003), i.e. the fraction of significant genes that were detected in our meta-analysis, that were not identified in any of the individual studies, highlighting this benefit of meta-analysis.

Consensus GSA of the meta-analysis results

We wanted to identify metabolic pathways and other biological processes that were affected by transcriptional changes in T2D. To do this, we performed consensus GSA using piano with GO-terms and pathways from iMyocyte2419 as gene-sets, and the meta-analysis *z*-scores as gene-level input. Both these analyses pointed to a transcriptional up-regulation of immune-related processes and down-regulation of genes involved in glycolysis, pyruvate metabolism, TCA cycle, oxidative phosphorylation, respiratory electron transport chain, mitochondrial proteins, beta-oxidation, and branched-chain amino acid (BCAA) metabolism. These results are in line with previous findings (Abdul-Ghani and DeFronzo, 2010; Donath and Shoelson, 2011; Lynch and Adams, 2014; Szendroedi et al., 2012).

In particular, reduced oxidative phosphorylation has been reported several times to be associated with T2D muscle (Mootha et al., 2003; Szendroedi et al., 2012). Other studies however have challenged this and report no change of oxidative phosphorylation and a normal mitochondrial function (Boushel et al., 2007; De Feyter et al., 2008; Frederiksen et al., 2008; Gallagher et al., 2010). Gallagher et al. speculated that the absence of insulin stimulation of subjects before sampling could explain why they did not identify any difference in oxidative phosphorylation. However, none of the studies included in our meta-analysis used stimulated subjects (this was part of our inclusion criteria) and we still detected a down-regulation of oxidative phosphorylation. This result was consistently found also when repeating the meta-analysis while including the Gallagher dataset (which did not detect oxidative phosphorylation) thus establishing down-regulated oxidative phosphorylation as one of the signatures of T2D muscle.

Further on, we also detected down-regulation of less studied gene-sets, including omega-6 fatty acid metabolism, vitamin E metabolism, nucleotide metabolism, and cysteine and methionine metabolism. There also appeared to be a down-regulation of GO-term gene-sets related to RNA-splicing, in line with the findings presented in one of the studies of the meta-analysis (Pihlajamäki et al., 2011).

Reporter metabolites and Kiwi – identifying a metabolic signature

Next, we wanted to exploit the topology of the myocyte metabolic network, iMyocyte2419, and integrate it with the transcriptional data from the six microarray studies, in order to identify a general metabolic signature of T2D in skeletal muscle. We carried out consensus GSA of the meta-analysis results using

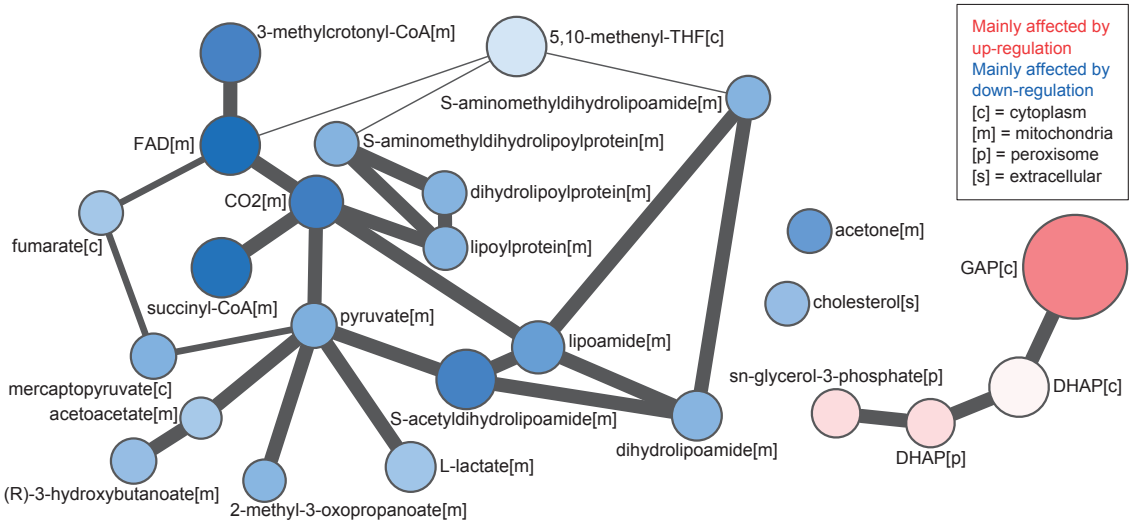

Figure 10. The metabolic signature of T2D in skeletal muscle. Consensus GSA of the meta-analysis results of six microarray datasets coupled with the Kiwi workflow identified a group of metabolites, and their connection in the myocyte metabolic network, that were significantly affected by transcriptional changes in T2D vs NGT. The maximum SPL was set to 3 and the p-value cutoff was set to 0.001.

metabolite gene-sets derived from iMyocyte2419 and passed these results to Kiwi to extract the network of significant metabolites. For this, we used the metabolite-metabolite network derived from iMyocyte2419, were we first had removed high-degree nodes (mainly co-factors) so that the paths connecting metabolites could not go through these metabolic hubs. The final identified metabolic signature, based on transcriptional changes associated with T2D in multiple different studies, is shown in Figure 10. In summary, this signature included metabolites affected by down-regulation, acting in the mitochondria, and involved in mitochondrial import and oxidation of pyruvate (pyruvate, CO_2, S-acetyldihydrolipoamide, lipoamide, and dihydrolipoamide), TCA cycle (fumarate, succinyl-CoA, CO_2, and FAD), BCAA degradation, folate one-carbon metabolism, and lipoylproteins and lipoamides. There was also a separate group of metabolites (GAP, DHAP, and sn-glycerol-3-phosphate) that were affected by transcriptional up-regulation, acting in the branch-point of glycolysis, pentose phosphate pathway, and lipid biosynthesis.

The reporter metabolites 3-methylcrotonyl-CoA, 2-methyl-3-oxopropanoate, dihydrolipoamide, lipoamide, and succinyl-CoA, are all intermediates of BCAA and were associated with transcriptional down-regulation in T2D in the meta-analysis, in line with our pathway and GO-term GSA, and with previous results (Lefort et al., 2010). As stated previously, high levels of plasma BCAA have been observed in connection to T2D and could play a role in the development of insulin resistance by interfering with insulin signaling or by causing mitochondrial dysfunction through accumulation of BCAA metabolites as a consequence of disrupted BCAA catabolism (Lynch and Adams, 2014). Insulin has been reported to lower plasma BCAA concentrations and to decrease levels of 3-methylcrotonyl-CoA (one of the reporter metabolites) in liver (Shin et al., 2014).

Less has been reported regarding the regulation of genes associated with 5,10-methenyl-THF in connection to T2D. This reporter metabolite is an intermediate

in folate one-carbon metabolism and acts as a carbon donor for nucleotide synthesis, methionine synthesis, purine synthesis, and DNA methylation. The regulation of 5,10-methenyl-THF is in line with our pathway GSA where we observed a down-regulation of both methionine and nucleotide metabolism. Two significantly differentially expressed genes are associated with 5,10-methenyl-THF. *MTHFD1* is down-regulated and implies a reduced interconversion between THF intermediates. On the other hand, *FTCD* is up-regulated, possibly indicating a contribution from histidine catabolism into THF metabolism. It has been shown that histidine has positive effects on T2D (Kimura et al., 2013; Lee et al., 2005; Stančáková et al., 2012), and perhaps an increased histidine catabolism in muscle could be linked to the negative effects of T2D.

The metabolic signature can predict individual disease states

We wanted to ensure the relevance of the identified metabolic signature of T2D in skeletal muscle. This subnetwork came out as a result of combining six different datasets, and we wanted to determine if it was prominent enough to be relevant on the level of individual subjects. To do this, we took the most significant genes underlying this network (12 genes with $p<1e-5$), associated with 20 out of the 25 metabolite gene-sets in the network, and asked if these genes alone had the power to predict if a subject was T2D or NGT. We used a random forest classification model to predict the disease state of each subject from its expression levels of the 12 genes, by training the model on the remaining subjects in that dataset (leave-one-out cross validation). For each study we could then plot the true positive rate against the false positive rate for different classification thresholds, known as a receiver operating characteristic (ROC) curve. The area under the ROC curve (AUC) can be used to assess the performance of a binary classification procedure, where an AUC of 1 represents a perfect classification and 0.5 represents a performance no better than random. This procedure was repeated 100 times to give a distribution of 100 AUC scores for each of the six studies. These values are shown in Figure 11 and are high for most datasets (close to perfect classification for two datasets, around 0.8 for three datasets, and around 0.6 for one dataset). To ensure that these results were specific for the metabolic signature and not just a general property of gene expression profiles, we reran the classification scheme 100 times for each study, while using randomly selected genes as classifiers. These AUC scores are also shown in Figure 11, covering a larger range of values but centered on around 0.5. This implies that the classifier genes have the ability to correctly predict the disease state of individual subjects across several studies suggesting that the metabolic signature is a common feature of T2D in skeletal muscle. Nevertheless, the performance was lower in the Patti dataset, pinpointing the complexity and impact of individual variation in connection to this disease.

Here, the point was not to establish a gene-set to be used for predicting disease states (in which case a different training/validation scheme would have been employed), but rather to show that the metabolic signature was relevant for individual phenotypes. Regardless, the 12 genes that we used as classifiers have later also been shown to be able to predict insulin resistance in independent datasets, using random general linear models (Chaudhuri et al., 2015).

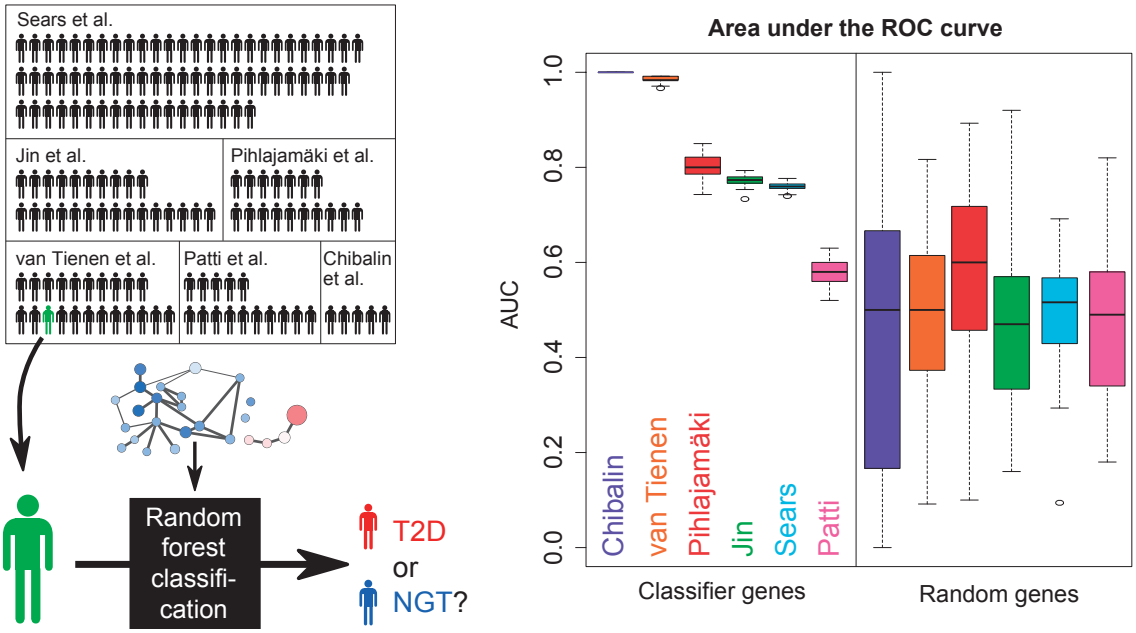

Figure 11. Classification of individual disease states. The metabolic signature has the power to predict the disease state (T2D or NGT) of individual subjects to a high accuracy in most datasets, and perform well compared to randomly selected genes.

Inherent properties of T2D and obese myocytes (Paper VI)

Dissecting the transcriptional profiles of T2D, obesity, and insulin

In **Paper III** we performed a general characterization of T2D skeletal muscle, through a meta-analysis of gene expression studies. The advantage with such an approach is that previously published data can be reused and analyzed in a new light gained from the combination of studies and increase in sample size. A disadvantage is that the experimental designs cannot be controlled and that information about the phenotypic characteristics of the subjects can be limited. Therefore, in **Paper VI**, we designed and carried out an experiment, which is described in Figure 12A, in order to study skeletal myocyte gene expression in a controlled manner. In this project, we used 24 human subjects divided into four groups, each consisting of 3 males and 3 females. Subjects in the first group (T2D) were diagnosed as T2D, but were not obese (BMI of 24.4±2.7). Subjects in the second group (OB) were obese (BMI of 35.2±3.6), but not diabetic. A third group (T2D&OB) consisted of subjects that were both T2D and obese (BMI of 33.2±2.8). Finally, there was a control group consisting of healthy non-obese (BMI of 24.0±0.6) subjects. This experimental setup represents a factorial design with two levels of each of the main factors T2D (either T2D or NGT) and OB (either obese or non-obese). The design allowed us to investigate and isolate the individual influence of T2D and obesity, respectively. This is of particular interest since a majority of T2D patients are also obese (Leibson et al., 2001; Mokdad et al., 2003), making it difficult to distinguish between the effects attributed purely to T2D,

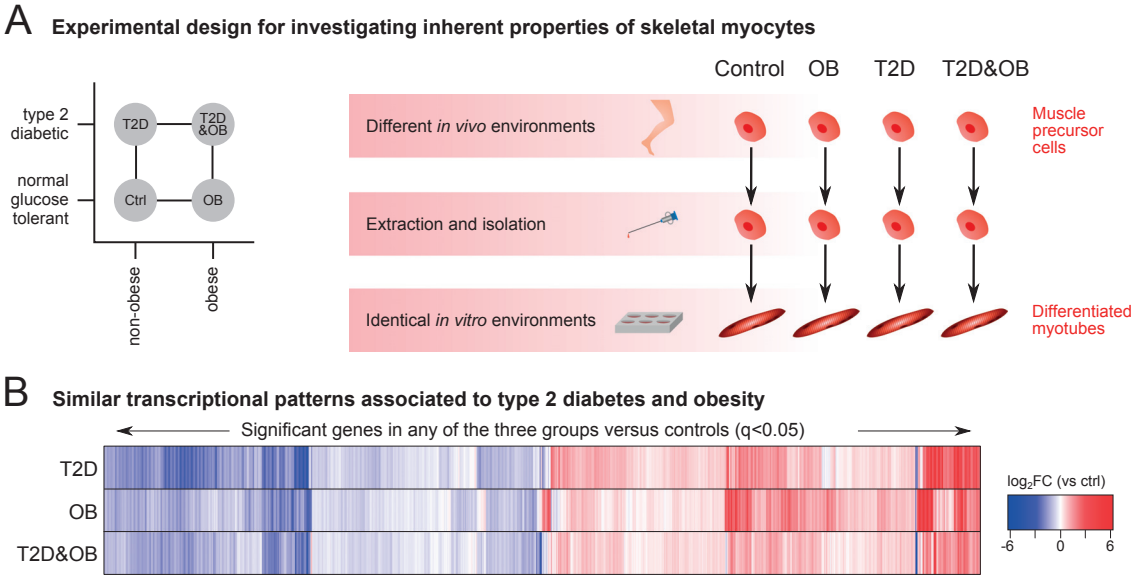

Figure 12. The transcription profiles associated with T2D and obesity are remarkably similar. A) A factorial design was used to study individual effects of T2D and obesity on gene expression. By using *in vitro* myocytes, acquired by culturing and differentiating muscle precursor cells isolated from the subjects, it was possible to identify inherent transcriptional signatures associated with T2D and obesity. B) By comparing the transcriptional changes identified between the controls and the remaining three groups (T2D, OB, and T2D&OB) it was found that the three groups displayed remarkably similar expression patterns.

obesity, or influenced by their interaction. To dissect this information can provide further insight into the mechanisms underlying this complex and multi-factorial disease.

Cells in the human body are influenced by the extracellular environment including circulating hormones, cytokines, and other agents. To avoid confounding the gene expression analysis with the possible influence of these factors and their variation between individuals we decided to use primary differentiated myotubes, an established *in vitro* model to study skeletal muscle. To do this, muscle biopsies were taken from each subject and muscle precursor cells were isolated form the biopsies. These were then cultured *in vitro* and differentiated into myotubes (*in vitro* myocytes). The *in vitro* myocytes were stimulated with insulin and samples for RNA-seq were taken at baseline and at 0.5, 1, and 2 hours after stimulation. As the *in vitro* myocytes were kept in the same controlled environment, any differences that we could identify between the groups would represent specific inherent properties of these myocytes, as a memory of the influence of the *in vivo* phenotype on the corresponding muscle precursor cells. These intrinsic characteristics are relevant to identify since they could represent more robust and hard-wired effects of T2D (or obesity), passed on from muscle precursor cells to differentiated myocytes, probably through genetic or epigenetic mechanisms, and present even without the direct influence from the diabetic (or obese) extracellular environment. As such, these properties could be of considerable value when designing new treatments for T2D, because they are likely hardwired exclusively in the pathological state of the skeletal muscle.

These inherent properties cannot be detected simply by analyzing tissue biopsy samples. Further on, in contrast to tissue samples, the *in vitro* myocyte RNA-seq samples are guaranteed to represent the myocyte specific transcriptome, without contamination from other cell types present in the tissue. The *in vitro* myocytes have been used before to study various aspects of T2D and it has been shown that several properties relevant to the T2D phenotype are retained in these cells (Bouzakri et al., 2003; Broholm et al., 2012; Gaster and Beck-Nielsen, 2004; Gaster et al., 2004; Green et al., 2011; Henry et al., 1996; McIntyre et al., 2004; Scheele et al., 2012; Thompson et al., 1996). In **Paper VI** we thus set out to characterize the inherent properties using genome-wide transcriptome analysis, focusing on the effects associated specifically and individually with T2D and obesity. This was done by using a linear model to analyze differential expression:

$$gene\ expression \sim intercept + T2D + OB + T2D{:}OB + time + gender + age$$

using the limma/voom pipeline (Law et al., 2014; Ritchie et al., 2015). Here the T2D and OB factors, and their interaction (T2D:OB), enable comparisons between the four groups. The influence on gene expression by insulin (time after stimulation), gender, and age is adjusted for and captured by the last factors in the model.

Similar transcriptional signatures associated with both T2D and obesity

We started by assessing the differential expression between controls and each of the three remaining groups (T2D, OB, and T2D&OB). All three groups were distinguished from controls, in terms of differential expression of a large number of genes. We clustered genes that were significant in at least one of the three comparisons, based on their fold changes, and what appeared was a surprising similarity of the changes taking place in all three groups compared to controls (Figure 12B). The pairwise Pearson correlation of the fold changes of all genes was also high between the three groups (0.67, 0.81, and 0.65). This implies that myocytes originating from muscle precursor cells from a T2D but non-obese person show very similar characteristics to those of myocytes originating from muscle precursor cells from an obese but NGT person. Apparently, muscle precursor cells independently influenced by T2D or obesity, promote similar transcriptional signatures in the myocytes they give rise to, probably through epigenetic mechanisms. Moreover, the induction of a similar transcriptional signature cannot be attributed to potential biases from the extracellular environment, insulin, age, and gender. It is possible that obesity induces an inherent transcriptional response that, being similar to that of T2D, provides a foundation to develop the disease under the influence of additional deleterious factors that can be present in the *in vivo* environment.

In line with the observed similarities between the three groups, we did not identify any significant difference in the transcriptional changes associated with insulin stimulation. This was concluded by introducing different combinations of interaction terms in the linear model, taking into account interaction between T2D, OB, and insulin (time). In none of the cases were these interaction terms significantly contributing to explaining the expression level of any single gene.

However, there might still be differences, others than those noted on the transcriptional level, present in the response to insulin signaling, which we did not measure.

An effect of the identification of similar transcriptional profiles associated with the T2D, OB, and T2D&OB groups is that there is significant interaction between T2D and OB. In other words, assessing the effect of T2D will be dependent on the level of obesity. For instance, fewer changes in gene expression were observed when comparing T2D vs NGT in the obese case compared with in the non-obese case. This, along with the evidence for different etiologies for T2D in obese and non-obese subjects (Arner et al., 1991), is important to take into account when designing or interpreting research that investigates the effect of T2D, and the BMIs of the subjects have to be carefully considered.

Exploring genetic and epigenetic influences on the transcriptional profiles

The similar transcriptional profiles of T2D and obesity are probably mediated through some combination of genetic and epigenetic effects, since the identified characteristics of the *in vitro* myocytes reflect properties passed on from the muscle precursor cells from the subjects in the different groups. Therefore, we wanted to exploit the RNA-seq data to try to uncover evidence for such influences and generate hypotheses for possible mechanisms behind these strikingly similar transcriptional changes.

At the genetic level we explored the association between SNPs and expression changes. To avoid drawing false conclusions from our relatively small dataset, we focused on known expressed quantitative trait loci (eQTL), which represents a significant association between a specific SNP and an expression change of a specific gene. Using the NCBI GTEx-eQTL browser (Lonsdale et al., 2013), which integrates eQTL information with phenotypic traits from GWAS studies, we could acquire eQTLs for T2D and obesity. This resulted in 186 eQTLs affecting the expression of 30 unique genes. Next, we compared these eQTL genes with our gene expression data and could thereby identify 7 genes showing significant differential expression between at least one of the three groups (T2D, OB, or T2D&OB) and controls. According to the database, the expression of these 7 genes were affected by 67 SNPs in total. To close the circle, we needed to validate whether any of the 67 SNPs were present in the subjects in our study. Using the RNA-seq data, which of course can only be used to infer the DNA-sequence of transcribed regions, we were able to determine (for at least four subjects in a group) the nucleotide sequences for the regions surrounding eight of the 67 SNPs. The results indicated some variation in nucleotide frequencies between the different groups and it is possible that some of these genetic variations explain part of the transcriptional signatures that we observed in connection to T2D and obesity. In particular we identified the transcription factors *PPARG* and *JAZF1*, which have the possibility to influence the expression of multiple other genes, as being both eQTL genes and differentially expressed in our data.

On the epigenetic level we explored the influence on transcription from histone modifications. To do this we performed GSA using piano and a histone

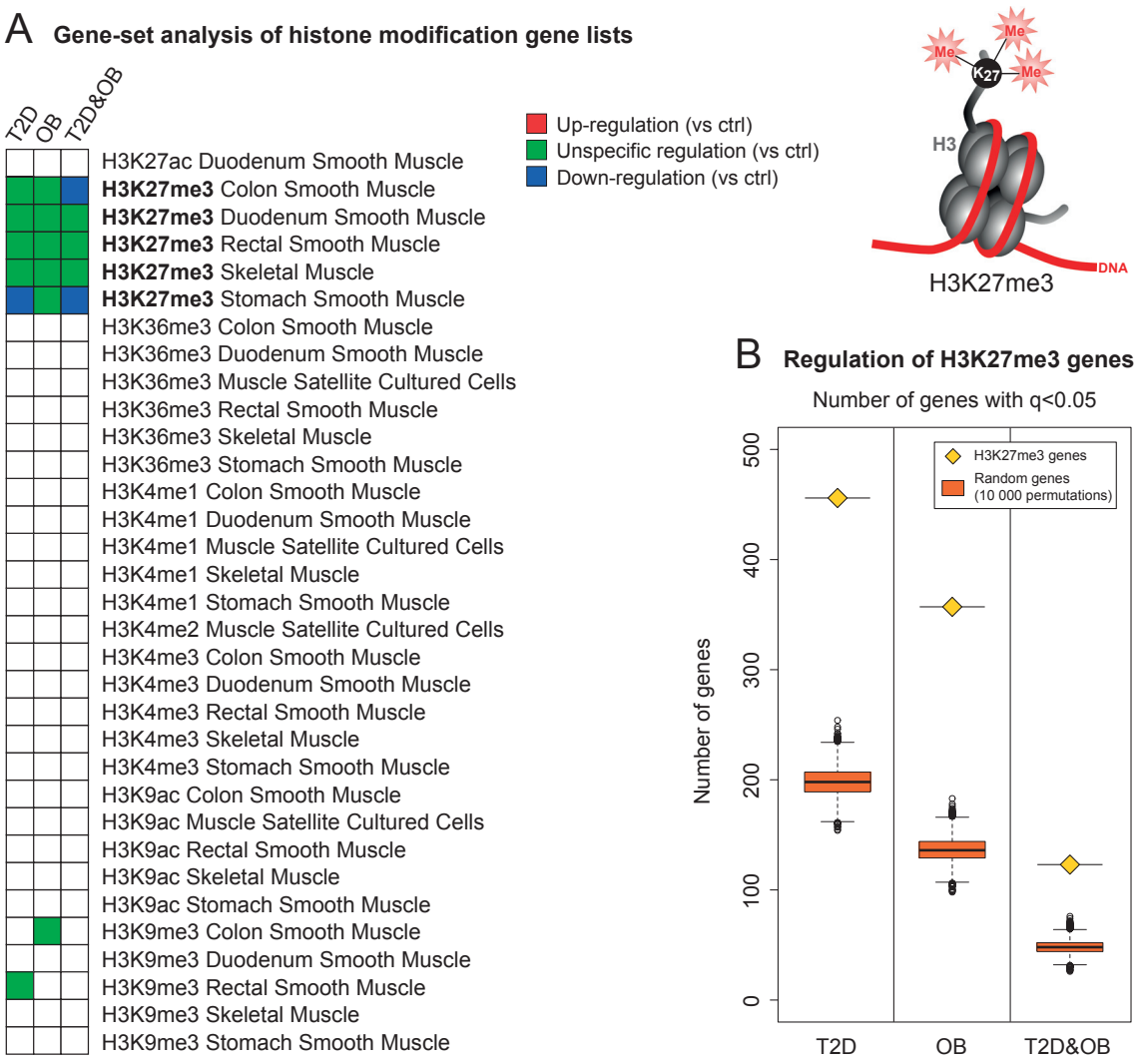

Figure 13. The transcription profiles may be influenced by changes to the H3K27me3 mark. A) A heatmap of the investigated histone modification gene-sets, showing a consistent significance of the H3K27me3 gene-set in all three groups. B) A boxplot showing the number of significant H3K27me3 genes compared to the distribution of number of significant genes from random permutation.

modification gene-set collection from the Epigenomics Roadmap project available through the Enrichr website (Chen et al., 2013). These gene-sets are based on ChIP-seq data, identifying for each gene if a certain histone modifications is present, across various cell types. We filtered this data for muscle related cells. In this project we used a specific GSA approach where we performed two parallel GSA runs, one with gene-level *q*-values (FDR adjusted *p*-values) as input, and one with gene-level fold changes as input. We then performed consensus GSA of these two runs to identify significant gene-sets based both on the statistical and biological level of change of the member genes. Through this analysis we identified one specific histone methylation mark, H3K27me3, as particularly interesting, as it was highly significant for all three groups compared to controls (Figure 13A). These results were consistent with a validation analysis of a different histone modification gene-set collection based on data from the ENCODE project (also downloaded from Enrichr). Figure 13B shows the number of significant genes belonging to the H3K27me3 gene-sets and compares this with the number of expected significant

genes, in a group of genes of that size, by randomly permuting the gene labels 10,000 times. As can be seen it is quite unlikely to observe these results by chance, indicating that there is a possibility that changes to this specific histone methylation mark in connection to T2D and obesity could have an influence on part of the similar transcriptional signatures that we observed in the T2D, OB, and T2D&OB groups compared to controls.

The H3K27me3 gene-sets show a general unspecific direction of regulation, i.e. a mix of up- and down-regulated genes. There is a subtle bias towards down-regulation however, so there is at least a subset of these genes that are down-regulated. The presence of H3K27me3 is known to repress the transcription of genes involved in development and differentiation (Boyer et al., 2006; Lee et al., 2006; Sen et al., 2008), and the demethylation of this histone mark is observed during myogenesis (Seenundun et al., 2010). In line with this, we found significant down-regulation, in the T2D group, of the myogenic marker genes *MYOD1*, *MYOG*, *TNNI1*, *MYH2*, and *MEF2C* (the last one also down-regulated in the OB and T2D&OB groups).

Characterizing the inherent transcriptional signatures

To characterize the functions of the expression patterns in the T2D, OB, and T2D&OB groups we performed GSA, using the same approach as for the histone modification gene-sets, i.e. basing the results on both the gene-level q-values and fold changes. We evaluated GO-terms, metabolic pathways from iMyocyte2419, and so-called hallmark gene-sets from the Molecular Signatures Database (Liberzon et al., 2015). The hallmark gene-sets are computationally and manually refined from a large set of founder sets and have been validated using gene expression data relevant to the phenotype or property that each gene-set represents. The GSA results are shown in Figure 14.

In line with the finding of influence from H3K27me3 on the transcriptional profiles, we identified several gene-sets in the hallmark and GO-term GSAs, affected by down-regulation in the three groups and related to development and differentiation. These gene-sets included myogenesis, muscle organ development, skeletal muscle tissue development, and muscle contraction. In fact, most of the down-regulated GO-terms were related to muscle function and structure. Collectively, this points to that a portion of the inherent transcriptional profiles associated with T2D and obesity includes down-regulation of genes connected to muscle development and function, possibly mediated through the H3K27me3 mark. On the other hand, we also observed significant up-regulation, in the three groups compared to controls, of gene-sets involved in the function and structure of the extracellular matrix (ECM), e.g. epithelial-mesenchymal transition, extracellular matrix, heparin binding, collagen binding, glycosaminoglycan binding, and integrin binding. This was interesting since ECM is important for muscle maintenance and development, and is required for myotube formation (Gillies and Lieber, 2011; Melo et al., 1996; Osses and Brandan, 2002; Stern et al.,

Figure 14. Gene-set analysis results. A) Significant GO-terms identified in the comparisons of the T2D, OB, and T2D&OB groups with controls, respectively. B) Significant hallmark gene-sets for the same comparisons as the GO-terms.

2009; Velleman et al., 2012). In what way these two opposite results are connected is difficult to interpret and requires additional research.

Another theme that we identified through the GSA was regulation of immune related functions and inflammation. These showed up-regulation in the OB group and unspecific regulation in the T2D group. This is in line with the known presence of chronic low-grade inflammation and activation of the immune system in association with T2D and obesity (Donath and Shoelson, 2011; Esser et al., 2014; Shoelson et al., 2006). Increased infiltration of macrophages and expression of

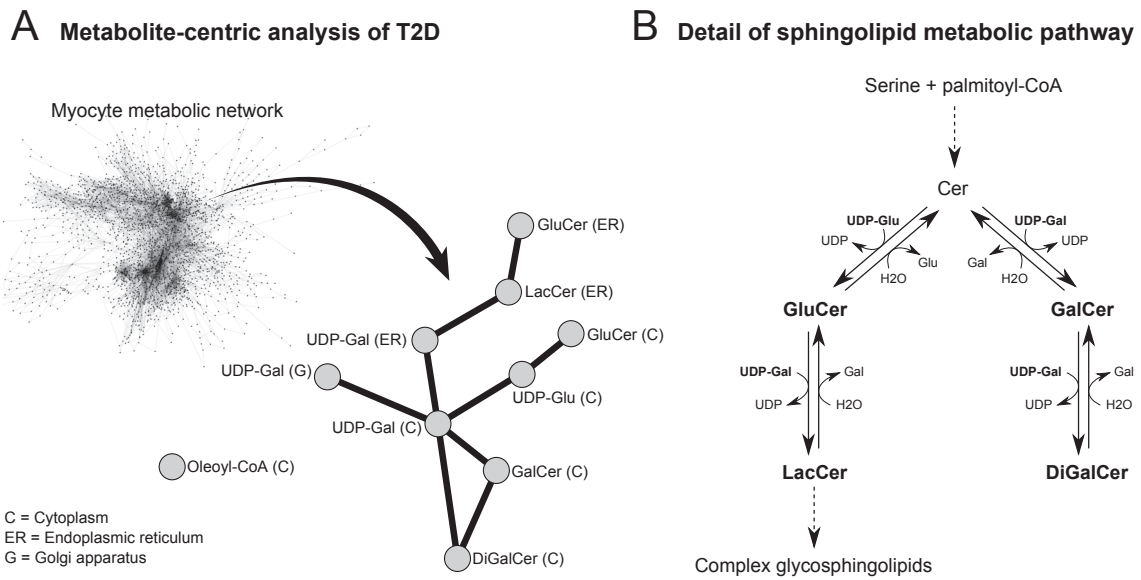

Figure 15. A sphingolipid subnetwork is regulated in T2D. A) The connected subnetwork of reporter metabolites that resulted from GSA of T2D gene expression. The myocyte GEM iMyocyte2419 was used as a basis for the topological analysis. B) The identified reporter metabolites are involved in a specific part of sphingolipid metabolism. Ceramide (Cer), glucosylceramide (GluCer), lactosylceramide (LacCer), galactosylceramide (GalCer), digalactosylceramide (DiGalCer), UDP-galactose (UDP-Gal), UDP-glucose (UDP-Glu).

inflammatory macrophage genes has been observed in T2D muscle (Fink et al., 2013; Weisberg et al., 2003). It was however surprising that we identified up-regulation of inflammation in myocyte specific transcription data, without the presence of immune cells.

T2D up-regulates a metabolite subnetwork involved in sphingolipid metabolism

In the previously described functional characterization we did not find any significant metabolic pathway gene-sets. Not to be constrained by classical pathway definitions we decided to complement this analysis by performing reporter metabolite analysis and network visualization, using piano and Kiwi. We used the network topology of iMyocyte2419 to extract the metabolite-gene associations required for the reporter metabolite GSA and the metabolite-metabolite network required for the visualization. We identified a group of significant reporter metabolites for the T2D group (none were found for the other groups) that were affected by transcriptional up-regulation. When we applied the Kiwi algorithm to these results it turned out that they belonged to a tightly connected subnetwork (Figure 15A) representing a specific part of sphingolipid metabolism (Figure 15B). These reactions involve the conversion of ceramide to galactosylceramide and digalactosylceramide, as well as to glucosylceramide and lactosylceramide, which can be converted to more complex glycosphingolipids. Sphingolipids function as both structural molecules and are involved in cell signaling. Sphingolipids have been implicated in T2D before, and an increase in sphingolipid production, plasma glycosphingolipid levels, and muscle ceramide levels have been observed in association with the disease (Adams et al., 2004; Gault et al., 2010; Haus et al., 2009; Russo et al., 2013; Summers and Nelson, 2005). Inhibition of the reaction from

ceramide to glucosylceramide (shown in Figure 15B) improved insulin sensitivity in animal models of T2D (Zhao et al., 2007). It has also been reported that ceramide levels in muscle correlate with insulin resistance (Amati et al., 2011; Straczkowski et al., 2007; Straczkowski et al., 2004). In summary, by using the network topology of the myocyte GEM and GSA coupled with the Kiwi algorithm we were able to identify up-regulation of a part of sphingolipid metabolism that is implicated in T2D and could not be detected by classical pathway analysis. Since this was detected as an inherent property of T2D myocytes, without the influence from circulating levels of insulin or plasma sphingolipids, it highlights the importance of sphingolipids in the development and progression of the disease in muscle.

Conclusions and outlook

With this thesis I had the ambition to contribute to the understanding of the manifestation of T2D in skeletal muscle, and identify possible mechanisms underlying the pathogenesis and development of insulin resistance. Along the way this also materialized in the development of several tools. The main driving force behind this was their immediate benefit for my own research, but I also believed that, if there was a chance that our algorithms could be of value to other researchers, it was worth the additional effort involved with bringing them to the level of user-friendly, documented software. Piano (**Paper I**) has improved the GSA workflow in several ways. It is a platform for running multiple different GSA algorithms in the same setup, the directionality classification improves the interpretation of the GSA results, and the consensus scoring approach enables a flexible and robust way to identify significant gene-sets. Kiwi (**Paper IV**) connects seamlessly to the output from piano and enables the integration of GSA with network analysis. In particular, the piano-Kiwi workflow has allowed us to exploit the myocyte metabolic network topology and integrate it with transcriptome data. These analysis steps provide a powerful way to quickly go from high-dimensional data and big networks, to the identification of implicated metabolite subnetworks, independent of classical pathway definitions. By using appropriate input data and relevant metabolic networks, this approach can easily be adopted and applied to other tissues and other diseases than T2D.

In the end of Part I, I discussed some considerations that I have found worthwhile taking into account when running GSA. I want to follow up here, with my thoughts on what should be improved next. There are certainly many different GSA algorithms and tools available. The handful that we tested in **Paper I** turned out to be quite consistent. There is of course always room for improvements, but I do not think that algorithm development for GSA is a main concern. Rather I would focus on what I believe is the weakest link in the GSA chain at the moment, namely the gene-set collections. Targeted efforts to improve the quality of gene-sets, in terms of valid gene members, reduced overlap, and broad unbiased coverage, would greatly benefit GSA and the interpretation of omics data. (What I mean with unbiased coverage is that databases from which gene-sets can be acquired typically rely on submissions from the research community and are therefore biased to what is being actively researched.) The improvement of gene-reaction associations in metabolic networks and of the topology of cell type- or context-specific GEMs, will also lead to better predictions of metabolites implicated in disease, using the piano-Kiwi workflow. The recently developed hallmark gene-set collection (Liberzon et al., 2015) that we used in **Paper VI** is an excellent example of improvement of gene-set quality, and it will be great to see a continued expansion of this collection in the future.

Research is a collaborative effort, directly or indirectly. I have had the opportunity to work in projects with several excellent scientists that have contributed with their specific expertise. But even a project carried out by a single individual is usually spurred by previous research and the results will hopefully benefit others in the field. Even though science is competitive, I have had very positive experience of cooperation and assistance from other researchers at conferences and that I have pursued with questions. It is reassuring to know that others, from their own unique angles, are attempting to tackle the unanswered questions regarding T2D. Our approach to study this complex disease is through systems biology, using high-throughput data, network modeling, and methods that rapidly process this information and distills it down to the essential components of interest. As mentioned earlier, this approach often represents data driven hypothesis generation, in contrast to hypothesis driven data generation (the former based on general questions like: "what are the differences between T2D and NGT?", whereas the latter typically involves more precise hypotheses and targeted experiments). Indeed we do have experimental data and our results indicate what genes are transcriptionally changed in association with T2D, but most of our conclusions about functions and mechanisms are inferred from transcriptional patterns. To be critical about my own work, our T2D studies lack experimental validation. However, this does not diminish the power of systems biology studies, because they make it possible to efficiently narrow down high-throughput data leading to the identification of likely targets, functions, or mechanisms, involved in the development of T2D. This information, together with results from other studies, is continuously assembled, piece by piece, as collective evidence, useful for us and others in the pursuit of learning more about T2D.

So, what did we learn about T2D from our studies? In **Paper III** we reconstructed a comprehensive myocyte-specific GEM using both transcriptome and proteome data as experimental evidence for the presence of reactions. Next, we analyzed gene expression profiles from 153 subjects, across 6 different studies, to establish a metabolic signature of skeletal muscle T2D. We identified a subnetwork of metabolites, mainly affected by down-regulation, involved in processes like mitochondrial oxidative metabolism, BCAA catabolism, and THF metabolism. These results were consistent with pathway and GO-term GSA, and it was possible to use the expression of the underlying genes to predict the disease state of individual subjects. In summary, this study provided holistic insight into the metabolic state of T2D muscle. In **Paper VI** we explored the inherent properties of skeletal muscle in association with T2D and obesity, by using *in vitro* myocytes from 24 subjects. We found a remarkable similarity between the transcriptional profiles of myocytes originating from T2D, OB, and T2D&OB subjects. It is possible that part of these transcriptional patterns are explained by genetic variations. We also identified a likely epigenetic candidate, H3K27me3, which could mediate the inherent transcription as a memory of the original *in vivo* phenotype. Characterizing the transcriptional changes, we found that myogenesis was dysregulated and muscle function was down-regulated in connection to T2D and obesity, whereas inflammation was up-regulated. We also identified an up-regulated metabolite subnetwork involved in sphingolipid metabolism. In

summary, this study provided a systematic characterization of the individual intrinsic effects of T2D and obesity in skeletal myocytes. Skeletal muscle insulin resistance is an important factor in the development of T2D that may appear long before elevation of blood glucose levels and diagnosis of the disease (DeFronzo and Tripathy, 2009). As such, deleterious changes in skeletal myocytes could be an early sign for the risk of developing T2D. The reporter metabolites identified in **Paper III** and **Paper VI**, through the piano-Kiwi workflow, could therefore be valuable biomarkers for the early detection of subjects at risk. If causal links can be proven, they could also represent potential drug targets.

How do the results from our two studies on T2D compare? Both studies look at T2D vs NGT in skeletal muscle. However, the results in **Paper III** are based on skeletal muscle tissue biopsies, whereas the results in **Paper VI** are from *in vitro* myocytes. Furthermore, we learned from the factorial design of T2D and obesity that BMI plays an important role in defining the differences between T2D and NGT. Indeed, the subject BMIs varied within and between the different datasets used in our meta-analysis. Nevertheless, one would expect some common biological themes to show up. In both studies we identified transcriptional upregulation, in T2D, of genes involved in immune- and inflammation-related processes. A chronic low-grade inflammation and immune system activation is associated with T2D (Donath and Shoelson, 2011) and macrophage infiltration has been observed in muscle of T2D subjects (Fink et al., 2013). Nevertheless, it was particularly interesting to find that these patterns also were inherently present in the transcriptional regulation of myocytes (without external influence, from e.g. the immune system). Further on, we found up-regulation of genes associated with the ECM, in the T2D *in vitro* myocytes. We cannot exclude that this observation is influenced by the fact that these cells are cultured. However, while revisiting the GSA results from the meta-analysis there was also an up-regulation in skeletal muscle tissue, of the GO-terms: regulation of cell shape, cell junction, basal part of cell, apical plasma membrane, and cell morphogenesis. If these processes point at the same mechanism as up-regulation of ECM, is however difficult to deduce. In the meta-analysis, we identified down-regulation of oxidative metabolism, specifically glycolysis, beta-oxidation, TCA cycle, and oxidative phosphorylation. We could not find any statistically significant changes of any metabolic pathways between the different *in vitro* myocytes, which could point to that these metabolic effects are context dependent, i.e. apparent when the myocytes are present in the tissue of the host but not inherently associated with the myocytes. We did however identify down-regulation of a metabolite subnetwork in the sphingolipid metabolism pathway. These results together point at reduced mitochondrial capacity and potential deleterious effects from metabolic overload in T2D, and are in line with previous results, including a study using a similar approach to ours (Zelezniak et al., 2010). Zelezniak et al. also detected 3-methylcrotonyl-CoA, part of BCAA catabolism, as a reporter metabolite in T2D subjects, which we also did in our meta-analysis. However, we did not see any effect on BCAA metabolism in the *in vitro* myocytes. One reason why we in **Paper VI** did not observe some of the changes that we did in the meta-analysis, or that have been reported in other studies, like e.g. down-regulation of BCAA metabolism or oxidative

phosphorylation, could be that these processes are not inherent to T2D myocytes, but rather a response to the diabetic extracellular environment, as mentioned earlier. This also highlights the strength of the *in vitro* myocyte model, i.e. that it enables us to delineate the hard-wired effects of T2D on skeletal muscle.

An apparent challenge is the variation of results from different studies using different human cohorts. Differences in data quality and sample size can only explain some of the occasionally low overlap of results between studies. In our meta-analysis in **Paper III**, we initially selected eight studies that compared muscle gene expression between T2D and NGT subjects. Nevertheless, two of the studies had a negative correlation with the remaining six. Even after careful examination of the subject characteristics data, microarray type, muscle type, or potential contamination of immune cells, we failed to find any explanation for this irregular pattern. (These two studies were not included in the final meta-analysis, see details in **Paper III**). Furthermore, even the top differentially expressed genes from the meta-analysis were not always consistent across studies, i.e. a gene could for instance be up-regulated in most of the groups, but down-regulated in one of the groups. Bigger human cohorts and systematic reporting of detailed subject characteristics will improve future studies of T2D, in particular when reanalyzing accumulated datasets. Systems biology and bioinformatics approaches can then be used to stratify subjects and patients in order to elucidate the complex heterogeneity of T2D.

Research is a collaborative effort. As I said in the background, I believe that holistic systems level research needs to be connected to molecular biology targeted experiments to eventually reach a full understanding of the development and potential treatment of T2D. We have provided some useful systems biology tools for the analysis of disease-related high-throughput data, and have identified promising mechanisms and targets connected to the pathogenesis of T2D. The next steps would be to validate these in separate experiments and determine whether they are a cause or consequence of the disease. I hope that our work, that I have presented here, will contribute to the collective knowledge and thereby make an impact on the quest to understand and cure T2D.

Acknowledgments

First and foremost, I would like express my sincerest thanks to my supervisor Jens Nielsen. I feel very lucky that you believed in me and took me on as a student and it has always been a pleasure to work with you. How you have managed to be closely involved in so many projects and still always been available for feedback, discussions and encouragements whenever I have needed it, is a mystery to me, but something I appreciate a lot. It is difficult to find anyone with a more positive and inspiring attitude than you. You have built an excellent multinational research environment, which together with your vast knowledge and experience has taught me so much about research, science and life. Apart from the scientific side of work I would also like to thank you for your generosity and engagement in social activities like Christmas mingles, ski trips, group outings and much more!

Intawat Nookaew, your happiness and friendliness was one of my first experiences of the sysbio group. You made a great impression on me while being supervisor for my master's thesis, and when I learnt that you would be my PhD co-supervisor I knew I would be in good hands. Your great ideas, support, tricks and tips, and the joy you spread around you are a few of the things I missed when you left the group.

This work would not have been possible without the contribution and support from my collaborators and co-authors, to whom I am very grateful: Camilla Scheele, Christa Broholm, Bente Klarlund Pedersen, Francesco Gatto, Adil Mardinoglu, Mathias Uhlén, Caroline Kampf, Anna Asplund, and Maria Pedersen. A special thanks also to Tobias Österlund, Rasmus Ågren, Francesco Gatto (who has read and commented on in principle all my manuscripts, always making them a lot better), Adil Mardinoglu, Subazini Thankaswamy, Erik Kristiansson, Rebecka Jörnsten, and Rahul Kumar for comments and assistance that were crucial for the completion of our publications. I have also had many other interesting scientific discussions and learnt a lot from Dina Petranovic, Ivan Mijakovic, Avlant Nilsson, Fredrik Karlsson, Tobias Österlund, Rasmus Ågren, Adil Mardinoglu, Natapol Pornputtapong, Saeed Shoiae, Kaisa Thorell, Amir Feizi, and Francesco Gatto.

I would also like to acknowledge the seamless and professional administrative support that I have received from Erica Dahlin, Martina Butorac, Anna Bolling, Josefin Jäwert, and Helena Janveden. Furthermore, a big thanks to Marie Nordqvist and Ximena Rozo for your patience and help during my few activities in the wet lab, and to Shaghayegh Hosseini for all the support and management around the dry lab infrastructure.

I have had the privilege to spend my days in a great workplace environment. The sysbio group is constantly changing, with frequent farewells and new welcomes, but the sysbio social culture remains, and has, as long as I have known it, constantly made it fun going to work. This was one of the reasons I was so happy I could stay after completing my master's thesis, and I am excited that I am allowed to stay in close affiliation with the group for a while more! I have had many entertaining fika breaks, lunches, office discussions, after works, parties and other encounters (including also indbio and math people) with Intawat, Dina, Christer, Joakim,

Calle, Rahul, Sakda, Verena, Martina, Gatto, Amir, Petri, Florian, Anastasia, Alexandra, Michi, Ed, Julle, Jens C, José, Clara, Boyang, Min, Sylvain, Raphael, Magnus, Avlant, Pouyan, Benjamín, Stefan, Marie, Shaq, Dimitra, Adil, Rasmus, Tobias, Fredrik K, Christoph, Natapol, Bouke, Saeed, Kaisa, Elias, Malin, Julia, Ximena, Eugene, Andreas, Nina, Ausra, Emma, Fredrik B, Anna J, Viktor, Johan and many, many more.

Amir and Gatto, we all started around the same time, although you guys managed to finish before me! We have had many adventures, exploring the world, traveling to courses and conferences together, and having intriguing conversations over wine and dinner here in Gothenburg. It has been awesome to have had you guys around all of my time as a PhD student, being deeply involved, both scientifically and socially. I have countless good memories from our hangouts, and I am sure there will be more to come!

A few people have made a particular impact on my life in and outside sysbio: Gatto, Amir, Petri, Florian, Anastasia, Alex, Michi, Ed, Julle and JC. You have been around for skiing, parties, squash, indoor climbing, outdoor bouldering, travel adventures, parties, after works, dinners, running, Liseberg, ice creams, parties, dancing, cocktail evenings, music festivals, choir performances, opera, surprise party – the list can be made long. You mean a lot and I hope we will continue to have crazy fun for a long time onwards! Remember: It is always nice to be hugged from behind.

I am fortunate to have many friends that are important to me outside of sysbio, but I will not attempt to list everyone here, because the list would be too long. I hope that you know who you are! I want to, however, mention a few friends who have encouraged my doctoral studies along the way. Eric and Anders, you have persistently shown great interest in every new publication I have had, without probably understanding a single word. Thanks for the entertaining weekend breaks from Chalmers, in Stockholm and Malmö – All in! Elin, David, Emma, we have been following similar (but different) paths for a long while. I have really enjoyed our “weekly” lunches, discussing PhD-student life and other topics.

Finally, my last paragraph in this thesis is dedicated to the most important people in my life, my family. Isa, Mamma, Pappa, Lukas, Marika, and my extended family, you are always there and you remind me of what is important in life.

References

Abdul-Ghani, M.A., and DeFronzo, R.A. (2010). Pathogenesis of insulin resistance in skeletal muscle. J. Biomed. Biotechnol. 2010, 476279.

Abel, E.D., Peroni, O., Kim, J.K., Kim, Y.-B., Boss, O., Hadro, E., . . . and Kahn, B.B. (2001). Adipose-selective targeting of the GLUT4 gene impairs insulin action in muscle and liver. Nature 409, 729-733.

Ackermann, M., and Strimmer, K. (2009). A general modular framework for gene set enrichment analysis. BMC Bioinf. 10, 47.

Adams, J.M., Pratipanawatr, T., Berria, R., Wang, E., DeFronzo, R.A., Sullards, M.C., and Mandarino, L.J. (2004). Ceramide Content Is Increased in Skeletal Muscle From Obese Insulin-Resistant Humans. Diabetes 53, 25-31.

Agren, R., Bordel, S., Mardinoglu, A., Pornputtapong, N., Nookaew, I., and Nielsen, J. (2012). Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. PLoS Comput. Biol. 8, e1002518.

Agren, R., Mardinoglu, A., Asplund, A., Kampf, C., Uhlen, M., and Nielsen, J. (2014). Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. Mol. Syst. Biol. 10.

Alexeyenko, A., Lee, W., Pernemalm, M., Guegan, J., Dessen, P., Lazar, V., . . . and Pawitan, Y. (2012). Network enrichment analysis: extension of gene-set enrichment analysis to gene networks. BMC Bioinf. 13, 226.

Almdal, T., Scharling, H., Jensen, J., and Vestergaard, H. (2004). The independent effect of type 2 diabetes mellitus on ischemic heart disease, stroke, and death: A population-based study of 13 000 men and women with 20 years of follow-up. Archives of Internal Medicine 164, 1422-1426.

Amati, F., Dubé, J.J., Alvarez-Carnero, E., Edreira, M.M., Chomentowski, P., Coen, P.M., . . . and Goodpaster, B.H. (2011). Skeletal Muscle Triglycerides, Diacylglycerols, and Ceramides in Insulin Resistance: Another Paradox in Endurance-Trained Athletes? Diabetes 60, 2588-2597.

Arner, P., Pollare, T., and Lithell, H. (1991). Different aetiologies of Type 2 (non-insulin-dependent) diabetes mellitus in obese and non-obese subjects. Diabetologia 34, 483-487.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., . . . and Harris, M.A. (2000). Gene Ontology: tool for the unification of biology. Nat. Genet. 25, 25.

Aspuria, P.-J., Lunt, S., Varemo, L., Vergnes, L., Gozo, M., Beach, J., . . . and Orsulic, S. (2014). Succinate dehydrogenase inhibition leads to epithelial-mesenchymal transition and reprogrammed carbon metabolism. Cancer & Metabolism 2, 21.

Banting, F.G., Best, C.H., Collip, J.B., Campbell, W.R., and Fletcher, A.A. (1922). Pancreatic Extracts in the Treatment of Diabetes Mellitus. Canadian Medical Association Journal 12, 141-146.

Barabasi, A.-L., Gulbahce, N., and Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. Nat Rev Genet 12, 56-68.

Becker, S.A., and Palsson, B.O. (2008). Context-specific metabolic networks are consistent with experiments. PLoS Comput. Biol. 4, e1000082.

Ben-Ami, H., Nagachandran, P., Mendelson, A., and Edoute, Y. (1999). Drug-induced hypoglycemic coma in 102 diabetic patients. Archives of Internal Medicine 159, 281-284.

Billings, L.K., and Florez, J.C. (2010). The genetics of type 2 diabetes: what have we learned from GWAS? Ann. N.Y. Acad. Sci. 1212, 59-77.

Björnholm, M., and Zierath, J.R. (2005). Insulin signal transduction in human skeletal muscle: identifying the defects in Type II diabetes. Biochem. Soc. Trans. 33, 354-357.

Bordbar, A., Feist, A., Usaite-Black, R., Woodcock, J., Palsson, B., and Famili, I. (2011). A multi-tissue type genome-scale metabolic network for analysis of whole-body systems physiology. BMC Syst. Biol. 5, 180.

Bordbar, A., and Palsson, B.O. (2012). Using the reconstructed genome-scale human metabolic network to study physiology and pathology. J. Intern. Med. 271, 131-141.

Boushel, R., Gnaiger, E., Schjerling, P., Skovbro, M., Kraunsøe, R., and Dela, F. (2007). Patients with type 2 diabetes have normal mitochondrial function in skeletal muscle. Diabetologia 50, 790-796.

Bouzakri, K., Roques, M., Gual, P., Espinosa, S., Guebre-Egziabher, F., Riou, J.-P., . . . and Vidal, H. (2003). Reduced Activation of Phosphatidylinositol-3 Kinase and Increased Serine 636 Phosphorylation of Insulin Receptor Substrate-1 in Primary Culture of Skeletal Muscle Cells From Patients With Type 2 Diabetes. Diabetes 52, 1319-1325.

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., . . . and Jaenisch, R. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature 441, 349-353.

Broholm, C., Brandt, C., Schultz, N.S., Nielsen, A.R., Pedersen, B.K., and Scheele, C. (2012). Deficient leukemia inhibitory factor signaling in muscle precursor cells from patients with type 2 diabetes. Am. J. Physiol. Endocrinol. Metab. 303, E283-E292.

Brownlee, M. (2001). Biochemistry and molecular cell biology of diabetic complications. Nature 414, 813-820.

Brownlee, M. (2005). The Pathobiology of Diabetic Complications: A Unifying Mechanism. Diabetes 54, 1615-1625.

Brunetti, P. (2007). The lean patient with type 2 diabetes: characteristics and therapy challenge. International Journal of Clinical Practice 61, 3-9.
Bumgarner, R. (2001). Overview of DNA Microarrays: Types, Applications, and Their Future. In Current Protocols in Molecular Biology (John Wiley & Sons, Inc.).
Butler, A.E., Janson, J., Bonner-Weir, S., Ritzel, R., Rizza, R.A., and Butler, P.C. (2003). β-Cell Deficit and Increased β-Cell Apoptosis in Humans With Type 2 Diabetes. Diabetes 52, 102-110.
Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C.A., . . . and Karp, P.D. (2014). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Res. 42, D459-D471.
Chaudhuri, R., Khoo, P.S., Tonks, K., Junutula, J.R., Kolumam, G., Modrusan, Z., . . . and James, D.E. (2015). Cross-species gene expression analysis identifies a novel set of genes implicated in human insulin sensitivity. Npj Systems Biology And Applications 1, 15010.
Chen, E., Tan, C., Kou, Y., Duan, Q., Wang, Z., Meirelles, G., . . . and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinf. 14, 128.
Chen, L., Magliano, D.J., and Zimmet, P.Z. (2012). The worldwide epidemiology of type 2 diabetes mellitus - present and future perspectives. Nat. Rev. Endocrinol. 8, 228-236.
Chevreul, M.E. (1815). Note sur le Sucre de Diabetes. In Annales de Chemie, vol 95 (Paris), pp. 319-320.
Chibalin, A.V., Leng, Y., Vieira, E., Krook, A., Björnholm, M., Long, Y.C., . . . and Zierath, J.R. (2008). Downregulation of Diacylglycerol Kinase Delta Contributes to Hyperglycemia-Induced Insulin Resistance. Cell 132, 375-386.
Cho, Y.S., Chen, C.-H., Hu, C., Long, J., Hee Ong, R.T., Sim, X., . . . and Seielstad, M. (2012). Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. Nat Genet 44, 67-72.
Choi, J.K., Yu, U., Kim, S., and Yoo, O.J. (2003). Combining multiple microarray studies and modeling interstudy variation. Bioinformatics 19, i84-i90.
Chomentowski, P., Coen, P.M., Radiková, Z., Goodpaster, B.H., and Toledo, F.G.S. (2010). Skeletal Muscle Mitochondria in Insulin Resistance: Differences in Intermyofibrillar Versus Subsarcolemmal Subpopulations and Relationship to Metabolic Flexibility. J. Clin. Endocrinol. Metab. 96, 494-503.
Chung, S.S.M., Ho, E.C.M., Lam, K.S.L., and Chung, S.K. (2003). Contribution of Polyol Pathway to Diabetes-Induced Oxidative Stress. Journal of the American Society of Nephrology 14, S233-S236.
Cnop, M., Vidal, J., Hull, R.L., Utzschneider, K.M., Carr, D.B., Schraw, T., . . . and Kahn, S.E. (2007). Progressive Loss of β-Cell Function Leads to Worsening Glucose Tolerance in First-Degree Relatives of Subjects With Type 2 Diabetes. Diabetes Care 30, 677-682.
Copeland, A.H. (1951). A reasonable social welfare function. (Seminar on Mathematics in Social Sciences, University of Michigan).
Craig, M.E., Femia, G., Broyda, V., Lloyd, M., and Howard, N.J. (2007). Type 2 diabetes in Indigenous and non-Indigenous children and adolescents in New South Wales. Medical Journal of Australia 186, 497.
Croft, D., Mundo, A.F., Haw, R., Milacic, M., Weiser, J., Wu, G., . . . and D'Eustachio, P. (2014). The Reactome pathway knowledgebase. Nucleic Acids Res. 42, D472-D477.
Cusi, K., Maezono, K., Osman, A., Pendergrass, M., Patti, M.E., Pratipanawatr, T., . . . and Mandarino, L.J. (2000). Insulin resistance differentially affects the PI 3-kinase– and MAP kinase–mediated signaling in human muscle. J. Clin. Invest. 105, 311-320.
Dabelea, D., Bell, R.A., D'Agostino Jr, R.B., Imperatore, G., Johansen, J.M., Linder, B., . . . and Mayer-Davis, E.J. (2007). Incidence of diabetes in youth in the United States. JAMA 297, 2716-2724.
de Borda, J.C. (1781). Mémoire sur les élections au scrutin. Histoire de l'Académie Royale des Sciences.
De Feyter, H.M., van den Broek, N.M.A., Praet, S.F.E., Nicolay, K., van Loon, L.J.C., and Prompers, J.J. (2008). Early or advanced stage type 2 diabetes is not accompanied by in vivo skeletal muscle mitochondrial dysfunction. Eur. J. Endocrinol. 158, 643-653.
de Leeuw, C.A., Mooij, J.M., Heskes, T., and Posthuma, D. (2015). MAGMA: Generalized Gene-Set Analysis of GWAS Data. PLoS Comput Biol 11, e1004219.
DeFronzo, R.A., Tobin, J.D., and Andres, R. (1979). Glucose clamp technique: a method for quantifying insulin secretion and resistance. American Journal of Physiology - Gastrointestinal and Liver Physiology 237, G214-G223.
DeFronzo, R.A., and Tripathy, D. (2009). Skeletal Muscle Insulin Resistance Is the Primary Defect in Type 2 Diabetes. Diabetes Care 32, S157-S163.
DeRisi, J., Penland, L., Brown, P.O., Bittner, M.L., Meltzer, P.S., Ray, M., . . . and Trent, J.M. (1996). Use of a cDNA microarray to analyse gene expression patterns in human cancer. Nat Genet 14, 457-460.
Dobson, M., and Fothergill, J. (1776). Experiments and Observations on the Urine in a Diabetes. In Medical Observations and Inquiries, vol 5 (London), pp. 298-316.
Donath, M.Y., and Shoelson, S.E. (2011). Type 2 diabetes as an inflammatory disease. Nat. Rev. Immunol. 11, 98-107.
Doria, A., Patti, M.-E., and Kahn, C.R. (2008). The emerging genetic architecture of type 2 diabetes. Cell Metab. 8, 186-200.

Dowse, G.K., Zimmet, P.Z., Finch, C.F., and Collins, V.R. (1991). Decline in Incidence of Epidemic Glucose Intolerance in Nauruans: Implications for the "Thrifty Genotype". American Journal of Epidemiology 133, 1093-1104.

Dresner, A., Laurent, D., Marcucci, M., Griffin, M.E., Dufour, S., Cline, G.W., . . . and Shulman, G.I. (1999). Effects of free fatty acids on glucose transport and IRS-1–associated phosphatidylinositol 3-kinase activity. J. Clin. Invest. 103, 253-259.

Du, X.L., Edelstein, D., Dimmeler, S., Ju, Q., Sui, C., and Brownlee, M. (2001). Hyperglycemia inhibits endothelial nitric oxide synthase activity by posttranslational modification at the Akt site. Journal of Clinical Investigation 108, 1341-1348.

Duarte, N.C., Becker, S.A., Jamshidi, N., Thiele, I., Mo, M.L., Vo, T.D., . . . and Palsson, B.Ø. (2007). Global reconstruction of the human metabolic network based on genomic and bibliomic data. Proc. Natl. Acad. Sci. USA 104, 1777-1782.

Duckworth, W., Abraira, C., Moritz, T., Reda, D., Emanuele, N., Reaven, P.D., . . . and Huang, G.D. (2009). Glucose Control and Vascular Complications in Veterans with Type 2 Diabetes. New England Journal of Medicine 360, 129-139.

Dunaif, A., and Finegood, D.T. (1996). Beta-cell dysfunction independent of obesity and glucose intolerance in the polycystic ovary syndrome. J. Clin. Endocrinol. Metab. 81, 942-947.

Eckel, R.H., Grundy, S.M., and Zimmet, P.Z. (2005). The metabolic syndrome. The Lancet 365, 1415-1428.

Efron, B., and Tibshirani, R. (2007). On testing the significance of sets of genes. Ann. Appl. Stat. 1, 107-129.

Esser, N., Legrand-Poels, S., Piette, J., Scheen, A.J., and Paquot, N. (2014). Inflammation as a link between obesity, metabolic syndrome and type 2 diabetes. Diabetes Res. Clin. Pr. 105, 141-150.

Ferrannini, E., Gastaldelli, A., Matsuda, M., Miyazaki, Y., Pettiti, M., Glass, L., and DeFronzo, R.A. (2003). Influence of Ethnicity and Familial Diabetes on Glucose Tolerance and Insulin Action: A Physiological Analysis. J. Clin. Endocrinol. Metab. 88, 3251-3257.

Fink, L.N., Oberbach, A., Costford, S.R., Chan, K.L., Sams, A., Blüher, M., and Klip, A. (2013). Expression of anti-inflammatory macrophage genes within skeletal muscle correlates with insulin sensitivity in human obesity and type 2 diabetes. Diabetologia 56, 1623-1628.

Fisher, R.A. (1932). Statistical methods for research workers. (Edinburgh: Oliver and Boyd).

Fodor, S.P., Read, J.L., Pirrung, M.C., Stryer, L., Lu, A.T., and Solas, D. (1991). Light-directed, spatially addressable parallel chemical synthesis. Science 251, 767-773.

Fonseca, N.A., Petryszak, R., Marioni, J., and Brazma, A. (2014). iRAP - an integrated RNA-seq Analysis Pipeline. bioRxiv doi:10.1101/005991

Forslund, K., Hildebrand, F., Nielsen, T., Falony, G., Le Chatelier, E., Sunagawa, S., . . . and Pedersen, O. (2015). Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. Nature 528, 262-266.

Fowler, M.J. (2008). Microvascular and Macrovascular Complications of Diabetes. Clinical Diabetes 26, 77-82.

Frederiksen, C.M., Højlund, K., Hansen, L., Oakeley, E.J., Hemmings, B., Abdallah, B.M., . . . and Gaster, M. (2008). Transcriptional profiling of myotubes from patients with type 2 diabetes: no evidence for a primary defect in oxidative phosphorylation genes. Diabetologia 51, 2068-2077.

Gallagher, I., Scheele, C., Keller, P., Nielsen, A., Remenyi, J., Fischer, C., . . . and Timmons, J. (2010). Integration of microRNA changes in vivo identifies novel molecular features of muscle insulin resistance in type 2 diabetes. Genome Medicine 2, 9.

Garcia-Albornoz, M., Thankaswamy-Kosalai, S., Nilsson, A., Varemo, L., Nookaew, I., and Nielsen, J. (2014). BioMet Toolbox 2.0: genome-wide analysis of metabolism and omics data. Nucleic Acids Res 42, W175-181.

Gaster, M., and Beck-Nielsen, H. (2004). The reduced insulin-mediated glucose oxidation in skeletal muscle from type 2 diabetic subjects may be of genetic origin—evidence from cultured myotubes. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease 1690, 85-91.

Gaster, M., Rustan, A.C., Aas, V., and Beck-Nielsen, H. (2004). Reduced Lipid Oxidation in Skeletal Muscle From Type 2 Diabetic Subjects May Be of Genetic Origin: Evidence From Cultured Myotubes. Diabetes 53, 542-548.

Gault, C.R., Obeid, L.M., and Hannun, Y.A. (2010). An overview of sphingolipid metabolism: from synthesis to breakdown. Advances in experimental medicine and biology 688, 1-23.

Gerstein, H.C., Miller, M.E., Byington, R.P., Goff Jr, D.C., Bigger, J.T., Buse, J.B., . . . and Action to Control Cardiovascular Risk in Diabetes Study Group (2008). Effects of intensive glucose lowering in type 2 diabetes. The New England journal of medicine 358, 2545-2559.

Gill, G.V., and Alberti, K.G.M.M. (1985). Hyperosmolar non-ketotic coma. Practical Diabetes International 2, 30-35.

Gillies, A.R., and Lieber, R.L. (2011). Structure and Function of the Skeletal Muscle Extracellular Matrix. Muscle & nerve 44, 318-331.

Gillies, C.L., Abrams, K.R., Lambert, P.C., Cooper, N.J., Sutton, A.J., Hsu, R.T., and Khunti, K. (2007). Pharmacological and lifestyle interventions to prevent or delay type 2 diabetes in people with impaired glucose tolerance: systematic review and meta-analysis. BMJ 334, 299.

Glaab, E., Baudot, A., Krasnogor, N., Schneider, R., and Valencia, A. (2012). EnrichNet: network-based gene set enrichment analysis. Bioinformatics 28, i451-i457.
Goeman, J.J., Van De Geer, S.A., De Kort, F., and Van Houwelingen, H.C. (2004). A global test for groups of genes: testing association with a clinical outcome. Bioinformatics 20, 93-99.
Goodpaster, B.H., Krishnaswami, S., Resnick, H., Kelley, D.E., Haggerty, C., Harris, T.B., . . . and Newman, A.B. (2003). Association Between Regional Adipose Tissue Distribution and Both Type 2 Diabetes and Impaired Glucose Tolerance in Elderly Men and Women. Diabetes Care 26, 372-379.
Green, C.J., Pedersen, M., Pedersen, B.K., and Scheele, C. (2011). Elevated NF- κ B Activation Is Conserved in Human Myocytes Cultured From Obese Type 2 Diabetic Patients and Attenuated by AMP-Activated Protein Kinase. Diabetes 60, 2810-2819.
Griffin, M.E., Marcucci, M.J., Cline, G.W., Bell, K., Barucci, N., Lee, D., . . . and Shulman, G.I. (1999). Free fatty acid-induced insulin resistance is associated with activation of protein kinase C theta and alterations in the insulin signaling cascade. Diabetes 48, 1270-1274.
Guidone, C., Manco, M., Valera-Mora, E., Iaconelli, A., Gniuli, D., Mari, A., . . . and Mingrone, G. (2006). Mechanisms of Recovery From Type 2 Diabetes After Malabsorptive Bariatric Surgery. Diabetes 55, 2025-2031.
Haffner, S.M., Lehto, S., Rönnemaa, T., Pyörälä, K., and Laakso, M. (1998). Mortality from Coronary Heart Disease in Subjects with Type 2 Diabetes and in Nondiabetic Subjects with and without Prior Myocardial Infarction. New England Journal of Medicine 339, 229-234.
Hales, C.N., and Barker, D.J.P. (2001). The thrifty phenotype hypothesis: Type 2 diabetes. British Medical Bulletin 60, 5-20.
Hao, T., Ma, H.-W., Zhao, X.-M., and Goryanin, I. (2010). Compartmentalization of the Edinburgh Human Metabolic Network. BMC Bioinf. 11, 393.
Hauner, H. (2002). The mode of action of thiazolidinediones. Diabetes-Met. Res. 18, S10-S15.
Haus, J.M., Kashyap, S.R., Kasumov, T., Zhang, R., Kelly, K.R., DeFronzo, R.A., and Kirwan, J.P. (2009). Plasma Ceramides Are Elevated in Obese Subjects With Type 2 Diabetes and Correlate With the Severity of Insulin Resistance. Diabetes 58, 337-343.
Hebenstreit, D., Fang, M., Gu, M., Charoensawan, V., van Oudenaarden, A., and Teichmann, S.A. (2011). RNA sequencing reveals two major classes of gene expression levels in metazoan cells. Mol. Syst. Biol. 7.
Hedges, L.V., and Olkin, I. (1985). Statistical Methods for Meta-analysis. (London: Academic Press).
Henry, R.R., Ciaraldi, T.P., Mudaliar, S., Abrams, L., and Nikoulina, S.E. (1996). Acquired Defects of Glycogen Synthase Activity in Cultured Human Skeletal Muscle Cells: Influence of High Glucose and Insulin Levels. Diabetes 45, 400-407.
Hillier, T.A., and Pedula, K.L. (2001). Characteristics of an Adult Population With Newly Diagnosed Type 2 Diabetes: The relation of obesity and age of onset. Diabetes Care 24, 1522-1527.
Hosack, D., Dennis, G., Sherman, B., Lane, H., and Lempicki, R. (2003). Identifying biological themes within lists of genes with EASE. Genome Biol. 4, R70.
Hu, F.B. (2011). Globalization of Diabetes: The role of diet, lifestyle, and genes. Diabetes Care 34, 1249-1257.
Hu, F.B., Li, T.Y., Colditz, G.A., Willett, W.C., and Manson, J.E. (2003). Television watching and other sedentary behaviors in relation to risk of obesity and type 2 diabetes mellitus in women. JAMA 289, 1785-1791.
Hu, F.B., Manson, J.E., Stampfer, M.J., Colditz, G., Liu, S., Solomon, C.G., and Willett, W.C. (2001). Diet, Lifestyle, and the Risk of Type 2 Diabetes Mellitus in Women. New England Journal of Medicine 345, 790-797.
Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 37, 1-13.
Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., . . . and Morgan, M. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. Nat Meth 12, 115-121.
Hummel, M., Meister, R., and Mansmann, U. (2008). GlobalANCOVA: exploration and assessment of gene group effects. Bioinformatics 24, 78-85.
Hung, J.-H., Yang, T.-H., Hu, Z., Weng, Z., and DeLisi, C. (2012). Gene set enrichment analysis: performance evaluation and usage guidelines. Briefings Bioinform. 13, 281-291.
Inzucchi, S.E., Bergenstal, R.M., Buse, J.B., Diamant, M., Ferrannini, E., Nauck, M., . . . and Matthews, D.R. (2012). Management of hyperglycaemia in type 2 diabetes: a patient-centered approach. Position statement of the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetologia 55, 1577-1596.
Iozzo, P., Pratipanawatr, T., Pijl, H., Vogt, C., Kumar, V., Pipek, R., . . . and DeFronzo, R.A. (2001). Physiological hyperinsulinemia impairs insulin-stimulated glycogen synthase activity and glycogen synthesis. Am. J. Physiol. Endocrinol. Metab. 280, E712-E719.
Itani, S.I., Ruderman, N.B., Schmieder, F., and Boden, G. (2002). Lipid-Induced Insulin Resistance in Human Muscle Is Associated With Changes in Diacylglycerol, Protein Kinase C, and I κ B- α . Diabetes 51, 2005-2011.
Jerby, L., Shlomi, T., and Ruppin, E. (2010). Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. Mol. Syst. Biol. 6.

Jin, W., Goldfine, A.B., Boes, T., Henry, R.R., Ciaraldi, T.P., Kim, E.-Y., . . . and Patti, M.-E. (2011). Increased SRF transcriptional activity in human and mouse skeletal muscle is a signature of insulin resistance. J. Clin. Invest. 121, 918-929.

Jones, A.G., and Hattersley, A.T. (2013). The clinical utility of C-peptide measurement in the care of patients with diabetes. Diabetic Medicine 30, 803-817.

Kahn, S.E., Cooper, M.E., and Del Prato, S. (2014). Pathophysiology and treatment of type 2 diabetes: perspectives on the past, present, and future. The Lancet 383, 1068-1083.

Kahn, S.E., Hull, R.L., and Utzschneider, K.M. (2006). Mechanisms linking obesity to insulin resistance and type 2 diabetes. Nature 444, 840-846.

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. 40, D109-D114.

Karlsson, F.H., Tremaroli, V., Nookaew, I., Bergstrom, G., Behre, C.J., Fagerberg, B., . . . and Backhed, F. (2013). Gut metagenome in European women with normal, impaired and diabetic glucose control. Nature 498, 99-103.

Karr, Jonathan R., Sanghvi, Jayodita C., Macklin, Derek N., Gutschow, Miriam V., Jacobs, Jared M., Bolival Jr, B., . . . and Covert, Markus W. (2012). A whole-cell computational model predicts phenotype from genotype. Cell 150, 389-401.

Kashyap, S.R., Belfort, R., Berria, R., Suraamornkul, S., Pratipranawatr, T., Finlayson, J., . . . and Cusi, K. (2004). Discordant effects of a chronic physiological increase in plasma FFA on insulin signaling in healthy subjects with or without a family history of type 2 diabetes. Am. J. Physiol. Endocrinol. Metab. 287, E537-E546.

Kelley, D.E., and Mandarino, L.J. (2000). Fuel selection in human skeletal muscle in insulin resistance: a reexamination. Diabetes 49, 677-683.

Kelly, T., Yang, W., Chen, C.S., Reynolds, K., and He, J. (2008). Global burden of obesity in 2005 and projections to 2030. Int J Obes 32, 1431-1437.

Khatri, P., and Drăghici, S. (2005). Ontological analysis of gene expression data: current tools, limitations, and open problems. Bioinformatics 21, 3587-3595.

Kim, J.K., Fillmore, J.J., Sunshine, M.J., Albrecht, B., Higashimori, T., Kim, D.-W., . . . and Shulman, G.I. (2004). PKC- θ knockout mice are protected from fat-induced insulin resistance. J. Clin. Invest. 114, 823-827.

Kim, S.Y., and Volsky, D.J. (2005). PAGE: parametric analysis of gene set enrichment. BMC Bioinf. 6, 144.

Kimura, K., Nakamura, Y., Inaba, Y., Matsumoto, M., Kido, Y., Asahara, S.-i., . . . and Inoue, H. (2013). Histidine Augments the Suppression of Hepatic Glucose Production by Central Insulin Action. Diabetes 62, 2266-2277.

Kohei, K. (2010). Pathophysiology of type 2 diabetes and its treatment policy. JMAJ 53, 41-46.

Kong, S.W., Pu, W.T., and Park, P.J. (2006). A multivariate approach for integrating genome-wide expression data and biological knowledge. Bioinformatics 22, 2373-2380.

Krook, A., Björnholm, M., Galuska, D., Jiang, X.J., Fahlman, R., Myers, M.G., . . . and Zierath, J.R. (2000). Characterization of signal transduction and glucose transport in skeletal muscle from type 2 diabetic patients. Diabetes 49, 284-292.

Laing, S.P., Swerdlow, A.J., Slater, S.D., Burden, A.C., Morris, A., Waugh, N.R., . . . and Patterson, C.C. (2003). Mortality from heart disease in a cohort of 23,000 patients with insulin-treated diabetes. Diabetologia 46, 760-765.

Law, C., Chen, Y., Shi, W., and Smyth, G. (2014). voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 15, R29.

Lee, C., Patil, S., and Sartor, M.A. (2015). RNA-Enrich: A cut-off free functional enrichment testing method for RNA-seq with improved detection power. Bioinformatics.

Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., . . . and Young, R.A. (2006). Control of Developmental Regulators by Polycomb in Human Embryonic Stem Cells. Cell 125, 301-313.

Lee, Tong I., and Young, Richard A. (2013). Transcriptional Regulation and Its Misregulation in Disease. Cell 152, 1237-1251.

Lee, Y.-t., Hsu, C.-c., Lin, M.-h., Liu, K.-s., and Yin, M.-c. (2005). Histidine and carnosine delay diabetic deterioration in mice and protect human low density lipoprotein against oxidation and glycation. Eur. J. Pharmacol. 513, 145-150.

Lefort, N., Glancy, B., Bowen, B., Willis, W.T., Bailowitz, Z., De Filippis, E.A., . . . and Mandarino, L.J. (2010). Increased Reactive Oxygen Species Production and Lower Abundance of Complex I Subunits and Carnitine Palmitoyltransferase 1B Protein Despite Normal Mitochondrial Respiration in Insulin-Resistant Human Skeletal Muscle. Diabetes 59, 2444-2452.

Leibson, C.L., Williamson, D.F., Melton, L.J., Palumbo, P.J., Smith, S.A., Ransom, J.E., . . . and Narayan, K.M.V. (2001). Temporal Trends in BMI Among Adults With Diabetes. Diabetes Care 24, 1584-1589.

Leslie, R.D., Volkmann, H.P., Poncher, M., Hanning, I., Orskov, H., and Alberti, K.G. (1986). Metabolic abnormalities in children of non-insulin dependent diabetics. British Medical Journal (Clinical research ed.) 293, 840-842.

Levy, J.C., Matthews, D.R., and Hermans, M.P. (1998). Correct Homeostasis Model Assessment (HOMA) Evaluation Uses the Computer Program. Diabetes Care 21, 2191-2192.

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, Jill P., and Tamayo, P. (2015). The Molecular Signatures Database Hallmark Gene Set Collection. Cell Systems 1, 417-425.
Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., and Mesirov, J.P. (2011). Molecular signatures database (MSigDB) 3.0. Bioinformatics 27, 1739-1740.
Liebl, A., Mata, M., and Eschwège, E. (2002). Evaluation of risk factors for development of complications in Type II diabetes in Europe. Diabetologia 45, S23-S28.
Lin, Y., and Sun, Z. (2010). Current views on type 2 diabetes. The Journal of endocrinology 204, 1.
Ling, C., and Groop, L. (2009). Epigenetics: A Molecular Link Between Environmental Factors and Type 2 Diabetes. Diabetes 58, 2718-2725.
Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., . . . and Moore, H.F. (2013). The Genotype-Tissue Expression (GTEx) project. Nat Genet 45, 580-585.
Lundberg, E., Fagerberg, L., Klevebring, D., Matic, I., Geiger, T., Cox, J., . . . and Uhlen, M. (2010). Defining the transcriptome and proteome in three functionally different human cell lines. Mol. Syst. Biol. 6.
Lynch, C.J., and Adams, S.H. (2014). Branched-chain amino acids in metabolic signalling and insulin resistance. Nat. Rev. Endocrinol. 10, 723-736.
Lyssenko, V., Jonsson, A., Almgren, P., Pulizzi, N., Isomaa, B., Tuomi, T., . . . and Groop, L. (2008). Clinical Risk Factors, DNA Variants, and the Development of Type 2 Diabetes. New England Journal of Medicine 359, 2220-2232.
Ma, H., Sorokin, A., Mazein, A., Selkov, A., Selkov, E., Demin, O., and Goryanin, I. (2007). The Edinburgh human metabolic network reconstruction and its functional analysis. Mol. Syst. Biol. 3, 135.
Maciejewski, H. (2013). Gene set analysis methods: statistical models and methodological differences. Briefings Bioinform.
Mansmann, U., and Meister, R. (2005). Goeman's Global Test versus an ANCOVA Approach. Methods Inf. Med. 44, 449-453.
Mardinoglu, A., Agren, R., Kampf, C., Asplund, A., Nookaew, I., Jacobson, P., . . . and Nielsen, J. (2013). Integration of clinical data with a genome-scale metabolic model of the human adipocyte. Mol. Syst. Biol. 9.
Mardinoglu, A., Agren, R., Kampf, C., Asplund, A., Uhlen, M., and Nielsen, J. (2014). Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. Nat. Commun. 5.
Mathers, C.D., and Loncar, D. (2006). Projections of Global Mortality and Burden of Disease from 2002 to 2030. PLoS Med 3, e442.
Matthews, D.R., Hosker, J.P., Rudenski, A.S., Naylor, B.A., Treacher, D.F., and Turner, R.C. (1985). Homeostasis model assessment: insulin resistance and β-cell function from fasting plasma glucose and insulin concentrations in man. Diabetologia 28, 412-419.
McIntyre, E.A., Halse, R., Yeaman, S.J., and Walker, M. (2004). Cultured Muscle Cells from Insulin-Resistant Type 2 Diabetes Patients Have Impaired Insulin, but Normal 5-Amino-4-Imidazolecarboxamide Riboside-Stimulated, Glucose Uptake. J. Clin. Endocrinol. Metab. 89, 3440-3448.
Meier, J.J., and Bonadonna, R.C. (2013). Role of Reduced β-Cell Mass Versus Impaired β-Cell Function in the Pathogenesis of Type 2 Diabetes. Diabetes Care 36, S113-S119.
Melo, F., Carey, D.J., and Brandan, E. (1996). Extracellular matrix is required for skeletal muscle differentiation but not myogenin expression. J. Cell. Biochem. 62, 227-239.
Mingrone, G., Panunzi, S., De Gaetano, A., Guidone, C., Iaconelli, A., Leccesi, L., . . . and Rubino, F. (2012). Bariatric Surgery versus Conventional Medical Therapy for Type 2 Diabetes. New England Journal of Medicine 366, 1577-1585.
Misra, A. (2003). Revisions of cutoffs of body mass index to define overweight and obesity are needed for the Asian-ethnic groups. Int J Obes Relat Metab Disord 27, 1294-1296.
Mokdad, A.H., Ford, E.S., Bowman, B.A., and et al. (2003). Prevalence of obesity, diabetes, and obesity-related health risk factors, 2001. JAMA 289, 76-79.
Mootha, V.K., Lindgren, C.M., Eriksson, K.F., Subramanian, A., Sihag, S., Lehar, J., . . . and Laurila, E. (2003). PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat. Genet. 34, 267-273.
Morino, K., Petersen, K.F., Dufour, S., Befroy, D., Frattini, J., Shatzkes, N., . . . and Shulman, G.I. (2005). Reduced mitochondrial density and increased IRS-1 serine phosphorylation in muscle of insulin-resistant offspring of type 2 diabetic parents. J. Clin. Invest. 115, 3587-3593.
Morris, A.P., Voight, B.F., Teslovich, T.M., Ferreira, T., Segrè, A.V., Steinthorsdottir, V., . . . and McCarthy, M.I. (2012). Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat. Genet. 44, 981-990.
Morrish, N.J., Wang, S.L., Stevens, L.K., Fuller, J.H., and Keen, H. (2001). Mortality and causes of death in the WHO multinational study of vascular disease in diabetes. Diabetologia 44, S14-S21.
Mozaffarian, D., Kamineni, A., Carnethon, M., Djoussé, L., Mukamal, K.J., and Siscovick, D. (2009). Lifestyle risk factors and new-onset diabetes mellitus in older adults: The cardiovascular health study. Archives of Internal Medicine 169, 798-807.
Muoio, D.M., and Newgard, C.B. (2008). Molecular and metabolic mechanisms of insulin resistance and beta-cell failure in type 2 diabetes. Nat. Rev. Mol. Cell Biol. 9, 193-205.

Murray, I.A.N. (1971). Paulesco and the Isolation of Insulin. Journal of the History of Medicine and Allied Sciences XXVI, 150-157.
Naeem, H., Zimmer, R., Tavakkolkhah, P., and Kuffner, R. (2012). Rigorous assessment of gene set enrichment tests. Bioinformatics 28, 1480-1486.
Nagaraj, N., Wisniewski, J.R., Geiger, T., Cox, J., Kircher, M., Kelso, J., . . . and Mann, M. (2011). Deep proteome and transcriptome mapping of a human cancer cell line. Mol. Syst. Biol. 7.
Nam, D., Kim, J., Kim, S.-Y., and Kim, S. (2010). GSA-SNP: a general approach for gene set analysis of polymorphisms. Nucleic Acids Res. 38, W749-W754.
Nam, D., and Kim, S.Y. (2008). Gene-set approach for expression pattern analysis. Briefings Bioinform. 9, 189-197.
Nathan, D.M., Turgeon, H., and Regan, S. (2007). Relationship between glycated haemoglobin levels and mean glucose levels over time. Diabetologia 50, 2239-2244.
Nishikawa, T., Edelstein, D., Du, X.L., Yamagishi, S.-i., Matsumura, T., Kaneda, Y., . . . and Brownlee, M. (2000). Normalizing mitochondrial superoxide production blocks three pathways of hyperglycaemic damage. Nature 404, 787-790.
Nogiec, C., Burkart, A., Dreyfuss, J.M., Lerin, C., Kasif, S., and Patti, M.-E. (2015). Metabolic modeling of muscle metabolism identifies key reactions linked to insulin resistance phenotypes. Molecular Metabolism 4, 151-163.
Oberhardt, M.A., Palsson, B.O., and Papin, J.A. (2009). Applications of genome-scale metabolic reconstructions. Mol. Syst. Biol. 5, 320.
Oliveira, A.P., Patil, K.R., and Nielsen, J. (2008). Architecture of transcriptional regulatory circuits is knitted over the topology of bio-molecular interaction networks. BMC Syst. Biol. 2, 17.
Orth, J.D., Thiele, I., and Palsson, B.O. (2010). What is flux balance analysis? Nat. Biotechnol. 28, 245-248.
Osses, N., and Brandan, E. (2002). ECM is required for skeletal muscle differentiation independently of muscle regulatory factor expression. American Journal of Physiology - Cell Physiology 282, C383-C394.
Pasquel, F.J., and Umpierrez, G.E. (2014). Hyperosmolar Hyperglycemic State: A Historic Review of the Clinical Presentation, Diagnosis, and Treatment. Diabetes Care 37, 3124-3131.
Patel, A., MacMahon, S., Chalmers, J., Neal, B., Billot, L., Woodward, M., . . . and The ADVANCE Collaborative Group (2008). Intensive blood glucose control and vascular outcomes in patients with type 2 diabetes.
Patil, K.R., and Nielsen, J. (2005). Uncovering transcriptional regulation of metabolism by using metabolic network topology. Proc. Natl. Acad. Sci. USA 102, 2685-2689.
Patti, M.E., Butte, A.J., Crunkhorn, S., Cusi, K., Berria, R., Kashyap, S., . . . and Mandarino, L.J. (2003). Coordinated reduction of genes of oxidative metabolism in humans with insulin resistance and diabetes: Potential role of PGC1 and NRF1. Proc. Natl. Acad. Sci. USA 100, 8466-8471.
Petryszak, R., Keays, M., Tang, Y.A., Fonseca, N.A., Barrera, E., Burdett, T., . . . and Brazma, A. (2015). Expression Atlas update—an integrated database of gene and protein expression in humans, animals and plants. Nucleic Acids Res.
Pihlajamäki, J., Lerin, C., Itkonen, P., Boes, T., Floss, T., Schroeder, J., . . . and Patti, Mary E. (2011). Expression of the Splicing Factor Gene SFRS10 Is Reduced in Human Obesity and Contributes to Enhanced Lipogenesis. Cell Metab. 14, 208-218.
Pinhas-Hamiel, O., and Zeitler, P. (2005). The global spread of type 2 diabetes mellitus in children and adolescents. The Journal of Pediatrics 146, 693-700.
Pratipanawatr, W., Pratipanawatr, T., Cusi, K., Berria, R., Adams, J.M., Jenkinson, C.P., . . . and Mandarino, L.J. (2001). Skeletal Muscle Insulin Resistance in Normoglycemic Subjects With a Strong Family History of Type 2 Diabetes Is Associated With Decreased Insulin-Stimulated Insulin Receptor Substrate-1 Tyrosine Phosphorylation. Diabetes 50, 2572-2578.
Qaseem, A., Humphrey, L.L., Sweet, D.E., Starkey, M., and Shekelle, P. (2012). Oral Pharmacologic Treatment of Type 2 Diabetes Mellitus: A Clinical Practice Guideline From the American College of Physicians. Annals of Internal Medicine 156, 218-231.
Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., . . . and Wang, J. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature 490, 55-60.
R Core Team (2005). R: A language and environment for statistical computing. (Vienna, Austria). [www.R-project.org/]
Ramachandran, A., Mary, S., Yamuna, A., Murugesan, N., and Snehalatha, C. (2008). High Prevalence of Diabetes and Cardiovascular Risk Factors Associated With Urbanization in India. Diabetes Care 31, 893-898.
Ramasamy, A., Mondry, A., Holmes, C.C., and Altman, D.G. (2008). Key Issues in Conducting a Meta-Analysis of Gene Expression Microarray Datasets. PLoS Med. 5, e184.
Rena, G., Pearson, E.R., and Sakamoto, K. (2013). Molecular mechanism of action of metformin: old or new insights? Diabetologia 56, 1898-1906.
Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43, e47-e47.

Robaina Estévez, S., and Nikoloski, Z. (2015). Context-Specific Metabolic Model Extraction Based on Regularized Least Squares Optimization. PLoS ONE 10, e0131875.
Romero, P., Wagg, J., Green, M., Kaiser, D., Krummenacker, M., and Karp, P. (2004). Computational prediction of human metabolic pathways from the complete human genome. Genome Biol. 6, R2.
Rosenbloom, A.L., Joe, J.R., Young, R.S., and Winter, W.E. (1999). Emerging epidemic of type 2 diabetes in youth. Diabetes Care 22, 345-354.
Russo, S.B., Ross, J.S., and Cowart, L.A. (2013). Sphingolipids in Obesity, Type 2 Diabetes, and Metabolic Disease. In Sphingolipids in Disease. E. Gulbins, and I. Petrache, eds. (Springer Vienna), pp. 373-401.
Ryu, J.Y., Kim, H.U., and Lee, S.Y. (2015). Reconstruction of genome-scale human metabolic models using omics data. Integrative Biology 7, 859-868.
Sacks, D.B., Arnold, M., Bakris, G.L., Bruns, D.E., Horvath, A.R., Kirkman, M.S., . . . and Nathan, D.M. (2011). Guidelines and Recommendations for Laboratory Analysis in the Diagnosis and Management of Diabetes Mellitus. Diabetes Care 34, e61-e99.
Saltiel, A.R., and Kahn, C.R. (2001). Insulin signalling and the regulation of glucose and lipid metabolism. Nature 414, 799-806.
Scheele, C., Nielsen, S., Kelly, M., Broholm, C., Nielsen, A.R., Taudorf, S., . . . and Pedersen, B.K. (2012). Satellite Cells Derived from Obese Humans with Type 2 Diabetes and Differentiated into Myocytes In Vitro Exhibit Abnormal Response to IL-6. PLoS ONE 7, e39657.
Schmidt, B.J., Ebrahim, A., Metz, T.O., Adkins, J.N., Palsson, B.Ø., and Hyduke, D.R. (2013). GIM3E: condition-specific models of cellular metabolism developed from metabolomics and expression data. Bioinformatics 29, 2900-2908.
Schmitz, O., Brock, B., and Rungby, J. (2004). Amylin Agonists: A Novel Approach in the Treatment of Diabetes. Diabetes 53, S233-S238.
Schwanhausser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., . . . and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature 473, 337-342.
Scully, T. (2012). Diabetes in numbers. Nature 485, S2-S3.
Sears, D.D., Hsiao, G., Hsiao, A., Yu, J.G., Courtney, C.H., Ofrecio, J.M., . . . and Subramaniam, S. (2009). Mechanisms of human insulin resistance and thiazolidinedione-mediated insulin sensitization. Proc. Natl. Acad. Sci. USA 106, 18745-18750.
Seenundun, S., Rampalli, S., Liu, Q.C., Aziz, A., Palii, C., Hong, S., . . . and Dilworth, F.J. (2010). UTX mediates demethylation of H3K27me3 at muscle-specific genes during myogenesis. The EMBO Journal 29, 1401-1411.
Segrè, A.V., Groop, L., Mootha, V.K., Daly, M.J., Altshuler, D., Consortium, D., and investigators, M. (2010). Common Inherited Variation in Mitochondrial Genes Is Not Enriched for Associations with Type 2 Diabetes or Related Glycemic Traits. PLoS Genet 6, e1001058.
Sen, G.L., Webster, D.E., Barragan, D.I., Chang, H.Y., and Khavari, P.A. (2008). Control of differentiation in a self-renewing mammalian tissue by the histone demethylase JMJD3. Genes & Development 22, 1865-1870.
Shaw, J.E., Sicree, R.A., and Zimmet, P.Z. (2010). Global estimates of the prevalence of diabetes for 2010 and 2030. Diabetes Res. Clin. Pr. 87, 4-14.
Shin, Andrew C., Fasshauer, M., Filatova, N., Grundell, Linus A., Zielinski, E., Zhou, J.-Y., . . . and Buettner, C. (2014). Brain Insulin Lowers Circulating BCAA Levels by Inducing Hepatic BCAA Catabolism. Cell Metab. 20, 898-909.
Shlomi, T., Cabili, M.N., Herrgard, M.J., Palsson, B.O., and Ruppin, E. (2008). Network-based prediction of human tissue-specific metabolism. Nat. Biotechnol. 26, 1003-1010.
Shoelson, S.E., Lee, J., and Goldfine, A.B. (2006). Inflammation and insulin resistance. J. Clin. Invest. 116, 1793-1801.
Shulman, G.I. (2000). Cellular mechanisms of insulin resistance. J. Clin. Invest. 106, 171-176.
Sjöström, L., Lindroos, A.-K., Peltonen, M., Torgerson, J., Bouchard, C., Carlsson, B., . . . and Wedel, H. (2004). Lifestyle, Diabetes, and Cardiovascular Risk Factors 10 Years after Bariatric Surgery. New England Journal of Medicine 351, 2683-2693.
Stančáková, A., Civelek, M., Saleem, N.K., Soininen, P., Kangas, A.J., Cederberg, H., . . . and Laakso, M. (2012). Hyperglycemia and a Common Variant of GCKR Are Associated With the Levels of Eight Amino Acids in 9,369 Finnish Men. Diabetes 61, 1895-1902.
Stephens, Z.D., Lee, S.Y., Faghri, F., Campbell, R.H., Zhai, C., Efron, M.J., . . . and Robinson, G.E. (2015). Big Data: Astronomical or Genomical? PLoS Biol 13, e1002195.
Stern, M.M., Myers, R.L., Hammam, N., Stern, K.A., Eberli, D., Kritchevsky, S.B., . . . and Van Dyke, M. (2009). The influence of extracellular matrix derived from skeletal muscle tissue on the proliferation and differentiation of myogenic progenitor cells ex vivo. Biomaterials 30, 2393-2399.
Stolar, M. (2010). Glycemic Control and Complications in Type 2 Diabetes Mellitus. The American Journal of Medicine 123, S3-S11.
Stouffer, S.A., Suchman, E.A., Devinney, L.C., Star, S.A., and Williams Jr, R.M. (1949). The American soldier: adjustment during army life. (Oxford, England: Princeton University Press).

Straczkowski, M., Kowalska, I., Baranowski, M., Nikolajuk, A., Otziomek, E., Zabielski, P., . . . and Gorska, M. (2007). Increased skeletal muscle ceramide level in men at risk of developing type 2 diabetes. Diabetologia 50, 2366-2373.

Straczkowski, M., Kowalska, I., Nikolajuk, A., Dzienis-Straczkowska, S., Kinalska, I., Baranowski, M., . . . and Gorski, J. (2004). Relationship Between Insulin Sensitivity and Sphingomyelin Signaling Pathway in Human Skeletal Muscle. Diabetes 53, 1215-1221.

Stump, C.S., Henriksen, E.J., Wei, Y., and Sowers, J.R. (2006). The metabolic syndrome: Role of skeletal muscle metabolism. Ann. Med. 38, 389-402.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., . . . and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA 102, 15545-15550.

Summers, S.A., and Nelson, D.H. (2005). A Role for Sphingolipids in Producing the Common Features of Type 2 Diabetes, Metabolic Syndrome X, and Cushing's Syndrome. Diabetes 54, 591-602.

Szendroedi, J., Phielix, E., and Roden, M. (2012). The role of mitochondria in insulin resistance and type 2 diabetes mellitus. Nat. Rev. Endocrinol. 8, 92-103.

Tarca, A., Draghici, S., Bhatti, G., and Romero, R. (2012). Down-weighting overlapping genes improves gene set analysis. BMC Bioinf. 13, 136.

Tarca, A.L., Bhatti, G., and Romero, R. (2013). A Comparison of Gene Set Analysis Methods in Terms of Sensitivity, Prioritization and Specificity. PLoS ONE 8, e79217.

Taylor, J., and Tibshirani, R. (2006). A tail strength measure for assessing the overall univariate significance in a dataset. Biostatistics 7, 167-181.

Thiele, I., Swainston, N., Fleming, R.M.T., Hoppe, A., Sahoo, S., Aurich, M.K., . . . and Palsson, B.O. (2013). A community-driven global reconstruction of human metabolism. Nat Biotech 31, 419-425.

Thompson, D.B., Pratley, R., and Ossowski, V. (1996). Human primary myoblast cell cultures from non-diabetic insulin resistant subjects retain defects in insulin action. J. Clin. Invest. 98, 2346-2350.

Tomasi, T., Sledz, D., Wales, J.K., and Recant, L. (1967). Insulin Half-Life in Normal and Diabetic Subjects. Experimental Biology and Medicine 126, 315-317.

Tripathy, D., Lindholm, E., Isomaa, B., Saloranta, C., Tuomi, T., and Groop, L. (2003). Familiality of metabolic abnormalities is dependent on age at onset and phenotype of the type 2 diabetic proband. Am. J. Physiol. Endocrinol. Metab. 285, E1297-E1303.

Tsai, C.A., and Chen, J.J. (2009). Multivariate analysis of variance test for gene set analysis. Bioinformatics 25, 897-903.

Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., . . . and Pontén, F. (2015). Tissue-based map of the human proteome. Science 347.

Utzschneider, K.M., Carr, D.B., Hull, R.L., Kodama, K., Shofer, J.B., Retzlaff, B.M., . . . and Kahn, S.E. (2004). Impact of Intra-Abdominal Fat and Age on Insulin Sensitivity and β-Cell Function. Diabetes 53, 2867-2872.

Vaag, A., Henriksen, J.E., and Beck-Nielsen, H. (1992). Decreased insulin activation of glycogen synthase in skeletal muscles in young nonobese Caucasian first-degree relatives of patients with non-insulin-dependent diabetes mellitus. J. Clin. Invest. 89, 782-788.

Wallace, T.M., Levy, J.C., and Matthews, D.R. (2004). Use and Abuse of HOMA Modeling. Diabetes Care 27, 1487-1495.

van Tienen, F.H.J., Praet, S.F.E., de Feyter, H.M., van den Broek, N.M., Lindsey, P.J., Schoonderwoerd, K.G.C., . . . and C., v.L.J. (2012). Physical Activity Is the Key Determinant of Skeletal Muscle Mitochondrial Function in Type 2 Diabetes. J. Clin. Endocrinol. Metab. 97, 3261-3269.

Wang, Y., Eddy, J., and Price, N. (2012). Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. BMC Syst. Biol. 6, 153.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nature reviews. Genetics 10, 57-63.

Weisberg, S.P., McCann, D., Desai, M., Rosenbaum, M., Leibel, R.L., and Ferrante, A.W., Jr. (2003). Obesity is associated with macrophage accumulation in adipose tissue. J. Clin. Invest. 112, 1796-1808.

Velleman, S.G., Shin, J., Li, X., and Song, Y. (2012). Review: The skeletal muscle extracellular matrix: Possible roles in the regulation of muscle development and growth. Canadian Journal of Animal Science 92, 1-10.

Whincup, P.H., Kaye, S.J., Owen, C.G., and et al. (2008). Birth weight and risk of type 2 diabetes: A systematic review. JAMA 300, 2886-2897.

Whiting, D.R., Guariguata, L., Weil, C., and Shaw, J. (2011). IDF Diabetes Atlas: Global estimates of the prevalence of diabetes for 2011 and 2030. Diabetes Res. Clin. Pr. 94, 311-321.

WHO (2011). Use of glycated haemoglobin (HbA1c) in the diagnosis of diabetes mellitus. (Geneva).

WHO (2014a). Global Health Estimates: Deaths by Cause, Age, Sex and Country. (Geneva).

WHO (2014b). Global status report on noncommunicable diseases 2014. (Geneva).

WHO, and IDF (2006). Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia. (Geneva).

Vlassis, N., Pacheco, M.P., and Sauter, T. (2014). Fast Reconstruction of Compact Context-Specific Metabolic Network Models. PLoS Comput Biol 10, e1003424.

Yanai, H., Adachi, H., Katsuyama, H., Moriyama, S., Hamasaki, H., and Sako, A. (2015). Causative anti-diabetic drugs and the underlying clinical factors for hypoglycemia in patients with diabetes. World Journal of Diabetes 6, 30-36.
Yang, Q., Graham, T.E., Mody, N., Preitner, F., Peroni, O.D., Zabolotny, J.M., . . . and Kahn, B.B. (2005). Serum retinol binding protein 4 contributes to insulin resistance in obesity and type 2 diabetes. Nature 436, 356-362.
Yang, W., Lu, J., Weng, J., Jia, W., Ji, L., Xiao, J., . . . and He, J. (2010). Prevalence of Diabetes among Men and Women in China. New England Journal of Medicine 362, 1090-1101.
Yizhak, K., Gaude, E., Le Dévédec, S., Waldman, Y.Y., Stein, G.Y., van de Water, B., . . . and Ruppin, E. (2014). Phenotype-based cell-specific metabolic modeling reveals metabolic liabilities of cancer. eLife 3.
Yoon, K.-H., Lee, J.-H., Kim, J.-W., Cho, J.H., Choi, Y.-H., Ko, S.-H., . . . and Son, H.-Y. (2006). Epidemic obesity and type 2 diabetes in Asia. The Lancet 368, 1681-1688.
Yu, C., Chen, Y., Cline, G.W., Zhang, D., Zong, H., Wang, Y., . . . and Shulman, G.I. (2002). Mechanism by Which Fatty Acids Inhibit Insulin Activation of Insulin Receptor Substrate-1 (IRS-1)-associated Phosphatidylinositol 3-Kinase Activity in Muscle. J. Biol. Chem. 277, 50230-50236.
Zajac, J., Shrestha, A., Patel, P., and Poretsky, L. (2010). The main events in the history of diabetes mellitus. In Principles of diabetes mellitus. L. Poretsky, ed. (New York: Springer), pp. 3-16.
Zelezniak, A., Pers, T.H., Soares, S., Patti, M.E., and Patil, K.R. (2010). Metabolic network topology reveals transcriptional regulatory signatures of type 2 diabetes. PLoS Comput. Biol. 6, e1000729.
Zhang, P., Zhang, X., Brown, J., Vistisen, D., Sicree, R., Shaw, J., and Nichols, G. (2010). Global healthcare expenditure on diabetes for 2010 and 2030. Diabetes Res. Clin. Pr. 87, 293-301.
Zhao, H., Przybylska, M., Wu, I.H., Zhang, J., Siegel, C., Komarnitsky, S., . . . and Cheng, S.H. (2007). Inhibiting Glycosphingolipid Synthesis Improves Glycemic Control and Insulin Sensitivity in Animal Models of Type 2 Diabetes. Diabetes 56, 1210-1218.
Zimmet, P., Alberti, K.G.M.M., and Shaw, J. (2001). Global and societal implications of the diabetes epidemic. Nature 414, 782-787.
Zimmet, P.Z. (1999). Diabetes epidemiology as a tool to trigger diabetes research and care. Diabetologia 42, 499-518.
Özcan, U., Cao, Q., Yilmaz, E., Lee, A.-H., Iwakoshi, N.N., Özdelen, E., . . . and Hotamisligil, G.S. (2004). Endoplasmic Reticulum Stress Links Obesity, Insulin Action, and Type 2 Diabetes. Science 306, 457-461.