Metabolome and Proteome Based Analysis of Muscular Dystrophies

Inaugural-Dissertation
to obtain the academic degree
Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry and Pharmacy
of Freie Universität Berlin

by

Tobias Opialla

from Berlin

August 2016

myograd INTERNATIONAL RESEARCH TRAINING GROUP FOR MYOLOGY PARIS | BERLIN
CHARITÉ UNIVERSITÄTSMEDIZIN BERLIN
BERLIN INSTITUTE OF HEALTH Charité & Max Delbrück Center
MDC MAX DELBRÜCK CENTER FOR MOLECULAR MEDICINE BERLIN-BUCH
BIMSB THE BERLIN INSTITUTE FOR MEDICAL SYSTEMS BIOLOGY

Die vorgelegte Arbeit wurde zwischen Juni 2012 und August 2016 zu gleichen Teilen am

Experimental and Clinical Research Center – a joint cooperation of the
Charité Medical Faculty and the Max Delbrück Center for Molecular Medicine

und am

Berlin Institute for Medical Systems Biology
at the Max-Delbrück-Center for Molecular Medicine

unter Anleitung von Dr. Stefan Kempa und Prof. Dr. Simone Spuler erstellt.

1. Gutachter: Dr. Stefan Kempa
2. Gutachterin: Prof. Dr. Simone Spuler

Datum der Disputation: 12. Oktober 2016

Acknowledgements

Well I must say: “I had a lot of fun!”
And I am extremely greatful for all opportunities and freedom I had during my thesis. Of course it is challenging to work in two labs in parallel, but it is definitely worth the effort. I got the best of two worlds and this is more, than one can ask for.

My greatest thanks goes to my two supervisors, Prof. Dr. Simone Spuler and Dr. Stefan Kempa, for providing me with the surroundings that enabled me to learn, laugh, play music and simply have a great time doing what I love: Science.

Science always means to collaborate, and so I am indebted to many many people.

Thank you Sarah, that you laid the ground work, I can now stand upon. I want to thank Matthias and Christin for many hours of tinkering, so that when I started, I could use your method. Also your help, introducing me to Lizzy and later setting up Phineas and Ferb was invaluable to me. I’d also like to thank for the provision of VANTED-templates by Olga and Matthias.

Doing bioinformatics was relatively new to me, when I started. But thanks to Chris and Henning, and also Christin, who gave me the first introduction to the world of R, this now is another tool I can use. I would like to thank Guido for always taking extreme good care of Toni and KITT, and introducing me to the world of proteomics, where also Chris was always a great help.

Also Alina, you were invaluable to me. Especially in the end of the thesis, when labwork was out of the question for me, it was really calming to know my samples in your capable hands. Another world that was quite new to me, when I started was to work with living things. I would like to thank Stefanie und Stephanie and Stefanie for always having another look at my cells, when I was unsure about their wellbeing. And Andreas and Tobias for showing me another world of beauty. Microscopy is truly an art.

Thank you Adrienne for helping me with the mice. Without your skill and ideas the data surely would have not have reached this quality level. And I truly hope, that this project can soon reach its full potential, when histology is included.

More thanks go to Séverine, who provided me with a great dataset. I had fun playing.

Verena, thank you so much for everything. Crazy lab times, screwing and unscrewing what felt like millions of falcons. Selecting mice. Always asking cool questions. Answering mine. And also for being a great friend.

Another huuuge thank you goes to Henning, for being the perfect lab mate. I think we think alike and I hope that we find a way to work together again at some point in the future. Working on the campus run was the most fun thing I experienced thus far in science. But also I now know my first three guitar chords and (somewhat) how to fly a plane.

Thank you Sabrina and Susanne, for helping me with all of the paperwork, that comes with being in two institutes at the same time.

I also want to thank all the people who agreed to donate their tissue, so that I could perform research with it. I hope it will help to ameliorate or perhaps someday heal the diseases you are suffering from. Thanks go to all people from the outpatient clinic for making this possible.

I’d like to thank the team from Zinsser to let me feel a bit like an engineer. Thank you

Fardad for stepping up, when you had to. I'm looking very much forward to finally playing with the robot.

In the final phase of my thesis, I had a bad accident. But thanks to the support of my colleagues and many people I got to know over the last year I can walk again. It is really good to know, to have such support. My thanks go to the whole team of Praxis Morand. You have magic hands. And quite flexible schedules. I thank the team from the Alpenstückle for providing me with another living room and work place.

The most long due thank you I owe to my parents and especially my mother. For inspiring me from early on to be curious and stay that way. Supporting me with all you could.

I am unendingly indebted to you Christine. For caring, supporting being there for me in crisis and joy.

Abstract

Muscle represents the means to move. This movement needs energy, which is generated by the central carbon metabolism. Respective to function, different types of muscle have developed. Two main categories exist: Red (TI) oxidative slow twitch muscles optimized for long lasting contractions and white (TII) glycolytic fast twitch muscles, which are differently affected by aging and disease. Metabolic profiles are altered in muscle disease, by changed enzyme levels or regulation.

I generated detailed metabolic and proteomic profiles from different muscle fiber types; *Extensor digitorum longus*, *Tibialis anterior* (both glycolytic), *Quadriceps* (glycolytic/oxidative) and *Soleus* (oxidative) of C57BL/6N mice.

I found a distinct molecular make up for each muscle. While current classification holds expectedly true, my systems biology approach allows for a more detailed look at the interplay between metabolites and proteins.

Dysferlinopathy, a hereditary muscular dystrophy, manifests in puberty, when glycolysis becomes prominent.

In her thesis S. Keller [1] subjected a mouse dysferlinopathy model (BLA/J) and human myotubes to MS-based metabolomics and proteomics and found a metabolic phenotype: Enzyme levels were mostly unchanged between diseased and control samples. Glycolysis intermediates including glucose-6-phosphate were reduced, polyol pathway members sorbitol and fructose were increased.

Based on work by M. Pietzke and Ch. Zasada *et al.* [2] I established label tracing methods including the glycogen pool in primary human myotubes.

This allowed to demonstrate function loss of hexokinase II in dysferlinopathy. Label incorporation (LI) was slower in glycolytic intermediates and glycogen. LI into sorbitol was unchanged.

Glycolysis impairment leads to more oxidative metabolism, causing oxidative stress, reflected in elevated pentose phosphate pathway poolsizes.

In the cell the membrane repair protein Dysferlin is located in dysferlin vesicles. I subjected fractions enriched for dysferlin vesicles by S. Kunz [3] to proteomics and found well established interaction partners, but also glycolytic enzymes never discussed before in the literature. This further establishes a metabolic role of dysferlin.

The techniques used in these studies require muscle tissue samples. Together with H. Kuich [4] we performed a proof of principle study on a single human volunteer during exercise to assess the reflection of muscle metabolism in easier obtainable blood. 10 µL blood samples from the fingertip, similar to a diabetics blood glucose test, sufficed to recapitulate well established knowledge of sports medicine and find a possible molecular explanation for the phenomenon “Hitting the Wall”. The samples show relatedness according to “rate of perceived exhaustion” possibly reflecting energy availability in the muscle. We could attribute the relations to nine key metabolites. This workflow might lead to easier muscle disease diagnostics in health care.

Zusammenfassung

Muskel ist der Hauptverbraucher von Energie im Körper. Diese Energie wird mittels des zentralen Kohlenstoff Metabolismus gewonnen. Zwei Hauptkategorien von Muskeln existieren: oxidative, langsam kontrahierende (Typ 1) und glykolytische, schnell kontrahierende (Typ 2).

Die verschiedenen Fasertypen zeigen unterschiedliche Anfälligkeiten für Krankheiten und den natürlichen Alterungsprozess.

Dies macht die Unterschiede zwischen Muskeln auf molekularer Ebene interessant. Ich generierte Metabolom- und Proteomdaten von Muskeln mit spezifischen Eigenschaften; *Extensor digitorum longus*, *Tibialis anterior* (beide glykolytisch), *Quadriceps* (gemischt) und *Soleus* (oxidativ) von Mäusen.

Es zeigte sich, das jeder Muskel eine spezifische molekulare Zusammensetzung besitzt. Die klassischen Kategorisierungsparameter bleiben bestehen, aber der systembiologische Ansatz erlaubt einen diffenrenzierteren Blick auf das Zusammenspiel von Metaboliten und Proteinen.
Dysferlinopathie ist eine Erbkrankheit, die sich während der Pubertät manifestiert, wenn sich der adulte glykolytische Stoffwechsel einstellt.

Meine Vorgängerin S. Keller [1] hatte Muskeln von BLA/J-Mäusen (ein Dysferlinopathie-Modell) und humane Myotuben analysiert und einen metabolischen Phänotyp der Krankheit gefunden: Glykolytische Enzyme waren unverändert, aber Glucose-6-phosphat und nachfolgende Metabolite waren vermindert, Sorbitol und Fructose erhöht.

Basierend auf Arbeit von Pietzke und Zasada *et al.* [2] etablierte ich eine Methode zur Messung der Labelinkorporation inklusive Glycogen in primären humanen Myotuben und konnte zeigen, dass tatsächlich die Funktion von Hexokinase II vermindert ist.

Verminderte Glykolyseaktivität führt zu einer Steigerung des oxidativen Metabolismus und erhöhtem oxidativen Stressniveau, was das beobachtete erhöhte Niveau an Pentosephosphatweg-Metaboliten zeigt.

In der Zelle liegt das Membranreparaturprotein Dysferlin in Dysferlinvesikeln vor. Diese Vesikel wurden von S. Kunz [3] isoliert. Im Vesikelproteom konnte ich etablierte Interaktionspartner und glykolytische Enzyme finden, die zwar z. T. in der Literatur bekannt sind, aber nie diskutiert wurden.

Diese Techniken basieren alle auf Muskelbiopsien. Gemeinsam mit H. Kuich [4] evaluierte ich im Rahmen einer Machbarkeitsstudie Blut als leichter erhältliches Probenmaterial, aus dem Muskelmetabolismus potentiell ablesbar ist. Aus 10 µL großen Blutproben, ähnlich einem Blutzuckertest, und gesammelt unter körperlicher Belastung, konnten wir etabliertes Wissen der Sportmedizin nachvollziehen und eine Erklärung für das Phänomen „Hitting the Wall“ finden. Insbesondere zeigten die Proben größte Ähnlichkeit gemäß der selbst empfundenen Erschöpfung, was potentiell Rückschlüsse auf den Energiezustand des Muskels zulässt. Wir konnten 9 Schlüsselmetabolite bestimmen, auf denen diese Verwandtschaft beruht. Diese Analyse-Pipeline könnte eine Translation von Metabolomics in den klinischen Alltag der Muskeldiagnostik ebnen.

Contents

Abstract 7

Zusammenfassung 9

List of Tables 15

List of Figures 15

List of Abbreviations 18

1 Muscle is the major site of metabolism 20

1.1 Central Carbon Metabolism 20
1.2 In muscle, energy is stored in various Forms 22
1.3 Muscle communicates with other organs via the blood stream 23
1.4 Different muscletypes show specialised metabolism 23
1.5 Myosin isoforms characterize fiber types 25
1.6 Muscle diseases are fiber type specific 26
1.7 The onset of dysferlinopathy coincides with changes in metabolism 26
1.8 Separation coupled mass spectrometry – a powerful tool in systems biology 28
1.9 Label tracing allows for measurement of metabolite poolsizes and formation rate at the same time 31
1.10 Inference of meaning from omics data by aggregation into concepts . . . 32
1.11 Integration of different omics datasets gives another information layer . . 33
1.12 A detailed look at muscular dystrophies with mass spectrometry 33

2 Methods 36

2.1 Study approval 36
2.2 Solutions and chemicals used 36
2.2.1 Solutions and chemicals used for metabolomics 37
2.2.2 Solutions and chemicals used for proteomics 38
2.3 Glycogen precipitation and determination with anthrone reaction 38
2.4 Test of linearity between signal and sample amount 38
2.5 Collection and extraction of EDL, QUAD, SOL and TA of C57BL/6N mice 39
2.6 Labeling of myotubes 40
2.6.1 Cell culture 40
2.6.2 Labeling of myotubes 41

2.6.3 Glycogen hydrolysis . . . 42
2.7 Extraction of blood samples obtained from a single healthy volunteer performing exercise . . . 42
2.8 Derivatization, GC–MS measurement and annotation of metabolites . . . 43
2.8.1 Derivatization . . . 43
2.8.2 GC–MS measurement . . . 43
2.8.3 Annotation . . . 44
2.9 Protein digestion, LC–MS measurement, protein identification . . . 44
2.9.1 Protein digestion of muscle tissue samples . . . 44
2.9.2 Protein digestion of dysferlin vesicles . . . 44
2.9.3 LC–MS measurement . . . 44
2.9.4 Protein annotation . . . 45
2.10 Statistics, plot preparation, software used . . . 45

3 Results **47**
3.1 Establishment of measuring metabolites from muscle tissue . . . 48
3.2 Comparing muscle fiber types . . . 52
3.2.1 Data handling . . . 53
3.2.2 Data analysis . . . 55
3.2.2.1 Separation by PCA . . . 57
3.2.2.2 Enriched terms from principal components . . . 60
3.2.2.3 Separation by principal component–discriminant function analysis (PC-DFA) . . . 68
3.2.2.4 Each muscletype has its own molecular makeup . . . 76
3.3 Label incorporation in myotubes from dysferlinopathy patients . . . 78
3.3.1 Establishment of pSIRM in muscle cells . . . 79
3.3.2 Uptake of substrate, with all carbon sources present . . . 80
3.3.3 Including glycogen into the pSIRM-Workflow . . . 82
3.3.4 pSIRM applied to primary human muscle cells of dysferlinopathy patients . . . 85
3.3.4.1 Poolsizes in myotubes from dysferlinopathy patients show a conserved metabolic phenotype . . . 85
3.3.4.2 Label incorporation is not changed in sorbitol but in glycogen and G6P . . . 89
3.4 Dysferlin interacts with membranous, mitochondrial and glycolytic proteins 91

3.4.1 Description of the dataset . . . 92
3.4.2 Dysferlin directly interacts with proteins involved in RNA-processing 94
3.4.3 Dysferlin vesicles are transport organelles within the cell . . . 95
3.4.4 Glycolytic proteins are enriched in the vesicular fraction . . . 95
3.5 Metabolomic profiles from single drops of capillary blood . . . 101
3.5.1 Subjective feeling throughout exercise . . . 101
3.5.2 Fuel source availability throughout exercise . . . 105
3.5.3 Metabolic insights when "Hitting the Wall" . . . 106
3.5.4 Metabolic states are reflected in subjective perceptions . . . 107

4 Discussion **110**
4.1 The established concepts of fiber types are reflected in the data . . . 110
4.2 Myosin heavy chain isoforms separate oxidative and predominantly glycolytic muscles . . . 112
4.3 Metabolite and enzyme levels do not always correlate . . . 113
4.4 Changes in poolsizes in dysferlinopathy are stable, irrespective of the fuel source . . . 117
4.5 Labeling experiments confirm an impairment of hexokinase II in dysferlinopathy . . . 118
4.6 The proteome of dysferlin vesicles reveals secondary interactions with metabolic enzymes . . . 121
4.7 Possible pseudo-fiber type switch dysregulates overall metabolism in dysferlinopathy — Reflections of oxygen availability . . . 123
4.8 Exercise and self-reported energy state are reflected in the metabolome of blood . . . 125
4.9 The minimal invasive sampling of blood from the fingertip might provide a tool in muscle disease diagnostics . . . 126
4.10 An individual could serve as its own control . . . 128
4.11 The combination of omics techniques allows to find more holistic insights into diseases . . . 129

5 Conclusions–Towards a more holistic medicine **132**

6 Publications **135**

References **136**

7 Appendix **151**
7.1 Example of a typical dotplot 151
7.2 Volcanoplot of the changes in poolsizes depicted in figure 3.3.8 151
7.3 Treemaps for GO-term enrichment in dysferlin pulldowns with different cutoffs 151

List of Tables

1 Properties of slow (Type I) and fast (Type II) muscles. 26
2 Extraction solutions used for linearity test 39
3 Volumes of PP dried to assess linearity between signal and sample amount 39
4 Labeling times used in tracing experiment of dysferlin defficient primary human myotubes 41
5 Characterization of individuals who donated their tissue to the pulsed stable isotope resolved metabolomics (pSIRM) experiment. 85
6 Proteins found to be direct interaction partners of dysferlin. 98
7 Proteins quantified in the current study, that are also mentioned by Flix, *et al.*, 2013 or de Morrée, *et al.*, 2010 99
8 GO-terms found to be enriched in vesicular fraction 100
9 Fiber type distribution in mouse muscle 112

List of Figures

1.8.1 Levels of omics 28
3.1.1 Analysis of linear correlation between sample amount and signal intensity. 51
3.2.1 Picture of a mouse hindlimb 53
3.2.2 Screeplot for data subsets 58
3.2.3 PCA plots for all subsets of data 59
3.2.4 Treemap of GO-terms enriched in PC1 62
3.2.5 Treemap of GO-terms enriched in PC3 63
3.2.6 Plot of *p*-values for term enrichment in PC1 65
3.2.7 Plot of *p*-values for term enrichment in PC2 66
3.2.8 Plot of *p*-values for term enrichment in PC3 67
3.2.9 Plot of linear discriminants of the respective data-subsets 69
3.2.10 Treemap of GO-terms enriched in LD1 70
3.2.11 Treemap of GO-terms enriched in LD3 71
3.2.12 Plot of *p*-values for term enrichment in LD1 73
3.2.13 Plot of *p*-values for term enrichment in LD2 74
3.2.14 Plot of *p*-values for term enrichment in LD3 75
3.2.15 Network of the CCM, data of muscle fiber types. 77
3.3.1 Experimental setup with only one carbon source present 80
3.3.2 Experimental setup with all carbon sources present 81

3.3.3 Labelincorporation in healthy myotubes after 30 min 81
3.3.4 Solubility of glycogen in different solvents 82
3.3.5 Distribution of glycogen in variations of standard metabolomics extraction method. 84
3.3.6 Label-incorporation in glycogen 84
3.3.7 Scheme of experimental setup for labeltracing in myotubes from patients that suffer from dysferlinopathy 86
3.3.8 Network representation of poolsize foldchange 88
3.3.9 Label incorporation into key metabolites of polyol pathway and glycolysis 90
3.4.1 Heatmap of dysferlin pulldowns 93
3.4.2 Volcanoplot of pulldowns with rAB vs. rabbit IgG 94
3.4.3 Volcanoplot of protein amounts in vescicular fraction vs. direct IP . . . 96
3.4.4 Significance values for glycolytic enzymes detected in samples enriched by different extraction protocols 97
3.5.1 Network of all metabolites measured in blood 102
3.5.2 Covariances for all metabolites measured, including unknowns 103
3.5.3 Running track and lap times 103
3.5.4 Plots of the metabolites used to assess credibility of the experiment . . . 104
3.5.5 Plots of free fatty acids and glycerol 105
3.5.6 Plots of TCA intermediates 106
3.5.7 Unsupervised hierarchical clustering of all features measured 107
3.5.8 Unsupervised hierarchical clustering of features characterizing the metabolic states 108
3.5.9 Circos plot of the selected metabolites characterizing the metabolic states described 109
4.2.1 Hierarchical clustering of the samples based on myosin heavy chain isoforms 112
4.2.2 Abundance index of all myosin-heavy-chain isoforms quantified in the four muscle types 114
4.5.1 Upper part of glycolysis in different muscle types 120
7.1.1 Exemplary Cleveland plot 152
7.2.1 Volcanoplot of foldchange in poolsize of the respective metabolites and corresponding *p*-values, normal controls vs. patients with dysferlinopathy 153
7.3.1 Treemaps of GO-Terms found enriched in pulldowns from total lysate of myotubes 154

7.3.2 Treemaps of GO-Terms found enriched in pulldowns from the vesicular fraction of myotubes 155

List of Abbreviations

2,3-BPG: 2,3-bisphospho-glycerate
2DRo: 2-deoxyribose
2OHG: 2-hydroxyglutarate
3OHB: 3-hydroxybutyrate
6PG: 6-phosphogluconate
AABA: 2-amino-butyrate
AcCoA: acetyl-CoA
ADP: adenosin di-phosphate
AGE: glycation end-product
αKG: α-ketoglutarate
ATP: adenosin tri-phosphate
CCM: central carbon metabolism
CK: creatine kinase
CoA: coenzyme A
DNA: deoxyribonucleicacid
F1,6BP: fructose-1,6-bisphosphate
F6P: fructose-6-phosphate
FAD: flavin adenine dinucleotide
FCS: foetal calf serum
FDR: false discovery rate
FFA: free fatty acid
G6P: glucose-6-phosphate
GC: gas chromatography
GO: gene ontology
GSH: glutathione
Glu: glutamate
HK2: hexokinase II
***i*Cit:** iso-citrate
IDH: isocitrate dehydrogenase
LGMD2B: limb girdle muscular dystrophy 2B
LOQ: limit of quantification
LP: lipid phase
MCW: methanol-chloroform-water

MDH: malate dehydrogynase
MS: mass spectrometry
MYH: myosin-heavy-chain
NADH: nicotinamide adenine dinucleotide
NADPH: nicotinamide adenine dinucleotide phosphate
PC-DFA: principal component–discriminant function analysis
PCA: principal component analysis
PCR: polymerase chain reaction
PCr: phosphocreatine
PPP: pentose phosphate pathway
PP: polar phase
pSIRM: pulsed stable isotope resolved metabolomics
PTS: phospho transferase system
RI: retention index
R5P: ribose-5-phosphate
Rib: ribose
RNA: ribonucleic acid
ROS: reactive oxygen species
RPE: rate of perceived exertion
Ru5P: ribulose-5-phosphate
SKMGM: skeletal muscle cell growth medium
TCA: tricarboxylic acid
TCAi: TCA-cycle intermediate
Xyl: xylose

1 Muscle is the major site of metabolism

Muscle, as a kinetic organ is a major site for metabolism, especially for turnover of sugar and fat. [5, *and references therein*]

The science of substrate interconversion, its educts and products, but also the enzymes involved is part of the fields of metabolomics and proteomics. [6] Often mass spectrometry based protocols are used to determine many species of molecules from a single sample. This gives a simultaneous overview of metabolic pathways used to generate energy equivalents in the form of ATP[1] as well as building blocks for cell maintenance. The combination of pathways along which energy is generated, but also building blocks like nucleotides or lipids are produced is called the central carbon metabolism (CCM). [7] The CCM is mainly comprised of glycolysis, tricarboxylic acid cycle (TCA-cycle, adjacent to oxidative phosphorylation), pentose phosphate pathway (PPP), urea cycle and β-oxidation.

In the following I will first briefly introduce the main parts of CCM and then show how muscle is specialized by performance these parts according to physical needs.

If parts of this network are adversely affected this can be cause and symptom of diseases. I will also briefly outline how this is especially important in muscle, which relies in its function on proper ATP supply.

Especially regarding fuel supply main muscle interactors in the body are liver and fat tissue and I will briefly give an example how these major metabolic contributors and storage depots work together via the bloodstream.

1.1 Central Carbon Metabolism

The central carbon metabolism (CCM) consists mainly of glycolysis, tricarboxylic acid cycle (TCA-cycle), β-oxidation, pentose phosphate pathway (PPP) and urea cycle.

Through these pathways the cell regulates overall anabolic or catabolic state, and it is also the “place” in the cell, where oxygen is consumed by its reduction to water. This regulation is instantaneous: Directly at the moment onset of muscular work , there is a so called step in oxygen consumption. [8]

Glycolysis or Embden-Meyerhof-Parnas pathway uses glucose to generate ATP. [9] It starts with hexokinase phosphorylating glucose, thereby reducing glucose concentration in the cell, favoring further import and also “trapping” it, because the much more polar glucose-6-phosphate (G6P) cannot diffuse back trough the cell membrane. During the reaction cascade to pyruvate, split into preparatory phase, where ATP is invested to drive

[1] adenosin tri-phosphate

the reactions, and payoff phase, overall two NADH[2] and two ATP molecules are generated per glucose molecule. [10]

NADH has to be reoxidized to NAD^+, otherwise glycolysis would stop. This can be achieved either trough anærobic fermentation of pyruvate to lactate, or by oxidation in the mitochondria.

Of note is a subpathway of glycolysis, the Rapoport–Luebering [11] pathway, by which 2,3-bisphospho-glycerate (2,3-BPG) is formed. 2,3-BPG regulates the oxygen affinity in erythrocytes. [12, 13]

Another alternative for the fate of glucose is to be reduced to Sorbitol, entering the polyol pathway [14]. After conversion to fructose and subsequent phosphorylation, it can then enter glycolyis at the level of fructose-1,6-bisphosphate (F1,6BP) or dihydroxy acetone phosphate and glyceraldehyde.

The **TCA-cycle**, or Szent-Györgyi-Krebs cycle [15, 16], is located in the mitochondria of the cell. In entry and inside this cycle, CO_2 is generated from the carbon of nutrients. It generates NADH and ubiquinol that are used in oxidative phosphorylation. Together with oxidative phosphorylation, the tricarboxylic acid (TCA)-cycle is the main source of adenosin tri-phosphate (ATP) in the cell.

Except from pyruvate, a major part of acetyl-CoA (AcCoA), that is fed into the TCA-cycle, is generated from fatty acids by **β-oxidation** (Knoop, 1904) [17, 18]. β-oxidation breaks down fatty acids in acetyl (C_2) units that are coupled with coenzyme A (CoA) to AcCoA. Per C_2 unit also reduction of one FAD[3] to $FADH_2$ and one NAD^+ to NADH take place, which are fed into oxidative phosphorylation.

The **Urea-cycle** (Krebs and Henseleit, 1932 [19]), represents the cells means to convert excess of toxic ammonia from degradation of amino acids to urea. The binding of carbamoyl phosphate to ornithine is located in the mitochondria. The formed citrulline is exported to the cytosol, where subsequently fumarate and urea are formed. The cycle is overall slightly exergonic, but the main amount of energy generated from amino acids stems from anaplerosis of e. g. TCA-cycle intermediates.

The **pentose phosphate pathway** (PPP) (Warburg and Christian, 1931 [20, 21] is mainly anabolic and uses glucose-6-phosphate (G6P) to generate ribose-5-phosphate (R5P), a building block of nucleotides, and NADPH[4] that is used to regenerate glutathione (GSH) [22, 23], and thus constitutes an important part in amelioration of oxidative stress. NADPH is also used as reduction equivalent in anabolic processes like fatty acid synthesis.

[2] nicotinamide adenine dinucleotide
[3] Flavin adenine dinucleotide
[4] Nicotinamide adenine dinucleotide phosphate

The PPP is tightly interconnected with upper glycolysis and besides the entry via glucose-6-phosphate an alternative pathway through gluconate formed by glucose oxidase and glucono-lactonase exists.

* * *

All these pathways are highly interconnected as exemplified for the polyol pathway. For example many TCA-cycle intermediates (TCAis) have pools in the mitochondria and the cytosol but also the PPP has many shared metabolites with glycolysis. For the various enzymes involved many tissue specific isoforms exist, that show different substrate affinities. These isoforms constitute a basic tendency to activity of certain CCM subpathways.

Depending on insulin and substrate level, muscle uses fat and/or glucose for energy homeostasis [24] but also to meet energy demands, which can vary from 0.015 kg to 1.5 kg ATP min^{-1} in an average sized human [5].

At the start of movement, muscle must have the means to readily produce new ATP-equivalents to ensure continuous function.

For this muscle has storage capacities for carbon sources.

1.2 In muscle, energy is stored in various Forms

It was once believed that during exercise, the substance of the muscle was used up and that a large amount of meat should be eaten to replace the lost tissue. [25] This belief is long overcome, but in muscle several energy pools with varying availability exist.

In the first seconds of movement, **phosphocreatine** (PCr) [26, 27] is used to re-phosphorylate adenosin di-phosphate (ADP). [28–30] This reaction is catalyzed by creatine kinase (CK). If muscle is injured CK leaks into the blood stream. This is used to monitor muscle break down in general.

Glucose, fueling glycolysis is then provided from **glycogen**, which constitutes about 1–2 % of muscle wet weight. [31–35] Glycogen, first identified by Claude Bernard in 1857 [36], consists of polymerized glucose, with either $\alpha(1 \rightarrow 4)$ glycosidic bonds, or it can also branch by forming $\alpha(1 \rightarrow 6)$ bonds. It has a very varying molecular weight, of mostly 250 MDa, that can go up to as high as 2.5 GDa. At the core of every glycogen particle, a glycogenin protein is situated. [10]

About 80 % of the glucose taken up into non contracting muscle is stored as glycogen. [37] Unlike the liver, muscle is generally considered unable to excrete glucose derived

from glycogen breakdown into the blood stream: muscle expresses glucose-6-phosphatase only at very low levels, and thus catabolises the glucose generated from glycogen breakdown exclusively by itself. [38, 39]

In terms of storage capacity, the most potent pool in the muscle is constituted of fat. Three **fatty acids**, bound together by a molecule of glycerol constitute one lipid molecule. Since lipids are not water soluble, they form so called liposomes and lipid droplets inside the cell. Their breakdown into AcCoA by β-oxidation is relatively slow. AcCoA can be used for glucose synthesis, which can in theory then be broken down by glycolysis, but since glucose synthesis and glycolysis do not run at the same time, fat can be used as energy source only, if oxygen supply in the tissue is sufficient.

In general, the higher the energy density of a certain energy pool, the more cumbersome and slower energy generation.

1.3 Muscle communicates with other organs via the blood stream

All the possibilities of energy supply mentioned above can only function, if the substrate uptake is not impaired, but also only if the other organs —especially the liver and adipose tissue— supply the substrates via the bloodstream.

During prolonged exercise the storage in muscle is depleted. The glycogen storage per tissue weight in liver is about 10 fold higher than in muscle. [40–43] Also, liver expresses glucose-6-phosphatase, [38, 39] which enables it to provide glucose, either generated from glycogen or gluconeogenesis, to the blood stream. This can be taken up by the muscle and used as fuel.

In certain metabolic conditions, like prolonged exercise or fasting, liver also provides ketone bodies, derived from excess AcCoA generated by β-oxidation of fatty acids, that are sent to the liver from adipose tissue via the blood stream.

One very well established example for the interplay between muscle and liver is the Cori-cycle. [10] Here excess lactate from anærobic fermentation of pyruvate in the muscle is excreted into the bloodstream and taken up by the liver, where it serves as substrate for gluconeogenesis. Glucose then is excreted by the liver to the blood stream and used by the muscle. This shifts metabolic load from muscle to liver in times of need.

1.4 Different muscletypes show specialised metabolism

When regarding at muscle as an organ, a pattern of sub-types with a preference for certain fuel storage pools and pathways emerges. Muscles themselves can be distinguished into

fiber types that express different proteins (isoforms of myosin heavy chain, cytoskeleton, calcium signaling, enzymes of the CCM) and show different levels of metabolites (e. g. lactate, TCA-cycle intermediates). [44].

Usually two types of fibers are distinguished: primarily oxidative (type I) and glycolytic (type II) fibers.

In humans types of fibers in a muscle spindle are mixed and thus difficult to separate for analysis. In smaller animals (e. g. mice), specialization of certain muscles is very high and some muscles have a highly dominant fiber type. [45, 46]

The fiber types have adapted to their primary functions. Type I fibers are used for long lasting contractions, like maintenance of body position or in endurance running. Type II fibers are fast twitching and often referred to as flight muscles.

Their physical properties, stemming from various isoforms of motor unit proteins are adapted to these functions. Type I muscles show a more rigid, less flexible structure, while fast fibers are suited for fast changes in overall physical form of muscle spindles.

Also metabolism in these muscles is highly tailored to their specific needs. In general no fiber type is capable of only one kind of metabolism, however they show a certain preference for and differences in size of the respective energy pools.

Type I muscles rely on oxidative metabolism. These muscles are also called red muscles, because of their higher levels of myoglobin, the oxygen transporter in muscle. They possess more mitochondria and rely on use of fatty acids. [47] Fatty acids have a very high energy density, about double the amount of carbohydrates ($39\,^{kJ}/_{g}$ vs. $17.2\,^{kJ}/_{g}$). So storage regarding to volume ratio is much higher in fat storage. But this in principle higher energy comes at the price, that supply of this energy through β-oxidation is slower than from carbohydrates through glycolysis and functions only in ærobic conditions. Thus these muscles can work at a relatively lower force level over longer periods of time, than type II muscles.

Type II muscles rely on glycolytic metabolism. They can produce higher force than type I muscles, but for shorter periods of time. Energy for this is saved in two pools, that are used during the first seconds or minutes of movement, the first being the phosphocreatine [28, 29], the second the glycogen pool. [48–50]

Phosphocreatine can be seen as a buffer for ATP: it can phosphorylate ADP very quickly, and thus provide ATP until energy becomes available by more sustainable means after the first few seconds of exercise. Phosphocreatine also shuttles ATP out of the mitochondrion.

When phosphocratine is used up, after seconds, another energy source needs to be provided to replenish ATP.The glycogen storage in muscle is much greater. About 80 % of the glucose taken up into non contracting muscle is stored as glycogen. [37] Unlike the liver, muscle is generally considered to be unable to excrete glucose derived from glycogen breakdown into the blood stream: muscle expresses glucose-6-phosphatase only at very low levels, and thus catabolises the glucose generated from glycogen breakdown exclusively by itself. [38, 39]

If oxygen supply is high enough, pyruvate is converted to AcCoA and fed into the TCA-cycle, located in the mitochondria. In mitochondria the majority of ATP-equivalents is generated while carbon of carbohydrates or fat is finally oxidized to CO_2. Aminoacids can also be used to generate energy, but do not comprise a major source of energy during normal physiological conditions and nutrient supply.

Regarding these interrelations, it is clear, that expression profiles but also activities of several proteins, especially enzymes have been used to characterize muscle fibers, or muscles of a certain type. [44]

The ATP produced by these pathways is used for maintainance of the tissue and cell, but predominantly it is used to facilitate movement: phosphorylation of myosin-heavy-chain (MYH) causes a conformational change which causes detachment of the myosin head from the thin filament. [51, 52]

1.5 Myosin isoforms characterize fiber types

Its physiological function –movement– muscle exerts by connection and release between myosin-heavy-chain (MYH) and the thin filament of the sarcomer, fueled by its ATPase activity. Since muscle forms differ in metabolism and rigidity, it is not surprising, that the constitutents of the sarcomer are also fiber type specific, especially MYH isoforms.

For fiber types further subclasses exist. Based on myosin heavy chain isoforms fiber-types are separated into for main classes, sorted by increasing reliance on glycolysis: TI, expressing predominantly MYH7, TIIA, expressing MYH2, TIIX, expressing MYH1, and TIIB, expressing MYH4. In the literature TIIX is also called TIID.

Also several other isoforms exist, but are predominantly expressed only in specialised muscles or regenerating and developing fibers respectively.

The fiber type, and also the dominant isoform of MYH is not fixed, but can change according to training [53–55].

Isoforms of MYH show variable ATPase activity at different pH. This is used in histology to differentiate fiber types. There are also fibers that show intermediate stainings in

histology, these are called mixed fibers. The mixed fibers can comprise a substantial amount of the muscle (≈30 %) [46], which shows that the normal classification is far from complete.

Table 1 lists the properties used to classically characterize muscle fiber types.

Table 1: Properties of slow (Type I) and fast (Type II) muscles.

Parameter	Type I	Type II
subforms	I	IIA| IIX/IID| IIB
color	red	white
predominant energy source	fat	sugar
myoglobin	more	less
Myh	–7	–2| –1| –4
mitochondria	more	less
contraction speed	slow	fast
durability	higher	lower
ATP consumption	1.5 $^{\text{mmol}}/_{\text{L s}}$	7 $^{\text{mmol}}/_{\text{L s}}$

1.6 Muscle diseases are fiber type specific

Also diseases of muscle are fiber type specific. Since the main function of muscle metabolism is to provide the energy needed to move, a muscle disease is likely to affect metabolism of muscle.

There is a whole class of so called metabolic myopathies, that fiber type specific are just by their nature. [56–60] But also other myopathies show a certain dominance in fiber types. Type II fibers are more affected by cachexia, in sepsis or the normally observed muscle loss in aging.

Since fiber type classification is mostly attributed to metabolic properties, it can also be the case, that the metabolism of fibers is altered to resemble a different fiber type.

This has been also attributed to a loss of type II fibers in dysferlinopathy (see below). [61]

1.7 The onset of dysferlinopathy coincides with changes in metabolism

Dysferlinopathy is an autosomal-recessive limb girdle muscular dystrophy, caused by mutations in the dysferlin gene that lead to a general deficiency. The patients present with

a wide range of disease phenotypes. [62, 63] The general prevalence is not known, but in the Jewish population of Lybia it is at least 1:1300 with a carrier rate of 10 %. [64]

Typically after a childhood where the individuals exhibit a rather active and often athletic inclined youth, coinciding with puberty the muscular strength deteriorates. [65] Typcial age of onset is 16–25 [66]. This leads to conditions where patients loose ambulation and are wheelchair bound 10–15 years after diagnosis. Until now no treatment exists.

Dysferlin is well recognized as involved in membrane repair [67, 68].

So one could argue that the reason for loss of muscle integrity in dysferlinopathy is impaired membrane maintenance. But the reason for late onset is unknown, and peculiar since during childhood the affected individuals often are highly athletic. [65]

The onset in late puberty coincides with a change in general metabolism: In childhood metabolism is mostly oxidative, while adults are more reliant on glycolysis. [8, 69–73]

Implications in metabolism are not discussed in the literature so far, except for a type I preponderance. [61]

In her thesis, Sarah Keller [1] investigated the metabolome and proteome of BLA/J mice —a model of dysferlinopathy [74]— and the metabolome of myotubes generated from primary patient material. She found, that most enzymes are unchanged between dysferlinopathy patients or mice and respective controls.

There was a slight increase of aldose reductase, that reduces glucose to sorbitol in *Tibialis anterior* which is in agreement with the findings in metabolite data. There was a strong decrease of isocitrate dehydrogenase 2 (IDH2), while other IDH isoforms were not altered. Opposed to isocitrate dehydrogenase (IDH)3, which uses NADH and was not altered, IDH2 uses NADPH as cofactor.

Changes in poolsizes can have several reasons. Since the enzyme levels in glycolysis were not changed, we expected a change in enzymatic activity, especially at the level of hexokinase II, the enzyme producing glucose-6-phosphate (G6P). To further investigate the metabolic changes in dysferlinopathy behavior, I transferred pulsed stable isotope resolved metabolomics (pSIRM), a mass spectrometry (MS) based technique invented in the lab of Stefan Kempa [2, 75–77] to function with primary human myotubes. (see also section 1.9)

* * *

1.8 Separation coupled mass spectrometry – a powerful tool in systems biology

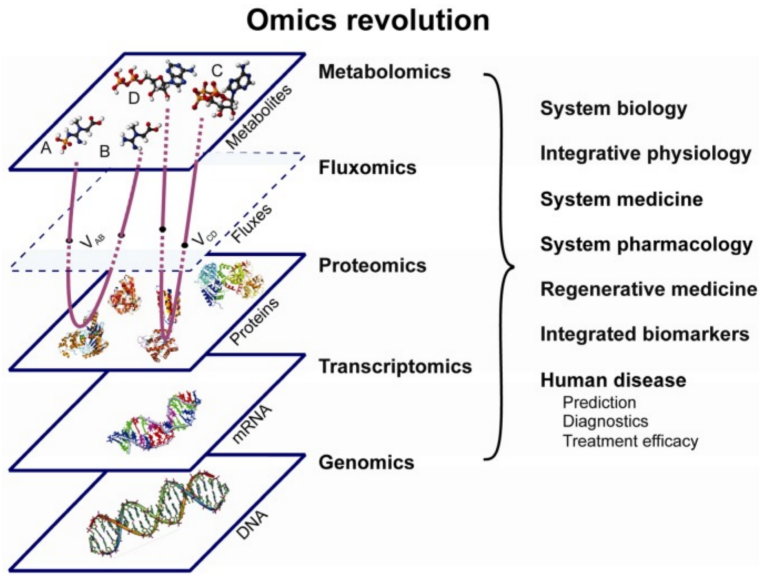


Figure 1.8.1: Levels of omics datasets. All depicted layers can regulate one another. Reproduced from [78]

To understand a system means to quantify as many parts, their interconversions and also the control, these parts exert on each other. In systems biology the goal is to understand the relations starting from DNA over RNA and proteins to the metabolic functions of enzymes, their production of metabolites and also the signaling exerted by phosphorylation and hormones — all in all to eventually understand the complexity of biology.

Omics-techniques are developed to contain information about everything in one of the layers of information mentioned above and shown in figure 1.8.1. [78] They normally allow for quantification of hundreds to thousands of features from the same sample. [79–83]

The layers involving **DNA** or **RNA** are measured by PCR-based techniques. Here the analytes are replicated until desirable signal strength is achieved.

For **proteins** and **metabolites** no such amplification techniques exist. Here often mass spectrometry based techniques are employed. [84] In mass spectrometry the analyte in question is ionized and subjected to a defined electric field[5] that allows to discern the molecular mass per charge ratio (m/z [Th[6]]), for example based on speed of linear motion (time of flight, TOF-detectors) or angular motion (Orbitrap [85, 86]).

[5] also magnetic fields or a combination of both electric and magnetic fields are employed.
[6] The unit of m/z is Thompson (Th)

Mass spectrometers are often coupled to chromatography, usually liquid chromatography (LC) or gas chromatography (GC). [87]

Chromatography separates the compounds contained in complex biological samples. Based on their physico-chemical properties the compounds are differentially retained by the chromatography column's stationary phase, so that they are ideally completely separated and can be injected into the mass spectrometer one after another.

Identification is then based on the mass spectrum and the time at which a compound is observed, i. e. the retention time.

The metabolomics data presented in this thesis were obtained by GC–MS, the proteomics data by LC–MS.

Not only in sample preparation, but also identification strategies of the compounds differs vastly between metabolites and proteins, thus they are presented separately in the following.

Proteins are usually enzymatically hydrolyzed and resulting peptides are subjected to LC–MS. Here the peptides are separated chromatographically and subjected to MS. The peptide ions can then be fragmented and their amino acid sequence is determined from the fragmentation pattern at MS/MS level and the peptide mass. This identification process is not definite but has some statistical error. Often a false discovery rate (FDR) of 1 % is chosen. [88] In more sophisticated setups the fragments obtained can be refragmented, to obtain deeper information levels.

Since peptides are quite defined in their composition by the proteinogenic amino acids, compared to metabolomics this automated identification is relatively robust. Peptide sequences obtained are then compared to a library of all proteins that the analyst deems to be theoretically included in the sample.

For definite protein identification unique peptides are needed. These peptides only can stem from one protein and allow for protein identification at gene or often even isoform level. Peptides are not always defining a protein unambiguously, especially isoforms of proteins share a high level of homologous sequences, thus a feature not always can be identified at isoform level. Sometimes ambiguity is higher and multiple different proteins constitute a protein group.

Based on the principle of Occam's razor [89, 90] peptides that cannot be unambiguously mapped to a single protein group in the data set are added the protein group that contains most peptides at this step in the identification procedure. [91] A standard tool for these identification and quantification procedures widely used in the field of proteomics is MaxQuant [91] including the Andromeda search engine [92].

Metabolites are either measured by NMR[7], LC–MS or GC–MS. All techniques have certain advantages and disadvantages, especially in regards to the vast heterogeneity of the class of "metabolites" that includes polar sugars and amino acids but also apolar lipids and sterols.

While **NMR** is generally the least sensitive technique it does not destroy the sample, which can still be subjected to other analysis techniques after data acquisition.

LC–MS shows superior sensitivity, but its separation power is not as high as in GC. For example, it is normally not possible to delineate hexoses, i. e. fructose and glucose. [93]. Although as in NMR, no sample derivatization is required for LC, the technolgy in general is less robust than NMR and GC–MS.

GC–MS is a robust technique, especially equipped to analyze central carbon metabolism (CCM) metabolites, partly because of its strength to separate sugar species. GC–MS offers important advantages over other metabolomics platforms: It is much more sensitive than NMR, and although it detects fewer features than LC–MS approaches, its better separation allows, for example, for the separate quantification of aldose and ketose sugars [93] that have been shown to have distinct physiological effects of great importance in diseases [94–96]. Furthermore, it allows for a more robust and reproducible quantification because of the near absence of ion suppression effects [97].

As the name says, in GC compounds traverse the separation column in the gas phase. Thus separation is mainly based on differences in boiling points. In order to reduce boling points of compounds like glucose to below pyrolysis temperatures, they are derivatized: chemical groups are added to compounds in a defined manner and for example reduce the relative amount of polar groups. Although the molecular wheight is increased, hydrophylic interactions are weakened, and thus the compounds evaporate more easily.

Compound identification in metabolomics is much more complex, than in proteomics. While in proteomics essentially all observable features are formed by proteinogenic amino acids, in metabolomics not only those amino acids but a multitude of sugars, carbonic acids, further amino acids, fatty acids and sugar phosphates, often differing only in isomeric arrangement are to be identified. For final identification manual visual inspection by highly trained specialists is required. [98] After several preprocessing steps, mass spectra of peaks are compared to mixtures of metabolites with defined composition that were measured as a standard in the same sample batch, or they are compared to databases of compounds measured using the same technical setup. [2, 75, 98, 99]

[7] Nuclear Magnetic resonance

Often many peaks of a chromatogram cannot be identified. These peaks still are likely to be valid signals of metabolites, they are simply not included in the available databases and termed “unknowns”. Some general functional groups can be identified from the unknowns fragmentation pattern in the mass spectrum. From those a subset of conceivable metabolite candidates can be inferred, e. g. sugars, phosphates or amines. But since final identification often hinges on isomeric differences, definite conclusions can only be made after a group of possible compounds is synthesized and their retention time and mass spectrum after derivatization is acquired.

The unknowns still can be included in databases[8] and compared to other samples if desired. This tracking allows backward compatibility if an unknown is identified later on.

1.9 Label tracing allows for measurement of metabolite poolsizes and formation rate at the same time

Not only the amount of a compound is of interest, but also its formation rate.

If the poolsize of a certain compound is high depending on treatment, this can have the reason of a blockage in subsequent processing steps, or an overproduction, or a combination of both.

The formation speed of a compound however allows to make statements about enzyme activity. This is important, since the amount of an enzyme does not necessarily correlate to product amount, as will be exemplified in section 4.3. Other factors as localization and allosteric regulation play an important role regarding metabolic activity. [100]

Since mass spectrometry data also includes information about isotope distribution, labeling of compounds is a possible means to infer production rate. Regarding effects on biology, stable isotopes of carbon and nitrogen generally do not cause differences in metabolism.

When substrates containing stable isotopes differing from naturally abundant isotopes (i. e. ^{13}C vs. naturally much higher abundant ^{12}C) are introduced to the cell, these are usually treated very similar to the naturally occurring variant.

The amount of a certain isotope species in a metabolite (or peptide) can be detected and quantified by mass spectrometry.

Pulsed stable isotope resolved metabolomics (pSIRM), a technique developed by Matthias Pietze and Christin Zasada in the lab of Stefan Kempa [2, 75, 76] allows to track

[8] Unknowns included in an unknown-database are called known unknowns, those not included new or unknown unknowns.

the formation rate at which new metabolites are formed. Contrary to classic approaches, this is done in the biological context of the whole cell, meaning that all possible interactors are present, simulating a relatively physiological context of the whole network.

The cells are introduced to a substrate made of stable isotopes, i. e. $^{13}C_u$-glucose, incubated for different intervals (labeling time) and subsequently harvested using ice-cold aqueous methanol. This stops metabolism instantly and allows also for sub-minute time resolution.

After separation of polar and apolar metabolites by liquid/liquid extraction with chloroform, the extracts are subjected to mass spectrometry.

Mass spectrometry allows to detect isotope incorporation in metabolites. Combined with gas chromatography (GC), which is able to even separate hexose-species from one another [93], this provides an extremely deep look into the dynamics of metabolism.

1.10 Inference of meaning from omics data by aggregation into concepts

After data is obtained, the researcher is tasked with data interpretation. This can be daunting, since an omics dataset normally is comprised of hundreds to thousands of analytes or features. Often several proteins and/or metabolites exert their effect in concert or belong to a certain signaling cascade that allows to infer certain subgroups of features as lined out already above in the CCM subpathways [10] or for example the general answer to a stimulus like insulin.

Besides the biochemical pathways as included in KEGG [101] or Reactome [102] also other aggregations exist.

The most prominent example is the gene ontology (GO)-database [103, 104]. Here the influences of certain genes on phenotype, regulation and interaction with other entities are aggregated in a hierarchical manner. GOCC (GO cellular component, e. g. “mitochondrium”, “nucleus”), GOMF (GO metabolic function, e. g. , “phosphatase”, “fatty acid binding”) and GOBP (GO biological process, e. g. “glycolysis”, “myelin assembly”) are the three topmost categories of GO-terms.

Since GO-terms have been derived from genes, the respective proteins can be matched accordingly. Metabolites however are not included in the GO-databases, although processes like “glycolysis” are contained.

GO-terms are hierarchically ordered. For example, GOCCs “mitochondrium, inner membrane” and “mitochondrium outer membrane” are part of GOCC “mitochondrium”.

This allows to construct networks and treemaps to infer higher level agglomeration of inferred concepts.

When the researcher is confronted with a dataset, he would usually extract all features that show significant alteration and use Fisher's exact test [105] to find concepts significantly overepresented in the altered features compared to all features.

Since not only GO-terms but many more agglomeration databases exist, in this thesis the term concept aims to encompass other databases as above mentioned Reactome pathways [102] or UniProt keywords [106].

The concepts obtained allow to direct further "diving" into the dataset, to find out what parts of a certain pathway are dysregulated or which organelles are particularly involved with proteins showing a difference between treatment groups.

1.11 Integration of different omics datasets gives another information layer

Integration of different omics datasets is quite challenging, especially when multiple layers of mutual regulation exist. Transcriptome and proteome are both strongly related to one another but still upregulation of a certain RNA-species not necessarily results in upregulation of respective protein expression. [107]

Similar relationships hold true for metabolites and metabolic enzymes. Among other influences, enzymatic activity is regulated by metabolites themselves (e. g. product inhibition) and protein abundance does not always reflect pathway activity that *can* correlate to metabolite poolsize.

Through integration of multi omics datasets, as presented in this thesis, also regulatory relationships can be inferred. This allows to unravel possible mechanisms of regulation or disease progression.

1.12 A detailed look at muscular dystrophies with mass spectrometry

In light of the substantial amount of mixed fiber types even in small mammals, and the overall fiber type specific phenotypes in muscle disease, it is interesting to know, how the interrelations of proteins and metabolites are reflected in the molecular make up.

Not necessarily at single fiber level, but more globally, in different muscles, because their physical form and use pattern are likely to be reflected in their specific composition.

Muscle fiber type is variable, depending on many factors like genetic predisposition, enervation, training state and age. [108–110]

In the literature only few omics studies comparing different muscles exist.

In 2012 Drexler *et al.* [111] compared *Extensor digitorum longus* and *Soleus* muscles of C57BL/6 mice and found 25 % of the quantified proteins to be significantly differentially expressed. They found differences in sarcomeric proteins and many kinases and enzymes in glutathione metabolism to be differentially expressed in the two muscle types compared.

In 2015 Rakus *et al.* [112] compared *Soleus* and *Tibialis anterior* of mice and found, besides the known differences in sarcomeric proteins, that enzymes comprising those pathways that are predominantly used by the respective muscletypes differing in their primary fuel source are differentially expressed. They also compared reaction rates of the extracted enzymes and found good correlation to their relative amounts. Rakus *et al.* concluded, that the upregulation in expression of a certain pathway is correlated with its activity.

In 2015 Deshmukh *et al.* [113] compared myotubes generated from C2C12 cells with *M. triceps* of C57BL/6 mice. In their comparison of two often used model systems they found 44 % of the proteome to be significantly different. This reflects the difficulties in transfer or research findings between different model systems.

Regarding metabolomics studies comparing different muscle types, after a thorough literature search I am not aware of the existence of any studies.

This warrants another look using up to date mass spectrometry based techniques to further characterize the molecular make up of muscle, and also, to my awareness for the first time, integrate metabolomics and proteomics data comparing muscle.

I chose four different muscles, that had been partially already been investigated using proteomics alone. Oxidative *Soleus* [111, 112], glycolytic *Tibialis anterior* [112], glycolytic *Extensor digitorum longus* [111] and *Quadriceps*, that has a mixed metabolism and showed the strongest phenotype in the investigations of dysferlinopathy by Sarah Keller [1].

Although for dysferlinopathy a type I fiber preponderance has been reported [61] and thus a metabolic implication of dysferlin is likely, in the work of her thesis Sarah Keller [1] was the first to report a metabolic phenotype of dysferlinopathy. Open questions remained, and with the novel technique of pSIRM, I wanted to further investigate this disease phenotype.

As explained above, the glycogen pool in muscle is of high general importance for muscle metabolism, as ≈80 % of glucose taken up into non-contracting muscle is routed

through the glycogen pool. [37] Because of its physico-chemical properties glycogen cannot be measured by the standard metabolomics workflow. I aimed to modify the method in a way, that also label incorporation in glycogen can be measured.

Another question was how dysferlin might exert its effect on glycolysis, that was obvious from Sarah Keller's data. During her thesis Séverine Kunz [3] developed a protocol to enrich dysferlin vesicles. We obtained proteome data from primary cells of dysferlin-normal healthy individuals including a broad set of background controls, in order to possibly elucidate implications of dysferlin on metabolic and especially glycolytic enzymes.

The techniques referred to are all based on muscle tissue, which is relatively hard to obtain. Biopsies inherently require at least a small injury of potentially diseased muscle, thus it is desirable to find less invasive techniques to investigate muscle. As all other organs, muscle receives fuel and excretes metabolites from and to the blood stream. Together with Henning Kuich [4] we performed an experiment with a single human volunteer performing exercise. He ran around the campus and I obtained small blood samples, similar to a diabetic's blood glucose test. We wanted to see if and how the metabolic changes during physical activity are reflected in the blood stream, which might possibly lead to new diagnostic tools in clinical investigations of muscle disease.

2 Methods

For this thesis, based on the work of especially Christin Zasada and Matthias Pietzke, an extraction protocol for stably labeled human myotubes had to be established. This process will be described in detail in the results section (section 3.3.1), but parameters for cell culture in general will be given here. The treatment of samples to compare label incorporation into dysferlin deficient myotubes will be given here, as this constitutes the finally established method. After extracts had been acquired in a manner as will be layed out below, general metabolite and protein data acquisition was always performed similarly, and will be described in this section.

Dysferlin vesicles were processed by Séverine Kunz, including in gel protein digestion. A central part of her thesis was the devolopment of dysferlin vesicle enrichment, that shall not be given here. For detailed methods, please refer to her thesis: Kunz, *et al.*, 2015 [3].

After extraction of mouse muscles, metabolite extracts were analyzed by GC–MS using the same parameters as the cells, and the blood samples. (section 2.8) The same holds true for dysferlin-vesicle and mice muscle protein extracts, which were analyzed by LC–MS (section 2.9).

2.1 Study approval

The local animal research committee approved all protocols (G0065/13, LaGeSo, Berlin, Germany).

Human muscle biopsy specimens and myoblast cell lines were obtained after informed consent prior to inclusion in the study (Ethical approval EA1/203/08, Charité Berlin, Germany). The volunteer in the campus run is a scientist well aware of the risks associated with human trials. He also gave written consent.

2.2 Solutions and chemicals used

PBS: Phosphate buffered saline, Gibco

Trypsin: 0.25 % trypsin-EDTA (1x), Gibco

FBS: Fetal bovine serum, Biochrom

GlutaMax: GlutaMAX^{TM}-I (100x), Gibco

SKMGM: Skeletal Muscle Cell Growth Medium (Basal Medium), unless stated otherwise supplemented with 10 % FBS, 1.5 GlutaMax and 40 $^{\mu g}/_{mL}$ gentamycin, proVitro

SKMGM knockout: SKMGM without glucose, pyruvate, glutamine, supplement (from manufacturer, includes growth factors, insulin), proVitro

Krebs-Ringer-HEPES (KRH) buffer: 140 mM sodium chloride, 5 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)

2.2.1 Solutions and chemicals used for metabolomics

Alcane mix: as retention index markers *n*-decane, *n*-dodecane, *n*-pentadecane, *n*-heptadecane, *n*-nonadecane, *n*-docosane, *n*-octacosane, *n*-dotriacontane, and *n*-hexatriacontane, are mixed to a final individual concentration of 2 $^{mg}/_{mL}$ together comprising the alcane-mix, were dissolved in hexane and combined to a final individual concentration of 2 % *v*/*v*.

Ident mix: the ident mix is a combination of 102 compounds combined in 4 combinations (A–D) in such a mannner, that they are once measured without interfering compounds at similar retention index, and once with interfering compounds. Please refer also to [2, 98].

Quant(ification) mix: the quant mix is composed of 63 compounds (stock concentrations 1 $^{mg}/_{mL}$). They are combined in a empirically determined manner (master mix 1:1), and then diluted in steps of 1:1, 1:2, 1:5, 1:10, 1:20, 1:50, 1:100 to 1:200. [2] For each batch of samples an aliquot was extracted in parallel and measured at the beginning of the sample set.

MCW: methanol|chloroform|water 5|2|1 *v*|*v*|*v*

$MeOH_{aq}$: aqueous methanol in given percentage, always *v*/*v*

MeOX: methoxyamine hydrochloride in pyridine, $c = 40\ ^{mg}/_{mL}$

MSTFA: (N-methyl-N-(trimethylsilyl)-trifluoroacetamide)

KOH: potassium hydroxide, 0.2 M

AcOH: 1 M acetic acid

potassium-acetate-buffer: KOH|AcOH 230|70

2.2.2 Solutions and chemicals used for proteomics

Urea buffer: 8 M urea in 100,mM Tris-HCl, pH 8.25

DTT: Dithiothreitol, reduction agent to reduce thioethers

IAA: Iodoacetamide, as alkylating agent of thiols formed by reduction of thioethers

2.3 Glycogen precipitation and determination with anthrone reaction

In one 10 cm dish of myoblasts, 150 µg of glycogen can be expected. [31, 33–35] To test precipiation behaviour in methanol, 150 µg aliquots of glycogen standard (from oyster, Sigma-Aldrich) were added to different concentrations of $MeOH_{aq}$ (0, 40, 50, 60, 70, 80, 90, 100 percent). Also 1 mL of chloroform was added to simulate extraction of cells.

Samples were treated as in cell extraction (see below, section 2.6.2). Phases were separated and dried and glycogen content was measured using anthrone reaction, which was modified to function in modern 96-well plate readers. This assay is very simple, and only requires to add anthrone, dissolved in concentrated sulfuric acid to the dissolved glycogen, followed by incubation at 100 °C for 10 min. This results in formation of a green color in the solution, that can be photometrically determined at λ = 620–630 nm. [42, 114–118]. In the literature, reaction is carried out in bigger volumes, and requires manual filling of cuvettes. I scaled the parameters down. Comparing to an external calibration curve, that showed high linearity (R^2 = 0.998 in a range between 2 and 420 µg), I could use this assay directly in a 96-well format.

2.4 Test of linearity between signal and sample amount

Tibialis anterior and *Quadriceps* muscles of one C57BL/6N mouse were smashed by biopulverizer treatment and extracted with MCW as follows:

- add MCW
- 15’ Ultrasound treatment
- add 25 % $MeOH_{aq}$to induce phase separation
- 15’ Ultrasound treatment
- centrifuge 10’ at 20.000 *g*
- separate samples
- add 100 µL cinnamic acid in MCW (c = 5 µg/mL) as internal standard

- dry in rotational vacuum concentrator

Volumes of extraction solvents can be found in table 2, the amounts of polar phase (PP) dried in table 3.

Table 2: Extraction solutions used for linearity test. PP: polar phase

	TA	QUAD
m_{Muscle} [mg]	91.53	102.9
V_{MCW} [µL]	1830.6	2058
$V_{25\,\%\text{MeOH}_{\text{aq}}}$ [µL]	915.3	1029
V_{PP} [µL]	2288.25	2572.5

Table 3: Volumes of PP dried to assess linearity between signal and sample amount

V_{PP} [µL]	10	20	50	100	200	500	1000
m_{Muscle} [mg]	0.4	0.8	2	4	8	20	40

The samples were derivatized, measured and annotated as described in section 2.8.

2.5 Collection and extraction of *Extensor digitorum longus*, *Quadriceps*, *Soleus* and *Tibialis anterior* of C57BL/6N mice

With assistance from Adrienne Rothe C57BL/6N male mice, 14 weeks of age ($n = 8$), (Charles River) were sacrificed by cervical dislocation and the muscles immediately extracted. We collected samples from 4 mice on two consecutive days, ensuring same time of day, to ameliorate influences by circadian rhythm.

Muscles used for **metabolomics analysis** were obtained first, briefly rinsed in PBS to remove eventual fur or other debris, weighed, photographed to document length and flash frozen in liquid nitrogen within within 4–8 minutes after sacrifice and were already cooled on an ice pack during imaging.

They were crushed via biopulverizer, and powder aliquots extracted by 20 min sonification in 40 µL/mg ice cold MCW containing 1.5 µg/mL cinnamic acid, in a cold room (≈7 °C). 20 µL/mg 25 %MeOH_{aq} were added, followed by another 20 min of sonification to ensure phase equilibrium. Debris was centrifuged (10 min at 20.000 *g*, 4 °C), phases separated and dried within 5 h using a rotational vacuum concentrator. Further processing was carried out as lined out below (section 2.8).

Muscles used for **proteomics analysis** were obtained second, cleaned, weighed, imaged and stored on cold PBS filter paper (Whatman) during further processing. With new razor blades, except from *Quadriceps*, six transversal sections were cut and mounted for histological sectioning with gum tragacanth on cork plates. Muscle sections were first frozen in *iso*-pentane at melting point, and then quickly transferred to liquid nitrogen. After histological sections, for experiments not included in this thesis, were obtained with assistance by Adrienne Rothe, the remaining sections were cleaned of gum tragacanth. Proteins were extracted with assistance by Alina Eisenberger using a tissue lyser (Precellys^{TM} 24, Bertin Technologies, Montigny le Bretonneux, France) with yttrium stabilized zirconoxid ceramic beads of 0.4–0.6 mm diameter (SiLibeads ceramic beads type ZY, Sigmund Lindner, Warmensteinach, Germany) at a speed of 6200 Hz for 20 sec runtime for two cycles. Protein content was determined using BCA-assay kit (Thermo Scientific) according to manufacturers specifications.

Equal volumes of protein extract were combined to keep relative protein concentration constant.

There might be a slight bias, introduced by different extraction efficiencies, since all samples were extracted in the same buffer amount and thinner muscle parts could be favored, but it is very unlikely, that the extraction solvent was oversaturated, often whole mouse muscles are extracted in similar volumes. [1]

2.6 Labeling of myotubes

2.6.1 Cell culture

Myoblasts were purified from muscle biopsies using magnetic antibody sorting as described in [68, 119]. Myoblast culture was carried out at 37 °C, 5 % CO_2 and 96 % relative humidity. Times between splitting of cells varied from 1 to 3 days, according to variable growth of primary cells determined by visual inspection.

For splitting, cells were washed with 5 mL PBS per 10 cm dish and detached by treatment with 1 mL trypsin for ≈2–3 minutes, diluted with fresh media and distributed onto new plates.

Myoblasts were cultured in skeletal muscle cell growth medium (SKMGM) partially in cooperation with Verena Schöwel and assistance by Stefanie Meyer-Liesener and Stefanie Belz until ≳80 % confluence was reached in 10 cm petri dishes.

To stimulate fusion into myotubes, cells were serum deprived by switch of cell culture media to Opti-MEM (Gibco). Myotube development was monitored visually, and was

generally achieved after 5–7 days.

2.6.2 Labeling of myotubes

Cells used for test experiments providing only one carbon source were donated by a 20 year old woman, biopsy was taken from *M. vastus lateralis*. Cells used for test experiments with all carbon sources present, were donated by a 55 year old woman, biopsy was taken from *M. vastus lateralis*. Muscle was diagnosed as non-pathological at the outpatient clinic for muscle disease, Charité Berlin-Buch.

In general, harvest was conducted as described in [2]. 24 h before labeling start myotubes were presented with the carbon source(s) present during the labeling, contained in re-established ^{12}C-SKMGM-knockout media (proVitro). Carbon source concentrations were chosen to simulate postprandial phyisological conditions in blood (glucose 1 g/L, palmitate 1000 mM, 3-hydroxybutyrate at 4 mM, a concentration, that is typical for a fasting state). Glutamine and pyruvate were replenished according to media specifications. I was asked by the company not to reveal specific concentration values. Since most of the data presented in this thesis, does not hinge on substrates except glucose and 3-hydroxybutyrate, I do not think this is necessary for the recapitulation of the results.

4 h before labeling start, fresh media was provided.

At labeling start, cells were washed with PBS and time was started at the moment, ^{13}C-labeling (or ^{12}C-background) media was put into the dish.

About 30 seconds before envisaged labeling time was up, cells were washed with Krebs-Ringer-HEPES buffer, containing the respective carbon sources, to wash away exrected metabolites, but not disrupt metabolism.

Table 4 gives the applied labeling times, see also figure 3.3.7.

Table 4: Labeling times used in tracing experiment of dysferlin defficient primary human myotubes. Time-points and number of replicates per patient ($n = 2 \times 2$) or control ($n = 3 \times 2$) for labeling and background-controls are given. Refer also to figure 3.3.7 OHB: 3-hydroxybutyrate

Media type	$n_{\text{individuals}}$	t_{labeling} [min]	3’	7’	15’	30’	60’	120’	10h
^{13}C-Glucose	5		2	2	2	2	2		2
^{12}C-Glucose	5					2			2
^{13}C-OHB	5				2	2	2	2	
^{12}C-OHB	5						2		

When labeling time was over, washing buffer was discarded, cells were placed on ice and immediately scratched with 5 mL 60 %MeOH$_{\text{aq}}$ containing 2 µg/mL cinnamic acid into

15 mL falcon tubes filled with 1 mL chloroform. They were vigorously shaken and flash frozen in liquid nitrogen.

In further processing, samples were shaken for 30 min at ≈4 °C to ensure phase equilibrium. After centrifugation, the polar (upper) and apolar (lower) phase as well as the glycogen containing interphase were separated. Extracts were dried in a rotational vacuum concentrator. To transfer the samples to 1.5 mL sample tubes for derivatization, they were dissolved in 20 % $MeOH_{aq}$ and dried again. The interphase was dissolved in water and split into two aliquots. Polar phase was submitted to GC–MS analysis (section 2.8).

2.6.3 Glycogen hydrolysis

After drying in a rotational vacuum concentrator, to remove free glucose both interphase aliquots were washed twice with 80 % ethanol (*v*/*v*). [120] One aliquot of each extract was hydrolyzed first by heating in 230 µL 0.2 M KOH 1 h 95 °C. Samples were briefly allowed to cool to ambient temperature, and neutralized by addition of 70 µL 1 M acetic acid. Amyloglucosidase and α-Amylase were added and glycogen was further hydrolyzed for 24 h at 37 °C. [121] To account for possible residual background glucose, the other aliquot was redissolved in potassium acetate buffer, dried together with its hydrolyzed counterpart, and used as matched background for subsequent calculation of label incorporation after GC–MS measurement (section 2.8).

2.7 Extraction of blood samples obtained from a single healthy volunteer performing exercise

In a proof of principle study to assess possible reflection of muscle work in the blood stream, Henning Kuich ran around the MDC-Campus, and I collected blood samples from the finger tip with my diabetics lancetting pen (BD OneTouch comfort, 0.2 mm, 33 G) and stabilized them directly in ice-cold MCW. While he ran, I accompanied him by bike and monitored his overall status, but also cheered him on, when he reached the point of "Hitting the Wall".[9]

He ran six laps of 2.2 km on the MDC-Campus. Because we wanted to monitor the phenomenon of "Hitting the Wall", which was experienced in lap 3, this lap was cut short by 200 m (10 %) in fear of missing this crucial time point (see figure 3.5.3).

[9] We had a lot of fun that day, but even more so, when we realized the potential of the data obtained. I still hope and believe, that someday we will bring the full potential of this to benefit humankind.

Whole blood samples of 10 µL were collected in 1 mL MCW, shaken (1000 rpm) at 4 °C for 10 minutes to lyse all cells in the blood. Then 500 µL water was added to induce phase separation. After further 10 minutes of shaking at 4 °C, 1000 rpm, to ensure phase equilibration, samples were centrifuged for 10 minutes at 4 °C, 20.000 *g*.

Two times 600 µL of polar phase and one time 200 µL of lipid phase were collected and dried for derivatization in a rotational vacuum concentrator.

2.8 Derivatization, GC–MS measurement and annotation of metabolites

2.8.1 Derivatization

In order to enhance the volatility of metabolites, they have to be derivatized by certain chemical compounds, that react with specific functional groups of the metabolites, e. g. hydroxyl or amine groups. Since minute amounts of water hinder the reaction, if the samples had been frozen, they were equilibrated to room temperature in a rotational vacuum concentrator for 15 min.

Derivatization was carried out as follows:

- add 10 µL MeOX (Ident/QuantMix: 20 µL)
- incubate 90 min at 30 °C on shaker
- add 30 µL MSTFA (Ident/QuantMix: 80 µL), including retention index markers (alcane mix)
- incubate 60 min at 37 °C on shaker
- centrifuge 10 min at 20000 *g*
- transfer supernatant to GC vial, $\gtrapprox$17 µL to GC vial
- seal vial

The samples where directly measured by GC–MS.

2.8.2 GC–MS measurement

Metabolite analyses were performed with a Pegasus III mass-spectrometer (LECO, St. Joseph, USA) equipped with an Agilent 6890N gas chromatograph and a VF-5ms column with 30 m length and 250 µm inner diameter (Agilent, Santa Clara, USA). 1 µL of sample was injected into a baffled liner (Gerstel, München, Germany) with a 1:5 split-ratio under a helium-flow of 1.2 mL/min. The oven was heated from 70 °C to 350 °C with 5 °C/min to

120 °C and 7 °C/min to 350 °C followed by 2 min hold time. Scan rates of 20 Hz and mass ranges of 70–600 Th were used.

2.8.3 Annotation

The GC–MS chromatograms were initially processed with the ChromaTOF software (LECO). The Ident mix and the Golm metabolome database (GMD) [99] were used to identify substances with respect to spectra-similarity and retention index using Maui–SILVIA [4, 98].

2.9 Protein digestion, LC–MS measurement, protein identification

2.9.1 Protein digestion of muscle tissue samples

Of the recombined extracts aliquots of 100 µg of proteins were reduced in 10 mM DTT for 30 minutes at room temperature and successively free cysteines were alkylated in 55 mM iodoacetamide for 20 minutes at room temperature in darkness. LysC digestion was performed by adding LysC (Wako) in a ratio 1:40 (*w*/*w*) to the sample and incubating it for 18 hours under gentle shaking at 30 °C. After LysC digestion, the samples were diluted 4 fold with 50 mM ammonium bicarbonate solution, 7 µL of immobilized trypsin (Applied Biosystems) were added and samples were incubated 4 h under rotation at 30 °C. Digestion was stopped by acidification with 10 µL of trifluoroacetic acid and trypsin beads were removed by centrifugation. A volume equal to 18 µg of digested protein were desalted on STAGE Tips, dried and reconstituted to 20 µL of 0.5 % acetic acid in water [122], to be subsequently analyzed by LC–MS (see below, 2.9).

2.9.2 Protein digestion of dysferlin vesicles

2.9.3 LC–MS measurement

5 µL of tryptic digest were injected in an UPLC system (Eksigent, Framingham, USA), using a 240 min gradient ranging from 5 % to 45 % of solvent B (80 % acetonitrile, 0.1 % formic acid; solvent A (5 % acetonitrile, 0.1 % formic acid), coupled to a Q Exactive Plus mass spectrometer analyzer (Thermo, San Jose, USA).

For the chromatographic separation a 30 cm long capillary (75 µm inner diameter) was packed with 3 µm C18 beads (Reprosil-AQ, Dr. Maisch) was used. On one end of the capillary nanospray tip was generated using a laser puller, allowing fretless packing. The nanospray source was operated with a spray voltage of 1.9 kV and an ion transfer

tube temperature of 260 °C. Data was acquired in data dependent mode, with one survey MS scan (resolution 70000 at 200 Th) followed by up to 10 MS/MS (resolution 7500 at 2500 Th). Once selected for fragmentation, ions were excluded from further selection for 30 sec, in order to increase new sequencing events.

2.9.4 Protein annotation

For **mouse muscle proteomes** raw data was analyzed using the MaxQuant proteomics pipeline version 1.5.3.30 [91] and the built in Andromeda search engine with the Uniprot database of mouse proteins [123], release 2016_01 including splice variants. Carbamidomethylation of cysteines was chosen as fixed modification, oxidation of methionine and acetylation of N-terminus were chosen as variable modifications. Two missed cleavage site were allowed and peptide tolerance was set to 20 ppm. The search engine peptide assignments were filtered at 1 % FDR at both the peptide and protein level, while other parameters were left as default, mathing between runs was enabled. For data processing and visualization, LFQ intensities [124] were used.

For **dysferlin vesicle proteomes** raw data was analyzed using the MaxQuant proteomics pipeline version 1.5.1.2 [91] and the built in Andromeda search engine with the Uniprot database of human proteins [123], release 2014_04 including splice variants. Carbamidomethylation of cysteines was chosen as fixed modification, oxidation of methionine and acetylation of N-terminus were chosen as variable modifications. Two missed cleavage site were allowed and peptide tolerance was set to 5 ppm. The search engine peptide assignments were filtered at 1 % FDR at both the peptide and protein level, while other parameters were left as default, mathing between runs was enabled. For data processing and visualization, normalized intensities were used.

2.10 Statistics, plot preparation, software used

For data analysis, R [125] and the following packages were used:

adgenet a R package for the multivariate analysis of genetic markers [126, 127]

Bioconductor Orchestrating high-throughput genomic analysis with Bioconductor [128]

KEGGREST Client-side REST access to KEGG [129]

KEGGgraph KEGGgraph, a graph approach to KEGG PATHWAY in R and bioconductor. [130]

RColorBrewer ColorBrewer Palettes [131]

Rcpp Seamless R and C++ Integration [132]

classInt Choose Univariate Class Intervals [133]

doBy Groupwise Statistics, LSmeans, Linear Contrasts, Utilities [134]

extrafont Tools for using fonts [135]

ggplot2 ggplot2 elegant graphics for data analysis [136]

ggrepel Repulsive Text and Label Geoms for 'ggplot2' [137]

gridExtra Miscellaneous Functions for "Grid" Graphics [138]

limma Limma powers differential expression analyses for RNA-sequencing and microarray studies [139]

plyr The Split-Apply-Combine Strategy for Data Analysis [140]

psych Procedures for Psychological, Psychometric, and Personality Research [141]

reshape Reshaping Data with the reshape Package [142]

scales Scale Functions for Visualization [143]

stringr Simple, Consistent Wrappers for Common String Operations [144]

svglite An 'SVG' Graphics Device [145]

treemap Treemap Visualization [146]

For general quality control of proteomics data, PTXQC was used. [147] Also Perseus was used for data analysis. [148]. The network from blood metabolomics was generated using Cytoscape [149], other networks were prepared using VANTED [150]. The circos plot was generated using circos [151]. All artwork was produced in Gimp and mainly inkscape [152] Some lists and tables were produced in Microsoft Excel 2013.

3 Results

In the framework of this thesis, metabolomics and proteomics data, obtained either by GC–MS or LC–MS was obtained from several different types of matrices originating from mice and humans.

From mice different muscle types, representing different metabolic types (oxidative, glycolytic, intermediate) were subjected to metabolic and proteomic analysis. Regarding proteomics, some comparative studies have been made, always comparing only two muscle types, either *Soleus* and *Extensor digitorum longus* [111] or *Soleus* and *Tibialis anterior* [112], but never comparing more than two muscles at the same time. To my knowledge, no comparative study, combining metabolomics and proteomics data of different muscle types exists. Since it is known, that often type II fibers are more prone to disease and show more pronounced atrophy in healthy processes as aging, a deeper understanding of the differences on a molecular level is of high interest. So far, muscle types have been discerned primarily by few features in separate studies. Most often they are described by the presence of certain myosin heavy chain isoforms, relying on a variation of activity depending on pH-level. But also differences in the expression or activity of certain glycolytic enzymes as lactate dehydrogynase or mitochondrial enzymes as malate- or succinate dehydrogynase. [153, 154, and references therein]

Also many other features differing between muscle types have been researched, but not on a systemic level, with data obtained from the same samples. This hinders comparability of the differences found between the muscle types and the possibility to detect unexpected relations, e. g. between different predominantly glycolytic muscles like *Extensor digitorum longus* and *Tibialis anterior*. The systems biology approach presented here allows for unbiased exploration of the feature space and possibly to draw unexpected conclusions in decoding disease mechanisms or preventative medicine.

To gain deeper understanding of muscle metabolism in disease, GC–MS based metabolomics was used to analyse samples obtained from patients suffering from dysferlinopathy. This genetic disease shows a late onset, starting from puberty, when also the metabolism of muscle starts to become more reliant on glycolytic metabolism. [8, 70–73]

The concurrence of disease onset and metabolic changes in muscle, as the organ of highest absolute metabolic rate, warrants for an investigation of metabolism in this disease. Based on work already conducted by Sarah Keller [1] and Verena Schöwel in the research groups of Simone Spuler and Stefan Kempa, I transferred pSIRM, previously established for cancer cells by Christin Zasada and Matthias Pietzke [2, 75–77] to function with

primary human muscle cells. I also modified the extraction procedure, in order to include label tracing in the glycogen pool.

To take biopsies and cultivate primary muscle cells is cumbersome and challenging, especially because the patients to be examined already have compromised muscles. Taking a biopsy puts also high strain on muscle health and is to be considered a major intervention in patients with muscular disorders. In order to circumvent this, tests with a less invasive technique, like liquid biopsies would be favorable. Blood is easily obtainable, but also reflects a mixture of the influences by the different metabolic organs. On the one side this is an advantage, since in the end, we are interested in the patients overall status, however it might complicate analysis, because the origin of metabolites can never be sure.
Overall, but especially during physical activity, muscle is the organ with highest absolute metabolic turnover. Under this presumption during exercise muscle should be the major contributor to changes in metabolite levels in the blood stream. In a proof of principle study with a trained individual, we tested, if and how physiological activity is reflected in blood during physical exercise. The sensitivity of modern mass spectrometry techniques allows to use only single droplets of blood, taken from a finger or the earlobe, similar to a blood glucose test and allows for longitudinal studies with high sampling frequency.

Gaining this deeper understanding starting from muscle types in mice, then looking at activity of the major metabolic enzymes in human sample material and in the end observing the physiological interplay of muscle in the body during activity allows for a very detailed look at factors that may interfere with its function.

3.1 Establishment of measuring metabolites from muscle tissue

Muscle tissue is highly complex and relatively rigid, reflecting its physiological function of force exertion. When analyzing tissue several imponderabilities have to be considered. For one, the complex matrix of muscle is likely to have an impact on extraction efficiency. Since I only wanted to compare muscle tissue with other muscle, this issue can be evaded by using approximately the same amount of tissue from each muscle.
Since the muscles I examined are of very different shape and size (see figure 3.2.1) sample homogeneity is certainly another important factor.

It has also to be kept in mind, that many metabolites are relatively unstable, so heating the samples has to be avoided. Proteins are more stable and proteomics samples can be subjected to a harsher treatment. When the samples were obtained with help of Adrienne Rothe, we always extracted the muscles to be subjected to metabolomics analysis first, in order to ensure sample integrity. From cervical dislocation till placing the muscle in liguid

nitrogen it took 4–8 minutes for the metabolomics samples.

Most extraction techniques for tissue involve lysis using ceramic beads [155, 156]. During this treatment the samples become relatively warm, so that sample deterioration could be an issue. As such a biopulverizer was used, which allows to obtain a powder of frozen tissue, keeping the sample temperature at well below −80 °C. [157] This ensures sample homogeneity and integrity during the preparation process.

For proteomics much less tissue is needed than for metabolomics. Since extraction with a biopulveriser is relatively cumbersome, and the muscle samples obtained were to be preserved for a different analysis not included in this thesis, except for *Quadriceps* the proteomics samples were transversely cut into six equally thick sections and mounted with gum tragacanth for histological sectioning. To ensure, that gum tragacanth was not causing a background signal in the bradford assay used to determine protein content after extraction, I extracted 100 mg of pure gum tragacanth, much more than could have been accidentally included in the actual muscle protein extracts, and performed a bradford assay of the extract. There was only negligible increase of signal, compared to pure extraction buffer.

The quadriceps samples were subjected to homogenization by biopulverizer.

After specimen for histology were obtained, the sections were removed from the mounting media and separately extracted using ceramic beads (also *Quadriceps*). These extracts were then recombined, to represent the whole muscle, and subjected to trypsinization.

Extracts of sections were recombined by equal volume amounts, to keep relative protein concentration at constant level. The central sections, where muscle diameter is higher, had higher protein concentrations, as expected, but since equal volumes were combined, not equal protein amounts, this effect was ameliorated in the recombined extracts.

Although obtaining metabolomics data was already established for many different matrices in the group of Stefan Kempa, there was only sparse data on the linear correlation of sample amount and measured intensity in muscle. Especially the tracking of unknowns was quite new in the group, and knowledge of the behavior of these compounds regarding linearity of signal strength had to be obtained.

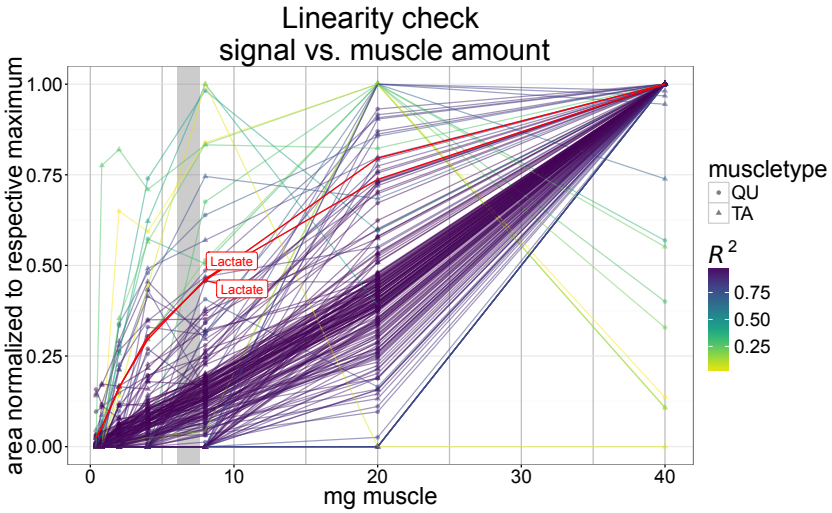
Unknowns, or unidentified features, are features, that cannot be mapped to a database entry. They still are valid metabolites, but either not included in the ident-mix or in the external databases used for annotation. In theory compound identification is possible to some extent from mass spectra, but for example hexose-species, differing only in very similar isomeric arrangement. In metabolomics, for unidentified substances, the mass spectra allow to infer certain functional groups, like amino or carboxyl groups. In the end,

doubtless identification is only possible if the compound is synthesized and retention index and mass spectrum measured on the same machine are congruent.

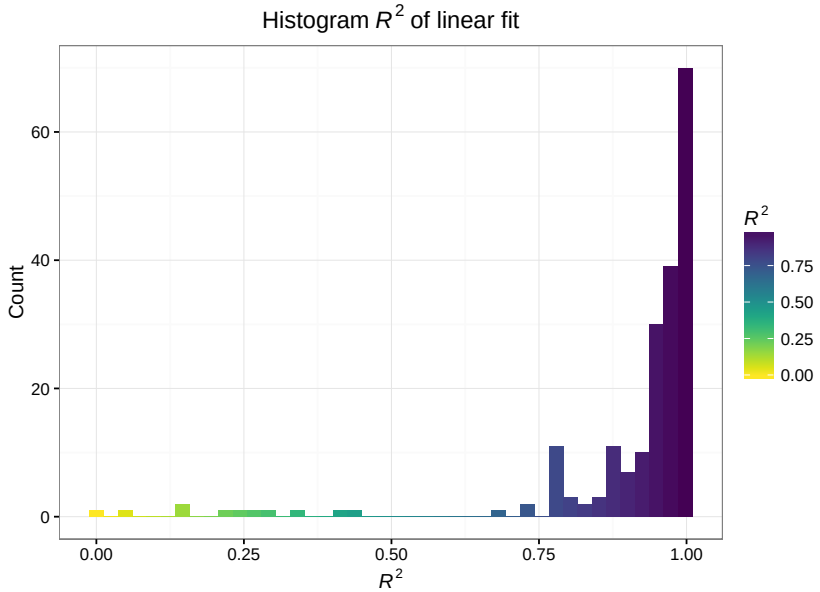
In the test data most of the compounds, including unidentified features were found to show linear behavior over a wide concentration range, as plotted in figure 3.1.1a. Here area normalized to its maximum per compound is plotted against the amount of muscle in the sample, color depicts R^2 of a linear fit, lactate, as a problematic case, is emphasized in red.

Some metabolites were only measurable at comparatively high concentrations, but showed a linear incline from there on. So if they are detected, they can be deemed quantifiable. The histogram in figure 3.1.1b shows the distribution of R^2 across all features measured in this experiment. This also shows, that overall linearity is quite good.

The gray shaded box denotes the amount of muscle used in the actual study. The limiting factor to consider was muscle mass of *Soleus* and *Extensor digitorum longus*, that weigh about 8–10 mg, and thus yield little sample material, which limits detectability of metabolites with low abundance, whereas some metabolites, e. g. lactate, quickly reach saturation. Lactate, which is a borderline metabolite, shows already some saturation in the range of muscle amount that was used in the main experiment. Since one whole muscle of SOL and EDL need to be used for a measurement in order to achieve satisfactory coverage of most metabolites, it seemed not reasonable to measure twice with different concentrations. Results based on lactate have to be treated carefully, however, lactate levels in different muscle type have already been extensively studied, and lactate amounts are not the focus of interest in the present experiments.



(a) To assess, if the metabolites especially the unidentified ones, show a linear relationship of signal strength and sample amount, intensities of features were normalized to their respective maximum and plotted versus the amount of sample extract. The lines are colored according to the coefficient of determination for a linear relationship Intensity~Sample amount. The grey bar indicates the amount of muscle tissue used in the actual experiment.



(b) Distribution of the coefficient of determination (R^2) for a linear fit of amount of sample used and normalized signal measured by GC–MS. By far, most of the metabolites, including the unidentified features, show a strong linear relationship reflected by $R^2 \cong 1$.

Figure 3.1.1: Analysis of linear correlation between sample amount and signal intensity.

3.2 Comparing muscle fiber types

Muscle is the organ that facilitates movement of the body, but also allows to keep body posture and facilitate vital functions like transporting blood through the body and bowel movements. As such it is of no surprise, that it has diverse phenotypes, i. e. in the heart and smooth or skeletal muscle. Skeletal muscle has been characterized as diverse at first by color. The redder a muscle appears the more it relies on oxidative metabolism: the red color stems from myoglobin, the oxygen transporting protein of muscle. [158]

Over time ever more modern techniques have been used to find differences in the different fiber types of muscle. Today four major types of muscle fibers (I, IIA, IID/IIX, IIB) have been established. They are mostly discerned by the respective isoform of myosin heavy chain and their ATPas activity at different pH levels.

Type I fibers are also described as oxidative, IIB as glycolytic, and IID/IIX and IIA as oxidative-glycolytic. Various intermediate forms exist and make up a considerable fraction (>20 %) of overall cross sectional area. (see also table 9) [46]

The classification in glycolytic and oxidative muscles emphasizes the major phenotypic differences of muscles, besides differences in rigidity: their metabolism and major energy source that is routed through the CCM.

In general, the larger the species, the more heterogeneous the fiber type distribution in a specific muscle becomes. Whereas in humans most muscles are strongly intermixed, fiber type distribution in e. g. rodents is more defined.

In the literature, *Soleus* has been often used as example for an oxidative muscle, whereas *Extensor digitorum longus* and *Tibialis anterior* have been used as examples of glycolytic muscles. [111, 112] In the group of Simone Spuler a difference in disease severity afffecting the different muscles has been observed, with *Quadriceps* being most severely affected in an animal model for dysferlinopathy (BLA/J-mice) [1, 159]. Data on systemic molecular comparison for muscle is scarce in general, but even more so for quadriceps.

The predominant animal models for muscle studies are mice and rats. Whereas rat *Soleus* consists almost only of type I fibers [160], in mice it is more heterogenous, with 40 % type I and type IIA each and 20 % type IIAX. It still has the highest percentage of type I fibers of any mouse muscle, the vast majority of the rest are oxidative glycolytic type IIA. Thus *Soleus* is the best choice, if conclusions about predominantly oxidative metabolism are to be drawn. For these reasons I chose the four muscles *Extensor digitorum longus* and *Tibialis anterior* as glycolytic, *Soleus* as oxidative and *Quadriceps* as mixed glycolytic-oxidative (figure 3.2.1), but also because it reflects the most severe disease phenotype in dysferlinopathy.

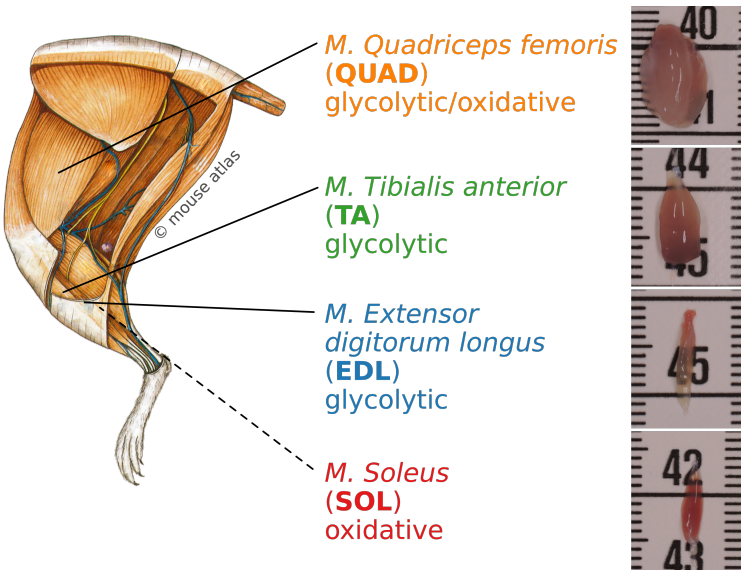


Figure 3.2.1: Picture of a mouse hindlimb. The four muscle types that were investigated are highlighted. *Extensor digitorum longus* and *Tibialis anterior* represent glycolytic muscles, the *Soleus* was chosen as oxidative muscle and has the highest proportin of type I fibers in mice, *Quadriceps* represents an oxidative-glycolytic muscle and shows the strongest disease phenotype in dysferlinopathy. The muscles vary in shape and size, which has to be considered when choosing extraction methods. Image modified from [161]

It is known that for example lactate produced by glycolytic muscle is transported to oxidative muscle via the blood stream, taken up and oxidized there. [162]
This shows in a striking manner, how specialized muscle can be, and warrants further investigation as to how the respective fiber type distribution and different functions of the muscles used in this study influence the molecular makeup of the muscle and how in turn disease susceptibility is affected by this.

I will lay out in the following, how the different sample groups and data types have been handled, ensued by the techniques used to gain further insights from this systemic approach and the integration of both metabolomics and proteomics data.

3.2.1 Data handling

In big "omics" datasets, there is always missing data. If only features (analytes) were kept, that show values for all samples, much information would be lost, especially since I am looking for differences between sample groups. Features showing a strong difference between sample groups, might be below limit of quantification (LOQ) in a certain sample group.

For imputation of missing data different strategies have been suggested [163], however,

a discussion of those strategies is beyond the scope of this thesis.

As mass spectrometry data is inherently count data, a log-normal-distribution of quantitative data is to be expected. [164] As such, all calculations, statistical tests etc. performed to arrive at the conclusions made in this thesis were made on log-transformed data. Generally it has been suggested to use the lower tail of values of a normal distribution, based on the means and standard deviations of the measured data. As an example, if for eight measured samples of the same sample group, only six numerical values were obtained, one would draw eight values from a normal distribution with the mean and standard deviation of the six values measured. The two values imputed are represented by the two lowest values of those eight values.

There are also features, that have no measured values for one or two sample groups. In general, if there is no value, this is usually the case because of suppression effects (in LC–MS) or too low concentration (GC–MS or LC–MS) [93]. Because the respective feature could be measured in other samples, one can expect the main reason for missing data to be low concentration in this particular sample.

Since I am looking for differences between sample groups (muscle types), especially those features are of interest, and should not be filtered out. In general, signals are deemed quantifiable, if they have a signal-to-noise ratio of $S/N < 10$. In order to be able to impute values for features where no value was obtained for the whole sample group, I divided the lowest measured value for a specific feature by 10 and used the standard deviation of the sample group that had the lowest measured value. In this way, the most interesting features are retained, but in a conservative manner. If I simply used a fixed value, I would drastically reduce noise in the data (standard deviation $s = 0$), which would be strongly biased towards those features being "significantly regulated".

Metabolomics Data

After aggregating derivatization products of the identified compounds and discarding metabolites found in less than at least half the samples for at least two different muscle types, a total of 167 metabolic features were annotated in the metabolomics data. Of those 167, 88 are unidentified features (unknowns), 79 could be identified.

Proteomics Data

In the proteomics samples 2457 protein groups were identified by MaxQuant [91]. Of those 2229 were retained after discarding protein groups found in less than half the samples for at least two muscle types.

Samples Retained

Overall we obtained eight samples for each of the following muscle type: *M. Extensor digitorum longus* (EDL), *M. Quadriceps femoris* (QUAD), *M. Soleus* (SOL), *M. Tibialis anterior* (TA). Muscles of one leg was used for proteomics, the other for metabolomics.

Of the proteomics samples the data of all but one sample was kept for analysis. The first sample of QUAD had to be discarded because of machine problems, resulting in many weakly identified proteins. Transfer of measurements between different measurement batches likely results in poor general data quality because of variances in machine performance. Due to restrictions of machine time it seemed not necessary to repeat all measurements for all samples.

Of the metabolomics samples the data of all but two samples was kept for analysis. In two samples of EDL there was an injection error, as indicated by very low levels of cinnamic acid (internal standard). Because of the small size of EDL in mice, no backups could be produced.

Overall the following samples were kept for analysis.

omics technique	**EDL**	**QUAD**	**SOL**	**TA**
metabolomics	6	8	8	8
proteomics	8	7	8	8
combined	6	7	8	8

3.2.2 Data analysis

Essentially I have two datasets, one containing metabolites, the other proteins. Of the metabolites several could be identified, however, there are about as many unidentified features. From a chemical viewpoint metabolites are much more heterogeneous than proteins. At the point of measurement protein samples consist only of peptides, while metabolites encompass sugars, sugar phosphates, amino acids, and carbonic acids.

The lipid compounds are separated during sample processing, but still the sample complexity is much higher. This hinders identification to some degree, and it is only possible to identify features, than can be found in databases. Transferability of these databases is only given to a certain extent, since using only the mass spectrum of the feature in question is not sufficient. For example hexoses or monosaccharides in general are quite similar to another and thus yield similar mass spectra. These compounds can only be identified conclusively by combining their mass spectrum with their retention index (RI). RIs offer only a limited comparability across machines. In order to identify

unknown features unambiguously one would need to obtain the presumed compound and measure it on the same machine.

In my study I used a set of compounds routinely measured with every sample batch. [2, 98] The features still not identified were compared to the Golm Metabolome Database (GMD) [99].

After this, still about half of the features in the data were not identified. It was not the scope of this thesis to try to identify unknown features or to enlarge the inhouse database of the KEMPA group. These unidentified features are still of high value to analysis. Since I was able to determine general linearity between signal intensity and sample amount, they are to be deemed quantifiable and thus changes in their relative amount are biologically meaningful.

In order to distill new insights from any omics set, it is not feasible to inspect every feature on its own, but to first get more general ideas about features that behave similar or to find out if general concepts of knowledge, as GO-terms [103] or pathways (e. g. KEGG-terms [101] or pathways from Reactome [102]) are enriched among the differentially expressed features.

In order to find those differentially expressed features that discern the sample groups, dimension reduction techniques are applied.

I used two different techniques, principal component analysis (PCA) [165] and principal component-discriminant function analysis (principal component–discriminant function analysis (PC-DFA)) [166]. Both methods show certain characteristics of the data. Whereas PCA does not separate intra group variability (biological variation of the mice, variation stemming from sample handling etc) from inter group variability (inherent differences between the muscles), PC-DFA separates these, by application of a two step procedure. First the groups of samples are determined by classical PCA. In a second step, the liniar discriminants between these predefined groups are found.

As such the principal components (PCs) resulting from PCA do not necessarily have inherent meaning, whereas the liniar discriminants (LDs) obtained from PC-DFA inherently show concepts that separate the groups defined by PCA.

When the vectors that separate the sample groups have been obtained, it is possible to find those features that have the strongest influence along the direction of this vector.

There will be a group of features that exert much stronger influence on a specific vector (PC or LD), since not all features will be significantly different between all sample groups. Thus, the values for the features can be subjected to clustering algorithms, separating “important” from “unimportant” features for each discriminating vector.

These “important” features can then be finally subjected to term-enrichment analysis.

Since there are several sub-datasets in the data obtained, I also wanted to see, if any hidden concepts might be revealed, when the aforementioned techniques are applied to those subsets only, or if a certain feature set might show a different similarity of the muscle types. For the unknown features of course, no term enrichment analysis was performed.

Different subsets of the data are of interest:

- proteomics data alone
- metabolomics data alone, of this
  - identified metabolites alone
  - unidentified features alone
- all data combined
- all data combined, without the unidentified features,
  for maximal inferabillity of knowledge

3.2.2.1 PCA separates the muscle types within the first three components

In PCA for all subsets the same general picture emerged:

In every subset samples generated from the same muscle show the highest degree of kinship. Of all subsets, proteomics data splits the sample groups with the greatest distance, while the datasets with a smaller feature space show lower confidence.

The screeplot of a PCA (see figure 3.2.2) is a representation of separation power for the principal components inherent in the dataset. For the different subsets of the data we see the same progression in general:

The first component (PC1), separating the predominantly oxidative muscle *Soleus* from the other, more glycolytic muscle types, contains most of explained variance (≈25 %). This is followed by PC2 and PC3 with eigenvalues of about the same size (≈8–12%). These two components separate the other three muscle types (*Quadriceps*,*Tibialis anterior* and *Extensor digitorum longus*). (see figure 3.2.3) In general, the degree of separation declines with the size of the featurespace used; the separation achieved by using the identified metabolites is slightly greater than by the unidentified metabolites.

The different principal components and factors are loaded by the differentially expressed proteins and metabolites. For principal components, the loadings can be used directly, for the factors importance is inferred from the loadings.

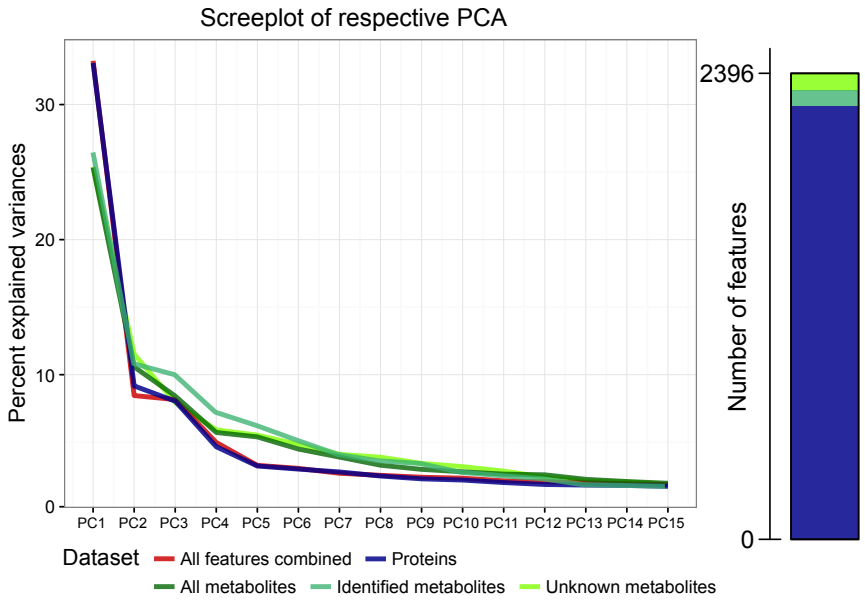


Figure 3.2.2: Screeplot of the different subsets used in the analysis. The first principal component (PC1) contains most of explained variance (≈25 %), followed by PC2 and PC3 with eigenvalues of about the same size (≈8–12%). PC1–PC3 combined separate the muscle types completely for nearly all sample sizes (see also figure 3.2.3.
The bar shows the number of features in the feature spaces used. $F_{proteins}$ is much larger than F_{knowns} and $F_{unknowns}$, which are of approximately equal size among themselves.

In order to get a deeper understanding of the underlying concepts the following strategy was applied:

↪ order the loadings by absolute values

↪ calculate the cumulative sum along the ordered absolute values

↪ perform hierarchical clustering, cutting at two clusters and thus separate the less influential features from the most influential features ⇒ $F_{influential}$

The plot of the features loading a specific component ordered by their values results in a sigmoid-shaped curve, with the most influential features at the tails. (see figure 7.1.1 on page 152 in the appendix). The end at which a feature lies is random, thus one cannot determine a direction of regulation between muscle fiber types from this value. Often the absolute values of the loadings are used to classify the importance of features. They are simply cut at the value of variance a single feature would give by chance, i. e. if there are 10 features, the features explaining more than 10 % of the variance in this particular component are deemed important. For datasets using few features, this has proved to be sufficient.

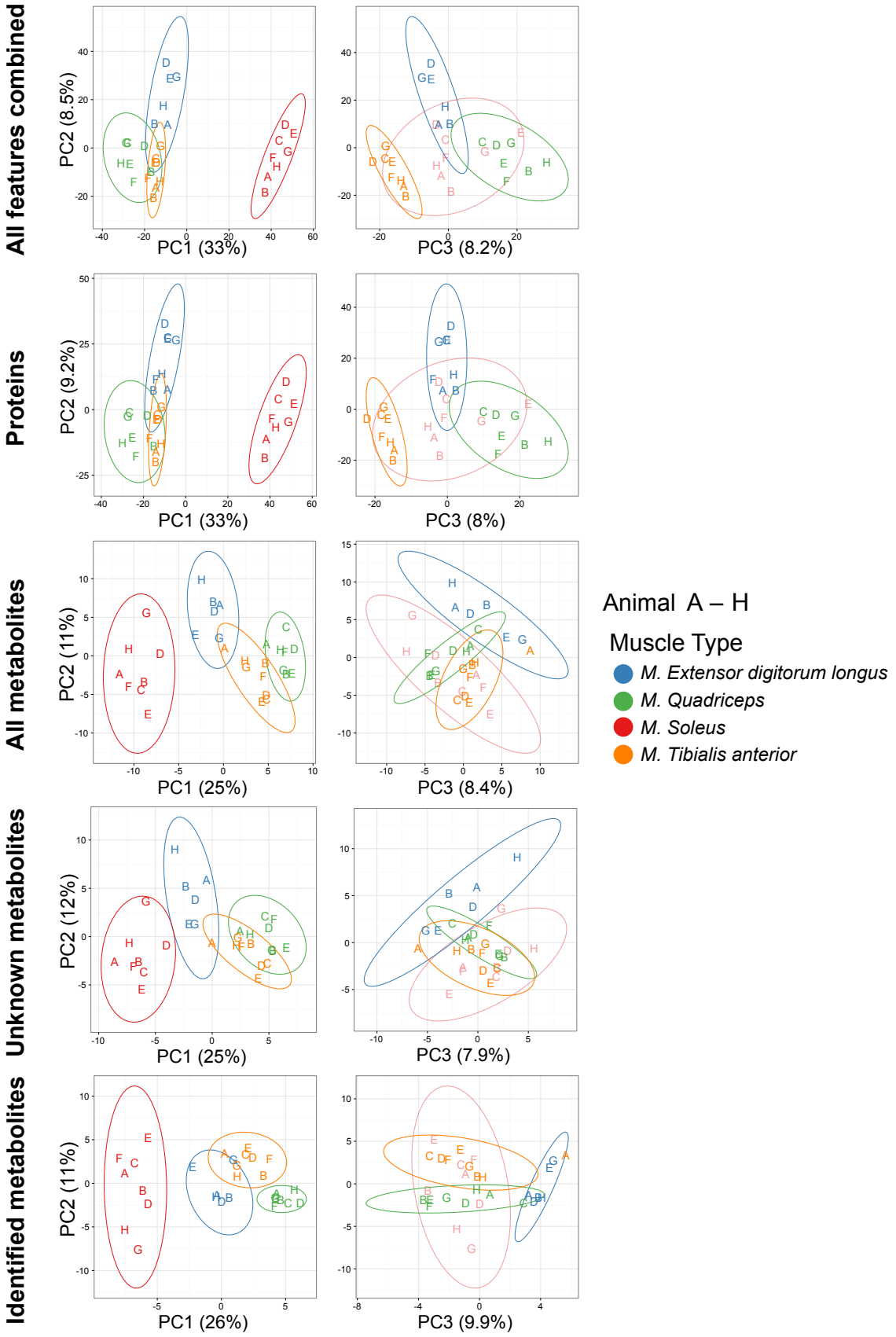


Figure 3.2.3: Overview of the separation of samples by the first 3 principal components. The samples are colored according to muscle type, letters refer to the respective animal. The ellipses around the sample groups depict a confidence area at a level of 95 %. In most subsets the muscle types are completely separated. PC1 separates SOL from the other muscletypes (left column), which are split by the combination of PC2 and PC3 (right column). To point this out more clearly in PC3 vs. PC2 samples from SOL where painted with less opacity. Separation power declines with size of the featurespace; using metabolites alone, QUAD and TA are not completely separated

In systems biology, this approach is not sufficient, since the feature space contains usually hundreds to thousands of features and gives too many false positives. In the present case this would mean almost half of all features to be influential on separation.

For separation of more influential features from less influential features I explored several strategies (cut by percentage of total loadings including FDR for multiple testing correction, hierarchical clustering, Otsu's method [167]). Hierarchical clustering using complete linkage of the absolute values of the loadings prooved to be the technique giving the most reasonable results.

In PC1, separating *Soleus* from the other muscles, the values of the different features showed almost uniform distribution. This means, that almost all features separate SOL from the other muscles. This does not allow to infer meaning, so a more strict cutoff strategy was chosen. By squaring the absolute values, the distribution is transformed to emphasize the higher values, i. e. forcing it to become more right-skewed. This transformation reduces the size of the space of important features ($F_{influential}$) from 1229 to 83, when applied to the full dataset.

3.2.2.2 The terms that are enriched in the regulated features show a focus on metabolic and muscle specific concepts

Features, that have been found to be highly regulated between muscle types ($F_{influential}$) which are separated by a specific principal component, can be subjected to term enrichment analysis, in order to find out, if general predefined concepts like GO-terms or pathways are differentially regulated between the muscle types. Overall the feature space of the influential features (proteins or metabolites) $F_{influential}$ was checked for enrichments of ontology concepts, using all features of the respective subset of data as background, e. g. if only metabolites were used in the PCA, the differentially regulated metabolites were tested for enrichment, with all metabolites as background dataset.

When the proteins are analysed separately, the ontology from GO [103, 104] can be used, since there are no GO-terms assigned to metabolites.

For combined features or metabolites alone the KEGG database [101], Reactome [102] and WikiPathways [168] as well as SMPDB [169] can be used.

These databases generally provide the possibility to either map protein identifiers (e. g. UniProt-IDs) or metabolite identifiers (e. g. KEGG-compound-IDs) but not at the same time.

KEGG can be considered the most widely used database of ontologies of metabolism. As such I performed the most detailed analysis using the KEGG database. I retrieved

all pathways the features in the respective dataset could be mapped to and performed an enrichment analysis using Fisher's exact test [105], to determine the enrichment of terms in a specific principal component. The returned *p*-values were corrected using the Benjamini-Hochberg-method (FDR). [170]

For the other databases mentioned I used IMPaLA [171] with the ConsensusPathDB [172] which offer the functionality to map both metabolites and proteins simultaneously. For proteins or genes, this tool only allows to map human identifiers. As such I used the UniProt database [106] to translate the UniProt-IDs for mouse proteins to their human orthologs.
It should be mentioned, that IMPaLA does not include multiple testing correction. The length of the whole results table was used to estimate overall background size for multiple testing correction.

3.2.2.2.1 GO-term enrichment in PCs from proteins

For the analysis of the regulated features of the proteins I used GOrilla [173], which also conserves hierarchy of GO-terms. After FDR correction for PC1 92, for PC2 none and for PC3 51 significantly enriched terms were found at a *p*-value <0.05). As GO-terms often are closely related to one another (e. g. TCA-cycle, mitochondrium and β-oxidation) it makes sense to aggregate these terms, to get a more general overview. For this I used REViGO [174]. In figures 3.2.4 and 3.2.5 the GO-terms found in PC1 and PC3 are shown.

In PC1, separating *Soleus* from the other muscles, the dominant theme is energy: "ADP metabolism", "carbohydrate metabolism" and "generation of precursor metabolites and energy". Also muscle contraction and muscle development are different between *Soleus* and the other muscles. Another theme is the response to ions ("calcium mediated signaling" and "ion membrane transport").

In PC2, separating EDL from QUAD and TA, no enrichment of any term was found at a *p*-value < 0.05. However a few pathways could be found enriched, with *p*-values slightly above the cutoff. See below.

In PC3, separating TA from QUAD and EDL and also QUAD from EDL for the most part, oxidative metabolism has the highest influece ("TCA-cycle", "acyl-CoA metabolism"). But also processes involving interaction with DNA seem to distinguish these muscle types.

The pathways enriched in the PCs of $F_{proteins}$ are outlined below in combination with metabolites.

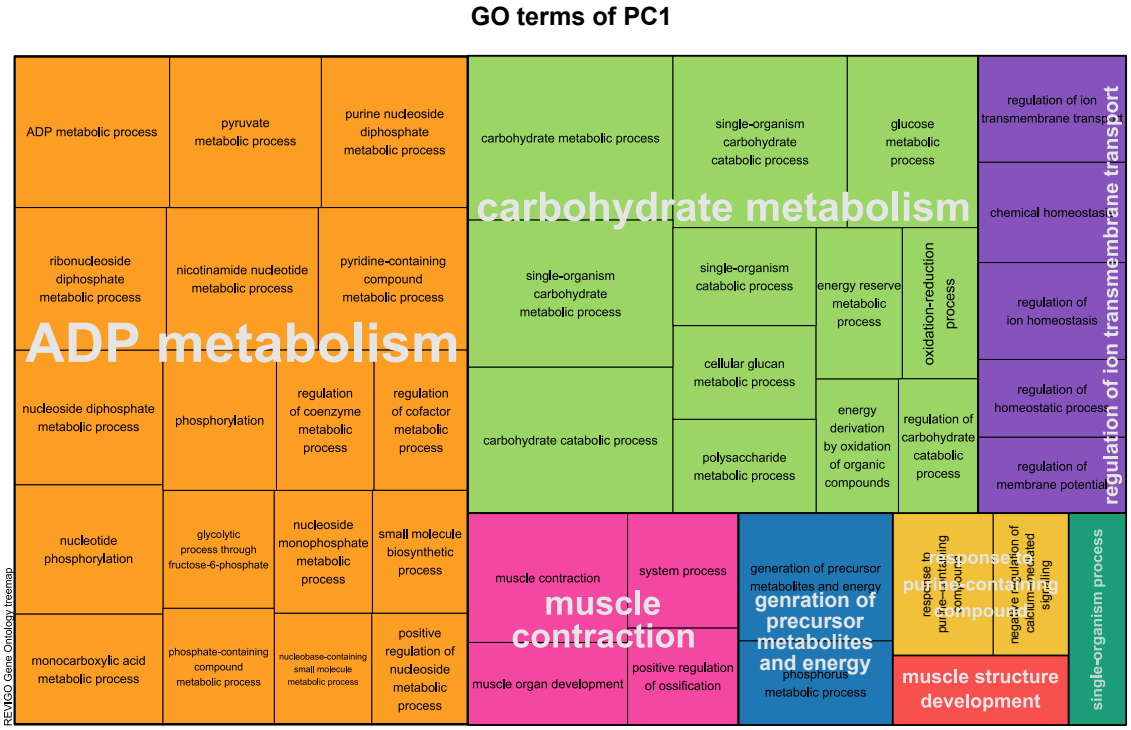


Figure 3.2.4: Treemap showing the significantly enriched GO-terms for PC1, which separates SOL from the other muscle types. Major themes are energy metabolism (CCM)-related, centering on carbohydrates, but also muscle contraction and ion transport through the membrane are different between SOL and the other muscle types.

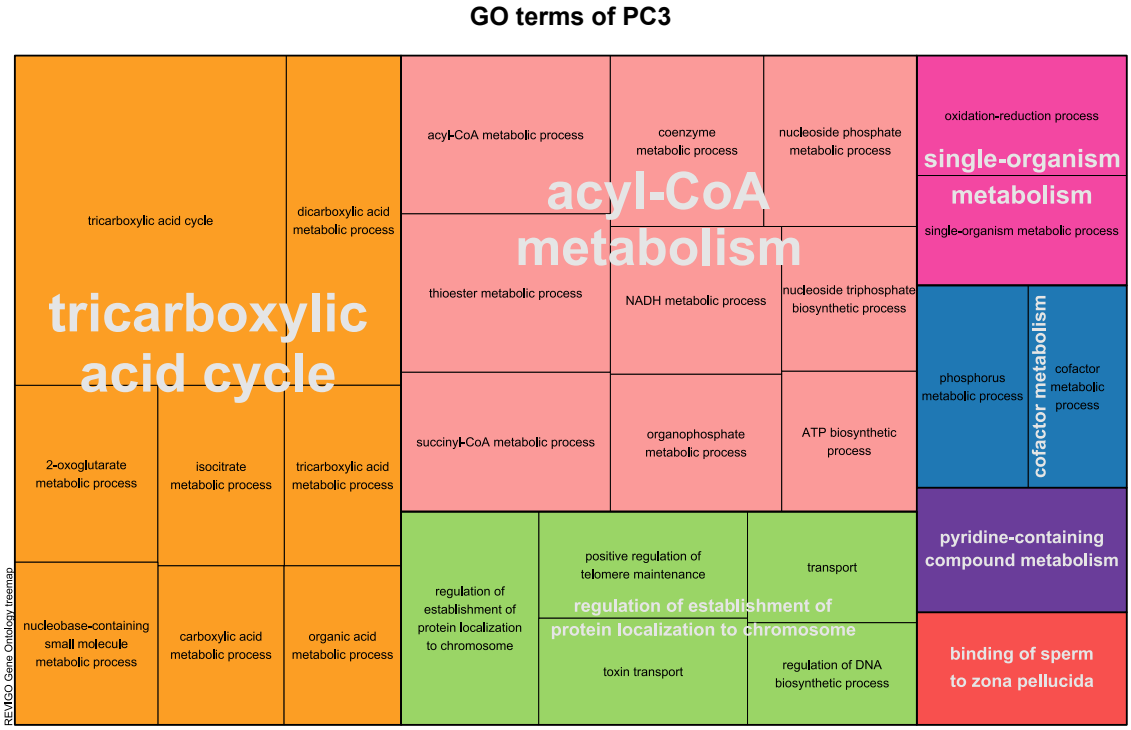


Figure 3.2.5: Treemap showing the significantly enriched GO-terms for PC3, which separates QUAD, EDL and TA with TA being most different from the other muscle types. Major themes are energy metabolism (CCM)-related, centering on mitochondrial (oxidative) metabolism but also DNA regulation is an important theme.

3.2.2.2.2 Metabolites

For the identified metabolites alone, there was no significant enrichment of any term found.

3.2.2.2.3 Enrichment in PCs when metabolites and proteins are combined

In order to not only simply give *p*-values for significantly enriched terms, but also to see, if selection criteria of "important" features $F_{influential}$ was stable, and also to see if and which additional insights can be obtained by combination of both proteomics and metabolomics data, *p*-values obtained from the respective subsets were plotted against each other. This is comparable to the 2D-annotation enrichment proposed by Cox and MANN. [175] The results from this analysis are outlined below.

Combining proteins and metabolites for pathway enrichment analysis gives some redundant pathways, enriched both in the feature space of proteins alone ($F_{proteins}$) and the combined feature space ($F_{combined}$) but also some, that are only significant if metabolites are included in the analysis. Of the redundant pathways no significant pathways are less significantly regulated in $F_{combined}$ (i. e. have a higher *p*-value), but many, if not most are more significantly enriched.

In PC1 (see figure 3.2.6), separating *Soleus* from the other muscles, as for the proteins alone the most dominant pathway is "Glycolysis/Gluconeogenesis". But also other pathways involving carbohydrate metabolism are differentially regulated comparing *Soleus* to the other muscles ("Starch and sucrose metabolism", "Galactose metabolism", "Pentose Phosphate Pathway" (PPP), "Insulin resistance", "Phosphotransferase system" (used by bacteria [176–178]) and others). Again also "Calcium signaling" is a significant term. There are no additional significantly enriched terms in the complete ConsensusPathDB, searched with IMPaLA.

The only term significantly enriched in PC2 (see figure 3.2.7), separating *Extensor digitorum longus* from *Quadriceps* and *Tibialis anterior* is "Metabolism" (from Reactome), in $F_{combined}$ as well as in $F_{proteins}$. "Alzheimer's disease" and "Electron transport Chain" (from WikiPathways) are slightly above the cutoff, as well as "Glucosinolate biosynthesis" and "2-Oxocarboxylic acid metabolism" for $F_{combined}$.

In PC3 (see figure 3.2.8), as for the GO-term analysis, TCA-cycle is the most significantly enriched term. Again also other terms related to conversion of oxygen are enriched ("Electron transport chain", "Oxidative phosphorylation", "Fumarase deficiency", and others). Also the terms for some neurological diseases emerge, but this is explainable by the fact that also in those terms, mitochondrial metabolism is included as well. "Glyoxylate

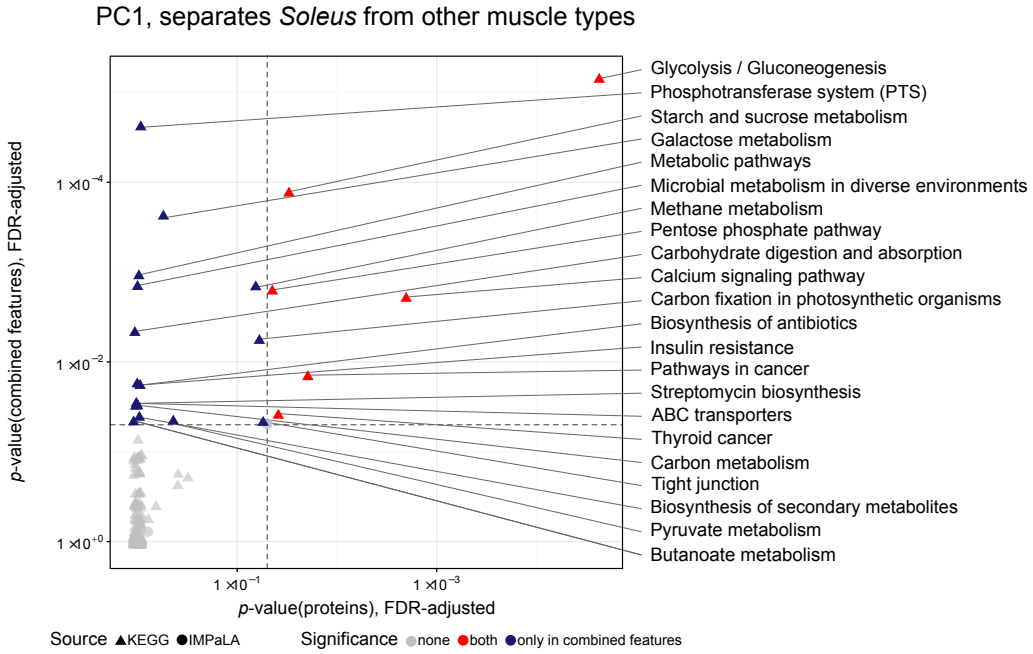


Figure 3.2.6: Plot of p-values for enrichment of metabolic ontologies retrieved from KEGG and IMPaLA in PC1, separating SOL from other muscle types. The emerging themes are revolving around carbohydrate metabolism (glycolysis, PPP, galactose, starch/sucrose metabolism). When metabolites are included, additionally "Insulin resistance" and interestingly the phosphotransferase system (PTS) are found. In the term PTS, which is of importance in microbial carbohydrate metabolism and its adaption to nutrient availability in the environment [176–178], many sugars but also sugar alcohols are included.
Dashed lines indicate a cutoff at $p = 0.05$.

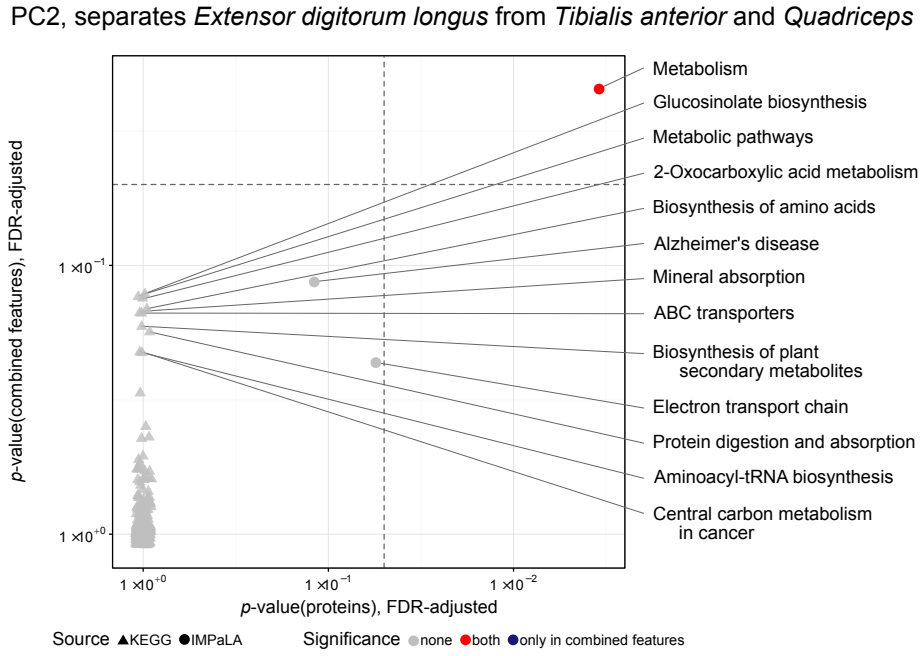


Figure 3.2.7: Plot of p-values for enrichment of metabolic ontologies retrieved from KEGG and IMPaLA in PC2, separating EDL not significantly from the other muscle types. The only concept found significantly enriched was the very general term "Metabolism" in general, also when only the proteins ($F_{proteins}$) were searched. Using also metabolites raises significance for some terms, but not to a significant level.
Dashed lines indicate a cutoff at $p = 0.05$.

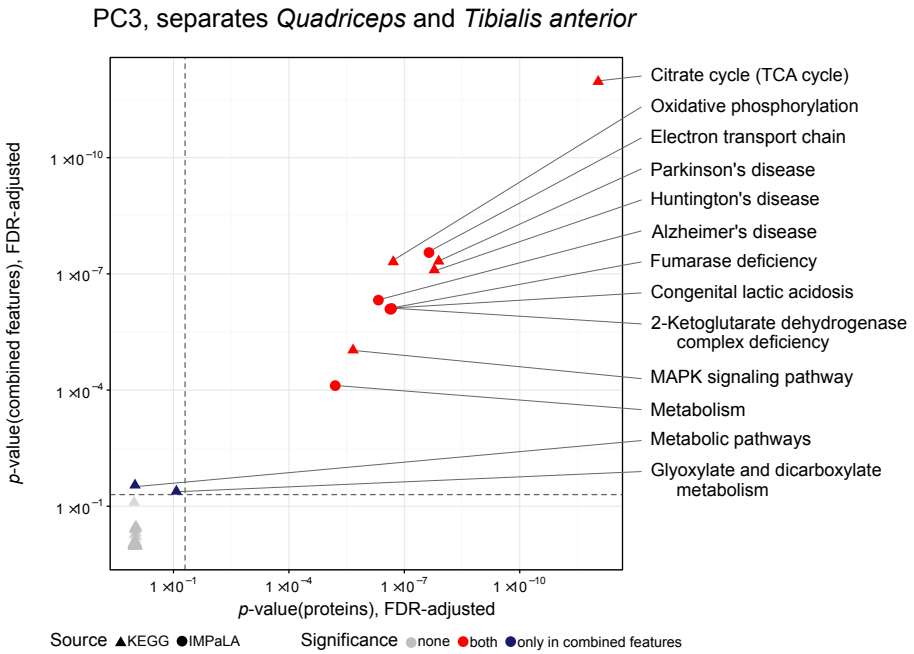


Figure 3.2.8: Plot of p-values for enrichment of metabolic ontologies retrieved from KEGG and IMPaLA in PC3, separating mostly QUAD and TA. Most concepts here revolve around oxidative metabolism, TCA-cycle being the most significant. The other terms are included, because they generally contain elements of mitochondrial metabolism.
Dashed lines indicate a cutoff at $p = 0.05$.

and dicarboxylate metabolism" is introduced by $F_{combined}$ only.

Overall the principle components separating the muscles are found to be formed by features that can be condensed under the general concepts of either oxidative or glycolytic metabolism, but also general muscular function.

3.2.2.3 Principal component-discriminant function analysis (PC-DFA)

Principal component-discriminant function analysis (PC-DFA) gives a less comparable picture among the data subsets than PCA. In most subsets the muscle types are fully discerned by either discriminants LD1 and LD2, latest in combination with LD3, but also LD2 and LD3 alone are mostly able to separate the sample groups from each other.

However, in PCA the splitting was comparable and discernable. In PC-DFA it is only comparable between the datasets that include either all features and the proteins in LD1 and LD3, with LD3 being reversed for proteins compared to all features combined.

In the following I will give a general description of the linear discriminants obtained and then determine, as which concepts are enriched in the features that cause the respective separation.

General description of the linear discriminants

Figure 3.2.9 shows the linear discriminants 1–3. This type of plot is generally comparable to the plot of the principal components (figure 3.2.3).

For the subsets consiting of *proteins* ($F_{proteins}$) and *all features combined* ($F_{combined}$) the following situation is found:

LD1 separates SOL from EDL, QUAD, and TA as did the first principal component in PCA.

LD2 separates QUAD from the other muscles, with EDL and TA being most similar.

LD3 separates EDL and TA, with LD3 being reversed when all features are combined compared to the proteins alone.

For subsets containing only *metabolites* we see a different separation.

LD1 here separates TA from SOL and also from EDL and QUAD, to different extent and order if the identified metabolites (F_{knowns}), or unidentified ($F_{unknowns}$) are used. In F_{knowns} SOL is more similar to EDL and QUAD than TA is, while in $F_{unknowns}$ a pattern similar to $F_{proteins}$ is visible.

LD2 separates EDL from the other muscles, with almost no variance for TA in F_{knowns}. EDL shows close to no variance among its replicates in all other LDs but LD2. For F_{knowns} TA shows no variance among its replicates in LD2.

LD3 In F_{knowns}, similar to LD1, LD3 separates TA from the other muscles, but to lesser extent, while in $F_{unknowns}$ QUAD is moderately separated from the other groups.

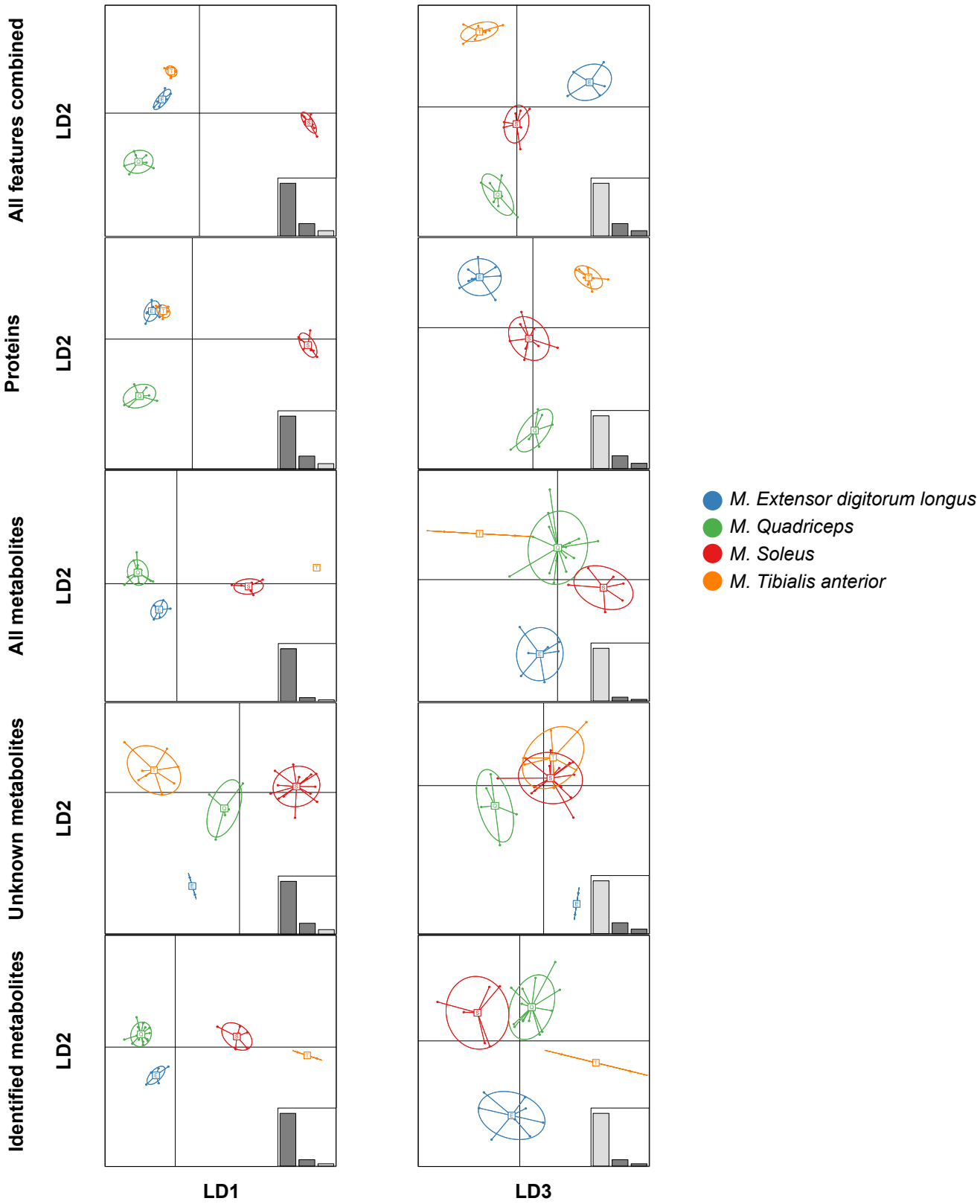


Figure 3.2.9: Plot of linear discriminants of the respective data-subsets. The insets show the relative eigenvalues of the linear discriminants (LDs) (similar to the screeplot of a PCA), shaded bars representing the LDs of the respective plot. For proteins and all features combined, there is a comparability, on how factors split the sample groups, even similar to PCA (figure 3.2.3). For metabolites the overall separation is less congruent.

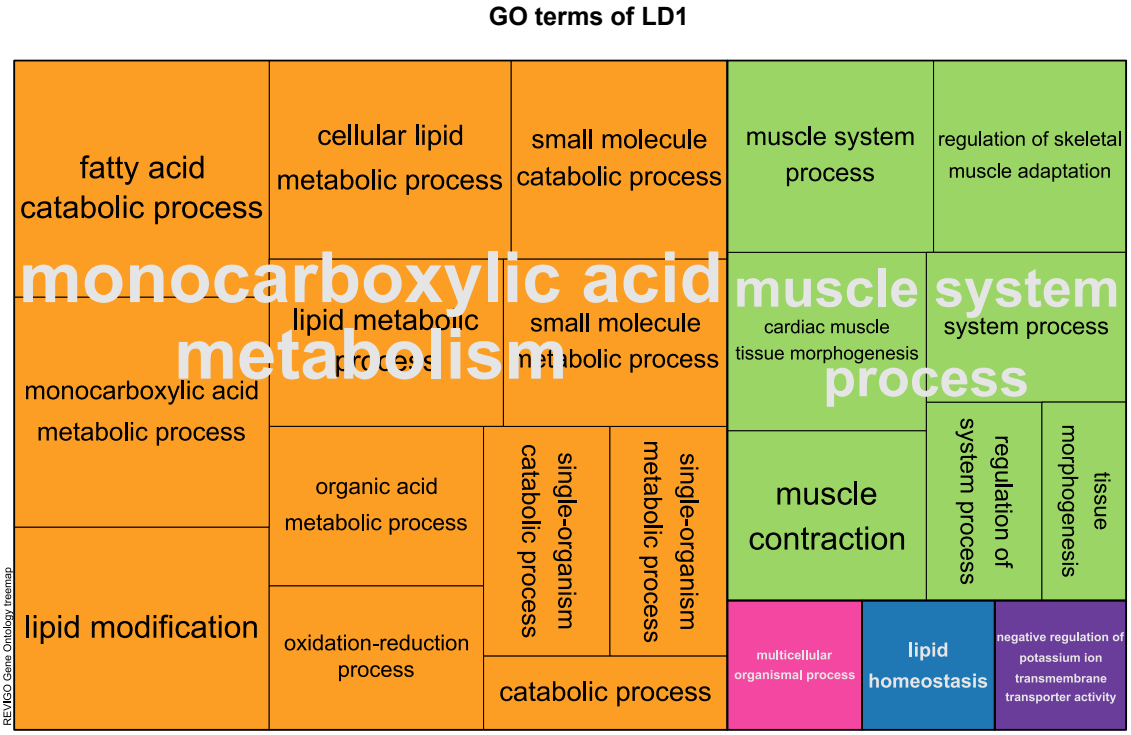


Figure 3.2.10: Treemap showing the significantly enriched GO-terms for LD1, separating SOL from the other muscle types. Major themes are carboxylic acid metabolism, that involves β-oxidation and overall lipid metabolism, as well as general processes revolving around muscle activity, -loss, -development and -adaption.

Overall, for $F_{combined}$ and $F_{proteins}$ a pattern similiar to PCA emerges, while the picture for metabolites is less clear and shows a completely different separation, comparing $F_{unknowns}$ and F_{knowns}.

In light of the given results, it seems not warranted to combine "interesting" features from $F_{proteins}$ and F_{knowns}. As such only the GO-term-enrichment for $F_{proteins}$ and the significantly enriched KEGG-terms in $F_{proteins}$ and $F_{combined}$ will be inspected.

3.2.2.3.1 Proteins

The same analysis pipeline (GOrilla→REViGO) as for $F_{proteins}$ from the PCA was used. (see page 61)

The biggest themes emerging for LD1 (figure 3.2.10), that separates SOL from the other muscles, are carboxylic acid metabolism, that involves β-oxidation and overall lipid metabolism, as well as general processes revolving around muscle activity, -loss, -development and -adaption.

In LD2, that separates QUAD from EDL and TA, no significantly enriched GO-terms could be found.

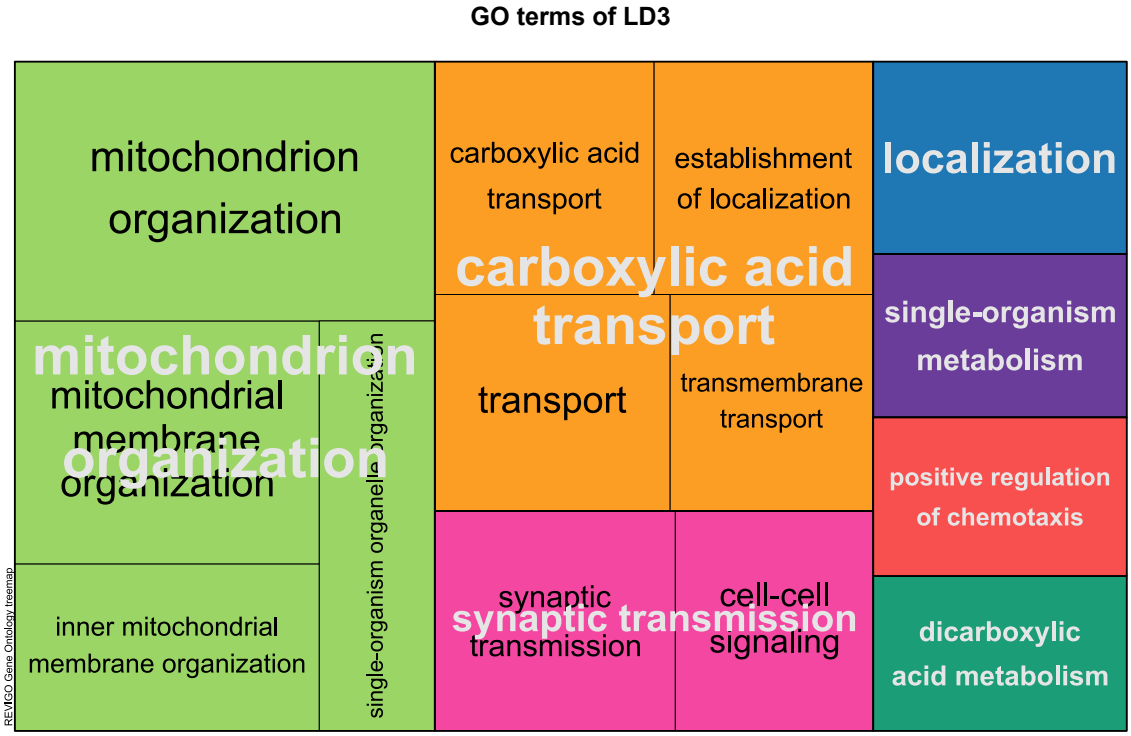


Figure 3.2.11: Treemap showing the significantly enriched GO-terms for LD3, separating EDL and TA from each other. The main themes include localization and transport, of mitochondria and, similar to LD1 carboxylic acid transport.

In LD3 (figure 3.2.11), separating EDL and TA from each other and to some degree from QUAD and SOL, the main themes are localization and transport, of mitochondria and, similar to LD1 carboxylic acid transport.

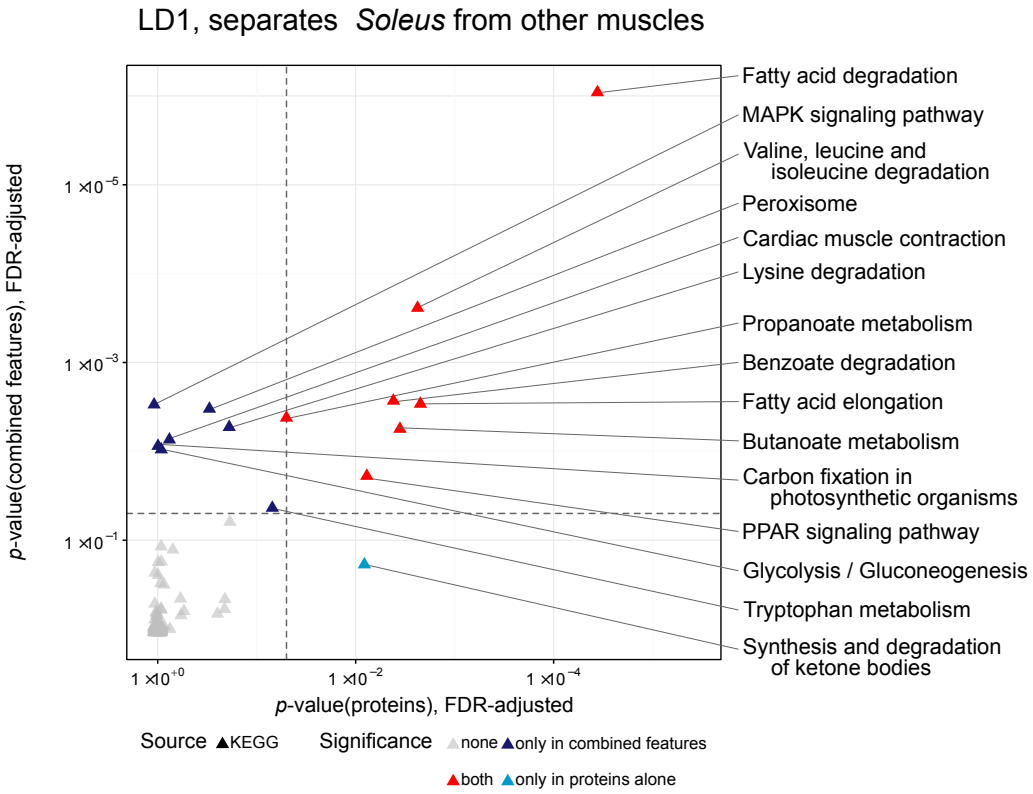


Figure 3.2.12: Plot of p-values for enrichment of metabolic ontologies retrieved from KEGG in LD1, separating SOL from other muscle types. Fatty acid degradation, elongation and oxidation ("Peroxisome" in $F_{combined}$) and short chain organic acid metabolism dominate. "Glycolysis/Gluconeogenesis" is only significantly enriched in $F_{combined}$, while "Synthesis and degradation of ketone bodies" is enriched only in $F_{proteins}$.
Dashed lines indicate a cutoff at $p = 0.05$.

3.2.2.3.2 KEGG-terms As in PCA for the metabolites alone, there was no significant term to be found at or close to the p-value-threshold of 0.05.

LD1 The KEGG terms associated with LD1 (figure 3.2.12), are similar to the themes found in GO-term analysis. Fatty acid degradation, elongation and oxidation ("Peroxisome" in $F_{combined}$) as well as short chain organic acid metabolism dominate. Also MAPK signaling ($F_{combined}$) and PPAR signaling in both feature spaces are found to be different between SOL and the other muscle types. "Glycolysis/Gluconeogenesis" is only significantly enriched in $F_{combined}$.
While there was no such case in PCA, in PC-DFA we see a term enriched only in $F_{proteins}$: "Synthesis and degradation of ketone bodies".

LD2 In contrast to GO-term analysis and PCA overall, LD2 (figure 3.2.13) shows two significant KEGG-terms:"TCA-cycle" and "Fatty acid biosynthesis", diverging between

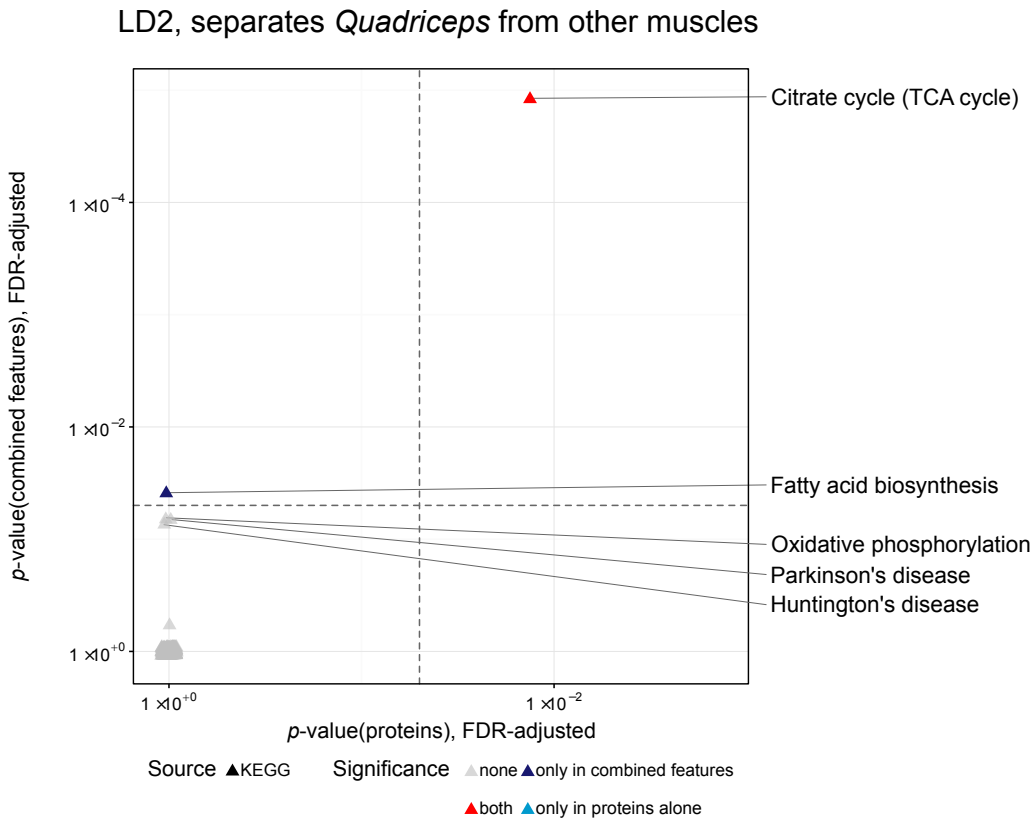


Figure 3.2.13: Plot of p-values for enrichment of metabolic ontologies retrieved from KEGG in LD2, separating between QUAD and the other muscles, with SOL being closest to QUAD. "TCA-cycle" and "Fatty acid biosynthesis" are the only terms significantly enriched, the latter only in $\mathrm{F}_{\mathrm{combined}}$.
Dashed lines indicate a cutoff at $p = 0.05$.

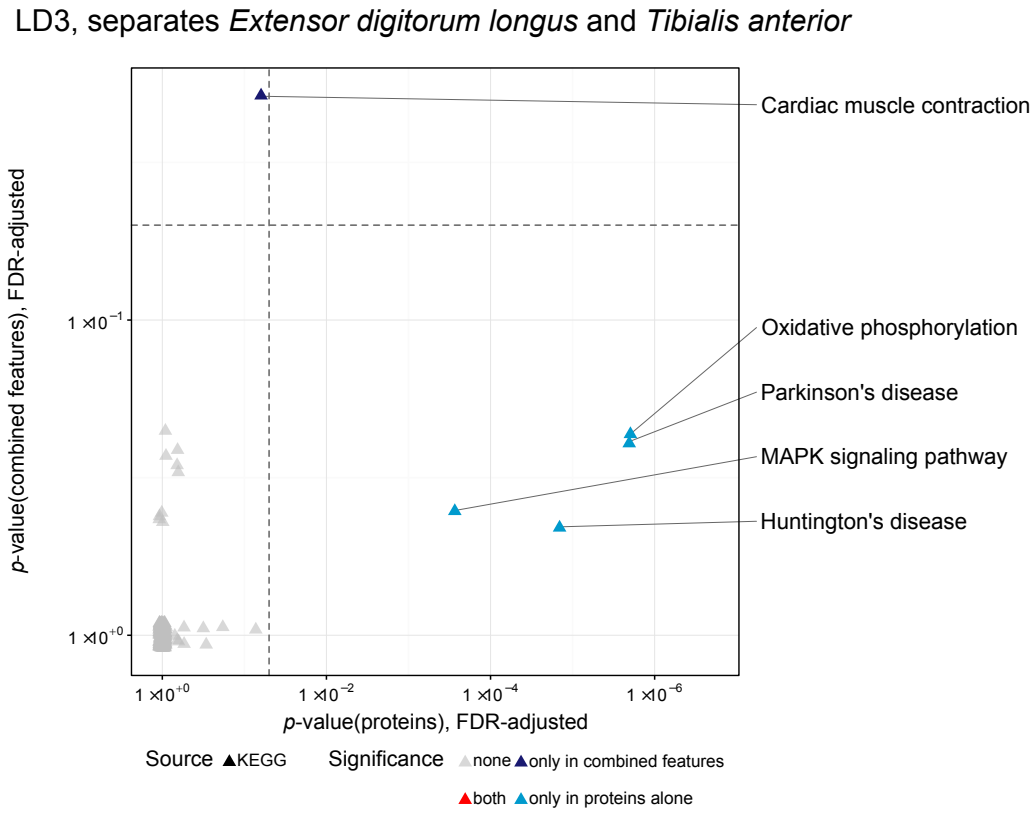


Figure 3.2.14: Plot of p-values for enrichment of metabolic ontologies retrieved from KEGG in LD3, separating mostly EDL and TA. No overlapping terms significantly enriched in both feature spaces compared were found. In $F_{combined}$ "Cardiac muscle contraction" was found to be enriched in this comparison, relating to the general theme of muscle. In $F_{proteins}$ pathways —relating to, or because they include elements of— oxidative phosphorylation are enriched. Dashed lines indicate a cutoff at $p = 0.05$.

QUAD and the other muscles, with SOL being closest to QUAD. “Fatty acid biosynthesis” is only enriched in $F_{combined}$. As in PC2 some other terms, likely due to their general theme “mitochondrial metabolism” are close to the selected p-value-threshold of 0.05 ($\approx$0.07) in $F_{combined}$.

LD3 LD3 (figure 3.2.14), separating mostly EDL and TA no overlapping terms significantly enriched in both feature spaces compared were found. In $F_{proteins}$ mitochondrial metabolism and also the “MAPK signaling pathway” were found to be regulated. In $F_{combined}$ “Cardiac muscle contraction” was found to be enriched in this comparison, relating to the general theme of muscle. This last finding is especially peculiar, since this pathway involves no “classical metabolites” in the sense of small molecules, measured in this study, but only proteins and ions (Ca^{2+}, H^+, K^+ and Na^+). In $F_{proteins}$ pathways —relating to, or because they include elements of— oxidative phosphorylation have been found.

Overall the themes separating the different muscle types are mostly CCM-related (TCA-cycle, PPP, Glycolysis), or muscle specific, the latter especially revealed by PC-DFA.

3.2.2.4 Each muscletype has its own molecular makeup

From the analysis of the dataset using dimension reduction techniques, it is clear, that the major differences between the muscles lie in their fuel selection and differences in the build up of the motor units. Figure 3.2.15 shows metabolites and proteins mapped to the central carbon metabolism. It is immediately obvious, that also the presumably similar glycolytic muscles EDL and TA have their own phenotype. Every aspect of the metabolism is tightly regulated in order to achieve maximal performance towards the specific roles all of these muscles have. This also shows, that the severity of muscle disease will likely affect each muscle in a different way.

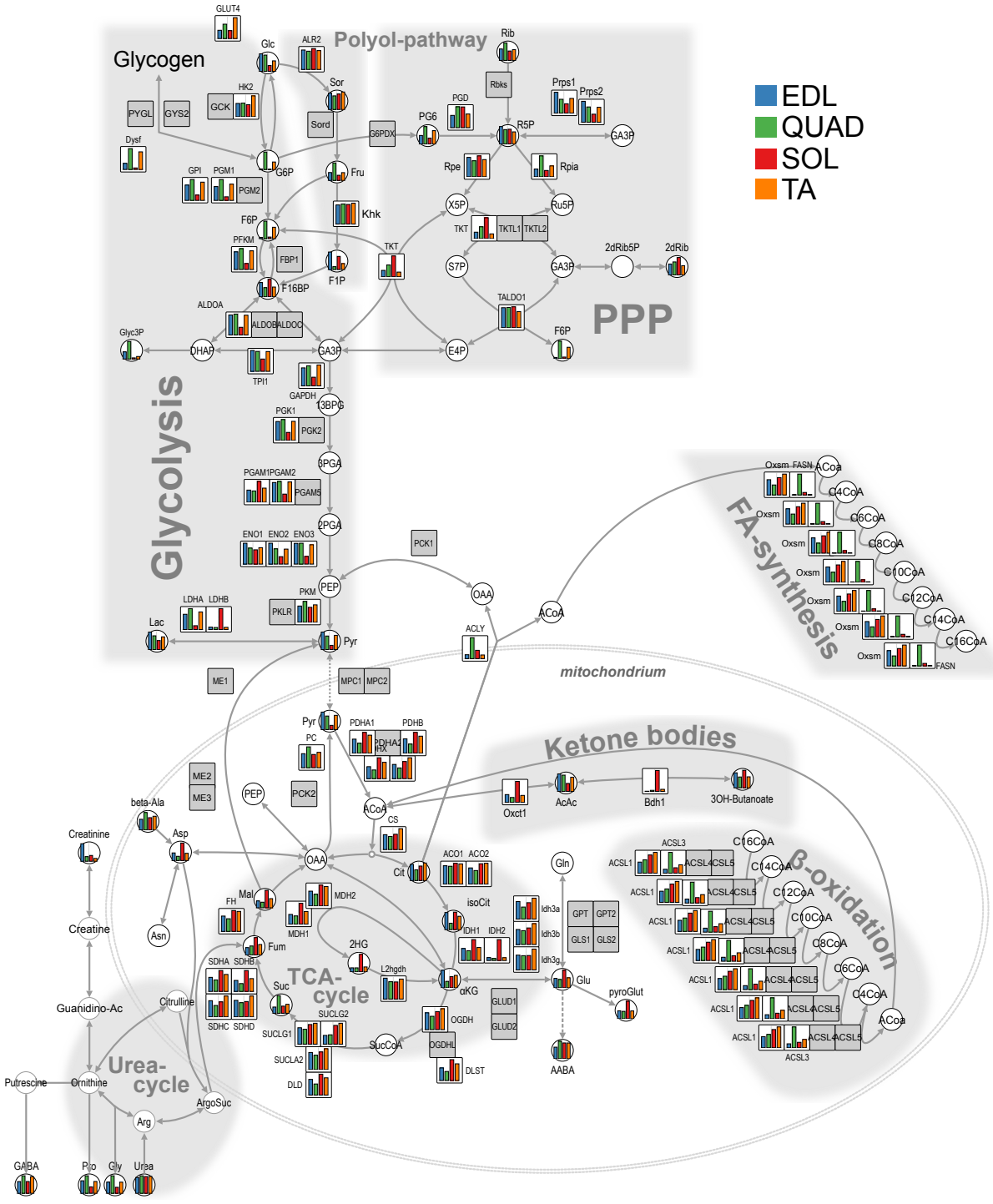


Figure 3.2.15: Network of the central carbon metabolism (CCM), with mapped metabolites and proteins. The distinct fine tuned differences of partly only small parts of pathways is immediately evident. The presumably similar glycolytic muscles TA and EDL are very distinct on a molecular level.

3.3 Label incorporation in myotubes from dysferlinopathy patients

Dysferlinopathy, or limb girdle muscular dystrophy 2B (LGMD2B) is an autosomal recessive genetic disease with a typical onset in early adulthood. [179] During childhood and their teenage years, the patients are typically quite active, many very athletic. [65] With disease progression they loose more and more strength and become non-ambulant after 10–20 years of diagnosis. In later stages they are typically wheelchair-bound and often dependent on respirators. Except for symptomatic care, there is no treatment available.

On a molecular level, the disease is characterized by a loss of function mutation of dysferlin. Dysferlin itself is highly important for membrane repair after injury of the muscle. [67, 68, 180]

Dysferlinopathy is a major research topic in the group of Simone Spuler. In data obtained from dysferlin deficient mice and primary human muscle cells from patients diagnosed with dysferlinopathy obtained by Sarah Keller [1], a metabolic phenotype was found, that points towards an impairment of hexokinase II function in dysferlinopathy. [1] The levels of Sorbitol and Fructose as members of the polyol pathway [14] were increased, while metabolites of glycolysis, downstream of hexokinase were decreased. A possible reason for this might have been a lower expression of hexokinase II (HK2), the enzyme, that phosphorylates glucose and forms G6P at the first step of glycolysis. However, proteomics analysis showed, that glycolytic enzyme levels are unaffected, in mouse and patient sample material.

Another reason for the low levels of G6P might be a higher uptake of G6P into the glycogen pool. It was also not clear, if the higher poolsizes of sorbitol and fructose are due to higher build-up rates or if this increase is observed because further processing of sorbitol is reduced.

In the Kempa group, a method to use pulsed stable isotope resolved metabolomics (pSIRM) in cancer cells had already been developed and established, when I started my work in the group. [2, 75]

pSIRM involves treatment of the cells with a substrate, that contains stable isotopes, usually ^{13}C. The cells take up the substrate and after a certain time, metabolism is quenched by extracting the cells with cold 50 % methanol$_{aq}$. The extract is added to a prepared amount of chloroform, and after phase equilibration and separation, lipid and polar phase are analyzed separately by GC–MS. The natural amount of ^{13}C makes up ≈1.1 % of total carbon. [181] Since it cannot be known, how much ^{13}C there is in a certain sample, or if

there is overlap in mass peaks with a (partially) co-eluting compound also some samples are treated the same, only that a ^{12}C substrate is used. This can be subtracted in the subsequent data analysis and allows for background correction.

The entire process is tailored to the end, that the influence on the cells is to be kept as small as possible. If, after some growth period there would be simply a switch of media including the ^{13}C-substrate, also a strong change in nutrient concentration might be introduced. Because of that, a media exchange some hours before labeling start, using ^{12}C-containing media is performed. This allows the cells to settle on the conditions applied by fresh media; at the same time substrates will not be depleted too much until the start of the labeling.

In general it is sufficient to produce a background control with ^{12}C only for one time point, somewhat in the middle of all labeling times used in a certain setup. If the times start to get stretched out too long (over several hours), obtaining more than one background time point can be advisable.

The application of this method has the potential to answer the question on activity of the enzymes involved in this observed increase of polyol pathway level metabolites on the one hand, and at the same time the decrease of G6P and further downstream metabolites of glycolysis.

In order to get a deeper understanding of this phenotype, I established pSIRM in primary human muscle cells, including a means to also trace label incorporation into glycogen.

3.3.1 Establishment of pSIRM in muscle cells

Primary human muscle cells require quite special conditions in order to grow in cultivation. Usually SKMGM (provitro), supplemented with foetal calf serum (FCS) and several growth factors is used in the group to grow muscle cells. The media innately contains 5 % FCS, another 10 % FCS are added shortly before usage.

As we wanted to see whether there is a prevalence for utilization of glucose, fatty acids or ketone bodies, the standard media could not be used, since it contains 1 g/L glucose and the usually used supplements, namely FCS contains fatty acids. If this medium had been used in pSIRM experiments, the label had been diluted, and great interference would have to be expected, when monitoring the fate of a single carbon source, since also intracellular label intensity would have been influenced by other carbon sources.

To overcome this a custom made SKMGM-knock-out-version that contained no glucose and no FCS was developed in cooperation with proVitro, .

This media could then be supplemented with the desired ^{13}C-Source; as serum compo-

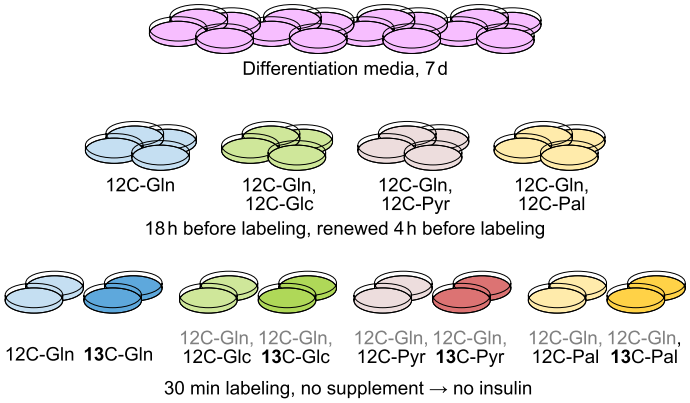


Figure 3.3.1: Experimental setup to show principal uptake of substrates into myotubes in cell culture. Glutamine was always present, because at that time the components of the supplement of SKMGM was only available in pre-mixed form at that time.

nent, dialyzed FCS was used, which contains no free fatty acids (FFAs) and no free amino acids.

I labeled myotubes derived from primary muscle cells of normal healthy controls (2 plates per condition) with two carbon sources only (glutamine was always present) for 30 min (see fig. 3.3.1) in order to test if the substrate is taken up at all. Before labeling, the myotubes were accustomed to the carbon sources for 18 h.

The data show, that all the carbon sources used (glucose, glutamine, palmitate, pyruvate) are taken up and metabolized by myotubes (see figure 3.3.3). Thus the myotubes can be somewhat forced to use the carbon source present.

3.3.2 Uptake of substrate, with all carbon sources present

In order to test substrate uptake of myotubes in cell culture under normal conditions (with all carbon sources present) I labeled myotubes (see figure 3.3.2) with complete media, using ^{13}C-glucose, ^{13}C-glutamine, ^{13}C-palmitate and ^{13}C-pyruvate for 30 min (3 plates each) and with ^{13}C-glucose also for 24 h (2 plates each), using 60 % methanol for the extraction, in order to be able to trace label-incorporation in glycogen (see below and figures 3.3.3, 3.3.6).

The data from this experiment show, what one would expect: looking at citrate, as a converging point of all carbon sources, we see that as soon as glucose (and insulin, which was always in the media) are present, predominantly glucose is taken up and metabolized, whereas from the other substrates label from pyruvate, that is metabolically very near to citrate, is higher. (see fig. 3.3.3)

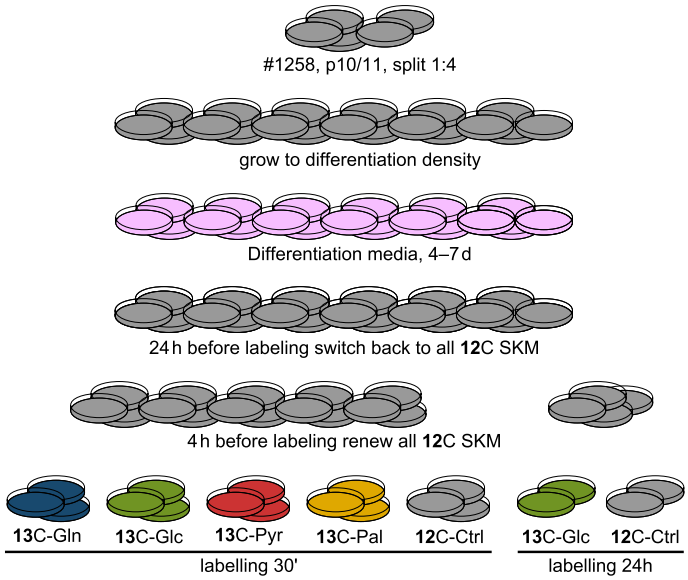


Figure 3.3.2: Experimental setup to show principal uptake of substrates into myotubes in cell culture. Gln: glutamine, Glc: glucose, Pyr: pyruvate, Pal: palmitate

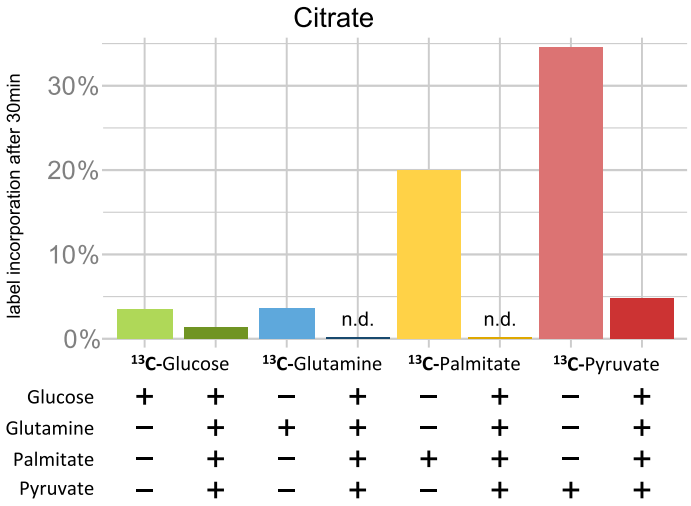


Figure 3.3.3: Labeling of myotubes for 30 min. When only one carbon source is present in preconditioning and during labeling, this carbon source is metabolized. (Glutamine was always in the media, however its impact can be considered of minor concern.) If all carbon sources are present, predominantly glucose is metabolized, as can be seen by the decrease in label in citrate originating from palmitate and pyruvate. Whereas a little amount of label-incorporation from pyruvate can be found, no label from palmitate was detected. It can be assumed, that the reduction in label-incorporation from glucose is due to the pyruvate still being used as a carbon source.
Gln: glutamine, Glc: glucose, Pyr: pyruvate, Pal: palmitate

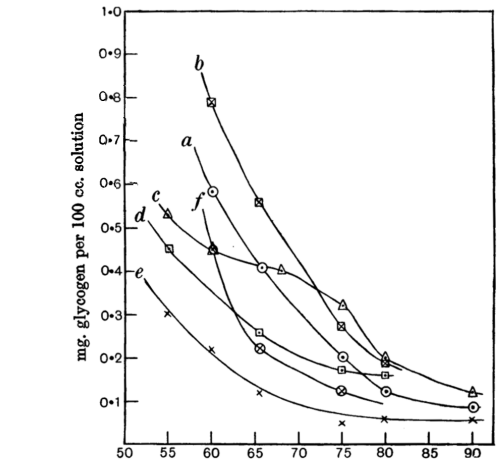


Fig. 2. The solubility of glycogen in aqueous alcohol.
(*a*) Alcohol + 0·6 % KCl. (*b*) Alcohol + 4 % CCl_3COOH.
(*c*) Alcohol + 21 % CH_3COOK. (*d*) Alcohol + 42 % CCl_3COOK.
(*e*) Alcohol + 15 % KCl. (*f*) Alcohol + 15 % KOH.

Figure 3.3.4: Solubility of glycogen in different solvents. Reproduced from Kerly, 1930 [182]

3.3.3 Including glycogen into the pSIRM-Workflow

Besides the free metabolites another interesting aspect regarding CCM is the glycogen pool. So I tried to find a method that could be integrated in our current metabolomics workflow, which involves phase separation between 50 % MeOH (PP) and chlorofom (lipid phase (LP)).

When glycogen is isolated from biological matrices, the workflows found in the literature usually involve relatively harsh treatment of the matrix. Often it is boiled for >1 h in solutions containing either 30 % potassium hydroxyde or 3 M perchloric acid, and then precipitated with alcohol.

This is not compatible with metabolomics of the CCM, since there are several relatively unstable compounds among those of pivotal interest (e. g. sugar phosphates).

The methods for glycogen extraction normally used in the literature focus only on the extraction of glycogen from a certain sample, and often on its extraction from animal matrices. In the present context, we wanted to extract it from cultured cells, which are likely extractable using much milder conditions.

It would be ideal, if the glycogen could be precipitated quantitatively, hydrolyzed, and then be submitted to analysis like the polar phase of the extract. This way all metabolites of interest could be directly measured for each sample.

The protocols used today mostly rely on work by Kerly [182] published in 1930, who researched solubility properties in different concentrations of ethanol using several additions (see also figure 3.3.4).

Unlike starch (its equivalent in plants), glycogen is soluble in water but very little in alcohol. [182]

According to Kerly after phase separation a part of the glycogen will be in the polar phase (50 % MeOH), i. e. not fully precipitated) and a part will be at the interface during the phase separation step.

From Kerly's work it can be estimated, that precipitation using methanol instead of ethanol will yield lower efficiency. However, if the alcohol concentration is too high, the solubility of other compounds of interest will be affected, and also there will be no phase separation from choloroform.

Precipitation of a compound means essentially to shift its solubility equilibrium. So, the absolute amount of the substance, and the volume of the extraction system have to be taken into account. Since no reliable absolute value on glycogen content in myotubes could be obtained from the literature, the expected amount of glycogen per dish has to be estimated.

According to the literature, the glycogen content in muscle is about 1–2 % of wet weight. [31–35]

A cell pellet of myotubes obtained from a 10 cm culture dish as typically used in the experiments to be conducted, weighs about 15 mg. If the rate of glycogen synthesis is comparable to the rate in vivo, about 150 µg of glycogen per culture dish can be expected.

The data on solubility available in the literature was determined using ethanol as alcohol. So the solubility in the two phase-system methanol$_{(aq)}$ | chloroform was tested for different methanol concentrations. Sample processing was done according to the current method, using 150 µg of glycogen standard per sample, and solvent amounts equivalent to the currently established method. [2]

The glycogen content was determined by photometric assay with anthrone. [42, 114, 115] In the recent literature, this assay is laid out with relatively high sample-volumes. [118] The volume was reduced, so that eppendorf-tubes and 96-well-plates could be used. In this "small volume version" linear range of the assay was found to be considerably larger, than stated by Roelandt *et al.* Except for sample volume and heating time (10 min, not 20 min, as stated by Seifter *et al.* [114]), the protocol of Roelandt was used.

The distribution of glycogen found is presented in figure 3.3.5.

As can be seen, if the current standard sample preparation with 50 % methanol is used, most of the glycogen will be in the polar phase. With 70 % methanol no phase separation could be observed. With 60 % methanol content in the polar phase, the envisaged analysis was feasible.

In order to check, if the higher methanol concentration would influence especially the measurability of sugar phosphates, three dishes of HeLa cells were harvested with 50 % and 60 % methanol each. No differences between the samples were observed (not shown).

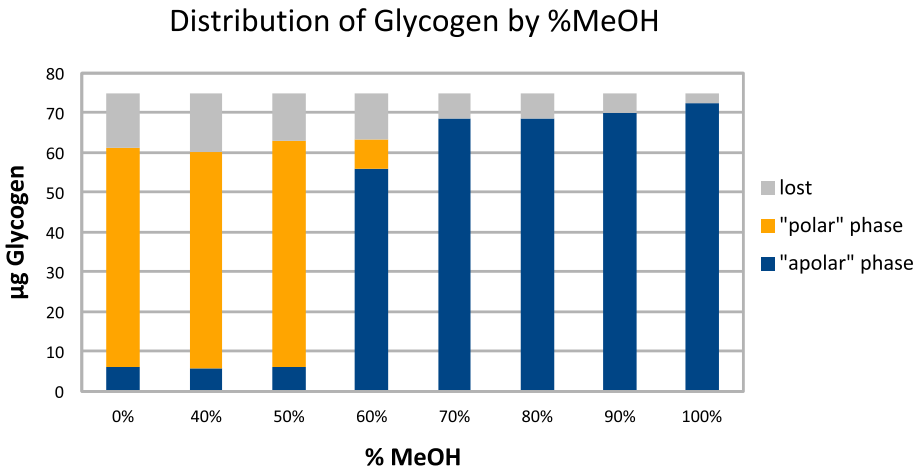


Figure 3.3.5: Distribution of glycogen in variations of standard metabolomics extraction method. Only half of the extract was used, as such, the overall amount sums up to 75 µg.

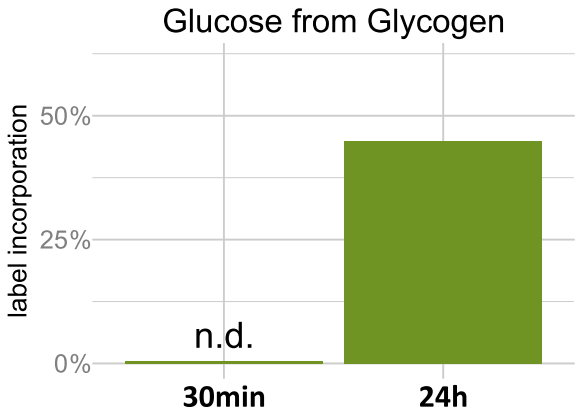


Figure 3.3.6: Label-incorporation in glycogen. After 30 min there is no label to be found, but after 24 h about 50 % of the glucose derived from glycogen is labeled. "Complete" media, with all carbon sources present was used.

As stated above, to test the method, and to see whether labelled glucose is incorporated into glycogen I treated myotubes derived from primary muscle cells of normal human controls with $^{13}C_u$-glucose for 30 min and for 24 h in media with all carbon sources present. Controls from the same individuum were treated the same way, using ^{12}C-glucose. The results are presented in figure 3.3.6.

After 30 min no labeling could be observed; however after 24 h the glycogen was labeled at a level of 50 %. (see fig. 3.3.6) This is interesting, since the myotubes were supplied with glucose and insulin. Thus one would expect no glycogen turnover under these conditions.

3.3.4 pSIRM applied to primary human muscle cells of dysferlinopathy patients

Previous work in the groups of Simone Spuler and Stefan Kempa performed by Verena Schöwel and Sarah Keller [1], showed in mice and samples from patients suffering from dysferlinopathy, that there is not only an impairment in membrane repair [67, 68] —as widely accepted in the field—, but also that dysferlinopathy seems to have a direct impact on glycolysis.

The data showed, that in dysferlinopathy the poolsizes of members of the polyol-pathway [14] —sorbitol and fructose— are increased, while the poolsizes of glucose-6-phosphate (G6P) and further downstream metabolites of G6P are decreased.

Together with Verena Schöwel we performed a pSIRM experiment to further delineate the metabolic changes in patients suffering from dysferlinopathy, using $^{13}C_u$-glucose or $^{13}C_u$-3-hydroxybutyrate as substrates.

We chose material of two patients, one male, one female and a set of 3 age (except one), sex and sampling location matched patients. The characteristics of these patients are given in table 5, a graphical representation of the experimental setup is given in figure 3.3.7.

Table 5: Characterization of individuals who donated their tissue to the pSIRM experiment.

Patient ID	sex	age[1]	location	disease causing mutations	dysferlin expression
DYSF_A	♀	36	*M. Triceps brachii*	c.1448 C>A heterozygous (Ex16) p.S483X c.*107T>A heterozygous	No dysferlin on protein level
DYSF_B	♂	25		c.2516 C>T homozygous (Ex25) p.P839L c.4167+1 G>T homozygous (Ex38)	No dysferlin on protein level
Control_A	♂	22		∅	Normal
Control_B	♂	51		∅	Normal
Control_C	♀	40		∅	Normal

[1] age in years at sampling

3.3.4.1 Poolsizes in myotubes from dysferlinopathy patients show a conserved metabolic phenotype regardless of carbon source

When poolsizes from a pSIRM-experiment are interpreted, it has to be considered, that these poolsizes reflect the sum of labeled and unlabeled intensities. The goal is to track the reaction rate, i. e. the speed at which label is incorporated into newly formed metabolites. Thus the cells are disturbed as little as possible. When exchanging media, usually this

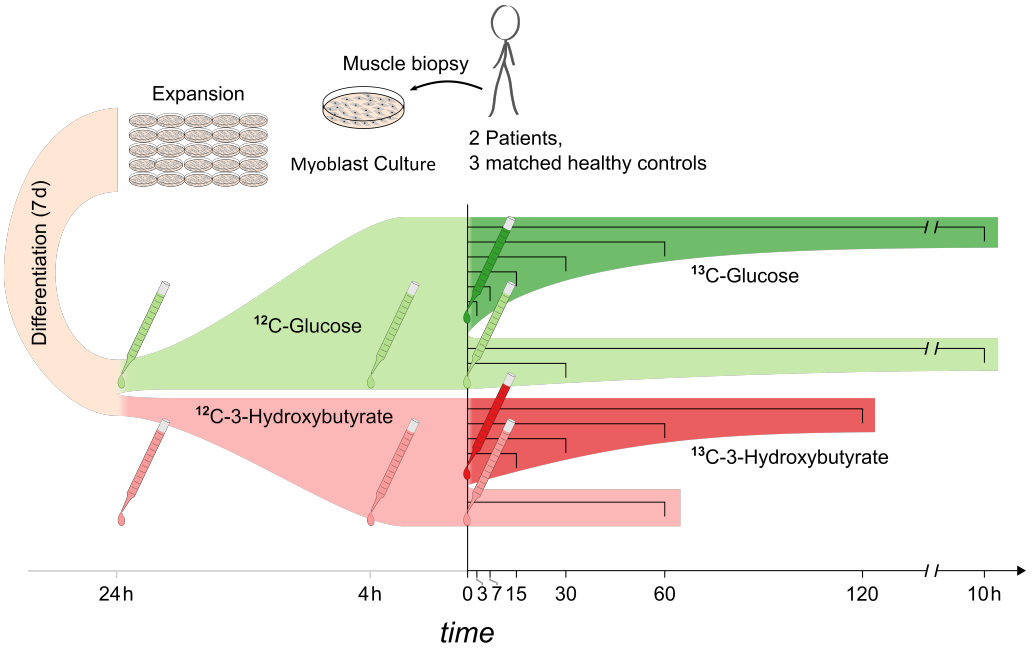


Figure 3.3.7: Scheme of the experimental setup for labeltracing in myotubes from patients that suffer from dysferlinopathy. After expansion, cells were serum starved and differentiated into myotubes for 7 d. 24 h before labeling start, they were accustomed to the carbon source to be present during the labeling time. 4 h before labeling time all cell media were renewed, in order to minimize metabolic perturbation from changes in substrate and product concentration. For controls at "labelling" start, another media exchange with ^{12}C-substrate containing media was performed. For the other experiments, they were labeled with the indicated ^{13}C-substrate at nearly logarithmic intervals. For illustrative purposes, the time axis below the scheme is highly irregular. Time is given in minutes or hours as indicated.

provides the cell with a much higher nutrient concentration and at the same time excreted compounds are removed from the cells surroundings. When working with myotubes, the switch is even more dramatic. Since the cultivated myoblasts need to fuse into myotubes, to give a model system more related to actual muscle, they are kept in differentiation media, that contains less FCS and thus different amounts of growths factors than normal SKMGM. This is why two media exchanges are performed before the begin of labeling: One switch 24 h before start of the labeling from differentiation media to SKMGM, containing the ^{12}C-isotopomers of carbon source(s) present during the labeling, and one switch 4 h before the actual labeling start — to dampen the metabolic pertubation caused by concentration change in substrate and excreted compounds, when labeling of the cells is started (see also figure 3.3.7).

In the current case the metabolite poolsizes changed very little among the early timepoints (3 min to 60 min), whereas some metabolite poolsizes had changed after 10 h (not shown). This is why the poolsizes discussed in the following are the means of the earlier timepoints (3 min to 60 min).

In figure 3.3.8 the fold changes of the metabolite poolsizes versus the controls have been plotted onto the network representation of the CCM, whereas in figure 7.2.1 in the appendix, the fold changes and respective p-values for metabolites measured in myotubes, presented with either glucose or 3-hydroxybutyrate (3OHB) are plotted as a volcano plot.

Overall for the metabolites that were detectable under both conditions applied, relative poolsizes were quite comparable across carbon sources. The major findings of poolsizes of metabolites mentioned above could be reproduced: in the pSIRM experiment, sorbitol and fructose are both upregulated, while G6P was downregulated (only detectable under glucose conditions).

Further analysis also revealed additional findings. The point were the sorbitol-pathway reconnects with glycolysis is at the level of F1,6BP, which is also significantly upregulated.

Several metabolites of the pentose-phosphate-pathway were also increased (6-phosphogluconate, 2-deoxyribose, ribose, ribulose-5-phosphate, xylose). This overall change is more pronounced when 3OHB is used as carbon source.

For the two treatment conditions applied, ornithine (both), urea (in 3OHB) (at $p_{\mathrm{Urea}} = 3.48 \times 10^{-5}$, $\mathrm{fc}_{\mathrm{Urea}} = 1.48$) and fumarate (in glucose) are upregulated ($p_{\mathrm{Ornithine}} = 1.13 \times 10^{-5}$, $p_{\mathrm{Fumarate}} = 4.45 \times 10^{-12}$). Since urea shows a fold change of 1.48, thus it is not included in the significant metabolites. (see figure 7.2.1)

Under glucose conditions the ketone-bodies 2-amino-butyrate $p_{\mathrm{AABA}} = 2.14 \times 10^{-8}$, $\mathrm{fc}_{\mathrm{AABA}} = 2.08$ and 3-hydroxy-butyrate $p_{\mathrm{3OHB}} = 2.79 \times 10^{-3}$, $\mathrm{fc}_{\mathrm{3OHB}} = 1.36$ are also

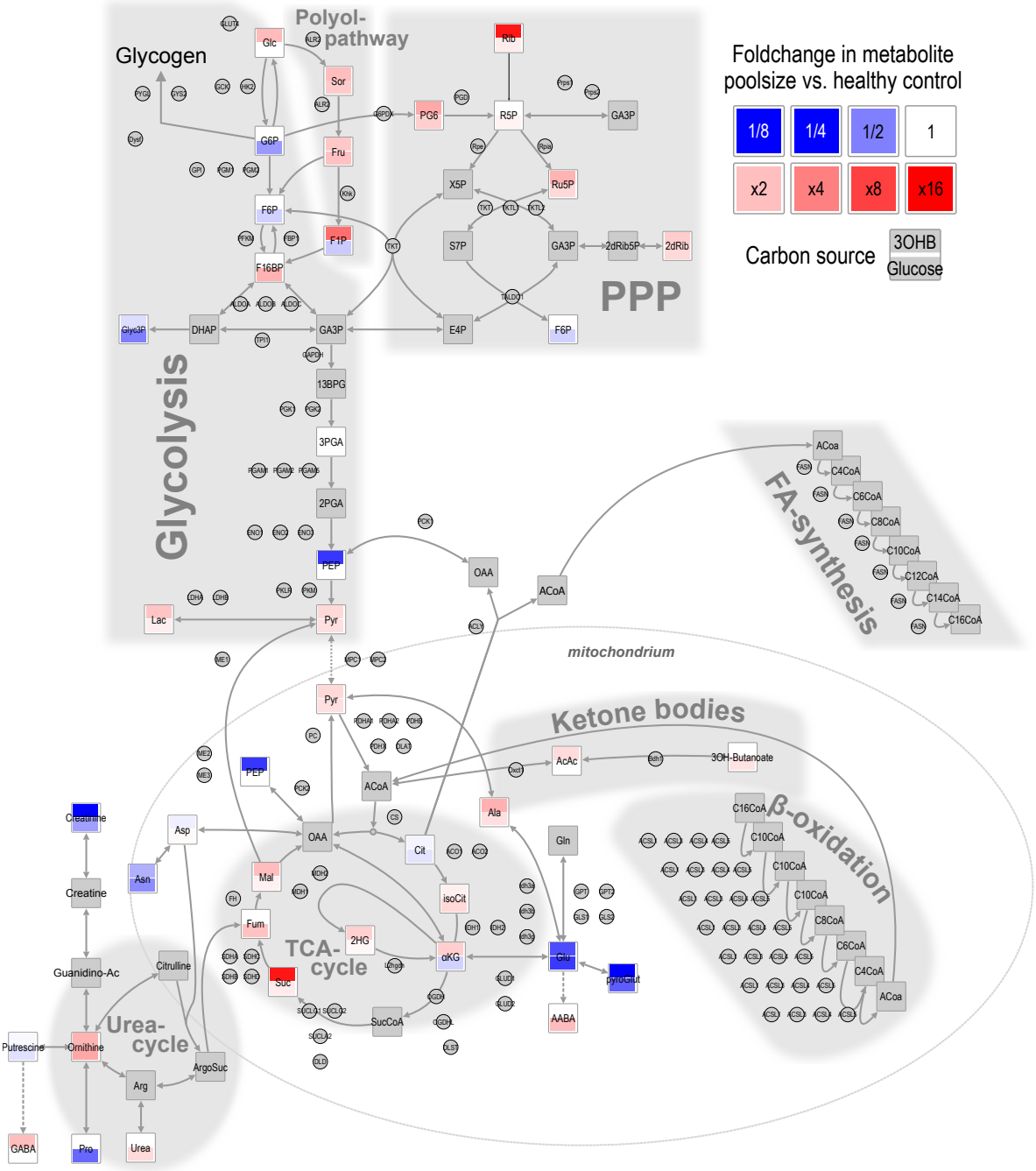


Figure 3.3.8: Network representation of foldchange($^{poolsize}/_{poolsize_{Control}}$) in myotubes, incubated with $^{13}C_u$-glucose or $^{13}C_u$-3-hydroxybutyrate. For a volcano plot of the same data, also including *p*-values, please refer to figure 7.2.1. If a metabolite was changed significantly in both treatment conditions, orientation of the regulation stayed the same. As found in the initial experiments by Sarah Keller [1] sorbitol and fructose as members of the polyol pathway are upregulated, while glucose-1/6-phosphate is downregulated. Additionally members of the pentose phosphate pathway (PPP) (xylose, gluconate-6-phosphate, ribulose-5-phosphate, ribose), and the sugars lactose and maltose (shown in figure 7.2.1) are also upregulated. Whereas, except fumarate and isocitrate, no intermediates of the TCA-cycle were found significantly changed under glucose conditions, α-ketoglutarate, 2-hydroxy-glutarate, succinate and malate were all found to be enriched in myotubes of dysferlinopathy patients when presented with 3-hydroxybutyrate as carbon source. The upregulation of the PPP is overall more pronounced in 3-hydroxybutyrate conditions. Creatinine which is spontaneously formed from creatine was only measured in the control myotubes, while ornithine is upregulated. glutamate (Glu) is strongly downregulated, which might also points to a higher running TCA-cycle. No glutamine was included in the media from 24 h before begin of the labeling experiments. The high demand for Glu could be the cause for the upregulation of the urea-cycle: Glu is replenished by increased proteolysis.

upregulated.

The creatin ⇌ creatin-P system provides muscle with energy during the first moments of exercise. [5]. Creatin and creatin-P are quite unstable substances and spontaneously form creatinine, when treated as during a typical sample treatment protocol for GC–MS based metabolomics.

Creatinine was only detectable in samples of the controls. As such, it can be considered as downregulated in samples from patients with dysferlinopathy. The *p*-value and fold change stem from imputed values.

3.3.4.2 Label incorporation is not changed in sorbitol but in glycogen and glucose-6-phosphate

The poolsizes of metabolites give only an indication regarding enzymatic activity. However, not only the mere amount of a metabolite is of interest, but also the speed at which it is produced or undergoes further metabolism. It depends on the combination of two influences, if an enrichment in poolsize of a certain metabolite is observed: (i) the enzyme producing the metabolite is more active and/or (ii) the enzyme, that uses the metabolite as a substrate is less active than in the control.

In order to discern this, one can look at adjacent metabolites. Enyzmatic reactions only occur if the overall Gibbs free energy is negative ($\Delta G < 0$). Often metabolic reaction cascades can be reversed, but this is usually only the case, if the overall balance of substrates and products is shifted, for example if the overall ATP content of a cell is high (low energy demand), and at the same time the educt concentration for gluconeogenesis (e. g. acetyl-CoA) is high, glycolysis is reversed, glucose is produced and stored for example in the glycogen pool. If we take care, that a pathway does not become oversaturated or that external stimuli, like insulin, do not change considerably during the experiment, we can expect a consistent direction for a certain reaction cascade or pathway.

That means, if an increase of adjacent metabolites is observed, but other metabolites at a fork in the overall reaction cascade is reduced, this is a strong indication, that the enzyme positioned at the bifurcation point is less active.

The information of the label incorporation from pSIRM can help to investigate these complex relationships even further.

If we see faster label incorporation in a certain metabolite than in the control and at the same time a similar or greater poolsize of this metabolite, one can conclude, that the enzyme, producing the metabolite in question is more active in this sample.

The data on label incorporation in the key metabolites sorbitol, glucose-6-phosphate

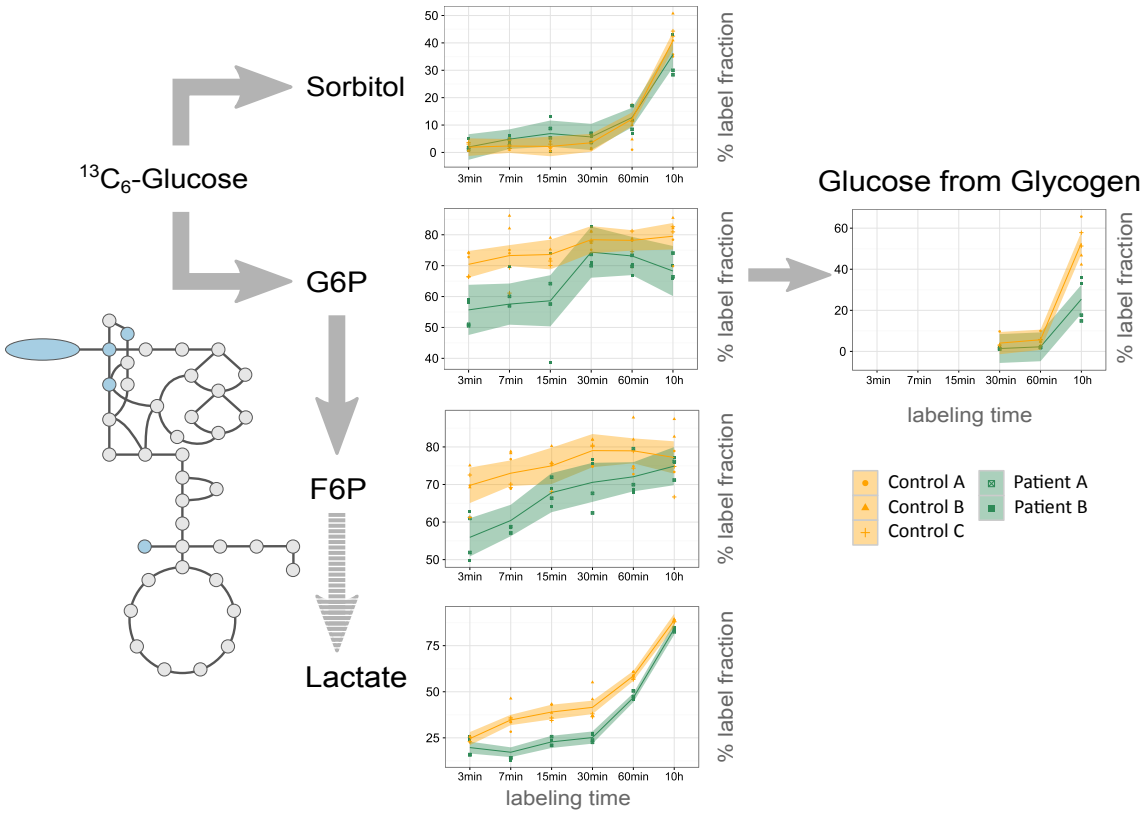


Figure 3.3.9: Label incorporation into glucose-6-phosphate (G6P), fructose-6-phosphate (F6P), sorbitol, lactate and hydrolyzed glycogen in myotubes from patients with dysferlinopathy, compared to normal healthy controls. (Same samples as described above for the poolsizes (figures3.3.8 and 7.2.1 in the appendix). Shaded areas show 95 % confidence interval based on *t*-test.
There is no difference between patients and cotrols in sorbitol, whereas in G6P, label was almost at saturation in controls already at the first timepoint 3 min. Samples from patients reach saturation only after ≈30 min. Also for lactate, further downstream in glycolysis, we see slower label incorporation, being leveled after 10 h.
In glycogen, label incorporation was only observed in the last timepoint (10 h). Here control samples show a higher label incorporation than samples generated from patients.

(G6P), fructose-6-phosphate (F6P) and lactate as well as the glycogen pool will be given in the following.

Label incorporation in samples treated with 3OHB was not detectable at levels to warrant any conclusions (data not shown). The possible reasons for this are discussed in section 4.4 on page 117.

Figure 3.3.9 shows the label incorporation into metabolites measured in myotubes labeled with $^{13}C_u$-glucose for up to 10 h.

There is no difference between patients and cotrols in sorbitol, whereas in G6P, label was almost at saturation in controls already at the first timepoint (3 min). Samples from patients reach saturation only after ≈30 min. Also for lactate, further downstream of

glycolysis, we see slower label incorporation, being leveled after 10 h.
In glycogen, label incorporation at all was only observed in the last timepoint (10 h). Here control samples show a higher label incorporation than samples generated from muscle biopsies of patients with dysferlinopathy.

Overall metabolite levels are relatively low in all samples (patients *and* controls). This might be due to the fact, that no glutamine was contained in the media. This is discussed in section 4.4 on page 117ff.

However, the data shows, that in patients sorbitol and its subsequent metabolites are enriched in poolsize, but at the same time relative label incorporation is as fast as in controls, whereas label incorporation is slower in G6P and subsequently lactate and the glycogen pool. The enzyme in muscle, that is located at the bifurcation point in question is hexokinase II. An influence on hexokinase II activity has never been shown thus far in the literature on dysferlinopathy; dysferlin is mainly seen as a membrane repair protein. However the disease phenotype may not only result from impaired injury repair, but overall lower energy availability. Also the poolsize data of the TCA-cycle intermediates (TCAis) shows a higher capacity for oxidative metabolism, and also the poolsize of the pentose phosphate pathway (PPP) intermediates highlights this. Higher TCA-cycle activity leads to higher levels in reactive oxygen species (ROS). In order to ameliorate this, more GSH reduction equivalents are needed; these are produced from the reactions of the PPP. Since the overall enzyme levels of the CCM are not changed in patients suffering from dysferlinopathy, it can be concluded, that dysferlin must have an influence on hexokinase II activity. This was investigated using a proteomics approach on dysferlin vesicles as laid out in the following section.

3.4 Dysferlin interacts with membranous, mitochondrial and glycolytic proteins

The data obtained from the pulsed stable isotope resolved metabolomics (pSIRM) experiment described above underpins that hexokinase II might be an (intermediate) interactor of dysferlin.

Dysferlin exerts its membrane repair potential by the formation of so called dysferlin vesicles. [67, 68, 183] These represent membrane patches, that are routed to the location of injury, where they fuse with the cell membrane [67, 68].

There have been some studies conducted on the overall makeup of dysferlin vesicles or dysferlin interactors [183–193], none of which showed interaction with glycolytic

enzymes.

Different clean up strategies have been used, some using milder, some harsher detergents to extract membranous proteins.

In the framework of her PhD-thesis Séverine Kunz [3] developed a procedure, that allows to separate intact dysferlin-vesicles from other organelles of the cell, especially the outer membrane. She could show this by electron microscopy and and also showed, that a dominant marker of cell membrane, sodium-potassium ATPase was absent in the dysferlin vesicle containing fraction. She purified dysferlin vesicles from samples of healthy controls in order to find proteins, that are integrated in these vesicles, but not necessarily direct interactors of dysferlin. This secondary interaction might hold some insights as to how dysferlin exerts other functions, also inside the cell, additionally to its well recognized function in membrane repair. [67, 68, 180] Since patients with dysferlinopathy show a changed metabolic phenotype, especially proteins involved in those processes are of interest.

In order to find the proteins that directly interact with dysferlin, Séverine Kunz also performed a standard immuno-precipitation (IP) of the whole-cell lysate, giving the potential to several potential false positives, serving as a relatively strict control.

The anti-dysferlin antibody rAB (JAI-1-49-3, Abcam) was coupled to protein A-coated magnetic beads, that facilitate the pulldown. Also, to filter out unspecific interactors even more stringently, both protocols (vesicle-fractionation and classic IP) were performed using rabbit IgG coupled to the beads. Every cleanup protocol was performed on 3 patients in 3 technical replicates each. ($n_{\text{samples per group}} = 9$)

After proteomics data was obtained, I analyzed the data.

3.4.1 Description of the dataset

In total 1182 proteins were found. After proteins, not measured in at least ⅔ of the samples of any group, or identified as reverse hits, I manually removed proteins marked as contaminants. In removal of contaminants I took care not to remove them unsupervised, since, for example, several structural proteins that are included in the standard list of contaminants in MaxQuant are of special interest in this experiment. I manually inspected the candidates of contaminants and removed only those, that showed the behaviour of a contanimant (only present in few samples with high standard deviation inside sample groups).

After filtering for the described criteria, I retained 752 proteins.

The proteins retained are shown as heat map in figure 3.4.1.

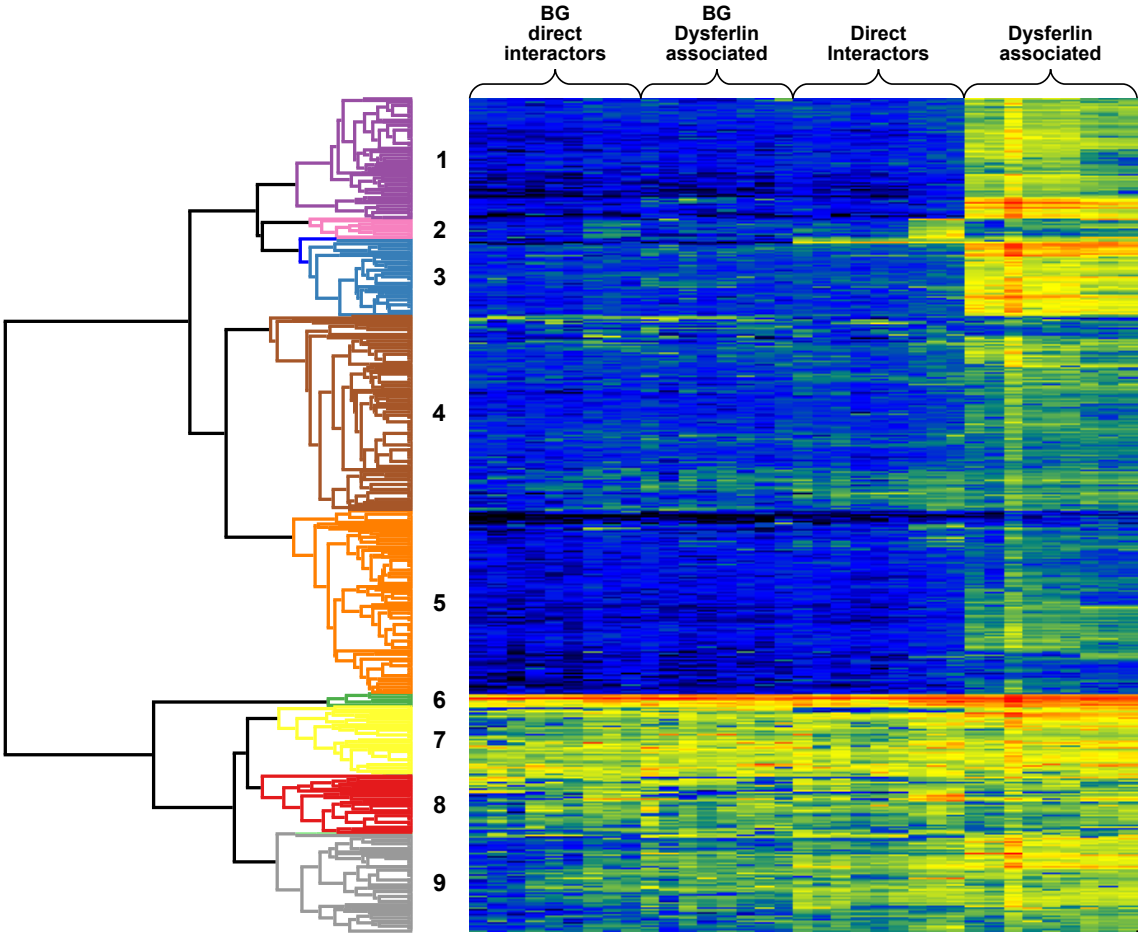


Figure 3.4.1: Heatmap of the samples including imputed values. The first two blocks represent background samples: the magnetic beads were coupled with IgG and show unspecific interactors. In the third block "direct interactors" samples from a pulldown after lysis with SDS buffer potentially including proteins of the whole cell lysate. The relatively harsh conditions applied lyse membranes; only direct interactors are pulled down. Since whole cell lysates are used, many proteins that are interacting at the cell membrane only might be pulled down. On the right side samples of the vesicular fraction are shown. In general those proteins in clusters 1 through 5 are of interest. They are much more enriched when dysferlin is used as bait protein compaired to IgG. Dysferlin itself is located in cluster 2, that contains proteins most highly enriched in both the total lysate and the vesicular fraction; these are the most likely candidates of direct interactors of dysferlin.

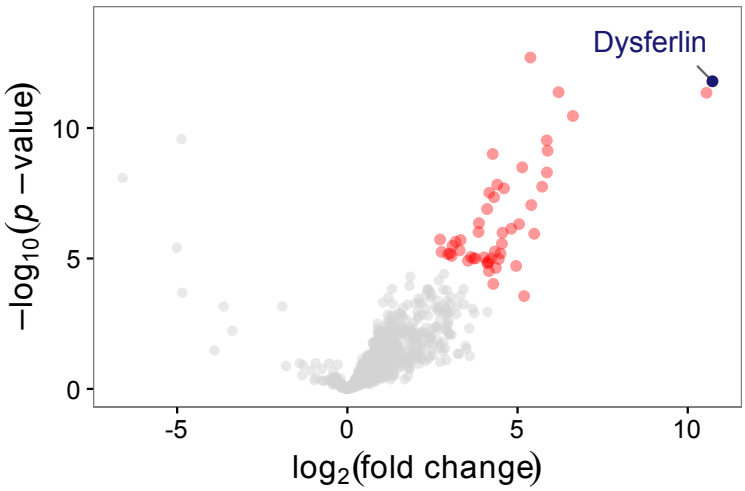


Figure 3.4.2: Volcanoplot of pulldowns with anti-dysferlin rAB vs. rabbit IgG. The bait protein, dysferlin is marked. The proteins in red separate themselves nicely from the rest and are thus deemed strong interaction partners of dysferlin. There are also some proteins stronger interacting with IgG than dysferlin. (left side of the plot) Those are not of interest for the analysis. Table 6 gives p-values and fold changes for the "strong" interaction partners of dysferlin.

I imputed missing data of the log-scaled values, using a Gaussian distribution with 30 % of standard deviation and a downshift of 1.8 of the mean of the whole data set.

3.4.2 Dysferlin directly interacts with proteins involved in RNA-processing

To determine direct interactors of dysferlin, I computed fold change of the measurements of direct interactors vs. their background as well as FDR-adjusted p-values of the fold changes. The log-transformed values were plotted against each other, resulting in the volcanoplot shown in figure 3.4.2.

This showed a particular cloud of proteins, separating themselves from the continuum of further points in the plot. Usually interaction is defined at a more or less arbitrary cutoff combination of fold change and p-value. Based on the volcano plot it seems reasonsable, that members of this cloud are the most important direct interactors of dysferlin. These proteins and their enrichment statistics can be found in table 6.

Of course it is still the case, that those proteins appearing more closely to the origin represent potentially valid interaction partners of dysferlin, however these interactions appear to be much weaker. Irrespective, if this cloud, or a much more lenient cutoff at $s_0 = 2$, FDR = 0.05 is chosen, the general themes found by GO-enrichment analysis using GOrilla [173] and REViGO [174] show a strong enrichment for terms revolving around RNA-processing. The respective treemaps are included in the appendix as figure 7.3.1.

These proteins include many known interactors recently reported, which shows in

general good agreement with the current literature. [183, 184] A comparison is given in table 7.

3.4.3 Dysferlin vesicles are transport organelles within the cell

When comparing the vesicular fraction to the background, RNA-processes are not present, as they were dominant in the whole cell lysate. Vesicle mediated transport is the most dominant theme, either, when a relatively lenient ($s_0 = 2$, FDR = 0.05) or the lowest fold change, and the highest p-value from the separate group of proteins found in the total lysate (figure 3.4.2) was used. Using the latter, more strict cutoff, leads expectedly to much less enriched GO-terms, but the overall theme of transport stays nevertheless dominant. The respective treemaps are shown in figure 7.3.2 in the appendix.

3.4.4 Glycolytic proteins are enriched in the vesicular fraction

To find proteins particularly enriched in the vesicular fraction, I averaged the background (IgG) values per patient and enrichment strategy ($n = 3$) and subtracted them from the respective sample values. If a value from a pulldown-sample was smaller than that of the corresponding average background, it was set as 'NA' due to the logarithmic properties of the data. From the values that remained in the pulldown-samples, I removed those, that were introduced by imputation, but still larger than their corresponding background value.

All replicates from the samples, which are not only replicate measurements of the same extract, but included every step including cell culture, carried out on different days, were then used to find proteins significantly changed ($s_0 = 2$, FDR = 0.05) between vesicular fraction and direct pulldown, without using imputed values. This approach puts emphasis on proteins reliably measured, by reducing the sample size used to calculate p-values, yielding high (less significant) values for proteins quantified in fewer samples. It might however also lead to discarding proteins, that are simply not found in one sample group, due to big enrichment differences between vesicular fraction and total cell lysate.

Perseus [148] was used to find enriched GO-terms. Only terms from the sub-ontology cellular component (GOCC) were found to be enriched.

The concepts found to be enriched are listed in table 8. The adjusted p-values given in the table, were not calculated with Perseus, since the algorithm of the Fisher-exact-test is faulty. Instead R was used.

The general themes to be found in the GO-terms are "membrane" and mitochondrium. Since these terms overlap for many proteins —members of "mitochondrial membrane" are

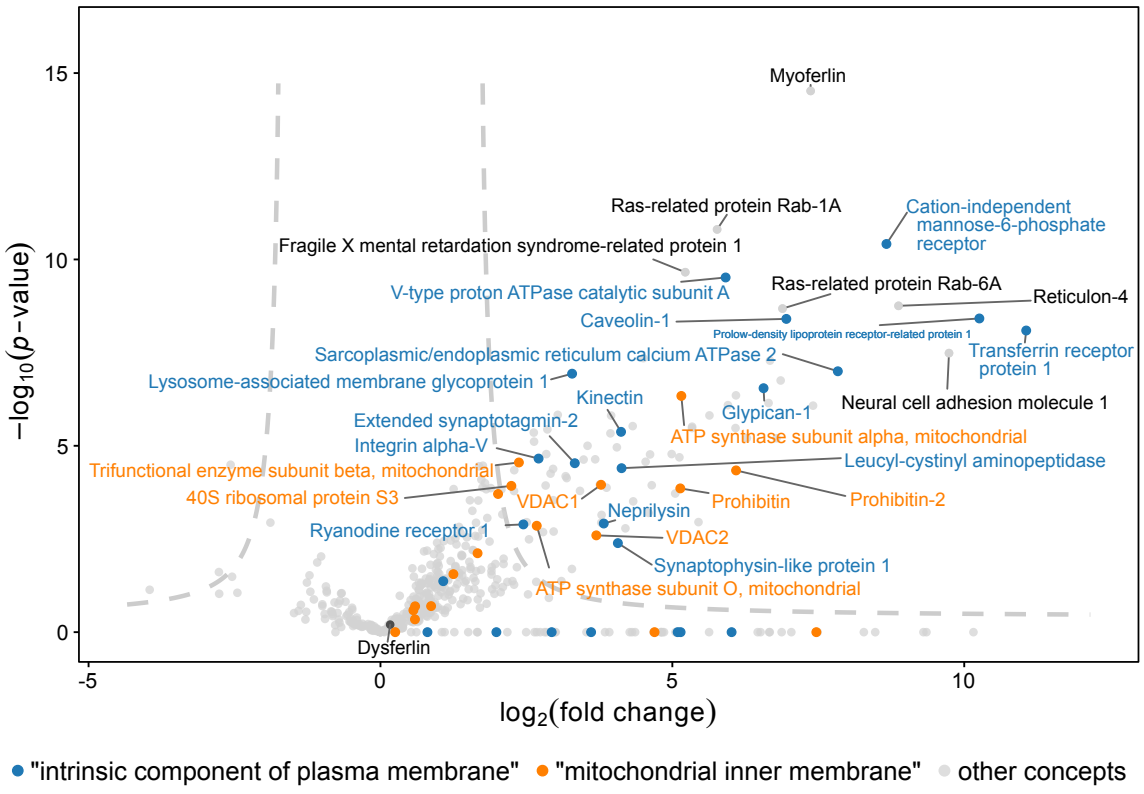


Figure 3.4.3: Volcanoplot of protein amounts in vescicular fraction vs. direct IP. The dashed lines depict a cutoff at fc $= 2$, $p < 0.05$. Mitochondrium associated proteins are shown in orange, membranous proteins in blue. The amount of dysferlin as the bait protein is not expected to change between the two preparation strategies used.

also in the term “plasma membrane”— for visualization purposes, I chose to highlight members of the term “mitochondrial inner membrane” as “mitochondrial” and proteins not in this category but still in “intrinsic component of plasma membrane” as “membrane”.

The proteins found in different amounts in the vesicular fraction and as direct interactors (after respective background correction) are plotted in figure 3.4.3. Many membranous proteins are enriched (caveolin 1, mannose-6-phosphate-receptor and others), but also proteins usually only found in mitochondria (VDAC1, VDAC2, ATP-synthase). Also proteins implicated in intracellular membrane trafficking (Ras-related protein Rab-1A, Rab-6A, Reticulon-4) are enriched in the vesicular fraction.

Also several proteins involved in glycolysis are found to be enriched higher in the vesicular fraction than in the total lysate: Aldolase A (ALDOA), lactate dehydrogenase A&B (LDHA, LDHB), glyceraldehyde-3-phosphate dehydrogenase (GAPDH), phosphoglycerate mutase (PGAM2 or 1 or 4) and pyruvate kinase (PKM). (see figure 3.4.4) Some are similarly enriched in both vesicular and total lysate fractions, like LDHA, but most are only significantly enriched in the vesicular fraction, and reach significant enrichment

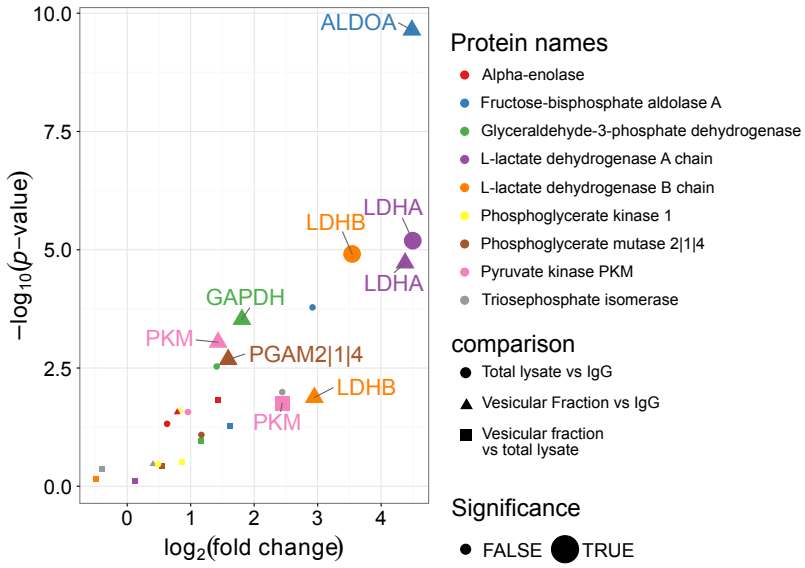


Figure 3.4.4: Significance values for glycolytic enzymes detected in samples enriched by different extraction protocols. Compared to their respective controls, enrichment statistics are much more significant in the vesicular fraction, and PKM is significantly higher enriched in direct comparison between proteins in vesicular fraction and total lysate.

levels. PKM shows the biggest difference between the glycolytic enymes, being significantly higher enriched in the direct comparison of vesicular fraction and total lysate pulldowns. (see figure 3.4.4. Although hexokinase II is not among them, several of them have also been found as potential interactors of dysferlin. Most peculiar is the presence of 6-phosphofructokinase (PFKM) in the work by de Morrée, *et al.*, 2010 [183] and Flix, *et al.*, 2013 [184], since PFKM is the key-regulator of glycolysis. Although in the work of de Morrée, *et al.*, 2010 the TCA-cycle is a representative concept in their cluster analysis, *metabolic implications of dysferlin where not discussed in the literature so far*.

Very recently, Li, *et al.*, 2016 [190] reported an implication of dysferlin in lipid droplets in the heart. To my knowledge this is the first implication of dysferlin in metabolic processes in the literature.

Table 6: Proteins found to be direct interaction partners of dysferlin.

Protein name	Gene name	$-\log_{10}(p)$	$\log_2$(fold change)
Dysferlin	DYSF	11.79	10.72
Ryanodine receptor 1	RYR1	11.35	10.55
Hypoxia up-regulated protein 1	HYOU1	10.46	6.63
Midkine	MDK	11.37	6.21
40S ribosomal protein S6	RPS6	9.13	5.88
60S ribosomal protein L4	RPL4	8.30	5.86
Multifunctional protein ADE2	PAICS	9.53	5.86
40S ribosomal protein S8	RPS8	7.75	5.72
40S ribosomal protein S2	RPS2	5.95	5.49
40S ribosomal protein S3a	RPS3A	7.05	5.40
60S ribosomal protein L29	RPL29	12.70	5.38
Focal adhesion kinase 1	PTK2	3.56	5.19
Nucleophosmin	NPM1	8.49	5.14
60S ribosomal protein L6	RPL6	6.32	5.04
T-complex protein 1 subunit beta	CCT2	4.71	4.96
60S ribosomal protein L8	RPL8	6.14	4.82
60S ribosomal protein L10 or L10-like	RPL10, RPL10L	7.69	4.60
60S ribosomal protein L7a	RPL7A	5.99	4.56
60S ribosomal protein L15	RPL15	5.57	4.54
L-lactate dehydrogenase A chain	LDHA	5.19	4.50
Adenylosuccinate lyase	ADSL	4.98	4.45
60S ribosomal protein L3	RPL3	7.83	4.40
40S ribosomal protein S9	RPS9	4.64	4.36
T-complex protein 1 subunit epsilon	CCT5	5.27	4.33
Serpin H1	SERPINH1	7.35	4.31
Myoferlin	MYOF	4.02	4.29
Calreticulin	CALR	9.01	4.27
T-complex protein 1 subunit gamma	CCT3	5.00	4.23
60S ribosomal protein L13	RPL13	7.52	4.17
40S ribosomal protein S4, X isoform	RPS4X	4.52	4.16
40S ribosomal protein S13	RPS13	4.90	4.15
60S ribosomal protein L14	RPL14	4.82	4.13
40S ribosomal protein S16	RPS16	4.84	4.11
Protein disulfide-isomerase A4	PDIA4	6.89	4.11
60S ribosomal protein L19	RPL19	5.04	4.02
60S ribosomal protein L23a	RPL23A	6.36	3.87
BAG family molecular chaperone regulator 2	BAG2	6.02	3.85
T-complex protein 1 subunit zeta	CCT6A	5.01	3.77
Filamin-C	FLNC	5.00	3.71
T-complex protein 1 subunit alpha	TCP1	5.06	3.63
L-lactate dehydrogenase B chain	LDHB	4.91	3.54
Endoplasmin	HSP90B1	5.71	3.33
Probable ATP-dependent RNA helicase DDX17 or 5	DDX17, DDX5	5.31	3.30
Polyadenylate-binding protein 1 or 3	PABPC1, PABPC3	5.64	3.18
78 kDa glucose-regulated protein	HSPA5	5.48	3.08
Clathrin heavy chain 1	CLTC	5.10	3.06
Ribose-phosphate pyrophosphokinase 1, 3 or 2	PRPS1, PRPS1L1, PRPS2	5.19	3.02
BRISC complex subunit Abro1	FAM175B	5.17	2.96
Protein RCC2	RCC2	5.25	2.77
Fragile X mental retardation syndrome-related protein 1	FXR1	5.73	2.72

Table 7: Proteins quantified in the current study, that are also mentioned by Flix, *et al.*, 2013 or de Morrée, *et al.*, 2010

Accession[1]	Name in Flix, *et al.*, 2013	measured	vesicle, strict[4]	lysate, strict[4]	vesicular, lenient[5]	lysate, lenient[5]	ves. vs lys., lenient[5]
P68133	Actin;alpha skeletal muscle	+	+		+		+
P30566	Adenylosuccinate lyase	+	+	+	+	+	
P61163	Alpha-centractin	+	+		+		+
O14958	Calsequestrin-2	+			+		+
P12277	Creatine kinase B-type	+			+		+
O75923	Desmin	+	+	+	+	+	
O75923	Dysferlin	+	+	+	+	+	
P21333	Filamin-A	+			+		
Q14315	Filamin-C	+	+	+	+	+	
P04075	Fructose-bisphosphate aldolase A	+	+	+	+	+	
P04406	Glyceraldehyde-3-phosphate dehydrogenase	+			+		
Q14974	Importin subunit beta-1	+			+	+	
P11055	Myosin-3	+	+		+		+
P35579	Myosin-9	+			+		+
Q09666	Neuroblast differentiation-associated protein AHNAK	+					
P14618	Pyruvate kinase isozymes M1/M2	+			+		+
P00558	Phosphoglycerate kinase 1	+					
Q6NZI2	Polymerase I and transcript release factor (PTRF)	+	+		+		+
Q13885	Protein S100-A1	+					
P68371	Ryanodine receptor 1	+					+
P08670	Stress-70 protein;mitochondrial	+			+		
Q09666	Tubulin beta chain	+					
O14958	Vimentin	+			+		+
P21817	Vinculin isoform 1	+	+	+	+	+	+
Q6ZMU5	Trim72/MG53[2]	+	+		+		+
Q03135	Caveolin 1[2]	+	+		+		+
P07355	Anexin A2[3]	+					
P47756	F-actin-capping protein subunit beta[3]	+	+		+	+	+
P55884	Eukaryotic translation initiation factor 3 subunit B[3]	+	+		+	+	+
P14625	Endoplasmin[3]	+	+	+	+	+	
P11021	78 kDa glucose-regulated protein[3]	+	+	+	+	+	+
P02545	Prelamin-A/C[3]	+			+		
Q9NZM1	Myoferlin[3]	+	+	+	+	+	+

[1] UniProtID from Flix, *et al.*, 2013.

[2] accession not explicitly stated, only mentioned in the text in Flix, *et al.*, 2013 and matched to UniProtID by gene name

[3] taken directly from de Morrée, *et al.*, 2010

[4] cutoff refers here to the proteins marked in red in figure 3.4.2, the smallest fold change and biggest p-value from these are taken and used as cutoff

[5] as more lenient cutoff $s_0 = 2$ and $p_{\text{FDR}} < 0.05$ have been chosen

Table 8: GO-terms found to be enriched (cutoff: fc = 2, $p < 0.05$) in proteins enriched in vesicular fraction vs. direct IP, ordered by significance. Two terms, used to demonstrate interaction with membranous and mitochondrial proteins are highlighted. see also figure 3.4.3.

Term ID	Description	p_{adjust}
GO:0044446	intracellular organelle part	1.80×10^{-4}
GO:0044422	organelle part	2.30×10^{-4}
GO:0044444	cytoplasmic part	1.45×10^{-3}
GO:0043231	intracellular membrane-bounded organelle	3.11×10^{-3}
GO:0031090	organelle membrane	4.89×10^{-3}
GO:0043229	intracellular organelle	4.99×10^{-3}
GO:0042645	mitochondrial nucleoid	5.35×10^{-3}
GO:0009295	nucleoid	5.35×10^{-3}
GO:0044459	plasma membrane part	6.45×10^{-3}
GO:0044424	intracellular part	7.63×10^{-3}
GO:0042470	melanosome	7.63×10^{-3}
GO:0048770	pigment granule	7.63×10^{-3}
GO:0005925	focal adhesion	1.56×10^{-2}
GO:0005887	integral component of plasma membrane	2.14×10^{-2}
GO:0016020	membrane	2.14×10^{-2}
GO:0005886	plasma membrane	2.14×10^{-2}
GO:0005743	**mitochondrial inner membrane**	2.42×10^{-2}
GO:0031226	**intrinsic component of plasma membrane**	3.16×10^{-2}
GO:0045259	proton-transporting ATP synthase complex	4.77×10^{-2}
GO:0016529	sarcoplasmic reticulum	5.07×10^{-2}

3.5 Metabolomic profiles from single drops of capillary blood

The experiment described in the following was devised and carried out together with Henning Kuich [4] to fully equal parts in close collaboration. The findings presented here were all derived from intense discussion among each other and our advisor Stefan Kempa.

When we profiled blood samples of a single volunteer performing exercise, we were able to cover a major portion of the central carbon metabolism, comprised of glycoloysis, tricarboxylic acid (TCA) cycle, fatty acid metabolism, ketone body metabolism, pentose phosphate pathway, and amino acid metabolism.

Figure 3.5.1 illustrates the four crucial elements of our methodology:

i The possibility of a high sampling frequency,

ii the great number of identified metabolites quantitatively measured by GC–MS,

iii the inclusion of erythrocytes, and

iv the extensive identification of metabolites of the central carbon metabolism (CCM) that lie at the junction of the interconnected pathways crucial to energy production and homeostatic maintenance.

A total of 277 metabolic features were detected in eight sampling time points of which 169 were quantified and 93 of which were identified (Figure 3.5.1), counting different derivatization products of the same compound as one compound. We calculated the coefficient of variation (COV) of the relative quantities of the metabolites and found that 85 % of the quantified metabolites displayed a COV between 0 and 0.1 (figure 3.5.2). The detection and quantification of the metabolites was found in a linear range of detection to a large extend (not shown).

3.5.1 Subjective feeling throughout exercise

At each sampling time point, the volunteer reported his perception of energy availability on a qualitative scale (Figure 3.5.4A), similar to the rate of perceived exertion (RPE) [195]. Progressive tiring in the first two laps culminated in “Hitting the Wall” in lap 3 (L3, crisis), which was the slowest lap despite being the shortest (see figure 3.5.3). L4 was the fastest lap characterized by a sensation of renewed energy and motivation (“runner’s high”). Subsequently, energy availability appeared stationary and comparable to that experienced during exercise initiation, although muscle fatigue became apparent, late in L5 and in L6. After a rest, the subject felt recovered despite muscle fatigue.

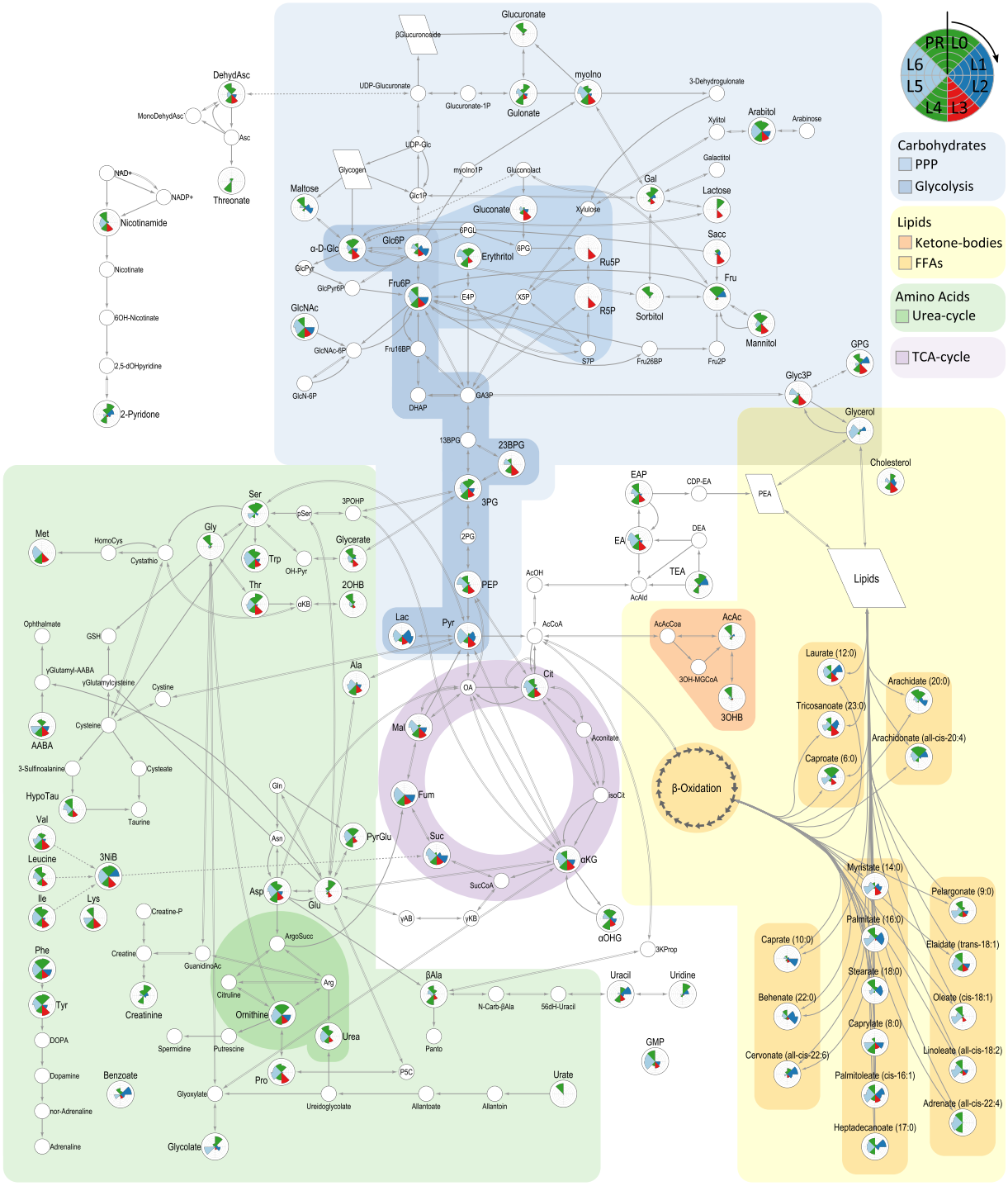


Figure 3.5.1: Network of all metabolites identified, shading depicts the metabolic pathways considered. Nightingale plots [194] for each metabolite depict poolsizes at each timepoint.

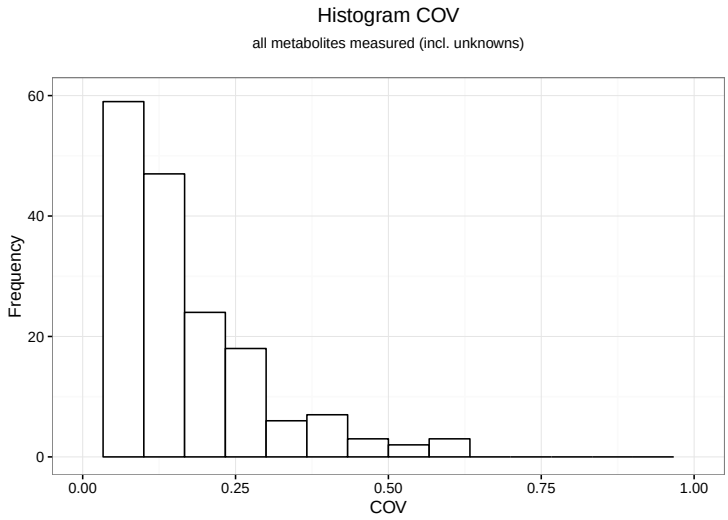
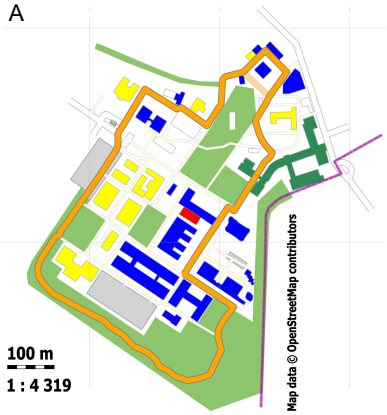


Figure 3.5.2: Covariances for all metabolites measured, including unknowns. The majority (85 %) shows a covariance between 0 and 0.1.

A



B

Lap	*t* 1/min	Distance 1/km
L1	10	2.2
L2	10	2.2
L3	13	2.0
L4	8	2.2
L5	10	2.2
L6	10	2.2

Figure 3.5.3: **A)** Running track (orange) on the campus of the Max-Delbrück-Center for Molecular Medicine, Berlin-Buch. **B)** overview of running time and and distance. Lap 3 despite being shortened in distance slightly (light orange part) took the longest time to complete.

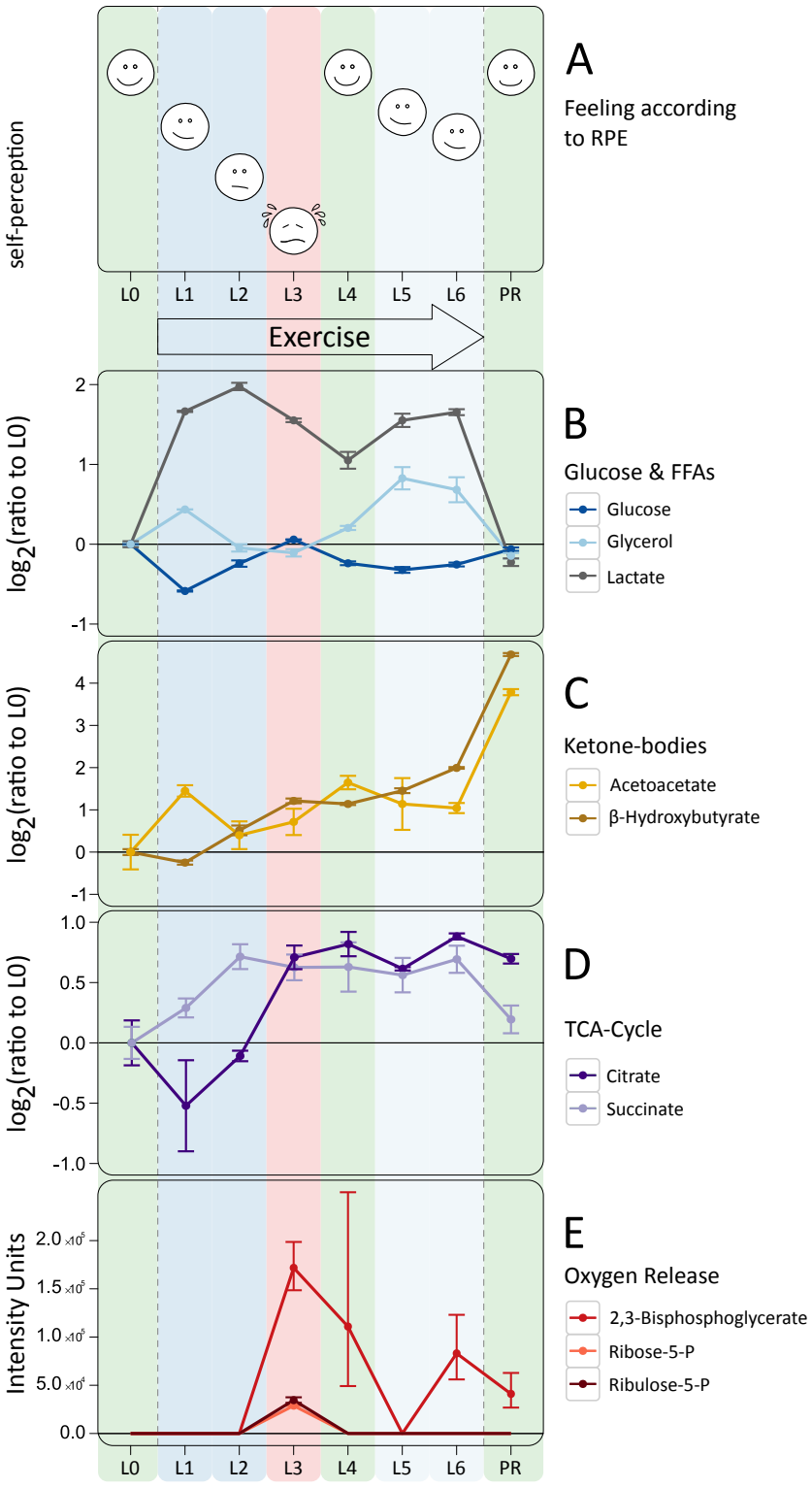


Figure 3.5.4: Plots of the metabolites used to assess credibility of the experiment. log_2-foldchange relative to the initial sample, errorbars show variability of two technical replicates.
A) feeling of the subject throughout the run according to RPE-scale [195], **B)** glucose, free fatty acids (represented by glycerol) and lactate, **C)** ketone bodies showing a drastic increase post run, **D)** TCAis represented by succinate (see also figure 3.5.6): citrate increases much later than the other TCAis, **E)** metabolites involved in oxygen transfer to the tissue

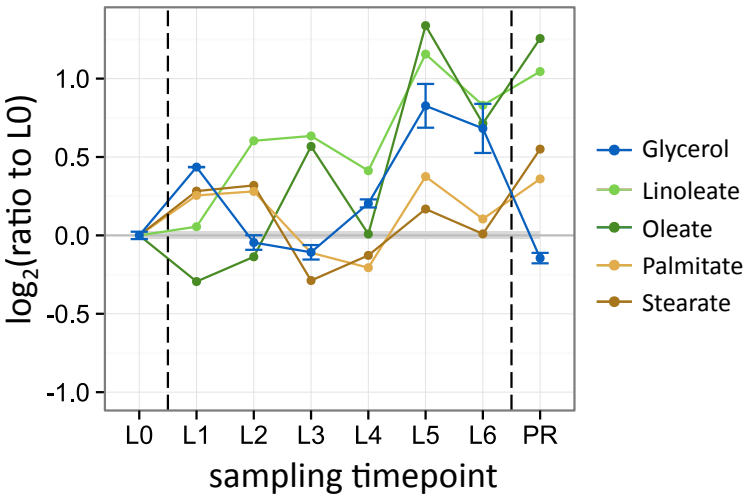


Figure 3.5.5: Plots of FFAs and glycerol, log_2-foldchange relative to the initial sample, errorbars show variability of two technical replicates, FFAs were measured only once. The major fatty acids in the human body [199] are shown. For the saturated fatty acids palmitate ($C_{16:0}$) and stearate ($C_{18:0}$) two release phases are observed, one before the "crisis" in L3, one after, correspoding to glycerol levels (blue). For the unsaturated fatty acids oleate ($C_{18:1}$) and linoleate ($C_{18:2}$) a more or less steady increase is observed with a higher release in the time after the crisis.

3.5.2 Fuel source availability throughout exercise

Exercise is primarily an energetic challenge. Our personalized approach inherently allows for only a low subject number, so we sought for a way to confirm our findings, and if our data from a single subject allows for interpretation at all. Changes in fuel source availability and their metabolic end products have been investigated before [196] and were used to assess the validity of our platform.

The initial phase of exercise appeared to be fueled primarily by glucose and its conversion to lactate (figure 3.5.4B). Glucose recovered to its original level by L3 ("crisis") while lactate levels declined, indicating the mobilization of glycogen stores from liver and an active Cori cycle [197]. After the crisis, glucose levels stabilized while lactate rose in L5 and L6, before being cleared to pre-exercise levels. The stable blood glucose levels concurrent with lactate production in laps 4 through 6 indicated a second stable phase of exercise. Glycerol, an indicator for fatty acid mobilization [198], and two of the most abundant FFAs, palmitate and stearate [199], similarly suggested two distinct phases during exercise. In response to exercise initiation, a first wave of glycerol and FFAs were released (figures 3.5.4B and 3.5.5). Their abundance was lowest during the crisis and subsequent high feeling, whereafter a second release was observed in L5 and L6.

Ketone bodies, which are commonly considered to serve as fuel only during starvation and to be pathological in high concentrations, increased in response to exercise initiation

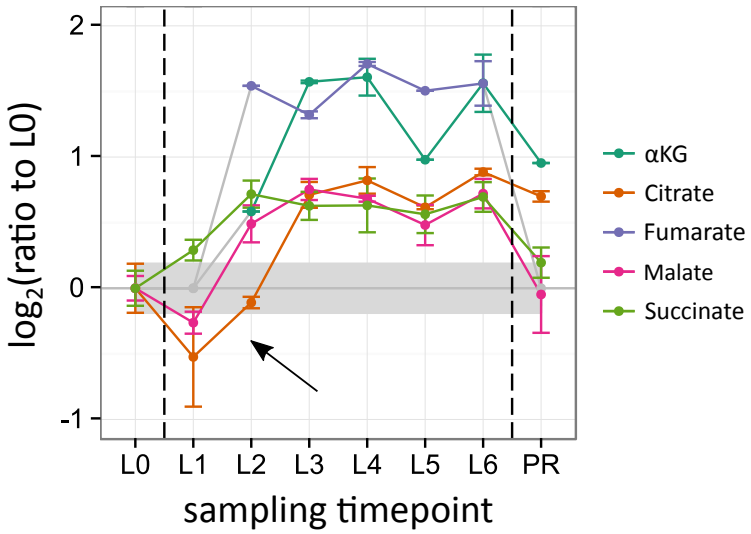


Figure 3.5.6: Plots of TCAis, log_2-foldchange relative to the initial sample, errorbars show variability of two technical replicates. All TCAis rise well above their initial level, except for citrate (see arrow).

but appeared more clearly and stably elevated in the second phase of exercise, before accumulating vastly after exercise completion (Figure 3.5.4).

3.5.3 Metabolic insights when "Hitting the Wall"

"Hitting the wall" is thought to occur when glycogen reserves are depleted and the energy requirement of the ongoing exercise can no longer be met [30, 200]. Our initial motivation was to determine the mechanistic basis for this presumed energy deficit previously experienced by our subject; therefore we sought to provoke its occurrence by exercising in a fasted state (in contrast to the carbohydrate loading strategy commonly employed by endurance runners).

All detectable TCA-cycle intermediates (TCAis) most likely reflect muscle tissue, since during exercise it represents the by far largest and most active tissue regarding energy turnover. Also, during endurance running about 50 % of the energy is generated from fat, being oxidized in the TCA. By L2 all of the detectable TCA intermediates except for citrate at the entrypoint of the TCA had risen well above baseline levels, indicating that its capacity for activity had been primed prior to the crisis in L3 (Figure 3.5.4D, Figure 3.5.6), but not a running TCA-cycle as citrate levels had comparatively not yet increased. Concordantly, the erythrocyte-specific metabolite 2,3–bisphosphoglycerate (2,3–BPG) [13] appeared and was at its highest level during the crisis (Figure 3.5.4). 2,3–BPG specifically binds to hemoglobin and lowers its oxygen affinity, causing oxygen

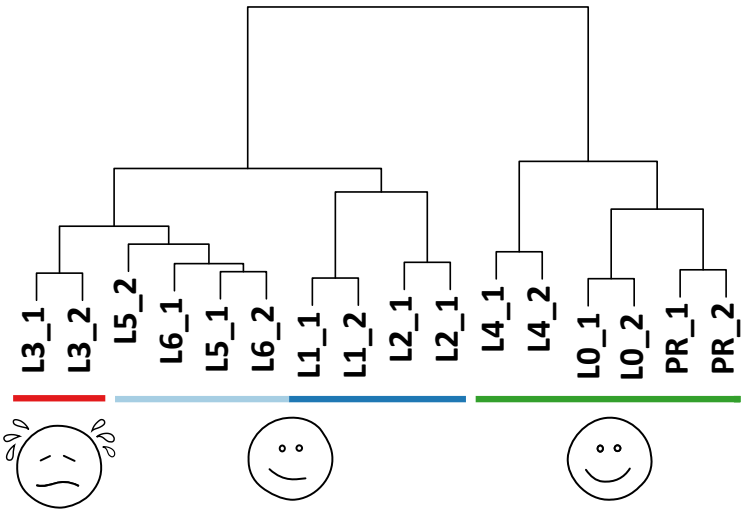


Figure 3.5.7: Unsupervised hierarchical clustering of all features measured. The samples show strongest kinship according to RPE.).

release in demanding tissue. [12] Simultaneously, the pentose phosphate pathway (PPP) intermediates ribose–5–phosphate (R5P) and ribulose–5–phosphate (Ru5P) were detected exclusively during the crisis (L3) (figure 3.5.4E).

3.5.4 Metabolic states are reflected in subjective perceptions

To investigate the global metabolic relationships of sampling time points, all polar metabolites were used for hierarchical clustering analysis (figure 3.5.7). This basic systems biology technique allows for interpretation of the data without being influenced from existing knowledge of metabolic networks and might lead to a gain of metabolic insights towards specific phenotypes in patients or other study subjects. The technical replicates of pre-exercise, post-exercise, and the high feeling were most related. The second major branch was composed of groupings of L1 and L2, as well as L5 and L6, respectively. Furthest from the cluster of sampling timepoints when the subject felt most “energetic” was the crisis (L3).

Astonishingly, metabolic states and their transitions as defined by the clustering of the metabolome of a single drop of blood appeared to be reflected in the volunteer’s subjective feeling throughout the exercise, and not, as one might expect for instance, the time course of the samples taken.

We then attempted to confirm the occurrence of these distinct metabolic states and transitions as suggested by the combination of sample clustering (representing biochemical measurements) and energy availability (representing subjective perception of exertion) by

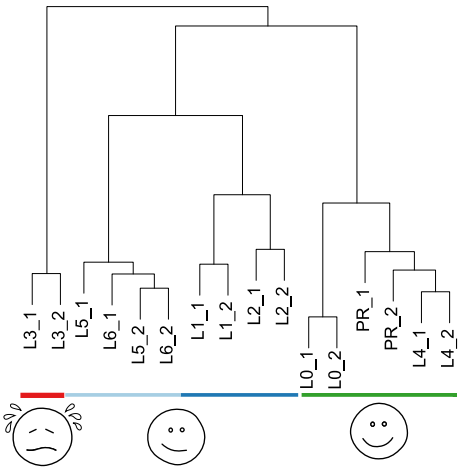


Figure 3.5.8: Unsupervised hierarchical clustering of features characterizing the metabolic states. See also figure 3.5.9.

identifying explanatory metabolites and relationships.

The two stable states during exercise (L1/L2 and L5/L6) were hypothesized to be mainly glycolytic and oxidative, respectively. The crisis during L3 had been mechanistically defined to be characterized by oxygen release and oxidative stress in erythrocytes (figure 3.5.4E). In order to identify the molecular basis for the "energetic" cluster composed of L0 and PR (stable), as well as L4 (part of a transition), we performed a factor analysis on a Pearson correlation matrix of all identified and unidentified metabolites. Several unidentified metabolites, triethanolamine, and creatinine were among the "energetic" metabolites, the latter of which potentially indicating energy availability in muscle cells due to its close relation to creatine and creatine-phosphate. [201]

Three essential elements seemed to define metabolic states and transitions in the subject:

i Metabolites correlating with a low rate of perceived exertion (RPE) (feeling "energetic"),

ii markers of oxygen release and oxidative stress,

iii the relative activity of fermentative glycolysis and the oxidative TCA-cycle.

When including only these parameters and relationships, principle component analysis (not shown) and hierarchical clustering successfully recapitulated the principle relationships as observed with all metabolites (figure 3.5.8).

These explanatory components of the metabolic response to exercise in the volunteer are summarily visualized in figure 3.5.9.

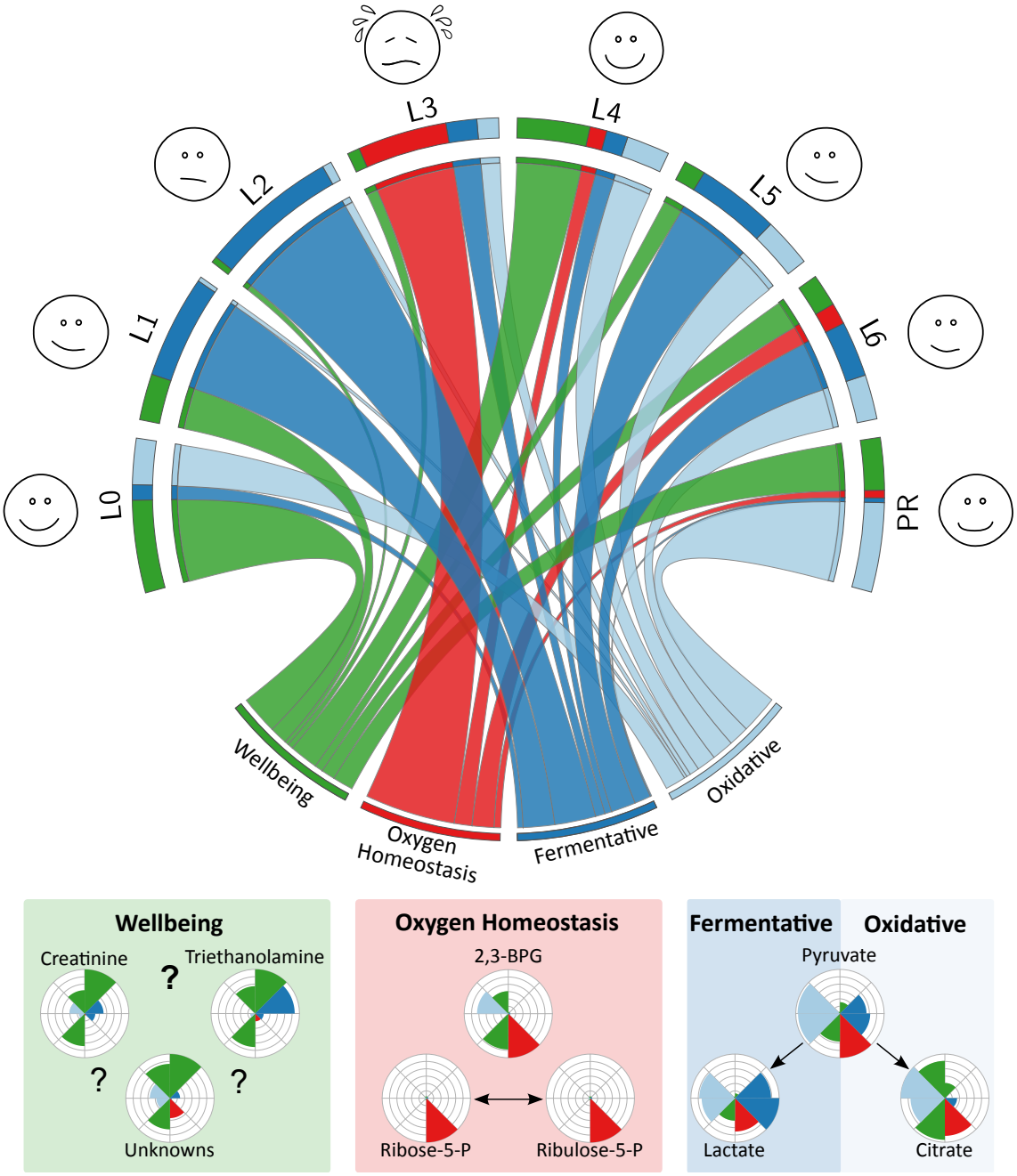


Figure 3.5.9: Top: Circos plot [151] of the selected metabolites characterizing the metabolic states described; below: Nightingale plots [194] of the metabolites separating the metabolic states.

4 Discussion

4.1 The established concepts of fiber types are reflected in the data

Since the first descriptions by Ranvier in the 1870's [202] muscles have been grouped into different categories, regarding their function and specialization, which reflects their morphological, molecular and biochemical properties.

ATPase-staining is a well established technique, but also determination of MYH isoforms on translational or transcriptional level has been used.

The concepts found to be enriched in the features separating the muscle types, composition of muscle types reflect their function: muscle needs energy to perform its primary task, facilitation of movement. This energy is primarily derived either through oxidative metabolism, with fatty acids as major fuel source, or anærob metabolism, using glucose with either lactate as output, if pyruvate is not shuttled into the TCA-cycle. [24]

Another important difference is rigidity of muscle. Oxidative muscle, which is primarily used in endurance movement and maintaining body position, is also physically more rigid, than normally less active fast twitching glycolytic muscle. This difference in physical rigidity stems from specific isoforms of those proteins that compose sarcomer and cytoskeleton. The sarcomer can be further divided into z-disc, m-band and thin- and thick filament. The proteins that finally use the vast amount of ATP muscle consumes [5, and references therein] can be found in the thick filament: the myosin heavy chain proteins.

In histology, their different performance at varying pH-level and since technological advancement allows for it, the amount of their isoforms at RNA-level are used to discern fiber types. There are over ten isoforms of myosin heavy chain, but mostly the genes MYH1, MYH2, MYH4 and MYH7, showing highest abundance in skeletal muscle in general, are described. Many are deemed specialised and only expressed in e. g. extra ocular muscles or during muscle development [203]. In the proteins of the dataset also several of those isoforms, deemed less ubiquitous expressed, have been quantified at isoform level besides all of those proteins that form the sarcromer.

So far, when studying muscle fiber types, either only few characteristics [46] or only two different kinds of muscle [111, 112] were compared. As far as omics-approaches comparing muscle fiber types exist, there is none, that integrates metabolomics and proteomics data.

In 2012 Drexler *et al.* [111] compared *Extensor digitorum longus* and *Soleus* muscles of C57BL/6 mice and found 25 % of the quantified proteins to be significantly differentially

expressed. They found differences in sarcomeric proteins, many kinases and enzymes in glutathione metabolism to be differentially expressed in the two muscle types compared.

In 2015 Rakus *et al.* [112] compared *Soleus* and *Tibialis anterior* of mice and found, besides the known differences in sarcomeric proteins, that enzymes comprising those pathways that are predominantly used by the respective muscletypes differing in their primary fuel source are differentially expressed. They also compared reaction rates of the extracted enzymes and found good correlation to their relative amounts. Rakus *et al.* concluded, that the upregulation in expression of a certain pathway is correlated with its activity.

In regard to sarcomer composition, and amount of metabolic enzymes I found similar results as described by Drexler *et al.* [111] and Rakus *et al.* [112].

In 2015 Deshmukh *et al.* [113] compared myotubes generated from C2C12 cells with *M. triceps* of C57BL/6 mice. They compared the two often used model systems and found 44 % of the proteome to be significantly different.

Mauri, *et al.*, 2015 [204] isolated single muscle fibers and performed proteomics analysis of single fibers of *Extensor digitorum longus* and *Soleus* from CD1-mice. They found differences in MYH distribution along the muscle fiber, as had been described for extraocular muscles before. [205] They also found distinct isoform-patterns in mitochondrial enzymes, that is likely to reflect metabolic regulation. This constitutes a very insteresting approach. However at the current state of technology, an integrative metabolomics and proteomics approach based on this technique is not possible, because of overall too low metabolite concentration and several hindrances regarding fiber preparation in a metabolomics conform manner.

These studies gave novel insights, but alike there are three "fast" muscle fiber types known in the literature (IIA, IIX, IIB), also mixed forms exist [45, 46, 160, 206–208] and there is no prototype of especially the fast muscle in mice. As prototype of slow muscle always *Soleus* is used, since it has the highest percentage of type I fibers in mice. Whereas in rats, *Soleus* is almost entirely composed of type I fibers, in mice only ≈42 % of the fibers of *Soleus* are characterized as type I. [45] It has been suggested, that mice show a lower percentage of type I fibers than bigger rodents, because of their lower body weight and thus lower effort to keep body posture. [207] Also intermediate forms of the "classic" fiber types, I (*Mhc7*), IIA (*Mhc2*), IIX (also known as IID, *Mhc1*), IIB (*Mhc4*), constitute a considerable part of the muscle which reflects also the findings in this study: although general characteristics validate the separation into the four "classic' types, this generalization leaves out the pontential interaction between neighbouring fibers and the

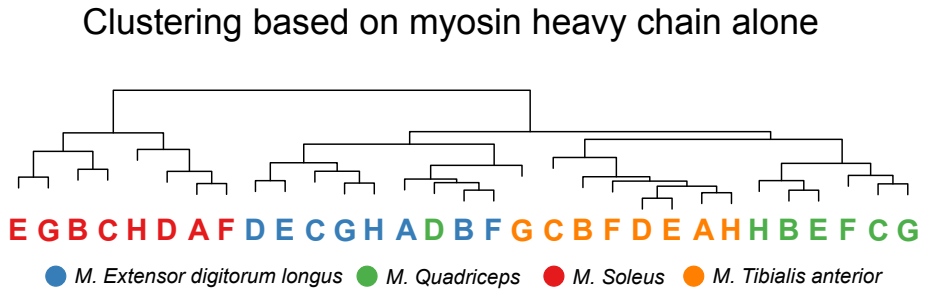


Figure 4.2.1: Hierarchical clustering of the samples only based on isoforms of myosin heavy chain. It is nearly possible to discern the muscle types only on this small feature set.

resulting molecular phenotype adapted to function, training and surrounding tissue. Table 9 gives the fiber type composition of muscles of C57BL/6 mice, as used in the present study, including mixed fiber types, that express more than one myosin isoform [45, 46, 207].

Table 9: Fiber type distribution in mouse muscle as obtained by ATPase-staining and distribution of MHC-isoforms as determined by western blot. Reproduced from [46]

Fiber type (ATPase-staining)	Muscle		
	SOL	EDL	TA
I	37.42 ± 8.20	0.44 ± 1.27	0.17 ± 1.25
IC/IIC	0.00 ± 0.00	3.55 ± 3.49	0.008 ± 1.04
IIA	38.62 ± 6.81	0.46 ± 0.68	1.12 ± 2.17
IIAD	18.74 ± 6.95	7.56 ± 4.51	3.42 ± 4.87
IID	5.69 ± 3.09	0.46 ± 1.34	2.25 ± 1.64
IIDB	0.00 ± 0.00	21.48 ± 7.33	33.83 ± 15.85
IIB	0.00 ± 0.00	66.01 ± 8.51	59.68 ± 9.95

MHC (Western blot)	Muscle		
	SOL	EDL	TA
I	41.50 ± 12.21	0.00 ± 0.68	0.00 ± 0.0
IIa	57.56 ± 13.32	10.82 ± 6.6	25.52 ± 15.15
IId	0.15 ± 2.82	0.00 ± 0.0	0.00 ± 0.0
IIb	0.00 ± 0.0	88.30 ± 6.6	74.48 ± 15.15

4.2 Myosin heavy chain isoforms separate oxidative and predominantly glycolytic muscles

Using only the myosin isoforms alone, it is nearly possible to separate the samples obtained from the different muscles by their origin (see figure 4.2.1).

Figure 4.2.2 shows the amount of each MYH found in the sample set. Although those isoforms usually used to describe fiber types, are by far the most predominant isoforms, also other isoforms like MYH3 (associated with muscle development and regeneration) or *Myh8* (predominantly expressed in extraocular muscles) were found to be expressed at significantly different levels. This does not mean, that these differences are necessarily of biological relevance, but it shows the power of the omics approach: proteins and metabolites of relatively high similarity can be reliably quantified at isoform level over several orders of magnitude. Drexler *et al.* [111] found similar results.

4.3 Metabolite and enzyme levels do not always correlate — examples with focus on mitochondrial metabolism

It is well known, that the long standing "paradigm" of biology, transcription leads to RNA, translation leads to proteins, proteins produce metabolites might hold generally true, but also many different ways of regulation and interaction are known.

Although I only measured poolsizes of proteins and metabolites, and not their formation rates or the overall phosphorylation status of proteins, that would reflect the intracellular signaling, often general agreement is expected as for example in the work by Rakus *et al.* [112].

I was able to discern all muscle types based on proteins and metabolites alone (figures 3.2.3 and 3.2.9, and infer concepts differentially regulated in the respective muscle groups figures 3.2.4 to 3.2.14).

In the following some remarks about peculiar differences between metabolite and protein levels will be given, mostly relating to parts of the CCM and overall energy levels, and the power an unbiased systems biology approach might hold over other methods.

If a fast muscle needs to be able to put out bursts of energy in a short period of time, this energy must be available at the time of need and thus stored in a different form, than in a slow contracting muscle that works continuously over longer periods of time. This is for example facilitated by the creatine-phosphate system in fast muscles and reflected in figure 3.2.15 with creatinine highly enriched in EDL. These different energy levels need different tuning of the energy sensors.

Although the results I receive agree in general with the findings by Rakus *et al.* [112], it should not be concluded that higher amounts of enzyme necessarily lead to congruent metabolite levels. As with the relation between RNA and proteins, this may hold true in many cases, but not all of them, especially since reaction rates depend stronly on substrate

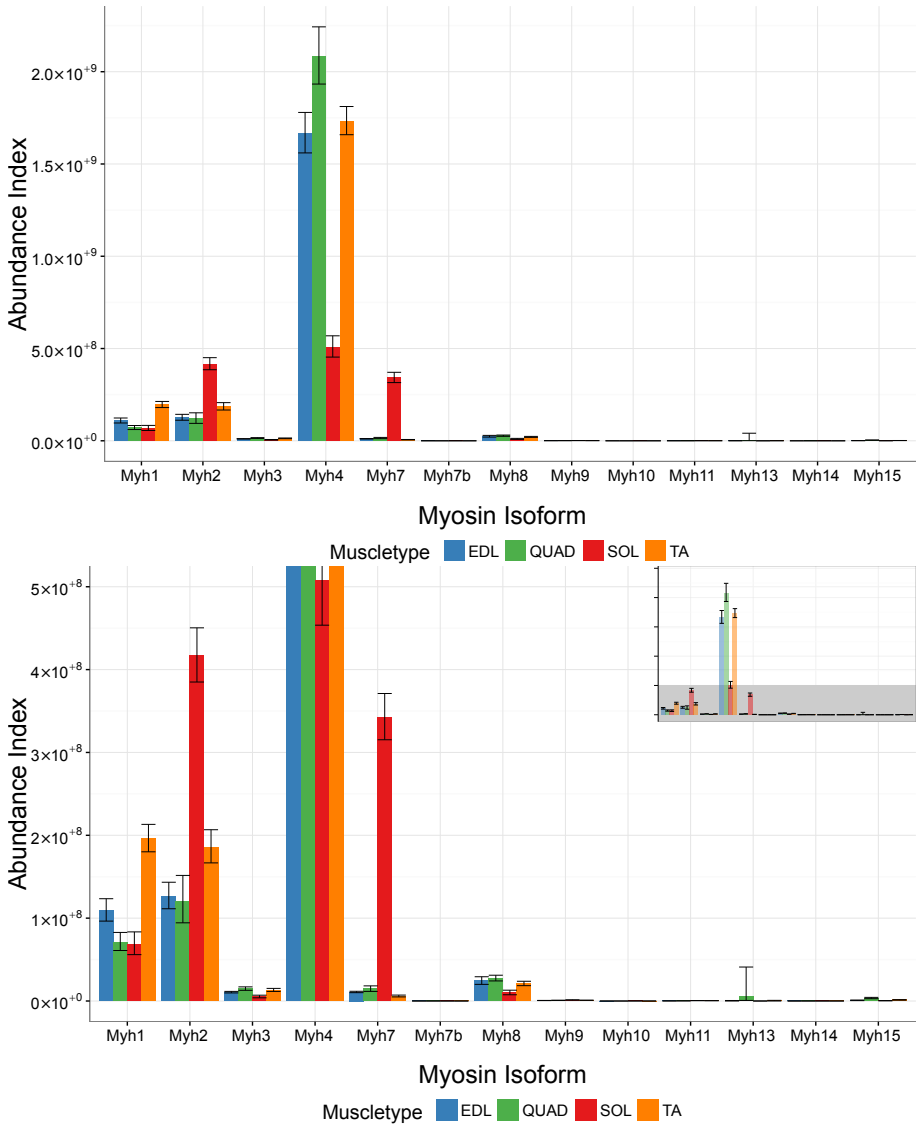


Figure 4.2.2: Abundance index of all myosin-heavy-chain isoforms quantified in the four muscle types. Top: total Abundance index, bottom: zoom into lower abundant isoforms as indicated in the inset

and product concentrations. [209, 210] Since the TCA-cycle is of major importance in (muscle) metabolism, I will discuss an example revolving around 2-hydroxyglutarate (2OHG), a metabolite that has received much attention recently. It was known for some time, that the D-stereoisomer was produced by mutant IDH1 and IDH2 functions as oncometabolite that can inhibit enzymes that use α-ketoglutarate [211–214]. The L-stereoisomer, that is produced by both the cellular and the mitochondrial isoform of malate dehydrogynase (MDH), has recently been shown to be a marker for hypoxia. [214–216].

Even though it is not possible to discern D-2OHG and L-2OHG by the GC–MS approach used in this study, it is very likely, that the majority of 2OHG is in fact L-2OHG, since this is the form usually produced under hypoxic conditions. 2OHG is highly enriched in SOL compared to the other muscles (see figure 3.2.15. Starting from iso-citrate (*i*Cit), IDH1 (slightly) and IDH2 (strongly) are enriched in *Soleus*. According to Rakus *et al.* [112] this should mean, that α-ketoglutarate (αKG) is to be enriched in SOL as well. However, this is not the case, the metabolite levels are about the same. Efflux from aKG can go to different pools: glutamate, (indirectly) succinate and 2OHG.

The abundance index of MDH1 is about 1.5-fold higher in SOL than in the other muscles, whereas MDH2 is enriched both for TA and SOL. So according to Rakus *et al.* [112] both in TA and SOL 2OHG should be upregulated. However, 2OHG is strongly enriched only in SOL, but not in TA. This demonstrates, that the assumption by Rakus *et al.* [112] holds true for some cases, but has to be treated with extreme care. If however the knowledge, that 2OHG is upregulated in hypoxia, is included in these considerations, the upregulation in the highly oxidation-active *Soleus* makes sense. These comments do not intend to dismiss or discredit the general idea of importance inferred by upregulated enzyme levels, but want to emphasize the importance of inclusion of metabolites into reasoning about cell states and metabolic activity. In the light of whole pathways combined, this is even more likely to be true, however, if the goal is to understand individual steps that may be influenced by diseases, the metabolite levels may give much deeper insight, than the proteome alone.

As mentioned above, the concepts, that split the muscle types are mostly metabolic. Even for the second principal component or discriminant, where only few concepts were found at all, the general theme of “Metabolism” shows the dominance of muscle as a tissue that uses up nutrients sent by the liver and adipose tissue, generating ATP through the CCM to facilitate movement.

It could be expected, that because of their strong reliance on oxidative metabolism and general higher mitochondrial content [47] SOL simply has the highest expression in all proteins related to mitochondria or amelioration of oxidative stress,e. g. glutathione

transferases.

This holds true for most proteins in these categories by far, but this is not always the case. There are certain isoforms that are signficantly higher expressed in glycolytic muscles (GSTM5), and a high number are expressed evenly. For "mitochondrial" proteins, there are even more drastic examples, as GPD2[10] (genral: $fc_{GPD2} \approx 8$, $p_{GPD2} < 10^{-20}$, the calcium carrier CMC1[11] (for TA: $fc_{CMC1} \approx 2.25$, $p_{CMC1} < 10^{-7}$) or SQRDL[12] (general: $fc_{SQRDL} \approx 12$, $p_{SQRDL} < 10^{-15}$). The mentioned examples are the most extreme cases of 18 out of 274 "mitochondrial" proteins.

But not only mitochondrial metabolism is highly different among the muscles investigated here. The appearance of the phospho transferase system (PTS) as significant term in PC1, that separates SOL from the other muscles reflects also the importance of sugars and sugar-alcohols.

It should be mentioned, that the results from IMPaLA have been obtained for protein IDs that were mapped to their human orthologs, and do not always necessarily reflect the same functions across species. However, most of the features that discern the muscles are members of the CCM which is highly conserved among species. The other big part, the sarcomer constituents can be regarded as highly conserved as well, at least in mammals or more general vertebrates.

In general muscles each seem to possess a metabolic phenotype tailored to their specific function, like movement or simply keeping a body posture that allows the other organs to function properly.

It is known, that the mere existence of lean body mass allows a person to lead a healthier and longer life and has positive influence on inclination towards disease in general.

Since type II fibers are much more susceptible to dysferlinopathy it is of special importance to gain a deeper understanding of the mechanism and specific properties of those muscles. An example on how this can be achieved through systems biology will be given below in combination with the discussion of the results obtained from application of pSIRM to myotubes generated from patients suffering from dysferlinopathy.

[10] Glycerol-3-phosphate dehydrogenase, mitochondrial
[11] Calcium-binding mitochondrial carrier protein Aralar1
[12] Sulfide:quinone oxidoreductase, mitochondrial

4.4 Changes in poolsizes in dysferlinopathy are stable, irrespective of the fuel source

Despite the levels of all proteins involved were similar between samples from mice or humans with dysferlinopathy, my predecessor Sarah Keller [1] had found a block of glycolysis at the level of hexokinase II. This was inferred from the result, that the poolsize of G6P and the downstream metabolites was reduced in disease, whereas the metabolites of an alternative pathway through sorbitol showed higher levels. Metabolite poolsize however does not allow to infer speed of a reaction, since every reaction can be reversed, and it is never clear, if accumulation occurs through blockade of the following reaction step, or by higher activity of the preceding step.

Another issue to be investigated was glucose uptake and possible routing through the glycogen pool. About 80 % of the glucose taken up into non contracting muscle is temporarily stored as glycogen. [37] As such it was of major importance to see if glycogen build up is impaired in dysferlinopathy. For this the method established by Matthias Pietzke and Christin Zasada was extended to also include precipitation and hydrolysis of glycogen, enabling us to trace label incorporation into glycogen.

After I successfully tested several ^{13}C-substrates, I tried to label myotubes generated from samples of patients suffering from dysferlinopathy with $^{13}C_u$-glucose or $^{13}C_u$-3-hydroxybutyrate.

Using glucose, label was detectable normally. For 3OHB no conclusive results regarding label incorporation could be obtained. This might be due to several reasons: In the body ketone-body levels rise in states of increased turnover of fatty acids in the liver, when insulin levels are low. However, the cell culture media SKMGM contains high levels of insulin, more than 10 fold higher, than physiological level.[13] This high level is needed to keep the primary muscle cells growing.

It was not the aim of this thesis to define a cell culture model mimicking a ketogenic nutrition state. Also, it is hard to know, at what point cells would stop to grow, if insulin levels are reduced, since it is likely, that this varies strongly with the respective patient.

The media was also deprived of glutamine, which is used to replenish TCA-cycle intermediates (TCAis). The absence of glutamine may have hindered his. When labeling with 3OHB, the myotubes were confronted with an unphysiological situation: high insulin levels, no glucose but 3OHB and no glutamine. This might have led to an extreme retardation of metabolism.

[13] The exact level is known to me, however I was asked by the manufacturer to keep it confidential.

If only poolsizes are regarded, interesting similarities, despite the change in fuel source, have been found. The occurrence of higher sorbitol and fructose levels are rather stable; in general, if a certain metabolite was up or down regulated in both carbon-sources, the direction of fold-change always stayed the same. This points towards an underlying stable adaptation to the metabolic situation, where glyoclysis seems to be impaired.

If glycolysis is impaired, the cell has to use other carbon sources. Oil-red-O-staining of cells of patients suffering from dysferlinopathy shows high levels of lipids[14]. This led to the assumption, that these cells predominantly thrive on lipids as carbon source. This theory is underpinned by the upregulation of TCAi and members of the pentose phosphate pathway (PPP). The PPP because this pathway provides reduction equivalents used to replenish glutathione, which in turn is used to ameliorate oxidative stress provoked by hightened TCA-activity.

Creatinine, as a marker for general energy availability in the cell is extremely down-regulated if 3OHB is provided as fuel source. This is not the case, if glucose is provided as fuel source. This points toward decreased energy availability in the unphysiological state, whereas under glucose treatment, the observed rerouting through the polyol-pathway seems to at least partially compensate for impaired glycolysis.

The upregulation of ornithine and urea shows heightened protein breakdown through proteolsysis, likely to compensate for lower energy availablity. This increased protein breakdown likely contributes to the overall disease symptoms.

The metabolite poolsizes overall substantiate the hypothesis, that in dysferlinopathy, there is an incomplete loss of function of hexokinase II.

4.5 Labeling experiments confirm an impairment of hexokinase II in dysferlinopathy

As mentioned above, regarding label incorporation, no conclusive results could be obtained from labeling with $^{13}C_u$-3-hydroxybutyrate. There was only very little label incorporation in general and a high unspecific background in the ^{12}C-3-hydroxybutyrate treated controls. General tendencies favor the hypothesis, that myotubes derived from patients that that suffer from dysferlinopathy show a faster running TCA-cycle, but overall the quality of the data does not warrant to draw this conclusion.

The data from labeling with$^{13}C_u$-glucose, fits the hypothesis of a loss of efficiency of hexokinase II. While there is no difference in label incorporation in sorbitol, G6P and

[14] in mice [217], but also in humans Spuler group, unpublished

downstream metabolites as lactate but also glyocgen show slower label incorporation. After a longer period of time, the same level of label incorporation is reached. Although for glycogen this has not been shown, because the overall poolsize of glycogen is rather big and labeling to saturation would take longer than experimentally feasible. This means that hexokinase II is in principle functional, but through the impairment of dysferlin it cannot function as efficient as in healthy individuals.

In this framework several factors can have an influence on the glycolytic flow through hexokinase II (HKII). The substrate of HKII, glucose, has to be taken up into the cell. In muscle this is predominantly achieved by use of either the insulin independent transporter GLUT1 or the insulin dependent uptake by GLUT4.

It was shown, for C2C12-derived myotubes, that they express far more GLUT1 than GLUT4. This contrasts the situation in muscle tissue, where by far GLUT4 is the more dominant means of glucose import into the muscle cell. [113].
If GLUT1 or GLUT4 is responsible for the uptake of glucose in our cell culture model is however of secondary relevance, since the label incorporation into sorbitol is not different between patients and healthy individuals. Also, as mentioned above, availability of insulin signal is certainly not the limiting factor.

Any metabolic cascade functions more efficiently if its members are in close vicinity, so that diffusion paths of intermediate products are short, and if the respective substrates are at high enough concentration. [209, 210] When glucose is inside the cell, it is phosphorylated by hexokinase II. This lowers the concentration of glucose itself and shifts the dynamic equilibrium to facilitate further import of glucose.

For the G6P formed by hexokinase II two main routes exist. Either it can pass through glyocolysis with glucose-6-phospate isomerase forming fructose-6-phosphate as the next step, or it can be funneled into the glycogen pool by means of phosphoglucomutase, that forms glucose-1-phosphate which is then polymerized as glycogen.

The downstream products of both these pathways show slower label incorporation. This allows for the conclusion, that the flow through hexokinase II is impaired in patients suffering from dysferlinopathy.

In her thesis, Sarah Keller [1] could show, that *Quadriceps* is the most affected muscle in dysferlinopathy.

Figure 4.5.1 shows all the enzymes mentioned above, except for GLUT1, which was not quantified in the muscles. The latter is not surprising, given its generally low abundance in tissue. [113]

Except for hexokinase II, the enzymes involved in glycolysis and glycogen formation

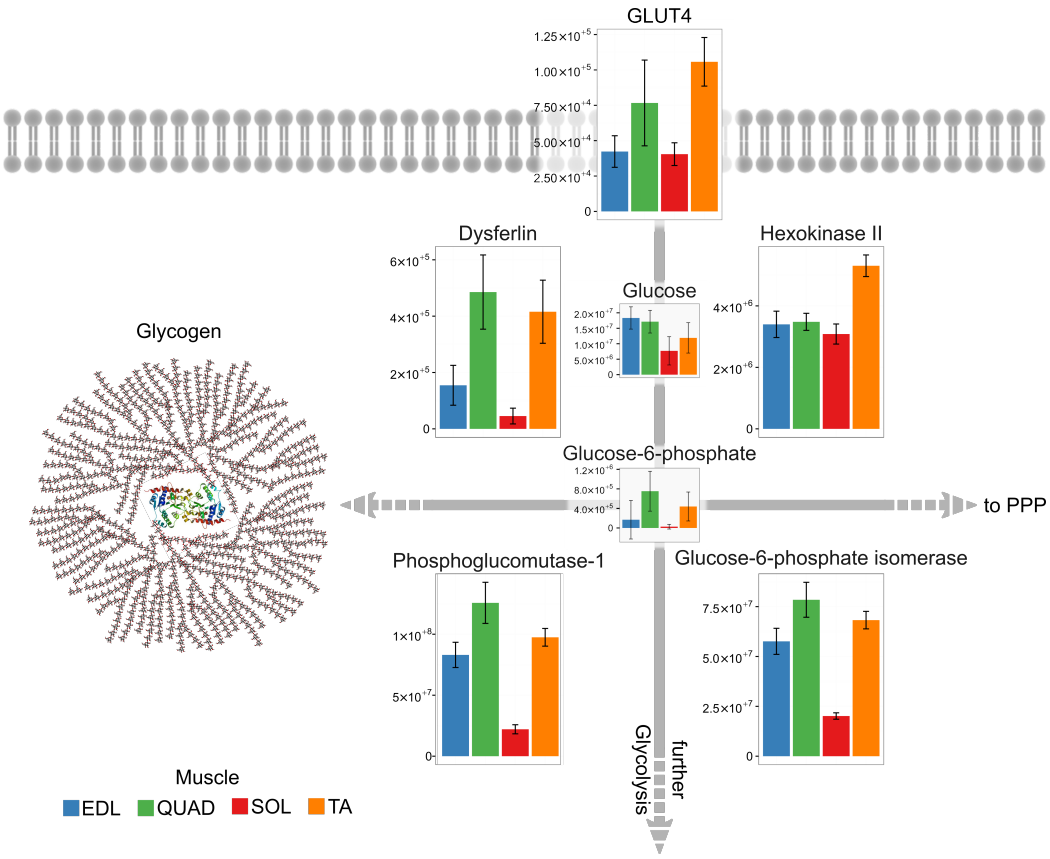


Figure 4.5.1: Upper part of glycolysis in different muscle types, relative abundance. Errorbars show 95 % confidence intervals based on logarithmic values. Hexokinase II is higher expressed in *Tibialis anterior* (TA) while dysferlin is similar expressed in *Quadriceps* (QUAD). In her thesis, Sarah Keller [1] found, that in BLA/J mice, a well recognized model of dysferlinopathy [74], QUAD shows a more pronounced metabolic phenotype than TA, whereas there were no differences in amounts of glycolytic enzymes compared to the respective wild type muscles. This could mean, that in dysferlinopathy the generally higher expression level of hexokinase II in TA might ameliorate the lesser efficiency of hexokinase II in dysferlinopathy.
Glycogen figure adapted from Haeggstroem, *et al.*, 2014 [218].

are higher expressed in glycolytic muscle (EDL, QUAD, TA). This reflects their general reliance on glucose as fuel source. Interestingly GLUT4 is expressed more strongly in QUAD and TA than SOL and EDL. This is peculiar in itself, when EDL is used as representative for glycolytic muscle and SOL as representative of an oxidative muscle. Nothing about this is mentioned by Drexler *et al.* [111], who compared EDL and SOL, and highlights again the particular phenotypes of different muscle phenotypes. Compared to QUAD and TA EDL has a much higer surface-to-volume ratio and thus EDL simply might need less GLUT4 in order to ensure adequate glucose import, or insulin signaling has a different potency on these muscles.

Hexokinase II is comparatively enriched in TA. This could be interpreted as a higher demand for this enzyme to ensure proper function of glycolysis. Dysferlin, which, according to the data presented here, has an influence on efficiency of hexokinase II is highly expressed in TA which has also a high level of hexokinase II, but also in QUAD which shows no enrichment for hexokinase II and is more prone to the symptoms of dysferlinopathy. This could be interpreted as follows:
TA shows lower symptoms of dysferlinopathy, since its higher levels of hexokinase II can ameliorate the lower efficiency of hexokinase II induced through dysferlinopathy, while QUAD loses its ability to perform glycolysis in a sufficient manner, because it already has lower levels of hexokinase II being more dependent on efficient performance of the first step in glycolysis.

4.6 The proteome of dysferlin vesicles reveals secondary interactions with metabolic enzymes

When regarding the proteomes of dysferlin vesicles (section 3.4) the major concepts among the proteins found in there are not surprising and reflect the widely accepted function of dysferlin as a membrane repair and vesicle forming protein. But also new, partly vesicle mediated interactions have been found, not recognized in the literature thus far. Very recently Li, *et al.*, 2016 [190] found a role for dysferlin in lipid droplets in heart tissue. This fits the observation in our data, that several mitochondrial proteins have been identified as interactors of dysferlin and underlines its obvious involvement in the central carbon metabolism (CCM) (see figure 3.4.3).

Not only an involvement with mitochondria could be found, but also several glycolytic enzymes are enriched among the —not exclusively vesicle mediated— dysferlin interactors, namely Aldolase A (ALDOA), lactate dehydrogenase A&B (LDHA, LDHB),

glyceraldehyde-3-phosphate dehydrogenase (GAPDH), phosphoglycerate mutase (PGAM2 or 1 or 4) and pyruvate kinase (PKM), as shown in figure 3.4.4 on page 97.

Although hexokinase II was not found as interactor of dysferlin, other glycolytic enzymes have been found to be enriched with dysferlin, which reflects an involvement of dysferlin in glycolysis.

Even though the direct comparison of pulldown-studies contains many pitfalls, since for example lysis parameters or detergents can have an extreme impact on the results, our data is in general good agreement with work by de Morrée, *et al.*, 2010, [183] and Flix, *et al.*, 2013 [184]. In these studies also several of the aforementioned glycolytic and mitochondrial enzymes have been found, but an involvement in glycolysis has not been discussed so far. De Morrée, *et al.*, 2010 attributed the enrichment of metabolic enzymes in general and mitochondrial enzymes in particular to "the high content of metabolic enzymes in muscle". While this is undebated here, thanks to the vast effort to have reliable controls, it can be stated already from the proteomics data alone, that dysferlin shows an involvement in central carbon metabolism.

This also reflects that dysferlin is significantly higher expressed in all glycolytic muscles than in *Soleus*, as had been found for *Extensor digitorum longus* also by Drexler, *et al.*, 2012 [111], but left undiscussed.

In both de Morrée, *et al.*, 2010 [183] and Flix, *et al.*, 2013 [184] also 6-phosphofructokinase (PFK) has been identified as interactor of dysferlin, also verified by blue native gel electrophoresis. Although it was not present in our dataset, in these studies also muscle tissue was used, which might explain this. This shows an even greater role for an effect of dysferlin on glycolysis activity, since PFK is the key regulator of glycolysis. [10]

In regard to the overall changes in metabolism demonstrated by the label incorporation experiments described above (3.3.4), the dysferlin vesicles proteome nicely reflects the overall changes in metabolism in dysferlinopathy patients.

In muscle hexokinase is located at the outer mitochondrial compartment [100, 219, 220], while other glycolytic enzymes have been located in the I-band of the sarcomer. [100, 221, 222] It has been speculated by Ohlendieck, *et al.*, 2010 [100], that this distinct localization has an impact on regulation of glycolysis. At harvest, the myotubes enriched for dysferlin-vesicles by Séverine Kunz were in a standard unstimulated state. It might be worthwhile, to repeat these experiments for starved and insulin stimulated cells.

Dysferlinopathy has been attributed with either a loss of type II fibers or a fiber type switch of type II to type I fibers. [61] Given the normal fiber classification, that is mainly based on metabolism, this observation might simply reflect a means by the muscle cell

to cope with the effects of diminished glycolytic efficiency. In her proteomics data of dysferlin deficient mice, Sarah Keller [1] found only very few alterations in overall protein expression between BLA/J and wild type muscle. There is a slight increase of aldose reductase in TA muscles of BLA/J mice. Aldose reductase reduces glucose to sorbitol, so this observation is congruent with the data presented here.

Regarding proteins, the striking observation was a strong decrease in specifically isocitrate dehydrogenase 2 (IDH2), that uses NADPH as cofactor, opposed to IDH3, which uses NADH and was not altered.

This again reflects the importance of the PPP in dysferlinopathy. Overall dysferlinopathy might simulate a phenotype of type I fibers, that could finally force the muscle into a more type I conform expression profile.

4.7 Possible pseudo-fiber type switch dysregulates overall metabolism in dysferlinopathy — Reflections of oxygen availability

To function properly, muscle needs enough oxygen, that is provided by means of myoglobin mediated diffusion in the long stretched multinucleated cell. Of course glycolytic muscles show less mitochondria than oxidative muscles, but they still also perform oxidative metabolism. [47]

Given the structural peculiarities of muscle cells, myoglobin mediated diffusion in these remarkably shaped cells has been vastly researched. [223] In healthy muscle, diffusion is not the *per se* limiting factor for any metabolic activity, but many fibers operate near diffusion limitation, especially in adults. [224] That means, if there are small pertubations in this delicate balance, the outcome may be dramatic.

Fiber diameters increase with potential for glycolysis which can function in anærobic state, given efficient lactate clearance. [225] In dysferlinopathy type I fiber preponderance has been observed [61], meaning, that type II fibers may switch to type I.

The surface of a fiber, thus the amount of membrane, increases in quadratic proportion to its diameter. Thicker glycolytic fibers need more membrane, whose repair is impaired in dysferlinopathy.

If repair is compromised, a first reaction might be overall reduction of membrane. Additionally production of sorbitol, as observed in dysferlinopathy consumes NADPH, which leads to production of glycation end-products (AGEs) and reactive oxygen species [226, 227]. AGEs themselves induce a membrane repair defect in muscle fibers similar to the

one seen in dysferlin-deficiency. [228]

The defect in membrane repair might lead to a shrinkage in fiber diameter, which in turn heightens oxygen concentration and availability. Higher oxygen levels shift the metabolic equilibrium to more oxidative metabolism and increase again oxidative stress, and in turn upregulate PPP. This was also found in the metabolite profiles of dysferlin deficient myotubes, shown in figure 3.3.8 on page 88.

Oxidative metabolism induces type 1 fiber characteristics in glycolytic muscle [229].

Also upregulated oxidative metabolism blocks glycolysis, albeit most prominent not at hexokinase-level but at 6-phosphofructokinase (PFK) level [10, 230], the latter being detected as interactor of dysferlin by de Morrée, *et al.*, 2010 [183] and Flix, *et al.*, 2013 [184]. Also pyruvate kinase regulates glycolysis. As in both de Morrée, *et al.*, 2010 [183] and Flix, *et al.*, 2013, its muscle specific isoform was detected in both subsets of dysferlin interactors (direct interaction and vesicular fraction). It was the only glycolytic enzyme significantly higher enriched in the vesicular fraction versus the complete lysate, as shown in figure 3.4.4. It is activated by rise in fructose-1,6-bisphosphate (F1,6BP) levels and inhibited by rise in alanine levels [10]. Interestingly F1,6BP, a metabolite normally produced by glycolysis, represents also a connection point of the polyol pathway, which is upregulated in dysferlinopathy. Consequently F1,6BP is also upregulated, which would activate pyruvate kinase. On the other hand also alanine is upregulated in dysferlinopathy, which would downregulate glycolysis.

As a result there are two different metabolic regulators of the same enzyme possibly producing conflicting signals.

The overall molecular makeup of fast glycolytic muscle is not suited to function under ATP-availability or at ATP-production speed at the level of an oxidative muscle (table 1) especially if regulation is overall inconsistent.

Given the novel insights presented here, and that many different characteristics are tailored to the specific molecular makeup of the muscle, the observed type I fiber preponderance [61] might actually be a *"pseudo-switch"* resulting from a multitude of failed or contradictory regulation states. These might lead to the observed disease phenotype of reduction in type II fibers or an incomplete "pseudo-switch" to type I fibers. Since still the muscle is partially equipped to import glucose at a high rate, but flux trough glycolysis is downregulated, glucose is rerouted through the polyol pathway which likely diminishes energy availability and results in membrane reduction and in turn further oxygen concentration. Given the overall pronounced molecular makeups at isoform level, the overall isoform equipment of the fiber does not fit anymore to its physical properties, leading to

many dysregolatory effects, finally manifesting in the phenotype of dysferlinopathy.

Over time epigenetics might adapt to the new metabolic profile, but since dysferlin also shows high interaction with RNA-processing 7.3.1, there are likely influences on this adaption.

Sarah Keller [1] found only very small differences between proteomes of dysferlinopathy deficient mice and wild type. But regulatory effects can also be exerted directly by (micro)-RNA. [231]

Overall in dysferlinopathy molecular makeup and physical properties do not fit anymore, which leads to disease.

This also fits to the onset of dysferlinopathy in puberty. Especially in adolescent males, when muscles grow, their volume increases, also at fiber type level. [232] Perhaps in childhood the aformentioned effects can be compensated but when this balance is tilted further, the clinical phenotype manifests.

This hypothesis could be possibly tested by a longitudinal study of children and young adults, using an exercise protocol, similar to the one employed in our proof-of-concept study, where we were able to show influences of oxygen availability and its consequences on metabolism wich will be discussed in section 4.8.

4.8 Exercise and self-reported energy state are reflected in the metabolome of blood

The data obtained from a single human volunteer performing endurance exercise enabled us to recapitulate well established knowledge but also to gain novel insights regarding the mechanism behind the phenomenon described as “Hitting the Wall”.

During exercise, an initial glycolytic phase was followed by a second phase in which free fatty acids (FFAs) and ketone bodies were abundant, while glycolysis contributed less strongly to meet the energetic demand. These two phases were separated by the “crisis” and subsequent “high” feeling.

All this might not represent novel findings, but we wanted to see, whether our findings in a single individual correspond to established knowledge, which is the case. Thus we can say, that our findings seem to be valid and might allow for deeper interpretation, especially regarding the fact, that the findings we reproduced here stem from multiple separate studies, while our method allows for similar results, but from a single experiment. This encouraged us to further study our dataset in order to gain a deeper understanding of the metabolic crisis when our subject was “Hitting the Wall”.

As erythrocytes lack mitochondria, the pentose phosphate pathway (PPP) is the only means of efficient NADPH regeneration, which is required for the regeneration of the antioxidant and main radical scavenger glutathione. [233] This points towards an increase in oxygen turnover by the erythrocytes. Taken together, while the TCA-cycle, most likely in muscle, appeared primed for activity by L2, the body of our subject suffered an acute oxygen shortage during the crisis. This has already been shown by Bowtell *et al.*, however only for measurements of the combined TCAi. [234, 235] Our approach enabled us to separate those and to find the lagging behavior of citrate compared to the other TCAi (figure 3.5.6).
Oxygen release mediated by 2,3-BPG caused a short pulse of oxidative stress in erythrocytes, and only after the establishment of sufficient oxygen distribution did citrate levels rise, indicating a fully active TCA-cycle and the resolution of the crisis (figure 3.5.4). The TCA-cycle activity remained elevated after the crisis in accordance with increased availability of both FFAs and ketone bodies during the second phase of exercise.

4.9 The minimal invasive sampling of blood from the fingertip might provide a tool in muscle disease diagnostics

This study was made possible by a reliable, informative, minimally invasive, and GC–MS based metabolomics approach adapted to minimal sample size using an optimized sample preparation and measurement procedure. The data obtained albeit from one experiment performed with a single individual, as is usually the case in personalized medicine, not only allows for the recapitulation of many known facts in exercise physiology in one single experiment, but its application in a longitudinal study provided a possible mechanistic explanation for an exercise-induced metabolic crisis in our subject.

It could therefore be sufficient to obtain real-time mechanistic assessments of many pathophysiological conditions and medical interventions that affect the metabolism of an individual subject and whose detailed mechanisms are unknown, especially since our method is untargeted and thus does not require early assumptions about the metabolites to be characterized. Assessing metabolism is of particular interest, because it represents both genetic predispositions and acute lifestyle influences. [236]

The sampling procedure, similar to a diabetic's blood glucose test, is minimally invasive and can be repeated frequently and easily without ill effect, encouraging the investigation of dynamic physiological responses to homeostatic challenges, diseases, or treatments. Specifically, the liquid-liquid sampling strategy preserves the metabolic homeostasis in

the analyzed specimen and favors a realistic perspective of metabolism. However, each method has limitations and sources of errors must always carefully be evaluated. Sampling capillary blood may be a source of variation by itself, but also the thickness of the blood and impurities due to the sampling procedure might affect the results. We tried to account for that by our normalization strategy, where we assumed, that not all metabolites would shift in the same direction from sampling point to sampling point, but that the overall AUC of the chromatogram would not change. Since we were able to reproduce well established knowledge our sample handling and analysis seems to be generally sound.

Opposite to such error is the easiness of the procedure allowing the sampling at nearly every place and condition. In case of exercise monitoring the obtained biological variation may clearly exceed the variation introduced by the procedure. For other conditions this has to be tested as prerequisite for the study design and choice of the method. The analysis of whole blood does not complicate the interpretation of the data, but rather both simplifies sampling and sample processing. At the same time it allows for novel insights by capitalizing on the diagnostic potential of erythrocytes that regulate oxygen homeostasis and are discarded in most diagnostic assays. This crucial metabolic parameter would usually require complementary methods.

The identification and quantification of key metabolites that serve as the branching points of interconnected, but often separately considered, metabolic pathways allows for a true systems interpretation and understanding (figure 3.5.1). We observed known metabolic changes throughout exercise, such as an initial reliance on glucose as the main fuel source, the subsequent activity of the Cori cycle, as well as fatty acid mobilization as indicated by increases in glycerol, palmitate, and stearate [162, 198].

In conservative methods, the investigator or clinician has to select for a comparatively small panel of metabolites. The potentially novel mechanism for the metabolic crisis of "Hitting the Wall" was identified by combining known facts about single metabolites and pathways, such as 2,3-bisphospho-glycerate (2,3-BPG) and ribose-5-phosphate (R5P) and ribulose-5-phosphate (Ru5P) as regulators and indicators of oxygen homeostasis [12, 13, 233]. Thus, by summarily considering the orchestrated interplay of different tissues reflected in the blood metabolome, new insights might arise from truly novel relationships or metabolites, but also the integration of previously separately considered and reported phenomena (figure 3.5.4).

This has allowed us to find indications of why the volunteer "feels bad" when "Hitting the wall" during exercise. The advance of research technology into its meaningful application in systems medicine is often particularly difficult with omics technologies. They require

extensive time consuming data analysis and the ability of experts to distill the vast datasets into their fathomable essentials that represent the explanatory components. Also, because of the high variablility of humans, due to nature and nurture often very big cohorts are averaged in (metabol-)omics studies. This is actually the opposite of personalized medicine and might in the end lead to a loss of information. In our case, the explanatory components were sufficient to replicate the relationships of sampling time points, indicating that it is possible to identify and elementarily understand systemic metabolic states in a humans or patient, respectively, just based on this individual alone, that serves as its own baseline. This condensation of the entire dataset into its essentials facilitated understanding and allows its efficient communication (see figure 3.5.9) to clinicians.

The advance of metabolomics from an explorative to a robust, reliable, and reproducible method that provides a consistent quantification over large sample batches is required for the transition into an applied, medically relevant technology. We suggest that GC–MS, coupled with our liquid-liquid sampling method that facilitates high time resolution and assessment of oxygen homeostasis, provides sufficient information for the transition from large cohort studies to personalized medicine and analysis.

The key to this transition is not to consider only whether observations in single individuals must generalize to whole populations while asking the question why and if cohort study findings are frequently not meaningful for individual patients. The demonstration that mechanistic observations can be made and metabolic states rationally defined in single individuals shows that studies on single individuals might deliver further insight in a patient's disease state. We have here presented a systems medicine platform that may soon allow clinicians to obtain biologically meaningful insights within days from sample collection from single individuals. Its focus lies on the personalized monitoring of dynamic responses to stimuli and longitudinal preventative monitoring and offers the possibility of personalized diagnosis and research. Unlike common cohort studies, this method is therefore aimed at identifying potential metabolic mechanisms instead of statistical differences of averages of large, heterogeneous populations (i. e. common biomarkers).

4.10 An individual could serve as its own control

Heterogeneity especially influences metabolomes as lifestyle and diet affect it more strongly on short time scales than proteomes, genomes, and epigenetic settings. Paired sample series are essential for comparability but simultaneously offer more meaningful information about the acute physiological states and responses than other technologies that produce snapshots.

In the envisaged setup the individual would serve as its own control. Given the fast responses and high dynamic of metabolism, it seems imperative, that the individuals physiology is challenged, e. g. by exercise or ingestion of certain food, which might briefly enlarge underlying deregulations of metabolic control.

Personalized monitoring will reduce issues of comparability between people of different ages, lifestyles, diets, and gender through paired controls, and might provide an alternative route to biological understanding of diseases by allowing for targeted $n = 1$ analyses, instead of the more common large n studies [237]. This could lead to a bench-at-bedside technology that might bring a wealth of $n = 1$ experiments analogous to studies of individuals with rare accidents or genetic mutations that have proven invaluable for a variety of medical fields. We therefore believe that this platform, from a single drop of capillary blood to the identification of mechanistic metabolic insight in humans, is a significant advance towards omics-based systems medicine and could be of great value to clinical research and diagnostics.

Given the central role of muscle in nutrient and energy turnover, especially in muscle disease perturbations of metabolism are to be expected. Similar to protein biomarkers of muscle damage, like creatine kinase and myoglobin [238, 239], especially under exercise conditions muscle is likely to be a main contributor to metabolic changes. The extent of this could be verified by microdialysis, that samples interstitial fluid. Comparing dialysates of muscle and fat tissue to blood samples obtained in various conditions could expand the current strategies in metabolic assessment by calorimetry in a metabolic chamber.

4.11 The combination of omics techniques allows to find more holistic insights into diseases

The adaption of physiology happens at various speeds on different layers of control. While the genetic makeup changes over (many) generations, transcription and translation can change in days to hours. However the fastest observable shifts are on metabolic and hormonal levels: in the instance we decide to move —or not to— metabolism adapts to the particular needs, given an overall healthy body. [240]

In minutes to hours also the proteomic makeup of the body changes, the body adapts to external challenges. If this is prolonged, fundamental physiological changes ensue, as for example in exercise training. [223] Muscles can change their molecular makeup so far on a molecular level, that they change their former classification type. [61, 108–110]

I could show, that muscles have a unique molecular makeup. Normally fibers are

classified in slow (type I) vs. fast (type II) with three subgroups of type II (IIA, IID or X and IIB). However, many mixed forms exist, and with novel techniques more and more subclasses might be revealed. I did not discern the muscles on single fiber level as was done by Murgia, *et al.*, 2015 [204], which is also interesting and suggests, that there might even be differences between and along fibers that would be classified into one of the four classic types mentioned, since the molecular makeup is highly adaptable and changes during training or in disease are perhaps even more subtle than can be detected by MYH isoform.

This is interesting for basic research, but in a clinical focus, the phenotype of muscle as a whole has to be addressed.

Dysferlinopathy starts to manifest during puberty, when glycolysis becomes a more important source of energy. [8, 70–73] The disease phenotype of dysferlinopathy is much more pronounced in glycolytic muscles than in oxidative muscles in animals [74] and a decrease in type II fibers is observed in humans. [61, 241]

This is attributed not only to a loss of type II fibers, but also a switch of fiber type towards type I fibers. [61]

With the data presented here, I could show that the point of regulatory impact in glycolysis is at the level of hexokinase. Not only are poolsizes of metabolites further downstream the glycolysis decreased and the polyol pathway increased, but also the formation rate of the glycolytic metabolites is altered in dysferlinopathy. The differences in poolsizes of glycolysis and TCA-cycle had also been found by Sarah Keller [1], on whose work this experiment is primarily based. I could additionally show, that there are differences in the poolsizes of PPP and TCA-cycle, interestingly also in 2-hydroxyglutarate (2OHG), a marker for oxidative stress, that also has a regulatory effect on glycolysis. [211, 214–216]

The change in poolsize can have many reasons since metabolites are normally substrates and products at the same time. pSIRM allows to infer the speed at which these metabolites are formed. I could show, that the rate at which glycolytic intermediates and glycogen are formed is altered in dysferlinopathy. Starting from glucose-6-phosphate (G6P) through lactate, label incorporation is slower in glycolysis, while it is not altered in the polyol pathway.

Dysferlinopathy seems to be caused by two different changes in the cell: on the one hand, membrane repair is slowed [67, 68], but also glycolysis, in fibers that are very reliant on this energy source is slowed down. Not only cell repair itself is compromised, but also energy for overall proper function is low.

The delineation of dysferlinopathy as metabolic disease, was only possible, because biop-

sies from patients had been taken, which always means to further traumatize a potentially diseased tissue of a patient.

If we could infer metabolism also from more easily obtainable material like blood, this would mean a great deal for clinical investigations.

In the study of the changes of the blood-metabolome of a single human volunteer, we were able to find known interrelations, but through the systemic and systems biology based approach of our technique, we could also find a possible explanation for the phenomenon of "Hitting the Wall". We were also surprised to find, that the state of self reported rate of perceived exertion (RPE) [195] is also highly reflected in the metabolome. This reflection is also based on many metabolites, that were not included in databases used for annotation and are thus "unknown". The revelations that can be expected, when databases will grow more and more in the future are surely exciting.

Except for the study on human blood, where we included the free fatty acids (FFAs), I only regarded those metabolites that are classified as "polar" (e. g. sugars, sugar-phosphates, amino-acids, carboxylic acids). Especially since dysferlin seems also to have an influence on lipid droplets [190] and because of its general involvement in membranous exchanges, analyzing lipid composition from the samples obtained is surely of interest for the future. Several interesting tendencies in the profiles of FFAs were observed, as periodic oscillation or distinct separation between the early and late phase of relase (figure 3.5.5). General pathways of lipid formation are known and for some species also some of their influences on biological systems, but overall lipidomics is a relatively young field and knowledge is mostly based on comparison of abundance in comparison to treatment. There are efforts on the way to aggregate this knowledge [242–244] but because of lack of general knowledge, and only one technical replicate of measurements is available we did not dare to interpret the possible mechanisms behind these patterns.

Another point to investigate will be the need for and kind of challenges that provoke the emergence of certain metabolic phenotypes. The power of our model of $n = 1$ studies is based on the absence of need for a control group, if challenges are applied, or if lifelong monitoring at regular intervals is used to monitor overall health status. Because our technique causes minimal harm, a combination of both is conceivable: a challenge regimen every few months, where frequent samples are obtained while external stimuli are applied. This would not only allow for high resolution in time, as is needed in metabolomics, but also the individual investigated served as its own, best matched, control.

5 Conclusions–Towards a more holistic medicine

Regarding the results from the experiments presented above, it can be said, that certain conclusions might only be drawn from the completeness of the data, a systems biology approach offers.

Only thanks to a metabolomic and proteomic dataset obtained from whole muscle of the same animals, I could unambiguously show that metabolite levels do not necessarily reflect protein levels and vice versa, similar to the relation between transcriptome and proteome [107]. This might be trivial, but it will be crucial to understand which direct relations between metabolites and proteins exist, since both regulate each other. Of course, there are other layers involved, still not discussed here, as hormonal regulation [245] or general cell signaling through phosphorylation cascades, but for this kind of data systems- or omics techniques are just starting to mature to a level, that allows for a complete integration. Also for metabolomics, let alone its integration with proteomics data, data analysis pipelines are far less standardized than for transcriptomic or proteomic data and their integration.

However interpretability of metabolomics data might even be superior to that of proteomics data, since many pathways of the more abundant (polar) metabolites are available, whereas, except for metabolic enzymes, for proteins in general only few established pathways exist. These are more aggregations of interaction partners, but general concepts are hard to discern. GO-terms, of course, provide a feasible classification of the proteins found in a certain subset of data, but this allows seldom for a direct inference of knowledge, i. e. the direction of regulation is not given by the simple membership of a protein to a GO-term.

This lack of interconnectedness of knowledge also is a hindrance in analyzing lipid species, because very little knowledge on regulation, influences or even pathways exists.

In the study of metabolites in blood during exercise, we were not only able to find known mechanisms, that had been established through decades of scientific work on cells, animals and large cohort studies, but through the systemic approach of our technique, we could also find a possible explanation for the phenomenon of “Hitting the Wall”.

Of course, we would not have been able to understand these interrelations if their building blocks, found by decades of scientific work on cells, animals and large cohort studies would not be available in the literature. But it seems that we are now able to interconnect this fragmented knowledge generated by ever more detailed fragmentation of biological systems to a more holistic systems understanding; especially in metabolism, which reflects

not only nature from the long standing predicament of nature vs nurture [236] but also the daily, or even minute-by-second influences on and in the human body opening the field to an even deeper understanding of those influences even in regards to the microbiome.

As often in the history of biology, having new techniques at our disposal, taking a new look grants new and more differentiated insights. Naturally the current classification of fiber types is valid, useful and merited, but at the end physiological makeup escapes from classification with increase in detail resolution offered by current mass spectrometry based techniques.

I only interpreted the results obtained from studying different muscle fiber types in the light of dysferlinopathy and its connection to glycolysis, but there is no telling in what insights an application of these techniques to patient tissues or regarding other diseases might yield. Also, as with every systems biology approach, data mining depends a lot on the interest of the investigator. With omics techniques, many different questions can be asked to the same dataset. This can lead to over interpretation of results, but for hypothesis generation, these techniques and the integration of the data is of unprecedented worth.

We were not only able to pinpoint the disturbance in metabolism of dysferlinopathy, but also, through further analysis of compartments of the cell formed *by* dysferlin, to give a possible mode of action for our observations. Also the data in the literature corroborates this, but the conclusions we found could not have been drawn from proteomics data alone and were not discussed thus far.

This again highlights the power of integration of multiple omics techniques, and I am looking forward as to where future advancements in technology and data integration might lead. In our proof of principle study of human exercise, we were able to delineate the sample relations to nine key metabolites. If this turns out to be a general applicable strategy, such a reduced dataset could be easily communicated to clinicians.

In order to make this advancement in technology feasible and accessible to medical personnel and scientists from other fields, a lot of development is still needed, especially regarding data analysis.

But also standardized sample handling is a major issue, especially in metabolomics. Since metabolites are often highly reactive, data quality hinges paramount on proper pre-analysis handling.

Because of the chemical diversity of metabolite classes, e. g. sugars, amino-acids, FFAs, lipids or nucleotide species, sample handling in metabolomics is quite laborious. Unlike DNA, RNA or proteins metabolites can not be precipitated to separate them from other components of a sample, but a liquid-liquid extraction is needed.

To truly bring the analysis of blood samples from the lab into the clinic, the aforementioned standardization is indispensible and also, if high frequent sampling techniques are applied the number of samples will surely increase dramatically compared to a research setting.

As a first step toward this goal, in collaboration with an industry partner, I was able to design a robot, that not only is capable of sample handling, but also can prepare the laborious control mixtures, that are used to verify identification of metabolites as well ensure comparable quantification over long periods of time.

"Impossible!" said the present.
–"Try it...", whispered the future.

6 Publications

Publication of the integrated metabolome and proteome data of mouse muscles will be pursued and the manuscript soon ready for publication.

* * *

The manuscript regarding dysferlins new role in glucose metabolism has recently been submitted in the following form:

Dysferlin guards glycolysis in skeletal muscle
Verena Schoewel*, **Tobias Opialla***, Séverine Kunz*, Sarah Keller*, Anja Maehler, Fabio E. Rojas Rusak, Helena Escobar, Eric Metzler, Matthias Pietzke, Christian Herrmann, Michael Boschmann, Guido Mastrobuoni, Andreas L. Birkenfeld, Hadi Al-Hasani, Mathias Treier, Stefan Kempa*, Simone Spuler*
**contributed equally*

* * *

The manuscript of the proof of principle study about reflection of muscle metabolism in the blood during exercise is very close to be ready for submission and will be sent out within a few days after submission of this thesis.

References

[1] Keller S, *GC/MS- and LC/MS-based metabolic and proteomic analysis of dysferlin-deficient muscle from patients and animal models*. Ph.D. thesis, Freie Universität Berlin (2014).

[2] Pietzke M *et al.*, Decoding the dynamics of cellular metabolism and the action of 3-bromopyruvate and 2-deoxyglucose using pulsed stable isotope-resolved metabolomics. *Cancer Metab* **2**, 9 (2014).

[3] Kunz S, *Characterization of the sarcolemma in limb-girdle muscular dystrophy*. Ph.D. thesis, Freie Universität Berlin (2015).

[4] Kuich JL HP, *Towards understanding organ-derived metabolic signatures in blood*. Ph.D. thesis, Freie Universität Berlin (2014).

[5] Birnbaum MJ, Skeletal muscle metabolism. In *Muscle*, (edited by Hill J) (Academic Press, Boston/Waltham, 2012), chap. 59, pages 841–853.

[6] Pearson H, Meet the human metabolome. *Nature* **446**, 8–8 (2007).

[7] Schulze A & Harris AL, How cancer metabolism is tuned for proliferation and vulnerable to disruption. *Nature* **491**, 364–373 (2012).

[8] Armstrong N & Barker AR, Oxygen uptake kinetics in children and adolescents: a review. *Pediatr Exerc Sci* **21**, 130–147 (2009).

[9] Bar-Even A *et al.*, Rethinking glycolysis: on the biochemical logic of metabolic pathways. *Nat Chem Biol* **8**, 509–517 (2012).

[10] Berg JM SLB Tymoczko JL, *The Glycolytic Pathway Is Tightly Controlled* (W H Freeman, 2002), chap. 16.2. 5 edn.

[11] Rapoport S & Luebering J, Glycerate-2,3-diphosphatase. *Journal of Biological Chemistry* **189**, 683–694 (1951).

[12] Benesch R & Benesch RE, The effect of organic phosphates from the human erythrocyte on the allosteric properties of hemoglobin. *Biochem Biophys Res Commun* **26**, 162–167 (1967).

[13] Chiba H & Sasaki R, Functions, of 2,3-bisphosphoglycerate and its metabolism. *Curr Top Cell Regul* **14**, 75–116 (1978).

[14] Lorenzi M, The polyol pathway as a mechanism for diabetic retinopathy: attractive, elusive, and resilient. *Exp Diabetes Res* **2007**, 61038 (2007).

[15] Krebs HA & Weitzman P, Krebs' citric acid cycle: half a century and still turning. Biochemical Society, London (1987).

[16] Nobel Media, Albert Szent-Györgyi – Biographical. Nobelprize.org (2014).

[17] Knoop F, Der abbau aromatischer fettsäuren im tierkaörper. *Beitr Chem Physiol Pathol* **6**, 150–162 (1904).

[18] Houten SM & Wanders RJA, A general introduction to the biochemistry of mitochondrial fatty acid β-oxidation. *Journal of Inherited Metabolic Disease* **33**, 469–477 (2010).

[19] Holmes FL, Hans Krebs and the discovery of the ornithine cycle. *Fed Proc* **39**, 216–225 (1980).

[20] Warburg O & Christian W, Über aktivierung der Robinsonschen Hexose-monophosphorsäure in roten Blutzellen und die Gewinnung aktivierender Fermentlösungen. *Biochem Z* **242**, 206–227 (1931).

[21] Horecker BL, The pentose phosphate pathway. *J Biol Chem* **277**, 47965–47971 (2002).

[22] Hopkins FG & Dixon M, On glutathione: II. A thermostabile oxidataion-reduction system. *Journal of Biological Chemistry* **54**, 527–563 (1922).

[23] Simoni RD, Hill RL & Vaughan M, The discovery of glutathione by f. gowland hopkins and the beginning of biochemistry at cambridge university. *Journal of Biological Chemistry* **277**, e13 (2002).

[24] Andres R, Cader G & Zierler KL, The quantitatively minor role of carbohydrate in oxidative metabolism by skeletal muscle in intact man in the basal state. measurements of oxygen and glucose uptake and carbon dioxide and lactate production in the forearm. *J Clin Invest* **35**, 671–682 (1956).

[25] Lewis S & Gutin B, Nutrition and endurance. *Am J Clin Nutr* **26**, 1011–1014 (1973).

[26] Eggleton P & Eggleton GP, The inorganic phosphate and a labile form of organic phosphate in the gastrocnemius of the frog. *Biochem J* **21**, 190–195 (1927).

[27] Fiske CH & Subbarow Y, The nature of the "inorganic phosphate" in voluntary muscle. *Science* **65**, 401–403 (1927).

[28] Esbjörnsson-Liljedahl M *et al.*, Metabolic response in type i and type ii muscle fibers during a 30-s cycle sprint in men and women. *J Appl Physiol* **87**, 1326–1332 (1999).

[29] Kushmerick MJ & Conley KE, Energetics of muscle contraction: the whole is less than the sum of its parts. *Biochemical Society Symposia* **675**, 227–231 (2002).

[30] De Feo P *et al.*, Metabolic response to exercise. *J Endocrinol Invest* **26**, 851–854 (2003).

[31] Karlsson J, Diamant B & Saltin B, Muscle metabolites during submaximal and maximal exercise in man. *Scand J Clin Lab Invest* **26**, 385–394 (1970).

[32] Dalrymple RH & Hamm R, A method for the extraction of glycogen and metabolites from a single muscle sample. *International Journal of Food Science & Technology* **8**, 439–444 (1973).

[33] Murat JC & Serfaty A, Simple enzymatic determination of polysaccharide (glycogen) content of animal tissues. *Clin Chem* **20**, 1576–1577 (1974).

[34] Passonneau JV & Lauderdale VR, A comparison of three methods of glycogen measurement in tissues. *Anal Biochem* **60**, 405–412 (1974).

[35] Fabiansson S & Reuterswärd A, Glycogen determination in post-mortem beef muscles. *Food Chemistry* **15**, 269–284 (1984).

[36] Bernard CM, Remarques sur la formation de la matière glycogène du foie. *Comptes rendus hebdomadaires des séances de l'Académie des sciences (Paris), Cr. hebd. Seanc. Acad. Sci., Paris* **44**, 1325–1331 (1857).

[37] Shulman GI *et al.*, Quantitation of muscle glycogen synthesis in normal subjects and subjects with non-insulin-dependent diabetes by 13C nuclear magnetic resonance spectroscopy. *N Engl J Med* **322**, 223–228 (1990).

[38] Surholt B & Newsholme EA, Maximum activities and properties of glucose 6-phosphatase in muscles from vertebrates and invertebrates. *Biochem J* **198**, 621–629 (1981).

[39] van Schaftingen E & Gerin I, The glucose-6-phosphatase system. *Biochem J* **362**, 513–532 (2002).

[40] van der Vies J, Two methods for the determination of glycogen in liver. *Biochem J* **57**, 410–416 (1954).

[41] van Handel E, Estimation of glycogen in small amounts of tissue. *Anal Biochem* **11**, 256–265 (1965).

[42] Roe JH & Dailey RE, Determination of glycogen with the anthrone reagent. *Anal Biochem* **15**, 245–250 (1966).

[43] Alonso MD *et al.*, A new look at the biogenesis of glycogen. *FASEB J* **9**, 1126–1137 (1995).

[44] Schiaffino S & Reggiani C, Skeletal muscle fiber types. In *Muscle*, (edited by Hill J) (Academic Press, Boston/Waltham, 2012), chap. 60, pages 855–867.

[45] Hämäläinen N & Pette D, The histochemical profiles of fast fiber types IIB, IID, and IIA in skeletal muscles of mouse, rat, and rabbit. *J Histochem Cytochem* **41**, 733–743 (1993).

[46] Augusto V, Padovani C & GER C, Skeletal muscle fiber types in C57BL6J mice. *Braz. J. morphol. Sci.* **21**, 89–94 (2004).

[47] Bassel-Duby R & Olson EN, Signaling pathways in skeletal muscle remodeling. *Annu Rev Biochem* **75**, 19–37 (2006).

[48] Gollnick PD *et al.*, Selective glycogen depletion in skeletal muscle fibres of man following sustained contractions. *J Physiol* **241**, 59–67 (1974).

[49] Hultman EH, Carbohydrate metabolism during hard exercise and in the recovery period after exercise. *Acta Physiol Scand Suppl* **556**, 75–82 (1986).

[50] Manabe Y *et al.*, Exercise training-induced adaptations associated with increases in skeletal muscle glycogen content. *FEBS Journal* **280**, 916–926 (2013).

[51] Huxley H & Hanson J, Changes in the cross-striations of muscle during contraction and stretch and their structural interpretation. *Nature* **173**, 973–976 (1954).

[52] Huxley AF & Niedergerke R, Structural changes in muscle during contraction: Interference microscopy of living muscle fibres. *Nature* **173**, 971–973 (1954).

[53] Coffey VG & Hawley JA, Training for performance: insights from molecular biology. *Int J Sports Physiol Perform* **1**, 284–292 (2006).

[54] Coffey VG & Hawley JA, The molecular bases of training adaptation. *Sports Med* **37**, 737–763 (2007).

[55] Kohn TA, Essén-Gustavsson B & Myburgh KH, Specific muscle adaptations in type ii fibers after high-intensity interval training of well-trained runners. *Scand J Med Sci Sports* **21**, 765–772 (2011).

[56] Elliot DL *et al.*, Metabolic myopathies: Evaluation by graded exercise testing. *Medicine* **68**, 163–172 (1989).

[57] Wortmann RL, Metabolic and mitochondrial myopathies. *Current Opinion in Rheumatology* **1**, 427–431 (1989).

[58] Wortmann RL, Metabolic myopathies. *Curr Opin Rheumatol* **3**, 925–933 (1991).

[59] Lammens M & Schoser B, Metabolic myopathies – an overview. *Pathologe* **30**, 370–378 (2009).

[60] Vorgerd M & Deschauer M, Metabolic and mitochondrial myopathies. *Z Rheumatol* **72**, 242–254 (2013).

[61] De la Torre C *et al.*, Proteomics identification of differentially expressed proteins in the muscle of dysferlin myopathy patients. *Proteomics Clin Appl* **3**, 486–497 (2009).

[62] Liu J *et al.*, Dysferlin, a novel skeletal muscle gene, is mutated in miyoshi myopathy and limb girdle muscular dystrophy. *Nat Genet* **20**, 31–36 (1998).

[63] Bashir R *et al.*, A gene related to Caenorhabditis elegans spermatogenesis factor fer-1 is mutated in limb-girdle muscular dystrophy type 2B. *Nat Genet* **20**, 37–42 (1998).

[64] Argov Z *et al.*, Muscular dystrophy due to dysferlin deficiency in Libyan Jews. Clinical and genetic features. *Brain* **123 (Pt 6)**, 1229–1237 (2000).

[65] Angelini C *et al.*, Dysferlinopathy course and sportive activity: clues for possible treatment. *Acta Myol* **30**, 127–132 (2011).

[66] Nguyen K *et al.*, Phenotypic study in 40 patients with dysferlin gene mutations: high frequency of atypical phenotypes. *Arch Neurol* **64**, 1176–1182 (2007).

[67] Marg A *et al.*, Sarcolemmal repair is a slow process and includes EHD2. *Traffic* **13**, 1286–1294 (2012).

[68] Schoewel V *et al.*, Dysferlin-peptides reallocate mutated dysferlin thereby restoring function. *PLoS One* **7**, e49603 (2012).

[69] Eriksson BO, Muscle metabolism in children–a review. *Acta Paediatr Scand Suppl* **283**, 20–28 (1980).

[70] Taylor DJ *et al.*, Ageing: effects on oxidative function of skeletal muscle in vivo. *Mol Cell Biochem* **174**, 321–324 (1997).

[71] Timmons BW, Bar-Or O & Riddell MC, Oxidation rate of exogenous carbohydrate during exercise is higher in boys than in men. *J Appl Physiol (1985)* **94**, 278–284 (2003).

[72] Stephens BR, Cole AS & Mahon AD, The influence of biological maturation on fat and carbohydrate metabolism during exercise in males. *Int J Sport Nutr Exerc Metab* **16**, 166–179 (2006).

[73] Timmons BW, Bar-Or O & Riddell MC, Influence of age and pubertal status on substrate utilization during exercise with and without carbohydrate intake in healthy boys. *Appl Physiol Nutr Metab* **32**, 416–425 (2007).

[74] Lostal W *et al.*, Efficient recovery of dysferlin deficiency by dual adeno-associated vector-mediated gene transfer. *Hum Mol Genet* **19**, 1897–1907 (2010).

[75] Pietzke M & Kempa S, Pulsed stable isotope-resolved metabolomic studies of cancer cells. *Methods Enzymol* **543**, 179–198 (2014).

[76] Kempa S, Pietzke M & Zasada C, Short term isotope pulse labeling method for analysing metabolic products in biological samples (2015). US Patent 20,150,330,969.

[77] Kempa S, Pietzke M & Zasada C, Short term isotope pulse labeling method for analysing metabolic products in biological samples (2015 G01N33/68 EP2944963).

[78] Nemutlu E *et al.*, ^{18}O-assisted dynamic metabolomics for individualized diagnostics and treatment of human diseases. *Croat Med J* **53**, 529–534 (2012).

[79] Fiehn O, Metabolomics–the link between genotypes and phenotypes. *Plant Mol Biol* **48**, 155–171 (2002).

[80] Trujillo E, Davis C & Milner J, Nutrigenomics, proteomics, metabolomics, and the practice of dietetics. *J Am Diet Assoc* **106**, 403–413 (2006).

[81] Dettmer K, Aronov PA & Hammock BD, Mass spectrometry-based metabolomics. *Mass Spectrom. Rev.* **26**, 51–78 (2007).

[82] Létisse F, Metabolomics and fluxomics: Tools for the functional analysis of metabolic systems. www (2008).

[83] Zamboni N & Sauer U, Novel biological insights through metabolomics and 13C-flux analysis. *Curr Opin Microbiol* **12**, 553–558 (2009).

[84] Yerlekar A & Kshirsagar M, A review on mass spectrometry: Technique and tools. *Int. Journal of Engineering Research and Applications* **4**, 12–23 (2014).

[85] Makarov A, Electrostatic axially harmonic orbital trapping: A high-performance technique of mass analysis. *Anal. Chem.* **72**, 1156–1162 (2000).

[86] Hu Q *et al.*, The orbitrap: a new mass spectrometer. *J Mass Spectrom* **40**, 430–443 (2005).

[87] Skoog D *et al.*, *Principles of Instrumental Analysis*. International student edition (Thomson Brooks/Cole, 2007).

[88] Cox J *et al.*, A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. *Nat. Protocols* **4**, 698–705 (2009).

[89] Nesvizhskii AI & Aebersold R, Interpretation of shotgun proteomic data the protein inference problem. *Molecular & Cellular Proteomics* **4**, 1419–1440 (2005).

[90] Nesvizhskii AI *et al.*, A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* **75**, 4646–4658 (2003).

[91] Cox J & Mann M, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotech* **26**, 1367–1372 (2008).

[92] Cox J *et al.*, Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res* **10**, 1794–1805 (2011).

[93] Lisec J *et al.*, Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nat Protoc* **1**, 387–396 (2006).

[94] Stanhope KL *et al.*, Consuming fructose-sweetened, not glucose-sweetened, beverages increases visceral adiposity and lipids and decreases insulin sensitivity in overweight/obese humans. *J Clin Invest* **119**, 1322–1334 (2009).

[95] Dunn WB *et al.*, Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chem Soc Rev* **40**, 387–426 (2011).

[96] DiNicolantonio JJ, O'Keefe JH & Lucan SC, Added fructose: a principal driver of type 2 diabetes mellitus and its consequences. *Mayo Clin Proc* **90**, 372–381 (2015).

[97] Koek MM *et al.*, Quantitative metabolomics based on gas chromatography mass spectrometry: status and perspectives. *Metabolomics* **7**, 307–328 (2011).

[98] Kuich PHJL, Hoffmann N & Kempa S, Maui-VIA: A user-friendly software for visual identification, alignment, correction, and quantification of gas chromatography-mass spectrometry data. *Front Bioeng Biotechnol* **2**, 84 (2014).

[99] Hummel J *et al.*, Decision tree supported substructure prediction of metabolites from GC–MS profiles. *Metabolomics* **6**, 322–333 (2010).

[100] Ohlendieck K, Proteomics of skeletal muscle glycolysis. *Biochim Biophys Acta* **1804**, 2089–2101 (2010).

[101] Kanehisa M *et al.*, KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* **44**, D457–D462 (2016).

[102] Fabregat A *et al.*, The reactome pathway knowledgebase. *Nucleic Acids Res* **44**, D481–D487 (2016).

[103] Consortium GO, Gene ontology consortium: going forward. *Nucleic Acids Res* **43**, D1049–D1056 (2015).

[104] Ashburner M *et al.*, Gene ontology: tool for the unification of biology. the gene ontology consortium. *Nat Genet* **25**, 25–29 (2000).

[105] Fisher RA, On the interpretation of χ^2 from contingency tables, and the calculation of P. *Journal of the Royal Statistical Society* **85**, 87–94 (1922).

[106] Consortium U & Bateman A, Uniprot: a hub for protein information. *Nucleic Acids Res* **43**, D204–D212 (2015).

[107] Schwanhäusser B *et al.*, Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).

[108] Buller AJ, Eccles JC & Eccles RM, Interactions between motoneurones and muscles in respect of the characteristic speeds of their responses. *The Journal of Physiology* **150**, 417–439 (1960).

[109] Gollnick PD *et al.*, Enzyme activity and fiber composition in skeletal muscle of untrained and trained men. *Journal of Applied Physiology* **33**, 312–319 (1972).

[110] Gollnick PD *et al.*, Effect of training on enzyme activity and fiber composition of human skeletal muscle. *Journal of Applied Physiology* **34**, 107–111 (1973).

[111] Drexler HCA *et al.*, On marathons and sprints: an integrated quantitative proteomics and transcriptomics analysis of differences between slow and fast muscle fibers. *Mol Cell Proteomics* **11**, M111.010801 (2012).

[112] Rakus D *et al.*, Absolute quantitative profiling of the key metabolic pathways in slow and fast skeletal muscle. *J Proteome Res* **14**, 1400–1411 (2015).

[113] Deshmukh AS *et al.*, Deep proteomics of mouse skeletal muscle enables quantitation of protein isoforms, metabolic pathways, and transcription factors. *Mol Cell Proteomics* **14**, 841–853 (2015).

[114] Seifter S & Dayton S, The estimation of glycogen with the anthrone reagent. *Arch Biochem* **25**, 191–200 (1950).

[115] Fong J, Schaffer FL & Kirk PL, The ultramicrodetermination of glycogen in liver. a comparison of the anthrone and reducing-sugar methods. *Archives of Biochemistry and Biophysics* **45**, 319–326 (1953).

[116] Carroll NV, Longley RW & Roe JH, The determination of glycogen in liver and muscle by use of anthrone reagent. *J Biol Chem* **220**, 583–593 (1956).

[117] Rousset M, Zweibaum A & Fogh J, Presence of glycogen and growth-related variations in 58 cultured human tumor cell lines of various tissue origins. *Cancer Res* **41**, 1165–1170 (1981).

[118] Roelandt P *et al.*, Differentiation of rat multipotent adult progenitor cells to functional hepatocyte-like cells by mimicking embryonic liver development. *Nat Protoc* **5**, 1324–1336 (2010).

[119] Philippi S *et al.*, Dysferlin-deficient immortalized human myoblasts and myotubes as a useful tool to study dysferlinopathy. *PLoS Curr* **4**, RRN1298 (2012).

[120] Alves LA, Almeida e Silva JaB & Giulietti M, Solubility of d-glucose in water and ethanol/water mixtures. *Journal of Chemical & Engineering Data* **52**, 2166–2170 (2007).

[121] Kempa S *et al.*, A plastid-localized glycogen synthase kinase 3 modulates stress tolerance and carbohydrate metabolism. *Plant J* **49**, 1076–1090 (2007).

[122] Rappsilber J, Ishihama Y & Mann M, Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and lc/ms sample pretreatment in proteomics. *Anal Chem* **75**, 663–670 (2003).

[123] The UniProt Consortium, Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Research* **40**, D71–D75 (2012).

[124] Cox J *et al.*, Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed maxlfq. *Mol Cell Proteomics* **13**, 2513–2526 (2014).

[125] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2016).

[126] Jombart T, adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405 (2008).

[127] Jombart T & Ahmed I, adegenet 1.3-1: new tools for the analysis of genome-wide snp data. *Bioinformatics* **27**, 3070–3071 (2011).

[128] Huber *et al.*, Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods* **12**, 115–121 (2015).

[129] Tenenbaum D, *KEGGREST: Client-side REST access to KEGG* (2016). R package version 1.12.2.

[130] Zhang JD & Wiemann S, KEGGgraph: a graph approach to KEGG PATHWAY in R and bioconductor. *Bioinformatics* **25**, 1470–1471 (2009).

[131] Neuwirth E, *RColorBrewer: ColorBrewer palettes* (2014). R package version 1.1-2.

[132] Eddelbuettel D & Francois R, Rcpp: Seamless R and C++ integration. *Journal of Statistical Software* **40**, 1–18 (2011).

[133] Bivand R, *classInt: Choose Univariate Class Intervals* (2015). R package version 0.1-23.

[134] Højsgaard S & Halekoh U, *doBy: Groupwise Statistics, LSmeans, Linear Contrasts, Utilities* (2016). R package version 4.5-15.

[135] Chang W, *extrafont: Tools for using fonts* (2014). R package version 0.17.

[136] Wickham H, *ggplot2: elegant graphics for data analysis* (Springer New York, 2009).

[137] Slowikowski K, *ggrepel: Repulsive Text and Label Geoms for 'ggplot2'* (2016). R package version 0.5.

[138] Auguie B, *gridExtra: Miscellaneous Functions for "Grid" Graphics* (2016). R package version 2.2.1.

[139] Ritchie ME *et al.*, limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* **43**, e47 (2015).

[140] Wickham H, The split-apply-combine strategy for data analysis. *Journal of Statistical Software* **40**, 1–29 (2011).

[141] Revelle W, *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois (2016). R package version 1.6.6.

[142] Wickham H, Reshaping data with the reshape package. *Journal of Statistical Software* **21**, 1–20 (2007).

[143] Wickham H, *scales: Scale Functions for Visualization* (2016). R package version 0.4.0.

[144] Wickham H, *stringr: Simple, Consistent Wrappers for Common String Operations* (2015). R package version 1.0.0.

[145] Wickham H *et al.*, *svglite: An 'SVG' Graphics Device* (2016). R package version 1.1.0.

[146] Tennekes M, *treemap: Treemap Visualization* (2016). R package version 2.4-1.

[147] Bielow C, Mastrobuoni G & Kempa S, Proteomics quality control: Quality control software for maxquant results. *J Proteome Res* **15**, 777–787 (2016).

[148] Tyanova S *et al.*, The Perseus computational platform for comprehensive analysis of (prote) omics data. *Nature Methods* (2016).

[149] Shannon P *et al.*, Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498–2504 (2003).

[150] Rohn H *et al.*, VANTED v2: a framework for systems biology applications. *BMC Systems Biology* **6**, 1–13 (2012).

[151] Krzywinski M *et al.*, Circos: an information aesthetic for comparative genomics. *Genome Res* **19**, 1639–1645 (2009).

[152] Wiki I, Frequently asked questions — inkscape wiki, (2015). [Online; accessed 2-August-2016].

[153] Close R, Properties of motor units in fast and slow skeletal muscles of the rat. *J Physiol* **193**, 45–55 (1967).

[154] Schiaffino S & Reggiani C, Fiber types in mammalian skeletal muscles. *Physiol Rev* **91**, 1447–1531 (2011).

[155] Wu H *et al.*, High-throughput tissue extraction protocol for NMR- and MS-based metabolomics. *Anal Biochem* **372**, 204–212 (2008).

[156] Dubacq S, Performing efficient sample preparation with hard tumor tissue: Precellys[reg] bead-beating homogenizer solution. *Nat Meth* **13**, – (2016).

[157] Blank JA *et al.*, Procedure for assessing myeloperoxidase and inflammatory mediator responses in hairless mouse skin. *J Appl Toxicol* **20 Suppl 1**, S137–S139 (2000).

[158] Ordway GA & Garry DJ, Myoglobin: an essential hemoprotein in striated muscle. *J Exp Biol* **207**, 3441–3446 (2004).

[159] Schoewel V, Histo_MMex38_60weeks_140716. Unpublished figure, manuscript in preparation.

[160] Matsuura T *et al.*, Skeletal muscle fiber type conversion during the repair of mouse soleus: potential implications for muscle healing after injury. *J Orthop Res* **25**, 1534–1540 (2007).

[161] Popesko P, *Colour Atlas of Anatomy of Small Laboratory Animals* (Saunders, 2003).

[162] Brooks GA, Mammalian fuel utilization during sustained exercise. *Comp Biochem Physiol B Biochem Mol Biol* **120**, 89–107 (1998).

[163] Lazar C *et al.*, Accounting for the multiple natures of missing values in label-free quantitative proteomics data sets to compare imputation strategies. *J Proteome Res* **15**, 1116–1125 (2016).

[164] Limpert E, Stahel WA & Abbt M, Log-normal distributions across the sciences: Keys and clues: On the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize log-normal distributions, which can provide deeper insight into variability and probability—normal or log-normal: That is the question. *BioScience* **51**, 341–352 (2001).

[165] Pearson K, Liii. on lines and planes of closest fit to systems of points in space. *Philosophical Magazine Series 6* **2**, 559–572 (1901).

[166] Gromski PS *et al.*, A tutorial review: Metabolomics and partial least squares-discriminant analysis–a marriage of convenience or a shotgun wedding. *Anal Chim Acta* **879**, 10–23 (2015).

[167] Otsu N, A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics* **9**, 62–66 (1979).

[168] Kutmon M *et al.*, Wikipathways: capturing the full diversity of pathway knowledge. *Nucleic Acids Res* **44**, D488–D494 (2016).

[169] Jewison T *et al.*, SMPDB 2.0: big improvements to the Small Molecule Pathway Database. *Nucleic Acids Res* **42**, D478–D484 (2014).

[170] Benjamini Y & Hochberg Y, Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)* pages 289–300 (1995).

[171] Kamburov A *et al.*, Integrated pathway-level analysis of transcriptomics and metabolomics data with IMPaLA. *Bioinformatics* **27**, 2917–2918 (2011).

[172] Kamburov A *et al.*, The ConsensusPathDB interaction database: 2013 update. *Nucleic Acids Res* **41**, D793–D800 (2013).

[173] Eden E *et al.*, GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48 (2009).

[174] Supek F *et al.*, REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**, e21800 (2011).

[175] Cox J & Mann M, 1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data. *BMC Bioinformatics* **13 Suppl 16**, S12 (2012).

[176] Kundig W, Ghosh S & Roseman S, Phosphate bound to histidine in a protein as an intermediate in a novel phospho-transferase system. *Proceedings of the National Academy of Sciences of the United States of America* **52**, 1067–1074 (1964).

[177] Deutscher J, Francke C & Postma PW, How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiol Mol Biol Rev* **70**, 939–1031 (2006).

[178] Lengeler JW, Pts 50: Past, present and future, or diauxie revisited. *J Mol Microbiol Biotechnol* **25**, 79–93 (2015).

[179] Weiler T *et al.*, Identical mutation in patients with limb girdle muscular dystrophy type 2B or Miyoshi myopathy suggests a role for modifier gene(s). *Human Molecular Genetics* **8**, 871–877 (1999).

[180] Han R, Muscle membrane repair and inflammatory attack in dysferlinopathy. *Skelet Muscle* **1**, 10 (2011).

[181] Riedel E & Janiak C, *Anorganische Chemie* (De Gruyter, 2002).

[182] Kerly M, The solubility of glycogen. *Biochem J* **24**, 67–76 (1930).

[183] de Morrée A *et al.*, Proteomic analysis of the dysferlin protein complex unveils its importance for sarcolemmal maintenance and integrity. *PLoS One* **5**, e13854 (2010).

[184] Flix B *et al.*, Dysferlin interacts with calsequestrin-1, myomesin-2 and dynein in human skeletal muscle. *Int J Biochem Cell Biol* **45**, 1927–1938 (2013).

[185] Di Fulvio S *et al.*, Dysferlin interacts with histone deacetylase 6 and increases alpha-tubulin acetylation. *PLoS One* **6**, e28563 (2011).

[186] Leung C *et al.*, Proteomic identification of dysferlin-interacting protein complexes in human vascular endothelium. *Biochem Biophys Res Commun* **415**, 263–269 (2011).

[187] van Haagen HHHBM *et al.*, In silico discovery and experimental validation of new protein-protein interactions. *Proteomics* **11**, 843–853 (2011).

[188] Azakir BA *et al.*, Dysferlin interacts with tubulin and microtubules in mouse skeletal muscle. *PLoS One* **5**, e10122 (2010).

[189] Feuk-Lagerstedt E *et al.*, Lipid raft proteome of the human neutrophil azurophil granule. *Proteomics* **7**, 194–205 (2007).

[190] Li L *et al.*, Comparative proteomics reveals abnormal binding of ATGL and dysferlin on lipid droplets from pressure overload-induced dysfunctional rat hearts. *Sci Rep* **6**, 19782 (2016).

[191] Cai C *et al.*, Membrane repair defects in muscular dystrophy are linked to altered interaction between mg53, caveolin-3, and dysferlin. *J Biol Chem* **284**, 15894–15902 (2009).

[192] Matsuda C *et al.*, The sarcolemmal proteins dysferlin and caveolin-3 interact in skeletal muscle. *Human molecular genetics* **10**, 1761–1766 (2001).

[193] Cacciottolo M *et al.*, Reverse engineering gene network identifies new dysferlin-interacting proteins. *J Biol Chem* **286**, 5404–5413 (2011).

[194] Cook ETS, *The life of Florence Nightingale* (Macmillan Publishers Limited, 1913), vol. 1, page 386.

[195] Borg GA, Psychophysical bases of perceived exertion. *Med Sci Sports Exerc* **14**, 377–381 (1982).

[196] Holloszy JO, Kohrt WM & Hansen PA, The regulation of carbohydrate and fat metabolism during and after exercise. *Front Biosci* **3**, D1011–D1027 (1998).

[197] Brooks GA *et al.*, Decreased reliance on lactate during exercise after acclimatization to 4,300 m. *J Appl Physiol (1985)* **71**, 333–341 (1991).

[198] Horowitz JF, Fatty acid mobilization from adipose tissue during exercise. *Trends Endocrinol Metab* **14**, 386–392 (2003).

[199] Risé P *et al.*, Fatty acid composition of plasma, blood cells and whole blood: relevance for the assessment of the fatty acid status in humans. *Prostaglandins Leukot Essent Fatty Acids* **76**, 363–369 (2007).

[200] Rapoport BI, Metabolic factors limiting performance in marathon runners. *PLoS Comput Biol* **6**, e1000960 (2010).

[201] Allen PJ, Creatine metabolism and psychiatric disorders: Does creatine supplementation have therapeutic value? *Neurosci Biobehav Rev* **36**, 1442–1462 (2012).

[202] Ranvier LA, *Leçons d'anatomie générale sur le système musculaire* (Renaut, M. J., Paris, 1880).

[203] Rossi AC *et al.*, Two novel/ancient myosins in mammalian skeletal muscles: MYH14/7b and MYH15 are expressed in extraocular muscles and muscle spindles. *J Physiol* **588**, 353–364 (2010).

[204] Murgia M *et al.*, Single muscle fiber proteomics reveals unexpected mitochondrial specialization. *EMBO Rep* **16**, 387–395 (2015).

[205] Park KA *et al.*, Myosin heavy chain isoform expression in human extraocular muscles: longitudinal variation and patterns of expression in global and orbital layers. *Muscle Nerve* **45**, 713–720 (2012).

[206] Pette D & Staron RS, Transitions of muscle fiber phenotypic profiles. *Histochem Cell Biol* **115**, 359–372 (2001).

[207] Pette D & Staron RS, Myosin isoforms, muscle fiber types, and transitions. *Microsc Res Tech* **50**, 500–509 (2000).

[208] Staron RS & Pette D, The continuum of pure and hybrid myosin heavy chain-based fibre types in rat skeletal muscle. *Histochemistry* **100**, 149–153 (1993).

[209] Michaelis L & Menten ML, Die kinetik der invertinwirkung. *Biochemische Zeitschrift* **49**, 333–369 (1913).

[210] Michaelis L *et al.*, The original Michaelis constant: translation of the 1913 Michaelis-Menten paper. *Biochemistry* **50**, 8264–8269 (2011).

[211] Losman JA & Kaelin WG Jr, What a difference a hydroxyl makes: mutant IDH, (R)-2-hydroxyglutarate, and cancer. *Genes Dev* **27**, 836–852 (2013).

[212] Leonardi R *et al.*, Cancer-associated isocitrate dehydrogenase mutations inactivate NADPH-dependent reductive carboxylation. *J Biol Chem* **287**, 14615–14620 (2012).

[213] Lu C *et al.*, IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature* **483**, 474–478 (2012).

[214] Harris AL, A new hydroxy metabolite of 2-oxoglutarate regulates metabolism in hypoxia. *Cell Metab* **22**, 198–200 (2015).

[215] Intlekofer AM *et al.*, Hypoxia induces production of L-2-hydroxyglutarate. *Cell Metab* **22**, 304–311 (2015).

[216] Oldham WM *et al.*, Hypoxia-mediated increases in L-2-hydroxyglutarate coordinate the metabolic response to reductive stress. *Cell Metab* **22**, 291–303 (2015).

[217] Grounds MD *et al.*, Lipid accumulation in dysferlin-deficient muscles. *The American Journal of Pathology* **184**, 1668 – 1676 (2014).

[218] Häggström M, Medical gallery of Mikael häggström 2014. *Wikiversity Journal of Medicine* **1** (2014).

[219] Parra J *et al.*, Enhanced catalytic activity of hexokinase by work-induced mitochondrial binding in fast-twitch muscle of rat. *FEBS Lett* **403**, 279–282 (1997).

[220] Vogt C *et al.*, Effects of insulin on subcellular localization of hexokinase ii in human skeletal muscle in vivo. *J Clin Endocrinol Metab* **83**, 230–234 (1998).

[221] Sigel P & Pette D, Intracellular localization of glycogenolytic and glycolytic enzymes in white and red rabbit skeletal muscle: a gel film method for coupled enzyme reactions in histochemistry. *J Histochem Cytochem* **17**, 225–237 (1969).

[222] Kowalski W, Gizak A & Rakus D, Phosphoglycerate mutase in mammalian striated muscles: subcellular localization and binding partners. *FEBS Lett* **583**, 1841–1845 (2009).

[223] Suarez RK, Shaken and stirred: muscle structure and metabolism. *J Exp Biol* **206**, 2021–2029 (2003).

[224] Kinsey ST, Locke BR & Dillaman RM, Molecules in motion: influences of diffusion on metabolic structure and function in skeletal muscle. *J Exp Biol* **214**, 263–274 (2011).

[225] van Wessel T *et al.*, The muscle fiber type-fiber size paradox: hypertrophy or oxidative metabolism? *Eur J Appl Physiol* **110**, 665–694 (2010).

[226] Greene DA *et al.*, Role of sorbitol accumulation and myo-inositol depletion in paranodal swelling of large myelinated nerve fibers in the insulin-deficient spontaneously diabetic bio-breeding rat. Reversal by insulin replacement, an aldose reductase inhibitor, and myo-inositol. *J Clin Invest* **79**, 1479–1485 (1987).

[227] Tang WH, Martin KA & Hwa J, Aldose reductase, oxidative stress, and diabetic mellitus. *Front Pharmacol* **3**, 87 (2012).

[228] Howard AC *et al.*, A novel cellular defect in diabetes: membrane repair failure. *Diabetes* **60**, 3034–3043 (2011).

[229] Hénique C *et al.*, Increasing mitochondrial muscle fatty acid oxidation induces skeletal muscle remodeling toward an oxidative phenotype. *FASEB J* **29**, 2473–2483 (2015).

[230] Olsen JV *et al.*, Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Sci Signal* **3**, ra3 (2010).

[231] Pasquinelli AE, MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. *Nat Rev Genet* **13**, 271–282 (2012).

[232] Glenmark B *et al.*, Muscle strength from adolescence to adulthood–relationship to muscle fibre types. *Eur J Appl Physiol Occup Physiol* **68**, 9–19 (1994).

[233] Thorburn DR & Kuchel PW, Regulation of the human-erythrocyte hexose-monophosphate shunt under conditions of oxidative stress. *European Journal of Biochemistry* **150**, 371–380 (1985).

[234] Bowtell JL & Bruce M, Glutamine: an anaplerotic precursor. *Nutrition* **18**, 222–224 (2002).

[235] Bowtell JL *et al.*, Tricarboxylic acid cycle intermediate pool size: functional importance for oxidative metabolism in exercising human skeletal muscle. *Sports Med* **37**, 1071–1088 (2007).

[236] Ingalls CP, Nature vs. nurture: can exercise really alter fiber type composition in human skeletal muscle? *J Appl Physiol* **97**, 1591–1592 (2004).

[237] Chen R *et al.*, Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell* **148**, 1293–1307 (2012).

[238] Brancaccio P, Lippi G & Maffulli N, Biochemical markers of muscular damage. *Clin Chem Lab Med* **48**, 757–767 (2010).

[239] Ohlendieck K, Proteomic identification of biomarkers of skeletal muscle disorders. *Biomark Med* **7**, 169–186 (2013).

[240] Özyener F *et al.*, Influence of exercise intensity on the on- and off-transient kinetics of pulmonary oxygen uptake in humans. *The Journal of Physiology* **533**, 891–902 (2001).

[241] Campanaro S *et al.*, Gene expression profiling in dysferlinopathies using a dedicated muscle microarray. *Hum Mol Genet* **11**, 3283–3298 (2002).

[242] Fahy E *et al.*, Update of the LIPID MAPS comprehensive classification system for lipids. *J Lipid Res* **50 Suppl**, S9–14 (2009).

[243] Sud M *et al.*, LMSD: LIPID MAPS structure database. *Nucleic Acids Res* **35**, D527–D532 (2007).

[244] The lipid maps lipidomics gateway.

[245] d'Albis A & Butler-Browne G, The hormonal control of myosin isoform expression in skeletal muscle of mammals (a review). *European Journal of Translational Myology (BAM)* **3**, 7–16 (1993).

7 Appendix

7.1 Example of a typical dotplot

7.2 Volcanoplot of the changes in poolsizes depicted in figure 3.3.8

7.3 Treemaps for GO-term enrichment in dysferlin pulldowns with different cutoffs

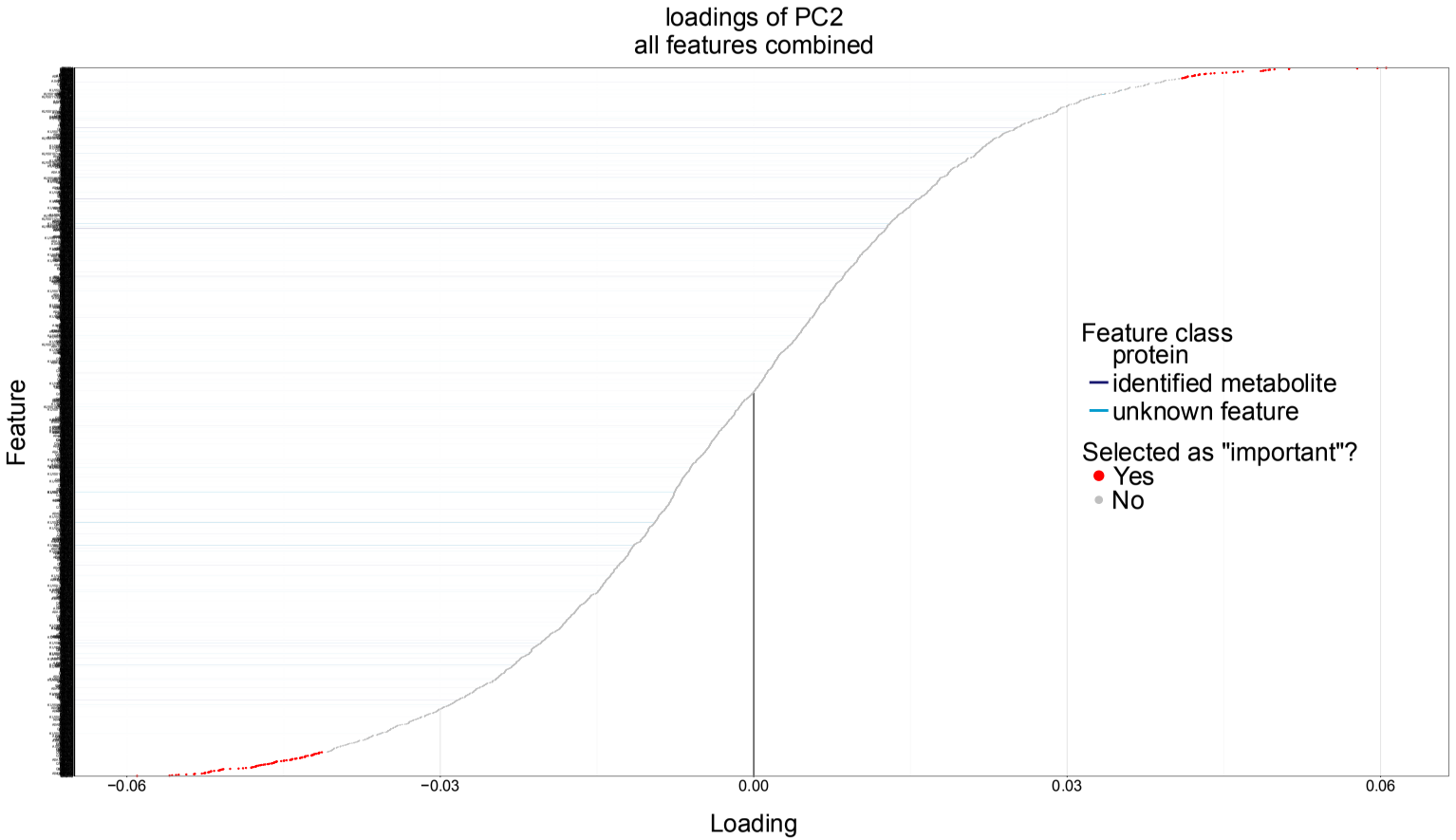


Figure 7.1.1: Exemplary CLEVELAND plot of the loadings of principal component 2, when all features (metablites and proteins) are used in a principal component analysis (PCA). The features deemed as "important" or strongly influencing the separation along the respective component, as determined by hierarchical clustering are marked in red.

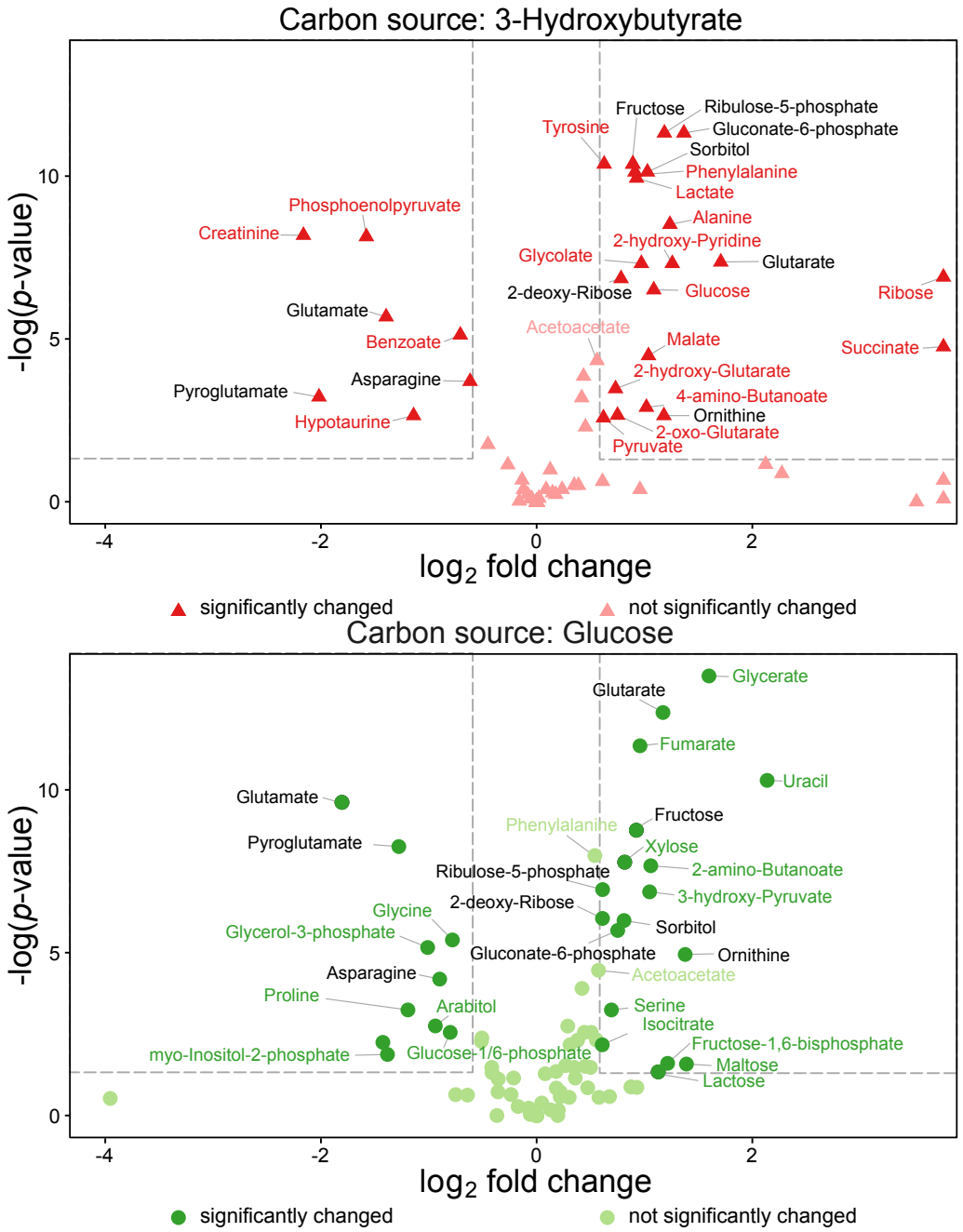


Figure 7.2.1: Volcanoplot of foldchange in poolsize of the respective metabolites and corresponding p-values, normal controls vs. patients with dysferlinopathy. The dashed lines indicate cutoffs at $p < 0.05$ and absolute fold change of > 1.5 .
Myotubes were incubated with $^{13}C_u$-3-hydroxybutyrate (top) or $^{13}C_u$-glucose (bottom). Metabolites found significantly changed in both experiments are labeled with black text. If a metabolite was changed significantly in both treatment conditions, orientation of the regulation stayed the same. As found in the initial experiments by Sarah Keller [1] sorbitol and fructose as members of the polyol pathway are upregulated, while glucose-1/6-phosphate is downregulated. Additionally members of the pentose phosphate pathway (PPP) (xylose, gluconate-6-phosphate, ribulose-5-phosphate, ribose), and the sugars lactose and maltose are also upregulated. Whereas, except fumarate and isocitrate, no intermediates of the TCA-cycle were found significantly changed under glucose conditions, α-ketoglutarate, 2-hydroxy-glutarate, succinate and malate were all found to be enriched in myotubes of dysferlinopathy patients when presented with 3-hydroxybutyrate as carbon source. The upregulation of the PPP is overall more pronounced in 3-hydroxybutyrate conditions. Creatinine which is spontaneously formed from creatine was only measured in the control myotubes, while ornithine is upregulated.

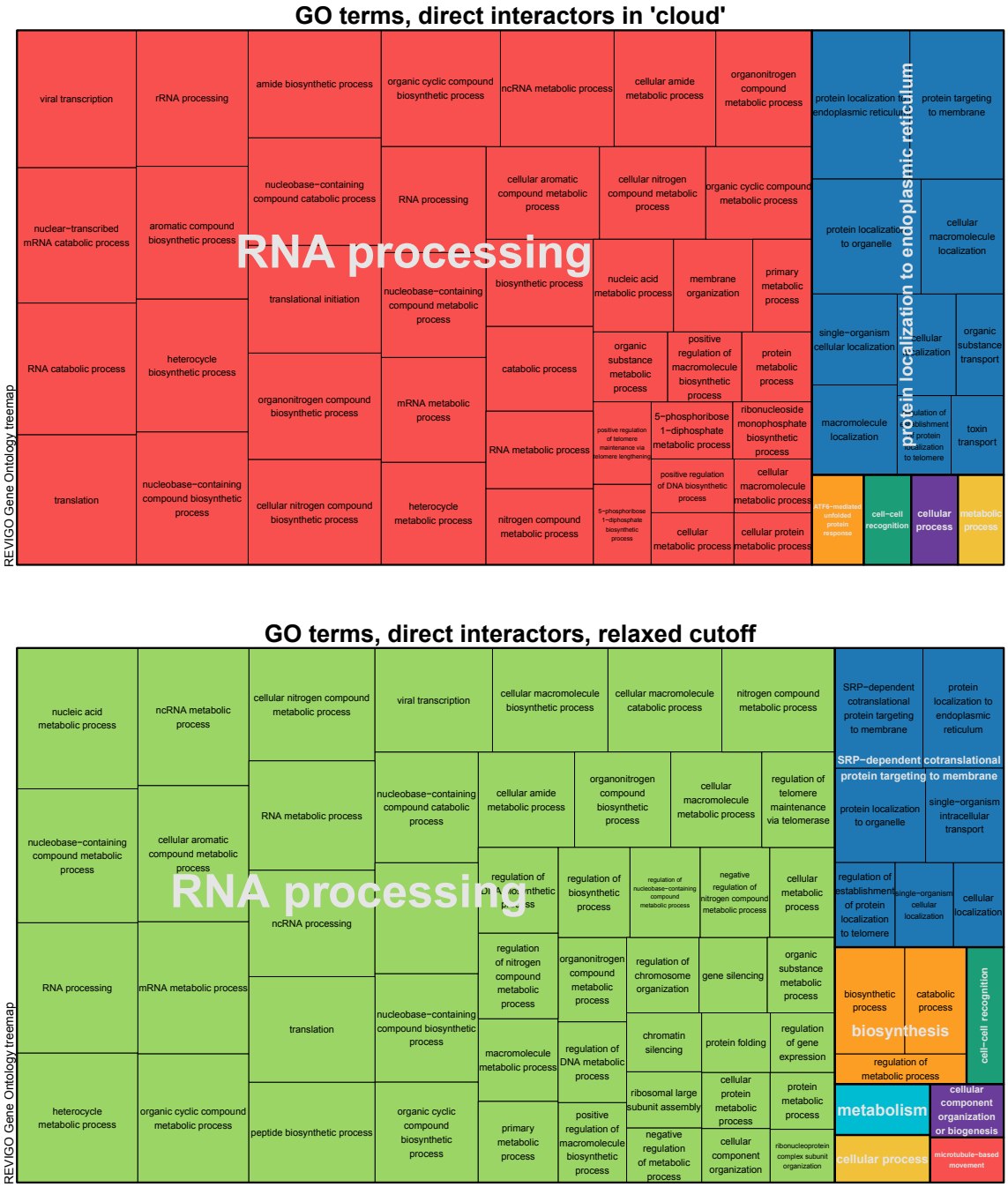


Figure 7.3.1: Treemaps generated with REViGO [174]. **Top:** GO-terms found enriched in the separate "cloud" from figure 3.4.2 with GOrilla [173]. **Bottom:** GO-terms found enriched when a much more relaxed cutoff ($s_0 = 2$, FDR = 0.05) is used. In both subsets, GO-terms enriched are mostly RNA-related. As a second major theme, vesicle formation/protein transport emerge.

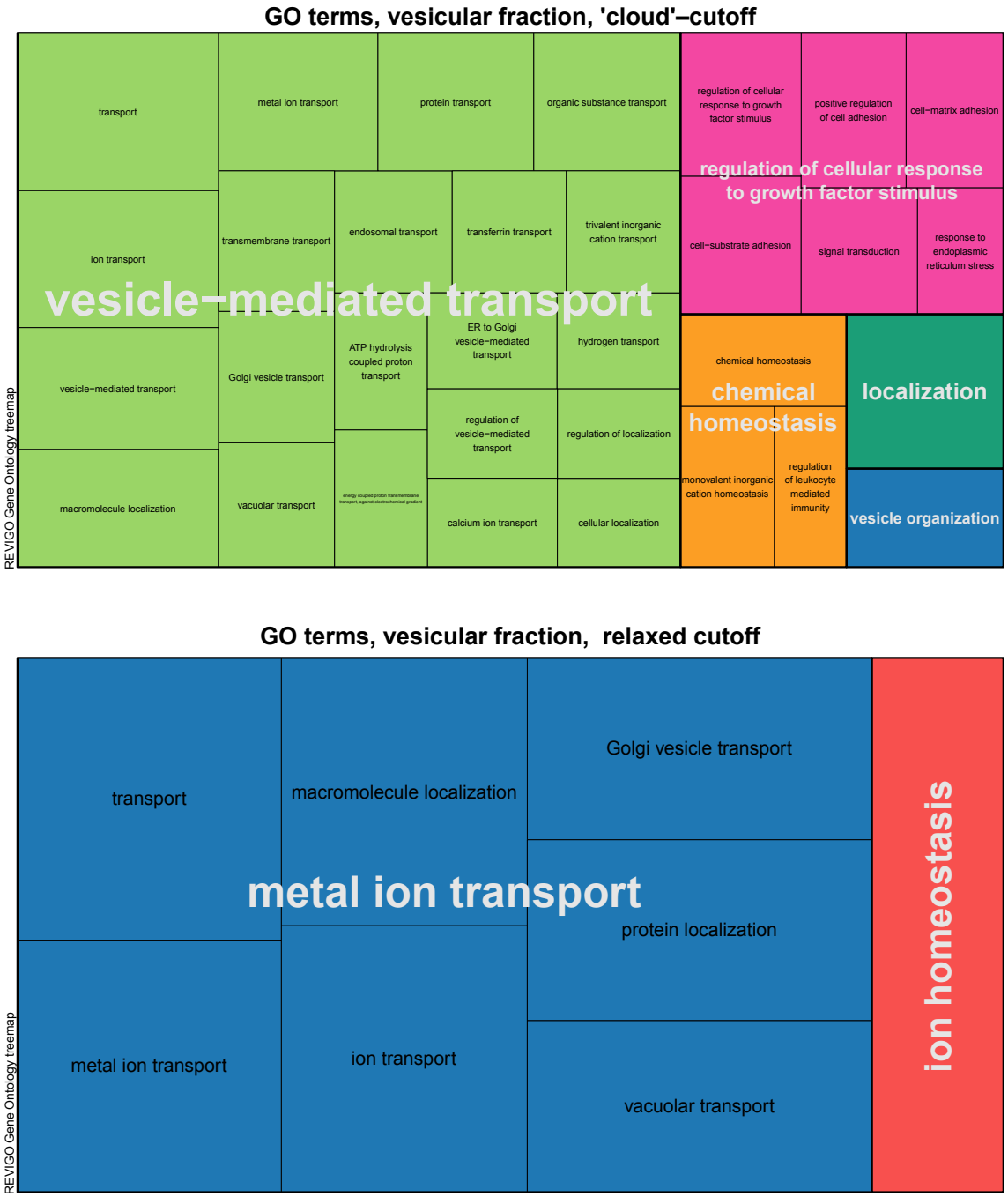


Figure 7.3.2: Treemaps generated with GOrilla [173] and REViGO [174]. **Top:** GO-terms found enriched when the lowest fold change and highest p-value of those proteins in the "cloud" from figure 3.4.2 were chosen as cutoff. **Bottom:** GO-terms found enriched when a much more relaxed cutoff ($s_0 = 2$, FDR = 0.05) is used. In both subsets, GO-terms enriched are mostly vesicle transport related. Under the more lenient cutoff (bottom) also the response to outsided stimuli is an important theme.