

Modelling Molecular Processes During Weight Loss

Regulation of Metabolic Flexibility



Samar H. K. Tareen

**Modelling molecular processes in weight loss:
Regulation of metabolic flexibility**

Samar H.K. Tareen

2020

**Modelling molecular processes in weight loss:
Regulation of metabolic flexibility**

Dissertation

to obtain the degree of Doctor at Maastricht University,
on the authority of the Rector Magnificus
Prof. Dr Rianne M. Letschert,
in accordance with the decision of the Board of Deans,
to be defended in public
on Wednesday, 25th of March 2020, at 16.00 hours

by

Samar Hayat Khan Tareen



The research presented in this thesis has been made possible with the support of the Dutch Province of Limburg.

Promotores

Prof. Dr Chris T.A. Evelo

Prof. Dr Ilja C.W. Arts

Prof. Dr Theo M. de Kok

Copromotor

Dr Martina Summer-Kutmon

Assessment Committee

Prof. Dr Ellen E. Blaak, Chair

Dr Armand Valsesia, Lausanne, Switzerland

Dr Marleen M.J. van Greevenbroek

Prof. Dr Natal A. van Riel, Amsterdam, UMC

Prof. Dr Ralf L. Peeters

© Samar H.K. Tareen, 2020.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior written permission of the author.

Cover and Layout: Samar H.K. Tareen

Printing: Ridderprint, www.ridderprint.nl

ISBN: 978-94-6380-716-6



Contents

1	General Introduction	1
2	Profiling cellular processes in adipose tissue during weight loss using time series gene expression	15
3	Exploring the cellular network of metabolic flexibility in the adipose tissue	47
4	Logical modelling reveals the PDC-PDK interaction as the regulatory switch driving metabolic flexibility at the cellular level	63
5	Stratifying cellular metabolism during weight loss: an interplay of metabolism, metabolic flexibility and inflammation	93
6	General Discussion	115
	Summary	123
	Samenvatting	127
	خلاصہ	131
	Valorisation	135
	Acknowledgements	139
	Curriculum Vitae	143
	List of Publications	147

1

General Introduction

Obesity and chronic illnesses

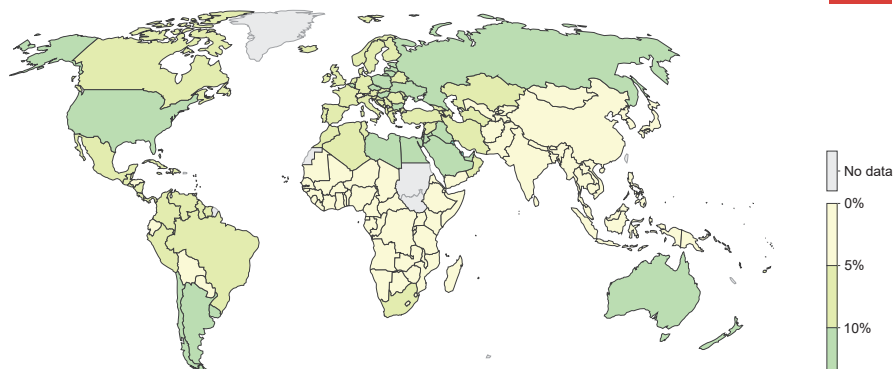
In recent decades, there has been an increase in the proportion of people becoming obese due to our increasingly more sedentary life style and high calorie diets. According to a WHO report [1], globally an estimated 11% of men and 15% women are classified as obese, that is, having a body mass index equal or greater than 30 kg/m². This is concerning as obesity is being associated with developments of an increasing number of non-communicable chronic illnesses. These illnesses predominantly include type 2 diabetes mellitus, several cardiovascular diseases, and metabolic syndrome [2–4], but obesity has also been found to be a contributing factor for several forms of cancer [5].

A number of different studies have shown the positive effects of diet induced weight loss, especially when coupled with regular exercise, to reduce the co-morbidities of obesity [6–8]. Studies have also shown the benefits of diet and exercise towards recovering from the aforementioned chronic illnesses. In addition, the type and composition of the diet, such as the proportion of fats and carbohydrates, also has an effect on weight loss [9]. However, one major issue for long-term health is the eventual weight regain of many obese individuals, post weight loss. As such, current research is focusing on the genetic and molecular causes of obesity and linking them with associated co-morbidities and chronic illnesses to design more targeted therapies towards personalised treatment of obesity and its co-morbidities. Studies show an impairment of insulin signalling along with changes in cellular metabolism as key molecular changes that take place in obesity [10–12].

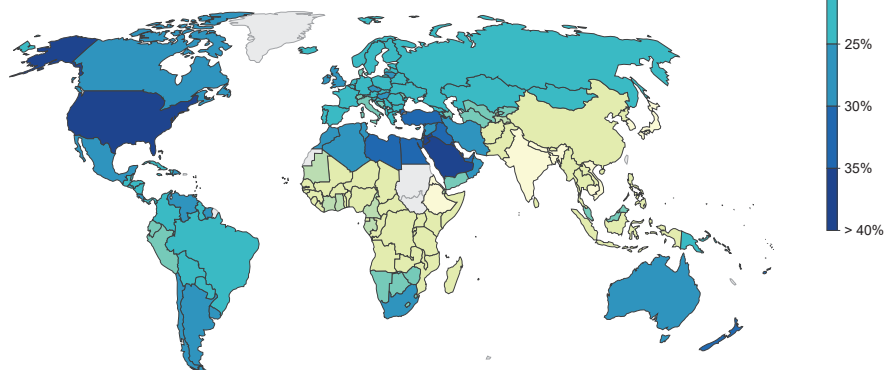
Metabolic flexibility and its role in obesity

Cellular energy metabolism, also referred to as cellular respiration, is a set of biochemical reactions that generate energy from various fuel molecules, also called substrates, such as glucose, fatty acids and amino acids. These substrates act as nutrients for the growth and/or functioning of cells. Primarily, the biochemical reactions involve the conversion of these substrates by oxidation into carbon dioxide and water while synthesising adenosine triphosphate (ATP). ATP is a molecule that powers biochemical reactions throughout the cell by releasing energy whenever it loses one or two of its phosphate groups. In aerobic respiration, oxygen acts as an electron receptor in the overall reaction. Under anaerobic (oxygen starved) conditions, other electron receptors are used; however, this variant is inefficient as it yields fewer ATP molecules per reaction as compared to aerobic respiration [14]. These biochemical reactions are termed as glycolysis and fatty acid oxidation for glucose and fatty acids respectively. Amino acids are catabolised into either glucose or fatty acids via gluconeogenesis or lipid synthesis respectively.

Share of adults defined as obese, 1975



Share of adults defined as obese, 2016



Source: WHO, Global Health Observatory

CC BY

Figure 1.1: WHO estimates showing the prevalence of obesity as a percentage of population in respective countries in 1971 and 2016. Illustration based on data sourced from [13].

Depending upon the availability of nutrients in the human body at any given time, different cells and tissues metabolise either glucose or fatty acids. In nutrient rich conditions, such as after a meal, glucose is the preferred substrate for metabolism. However, as the glucose reservoirs in the human body deplete, cells and tissues start switching over to fatty acid metabolism to reserve the limited amount of glucose for cells and tissues unable to metabolise fatty acids, such as erythrocytes, endothelial cells and neurons in the brain. This ability to switch between these substrates is termed as metabolic flexibility [15]. Metabolic flexibility exists both at the intracellular level, where glycolysis and fatty acid oxidation fuel metabolism, as well as across the human body where various tissues and organs communicate with each other to maintain usable levels of glucose and fatty acids. The metabolic flexibility across the human body is also referred to as the Randle Cycle [15, 16], named after Sir Philip John Randle who described it back in 1966. In the Randle Cycle, metabolically active tissues and organs such as the adipose tissue, the liver and skeletal muscles play an important role in maintaining circulating levels of glucose and fatty acids through signalling cascades and hormones.

Studies in obesity and chronic diseases have often found an impairment in metabolic flexibility affecting cellular energy metabolism [17, 18]. This impairment manifests itself as impaired cellular signalling leading to ineffective utilisation of metabolic substrates. In the adipose tissue, this leads to inefficient release of fatty acids to circulation. Cellular metabolism in obese individuals also showed a marked deviation in the expression of rate limiting enzymes in glycolysis and fatty acid oxidation compared to baseline. For example, in this particular study [19], gene expression data of the adipose tissue between lean and obese individuals showed the impaired expression of several enzymes in obese individuals. As such, the impaired expression of these enzymes affects cellular metabolism, and leads to subsequent unavailability of substrates for other tissues.

An application of systems biology

The complex interplay of obesity and chronic illnesses, as well as the mechanistic changes occurring at the cellular level are a prime target for analysis using approaches from systems biology. Systems biology is the application of computational and mathematical approaches to model biological components as whole systems in order to study and analyse emergent behaviours and properties [20]. The approaches allow the synergistic application of multiple types of data in order to perform targeted analyses and generate new information.

Multi-omics data

In modern life sciences research, the study of the various aspects and/or levels of the cellular information, function and dynamics have been divided into their own targeted fields, referred to as ‘-omics’. These include the genomics that relates to the evolution and genetics aspect of the cell, the transcriptomics that deals with the expression of the genes, the proteomics that deals with the location and function of the proteins, and the metabolomics that deals with the biochemical metabolites and their utilisation in the cell. Individually, these various ‘-omics’ have provided a plethora of research data and advancement. However, collectively, they represent a synergistic picture of the cellular dynamics and functions which are difficult to discern from study of these ‘-omics’ individually. In this thesis, we made use of transcriptomics and proteomics data.

Transcriptomics

Transcriptomics is the study, collection and the analysis of the gene expression of cells or tissues at a given time. The transcriptomics reveals which genes of the cell are being transcribed to be translated into proteins, and represent the state of the cell at the give time in response to an activity or stimulus. The transcriptome of the cell or tissue can also be measured in a successive series of data points at different time points, providing a crude motion picture of the activity of the cell or tissue. Various high-throughput technologies exist to capture these snapshots, the major ones being microarray and RNA-sequencing [21], providing expression data for tens of thousands of genes at a time. For a detailed explanation of the various technologies, Lowe *et al* is referred [21].

Proteomics

Similar to transcriptomics, proteomics is the large-scale study of the protein content of the cell, its localisation as well as its function. While the transcriptomics reveals the expression of the genome, the proteomics represents the translation into biomolecules that are responsible for the biochemical processes taking place in the cell. Similar to transcriptomics, the proteomics can also be measured as different time points to represent the evolution of the activity of the cell with respect to time, and maybe in response to a stimulus. However, unlike transcriptomics, proteomics have not reached the same scale of high-throughput data generation. This gap is primarily because of how the quantification technologies of proteomics work – in that the quantity, structure and function needs to be determined in a lengthy multistep procedure [22]. Multiple technologies for proteomics data collection exist, as reviewed in [22].

Data analysis

Various statistical analysis methods and pipelines have been developed for the analysis of high-throughput ‘-omics’ data [23]. One of the most common analyses performed on transcriptomics data is differential expression analysis [24]. This analysis is used to generate contrasts between two sets of data points to find the differences in expression between them. These sets can be two different cases, two different studies, or two different time points in the same study (providing a trend for gene expression changes in response to stimuli). Differential expression analysis generates a list of differentially expressed genes that were found to be significantly changed between the two sets of data points. These genes can then be used to probe databases to analyse different biological processes and pathways that might be affected by the stimuli or conditions of the study, allowing for further targeted research and analyses. In addition to differential expression analyses, (multi)-omics data can also be utilised to generate clusters of individuals or data points based on similarity to generate different biological profiles [25]. In this thesis, we have utilised multi-omics data analyses for differential expression analyses (Chapter 2, 3 and 5) as well as for stratifying biological profiles (Chapter 5).

Network biology

Network biology is the application of graph theory towards the curation, visualisation and analysis of biological multi-omics data [26]. One of the advantages of multi-omics data, as described previously, is the benefit of synergistic analyses via the integration of data. Network biology represent one of the best frameworks for data integration and integrated analyses by connecting the different components of the biological system together in the form of pathways and interactions. From generating abstracted regulatory networks for exploring the dynamics of regulation of biological components [27], to data-driven co-expression networks for finding groups of genes being regulated or targeted together [28], to physically interacting protein-protein and protein-ligand networks [29], network biology allows the merger of different levels of data at different time scales. In addition, network biology capitalises on the analytical methods available to graph theory and applies them in a biological context for generating new information such as new biological behaviours, new protein functions or new structural topology [30–32]. Figure 1.2 shows an example of the application of network biology, showing a correlation network of gene expression with genes clustered together based on their fold change. The blue-red colour spectrum provides the range of fold change, while the edge style shows the correlation value between gene pairs.

In this thesis, we have utilised network biology in multiple applications including generating time series gene co-expression networks (Chapter 2),

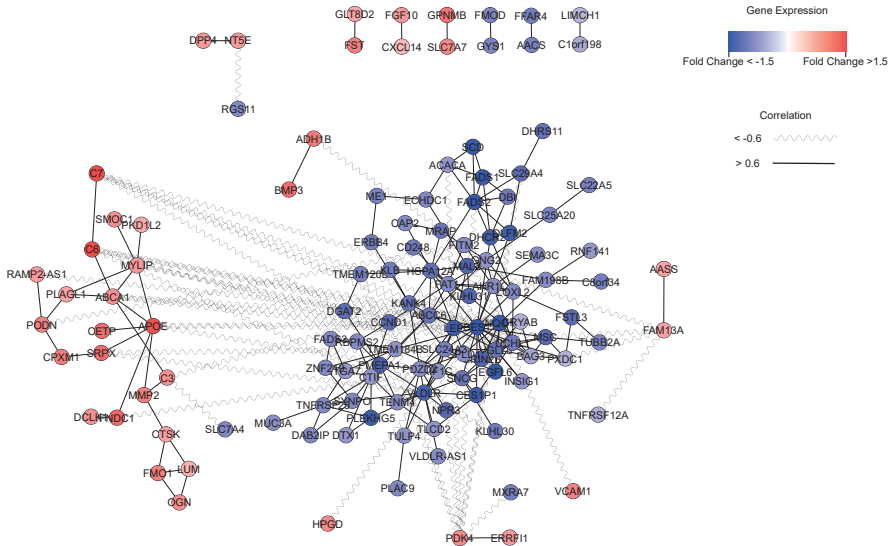


Figure 1.2: Correlation network of genes showing their expression fold changes as the blue-red node colour spectrum. The edge style shows the type of correlation between the genes, and the clustering broadly shows groups of genes having similar expression fold changes.

multi-scale multi-level abstracted visualisation of cellular processes (Chapter 3), overlaying existing/published information (Chapter 3), and generating theoretical models for querying biological systems for which we do not have the data (Chapter 4).

Datasets analysed in this thesis

We have analysed transcriptomics and proteomics expression data from a weight loss study conducted at Maastricht University [33]. The study was originally conducted to explore the ‘yoyo’ effect – the weight regain post weight loss in the context of high caloric restriction. The study compared two weight loss diets, a low calorie diet at 1250 kcal/day and a very low calorie diet at 500 kcal/day. The original study found that the intensity of the caloric restriction affected the changes induced in the gene expression of the adipose tissue, however this did not produce any differences in the weight regain post weight loss as both diet groups regained above 50% of their lost weight during dietary intervention. Given that diet induced weight loss is one of the most effective means of reducing obesity and the chances of developing its co-morbidities, this study became quite suitable for our research as we would be able to utilise the time series expression data to map cellular processes

involved in weight loss and caloric restriction. In addition, we also had access to phenotypic/clinical measurements of the individuals, allowing us to find patterns linking mechanistic changes at the cellular level to the phenotypic changes at the organism level.

In addition, the study presented in this thesis also relies on several publicly available datasets. These datasets are the GTEx baseline RNA-seq transcriptomics expression data (ArrayExpress: E-MTAB-5214), an obesity study data (Gene Expression Omnibus: GSE55200), and the BLUEPRINT epigenome project (ArrayExpress: E-MTAB-3827). Collectively, these studies provided additional baseline and contrasting expression information for the inference of results obtained from the analysis of the Yoyo study.

Thesis objectives and study design

The primary objective of this thesis was to explore, and identify cellular processes being affected during weight loss as possible targets for future research for personalised treatment of obesity, and by extension its co-morbidities. We followed an exploratory data driven study design where we started with analyses unhindered by strict hypotheses and presumptions. The study direction evolved based on the evidences provided by the various statistical and network based analyses conducted. These evidences were checked against known literature in parallel to the analyses with the purpose to minimise the chances of biased interpretation and to explore more avenues for subsequent analyses.

Thesis outline

In Chapter 2, we present a systems biology pipeline in which we capitalise on the variability in human data to generate groups of related co-expressing genes being affected by caloric restriction in the subcutaneous adipose tissue. The objective is to explore cellular processes affected by caloric restriction that can be targeted for further analysis in the context of obesity. We started with the microarray transcriptomics data for the Yoyo study and performed differential expression analysis to isolated differentially expressed genes that were then used as input for the dynamically co-expressed neighbourhood (DCeN) algorithm [34]. As opposed to general co-expression algorithms, DCeN forms patterns of gene co-expression within the individuals first, before averaging the correlation values, thereby avoiding the dampening effect of using average expression across individuals to calculate the correlations. This approach provided us with several clusters of closely co-expressing and correlated genes, with each cluster associated with unique functions in the backdrop of metabolism.

In Chapter 3, we focus on one of the clusters of correlated genes from Chapter 2. This cluster represents various rate limiting enzymes associated with cellular metabolism and was the first instance where we came across metabolic flexibility. The objective of this chapter is to review current known literature about cellular processes involved with metabolic flexibility. In addition to reviewing cellular processes involved with metabolic flexibility, we also constructed and provided a network resource for overlaying existing information. The utility of the network resource was demonstrated by overlaying baseline gene expression data of the cellular metabolism from the GTEx dataset, and comparing it to a differential expression analysis of a publicly available dataset to immediately highlight components having an expression pattern divergent from baseline.

In Chapter 4, we modelled cellular metabolic flexibility using a discretised modelling framework called René Thomas Kinetic Logic Formalism. The hypothesis for this study is that the inhibition of pyruvate dehydrogenase complex (PDC) by the pyruvate dehydrogenase kinase (PDK) isoenzymes during cellular metabolism is a key mechanism for cellular metabolic flexibility. Literature estimates suggest that metabolic flexibility, specifically the switching of substrates, is a sub-second process, for which we currently do not have accurate measurements, nor any method to observe in real-time. As such, the discretised framework allowed us to model the system in a time and quantity abstracted system, thereby focusing on the phenotypes generated by the model instead. The results from our model suggested that not only the regulation of PDC by PDKs affecting cellular metabolic flexibility a very likely case, but also that irregularities introduced in other parts of the model also transmitted their effect through this negative regulatory interaction.

Chapter 5 represents a return to data driven analysis and the Yoyo study, where we utilise our information from Chapter 3 and 4 to generate a list of genes/proteins involved in metabolic flexibility. The objective in this chapter is to stratify metabolic profiles during caloric restriction based on the genes/proteins and, by extension, their cellular processes involved with metabolic flexibility. We apply a new systems biology pipeline where we use the Affinity Network Fusion (ANF) algorithm [35] to stratify samples of the individuals from the Yoyo study based on their metabolic profile generated by their transcriptomics and proteomics data of the metabolic flexibility genes/proteins. The algorithm provided us with two clusters of samples, which when arranged chronologically with respect to the Yoyo study showed the respective individuals of the samples either changing or maintaining their clustering as they progress through the caloric restriction. Differential expression analysis revealed that one of the clusters had a lower expression of cellular metabolism associated genes across the board compared to the other cluster. In addition, this cluster also showed a higher expression of

genes associated with inflammation, indicating a peculiar inverse relationship between metabolism and inflammation, warranting further investigation in future studies.

The thesis ends with a general discussion on the interplay of cellular metabolism, metabolic flexibility and inflammation in the adipose tissue. In addition, the usefulness and utility of the combinatorial approaches in systems biology workflows and multi-omics data analysis, as well as the methodological challenges faced in this study are also discussed.

References

- [1] “Global status report on noncommunicable diseases 2014,” tech. rep., World Health Organization, Geneva, Switzerland, 2014.
- [2] J. Kaur, “A comprehensive review on metabolic syndrome,” *Cardiology Research and Practice*, vol. 2014, 2014.
- [3] C. Power, S. M. P. Pereira, C. Law, and M. Ki, “Obesity and risk factors for cardiovascular disease and type 2 diabetes: Investigating the role of physical activity and sedentary behaviour in mid-life in the 1958 british cohort,” *Atherosclerosis*, vol. 233, no. 2, pp. 363 – 369, 2014.
- [4] L. F. Van Gaal, I. L. Mertens, and C. E. De Block, “Mechanisms linking obesity with cardiovascular disease,” *Nature*, vol. 444, pp. 875–880, 2006.
- [5] T. W. Stone, M. McPherson, and L. Gail Darlington, “Obesity and cancer: Existing and new hypotheses for a causal connection,” *EBioMedicine*, vol. 30, pp. 14–28, 2018.
- [6] “Diet, nutrition and the prevention of chronic diseases. report of the joint WHO/FAO expert consultation (WHO technical report series), no. 916 (trs 916),” tech. rep., World Health Organization, Geneva, Switzerland, 2003.
- [7] S. Klein, N. F. Sheard, X. Pi-Sunyer, A. Daly, J. Wylie-Rosett, K. Kulkarni, and N. G. Clark, “Weight management through lifestyle modification for the prevention and management of type 2 diabetes: Rationale and strategies,” *Diabetes Care*, vol. 27, no. 8, pp. 2067–2073, 2004.
- [8] C. J. Lavie, R. V. Milani, and H. O. Ventura, “Obesity and cardiovascular disease risk factor, paradox, and impact of weight loss,” *Journal of the American College of Cardiology*, vol. 53, no. 21, pp. 1925–1932, 2009.
- [9] I. Shai, D. Schwarzfuchs, Y. Henkin, D. R. Shahar, S. Witkow, I. Greenberg, R. Golan, D. Fraser, A. Bolotin, H. Vardi, O. Tangi-Rozental, R. Zuk-Ramot, B. Sarusi, D. Brickner, Z. Schwartz, E. Sheiner, R. Marko, E. Katorza, J. Thiery, G. M. Fiedler, M. Blüher, M. Stumvoll, and M. J. Stampfer, “Weight loss with a low-carbohydrate, mediterranean, or low-fat diet,” *New England Journal of Medicine*, vol. 359, no. 3, pp. 229–241, 2008.
- [10] K. L. Campbell, K. E. Foster-Schubert, K. W. Makar, M. Kratz, D. Hagman, E. A. Schur, N. Habermann, M. Horton, C. Abbenhardt, L.-Y. Kuan, L. Xiao, J. Davison, M. Morgan, C.-Y. Wang, C. Duggan, A. McTiernan, and C. M. Ulrich, “Gene expression changes in adipose tissue with diet- and/or exercise-induced weight loss,” *Cancer Prevention Research*, vol. 6, no. 3, pp. 217–231, 2013.
- [11] P. A. Kern, S. Ranganathan, C. Li, L. Wood, and G. Ranganathan, “Adipose tissue tumor necrosis factor and interleukin-6 expression in human obesity and insulin resistance,” *American Journal of Physiology - Endocrinology and Metabolism*, vol. 280, no. 5, pp. E745–E751, 2001.
- [12] C. Leyvraz, C. Verdumo, M. Suter, A. Paroz, J.-M. Calmes, P. M. Marques-Vidal, and V. Giusti, “Changes in gene expression profile in human subcutaneous adipose tissue during significant weight loss,” *Obesity Facts*, vol. 5, pp. 440–451, 2012.
- [13] H. Ritchie and M. Roser, “Obesity & bmi,” *Our World in Data*, 2019. <https://ourworldindata.org/obesity>.

- [14] P. R. Rich, "The molecular machinery of keilin's respiratory chain," *Biochemical Society Transactions*, vol. 31, no. 6, p. 1095, 2003.
- [15] L. Hue and H. Taegtmeyer, "The randle cycle revisited: a new head for an old hat," *American Journal of Physiology - Endocrinology And Metabolism*, vol. 297, no. 3, pp. E578–E591, 2009.
- [16] R. M. Denton and P. J. Randle, "Citrate and the regulation of adipose-tissue phosphofructokinase," *Biochemical Journal*, vol. 100(2), no. 0264-6021 (Print), pp. 420–423, 1966.
- [17] J. E. Galgani, C. Moro, and E. Ravussin, "Metabolic flexibility and insulin resistance," *American journal of physiology. Endocrinology and metabolism*, vol. 295, no. 5, pp. E1009–E1017, 2008.
- [18] B. H. Goodpaster and L. M. Sparks, "Metabolic flexibility in health and disease," *Cell metabolism*, vol. 25, no. 5, pp. 1027–1036, 2017.
- [19] F. Badoud, K. P. Lam, A. DiBattista, M. Perreault, M. A. Zulyniak, B. Cattrysse, S. Stephenson, P. Britz-McKibbin, and D. M. Mutch, "Serum and adipose tissue amino acid homeostasis in the metabolically healthy obese," *Journal of Proteome Research*, vol. 13, no. 7, pp. 3455–3466, 2014.
- [20] I. Tavassoly, J. Goldfarb, and R. Iyengar, "Systems biology primer: the basic methods and approaches," *Essays In Biochemistry*, vol. 62, no. 4, p. 487, 2018.
- [21] R. Lowe, N. Shirley, M. Bleackley, S. Dolan, and T. Shafee, "Transcriptomics technologies," *PLOS Computational Biology*, vol. 13, no. 5, p. e1005457, 2017.
- [22] B. Aslam, M. Basit, M. A. Nisar, M. Khurshid, and M. H. Rasool, "Proteomics: Technologies and their applications," *Journal of Chromatographic Science*, vol. 55, no. 2, pp. 182–196, 2017.
- [23] C. Manzoni, D. A. Kia, J. Vandrovцова, J. Hardy, N. W. Wood, P. A. Lewis, and R. Ferrari, "Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences," *Briefings in bioinformatics*, vol. 19, no. 2, pp. 286–302, 2016.
- [24] M. Kebschull, M. J. Fittler, R. T. Demmer, and P. N. Papapanou, *Differential Expression and Functional Analysis of High-Throughput -Omics Data Using Open Source Tools*, pp. 327–345. New York, NY: Springer New York, 2017.
- [25] G. Tini, L. Marchetti, C. Priami, and M.-P. Scott-Boyer, "Multi-omics integration – a comparison of unsupervised clustering methodologies," *Briefings in Bioinformatics*, 2017.
- [26] A.-L. Barabási and Z. N. Oltvai, "Network biology: understanding the cell's functional organization," *Nature Reviews Genetics*, vol. 5, no. 2, pp. 101–113, 2004.
- [27] R. Albert and J. Thakar, "Boolean modeling: a logic-based dynamic approach for understanding signaling and regulatory networks and for making useful predictions," *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, vol. 6, no. 5, pp. 353–369, 2014.
- [28] P. Langfelder and S. Horvath, "WGCNA: an r package for weighted correlation network analysis," *BMC Bioinformatics*, vol. 9, no. 1, pp. 1–13, 2008.

-
- [29] L. Giot, J. S. Bader, C. Brouwer, A. Chaudhuri, B. Kuang, Y. Li, Y. L. Hao, C. E. Ooi, B. Godwin, E. Vitols, G. Vijayadamodar, P. Pochart, H. Machineni, M. Welsh, Y. Kong, B. Zerhusen, R. Malcolm, Z. Varrone, A. Collis, M. Minto, S. Burgess, L. McDaniel, E. Stimpson, F. Spriggs, J. Williams, K. Neurath, N. Ioime, M. Agee, E. Voss, K. Furtak, R. Renzulli, N. Aanensen, S. Carrolla, E. Bickelhaupt, Y. Lazovatsky, A. DaSilva, J. Zhong, C. A. Stanyon, R. L. Finley, K. P. White, M. Braverman, T. Jarvie, S. Gold, M. Leach, J. Knight, R. A. Shimkets, M. P. McKenna, J. Chant, and J. M. Rothberg, "A protein interaction map of *Drosophila melanogaster*," *Science*, vol. 302, no. 5651, p. 1727, 2003.
- [30] E. Bullmore and O. Sporns, "Complex brain networks: graph theoretical analysis of structural and functional systems," *Nature Reviews Neuroscience*, vol. 10, p. 186, 2009.
- [31] K.-I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal, and A.-L. Barabási, "The human disease network," *Proceedings of the National Academy of Sciences*, vol. 104, no. 21, p. 8685, 2007.
- [32] R. Z. Paracha, J. Ahmad, A. Ali, R. Hussain, U. Niazi, S. H. K. Tareen, and B. Aslam, "Formal modelling of toll like receptor 4 and jak/stat signalling pathways: Insight into the roles of socs-1, interferon- β and proinflammatory cytokines in sepsis," *PLOS ONE*, vol. 9, no. 9, p. e108466, 2014.
- [33] R. G. Vink, N. J. T. Roumans, L. A. J. Arkenbosch, E. C. M. Mariman, and M. A. van Baak, "The effect of rate of weight loss on long-term weight regain in adults with overweight and obesity," *Obesity*, vol. 24, no. 2, pp. 321–327, 2016.
- [34] L. L. Elo and B. Schwikowski, "Analysis of time-resolved gene expression measurements across individuals," *PLoS ONE*, vol. 8, pp. 1–8, 12 2013.
- [35] T. Ma and A. Zhang, "Affinity network fusion and semi-supervised learning for cancer patient clustering," *Methods*, vol. 145, pp. 16–24, 2018.

2

Profiling cellular processes in adipose tissue during weight loss using time series gene expression

Samar H.K. Tareen¹, Michiel E. Adriaens¹, Ilja C.W. Arts^{1,2}, Theo M. de Kok^{1,3}, Roel G. Vink⁴, Nadia J.T. Roumans⁴, Marleen A. van Baak⁴, Edwin C.M. Mariman⁴, Chris T. Evelo^{1,5} and Martina Kutmon^{1,5}

¹ Maastricht Centre for Systems Biology (MaCSBio), Maastricht University, the Netherlands

² Department of Epidemiology, CARIM School for Cardiovascular Diseases, Maastricht University, the Netherlands

³ Department of Toxicogenomics, GROW School of Oncology and Developmental Biology, Maastricht University, the Netherlands

⁴ Department of Human Biology, NUTRIM Research School, Maastricht University, the Netherlands

⁵ Department of Bioinformatics – BiGCaT, NUTRIM Research School, Maastricht University, the Netherlands

Published in: *Genes* (2018) 9:11, 525.

doi: 10.3390/genes9110525

Abstract

Obesity is a global epidemic identified as a major risk factor for multiple chronic diseases and, consequently, diet-induced weight loss is used to counter obesity. The adipose tissue is the primary tissue affected in diet-induced weight loss, yet the underlying molecular mechanisms and changes are not completely deciphered. In this study, we present a network biology analysis workflow which enables the profiling of the cellular processes affected by weight loss in the subcutaneous adipose tissue. Time series gene expression data from a dietary intervention dataset with two diets was analysed. Differentially expressed genes were used to generate co-expression networks using a method that capitalises on the repeat measurements in the data and finds correlations between gene expression changes over time. Using the network analysis tool Cytoscape, an overlap network of conserved components in the co-expression networks was constructed, clustered on topology to find densely correlated genes, and analysed using Gene Ontology enrichment analysis. We found five clusters involved in key metabolic processes, but also adipose tissue development and tissue remodelling processes were enriched. In conclusion, we present a flexible network biology workflow for finding important processes and relevant genes associated with weight loss, using a time series co-expression network approach that is robust towards the high inter-individual variation in humans.

Introduction

In recent years, obesity has become a global epidemic with a World Health Organisation (WHO) estimate of 11% men and 15% women worldwide being obese in 2014 [1], with clear indications that this will continue to rise in the foreseeable future. Obesity has been identified as a major risk factor for multiple diseases and conditions such as type 2 diabetes mellitus (T2DM), cardiovascular diseases (CVD) and the metabolic syndrome (MetS) [2–4]. Consequently, a number of studies [5–7] have recommended weight loss through diet and physical activity to counter obesity and its co-morbidities.

In previous studies, various efforts have been made to understand the molecular biology behind obesity and the effect of weight loss [8–10]. These studies have focused on gene expression profiles and targeted pathways associated with the obese system, providing transcriptomic snapshots to better understand the functioning of the system in a particular state or in responses to stimuli by comparing the expression of the genes across the genome.

In recent years, gene expression profiling has been extended into time series, generating gene expression snapshots at different time points. The time points collectively show how gene expression changes over time, particularly in response to different stimuli or interventions. The time series also yield differential gene expression data which provides expression contrasts between pairs of time points, which are used to glean which processes may or may not be active at a particular time point via cellular pathways [11]. However, it has been difficult to accurately associate cellular processes and pathways with time series gene expression profiles, partly due to high inter-individual variability in human transcriptomic datasets [12, 13].

In this article, we present a network biology analysis workflow which ties time series gene expression analysis with gene expression pattern correlation and downstream enrichment techniques to identify biological processes and pathways in the adipose tissue as possible regulatory candidates linking the obese system and chronic diseases with the beneficial effects of weight loss. The analyses give us a detailed view of what is happening in the subcutaneous adipose tissue during weight loss, irrespective of the intensity and duration of caloric restriction. Furthermore, the analysis provides us with closely correlated clusters of genes based on their gene expression patterns. These clusters represent possible areas of cross-talk between different biological processes, increasing our understanding of the functioning of the subcutaneous adipose tissue, as well as providing us with new areas of detailed research in the context of obesity and chronic diseases. In addition, the described network biology workflow is able to perform these analyses on human datasets containing high variation between the participants of the study, allowing for application on small sized studies as well.

Materials and Methods

Analysis workflow

The steps in our network biology analysis workflow is shown in Figure 2.1 and explained in detail in the following subsections. This workflow follows a data driven approach, taking raw expression data from microarrays to computable networks and information relating to biological processes. The data pre-processing (step 1) and constructing of correlation networks (step 2) are performed in ArrayAnalysis [14] and R [15]. The network analysis steps (3–5) are performed in Cytoscape [16] using core and app functionality.

Dataset

Raw transcriptomics data was obtained from the ‘Yoyo study’ [17] (Clinical Trial ID: NCT01559415, www.clinicaltrials.gov). The study was a human weight loss and subsequent weight regain study, comparing two weight loss diets: a low calorie diet (LCD) of 1,250 kcal/day for 12 weeks, and a very low calorie diet (VLCD) of 500 kcal/day for five weeks. Participants of each diet then underwent a four week weight maintenance period, with a nine-month follow up. All participants in the study were overweight and/or obese Caucasian individuals with body mass index (BMI) from 28 kg/m² to 35 kg/m² aged from 32 to 67 years (median age 51). The exact details of the study design can be found in [17].

The transcriptomics data was available on the Gene Expression Omnibus (ID: GSE77962). The data consisted of Affymetrix Human Gene ST 1.1 microarray platform expression data of the subcutaneous adipose tissue for 57 individuals. The complete gene expression data for the first three time points of the study (before weight loss, after weight loss, after weight maintenance) was available for only 46 subjects (22 LCD and 24 VLCD). The transcriptomics data was not available for the final time point of the study.

Data quality control and filtration of background expression

ArrayAnalysis.org was used for the quality control (QC) analysis and subsequent normalisation of the raw expression data [14]. ArrayAnalysis.org is an online pipeline for QC and normalisation of microarray expression data. In our study, we used the default settings for Affymetrix Human Gene ST 1.1 microarray platform and normalised using robust multi-array averaging (RMA) normalisation. The normalised expression data was then filtered for noise, generated via background gene expression, by removing genes which had median expression in the dataset equal-to or lower-than the median Y-chromosome gene expressions in female individuals in the dataset.

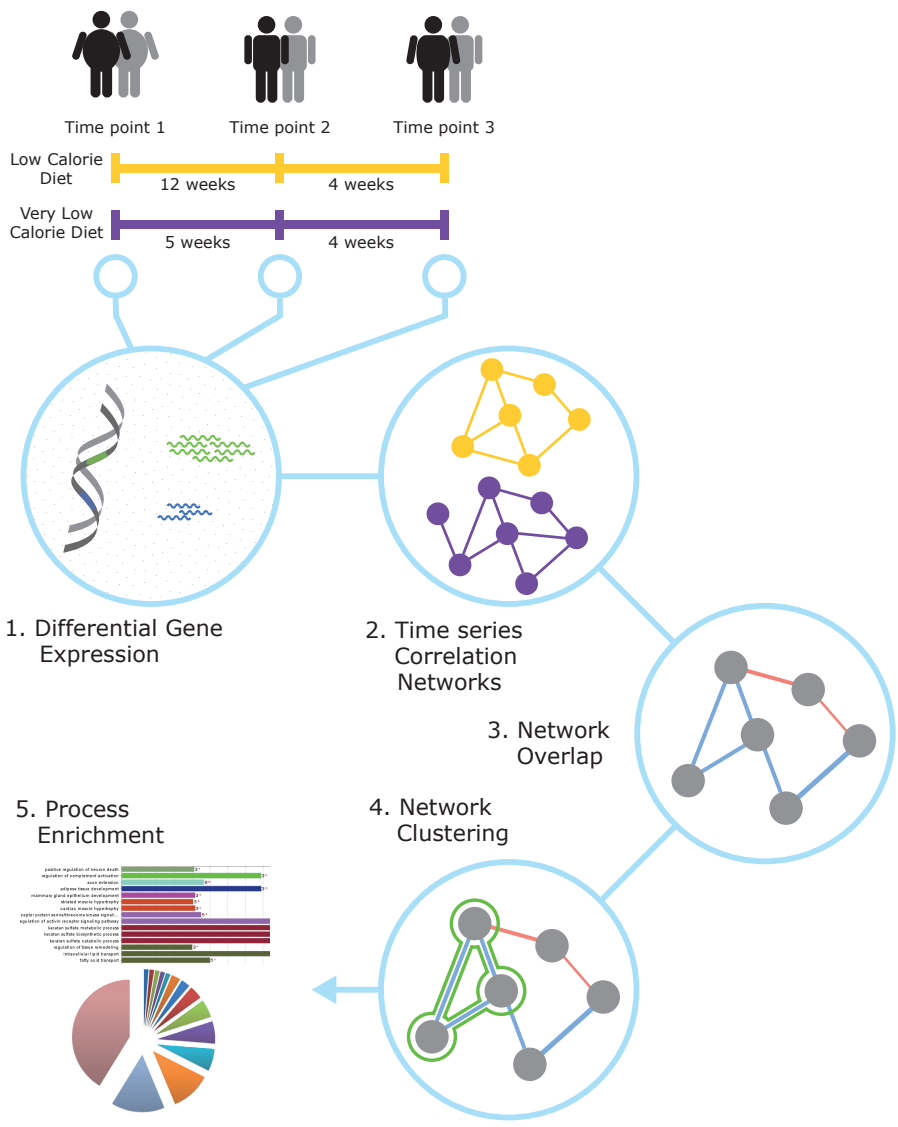


Figure 2.1: Network biology analysis workflow. (1) The time series expression data from the time points is normalised and differential expression analysis is performed. (2) Correlation networks are constructed on the time series for each diet respectively. (3) The overlap network is generated from the two networks showing the correlations which are shared between the two diets. (4) Community clustering is performed to find clusters of genes which are showing the most similar expression patterns. (5) The overlap network and the gene clusters are then used for process enrichment to find the affected cellular processes.

Differential expression analysis

The filtered expression data was then used for differential expression analysis of the genes across the three time points, separately for each diet. This analysis allowed us to compare how the gene expressions changed after weight loss and after weight maintenance. Comparing the number of significantly differentially expressed genes using the criteria $|FC| \geq 1.2 \wedge p\text{-value} < 0.05$ (where $|FC|$ is the absolute fold change of each gene) allowed us to estimate how significantly the diets are affecting the adipose tissue processes. It also showed how the expression intensities and processes differ between the two diets. Data was corrected for multiple testing using the q-value method [18].

Network inference and clustering

In order to study and visualise the gene expression over time, we opted to generate co-expression networks. To maintain correlation values of gene pairs across the three time points, we applied and modified the correlation method used in the Dynamically Co-expressed Neighbourhood (DCeN) algorithm [19]. The DCeN algorithm was developed for analysing time series gene expression data and calculates the correlations within individuals first which are then averaged per group. In our modification (see R implementation in Supplementary File, section Source Code), we calculate the signed correlations for generating the network for each diet based on a single input data group, using an absolute correlation value cut-off at 0.6 ($|\text{corr.}| \geq 0.6$) in at least 25% of the samples in each group, leaving only strong correlations. This is performed for each diet, after which an overlap network is constructed by taking the intersection of the edges present in the respective correlation networks.

Frobenius norm was used to address the inter-individual gene expression and expression pattern variability within the respective diets. The Frobenius norm measures the square root of the sum of square differences between the correlation matrix of the diet and the correlation matrix of the individual. The difference presents a distance dimension to compare how “distant” an individual is compared to the whole group in the diet itself. Mathematically, for two correlation matrices A and B with n gene correlations, the Frobenius norm will be calculated as

$$\|A - B\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^n |a_{ij} - b_{ij}|^2}. \quad (2.1)$$

Topological clustering was done on the co-expression networks using the GLay community cluster algorithm [20] through the ClusterMaker app [21] in Cytoscape [16]. Since the networks were already based on correlated gene expression patterns over time, closely connected gene clusters in the network

represented groups of genes having a highly similar or highly dissimilar expression pattern over time, and thus provided links between the processes in which the genes are involved. The \log_2 fold change of genes in the respective topological gene clusters of each diet was also plotted for each cluster to visually represent the clustering patterns over time, shown in Figures S2 and S3.

Gene ontology and pathway enrichment

Gene Ontology (GO) enrichment was performed on both the complete networks as well as the network clusters using the ClueGO app [22] in Cytoscape. For the gene clusters, the settings in ClueGO were modified to allow any number or percentage of genes in any level of GO (0–20 in ClueGO) with the p -values of the pathways ≤ 0.05 . For the complete network, the settings only differed in having at least 3 in number, or at least 4%, of genes in the respective GO terms. Pathway enrichment was performed using the over-representation analysis module at ConsensusPathDB [23], selecting all available pathway databases using a minimum gene overlap of 2 and a p -value ≤ 0.01 .

Software and libraries

ArrayAnalysis.org [14] was used for QC and normalisation, using custom chip definition file (CDF) annotation from BrainArray (version 19.0.0, ENSG). R [15] v3.2.3 was used with limma v3.26.5 package [24] for the differential gene expression analysis and qvalue v2.2.2 package [18] for the false discovery rate (FDR) analysis. Cytoscape [16] v3.4.0 was used for visualisation of the networks. Clustering was done using the ClusterMaker2 app [21] v0.9.5 in Cytoscape. ClueGO app [22] v2.3.2 in Cytoscape was used for GO enrichment, and ConsensusPathDB [25, 26] version 31 was used for the pathway enrichment.

Data availability

All relevant data is within the paper and its Supporting Information files. Gene expression data for the Yoyo study is accessible at Gene Expression Omnibus (accession number GSE77962).

Results

Data normalisation and filtration

The microarray gene expression data was obtained from a weight loss study comparing a low calorie diet (LCD) and a very low calorie diet (VLCD) [27]. In addition, 138 samples were used from the study – 46 individuals across three

time points. The three time points measured gene expression in the adipose tissue before weight loss, after weight loss, and after weight maintenance, respectively. Starting with the raw expression data of the 46 individuals, the QC analysis results showed two microarray samples as outliers (Figure S1). These outliers were two individuals from the VLCD diet group because their data at time point 1 (before weight loss) differed from the data cluster of the remaining VLCD members, whereas the data at the other two time points did not. This inconsistency led to the removal of the data from these two individuals across all three time points to remove any skewness in expression intensities or patterns introduced by the possibly erroneous data at these first time point samples. The normalised expression data including 21,641 genes was filtered for background gene expression leaving a total of 18,113 unique genes.

Differential expression

The differential expression analysis was performed within individuals, between expressions after weight loss and before weight loss (time points 2-1); after weight maintenance and before weight loss (time points 3-1); and after weight maintenance and after weight loss (time points 3-2). The time points 3-2 analysis for LCD came up with very high local false discovery rate (FDR) for p -value < 0.05 , implying a high chance of false positives even in significant results. Thus, this time point difference was not included in further analyses for either diet. In Figure 2.2, results of the differential gene expression analyses between the different time points in the different diets are shown and compared. A total of 286 genes were significantly differentially expressed between time points 1 and 2 for LCD, and 1,793 genes for VLCD in the same period. These numbers goes down to 220 for LCD and 399 for VLCD respectively between time points 1 and 3. The difference in number of significant differentially expressed genes between LCD and VLCD during the weight loss period indicates that the pattern of expression in VLCD is more perturbed than in LCD in response to stronger caloric restriction. Furthermore, not all genes in the two comparisons within respective diets were the same, with some genes being significantly perturbed in either the weight loss period, or the weight maintenance period, but not in both.

Correlation and overlap networks

The differentially expressed genes of each diet were then used to create expression data subsets for each diet from the filtered dataset. These subsets were fed into the modified DCeN algorithm (complete procedure given in the Materials and Methods section) to generate the correlation matrices, using an absolute Pearson correlation of ≥ 0.6 as the cut-off. The generated co-expression

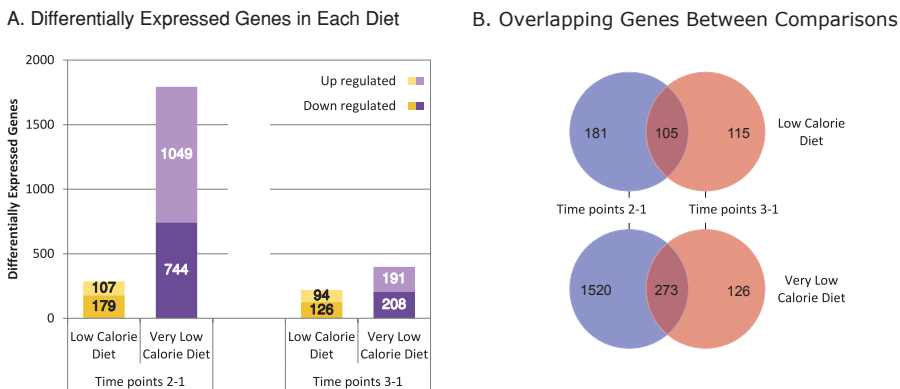


Figure 2.2: The number of differentially expressed genes in each diet. **(A)** The number of up and downregulated genes along the two time point comparisons. Time points 2-1: After Weight Loss—Before Weight Loss, and time points 3-1: After Weight Maintenance—Before Weight Loss; **(B)** the number of differentially expressed genes overlapping between the two comparisons within each diet.

network for the LCD diet contained 123 genes (nodes) and 250 correlations between these genes (edges). Forty-one of the 250 edges have negative correlation values. The positive correlations indicate highly similar expression patterns across the three time points, while the negative correlations indicate mirrored expression patterns. The LCD correlation network is shown in Figure S2. For the VLCD diet, the network contained 1382 nodes and 35,791 edges, of which 11,270 edges showed negative correlations between their respective genes. The VLCD correlation network is shown in Figure S3.

To find the common mechanisms, irrespective of the intensity and duration of caloric restriction, we generated an overlap network of the two diets. The overlap network is constructed by finding the gene pairs (edges) that are common between the two networks and correlated in the same direction (positive or negative). The overlap network shown in Figure 2.3 consisted of a single large connected component and five separate pairs correlated genes. The complete network contained 71 genes and 127 correlations between those genes, of which only 13 correlations were negative. The flexibility of the workflow to use the high level of sample variability was checked by comparing the correlation matrices of each individual with the correlation matrices of their respective diet group using the Frobenius norm of the distance matrix. The plot of the calculated Frobenius distances showing each individual in the respective diet is provided as Figure S4. The plot includes certain individuals having a higher distance from the respective diets, which shows that these individuals, despite the variability in expression data, were still part of the construction of the correlation network of the respective diet.

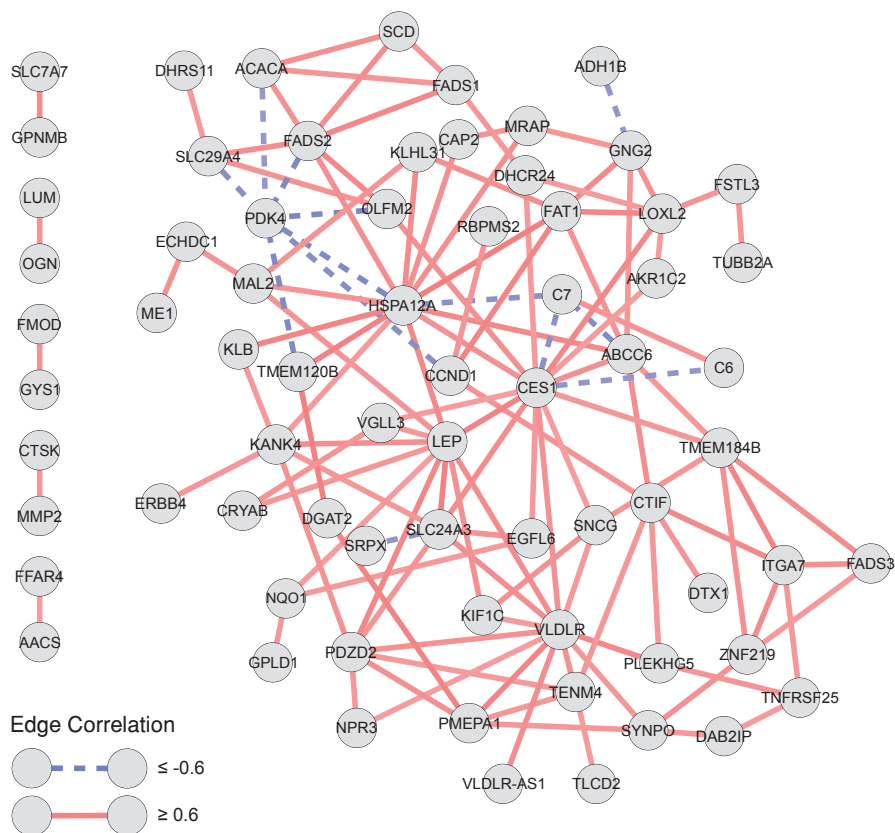


Figure 2.3: Overlap network showing the intersection of the edges of the low calorie diet (LCD) and the very low calories diet (VLCD) correlation networks. The intersection only depends on the sign/direction of the correlation (positive or negative), and not the exact value of the correlation.

Enrichment and clustering results

The overlap network was then used for Gene Ontology (GO) enrichment analysis. The most significant classes ranged from metabolic processes such as long chain fatty-acyl-CoA metabolic process (GOID: 35336) and keratan sulphate catabolic process (GOID: 42340), to non-metabolic processes such as adipose tissue development (GOID: 60612) and regulation of tissue remodelling (GOID: 34103). Using the topological clustering method described in the Materials and Methods section, the overlap network was clustered into five clusters, and a set of five un-clustered paired genes shown in Figure 2.4. Cluster 1 consists of eight genes, followed by 17 in Cluster 2, 14 in Cluster 3, 17 in Cluster 4, and five in Cluster 5, with the five unconnected paired correlations remaining

unclustered. Clusters 1 and 4 provided additional significant enrichment terms, such as thioester biosynthetic process (GOID: 35384) and regulation of signal transducer and activator of transcription STAT protein input into nucleus (GOID: 2000364) respectively, adding a new dimension of process regulation information to the network. The five un-clustered gene pairs were enriched together and predominantly targeted the keratan sulphate catabolic process (GOID: 42340). Clusters 2 and 3 did not provide any significant enriched terms. Figure 2.5 shows these resultant GO terms as pie charts. Of the 71 genes in the overlap network, carboxylesterase 1 (CES1), heat shock protein family A member 12A (HSPA12A), very low density lipoprotein receptor (VLDLR) and leptin (LEP) showed a high degree centrality of 15, 14, 12 and 11, respectively. Expectedly, these high degree genes are parts of clusters, with CES1, VLDLR, and LEP belonging to Cluster 4, while HSPA12A acts as the hub node in Cluster 3.

Discussion

In this article, we presented a network biology workflow to find genes of interest using gene expression data for targeted analysis of processes and pathways affected during weight loss. Our data-driven workflow is applicable to any time series expression data. One of the reasons for the construction of the workflow was the previously reported poor performance of existing correlation calculation methods on human time series expression data [19]. Essentially, most standard techniques generate correlations between the average expression of the genes across the whole group. This methodology works perfectly for homogeneous datasets/groups, but, for heterogeneous groups, this methodology fails as the average gene expression of the group is dampened because of the high variability in the measurements. The dynamically co-expressed neighbourhoods (DCeN) algorithm created by Elo and Schwikowski was specifically made to address this issue by calculating gene–gene correlations within individual members of the dataset, and then averaging the correlations, thus preserving the variability in the data while also generating meaningful correlations [19]. In our article, we modified the DCeN algorithm to focus on a single expression group at a time to generate the correlation matrices of the diets individually.

Based on the high number of differentially expressed genes in VLCD, the resultant network for VLCD was also much larger in terms of the number of genes (nodes) and correlated expression patterns (edges) than the LCD network. The difference in size of the networks, coupled with the number of differentially expressed genes, clearly shows that the adipose tissue gene expression is affected in both diets and that VLCD causes a larger perturbation of the adipose tissue. The overlap network, constructed via the edge intersection

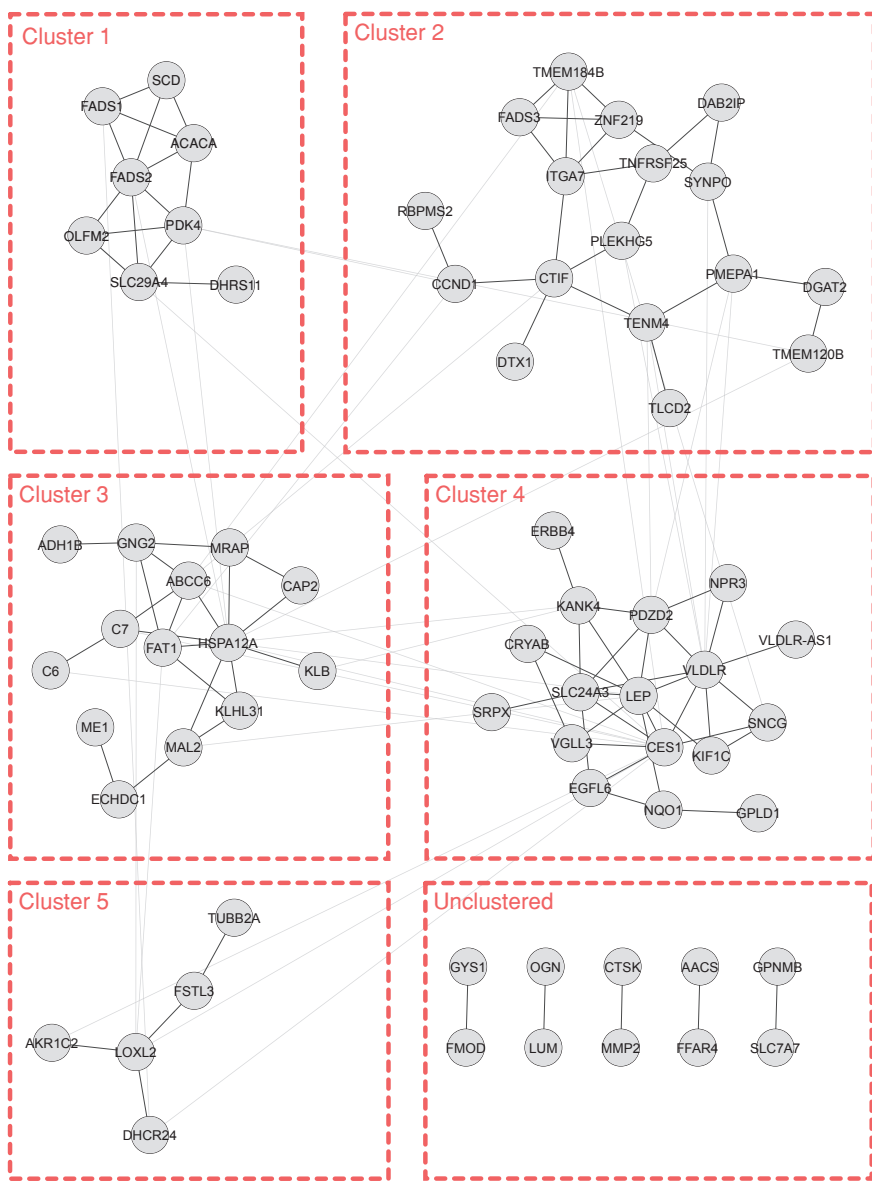


Figure 2.4: GLayer community clusters of the overlap network. Faded edges show the edges removed by the algorithm to generate the “community” of genes based on the topology.

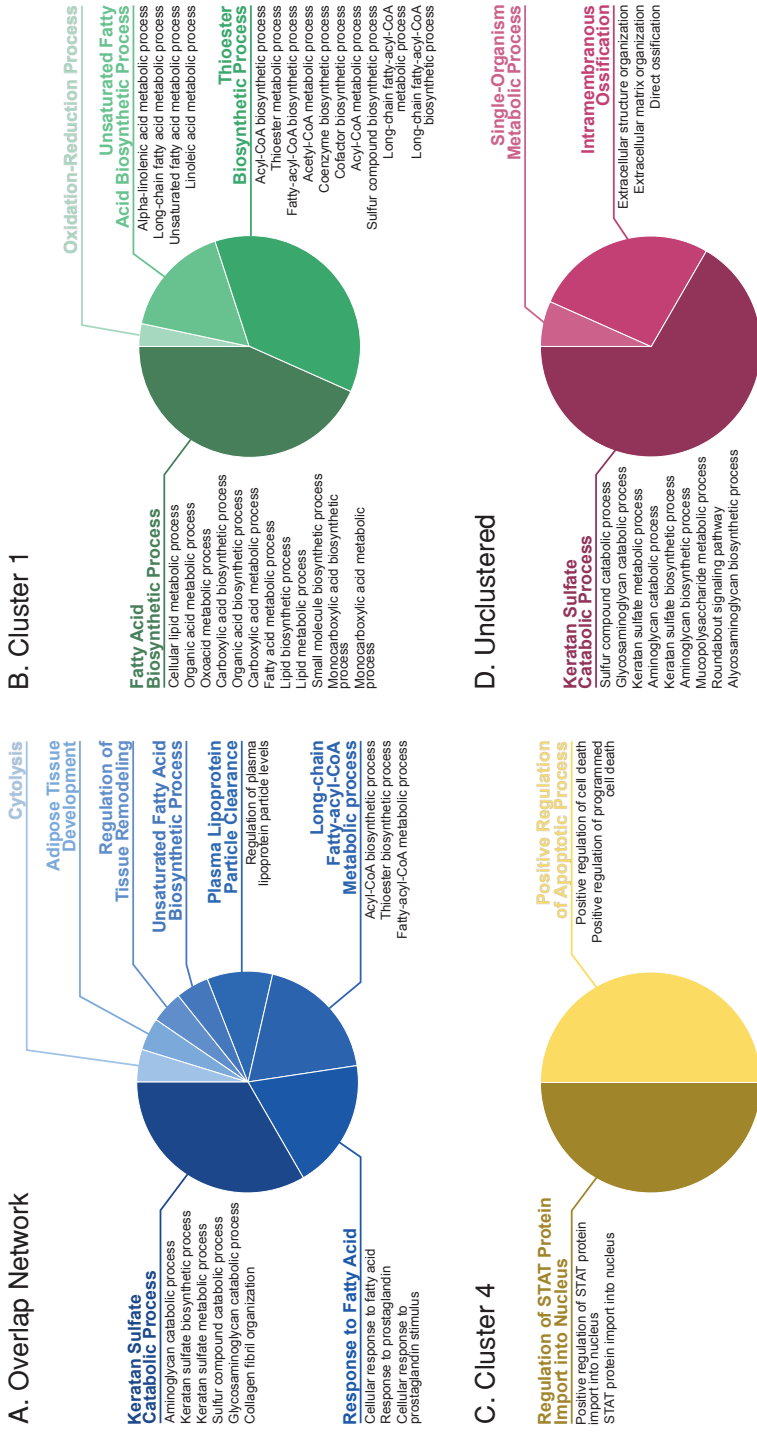


Figure 2.5: Gene ontology term pie charts constructed using the ClueGO results. Each pie chart contains major GO terms as slices, with sub-terms listed underneath each. (A) GO term pie chart for the overlap network; (B) GO term pie chart for Cluster 1; (C) GO term pie chart for Cluster 4; (D) GO term pie chart for un-clustered gene pairs.

of the co-expression networks of the diets, provides a set of genes having the same expression pattern in both diets. This set of genes and their expression pattern represents a behaviour, which, based on our results, is likely to always occur with weight loss irrespective of the caloric restriction, or at least for diets between 1,250 kcal/day and 500 kcal/day. In that respect, this set is ideal for construction of adipose tissue regulatory behaviours involved in weight loss, and can be then extended to other differentially expressed genes found in the respective diets, as can be seen by the processes found and highlighted in the results of the GO enrichment analysis.

Additionally, the expression patterns of the genes were plotted for each cluster using their \log_2 fold change within each diet. The plots, provided as Figure S5, show either the similarity or the mirroring of the expression pattern of each gene with respect to the rest of the genes within the cluster. Interestingly, the plots show that much of the expression pattern clustering is dependent on the change in expression pattern between the first two time points (during weight loss) with variations seen in the patterns between the last two time points (during weight maintenance). Across the 71 genes of the overlap network, PDK4, C6, C7, ADH1B, SRPX, GPNMB, SLC7A7, MMP2, OGN, LUM, and CTSK were found to be upregulated after weight loss, with the remaining genes downregulated. After weight maintenance, however, the gene expression pattern has variations with many gene expression patterns showing a reversal in trend but some genes also maintaining their expression patterns achieved during weight loss such as MMP2, OGN and LUM.

In addition, we also explored pathway enrichment methods which gave functionally very similar results to GO enrichment. The complete list of enriched pathways is provided as Table S5. Most of the enriched pathways and processes relate to metabolism and have been found crucial in obesity, T2DM, CVD and MetS [4, 28]. The enriched results include fatty acid biosynthesis pathways such as the downregulation of the acetyl-CoA metabolic network, previously described in [29]. We also confirmed the involvement of signalling pathways such as AMPK [30, 31] and PPAR signalling [32, 33]. Omega-3 and Omega-6 fatty acid metabolism pathways, as well as the complement system/cascade, were also found to be enriched in our results. Omega-3 have been proposed to be involved in lipid metabolism and adipokine regulation, both affected themselves in obesity and metabolic syndromes [34]. Omega-6 fatty acids have been found to be involved in anti-inflammatory effects [35]. However, high Omega-6 to Omega-3 ratio has also been found to increase the risk for obesity [36]. The complement cascade/system has been found to be dysregulated in obesity and associated co-morbidities, linking inflammatory effects, insulin resistance and impaired metabolism [37]. Additionally, other pathways such as ‘reversal of insulin resistance by leptin’ were also enriched for the overlap network, indicating that certain beneficial pathways are triggered

irrespective of the level and duration of caloric restriction. However, the overlap network only indicates a core set of pathways, and the intensity of the effect of the respective pathways and biological processes will differ when taken together with the extended networks unique to each diet.

Detailed literature study of the 71 genes in the overlap network yielded additional details regarding the involvement of these genes and their products in adipose tissue metabolism. Interestingly, the clusters of these genes showed unifying themes for most of the members of the respective clusters, indicating possible points of cross-talk of different cellular processes in the larger backdrop of adipose tissue metabolism and energy homeostasis. Additionally, some genes and their products have not been profiled for any particular cellular task or process within the context of obesity or adipose tissue metabolism. However, since these genes were found to be part of their respective clusters, the possible cross-talks of these genes allows for new avenues of research. The gene symbols of all the clustered and unclustered genes of the overlap network are provided as Table S4. In Cluster 1, the pyruvate dehydrogenase kinase 4 (PDK4) and acetyl-coenzyme A carboxylase 1 (ACACA) genes are clustered together, representing the upstream regulators of the tricarboxylic acid (TCA) cycle. The protein kinase PDK4 has previously been identified as an inhibitory regulator of pyruvate dehydrogenase complex (PDC) [38] which converts pyruvate and co-enzyme A (CoA) into acetyl-CoA for the TCA cycle. On the other hand, ACACA uses acetyl-CoA for fatty acid biosynthesis by converting it to malonyl-CoA [39]. Stearoyl-CoA desaturase (SCD), fatty acid desaturase 1 and 2 (FADS1 and FADS2 respectively), were also clustered with PDK4 and ACACA. The protein SCD has been shown to take part in triglyceride storage in white adipose tissue [40]. Furthermore, SCD, FADS1 and FADS2 are involved in the desaturation of fatty acids, and, as such, are linked to de novo fatty acid synthesis.

The genes constituting the second cluster of the overlap network (Cluster 2 in Figure 2.4) are also predominantly linked with metabolism, lipogenesis and lipolysis, in particular diacylglycerol O-acyltransferase 2 (DGAT2), zinc finger protein 219 (ZNF219), and fatty acid desaturase 3 (FADS3). However, considering the filtering of the genes and the study design of the source data, which is to be expected. Next to genes related to this common theme, links to tissue proliferation and differentiation were found in the cluster as well. G1/S-specific cyclin-D1 (CCND1), integrin alpha-7 precursor (ITGA7), transmembrane protein 120B (TMEM120B) and transmembrane protein 184B (TMEM184B) all have been associated with tissue proliferation and differentiation [41, 42].

The genes in Cluster 3 represent closer links to inter-cell signalling and immune system response apart from adipose tissue metabolism. Complement

component C6 and C7 play a role in innate immune response by forming the terminal membrane attack complex (MAC) [43, 44]. FAT atypical cadherin 1 (FAT) is involved in cell-cell signalling and has been shown to be affected by diet [45]. Adenylate cyclase-associated protein 2 (CAP2), kelch-like family member 31 (KLHL31) and klotho beta (KLB) are also involved in signal transduction through hedgehog, IFN-JAK-STAT and MAPK pathways [46, 47]. Except for these and alcohol dehydrogenase 1B (ADH1B), malic enzyme 1 (ME1) and ethylmalonyl-CoA decarboxylase (ECHDC1), other expressed genes in the cluster have little known information in the context of obesity and chronic diseases.

Relative to the previous three clusters, the genes in Cluster 4 show a more diverse profile in terms of intra-, inter- and extra-cellular biological processes. Perhaps the most important gene expression in the cluster is that of leptin (LEP), an adipokine responsible for organism-wide signalling through multiple signalling pathways [48, 49]. Through these signalling pathways, LEP has been implicated in regulation of tissue and organism-wide metabolism, energy homeostasis, fat storage and inflammation [50, 51]. The organism-wide effects are extended by a very low density lipoprotein receptor (VLDLR), which facilitates removal of circulating very low density lipoproteins [52], and has also been found to induce adipose tissue inflammation [53]. Other genes found in this cluster that are known to be involved in signal transduction include synuclein gamma (SNCG) [54], and erb-B2 receptor tyrosine kinase 4 (ERBB4) [55, 56].

Cluster 5 was the smallest of all clusters, consisting of only five members: follistatin like 3 (FSTL3), 24-dehydrocholesterol reductase (DHCR24), lysyl oxidase like 2 (LOXL2), tubulin beta 2A class IIa (TUBB2A) and aldo-keto reductase family 1 member C2 (AKR1C2). FSTL3 is a high-affinity inhibitor of transforming growth factor beta (TGF β) family members, including activin A and myostatin, and thus is involved in modulating glucose homeostasis [57]. DHCR24 binds to and protects p53 from degradation, shielding the cell from oxidative stress [58]. LOXL2 and TUBB2A appear to be involved with cellular and tissue morphology via extracellular matrix cross-linking [59] and microtubule component synthesis, respectively. AKR1C2 is known to be associated with central obesity and association with long-term weight gain in men has been suggested [60, 61]. In addition, changes of the AKR1C2 protein in the adipose tissue during weight maintenance were found to positively correlate with body weight, waist, BMI and plasma low density lipoprotein (LDL), and to correlate negatively with plasma LEP [62]. Nevertheless, the exact role of the expression of this gene in terms of cellular processes related to obesity is unknown [60].

In summary, our study found several pathways known to be involved with

obesity and associated co-morbidities, along with clusters of genes representing different but related molecular and biochemical processes in the adipose tissue cells. Many of these pathways have been known to be associated with several dysregulations in obesity; however, the interactions and cross-talk between these pathways and cellular functions need to be further explored to better understand the cellular and tissue wide dysregulation in obesity and associated chronic diseases. Collectively, the presented results provide directions for future research and exploration of obesity related chronic diseases through predictive modelling and analyses.

Conclusions

In conclusion, our article presents a workflow for finding candidate regulatory genes and processes using differential gene expression and expression pattern over time using a time series dataset. We have shown that this workflow is able to isolate several biological processes and pathways having known links with obesity, T2DM, CVD, and metabolic syndrome, allowing future analyses and predictive modelling focused on these particular biological processes. The workflow presented here is flexible by design and uses only freely available tools that can be easily connected. It is applicable to any form of time series expression data such as RNA-seq data and non-coding RNA expression arrays and it effectively accommodates inter-sample variability when constructing the correlation networks. This allows network based analyses of human intervention studies and datasets for which this was previously difficult.

References

- [1] "Global status report on noncommunicable diseases 2014," tech. rep., World Health Organization, Geneva, Switzerland, 2014.
- [2] C. Power, S. M. P. Pereira, C. Law, and M. Ki, "Obesity and risk factors for cardiovascular disease and type 2 diabetes: Investigating the role of physical activity and sedentary behaviour in mid-life in the 1958 british cohort," *Atherosclerosis*, vol. 233, no. 2, pp. 363 – 369, 2014.
- [3] L. F. Van Gaal, I. L. Mertens, and C. E. De Block, "Mechanisms linking obesity with cardiovascular disease," *Nature*, vol. 444, pp. 875–880, 2006.
- [4] J. Kaur, "A comprehensive review on metabolic syndrome," *Cardiology Research and Practice*, vol. 2014, 2014.
- [5] C. J. Lavie, R. V. Milani, and H. O. Ventura, "Obesity and cardiovascular disease risk factor, paradox, and impact of weight loss," *Journal of the American College of Cardiology*, vol. 53, no. 21, pp. 1925–1932, 2009.
- [6] S. Klein, N. F. Sheard, X. Pi-Sunyer, A. Daly, J. Wylie-Rosett, K. Kulkarni, and N. G. Clark, "Weight management through lifestyle modification for the prevention and management of type 2 diabetes: Rationale and strategies," *Diabetes Care*, vol. 27, no. 8, pp. 2067–2073, 2004.
- [7] "Diet, nutrition and the prevention of chronic diseases. report of the joint WHO/FAO expert consultation (WHO technical report series), no. 916 (trs 916)," tech. rep., World Health Organization, Geneva, Switzerland, 2003.
- [8] C. Leyvraz, C. Verdumo, M. Suter, A. Paroz, J.-M. Calmes, P. M. Marques-Vidal, and V. Giusti, "Changes in gene expression profile in human subcutaneous adipose tissue during significant weight loss," *Obesity Facts*, vol. 5, pp. 440–451, 2012.
- [9] K. L. Campbell, K. E. Foster-Schubert, K. W. Makar, M. Kratz, D. Hagman, E. A. Schur, N. Habermann, M. Horton, C. Abbenhardt, L.-Y. Kuan, L. Xiao, J. Davison, M. Morgan, C.-Y. Wang, C. Duggan, A. McTiernan, and C. M. Ulrich, "Gene expression changes in adipose tissue with diet- and/or exercise-induced weight loss," *Cancer Prevention Research*, vol. 6, no. 3, pp. 217–231, 2013.
- [10] P. A. Kern, S. Ranganathan, C. Li, L. Wood, and G. Ranganathan, "Adipose tissue tumor necrosis factor and interleukin-6 expression in human obesity and insulin resistance," *American Journal of Physiology - Endocrinology and Metabolism*, vol. 280, no. 5, pp. E745–E751, 2001.
- [11] Z. Bar-Joseph, A. Gitter, and I. Simon, "Studying and modelling dynamic biological processes using time-series gene expression data," *Nature Review Genetics*, vol. 13, pp. 552–564, 8 2012.
- [12] M. A. García-Campos, J. Espinal-Enríquez, and E. Hernández-Lemus, "Pathway analysis: State of the art," *Frontiers in Physiology*, vol. 6, p. 383, 2015.
- [13] P. Khatri, M. Sirota, and A. J. Butte, "Ten years of pathway analysis: Current approaches and outstanding challenges," *PLOS Computational Biology*, vol. 8, pp. 1–10, 02 2012.
- [14] L. M. T. Eijssen, M. Jaillard, M. E. Adriaens, S. Gaj, P. J. de Groot, M. Müller, and C. T. Evelo, "User-friendly solutions for microarray quality control and pre-processing on arrayanalysis.org," *Nucleic Acids Research*, vol. 41, no. W1, pp. W71–W76, 2013.

-
- [15] R Development Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. ISBN 3-900051-07-0.
- [16] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker, "Cytoscape: A software environment for integrated models of biomolecular interaction networks," *Genome Research*, vol. 13, no. 11, pp. 2498–2504, 2003.
- [17] R. G. Vink, N. J. T. Roumans, L. A. J. Arkenbosch, E. C. M. Mariman, and M. A. van Baak, "The effect of rate of weight loss on long-term weight regain in adults with overweight and obesity," *Obesity*, vol. 24, no. 2, pp. 321–327, 2016.
- [18] J. D. Storey, with contributions from, A. J. Bass, A. Dabney, and D. Robinson, *qvalue: Q-value estimation for false discovery rate control*, 2015. R package version 2.8.0.
- [19] L. L. Elo and B. Schwikowski, "Analysis of time-resolved gene expression measurements across individuals," *PLoS ONE*, vol. 8, pp. 1–8, 12 2013.
- [20] G. Su, A. Kuchinsky, J. H. Morris, D. J. States, and F. Meng, "Glay: community structure analysis of biological networks," *Bioinformatics*, vol. 26, no. 24, pp. 3135–3137, 2010.
- [21] J. H. Morris, L. Apeltsin, A. M. Newman, J. Baumbach, T. Wittkop, G. Su, G. D. Bader, and T. E. Ferrin, "clustermaker: a multi-algorithm clustering plugin for cytoscape," *BMC Bioinformatics*, vol. 12, no. 1, p. 436, 2011.
- [22] G. Bindea, B. Mlecnik, H. Hackl, P. Charoentong, M. Tosolini, A. Kirilovsky, W.-H. Fridman, F. Pagès, Z. Trajanoski, and J. Galon, "Cluego: a cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks," *Bioinformatics*, vol. 25, no. 8, pp. 1091–1093, 2009.
- [23] A. Kamburov, C. Wierling, H. Lehrach, and R. Herwig, "Consensuspathdb—a database for integrating human functional interaction networks," *Nucleic Acids Research*, vol. 37, no. suppl 1, pp. D623–D628, 2009.
- [24] M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, and G. K. Smyth, "limma powers differential expression analyses for RNA-sequencing and microarray studies," *Nucleic Acids Research*, vol. 43, no. 7, p. e47, 2015.
- [25] A. Kamburov, U. Stelzl, H. Lehrach, and R. Herwig, "The consensuspathdb interaction database: 2013 update," *Nucleic Acids Research*, vol. 41, no. D1, pp. D793–D800, 2013.
- [26] A. Kamburov, K. Pentchev, H. Galicka, C. Wierling, H. Lehrach, and R. Herwig, "Consensuspathdb: toward a more complete picture of cell biology," *Nucleic Acids Research*, vol. 39, no. suppl 1, pp. D712–D717, 2011.
- [27] R. G. Vink, N. J. Roumans, P. Fazelzadeh, S. H. Tareen, M. V. Boekschoten, M. A. van Baak, and E. C. Mariman, "Adipose tissue gene expression is differentially regulated with different rates of weight loss in overweight and obese humans," *International Journal of Obesity*, 12 2016.
- [28] J. A. Menendez, A. Vazquez-Martin, F. J. Ortega, and J. M. Fernandez-Real, "Fatty acid synthase: Association with insulin resistance, type 2 diabetes, and cancer," *Clinical Chemistry*, vol. 55, no. 3, pp. 425–438, 2009.

-
- [29] H. Dharuri, P. A. C. 't Hoen, J. B. van Klinken, P. Henneman, J. F. J. Laros, M. A. Lips, F. el Bouazzaoui, G.-J. B. van Ommen, I. Janssen, B. van Ramshorst, B. A. van Wagenveld, H. Pijl, K. Willems van Dijk, and V. van Harmelen, "Downregulation of the acetyl-coa metabolic network in adipose tissue of obese diabetic individuals and recovery after weight loss," *Diabetologia*, vol. 57, no. 11, pp. 2384–2392, 2014.
- [30] R. W. Mackenzie and B. T. Elliott, "Akt/pkb activation and insulin signaling: a novel insulin signaling pathway in the treatment of type 2 diabetes," *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy*, vol. 7, pp. 55–64, 2 2014.
- [31] D. G. Hardie, "Role of amp-activated protein kinase in the metabolic syndrome and in heart disease," *FEBS Letters*, vol. 582, no. 1, pp. 81 – 89, 2008.
- [32] S. Tyagi, P. Gupta, A. Saini, C. Kaushal, and S. Sharma, "The peroxisome proliferator-activated receptor: A family of nuclear receptors role in various diseases," *Journal of Advanced Pharmaceutical Technology and Research*, vol. 2, no. 4, pp. 236–240, 2011.
- [33] A. Leonardini, L. Laviola, S. Perrini, A. Natalicchio, and F. Giorgino, "Cross-talk between ppar and insulin signaling and modulation of insulin sensitivity," *PPAR Research*, vol. 2009, pp. 1–12, 2009.
- [34] K. Albracht-Schulte, N. S. Kalupahana, L. Ramalingam, S. Wang, S. M. Rahman, J. Robert-McComb, and N. Moustaid-Moussa, "Omega-3 fatty acids in obesity and metabolic syndrome: a mechanistic update," *The Journal of Nutritional Biochemistry*, vol. 58, pp. 1 – 16, 2018.
- [35] E. Tortosa-Caparrós, D. Navas-Carrillo, F. Marín, and E. Orenes-Piñero, "Anti-inflammatory effects of omega 3 and omega 6 polyunsaturated fatty acids in cardiovascular disease and metabolic syndrome," *Critical Reviews in Food Science and Nutrition*, vol. 57, no. 16, pp. 3421–3429, 2017. PMID: 26745681.
- [36] A. P. Simopoulos, "An increase in the omega-6/omega-3 fatty acid ratio increases the risk for obesity," *Nutrients*, vol. 8, no. 3:128, 2016.
- [37] J. M. Moreno-Navarrete and J. M. Fernández-Real, "The complement system is dysfunctional in metabolic disease: Evidences in plasma and adipose tissue from obese and insulin resistant subjects," *Seminars in Cell & Developmental Biology*, 2017.
- [38] R. R. Attia, P. Sharma, R. C. Janssen, J. E. Friedman, X. Deng, J. S. Lee, M. B. Elam, G. A. Cook, and E. A. Park, "Regulation of pyruvate dehydrogenase kinase 4 (PDK4) by ccaat/enhancer-binding protein β (c/ebp β)," *Journal of Biological Chemistry*, vol. 286, no. 27, pp. 23799–23807, 2011.
- [39] C. L. Colbert, C.-W. Kim, Y.-A. Moon, L. Henry, M. Palnitkar, W. B. McKean, K. Fitzgerald, J. Deisenhofer, J. D. Horton, and H. J. Kwon, "Crystal structure of spot 14, a modulator of fatty acid synthesis," *Proceedings of the National Academy of Sciences*, vol. 107, no. 44, pp. 18820–18825, 2010.
- [40] N. S. Kalupahana, T. Jayalath, S. Wang, and N. Moustaid-Moussa, "Regulation and metabolic functions of white adipose tissue stearyl-coa desaturase," in *Stearyl-CoA Desaturase Genes in Lipid Metabolism*, pp. 49–60, Springer, 2013.

-
- [41] Y. Haim, M. Blüher, N. Slutsky, N. Goldstein, N. Klötting, I. Harman-Boehm, B. Kirshtein, D. Ginsberg, M. Gericke, E. G. Jurado, J. Kovsan, T. Tarnovscki, L. Kachko, N. Bashan, Y. Gepner, I. Shai, and A. Rudich, "Elevated autophagy gene expression in adipose tissue of obese humans: A potential non-cell-cycle-dependent function of e2f1," *Autophagy*, vol. 11, no. 11, pp. 2074–2088, 2015. PMID: 26391754.
- [42] D. G. Batrakou, J. I. de las Heras, R. Czapiewski, R. Mouras, and E. C. Schirmer, "Tmem120a and b: Nuclear envelope transmembrane proteins important for adipocyte differentiation," *PLOS ONE*, vol. 10, pp. 1–21, 05 2015.
- [43] J. Phielers, R. Garcia-Martin, J. D. Lambris, and T. Chavakis, "The role of the complement system in metabolic organs and metabolic diseases," *Seminars in Immunology*, vol. 25, pp. 47–53, 05 2013.
- [44] S. Kaye, A. I. Lokki, A. Hanttu, E. Nissilä, S. Heinonen, A. Hakkarainen, J. Lundbom, N. Lundbom, L. Saarinen, O. Tynninen, M. Muniandy, A. Rissanen, J. Kaprio, S. Meri, and K. H. Pietiläinen, "Upregulation of early and downregulation of terminal pathway complement genes in subcutaneous adipose tissue and adipocytes in acquired obesity," *Frontiers in Immunology*, vol. 8, p. 545, 2017.
- [45] A. Perfilyev, I. Dahlman, L. Gillberg, F. Rosqvist, D. Iggman, P. Volkov, E. Nilsson, U. Risérus, and C. Ling, "Impact of polyunsaturated and saturated fat overfeeding on the dna-methylation pattern in human adipose tissue: a randomized controlled trial," *The American Journal of Clinical Nutrition*, vol. 105, no. 4, pp. 991–1000, 2017.
- [46] A. Moisan, Y.-K. Lee, J. D. Zhang, C. S. Hudak, C. A. Meyer, M. Prummer, S. Zoffmann, H. H. Truong, M. Ebeling, A. Kiialainen, R. Gérard, F. Xia, R. T. Schinzel, K. E. Amrein, and C. A. Cowan, "White-to-brown metabolic conversion of human adipocytes by jak inhibition," *Nature Cell Biology*, vol. 17, pp. 57–67, 01 2015.
- [47] D. V. Lee, D. Li, Q. Yan, Y. Zhu, B. Goodwin, R. Calle, M. B. Brenner, and S. Talukdar, "Fibroblast growth factor 21 improves insulin sensitivity and synergizes with insulin in human adipose stem cell-derived (hasc) adipocytes," *PLOS ONE*, vol. 9, pp. 1–11, 11 2014.
- [48] Y. Zhou and L. Rui, "Leptin signaling and leptin resistance," *Frontiers of Medicine*, vol. 7, no. 2, pp. 207–222, 2013.
- [49] R. Yang and L. A. Barouch, "Leptin signaling and obesity," *Circulation Research*, vol. 101, no. 6, pp. 545–559, 2007.
- [50] G. R. Hajer, T. W. van Haeften, and F. L. Visseren, "Adipose tissue dysfunction in obesity, diabetes, and vascular diseases," *European Heart Journal*, vol. 29, no. 24, p. 2959, 2008.
- [51] K. Makki, P. Froguel, and I. Wolowczuk, "Adipose tissue in obesity-related inflammation and insulin resistance: Cells, cytokines, and chemokines," *International Scholarly Research Notices Inflammation*, vol. 2013, no. 139239, p. 12, 2013.
- [52] M. Clemente-Postigo, M. I. Queipo-Ortuño, D. Fernandez-Garcia, R. Gomez-Huelgas, F. J. Tinahones, and F. Cardona, "Adipose tissue gene expression of factors related to lipid processing in obesity," *PLOS ONE*, vol. 6, pp. 1–8, 09 2011.
- [53] A. Nguyen, H. Tao, M. Metriane, and T. Hajri, "Very low density lipoprotein receptor (vldlr) expression is a determinant factor in adipose tissue inflammation and adipocyte-macrophage interaction," *Journal of Biological Chemistry*, vol. 289, no. 3, pp. 1688–1703, 2014.

-
- [54] T. N. Dunn, T. Akiyama, H. W. Lee, J. B. Kim, T. A. Knotts, S. R. Smith, D. D. Sears, E. Carstens, and S. H. Adams, "Evaluation of the synuclein- γ (sn γ) gene as a ppar γ target in murine adipocytes, dorsal root ganglia somatosensory neurons, and human adipose tissue," *PLOS ONE*, vol. 10, pp. 1–19, 03 2015.
- [55] G.-X. Wang, X.-Y. Zhao, and J. D. Lin, "The brown fat secretome: metabolic functions beyond thermogenesis," *Trends in Endocrinology & Metabolism*, vol. 26, no. 5, pp. 231–237, 2015.
- [56] Y. Ma, M. Gao, and D. Liu, "Preventing high fat diet-induced obesity and improving insulin sensitivity through neuregulin 4 gene transfer," *Scientific Reports*, vol. 6, 05 2016.
- [57] M. L. Brown, L. Bonomi, N. Ungerleider, J. Zina, F. Kimura, A. Mukherjee, Y. Sidis, and A. Schneyer, "Follistatin and follistatin like-3 differentially regulate adiposity and glucose homeostasis," *Obesity*, vol. 19, no. 10, pp. 1940–1949, 2011.
- [58] S. Z. Berisha, D. Serre, P. Schauer, S. R. Kashyap, and J. D. Smith, "Changes in whole blood gene expression in obese subjects with type 2 diabetes following bariatric surgery: a pilot study," *PLOS ONE*, vol. 6, pp. 1–9, 03 2011.
- [59] Y.-M. Kim, E.-C. Kim, and Y. Kim, "The human lysyl oxidase-like 2 protein functions as an amine oxidase toward collagen and elastin," *Molecular Biology Reports*, vol. 38, no. 1, pp. 145–149, 2011.
- [60] D. J. Wake, M. Strand, E. Rask, J. Westerbacka, D. E. W. Livingstone, S. Soderberg, R. Andrew, H. Yki-Jarvinen, T. Olsson, and B. R. Walker, "Intra-adipose sex steroid metabolism and body fat distribution in idiopathic human obesity," *Clinical Endocrinology*, vol. 66, no. 3, pp. 440–446, 2007.
- [61] F. G. Bouwman, J. M. A. Boer, S. Imholz, P. Wang, W. M. M. Verschuren, M. E. T. Dollé, and E. C. M. Mariman, "Gender-specific genetic associations of polymorphisms in ace, akr1c2, fto and mmp2 with weight gain over a 10-year period," *Genes & Nutrition*, vol. 9, no. 6, p. 434, 2014.
- [62] F. G. Bouwman, M. Claessens, M. A. van Baak, J.-P. Noben, P. Wang, W. H. M. Saris, and E. C. M. Mariman, "The physiologic effects of caloric restriction are reflected in the in vivo adipocyte-enriched proteome of overweight/obese subjects," *Journal of Proteome Research*, vol. 8, no. 12, pp. 5532–5540, 2009. PMID: 19827857.

Supplementary Materials

The following supplementary material is hosted by *Genes* at <https://doi.org/10.3390/genes9110525> (PDF, 1027KB).

Supplementary Figures

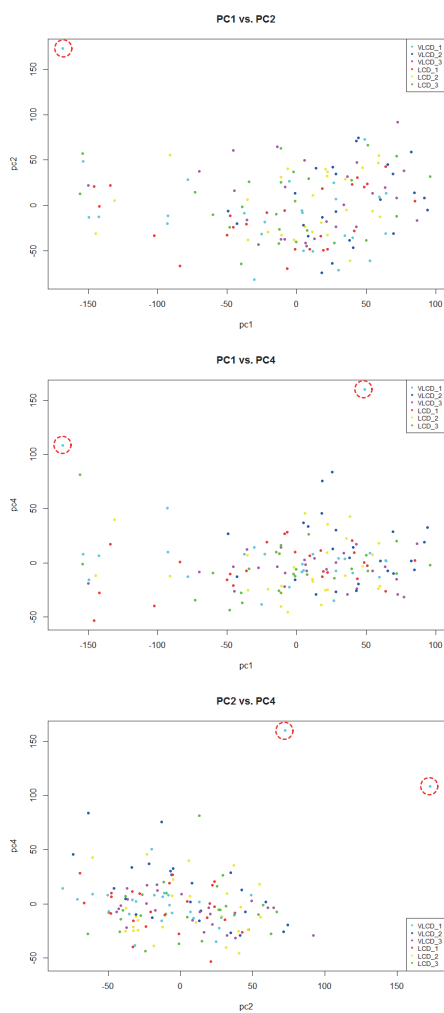


Figure 2.S1: Principal component analysis of the raw gene expression data of the 46 individuals, showing the outlying samples highlighted in red circles.

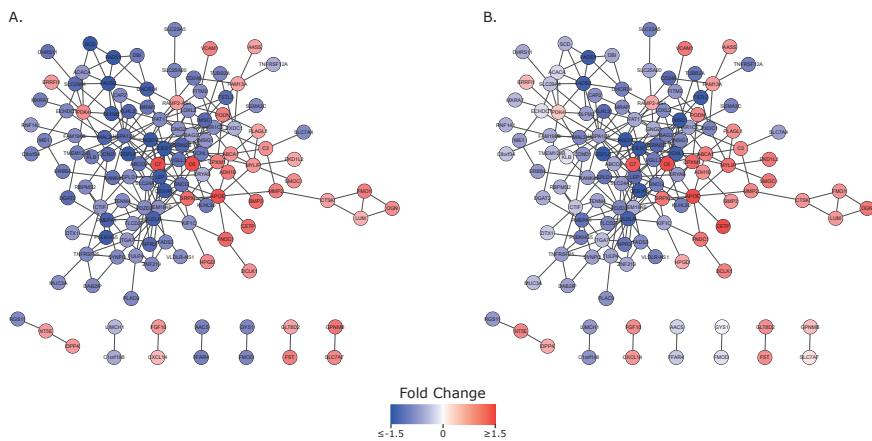


Figure 2.S2: Correlation network generated for the low calorie diet (LCD). A) Fold change for time points 2-1. B) Fold change for time points 3-1.

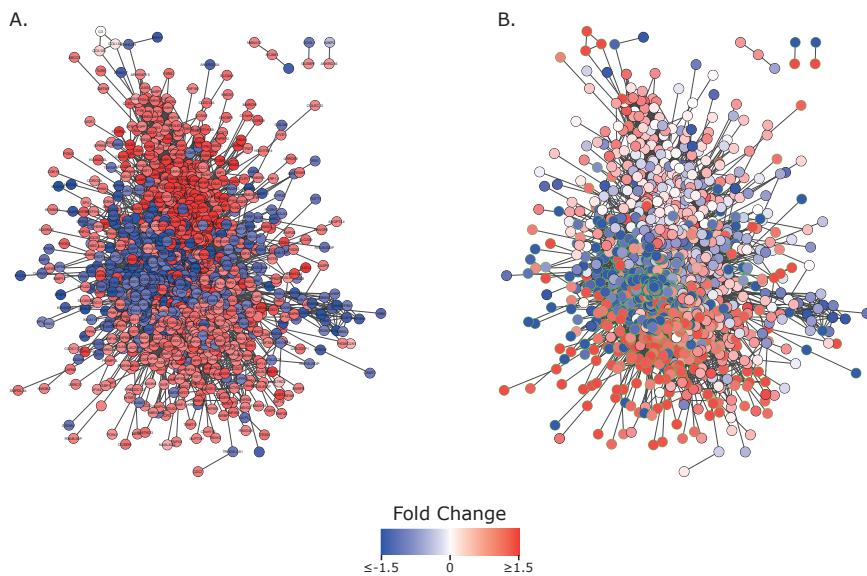


Figure 2.S3: Correlation network generated for the very low calorie diet (VLCD). A) Fold change for time points 2-1. B) Fold change for time points 3-1.

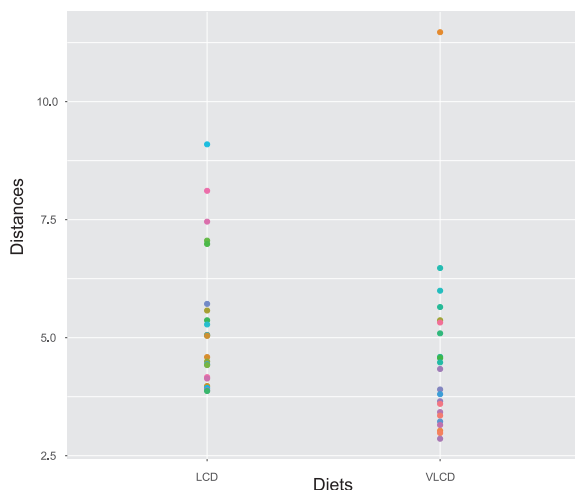


Figure 2.S4: Frobenius norm plot showing the norm distance of the correlation network of the 44 individuals with the respective diet-wide correlation network.

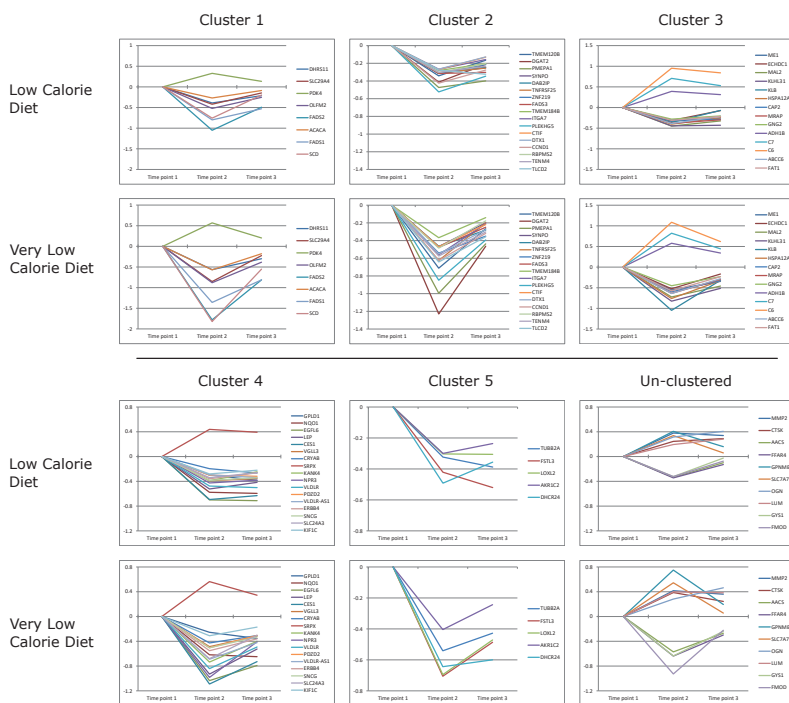


Figure 2.S5: Gene expression patterns of the respective genes in each topological cluster of the overlap network, based on the fold change in the respective diets. Also shows the five un-clustered paired correlated genes.

Supplementary Tables

Table 2.S1: Results from the robustness analysis using the minimum %age individuals having the same correlation direction (for $|\text{corr.}| \geq 0.6$).

%age min. individuals		LCD	VLCD
25	Total edges	127	127
	%age	100.00	100.00
30	Total edges	127	126
	%age	100.00	99.21
35	Total edges	127	126
	%age	100.00	99.21
40	Total edges	125	126
	%age	98.43	99.21
45	Total edges	119	125
	%age	93.70	98.43
50	Total edges	108	120
	%age	85.04	94.49

Table 2.S2: Results from the leave-n-out cross-validation analysis for $n=1$ and 2 (max 10% of total samples) in the overlap network. The analysis here represents loss of sensitivity. The correlation criteria is $|\text{corr.}| \geq 0.6$ in at least 25% of the individuals having the same direction of the correlation.

Leave-n-out, n=	# of combinations in each diet	Criteria: $ \text{corr.} \geq 0.6$ and min. 25% ind.s	LCD	VLCD
0	1	Total edges	127	127
		%age	100.00	100.00
1	22	Total edges	85	117
		%age	66.93	92.13
2	231	Total edges	52	104
		%age	40.94	81.89

Table 2.S3: Results from the leave-n-out cross-validation analysis for n= 1 and 2 (max 10% of total samples) in the correlation networks of the respective diets. The analysis here represents loss of specificity. The correlation criteria is $|\text{corr.}| \geq 0.6$ in at least 25% of the individuals having the same direction of the correlation.

Leave-n-out, n=	# of combinations in each diet	Criteria: $ \text{corr.} \geq 0.6$ and min. 25% ind.s	LCD	VLCD
0	1	Total edges	149	1125
1	22	New edges (avg. of all combinations)	20	54
2	231	New edges (avg. of all combinations)	~36	~82

Table 2.S4: List of genes in the respective clusters from the overlap network, represented here in HGNC symbols.

Clusters	Genes
1	ACACA, DHRS11, FADS1, FADS2, OLFM2, PDK4, SCD, SLC29A4
2	CCND1, CTIF, DAB2IP, DGAT2, DTX1, FADS3, ITGA7, PLEKHG5, PMEPA1, RBPMS2, SYNPO, TENM4, TLCD2, TMEM120B, TMEM184B, TNFRSF25, ZNF219
3	ABCC6, ADH1B, C6, C7, CAP2, ECHDC1, FAT1, GNG2, HSPA12A, KLB, KLHL31, MAL2, ME1, MRAP
4	GPLD1, KANK4, KIF1C, LEP, NPR3, NQO1, PDZD2, SLC24A3, SNCG, SRPX, VGLL3, VLDLR, VLDLR-AS1
5	AKR1C2, DHCR24, FSTL3, LOXL2, TUBB2A
Un-clustered	AACS, CTSK, FFAR4, FMOD, GPNMB, GYS1, LUM, MMP2, OGN, SLC7A7

Table 2.S5: Results from the ConsensusPathDB analysis on the overlap network. BioCarta leptin pathway discussed in the discussion section of the article is highlighted in red.

pathway	p-value	q-value	source	members_input _overlap	members_input _overlap_genes	size	effective _size
Omega-3 fatty acid metabolism	4.84E-05	0.00781987	EHMN	ENSG00000221968; ENSG00000134824; ENSG00000149485	3995; 3992; 9415	16	16
Fatty Biosynthesis	0.000130585	0.00781987	Wiki pathways	ENSG00000099194; ENSG00000093144; ENSG00000278540	55862; 31; 6319	22	22
Biosynthesis of unsaturated fatty acids - Homo sapiens (human)	0.000149689	0.00781987	KEGG	ENSG00000099194; ENSG00000134824; ENSG00000149485	3992; 6319; 9415	23	23
AMPK signaling pathway - Homo sapiens (human)	0.000235714	0.00781987	KEGG	ENSG00000104812; ENSG00000099194; ENSG00000174697; ENSG00000110092; ENSG00000278540	6319; 31; 595; 3952; 2997	124	123
Omega-6 fatty acid metabolism	0.000272466	0.00781987	EHMN	ENSG00000221968; ENSG00000134824; ENSG00000149485	3992; 3995; 9415	28	28
Keratan sulfate biosynthesis	0.000272466	0.00781987	Reactome	ENSG00000139329; ENSG00000106809; ENSG00000122176	4060; 4969; 2331	28	28
Alpha Linolenic Acid and Linoleic Acid Metabolism	0.000304106	0.00781987	SMPDB	ENSG00000134824; ENSG00000149485	9415; 3992	6	6
Terminal pathway of complement	0.000424502	0.008930523	Reactome	ENSG00000112936; ENSG00000039537	730; 729	7	7

Keratan sulfate/keratin metabolism	0.000446526	0.008930523	Reactome	ENSG00000139329; ENSG00000106809; ENSG00000122176	4060; 4969; 2331	33	33
Nuclear Receptors Meta-Pathway	0.000514786	0.009234772	WikiPathways	ENSG00000065833; ENSG00000099194; ENSG00000198848; ENSG00000110092; ENSG00000004799; ENSG00000112186; ENSG00000181019	4199; 595; 1066; 6319; 1728; 10486; 5166	314	313
Linoleic acid (LA) metabolism	0.000564347	0.009234772	Reactome	ENSG00000134824; ENSG00000149485	3992; 9415	8	8
reversal of insulin resistance by leptin	0.000723467	0.010852004	BioCarta	ENSG00000174697; ENSG00000278540	3952; 31	9	9
alternative complement pathway	0.000901689	0.012484929	BioCarta	ENSG00000112936; ENSG00000039537	730; 729	10	10
Extracellular matrix organization	0.001202898	0.015465837	Reactome	ENSG00000135424; ENSG00000143387; ENSG00000134013; ENSG00000139329; ENSG00000087245; ENSG00000122176	4017; 3679; 2331; 4060; 4313; 1513	264	263
lectin induced complement pathway	0.001314758	0.01549266	BioCarta	ENSG00000112936; ENSG00000039537	730; 729	12	12
alpha-linolenic acid (ALA) metabolism	0.001549266	0.01549266	Reactome	ENSG00000134824; ENSG00000149485	3992; 9415	13	13
alpha-linolenic (omega3) and linoleic (omega6) acid metabolism	0.001549266	0.01549266	Reactome	ENSG00000134824; ENSG00000149485	3992; 9415	13	13

icosapentaenoate biosynthesis (metazoa) II	0.001549266	0.01549266	0.01549266	HumanCyc	ENSG00000134824; ENSG00000149485	3992; 9415	13	13
classical complement pathway	0.001802198	0.017073456	0.017073456	BioCarta	ENSG00000112936; ENSG00000039537	730; 729	14	14
ECM proteoglycans	0.002000476	0.018004287	0.018004287	Reactome	ENSG00000139329; ENSG00000135424; ENSG00000122176	4060; 3679; 2331	56	55
Complement Activation, Classical Pathway	0.00266988	0.022884683	0.022884683	WikiPathways	ENSG00000112936; ENSG00000039537	730; 729	17	17
AMPK Signaling	0.003665435	0.026814921	0.026814921	WikiPathways	ENSG00000104812; ENSG00000174697; ENSG00000278540	3952; 31; 2997	68	68
PPAR signaling pathway - Homo sapiens (human)	0.003819792	0.026814921	0.026814921	KEGG	ENSG00000065833; ENSG00000099194; ENSG00000134824	6319; 9415; 4199	69	69
Linoleate metabolism	0.004306497	0.026814921	0.026814921	EHMN	ENSG00000221968; ENSG00000134824; ENSG00000149485	9415; 3992; 3995	72	72
Pyruvate Dehydrogenase Complex Deficiency	0.004469153	0.026814921	0.026814921	SMPDB	ENSG00000065833; ENSG00000278540	4199; 31	22	22
Primary hyperoxaluria II, PH2	0.004469153	0.026814921	0.026814921	SMPDB	ENSG00000065833; ENSG00000278540	4199; 31	22	22
Pyruvate kinase deficiency	0.004469153	0.026814921	0.026814921	SMPDB	ENSG00000065833; ENSG00000278540	4199; 31	22	22
Leigh Syndrome	0.004469153	0.026814921	0.026814921	SMPDB	ENSG00000065833; ENSG00000278540	4199; 31	22	22
Pyruvate Metabolism	0.004469153	0.026814921	0.026814921	SMPDB	ENSG00000065833; ENSG00000278540	4199; 31	22	22

Pyruvate Decarboxylase E1 Component Deficiency (PDHE1 Deficiency)	0.004469153	0.026814921	SMPDB	ENSG00000065833; ENSG00000278540	4199; 31	22	22
Leptin signaling pathway	0.005011674	0.028826678	Wikipathways	ENSG00000174697; ENSG00000110092; ENSG00000278540	595; 3952; 31	76	76
Metabolism	0.005124743	0.028826678	Reactome	ENSG00000151632; ENSG00000062282; ENSG00000106809; ENSG00000116133; ENSG00000196616; ENSG00000149485; ENSG00000004799; ENSG00000134824; ENSG00000139329; ENSG00000104812; ENSG00000278540; ENSG00000122176; ENSG00000186469; ENSG00000181019	1728; 2331; 4060; 1718; 2997; 31; 125; 5166; 4969; 1646; 54331; 9415; 84649; 3992	1481	1472
Regulation of Complement cascade	0.005308714	0.02895662	Reactome	ENSG00000112936; ENSG00000039537	730; 729	24	24
Propanoate metabolism - Homo sapiens (human)	0.007186387	0.038045578	KEGG	ENSG00000093144; ENSG00000278540	55862; 31	28	28
Activation of Matrix Metalloproteinases	0.008222042	0.042284788	Reactome	ENSG00000143387; ENSG00000087245	1513; 4313	30	30
Collagen degradation	0.009893155	0.049465773	Reactome	ENSG00000143387; ENSG00000087245	1513; 4313	33	33

3

Exploring the cellular network of metabolic flexibility in the adipose tissue

Samar H.K. Tareen¹, Martina Kutmon^{1,2}, Michiel E. Adriaens¹, Edwin C.M. Mariman³, Theo M. de Kok^{1,4}, Ilja C.W. Arts^{1,5} and Chris T. Evelo^{1,2}

¹ Maastricht Centre for Systems Biology (MaCSBio), Maastricht University, the Netherlands

² Department of Bioinformatics – BiGCaT, NUTRIM Research School, Maastricht University, the Netherlands

³ Department of Human Biology, NUTRIM Research School, Maastricht University, the Netherlands

⁴ Department of Toxicogenomics, GROW School of Oncology and Developmental Biology, Maastricht University, the Netherlands

⁵ Department of Epidemiology, CARIM School for Cardiovascular Diseases, Maastricht University, the Netherlands

Published in: *Genes & Nutrition* (2018) 13:17.

doi: 10.1186/s12263-018-0609-3

Abstract

Background: Metabolic flexibility is the ability of cells to change substrates for energy production based on the nutrient availability and energy requirement. It has been shown that metabolic flexibility is impaired in obesity and chronic diseases such as type 2 diabetes mellitus, cardiovascular diseases and metabolic syndrome, although, whether it is a cause or an effect of these conditions remains to be elucidated.

Main Body: In this paper we have reviewed the literature on metabolic flexibility and curated pathways and processes resulting in a network resource to investigate the interplay between these processes in the subcutaneous adipose tissue. The adipose tissue has been shown to be responsible, not only for energy storage, but also for maintaining energy homeostasis through oxidation of glucose and fatty acids. We highlight the role of pyruvate dehydrogenase complex–pyruvate dehydrogenase kinase (PDC-PDK) interaction as a regulatory switch which is primarily responsible for changing substrates in energy metabolism from glucose to fatty acids and back. Baseline gene expression of the subcutaneous adipose tissue, along with a publically available obesity data set, are visualised on the cellular network of metabolic flexibility to highlight the genes that are expressed and which are differentially affected in obesity.

Conclusion: We have constructed an abstracted network covering glucose and fatty acid oxidation, as well as the PDC-PDK regulatory switch. In addition, we have shown how the network can be used for data visualisation and as a resource for follow-up studies.

Metabolic flexibility

Metabolic flexibility is defined as the ability of an organism to adapt its substrate for energy production in cellular respiration, based on the availability of the substrates [1]. The primary substrates are glucose and fatty acids, which are converted to Acetyl-Coenzyme A (Acetyl-CoA) for use in the tricarboxylic acid cycle (TCA cycle). Cellular respiration for most tissues and organs utilises only one energy substrate at a given time; glucose during the fed state and fatty acids during the fasted state (exceptions include the brain for example). However, it has been observed that under stress and severe energy deprivation conditions this exclusivity can be broken and both glucose and fatty acids are consumed for energy production [1].

Given that metabolic flexibility is associated with maintaining a dynamic and shifting balance between the two sources of energy, it may have a prominent role in the development of metabolic diseases and associated conditions. The inability or impairment of the organism to change its source as per requirements is called metabolic inflexibility. A number of recent studies have started focusing on its association with conditions pertaining to malfunctioning metabolism, including obesity, type 2 diabetes mellitus (T2DM), cardiovascular diseases (CVD) and metabolic syndrome (MetS) [2–5]. Considering the implication of metabolic flexibility in disease development, we focus on curating the underlying cellular/molecular mechanisms in this study, specifically in the adipose tissue as several adipose tissue gene expression markers have linked it with reduced metabolic flexibility [6].

Adipose tissue holds a central role in metabolic flexibility and energy metabolism with major regulatory mechanisms and roles, both tissue- and organism-wide [7, 8]. Although adipose tissue stores the majority of the fat in the body, most of the fat is synthesised *de novo* by the liver. The adipose tissue ends up storing both the synthesised fat released by the liver, as well as dietary fat [9][9]. In addition, the adipose tissue only takes up 10-15% of circulating glucose [10]. However, this interplay and balance between glucose uptake as well fatty acid uptake and later release is the result of metabolic flexibility in the adipose tissue. Indeed, metabolic inflexibility in the adipose tissue has been known to cause impaired adipokine signalling, as well as impaired non-esterified fatty acid (NEFA) clearance from circulation, triggering NEFA mediated signalling cascades in other tissues (reviewed in [11, 12]). Thus, the impairment of metabolic flexibility in the adipose tissue can cause systemic effects with regards to energy provision and related processes.

In this review, we summarize the cellular mechanisms pertaining to metabolic flexibility in a network of interacting molecular species and processes. The major benefit of this approach is that it allows further study of the various cellular processes involved in metabolic flexibility to pinpoint crucial elements

in said systems. Similar approaches have previously been employed, for example in [13] where data and existing knowledge were collectively used to identify seemingly unrelated processes involved in adipogenesis in culture. In our review, we employ existing knowledge in terms of known pathways to curate a network representing cellular metabolic flexibility in the adipocytes. Subsequently, baseline expression data of the subcutaneous adipose tissue [14] along with expression data from a publicly available obesity dataset [15] are mapped onto the network as a use case showing the expression of the components of the network under baseline/non-diseased and obese conditions.

Biochemical pathways of metabolic flexibility

In this review, we have curated an abstracted network representing pathways of cellular metabolic flexibility through literature review and querying the WikiPathways database [16]. We started with biochemical reactions involved in glucose and fatty acid oxidation in the adipose tissue, namely the glycolysis and fatty acid β -oxidation processes, and expanded them to link them to each other via the TCA cycle. Next, rate-limiting enzymes as well as transport, signalling and regulatory proteins were included to expand upon the biochemical processes, along with their respective interactions with other components already in the network. This was followed by the addition of fatty acid synthesis downstream of the TCA cycle as a feedback mechanism to fatty acid β -oxidation. Furthermore, cellular signalling cascades known to affect cellular oxidation were also added.

Finally, to give a simplified overview and ease its understanding, the network was abstracted by only leaving in rate-limiting steps, major metabolites between said steps, and associated regulatory proteins. The exact procedure and order of reduction differs from network to network, however, the basic idea remains the same, i.e. to represent multiple nodes and/or edges by a single node and/or edge. Figure 3.1 illustrates this procedure. As an example, consider the procedure of fatty acid breakdown to release multiple Acyl-CoA molecules, which is a multi-step process involving multiple sets of enzymes and reactions. However, unless we are specifically targeting a step within this procedure, or one of the steps is a rate-limiting step under scrutiny, we can represent the whole breakdown process in an abstracted manner using a Fatty Acid node, linked to an Acyl-CoA node with an edge.

The resultant abstracted cellular network of metabolic flexibility is shown in Figure 3.2. The colour coded sections identify the major pathways with, (i) green for glycolysis related components, (ii) orange for fatty acid β -oxidation, (iii) yellow for fatty acid synthesis, (iv) cyan for the TCA cycle, and (v)

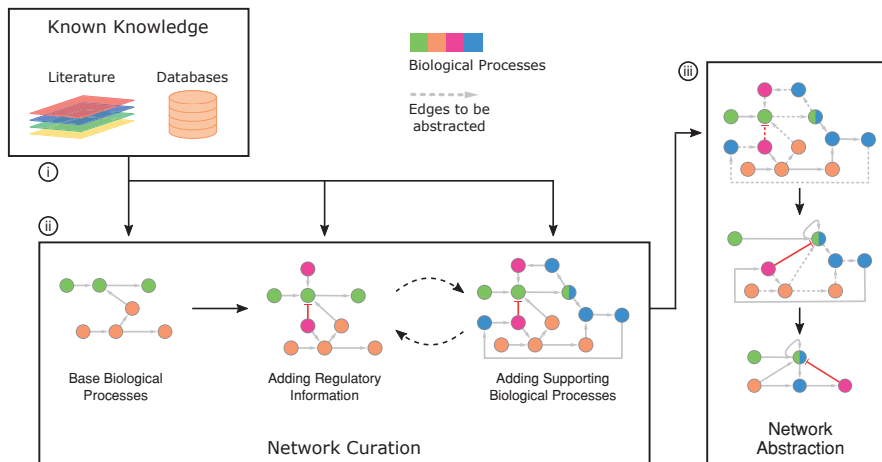


Figure 3.1: Methodology overview showing the workflow to construct the abstracted network. i) Known knowledge in the form of published literature and databases is queried regarding cellular metabolism. ii) Base biological processes are isolated, and then expanded by adding regulators and other related processes as long as they are related to cellular metabolism. iii) The expanded network is then abstracted by merging edges such that only major components and rate-limiting steps remain.

magenta for regulators of metabolic flexibility. In the abstracted network, we also see how these pathways are interacting with each other, in particular how the various products of the TCA cycle are playing roles in activating or inhibiting different pathways through feedback mechanisms. We define any interaction that activates or continues a process in the network as a positive interaction. In the network shown in Figure 3.2, these positive interactions cover transcriptional activation, allosteric activation, biochemical reactions (substrate consumption and/or product formation), protein complex formation, and species transportation. Negative interactions exclusively refer to inhibitory interactions, whether they are allosteric inhibition or transcriptional inhibition. The following subsections explain these pathways in detail.

Cellular energy production

The TCA cycle, also referred to as the citric acid cycle or Krebs cycle, is the primary biochemical pathway for cellular energy production and respiration in all aerobic cells [17]. The cycle starts with the conversion of Acetyl-CoA into citrate, continues through a series of biochemical reactions where it reduces NAD^+ to NADH, produces FADH_2 and CO_2 , and ends with the reconversion to citrate, thus starting the next iteration of the cycle. The NADH and FADH_2 are then converted into ATPs via oxidative phosphorylation. Thus,

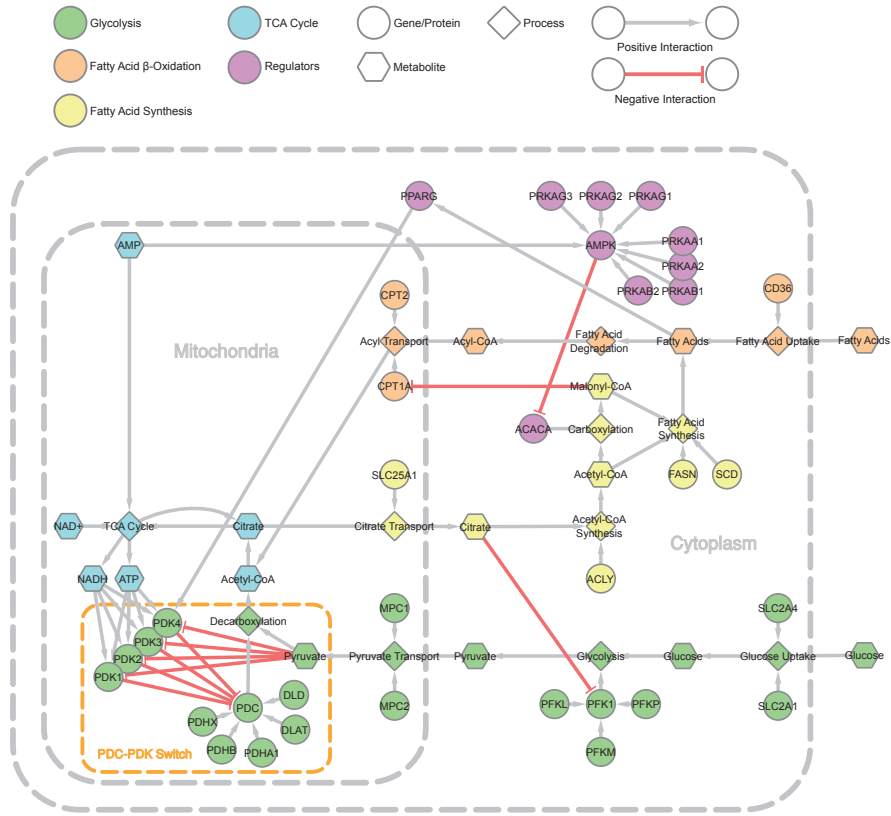


Figure 3.2: Abstracted cellular network of metabolic flexibility in the adipose tissue. The network consists of different pathways and processes, which are grouped together into five major, colour coded categories: (i) green for glycolysis, (ii) orange for fatty acid β -oxidation, (iii) yellow for fatty acid synthesis, (iv) cyan for TCA cycle, and (v) magenta for regulators of metabolic flexibility.

by consuming Acetyl-CoA, the cycle produces cellular energy in the form of ATPs and replenishes NADH concentrations in the cell. The TCA cycle can be viewed in detail at WikiPathways [16] (Pathway ID: WP78 [18]), and is shown as the cyan module in Figure 3.2. Glucose or fatty acids are consumed upstream of the TCA cycle for the generation of Acetyl-CoA, but are regulated by the TCA cycle downstream as well, forming the basis of metabolic flexibility.

Glucose uptake and oxidation

Glucose is the most readily utilisable resource for the production of Acetyl-CoA for the TCA cycle. The process starts with the uptake of glucose into the cell, which can be insulin-dependent via the SLC2A4 (also known as GLUT4)

glucose transporters, or insulin-independent via the SLC2A1 (also known as GLUT1) transporters. Glucose is then phosphorylated by phosphofructokinases (PFK1 and PFK2) and converted into pyruvate, a precursor to Acetyl-CoA, via glycolysis in the cytoplasm. PFK1 itself is composed of four subunits consisting of three subtypes; PFKL (liver type), PFKM (muscle type), and PFKP (platelets type), the combination depending on the tissue [1]. The pyruvate converted by the PFKs is then transported into the mitochondria by the mitochondrial pyruvate carriers (MPC1 and MPC2), where it is converted to Acetyl-CoA via the pyruvate dehydrogenase complex (PDC) [19]. The complete pathway including intermediate metabolites and enzymes is available at WikiPathways (Pathway ID: WP534 [18, 20]), with the abstracted pathway shown as the green module in Figure 3.2. Interestingly, PFKs are allosterically inhibited by citrate, a primary component of the TCA cycle, when the citrate is transported into the cytoplasm via the citrate carrier (SLC25A1) [21, 22].

Fatty acid β -oxidation

After glucose, fatty acids, usually in the form of triglycerides, are the preferred source to generate energy via cellular oxidation. The triglycerides are first processed by lipoprotein lipase (LPL) outside the adipocytes to yield glycerol and separated fatty acid chains after which the fatty acids are taken up into the cell by fatty acid transporters such as CD36 [23, 24]. The fatty acid chains are converted into Acyl-CoA by Acyl-CoA synthetase family of enzymes, which is then processed further by various enzymes yielding multiple Acetyl-CoA molecules per fatty acid chain (increasing the yield of the TCA cycle per gram of fatty acid) [23, 24]. The rate-limiting step in β -oxidation is controlled by hydroxylacyl-CoA dehydrogenase (HADH) for small- and medium-length fatty acids. However, in the adipose tissue, the rate limiting step is the transport of Acyl-CoA into the mitochondria after the breakdown of long-chain fatty acids, conducted by carnitine palmitoyltransferases (CPT1A and CPT2) [25]. It has been shown that citrate from TCA cycle can escape into the cytoplasm from the mitochondria, where it is converted to Acetyl-CoA by ATP-citrate lyase (ACLY), which can then be converted to Malonyl-CoA by Acetyl-CoA carboxylase (ACACA). Malonyl-CoA is known to restrict the uptake of fatty acids into the mitochondria by inhibiting CPT1A, thereby redirecting fatty acids towards esterification and storage, and creating a feedback mechanism from the TCA cycle [1]. The complete pathway of fatty acid β -oxidation can be viewed in WikiPathways (Pathway ID: WP143 [1, 26]), and the abstracted representation is shown as the orange module in Figure 3.2.

Energy storage in the adipose tissue

Excess energy is stored in the form of fatty acids by the conversion of Acetyl-CoA into fatty acids by fatty acid synthase (FASN) and stearoyl-CoA desaturase (SCD) [27]. This conversion can be in response to both higher presence of glucose and the resultant higher output of the glycolytic pathway, and thus the previously mentioned Malonyl-CoA mediated esterification. FASN utilises both Acetyl-CoA and Malonyl-CoA for the production of fatty acids in the cytoplasm, which can then be desaturated by SCD and stored as triglycerides. The cytoplasmic Acetyl-CoA can also be provided by the aforementioned cytoplasmic citrate via its conversion by ACLY. These fatty acids are then either stored as fat droplets in the adipocytes, or are converted into free fatty acids and excreted to be transported to other tissues and organs via plasma albumin [28]. The detailed pathway of fatty acid biosynthesis is available at WikiPathways (Pathways IDs: WP357 [29]), whereas the yellow module in Figure 3.2 shows the abstracted representation. Recent studies have cited adverse effects of high quantities of dietary fructose as it has been shown that it promotes *de novo* lipogenesis, contributing to higher circulating triglycerides, and thus obesity associated chronic diseases [5, 30]. Whether this contribution has any effects on cellular metabolic flexibility remains to be elucidated.

The PDC-PDK regulatory switch

In glucose oxidation, PDC controls the final step of the conversion of pyruvate to Acetyl-CoA for the TCA cycle, and it has been shown that inhibition of PDC moves the source of energy production from glucose to fatty acids [31]. PDC is composed of three subunits, E1, E2 and E3. Subunit E1 is composed of pyruvate dehydrogenase E1 component subunit alpha (PDHA1) and pyruvate dehydrogenase E1 component subunit beta (PDHB). Subunit E2 consists of dihydrolipoyllysine-residue acetyltransferase (DLAT) while dihydrolipoyl dehydrogenase (DLD) comprises subunit E3. Finally, pyruvate dehydrogenase protein X component (PDHX) anchors the E2 and E3 subunits together, forming functional PDC. One of the major regulators of PDC is the pyruvate dehydrogenase kinase (PDK) family of proteins which have been shown to deactivate the functioning of PDC by phosphorylating it at specific positions [31]. To date, four PDK isoenzymes (1-4) have been identified [31, 32]. Of these, PDK2 and PDK4 have been found to be ubiquitously expressed, especially in tissues and organs with high glucose and fatty acid oxidation rates, for example adipose tissue, liver, heart and other muscle tissues [27]. The adipose tissue has been shown to have a dominant expression of PDK4. The myocardium, on the other hand, expresses PDK1 leading to a stricter regulation of PDC [1]. PDK3 expression is the most restricted, and has been found predominantly in the

brain, testes and kidneys only [1, 31].

In the adipose tissue, the expression of PDK4 has been shown to regulate the conversion of pyruvate into Acetyl-CoA by inhibiting PDC activity [27]. However, PDK4 is allosterically inhibited by pyruvate when in high concentrations, creating a feedback mechanism [27]. In conditions where glucose concentrations drop, less Malonyl-CoA is available from the TCA cycle to mediate esterification of fatty acids, allowing fatty acids to be converted into Acetyl-CoA via β -oxidation [1]. These effects place the PDC-PDK4 protein interaction as a substrate switch, effectively changing the energy source from glucose oxidation to fatty acid β -oxidation. Thus the switch regulates which energy source to metabolise with regards to the nutrient state (glucose or fatty acid availability), as well as whether to focus efforts towards energy production or storage. In addition to pyruvate allosteric inhibition, PDK4 has been shown to have other allosteric interactions with ATP and NADH inducing PDK4 mediated inhibition of PDC. These interactions are highlighted visually by the orange dashed box in Figure 3.2.

PPAR γ signalling and regulatory effects

In addition to the allosteric interactions, PDKs are also regulated transcriptionally via the transcription factors forkhead box protein O (FoxO), Peroxisome proliferator activated receptors (PPARs), and estrogen related receptor α (ERR α) [27]. Of particular interest in the context of the adipose tissue is PPAR γ and PPAR γ coactivator 1 α (PGC1 α) expressions as they affect the transcription of PDK4, improving its expression [33, 34]. PPAR γ is also a cellular fatty acid sensor having a subset of free fatty acids as its ligands [27, 35], and is primarily associated with adipogenesis [33]. PPAR γ is shown as part of the magenta module in Figure 3.2.

AMPK-mediated override

Under unstressed conditions, either glucose or fatty acids are exclusively utilised as substrate for the TCA cycle. Recently, however, it has been shown that under conditions pertaining to energy stress, caused by either nutrient deprivation, or exercise and physical activity, this substrate exclusivity is overridden by the AMP-activated protein kinase (AMPK) signalling cascade, allowing both glucose and fatty acids to be utilised for energy production [1]. This override is possible as AMPK relies on high concentration of accumulated AMP relative to ATP for its activity. AMPK is a heterotrimeric sensor for cellular energy homeostasis, consisting of a catalytic component (α 1 or α 2), and two regulatory components (β 1 or β 2; γ 1, γ 2 or γ 3) [36]. The subcomponents α 1, α 2, β 1, β 2, γ 1, γ 2, and γ 3 are respectively referred to as PRKAA1, PRKAA2,

PRKAB1, PRKAB2, PRKAG1, PRKAG2, and PRKAG3 in Figure 3.2. It has been shown that under nutrient stressed conditions, AMPK mediated inactivation of ACACA and indirect activation of PFK removes the inhibition of glucose oxidation, and allows pyruvate to contribute to Acetyl-CoA formation [1]. AMPK, its subunits, and ACACA are collectively shown as part of the magenta module in Figure 3.2.

Mapping gene expressions in metabolic flexibility

Our abstracted cellular network of metabolic flexibility (Figure 3.2) can be used as a network resource to visualise and analyse expression data for a quick overview of the expression levels of key factors in metabolic flexibility. To demonstrate this utility, we process and visualise a baseline and a publicly available obesity study dataset separately. The cellular network of metabolic flexibility with the visualised data is provided as Additional File 1 for use as a network resource. Additionally, we opted to retain the genes/proteins usually known to have negligible expression in the adipose tissue, in the network to see if their expression is affected in the obesity dataset.

The \log_2 baseline expression in transcripts per million (TPM) [37] for the adipose tissue is visualised in Figure 3.3 using the Genotype Tissue Expression (GTEx) Homo Sapiens baseline datasets from Expression Atlas (Array Express ID: E-MTAB-5214; [14]). The GTEx project analyses global RNA expression from RNA-seq data of non-diseased tissue from humans and provides a reference of baseline measurements of human gene expression and regulation. The expression trends in the baseline expression correspond to the expression patterns reported in the studies cited earlier, for example the predominant expression of PDK4 and PRKAA1 reported in adipose tissue [1, 27].

Figure 3.4 shows the \log_2 fold changes in the cellular network of metabolic flexibility in obese individuals as compared to lean individuals. For this visualisation, we used a relatively recent publicly available dataset (GSE55200 [15]) which contains the subcutaneous adipose tissue transcriptomics (microarray) expression from 7 lean and 16 obese individuals. The original study collected the tissue samples from lean healthy, metabolically healthy obese, and metabolically unhealthy obese individuals to examine the differences in expression between the groups. In the visualisation the metabolic processes appear to be impaired in the obese individuals compared to lean individuals, primarily because most genes are being down regulated. The PDC-PDK switch is also affected in the obese individuals compared to lean, although, PDK3 expression is increased in the obese individuals (while those of other PDKs are decreased). This seems peculiar considering PDK3 has little expression as per the GTEx baseline expression dataset, indicating that some signalling

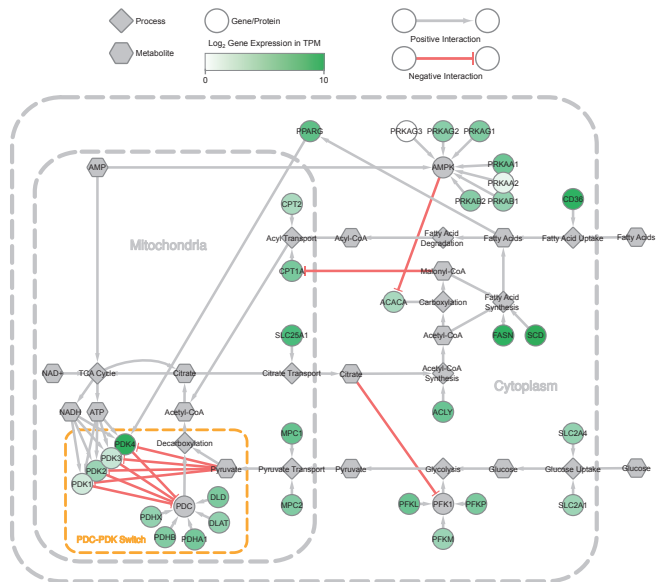


Figure 3.3: Baseline gene expression of the network in the adipose tissue. Expression is median \log_2 TPM expression of GTEx Homo sapiens baseline dataset from Expression Atlas.

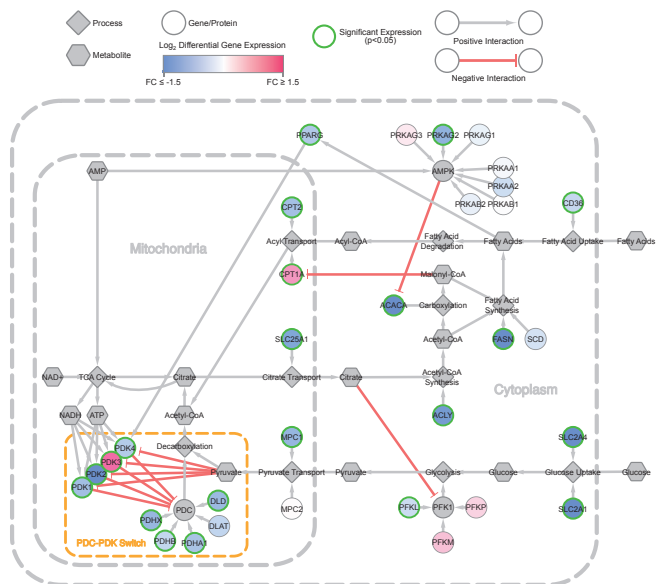


Figure 3.4: Differential expression of the metabolic flexibility network between obese vs lean healthy individuals. Data shown is GSE55200 from the gene expression omnibus. FC means fold change.

or other processes are targeting PDK3 in obese individuals. These signalling processes targeting PDK3 can then be further explored or extend the network for mathematical modelling (using quantitative or qualitative data from the visualised dataset), testing various hypotheses regarding the expression of PDK3 in obesity, which can then be validated using wet-lab methods. In addition, the glucose and fatty acid transporters also appear to be affected, again signalling a possible impairment of the metabolic processes involved with energy production and homeostasis.

Conclusions

Our abstracted cellular network of metabolic flexibility highlights the key components involved in metabolic flexibility, providing a resource for directed pathway and system dynamics analyses in the future. Considering the complex interplay between the various cellular processes associated with metabolic flexibility, it is clear that metabolic flexibility is affected in obesity and associated comorbidities, in particular the PDC-PDK switch governing the substrate utilisation in cellular respiration. Thus the cellular network of metabolic flexibility allows us to target various components (enzymes, biological processes etc.) for further analyses in the context of obesity and the development of chronic diseases.

Availability of data and materials

The datasets analyses in the study are available at array express and gene expression omnibus (GEO). Specifically:

- E-MTAB-5214: [<https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-5214>] [14]
- GSE55200: [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE55200>] [15]

References

- [1] L. Hue and H. Taegtmeyer, "The randle cycle revisited: a new head for an old hat," *American Journal of Physiology - Endocrinology And Metabolism*, vol. 297, no. 3, pp. E578–E591, 2009.
- [2] S. N. Vallerie and K. E. Bornfeldt, "Metabolic flexibility and dysfunction in cardiovascular cells," *Arteriosclerosis, Thrombosis, and Vascular Biology*, vol. 35, no. 9, pp. e37–e42, 2015.
- [3] J. J. Dubé, P. M. Coen, G. DiStefano, A. C. Chacon, N. L. Helbling, M. E. Desimone, M. Stafanovic-Racic, K. C. Hames, A. A. Despines, F. G. S. Toledo, and B. H. Goodpaster, "Effects of acute lipid overload on skeletal muscle insulin resistance, metabolic flexibility, and mitochondrial performance," *American Journal of Physiology - Endocrinology And Metabolism*, vol. 307, no. 12, pp. E1117–E1124, 2014.
- [4] M. K. C. Hesselink, V. Schrauwen-Hinderling, and P. Schrauwen, "Skeletal muscle mitochondria as a target to prevent or treat type 2 diabetes mellitus," *Nat Rev Endocrinol*, vol. 12, no. 11, pp. 633–645, 2016.
- [5] V. Varma, L. G. Boros, G. T. Nolen, C.-W. Chang, M. Wabitsch, R. D. Beger, and J. Kaput, "Metabolic fate of fructose in human adipocytes: a targeted 13c tracer fate association study," *Metabolomics*, vol. 11, no. 3, pp. 529–544, 2015.
- [6] L. M. Sparks, B. Ukropcova, J. Smith, M. Pasarica, D. Hymel, H. Xie, G. A. Bray, J. M. Miles, and S. R. Smith, "Relation of adipose tissue to metabolic flexibility," *Diabetes Research and Clinical Practice*, vol. 83, no. 1, pp. 32–43, 2009.
- [7] P. Arner and A. Kulyte, "MicroRNA regulatory networks in human adipose tissue and obesity," *Nat Rev Endocrinol*, vol. 11, no. 5, pp. 276–288, 2015.
- [8] T. S. Nielsen, N. Jessen, J. O. L. Jørgensen, N. Møller, and S. Lund, "Dissecting adipose tissue lipolysis: molecular regulation and implications for metabolic disease," *Journal of Molecular Endocrinology*, vol. 52, no. 3, pp. R199–R222, 2014.
- [9] T. Hara, D. Kashiwara, A. Ichimura, I. Kimura, G. Tsujimoto, and A. Hirasawa, "Role of free fatty acid receptors in the regulation of energy metabolism," *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, vol. 1841, no. 9, pp. 1292–1300, 2014.
- [10] B. B. Kahn, "Glucose transport: Pivotal step in insulin action," *Diabetes*, vol. 45, pp. 1644–1654, 1996.
- [11] L. Storlien, N. D. Oakes, and D. E. Kelley, "Metabolic flexibility," *Proceedings of the Nutrition Society*, vol. 63, no. 2, pp. 363–368, 2007.
- [12] G. Fantuzzi, "Adipose tissue, adipokines, and inflammation," *Journal of Allergy and Clinical Immunology*, vol. 115, no. 5, pp. 911–919, 2005.
- [13] I. Nassiri, R. Lombardo, M. Lauria, M. J. Morine, P. Moyseos, V. Varma, G. T. Nolen, B. Knox, D. Sloper, J. Kaput, and C. Priami, "Systems view of adipogenesis via novel omics-driven and tissue-specific activity scoring of network functional modules," *Scientific Reports*, vol. 6, p. 28851, 2016.
- [14] GTEx_Consortium, "The genotype-tissue expression (gtex) pilot analysis: Multitissue gene regulation in humans," *Science*, vol. 348, no. 6235, pp. 648–660, 2015.

-
- [15] F. Badoud, K. P. Lam, A. DiBattista, M. Perreault, M. A. Zulyniak, B. Cattrysse, S. Stephenson, P. Britz-McKibbin, and D. M. Mutch, "Serum and adipose tissue amino acid homeostasis in the metabolically healthy obese," *Journal of Proteome Research*, vol. 13, no. 7, pp. 3455–3466, 2014.
- [16] M. Kutmon, A. Riutta, N. Nunes, K. Hanspers, E. L. Willighagen, A. Bohler, J. Mélius, A. Waagmeester, S. R. Sinha, R. Miller, S. L. Coort, E. Cirillo, B. Smeets, C. T. Evelo, and A. R. Pico, "Wikipathways: capturing the full diversity of pathway knowledge," *Nucleic Acids Research*, vol. 44, no. D1, pp. D488–D494, 2016.
- [17] H. A. Krebs, "The history of the tricarboxylic acid cycle," *Perspectives in Biology and Medicine*, vol. 14, no. 1, pp. 154–172, 1970.
- [18] Kdahlquist, A. Pico, M. van Iersel, T. Kelder, and et al., "Tca cycle (homo sapiens)." <https://www.wikipathways.org/index.php?title=Pathway:WP78&oldid=90661>, Accessed on: 15-06-2018.
- [19] D. K. Bricker, E. B. Taylor, J. C. Schell, T. Orsak, A. Boutron, Y.-C. Chen, J. E. Cox, C. M. Cardon, J. G. Van Vranken, N. Dephoure, C. Redin, S. Boudina, S. P. Gygi, M. Brivet, C. S. Thummel, and J. Rutter, "A mitochondrial pyruvate carrier required for pyruvate uptake in yeast, drosophila, and humans," *Science*, vol. 337, pp. 96–100, 2012.
- [20] Kdahlquist, K. Hanspers, M. van Iersel, S. Coort, and et al., "Glycolysis and gluconeogenesis (homo sapiens)." <https://www.wikipathways.org/index.php?title=Pathway:WP534&oldid=94762>, Accessed on: 15-06-2018.
- [21] R. M. Denton and P. J. Randle, "Citrate and the regulation of adipose-tissue phosphofructokinase," *Biochemical Journal*, vol. 100(2), no. 0264–6021 (Print), pp. 420–423, 1966.
- [22] G. V. Gnoni, P. Priore, M. J. H. Geelen, and L. Siculella, "The mitochondrial citrate carrier: Metabolic role and regulation of its activity and expression," *IUBMB Life*, vol. 61, no. 10, pp. 987–994, 2009.
- [23] R. W. Schwenk, G. P. Holloway, J. J. F. P. Luiken, A. Bonen, and J. F. C. Glatz, "Fatty acid transport across the cell membrane: Regulation by fatty acid transporters," *Prostaglandins, Leukotrienes and Essential Fatty Acids (PLEFA)*, vol. 82, no. 4, pp. 149–154, 2010.
- [24] I. J. Goldberg, R. H. Eckel, and N. A. Abumrad, "Regulation of fatty acid uptake into tissues: lipoprotein lipase- and cd36-mediated pathways," *Journal of Lipid Research*, vol. 50, no. Supplement, pp. S86–S90, 2009.
- [25] J. D. Warfel, B. Vandanmagsar, O. S. Dubuisson, S. M. Hodgson, C. M. Elks, E. Ravussin, and R. L. Mynatt, "Examination of carnitine palmitoyltransferase 1 abundance in white adipose tissue: implications in obesity research," *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology*, vol. 312, pp. R816–R820, 2017.
- [26] M. Adriaens, P. Ahles, F. Jagers, C. Evelo, and et al., "Fatty acid beta oxidation (homo sapiens)." <https://www.wikipathways.org/index.php?title=Pathway:WP143&oldid=94768>, Accessed on: 15-06-2018.
- [27] S. Zhang, M. W. Hulver, R. P. McMillan, M. A. Cline, and E. R. Gilbert, "The pivotal role of pyruvate dehydrogenase kinases in metabolic flexibility," *Nutrition & Metabolism*, vol. 11, no. 1, p. 10, 2014.

-
- [28] A. A. Spector, "Plasma lipid transport," *Clinical Physiology and Biochemistry*, vol. 2(2-3), no. 0252-1164 (Print), pp. 123–134, 1984.
- [29] C. Evelo, M. Adriaens, A. Pico, J. Saito, and et al., "Fatty acid biosynthesis (homo sapiens)." <https://www.wikipathways.org/index.php?title=Pathway:WP357&oldid=94197>, Accessed on: 15-06-2018.
- [30] K. L. Stanhope, J.-M. Schwarz, and P. J. Havel, "Adverse metabolic effects of dietary fructose: Results from recent epidemiological, clinical, and mechanistic studies," *Current opinion in lipidology*, vol. 24, no. 3, pp. 198–206, 2013.
- [31] M. J. Holness and M. C. Sugden, "Regulation of pyruvate dehydrogenase complex activity by reversible phosphorylation," *Biochemical Society Transactions*, vol. 31, no. 6, pp. 1143–1151, 2003.
- [32] M. M. Bowker-Kinley, P. Davis Wu, R. A. Wu, P. Fau Harris, K. M. Harris, R. A. Fau Popov, and K. M. Popov, "Evidence for existence of tissue-specific regulation of the mammalian pyruvate dehydrogenase complex," *Biochemical Journal*, vol. 329(1), no. 0264-6021 (Print), pp. 191–196, 1998.
- [33] J.-E. Lee and K. Ge, "Transcriptional and epigenetic regulation of ppar γ expression during adipogenesis," *Cell & Bioscience*, vol. 4, no. 1, p. 29, 2014.
- [34] T. Cadoudal, E. Distel, S. Durant, F. Fouque, J.-M. Blouin, M. Collinet, S. Bortoli, C. Forest, and C. Benelli, "Pyruvate dehydrogenase kinase 4 regulation by thiazolidinediones and implication in glyceroneogenesis in adipose tissue," *Diabetes*, vol. 57, no. 9, pp. 2272–2279, 2008.
- [35] T. Varga, Z. Czimmerer, and L. Nagy, "Ppars are a unique set of fatty acid regulated transcription factors controlling both lipid metabolism and inflammation," *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, vol. 1812, no. 8, pp. 1007–1022, 2011.
- [36] M. Daval, F. Fougelle, and P. Ferré, "Functions of amp-activated protein kinase in adipose tissue," *The Journal of Physiology*, vol. 574, no. 1, pp. 55–62, 2006.
- [37] G. P. Wagner, K. Kin, and V. J. Lynch, "Measurement of mrna abundance using rna-seq data: Rpkms measure is inconsistent among samples," *Theory in Biosciences*, vol. 131, no. 4, pp. 281–285, 2012.

Supplementary Materials

The cellular network of metabolic flexibility is provided as a resource in the form of a Cytoscape session file (Additional File 1). Hosted by *Genes & Nutrition* at: <https://doi.org/10.1186/s12263-018-0609-3> (7Zip, 42KB)

4

Logical modelling reveals the PDC-PDK interaction as the regulatory switch driving metabolic flexibility at the cellular level

Samar H.K. Tareen¹, Martina Kutmon^{1,2}, Ilja C.W. Arts^{1,3}, Theo M. de Kok^{1,4},
Chris T. Evelo^{1,2} and Michiel E. Adriaens¹

¹ Maastricht Centre for Systems Biology (MaCSBio), Maastricht University, the Netherlands

² Department of Bioinformatics – BiGCaT, NUTRIM Research School, Maastricht University, the Netherlands

³ Department of Epidemiology, CARIM School for Cardiovascular Diseases, Maastricht University, the Netherlands

⁴ Department of Toxicogenomics, GROW School of Oncology and Developmental Biology, Maastricht University, the Netherlands

Published in: *Genes & Nutrition* (2019) 14:27.

doi: 10.1186/s12263-019-0647-5

Abstract

Background: Metabolic flexibility is the ability of an organism to switch between substrates for energy metabolism, in response to the changing nutritional state and needs of the organism. On the cellular level, metabolic flexibility revolves around the tricarboxylic acid cycle by switching acetyl coenzyme A production from glucose to fatty acids and vice versa. In this study, we modelled cellular metabolic flexibility by constructing a logical model connecting glycolysis, fatty acid oxidation, fatty acid synthesis and the tricarboxylic acid cycle, and then using network analysis to study the behaviours of the model.

Results: We observed that the substrate switching usually occurs through the inhibition of pyruvate dehydrogenase complex (PDC) by pyruvate dehydrogenase kinases (PDK), which moves the metabolism from glycolysis to fatty acid oxidation. Furthermore, we were able to verify four different regulatory models of PDK to contain known biological observations, leading to the biological plausibility of all four models across different cells and conditions.

Conclusion: These results suggest that the cellular metabolic flexibility depends upon the PDC-PDK regulatory interaction as a key regulatory switch for changing metabolic substrates.

Introduction

Metabolic flexibility is the ability of an organism to switch between substrates for energy metabolism, adapting to the changing nutritional state and needs of the organism [1]. In complex organisms, such as humans, the various cells and tissues utilising either glucose or fatty acids and its derivatives to fuel metabolism maintain metabolic flexibility. This flexibility revolves around the tricarboxylic acid (TCA) cycle in oxidative metabolism, where several biochemical processes interact with each other to use either glycolysis or fatty acid oxidation to fuel metabolism. Due to various tissues and organs having different energy requirements, metabolic flexibility in complex organisms also includes the delicate balance between these tissues and organs in utilising the correct substrate at correct times, so as not to starve off limited supply of nutrients critical to the functioning of other organs and tissues [1]. This cellular and tissue/organ level metabolic flexibility works collectively to manage the nutrient state and needs of the organism, and by design enforces the utilisation of a single substrate on the cellular level. Recent studies have found impaired metabolic flexibility to be associated with obesity and related co-morbidities, chiefly type 2 diabetes mellitus and cardiovascular diseases [2–4].

In a recent review [5], we explored the various cellular processes involved in maintaining cellular metabolic flexibility in the adipose tissue. The two major energy production mechanisms, glycolysis and fatty acid oxidation, are tied to the TCA cycle – glycolysis through the production of pyruvate and its conversion to acetyl coenzyme A (acetyl-CoA), and fatty acids through their breakdown to acyl coenzyme A and transportation into the mitochondria through the carnitine transport mechanism for eventual conversion to acetyl-CoA. This acetyl-CoA is converted to citrate, which starts the TCA cycle converting adenosine monophosphate (AMP) to adenosine triphosphate (ATP) and oxidised nicotinamide adenine dinucleotide (NAD⁺) to its reduced form NADH, improving the energy state of the cell. Excess energy, in the form of citrate escapes to the cytoplasm from the mitochondria, where it shuts down glycolysis and/or fatty acid oxidation and contributes to the re-synthesis of cellular fatty acids. Some additional cellular processes also assist in the regulation of cellular metabolic flexibility, namely the adenosine monophosphate-activated protein kinase (AMPK) signalling cascade and the peroxisome proliferator-activated receptor gamma (PPAR γ) nuclear receptor mediated transcriptional regulation [5].

In another study of ours [6], we used published data to generate clusters of correlated genes preserved in the majority of individuals that participated in a weight loss study [7]. We observed that one of the generated clusters was primarily involved with the upstream regulation of the TCA cycle. This observation suggests additional links between the regulation of cellular

metabolic flexibility and obesity related co-morbidities, considering that weight loss is the predominant method of countering obesity and its ill effects. In our review [5], we also highlighted an inhibitory regulatory interaction directed from pyruvate dehydrogenase kinase (PDK) to pyruvate dehydrogenase complex (PDC) as a key nutrient switching mechanism between glucose and fatty acids, especially since we observed it to be affected in obesity.

Based on these observations, we hypothesise that this regulatory interaction, termed by us as the PDC-PDK regulatory switch, is a key regulator of cellular metabolic flexibility. As such, we focused on the changing of the metabolic substrate in response to the PDC-PDK regulatory interaction in this study. We have used logical modelling (i.e., a predicate logic based modelling framework) to construct a regulatory model to test this hypothesis, and show how the various perturbations designed to derail cellular metabolic flexibility are propagated through the malfunctioning of the PDC-PDK regulatory switch. We opted to use logical modelling, as opposed to more complex and complicated quantitative modelling primarily because of the high resolution of data required to accurately model cellular processes quantitatively [8]. Molecular interactions in the cell, such as protein associations, occur rapidly at rates from less than $10^3 \text{ M}^{-1}\text{s}^{-1}$ to greater than $10^9 \text{ M}^{-1}\text{s}^{-1}$ [9] and require specialised experiments and lengthy simulations to determine accurately [10]. Many of the cellular interactions that we modelled in this study currently have little-to-no accurate *in vivo* measurements available on smaller time scales at which they occur to properly construct and train a quantitative regulatory model. These interactions include, but are not limited to, protein-protein interactions, site-specific phosphorylation, allosteric interactions, chemical associations and dissociations. As logical modelling does not rely on material and spatiotemporal quantification of entities, it has been successfully applied previously in scenarios with sparse data availability [8, 11–13].

Materials and methods

We applied a well-established pipeline of logical modelling and analysis of biological pathways and networks using discrete/qualitative models [11, 14–16]. The pipeline started with the construction of the biological regulatory network, after which its parameters were defined via logic circuits (which can also be represented in a tabulated manner). Collectively, the network and the parameters constitute a single model, with multiple parameter sets representing distinct models of the same regulatory network. Using the parameters, a new network called a state transition graph (STG) was constructed, representing all possible behaviours of the regulatory network in the discrete state space. This network was then further analysed for relevant biological behaviours, both for system verification and for predictions. Figure 4.1 represents the overview of

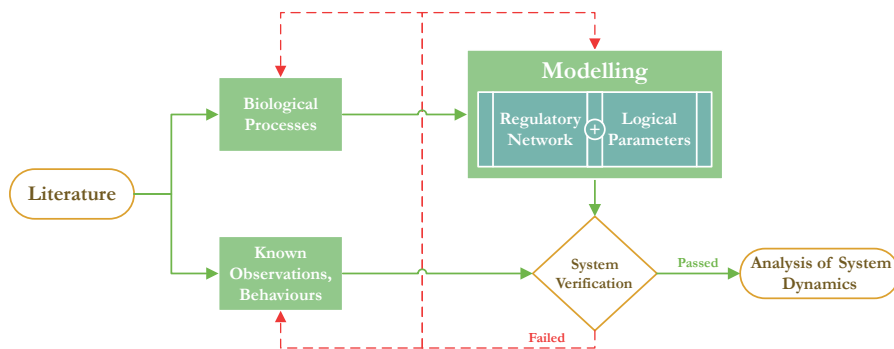


Figure 4.1: Workflow of the methodology. Biological processes and their known observations were extracted from literature. These processes were then used to construct the regulatory network with the logical parameters selected based on the biochemical reactions and known interactions. The regulatory network then underwent system verification where it was tested to check if it could exhibit known biological observations and behaviours or not. If the verification failed, then troubleshooting was performed by checking the model for errors, changing the system definitions extracted from the literature, and/or checking if the known observations were in conflict with the system. If the system verification passed, then the dynamics generated by the model were analysed for biologically meaningful behaviours.

this methodology, and we have used a toy example to illustrate the procedure step-by-step in the subsequent sections.

Logical Modelling

In logical modelling, the model represents a system by using discrete values of 0 and 1 for OFF and ON states of the entities comprising the system. The dynamics of the system are then defined by step functions that change these values. In our study, we have employed the René Thomas Kinetic Logic formalism [17] and refer to the work of Paracha *et al* [14] for the mathematical definitions and constraints of the formalism. Of note is the distinction that we defined 1 as availability and 0 as unavailability of an entity in this study. This availability and unavailability, however, does not imply any concentration of the said entity, only whether the entity is able to perform its functions or not. We used this interpretation primarily to model allosteric inhibitory interactions which are otherwise difficult to model in a concentration-based interpretation (since the concentration-based interpretation would imply that the allosteric inhibition and allosteric activation always decreases and increases the production of the target entity respectively, which may not always be the case in reality). In the following sections, we show the modelling and analysis of a toy example to ease the reader into the application of the formalism.

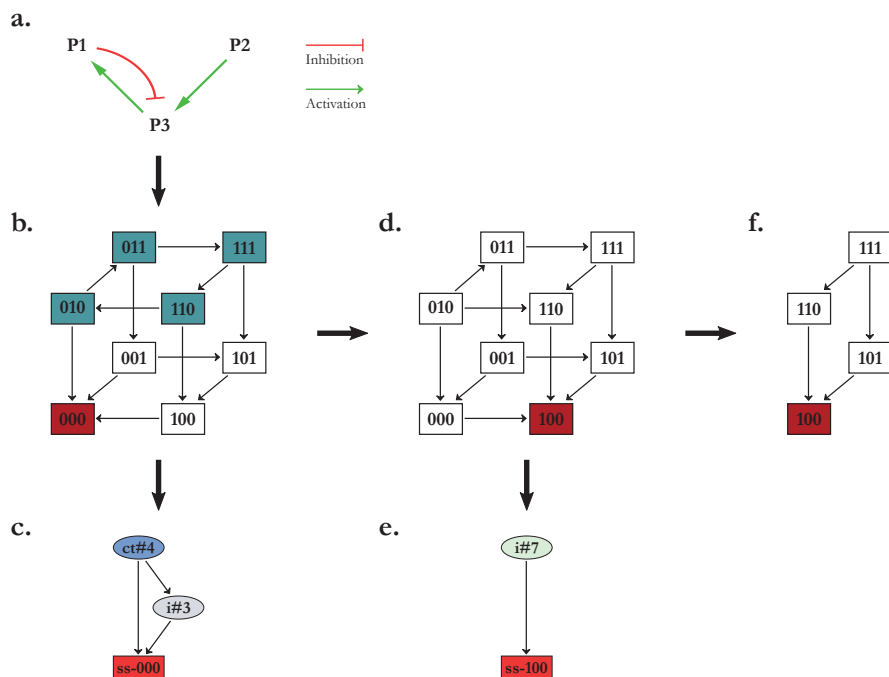


Figure 4.2: Step-by-step analysis of the toy example. a) The regulatory network of the toy example. P1 and P2 activate P3, whereas P1 inhibits P3. b) The state transition graph (STG) of the toy example. c) The hierarchical transition graph (HTG) of the toy example. d) The STG of the toy example with P1 showing ectopic activity. e) The HTG of P1 ectopic activity. f) The STG of P1 ectopic activity when the system is initialised with all entities as active (i.e. at level 1). The STGs and HTGs were generated using GINsim [18].

Regulatory Network

In the toy example, we have a regulatory network consisting of three entities, P1, P2 and P3. P1 and P3 have a reciprocal relationship where P3 is an activator of P1, and P1 the inhibitor of P3. An activator implies that the source entity has a positive effect on the concentration and/or activity of the target entity, whereas an inhibitor has a negative effect. We also see P2 as an activator of P3, and by extension, having an indirect effect on the activity of P1 mediated via P3. The regulatory network for the toy example is shown in Figure 4.2a. One constraint in the René Thomas formalism is that the maximum discrete level attainable by an entity is constrained by the total number of its target entities [17]. In the toy example, all three entities have a single target entity, and are thus constrained to a maximum discrete level of 1 each.

Table 4.1: Logical parameters of the toy example. The format used here represents the presence of respective entities in the system when they are listed in the parameter set.

Entity	Parameter Set	Target Value
P1	{ }	0
	{P3}	1
P2	{ }	0
	{ }	0
P3	{P1}	0
	{P2}	1
	{P1, P2}	0

Logical Parameters

As multiple entities can affect the same target entity simultaneously, a set of logical parameters was selected which define the rules and precedence governing the evolution of the target entities in the network. These logical parameters are based on the behaviours observed, measured or inferred using data and experiments. In the toy example, we define the logical parameters based on the interactions that we have in the regulatory network,

1. P2 is the activator of P3.
2. P3 is the activator of P1.
3. P1 is the inhibitor of P3.

We see that the logical parameters for P1 are trivial as it has only one activator and no inhibitor. Generally, in the absence of their activators, the entities are assumed to degrade over time. Thus, P1 would be reduced to level 0 in the absence of P3, whereas P2 would reduce to 0 as its activator has not been modelled in the regulatory network. It is also possible to assume an implied activator that would activate P2, but generally, such assumptions are made only if an inhibitor of the entity in question has been explicitly modelled in the regulatory network to balance the activity of the said entity. Since we do not have any such an inhibitor for P2 in the regulatory network, we assume otherwise. Lastly, the logical parameters for P3 are non-trivial since we have both an activator (P2) and an inhibitor (P1) that can act simultaneously on P3. Here we model a precedence for the inhibitor, and assume that P3 would always be inhibited by P1 whenever P1 is present in the system, otherwise P3 would rely on P2 to become activated. These parameters are tabulated as Table 4.1.

State Transition Graphs (STGs)

The logical parameters govern the behaviour of the regulatory network, from which it is possible to generate a graph of all possible behaviours. This graph is called a State Transition Graph (STG), consisting of states (nodes) and transitions (edges). Each state represents a particular configuration of the complete regulatory network, where a configuration is defined as a unique combination of the discrete levels of the entities of the regulatory network. Thus, any two states would have a different discrete level for at least one entity. The maximum number of states is defined by the formula lvl^n , where lvl represents the maximum discrete level plus 1 (essentially the total number of levels available for a given entity), and n represents the number of entities having that level. In the toy example, this formula would be 2^3 , totalling eight states, as we can see in Figure 4.2b.

The transitions represent the changes in the discrete levels of the entities, and thus the transitioning of the regulatory network from one configuration to another. The transitions are constrained by the logical parameters, and can only exist between two states if the source state satisfies the discrete levels of the target state through the logical parameters. This imparts directionality to the behaviours represented in the STG, generating cyclic and acyclic paths in the graph. A behaviour can then be defined as a path in the STG, essentially a series of states connected by transitions between them. Finally, we utilise asynchronous transitions, which only allow one entity to change its discrete level between two successive states. The logical parameters given in Table 4.1 were used to derive the STG of the toy example shown in Figure 4.2b. Each state is labelled with three numbers, representing the discrete levels of P1, P2, and P3 in that order (for example, state 101 refers to P1 and P3 having level 1, and P2 having level 0).

System Verification

Using the STGs generated by a regulatory network, it is possible to reverse engineer sets of logical parameters that allow certain known behaviours of the system to exist in the regulatory network. By extension, this allowed our modelled system to be verified against known biological observations or to find logical parameters that satisfy those conditions. We utilised a model checking technique called computational tree logic (CTL) to identify known behaviours [19]. CTL allowed us to use predicate logic along with quantifiers to formulate behaviours, and test which sets of logical parameters allow such behaviours to exist within their STGs. Specifically, for a given predicate formula ϕ , these quantifiers are;

- $AG\phi$: From a given state, all states (G) along all paths (A) must satisfy ϕ

- $AF\phi$: From a given state, at least one future state (F) along all paths (A) must satisfy ϕ
- $EG\phi$: From a given state, all states (G) along at least one path (E) must satisfy ϕ
- $EF\phi$: From a given state, at least one future state (F) along at least one path (E) must satisfy ϕ
- $AX\phi$: From a given state, all (A) immediate successor states (X) must satisfy ϕ
- $EX\phi$: From a given state, at least one (E) successor state (X) must satisfy ϕ

For the toy example, we formulate the CTL formula as,

$$((P2 = 1) \rightarrow EX(EG(P2 = 1))) \wedge ((P2 = 1) \rightarrow EX(EG(P2 = 0)))$$

The formula states that when P2 is at discrete level 1, there exists at least one path from at least one successor state, which maintains P2 at level 1, AND there exists at least one path from at least one successor state which reduces the level of P2 to 0 indefinitely. When applied, this formula is tested against all STGs produced by all possible logical parameter sets for the toy example regulatory network. Only 2 logical parameter sets out of the total 36 are able to satisfy this property, one of which is already given as Table 4.1. Additional file 1 provides the source file for the toy example system verification, and includes the definition of the network, the logical parameters, and the CTL formula.

Network Analysis

Once the system was verified, we proceeded with the analysis of the behaviours provided in the STG. The results of such an analysis for the toy example is visualized in Figure 4.2. The blue shaded states represent a cyclic behaviour where the system can keep transitioning from one state after the other, successively and indefinitely. We can also see that the entity P2 maintains its level 1 as long as the system remains within this cyclic behaviour. As soon as P2 changes its level to 0, the system transitions away from the cyclic behaviour into several separate acyclic behaviours, all of which led the system to a deadlock state 000. A deadlock state, also referred to as a stable steady state or fixed point, is defined as a state in an STG, which has no exit transitions, and implies that the system gets stuck in this state. The acyclic states are coloured white, whereas the deadlock state is coloured red in Figure 4.2b. Cyclic behaviours/trajectories represent periodic or recurring biological processes, such as circadian rhythms, while acyclic behaviour/trajectories represent one-way propagations, such as signalling cascades. In Figure 4.2b, we can see that the maintenance of P2 at level 1 is required to keep the system in a periodic behaviour.

Hierarchical Transition Graphs (HTGs)

One of the drawbacks of logical modelling is state-space explosion – the size of the STG increases exponentially with linear increase in the size of the regulatory network. For example, increasing our regulatory network to four entities would yield an STG of sixteen states, while six entities would create an STG of sixty-four states. Subsequently, the STG of larger regulatory networks becomes very complex, and extremely tedious and error-prone for manual analysis. However, it is possible to analyse large STGs by finding sub-networks and patterns (such as strongly connected components (SCCs), essentially linked cyclic paths) contained within the STG itself. Towards this end, GINsim allows us to collapse these sub-networks and patterns in the network to generate Hierarchical Transition Graphs (HTGs) [18, 20, 21]. The collapsed substructures are then represented as,

- **Transient SCC:** a node containing a strongly connected component, which also has outgoing transitions to other components or parts in the HTG. These nodes are labelled as 'ct#' followed by the number of states contained within, e.g., 'ct#4' in Figure 4.2c.
- **Terminal SCC:** a node containing a strongly connected component, which does not have any outgoing transitions to other components. These nodes are labelled as 'ca#' followed by the number of states.
- **Irreversible Component:** a node containing states and transitions that do not have any cycles in them. Such components represent unidirectional flow in the behaviours being represented by the HTG. These nodes are labelled as 'i#' followed by the number of states.
- **Rooted Irreversible Component:** an irreversible component that includes at least one state with no incoming transitions. These nodes are labelled the same as irreversible components, i.e., 'i#' followed by the number of states.
- **Stable State:** a node containing a single deadlock state which the system is unable to exit upon entering i.e., it has no outgoing transitions. These nodes are labelled as 'ss-' followed by the label of the state/configuration itself. For example, the nodes 'ss-000' and 'ss-100' in Figure 4.2c and 4.2d respectively.

We then proceeded to collapse the acyclic paths into single nodes, and finally merged all edges between these collapsed nodes based on the edges present between the respective states in the STG. Figure 4.2c shows the HTG for the toy example, where the cyclic path has been collapsed to the light blue node 'ct#4', and the acyclic path has been collapsed to the grey node 'i#3', with the deadlock state represented as its own node 'ss-000'. Thus, the HTG provides a structural representation of the system by illustrating the connections between various sets of behaviours found in the underlying STG. In addition, HTGs are by definition

acyclic due to the collapsing procedure.

Perturbation Analysis

In addition to the HTG, we also performed the perturbation analysis, where we restricted the parameters of one or more entities to represent knockouts or ectopic activities. The STG constructed with these restrictions establishes the propagation of their effects through the rest of the system. Figure 4.2d shows the STG when entity P1 is restricted to ectopic activity. We can immediately see that the cyclic behaviour found in the STG in Figure 4.2b is no longer available, and that the deadlock state has moved from 000 to 100. In addition, its HTG (Figure 4.2e) shows only two nodes, a rooted 'i#7' irreversible node, and the 'ss-100' deadlock node. The lack of incoming transitions to the rooted state shows unique conditions from which a modelled system is able to recover, but requires external, un-modelled influences/regulation to achieve.

It is possible to couple a perturbation with a defined initial state to fine-tune the behaviours of the system in response to known restrictions, or predict outcomes of new restrictions. In Figure 4.2f, we see an STG generated when the initial state was defined as 111 with the P1 ectopic activity perturbation. The STG shows the routes available to the system using the logical parameters from Table 4.1 under the ectopic expression of P1, both of which lead to the deadlock state 100.

Software

Logical modelling, network and perturbation analyses were performed using GINsim v3.0 [20]. The system verification was done using SMBioNet v3.1 [22]. Cytoscape [23] was used to visualise the networks.

Results

The Regulatory Network of Cellular Metabolic Flexibility

In a review done previously [5], we had explored the pathways involved with cellular metabolic flexibility and had constructed a network of these pathways representing cellular metabolic flexibility. This network links both glucose and fatty acid oxidation with the TCA cycle as energy production methods, along with fatty acid (re)synthesis as an energy production and/or storage method. As mentioned previously in the Introduction section, metabolic flexibility at the cellular level enforces that only glucose or fatty acids are utilised for energy production at any given time. Exceptions to this enforcement have only been observed in situations involving high cellular stress and depleted nutrient conditions, such as in ischemic hearts [1].

Figure 4.3 illustrates the complete biological regulatory network model. We have abstracted the larger cellular network of metabolic flexibility from the review to reduce the number of entities and thus reduce the complexity introduced by state-space explosion. Our regulatory network consists of ten entities, namely Glucose, Pyruvate, Pyruvate Dehydrogenase Kinase (PDK), Pyruvate Dehydrogenase Complex (PDC), Acetyl Coenzyme A (Acetyl-CoA), Citrate, Malonyl Coenzyme A (Malonyl-CoA), circulating fatty acids, cellular fatty acids (Fatty Acids), and adenosine monophosphate-activated protein kinase (AMPK). The interactions and processes represented by the edges of the biological network are explained in Table 4.2, and we direct the readers to the original review article [5] for the detailed explanation of the biological pathways and interactions involved in metabolic flexibility. Additional file 2 also shows a generic procedure for abstraction/reduction of regulatory networks. For a more detailed explanation of the reduction algorithm including the formal definitions and proofs, we refer the reader to Saadatpour et al. [24].

System Verification of the Logical Parameters Governing Cellular Metabolic Flexibility

The selection of the logical parameters for the regulatory network was done manually as most of the interactions present in the system are well-studied biological processes. The exception was the regulation of PDK as it is relatively less known and does not rely on stringent biochemical reactions, allowing for multiple regulatory possibilities. What is known is the inhibition of PDK isoenzymes via pyruvate, and the activation of PDK isoenzymes through either TCA cycle products or peroxisome proliferator-activated receptor gamma (PPAR γ) triggered by fatty acids in the cytoplasm [5]. Retaining these interactions, we get six combinations of logical parameters. We selected four sets of logical parameters for PDK, generating four models of our regulatory network (the remaining two sets of logical parameters lead to either no effect of inhibition, or no effects of activation). These models are,

- Model 1: The inhibitor, Pyruvate, always blocks PDK activity when it is present in the system.
- Model 2: The inhibitor, Pyruvate, only blocks PDK activity when at least one of its activators, Acetyl-CoA (mediating activation via increased BADH and ATP) or Fatty Acids (mediating activation via PPAR γ), is absent from the system. Thus, the activators collectively override Pyruvate mediated inhibition.
- Model 3: Acetyl-CoA mediated activation of PDK bypasses Pyruvate mediated inhibition.
- Model 4: Fatty Acid mediated activation of PDK bypasses Pyruvate mediated

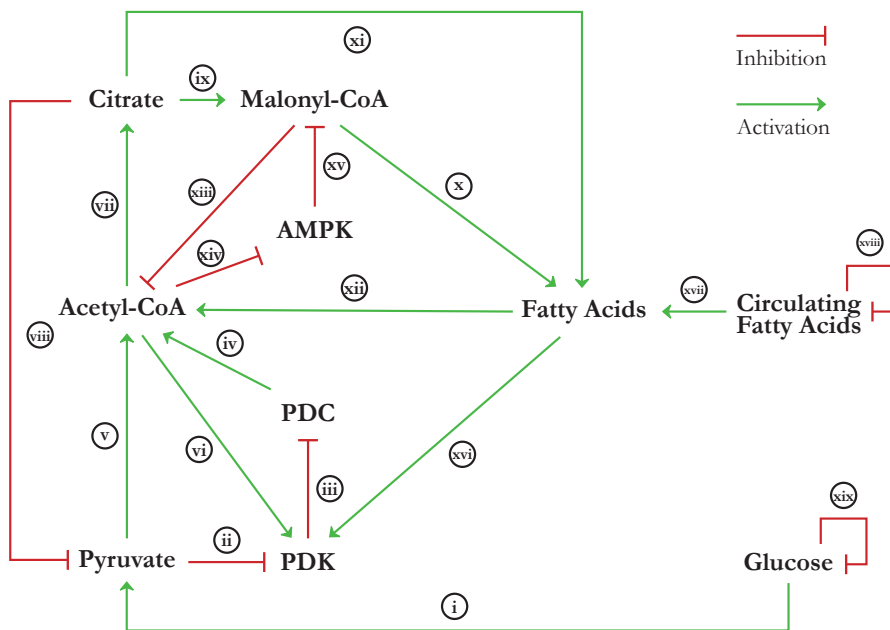


Figure 4.3: Biological regulatory network of cellular metabolic flexibility. The regulatory network consists of ten entities representing the biological processes involved in cellular metabolic flexibility. The entities interact with one another through various processes, abstractly represented here as activation or inhibition interactions. The interactions are labelled with Roman numerals, and are explained in Table 4.2.

inhibition.

Figure 4.4 shows the circuit diagram representations of the logical parameters for each entity, including the circuit diagrams for the four PDK parameter sets. The tabulated logical parameters are provided as Additional file 3. The four models were then tested in SMBioNet [22] for system verification and parameter selection. The following biological properties were tested,

- Glucose Oxidation: Ensuring that glucose is oxidised to pyruvate, which contributes to acetyl-CoA production
- Fatty acid Oxidation: Ensuring that fatty acid oxidation to acetyl-CoA takes place in the absence of malonyl-CoA
- Known PDK interactions: Ensuring that PDK inhibits glucose oxidation, allowing fatty acid oxidation to take place
- Absence of PDK: Ensuring glucose oxidation resumes in the absence of PDK, creating malonyl-CoA

Table 4.2: Edge list and explanation of the interactions in biological regulatory network in Figure 4.3.

Edge Label	Interaction Explanation
i	Represents the process of glucose uptake and its multi-step conversion via various enzymes to Pyruvate [25, 26]
ii	Represents the allosteric inhibition of the PDK enzymes by Pyruvate [27].
iii	Represents the inhibition of PDC by PDKs via site-specific phosphorylation [1, 27].
iv	Represents the involvement of PDC in converting Pyruvate into Acetyl-CoA via decarboxylation [25, 26].
v	Represents the consumption of Pyruvate to create Acetyl-CoA via PDC mediated decarboxylation [28].
vi	Represents the allosteric activation of PDKs via NADH and ATP produced during the TCA cycle fuelled by Acetyl-CoA [1].
vii	Represents the conversion of Acetyl-CoA to Citrate in the mitochondria, part of which is transported into the cytoplasm [29, 30].
viii	Represents the inhibition of phosphofructokinases (PFKs) by cellular Citrate, thereby inhibiting the production of Pyruvate from Glucose [31, 32].
ix	Represents the conversion of Citrate to Malonyl-CoA through the Acetyl-CoA carboxylase 1 (ACACA) mediated carboxylation [1].
x	Represents the utilisation of Malonyl-CoA for fatty acid synthesis [29, 30].
xi	Represents the reconversion of Citrate to Acetyl-CoA in the cytoplasm to be used for fatty acid synthesis alongside Malonyl-CoA [29, 30].
xii	Represents the breakdown of fatty acids to Acyl-CoA, transport into the mitochondria via the carnitine transport process and conversion to Acetyl-CoA for the TCA cycle [33, 34].
xiii	Represents the inhibition of the carnitine transport process by Malonyl-CoA, thereby affecting Acetyl-CoA production [1].
xiv	Represents the negative effect of Acetyl-CoA on AMPK activity via higher ATP and lower AMP concentrations [1, 35].
xv	Represents the inhibition of Malonyl-CoA production by the AMPK mediated inhibition of ACACA [1, 35].
xvi	Represents the increased activity of PDKs by cellular fatty acids via Peroxisome Proliferator-Activated Receptor gamma (PPAR γ) signalling [27, 36–38].
xvii	Represents the uptake of circulating fatty acids into the cell [39, 40].
xviii	Highly abstracted representation of circulating fatty acid regulation outside the cell.
xix	Highly abstracted representation of circulating glucose regulation outside the cell.

Circuit Diagrams of the Logical Parameters

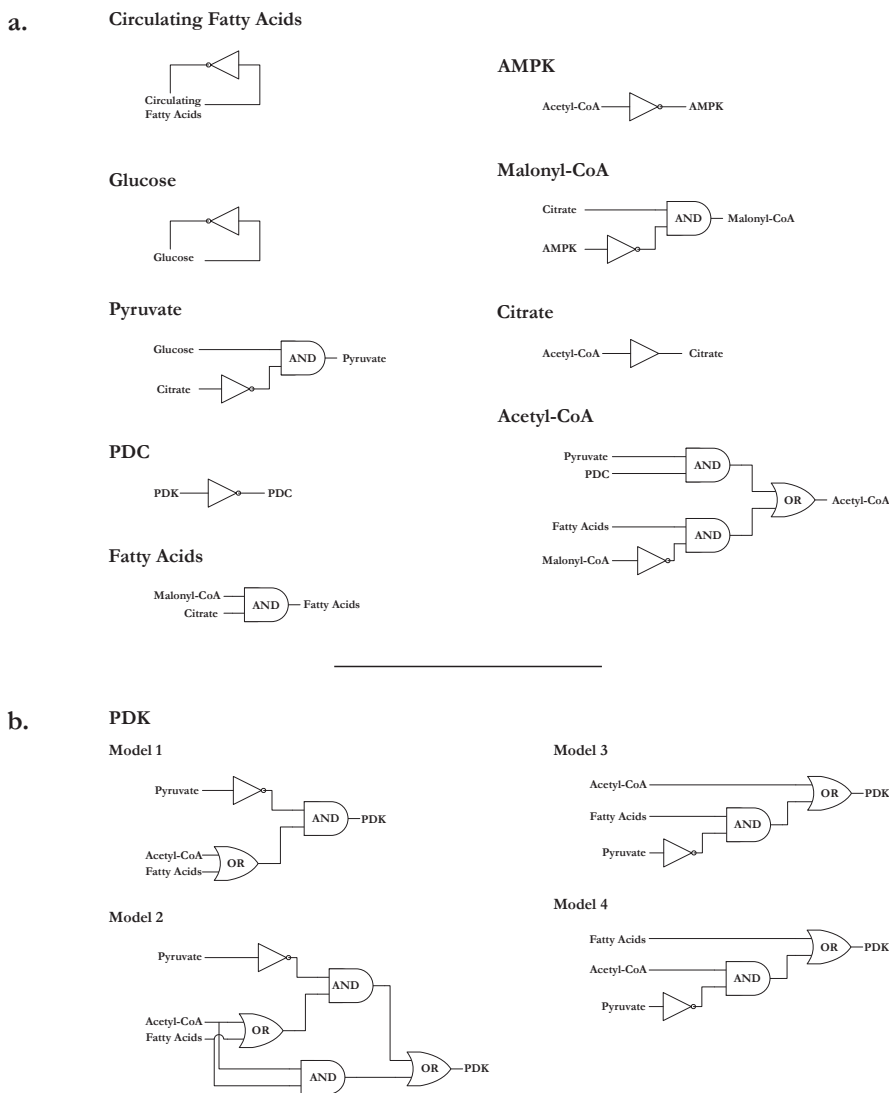


Figure 4.4: Circuit diagrams of the logical parameters for the regulatory network of cellular metabolic flexibility. a) The circuit diagrams representing the entities other than PDK. Each entity has a single circuit diagram representing the respective set of parameters. b) Shows the four models of PDK regulation, differing on how the activators (Fatty Acids and Acetyl-CoA) are able to affect the activation of PDK in the presence of the inhibitor (Pyruvate). The tabulated logical parameters for all entities are provided as Additional file 3.

Table 4.3: CTL formulae used for system verification of the regulatory network of cellular metabolic flexibility.

Property	CTL Formula
Glucose Oxidation	$((Glucose = 1 \wedge Pyruvate = 0 \wedge Citrate = 0) \rightarrow EX(Glucose = 1 \wedge Pyruvate = 1))$ \wedge $((Pyruvate = 1 \wedge PDC = 1 \wedge Acetyl-CoA = 0) \rightarrow EX(Pyruvate = 1 \wedge PDC = 1 \wedge Acetyl-CoA = 1))$
Fatty acid Oxidation	$((Fatty\ Acids = 1 \wedge Malonyl-CoA = 0 \wedge Acetyl-CoA = 0) \rightarrow EX(Fatty\ Acids = 1 \wedge Malonyl-CoA = 0 \wedge Acetyl-CoA = 1))$
Presence of PDK	$((PDK = 1 \wedge PDC = 1 \wedge Pyruvate = 1 \wedge (Fatty\ Acids = 0 \mid Malonyl-CoA = 1) \wedge Acetyl-CoA = 1) \rightarrow EF(PDK = 1 \wedge PDC = 0 \wedge (Fatty\ Acids = 0 \mid Malonyl-CoA = 1) \wedge Acetyl-CoA = 0))$ \wedge $((PDK = 1 \wedge (PDC = 0 \mid Pyruvate = 0) \wedge Fatty\ Acids = 1 \wedge Malonyl-CoA = 0 \wedge Acetyl-CoA = 0) \rightarrow EX(PDK = 1 \wedge (PDC = 0 \mid Pyruvate = 0) \wedge Fatty\ Acids = 1 \wedge Malonyl-CoA = 0 \wedge Acetyl-CoA = 1))$
Absence of PDK	$((PDK = 0 \wedge PDC = 1 \wedge Pyruvate = 1 \wedge (Fatty\ Acids = 0 \mid Malonyl-CoA = 1) \wedge Acetyl-CoA = 0) \rightarrow EX(PDK = 0 \wedge PDC = 1 \wedge Pyruvate = 1 \wedge (Fatty\ Acids = 0 \mid Malonyl-CoA = 1) \wedge Acetyl-CoA = 1))$ \wedge $((PDK = 0 \wedge PDC = 1 \wedge Pyruvate = 1 \wedge Malonyl-CoA = 0 \wedge Acetyl-CoA = 1) \rightarrow EF(PDK = 0 \wedge PDC = 1 \wedge Pyruvate = 1 \wedge Malonyl-CoA = 1 \wedge Acetyl-CoA = 1))$

These biological properties are codified as CTL formulae in Table 4.3, and their formulation is explained in Additional file 4. All four models of PDK regulation were able to satisfy these properties, implying that the four logical parameter sets modelled for PDK are biologically plausible. The SMBioNet source file for the system verification of the regulatory network is provided in Additional file 5.

Network Analysis of the Behaviours Exhibited by the Models

Since all four parameter sets of PDK regulation passed system verification on known biological observations, we analysed all four models. The STG (state transition graph; see Methods section) of each model consisted of 1,024 states,

and 6,144 transitions between the edges. The only difference between the STGs of the four models were the transitions between the states governing the regulation of PDK, which was expected as per the logical parameters. The STG of model 1 is provided in Additional file 6, and shows the size, density, and complexity of the network and behaviours contained within.

We then proceeded to collapse the four STGs into their respective HTGs (hierarchical transition graphs; see Methods section), to compare the behavioural patterns and substructures in the models. We observed that due to the regulation of Glucose and Circulating Fatty Acids as inhibitory self-loops, the majority of the states from the STGs, 992 to be exact, are collapsed into a single node in the HTG with the remaining nodes showing fluctuations of Glucose and Circulating Fatty Acids without any effect on the cellular environment. Essentially, the various groups of system dynamics were being integrated together by the fluctuation of the glucose and circulating fatty acid input nodes. The self-inhibitory loop on each of the input nodes would switch them between 0 and 1, thereby linking all different strongly connected components together to form a single strongly connected component. Although this behaviour does show how interconnected the biological behaviours are, it makes the analysis of these behaviours that much complex. To remedy this situation, we opted to remove the self-inhibitory loops and restricted the input nodes to the four combinations of 0 and 1 to see how the system behaves for these particular input conditions. The only difference is that the edges that connected the pairs of states differing in only the level of either glucose or circulating fatty acids are no longer connected due to the absence of the self-inhibitory loop governing the change in level, neatly dividing the previously large 1,024 state STG into four smaller 256 state STGs for detailed analysis. We then proceeded with comparing the HTGs of each model with the respective input combination discussed below. Figure 4.5 shows the HTGs of Model 1 for all four input combinations.

Input 1: Glucose only

This input combination models the availability of circulating glucose and the unavailability of circulating fatty acids. All four models of PDK regulation collapse the STG into a two node HTG containing all 256 states pertaining to this input combination. The root node is an irreversible component that does not contain any incoming edges, and consists of either 31, 33, 29 or 39 states for models 1, 2, 3 and 4 respectively. This root node shows a situation from which the system is able to recover to stable behaviours, but is unable to return to the original situation (hence the irreversible component). The remaining states are collapsed into the leaf node, which represents a terminal SCC. The system can remain in the terminal SCC indefinitely, following the cyclic behaviours it

Hierarchical Transition Graphs of Model 1

One each for the four input conditions

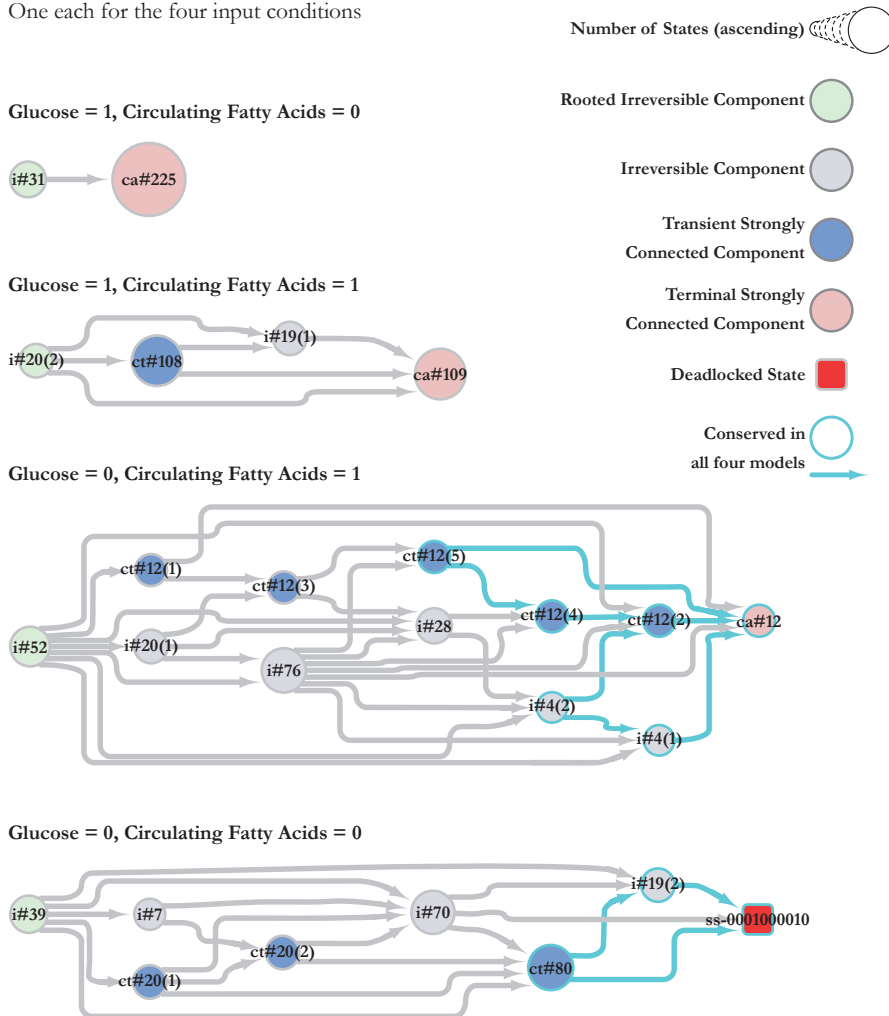


Figure 4.5: Hierarchical transition graphs (HTGs) of Model 1. Each node is labelled with a set of letters denoting the type of the node, followed by the number of state that node is representing. For states having the same type and number of states, an added number in parentheses is added to the name to differentiate them. The size of the node represents the number of states contained within the node. The irreversible components ('i#') represent states which do not contain any cycles or homeostatic behaviours. The strongly connected components ('ct#' and 'ca#') represented cyclic or homeostatic behaviours. The deadlocked state ('ss-') represents a single state where the system dynamics seize to function. The nodes and edges in cyan represent the nodes and edges which are conserved in all four models of PDK regulation. The HTGs of the remaining models 2, 3 and 4 are provided as Additional file 7.

represents. In case the input conditions change, the system would then be able to move to complementary state in one of the HTGs representing the new input combination.

Input 2: Both Glucose and Circulating Fatty Acids

In this input combination, both glucose and fatty acids are available for cellular metabolism. Models 1, 2 and 3 generate a four node HTG each, all consisting of two irreversible components (one of which is a root node), and two SCCs (one of which is a terminal SCC). Model 4 generates ten nodes, consisting of five irreversible components and five SCCs. We observed that the smaller HTGs of the first three models represent very similar sets of dynamics because the logical parameters for these models tie the regulation of PDK completely or partially to the TCA cycle. In model 4, this dependency is nullified as the Fatty Acid mediated activation via PPAR γ signalling is able to bypass the TCA cycle. This results in stable cyclic behaviours in the terminal SCC untying from each other and dispersing into smaller groups of cyclic behaviours, accompanied by irreversible components. However, the non-terminal SCC remains the same as those of models 1, 2 and 3. In essence, the larger structure of the HTGs remains intact.

Input 3: Circulating Fatty Acids only

In this input combination, only circulating fatty acids are available for metabolism. All four models show strong divergence in their respective HTGs, with different types, number, and sizes of nodes. The only conserved pattern in the HTGs is the terminal SCC, along with three more non-terminal SCCs leading to the terminal SCC. This sub-network maintains both the number of states, as well as the edges between the respective SCCs across all four models, suggesting that although the change in PDK regulation has a strong effect in the upstream behaviours in the HTGs, the system converges to the same behaviours and patterns.

Input 4: No Glucose or Circulating Fatty Acids

This input combination represents an extreme scenario where neither circulating glucose nor circulating fatty acids are available to fuel metabolism. As such, the system moves towards cell death, which can be seen as the only deadlocked state (also known as stable state) in the all four HTGs, labelled as 'ss-0001000010'. Models 2, 3 and 4 show very similar behaviour patterns in terms of the node types and the edges between them, whereas model 1 shows some behaviours as separate nodes instead. However, as with the previous input combination, this one also has a sub-network conserved between all four HTGs consisting of an

SCC, an irreversible component, and the deadlocked state.

Perturbation of Regulatory Components for Impact Propagation

In the perturbation analysis, we opted to perturb the non-metabolite inhibitors of the regulatory network to either ectopic or knockout levels to observe how their effects propagate through the system. The reason for selecting non-metabolite inhibitors is that the metabolite inhibitors are derived from the metabolic processes themselves, thus perturbing those would create a self-fulfilling prophecy in terms of gauging the effects of the regulators on cellular metabolic flexibility. The only two entities in our regulatory network fulfilling this criterion are PDK and AMPK.

We started with perturbing PDK to knockout level by restricting it to level 0. We then initialised the system with all entities at level 0 and tested it with all four input combinations. We observed the same behaviours as those in the network analysis for all input combinations, except for when both circulating glucose and circulating fatty acids are available for metabolism. For this input combination, we observed that when PDK is locked to level 0 the STG shows that the Acetyl-CoA is derivable from both glucose and fatty acid sources simultaneously, contrary to the known biological properties which were also checked via CTL model checking. When we allowed the system to change the PDK level to 1, we immediately observe only fatty acid driven Acetyl-CoA production.

We then proceeded to test PDK at ectopic level by restricting it to level 1 and testing again with the four input combinations (while the remaining entities of the system are initialised at level 0). We observed that, like previously, the behaviours were similar to those observed in the network analysis except for one input combination, this time it being the availability of circulating glucose and the absence of circulating fatty acids. The STG of this combination showed that there was no Acetyl-CoA production as the ectopic activity of PDK was barring Pyruvate conversion to Acetyl-CoA, and there were no available fatty acids to fuel metabolism. When we allowed the system to change the PDK level to 0, glucose driven metabolism resumed.

Lastly, we perturbed AMPK using the same method used for PDK. We did not observe any changes when AMPK was restricted to level 0. However, when restricted to level 1, we observed that changes in the behaviours generated when both the circulating glucose and circulating fatty acids are available. For this input combination, the ectopic activity of AMPK led to a fatty acid preferred metabolism, which, in turn, led to ectopic PDK levels, leading solely to fatty acid driven metabolism, instead of any switching behaviour. Here, we observed that AMPK perturbation still acted through PDK mediated regulation.

Discussion

In this study, we have utilised a logical modelling and network analysis workflow to assess the hypothesis of the PDC-PDK regulatory switch being a key regulatory mechanism behind cellular metabolic flexibility. We start with the construction of the regulatory network by abstracting the cellular network of metabolic flexibility reviewed earlier [5]. The regulatory network, consisting of ten entities and nineteen interactions, covers both the glucose and fatty acid oxidative metabolism pathways, merging them with the fatty acid synthesis pathway along with AMPK and PPAR γ signalling pathways. The selection of the logical parameters for all but one of the entities was relatively straightforward as the biological processes being represented have been well studied [25, 26, 29, 34]. The remaining entity, PDK, had four possible sets of logical parameters as it has competing activators and inhibitors in the regulatory network. As opposed to other entities, the competitive regulation of PDK is difficult to decipher because the interactions are not direct outputs of biochemical reactions. On top of this, various regulators of PDK have different intensities of regulations for the four PDK isoenzymes in different tissues [41].

The interactions in the regulatory network are both simplified and abstracted so we relied on system verification using model checking to test four possible logical parameter sets for PDK regulation. CTL formulae were used to formulate both oxidative metabolisms and the known behaviours of the PDC-PDK regulatory interactions. The model checker, SMBioNet [22] verified all four logical parameter sets of PDK to contain the formulated known behaviours, leading us to proceed further with four models differing on the regulation of PDK. The verification of all four logical parameter sets shows the biological plausibility of the four different types of PDK regulation, in line with the multiple intensities and tissue specific regulation discussed in [41]. In addition, the various STGs and HTGs generated by the four models show that the system eventually settled into very similar dynamics for the different sets of inputs, again supporting the biological plausibility of the four models. The STGs and HTGs differed only in the upstream regulation of the dynamics because of the difference in PDK regulation. However, the simplicity of the abstraction used in our regulatory network, both in terms of representing the four PDK isoenzymes as a single entity as well as merging the various regulations (such as acetyl-CoA, NADH and ATP mediated activation of various PDKs into a single edge), limits our model to being non-tissue specific. This limitation also affects the elucidation of the contexts involved and/or required by the four models to exhibit the behaviours presented in this study. As the base model itself is not completely tissue specific, it is possible that these four models broadly represent different paths of metabolic flexibility in different tissues. However, it is equally likely that these four models can exist in the same tissue

but at different times because of other effects not modelled in this study (such as epigenetic effects over time).

As an additional verification step, we constructed the larger cellular metabolic flexibility network from our review [5] in the software GINsim, consisting of 63 entities and 81 edges. This network included one modification from the network depicted in our review – two inhibitory edges, one from ATP to AMP and the other from NADH to NAD⁺, were added to account for their cycling in the cell. We then proceeded to find the deadlocked states in the larger network to compare with the network we had manually constructed in this study. We found that the larger network generated only one deadlocked state. However, comparing only the ten entities modelled in the network in this study, we found that the levels of these ten entities were the same for both deadlocked states. We also used the reduction tool offered in GINsim that computationally reduces a model. We reduced the larger network, deselecting the ten entities found in our manually constructed model in this study, making the reduction algorithm preserve them. The reduced model generated by GINsim consisted of eleven entities, with the additional entity being the node for TCA cycle. This reduced model also generated a single deadlocked state, as was expected because the reduction algorithm preserves the mathematical constraints in the model. This deadlocked state only showed two entities deadlocked at level 1, PDK and AMPK, the same as the ones from our manually constructed model (shown in the “Input 4:” sub-sub-section of the Results section). These comparisons serve as an alternative verification method for the manual construction of the network presented in this study. The additional models are provided as Additional file 8.

The perturbation analysis, done by restricting certain components to knockout or ectopic expressions, allowed us to test the propagation of regulation in the regulatory network when the negative regulators in the system malfunctioned. We elected to perturb the non-metabolite negative regulators, namely PDK and AMPK, because we wanted to test the regulation of the switching of metabolism from glucose to fatty acids independent of the increase or decrease in metabolite concentrations implied in logical modelling. The results of the perturbation analysis reveal that malfunctioning of PDK has a direct effect on the switching of metabolism, which is in line with our hypothesis. What is more interesting is that the perturbation of the only other non-metabolite inhibitor modelled in our regulatory network, AMPK, also propagated its effects through persistent activation or availability of PDK, providing additional support to our hypothesis. The results of the perturbation analysis, coupled with the conserved dynamics in all four models reinforces the hypothesis that the PDK isoenzymes are a key regulatory element of cellular metabolic flexibility via the PDK-PDK interaction.

This propagation of the regulatory effect places the PDK enzymes squarely in the middle of perturbed metabolism, as can be seen in cancer studies [42, 43] where PDKs were found to be over-expressed. The reprogramming of cellular metabolism has been identified as a new hallmark of cancer where the cellular metabolism of the cancerous cells moves away from complete glucose oxidation to just glycolysis [44]. In the aforementioned cancer studies, PDK expression was suppressed via treatment with Dichloroacetate (DCA), an inexpensive small molecule suppressor, to switch metabolism over to glucose oxidation. Similar treatments in other studies has shown that the DCA treatment caused apoptosis in cancer cells [45–48]. However, on the other side of the spectrum, a study targeting Alzheimer’s disease via rat central nervous system cell line models found that the overexpression of PDK1 (along with lactate dehydrogenase A) conferred a resistance to Ameloid β and other toxins, thereby mitigating some of the mechanisms underlying Alzheimer’s disease progression [49]. When taken collectively, these studies indicate that the balance of PDK enzyme expressions play an important role in the health of various cell types, thus, relying on the cellular metabolic flexibility through the PDC-PDK regulatory interaction. In addition, this maintenance of cellular metabolic flexibility as well as the tweaking of the PDC-PDK switch can be further used to supply new drug targets for the aforementioned ailments and conditions.

In cell types reliant on a single substrate for oxidative metabolism, the role of the metabolic switch is slightly different, likely to regulate the rate of oxidative metabolism instead of metabolic flexibility. One example of such cell types are Endothelial cells which utilise glucose. We took the network resource developed in our previous study [5] and visualised baseline RNA-Seq expression data for endothelial cells from the BLUEPRINT Epigenome project [50] (available from ArrayExpress as E-MTAB-3827). These endothelial cells were extracted from the umbilical vein during proliferating and resting states (the visualisation is provided as Additional file 9). We observed that the expression of PDKs is very low, and no expression of PPAR γ is taking place, indicating the absence of cellular fatty acid mediated PPAR γ signalling. Interestingly, there is expression of Stearoyl-CoA desaturase 1 (SCD), an enzyme from the fatty acid (re)synthesis pathway. However, studies have shown that SCD in endothelial cells plays a vital role in mitigating laminar stress [51, 52], thereby justifying its expression and indicating the limited role of the PDC-PDK regulatory switch in such cell types.

In essence, all four models of PDK regulation analysed in this study, coupled with the network and perturbation analyses, strongly suggest the PDK mediated inactivation of PDC as a key switching mechanism of cellular metabolism from glucose to fatty acids, and therefore, a key regulator of cellular metabolic flexibility. However, the model has its limitations, not least of which is that logical modelling is time and quantity independent. This means that

that although our models suggests the PDC-PDK switch as a key regulator, it does not tell us anything about the intensity and duration of this regulation. In addition, wet-lab experimentation is still required to validate our findings, the design of which is difficult in itself since many of the interactions and regulations modelled in the system happen at extremely short time scales, the accurate measurement of which is tricky at best. These design constraints themselves are one of the reasons why quantitative data was sparse and lacking at the time of this study, resulting in the logical modelling conducted in this study. While our study focuses on the substrate switching in metabolism and specifically the enzymatic regulation of this substrate switching, there are other factors/mechanisms of the regulation of cellular metabolism that are not considered in our current model. One of these is the ratio of deuterium to hydrogen in the cellular environment, which has been shown to affect cellular metabolism [53, 54]. Although this aspect is out of the scope of the current study, it would be interesting to further investigate its impact on the regulation of cellular metabolism and metabolic flexibility in future studies.

Conclusion

In this study, we have modelled and analysed cellular metabolic flexibility using logical modelling and network analysis. The results of our models strongly suggest that the PDC-PDK regulatory switch plays an important role in the regulation of cellular metabolic flexibility, revolving around the TCA cycle and the oxidative metabolism of glucose and fatty acids. The results support the hypothesis that this regulatory switch relies on the regulation of PDK itself, and thus PDK regulation acts as the pivot balancing cellular metabolic flexibility between available nutrients.

Availability of data and material

The Blueprint Consortium data visualised in this study (Additional File 7) is available from ArrayExpress as E-MTAB-3827 [<https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-3827/>].

References

- [1] L. Hue and H. Taegtmeyer, "The randle cycle revisited: a new head for an old hat," *American Journal of Physiology - Endocrinology And Metabolism*, vol. 297, no. 3, pp. E578–E591, 2009.
- [2] S. N. Vallerie and K. E. Bornfeldt, "Metabolic flexibility and dysfunction in cardiovascular cells," *Arteriosclerosis, Thrombosis, and Vascular Biology*, vol. 35, no. 9, pp. e37–e42, 2015.
- [3] J. J. Dubé, P. M. Coen, G. DiStefano, A. C. Chacon, N. L. Helbling, M. E. Desimone, M. Stafanovic-Racic, K. C. Hames, A. A. Despines, F. G. S. Toledo, and B. H. Goodpaster, "Effects of acute lipid overload on skeletal muscle insulin resistance, metabolic flexibility, and mitochondrial performance," *American Journal of Physiology - Endocrinology And Metabolism*, vol. 307, no. 12, pp. E1117–E1124, 2014.
- [4] M. K. C. Hesselink, V. Schrauwen-Hinderling, and P. Schrauwen, "Skeletal muscle mitochondria as a target to prevent or treat type 2 diabetes mellitus," *Nat Rev Endocrinol*, vol. 12, no. 11, pp. 633–645, 2016.
- [5] S. H. K. Tareen, M. Kutmon, M. E. Adriaens, E. C. M. Mariman, T. M. de Kok, I. C. W. Arts, and C. T. Evelo, "Exploring the cellular network of metabolic flexibility in the adipose tissue," *Genes & Nutrition*, vol. 13, no. 1, p. 17, 2018.
- [6] S. H. K. Tareen, M. E. Adriaens, I. C. W. Arts, T. M. de Kok, R. G. Vink, N. J. T. Roumans, M. A. van Baak, E. C. M. Mariman, C. T. Evelo, and M. Kutmon, "Profiling cellular processes in adipose tissue during weight loss using time series gene expression," *Genes*, vol. 9, no. 11:525, 2018.
- [7] R. G. Vink, N. J. T. Roumans, L. A. J. Arkenbosch, E. C. M. Mariman, and M. A. van Baak, "The effect of rate of weight loss on long-term weight regain in adults with overweight and obesity," *Obesity*, vol. 24, no. 2, pp. 321–327, 2016.
- [8] N. Le Novère, "Quantitative and logic modelling of molecular and gene networks," *Nature Reviews Genetics*, vol. 16, p. 146, 2015.
- [9] G. Schreiber, G. Haran, and H. X. Zhou, "Fundamental aspects of protein-protein association kinetics," *Chemical Reviews*, vol. 109, no. 3, pp. 839–860, 2009.
- [10] M. Schlosshauer and D. Baker, "Realistic protein–protein association rates from a simple diffusional model neglecting long-range interactions, free energy barriers, and landscape ruggedness," *Protein Science*, vol. 13, no. 6, pp. 1660–1669, 2009.
- [11] W. Abou-Jaoudé, P. Traynard, P. T. Monteiro, J. Saez-Rodriguez, T. Helikar, D. Thieffry, and C. Chaouiya, "Logical modeling and dynamical analysis of cellular networks," *Frontiers in Genetics*, vol. 7, no. 94, 2016.
- [12] M. K. Morris, J. Saez-Rodriguez, P. K. Sorger, and D. A. Lauffenburger, "Logic-based models for the analysis of cell signaling networks," *Biochemistry*, vol. 49, no. 15, pp. 3216–3224, 2010.
- [13] J. Saez-Rodriguez, L. Simeoni, J. A. Lindquist, R. Hemenway, U. Bommhardt, B. Arndt, U.-U. Haus, R. Weismantel, E. D. Gilles, S. Klamt, and B. Schraven, "A logical model provides insights into t cell receptor signaling," *PLOS Computational Biology*, vol. 3, no. 8, p. e163, 2007.
- [14] R. Z. Paracha, J. Ahmad, A. Ali, R. Hussain, U. Niazi, S. H. K. Tareen, and B. Aslam, "Formal modelling of toll like receptor 4 and jak/stat signalling pathways: Insight into the roles of socs-1, interferon- β and proinflammatory cytokines in sepsis," *PLOS ONE*, vol. 9, no. 9, p. e108466, 2014.

-
- [15] J. Ahmad, U. Niazi, S. Mansoor, U. Siddique, and J. Bibby, "Formal modeling and analysis of the mal-associated biological regulatory network: Insight into cerebral malaria," *PLOS ONE*, vol. 7, no. 3, p. e33532, 2012.
- [16] R. Albert and J. Thakar, "Boolean modeling: a logic-based dynamic approach for understanding signaling and regulatory networks and for making useful predictions," *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, vol. 6, no. 5, pp. 353–369, 2014.
- [17] R. Thomas and R. D'Ari, *Biological Feedback*. CRC Press, Inc., 1990.
- [18] A. Naldi, D. Berenguier, A. Fauré, F. Lopez, D. Thieffry, and C. Chaouiya, "Logical modelling of regulatory networks with ginsim 2.3," *Biosystems*, vol. 97, no. 2, pp. 134–139, 2009.
- [19] S. Peres and C. Jean-Paul, "Contribution of computational tree logic to biological regulatory networks: Example from *pseudomonas aeruginosa*," in *Computational Methods in Systems Biology* (C. Priami, ed.), (Berlin, Heidelberg), pp. 47–56, Springer Berlin Heidelberg, 2003.
- [20] A. Naldi, C. Hernandez, W. Abou-Jaoudé, P. T. Monteiro, C. Chaouiya, and D. Thieffry, "Logical modeling and analysis of cellular regulatory networks with ginsim 3.0," *Frontiers in Physiology*, vol. 9, p. 646, 2018.
- [21] D. Bérengruier, C. Chaouiya, P. T. Monteiro, A. Naldi, E. Remy, D. Thieffry, and L. Tichit, "Dynamical modeling and analysis of large cellular regulatory networks," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 23, no. 2, p. 025114, 2013.
- [22] Z. Khalis, J.-P. Comet, A. Richard, and G. Bernot, "The smbionet method for discovering models of gene regulatory networks," *Genes, genomes and genomics*, vol. 3, no. 1, pp. 15–22, 2009.
- [23] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker, "Cytoscape: A software environment for integrated models of biomolecular interaction networks," *Genome Research*, vol. 13, no. 11, pp. 2498–2504, 2003.
- [24] A. Saadatpour, R. Albert, and T. C. Reluga, "A reduction method for boolean network models proven to conserve attractors," *SIAM Journal on Applied Dynamical Systems*, vol. 12, no. 4, pp. 1997–2011, 2013.
- [25] Kdahlquist, A. Pico, M. van Iersel, T. Kelder, and et al., "Tca cycle (homo sapiens)." <https://www.wikipathways.org/index.php?title=Pathway:WP78&oldid=90661>, Accessed on: 15-06-2018.
- [26] Kdahlquist, K. Hanspers, M. van Iersel, S. Coort, and et al., "Glycolysis and gluconeogenesis (homo sapiens)." <https://www.wikipathways.org/index.php?title=Pathway:WP534&oldid=94762>, Accessed on: 15-06-2018.
- [27] S. Zhang, M. W. Hulver, R. P. McMillan, M. A. Cline, and E. R. Gilbert, "The pivotal role of pyruvate dehydrogenase kinases in metabolic flexibility," *Nutrition & Metabolism*, vol. 11, no. 1, p. 10, 2014.
- [28] H. A. Krebs, "The history of the tricarboxylic acid cycle," *Perspectives in Biology and Medicine*, vol. 14, no. 1, pp. 154–172, 1970.
- [29] C. Evelo, M. Adriaens, A. Pico, J. Saito, and et al., "Fatty acid biosynthesis (homo sapiens)." <https://www.wikipathways.org/index.php?title=Pathway:WP357&oldid=94197>, Accessed on: 15-06-2018.
- [30] A. A. Spector, "Plasma lipid transport," *Clinical Physiology and Biochemistry*, vol. 2(2-3), no. 0252-1164 (Print), pp. 123–134, 1984.

-
- [31] R. M. Denton and P. J. Randle, "Citrate and the regulation of adipose-tissue phosphofructokinase," *Biochemical Journal*, vol. 100(2), no. 0264-6021 (Print), pp. 420-423, 1966.
- [32] G. V. Gnoni, P. Priore, M. J. H. Geelen, and L. Siculella, "The mitochondrial citrate carrier: Metabolic role and regulation of its activity and expression," *IUBMB Life*, vol. 61, no. 10, pp. 987-994, 2009.
- [33] J. D. Warfel, B. Vandanmagsar, O. S. Dubuisson, S. M. Hodgeson, C. M. Elks, E. Ravussin, and R. L. Mynatt, "Examination of carnitine palmitoyltransferase 1 abundance in white adipose tissue: implications in obesity research," *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology*, vol. 312, pp. R816-R820, 2017.
- [34] M. Adriaens, P. Ahles, F. Jagers, C. Evelo, and et al., "Fatty acid beta oxidation (homo sapiens)." <https://www.wikipathways.org/index.php?title=Pathway:WP143&oldid=94768>, Accessed on: 15-06-2018.
- [35] M. Daval, F. Fougere, and P. Ferré, "Functions of amp-activated protein kinase in adipose tissue," *The Journal of Physiology*, vol. 574, no. 1, pp. 55-62, 2006.
- [36] J.-E. Lee and K. Ge, "Transcriptional and epigenetic regulation of ppar γ expression during adipogenesis," *Cell & Bioscience*, vol. 4, no. 1, p. 29, 2014.
- [37] T. Cadoudal, E. Distel, S. Durant, F. Fouque, J.-M. Blouin, M. Collinet, S. Bortoli, C. Forest, and C. Benelli, "Pyruvate dehydrogenase kinase 4 regulation by thiazolidinediones and implication in glyceroneogenesis in adipose tissue," *Diabetes*, vol. 57, no. 9, pp. 2272-2279, 2008.
- [38] T. Varga, Z. Czimmerer, and L. Nagy, "Ppars are a unique set of fatty acid regulated transcription factors controlling both lipid metabolism and inflammation," *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, vol. 1812, no. 8, pp. 1007-1022, 2011.
- [39] R. W. Schwenk, G. P. Holloway, J. J. F. P. Luiken, A. Bonen, and J. F. C. Glatz, "Fatty acid transport across the cell membrane: Regulation by fatty acid transporters," *Prostaglandins, Leukotrienes and Essential Fatty Acids (PLEFA)*, vol. 82, no. 4, pp. 149-154, 2010.
- [40] I. J. Goldberg, R. H. Eckel, and N. A. Abumrad, "Regulation of fatty acid uptake into tissues: lipoprotein lipase- and cd36-mediated pathways," *Journal of Lipid Research*, vol. 50, no. Supplement, pp. S86-S90, 2009.
- [41] M. M. Bowker-Kinley, P. Davis, W. F. Wu, R. A. Wu, P. F. Harris, K. M. Harris, R. A. Popov, and K. M. Popov, "Evidence for existence of tissue-specific regulation of the mammalian pyruvate dehydrogenase complex," *Biochemical Journal*, vol. 329(1), no. 0264-6021 (Print), pp. 191-196, 1998.
- [42] S. Sradhanjali, D. Tripathy, S. Rath, R. Mittal, and M. M. Reddy, "Overexpression of pyruvate dehydrogenase kinase 1 in retinoblastoma: A potential therapeutic opportunity for targeting vitreous seeds and hypoxic regions," *PLOS ONE*, vol. 12, no. 5, p. e0177744, 2017.
- [43] N. Ho and B. L. Coomber, "Pyruvate dehydrogenase kinase expression and metabolic changes following dichloroacetate exposure in anoxic human colorectal cancer cells," *Experimental Cell Research*, vol. 331, no. 1, pp. 73-81, 2015.
- [44] D. Hanahan and R. Weinberg, "Hallmarks of cancer: The next generation," *Cell*, vol. 144, no. 5, pp. 646-674, 2011.

-
- [45] G. M. Saed, N. M. Fletcher, Z. L. Jiang, H. M. Abu-Soud, and M. P. Diamond, "Dichloroacetate induces apoptosis of epithelial ovarian cancer cells through a mechanism involving modulation of oxidative stress," *Reproductive Sciences*, vol. 18, no. 12, pp. 1253–1261, 2011.
- [46] E. D. Michelakis, G. Sutendra, P. Dromparis, L. Webster, A. Haromy, E. Niven, C. Maguire, T. L. Gammer, J. R. Mackey, D. Fulton, B. Abdulkarim, M. S. McMurtry, and K. C. Petruk, "Metabolic modulation of glioblastoma with dichloroacetate," *Science Translational Medicine*, vol. 2, no. 31, p. 31ra34, 2010.
- [47] C.-W. Lu, S. Lin, C.-W. Chien, S.-C. Lin, C.-T. Lee, B.-W. Lin, J. C. Lee, and S.-J. Tsai, "Overexpression of pyruvate dehydrogenase kinase 3 increases drug resistance and early recurrence in colon cancer," *The American journal of pathology*, vol. 179 3, pp. 1405–14, 2011.
- [48] J. Y. Y. Wong, G. S. Huggins, M. Debidda, N. C. Munshi, and I. De Vivo, "Dichloroacetate induces apoptosis in endometrial cancer cells," *Gynecologic Oncology*, vol. 109, no. 3, pp. 394–402, 2008.
- [49] J. T. Newington, T. Rappon, S. Albers, D. Y. Wong, R. J. Rylett, and R. C. Cumming, "Overexpression of pyruvate dehydrogenase kinase 1 and lactate dehydrogenase a in nerve cells confers resistance to amyloid β and other toxins by decreasing mitochondrial respiration and reactive oxygen species production," *Journal of Biological Chemistry*, vol. 287, no. 44, pp. 37245–37258, 2012.
- [50] D. Adams, L. Altucci, S. E. Antonarakis, J. Ballesteros, S. Beck, A. Bird, C. Bock, B. Boehm, E. Campo, A. Caricasole, F. Dahl, E. T. Dermitzakis, T. Enver, M. Esteller, X. Estivill, A. Ferguson-Smith, J. Fitzgibbon, P. Flicek, C. Giehl, T. Graf, F. Grosveld, R. Guigo, I. Gut, K. Helin, J. Jarvius, R. Küppers, H. Lehrach, T. Lengauer, Å. Lernmark, D. Leslie, M. Loeffler, E. Macintyre, A. Mai, J. H. A. Martens, S. Minucci, W. H. Ouwehand, P. G. Pelicci, H. Pendeville, B. Porse, V. Rakyán, W. Reik, M. Schrappe, D. Schübeler, M. Seifert, R. Siebert, D. Simmons, N. Soranzo, S. Spicuglia, M. Stratton, H. G. Stunnenberg, A. Tanay, D. Torrents, A. Valencia, E. Vellenga, M. Vingron, J. Walter, and S. Willcocks, "Blueprint to decode the epigenetic signature written in blood," *Nature Biotechnology*, vol. 30, p. 224, 2012.
- [51] X. M. Qin, J. Tian, P. Zhang, Y. Fan, L. Chen, Y. fei Guan, Y. Fu, Y. Zhu, S. Chien, and N. Wang, "Laminar shear stress up-regulates the expression of stearoyl-coa desaturase-1 in vascular endothelial cells.," *Cardiovascular research*, vol. 74 3, pp. 506–14, 2007.
- [52] X. Liu, M. S. Strable, and J. M. Ntambi, "Stearoyl coa desaturase 1: Role in cellular inflammation and stress," *Advances in Nutrition*, vol. 2, no. 1, pp. 15–22, 2011.
- [53] A. Olgun, "Biological effects of deuteronation: Atp synthase as an example," *Theoretical Biology and Medical Modelling*, vol. 4, p. 9, Feb 2007.
- [54] C. Lewis, S. J. Parker, B. P. Fiske, D. McCloskey, D. Y. Gui, C. R. Green, N. I. Vokes, A. M. Feist, M. G. Vander Heiden, and C. M. Metallo, "Tracing compartmentalized NADPH metabolism in the cytosol and mitochondria of mammalian cells," *Molecular Cell*, vol. 55, pp. 253–263, Jul 2014.

Supplementary Materials

The following supplementary material is hosted by *Genes & Nutrition* at <https://doi.org/10.1186/s12263-019-0647-5>.

Additional file 1

Archive containing the SMBioNet and GINsim source files for toy example. (Zip, 1.66KB)

Additional file 2

PDF document showing a generic abstraction/reduction process for regulatory networks. (PDF, 573KB)

Additional file 3

Tabulated logical parameters for the regulatory network of cellular metabolic flexibility. (PDF, 60.1KB)

Additional file 4

Explanation and conversion of the biological properties into computation tree logic (CTL). (PDF, 97.5KB)

Additional file 5

Archive containing the SMBioNet and GINsim source file for the regulatory network of cellular metabolic flexibility. (Zip, 10.1KB)

Additional file 6

Cytoscape session file of the STGs of Models 1, 2, 3 and 4. (Cys, 967KB)

Additional file 7

Cytoscape session file of the HTGs of Models 1, 2, 3 and 4. (Cys, 82.4KB)

Additional file 8

Archive containing alternative verification models from the review article constructed in GINsim. (Zip, 11.3KB)

Additional file 9

Cytoscape session file visualising the RNA-seq expression data of endothelial cells. (Cys, 58.6KB)

5

Stratifying cellular metabolism during weight loss: an interplay of metabolism, metabolic flexibility and inflammation

Samar H.K. Tareen¹, Martina Kutmon^{1,2}, Theo M. de Kok^{1,3}, Edwin C.M. Mariman⁴, Marleen A. van Baak⁴, Chris T. Evelo^{1,2}, Michiel E. Adriaens^{1,6}, Ilja C.W. Arts^{1,5,6}

¹ Maastricht Centre for Systems Biology (MaCSBio), Maastricht University, the Netherlands

² Department of Bioinformatics – BiGCaT, NUTRIM Research School, Maastricht University, the Netherlands

³ Department of Toxicogenomics, GROW School of Oncology and Developmental Biology, Maastricht University, the Netherlands

⁴ Department of Human Biology, NUTRIM Research School, Maastricht University, the Netherlands

⁵ Department of Epidemiology, CARIM School for Cardiovascular Diseases, Maastricht University, the Netherlands

⁶ *These authors jointly supervised this work*

Published in: *Nature Scientific Reports* (2020) 10, 1651.

doi: 10.1038/s41598-020-58358-z

Abstract

Obesity is a global epidemic, contributing significantly to chronic non-communicable diseases, such as type 2 diabetes mellitus, cardiovascular diseases and metabolic syndrome. Metabolic flexibility, the ability of organisms to switch between metabolic substrates, is found to be impaired in obesity, possibly contributing to the development of chronic illnesses. Several studies have shown the improvement of metabolic flexibility after weight loss. In this study, we have mapped the cellular metabolism of the adipose tissue from a weight loss study to stratify the cellular metabolic processes and metabolic flexibility during weight loss. We have found that for a majority of the individuals, cellular metabolism was downregulated during weight loss, with gene expression of all major cellular metabolic processes (such as glycolysis, fatty acid β -oxidation etc.) being lowered during weight loss and weight maintenance. Parallel to this, the gene expression of immune system related processes involving interferons and interleukins increased. Previously, studies have indicated both negative and positive effects of post-weight loss inflammation in the adipose tissue with regards to weight loss or obesity and its co-morbidities; however, mechanistic links need to be constructed in order to determine the effects further. Our study contributes towards this goal by mapping the changes in gene expression across the weight loss study and indicates possible cross-talk between cellular metabolism and inflammation.

Introduction

Obesity has, in recent years, been classified as a global epidemic. WHO estimates have shown a steady increase in obesity with trends expected to continue [1]. This makes obesity a public health concern as it has been linked to cardiovascular diseases, type 2 diabetes and metabolic syndrome [2–4]. Therefore, weight loss and exercise is encouraged to counter obesity and, by extension, to prevent the progression of its co-morbidities [5, 6]. A focus in recent research has been to find the mechanistic links between obesity and associated co-morbidities by studying the underlying molecular processes [7, 8]. In one of our previous studies [9], we explored the expression of genes changing in the subcutaneous adipose tissue over time in response to caloric restriction. We observed that one of the clusters of co-expressed genes, which was being differentially expressed or differentially regulated over caloric restriction, was involved with a number of processes associated with the tricarboxylic acid (TCA) cycle and cellular metabolism. Interestingly, these processes are also part of what collectively constitutes the system of metabolic flexibility – the ability of organisms to switch metabolic substrates depending on nutrient conditions and needs [10]. In our study, this amounted to switching between glucose metabolism and fatty acid metabolism as these are the chief metabolic substrates for the majority of the tissue and cell types in the human body [11].

Studies have found an association between impaired metabolic flexibility (also known as metabolic inflexibility) and type 2 diabetes mellitus, cardiovascular diseases and metabolic syndrome [12–15]. Other studies have narrowed down the effects of metabolic inflexibility in the adipose tissue on the impairment of adipokine signalling as well as the clearance of circulating non-esterified fatty acids (as reviewed in [11] and [16]). Corpeleijn *et al* showed an improvement in metabolic flexibility after weight loss, by an increase in fatty acid uptake and oxidation in skeletal muscles, indicating a positive association between the two [17, 18]. Yet, several studies also show that formerly obese individuals exhibit metabolic inflexibility post weight loss in response to high-fat diets [19]. There is also evidence that individuals may be genetically predisposed to be more or less metabolically flexible [19] although the extent of this predisposition is not currently well established.

In this article, we utilise a metabolic flexibility gene set based on cellular metabolism to cluster transcriptomics and proteomics expression data from a weight loss study. The purpose of this exercise is to identify and analyse expression profiles of individuals clustered by cellular metabolism centring on metabolic flexibility. In our previous studies [9, 20, 21], we have shown how the cellular regulation of the tricarboxylic acid cycle, as well as the switching of substrates between glucose and fatty acids have a central role to play in

obesity and chronic ailments such as type 2 diabetes mellitus and metabolic syndrome. Capitalising on this work, we curated a list of 291 genes involved in, or associated with the regulation of, cellular metabolism and substrate switching. The data was then clustered on samples of the individuals to identify pathways and processes associated with the change in metabolic profiles.

Materials and methods

Dataset

We used expression data from the 'Yoyo study' [22] (Clinical Trial ID: NCT01559415, www.clinicaltrials.gov), a human weight loss study exploring the yo-yo effect seen as subsequent weight regain. The study covers two diets: a low calorie diet (LCD) of 1,250 kcal/day and a very low calorie diet (VLCD) of 500 kcal/day. The study design has four time points for data collection: the first being before dietary intervention (time point 1), followed by one at the end of weight loss diet (12 weeks for LCD, 5 weeks for VLCD) (time point 2), a third after a 4-week weight maintenance period (time point 3), and follow-up 9 months later (time point 4). All the participants of the study were Caucasian with a BMI between 28 kg/m² and 35 kg/m², aged between 32 and 67 years old (median age of 51 years). Additionally, the amount of weight lost by the individuals in both diets was almost the same. For further details, we refer to the original study publication [22].

Transcriptomics

The Yoyo study has microarray transcriptomics measurements for 46 individuals for the first three data collection points of the study, with samples from the subcutaneous adipose tissue of the individuals. The transcriptomics data is available on Gene Expression Omnibus under ID: GSE77962. The array platform is Affymetrix Human Gene ST 1.1 arrays.

In our previous analysis [9] we have analysed and normalised the transcriptomics data where we removed outliers, reducing the number of individuals to 44. We have also performed background noise correction on the measurements by removing all genes with median expression equal to or less than the expression of Y-chromosome genes in female individuals, giving us measurements for a total of 18,113 genes/transcripts.

Proteomics

The proteomics data from the Yoyo study are from the adipose tissue samples at time points 1 through 3 using liquid chromatography-mass spectrometry (LC-MS/MS) [23]. Both the samples and controls were labelled with TMT

isobaric mass tagging labelling reagent (10plex, Thermo Scientific) and measured in the Dionex ultimate 3000 nanoflow HPLC instrument. The mass spectrometry data was then queried against UniProt using Sequest HT Proteome Discoverer 2.1 (Thermo Scientific) [23]. In this study, we use the corrected proteomics data (corrected for between and within runs) generated in the original study, and refer the readers to it for details regarding the parameters of the data generation. The proteomics data is freely available from the authors upon request.

Phenotypic measurements

The Yoyo study includes phenotypic and clinical measurements of the individuals. The anthropometric measurements [22] include sex, age, height, weight, body mass index (BMI), fat mass, fat-free mass, hip circumference, waist size, systolic and diastolic blood pressure, and physical activity score (calculated via the Baecke questionnaire for habitual physical activity [24]). The anthropometric measurements, apart from age and height, were measured at each data collection point in the original study.

Likewise, fasting forearm venous plasma [25] concentrations of glucose, insulin, free fatty acids, triglycerides and total cholesterol levels, and HOMA-IR index were measured for each data collection point. The adipose tissue arteriovenous flux measurements [26], however, were measured only in a subset of individuals (13 of the 38). These measurements were taken at time points 1 and 3, and included flux measurements for glucose, free fatty acids, total glycerol, free glycerol, triacylglycerol and lactate, as well as measurements of total blood flow, plasma blood flow and insulin. An additional measurement available from the Yoyo study is the *s*-value, which is a score for weight regain and/or maintenance defined as the ratio of weight loss maintained at follow up to the weight lost immediately following intervention. Mathematically, this is represented as $s\text{-value} = (W_{T4} - W_{T3}) / (W_{T2} - W_{T1})$ where W_{Tn} is the weight of the individual at time point n . The phenotypic measurements described here are also freely available from the authors upon request.

Metabolic flexibility gene list

In a previous study [20], we explored the cellular processes involved in cellular metabolic flexibility and presented them as a combined network of rate limiting steps involved in these processes. In the current study, we utilised the network and processes from this review to curate a list of genes/proteins involved in the regulation of cellular metabolic flexibility. This list is provided as Additional file 1 and consists of 291 genes/proteins.

Affinity network fusion

For the clustering of the transcriptomics and proteomics expression data across time points, we utilise affinity network fusion (ANF) [27]. ANF itself is built on top of the similarity network fusion algorithm [28]. These methods utilise multi-omics data to construct clusters of individuals that are more similar to each other than to others across their data spectrum. ANF is an optimisation over the original algorithm and incorporates weights for the various sources of the multi-omics data when constructing the affinity network to use with spectral clustering.

In our study, we utilised ANF to generate clusters of samples based on the transcriptomics and proteomics expression data filtered on the genes/proteins in the metabolic flexibility gene list. 38 individuals were used in the clustering totalling 106 samples across the three time points; 8 individuals were missing one measurement each at any of the three time points. The expression data used by the ANF algorithm consisted of the transcriptomics with 240 genes measured out of the 291 initially selected, and the proteomics with only 27 proteins. As the usable proteomics data were much smaller in amount and much larger in dynamic range compared to the transcriptomics, we opted to weight the two expression data types as the percentage coverage of the metabolic flexibility gene list, i.e., 240/291 for transcriptomics and 27/291 for proteomics.

The ANF algorithm also uses the k-nearest neighbour parameter to construct robust affinity networks. In our study, we tested $k = \{3, 4, 5, 6, 7, 8, 9, 10\}$ as the possible nearest neighbour values. In all cases, the algorithm produced two clusters, with only one sample changing cluster membership from $k = 8$ onwards. The cluster membership list is provided as Additional file 2. We chose the clustering at $k = 3$ for further analysis due to the robustness of the clustering given that the cluster membership did not change for several increasing k-values after 3.

Statistical analysis of phenotypic measurements

The means of the various phenotypic measurements of the respective samples of each cluster were compared using t-tests in R [29], whereas a chi-square test was used for the comparison of sex. The obtained p-values were corrected for multiple testing using the Benjamini & Hochberg approach, that controls the false discovery rate (FDR) [30].

Differential expression analysis

For the differential expression analysis between samples of the clusters at different time points, we used the whole transcriptomics data set analysed

via Limma v3.40.2 [31] package in R v3.5.1 [29]. The criteria for defining differentially expressed genes was $|\text{Fold change}| \geq 1.2$ and $p\text{-value} < 0.05$ [32]. For differential comparisons between samples from the same set of individuals, the individual was used as a blocking factor in the linear model fit. For comparisons between sets with a mixture of unique and repeated individuals, the correlation between samples from the same individuals was modelled using the duplicate correlation function in limma.

Gene ontology enrichment

Gene ontology enrichment was performed in Cytoscape v3.7.1 [33] using the ClueGO app for Cytoscape v2.5.4 [34]. Enrichment was performed for biological process ontologies (updated: 10-04-2019) with all evidences used for input except for 'inferred from electronic annotation' (All_without_IEA checkbox). Furthermore, Gene Ontology term fusion was allowed and pathway selection was restricted to pathways with $p\text{-value} \leq 0.05$. All other parameters in ClueGO remained the same.

Pathway overrepresentation analysis

For pathway overrepresentation analysis, we used PathVisio v3.3.0 [35]. PathVisio is a software that uses expression data to map and visualise various pathways from the WikiPathways database [36] while also using z-score to analyse which pathways are overrepresented in the database based on the mapped expression data. This allows us to focus on pathways in relevant biological context to infer the meaning behind the expression data. In this study, the pathway selection criteria used was a z-score ≥ 1.96 (critical score for 95% confidence interval), a $p\text{-value} < 0.05$ and at least three significantly changed genes (significance criteria the same as in differential expression analysis). For a tutorial on PathVisio, we redirect the reader to [37].

Results

Stratification of the metabolic profiles

Using the affinity network fusion (ANF) algorithm for clustering we generated a two-cluster system with samples across individuals and time points clustered with each other. The primary purpose of this clustering was to cluster the transcriptomics and proteomics expression patterns of the 291 metabolic flexibility related genes and proteins. The clustering thus generated two expression profiles, with individuals changing cluster membership over the duration of the study (visualised in Figure 5.1A). Figure 5.1B shows the respective individuals in each cluster membership pattern found. In this

subfigure, we can see that at the beginning of the caloric restriction (TP1) the majority of individuals are in Cluster B compared to Cluster A. Immediately after weight loss, at time point 2 (TP2), the cluster membership for Cluster A surges to become larger than Cluster B, with both clusters settling close to even after weight maintenance (TP3). This indicates that the metabolic profile of a large number of individuals changed during weight loss with some of them maintaining that change whereas others reverted to their original profile.

We next compared the phenotypic/clinical measurements between the clusters with 101 samples instead of 106 due to missing/incomplete data for two individuals. We identified significant differences in sex, height, BMI, fat mass, fat free mass, waist size and hip circumference. From the forearm venous plasma measurements, total cholesterol, insulin, free fatty acids, triglycerides levels as HOMA-IR index were found to be significantly different between the clusters. However, none of the adipose tissue arteriovenous flux measurements were found to be significant post multiple testing correction. Lastly, the *s*-value, defined as a value for weight regain and/or maintenance, was also not found to be significantly different between the two clusters. These results are collectively provided as Additional file 3.

Differential expression analysis of the 18,113 gene transcriptomics data between the samples of the two clusters showed that only 1,343 genes were differentially expressed between the two clusters ($|\text{Fold change}| \geq 1.2$ and $p\text{-value} < 0.05$). 150 of these differentially expressed genes had an absolute fold change greater than 1.5, and 28 of those had it greater than 2. The number of differentially expressed genes have been tabulated in Table 5.1. A gene ontology enrichment analysis [34] on the 1,343 genes revealed that apart from the expected metabolic processes (fatty acid and other lipid metabolism, oxidoreductase activity, chemical homeostasis etc.) there were processes related to tissue morphology and inflammatory response also being enriched. Figure 5.2 shows a pie chart of the gene ontology terms that were enriched.

A deeper look at the differentially expressed genes in the context of biological pathways (using PathVisio [35] for pathway overrepresentation analysis) also revealed several metabolic pathways being differentially affected between the clusters. It appears that the overall cellular metabolism in Cluster A is decreased across the board; in the pathway overrepresentation results, the majority of the genes in the electron transport chain, fatty acid biosynthesis, fatty acid beta oxidation, glycolysis, and citric acid pathways are expressed lower (at least with a fold change of 1.2) in Cluster A compared to Cluster B. However, these are to be expected due to the way the ANF algorithm constructed the clusters based on the 291 metabolic flexibility-, and thus metabolism-related genes. On the other hand, we observed that pathways associated with the immune system and tissue restructuring were expressed higher in Cluster A compared to Cluster B.

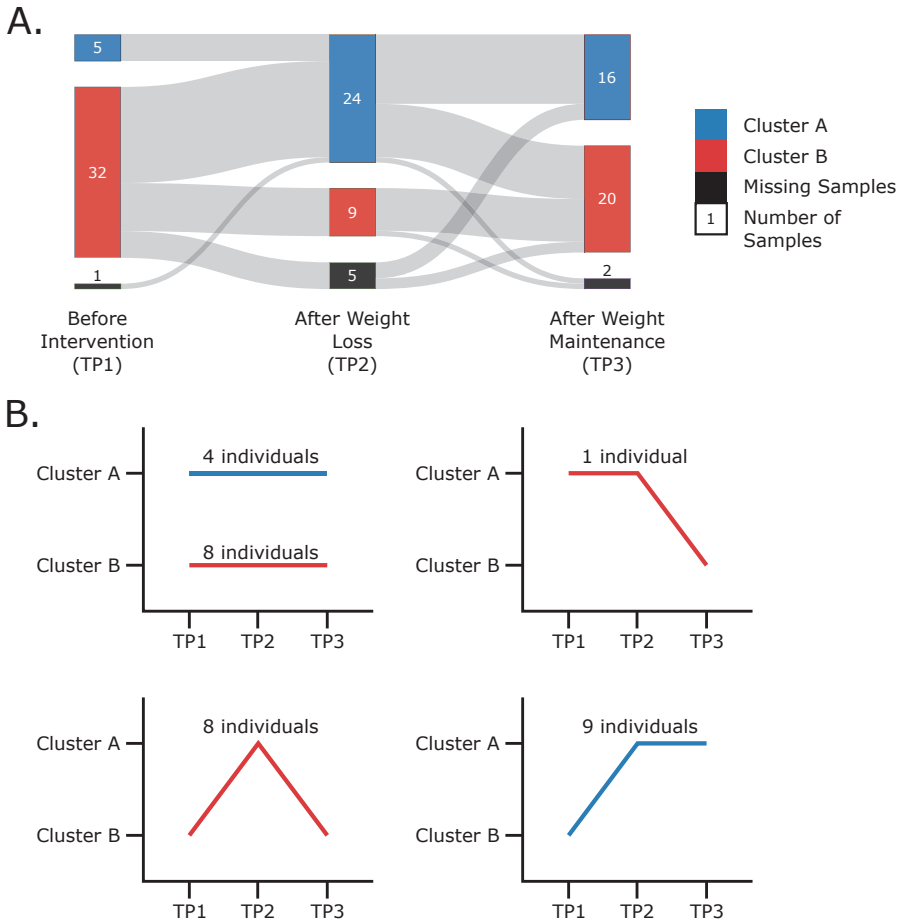


Figure 5.1: A) Sankey diagram showing the proportion of individuals changing or not changing clusters across the time points of the dietary intervention. Eight individuals had missing samples at different time points and thus, were not clustered at those time points. These samples are shown as black blocks at the respective time points. B) The various cluster membership patterns of the thirty individuals observed across the three time points, as well as the respective number of individuals following said pattern.

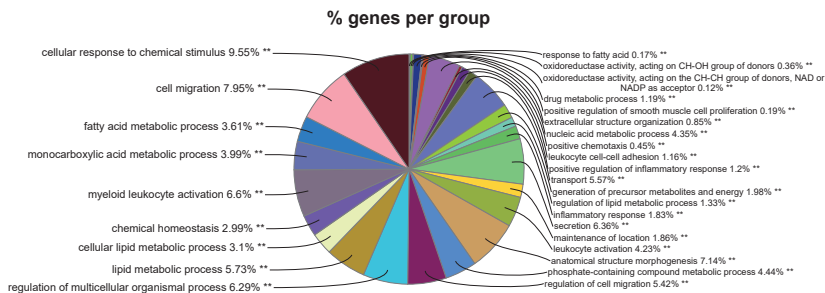


Figure 5.2: Pie chart showing the proportions of genes enriched for each of the gene ontology term for the differentially expressed genes between Cluster A and B. The differentially expressed genes were significant at p -value < 0.05 and an absolute fold change of at least 1.2.

Table 5.1: The number of differentially expressed genes across different groups for absolute fold changes of 1.2, 1.5 and 2. All counted genes are significant at p -value < 0.05 . Cluster A_st or Cluster B_st: individuals staying in the respective cluster A or B throughout the dietary intervention. Comparisons 1–5 are illustrated in Figure 5.3

	$ \text{FC} \geq 1.2$	$ \text{FC} \geq 1.5$	$ \text{FC} \geq 2$
Cluster A - Cluster B	1343	150	28
Comparison 1	1286	178	33
Comparison 2	282	24	1
Comparison 3	897	136	25
Comparison 4	669	75	16
Comparison 5	660	49	4
Cluster A_st - Cluster B_st	2838	460	97

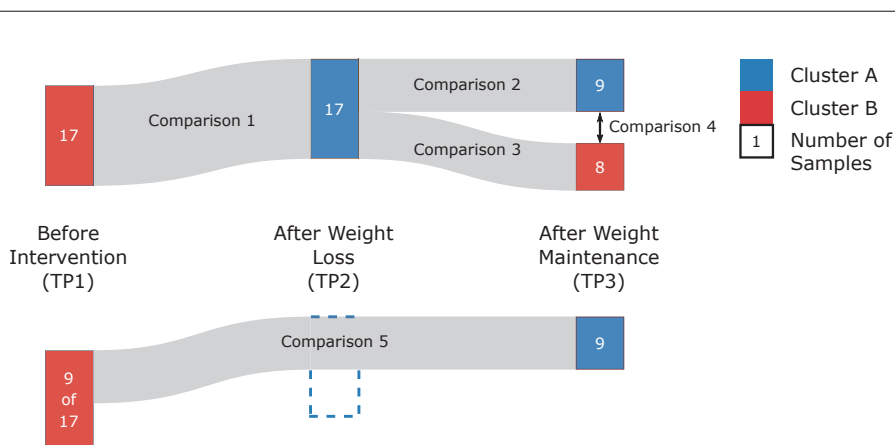


Figure 5.3: Illustration of the 17 subjects which switched from Cluster B to Cluster A during weight loss, after which 8 of the subjects reverted to Cluster B. The five comparisons used to break down and analyse the changes in the gene expression are also labelled.

In particular, Complement C6 and C7, inflammatory interleukins IL1B and IL8, and matrix metalloproteinases MMP2 and MMP9 were found to be expressed higher in Cluster A. The differential expression analysis results, gene ontology enrichment results and pathway overrepresentation results are provided as Additional files 4, 5 and 6 respectively.

Cluster membership patterns

Cluster membership changers

Figure 5.1B shows the number of individuals in each cluster membership pattern found. We observed that a total of 17 individuals change their cluster membership from Cluster B to Cluster A when moving from time point 1 (TP1) to time point 2 (TP2); 9 of these individuals then stay in Cluster A until time point 3 (TP3), while the remaining 8 revert to their original cluster. We further analysed this pattern in the following comparisons (illustrated in Figure 5.3; numbers of differentially expressed genes provided in Table 5.1).

Comparison 1 We first performed a paired differential expression analysis of all 17 individuals of this pattern between TP2 and TP1 to see the first changes that occurred when the individuals changed their cluster membership. A total of 1,286 differentially expressed genes were found in this comparison at absolute fold change equal to or greater than 1.2 and p-value less than 0.5. Gene ontology enrichment of these showed a similar pattern to the gene ontology enrichment when comparing the two clusters as a whole – the processes were broadly categorised into metabolic, immune/inflammatory and tissue

morphology. This is understandable as these changes would be the reason for these 17 individuals being clustered in Cluster A at TP2 as opposed to Cluster B at TP1. Similarly, the pathway enrichment results also follow the same pattern as that of the whole cluster comparison. Genes in metabolic processes such as fatty acid beta oxidation, electron transport chain complexes, glycolysis, lipogenesis, gluconeogenesis and amino acid metabolism were down regulated at TP2 Cluster A compared to when these individuals were in Cluster B at TP1. Genes involved in inflammation, on the other hand, were found to be upregulated in the same pattern as the whole cluster comparison.

Comparison 2 We performed a paired differential expression analysis between TP3 and TP2 for the 9 individuals that retained their new cluster to see what changes occurred even when the cluster membership did not change. In this comparison, we found 282 differentially expressed genes. Considering that this comparison is between the 9 individuals which stay in Cluster A at both TP2 and TP3, the low number is expected. Consequently, the gene ontology enrichment and pathway overrepresentation analysis also do not show any major differences between these samples. Of the minor differences found, peroxisome proliferator activated receptor gamma (PPAR γ), fatty acid synthase (FASN) and stearyl-CoA desaturase (SCD) were found to be upregulated in these individuals at TP3 compared to TP2. This is interesting, especially in the light of our previous studies covering the role of these proteins in metabolic flexibility. However, the lack of strong results from the two analyses overall make it difficult to infer the effects of their upregulation concretely.

Comparison 3 We performed a paired differential expression analysis between TP3 and TP2 for the 8 individuals that changed back to the original cluster to see what changes occurred when these individuals returned to their original cluster. In this comparison, we found 897 differentially expressed genes. The gene ontology enrichment analysis for these genes showed that these were also tied to cellular metabolic processes, with two immune system related terms also enriched. Compared to the previous two parts, the individuals in this analysis showed a reversed trend where the immune system related pathways were being downregulated at TP3 compared to TP2. The human complement system pathway and toll-like receptor (TLR) associated pathways showed downregulation of respective components. The matrix metalloproteinase MMP9 was also strongly downregulated. On the other hand, upregulated genes showed an upregulation of metabolism related processes across the board with lipid metabolism and biosynthesis showing the highest z-scores in pathway overrepresentation. Sterol regulatory element-binding protein (SREBP) signalling also showed strong upregulation compared to both TP2 in these individuals, as well as the previous two parts. Also in contrast to the previous two parts, as well as the whole cluster differential expression, was the

upregulation of leptin (LEP) and downregulation of leptin receptor (LEPR). In addition, glycolysis and gluconeogenesis enzymes were mostly found to be slightly upregulated.

Comparison 4 We performed a differential expression analysis between the 17 individuals at TP3 to see which changes were retained by these individuals by the final time point. In this comparison, a total of 669 genes were found to be differentially expressed genes between 17 individuals, 8 at TP3 in Cluster B and 9 at TP3 in Cluster A. The gene ontology enrichment showed a limited number of terms, all associated with cellular metabolism. Pathway overrepresentation showed genes from several immune system related processes, such as interferon signalling and interleukin signalling to be downregulated. Metabolism associated processes, on the other hand, were found to be upregulated. Overall, the results of this analysis mirror those of the previous part.

Comparison 5 We performed a paired differential expression analysis between the 9 individuals at TP3 and TP1 to see the changes in metabolic processes, if any, between their old cluster at TP1 and their new cluster at TP3. In this comparison, we observed a total of 660 differentially expressed genes, with the gene ontology enrichment and pathway overrepresentation results following the same pattern as that in Comparison 1. The reduction in the number of differentially expressed genes only affected the intensity of the gene expression and the z-score in the overrepresentation analysis. Due to this, the pathways results are difficult to interpret as considerable parts of many pathways have varied expression patterns between the individuals and high p-values.

Cluster membership maintainers

We also observed that a number of individuals do not change their cluster membership at all and stay within Cluster A (4 individuals) or Cluster B (8 individuals) throughout the study. The differential gene expression analysis between individuals staying in Cluster A and individuals staying in Cluster B yields the largest number of differentially expressed genes, at 2,838, compared to the previous analyses. Gene ontology enrichment analysis showed a combination of metabolic, inflammatory and tissue morphology terms that were enriched. The pathway overrepresentation showed that individuals staying in Cluster A persistently had a lower overall expression of cellular metabolism with various enzymes across lipid metabolism, SREBP signalling, mitochondria electron transport chain, TCA cycle, and glycolysis being downregulated compared to individuals in Cluster B. Immune system processes, such as components of the human complement system and interleukin signalling were found to be upregulated in Cluster A. In addition, matrix metalloproteinases MMP2, MMP7 and MMP9 were also found to be upregulated in Cluster A.

The differential expression analyses output, gene ontology enrichments output, as well as the pathway overrepresentation analysis output for the aforementioned results are provided as Additional files 4, 5 and 6 respectively.

Discussion

Lowered cellular metabolism

In the study presented in this article, we have analysed transcriptomics and proteomics expression data of the subcutaneous adipose tissue to stratify the individuals based on the expression of their genes and proteins involved in cellular metabolism, to construct metabolic profiles based on the roles of these genes and proteins in cellular metabolic flexibility. We clustered the samples of the metabolic flexibility related transcriptomics and proteomics data to generate two clusters, A and B, which showed marked difference in their whole gene expression profile. The clustering pattern generated by the ANF algorithm showed an overall lowered expression of metabolism related genes in Cluster A compared to Cluster B. Interestingly, a larger number of individuals started in Cluster B, changed their metabolic profile in response to caloric restriction and joined Cluster A, and then about half of them reverted to their original profile. This is likely related to the fact that the individuals in the study were overweight before dietary intervention, and their response to the caloric restriction is reflected in them changing clusters. However, it is interesting that several individuals maintained their new lowered metabolic gene expression beyond caloric restriction into weight maintenance, although, given the time period between the data collection points (4 weeks), it is difficult to assess the long term impact and/or changes in metabolic profile. The phenotypic/clinical measurements did not show any major difference between the two clusters, apart from the significant anthropometric measurement of sex, height, BMI, fat mass, fat free mass, waist size and hip circumference. These are explained by the clustering pattern that we observe where there is a disproportionate number of before weight loss samples in Cluster B compared to Cluster A, and vice versa after weight loss. As most of these anthropometric measurements are related, and are partially interrelated, their simultaneous significance is not surprising in light of the clustering pattern. Consequently, we also did not see any difference in the weight maintenance score (*s*-values) of the two clusters either, indicating that the impact of the metabolic profiles built on the metabolic flexibility related genes is much subtler than we had expected. However, we did see a significant difference (*p*-value = ~0.03) in the *s*-values for the nine subjects which changed their profile over the course of the dietary intervention from Cluster B (TP1) to Cluster A (TP2), when compared to the *s*-values of the eight subjects which stayed in Cluster B throughout the dietary intervention.

Increased inflammatory response

Comparing the samples of the two clusters as a whole shows that apart from the difference in cellular metabolism, expected as the clustering was based on the 291 cellular metabolism associated genes, the two clusters also show a difference in immune related processes in the adipose tissue expression data. Interestingly, the immune system processes, such as interferon signalling, toll-like receptor signalling and interleukin signalling were expressed higher in Cluster A compared to Cluster B, while the metabolic processes were being downregulated in Cluster A compared to Cluster B. The immune system response in terms of inflammation is associated with obesity [38–42], as such we find it interesting that immune processes were being upregulated in Cluster A, compared to Cluster B, especially since there is a larger proportion of TP2 samples in Cluster A (TP2 is the time point immediately after weight loss in the data). It is possible that, because of weight loss, the adipocytes in the adipose tissue shrink in size, and thus the adipose tissue samples post weight loss contain a larger proportion of immune cells as compared to adipocytes. However, this is unlikely, as the differential expression analysis between Cluster A at TP3 and Cluster B at TP3 shows a lower immune system associated gene expression in Cluster B samples compared to those of Cluster A. This result suggests that the immune system expression is not tied to the relative proportion of cells in the collected samples; otherwise, the immune system related gene expression would not have been significantly different between the two clusters at TP3. However, this is purely deductive reasoning, and a study on the histological/morphological data of the subcutaneous adipose tissue during weight loss would be required to conclusively (in)validate this perspective.

Implication in the context of metabolic flexibility

Some recent studies in mice have shown an upregulation of macrophage related immune activity in the mouse adipose tissue after weight loss [43, 44]. Indeed, a previous report from the Yoyo study data also found leukocyte integrin gene activity in relation to weight regain post weight loss using a set of extracellular matrix related genes [45]. These studies, collectively, indicate that the continued immune activity after weight loss may explain and/or contribute to the variation seen in the weight regain after weight loss. Other studies cite persistent low-grade inflammation as a cause for chronic non-communicable diseases [46–48]. It has also been shown that localised high expression of tumour necrosis factor (TNF) and interleukin-6 (IL-6) are associated with obesity induced insulin insensitivity [49]. Given that we know that metabolic inflexibility is associated with chronic illnesses such as type 2 diabetes mellitus, cardiovascular diseases and metabolic syndrome [2–4, 50], it is possible that there exists a link through the persistent low-grade inflammation. This is

certainly possible as in one of our previous studies [21], we showed how PPAR γ plays a role in regulating cellular metabolic flexibility, while it is also known that PPAR γ has connections to the immune system [51–53].

As mentioned earlier, we see an increase in immune response after weight loss and/or weight maintenance in the majority of the individuals of the data analysed. This observation complicates the interpretation as weight loss is expected to improve or restore metabolic flexibility, given the reduction of obesity. Furthermore, based on how the ANF algorithm produces the clustering, we expect to find at least the metabolic flexibility genes to be different in the two clusters, introducing a minor, yet inherent bias. In the data we analysed in this study, we did not have the time scale resolution required to decipher more concrete links between metabolic flexibility, adipose tissue metabolism and the immune response. However, our study still provides hints towards this interaction. From a strictly statistical perspective, a considerable portion of the differential expression results in this study become insignificant when adjusted for multiple correction. Although this insignificance has little impact of enrichment analyses and their results, it still shows that much work needs to be done in order to untangle these cellular interactions and their effects. As such, the interplay of these systems, at the very least in the adipose tissue, needs to be further elucidated in more targeted studies as an important step towards combatting and containing obesity and associated co-morbidities. Although we cannot comment on the technology that might be required for accurately measuring the rate of change of cellular metabolism, perhaps it might be possible to indirectly infer such a rate using cell line experiments. In addition, given that the data utilised in this article was a dietary intervention dataset, it would be interesting to analyse a dataset comparing obese and lean individuals on the same data types and parameters to observe if any combination of parameters stand out as phenotypically defining metabolic inflexibility.

Conclusions

Cellular metabolism and metabolic flexibility have been shown to be associated with obesity and the development of chronic illnesses such as type 2 diabetes mellitus and metabolic syndrome. In our study, we clustered gene expression samples from a weight loss study into two clusters, based on 291 genes associated with cellular metabolic flexibility. Our analyses showed that the majority of the individuals had their metabolism associated genes downregulated after weight loss and weight maintenance, but also had an upregulation of immune system associated genes. A higher expression of the immune system has previously been associated with the impairment of metabolism post-weight loss in mice. Our study suggests a similar pattern in the human adipose tissue, opening the way for

targeted studies to elucidate the interactions of metabolism, metabolic flexibility and the immune system in the context of chronic diseases. Furthermore, it was observed that individuals which had changed their metabolic profiles in response to caloric restriction (as reflected by the change in Cluster membership from B to A) had a significant retention of lost weight compared to individuals which had not changed their cluster membership (i.e., remained in Cluster B).

Data Availability

The expression data analysed in the study is available at gene expression omnibus (GEO) under accession ID: GSE77962. The proteomics data and clinical/phenotypic measurements are freely available from the authors upon request.

References

- [1] "Global status report on noncommunicable diseases 2014," tech. rep., World Health Organization, Geneva, Switzerland, 2014.
- [2] J. Kaur, "A comprehensive review on metabolic syndrome," *Cardiology Research and Practice*, vol. 2014, 2014.
- [3] C. Power, S. M. P. Pereira, C. Law, and M. Ki, "Obesity and risk factors for cardiovascular disease and type 2 diabetes: Investigating the role of physical activity and sedentary behaviour in mid-life in the 1958 british cohort," *Atherosclerosis*, vol. 233, no. 2, pp. 363 – 369, 2014.
- [4] L. F. Van Gaal, I. L. Mertens, and C. E. De Block, "Mechanisms linking obesity with cardiovascular disease," *Nature*, vol. 444, pp. 875–880, 2006.
- [5] S. Klein, N. F. Sheard, X. Pi-Sunyer, A. Daly, J. Wylie-Rosett, K. Kulkarni, and N. G. Clark, "Weight management through lifestyle modification for the prevention and management of type 2 diabetes: Rationale and strategies," *Diabetes Care*, vol. 27, no. 8, pp. 2067–2073, 2004.
- [6] C. J. Lavie, R. V. Milani, and H. O. Ventura, "Obesity and cardiovascular disease risk factor, paradox, and impact of weight loss," *Journal of the American College of Cardiology*, vol. 53, no. 21, pp. 1925–1932, 2009.
- [7] K. L. Campbell, K. E. Foster-Schubert, K. W. Makar, M. Kratz, D. Hagman, E. A. Schur, N. Habermann, M. Horton, C. Abbenhardt, L.-Y. Kuan, L. Xiao, J. Davison, M. Morgan, C.-Y. Wang, C. Duggan, A. McTiernan, and C. M. Ulrich, "Gene expression changes in adipose tissue with diet- and/or exercise-induced weight loss," *Cancer Prevention Research*, vol. 6, no. 3, pp. 217–231, 2013.
- [8] C. Leyvraz, C. Verdumo, M. Suter, A. Paroz, J.-M. Calmes, P. M. Marques-Vidal, and V. Giusti, "Changes in gene expression profile in human subcutaneous adipose tissue during significant weight loss," *Obesity Facts*, vol. 5, pp. 440–451, 2012.
- [9] S. H. K. Tareen, M. E. Adriaens, I. C. W. Arts, T. M. de Kok, R. G. Vink, N. J. T. Roumans, M. A. van Baak, E. C. M. Mariman, C. T. Evelo, and M. Kutmon, "Profiling cellular processes in adipose tissue during weight loss using time series gene expression," *Genes*, vol. 9, no. 11:525, 2018.
- [10] L. Hue and H. Taegtmeier, "The randle cycle revisited: a new head for an old hat," *American Journal of Physiology - Endocrinology And Metabolism*, vol. 297, no. 3, pp. E578–E591, 2009.
- [11] L. Storlien, N. D. Oakes, and D. E. Kelley, "Metabolic flexibility," *Proceedings of the Nutrition Society*, vol. 63, no. 2, pp. 363–368, 2007.
- [12] J. J. Dubé, P. M. Coen, G. DiStefano, A. C. Chacon, N. L. Helbling, M. E. Desimone, M. Stafanovic-Racic, K. C. Hames, A. A. Despines, F. G. S. Toledo, and B. H. Goodpaster, "Effects of acute lipid overload on skeletal muscle insulin resistance, metabolic flexibility, and mitochondrial performance," *American Journal of Physiology - Endocrinology And Metabolism*, vol. 307, no. 12, pp. E1117–E1124, 2014.
- [13] M. K. C. Hesselink, V. Schrauwen-Hinderling, and P. Schrauwen, "Skeletal muscle mitochondria as a target to prevent or treat type 2 diabetes mellitus," *Nat Rev Endocrinol*, vol. 12, no. 11, pp. 633–645, 2016.

-
- [14] S. N. Vallerie and K. E. Bornfeldt, "Metabolic flexibility and dysfunction in cardiovascular cells," *Arteriosclerosis, Thrombosis, and Vascular Biology*, vol. 35, no. 9, pp. e37–e42, 2015.
- [15] V. Varma, L. G. Boros, G. T. Nolen, C.-W. Chang, M. Wabitsch, R. D. Beger, and J. Kaput, "Metabolic fate of fructose in human adipocytes: a targeted 13c tracer fate association study," *Metabolomics*, vol. 11, no. 3, pp. 529–544, 2015.
- [16] G. Fantuzzi, "Adipose tissue, adipokines, and inflammation," *Journal of Allergy and Clinical Immunology*, vol. 115, no. 5, pp. 911–919, 2005.
- [17] E. Corpeleijn, W. H. M. Saris, and E. E. Blaak, "Metabolic flexibility in the development of insulin resistance and type 2 diabetes: effects of lifestyle," *Obesity Reviews*, vol. 10, no. 2, pp. 178–193, 2009.
- [18] E. Corpeleijn, M. Mensink, M. E. Kooi, P. M. H. J. Roekaerts, W. H. M. Saris, and E. E. Blaak, "Impaired skeletal muscle substrate oxidation in glucose-intolerant men improves after weight loss," *Obesity*, vol. 16, no. 5, pp. 1025–1032, 2008.
- [19] A. Astrup, "The relevance of increased fat oxidation for body-weight management: metabolic inflexibility in the predisposition to weight gain," *Obesity Reviews*, vol. 12, no. 10, pp. 859–865, 2011.
- [20] S. H. K. Tareen, M. Kutmon, M. E. Adriaens, E. C. M. Mariman, T. M. de Kok, I. C. W. Arts, and C. T. Evelo, "Exploring the cellular network of metabolic flexibility in the adipose tissue," *Genes & Nutrition*, vol. 13, no. 1, p. 17, 2018.
- [21] S. H. Tareen, M. Kutmon, I. C. Arts, T. M. de Kok, C. T. Evelo, and M. E. Adriaens, "Logical modelling reveals the pdc-pdk interaction as the regulatory switch driving metabolic flexibility at the cellular level," *Genes & Nutrition*, vol. 14, no. 1, p. 27, 2019.
- [22] R. G. Vink, N. J. T. Roumans, L. A. J. Arkenbosch, E. C. M. Mariman, and M. A. van Baak, "The effect of rate of weight loss on long-term weight regain in adults with overweight and obesity," *Obesity*, vol. 24, no. 2, pp. 321–327, 2016.
- [23] N. J. T. Roumans, R. G. Vink, F. G. Bouwman, P. Fazelzadeh, M. A. van Baak, and E. C. M. Mariman, "Weight loss-induced cellular stress in subcutaneous adipose tissue and the risk for weight regain in overweight and obese adults," *International Journal Of Obesity*, vol. 41, p. 894, 2016.
- [24] J. A. Baecke, J. Burema, and J. E. Frijters, "A short questionnaire for the measurement of habitual physical activity in epidemiological studies," *The American Journal of Clinical Nutrition*, vol. 36, no. 5, pp. 936–942, 1982.
- [25] R. G. Vink, N. J. Roumans, P. Fazelzadeh, S. H. Tareen, M. V. Boekschoten, M. A. van Baak, and E. C. Mariman, "Adipose tissue gene expression is differentially regulated with different rates of weight loss in overweight and obese humans," *International Journal of Obesity*, 12 2016.
- [26] H. Goossens Gijs, A. Bizzarri, N. Venteclef, Y. Essers, P. Cleutjens Jack, E. Konings, W. E. Jocken Johan, M. Čajlaković, V. Ribitsch, K. Clément, and E. Blaak Ellen, "Increased adipose tissue oxygen tension in obese compared with lean men is accompanied by insulin resistance, impaired adipose tissue capillarization, and inflammation," *Circulation*, vol. 124, no. 1, pp. 67–76, 2011.

-
- [27] T. Ma and A. Zhang, "Affinity network fusion and semi-supervised learning for cancer patient clustering," *Methods*, vol. 145, pp. 16–24, 2018.
- [28] B. Wang, A. M. Mezlini, F. Demir, M. Fiume, Z. Tu, M. Brudno, B. Haibe-Kains, and A. Goldenberg, "Similarity network fusion for aggregating data types on a genomic scale," *Nature Methods*, vol. 11, p. 333, 2014.
- [29] R Development Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. ISBN 3-900051-07-0.
- [30] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995.
- [31] M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, and G. K. Smyth, "limma powers differential expression analyses for RNA-sequencing and microarray studies," *Nucleic Acids Research*, vol. 43, no. 7, p. e47, 2015.
- [32] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov, "Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles," *Proceedings of the National Academy of Sciences*, vol. 102, no. 43, p. 15545, 2005.
- [33] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker, "Cytoscape: A software environment for integrated models of biomolecular interaction networks," *Genome Research*, vol. 13, no. 11, pp. 2498–2504, 2003.
- [34] G. Bindea, B. Mlecnik, H. Hackl, P. Charoentong, M. Tosolini, A. Kirilovsky, W.-H. Fridman, F. Pagès, Z. Trajanoski, and J. Galon, "Cluego: a cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks," *Bioinformatics*, vol. 25, no. 8, pp. 1091–1093, 2009.
- [35] M. Kutmon, M. P. van Iersel, A. Bohler, T. Kelder, N. Nunes, A. R. Pico, and C. T. Evelo, "Pathvisio 3: An extendable pathway analysis toolbox," *PLOS Computational Biology*, vol. 11, no. 2, p. e1004085, 2015.
- [36] M. Kutmon, A. Riutta, N. Nunes, K. Hanspers, E. L. Willighagen, A. Bohler, J. Mélius, A. Waagmeester, S. R. Sinha, R. Miller, S. L. Coort, E. Cirillo, B. Smeets, C. T. Evelo, and A. R. Pico, "Wikipathways: capturing the full diversity of pathway knowledge," *Nucleic Acids Research*, vol. 44, no. D1, pp. D488–D494, 2016.
- [37] PathVisio development team. <https://www.pathvisio.org/documentation/tutorials/>, Accessed on: 10-06-2019.
- [38] M. A. Exley, L. Hand, D. O'Shea, and L. Lynch, "Interplay between the immune system and adipose tissue in obesity," *Journal of Endocrinology*, vol. 223, no. 2, p. R41, 2014.
- [39] J. Freese, R. Klement, B. Ruiz-Núñez, S. Schwarz, and H. Lotzterich, "The sedentary (r)evolution: Have we lost our metabolic flexibility? [version 2; peer review: 2 approved, 1 approved with reservations]," *F1000Research*, vol. 6, no. 1787, 2018.
- [40] M. Mraz and M. Haluzik, "The role of adipose tissue immune cells in obesity and low-grade inflammation," *Journal of Endocrinology*, vol. 222, no. 3, p. R113, 2014.

-
- [41] E. Stolarczyk, "Adipose tissue inflammation in obesity: a metabolic or immune response?," *Current Opinion in Pharmacology*, vol. 37, pp. 35–40, 2017.
- [42] F. M. Wensveen, S. Valentić, M. Šestan, T. Turk Wensveen, and B. Polić, "Interactions between adipose tissue and the immune system in health and malnutrition," *Seminars in Immunology*, vol. 27, no. 5, pp. 322–333, 2015.
- [43] D. S. Kyung, H. R. Sung, Y. J. Kim, K. D. Kim, S. Y. Cho, J. H. Choi, Y. H. Lee, I. Y. Kim, and J. K. Seong, "Global transcriptome analysis identifies weight regain-induced activation of adaptive immune responses in white adipose tissue of mice," *International Journal Of Obesity*, vol. 42, p. 755, 2017.
- [44] B. Zamarron, T. Mergian, G. Martinez-Santibanez, and C. Lumeng, "Impact of weight loss on obese adipose tissue immune cell function (cam1p.154)," *The Journal of Immunology*, vol. 194, no. 1 Supplement, p. 48.11, 2015.
- [45] N. J. T. Roumans, R. G. Vink, P. Fazelzadeh, M. A. van Baak, and E. C. M. Mariman, "A role for leukocyte integrins and extracellular matrix remodeling of adipose tissue in the risk of weight regain after weight loss," *The American Journal of Clinical Nutrition*, vol. 105, no. 5, pp. 1054–1062, 2017.
- [46] K. de Punder and L. Pruimboom, "Stress induces endotoxemia and low-grade inflammation by increasing barrier permeability," *Frontiers in Immunology*, vol. 6, p. 223, 2015.
- [47] A. H. Miller and C. L. Raison, "The role of inflammation in depression: from evolutionary imperative to modern treatment target," *Nature Reviews Immunology*, vol. 16, p. 22, 2015.
- [48] J. Warnberg, E. Nova, J. Romeo, L. A. Moreno, M. Sjostr om, and A. Marcos, "Lifestyle-related determinants of inflammation in adolescence," *British Journal of Nutrition*, vol. 98, no. S1, pp. S116–S120, 2007.
- [49] P. A. Kern, S. Ranganathan, C. Li, L. Wood, and G. Ranganathan, "Adipose tissue tumor necrosis factor and interleukin-6 expression in human obesity and insulin resistance," *American Journal of Physiology - Endocrinology and Metabolism*, vol. 280, no. 5, pp. E745–E751, 2001.
- [50] R. L. Smith, M. R. Soeters, R. C. I. Wust, and R. H. Houtkooper, "Metabolic flexibility as an adaptation to energy resources and requirements in health and disease," *Endocrine Reviews*, vol. 39, no. 4, pp. 489–517, 2018.
- [51] A. Croasdell, P. F. Duffney, N. Kim, S. H. Lacy, P. J. Sime, and R. P. Phipps, "Ppar γ and the innate immune system mediate the resolution of inflammation," *PPAR Research*, vol. 2015, p. 20, 2015.
- [52] L. Széles, D. T or ocsik, and L. Nagy, "Ppar γ in immunity and inflammation: cell types and diseases," *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, vol. 1771, no. 8, pp. 1014–1030, 2007.
- [53] X. Zhang and H. A. Young, "PPAR and immune system – what do we know?," *International Immunopharmacology*, vol. 2, no. 8, pp. 1029–1044, 2002.

Supplementary Materials

The following supplementary material is hosted by *Nature Scientific Reports* at <https://dx.doi.org/10.1038/s41598-020-58358-z>.

Additional file 1

Gene list showing the 291 cellular metabolic flexibility related genes used for ANF clustering. (Zip, 2.8KB)

Additional file 2

Tab delimited file showing the clustering for each sample across the different k-values used in ANF clustering. (Zip, 9.5KB)

Additional file 3

Archive containing the MS Excel sheets of the statistical analyses of phenotypic/clinical measurements. (Zip, 32.6KB)

Additional file 4

Archive containing the differential expression results for the comparisons listed in Table 5.1. (Zip, 5.21MB)

Additional file 5

Archive containing the gene ontology enrichment results on the differentially expressed genes from the comparisons in Table 5.1. (Zip, 6.66MB)

Additional file 6

Archive containing the pathway overrepresentation analysis results on the differentially expressed genes from the comparisons in Table 5.1. (Zip, 82.9KB)

6

General Discussion

The research presented in this thesis shows the application of systems biology towards the analyses of complex mechanistic components of biological systems. These approaches rely on a combination of sources and types of data to model and study emergent behaviours and properties of complex biological systems. Naturally, weight loss, obesity and associated chronic illnesses provided the perfect challenge for the application of these methods, as well as for developing new pipelines to streamline the analytical procedures. The primary objective of this thesis is to explore and identify cellular processes being affected during weight loss as possible targets for future research for personalised treatment of obesity, and by extension its co-morbidities.

The dynamics of metabolic flexibility

The primary function of metabolic flexibility is to ensure the availability of nutrients for the various cells and tissues depending on their type and requirements, given the nutrient state or condition of the organism [1]. This is achieved by switching between the available energy sources for cellular metabolism. Generally, the cells will metabolise glucose via oxidative glycolysis given that glucose is readily available for energy metabolism. Only during the fasting state, when glucose levels begin dropping, cells switch over to fatty acid metabolism to retain the limited glucose resource for neurons, erythrocytes, endothelial cells and other cells that are unable to metabolise lipids [1]. There have already been several indications to the impairment of metabolic flexibility in the obese system and associated chronic illnesses, but the mechanisms of the impairment have not been thoroughly explored. Indeed, the biochemical processes of oxidative glycolysis, fatty acid oxidation, fatty acid (re)synthesis, and the tricarboxylic acid cycle themselves constitute the energy productive core of the cellular metabolic landscape. However, these processes are strongly affected in the obese system when compared to lean subjects and healthy adipose tissue baseline measurements, as can be observed in Chapter 3. Several of the essential enzymes involved in cellular metabolism are observed to have impaired gene expression in the obese system, especially the pyruvate dehydrogenase kinases (PDKs) as they showed a deviation from their expression patterns in the subcutaneous adipose tissue. Interestingly, these essential enzymes are also affected during caloric restriction and dietary interventions for weight loss to reduce obesity, as can be observed in Chapter 2, representing a central role of metabolic flexibility in the familiar environment of cellular metabolism.

In this landscape of metabolic flexibility and the cellular machinery involved in the underlying biochemical processes, some parts of the machinery do stand out compared to others for having a far-reaching effect. These parts were observed to be the negative regulation of pyruvate dehydrogenase complex

(PDC) by site-specific phosphorylation done by the pyruvate dehydrogenase kinases (PDKs), as elaborately discussed in Chapter 4. Modelling a condensed form of the cellular metabolic flexibility processes did reveal that perturbations introduced in the system propagated via this negative regulation, indicating a possible central role of this negative regulation for regulating cellular metabolic flexibility. In addition, PDKs are known to be targets of several nuclear receptors, such as the peroxisome proliferator-activated receptor (PPAR) family [2], opening them to be affected by extracellular signals, possibly indicating a mechanistic aspect of inter-organ metabolic flexibility. Furthermore, adenosine monophosphate-activated protein kinase (AMPK) also plays a regulatory role in fatty acid oxidation, understandable as AMPK is a known sensor for energy levels in the cell [1]. AMPK is also known to be involved in inflammation, indicating a cross-talk between inflammation and metabolic flexibility, at least on the molecular level.

It is known that chronic low-grade inflammation of the adipose tissue is a side effect of obesity [3, 4], and it has been shown that inflammation is reduced post weight loss [5]. Yet, Chapter 5 showed an upregulation of the expression of inflammation-associated interleukins, toll-like receptors and tumour necrosis factor in a large number of the samples from the Yoyo dietary intervention study. The majority of these samples were post weight loss and/or post weight maintenance. In addition, the majority of the samples that showed a lower expression of these inflammation-associated genes were from before the dietary intervention, indicating an interesting finding – it has been shown in some studies in mice that inflammation was observed after weight loss [6, 7]. Collectively, the studies and the findings in Chapter 5 indicate a possible cross-talk between cellular metabolic flexibility and inflammation, especially considering that the samples in Chapter 5 were clustered based on the transcriptomics and proteomics expression data for a list of gene and proteins involved with cellular metabolic flexibility.

Towards a clinical profile of metabolic flexibility

In addition to the lack of mechanistic insights in the functioning and role of metabolic flexibility in obesity and associated chronic illnesses, a set of clinical measurements for assessing metabolic flexibility and inflexibility have also been not clearly defined [8]. Chapter 5 shows two metabolic profiles based on the expression of genes and proteins involved in cellular metabolic flexibility. Yet, when these profiles are tested on the clinical measurements available in the Yoyo study, namely the anthropometric, blood plasma, and arteriovenous adipose tissue flux measurements, only a few measurements showed any significance differences between the two profiles. These significant measurements were height, body mass index (BMI), fat mass, fat free mass, waist size and hip

circumference – expected as the metabolic profiles on which the clustering was based had a disproportionate distribution of pre-dietary intervention and post-dietary intervention samples. None of the metabolites and hormones measured in the blood plasma and arteriovenous flux measurement of the adipose tissue showed any significant differences between the two profiles, nor did the two intensities of caloric restriction.

It is important to note that the metabolic profiles were generated from samples from a single tissue type – the subcutaneous adipose tissue. Metabolic flexibility is an interconnected energy and nutrient balancing system that also integrates the metabolic and signalling processes of several tissues and organs to maintain nutrient availability throughout the whole organism [1]. Indeed, a manifestation of this energy and nutrient balance is the cycling of glucose and fatty acids in the form of the Randle cycle [1]. The metabolic profiles generated by the subcutaneous adipose tissue only represent one part of the whole organism-level metabolic flexibility. The possibility of the missing data, from other tissues and organs of the subjects that underwent dietary intervention, providing improved results to find significant differences in the clinical measurements of the subjects cannot be overlooked. One such example would be data showing inter-tissue signalling and communication through hormonal and metabolites signalling cascades that can be used to analyse the interdependence and interactivity of these tissues for maintaining metabolic flexibility. In addition, given that the data utilised in this article study was a dietary intervention dataset, it would be interesting to analyse a dataset comparing obese and lean subjects on the same data types and parameters to observe if any combination of parameters stand out as phenotypically defining metabolic inflexibility. These insights show a need for detailed studies, targeting multiple tissues for multiple types of data to generate a detailed clinical profile of metabolic flexibility to study its regulation, impairment and effects in the development of obesity and associated chronic illnesses.

The synergy of multi-omics analyses

In recent decades, life sciences and medicine research have had an explosion of different subfields pushing the boundaries of science in their own niche areas. Boosted by ever improving technological innovations as well as high-throughput data generation capabilities, these fields have fragmented into focusing on different aspects of the underlying biology – commonly referred to as ‘-omes’ [9]. However, life does not care for arbitrary lines and boundaries drawn by scientists, or humans in general. One may focus on the transcriptome for finding regulatory genes, or the proteome for possible ligands of an enzyme, or the metabolome for the flux of metabolites, the fact remains that all these systems work simultaneously in the living cells, building on top of one another

in an interconnected and interdependent manner. For example, it is quite well known that gene expression does not always translate directly into protein expression and activity [10]. Thus arises the need for research and analyses that draw upon the multiple facets of available data to construct and analyse complex systems in an integrated manner.

The research presented in this thesis follow a pattern of iterative data source aggregation. It starts with the exploratory analysis of the transcriptomics data obtained from the subcutaneous adipose tissue in Chapter 2, expanding into the realm of known knowledge from databases and literature in Chapter 3, integrating known knowledge to construct theoretical models in Chapter 4, and then adding multiple sources and types of data to generate a more refined picture of metabolic flexibility and its interconnectedness with weight loss, obesity and inflammation in Chapter 5. This aggregation of data sources results in a richer and more comprehensive analysis of the complex system being studied. In essence, there is a growing need to move towards holistic instead of reductionist analyses to both understand and truly appreciate the complexity of living systems.

The scalability and analytical power of network biology

Holistic data analyses require robust frameworks and methodologies to accommodate the different constraints of each source and/or type of data being analysed. These frameworks need to be able to represent the different layers of biological data, from interactions of molecules in biochemical reactions to interactions between cells, tissues and organs. One successful application to fulfil these requirements has been that of network biology. Network biology is the application of the methods and frameworks of graph theory to curate, visualise and analyse biological systems in an integrated manner [11, 12]. Indeed, the versatility of a network representation to visualise biological entities and overlay existing knowledge streamlines the analyses and interpretation of the data, as can be seen in Chapter 2, and 3. The network structure itself represents a form of data and can be capitalised upon to find groups of interacting components or infer patterns of behaviours and phenotypes that a system is exhibiting, as can be seen in Chapter 2 and 4.

A network representation, with biological entities represented as objects (node/vertices) and their connection represented with a simple line between them (links/edges), provides both simplicity and adaptability to the framework simultaneously, allowing multi-scale and multi-level visualisation and analysis of complex biological systems, as can be seen in the works contained in this thesis. As such, networks are not only a data format or visualisation,

but a simplified and elegant model of the underlying system as well, capable of containing large sets of dynamics and behaviours and providing new information and connections that are often missed when processing experimental data.

Methodological challenges faced in this work

One drawback of holistic analyses, however, is that they require large amounts of generated and meticulously curated data. This problem is exacerbated when studying humans because of the time and cost requirements associated with large studies for multi-omics data generation. Furthermore, if the study requires invasive procedures for sample collection, the ethical constraints and requirements, although justified, still form barriers. Reducing the number of subjects involved in studies ameliorates these problems, but then the high variability found between human subjects is amplified relative to the sample size. Depending upon the tissue or organ, data from human subjects can be quite divergent [13]. This problem was observed when the transcriptomics data from the Yoyo study was analysed in Chapter 1, requiring approaches other than the standard WGCNA method [14] for gene co-expression analyses.

In addition to high variability, challenges in terms of the lack of data were also faced in two other aspects. First was missing data from the Yoyo study where several subjects were not measured at several data points, reducing the number of available samples and data points to be used in the analyses. It is possible to infer missing values in the data; however, working with a few dozen subjects also lowers the confidence of the inferred data considering the small sample size, thereby forcing a removal of subjects with missing data from the analyses. These issues warrant closer collaborations between wet-lab and dry-labs as both stand to gain from one another's expertise in generating large amounts of targeted high-quality data and the subsequent feedback of downstream analyses. Essentially, collaborations should start before, not after, data generation. Another benefit of these closer collaborations would be the targeting and collection of the large number of biological parameters of the biological system, which are often overlooked when doing wet-lab analyses, but are crucial for a detailed *in silico* representation and analysis of the complex biological systems.

Lasting thoughts

Living organisms are a multi-scale and multi-level system of densely interconnected components, that is, they operate at different time scales and between different layers of components simultaneously, in a mechanistic and almost harmonious manner. The general approach to analysing biological

systems up until very recently has been to analyse them in parts, due to technological limitations. However, with the advent of newer technologies and faster machines, it is now possible to move away from the traditional reductionist analytical approaches towards holistic analyses, such as those being offered by systems biology, aiming to understand, model and predict biological functioning, thus, embracing the complexity of biology and appreciating its beauty in its entirety.

References

- [1] L. Hue and H. Taegtmeier, "The randle cycle revisited: a new head for an old hat," *American Journal of Physiology - Endocrinology And Metabolism*, vol. 297, no. 3, pp. E578–E591, 2009.
- [2] S. Zhang, M. W. Hulver, R. P. McMillan, M. A. Cline, and E. R. Gilbert, "The pivotal role of pyruvate dehydrogenase kinases in metabolic flexibility," *Nutrition & Metabolism*, vol. 11, no. 1, p. 10, 2014.
- [3] M. A. Exley, L. Hand, D. O'Shea, and L. Lynch, "Interplay between the immune system and adipose tissue in obesity," *Journal of Endocrinology*, vol. 223, no. 2, p. R41, 2014.
- [4] F. M. Wensveen, S. Valentić, M. Šestan, T. Turk Wensveen, and B. Polić, "Interactions between adipose tissue and the immune system in health and malnutrition," *Seminars in Immunology*, vol. 27, no. 5, pp. 322–333, 2015.
- [5] V. E. Bianchi, "Weight loss is a critical factor to reduce inflammation," *Clinical Nutrition ESPEN*, vol. 28, pp. 21–35, 2018.
- [6] D. S. Kyung, H. R. Sung, Y. J. Kim, K. D. Kim, S. Y. Cho, J. H. Choi, Y. H. Lee, I. Y. Kim, and J. K. Seong, "Global transcriptome analysis identifies weight regain-induced activation of adaptive immune responses in white adipose tissue of mice," *International Journal Of Obesity*, vol. 42, p. 755, 2017.
- [7] B. Zamarron, T. Mergian, G. Martinez-Santibanez, and C. Lumeng, "Impact of weight loss on obese adipose tissue immune cell function (cam1p.154)," *The Journal of Immunology*, vol. 194, no. 1 Supplement, p. 48.11, 2015.
- [8] B. H. Goodpaster and L. M. Sparks, "Metabolic flexibility in health and disease," *Cell metabolism*, vol. 25, no. 5, pp. 1027–1036, 2017.
- [9] R. P. Horgan and L. C. Kenny, "'omic' technologies: genomics, transcriptomics, proteomics and metabolomics," *The Obstetrician & Gynaecologist*, vol. 13, no. 3, pp. 189–195, 2011.
- [10] F. Edfors, F. Danielsson, B. M. Hallström, L. Käll, E. Lundberg, F. Pontén, B. Forsström, and M. Uhlén, "Gene-specific correlation of rna and protein levels in human cells and tissues," *Molecular Systems Biology*, vol. 12, no. 10, p. 883, 2016.
- [11] A.-L. Barabási and Z. N. Oltvai, "Network biology: understanding the cell's functional organization," *Nature Reviews Genetics*, vol. 5, no. 2, pp. 101–113, 2004.
- [12] T. Ideker and R. Nussinov, "Network approaches and applications in biology," *PLOS Computational Biology*, vol. 13, no. 10, p. e1005771, 2017.
- [13] J. D. Storey, J. Madeoy, J. L. Strout, M. Wurfel, J. Ronald, and J. M. Akey, "Gene-expression variation within and among human populations," *American journal of human genetics*, vol. 80, no. 3, pp. 502–509, 2007.
- [14] P. Langfelder and S. Horvath, "WGCNA: an R package for weighted correlation network analysis," *BMC Bioinformatics*, vol. 9, no. 1, pp. 1–13, 2008.

Summary

One of the fastest growing and most persistent public health problems affecting the health of the global human population is the epidemic of overweight and obesity, with WHO estimates from 2018 showing the tripling of obesity since 1975 with no signs of any reduction in the rate. This increase is alarming, considering that obesity elevates the risk of developing chronic illnesses such as cardiovascular diseases, type 2 diabetes mellitus, strokes and even some forms of cancer, making obesity one of the major, yet easily accessible, gateways to persistent chronic illnesses. The modern increasingly sedentary lifestyle coupled with an increasingly calorie rich and unbalanced diet is the chief cause of obesity with many rich and developed countries showing the largest proportions of obesity.

The general directions to counter overweight and obesity include a calorie restricted diet and regular physical exercise; however, these are not always effective as many individuals suffer from subsequent weight regain following a diet or exercise plan. Consequently, modern research is tackling obesity on several fronts, from studies on the effects of lifestyles, in terms of different types of diets and physical exercises, to studies focusing on the molecular changes during the development of obesity itself as well as associated chronic illnesses. These efforts have resulted in many a change in our understanding of the human body, not least of which is our recognition of the adipose tissue as an active and major regulatory organ, while previously the adipose tissue was thought of as a passive energy storage tissue. Additionally, the localisation, and thus type, of the adipose tissue also defines the role of the adipose tissue and the extent of its regulatory effect.

This thesis presents an exploratory research of the molecular changes occurring in the subcutaneous (below the skin) adipose tissue during obesity and weight loss to provide a map of molecular processes changing and performing different roles in the two respective conditions. Expression data showing the expression of genes of the subcutaneous adipose tissue from a weight loss study is explored in **chapter 2** to provide a list of genes the expressions of which change over time in response to diet. The samples of the tissue in the weight loss study were acquired from the abdominal subcutaneous adipose tissue. This list of genes provides the foundation on which further analyses are performed to elucidate the changes occurring in the subcutaneous adipose tissue as weight loss progresses. Of the various changes in the expression of genes, and by extension their cellular processes, we observed that the genes involved in the regulation of cellular metabolism had their expression changing in response to the weight loss diet. A search in existing literature regarding these genes presented us with the concept of metabolic flexibility in which an organism changes the nutrients it is metabolising for energy generation in response to the availability of said nutrients. Research has shown that the impairment of metabolic flexibility is associated with obesity and

several of the chronic diseases. **Chapter 3** provides a condensed visual map of the various cellular processes and the respective genes (and protein complexes) that regulate metabolic flexibility, as well as their expression profiles when comparing the expression data of obese individuals with lean individuals.

With current wet-lab technology it is difficult to either directly observe or measure cellular metabolic flexibility in near real-time as the switch from one nutrient type to another is a rapid process. Thus, we utilised theoretical biology in terms of *in silico* modelling of cellular metabolic flexibility in **chapter 4** to analyse the system. The various models showed a prominent role for the pyruvate dehydrogenase kinases (PDKs) in regulating the switching of cellular metabolism from glucose to fatty acids and vice versa. In addition, different combinations of the regulators of the PDK enzymes were also tested, showing that the difference in regulators eventually led the cellular system to converge on the same set of results, indicating that the different combinations may be the different methods of the regulation of PDK enzymes in effect in different cells and tissues.

However, *in silico* modelling has its limitations, predominantly that models are only as good as our understanding of the system in question and the rules that underlie its functions. In order to continue exploring the role of cellular metabolic flexibility in the context of obesity and weight loss, we returned to the data from the weight loss study in **chapter 5**. In addition to gene expression data, we also utilised protein expression, blood plasma and anthropometric measurements to stratify the individuals undergoing weight loss in to groups based on the expression data of the genes and proteins involved in metabolic flexibility. Our analysis yielded two groups which showed opposing trends with one group showing a lower expression of metabolic flexibility associated genes and proteins, and the other group the opposite. Additionally, our analysis also revealed that the two groups showed opposing trends in the expression of genes and proteins associated with inflammation during weight loss, with the expression patterns of inflammation-associated genes and proteins following an inverted trend compared to metabolic flexibility-associated genes and proteins in the respective groups. Essentially, this represents an association, if not a role, of inflammation with the regulation of metabolic flexibility, and with obesity and chronic illnesses in the bigger picture.

Collectively, the research presented in this thesis has provided a detailed picture of cellular metabolic flexibility in obesity and weight loss, as well as the changes taking place in the subcutaneous adipose tissue during weight loss as a whole. The chapters have identified genes, proteins and biological processes involved with metabolic flexibility, allowing their use as a guide map for successive targeted research in the future, as we progress towards tackling the epidemic of obesity and associated chronic illnesses.

Samenvatting

Een van de snelst groeiende en meest hardnekkige volksgezondheidsproblemen die de gezondheid van de wereldbevolking beïnvloeden is de epidemie van overgewicht en obesitas. Volgens schattingen van de WHO uit 2018 is het aantal gevallen van obesitas sinds 1975 verdrievoudigd zonder enige tekenen van verlaging van deze toename. Deze toename is alarmerend, gezien het feit dat obesitas het risico op het ontwikkelen van chronische ziekten zoals hart- en vaatziekten, diabetes mellitus type 2, beroertes en zelfs sommige vormen van kanker verhoogt, waardoor obesitas een van de belangrijkste routes is dat leidt tot de ontwikkeling van chronische ziekten. De moderne, sedentaire levensstijl in combinatie met een steeds calorierijker en onevenwichtig dieet is de hoofdoorzaak van obesitas, waarbij populaties in rijke en ontwikkelde landen de grootste proporties van obesitas vertonen.

De algemene aanbevelingen om overgewicht en obesitas tegen te gaan, zijn een caloriearm dieet en regelmatige lichaamsbeweging; deze zijn echter niet altijd effectief, omdat veel mensen last hebben van gewichtstoename na een dieet of trainingsschema. Dientengevolge pakt modern onderzoek zwaarlijvigheid op verschillende fronten aan, van studies naar de effecten van levensstijlen, in termen van verschillende soorten diëten en fysieke oefeningen, tot studies die zich richten op de moleculaire veranderingen tijdens de ontwikkeling van zwaarlijvigheid zelf en bijbehorende chronische ziekten. Deze inspanningen hebben geleid tot veel veranderingen in ons begrip van het menselijk lichaam, niet in het minst van onze erkenning van het vetweefsel als een actief en belangrijk regulerend orgaan, terwijl voorheen het vetweefsel werd beschouwd als een passief energieopslagweefsel. Bovendien bepaalt de lokalisatie, en dus het type, van het vetweefsel ook de rol van het vetweefsel en de omvang van het regulerende effect ervan.

Dit proefschrift presenteert een verkennend onderzoek naar de moleculaire veranderingen die optreden in het subcutane (onderhuidse) vetweefsel tijdens obesitas en gewichtsverlies om de veranderende moleculaire processen in kaart te brengen die in de twee respectieve omstandigheden verschillende rollen vertegenwoordigen. Expressiedata die de expressie van genen van het subcutane vetweefsel uit een onderzoek naar gewichtsverlies laten zien, worden in **hoofdstuk 2** onderzocht, om op deze manier een lijst van genen te verschaffen waarvan de expressies in de loop van de tijd veranderen als reactie op een dieet. De monsters van het weefsel in het onderzoek naar gewichtsverlies werden verkregen uit het onderhuidse buikweefsel. De betreffende lijst met genen uit de expressiedata hebben de basis gevormd voor verdere analyses om de veranderingen in genexpressie op te helderen die zich voordoen in het onderhuidse vetweefsel naarmate het gewichtsverlies vordert. Van de verschillende veranderingen in de expressie van genen en hun cellulaire processen hebben we vastgesteld dat de genen die betrokken zijn bij de regulatie van cel metabolisme hun expressie veranderen als reactie op

het dieet voor gewichtsverlies. Een zoekopdracht in bestaande literatuur over deze genen presenteerde ons het concept van metabolische flexibiliteit waarin een organisme de metabolisatie van voedingsstoffen veranderd ten bate van de energieopwekking in reactie op de beschikbaarheid van de betreffende voedingsstoffen. Onderzoek heeft aangetoond dat de aantasting van metabole flexibiliteit wordt geassocieerd met obesitas en verschillende chronische ziekten. **hoofdstuk 3** biedt een beknopt visueel overzicht van de verschillende cellulaire processen en de respectieve genen (en eiwitcomplexen) die de metabolische flexibiliteit reguleren, evenals hun expressieprofielen bij het vergelijken van de expressiegegevens van zwaarlijvige individuen met magere individuen.

Met de huidige 'wet-lab-technologie' is het moeilijk om cellulaire metabole flexibiliteit zowel direct te observeren als in real-time te meten, simpelweg omdat de overstap van het ene voedingsmiddeltype naar het andere een snel proces is. Dus gebruikten we theoretisch biologie in termen van in silico modellering van cellulaire metabolische flexibiliteit in **hoofdstuk 4** om het systeem te analyseren. De verschillende modellen toonden een prominente rol voor de pyruvaatdehydrogenase kinasen (PDK's) bij het reguleren van de overschakeling van cellulair metabolisme van glucose naar vetzuren en vice versa. Bovendien hebben we verschillende combinaties van de regulatoren van de PDK-enzymen getest, hieruit bleek dat het verschil in regulatoren ons uiteindelijk leidde naar dezelfde set resultaten, wat aangeeft dat de verschillende combinaties effectief een voorstelling zijn van de verschillende methoden van regulering van PDK-enzymen in verschillende cellen en weefsels.

In silico modellering heeft echter zijn beperkingen, met name dat modellen slechts zo goed zijn als ons begrip van het systeem in kwestie en de regels die ten grondslag liggen aan zijn functies. Om de rol te blijven verkennen van cellulaire metabole flexibiliteit in de context van obesitas en gewichtsverlies, keerde wij terug naar de gegevens van het gewichtsverliesonderzoek in **hoofdstuk 5**. Naast genexpressiegegevens, gebruikten we ook eiwitexpressie, bloedplasma en antropometrische metingen om de individuen die gewichtsverlies ondergaan te stratificeren in groepen op basis van de expressiegegevens van de genen en eiwitten die betrokken zijn bij metabole flexibiliteit. Onze analyse leverde twee groepen op die tegengestelde trends vertoonde, waarbij één groep een lagere uitdrukking vertoonde van metabole flexibiliteit geassocieerde genen en eiwitten, en de andere groep het tegenovergestelde. Bovendien bleek uit onze analyse dat de twee groepen tegengestelde trends vertoonden in de expressie van genen en eiwitten die betrokken zijn bij ontstekingsprocessen gedurende gewichtsverlies, met de expressiepatronen van ontsteking gerelateerde genen en eiwitten volgens een omgekeerde trend vergeleken met metabole flexibiliteit geassocieerde genen en eiwitten in de respectieve groepen. In wezen vertegenwoordigt dit een associatie, zo niet een rol, van ontstekingsprocessen

in de regulatie van metabole flexibiliteit alsmede met obesitas en chronische ziekten in het grotere geheel.

Gezamenlijk heeft het onderzoek in dit proefschrift een gedetailleerd beeld opgeleverd van cellulaire metabole flexibiliteit bij obesitas én gewichtsverlies, evenals de veranderingen die plaatsvinden in het onderhuidse vetweefsel tijdens het proces van gewichtsverlies als geheel. De hoofdstukken hebben genen, eiwitten en biologisch geïdentificeerd processen in kaart gebracht die te maken hebben met metabole flexibiliteit, welke als leidraad kunnen worden gebruikt voor vorderingen in toekomstig onderzoek in de aanpak van de epidemie van obesitas en bijbehorende chronische ziekten.

خلاصہ

ایک غذائی جزو سے دوسرے جزو میں تبدیلی کی کیمیائی رد عمل کے تیز ہونے کی وجہ سے وقتِ حاضر کی تجربہ گاہ کی ٹیکنالوجی سے خلیات کی تھوٹی چمک کا بیک وقت مشاہدہ کرنا یا حساب لگانا مشکل ہے۔ چنانچہ ہم نے باب ۴ میں نظریاتی حیاتیات کے ذریعے خلیاتی تحول کی کمپیوٹیشنل ماڈلنگ کی۔ مختلف ماڈلوں میں خلیاتی تحول کا گلوکوز سے فیٹی ایسڈ اور واپس بدلنے میں پائروویٹ ڈیہائیڈروجنیز کا نینز (پی ڈی کے) کا نمایاں کردار واضح ہوا۔ اضافی طور پر پی ڈی کے انزائم کے اپنے ضابطے میں ملوث اشیاء کے مختلف مجموعوں کو بھی آزمایا گیا جس سے ظاہر ہوا کہ مختلف مجموعے خلیاتی نظام کو یکساں نتیجے پر پہنچاتے ہیں، اشارتاً یہ کہ یہ مجموعے مختلف خلیات اور مختلف بافت میں اثر انداز ہیں۔

بہر صورت کمپیوٹر ماڈلنگ کی اپنی حدود ہیں غالباً یہ کہ ماڈل صرف اتنے ہی ٹھیک ہوتے ہیں جتنے زیرِ غور نظام کی حدود اور اُنکے متعلق ہماری اپنی سمجھ۔ چنانچہ موٹاپے اور وزن گھٹانے میں خلیات کی تھوٹی چمک کے کردار کی تحقیق جاری رکھنے کے لیے باب ۵ میں وزن گھٹنے کے مطالعہ میں دیے گئے معطیات کو رجوع کیا۔ تعبیرِ وراثہ کے ساتھ ہم نے لحمیات کی تعبیر، خون مایہ اور بشری پیمائش کا استعمال کرتے ہوئے وزن گھٹاتے افراد کو تھوٹی چمک میں ملوث وراثے اور لحمیات کی تعبیر کے نقشے کی بنا پر گروہوں میں تقسیم کیا۔ ہمارے تجزیے کے مطابق دو گروہ سامنے آئے۔ ایک گروہ میں تھوٹی چمک میں ملوث وراثوں اور لحمیات کی تعبیر میں کمی کا رجحان تھا جبکہ دوسرا گروہ اسکے برخلاف پایا گیا۔ اضافی طور پر ہمارے تجزیے میں سوزش سے متعلق وراثوں اور لحمیات کے رجحان میں بھی دونوں گروہوں میں تضاد پایا گیا جس میں سوزش کے وراثے اور لحمیات کے رجحان تھوٹی چمک کے وراثوں اور لحمیات کے رجحان کی مخالفت کرتے پائے گئے۔ بنیادی طور پر یہ تجزیے سوزش اور تھوٹی چمک کے درمیان کم از کم موٹاپے اور دائمی امراض کے سیاق میں وابستگی کی طرف اشارہ کرتے ہیں۔

مجموعی طور پر اس مقالے میں پیش کی گئی تحقیق موٹاپے اور وزن کے گھٹنے میں خلیاتی تحول کی چمک اور زیرِ جلد چربی بافت میں ہونے والی تبدیلیوں کے بارے میں ایک مفصل تصویر کھینچتی ہے۔ دیے گئے ابواب میں تھوٹی چمک میں ملوث وراثوں، لحمیات اور خلیاتی طریق کی نشاندہی کی گئی ہے تاکہ ان کا موٹاپے اور دائمی امراض پر قابو پانے کی مستقبل کی تحقیقات میں استعمال کیا جاسکے۔

دنیا کی کل انسانی آبادی کی صحت کو پیش مستقل اور تیزی سے پھیلنے والے مسائل میں سے ایک مسئلہ وزن کے عدم توازن اور موٹاپے کا ہے۔ ڈبلیو ایچ او کے اندازے کے مطابق ۱۹۷۵ سے ۲۰۱۸ کے مابین عالمی موٹاپے میں تین گنا اضافہ ہوا ہے اور اس میں کمی کا کوئی رجحان نظر نہیں آتا۔ یہ اضافہ انتہائی پریشان کن ہے کیونکہ موٹاپے کی موجودگی کی دائمی امراض کے خطرے کو بڑھاتی ہے جیسا کہ امراضِ قلب، ذیابیطیس، فالج، یہاں تک کہ سرطان کی کئی اقسام۔ اس وجہ سے موٹاپا دائمی امراض تک کا اہم مگر آسان راستہ ہے۔ دور حاضر کا آرام طلب طرز زندگی بچ کیوری سے لبریز غیر متوازن غذا موٹاپے کی سب سے اہم وجہ ہے، لہذا امیر و ترقی یافتہ ممالک میں موٹاپا بڑھتے ہوئے تناسب سے دیکھنے کو ملتا ہے۔

عموماً وزن کے عدم توازن اور موٹاپے پر قابو پانے کے لیے غذا میں کیوری کی کمی اور ورزش کی ہدایت کی جاتی ہے البتہ یہ طریقے ہمیشہ اثر انداز نہیں ہوتے اور کثیر تعداد میں افراد کا وزن مقررہ غذا اور ورزش کے ختم ہونے کے بعد دوبارہ بڑھ جاتا ہے۔ نتیجتاً دور حاضر میں موٹاپے کو قابو کرنے کے لیے کئی پہلوؤں پر تحقیقات کی جا رہی ہے جیسے کہ طرز زندگی پہ اثرات پر مبنی مختلف غذاؤں اور باقاعدہ ورزش پر مطالعہ سے لے کر موٹاپے کے دوران ہونے والی سالماتی تبدیلیاں اور ان کا دائمی امراض سے تعلق پر مطالعہ وغیرہ۔ ان کوششوں سے انسانی جسم کے متعلق ہماری سوچ میں کافی تبدیلی آئی ہے، جس میں خصوصی طور پر چربی بلی بافت کو ایک غیر فعال توانائی کے ذخیرے کی بجائے فعال اور اہم انضباطی عضو کے طور پر جاننا ہے۔ علاوہ ازیں چربی بلی بافت کی مقام بندی پر مبنی اقسام اس کے کردار اور انضباطی اثرات کا تعین بھی کرتے ہیں۔

یہ مقالہ موٹاپے اور وزن گھٹانے کے دوران زیر جلد چربی بلی بافت میں ہونے والی سالماتی تبدیلیوں کا ایک تحقیقی مطالعہ پیش کرتا ہے تاکہ ان دونوں صورتوں میں ہونے والے سالماتی طریق میں تبدیلیوں اور کارکردگی کا نقشہ فراہم کیا جاسکے۔ باب ۲ میں زیر جلد چربی بلی بافت میں تعبیر وراثہ کی تحقیق پیش کی گئی ہے تاکہ ان وراثوں کی فہرست فراہم کی جائے جن کی تعبیر میں وزن کے گھٹنے کے دوران تبدیلیاں ہوئی ہوں۔ وزن کے گھٹنے کے مطالعے میں بافت کے نمونے ٹکمی زیر جلد چربی بلی بافت سے حاصل کیے گئے تھے۔ وراثوں کی یہ فہرست مزید تجزیوں کی بنیاد فراہم کرتی ہے تاکہ وزن گھٹنے کے نتیجے میں زیر جلد چربی بلی بافت میں ہونے والی تبدیلیوں کو واضح کیا جاسکے۔ تعبیر وراثہ اور اس سے منسلک خلیاتی طریق میں ہم نے یہ مشاہدہ کیا کہ وزن گھٹنے کے لیے پرہیزی غذا کے نتیجے میں خلیاتی تحول کے ضابطے میں ملوث وراثوں کی تعبیر تبدیل ہو رہی تھی۔ موجودہ مقالوں میں ان وراثوں سے متعلق تفتیش نے ہمیں تھوٹی پلک کا تصور پیش کیا جس میں جاندار اپنے تحول کے غذائی اجزاء ان اجزاء کی دستیابی کی بنیاد پر بدلتا ہے۔ تحقیق نے یہ ظاہر کیا ہے کہ اس تھوٹی پلک میں خرابی موٹاپے اور کئی دائمی امراض سے منسلک ہے۔ تھوٹی پلک کے خلیاتی طریق اور متعلقہ وراثے (بچ لحمیات) کا مرکز نقشہ باب ۳ میں پیش کیا گیا ہے، جس کے ساتھ موٹے اور لاغر افراد کے وراثوں کی تعبیر کا موازنہ بھی موجود ہے۔

Valorisation

Metabolic foundations of obesity and weight loss

Obesity is at epidemic levels, increasing worldwide in both developed and developing countries. Not only does obesity cause long-term physical stress on the body leading to joint pain and arthritis, it is also associated with the development of chronic illnesses such as type 2 diabetes mellitus, several cardiovascular diseases such as atherosclerosis, myocardial infarctions, arrhythmias and sudden cardiac arrests, and several cancers of organs and tissues associated with the gastrointestinal tract. Although weight loss via diet and exercise is one of the most successful methods of controlling obesity, many individuals regain about half of their lost weight shortly after the diet. In addition, our knowledge of the changes occurring inside our cells and organs is limited by both technology and the invasiveness of sampling procedures.

Due to these limitations, our understanding of the physiological changes in the human body had been limited for quite some time. An excellent example of this is the change in our understanding and perception of the adipose tissue over the past two decades. Previously, the adipose tissue was considered only as an energy storage organ for storing fats. As such, even though it was a major contributor to obesity, its role was considered a passive one. However, with advancement in technology, we have discovered the active hormonal and regulatory role that the adipose tissue plays in obesity, weight loss, and energy metabolism. The study presented in this thesis provides a glimpse of those changes in the adipose tissue in the case of weight loss in obese subjects. The study elucidates the role of adipocyte cellular metabolism in managing different nutrients for producing energy and provides a contrast of how the molecular processes differ between obesity and weight loss. These differences paint a picture of how obesity may arise on the molecular level and how cellular metabolism may affect weight loss, allowing future studies of a more targeted and personalised design, both in diet and exercise plans as well as drug development, to counter the effects of obesity and induce efficient weight loss. Additionally, the study presented here also pushes for a metabolism-based understanding of obesity and weight loss, allowing both researchers and the larger society to move beyond calorie counting when it comes to weight loss, and, in the future, personalise their own weight loss plans based on their understanding of their own metabolism.

Tools and resources for extended research

One of the problems facing the research community, and arguably the research domain in general, is the difficulty in replicating existing research. Many a times, produced data and reported results are difficult to use or replicate due to technological limitations, sampling biases and a general lack

of details of the procedure. To address these issues, the FAIR principles have been developed, promoting research and data to be Findable, Accessible, Interoperable and Reusable. Although not mandatory yet, these principles provide an abstract framework for research to be both available and useful to the wider research community, even across research domains. In compliance with these principles, the study presented in this thesis has been published with extensive supplementary materials. In addition, all computational pipelines and software used have been detailed with their versions and libraries in the respective research articles to support reproducibility, and any new or modified programming scripts have been provided as supplementary material.

Furthermore, the computational pipelines and analysis procedures utilised in this thesis have been constructed in a generalised manner. That is, the procedures can be used on similar data types in different biological settings to answer different research questions. Such approaches allow for the reusability and iterative improvement of the methodology in tandem with pushing the boundaries of science.

Marching towards virtual humans

Research is becoming increasingly costly and time consuming as we proceed with pushing the boundaries of knowledge and science. This produces a drain on the economy, as research cannot always promise a product or profit. This drain is especially true for medical sciences where a drug or treatment can take upwards of 15 years to develop, study, and trial for safety. However, medical sciences have also benefitted magnificently with the advancement of computers and digital technology. We have digitised our data to store millions of samples of thousands of studies, allowing unparalleled access and reusability. We are still tied to costly procedures, as we still require wet-lab analyses and experimentation to find medical solutions conclusively. To remedy this, computational modelling, simulation and analyses are taking precedence to cut down on time and cost. This thesis contains an application of this approach where a model of cellular metabolism was used due to technological limitations and a lack of available data. Such models are easy to manipulate and adapt for a variety of conditions, allowing researchers to simulate their experiments or procedures before spending limited and crucial economic resources. With advancement in computer engineering providing efficient and economical computational resources, the fields of theoretical and systems biology are steadily marching towards virtualising human systems and data to construct detailed simulations of humans, from molecular processes to personalities and behaviours. Such a technologies hold great potential for targeted and personalised medication for a fraction of the costs associated with current research and development procedures.

Acknowledgements



Where to even begin? The PhD journey has definitely been quite an undertaking and, as with all such endeavours, one does not accomplish them standing alone. Here I will try my level best to acknowledge all the people, the effort and support of whom have eventually lead me to this point.

Starting with the person without the contribution of whom I would not be writing this thesis at all, Tina. The simple truth is, without your support this PhD would be stuck in limbo. You went beyond your role as the daily-/co-supervisor in my PhD project. You were both a mentor and a friend, patiently listening to me regardless of whether I was excitedly blazing through my ideas or when I was being crushed under the stresses of research. Chasing down senior academics and getting through academic bureaucracy (read: procedures), you really helped out at many key junctures in the project. You were always there with a smile, ever ready to find solutions, and you did all of that while juggling supervision, teaching, student projects, and a family! Not to mention that you make some delicious chocolate mousse. You deserve your assistant professorship, and I wish the very best for you and your family.

Michiel, unofficially my *second* daily supervisor, your contribution, specifically with your expertise in the statistical side of things, was something that really brought out the flavour in the rather dry field of computational analyses. Suffice to say, the project would have gone in a very different direction had you not been there to keep me on track and focusing on things that matter. That, and your uncanny ability to immediately find relevant research articles! I wish you the best with your assistant professorship as well.

Chris, Ilja and Theo, my professors and supervisors, I really appreciate your guidance during my PhD. You helped me understand the intricacies of academic research, especially at the higher levels of the academic ladder. All three of you have collectively helped me somewhat decide the directions I want to take in my post PhD research, both in terms of things to pursue and things to avoid. Chris, frankly it seemed like you were always busy with conferences and meetings, yet somehow you did show up whenever we really needed you. I do fondly remember the occasional lunch meetings we had, mostly because we would go off a tangent and spend the whole meeting discussing interesting research topics instead of my PhD project, but I appreciate them nonetheless because they kept my curiosity and thirst for knowledge sated. Ilja, I am very much thankful for your help towards the end of my PhD, specifically with the formalities and official work required for completing the project. Theo, although toxicogenomics kept

you busy and away from our regular meetings, I do appreciate the constructive, and occasionally very direct, criticism you provided over the years.

Edwin and Marleen, thank you for your help with the PhD project – I mean, it would certainly be difficult to do all this research without the data! I appreciate the help, biological insights and interpretations that both of you provided while complying to the best of your abilities with our occasionally eleventh hour meeting schedules.

Ah, Pierre, my trusted friend, I believe the list of things we haven't talked about is definitely shorter than of the ones we have. Whether it was those lazy afternoons spent sipping Limonata at Rivazza or munching on baguette deluxe at Banditos, we always had the liveliest of discussions, even when discounting all the trolling I put you through. And then the occasional evenings spent sipping chai while playing cards or listening to your questionable taste in South Asian music. Suffice to say, there's a reason I chose you to be one of the paranympths, and I feel honoured that you accepted it. There's just one tiny thing I would like to add though, Seahawks...are overrated. 49ers FTW!

Rachel, my second paranympth, and pretty much a combo deal with Pierre. Credit where it's due, thank you so much for translating the summary to Dutch! I really have enjoyed the time we have spent at the university. The insights you have had with different socio-economic problems were always a delight to listen to and discuss. I wish you the best with your ongoing project and your secret ideas. I am honoured to have you as my paranympth.

Isma, thank you for hearing out my almost daily rants – I am truly sorry for those. I appreciate your well-thought suggestions and takes on many different things, and I am also thankful for your suggested edits for the Urdu summary. I wish you the best with finishing your thesis and your postdoctoral research! Maryam and Chaitra, it was a pleasure to have spent time with you at MaCSBio, and I hope your research projects go smoothly and splendidly. Shauna, I wish you the best with your postdoctoral research. Bob, it was a pleasure to have shared the same office space as you. From the various organisms you "accidentally" grew on the food in your desk drawers to your hilarious thought experiments over lunch, you provided the much-needed break to keep us PhD researchers sane. I still think you should patent the "shakey-shakey" before someone steals the idea.

Bart, Balázs, David and Charlie, thanks for the near-daily office chats and research discussions. Don't let the the PhD projects drag you away from keeping the office lively and friendly. Claudia, suffice to say, MaCSBio would be going in circles without you. You have definitely kept the place running, whether it was helping with forms, or with equipment and events. Thank you for all your help with navigating the formalities and requirements of the PhD. Peter, I will miss our chats about hardware and software troubles, and my

year-long rant about not having access to an Adobe licence. Dries, it was nice having you around, and I really enjoyed the Christmas and New Year's parties. Hope you have many more! Mirella, it was always a pleasure to walk into the office and see a friendly face. Ilona, although you only dropped by occasionally, it was enjoyable to talk with you. Elisa and Wael, I am glad to have met you and wish the best for the whole family – say hi to Mila and Nadine for me!

Hilmi, although you were at MaCSBio for a short time, I enjoyed it. Whether it was talking about research and development, or politics and economics, it was always a delight to have your unique blend of insight, humour and trademarked apathy. Keep challenging yourself with your photography brother! Jihad and Faizan, the weekly trips to the masjid were a much-needed decompression and I wish the best for both of you. Hans, thanks so much for being an awesome friend and for helping me with all the Gemeente letters.

Mum and Dad, nothing I say will ever come close to what both of you have done and what each of you has sacrificed for me to be at this point in my life. Any thanks I give will always fall short, but I will nonetheless say that I am sincerely grateful from the bottom of my heart to both of you! Speaking of family, I would also like to thank my sister-in-law Raabya for taking a look and suggesting improvements to the Urdu summary in this thesis.

Finally, I will take this opportunity to also thank two specific people from my past. My career choice, my passion for research, my insatiable curiosity of the world around me, and my obsession with finding what makes the world tick can be traced back to these two individuals – thank you Dr Jamil Ahmad, my MS supervisor, for nurturing my passion for research, and Mr Khalid Khan, my highschool teacher, for sparking my interest in biology in the first place.

Curriculum Vitae

Samar Hayat Khan Tareen was born on the 14th of August, 1989 in Abu Dhabi, United Arab Emirates. Around the age of 6, his family moved back to Pakistan, to the historic frontier town, regional trade hub, and garrison city of Peshawar. From a very young age, Samar had exposure to different forms of electronic and printed media, culminating in a voracious appetite for historical and scientific documentaries. He completed his schooling in Peshawar in the pre-medical group at the Faculty of Science at Peshawar Model School and Degree College in 2007.

By this time, he had already established an active interest in biological sciences, as well as a hobby of tinkering with computers. Needless to say, bioinformatics appealed to him as the way forward, and he opted for a bachelor's degree in bioinformatics in 2007 at COMSATS University (then known as COMSATS Institute of Information Technology; CIIT) in Islamabad, Pakistan. During his bachelor's degree, he did an internship at the National Institute of Health in Islamabad, working with microbial contamination of underground water sources across Pakistan.

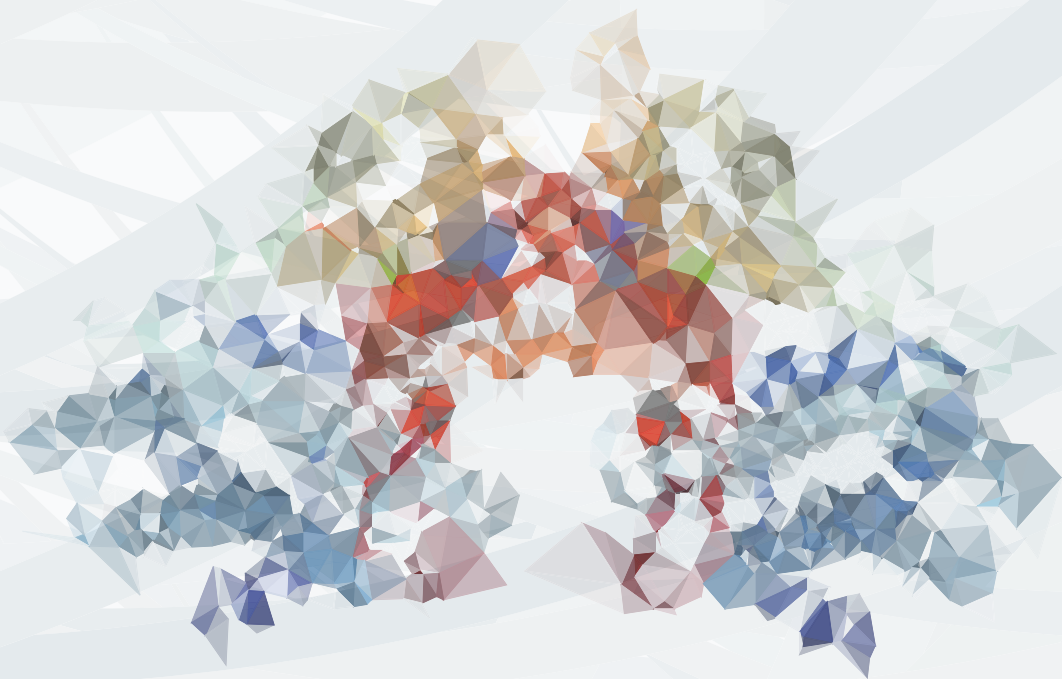
After attaining his bachelor's degree in 2011, Samar opted for a gap year but found himself back in academia as a master's student in 2012 at the Research Centre for Modelling and Simulation, National University of Sciences and Technology (NUST) in Islamabad, Pakistan. He worked with various network based modelling frameworks and their application in dynamic biological systems at the molecular level. After completing his master's degree in 2014, he was a visiting faculty member at the Shifa Tameer-e-Millat University in Islamabad for a few months before moving to Maastricht, the Netherlands, in early 2015 to pursue a PhD in systems biology. Here, he worked in the domain of obesity and chronic illnesses, focusing on the molecular changes in the adipose tissue during weight loss, narrowing down to the regulation of cellular metabolism during weight loss and its impact on other cellular processes via cross-talk.

As of November 2019, Samar has moved to Cambridgeshire, the United Kingdom, and is working at the Babraham Institute as a postdoctoral bioinformatician and systems biologist in immunology, specifically characterising regulatory T-cells and their roles in various tissues.

List of Publications

- **S.H.K. Tareen**, M. Kutmon, T.M. de Kok, E.C.M. Mariman, M.A. van Baak, C.T. Evelo, M.E. Adriaens and I.C.W. Arts, *Stratifying cellular metabolism during weight loss: an interplay of metabolism, metabolic flexibility and inflammation*
Nature Scientific Reports, vol. 10, p. 1651, 2020.
doi: 10.1038/s41598-020-58358-z
- **S.H.K. Tareen**, M. Kutmon, I.C.W. Arts, T.M. de Kok, C.T. Evelo, and M.E. Adriaens, *Logical modelling reveals the PDC-PDK interaction as the regulatory switch driving metabolic flexibility at the cellular level*
Genes & Nutrition, vol. 14, no. 27, 2019.
doi: 10.1186/s12263-019-0647-5
- A. Siddiq, E. Cirillo, **S.H.K. Tareen**, A. Ali, M. Kutmon, L.M.T. Eijssen, J. Ahmad, C.T. Evelo, and S.L. Coort, *Biological pathways leading from ANGPTL8 to diabetes mellitus – A co-expression network based analysis*
Frontiers in Physiology, vol. 9, p. 1841, 2018.
doi: 10.3389/fphys.2018.01841
- **S.H.K. Tareen**, M.E. Adriaens, I.C.W. Arts, T.M. de Kok, R.G. Vink, N.J.T. Roumans, M.A. van Baak, E.C.M. Mariman, C.T. Evelo, and M. Kutmon, *Profiling cellular processes in adipose tissue during weight loss using time series gene expression*
Genes, vol. 9, no. 11, p. 525, 2018.
doi: 10.3390/genes9110525
- **S.H.K. Tareen**, M. Kutmon, M.E. Adriaens, E.C.M. Mariman, T.M. de Kok, I.C.W. Arts, and C.T. Evelo, *Exploring the cellular network of metabolic flexibility in the adipose tissue*
Genes & Nutrition, vol. 13, no. 17, 2018.
doi: 10.1186/s12263-018-0609-3
- A. Siddiq, E. Cirillo, **S.H.K. Tareen**, A. Ali, M. Kutmon, L.M.T. Eijssen, J. Ahmad, C.T. Evelo, and S.L. Coort, *Visualizing the regulatory role of angiopoietin-like protein 8 (ANGPTL8) in glucose and lipid metabolic pathways*
Genomics, vol. 109, no. 5-6, pp. 408–418, 2017.
doi: 10.1016/j.ygeno.2017.06.006
- R.G. Vink, N.J.T. Roumans, P. Fazlzadeh, **S.H.K. Tareen**, M.V. Boekschoten, M.A. van Baak, and E.C.M. Mariman, *Adipose tissue gene expression is differentially regulated with different rates of weight loss in overweight and obese humans*
International Journal of Obesity, vol. 41, no. 2, pp. 309–316, 2016.
doi: 10.1038/ijo.2016.201
- S. Khalid, R. Hanif, **S.H.K. Tareen**, A. Siddiq, Z. Bibi, and J. Ahmad, *Formal modeling and analysis of ER- α associated biological regulatory network in breast cancer*
PeerJ, vol. 4, p. e2542, 2016.
doi: 10.7717/peerj.2542
- Z.-u.-R. Farooqi, **S.H.K. Tareen**, J. Ahmed, and N.-u.-S.S. Zaidi, *Logical analysis of regulation of Interleukin-12 expression pathway regulation during HCV infection*
Protein and Peptide Letters, vol. 23, no. 6, pp. 581–589, 2016.
doi: 10.2174/0929866523666160419150136

-
- Z. Bibi, J. Ahmad, A. Ali, A. Siddiqua, S. Shahzad, **S.H.K. Tareen**, H.A. Janjua, and S. Khusro, *On the modeling and analysis of the biological regulatory network of NF-kb activation in HIV-1 infection*
Complex Adaptive Systems Modeling, vol. 4, no. 1, 2016.
doi: 10.1186/s40294-015-0013-4
 - **S.H.K. Tareen**, J. Ahmad, and O. Roux, *Parametric linear hybrid automata for complex environmental systems modeling*
Frontiers in Environmental Science, vol. 3, p. 47, 2015.
doi: 10.3389/fenvs.2015.00047
 - B. Aslam, J. Ahmad, A. Ali, R.Z. Paracha, **S.H.K. Tareen**, S. Khusro, T. Ahmad, S.A. Muhammad, U. Niazi, and V. Azevedo, *Structural modeling and analysis of dengue-mediated inhibition of interferon signaling pathway*
Genetics and Molecular Research, vol. 14, no. 2, pp. 4215–4237, 2015.
doi: 10.4238/2015.April.28.4
 - A. Obaid, J. Ahmad, A. Naz, F.M. Awan, R.Z. Paracha, **S.H.K. Tareen**, S. Anjum, A. Raza, J. Baumbach, and A. Ali, *Modeling and analysis of innate immune responses induced by the host cells against hepatitis C virus infection*
Integrative Biology, vol. 7, no. 5, pp. 544–559, 2015.
doi: 10.1039/c4ib00285g
 - **S.H.K. Tareen** and J. Ahmad, *Modelling and analysis of the feeding regimen induced entrainment of hepatocyte circadian oscillators using petri nets*
PloS ONE, vol. 10, no. 3, p. e0117519, 2015.
doi: 10.1371/journal.pone.0117519
 - B. Aslam, J. Ahmad, A. Ali, R.Z. Paracha, **S.H.K. Tareen**, U. Niazi, and T. Saeed, *On the modelling and analysis of the regulatory network of dengue virus pathogenesis and clearance*
Computational biology and chemistry, vol. 53, no. B, pp. 277–291, 2014.
doi: 10.1016/j.combiolchem.2014.10.003
 - R.Z. Paracha, J. Ahmad, A. Ali, R. Hussain, U. Niazi, **S.H.K. Tareen**, and B. Aslam, *Formal modelling of Toll Like Receptor 4 and JAK/STAT signalling pathways: Insight into the roles of SOCS-1, Interferon- β and proinflammatory cytokines in Sepsis*
PloS ONE, vol. 9, no. 9, p. e108466, 2014.
doi: 10.1371/journal.pone.0108466



Crystal structure of PDK4 and ligands retrieved from NGL Viewer (AS Rose et al. 2018 NGL viewer: web-based molecular graphics for large complexes. *Bioinformatics* doi: 10.1093/bioinformatics/bty419), and RCSB PDB. Structure converted to low-poly graphics using Polytomizator (Created by Stone Tao; <https://stonet2000.github.io/Polytomizator/>).

