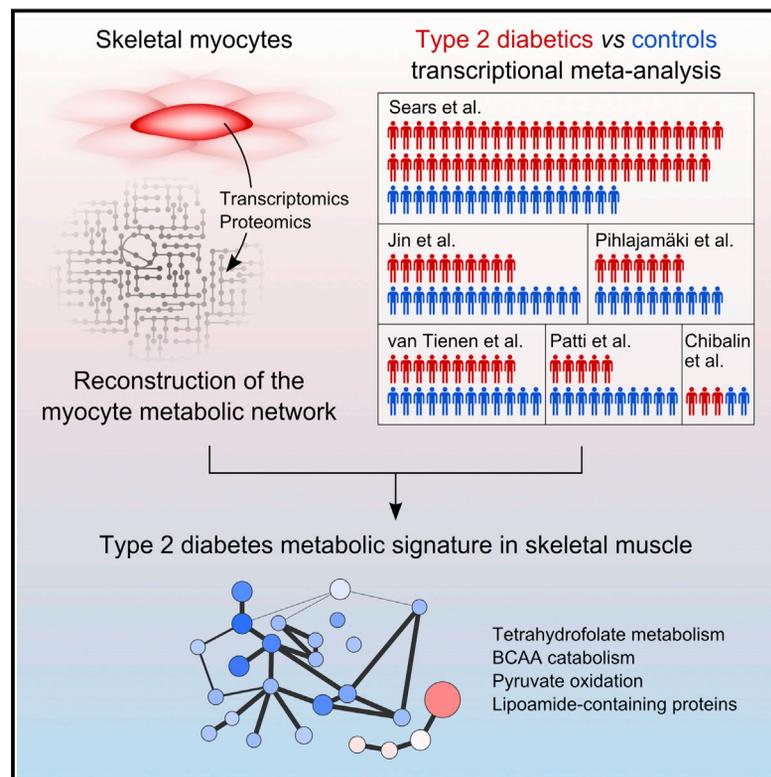


Proteome- and Transcriptome-Driven Reconstruction of the Human Myocyte Metabolic Network and Its Use for Identification of Markers for Diabetes

Graphical Abstract



Authors

Leif Väre, Camilla Scheele, ..., Bente Klarlund Pedersen, Jens Nielsen

Correspondence

nielsenj@chalmers.se

In Brief

Type 2 diabetes is associated with an altered metabolism in skeletal myocytes. To elucidate these metabolic changes at a systems level, Väre et al. reconstructed the myocyte metabolic network and integrated it with transcription data from six different studies, identifying a consistent metabolic signature of diabetic muscle.

Highlights

- Reconstruction of a comprehensive myocyte-specific genome-scale metabolic model
- Meta-analysis of type 2 diabetes transcription in skeletal muscle from six studies
- Transcriptional regulation of metabolism around pyruvate, BCAAs, and THF
- Myocyte metabolic model identifies a metabolic signature of diabetes

Accession Numbers

GSE63887



Proteome- and Transcriptome-Driven Reconstruction of the Human Myocyte Metabolic Network and Its Use for Identification of Markers for Diabetes

Leif Våremo,¹ Camilla Scheele,^{2,3} Christa Broholm,² Adil Mardinoglu,¹ Caroline Kampf,⁴ Anna Asplund,⁴ Intawat Nookaew,^{1,7} Mathias Uhlén,^{5,6} Bente Klarlund Pedersen,² and Jens Nielsen^{1,6,*}

¹Department of Biology and Biological Engineering, Chalmers University of Technology, 41296 Gothenburg, Sweden

²Centre of Inflammation and Metabolism and Centre for Physical Activity Research, Department of Infectious Diseases, Rigshospitalet, University of Copenhagen, 2100 Copenhagen Ø, Denmark

³Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, 2200 Copenhagen N, Denmark

⁴Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, 75185 Uppsala, Sweden

⁵Department of Proteomics, School of Biotechnology, AlbaNova University Center, Royal Institute of Technology (KTH), 10691 Stockholm, Sweden

⁶Science for Life Laboratory, Royal Institute of Technology (KTH), 17121 Stockholm, Sweden

⁷Present address: Comparative Genomics Group, Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

*Correspondence: nielsenj@chalmers.se

<http://dx.doi.org/10.1016/j.celrep.2015.04.010>

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

SUMMARY

Skeletal myocytes are metabolically active and susceptible to insulin resistance and are thus implicated in type 2 diabetes (T2D). This complex disease involves systemic metabolic changes, and their elucidation at the systems level requires genome-wide data and biological networks. Genome-scale metabolic models (GEMs) provide a network context for the integration of high-throughput data. We generated myocyte-specific RNA-sequencing data and investigated their correlation with proteome data. These data were then used to reconstruct a comprehensive myocyte GEM. Next, we performed a meta-analysis of six studies comparing muscle transcription in T2D versus healthy subjects. Transcriptional changes were mapped on the myocyte GEM, revealing extensive transcriptional regulation in T2D, particularly around pyruvate oxidation, branched-chain amino acid catabolism, and tetrahydrofolate metabolism, connected through the down-regulated dihydrolipoamide dehydrogenase. Strikingly, the gene signature underlying this metabolic regulation successfully classifies the disease state of individual samples, suggesting that regulation of these pathways is a ubiquitous feature of myocytes in response to T2D.

INTRODUCTION

Skeletal muscle is one of the most abundant tissues in the human body, accounting for more than a third of our body weight (Lorenz and Campello, 2001), and is primarily composed of myo-

cytes. Myocytes communicate with other tissues and organs, such as liver and adipose, through secretion of myokines (e.g., myostatin, follistatin, and interleukins) (Pedersen and Febbraio, 2012), and the metabolism of myocytes has a major impact on whole-body homeostasis. For instance, myocytes are responsible for roughly 75% of the insulin-stimulated clearance of glucose from the blood after a meal (Stump et al., 2006). As such, myocytes are highly susceptible to insulin resistance, and the dysfunction of myocyte metabolism is implicated in lifestyle diseases such as obesity, metabolic syndrome, and type 2 diabetes (T2D), all of which are to date globally ubiquitous (Chen et al., 2012).

These medical conditions are complex in their mode of action even in the context of a single tissue, like muscle, in that alterations at the systems level are required to contribute to their progression. This complexity can be explored through the generation of state-of-the-art genome-wide “omics” data (e.g., transcriptomics and proteomics) that provide a global and unbiased insight into myocyte function, dysfunction, and metabolism. Nevertheless, in order to gain as much knowledge as possible from such data, it is essential to have informative scaffolds for functional and mechanistic interpretation of genome-wide omics data. In this context, genome-scale metabolic models (GEMs) are useful representations of metabolism, and in the broad area of human disease, they have successfully been used for prediction and interpretation through simulation and data integration (Bordbar et al., 2014; Mardinoglu et al., 2013b; Våremo et al., 2013b). GEMs are, simply put, a list of mass-balanced metabolic reactions that are connected into a network structure through their shared metabolites. Each reaction is assigned to a specific compartment, e.g., the cytoplasm or mitochondria, and associated to genes known to code for proteins involved in that reaction. During the last decade, there has been an increased development and application of human GEMs resulting in generation of several so-called context-specific GEMs. These GEMs capture the active subset of metabolism present in a particular

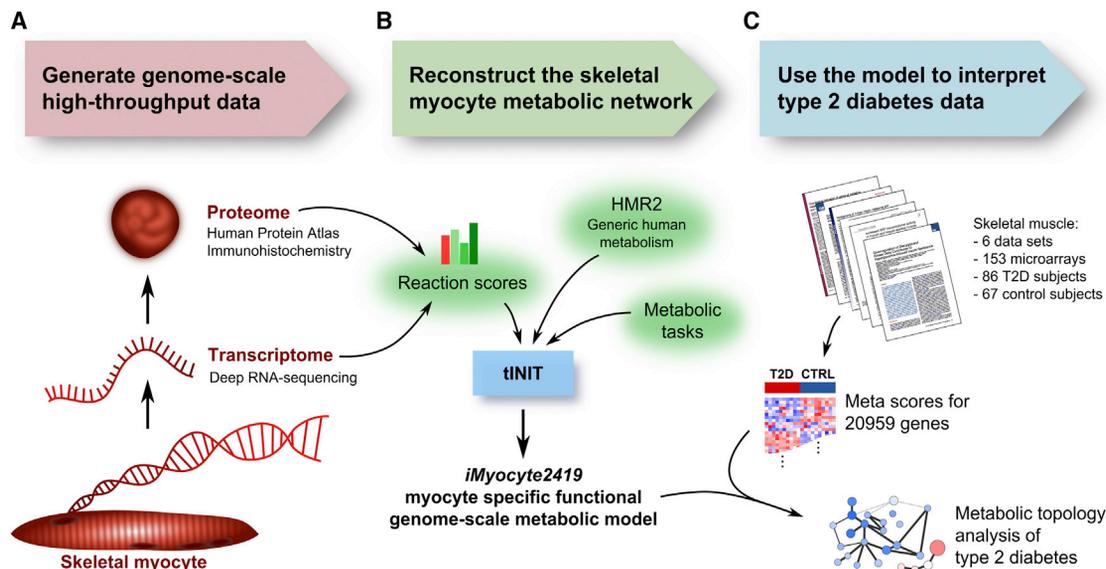


Figure 1. Overview of the Workflow

(A) Myocyte-specific transcriptome data were generated using RNA-seq and compared to immunohistochemistry-based proteome data from the Human Protein Atlas in order to evaluate the correlation and consistency between the transcript and protein levels.

(B) The proteome and transcriptome data were used to score the reactions in the generic human metabolic network HMR2 in order to evaluate the likelihood of a metabolic reaction being present in myocytes. These scores were used as input to the tINIT algorithm, together with a list of metabolic tasks that the myocyte GEM should be able to perform. The output of tINIT is a functional myocyte GEM that was manually curated, quality controlled, and compared to other existing GEMs.

(C) The network topology of iMyocyte2419 was used to contextualize and interpret T2D transcriptome data from six different studies, comprising in total 153 subjects.

context, such as a disease state (Agren et al., 2014) or specific tissues and cell types (Bordbar et al., 2011; Mardinoglu et al., 2013a, 2014).

Reconstructing context-specific (e.g., cell-type-specific) GEMs requires high-quality genome-wide data and a comprehensive human metabolic network to use as a template. To this end, we previously constructed the Human Metabolic Reaction (HMR) database (Agren et al., 2012), which has been continuously updated (Mardinoglu et al., 2013a) and is currently in version 2.0 (HMR2) (Mardinoglu et al., 2014). Reconstruction of a context-specific GEM from the generic metabolic network is a process of determining, for each metabolic reaction, whether or not it should be present, while simultaneously maintaining a connected network that represents a functioning model (Agren et al., 2012). The evidence for the presence of a reaction can be determined by experimentally measuring the existence of its components, i.e., the relevant metabolites and enzymes, and high-quality data with genome-wide coverage are therefore essential for the reconstruction process. We have previously constructed manually curated GEMs for the metabolically active tissues liver (Mardinoglu et al., 2014) and adipose (Mardinoglu et al., 2013a). Like skeletal muscle, these tissues are implicated in complex human metabolic diseases such as T2D, and there is a lot of metabolic interplay among the three tissues. It is therefore of great interest to reconstruct the skeletal myocyte metabolism in a comprehensive GEM.

Here, we generated cell-type-specific RNA-sequencing (RNA-seq) data for human myocytes and studied the correlation of this

data with antibody-based immunohistochemistry for myocytes from the Human Protein Atlas (Uhlen et al., 2010). These data represent an extensive description of myocyte gene expression at both the transcript and protein levels and enabled us to reconstruct a comprehensive cell-type-specific skeletal myocyte GEM. We next used the model to analyze and contextualize T2D transcription data collected from six different studies comprising in total 153 subjects. The meta-analysis of these six studies provided a systems-level description of metabolic regulation in T2D. Finally, we used the genes that underlie the identified metabolic changes to classify individual T2D samples from controls. An outline of the workflow is presented in Figure 1.

RESULTS

Cell-Type-Specific RNA-Seq Data for Human Skeletal Myocytes

To quantify the skeletal myocyte-specific transcriptome, muscle precursor cells were isolated from the vastus lateralis muscle of three male and three female individuals (Table 1) and differentiated into myotubes. These in vitro myocytes maintain genetic and stable epigenetic components from their human donor, resulting in a conservation of the in vivo phenotype in the in vitro system (Broholm et al., 2012; Green et al., 2011; Scheele et al., 2012). Thus, while cell lines rely on the fact that the gene expression in one single cell is representative of a cell population in a given tissue or representative of a given human population, our model system also takes biological variation into account. This,

Table 1. Characteristics of the Six Subjects

	Female 1	Female 2	Female 3	Male 1	Male 2	Male 3
Age (years)	50	51	66	49	51	51
Body mass index (kg/m ²)	23.0	24.0	24.8	23.7	23.9	24.4
Plasma glucose (mmol/l)	4.9	5.0	5.5	5.6	6.4	4.9
Plasma insulin (pmol/l)	25	17	36	26	24	20
HOMA-IR	0.47	0.38 ^a	0.69	0.5	0.48	0.38
Plasma total cholesterol (mmol/l)	5.0	6.2	5.4	4.3	5.0	6.8
Plasma HDL cholesterol (mmol/l)	2.0	3.0	1.7	1.0	1.3	2.0
Plasma LDL cholesterol (mmol/l)	2.7	3.0	3.1	2.7	3.2	4.2
Plasma triglyceride (mmol/l)	1.16	0.72	0.97	1.20	1.38	1.40
Sequencing depth (M paired-end reads)	48.0	84.9	62.6	45.8	68.9	49.5

Phenotypic characteristics, selected plasma levels, and sequencing depth for the six samples.

^aThe minimum allowed input (20 pmol/l) for plasma insulin was used for HOMA-IR calculation.

along with the fact that the culture only contains myocytes, in contrast to heterogeneous tissue biopsies, enables the reconstruction and analysis of myocyte-specific metabolism.

Isolated myocyte mRNA was sequenced at an average depth of 60 million paired-end reads per sample, and relative abundances in the form of FPKM (fragments per kilobase per million mapped reads) values were calculated for each protein-coding gene. The transcriptomes of the six samples showed high pairwise correlations (Figure 2A), and the vast majority of expressed genes (with FPKM >1) were expressed in all samples (Figure 2B). We identified 35 differentially expressed genes between males and females, including *NDUFA4L2*, *COX7A1*, *CD36*, *GPR116*, and *IL6*, indicating a higher activity of lipid and energy metabolism in female myocytes (see Supplemental Results and Table S1).

Correlation between Myocyte Transcriptome and Proteome

The immunohistochemistry-based protein abundance data from the Human Protein Atlas (HPA) version 12 (Uhlen et al., 2010) contains protein abundance levels (not detected, low, medium, or high) for 16,378 genes in skeletal myocytes. In HPA, the reliability of the abundance level of a given protein is either “supportive” or “uncertain,” depending on the extent and quality of evidence, and currently, 12,260 genes, 75% of the genes measured in skeletal myocytes, are associated with abundance levels with uncertain reliability (Figure 2C).

A positive correlation between mRNA and protein abundances has been reported (Lundberg et al., 2010; Nagaraj et al., 2011; Schwanhäusser et al., 2011), although transcript levels cannot fully explain the variation of protein concentrations in a cell, due to post-transcriptional events (Vogel and Marcotte, 2012). However, mRNA abundances are generally believed to be adequate for simply predicting protein presence/absence (Ramakrishnan et al., 2009).

We used the myocyte-specific RNA-seq data for evaluating if the myocyte proteome and transcriptome are correlated. We compared the FPKM-value distributions (median of the six samples) of genes with different protein abundance levels and separated the supportive and uncertain data. For the uncertain protein abundance levels, there was no clear difference between

the FPKM-value distributions of the non-detected and detected (low, medium, or high) proteins in HPA, but more prominently, for the more reliable supportive protein abundance levels, the detected proteins generally had distinctively higher FPKM values (Figure 2D), indicating a clear relation between expressed proteins and high transcript levels.

Refining the Protein Abundance Scores Using RNA-Seq Data

Since the observed difference in transcript levels between detected (low, medium, or high) and non-detected proteins is not apparent for the proteins with uncertain protein abundance levels, a portion of these proteins has likely been assigned an incorrect abundance level. To investigate this, we checked the FPKM-value distribution specifically for proteins with abundance levels that were updated, based on improved protein-level information (according to the standards of HPA), from being non-detected in previous versions of HPA to detected in the current version. The FPKM-value distribution of these proteins showed a similar pattern of higher values (Figure 2D), supporting the rationality behind an approach to use FPKM values to correct the uncertain protein abundance levels.

An FPKM ≥ 1 has previously been used as a threshold for determining protein presence (Fagerberg et al., 2014; Hebenstreit et al., 2011). This was based on the observation of a bimodal distribution of expression abundances that separated genes into two groups, referred to as lowly expressed (LE) and highly expressed (HE) genes (Hebenstreit et al., 2011). Our myocyte RNA-seq data also showed this pattern, and importantly, the FPKM values of the detected proteins with supportive reliability (most likely expressed on a protein level) matched the HE distribution (Figure 2E). For our data, the HE and LE group were also divided at an FPKM value around 1, so this value was selected as the cutoff.

We considered the HPA protein-level data as the primary source of enzyme evidence but used the RNA-seq data to reduce potential false negatives (as indicated in Figure 2C), i.e., assigning the abundance level low (detected) to the missing proteins in HPA and the uncertain non-detected proteins, if their FPKM ≥ 1 . Based on this, 2,460 genes with

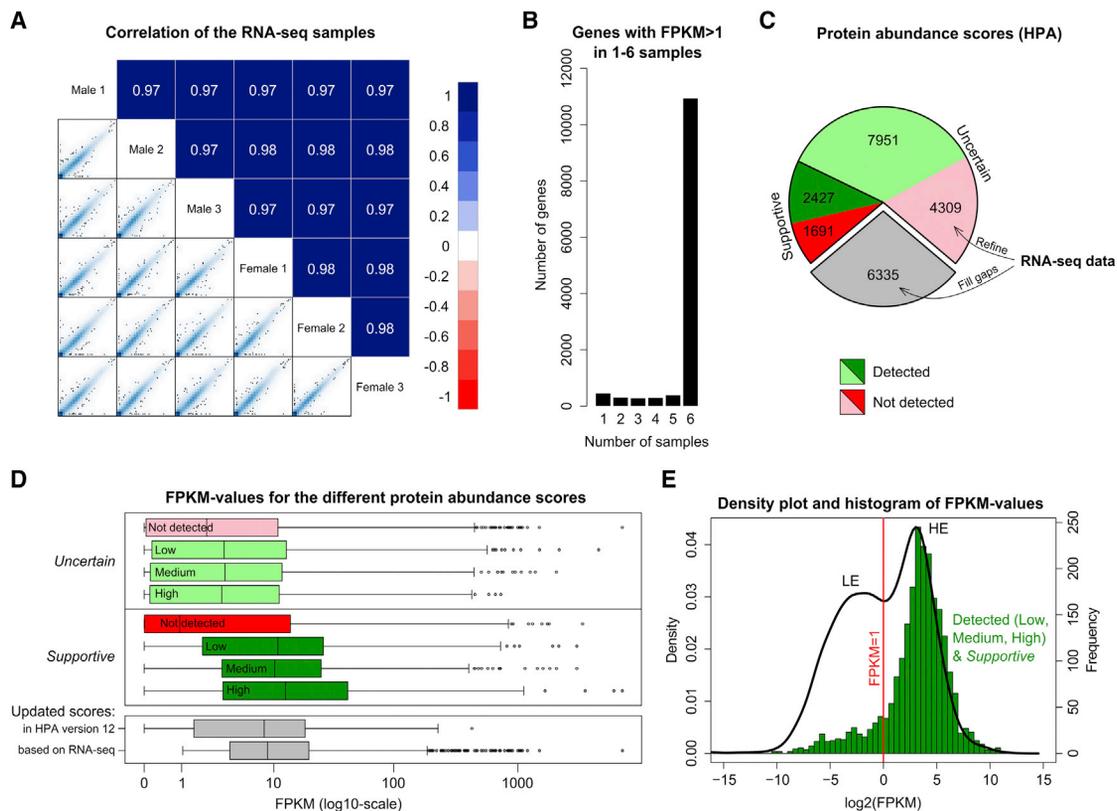


Figure 2. Comparison of Proteome and Transcriptome Data

(A) Pairwise Spearman correlations and scatterplots of $-\log_{10}(\text{FPKM values})$, for the six subjects, show a high consistency of their transcriptomes. (B) A majority of genes with an FPKM value larger than 1 are expressed in all six samples, showing a high consistency in whether or not a gene is expressed. (C) The pie chart shows a summary of the coverage of protein-coding genes in the Human Protein Atlas (HPA). A majority of the genes are associated with protein abundance levels with uncertain reliability, and more than 6,000 genes remain to be measured. (D) The (median) FPKM-value distributions are displayed separately for the genes of the different HPA abundance groups. For the genes with supportive protein abundance levels, there is a clear pattern of higher FPKM values for detected proteins (green) than non-detected proteins (red). This pattern is not apparent for the genes with uncertain abundance levels, indicating that a portion of these are incorrectly scored. Indeed, genes that were incorrectly scored in previous HPA versions but subsequently updated when new evidence was acquired according to the standards of HPA also show a pattern of higher FPKM values. (E) The FPKM-value distribution of all genes show two peaks, representing lowly expressed (LE) genes and highly expressed (HE) genes. The HE genes have been shown to be more likely to be translated into functional proteins. In line with this, the histogram of detected and supportive genes (in HPA) fits the HE part of the density plot. The two peaks are divided at an FPKM of roughly 1, and this was chosen as a cutoff, so that non-detected genes (in HPA) with uncertain protein abundance levels were updated to be detected, if their FPKM-value was larger than 1. The lowermost boxplot in (D) shows the FPKM values of these genes. See also Table S2.

uncertain non-detected protein abundance levels and $\text{FPKM} \geq 1$ were relabeled as detected (their FPKM values are summarized in the lowermost boxplot in Figure 2D). Additionally, 3,221 genes with protein-level information missing in HPA and with $\text{FPKM} \geq 1$ were also added as detected. Table S2 gives a summary of the updated genes and their number of associated reactions.

Reconstructing iMyocyte2419, a Comprehensive Myocyte GEM

The updated protein abundance levels were used to score the reactions in HMR2 so that a reaction received a positive score if at least one associated enzyme was determined detected in myocytes. During the reconstruction process, we also included known functionality of the myocytes by defining a list of 247 metabolic functions (or tasks) that the myocyte GEM should be

able to perform (Table S3), e.g., simulating uptake of fatty acids and cholesterol from lipoproteins. Using the tINIT (task-driven Integrative Network Inference for Tissues) algorithm (Agren et al., 2012, 2014), which is part of the RAVEN toolbox (Agren et al., 2013), a draft myocyte GEM was generated based on the positive and essential reactions, while ensuring a connected network where all reactions could carry flux. Next, a relatively low amount of reactions (22) were added so that the final model could perform all the required tasks. Finally, relevant exchange reactions were reintroduced and negative gene associations were removed. The resulting skeletal myocyte GEM, iMyocyte2419, consists of 5,590 reactions in eight different compartments, 4,448 metabolites (2,396 unique), and 2,419 genes. iMyocyte2419 can be downloaded in the systems biology markup language (SBML) format from the repository of curated models at <http://www.metabolicatlas.org>.

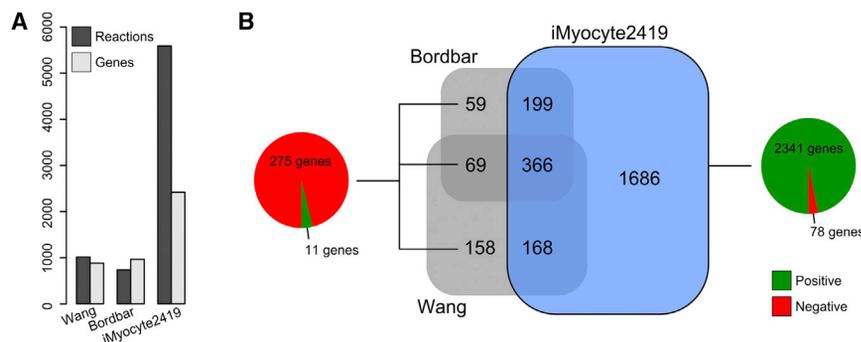


Figure 3. Comparison and Validation of the Myocyte GEM

(A) The myocyte GEM, iMyocyte2419, was compared to two previously published, yet smaller-scale, muscle GEMs. The barplot shows the number of genes and reactions in each model. (B) The Venn diagram shows the gene overlap among the three models. The pie charts show the gene scores (based on the RNA-seq and HPA data) for the indicated genes. As can be seen, the majority of the excluded genes in iMyocyte2419 have negative scores (non-detected), whereas only 78 out of 2,419 genes in iMyocyte2419 have a negative score.

See also [Tables S4](#) and [S5](#) for more details.

iMyocyte2419 Has Increased Metabolic Coverage Compared to Previous Models

Two small-scale GEMs for skeletal muscle, based on tissue-level omics data, have previously been published ([Bordbar et al., 2011](#); [Wang et al., 2012](#)). In terms of numbers of reactions and genes, both of these GEMs are significantly smaller than iMyocyte2419 ([Figure 3A](#)). Nevertheless, it is interesting to compare the overlap of the three GEMs ([Figure 3B](#)). In terms of gene content, 82% of the Bordbar GEM and 70% of the Wang GEM are encompassed by iMyocyte2419. There are 286 genes ([Table S4](#)) in the muscle GEMs that are excluded from iMyocyte2419, and the reason is that the majority of these have negative scores (not detected) based on the RNA-seq and proteomics data. Of the 11 genes with positive scores (detected), 4 are not included in HMR2, 4 were discarded during reconstruction in favor of preserving connectivity of the network, and 3 are actually included in iMyocyte2419 but under different Ensembl IDs. In contrast, the majority of the genes included in iMyocyte2419 are positively scored, whereas only 78 out of 2,419 genes have a negative score.

Although it is expected to have a small number of negatively scored genes in the final model, due to the constraints that are put to ensure a connected and functional network, we investigated why these genes were included. Out of the 78 genes with negative scores in iMyocyte2419, 58 are detected at the transcript level (FPKM value above 0) and 38 of these have FPKM values above 1 ([Table S5](#)). The 78 genes are associated with 292 reactions (5.2% of total reactions) that were not associated with any other positive genes. These reactions were required to be included in iMyocyte2419 either because they were part of pathways where several other proteins had a positive score (connectivity) or to successfully simulate all of the required metabolic tasks (functionality). Since the myocyte GEM provides a network-based representation of myocyte metabolism, the 78 genes with a negative score in HPA may actually be false negatives. To investigate this, we reevaluated the existing protein-level evidence in HPA for the 78 negative genes and, based on this, it was determined that 14 genes ([Table S5](#)) would be updated in HPA (version 13) as being detected in myocytes. This shows how the contextual analysis of proteins, provided by the GEM reconstruction, can assist in data curation.

Meta-Analysis of T2D Muscle Gene Expression

The topology of iMyocyte2419 provides a scaffold for interpreting high-throughput data in the context of myocyte metabolism. To establish a consensus result of the metabolic changes that occur in skeletal muscle of T2D subjects, we queried the Gene Expression Omnibus (GEO) data repository ([Edgar et al., 2002](#)) and ArrayExpress ([Rustici et al., 2013](#)) for relevant microarray datasets. This resulted in eight datasets that fitted our criteria (see [Experimental Procedures](#) for details).

Meta-analysis is used to combine multiple studies with similar hypotheses into one unified result. [Ramasamy et al. \(2008\)](#) studied the key issues in microarray meta-analysis and recommended the method proposed by Choi et al. ([Choi et al., 2003](#)), which we used here. Briefly, a Z score is calculated for each gene, and significant genes altered in T2D muscle can then be identified by selecting genes with absolute Z scores larger than a given threshold.

First, we calculated gene-wise Z scores for the eight individual datasets in order to calculate the Pearson correlation between each possible pair of studies. Surprisingly, even though the datasets were selected to include baseline skeletal muscle samples from people with T2D and healthy control subjects, two datasets (Gallagher and Wu) clearly stood out by showing negative correlation with the majority of the remaining datasets ([Figure 4A](#)). This pattern could not, as far as we could assess, be explained by the subject characteristics (e.g., body mass index, age, or fasting glucose levels), microarray platforms, or muscle types used in these datasets ([Table S6](#)). We also checked for possible sample contamination of immune cells by using the R-package ESTIMATE ([Yoshihara et al., 2013](#)), but again, no pattern was found that could explain the negative correlation of these two datasets with the rest ([Figure S1](#)). In order to perform a meta-analysis of a set of studies that represent a consistent gene expression phenotype, we decided to focus on the bigger group of six studies ([Table 2](#)) and exclude the Gallagher and Wu datasets from our meta-analysis. As it turned out, the meta-analysis of these six studies, comprising in total 86 T2D samples and 67 controls, recapitulated several known biological processes attributed to T2D (as described below), supporting the selection of these studies for further analysis.

One major advantage with meta-analysis is the increase in statistical power as the sample size increases, and this can be illustrated by the integrated discovery rate (IDR) ([Choi et al., 2003](#)),

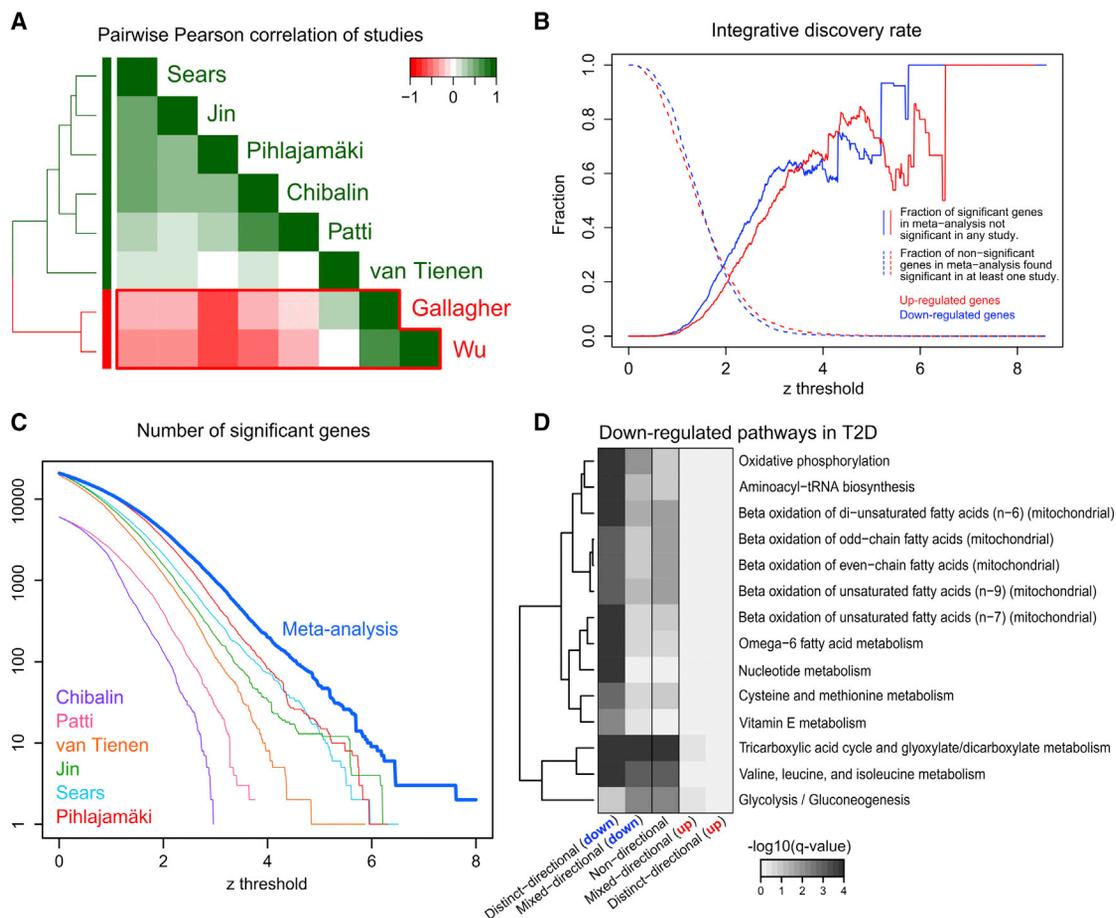


Figure 4. Meta-Analysis of Type 2 Diabetes Gene Expression

(A) A heatmap of pairwise Pearson correlations, based on the genes with $|Z| > 1$ in both compared studies. The Gallagher and Wu datasets showed a negative correlation with most of the remaining studies and cluster separately in the hierarchical clustering. In order to analyze an overall consistent phenotype, these two studies were not considered in the downstream analysis. See also Figure S1 and Tables 2 and S6 for details about the studies.

(B) The plot shows the fraction of significant genes, in the meta-analysis, that were not significant in any individual study, for different Z score cutoffs. The dashed lines similarly show the fraction of non-significant genes, in the meta-analysis, that were found significant in at least one of the individual studies. At a Z score cutoff around 3, in principle, no genes are lost from the individual studies, whereas the increased statistical power of the meta-analysis identifies a large fraction of significant genes that were not possible to detect in the individual studies.

(C) The plot shows the number of significant genes, in the meta-analysis and in each individual study, at different Z score cutoffs, showing the power of the meta-analysis.

(D) A gene-set analysis (GSA) of the meta-analysis results, using the pathways of iMyocyte2419 as gene sets, show pathways significantly affected by transcriptional regulation. The q-values (FDR-adjusted p values) are medians from the GSA methods used in the consensus GSA. In T2D, all pathways are either distinctly downregulated, or are affected by a subset of downregulated genes.

See also Figure S2 for the GO-term analysis and Tables S7, S8, and S9 for the meta-analysis Z scores and the complete GSA results.

i.e., the fraction of significant genes detected in the meta-analysis (by integrating studies) that is not detected in any of the individual datasets. Figure 4B shows the IDR for upregulated and downregulated genes, as well as the fraction of non-significant genes in the meta-analysis that were significant in any of the individual datasets, i.e., the potential loss of information, study-specific effects, or false positives in the individual datasets. At a Z score threshold of ~ 3 , close to 50% of the significant genes in the meta-analysis would not be identified by analyzing any of the individual datasets and practically no significant genes identified by any of the individual datasets were lost in the meta-analysis. In Figure 4C, the number of significant genes at different Z

score thresholds shows that the meta-analysis always detected more genes than any of the individual studies.

Integrating the Meta-Analysis Results with iMyocyte2419 Reveals Pathways Implicated in T2D

We used the gene-reaction associations to investigate if any specific metabolic processes were significantly affected by T2D gene expression. This was done through consensus gene-set analysis (GSA) (Väremo et al., 2013a) with the meta-analysis gene-wise Z scores as input data (Table S7) and the iMyocyte2419 pathways as gene sets. Figure 4D shows the metabolic pathways most significantly affected by differential gene

Table 2. Summary of the Microarray Datasets Used in the Meta-Analysis

Author	T2D	NGT	Platform	Genes	Muscle Type	Dataset ID	PubMed ID
van Tienen et al. (2012)	10	12	HG-U133 Plus 2	20899	vastus lateralis	GSE19420	22802091
Sears et al. (2009)	51	18	HG-U133 Plus 2	20899	vastus lateralis	GSE13070	19841271
Pihlajamäki et al. (2011)	7	10	HG-U133 Plus 2	20899	rectus abdominis	GSE22435	21803291
Jin et al. (2011)	10	15	HG-U133 Plus 2	20899	vastus lateralis	GSE25462	21393865
Patti et al. (2003)	5	10	Hu6800	6060	vastus lateralis	GSE21340	12832613
Chibalin et al. (2008)	3	2	Hu6800	6060	vastus lateralis	E-MEXP-1270	18267070

T2D, number of type 2 diabetes samples; NGT, number of normal glucose-tolerant (control) samples; platform, Affymetrix microarray type; genes, number of unique Ensembl gene IDs; muscle type, the origin of the biopsy; dataset ID, GEO accession number (except for the Chibalin dataset, which comes from ArrayExpress).

expression in T2D, all showing patterns of downregulation. These include several processes suggested to be implicated in T2D, such as oxidative phosphorylation, β -oxidation, tricarboxylic acid cycle, glycolysis, and branched-chain amino acid (BCAA) (valine, leucine, and isoleucine) metabolism (Abdul-Ghani and DeFronzo, 2010; Lynch and Adams, 2014; Szendroedi et al., 2012). Additionally, other less studied gene sets also appear to be affected by transcriptional downregulation, namely omega-6 fatty acid metabolism, vitamin E metabolism, nucleotide metabolism, and cysteine and methionine metabolism.

In a similar manner, we used Gene Ontology (GO) terms as gene sets, and this also pointed at regulation of metabolism, in particular downregulation of genes coding for proteins acting in the mitochondria, respiratory electron transport chain, oxidoreductase activity, pyruvate metabolism, tricarboxylic acid cycle, glucose metabolic process, and BCAA catabolic process (Figure S2). Additionally, GO terms related to RNA splicing (RNA helicase activity, RNA splicing, and mRNA processing) were enriched by downregulated genes. A decreased expression of splicing factors in muscle of obese subjects has previously been reported by one of the studies included in the meta-analysis (Pihlajamäki et al., 2011), and interestingly, this signature is recapitulated also when taking all six studies into account. Finally, the GO-term analysis also revealed upregulation of several immune-related processes, which is in line with the evidence of inflammation being associated with the development of T2D (Donath and Shoelson, 2011). The complete pathway and GO-term GSA results are presented in Tables S8 and S9.

Topology-Based Analysis of the Meta-Analysis Results Using iMyocyte2419

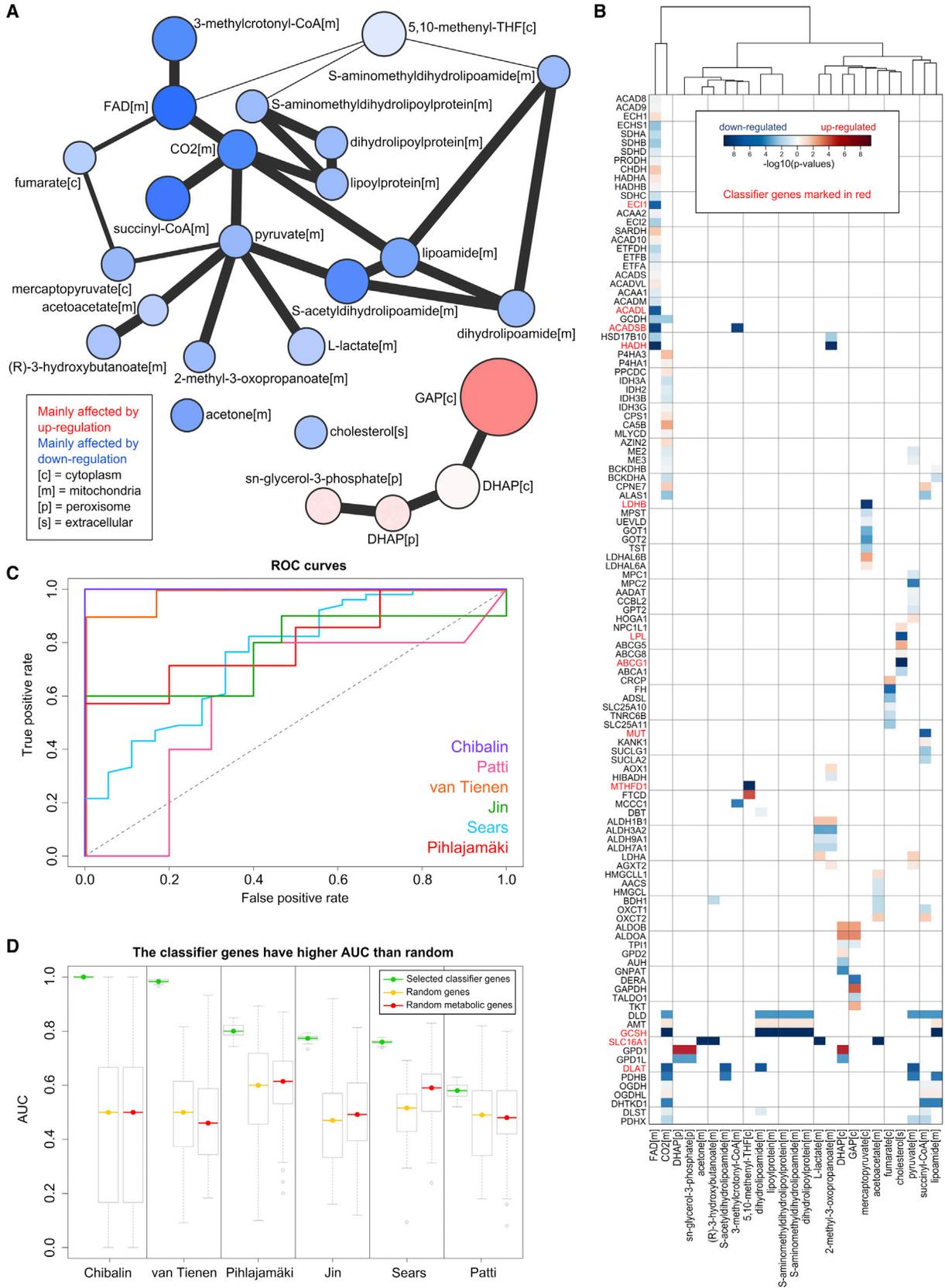
We next used the transcriptional data to identify reporter metabolites (Patil and Nielsen, 2005), which are metabolite gene sets defined by the gene-reaction associations provided by iMyocyte2419, so that for a given metabolite, a gene set consists of all genes associated with reactions in which the metabolite is participating. Consensus GSA was performed (complete results are shown in Table S10), and the GSA visualization tool Kiwi was used to process the results in order to highlight the most significant metabolites and their interconnectivity in the myocyte metabolic network. In order to demonstrate the value of a myocyte-specific GEM, we also performed the same analysis using the generic human metabolic network, HMR2, which yielded similar, but not identical, results (Figure S3). The two results

have a Pearson correlation of 0.89 (based on the non-directional median adjusted p values). This difference is due to the changes made to the gene-metabolite edges to reflect the myocyte-specific metabolic topology, and this shows the value of cell-type-specific models for integrative analysis.

As can be seen in the network plot (Figure 5A), most of the implicated metabolites are mitochondrial (marked with [m]) and affected by transcriptional downregulation (blue nodes) in T2D. In addition to the general downregulation of mitochondrial processes, the network plot is recapitulating on a more detailed level some of the implicated parts of metabolism that were identified in the pathway and GO-term analyses. The mitochondrial import of pyruvate and its conversion to acetyl-coenzyme A (acetyl-CoA) is affected by transcriptional downregulation indicated by the reporter metabolites: pyruvate, CO₂, S-acetyldihydrolipoamide, lipoamide, and dihydrolipoamide. The underlying genes associated with the reporter metabolites are displayed in Figure 5B, which illustrates significant downregulation of the mitochondrial pyruvate carrier *MPC2*, as well as *PDHB*, *DLAT*, and *DLD* (all taking part in pyruvate oxidation). There is also an apparent reduction of TCA cycle activity, shown by the reporter metabolites fumarate, succinyl-coenzyme A (succinyl-CoA), CO₂, and FAD, as well as FADH₂, H⁺, NADH, and NAD⁺ (high-degree metabolites that were excluded from the plot to preserve readability).

In line with the observed downregulation of BCAA metabolism in the pathway and GO-term analysis, the reporter metabolites 3-methylcrotonyl-coenzyme A, 2-methyl-3-oxopropanoate, dihydrolipoamide, lipoamide, and succinyl-CoA, intermediates of BCAA degradation, are all surrounded by downregulated genes. In particular, dihydrolipoamide dehydrogenase (*DLD*) is part of the branched-chain alpha-keto acid dehydrogenase complex (*BCKDC*), 3-methylcrotonyl-CoA carboxylase (*MCCC1*) is catalyzing a key step in leucine degradation, short/branched chain specific acyl-CoA dehydrogenase (*ACADSB*) is catalyzing an intermediate step in valine and isoleucine catabolism, and the aldehyde dehydrogenase *ALDH3A2* and methylmalonyl CoA mutase (*MUT*) are catalyzing some of the last steps of the breakdown of valine and isoleucine to form succinyl-CoA.

The reporter metabolite 5,10-methenyl-THF is an intermediate in folate one-carbon metabolism. It is one of several tetrahydrofolate (THF) derivatives (also including 10-formyl-THF, 5,10-methylene-THF, and 5-methyl-THF), which act as one-carbon group donors in nucleotide synthesis, methionine synthesis,



(legend on next page)

purine synthesis, and DNA methylation. Interestingly, two genes, *MTHFD1* and *FTCD*, are both associated with this metabolite but are significantly regulated in different directions. The cytosolic methylenetetrahydrofolate dehydrogenase (*MTHFD1*) is significantly downregulated in T2D and is catalyzing the interconversions of 5,10-methylene-THF, 5,10-methenyl-THF, and 10-formyl-THF. Formimidoyltransferase cyclodeaminase (*FTCD*), on the other hand, is significantly upregulated and functions to contribute with one-carbon units from histidine degradation to the folate pool.

Interestingly, the emergent downregulation surrounding the lipoylproteins and lipoamides connects several of the implicated parts of metabolism identified above, as they play an important role in pyruvate, BCAA, and THF metabolism. The key enzyme that is connecting these parts (as can be identified by the heatmap in Figure 5B) is the transcriptionally downregulated DLD, which is located in the mitochondrial matrix, where it is a component of the pyruvate dehydrogenase complex, the branched-chain alpha-keto acid dehydrogenase complex, and the glycine decarboxylase complex.

Finally, a separate subnetwork affected by transcriptional regulation, which emerges from the reporter metabolite analysis, is that of glyceraldehyde 3-phosphate (GAP), dihydroxyacetone phosphate (DHAP), and sn-glycerol-3-phosphate. In particular, GAP is the most statistically significant reporter metabolite, connected to increased transcriptional activity, driven by the significantly upregulated genes transketolase (*TKT*), aldolase A and B (*ALDOA* and *ALDOB*), and GAP dehydrogenase (*GAPDH*). Also glycerol-3-phosphate dehydrogenase 1 (*GPD1*), connected to DHAP and sn-glycerol-3-phosphate, is upregulated, but there is also an influence of downregulated genes (*DERA*, *GNPAT*, and *GPD1L*). In summary, this points toward an active transcriptional regulation around the branch point of glycolysis, the pentose phosphate pathway and lipid biosynthesis, in T2D skeletal muscle.

The Identified Metabolic Signatures Have the Power to Classify T2D Samples

The subnetwork of reporter metabolites establish a metabolic signature of T2D in muscle, across six different studies. In order to determine if this signature was prominent enough to be recognized at the level of individual subjects, we performed a sample classification approach. To do this, we selected the most significant ($p < 1e-5$) genes among the ones connected to the reporter

metabolites (marked in red in Figure 5B). These 12 genes are associated with 20 out of 25 metabolite gene sets, thus generally representing the affected identified metabolic subnetwork.

First, we investigated to what extent each gene influenced the myocyte function by blocking the associated reactions in iMyocyte2419 and testing what metabolic tasks (Table S3) the GEM would fail to carry out. The 12 genes disrupted in total 105 of the 247 metabolic tasks (Table S11), and the biggest effect in terms of number of failed tasks was observed for *ACADL* and *HADH*, both disrupting 55 tasks.

Next, for each study, the disease state (healthy or diabetic) of each sample was predicted based on a random forest model trained on the expression values of the 12 genes in the remaining samples. The performance of the classification is shown by receiver-operating characteristic (ROC) curves for the six studies (Figure 5C). The area under the ROC curve (AUC) is shown in Figure 5D, while repeating the classification 100 times for each study, and can be compared to the AUC scores of using random genes (orange) or random metabolic genes (red) as classifiers. The AUC scores are larger than random in all six datasets, illustrating the relevance of these 12 representative genes on a sample-wise level, across several studies. This suggests that the expression patterns of these genes and the corresponding effects in metabolism are a ubiquitous feature of T2D in muscle.

DISCUSSION

Here, we reconstructed the skeletal myocyte metabolism and by that provide a comprehensive myocyte GEM. This was in large part enabled by the high-quality proteome data from HPA and the RNA-seq data generated in this study, together providing an extensive description of myocyte-specific gene expression. Human genome-wide data are often collected from tissue samples, but in order to study the myocyte metabolism, our goal was to make a cell-type-specific, rather than tissue-specific, GEM. In this context, the myocyte-specific RNA-seq data were invaluable for the reconstruction process. The use of in vitro myocytes could arguably be a step away from in vivo biology; however, data show that these cells are able to recapitulate known phenotypes of their donors (Broholm et al., 2012; Green et al., 2011; Scheele et al., 2012), including known differences between males and females, such as higher lipid and energy metabolism in females, as shown here.

Figure 5. Metabolic Subnetworks Implicated in Type 2 Diabetes

(A) Using the tool Kiwi, the network of significant reporter metabolites was produced. An edge between two metabolites means that they are less than three steps apart in the metabolite-metabolite network (extracted from iMyocyte2419), and the thickest edges connect the metabolites that are closest to each other. In order to avoid too unspecific connections, metabolite nodes with a degree above 73 (mainly co-factors) were removed from the network prior to calculating the metabolite-metabolite distances. The node size corresponds to the gene-set (reporter metabolite) significance (median of non-directional p value), and the color captures the general direction of change. See also Table S10 for the complete reporter metabolite results and Figure S3 for a comparison to the same analysis while using HMR2.

(B) The heatmap shows the significance and direction of regulation of the genes connected to the metabolites in (A).

(C) The most significant genes, underlying the metabolic subnetwork (marked in red in the heatmap) were used to classify the disease state (healthy or diabetic) of individual samples, using a random forest model. The performance of the classification is shown by (a representative) receiver-operating characteristic (ROC) curve for each study of the meta-analysis. See also Table S11.

(D) The area under the ROC curve (AUC) was calculated 100 times for each study by repeating the random forest classification (green bars) and can be compared to the AUC scores of classification based on random or random metabolic genes (yellow and red bars). The high AUC scores, and the fact that they are larger than the AUC scores from random genes, show that the metabolic signature shown in (A) is relevant also on the level of individual T2D subjects.

The myocyte GEM, iMyocyte2419, is a valuable resource that can be used in future research on muscle metabolism, providing a scaffold for data integration and support for analyzing genome-wide data, as demonstrated in this study. Furthermore, the GEM can be used for simulation of muscle metabolism, and there is an increasing interest in studying the metabolic interactions between cells and tissues and the implication this has on blood metabolite levels and whole-body metabolism. An approach to accomplish this task is to connect multiple GEMs (Bordbar et al., 2011), which requires a proper simulation framework and high-quality cell-type-specific GEMs. The myocyte GEM, together with our previous published models of hepatocyte and adipocyte metabolism (Mardinoglu et al., 2013a, 2014), constitutes an essential step in this direction.

The value of our reconstructed GEM for analysis of transcription data was here clearly demonstrated by the identification of metabolic anomalies associated with T2D, some consistent with earlier findings but several not reported before. A set of genes associated with oxidative phosphorylation have previously been shown to be downregulated in human diabetic muscle tissue (Mootha et al., 2003), and this, together with mitochondrial dysfunction, has repeatedly been reported as one of the signatures of T2D muscle (Szendroedi et al., 2012). Conversely, other studies report a normal mitochondrial function in T2D (Boushel et al., 2007) and no difference in oxidative phosphorylation (De Feyter et al., 2008) or the expression of associated genes (Frederiksen et al., 2008). The Gallagher study (Gallagher et al., 2010), which we briefly considered here, found no difference in expression of genes in the oxidative phosphorylation gene set, as opposed to Mootha et al. (2003), and speculated that this may be because the subjects in the latter study were exposed to insulin prior to biopsy sampling, suggesting that a reduced oxidative phosphorylation could be explained by a dissimilar acute insulin response rather than a baseline difference.

In order to not be confounded by data from a single study, we performed a meta-analysis in order to establish a consistent result across multiple datasets. This analysis covered T2D gene expression in human muscle from six studies covering 153 subjects, and the findings reported here are thus supported on a wider base. Following the speculations of Gallagher et al. (2010), during the selection of the studies for the meta-analysis, we excluded all studies where subjects had been exposed to any insulin stimuli (e.g., a hyperinsulinemic-euglycemic clamp) prior to biopsy sampling. Regardless, the results from our pathway analysis showed a significant downregulation of oxidative phosphorylation. Thus, at least across the six studies of the meta-analysis, there is an association between T2D and reduced expression of oxidative phosphorylation genes, although it still remains to be investigated if this is a cause or an effect of T2D. The Gallagher study, which reported no difference in oxidative phosphorylation between diabetics and controls, was excluded from the main meta-analysis (together with the Wu dataset) due to negative correlation with the other six studies, but while running our analyses, no pathways were found to be significantly changed (at $q < 0.05$) in a separate meta-analysis of these two datasets. Furthermore, when including these two datasets in the meta-analysis, oxidative phosphorylation was still found significantly downregulated.

The topology of iMyocyte2419 gave us a unique opportunity to further dissect the meta-analysis results to the level of individual metabolites in a network-based approach. This analysis revealed connected metabolic subnetworks affected by T2D transcriptional regulation, in particular around pyruvate oxidation, BCAA catabolism, and tetrahydrofolate metabolism. These results were found to be very robust markers of T2D, and based on the underlying top significantly differentially expressed genes, we were able to classify the disease state of individual samples, yielding high AUC scores (Figure 5D).

The downregulation surrounding 5,10-methenyl-THF has, to our knowledge, not been reported before in human diabetic skeletal muscle. Interestingly, two of the pathways, methionine and nucleotide metabolism, that THF-derivatives are precursors for were also identified as affected by transcriptional downregulation in the pathway analysis. There is also an influence of both upregulation and downregulation around 5,10-methenyl-THF. The downregulated *MTHFD1* suggests a decreased activity around interconversion of THF intermediates, whereas there is an upregulation of *FTCD* that could indicate a contribution into THF metabolism from histidine catabolism. Plasma histidine has been reported to be associated with positive effects on diabetes, such as reduced blood glucose levels (Stancáková et al., 2012), triglycerides, and proinflammatory cytokines (Lee et al., 2005). Histidine has also been shown to reduce gluconeogenesis in liver (possibly by activating IL6/STAT3 signaling, but without increasing insulin levels), and it was therefore suggested as a candidate target for T2D treatment (Kimura et al., 2013). Speculatively, increased histidine catabolism in muscle could reduce plasma histidine, limiting its reported positive effects on T2D. When we specifically checked the expression of genes in our meta-analysis belonging to the histidine degradation pathway, we did coincidentally find an upregulation ($p < 0.07$) of both histidine ammonia-lyase (*HAL*) and urocanate hydratase 1 (*UROC1*). *UROC1* has previously been found to be demethylated in the placenta of rats with gestational diabetes (Petropoulos et al., 2014).

Increased plasma levels of BCAAs have been associated with T2D, and it has been hypothesized that BCAAs can promote insulin resistance either by activating mTORC1 signaling, which in turn can lead to inhibition of insulin signaling, or by leading to toxic accumulation of BCAA metabolites through disruptive BCAA metabolism, which could lead to mitochondrial dysfunction (Lynch and Adams, 2014). In line with our results, a decreased expression of BCAA catabolism in muscle has been reported before (Lefort et al., 2010). It has also been shown that insulin lowers plasma BCAA levels and increases the expression in liver of *DLD* and 3-methylcrotonyl-CoA carboxylase (Shin et al., 2014), both genes that we identified with opposite patterns in the meta-analysis, i.e., downregulated in diabetic muscle. These authors also reported that insulin lowered liver levels of 3-methylcrotonyl-CoA, an intermediate in BCAA catabolism and a metabolite that was identified in our analysis as a reporter metabolite in diabetic muscle. Interestingly, our network analysis also identified lipoamide-containing proteins, connected to the downregulation of *DLD*, as contributors not only to decreased BCAA catabolism but also to the downregulation that we observed around pyruvate and THF metabolism. *DLD*

is catalyzing the reaction of dihydrolipoamide to lipoamide, which is needed as an intermediate step in the carboxylation of pyruvate, branched-chain alpha-keto acids (which is a step in BCAA degradation), and glycine (at the same time forming 5,10-methylene-THF from THF).

In conclusion, our extensive analysis, using multiple transcriptome datasets, comprehensive metabolic modeling, consensus GSA, and network topology analysis, provided a holistic insight into the metabolic state of type 2 diabetic human muscle. The identified metabolic signature of T2D was confirmed by classification of the sample disease states, across multiple cohorts, based on gene expression patterns. Finally, the myocyte-specific GEM iMyocyte2419, which was reconstructed based on myocyte-specific transcriptome and proteome data, provides a valuable and useful resource for future studies.

EXPERIMENTAL PROCEDURES

Subject Selection and Characteristics

Six normal glucose tolerant and non-obese males and females (aged 49–66 years, body mass index 23.0–24.8 kg/m²) were selected from a previous cohort (Green et al., 2011; Møller et al., 2014). All participants gave written informed consent before inclusion, and the study was performed according to the Declaration of Helsinki and approved by The Regional Committee on Biomedical Research Ethics in Denmark (KF 01-141/04). HOMA-IR values were calculated using the HOMA2 Calculator 2.2.3 (<https://www.dtu.ox.ac.uk/homacalculator/>).

Culturing of Myocytes and Sampling of RNA

Muscle precursor cells (satellite cells) were isolated from human skeletal muscle biopsy specimens obtained from the vastus lateralis muscle using a biopsy needle with suction (Bergström, 1975) as described in detail previously (Green et al., 2011). Isolated satellite cells were cultured and fully differentiated (see Supplemental Experimental Procedures). Total RNA was extracted using Trizol (Life Sciences) according to the manufacturer's instructions. Frozen samples were shipped to the sequencing facility where they were purified by poly(A) enrichment using Illumina TruSeq RNA.

RNA-Seq and Data Analysis

The raw sequence reads were trimmed from adaptor sequences and aligned to the human genome using STAR (Dobin et al., 2013). FPKM values and gene counts were calculated using Cufflinks (Trapnell et al., 2010) and HTSeq-count (Anders et al., 2015). Differential expression was assessed using edgeR (Robinson et al., 2010). See Supplemental Experimental Procedures for more details.

Reconstructing the Myocyte GEM

To reconstruct the myocyte GEM, the gene scores acquired from the Human Protein Atlas and the RNA-seq data were used by the tINIT algorithm, together with HMR2 as reference model and the metabolic tasks defined in Table S3. See further details in Supplemental Experimental Procedures.

Meta-Analysis

The GEO database and ArrayExpress were queried for datasets that contained subjects with T2D or insulin resistance, and that had not been given any stimuli (such as insulin or drug) prior to biopsy sampling, according to the description given in the corresponding publications, which resulted in eight datasets. Six of these (see Table 2) were used in the final analysis, and two (GEO accession numbers GSE22309 and GSE18732) were excluded due to negative correlation with the others. The datasets were preprocessed (see Supplemental Experimental Procedures) and the meta-analysis was carried out according to Choi et al. (2003), and as implemented in the Bioconductor R package GeneMeta, using a random-effects model, yielding one Z score for each gene.

Pathway and GO-Term Analysis

The pathway and GO-term analyses were performed using the consensus GSA approach implemented in the Bioconductor R package piano (Våremo et al., 2013a). Each GSA method generates its own gene-set p values based on gene permutation. The consensus approach then yields a consensus score for each gene set, and the median p value for each gene set is reported (see details in Supplemental Experimental Procedures). The heatmap in Figure 4D shows $-\log_{10}$ (median of false discovery rate [FDR]-adjusted p values) of the pathways with a median adjusted $p < 0.005$ (in at least one of the directionality classes), and the heatmap in Figure S2 shows $-\log_{10}$ (median of FDR-adjusted p values) of the GO terms with a median adjusted $p < 0.05$ (in the non-directional class).

Reporter Metabolite Analysis

Reporter metabolite analysis was performed using piano, as described above, and the metabolite-reaction-gene associations of iMyocyte2419 (and HMR2) to create the metabolite gene sets. The Python package Kiwi (Våremo et al., 2014) was used to make the network plot (Figure 5A). For this, the metabolite-metabolite network was reduced to metabolites with degrees ≤ 73 (removing mainly co-factors). The p value cutoff was set to 0.001, and the shortest path-length (SPL) cutoff was set to 3. The network plot layout was manually adjusted in Cytoscape.

Random Forest Classification

The random forest classification was carried out, using the R package randomForest, by leave-one-out cross-validation, so that the disease state of each sample was predicted using a model based on gene expression data from the remaining samples. See Supplemental Experimental Procedures for further details.

ACCESSION NUMBERS

The accession number for the raw and processed RNA-seq data reported in this paper is GEO: GSE63887.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Results, Supplemental Experimental Procedures, three figures, and 11 tables and can be found with this article online at <http://dx.doi.org/10.1016/j.celrep.2015.04.010>.

AUTHOR CONTRIBUTIONS

L.V. analyzed all data, reconstructed and validated the myocyte genome-scale metabolic model, carried out the meta-analysis and the random forests classification, and drafted the manuscript. C.S. and C.B. isolated satellite cells, were involved in the study design, and carried out the myocyte culturing and RNA sample collection. A.M. assisted in the reconstruction and validation of the myocyte genome-scale metabolic model and in the interpretation of the meta-analysis results. C.K. and A.A. carried out the reevaluation of the protein classification in the Human Protein Atlas, corresponding to the negative genes in the myocyte genome-scale metabolic model. I.N. assisted in the RNA-seq data analysis and the meta-analysis. J.N. conceived the study. M.U., B.K.P., and J.N. designed and coordinated the study. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

This work was funded by the Knut and Alice Wallenberg Foundation and the Bill and Melinda Gates Foundation. The publishing fee was funded by Chalmers Library. The Centre of Inflammation and Metabolism (CIM) is supported by a grant from the Danish National Research Foundation (DNRF55). The Centre for Physical Activity Research (CFAS) is supported by a grant from Trygfondene. CIM is part of the UNIK Project Food, Fitness & Pharma for Health and Disease, supported by the Danish Ministry of Science, Technology, and Innovation. CIM is a member of DD2, the Danish Center for Strategic Research in Type 2

Diabetes (the Danish Council for Strategic Research, grant no. 09-067009 and 09-075724). The Novo Nordisk Foundation Center for Basic Metabolic Research (<http://www.metabol.ku.dk>) is supported by an unconditional grant from the Novo Nordisk Foundation to University of Copenhagen. The authors would like to thank Francesco Gatto for valuable ideas and discussions. The computations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at C3SE.

Received: December 9, 2014

Revised: February 6, 2015

Accepted: April 3, 2015

Published: April 30, 2015

REFERENCES

- Abdul-Ghani, M.A., and DeFronzo, R.A. (2010). Pathogenesis of insulin resistance in skeletal muscle. *J. Biomed. Biotechnol.* *2010*, 476279.
- Agren, R., Bordel, S., Mardinoglu, A., Pornputtpong, 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., Liu, L., Shoaie, S., Vongsangnak, W., Nookaew, I., and Nielsen, J. (2013). The RAVEN toolbox and its use for generating a genome-scale metabolic model for *Penicillium chrysogenum*. *PLoS Comput. Biol.* *9*, e1002980.
- 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*, 721.
- Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* *31*, 166–169.
- Bergström, J. (1975). Percutaneous needle biopsy of skeletal muscle in physiological and clinical research. *Scand. J. Clin. Lab. Invest.* *35*, 609–616.
- Bordbar, A., Feist, A.M., Usaite-Black, R., Woodcock, J., Palsson, B.O., 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., Monk, J.M., King, Z.A., and Palsson, B.O. (2014). Constraint-based models predict metabolic and associated cellular functions. *Nat. Rev. Genet.* *15*, 107–120.
- 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.
- 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.
- 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.
- Chibalin, A.V., Leng, Y., Vieira, E., Krook, A., Björholm, M., Long, Y.C., Kotova, O., Zhong, Z., Sakane, F., Steiler, T., et al. (2008). Downregulation of diacylglycerol kinase delta contributes to hyperglycemia-induced insulin resistance. *Cell* *132*, 375–386.
- Choi, J.K., Yu, U., Kim, S., and Yoo, O.J. (2003). Combining multiple microarray studies and modeling interstudy variation. *Bioinformatics* *19* (Suppl 1), i84–i90.
- 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.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* *29*, 15–21.
- Donath, M.Y., and Shoelson, S.E. (2011). Type 2 diabetes as an inflammatory disease. *Nat. Rev. Immunol.* *11*, 98–107.
- Edgar, R., Domrachev, M., and Lash, A.E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* *30*, 207–210.
- Fagerberg, L., Hallström, B.M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., Habuka, M., Tahmasebpoor, S., Danielsson, A., Edlund, K., et al. (2014). Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol. Cell. Proteomics* *13*, 397–406.
- Frederiksen, C.M., Højlund, K., Hansen, L., Oakeley, E.J., Hemmings, B., Abdallah, B.M., Brusgaard, K., Beck-Nielsen, H., 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.J., Scheele, C., Keller, P., Nielsen, A.R., Remenyi, J., Fischer, C.P., Roder, K., Babraj, J., Wahlestedt, C., Hutvagner, G., et al. (2010). Integration of microRNA changes in vivo identifies novel molecular features of muscle insulin resistance in type 2 diabetes. *Genome Med.* *2*, 9.
- 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.
- 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*, 497.
- Jin, W., Goldfine, A.B., Boes, T., Henry, R.R., Ciaraldi, T.P., Kim, E.-Y., Emeican, M., Fitzpatrick, C., Sen, A., Shah, A., et al. (2011). Increased SRF transcriptional activity in human and mouse skeletal muscle is a signature of insulin resistance. *J. Clin. Invest.* *121*, 918–929.
- Kimura, K., Nakamura, Y., Inaba, Y., Matsumoto, M., Kido, Y., Asahara, S., Matsuda, T., Watanabe, H., Maeda, A., Inagaki, F., et al. (2013). Histidine augments the suppression of hepatic glucose production by central insulin action. *Diabetes* *62*, 2266–2277.
- 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., Brophy, C., Meyer, C., Højlund, K., Yi, Z., 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.
- Lorenz, T., and Campello, M. (2001). Biomechanics of skeletal muscle. In *Basic Biomechanics of the Musculoskeletal System*, Third Edition, M. Nordin and V.H. Frankel, eds. (Philadelphia: Lippincott Williams & Wilkins), pp. 148–174.
- Lundberg, E., Fagerberg, L., Klevebring, D., Matic, I., Geiger, T., Cox, J., Algenäs, C., Lundeberg, J., Mann, M., and Uhlen, M. (2010). Defining the transcriptome and proteome in three functionally different human cell lines. *Mol. Syst. Biol.* *6*, 450.
- Lynch, C.J., and Adams, S.H. (2014). Branched-chain amino acids in metabolic signalling and insulin resistance. *Nat. Rev. Endocrinol.* *10*, 723–736.
- Mardinoglu, A., Agren, R., Kampf, C., Asplund, A., Nookaew, I., Jacobson, P., Walley, A.J., Froguel, P., Carlsson, L.M., Uhlen, M., and Nielsen, J. (2013a). Integration of clinical data with a genome-scale metabolic model of the human adipocyte. *Mol. Syst. Biol.* *9*, 649.
- Mardinoglu, A., Gatto, F., and Nielsen, J. (2013b). Genome-scale modeling of human metabolism - a systems biology approach. *Biotechnol. J.* *8*, 985–996.
- 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*, 3083.
- Møller, J.B., Pedersen, M., Tanaka, H., Ohsugi, M., Overgaard, R.V., Lynge, J., Almind, K., Vasconcelos, N.-M., Poulsen, P., Keller, C., et al. (2014). Body composition is the main determinant for the difference in type 2 diabetes pathophysiology between Japanese and Caucasians. *Diabetes Care* *37*, 796–804.

- Mootha, V.K., Lindgren, C.M., Eriksson, K.F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., et al. (2003). PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **34**, 267–273.
- Nagaraj, N., Wisniewski, J.R., Geiger, T., Cox, J., Kircher, M., Kelso, J., Pääbo, S., and Mann, M. (2011). Deep proteome and transcriptome mapping of a human cancer cell line. *Mol. Syst. Biol.* **7**, 548.
- 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., Miyazaki, Y., Kohane, I., Costello, M., Saccone, R., et al. (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.
- Pedersen, B.K., and Febbraio, M.A. (2012). Muscles, exercise and obesity: skeletal muscle as a secretory organ. *Nat. Rev. Endocrinol.* **8**, 457–465.
- Petropoulos, S., Guillemin, C., Ergaz, Z., Dimov, S., Suderman, M., Weinstein-Fudim, L., Ornoy, A., and Szyf, M. (2014). Gestational diabetes alters offspring DNA methylation profiles in human and rat: identification of key pathways involved in endocrine system disorders, insulin signaling, diabetes signaling and IL-K signaling. *Endocrinology*, Published online December 16, 2014. <http://dx.doi.org/10.1210/en.2014-1643>.
- Pihlajamäki, J., Lerin, C., Ikonen, P., Boes, T., Floss, T., Schroeder, J., Dearie, F., Crunkhorn, S., Burak, F., Jimenez-Chillaron, J.C., et al. (2011). Expression of the splicing factor gene SFRS10 is reduced in human obesity and contributes to enhanced lipogenesis. *Cell Metab.* **14**, 208–218.
- Ramakrishnan, S.R., Vogel, C., Prince, J.T., Li, Z., Penalva, L.O., Myers, M., Marcotte, E.M., Miranker, D.P., and Wang, R. (2009). Integrating shotgun proteomics and mRNA expression data to improve protein identification. *Bioinformatics* **25**, 1397–1403.
- 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.
- Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140.
- Rustici, G., Kolesnikov, N., Brandizi, M., Burdett, T., Dylag, M., Emam, I., Farne, A., Hastings, E., Ison, J., Keays, M., et al. (2013). ArrayExpress update—trends in database growth and links to data analysis tools. *Nucleic Acids Res.* **41**, D987–D990.
- Scheele, C., Nielsen, S., Kelly, M., Broholm, C., Nielsen, A.R., Taudorf, S., Pedersen, M., Fischer, C.P., 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.
- Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature* **473**, 337–342.
- Sears, D.D., Hsiao, G., Hsiao, A., Yu, J.G., Courtney, C.H., Ofrecio, J.M., Chapman, J., and Subramaniam, S. (2009). Mechanisms of human insulin resistance and thiazolidinedione-mediated insulin sensitization. *Proc. Natl. Acad. Sci. USA* **106**, 18745–18750.
- Shin, A.C., Fasshauer, M., Filatova, N., Grundell, L.A., Zielinski, E., Zhou, J.Y., Scherer, T., Lindtner, C., White, P.J., Lapworth, A.L., et al. (2014). Brain insulin lowers circulating BCAA levels by inducing hepatic BCAA catabolism. *Cell Metab.* **20**, 898–909.
- Stancáková, A., Civelek, M., Saleem, N.K., Soininen, P., Kangas, A.J., Cederberg, H., Paananen, J., Pihlajamäki, J., Bonnycastle, L.L., Morken, M.A., et al. (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.
- 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.
- 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.
- Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515.
- Uhlen, M., Oksvold, P., Fagerberg, L., Lundberg, E., Jonasson, K., Forsberg, M., Zwahlen, M., Kampf, C., Wester, K., Hober, S., et al. (2010). Towards a knowledge-based Human Protein Atlas. *Nat. Biotechnol.* **28**, 1248–1250.
- 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., de Coo, I.F.M., Nicolay, K., Prompers, J.J., Smeets, H.J.M., and van Loon, 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.
- Våremo, L., Nielsen, J., and Nookaew, I. (2013a). Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. *Nucleic Acids Res.* **41**, 4378–4391.
- Våremo, L., Nookaew, I., and Nielsen, J. (2013b). Novel insights into obesity and diabetes through genome-scale metabolic modeling. *Front. Physiol.* **4**, 92.
- Våremo, L., Gatto, F., and Nielsen, J. (2014). Kiwi: a tool for integration and visualization of network topology and gene-set analysis. *BMC Bioinformatics* **15**, 408.
- Vogel, C., and Marcotte, E.M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* **13**, 227–232.
- Wang, Y., Eddy, J.A., and Price, N.D. (2012). Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. *BMC Syst. Biol.* **6**, 153.
- Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., Treviño, V., Shen, H., Laird, P.W., Levine, D.A., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612.