Modelling the dynamics in time-series of metabolomic and transcriptomic data

Citation for published version (APA):

O'Donovan, S. (2020). *Modelling the dynamics in time-series of metabolomic and transcriptomic data: insulin resistance and liver toxicity*. [Doctoral Thesis, Maastricht University]. ProefschriftMaken. <https://doi.org/10.26481/dis.20200928so>

Document status and date:

Published: 01/01/2020

DOI:

[10.26481/dis.20200928so](https://doi.org/10.26481/dis.20200928so)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Modelling the dynamics in time series of metabolomic and transcriptomic data

Insulin resistance & liver toxicity

Shauna O'Donovan

Doctoral thesis

**MODELLING THE DYNAMICS IN
TIME-SERIES OF METABOLOMIC AND
TRANSCRIPTOMIC DATA**

INSULIN RESISTANCE AND LIVER TOXICITY

Shauna Danielle O'Donovan

Maastricht Centre for Systems Biology (MaCSBio),
Maastricht University

2020

© Shauna Danielle O'Donovan

Maastricht Centre for Systems Biology (MaCSBio),
Maastricht University, Maastricht 2020.

For articles published or accepted for publication, the copyright has been transferred to the respective publisher. 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	Shauna O'Donovan
Production	Proefschriftmaken.nl
ISBN	978-94-6380-938-2

MODELLING THE DYNAMICS IN TIME-SERIES OF METABOLOMIC AND TRANSCRIPTOMIC DATA

INSULIN RESISTANCE AND LIVER TOXICITY

Dissertation

To obtain the degree of Doctor at Maastricht University,
on the authority of the Rector Magnificus, Prof. Dr. R.M. Letschert,
in accordance with the decision of the Board of Deans,
to be defended in public
on Monday 28th of September 2020, at 14.00 hours

by

Shauna Danielle O'Donovan

Promotors

Prof.Dr. Theo M.C.M de Kok
Prof.Dr.Ir. Ralf L.M. Peeters
Prof.Dr.Ir. Natal A.W. van Riel (Eindhoven University of Technology)
Prof.Dr.Ir Ilja C.W. Arts

Assessment Committee

Prof.Dr. Hubert J.M. Smeets (chairperson)
Prof.Dr. Patrick A.J. Schrauwen
Dr. Maria Suarez Diez (Wageningen University and Research)
Prof.Dr. Mathieu Vinken (Vrije Universiteit Brussel, Belgium)
Dr. Ronald L. Westra

The work contained in this thesis was supported by
a grant from the Dutch Province of Limburg

Contents

1	Introduction	1
2	Improved quantification of muscle insulin sensitivity using oral glucose tolerance test data: the MISI Calculator	21
3	A computational model of postprandial adipose tissue lipid metabolism derived using human arteriovenous stable isotope tracer data	45
4	Use of deep learning methods to translate drug-induced gene expression changes from rat to human primary hepatocytes exposed in vitro and in vivo	81
5	Use of domain adaptation to predict human in vivo drug-induced gene expression changes using rat in vitro and in vivo gene expression data	121
6	General Discussion	149
	Valorisation	167
	Summary	175
	Acknowledgements	181
	Publications	185
	About the author	187

1

Introduction

The human organism can be thought of as a system of systems, organised hierarchically and biologically from the genome through the transcriptome, proteome, to the metabolome at the molecular level. The systems is also organised spatially, from individual cells, which combine to form the diverse network of tissues and organs that make up the human body [1-4] (Figure 1.1). Each layer of this system is dynamically embedded in the others via a complex network of regulatory and signalling mechanisms [5], from whole-body inter-tissue signalling by hormones and cytokines transported via the blood to the regulation of gene expression by transcription factors modulating intra-cellular protein levels. This multi-scale regulatory system is in almost constant flux, adapting behaviours across several spatiotemporal domains in response to environmental stimuli in an effort to maintain homeostasis.

Advancements in high-throughput omics technologies now permits us to identify and quantify a vast array of biological entities from each level of the system (i.e. RNA transcripts, proteins, and metabolites). In the last decade, the use of omics technologies, particularly transcriptomic platforms such as microarrays and RNASeq, has become ubiquitous across a range of biological disciplines, providing genome-wide insight as to how the system changes under various physiological and pathological conditions. However, processes, such as gene expression, are inherently dynamic. For example, when cells are faced with an external stressor, such as an exposure to a harmful compound, they respond by inducing a new transcription programme in order to synthesise the proteins necessary to eliminate the compound from the system. Typically, the new expression pattern will start

by activating a few transcription factors, which in turn may either initiate or repress the transcription of many other genes that act in response to the new condition. While characterisation of a snapshot of a perturbed biological system using omics technologies can reveal some of the genes, proteins, and metabolites that are specifically altered under the new stable condition, static sampling fails to capture the complex sequence of interactions necessary to arrive at this new state [6]. Time series of data are essential to not only identify the complete set of genes, proteins, and metabolites that are changed under the new conditions, but also to unravel the dynamic interactions between individual components of the system. In this way, time series of omics data has the potential to allow us to go beyond simple associations and establish causative mechanism for disease states that may have therapeutic implications [6-8].

Multi-variate time series data

The generation of time series of data in order to analyse dynamic response and interactions of several biological entities is becoming increasingly prevalent across several biological disciplines.

Toxicological perturbations to decipher system dynamics.

The field of toxicology primarily concerns itself with the evaluation of any type of chemical, environmental factor, or pharmaceutical compound for potential harmful effects in humans. For more than 40 years, the 'gold standard' for assessing a compound for potential adverse outcome in humans has been the two-year rodent *in vivo* bio assay [9]. However, this method is laborious, time-consuming and not in-line with the growing emphasis on the '3Rs' principle: the reduction, replacement and refinement of animal use in experimentation [10]. With the advent of high-throughput omics technologies, *in vitro* toxicogenomics bio-assays have emerged as a viable alternative to the *in vivo* rodent models. Not only have multiple studies indicated that *in vitro* toxicogenomics assays are capable of discriminating between compounds from different subclasses of carcinogenicity [11-14], but these assays may also provide information on genome-wide alterations in mRNA levels following exposure to a compound, allowing for potential insight into the mechanisms that underlie specific types of toxic responses [15]. The evolution of toxicant induced gene regulatory networks occurs both over time and over the compound dosage.

Consequently, within the field of toxicogenomics it is common practice to measure time series of gene expression following a single exposure, or several consecutive exposures, to a compound at multiple dosages [16,17]. By measuring a time series of data, it is not only possible to capture genes that respond earlier, or later, to the exposure but also determine if these alterations in the gene expression are temporary or sustained over a longer

Figure 1.1 : Networks of the human system

The human body is composed of a complex system of distinct organs and tissues that interact via a complex regulatory system of hormones and cytokines to respond to environmental stimuli. Each tissue is composed of a diverse population of cells co-operating to perform particular functions. Within each cell, there is also a complex machinery regulating the flow on information from the genome through the transcriptome to the proteome, modulating cellular metabolic responses. The structure of this thesis also follows this hierarchical organisation of systems within the human body, exploring methods to capture the dynamic responses of each system to external perturbations. In Chapter 2, we quantify skeletal muscle specific insulin resistance from postprandial plasmal glucose trajectory. In Chapter 3, we focus on the tissue specific response to a meal, constructing a computational model of adipose tissue metabolism. Finally, in Chapters 4 and 5 we look at responses to an exposure at the cellular level, exploring the potential of deep learning architectures to analyse temporal responses in gene expression.

period of time at higher dosages. In this way measurement of time series of gene expression data allow for the inference of causal events, facilitating a more mechanistic insight into the modes of toxicological action at a cellular level [18,19]. However, while there are now multiple large publicly available data bases containing time series of gene expression in response to an exposure to an array of compounds [16,17], the development of methods to effectively analyse the temporal gene responses has lagged somewhat behind.

Metabolic challenge tests to quantify the functional capacity of system.

While we are usually more interested in identifying the mechanisms underlying how a biological system responds to an environmental stimulus, the ability of a system to cope with an external challenge or stress can also give an indication of the ‘fitness’ or resilience of the underlying biological system. In fact, recently a new definition of health was proposed which utilises this principle, emphasising the ability of an organism to adapt to physical, emotional, and social challenges in life as a crucial aspect of human health [20].

Chronic diet-related diseases, such as Metabolic Syndrome, type 2 diabetes mellitus (T2DM), and non-alcoholic fatty liver disease (NAFLD), result from sustained disturbances in the delicate interplay between metabolic tissues such as the liver, skeletal muscle, and adipose tissue. These conditions are typically characterised by a loss of glycaemic control, which is often accompanied by disturbances in the storage and release of non-esterified fatty acids (NEFA) [21] and a decrease of insulin sensitivity across multiple metabolic tissues [22-24]. It has long been recognised that dysregulation in glucose homeostasis is evident in the postprandial state long before it is detectable in the tightly regulated, fasting state, and overt clinical disease is present [25-28].

Oral glucose tolerance test

For clinical and research purposes, an oral glucose tolerance test (OGTT), consisting of oral consumption of a 75g glucose oral load, is most commonly used to assess the body’s capacity to clear glucose from the blood [25,27]. Postprandial excursions in plasma glucose and insulin concentrations can be assessed by collecting blood samples at regular time intervals following consumption of the meal. In healthy individuals, with

normal glucose tolerance, plasma glucose insulin levels rise following dietary intake and usually return to fasting levels within 2-3 hours [25] (Figure 1.2, green). As individuals develop insulin resistance higher concentrations of insulin are required to maintain glycaemic control and plasma glucose levels may take longer to return to fasting levels following the glucose bolus (Figure 1.2, yellow). Eventually, it is no longer possible to sustain such high levels of insulin secretion and individuals display a loss of glycaemic control, which also manifests itself as increased plasma glucose concentrations in the fasting state (Figure 1.2, red). The two-hour postprandial plasma glucose concentration is commonly used to assess postprandial glucose trajectories and to classify diabetes and pre-diabetes [25,26,29].

Figure 1.2 : Plasma glucose and insulin trajectories during an OGTT.

This figure depicts a range a possible A) plasma glucose and B) their corresponding insulin trajectories in response to a standard 75g OGTT. Sampling time points are indicated with an X. A typical healthy OGTT response is shown in green, an insulin resistant or impaired glucose tolerance state response is depicted in yellow and a diabetic state is represented in red. Note, it is possible to detect impairments in glycaemic control following the glucose load long before fasting glucose concentrations increase.

Mixed meal challenge tests

In light of implication of disturbances in NEFA metabolism [21] as well as elevated circulating levels of triglycerides and branch chain amino acids [30-32], with the development of insulin resistance, NAFLD, and T2DM studies are increasingly measuring time courses of several metabolic species as well as other glucoregulatory hormones during challenge tests. In addition, alternative nutritional challenges such as or lipid tolerance tests and mixed meal tolerance tests (MMT), which consist of a high caloric mixture of carbohydrates, proteins, and fats, are gaining popularity [33,34]. Not only do the mixed meals induce physiologically more relevant postprandial responses, as their composition reflects that of a standard meal, but they also allow for the assessment of postprandial excursions of triglycerides, NEFA, and individual amino acids and their interplay with glucose and glucoregulatory hormones. Furthermore, the postprandial triglyceride responses to high fat challenges has been shown to be a strong determinant of cardiovascular disease risk values [35,36].

However, the analysis and interpretation of the high dimensional multivariate time series generated by MMTs presents additional challenges. The majority of summary measures for analysing challenge test data have been developed specifically for use on OGTT responses, and are not applicable to multivariate MMT data.

Continuous glucose monitoring

Technological advances in wearable devices for glucose monitoring in recent years mean that it is now possible to generate extensive time series of interstitial glucose levels over the course of several days [37]. Continuous glucose monitor (CGM) data provides the most physiologically relevant assessment of an individual's glycaemic control, monitoring glucose excursions in response to their own diet for multiple days [38,39]. However, as the monitors typically record interstitial glucose levels approximately every five minutes over five, or more, days; analysing and interpreting the resulting high dimensional time series of glucose values can be onerous. Moreover, as the consumed meals have not been standardised, applying cut-off values, such as those identified for OGTT data, can be uninformative.

Analysing multivariate time series

While there has been a surge in the collection of multi-variate time series data across several scientific fields and disciplines, the development of methods to effectively analyse the temporal component of what can often be very high dimension data has lagged somewhat behind [1,6,40]. Consequently, there is a patchwork of often field specific methods to evaluate multivariate time series.

Independent analysis of time points

Many of the specialised methods developed for processing and analysing specific omics measurements, such as gene expression data, have been developed for use on static measurements. Given the lack of sufficient methods for the analysis of time course measurements, it is not uncommon to see individual time points of gene expression data being analysed independently[14,41].

The most commonly applied methods for assessing challenge test data consider the individual time points of an OGTT separately. A series of clinically relevant cut-off values based on the fasting and two-hour post-load plasma glucose concentration following a 75g oral glucose bolus have been established [25,26]. The WHO defined the threshold of diabetes as having a fasting plasma glucose level greater than 7 mmol/l or a two-hour post-load concentration greater than 11.1 mmol/l [25]. Multiple studies have also reported on the one-hour post-load glucose concentration as having a stronger association with incidence of T2DM and cardio-vascular disease than the two-hour value [42,43]. While these cut-offs are sufficient for classifying conditions such as diabetes and pre-diabetes, they fail to capture the diversity in postprandial glucose trajectories, which are indicative of a range of underlying pathologies including insulin resistance, impaired insulin secretion, and inefficacious transition from the gut.

Area under the curve

One of the most frequently used methods to analyse and compare multi-variate time courses of data is the area under the curve [27]. Area under the curve is easily generalisable to different perturbations and temporal scales. The incremental area under the glucose curve over the two-hour period following consumption of a meal is regularly used as

a measure of postprandial glycaemic response in OGTT and continuous glucose monitoring data [38]. While this method can capture the magnitude of the glucose curve over the postprandial period, it does not discriminate between fast and slow responses [39,44]. Consequently, it is possible for very different glucose trajectories to yield similar postprandial glycaemic response values as depicted in Figure 1.3. The use of area under the curve to analyse multivariate mixed meal challenge test data means that the dynamic trajectory of each metabolite is analysed separately, neglecting the dependencies between the metabolite responses [27]. Moreover, the area under the curve method is not sufficiently sensitive to detect changes in peak location. It requires quite substantial differences between metabolite response curves in order to identify a significant difference.

Figure 1.3 : Area under the curve for glucose response curves

This figure depicts two very different average OGTT glucose response curves. While the curves have visibly different trajectories, with differing peak heights, peak times, and peak duration the curves have also identical 2-hour area under the curve values. Demonstrating that area under the curve is not a particularly sensitive metric of time series data.

Insulin sensitivity indices

Several simple summary measures have been constructed to quantify properties such as insulin resistance, or rate of insulin secretion, from the post-load trajectories of glucose and insulin during and OGTT [45]. For example, the Matsuda index quantifies whole body insulin sensitivity as a function of the mean postprandial glucose and insulin concentrations during an OGTT [46]. In recent years, several indices of tissues specific insulin resistance have also been derived [47,48]. However, as these indices typically rely on AUC or mean response values, they also fail to capture the dynamic properties underlying the postprandial glucose and insulin trajectories.

Mathematical models

Mathematical models consist of systems of coupled non-linear ordinary differential equations (ODEs) that describe how the concentrations of various biological entities in the system change over time. Mathematical models have been applied across a range of diverse biological disciplines to study the evolution of complex biological phenomena over time, from models of population dynamics in ecology, to large pharmacokinetic models used to study drug metabolism [49]. Model equations are typically constructed using a bottom up approach, relying on *a priori* information of the underlying biological systems to define how many of the key players interact. These ODE based models are particularly amenable to reconstruction as computer simulations. In this way, large and complex biological systems can be represented in a computer readable format and the response of the biological system under different conditions can be simulated *in silico* (Figure 1.4). Mathematical models allow researchers to investigate how complex regulatory processes are connected and simulate how disruptions of these processes may contribute to the development of adverse outcomes such as disease [50].

The behaviours described by an ODE based model are governed by two components; the equations themselves, which define which biological species interact, and the set of parameter values, which define the rates at which each reaction can occur [51,52]. Traditionally, parameter values were taken from literature, often approximated from experiments conducted *in vitro* or in model organisms. However, the availability of computer-based numerical solvers means that now, even large mathematical models can be

resolved rapidly, allowing parameter values to be directly estimated by fitting model simulation to measured time series of for instance postprandial metabolomics data in humans [51]. In this way, parameter sets describing particular states of the underlying biological system can be identified directly from human data. ODE-based approaches can also be used to describe gene expression responses following an *in vitro* exposure to a toxic compound [35]

Dose-Time-Network Identification (DTNI) incorporates ODEs to describe how the expression of each gene evolves both over both time and dose when inferring gene regulatory networks for exposure data [35]. By integrating

Figure 1.4 : Mathematical models

Mathematical models consist of systems of differential equations describing how a collection of biological entities interact and their behaviour evolves over time. Mathematical models facilitate the representation of pre-existing knowledge about a biological system in a computer understandable format allowing for the *in silico* simulation of the biological system under different conditions. Moreover, with improved numerical solvers, mathematical models can now be parameterised using experimental data. In this way, not only can mathematical models be individualised, but they can also be used to quantify dynamic responses in multivariate time series of data in a physiologically interpretable way.

both the temporal and dose dependent information DTNI has been shown to provide more in-depth information on the mode-of-action of individual compounds, in particular key events and potential molecular initiating event, than existing methods using just the temporal information alone.

The Dalla-Man meal simulation model is a hallmark model in the field of metabolic research [53]. The model described the systemic interplay between glucose and insulin in the postprandial state. This model was parameterised both for normal healthy individuals and a population of individuals with T2DM using gold-standard triple glucose tracer data. The Dalla-Man meal simulator has been demonstrated to be of value as a simulator [54], having been approved for use as an alternative to animal testing in pre-clinical testing of closed loop control systems for Type 1 Diabetes Mellitus [55].

Estimation of parameters values for ODE based mathematical models from measured data has previously been applied to quantify dynamic responses from multivariate time series of metabolic data. Bergman et al.'s glucose minimal was first proposed in 1969 as a method for quantifying insulin sensitivity from time series of plasma glucose and insulin concentrations during a frequently sampled intravenous glucose tolerance test [56]. Computational models not only allow for the quantification of dynamic properties of time series data, but they also allow for temporal information of multiple, related metabolic species to be evaluated simultaneously providing a more complete appraisal of a multivariate time series of data.

Machine learning

With the advent of the big data era in biological research, machine learning has emerged as a promising method for analysing high dimensional data. Unlike classical bottom up approaches, which rely on pre-existing knowledge of the underlying system, machine learning techniques are data driven, utilising data patterns and inference to make predictions. Machine learning approaches have already been successfully applied to analyse multivariate time series of data. In the field of toxicogenomics, a range of pattern recognition methods, such as dynamic time warping and STEM clustering, have been applied to relate time series of gene expression data following exposure to relevant phenotypes [19,47]. In metabolic research, Hulman et al. have recently applied latent class trajectory analysis to categorise glucose trajectories during an OGTT, identifying four distinct physiologically relevant subclasses of glucose responses. In addition, Hall et al. have made use of spectral clustering

to develop a more comprehensive measure of glycaemic variability from continuous glucose monitoring data [39].

Deep Neural Networks

Artificial Neural Networks (ANNs) are a form computation model inspired by our understanding of the function and structure of biological networks of neurons in the brain [58]. The basic structure of an ANN consists of several nodes, or neurons, that are arranged into multiple connected layers (Figure 1.5). At each layer, nodes process the information received from the previous layer as a function of their received input and connection weights and respond accordingly, i.e. become active or remain at zero. Through training the connection weights are tuned, applying heavier weights to the more informative features. In this way, the network transforms the input data into increasingly abstract feature representations for use in classification or predicting [59].

In recent years, with advances in computational power coupled with the availability of sufficiently large training data sets Deep Neural Networks (DNNs) have emerged as a highly successful branch of machine learning. DNNs have produced outstanding results on a range of applications including natural language [60] and image processing [61] and unlike many other machine learning methods, ANNs and DNNs are particularly suited to high dimensional, highly correlated data. With the advent in high-throughput genomics technologies there has been an explosion in the generation of big data in several areas of biology research. DNNs have already been successfully applied to gene expression data, predicting genome wide alterations in gene expression as a results of histone modifications [62] and predicting pharmacological properties of a novel candidate drug given gene expression data [63]. As the volume of available multivariate time series of data increases, DNNs provide an unprecedented opportunity to analyse and explore the high dimensional data, effectively capturing relevant patterns in the dynamic responses.

Transfer learning

Deep learning and machine learning methods typically require large volumes of training data. Currently, in several biological disciplines, it may be too laborious, costly, or even infeasible to generate sufficient quantities of data to train a reliable predictor. Transfer learning is a branch

of machine learning in which knowledge gained from solving one problem is then utilised while solving another, similar, problem [64]. Transfer learning is increasingly being viewed as a solution for the data bottleneck, allowing data from well-studied model organism or systems to be utilised to train models for the target system. A transfer learning approach has already been successfully applied to train a network to classify and segment histopathological images of brain tumours, colon and breast cancer cases using a large publicly available database of natural images [65,66].

Figure 1.5 : Schematic of an artificial neural network

ANNs consist of multiple layers of nodes, or neurons. An input layer (blue), several hidden layers (green), and an output layer (yellow). At each hidden layer, nodes process the information received from the previous layer as a function of the node activations and their respective connection weight (edge weight) and responds accordingly, i.e. become active or remain at zero (inactive). During the process of training, the weight assigned to each connection (thickness/colour of connection edge) is tuned in such a way that most informative features will receive a heavier weight. In this way, the network finds the most favourable representation of the input data.

Aim and outline of thesis

Systems Biology is an emergent research discipline combining experimental and computational biology to uncover the complex interaction between individual component of biological systems and how these interactions give rise to the function and behaviour of the system. In this thesis, we apply computational systems biology approaches to identify disease states by capturing alterations in the dynamic response to an external stimulus across several spatial domains from whole body responses, through to individual tissues, and onto a cellular level (Figure 1). While the challenges of analysing time series data are common across many biological disciplines, in this thesis we have focused on two specific cases; the analysis of meal challenge test data, commonly used as a measure of metabolic health in nutritional and metabolic research, and hepatic gene expression in response to an exposure in toxicological assays.

The research in **Chapter 2** focuses on the standardisation of Abdul-Ghani's muscle insulin sensitivity index through the introduction of a downloadable MISI Calculator. We evaluate accuracy of the index when applied to OGTT data with different sampling schedules and introduce a mathematically more robust method for its calculation. In **Chapter 3**, we employ computational modelling techniques to examine tissue specific insulin resistance at a more mechanistic level. Arteriovenous measurements of multiple metabolites across the abdominal subcutaneous adipose tissue collected at regular intervals during a mixed meal challenge test are used to construct a mathematical model of postprandial adipose tissue metabolism. We then apply our adipose tissue model to quantify rates of insulin-mediated reactions from multivariate time series data collected before and after weight loss in the Yoyo study population indicating an improvement in adipose tissue insulin sensitivity following caloric restriction.

Time series of data are also frequently generated in exposure studies when determining the potential toxicity of a compound. However, unlike skeletal muscle and adipose tissue the liver is not readily accessible for data collection. Animal and hepatic cell line models are typically used to study the time course of liver exposures. This being said, relating animal and *in vitro* results to the human *in vivo* situation presents an additional challenge. In **Chapter 4** we demonstrate the ability of deep learning models to predict time series of human *in vitro* gene expression following exposure to a previously unseen compound given a measured time series of rat *in vitro* gene expression, bench-

marking the deep learning models against more traditional machine learning approaches. The resulting models were also used to relate time series of *in vivo* gene expression to *in vitro* in rats using data from Open TG-GATEs. In **Chapter 5**, we build on the deep learning methods implemented in **Chapter 4**, applying transfer learning to train a deep learning model to predict time series of human *in vivo* gene expression given a time series of human *in vitro* gene expression by leveraging a large data set of rat *in vitro* and *in vivo* gene expression. An overall evaluation of the findings can be found in the General Discussion in **Chapter 6**.

References

1. Liang, Y., Kelemen, A. Dynamic modelling and network approaches for omics time course data: overview of computational approaches and applications. *Briefings in Bioinformatics*, 19(5): 1051-1068 (2017).
2. Guo, N.L. Network medicine: new paradigm in the Omics era. *Anat. Physiol.* 1(1):1000e106 (2011).
3. Machado, D., Costa, R.S., Rocha, M. et al. Modelling formalisms in systems biology. *AMB Express*. 1:45 (2011).
4. Barabási, A-L, Oltvai, Z. Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5:101-13 (2004).
5. Leca, P., Nguyen, T.P., Priami, C. et al. Network inference from time-dependent Omics data. *Methods Mol. Biol.* 719:435-55 (2011).
6. Bar-Joseph, Z. Analysing time series gene expression data. *Bioinformatics* 20(16): 2493-2503 (2004).
7. Bar-Joseph, Z., Gitter, A., Simon, I., Studying and modelling dynamic processes using time-series gene expression data. *Nat. Rev. Genet.* 13(8):552-64 (2012).
8. Hendrix, D.M., Souza, T., Jennen, D.G.J., Kleinjans, J.C.S. DTNI: a novel toxicogenomics data analysis tool for identifying molecular mechanisms underlying the adverse effects of toxic compounds. *Arch. Toxicol.* 91(6):2343-2352 (2017).
9. Chhabra, R.S., Huff, J.E., Schwetz, B.S., Selkrk, J. An overview of pre-chronic and chronic toxicity/carcinogenicity experimental study designs and criteria use by the National Toxicology Program. *Environ. Health Perspect.* 86:313-21 (1990).
10. Stokes, W.S., Animals and the 3Rs in toxicology research and testing: The way forward. *Hum. Exp. Toxicol.* 34(12): 1297-1303 (2015).
11. van Delft, J.H., van Agen, E., van Breda, S.G., Herwijnen, M.H., Staal, Y.C., Kleinjans, J.C. Discrimination of genotoxic from non-genotoxic carcinogens by gene expression profiling. *Carcinogenesis*, 25: 1265-76 (2004).
12. Ellinger-Ziegelbauer, H., Aubrecht, J., Kleinjans, J.C., Ahr, H.J., Application of toxicogenomics to study mechanisms of genotoxicity and carcinogenicity. *Toxicol. Lett.*, 186: 36-44 (2009).
13. Hu, T., Gibson, D.P., Carr, G.J., Torontali, S.M., Tiesman, J.P., Chaney, J.G., Aardem, M.J., Identification of a gene expression profile that discriminates indirect-acting genotoxins from direct-acting genotoxins. *Mutat. Res.*, 549: 5-27 (2004).
14. Mathijs, K., Brauers, K.J., Jennen, D.G., Lizarraga, D., Kleinjans, J.C., van Delft, J.H., Gene expression profiling in primary mouse hepatocytes discriminates true from false-positive genotoxic compounds. *Mutagenesis*, 25, 561-568 (2010).
15. Gautier, L., Taboureaux, O., Audouze, K. The effect of network biology on drug toxicology. *Exp. Opin. Drug. Metab. Toxicol.* 9(11):1409-1418 (2013).
16. Igarashi, Y., Nakatsu, N., Yamashita, O., Ono, A., Urushidani, T., Yamada, H., Open TG-GATES: a large-scale toxicogenomics database. *Nucleic Acids Res.* 43: 21-7.
17. Ganter, B., Tugendreich, S., Pearson, C.I., Ayaboglu, E., Baumhueter, S., Bostian, K.A. et al. Development of a large-scale chemogenomics database to improve drug candidate selection and to understand mechanisms of chemical toxicity and action. *J. Biotechnol.* 119: 219-244 (2005).
18. Hendrix, D.M., Souza, T., Jennen, D.G.J., Kleinjans, J.C.S. DTNI: a novel toxicogenomics data analysis tool for identifying molecular mechanisms underlying the adverse effects of toxic compounds. *Arch. Toxicol.* 91(6):2343-2352 (2017).
19. Cavill, R., Kleinjans, J., Briede, J.J. DTW4Omics : Comparing patterns in biological timer series. *PLOS One* 8(8):e71823 (2013).

-
20. Huber, M., Knottnerus, J.A., Green, L. et al. How should we define healthy? *Bmj*. 343:d4163 (2011)
 21. Unger, R.H. Minireview: weapons of lean body mass destruction: the role of ectopic lipids in the metabolic syndrome. *Endocrinology* 144:5159–5165 (2003).
 22. Kahn, S.E., Hull, R.L., Utzschneider, K.M. Mechanisms linking obesity to insulin resistance and type 2 diabetes. *Nature* 444:840–846 (2006).
 23. Eckel, R.H., Grundy, S.M., Zimmet, P.Z. The metabolic syndrome. *Lancet* 365:1415–1428 (2005).
 24. Reaven, G.M. Role of insulin resistance in human disease. *Diabetes* 37:1595–1607 (1988).
 25. World Health Organization: Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia. Report of a WHO/IDF consultation. 2006, Geneva: World Health Organization
 26. American Diabetes Association: Standards of Medical Care in Diabetes-2011. *Diabetes Care*. 34 (Suppl 1): S11–S61. (2011).
 27. Vis, D.J., Westerhuis, J.A., Jacobs, D.M., van Dynhoven, J.P.M., Wopereis, S. et al. Analyzing metabolomics-based challenge tests. *Metabolomics*. 11(1):50–63 (2015).
 28. van Ommen, B., Keijer, J., Kleemann, R., Elliott, R., Drevon, C. A., McArdle, H., et al. The challenges for molecular nutrition research 2: Quantification of the nutritional phenotype. *Genes Nutr.*, 3, 51–59 (2008).
 29. Fray Frosli, K., Roislien, J., Ovigstad, E, Godang, K., Bollerslev, J. et al. Shape information from glucose curves. Functional data analysis compared with traditional summary measures. *BMC Med. Res.* 13:6 (2013).
 30. Würtz P, Soininen P, Kangas AJ et al. Branched-chain and aromatic amino acids are predictors of insulin resistance in young adults. *Diabetes Care*. 2013; 36: 648–655.
 31. Batch BC, Shah SH, Newgard CB et al. Branched chain amino acids are novel biomarkers for discrimination of metabolic wellness. *Metabolism*. 2013; 62: 961–969.
 32. Bloomgarden, Z. Diabetes and branched-chain amino acids: What is the link? *J. Diabetes*. 10(5):350–352 (2018).
 33. Stroeve, J.H.M., van Wietmarschen H., Kremer, B.H.A, van Ommen, B., Wopereis, S. Phenotypic flexibility as a measure of health: the optimal nutritional response test. *Genes Nutr.* 10(3):13 (2015).
 34. Woperies, S., Stroeve, J.H.M, Stafleu, A., et al. Multi-parameter comparison of a standardized mixed meal tolerance test in healthy and type 2 diabetes subjects: the PhenFlex challenge. *Genes Nutr.* 12:21 (2017).
 35. Freiberg, J.J., Tybjaerg-Hansen, A., Jensen, J.S., Nordestgaard, B.G. Nonfasting triglycerides and risk of ischemic stroke in the general population. *JAMA*. 300(18):2142–52 (2008).
 36. Mora, S., Rifai, N., Buring, J.E., Ridker, P.M. Fasting compared with non-fasting lipids and apolipoproteins for predicting incident of cardiovascular events. *Circulation*. 118(18):S568–S568 (2008).
 37. Bailey, T.S., Ahmann, A., Brazg, R., Christiansen, M., Garg, S. et al. Accuracy and acceptability of the 6-day Enlite continuous subcutaneous glucose sensor. *Diabetes Technol. Ther.* 16:277–283 (2014).
 38. Zeevi, D., Korem, T., Zmora, N., Israeli, D., Rothschild, D. et al. Personalized nutrition by prediction of glycemic responses. *Cell*. 163(5):1079–1094 (2015).
 39. Hall, H., Perelman, D., Breschi, A., Limcaoco, P., Kellogg, R. et al. Glucotypes reveal new patterns of glucose dysregulation. *PLOS Biol.* 16(7):e2005143 (2018).
 40. Grigorov, M.G. (2011) Analysis of Time Course Omics Datasets. In: Mayer B. (eds) Bioinformatics for Omics Data. *Methods in Molecular Biology (Methods and Protocols)*, vol 719. Humana Press

41. Caiment, F., Tsamou, M., Jennen, D., Kleinjans, J. Assessing compound carcinogenicity in vitro using connectivity mapping. *Carcinogenesis*. 35(1):201-207 (2013).
42. Jagannathan, R., Bergman, M. Use of 1-h post-load plasma glucose concentration to identify individuals at high risk of developing type 2 diabetes. *Diabet. Med.* 34:877-878 (2017).
43. Pareek, M., Bhatt, D.L., Nielsen, M.L., et al. Enhanced predictive capability of a 1-hour oral glucose tolerance test: a prospective population-based cohort study. *Diabetes Care*. 41:171-177 (2018).
44. Hulman, A., Witte, D.R., Vistisen, D., Balkau, B., Dekker, J.M. et al. Pathophysiological Characteristics Underlying different glucose response curves: A latent class trajectory analysis from the prospective EGIR-RISC Study. *Diabetes Care*. 41(8):1740-1748 (2018).
45. Gutch, M., Kumar, S., Razi, S.M., Gupta, K.K., Gupta, A. Assessment of insulin sensitivity/resistance. *Indian J Endocrinol Metab.* 19(1):160-164 (2015).
46. Matsuda, M., DeFronzo, R. A. Insulin sensitivity indices obtained from oral glucose tolerance testing: comparison with the euglycaemic insulin clamp. *Diabetes Care*. 22, 1462-1470 (1999).
47. Abdul-Ghani, M.A., Matsuda, M., Balas, B., DeFronzo, R.A., Muscle and liver insulin resistance indexes derived from the oral glucose tolerance test. *Diabetes Care*. 30(1):89-94 (2007).
48. Gastaldelli, A., Harrison, S.A., Belfort-Aguilar, R., Hardies, L.J., Balas, B. et al. Importance of changes in adipose tissue insulin resistance to histological response during thiazolidinedione treatment of patients with nonalcoholic steatohepatitis. *Hepatology*. 50(4):1087-93 (2009).
49. Sheiner, L. B., Ludden, T. M. Population pharmacokinetics/dynamics. *The Annual Review of Pharmacology and Toxicology*, 32: 185-209. (1992).
50. Fischer, H.P. Mathematical modelling of complex biological systems: from parts lists to understanding systems behaviour. *Alcohol Res Health*. 31(1):49-59 (2008).
51. van Riel, N.A.W. Dynamic modelling and analysis of biochemical networks: mechanism-based models and model-based experiments. *Brief. Bioinform.* 7(4):364-374 (2006).
52. Tiemann, C.A., Vanlier, J., Hilbers, P.A.J., van Riel, N.A.W. Parameter adaptations during phenotype transitions in progressive diseases. *BMC Syst. Biol.* 5:174 (2011).
53. Dalla Man, C., Rizza, R.A., Cobelli, C., Meal simulation model of the glucose-insulin system. *IEEE Trans. Biomed. Eng.* 54(10):1740-9 (2007).
54. Ajmera, I., Swat, M., Laibe, C., Novère, N.L., Chelliah, V. The impact of mathematical modeling on the understanding of diabetes and related complications. *CPT Pharmacomet Syst Pharmacol.* 2: e54 (2013).
55. Kovatchev, B.P., Breton, M., Man, C.D., Cobelli, C. In Silico Preclinical Trials: A Proof of Concept in Closed-Loop Control of Type 1 Diabetes. *J Diabetes Sci Technol Online*. 3: 44-55 (2009).
56. Bergman, R.N., Ider, Y.Z., Bowden, C.R., Cobelli, C. Quantitative estimation of insulin sensitivity. *Am. J. Physiol.* 236(6): E677-77 (1979).
57. Hendrickx, D.M., Jennen, D.G., Briede, J.J., Cavill, R., de Kok, T.M., Kleinjans, J.C. Pattern recognition methods to relate time profiles of gene expression with phenotypic data: a comparative study. *Bioinformatics*. 31(13):2115-22 (2015).
58. LeCun, Y., Bengio, Y., Hinton, G. Deep learning. *Nature* 521, 436-444 (2015).
59. Lee, H., Grosse, R., Ranganath, R., and Ng, A. Y. (2011). Unsupervised learning of hierarchical representations with convolutional deep belief networks. *Commun. ACM* 54, 95-103.
60. Socher, R., Lin, C.C.Y., Ng, A.Y., Manning, C.D. Parsing natural scenes and natural language with recursive neural networks. In: *Proceedings of the 28th International Confer-*

-
- ence on Machine Learning (ICML-11), 129–136.
61. Sutskever, I., Vinyals, O., and Le, Q. V. Sequence to sequence learning with neural networks,” in *Advances in Neural Information Processing Systems 27 (NIPS 2014)*.
 62. Singh, R., Lanchantin, J., Robins, G., Qi, Y., DeepChrome: deep-learning for predicting gene expression from histone modifications, *Bioinformatics*, 32, i639-i648, (2016).
 63. Aliper, A., Plis, S., Artemov, A., Ulloa, A., Mamoshina, P., Zhavoronkov, A., Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. *Mol Pharm* 13, 2524-2530. (2016).
 64. Pan, S. J., Yang, Q. A survey on transfer learning. *IEEE Trans. Know. Data Eng.* 22, 1345–1359 (2010).
 65. Xu, Y., Jia, Z., Wang, L.B., Ai, Y., Zhang, F. et al. Large scale tissue histopathology image classification, segmentation, and visualizations via deep convolutional activation features. *BMC Bioinform.* 18:281 (2017).
 66. Wahab, N., Khan, A. Transfer learning based deep CNN for segmentation and detection of mitoses in breast cancer histopathological images. *Microscopy* 68(3):216-233 (2019).

2

Improved quantification of muscle insulin sensitivity using oral glucose tolerance test data: the MISI Calculator

Shauna O'Donovan¹, Michael Lenz¹, Gijs H. Goossens², Carla J.H. van der Kallen^{3,4}, Simone J.M.P. Eussen⁵, Coen D.A. Stehouwer^{3,4}, Marleen M. van Greevenbroek^{3,4}, Miranda T. Schram^{3,4,6}, Simone J. Sep^{3,4,7,8}, Ralf L.M. Peeters^{1,9}, Ellen E. Blaak², Natal A.W. van Riel^{1,10}, Theo M.C.M. de Kok^{1,11}, Ilja C.W. Arts^{1,4,5}.

¹Maastricht Center for Systems Biology (MaCSBio), Maastricht University. ²Dept. of Human Biology, NUTRIM School of Nutrition and Translational Research in Metabolism, Maastricht University. ³Dept. of Internal Medicine, Maastricht University Medical Center+. ⁴Cardiovascular Research Institute Maastricht (CARIM), Maastricht University. ⁵Dept. of Epidemiology, Maastricht University. ⁶Heart and Vascular Center, Maastricht University Medical Center+. ⁷School of Public Health and Primary Care (CAPHRI), Maastricht University. ⁸Dept. of Rehabilitation Medicine, Maastricht University. ⁹Dept. of Data Science & Knowledge Engineering, Maastricht University. ¹⁰Dept. of Biomedical Engineering, Eindhoven University of Technology. ¹¹Dept. of Toxicogenomics, GROW School of Oncology and Developmental Biology, Maastricht University.

Scientific Reports (2019) 9(1):9388

Abstract

The Muscle Insulin Sensitivity Index (MISI) has been developed to estimate muscle-specific insulin sensitivity based on oral glucose tolerance test (OGTT) data. To date, the score has been implemented with considerable variation in literature and initial positive evaluations were not reproduced in subsequent studies. In this study, we investigate the computation of MISI on oral OGTT data with differing sampling schedules and aim to standardise and improve its calculation. Seven time point OGTT data for 2631 individuals from the Maastricht Study and seven time point OGTT data combined with a hyperinsulinemic-euglycaemic clamp for 71 individuals from the PRESERVE Study were used to evaluate the performance of MISI. MISI was computed on subsets of OGTT data representing four and five time point sampling schedules to determine minimal requirements for accurate computation of the score. A modified MISI computed on cubic splines of the measured data, resulting in improved identification of glucose peak and nadir, was compared with the original method yielding an increased correlation ($\rho = 0.576$) with the clamp measurement of peripheral insulin sensitivity as compared to the original method ($\rho = 0.513$). Finally, a standalone MISI calculator was developed allowing for a standardised method of calculation using both the original and improved methods.

Introduction

Insulin resistance is an important feature of the metabolic syndrome, which has been associated with approximately a threefold increased risk for the development of type 2 diabetes mellitus and cardiovascular disease [1]. Insulin resistance can occur in several tissues including the liver, skeletal muscle, and adipose tissue [2,3]. The severity of insulin resistance has also been found to vary between the tissues leading to a range of phenotypes [4,5,6] including dyslipidemia [7], disturbed glucose metabolism, and hyperinsulinemia [8]. Interventions, such as reduced caloric intake, have been shown to reduce whole-body insulin resistance [9]. However, several studies have demonstrated that certain interventions exert more or less tissue-specific effects, metformin administration improves hepatic insulin sensitivity [10] and data suggests that increased physical activity predominantly affects skeletal muscle insulin sensitivity [11]. This tissue-specificity of interventions emphasises the importance of being able to accurately quantify tissue-specific insulin sensitivity. Currently, a hyperinsulinemic-euglycaemic clamp [12] combined with a glucose tracer infusion [13,14] is the gold standard for quantifying peripheral insulin resistance. Unfortunately, this method is costly, invasive, and time-consuming, making it difficult to implement in a clinical setting or on a large scale. These difficulties have been a motivating factor in the development of alternative methods.

Several metrics exist which attempt to quantify insulin sensitivity in the body as a whole using glucose and insulin measurements either after fasting (HOMA-IR [15], QUICKI [16]) or during an Oral Glucose Tolerance Test (OGTT) (Matsuda [17], Stumvoll [18]). These scores are widely used throughout metabolic and epidemiological research. In their 2007 paper, Abdul-Ghani et al. developed and validated tissue specific insulin sensitivity scores for muscle and liver [19]. The muscle insulin sensitivity index (MISI) has been validated through comparison with a hyperinsulinemic-euglycaemic clamp combined with a glucose tracer infusion (n=155, Spearman correlation of $\rho = 0.78, p < 0.0001$).

The validity of the MISI score was subsequently questioned by several authors, reporting still significant, but considerably lower correlations of 0.542 [20], 0.55 [21], and 0.41 [22] with the hyperinsulinemic clamp. Moreover, Wilson and Ross [21] also investigated the standard error of estimate of MISI, their results indicating a lack of precision with the score. While these lower

correlations have been attributed to differences in the heterogeneity of study populations, particularly with regard to glucose tolerance status and BMI range [23,24], the circumstances in which the application of the MISI score is appropriate and the factors contributing to such low correlation values in other studies still requires further evaluation.

Here, we investigated the dependence of MISI on the number of measurement time points during an OGTT, identifying minimal requirements for calculation in order to ensure an accurate estimation of MISI. Furthermore, we evaluated both the original and proposed modified MISI score with improved numerics (introduced in this article) based on a comparison to a hyperinsulinemic-euglycaemic clamp and describe numerical issues in the original MISI score calculations that contribute to the reported low correlations with the clamp.

In addition, as the original article [19] did not specify measurement units or a precise method of calculation, the MISI score has been calculated with substantial variation in literature to date [25]. As a result MISI is most frequently used to rank individuals into relative muscle insulin sensitivity tertiles [26], making comparison of MISI scores across studies impossible. Therefore, we have created a downloadable stand-alone MISI calculator for the automated and standardised calculation of MISI using both the original method and proposed modified method. The calculator also automatically flags OGTT curves for which the calculated MISI value may not be meaningful. This is of particular importance when automating the computation of MISI for large data sets which are increasingly becoming available.

Research Design & Methods

Standardising original MISI calculation

The MISI formula is based on the assumption that the decay of plasma glucose during an OGTT is primarily due to uptake by skeletal muscle, as endogenous production of glucose from liver will be inhibited by the post load increase in insulin [19]. It is calculated as

$$MISI = \frac{\frac{dG}{dt}}{I} \quad (2.1)$$

where $\frac{dG}{dt}$ denotes the slope of the regression line from the peak of the plasma

glucose curve to its nadir and $\bar{I}$ is the average insulin concentration over the duration of the OGTT. There has been some variation in the implementation of MISI to date, most notably the omission of 0 minute time point when computing $\bar{I}$ [25]. Hence, we propose a standardised original MISI were $\frac{dG}{dt}$ is measured in $\mu\text{mol/L/min}$ and computed as the line of best fit from the peak to the nadir, excluding any glucose rebound, and $\bar{I}$ is the mean measured insulin over the full 120 minutes in pmol/L .

Figure 2.1 : Overview of exclusion criteria for Maastricht Study data and results of analysis in PRESERVE and Maastricht studies.

(a) Flow chart of exclusion of data from the Maastricht Study for analysis. **(b)** Bar plots of time of glucose peak in Maastricht study OGTT data, white bars show seven time point data; black bars indicate the five time point subset. Dashed line shows distribution of peak glucose time predicted by cubic splining of five time point data. **(c)** $\frac{dG}{dt}$ five (dashed line) computed on a five time point subset (black circles) differs from $\frac{dG}{dt}$ seven (dot-dashed line) computed on a seven time point OGTT (black circles plus empty circles). $\frac{dG}{dt}$ spline (dotted line) computed on a cubic spline of the five time point data (solid line) more closely represents the seven time point $\frac{dG}{dt}$. **(d)** Scatter plot of PRESERVE Study ($n=69$) clamp ssGIR with both the standard MISI (open circles) and cubic spline MISI (crosses). Correlation for standard MISI and modified MISI with ssGIR value are represented by the solid and dashed lines of best fit respectively.

Modified MISI

The original MISI calculation is heavily dependent on the sampling times during an OGTT. Specifically, the calculation of dG/dt is critical since it relies on the identification of the maximum and minimum plasma glucose concentrations over the duration of the OGTT, which is currently limited to the discrete set of measured time points; with the true peak/nadir often occurring between two sampled time points. The computation of $\bar{I}$ (mean insulin concentration) is also sensitive to the sampling times, particularly in the case of unequal sampling intervals, which results in an increased weight of the more densely sampled intervals of the insulin curve. Such unequal sampling intervals are common with four (e.g. 0,30,60,120 min) or seven (e.g. 0,15,30,45,60,90,120 min) time point OGTTs. In order to overcome these problems, we propose to interpolate the OGTT data with a cubic spline, inferring smooth glucose and insulin curves that allow for improved prediction of the peak glucose time and concentration while also accounting for unequal sampling intervals. Cubic splining is achieved by the piece-wise fitting of third order polynomials between the measured data points such that the first and second derivatives at each measured point are continuous. In addition, we assume that the first and second derivatives at the start of the glucose curve (0min) are zero, imposing a steady state condition. The regression line used to determine dG/dt is now fitted to the full inferred glucose curve rather than a limited number of sampled points (Figure 2.1 c), with a prediction of the true glucose peak and nadir. $\bar{I}$ is also computed on the full inferred insulin curve, thereby improving MISI calculation. Herein, this cubic spline method will be referred to as “modified MISI”.

Flagging of problematic OGTT curves

Comparison of five and seven time point OGTT data from the Maastricht Study [27] identified several occasions where it is either not possible to calculate MISI or where, given the number of time points, the calculated MISI value might not be biologically meaningful. The criteria for flagging OGTT curves are as follows (Figure 2):

1. The glucose curve has a peak at 120min (no unique solution for $\frac{dG}{dt}$)
2. The glucose curve appears to be flat, defined as peak glucose concentration (G_{peak}) being less than 0.5 mmol/L greater than the fasting value (G_0): $G_{peak} < G_0 + 0.5 \text{ mmol/L}$
3. The glucose curve has a non-negligible rebound, i.e. the nadir is not

the global minimum for the glucose curve over 120 minutes. Defined as having a glucose rebound more than 0.5 mmol/L greater than global minimum.

Datasets and measurements

In order to examine the effect of different sampling schedules and population composition on the computation of MISI and modified MISI we use data from the Maastricht Study [27] and PRESERVE Study [28].

The Maastricht Study

The Maastricht Study is an observational prospective population-based cohort study. The rationale and methodology have been described previously [27]. In brief, the study focuses on the etiology, pathophysiology, complications, and comorbidities of type 2 diabetes and

Figure 2.2 : Criteria for flagging glucose curves when calculating MISI.

(a) no unique solution for calculating $\frac{dG}{dt}$ when the glucose peak occurs at 120 minutes. (b) Glucose curves with a glucose peak less than 0.5 mmol/l greater than the fasting value have often failed to capture the glucose peak and tend to yield low MISI values (dotted line), whereas the true value may be quite high (dashed line computed on seven time points). (c) Larger glucose rebound may not be negligible, strictly using the nadir may result in biologically irrelevant MISI scores.

is characterised by an extensive phenotyping approach. All individuals aged between 40 and 75 years and living in the southern part of the Netherlands were eligible for participation. Participants were recruited through mass media campaigns and from the municipal registries and the regional Diabetes Patient Registry via mailings. Recruitment was stratified according to known type 2 diabetes status, with an oversampling of individuals with type 2 diabetes, for reasons of efficiency. The present report includes cross-sectional data from the first 3451 participants, who completed the baseline survey between November 2010 and September 2013. The examinations of each participant were performed within a time window of three months. The study has been approved by the institutional medical ethical committee (NL31329.068.10) and the Minister of Health, Welfare and Sports of the Netherlands (Permit 131088-105234-PG). All participants gave written informed consent. Provided participants were not using exogenous insulin and had a fasting plasma less than 11 mmol/L they underwent a seven time point 75g OGTT (0,15,30,45,60,90,120 mins). Removing individuals with missing time points, complete seven time point OGTT data was available for 2631 individuals, (Figure 1a), 1584 with normal glucose tolerance, 124 with impaired fasting glucose, 305 with impaired glucose tolerance, and 616 with type 2 diabetes, cut-offs defined by WHO 2006 criteria [29]. Subjects had a wide range of BMI (14.38 – 52.25 kg/m²) [mean 26.81 ± standard deviation 4.32], a mean age of 59.7 ± 8.2 years and were 52.1% male (Supplemental Table S1), For subsequent analysis all flagged glucose curves were excluded, with OGTT data for 1891 individuals used to compare common OGTT sampling schedules.

PRESERVE study

The PRESERVE Study is a randomised controlled, double-blind, two-centre study in which 79 individuals with impaired fasting glucose and/or impaired glucose tolerance were randomised to receive 320mg of valsartan or placebo for 26 weeks in order to evaluate the effects of the angiotensin II type 1 receptor blocker on insulin sensitivity and β -cell function [28] (Clinical trial reg. no. ISRCTN42786336). The study was conducted in accordance with the Declaration of Helsinki and was approved by the local ethics committee. All participants gave written informed consent. At baseline, all individuals underwent a seven time point 75g OGTT (t0,10,20,30,60,90,120 mins) and a hyperinsulinemic-euglycaemic clamp with an insulin infusion rate of 40 mU/min/m². The steady state glucose infusion rate over the final 30 min of the hyperinsulinemic-euglycaemic

clamp (ssGIR, $\text{mg} \cdot \text{min}^{-1} \cdot \text{kg}^{-1}$) was used to calculate peripheral insulin sensitivity. Data for 71 out of the 79 individuals was used, excluding five individuals with missing values in OGTT or clamp data, and three individuals for whom MISI could not be computed as their peak glucose concentration occurred at 120 minutes. Of the remaining 71 individuals (45% male, average age of 58.2 ± 6.9 years, mean BMI of $29.8 \pm 4.5 \text{ kg/m}^2$) 39 had impaired fasting glucose, 11 had impaired glucose tolerance, and 21 had both impaired fasting glucose and impaired glucose tolerance, using glucose metabolism cut-offs as defined by WHO 2006 [29]. Two further individuals were classed as statistical outliers for MISI due to acute hypoglycaemia at 90 minutes ($<2.5 \text{ mmol/L}$), resulting in very high MISI values. Results are provided both including and excluding these individuals (Supplemental table S2).

The MISI calculator

The MISI calculator was programmed in MATLAB 2014b, The MathWorks Inc., Natick, Massachusetts, United States of America. and the stand alone graphical user interface (GUI) was generated to install Matlab Runtime, such that no Matlab installation is required for the use of the calculator (Figure 2.3). Both the source code and the stand-alone calculator can be downloaded from www.maastrichtuniversity.nl/misi. A tutorial and example data files are also provided.

Calculations and statistics

In order to evaluate the performance of both the standard and modified MISI the following analyses were performed using each method and compared. In order to test the effect of sampling frequency on performance MISI calculated on subsets of data representing several standard OGTT sampling schedules derived from the Maastricht Study OGTT data ($n=1891$) were compared. Seven time points ($t_0, 15, 30, 45, 60, 90, 120$ min), five time points ($t_0, 30, 60, 90, 120$ min) and several four time point subsets omitting various time points from the five time point subset were compared. Spearman correlations with standard five time point sampling schedule with permutation based p-values and reclassification probabilities were computed, as this was the sampling frequency validated in the original article [19]. Reclassification probabilities between two scores were calculated as the number of samples that were classified into two

Chapter 2. Improved quantification of muscle insulin sensitivity using oral glucose tolerance test data: the MISI Calculator

Figure 2.3 : MISI score calculator indicating input and output for its use.

The MISI Calculator takes time series of glucose and insulin measurements collected during an OGTT for multiple individuals stored in the form of Excel files as input. The calculator can convert measurement units of both glucose and insulin to SI units used for calculation. The user can specify flagging criteria and method for the calculation of MISI. The user can choose to save plots of the glucose and insulin curves when calculating the MISI score or view the flagged glucose and insulin curves in the MISI Calculator app. Output of the calculator is an Excel file containing the user-supplied ID number and the calculated MISI value. If MISI was not calculated, either due to flagging or missing values, an explanation is provided along with a suggested MISI value where possible.

different insulin resistance tertiles by each score, divided by the total number of samples. This approach was repeated within the PRESERVE Study data set ($n=71$), comparing the MISI computed using both methods on five time point and the various four time point subsets to the ssGIR of the hyperinsulinemic-euglycaemic clamp. It has been reported that the heterogeneity of the population being examined may have an impact on the strength of the correlation between MISI and the clamp-based insulin sensitivity [20,23] with a wider spread in values yielding a higher correlation [24]. To further investigate this relationship and to test the robustness of our correlation results, a bootstrapping-like approach was used, in which 1000 subsets of various sizes (50-70 individuals) were randomly drawn from

the 71 individuals of the PRESERVE dataset and evaluated based on their Spearman correlation between both the standard and modified MISI and the clamp-based insulin sensitivity.

Results

Modified MISI

Analysis of the Maastricht Study seven time point OGTT demonstrated that 27% of the total population studied (722 of 2631 individuals) had a peak glucose concentration at 45 mins (Figure 2.1 b, white bars). Had the OGTT been sampled at just five time points (t0,30,60,90,120 min) (Figure 2.1 b, black bars), the majority of individuals with a peak glucose at 45 min would be distributed between having a glucose peak at 30 min (366 of 722 individuals) or 60 min (324 of 722 individuals). With some individuals now having a peak at 90 min or later (32 of 722), introducing potential inaccuracy in the resulting MISI values as the true concentration of the glucose peak will be underestimated. Cubic splining of this five time point OGTT subset allowed for the peak glucose concentration to be inferred. Out of the 722 individuals with a peak at 45 minutes in the seven time point data, 409 were assigned a peak glucose concentration between 37.5 mins and 52.5 mins, enabling the recovery of information which may potentially have been lost as a result of less frequent sampling.

A five time point subset was generated from the PRESERVE Study OGTT data using 0, 30, 60, 90, and 120 min time points to reflect a standard OGTT. The clamp ssGIR was compared to both the standard and modified MISI computed on the five time point subset using Spearman correlation. The modified method yielded a non-significant increase in correlation with ssGIR over the standard method $\rho = 0.576$ versus $\rho = 0.513$ excluding the two statistical outliers (n=69). Retaining the two outliers the same trend is observed with the modified MISI yielding a correlation of $\rho = 0.526$ with the ssGIR versus $\rho = 0.469$ for the standard method (n=71). ($p < 0.001$ for all correlation coefficients).

MISI applied to OGTT with varying temporal information

In order to determine the minimum requirements for calculation, both the standard and modified MISI were computed on various four

time point subsets of OGTT data and compared with the ssGIR of the hyperinsulinemic euglycaemic clamp for 66 individuals in the PRESERVE study. Results, shown in Table 2.1, indicate poor correlation between standard MISI computed on four time points and the clamp ssGIR, particularly when the 30 or 60 minute time point was absent ($\rho = 0.440$ and $\rho = 0.356$ respectively). The modified MISI calculated on four time points yields a higher correlation with the clamp ssGIR than the standard MISI. When the 60 minute time point was missing the correlation coefficient increases from $\rho = 0.356$ for the standard MISI to $\rho = 0.618$ for the modified MISI.

Within epidemiological and intervention studies to date MISI has typically been used to rank individuals into muscle insulin sensitivity tertiles [26,30]. Individuals within the Maastricht Study (n=1891) were ranked into muscle insulin sensitivity tertiles using the standard MISI computed on the four time point subsets of data and compared with classification defined using the five time point data. Again, provided the 120 min time point was included the reclassification probability was less than 33%. Reclassification probability of MISI computed on seven time point data with MISI computed on five time point OGTT data was 15.2%. Looking more closely at the effect of reclassification of the composition of the muscle insulin sensitivity tertiles defined by both standard and modified MISI in the Maastricht Study data set (n=1891) (Supplementary table S2.3). It is of note that the least insulin sensitive group as defined by modified MISI has an increased mean whole body insulin resistance (as defined by HOMA) and contains a lower proportion of normal glucose tolerance (NGT) individuals than the group defined using standard MISI (55.2% for modified MISI versus 56.7% for standard MISI calculated on seven time points). Moreover, the reverse can be seen in the most insulin sensitive groupings; with the modified MISI group having a lower mean HOMA value and a greater number of NGT individuals than the group defined using standard MISI (65.1% for modified MISI versus 60.8% for standard MISI calculated on seven time points). These differences are statistically non-significant (using a two tailed t-test).

Table 2.1: Spearman correlations and reclassification probabilities for MISI computed using both the standard and modified methods on OGTT data for various common sampling schedules

	Maastricht Study (n=1891)				PRESERVE Study (n=66)			
Correlation time points	5 time point MISI		7 time point MISI		5 time point MISI		7 time point MISI	
	std	mod	std	mod	std	mod	std	mod
4(30,60,90,120)	0.999	0.945	0.924	0.819	0.999	0.944	0.473	0.457
4(0,60,90,120)	0.928	0.881	0.878	0.794	0.897	0.826	0.440	0.464
4(0,30,90,120)	0.817	0.842	0.775	0.763	0.831	0.852	0.356	0.618
4(0,30,60,120)	0.790	0.703	0.748	0.664	0.886	0.755	0.481	0.661
5(0,30,60,90,120)	1	1	0.925	0.864	1	1	0.476	0.561
Reclassification time points	5 time point MISI		7 time point MISI		5 time point MISI		7 time point MISI	
	std	mod	std	mod	std	mod	std	mod
4(30,60,90,120)	0.015	0.202	0.151	0.277	0.030	0.152	0.470	0.576
4(0,60,90,120)	0.160	0.216	0.202	0.302	0.106	0.212	0.515	0.485
4(0,30,90,120)	0.326	0.292	0.348	0.347	0.349	0.227	0.500	0.485
4(0,30,60,120)	0.331	0.406	0.365	0.412	0.167	0.333	0.591	0.394
5(0,30,60,90,120)	0	0	0.145	0.220	0	0	0.485	0.485

First two columns show comparison with MISI computed on five and seven time points within the Maastricht Study (n=1891). The second two columns show comparison with MISI computed on five time points and the steady state glucose infusion rate of the hyperinsulinemic-euglycaemic clamp (ssGIR) for individuals within the PRESERVE Study. The upper half of the table shows correlation coefficients computed by Spearman correlation, the lower section shows net reclassification probabilities

Effect of population on correlation with clamp

Several studies have reported much lower correlations between MISI and the clamp ssGIR than observed in the original paper [20,21,22]. These lower values have largely been attributed to the reduced heterogeneity of the populations being used, with a narrower range of muscle insulin sensitivity when compared to the original study [23,24]. Testing the sensitivity of the correlation between the MISI and the clamp ssGIR to population composition, we randomly selected 1000 sub-populations of increasing size from the PRESERVE study and compared the Spearman correlations achieved between MISI and the clamp ssGIR (Figure 2.4). Looking at the range of the 95% quantiles, correlations of 0.36 to 0.60 with a median of 0.47 were possible when randomly selecting sub-populations

Figure 2.4: Spearman correlation values between standard and modified MISI and clamp ssGI for randomly selected sub-populations of PRESERVE study

Median spearman correlation between clamp ssGI and both standard MISI (solid line) and cubic spline MISI (dashed line) for 1000 randomly selected subgroups of varying size from the PRESERVE Study. 95% quantiles are indicated by the shaded region to either side of the medians.

of 50 individuals from the data set of 71 individuals using the standard method, demonstrating the variability of the Spearman correlation between MISI and the clamp. For the modified method there was an improved median correlation of 0.53 for the same sub-populations with 95% quantiles of 0.42 to 0.64.

Flagged glucose curves

The original article already mentions that MISI is not appropriate for individuals with type 2 diabetes due to the peak occurring at 120 minutes [19]. There were 154 individuals (6% of total data set) with a peak at 120 minutes within the Maastricht Study, of whom 107 (69.5%) had type 2 diabetes. As there is no unique solution for dG/dt for a glucose curve with a peak at 120 minutes (Figure 2.2 a) it is not possible to compute MISI.

However, it would be expected that these individuals would fall into the least muscle insulin sensitive tertile.

In addition, there were 47 individuals (2%) where the peak glucose concentration was less than 0.5 mmol/L greater than the fasting value (Figure 2.2 b). These curves often appeared quite flat when using a standard five time point sampling frequency and were assigned low MISI values, indicating low muscle insulin sensitivity. However, on examining the seven time point data most of these individuals would be classified among the most insulin sensitive, as their glucose peak occurred at 15 minutes, and glucose levels had returned to near fasting concentration by 30 minutes. These individuals tended to have lower BMI than the mean for the whole data set ($24.1 \pm 3.1 \text{ kg/m}^2$) with 45 of the 47 individuals being classed as normal glucose tolerant (Supplemental Table S2.1). Conversely, there were 58 individuals (2%) with large glucose rebounds. Abdul-Ghani et al.'s original article [19] specifies that any glucose rebound should be excluded in the calculation of $\frac{dG}{dt}$, leading to some issues when automating the computation of MISI. This especially becomes problematic with more frequently sampled glucose curves, where fluctuations in the glucose curves as a result of measurement noise may introduce erroneous local minima (Figure 2.2 c). In order to compensate for this, our calculator excludes the rebound if it is less than 0.5 mmol/L greater than the minimum. If the rebound is larger than .5 mmol/L the glucose curve will be flagged; allowing the user to choose to accept the calculated value following visual inspection of the glucose curve.

MISI Calculator

The MISI calculator allows for the automated calculation of MISI in a standardised way. It consists of a graphical user interface which guides the user through uploading glucose and insulin OGTT data and specification of measurement time points (Figure 2.3). It automatically filters for missing values and converts glucose and insulin measurements to SI units. The calculator allows the user to specify which criteria they wish to flag glucose curves for manual inspection and computes MISI using the standard or modified methods. The output file contains either the calculated MISI values, or an explanation as to why the calculation of MISI is not meaningful and, where possible, a suggested MISI value for each individual. The calculator also allows the user to visually inspect the flagged glucose curves and suggested MISI values.

Discussion

Insulin resistance is often present in multiple tissues long before the onset of overt clinical disorders such as type 2 diabetes mellitus and cardio-vascular diseases. Abdul-Ghani et al. [19] proposed a score to quantify muscle-specific insulin sensitivity using five time point OGTT data (t0,30,60,90,120 min) which was validated against a hyperinsulinemic-euglycaemic clamp combined with a glucose tracer infusion.

Several studies have failed to reproduce the correlation between MISI and the clamp reported in the original study [20,21,22]; these lower correlations are typically attributed to a reduced heterogeneity in the population composition. Analysis of correlations obtained for MISI computed on randomly selected sub-populations from the PRESERVE study, a relatively homogenous population with respect to BMI and glucose tolerance status when compared to the original study population [19], with the clamp ssGIR highlighted the range of correlation coefficients which could be obtained through small variations in the population composition. Correlation for the standard MISI computed on five time point OGTT data with the clamp ssGIR for the whole data was 0.513, which is considerably lower than the 0.79 reported in the original study, but is comparable with values achieved by other studies [20,21,22]. Our proposed modified MISI method yields an improved correlation with the clamp of 0.576. While the increase in correlation is non-significant, the modified method is also numerically more robust than the original method. Application of the original index to a standard five time point OGTT would be expected to under-estimate the glucose peak and nadir, as demonstrated in Figure 2.1 b. The use of cubic splining in our modified MISI method allows for inference of the glucose peak and nadir while also correcting for unequal sampling intervals when calculating the mean insulin concentration. Spearman rank based correlation was used as the relationship between clamp ssGIR and MISI was observed to be non-linear. While the hyperinsulinemic-euglycaemic clamp performed during the PRESERVE study was not combined with a glucose tracer infusion, as the insulin infusion rate was the same as in Abdul-Ghani's original paper (40 mU/min/m²)[19] we assumed this is sufficient to fully suppress endogenous glucose production. We also acknowledge the need to further validate our modified MISI method, as well as the original index, on a range of heterogeneous populations. The composition of the PRESERVE study population, consisting only of individuals with impaired fasting glucose and/or impaired glucose tolerance, may limit the

generalisability of our results.

Comparison of MISI computed on five versus seven time point OGTT data highlighted several, previously unreported, situations in which the numerical value obtained when automating the calculation of MISI may not be biologically relevant. Abdul-Ghani addressed issues with glucose curves with a peak at 120, advising the score not be applied to individuals with type 2 diabetes mellitus [19]. Our analysis also revealed several additional, potentially problematic, curves requiring manual inspection; namely ‘flat’ glucose curves, which may arise from a failure to capture a rapid response to the oral glucose bolus due to infrequent sampling, and curves, where due to erroneous identification of the nadir, the glucose rebound may not be negligible. While more frequent sampling should improve the accuracy of MISI as it will result in improved detection of the glucose peak. However, an increased number of time points also increases the chances of detecting a premature nadir. Our calculator flags glucose curves which display potentially problematic traits such as large rebounds or modest peaks, allowing for manual inspection of the MISI value. This is particularly useful when applying MISI to large data sets, as in the case of the Maastricht study, where manual inspection of each glucose curve is prohibitive due to population size.

As part of our analyses we also evaluated the use of the oral glucose minimal model to infer the full glucose curve from measured insulin data (work not shown). The oral glucose minimal model [31] is a physiology based mathematical model which describes both insulin dependent and independent glucose dynamics following oral administration of a glucose load. Fitting of the oral glucose minimal model to the discretely sampled OGTT glucose data for each individual, using the measured insulin data as input, allowed for inference of the full glucose curve; including a prediction of the true glucose peak and nadir, as with cubic splining. However, as it was necessary to estimate a unique model parameter set for each individual in order to infer the glucose curve, this method is computationally much more heavy than cubic splining. Given the longer computation time necessary for use of the oral glucose minimal model, which becomes prohibitive for use in larger cohort studies, it was decided to use the cubic spline method.

As the authors of the original paper [19] did not provide a detailed method for the calculation of MISI there has been considerable variation in the implementation of MISI in research to date [25, 32]. As no set measurement units are specified, the MISI values reported in literature differ by several orders of magnitude [32-35]. Ergo, there is currently no cut-off as to what

MISI value can be classed as muscle insulin resistance. Currently, MISI is most often used to rank individuals into muscle insulin sensitivity tertiles. Consequently, classification of muscle insulin resistance is relative to each data set making comparison of results across studies difficult. Variation in the computational method can also be seen in the literature, with some authors choosing to remove certain time points when computing MISI. Here, we have demonstrated that the exclusion of time points can have an impact on the resulting ranking akin to the reclassification probability for the four time point subsets reported in Table 1. Spearman correlations and reclassification probabilities for each of the four time point subsets with the ssGIR of the clamp in the PRESERVE study indicate that caution should be used when applying the standard MISI to four time point OGTT sampling schedules, in particular when the 30 or 60 minute time point is absent. The higher correlation coefficients obtained between the modified MISI method and the clamp ssGIR suggests the use of cubic splining offers the modified MISI method a robustness to the limited data and unequal sampling schedules of the four time point subsets of OGTT data. Based on these results we would advise using the modified method when computing MISI on four or five time point OGTT data, and not computing MISI using either method if the 120 min time point is not available.

In order to prevent these issues and standardise the computation of MISI we have developed the MISI Calculator. This calculator provides a user friendly tool, allowing the user to compute MISI using the standard method or the modified method, which is recommended when using data with limited temporal information or with unequally spaced sampling intervals, while also standardising both the computational method and units used. In the long term, we envision this standardised computation of MISI will allow for easier comparison of MISI across studies and the identification of a cutoff value of MISI indicative of clinical muscle insulin resistance.

References

1. Ford, E.S. Risk for all-cause mortality, cardiovascular disease, and diabetes associated with metabolic syndrome: a summary of the evidence. *Diabetes Care*, 28,1769-1778 (2005).
2. Stinkens, R., Goossens, G. H., Jocken, J. W. E, Blaak, E. E., Targeting fatty acid metabolism to improve glucose metabolism. *Obes. Rev.* 16, 15–757 (2015).
3. Wilcox, G. Insulin and Insulin Resistance. *Clin. Biochem. Rev.* 26, 19–39 (2005).
4. Unnikrishnan, A. G. Tissue-specific insulin resistance. *Postgrad. Med. J.* 80, 435 (2004).
5. Rask-Madsen, C., Kahn, C. R. Tissue-specific insulin signaling, metabolic syndrome and cardiovascular disease. *Arterioscler. Thromb. Vasc. Biol.* 32, 2052–2059 (2012).
6. Goossens, G. H., Moors, C. C. M., Jocken, J. W., van der Zijl, N. J., Jans, A. et al. Altered skeletal muscle fatty acid handling in subjects with impaired glucose tolerance as compared to impaired fasting glucose. *Nutrients*. 8, 164 (2016).
7. Bruning, J. C., Michael, M. D., Winnay, J. N., Hayashi, T., Horsch, D., Accili, D. et al. A muscle-specific insulin receptor knockout exhibits features of the metabolic syndrome of NIDDM without altering glucose tolerance. *Mol. Cell.* 2, 559–569 (1998).
8. Michael, M. D., Kulkarni, R. N., Postic, C., Previs, S.F., Shulman, G. I., Magnuson, M.A. et al. Loss of insulin signalling in the hepatocytes leads to severe insulin resistance and progressive hepatic dysfunction. *Mol. Cell.* 6, 87–97 (2000).
9. Larson-Meyer, D. E., Heilbronn, L. K., Redman, L. M., Newcomer, B. R., Frisard, M.I., Anton, S. et al. Effect of calorie restriction with or without exercise on insulin sensitivity, b-cell function, fat cell size, and ectopic lipid in overweight subjects. *Diabetes care.* 29, 1337–1344 (2006).
10. Gómez-Sámano, M. A., Gulias-Herrero, A., Cuevas-Ramos, D., Brau-Figueroa, H., Mehta, R., Vargas-Gutierrez, D. et al. Metformin and improvement of the hepatic insulin resistance index independent of anthropometric changes. *Endocr. Pract.* 8, 8-16 (2012).
11. Lessard, S. J., Rivas, D. A., Chen, Z. P., Bonen, A. Febrario, M. A., Reeder, D. W. et al. Tissue-specific effects of rosiglitazone and exercise in the treatment of lipid-induced insulin resistance. *Diabetes.* 6, 1856-1864 (2007).
12. DeFronzo, R. A., Tobin, J. D., Andres, R., Glucose clamp technique: a method for quantifying insulin secretion and resistance. *Am. J. Physiol.* 237, E214-E223 (1979).
13. DeFronzo, R. A., Simonson, D., Ferrannini, E., Hepatic and peripheral insulin resistance: a common feature of type 2 (non-insulin-dependent) and type 1 (insulin-dependent) diabetes mellitus. *Diabetologia.* 23, 331-319 (1982).
14. Reijnders, D., Goossens, G. H., Hermes, G. D., Neis, E. P., van der Beek, C. M., Most, J. et al Effects of gut microbiota manipulation by antibiotics on host metabolism in obese humans: a randomized double-blind placebo-controlled trial. *Cell Metab.* 24, 63-74 (2016).
15. Matthews, D. R., Hosker, J. P., Rudenski, A. S., Naylor, B. A., Treacher, D. F., Turner, R.C. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia*, 28, 412-419 (1985).
16. Katz, A., Nambi, S. S., Mather, K., Baron, A. D., Follmann, D. A., Sullivan, G. et al. Quantitative insulin sensitivity check index: a simple, accurate method for assessing insulin sensitivity in humans. *J. Clin. Endocrinol. Metab.* 85, 2402-10 (2000).
17. Matsuda, M., DeFronzo, R. A. Insulin sensitivity indices obtained from oral glucose tolerance testing: comparison with the euglycaemic insulin clamp. *Diabetes Care.* 22, 1462-1470 (1999).
18. Stumvoll, M., Mitrakou, A., Pimenta, W., Jenssen, T., Yki-Jarvinen, H., van Haeften, T. et al. Use of the oral glucose tolerance test to assess insulin release and insulin sensitivity. *Diabetes Care.* 23, 295-301 (2000).
19. Abdul-Ghani, M. A., Matsuda, M. A., Balas, B., DeFronzo, R. A. Muscle and liver insulin

- resistance indexes derived from the oral glucose tolerance test. *Diabetes Care*. 30, 89-94 (2007).
20. Bastard, J. P., Faraj, M., Karelis, A. D., Lavasseur, J., Garrel, D., Prud'homme, D. et al. Muscle and liver insulin resistance indexes derived from the oral glucose tolerance test; Response to Abdul-Ghani et al. *Diabetes Care* 30, e83-83 (2007).
 21. Wilson, M., Ross, R. Measurement of muscle insulin sensitivity in obese men. *Diabetes Metab.* 40, 161-164 (2014).
 22. Kim, M. K. Clinical characteristics of OGTT- derived hepatic- and muscle insulin resistance in healthy young men. *J. Exerc. Nutrition Biochem.* 18, 385-92 (2014).
 23. Abdul-Ghani, M. A., Matsuda, M. A., Balas, B., DeFronzo, R. A. Muscle and liver insulin resistance indexes derived from the oral glucose tolerance test: response to bastard et al. *Diabetes Care*. 30, 2337 (2007).
 24. Bland, M. J., Altman, D. G. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 327, 307-10 (1986).
 25. Croymans, D. M., Paparisto, E., Lee, M. M., Brandt, N., Le, B. K., Lohan, D. et al Resistance training improves indices of muscle insulin sensitivity and β -cell function in overweight/obese, sedentary young men. *J. Appl. Physiol.* 115, 1245-1253 (2013).
 26. Abdul-Ghani, M. A., Matsuda, M., DeFronzo, R. A. Strong association between insulin resistance in liver and skeletal muscle in non-diabetic subjects. *Diabet. Med.* 25, 1289-1294 (2008).
 27. Schram, M. T., Sep, S. J. S., van der Kallen, C. J., Dagnelie, P. C., Koster, A., Schaper, N. et al. The Maastricht Study: an extensive phenotyping study on determinants of type 2 diabetes, its complications and its comorbidities. *Eur. J. Epidemiol.* 29, 439-51 (2014).
 28. van der Zijl, N. J., Moors, C. C., Goossens, G. H., Hermans, M. M., Blaak, E. E., Diamant, M. Valsartan improves β -cell function and insulin sensitivity in subjects with impaired glucose metabolism. *Diabetes Care*. 34, 845-851 (2011).
 29. WHO. Definition and Diagnosis of diabetes mellitus and intermediate hyperglycemia. (2006) available at <https://www.who.int/diabetes/publications/diagnosis%20diabetes2006/en/>
 30. Blanco-Rojo, R., Alcalá-Díaz, J. F., Wopereis, S., Perez-Martínez, P., Quintana-Navarro, G. M., Marin, C. et al. The insulin resistance phenotype (muscle or liver) interacts with the type of diet to determine changes in disposition index after 2 years of intervention: the CORDIOPREV-DIAB randomised clinical trial. *Diabetologia*. 59, 67-76 (2015).
 31. Cobelli, C., Dalla Man, C., Toffolo, G., Basu, R., Vella, A., Rizza, R, The oral minimal model method. *Diabetes*. 63,1203-1213 (2014).
 32. de Matos, M. A., Vieira, D. V., Pinhal, K. C., Lopes, J. F., Dias-Peixoto, M. F., Pauli, J. R. et al. High-intensity interval training improves markers of oxidative metabolism in skeletal muscle of individuals with obesity and insulin resistance. *Front. Physiol.* 9, 1451 (2018).
 33. Fiorentino, T. V., Succurro, E., Arturi, F., Giancotti, A., Peronace, C., Quirino, A. et al. Serum IgG2 levels are specifically associated with whole-body-insulin mediated glucose disposal in non-diabetic offspring of type 2 diabetic individuals: a cross-sectional study. *Scientific Reports*. 8, 13616 (2018).
 34. Bowden-Davies, K. A., Sprung, V. S., Norman, J. A., Thompson, A., Mitchell, K. A., Halford, J. C. G. et al. Short-term decreased physical activity with increased sedentary behaviour causes metabolic derangements and altered body composition: effects in individuals with and without a first degree relative with type 2 diabetes. *Diabetologia* 61, 1281-1294 (2018).
 35. Forrest, L., Sedmak, C., Skider, S., Grewal, S., Harman, S. M., Blackman, M. R., et al. Effects of growth hormone on hepatic insulin sensitivity in healthy older adults. *Endocrine* 63, 497-506 (2019).

Supplementary Material

Supplementary Table S2.1: Demographics of relevant subsets of individuals from the Maastricht Study used in this study

	Maastricht Study data set	Maastricht Study (full seven time point OGTT)	Maastricht Study (removed curves flagged at 4 time points	Flat glucose curves (5 time point OGTT)	Peak at 120 minutes (5 time point OGTT)	Large glucose rebound (5 time point OGTT)
number of individuals	3451	2631	1891	47	154	58
Age (years)	59.8 ± 8.3 (40 - 76)	59.7 ± 8.2 (40 - 76)	60.1 ± 8.1 (40 - 76)	54.2 ± 8.4 (40 - 73)	62.5 ± 7.8 (41 - 75)	57.5 ± 7.9 (41 - 74)
Sex (n, % male)	1775 50.1%	1373 52.2%	1066 56.4%	9 19.2%	69 44.8%	17 29.3%
BMI (kg/m ²)	27.1 ± 4.6 (14.4 - 52.3)	26.8 ± 4.3 (14.4 - 52.3)	27.1 ± 4.3 (14.4 - 52.3)	24.2 ± 3.1 (18.2 - 31.3)	27.4 ± 4.6 (17.9 - 41.5)	25.3 ± 3.8 (17.9 - 36.1)
Normal glucose tolerance	1924 55.8%	1584 60.2%	1164 61.6%	45 95.7%	27 17.5%	54 93.1%
Impaired fasting glucose	143 4.0%	124 4.7%	105 5.6%	1 2.1%	0 0.0%	0 0.0%
Impaired glucose tolerance	368 10.7%	305 11.6%	257 13.6%	0 0.0%	20 13.0%	3 5.2%
Type 2 diabetes	975 28.3%	616 23.4%	365 19.3%	1 2.1%	107 69.5%	1 1.7%
Type 1 diabetes	37 1.1%	- -	- -	- -	- -	- -
Other type diabetes	4 0.1%	- -	- -	- -	- -	- -
Oral glucose lowering	716 20.8%	446 16.9%	254 13.4%	0 0.0%	89 57.8%	1 1.7%
Insulin use	257 7.5%	- -	- -	- -	- -	- -

For Age and BMI both the mean ± the standard deviation are shown along with the range of values from minimum to maximum. For other variables the number of relevant individuals and the percentage of total population are supplied. Column one: full data set of 3451 participants, who completed the baseline survey between November 2010 and September 2013. Columns two: removing individuals who either did not undergo an OGTT, due to use of exogenous insulin or fasting glucose ≥ 11 mmol/L (n=300), or had one or more missing values for glucose or insulin during the OGTT (n=520) full seven time point OGTT data is available for 2631 individuals. This subset of individual is used to evaluate the location of the glucose peak in five versus seven time point OGTT data. Column three: all individuals with glucose curves which were flagged on four time point data were removed resulting in a final data set of 1891 individuals used for comparing reclassification probabilities with MISI ranking derived on five and seven time point OGTT data. Columns four through six: subsets of individuals whose glucose curves were flagged using five time point data.

Supplementary Table S2.2 : Demographics of participants from the PRESERVE Study

	PRESERVE Study (including outliers)		PRESERVE Study (excluding outliers)	
Numer of individuals	71		69	
Age (years)	58.3 ± 6.9 (38 - 69)		58.3 ± 6.9 (38 - 69)	
Sex (% male)	54.9%		55.1%	
BMI (kg/m ²)	29.8 ± 4.5 (22.7 - 44.4)		29.8 ± 4.5 (22.7 - 44.4)	
Impaired fasting glucose (IFG)	39	54.9%	37	53.6%
Impaired glucose tolerance (IGT)	11	15.5%	11	15.9%
Both (IFG & IGT)	21	29.6%	21	30.4%

Demographics for participants used from the PRESERVE Study, both including and excluding two MISI score outliers. Age and BMI are given in mean ± standard deviation. Remaining variables indicate the glucose tolerance status of the participants described by the number of individuals and percentage of total population.

Supplementary Table S2.3 : Demographics of highest and lowest muscle insulin sensitivity tertile defined by standard and modified MISI

Reclassification of muscle insulin sensitivity tertiles (Maastricht Study n=1891)				
	Low muscle insulin sensitivity			
	Five time points		Seven time points	
	Original MISI	Modified MISI	Original MISI	Modified MISI
BMI	28.4 ± 4.6	28.5 ± 4.6	28.5 ± 4.7	28.6 ± 4.5
HOMA-IR	4.11 ± 3.22	4.21 ± 3.24	4.18 ± 3.23	4.25 ± 3.23
AGE (years)	60.6 ± 8.1	60.6 ± 8.0	60.4 ± 8.2	60.6 ± 8.1
# NGT	361 (57.3%)	348 (55.2%)	357 (56.7%)	348 (55.2%)
# IFG	30 (4.8%)	32 (5.1%)	29 (4.6%)	32 (5.1%)
# IGT	140 (22.2%)	141 (22.4%)	144 (22.9%)	136 (21.6%)
# T2DM	99 (15.7%)	109 (17.3%)	100 (15.9%)	114 (18.1%)

	Highest muscle insulin sensitivity			
	Five time points		Seven time points	
	Original MISI	Modified MISI	Original MISI	Modified MISI
BMI	26.0 ± 3.9	25.8 ± 3.8	26.0 ± 3.9	25.7 ± 3.7
HOMA-IR	2.21 ± 1.56	2.13 ± 1.45	2.21 ± 1.61	2.13 ± 1.52
AGE (years)	59.9 ± 8.1	59.9 ± 8.0	59.9 ± 8.0	59.9 ± 7.9
SEX (% female)	39.1%	40.0%	39.9%	41.1%
# NGT	389 (61.8%)	402 (63.8%)	383 (60.8%)	410 (65.1%)
# IFG	44 (7.0%)	41 (6.5%)	40 (6.4%)	42 (6.7%)
# IGT	32 (5.1%)	31 (4.9%)	40 (6.4%)	29 (4.6%)
# T2DM	166 (26.4%)	157 (24.9%)	168 (26.7%)	150 (23.8%)

Demographics of highest and lowest muscle insulin sensitivity tertiles defined by both standard MISI and modified MISI calculated on five time point (0,30,60,90,120 minute) and seven time point (0,15,30,45,60,90,120) OGTT data from the Maastricht Study. BMI, HOMA, and AGE are given as the mean value for the tertile ± the standard deviation. Sex is given as percentage male. Glucose tolerance status is indicated as the number of individual of the glucose tolerance status and percentage of the full grouping. Glucose tolerance statuses are as follows; normal glucose tolerance (NGT), impaired fasting glucose (IFG), impaired glucose tolerance (IGT), and type 2 diabetes mellitus (T2DM). Each tertile consists of 630 individuals.

3

A computational model of postprandial adipose tissue lipid metabolism derived using human arteriovenous stable isotope tracer data

Shauna O'Donovan^{1,2}, Michael Lenz^{1,3,4}, Roel G Vink⁵, Nadia J.T. Roumans⁵, Theo M.C.M. de Kok^{1,6}, Edwin C.M. Mariman⁵, Ralf L.M. Peeters^{1,9}, Natal A.W. van Riel^{1,10}, Marleen A. van Baak⁵, Ilja C.W. Arts^{1,4,5}.

¹Maastricht Center for Systems Biology (MaCSBio), Maastricht University. ²Division of Human Health and Nutrition, Wageningen University. ³Institute of Organismic and Molecular Evolution, Johannes Gutenberg University Mainz. ⁴Preventive Cardiology and Preventative Medicine - Center for Cardiology, University Medical Center of the Johannes Gutenberg University Mainz. ⁵Dept. of Human Biology, NUTRIM School of Nutrition and Translational Research in Metabolism, Maastricht University. ⁶Dept. of Toxicogenomics, GROW School of Oncology and Developmental Biology, Maastricht University. ⁷Dept. of Data Science & Knowledge Engineering, Maastricht University. ⁸Dept. of Biomedical Engineering, Eindhoven University of Technology. ⁹Dept. Epidemiology, CARIM School for Cardiovascular Disease, Maastricht University.

PLOS Computational Biology (2019) 15(10): e1007400.

Abstract

Given the association of disturbances in non-esterified fatty acid (NEFA) metabolism with the development of Type 2 Diabetes and Non-Alcoholic Fatty Liver Disease, computational models of glucose-insulin dynamics have been extended to account for the interplay with NEFA. In this study, we use arteriovenous measurement across the subcutaneous adipose tissue during a mixed meal challenge test to evaluate the performance and underlying assumptions of three existing models of adipose tissue metabolism and construct a new, refined model of adipose tissue metabolism. Our model introduces new terms, explicitly accounting for the conversion of glucose to glyceraldehyde-3-phosphate, the postprandial influx of glycerol into the adipose tissue, and several physiologically relevant delays in insulin signalling in order to better describe the measured adipose tissues fluxes. We then applied our refined model to human adipose tissue flux data collected before and after a diet intervention as part of the Yoyo study, to quantify the effects of caloric restriction on postprandial adipose tissue metabolism. Significant increases were observed in the model parameters describing the rate of uptake and release of both glycerol and NEFA. Additionally, decreases in the model's delay in insulin signalling parameters indicates there is an improvement in adipose tissue insulin sensitivity following caloric restriction.

Introduction

The adipose tissue plays a key role in the regulation of plasma triglyceride and non-esterified fatty acid (NEFA) concentrations [1,2]. In the fasting state, lipolysis of triglycerides stored in the adipose tissue delivers NEFA to the plasma, where it can be taken up by other tissues, including skeletal muscle, heart, and liver [3]. In the postprandial state dietary derived chylomicron-triglycerides are removed from the circulation by insulin stimulated hydrolysis by lipoprotein lipase (LPL) at the endothelial cell wall [4]. NEFA released by LPL lipolysis is primarily taken up by the adipose tissue and may subsequently be stored as triglyceride [5]. Disturbances in the regulation of lipid metabolism by the adipose tissue can lead to raised plasma NEFA concentrations resulting in ectopic fat deposition. The accumulation of fat in insulin sensitive tissues such as skeletal muscle and liver has been related to the development of insulin resistance and, consequently, increases the risk for developing Type 2 Diabetes and cardiovascular disease [6-9]. The contribution of disturbed NEFA metabolism to the development of insulin resistance highlights the need to include adipose tissue NEFA metabolism when investigating disturbances in the glucose-insulin regulatory system during disease development [10].

Several computational models exist which describe adipose tissue NEFA metabolism; from smaller models focusing on NEFA dynamics between the plasma and adipose tissue alone [11,12] to larger models describing several metabolites across multiple tissues [10-14]. Typically, these models have been parameterised using values reported in literature and their performance assessed by comparison with measured plasma metabolite concentrations during an oral glucose or mixed meal challenge test. However, with traditional venous sampling it is not possible to differentiate between the contributions of individual tissues (i.e. liver or skeletal muscle) to the whole body plasma metabolite concentration. Consequently, it is not possible to fully assess the structure of the model and assumptions upon which the model has been constructed. Therefore, more detailed measurements are needed, providing information about contributions of individual tissues to postprandial changes in plasma NEFA concentrations.

In recent years, new quantitative knowledge has been gained on human adipose tissue NEFA dynamics in both the fasting and postprandial state through the use of stable isotope tracers coupled with arteriovenous sampling [5,15,16]. The simultaneous sampling of blood from an artery

(or arterialised hand vein) and a vein draining a specific tissue depot (e.g. the abdominal subcutaneous adipose tissue) allows for the calculation of metabolite fluxes across this tissue. In combination with the use of stable isotope tracers, it is now possible to further untangle postprandial lipid metabolism and to quantify rates of appearance of triglyceride in the plasma, rates of LPL lipolysis of lipoprotein derived triglyceride, and spill-over of LPL derived NEFA into the plasma, while also providing insights into the behaviour of secondary metabolites such as glycerol.

Arteriovenous measurements provide invaluable information in the evaluation and refinement existing models of adipose tissue fatty acid metabolism. One example where such data has been successfully used is the 2008 study of Kim et al. [12], who modelled adipose tissue metabolism with a focus on the regulation of lipolysis. However, their model focused primarily on the fasting state, as well as the effect of an epinephrine infusion, and did not consider the postprandial state nor the influence of insulin.

In the present study, we used human *in vivo* measurements of net triglyceride, NEFA, glucose, and glycerol fluxes across the abdominal subcutaneous adipose tissue along with arterial insulin collected in the fasting and postprandial state during a high fat mixed meal challenge test to evaluate three existing models of postprandial adipose tissue dynamics. The inclusion of a palmitate stable isotope tracer in the meal provides a first opportunity to evaluate model terms describing the postprandial spill-over of LPL derived NEFA on human data. Ultimately, as no existing model could sufficiently describe the calculated metabolite fluxes, we construct a new refined model of postprandial adipose tissue metabolism which could be parameterised by data; introducing novel terms explicitly accounting for the conversion of glucose to G-3-P for use in re-esterification and the postprandial uptake of glycerol into the adipose tissue. The resulting model was then used to quantitatively estimate the impact of caloric restriction induced weight loss on subcutaneous adipose tissue fatty acid dynamics in a population of sixteen overweight or obese individuals that participated in the Yoyo study [17]. The model also allows for the prediction of the dynamic response of several reactions, which were not directly measured, to the ingestion of a meal (e.g. uptake and release of NEFA and glycerol by the adipose compartment), providing additional insights into the alterations in adipose tissue metabolism following caloric restriction which contribute to the observed changes in the calculated adipose tissue fluxes.

Materials and methods

Study design and work-flow

The study design and work-flow is visualised in Figure 3.1 and detailed below. In summary, three existing models of postprandial adipose tissue glucose, fatty acid, and triglyceride dynamics were evaluated using arteriovenous measurements across the abdominal subcutaneous adipose tissue. Subsequently, a refined model was constructed by either selecting the best fitting model term of the three existing models or, in cases where none of the existing models allowed for an adequate fit to the data, new or modified physiological mechanisms were introduced. Finally, model parameters were estimated before and after weight loss in the Yoyo study [17] and compared, taking their confidence intervals into account.

Yoyo study and arteriovenous measurements

The Yoyo study is a dietary intervention study designed to investigate the effect of the rate of weight loss (fast or slow) on subsequent weight regain [18]. Sixty-one overweight ($\text{BMI} > 25\text{kg/m}^2$) and obese ($\text{BMI} > 30\text{kg/m}^2$) individuals ($\text{BMI} 28\text{-}35\text{ kg/m}^2$) were randomly assigned to one of two diet interventions designed to achieve a weight loss of approximately 10%, a low calorie diet of 1250 kcal/day over twelve weeks (slow weight loss) or a very low calorie diet of 500 kcal/day for five weeks (rapid weight loss). Following the diet intervention all participants underwent a four week weight stabilisation period with a diet based on the energy requirements of each participant. As part of this study, sixteen individuals underwent a high fat mixed-meal challenge test (milk shake) containing 100 mg of [$\text{U-}^{13}\text{C}$] palmitate tracer at baseline and following the weight stabilisation period [17]. Fasting samples were collected at -30 and 0 minutes. The meal was consumed at 0 minutes, participants were asked to consume the shake within 10 minutes, and samples were collected at 60, 120, 180, 240, 300 minutes postprandially from an arterialised dorsal hand vein and the superficial epigastric vein draining the abdominal subcutaneous tissue. Abdominal subcutaneous adipose tissue blood flow was measured using the ^{133}Xe washout technique [19,20].

Figure 3.1: Work-flow for comparison of existing model terms and construction of refined model

Jelic, Pratt, and Sips model terms are compared for each metabolite flux separately using the postprandial arteriovenous tracer measurements from the Yoyo study. In addition, as the refined model is constructed with the aim that parameters could be estimated directly from measured data, existing model terms are also compared using identifiability analysis, sensitivity analysis, and the Akaike Information Criterion. The refined model of adipose tissue metabolism is constructed either by selecting the best fitting of the existing model terms, or by introducing novel physiological mechanisms. The resulting refined model is then parameterised using data from both the baseline and following weight stabilisation from the Yoyo study, and the values compared.

Ethics statement

All subjects gave their written informed consent before participation in the Yoyo Study. The Yoyo Study was performed according to the Declaration of Helsinki and was approved by the Medical Ethics Committee of Maastricht University Medical Centre (METC 11-3-066) with the approval number NL38099.068.11

Calculations

Individual metabolite fluxes across the abdominal subcutaneous adipose tissue were computed for NEFA, triglyceride, glucose, and glycerol by multiplying the arteriovenous plasma concentration difference of each measured metabolite by the adipose tissue blood flow [17].

Plasma tracer concentrations in NEFA and TG fractions were computed as described in the original study [17]. The rate of fractional spill-over of NEFA derived from LPL lipolysis of chylomicron triglyceride was calculated as one minus the rate of fractional extraction as described by Bickerton et al. [5]

Two indices of insulin resistance were calculated using the arterial metabolite measurements. HOMA-IR is a measure of whole body insulin resistance, and is calculated as fasting plasma glucose (mmol/l) times the fasting plasma insulin (μ U/ml) divided by 22.4 [21]. Several studies have evaluated appropriate HOMA-IR cut-off values for determining insulin resistance, with identified cut-off values ranging from 1.85 to 2.01 depending on the study population [22]. ADIPO-IR is a surrogate measure of adipose tissue specific insulin resistance and is calculated as fasting plasma NEFA concentration (mmol/l) times the fasting plasma insulin concentration (pmol/l) [23].

Existing models

Models of postprandial lipid metabolism were assessed to identify models that described insulin mediated adipose tissue specific dynamics of glucose, triglyceride, NEFA, and/or glycerol. Three published models describing complementary aspects of adipose tissue metabolism were included in the analysis, namely the Jelic [11], Pratt [13], and Sips [10] models.

The Pratt Model is a large, multi-compartmental, computational model describing the dynamics of several metabolite species including glucose, NEFA, glycerol, pyruvate, and both endogenous and dietary triglyceride in the fasting and postprandial states in plasma, skeletal muscle, adipose tissue, and liver [13]. The Jelic model is a two compartment, physiology-based, mathematical model describing NEFA dynamics across the plasma and a lumped interstitial adipose space in the postprandial state [11]. The Jelic model also accounts for two physiologically relevant delays in insulin signalling not present in the Pratt model. The model parameters for both the Jelic and Pratt models have been taken from literature and both models

have been validated using oral glucose tolerance test (OGTT) and mix meal challenge test data in lean and abdominally obese individuals. The Sips model [10] is an extension of the Dalla-Man model [24] of postprandial glucose-insulin interplay to include NEFA kinetics. Unlike the previous two models, all parameters in the Sips model have been estimated from data. Where possible parameter values were maintained at the values provided for the lean healthy individual in the Dalla-Man model, as these parameters had been validated using gold standard triple-glucose tracer data [24]. The remaining 21 parameters were estimated from experimental data consisting of plasma measurements of several metabolites from individuals under various clamp conditions [25,26] and also a frequently sampled oral glucose tolerance test (OGTT) and oral lipid tolerance test [10].

Model evaluation and refinement

The three existing models (Jelic, Pratt, Sips) were decomposed into relevant sub-units describing each measured metabolite's dynamics across and within the adipose tissue using a so-called divide and conquer approach [27] (Figure 3.1, Supplementary Table S3.1). These model terms were then evaluated by comparison to the calculated metabolite flux from the arteriovenous data (average $\mu = (\mu_1, \dots, \mu_7)$ and standard deviation $\sigma = (\sigma_1, \dots, \sigma_7)$) according to the following error function $C(p)$:

$$C(p) = \sum_{i=1}^7 \left(\frac{M(p, t_i) - \mu_i}{\sigma_i} \right)^2 \quad (3.1)$$

Here $M(p, t_i)$ is the model prediction at time t_i ,
 $t = (-30, 0, 60, 120, 180, 240, 300)$, for a parameter set p .

Optimal parameter sets for each term were obtained through non-linear regression, by finding the parameter set which best described the measured flux, minimising the error term $C(p)$. Optimisation of parameters was performed using *lsqnonlin* (MATLAB 2014b, The MathWorks Inc., Natick, Massachusetts, United States) a local, gradient-based least square solver. The parameter values supplied in the original model studies were used as initial values for the parameter search. Parameterisation of the final, refined model for comparison of baseline and following weight stabilisation was performed using a combination of a global and local search algorithm. Controlled Random Search [28] with 250 randomly selected initial parameter sets was utilised to search the parameter space in order to provide a good initial value for *lsqnon-*

lin.

Terms describing the uptake and release of metabolites from tissues other than the adipose tissue, for example appearance from the gut and insulin secretion, are not described in this model. Available measured arterial concentrations of insulin, glucose, NEFA, triglyceride, and glycerol are supplied as dependent inputs to the model terms [29].

In the cases when the existing models were not capable of describing the experimentally estimated fluxes, underlying assumptions and model terms were evaluated based on existing biological knowledge and modified accordingly to provide an improved description of postprandial metabolite dynamics across the adipose tissue.

Additionally, all model terms were examined statistically, using the Akaike Information Criterion (AIC_c) corrected for small sample sizes [30], to select the most parsimonious model that is both biologically sound and can adequately describe the measured flux data.

The above procedure was performed to evaluate terms describing the triglyceride, glucose, glycerol, and NEFA fluxes across the adipose tissue. Terms describing the fractional spill-over of NEFA by LPL lipolysis could also be assessed for the first time using the $[U-^{13}C]$ palmitate tracer data.

Parameter identifiability

Model terms were also evaluated for parameter identifiability, as a primary objective of this study was to provide a model that could be parameterised by experimental data and used to quantify adipose tissue metabolism from meal challenge test data. Parameter identifiability was evaluated using Profile Likelihood Analysis [31], whereby one parameter was varied iteratively from its optimal value and the remaining parameters were re-estimated. For an identifiable parameter the error measure $C(p)$ would be expected to increase as the parameter value deviated from its optimum [31,32]. 95% confidence intervals for the estimated parameter values have been estimated using the parameter value covariance matrix approximated using the Jacobian matrix provided as output of the lsqnonlin algorithm used in the parameter estimation procedure.

Figure 3.2 : Structure of refined mathematical model of adipose tissue metabolism

The refined model of adipose tissue metabolism consists of a two compartmental model, describing dynamics between the plasma and a lumped interstitial adipose space. The insulin stimulated LPL lipolysis of circulating triglycerides releases glycerol and NEFA. The hydrolysed NEFA passes into the adipose space, with an insulin-dependent fraction spilling over into the plasma. The insulin inhibited lipolysis of the triglycerides stored within the adipose tissue releases NEFA and glycerol. Glucose passes into the adipose space at an insulin dependent and independent rate. Glucose is converted to G-3-P and provides the glycerol backbone necessary for re-esterification of NEFA for storage as triglyceride within the adipose space. Novel model terms, introduced in this analysis, are shown in red, existing models are shown in black. Reactions that are stimulated by insulin are depicted with a green insulin symbol. Reactions inhibited by insulin are shown with a red symbol.

Results

The three existing models of postprandial adipose tissue metabolism implemented in this study were not capable of describing the calculated postprandial adipose tissue metabolite fluxes. Consequently, a refined model of adipose tissue postprandial metabolism was constructed. The refined model is formulated as a two compartment model consisting of a

blood plasma compartment and a lumped interstitial adipose compartment reflecting the sampling method used in the arteriovenous data (Figure 3.2). In the following, we show the results of the model evaluation and refinement separately for each measured metabolite using the adipose tissue flux data before the weight loss intervention.

Triglyceride flux

Circulating plasma triglyceride is hydrolysed by LPL at the endothelial wall releasing NEFA and glycerol. The Pratt model describes this reaction as being proportional to the concentration of triglyceride present in circulating lipoproteins (both chylomicron and VLDL) and occurring at a basal and insulin stimulated rate using linear terms [13], with the plasma insulin concentration having an instantaneous effect. The Jelic model accounts for the saturation of enzyme mediated LPL lipolysis using Michaelis-Menten kinetics, with the rate of hydrolysis of circulating triglyceride being dependent on the plasma triglyceride concentration and stimulated by a delayed insulin effect [11]. The transcription of LPL and subsequent secretion of the LPL protein to the endothelial wall are known to be stimulated by insulin, however, these processes take some time [4]. As a result, the Jelic model introduced a three-fold insulin delay term, with the time delay parameter set to 240 mins [11]. The Sips model makes use of the Jelic model term and parameters [10].

Using the parameter values as specified in the original publications, neither the Jelic nor Pratt models are capable of describing the mean measured triglyceride flux across the adipose tissue (Figure 3.3 A). Allowing the parameters to be estimated from the data, the Jelic model provides an improved fit. A more detailed investigation demonstrated that the insulin delay is the essential component missing in the Pratt model (Supplementary Figures S3.2 S3.3). As a result, the optimised triglyceride flux term ($[TG]_{flux}$) shown below was derived, making use of the insulin dependent linear approximation from the Pratt model while introducing the insulin delay term of the Jelic model.

$$TG_{flux} = K_{ad}[TG_{art}][I_{LPL}] \quad (3.2)$$

Where K_{ad} is the rate parameter for LPL lipolysis, $[TG_{art}]$ is the arterial triglyceride concentration and $[I_{LPL}]$ is the delayed insulin signal, modelled with a three compartmental delay (Eq [3.12-3.15]) with a time delay

parameter τ_{LPL} also estimated from data.(Figure 3.4)

Fractional spill-over

NEFA released by LPL lipolysis of circulating triglyceride is taken up by the adipose tissue. However, this process is inherently leaky with a proportion of the released NEFA spilling-over directly into the plasma [33,34]. Arteriovenous studies combined with NEFA stable isotope tracers

Figure 3.3 : Comparison of the Jelic, Pratt, Sips, and refined model fit to baseline flux data from the Yoyo Study

Model terms from the Pratt (green), Jelic (blue), Sips (yellow) simulated using the parameter values provided in their respective publications and refined model (red) describing the (A) postprandial LPL mediated lipolysis of circulating triglycerides, (B) the fractional spill over of LPL derived NEFA, (C) the uptake of glucose into the adipose tissue, (D) the glycerol and (E) NEFA efflux from the abdominal subcutaneous adipose tissue are shown. Metabolite fluxes are calculated as the arteriovenous difference in a metabolite across the adipose tissue multiplied by the adipose tissue blood flow. The mean calculated adipose tissue fluxes are shown with the black crosses $\pm$ the standard error of the mean. Negative flux values indicate a net release of the metabolite from the adipose space, positive values indicate a net uptake.

have previously demonstrated that the fractional spill-over of NEFA from LPL lipolysis increases later in the postprandial period [5,33]. The Jelic model does not describe the spill-over of NEFA derived from LPL lipolysis [11]. The Pratt model describes fractional spill-over at a constant rate of 25% [13]. The Sips model describes the fractional spill-over of LPL derived NEFA as occurring at a basal and insulin inhibited rate, using the same delayed insulin signal as LPL lipolysis with a 240 min time delay parameter [10].

The Sips term could not describe the increasing postprandial fractional spill-over measured in the Yoyo Study data (Figure 3.3 B). The 240 min insulin time delay served to over-damp the insulin signal. As a result, the insulin delay was removed. Further analysis of the Sips term with use of the Akaike Information Criterion led to the removal of the basal fractional spill-over rate, yielding the following optimised fractional spill-over term.

$$\% \text{spill-over} = \frac{1}{100} \left(D_{\text{spill}} \frac{I_B}{[I_{\text{art}}]} \right) \quad (3.3)$$

Here I_B is the basal plasma insulin concentration, $[I_{\text{art}}]$ the arterial insulin concentration and D_{spill} an estimated constant (Figure 3.4 B).

Glucose flux

Glucose enters the adipose tissue along the concentration gradient facilitated by insulin dependent (GLUT4) and insulin independent (GLUT1) transporters [35]. Within the adipocyte, glucose is quickly converted to glucose-6-phosphate (G-6-P) which serves to trap the glucose within the adipocyte. Glucose is the primary source of the glyceraldehyde-3-phosphate (G-3-P) backbone needed for re-esterification of adipocyte NEFA [36].

Both the Pratt [13] and the Sips model [10] describes the uptake of glucose by the adipose tissue as occurring at an insulin dependent and independent rate. However, the Pratt model uses direct plasma insulin stimulation rather than accounting for a delay in the effect of insulin signalling which induces the translocation of GLUT4 transporters from the transport vesicles of the cell to the membrane. The Jelic model does not account for glucose dynamics. With the introduction of a time delay in insulin stimulation the linear term from the Pratt model was capable of describing the glucose flux. In addition, it was necessary to explicitly account for the conversion of glucose to G-6-P in order to remedy an erroneous prediction of a postprandial glucose efflux

from the interstitial adipose space (Figure 3.3 C).

$$G_{flux} = -GLUT1[G_{art}] - GLUT4[G_{art}][I_{AT}] \quad (3.4)$$

Where $[G_{art}]$ is the arterial glucose concentration, $[I_{AT}]$ is the adipose tissue delayed insulin signal described using a threefold delay with a time delay parameter τ_{AT} (Eq. 3.12-3.15), GLUT1 and GLUT4 are constants estimated from the data (Figure 3.4 C).

Figure 3.4 : Fitting of the refined model to adipose tissue flux data measured at baseline and following caloric restriction from the Yoyo Study

The refined model was fit to postprandial adipose tissue flux measurements of (A) LPL mediated lipolysis of circulating triglycerides, (B) the fractional spill over of LPL derived NEFA, (C) influx of glucose, and efflux of both (D) glycerol, and (E) NEFA to the adipose tissue at baseline and following a period of caloric restriction. Blue crosses indicate the mean adipose tissue flux at baseline, the red crosses show the mean flux following caloric restriction, error bars indicate the standard error of the respective means. Refined model prediction at baseline is shown in blue, and the model fit following the diet intervention is shown in red. Negative flux values indicate a net release of the metabolite from the adipose space, positive values indicate a net uptake.

Glycerol flux

Glycerol is released into the plasma by LPL lipolysis of circulating triglyceride at the endothelial wall [35]. Glycerol is also released by lipolysis of triglyceride stored within the adipose tissue (denoted as ATL lipolysis [37]). It is commonly assumed that due to the inactivity of glycerol kinase in the adipose tissue [38], all glycerol released by ATL lipolysis enters the plasma for transportation to the liver. Postprandial glycerol dynamics are not accounted for in the Jelic [11] nor Sips models [10]. The Pratt model accounts for the release of glycerol by ATL lipolysis within the adipose tissue which appears directly in the liver compartment [13]. Glycerol release into plasma by LPL lipolysis is not described in this model.

Analysis of the postprandial glycerol and triglyceride fluxes from the YoYo Study indicate the uptake of glycerol into the interstitial adipose space in the postprandial state (Figure 3.5). As a result, the Pratt glycerol term was extended to account for the concentration gradient diffusion of plasma glycerol into and from the interstitial adipose compartment.

$$GLY_{flux} = K_{ad}[TG_{art}][I_{LPL}] + p_{GLY}([GLY_{AT}] - ([GLY_{art}] + K_{ad}[TG_{art}][I_{LPL}])) \quad (3.5)$$

$$\frac{d[GLY_{AT}]}{dt} = -p_{GLY}([GLY_{AT}] - [GLY_{art}] + K_{ad}[TG_{art}][I_{LPL}]) + B_{ATL} + \frac{ATL_{max}}{1 + \frac{[I_{AT}]}{K_{ATL}}} \quad (3.6)$$

Where ATL lipolysis is described as occurring at a basal (B_{ATL}) and saturable insulin inhibited rate using terms from the Jelic model (ATL_{max} , K_{ATL}) [11]. Where $[I_{AT}]$ is the delayed interstitial adipose compartment insulin signal as described for the glucose flux above. (Figure 3.3 D, Figure 3.4 D)

NEFA flux

In the fasting state NEFA is released by the lipolysis of stored triglyceride within the adipocyte and passes into the plasma for delivery to other tissues. The lipolysis of stored triglyceride is inhibited by increasing insulin concentration in the postprandial state [39,40] which also stimulates the rate of re-esterification of free NEFA within the adipocyte for storage as triglyceride [41]. In addition, insulin stimulates the LPL lipolysis of

circulating triglyceride, removing excess dietary triglyceride from the plasma. While the majority of NEFA released by LPL lipolysis passes into the adipocyte for re-esterification a fraction spills-over directly into the plasma [33,34].

Terms describing the fractional spill-over, LPL and ATL lipolysis have previously been derived. Terms describing the rate of re-esterification were compared using the Yoyo study NEFA flux. The Pratt model describes the rate of re-esterification as a linear term, directly stimulated by plasma insulin and adipose glucose concentration [13]. The Jelic model uses saturable kinetics coupled with a delayed insulin stimulation [11]. Both models used passive diffusion of NEFA between the plasma and interstitial adipose compartment. The Sips model does not explicitly describe the adipose compartment, instead using a linear term which accounts for net lipolysis and re-esterification of NEFA within the adipose tissue [10]. While the

Figure 3.5 : Postprandial uptake of glycerol by adipose space

(A) Mean postprandial triglyceride influx (blue) (equivalent to the release of glycerol by LPL lipolysis (red)) and total glycerol efflux (yellow) across the abdominal subcutaneous adipose tissue at baseline in the Yoyo study ($\pm$ standard error of the mean). (B) Mean ($\pm$ standard error of the mean) of glycerol flux minus triglyceride flux in postprandial phase measured at baseline in the Yoyo study indicating a net uptake of glycerol by the abdominal subcutaneous adipose tissue in the postprandial condition (green shaded region). This is in contradiction to the commonly held assumption that the glycerol efflux equals the direct sum of glycerol released by LPL lipolysis of circulating triglyceride and glycerol released by ATL lipolysis within the adipose tissue.

Jelic term could produce a good description of the NEFA flux when fit to the data (Supplementary Fig S3.2), it assumed an unlimited supply of G-3-P and required the estimation of five parameters from the data (Supplementary table S3.1). The linear re-esterification term of the Pratt model was modified to include the delayed interstitial adipose compartment insulin signal. In addition, terms explicitly accounting for the production of the G-3-P backbone necessary for re-esterification from glucose taken into the adipose tissue were introduced, providing a sufficient description of the measured postprandial NEFA flux (Figure 3.3 E, Figure 3.4 E).

$$NEFA_{flux} = \frac{3}{100} \left(D_{spill} \frac{I_B}{I_{art}} \right) K_{ad} [TG_{art}] [I_{LPL}] - p_{NEFA} ([NEFA_{PL}] - [NEFA_{AT}]) \quad (3.7)$$

$$\begin{aligned} \frac{d[NEFA_{AT}]}{dt} = & \frac{3}{100} \left(1 - D_{spill} \frac{I_B}{I_{art}} \right) K_{ad} [TG_{art}] [I_{LPL}] + p_{NEFA} ([NEFA_{PL}] - [NEFA_{AT}]) \\ & + 3 \left(B_{ATL} + \frac{ATL_{max}}{1 + \frac{[I_{AT}]}{K_{ATL}}} \right) - 3 (K_{reester} [I_{AT}] [NEFA_{AT}] [G-3-P_{AT}]) \end{aligned} \quad (3.8)$$

$$\text{With } [NEFA_{PL}] = [NEFA_{art}] + \frac{3}{100} \left(D_{spill} \frac{I_B}{I_{art}} \right) K_{ad} [TG_{art}] [I_{LPL}]$$

Where the adipose tissue concentration of G-3-P ($[G-3-P_{AT}]$) calculated using the following equations.

$$\frac{d[G-3-P_{AT}]}{dt} = [G-3-P_{pro}] - K_{reester} [I_{AT}] [NEFA_{AT}] [G-3-P_{AT}] \quad (3.9)$$

$$\frac{d[G-6-P]}{dt} = \frac{1}{\tau_{G-3-P}} (2 \text{frac}_{use} (GLUT1 [G_{art}] + GLUT4 [G_{art}] [I_{AT}]) - [G-6-P]) \quad (3.10)$$

$$\frac{d[G-3-P_{pro}]}{dt} = \frac{1}{\tau_{G-3-P}} ([G-6-P] - [G-3-P_{pro}]) \quad (3.11)$$

Re-esterification occurs at an insulin stimulated linear rate $K_{reester}$ dependent on the available concentrations of G-3-P and NEFA. A portion of the glucose taken up by the adipose space (frac_{use}) is converted to G-6-P and then to adipose tissue G-3-P ($[G-3-P_{AT}]$) which will be available for use in re-esterification, the remaining glucose leaves the system for use by other cellular functions which are not explicitly described here. The production of adipose tissue G-3-P is described by the term $[G-3-P_{pro}]$ with a two compartmental delay governed by the delay constant τ_{G-3-P} .

Insulin delays

Insulin stimulates the translocation, secretion, and in some cases transcription of several enzymes and transport proteins involved in adipose tissue metabolism. These processes occur over time spans of several minutes to several hours. The effectiveness of the measured plasma insulin signal may be further damped by the presence of insulin resistance, a condition where the responsiveness of tissues to plasma insulin is reduced. Rather than describing the full sequence of reactions involved the dampening of the plasma insulin signal is approximated using a three compartmental delay, as in the Jelic model [11].

$$\frac{d[I_1]}{dt} = \frac{1}{\tau}([I_{art}] - [I_1]) \quad (3.12)$$

$$\frac{d[I_2]}{dt} = \frac{1}{\tau}([I_1] - [I_2]) \quad (3.13)$$

$$\frac{d[I_{delay}]}{dt} = \frac{1}{\tau}([I_2] - [I_{delay}]) \quad (3.14)$$

Where $[I_{art}]$ is the measured arterial insulin concentration and τ is the respective time delay parameter. Two insulin time delays are simulated, a long time delay for LPL lipolysis ($\tau = \tau_{LPL}$) and a shorter interstitial adipose compartment time delay ($\tau = \tau_{AT}$). The structure of the insulin delay terms are not altered from the original model [11]. However, the time delay parameters are not fixed, as in the Jelic and Sips models, and are estimated from the data.

Application of adipose model - quantification of caloric-restriction

The resulting refined model was then used to quantify the effect of caloric restriction using meal challenge test data collected before and after a diet intervention as part of the Yoyo study.

Parameter values before and after caloric restriction

Parameter sets were estimated from data collected at baseline and after a weight stabilisation period following caloric restriction (Table 1, the complete parameter sets are provided in Supplementary Table S2). The rate parameters for both glycerol and NEFA concentration gradient based

diffusion from the plasma to the adipose space (P_{GLY} and P_{NEFA}) increase significantly following caloric restriction, indicating an increase in the rate of glycerol and NEFA transport in and out of the adipose tissue. A large, although not significant, decrease can be seen in the insulin delay parameter for LPL lipolysis of circulating triglyceride, τ_{LPL} , (157 mins at baseline to 113 mins following weight stabilisation) which is not accompanied by a strong change in the value for K_{ad} , the rate parameter for that term. In addition, there is a four minute decrease in the adipose tissue insulin delay parameter τ_{AT} . The reduction in the insulin delay parameters indicates a less damped response to insulin signalling following caloric restriction, indicative of improved insulin sensitivity.

Decomposition of glycerol and NEFA model predictions

The estimated values for P_{GLY} and P_{NEFA} increased significantly following caloric restriction, but it remained unclear if the observed differences in the glycerol and NEFA fluxes following caloric restriction (Figure 3.4) could be entirely attributed to these increases in the rates of uptake and release of both metabolites. In Figure 3.6, our refined model's predictions are used to decompose the calculated adipose glycerol and

Table 3.1: Parameter values estimated for data collected at baseline and following caloric restriction

Parameters	Role/Function	Baseline	Following caloric restriction
K_{ad}	Linear kinetic parameter LPL lipolysis.	0.0096 (0.0068 , 0.0213)	0.0087 (0.0057 , 0.0117)
τ_{LPL}	LPL insulin delay	156.92 (68.3 , 245.5)	112.76 (27.05 , 198.48)
τ_{AT}	Adipose insulin delay.	21.19 (-16.28 , 58.66)*	17.13 (-10.34 , 44.61)*
P_{GLY}	Rate parameter for uptake/release of glycerol.	0.249 (0.109 , 0.389)	17.13 (0.283 , 0.566)
P_{NEFA}	Rate parameter for uptake/release of NEFA.	0.0444 (0.0316 , 0.0571)	0.0803 (0.041 , 0.13)

Parameter values estimated by fitting the refined model to the calculated adipose tissue metabolite fluxes at baseline and following weight stabilisation for a selection of parameters are shown, 95% confidence intervals for parameter estimates are displayed in parentheses below the estimated value. The complete set of parameter values can be found in Supplementary Table S3.2. *All parameters are bound below by zero during parameter estimation.

NEFA fluxes over the duration of the mixed meal challenge test into their constituent reactions. This allows for visualisation of how the rates of specific, unmeasured reactions change following caloric restriction. The model predicts an increase in the rate of ATL lipolysis within the adipose tissue following caloric restriction (Figure 3.6 B and E, red lines), with a predicted 56% increase in the rate of lipolysis of stored triglyceride ($0.229 \mu\text{mol}/100 \text{ ml tissue}/\text{min}$ at baseline to $0.359 \mu\text{mol}/100 \text{ ml tissue}/\text{min}$ following caloric restriction), resulting in the observed increase in efflux of both NEFA and glycerol from the adipose tissue in the fasting state (Figure 3.6 A and D, purple

Figure 3.6 : Decomposition of model glycerol and NEFA flux predictions into constituent reactions

Comparison of model predictions of rates of reactions contributing to the (A) glycerol flux in the plasma compartment, (B) rates of glycerol rate of reactions in adipose compartment, (C) plasma and adipose compartment glycerol concentrations, (D) NEFA flux in the plasma compartment, (E) rates of NEFA reaction in adipose compartment, and (F) plasma and adipose compartment NEFA concentrations. Solid lines represent baseline estimates and dashed lines represents estimates following weight stabilisation.

lines). An increase in ATL lipolysis would be expected following such a period of caloric restriction as the triglyceride stores in the adipose tissue are hydrolysed to supply NEFA for use in other tissues. In the later postprandial phase, the lower concentrations of circulating triglyceride following caloric restriction (total AUC 642.2 μmol a baseline to just 463 μmol following the diet intervention), results in the reduced efflux of glycerol and NEFA from LPL lipolysis after 120 minutes (Figure 3.6 A and D, yellow lines).

Interestingly, the peak in the rate of LPL occurs approximately 40 mins earlier (230 mins at baseline versus 190 mins following caloric restriction) reflecting the estimated decrease in τ_{LPL} (Figure 3.6 A and E, yellow lines). Note, there is no decrease in the peak time of measured arterial insulin (Figure 3.4 E), however there is a decrease in both the fasting and postprandial plasma insulin concentrations (total AUC 6821.5 μU of insulin at baseline to 5291.8 μU following caloric restriction). Despite the decrease in circulating concentrations of insulin the postprandial inhibition of ATL lipolysis and stimulation of re-esterification by insulin occur at faster rates. The magnitude of slope describing the rate of postprandial inhibition of ATL lipolysis increases from 0.019 at baseline to 0.0023 following caloric restriction. Concurrently, the rate of stimulation of re-esterification increases from 0.0055 at baseline to 0.0078 following the diet intervention. These increases are indicative of improved responsiveness to circulating insulin, as reflected with the decrease in the estimated values for both insulin delay parameters.

Alternative measures of insulin resistance

Both whole body and adipose tissue specific insulin resistance were assessed using the HOMA-IR and ADIPO-IR indices respectively. HOMA-IR values decreased significantly from a mean values of 2.27 at baseline, which would be considered insulin resistant, to 1.51 following the diet intervention ($p=0.036$). ADIPO-IR values also decreased significantly from 47.27 at baseline to 28.90 following caloric restriction ($p=0.006$). Supporting the model prediction of an improvement in insulin sensitivity.

Discussion

We present a refined mathematical model of postprandial adi-

pose tissue insulin mediated dynamics of glucose, NEFA, triglyceride, and glycerol. Our refined model elucidates the strong effect of insulin on adipose tissue lipid metabolism, building on the work of Jelic [11], Pratt [13], and Sips [10], with all measured adipose tissue metabolite fluxes being stimulated or inhibited by insulin. Our model also introduces several novel terms. We explicitly account for the conversion of glucose to G-3-P, which serves to trap glucose within the adipose space. Thereby, counteracting the erroneous predictions of a postprandial efflux of glucose from the adipose compartment in the existing models due to glucose accumulation. We also introduce a term accounting for the concentration gradient dependent uptake and release of glycerol. This mechanism had been postulated by Coppack et al. in their 2005 model [42], but has not been accounted for in the three subsequently published models that we have compared. Moreover, inclusion of a [U- ^{13}C] palmitate stable isotope tracer in the ingested meal allowed us to estimate the postprandial fractional spill-over of LPL derived NEFA into the plasma. The corresponding model term has, to the best of our knowledge, not previously been validated using experimental data.

Arteriovenous measurements in the fasting and postprandial conditions across the abdominal subcutaneous tissue were employed to evaluate three existing models of postprandial adipose tissue metabolism and the assumptions upon which they have been constructed. While the existing models included in this analysis perform well in describing dynamics between triglycerides, NEFA, and glucose at the whole-body level, none of the existing models could describe the adipose tissue specific fluxes well, leading to the construction of our refined model.

Methods such as area under the curve (AUC) are most often employed to compare time series of metabolite concentrations during challenge tests, with each metabolite evaluated independently of the others. However, AUC fails to capture the dynamic properties of the postprandial metabolite curves. It is possible for several different response curves to have the same AUC value. With computational models the dynamic responses of all measured metabolites are used to parameterise a model in a physiologically meaningful way. Consequently, computational models could also prove to be a powerful tool for the interpretation and quantification of dynamic time series of data, as in the case of meal challenge test data. Our refined model was applied to time series data collected before and after a weight loss intervention study [17] to investigate the effects of caloric restriction on adipose tissue metabolism. Comparison of parameter values estimated from data collected before and after caloric restriction indicated a significant increase in the rate coefficient for

the concentration gradient based transport of both glycerol and NEFA in and out of the adipose compartment. Use of the model to decompose the glycerol and NEFA flux predictions into their constituent reactions predicted an increase in the rate of ATL lipolysis within the adipose tissue following caloric restriction, resulting in the measured increased efflux of NEFA from the adipose tissue in the fasting state. This would be expected following a period of caloric restriction as triglyceride stored in the adipose tissue is hydrolysed and the resulting NEFA is transported for use in other tissues. A large, but non-significant, decrease was observed in the insulin time delay parameter for LPL lipolysis of circulating triglyceride, τ_{LPL} , from 157 minutes at baseline to 113 minutes following caloric restriction, the effect of which can be observed with the peak in the rate of LPL lipolysis occurring approximately 40 minutes earlier following weight stabilisation. This is accompanied by a non-significant four minute decrease in the general adipose tissue insulin delay τ_{AT} . These reductions indicate a less damped response to insulin stimulation following the diet intervention, which is in line with previous studies reporting a reduction in whole body insulin resistance following prolonged caloric restriction [43]. Whole body insulin sensitivity for the sixteen individuals, measured using HOMA-IR, decreased significantly following caloric restriction, from a mean of 2.27 at baseline to 1.51 following the diet intervention ($p=0.036$). In addition, adipose tissue specific insulin resistance, assessed using ADIPO-IR, decreased significantly following the diet intervention, further supporting our model's prediction of improved insulin sensitivity following caloric restriction. It is also of note that the LPL time delay estimates are more than one, and close to two hours, shorter than the fixed 240 minutes time delay proposed by Jelic et al. [11] and subsequently used in the Sips model [10]. While changes could be observed in several other parameter values before and after caloric restriction, these differences were not significant. Previous analysis of the Yoyo study data using more traditional techniques (incremental are under the curve) found only a significant decrease in the fasting and postprandial triglyceride flux, however this was accompanied by reduced arterial triglyceride concentration [17]. Thus, we see the additional insights which can be gained through the fitting of computational models in the quantification of dynamic time series data [44].

Multiple arteriovenous studies have used the glycerol flux to infer information regarding rate of ATL lipolysis under the assumption that glycerol release by LPL lipolysis, equivalent to the instantaneous triglyceride flux, remains within the plasma, and due to the negligible activity of glycerol kinase within the adipose tissue [38], all glycerol released by ATL lipolysis enters

the plasma for transport to the liver. Thereby, the rate of ATL lipolysis can be estimated as the difference between the measured glycerol and triglyceride fluxes [45,46]. However, analysis of the postprandial glycerol and triglyceride fluxes from the Yoyo Study indicates that there is an influx of glycerol into the interstitial adipose space in the postprandial state, in contradiction with this assumption (Figure 3.5). The introduction of a term accounting for the concentration gradient dependent uptake and release of glycerol by the adipose tissue, which was first proposed by Coppack et al. [41], allows the model to describe the measured postprandial influx of glycerol. In addition, in order to achieve the best fit of the flux data, in particular the measured postprandial influx of glycerol from the plasma to the adipose tissue, it was necessary to introduce a sink term in adipose tissue glycerol which cannot be accounted for with current biological understanding. In future work it may be possible with the integration of adipose tissue specific measurements of other omics data to determine the cause of the glycerol disappearance.

In combination with arterio-venous measurements, the use of stable isotope tracers allow for the quantification of reactions that are not directly measurable, such as rates of appearance and turn over. The meal administered in the Yoyo study includes 100mg of [U-¹³C] palmitate stable isotope tracer which allows for the estimation of the rate of fractional spill-over of NEFA from LPL lipolysis of dietary triglycerides using the method as described by Bickerton et al. [5]. As the Yoyo study utilises a single tracer we do not have the data to evaluate chylomicron and VLDL triglyceride separately. Therefore, we have grouped the terms and used the available measurements of fractional spill-over of chylomicron derived NEFA to be representative of fractional spill-over from total circulating triglyceride pool. Moreover, we assume the dynamics of palmitate is representative for the generic NEFA pool. Incorporation of additional fatty acid tracers under different challenge conditions would allow for the extension of our model to describe the complex systemic interplay between different lipoprotein classes and NEFA species. Inclusion of a palmitate stable isotope tracer in a meal has been shown to label dietary derived triglycerides and NEFA for approximately the first 120 to 180 minutes [5,47]. Due to recycling of NEFA both within the adipose tissue and eventual incorporation of labelled NEFA into VLDL, estimates of fractional spill-over become less reliable beyond this time. This is reflected with the increases in the measured standard error of mean for later time points in part B of Figure 3.3. To this end, error measurements have been weighted by the standard deviation within the 16 individuals such that a time point with a larger standard deviation will have a lower priority when estimating

parameter values.

In order to produce a model which could be parameterised from data, certain model terms were simplified from those presented in the original models, e.g. use of a linear term in place of Michaelis-Menten dynamics to describe LPL lipolysis of circulating triglycerides. A linear approximation is sufficient in this situation to describe this reaction, as saturation is not expected to occur given the measured arterial triglyceride concentrations within the Yoyo study. The use of Profile Likelihood Analysis indicates that six of the fourteen model parameters are identifiable given the data, with a further six having upper or lower bounds (Supplementary Figure S3.1). For future implementation of the model, fixation of parameter values for which reliable experimental estimates are available would improve the estimation of parameters of interest for the biological questions being asked. As with previous computational models of adipose tissue metabolism, we have considered whole body adipose tissue as one homogenous unit. While the arteriovenous sampling of the abdominal subcutaneous tissue technically restricts our model evaluations to a single adipose tissue depot, we believe that our model sufficiently captures postprandial adipose tissue dynamics given that the measured depot, the abdominal subcutaneous adipose tissue, is considered the primary site for the storage and release of NEFA into systemic circulation.

Arteriovenous measurements across other tissue depots, such as skeletal muscle, presents an opportunity to further evaluate and, where necessary, refine existing model terms describing contributions of other tissues to glucose and NEFA homeostasis, as in the case of adipose tissue metabolism in this study. In this study we present a work-flow where, through the use of dependent inputs in combination with a divide and conquer approach, the model evaluation procedure can be reduced to a non-linear regression problem, greatly reducing the computational time.

Given the association of ectopic fat deposition with the development of several metabolic disorders, such as Non-Alcoholic Fatty Liver Disease and Type 2 Diabetes, the study of the dysregulation of adipose tissue metabolism has garnered much attention in recent years. Our refined model is capable of simulating *in silico* the dynamics of human *in vivo* adipose tissue metabolism in both the fasting and postprandial state. Through the variation of model parameters or modulation of dependent inputs, such as the arterial insulin concentration, it is possible to simulate the dynamics in adipose tissue metabolism in response to different external stimuli and gain insight into potential sources of dysregulation in adipose metabolism. Furthermore,

embedding our refined adipose tissue model into a larger whole body model, with terms describing the contributions of other tissues, would allow for the investigation how the model simulated dysregulation in adipose tissue metabolism would impact on the glucose and NEFA dynamics in other tissues and whole body metabolism.

In conclusion, we present a refined physiology-based computational model of adipose tissue metabolism which has been shown, using arteriovenous measurements across the abdominal subcutaneous adipose tissue, to outperform several exiting models of adipose tissue metabolism. Our model elucidates the strong influence of insulin signalling on adipose tissue dynamics, particularly the cycling between storage and release of NEFA in the fasting and postprandial states. Application of our model to data collected before and after a diet intervention allows for quantification of the effect of caloric restriction on adipose tissue metabolism. Estimated parameter values indicate that the delays in insulin effectiveness in the system are not fixed, with the estimated time delays in insulin signalling not only differing from the values used in previous models, but also decreasing following the diet intervention suggesting an improvement in adipose tissue insulin sensitivity following caloric restriction.

References

1. Frayn, K.N. Adipose tissue as a buffer for daily lipid flux. *Diabetologica*. 45:1201-1210 (2002).
2. Frayn, K.N., Karpe, F., Fielding, B.A., MacDonald, I.A., Coppack, S.W. Integrative physiology of human adipose tissue. *Int J Obes Relat Metab Disord*. 27:875-888 (2003).
3. Samra, J.S., Clark, M.L., Humphreys, S.M., MacDonald, I.A., Frayn, K.N. Regulation of lipid metabolism in adipose tissue during early starvation. *Am J physiol* 271:E541-E546 (1996).
4. Cryer, A. Tissues lipoprotein lipase activity and its action in lipoprotein metabolism. *Int J Biochem* 13:525-541 (1981).
5. Bickerton, A.S.T., Roberts, R., Fielding, B.A., Hodson, L., Blaak, E.E., Wagenmakers, A.J.M., Gilbert, M., Karpe, F., Frayn, K.N. Preferential Uptake of Dietary Fatty Acids in Adipose Tissue and Muscle in the Postprandial Period. *Diabetes* 56(1):168-176 (2007).
6. Unger, R.H. Minireview: weapons of lean body mass distruction: the role of ectopic lipids in the metabolic syndrome. *Endocrinology*. 144:5159-5165 (2003).
7. Lewis, G.F., Carpentier, A., Adeli, K., Giacca, A. Disordered fat storage and mobilization in the pathogenesis of insulin resistance and type 2 diabetes. *Endocr. Rev.* 23:201-229 (2002).
8. Bays, H., Mandarino, L., DeFronzo, R.A. Role of the adipocyte, free fatty acids, and ectopic fat in pathogenesis of type 2 diabetes mellitus: peroxisomal proliferator-activated receptor agonists provide a rational therapeutic approach. *J Clin Endocrinol Metab*. 89:463-478 (2004).
9. Bergman, R.N., Ader, M. Free fatty acids and pathogenesis of type diabetes mellitus. *Trends Endocrinol Metab*. 11:351-356 (2000).
10. Sips, F.L.P., Nyman, E., Adiels, M., Hilbers, P.A.J., Strålfors, P., van Riel, N.A.W., Cedersund, G. Model-Based Quantification of the Systemic Interplay between Glucose and Fatty Acids in the Postprandial State. *PLoS ONE*. 10(9): e0135665 (2015).
11. Jelic, K., Hallgreen, C.E., Colding-Jørgensen, M. A Model of NEFA Dynamics with Focus on the Postprandial State. *Ann Biomed Eng*. 37: 1897 (2009).
12. Kim, J., Saidel, G.M., Kalhan, S.C. A computational model of adipose tissue metabolism: Evidence for intracellular compartmentation and differential activation of lipases. *J Theor Biol*. 251(3):523-540 (2008).
13. Pratt, A.C., Wattis, J.A., Salter, A.M. Mathematical modelling of hepatic lipid metabolism. *Math Biosci* 262:167-181 (2015).
14. Pearson, T., Wattis, J.A.D., King, J.R., MacDonald, I.A., Mazzatti, D.J. The effects of insulin resistance on individual tissues: an application of a mathematical model of metabolism in humans. *Bull Math Bio*. 78:1189-1217 (2016).
15. Bickerton, A.S.T., Roberts, R., Fielding, B.A., Tornqvist, H., Blaak, E.E., Wagenmakers, A.J.M., Gilbert, M., Humphreys, S.M., Karpe, F., Frayn, K.N. Adipose tissue fatty acid metabolism in insulin-resistant men. *Diabetologica*. 51:1466-1474 (2008).
16. McQuiad, S.E., Hodson, L., Neville, M.J., Dennis, A.L., Cheeseman, J., Humphreys, S.M., Ruge, T., Gilbert, M., Fielding, B.A., Frayn, K.N., Karpe, F. Downregulation of adipose tissue fatty acid trafficking in obesity. *Diabetes*. 60(1):47-55 (2011).
17. Vink, R.G., Roumans, N.J., van der Kolk, B.W., Fazelzadeh, P., Boekschoten, M.V., Mariman, E.C., van Baak, M.A. Adipose Tissue Meal-Derived Fatty Acid Uptake Before and After Diet-Induced Weight Loss in Adults with Overweight and Obesity. *Obesity*. 25:1391-1399 (2017).
18. Vink, R.G., Roumans, N.J., Arkenbosch, L.A., Mariman, E.C., van Baak, M.A. The effect

- of rate of weight loss on long-term weight regain in adults with overweight and obesity. *Obesity*. 24(2):321-327 (2016).
19. Blaak, E.E., van Baak, M.A., Kemerink, G.J., Pakbiers, M.T.W., Heidendal, G.A.K., Saris, W.H.M. β -adrenergic stimulation and abdominal subcutaneous fat blood flow, lean, obese, and reduced-obese subjects. *Metabolism*. 44(2):183-187 (1995).
 20. Samra, J.S., Frayn, K.N., Giddings, J.A., Clark, M.L., MacDonald, I.A. Modification and validation of a commercially available portable detector for measurement of adipose tissue blood flow. *Clin Physiol*. 15(3):241-248 (1995).
 21. Matthews, D.R., Hosker, J.P., Rudenski, A.S., Naylor, B.A., Treacher, D.F., Turner, R.C. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia*. 28(7):412-9 (1985).
 22. Gayoso-Diz, P., Otero-Gonzalez, A., Rodriguez-Alvarez, M.X., Gude, F., Garcia, F., de Francisco, A., Gonzalez-Quintela, A. Insulin resistance (HOMA-IR) cut-off values and the metabolic syndrome in a general adult population: effect of gender and age: EPIRCE cross-sectional study. *BMC Endocr Disord*. 13:47 (2013).
 23. Gastaldelli, A., Cusi, K., Pettiti, M., Hardies, J., Miyazaki, Y., Berria, R., Buzzigolo, E., Sironi, A.M., Ersosimo, E., Ferrannini, E., DeFronzo, R.A. Relationship between hepatic/visceral fat and hepatic insulin resistance in nondiabetic and type 2 diabetic subjects. *Gastroenterology*. 133(2):496-506 (2007).
 24. Dalla Man, C., Rizza, R.A., Cobelli, C. Meal simulation model of the glucose-insulin system. *IEEE Trans Biomed Eng*. 54(10):1740-9 (2007).
 25. Boden, G., Chen, X., Ruiz, J., White, J.V., Rossetti, L. Mechanisms of fatty acid-induced inhibition of glucose uptake. *J Clin Invest*. 93: 2438-2446 (1994).
 26. Ferrannini, E., Barrett, E.J., Bevilacqua, S., DeFronzo, R.A. Effect of fatty acids on glucose production and utilization in man. *J Clin Invest*. 72: 1737-1747 (1983).
 27. Kim, S. Parameter Estimation Using Divide-and-Conquer Methods for Differential Equation Models. *J Biom Biostat*. 7(3):305 (2016).
 28. Price, W.L. A Controlled Random Search Procedure for Global Optimization. *The Computer Journal*. 4(1):367-370 (1977).
 29. van Riel, N.A., Sontag, E.D. Parameter estimation in models combining signal transduction and metabolic pathways: the dependent inputs approach. *Syst Biol (Stevenage)* 153(4):263-74.
 30. Hurvich, C.M., Tsai, C.L. Regression and time series model selection in small samples. *Biometrika*. 76(2):297-307 (1989).
 31. Murphy, S.A., van der Vaart, A.W. On Profile Likelihood. *J Am Stat Assoc*. 95(450):449-465 (2000).
 32. Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., Timmer, J. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*. 25(15):1923-1929 (2009).
 33. Evans, K., Burdge, G.C., Wootton, S.A., Clark, M.L., Frayn, K.N. Regulation of dietary fatty acid entrapment in subcutaneous adipose tissue and skeletal muscle. *Diabetes*. 51:2684-2690 (2001).
 34. Fielding, B.A., Frayn, K.N. Lipoprotein lipase and the disposition of dietary fatty acids. *Br J Nutr*. 80:495-502 (1998).
 35. Shepherd, P.R., Kahn, B.B. Glucose Transporters and Insulin Actions – Implications for Insulin Resistance and Diabetes Mellitus. *N Engl J Med*. 341:248-257 (1999).
 36. Frasson, D., Boschini, R.P., Chaves, M.E.S.M., Dos Santos, S.D.P., Gomes, R.R., Valentim, M.A.R., Garofalo, L.C.C., Navegantes, R.H., Migliorini, R.H., Kettelhut, I.D.C. The sympathetic nervous system regulates the glycerol-3P generation pathways in white adipose tissue of fasted, diabetic and high-protein diet-fed rats. *Metabolism*. 61:1473-1485 (2012).

-
37. Zimmermann, R., Strauss, J.G., Haemmerle, G., Schoiswohl, G., Birner-Gruenberger, R., Riederer, M., Lass, A., Neuberger, G., Eisenhaber, F., Hermetter, A., Zechner, R. Fat mobilization in adipose tissue is promoted by adipose triglyceride lipase. *Science*. 306:1383–1386 (2004).
 38. Lin, E.C. Glycerol utilization and its regulation in mammals. *Annu Rev Biochem*. 46:776–795 (1977).
 39. Zierath, J.R., Livingston, J.N., Thorne, A., Bolinder, J., Reynisdottir, S., Lonnqvist, F., Arner, P. Regional difference in insulin inhibition of non-esterified fatty acid release from human adipocytes: relation to insulin receptor phosphorylation and intracellular signalling through the insulin receptor substrate-1 pathway. *Diabetologia*. 41:1343–1354 (1998).
 40. Stumvoll, M., Wahl, H.G., Machicao, F., Haring, H. Insulin sensitivity of glucose disposal and lipolysis: no influence of common genetic variants in IRS-1 and CAPN10. *Diabetologia*. 45:651–656 (2002).
 41. Coppack, S.W., Frayn, K.N., Humphreys, S.M., Dhar, H., Hockaday, T.D. Effects of insulin on human adipose tissue metabolism in vivo. *Clin. Sci. (Colch)* 77:663–670. (1989)
 42. Coppack, S.W., Chinkes, D.L., Miles, J.M., Patterson, B.W., Klein, S. A Multicompartmental Model of IN Vivo Adipose Tissues Glycerol Kinetics and Capillary Permeability in Lean and Obese Humans. *Diabetes*. 54(7):1934–1941 (2005).
 43. Dube, J.J., Amati, F., Toledo, F.G., Stefanovic-Racic, M., Rossi, A., Coen, P., Goodpaster, B.H. Effects of weight loss and exercise on insulin resistance, and intramyocellular triacylglycerol, diacylglycerol and ceramide. *Diabetologia*. 54:1147–1156 (2011).
 44. Rozendaal, Y.J., Maas, A.H., van Pul, C., Cottaar, E.J.E., Haak, H.R., Hilber, P.A.J., van Riel, N.A. Model-based analysis of postprandial glycemic response for different types of food. *Clin Nutr Exp*. 19:32–45 (2018).
 45. Frayn, K.N., Humphreys, S.M. Metabolic characteristics of human subcutaneous abdominal adipose tissue after overnight fast. *Am J Physiol Endocrinol Metab*. 302(4):E468–E475 (2014).
 46. Dimitriadis, G., Mitrou, P., Lambadiari, V., Boutai, E., Maratou, E., Koukkou, E., Tzanela, M., Thalassinou, N., Raptis, S.A. Glucose and lipid fluxes in the adipose tissue after meal ingestion in hyperthyroidism. *J Clin Endocrinol Metab*. 91(3):1112–8 (2006).
 47. Heath, R.B., Karpe, F., Milne, R.W., Brudge, G.C., Wootton, S.A., Frayn, K.N. Selective partitioning of dietary fatty acids into the VLDL TG pool in the early postprandial period. *J Lipid Res*. 44:2065–2072 (2003).

Supplementary Material - Chapter 3

Supplemental Figure S3.1 : Results of Profile Likelihood Analysis for each parameter of the refined model

The parameter values estimated from the baseline data are shown with a red cross and the value of $C(p)$ resulting from iteratively adjusting the parameter value and re-estimating the parameter values indicated by the blue line. K_{ad} , τ_{LPL} , D_{spill} , τ_{AT} , P_{GLY} , P_{NEFA} are identifiable parameters. Profile likelihood for several other parameters ($GLUT1$, $GLUT2$, B_{ATL} , ATL_{max} , and τ_{G3P}) indicate the existence of an upper bound, these parameters have been bound below by zero in the parameter estimation procedure. The parameter describing the fractional usage of glucose for G-3-P production in re-esterification are bound above by one. Finally two parameters (K_{ATL} , and $K_{reester}$) appear to be practically non-identifiable.

Supplemental Figure S3.2 : Results of fitting the Jelic, Pratt, Sips, and refined models to the baseline adipose tissue flux data.

Model simulation using parameter values estimated by fitting each model to the measured adipose (A) triglyceride flux, (B) fractional spill over of LPL derived NEFA, (C) glucose influx, (D) glycerol and (E) NEFA efflux are shown, Jelic (blue), Pratt (green), Sips (blue), and refined model (red). Mean baseline calculated adipose tissue flux values $\pm$ the standard error of the mean from the Yoyo study are shown in black. The Jelic model is capable of producing quite of good fit of the triglyceride and NEFA fluxes. The Jelic model does not include terms to describe the fractional spill over, glucose, or glycerol fluxes. While the Pratt model can produce a qualitatively good fit of the glucose, NEFA, and glycerol fluxes it cannot produce a good fit of the triglyceride flux (A). The Pratt model also assumes that the contribution of insulin dependent lipolysis to the overall lipolysis of circulating triglyceride is negligible, consequently, the rate of LPL lipolysis is primarily determined by the circulating triglyceride concentration.

Supplemental Figure S3.3 : Increasing the influence of insulin dependent LPL lipolysis in the Pratt model

In the above figure we increase the weight of the contribution of insulin dependent LPL lipolysis, such that insulin dependent LPL lipolysis accounts for 1% (blue line) and 10% (yellow line) of the total adipose tissue triglyceride flux while maintaining the other parameters at the values provided in the original publication. As the Pratt model uses direct plasma insulin stimulation rather than accounting for delays in insulin signalling, as in the Jelic, Sips, and refined models, the model simulated triglyceride flux beings to peak too early under the influence of plasma insulin. The refined model makes use of the LPL lipolysis term from the Pratt model, but introduces the three compartmental delay from the Jelic and Sips models. With this delay the insulin signalling, the refined model can produce a good fit to the triglyceride flux data at baseline and following weight stabilisation (red line).

Supplementary Table S1: Comparison of terms from the Jellic, Pratt, Sips, and refined adipose tissue models.

Term	Jellic Model [1]	Pratt Model [2]	Sips Model [3]	Refined Model
LPL lipolysis	$= -\frac{LPL_{max} \cdot [TG_{PL}]}{K_{LPL} + [TG_{PL}]} \cdot \frac{P \cdot [I_{LPL}]}{K_P + [I_{LPL}]}$ <ul style="list-style-type: none"> Accounts for enzyme kinetics. Insulin action is modelled with a third order delay. $[I_{LPL}]$ (240 min delay) 	$= -K_a(1 + K_{ai}[I_{PL}][TG_{PL}])$ <ul style="list-style-type: none"> No enzyme kinetics. (linear dynamics) No insulin delay. (plasma insulin stimulation $[I_{PL}]$) 	$= -\frac{LPL_{max} \cdot [TG_{PL}]}{K_{LPL} + [TG_{PL}]} \cdot \frac{P \cdot [I_{LPL}]}{K_P + [I_{LPL}]}$ <ul style="list-style-type: none"> As in Jellic model. 	$= -k_{ad}[I_{LPL}][TG_{PL}]$ <ul style="list-style-type: none"> No enzyme kinetics (do not have data to accurately estimate parameters.) Introduces Jellic model insulin delay. $[I_{LPL}]$
ATL lipolysis	$= B_{ATL} + \frac{ATL_{max}}{1 + \frac{[I_{AT}]}{K_{ATL}}}$ <ul style="list-style-type: none"> Occurs at basal and saturable rate. Inhibited by delayed insulin $[I_{AT}]$ 	$= \frac{3\beta_L}{1 + K_{Fi}[I_{PL}]^2}$ <ul style="list-style-type: none"> Saturable rate. Inhibited by plasma insulin. $[I_{PL}]$ Released NEFA appears directly in plasma compartment. 	$= p_B([I_{d7}] - I_b)[NEFA]$ <ul style="list-style-type: none"> Linear term governed by difference between delayed insulin $[I_{d7}]$ and basal insulin. I_b 	$= B_{ATL} + \frac{ATL_{max}}{1 + \frac{[I_{AT}]}{K_{ATL}}}$ <ul style="list-style-type: none"> Occurs at basal and saturable rate. Inhibited by delayed insulin. $[I_{AT}]$
Re-esterification	$= -\frac{RST_{max} \cdot [NEFA_{AT}]}{K_{RST} + [NEFA_{AT}]} \cdot \frac{Q \cdot [I_{AT}]}{K_Q + [I_{AT}]}$ <ul style="list-style-type: none"> Saturable rate. Stimulated by delayed insulin. $[I_{AT}]$ Independent of glucose/G-3-P concentration. 	$= -K_{aa}[I_{PL}][NEFA_{AT}][G_{AT}]$ <ul style="list-style-type: none"> Linear rate. Stimulated by plasma insulin. $[I_{AT}]$ concentration. 	<p>Not present</p> <ul style="list-style-type: none"> Re-esterification is not explicitly described. Coupled with term describing NEFA flux. 	$= -3K_{rest}[I_{AT}][NEFA_{AT}][G-3-P_{AT}]$ <ul style="list-style-type: none"> Linear rate. Stimulated by delayed insulin. $[I_{AT}]$ Dependent on adipose G-3-P concentration.
Fractional spill-over	<p>Not present</p> <ul style="list-style-type: none"> All NEFA released by LPL lipolysis enters interstitial adipose compartment. 	$\text{spill-over} = K_{lp}$ <ul style="list-style-type: none"> Occurs at a constant rate $K_{lp} = 0.25$. 	$\text{spill-over} = \frac{A_{spill} + B_{spill}}{100} \cdot \frac{I_b}{[I_{d6}]}$ <ul style="list-style-type: none"> Insulin independent and delayed insulin $[I_{d6}]$ dependent term. 	$\text{spill-over} = \frac{D_{spill}}{100} \cdot \frac{I_b}{[I_{PL}]}$ <ul style="list-style-type: none"> Plasma insulin dependent. $[I_{PL}]$

Term	Jelic Model [1]	Pratt Model [2]	Sips Model [3]	Refined Model
Glucose flux	<p>Not present</p> <ul style="list-style-type: none"> Does not describe glucose dynamics. 	$= -d_{BA}(1 + K_{ga}[I_{PL}])([G_{PL}] - [G_{AT}])$ <ul style="list-style-type: none"> Occurs at plasma insulin dependent $[I_{PL}]$ and insulin independent rate. Rate dependent on relative concentration of adipose and plasma glucose. 	$= -\frac{(K_{uid1} + K_{uid3} \frac{[I_{IP}]}{[NEFA_{PL}]})([G_t]}{K_{m,uid} + [G_t]}$ <ul style="list-style-type: none"> No explicit adipose glucose flux. Occurs at insulin and NEFA dependent and independent rate. Delayed insulin stimulation $[I_{d3}]$ Delayed NEFA inhibition 	$= -(GLUT_1 + GLUT_4[I_{AT}])([G_{PL}]$ <ul style="list-style-type: none"> Occurs at insulin dependent and insulin independent rate. Assumes adipose glucose concentration is negligible. Delayed insulin stimulation $[I_{AT}]$
NEFA flux	$= -PP([NEFA_{PL}] - [NEFA_{AT}])$ <ul style="list-style-type: none"> Concentration gradient dependent transport. Fractional spill-over not described. 	$= -k_{na}[NEFA_{PL}] + \frac{3\beta I}{1 + K_{ft}[I_{PL}]^2}$ $+ 3K_a \cdot K_{lp}(1 + K_{ai}[I_{PL}])([TG_{PL}]$ <ul style="list-style-type: none"> Linear term describing NEFA uptake only. NEFA release by ATL lipolysis contributes directly to plasma NEFA. Fractional spill-over of LPL derived NEFA contributes to flux. 	$= -(P_A + P_B([I_{DT}] - I_b))[NEFA_{PL}]$ $+ \frac{spill}{V_{NEFA}} \left(\frac{L_{PL}L_{max} \cdot [TG_{PL}]}{K_{LPL} + [TG_{PL}]} \cdot \frac{P \cdot [I_{LPL}]}{K_P + [I_{LPL}]} \right)$ <ul style="list-style-type: none"> Linear term describing NEFA uptake. NEFA released by ATL lipolysis contributes directly to NEFA flux. Fractional spill-over of LPL derived NEFA contributes to flux. 	$= -P_{NEFA}([NEFA_{PL}] - [NEFA_{AT}])$ $+ 3 \left(\frac{I_b}{100} \right) K_{ad}[I_{LPL}][TG_{PL}]$ <ul style="list-style-type: none"> Concentration gradient dependent transport. Fractional spill-over of LPL derived NEFA contributes to flux.
Glycerol flux	<p>Not present</p> <ul style="list-style-type: none"> There is no description of glycerol dynamics. 	<p>Not present</p> <ul style="list-style-type: none"> Glycerol released by ATL lipolysis appears directly in the liver compartment. Glycerol released by LPL lipolysis not described. 	<p>Not present</p> <ul style="list-style-type: none"> There is no description of glycerol dynamics. 	$= -P_{GLY}([GLY_{AT}] - [GLY_{PL}])$ $+ K_{ad}[I_{LPL}][TG_{PL}]$ <ul style="list-style-type: none"> Concentration gradient dependent transport of glycerol. Glycerol released from LPL lipolysis contributes to flux.

Terms describing individual metabolite fluxes across the adipose tissue were extracted from the Jelic, Pratt, and Sips model and compared using the Yoyo study arterio-venous measurements. The resulting refined model terms are shown in the fifth column.

References:

1. Jelic K, Hallgreen C E, Colding-Jørgensen M. "A Model of NEFA Dynamics with Focus on the Postprandial State", *Ann Biomed Eng* (2009) 37: 1897.
2. Pratt AC, Wattis JA, Salter AM. "Mathematical modelling of hepatic lipid metabolism", *Math Biosci* (2015) 262:167-181.
3. Sips FLP, Nyman E, Adiels M, Hilbers PAJ, Strålfors P, van Riel NAW, Gunnar Cedersund. "Model-Based Quantification of the Systemic Interplay between Glucose and Fatty Acids in the Postprandial State." *PLoS ONE* (2015) 10(9): e0135665.

Supplemental Table S 3.2 : Parameter values estimated for data collected at baseline and following caloric restriction

Flux	Role/Function	Parameter	Baseline	Following caloric restriction
Triglyceride	Linear kinetic parameter.	K_{ad}	0.0096 (0.0068, 0.0213)	0.0087 (0.0057, 0.0117)
	LPL insulin delay	τ_{LPL}	156.92 (68.3, 245.5)	112.76 (27.05, 198.48)
Spill-over	Fractional spill-over	D_{spill}	43.17 (37.98, 48.37)	46.91 (34.67, 59.15)
Glucose	Insulin independent uptake	GLUT1	0.017 (-0.002, 0.036)*	2.16×10^{-7} (-0.03, 0.03) *
	Insulin dependent uptake.	GLUT4	6.106×10^{-4} (-4.02×10^{-4} , 0.002)*	0.0017 (-2.72×10^{-4} , 0.0037)*
Glycerol	Adipose insulin delay.	τ_{AT}	21.19 (-16.28, 58.66)*	17.13 (-10.34, 44.61)*
	Rate parameter for uptake/release.	P_{GLY}	0.249 (0.109, 0.389)	17.13 (0.283, 0.566)
	Basal ATL lipolysis.	B_{ATL}	0.001 (-0.467, 0.469)*	2.1×10^{-4} (-0.781, 0.781) *
	V_{max} of enzyme mediated ATL lipolysis.	ATL_{max}	1.595 (-56.352, 59.542)*	0.559 (-0.384, 1.501)*
	K_m of enzyme mediated ATL lipolysis.	K_{ATL}	1.631 (-71.268, 74.531)*	11.312 (-78.68, 101.31)*
NEFA	Rate of re-esterification.	K_{reest}	3.019×10^{-4} (-9.079×10^{-4} , 0.0015)*	0.011 (-0.0101, 0.0122)*
	Rate parameter for uptake/release.	P_{NEFA}	0.0444 (0.0316, 0.0571)	0.0803 (0.041, 0.13)
	Time delay for production of G-3-P from glucose.	τ_{G-3-P}	33.095 (-28.589, 94.779)*	33.24 (-75.768, 142.248)*
	Fraction of adipose glucose converted	$frac_{cuse}$	0.919 (0.279, 1.559)†	0.706 (0.048, 1.364)†

Complete set of parameter values estimated by fitting the refined model to the calculated adipose tissue metabolite fluxes at baseline and following weight stabilisation. 95% confidence intervals for parameter estimates are displayed in parentheses below the estimated value. The coloured boxes indicate the model term, or terms, in which each parameter appears. *All parameters were bound below by zero during parameter estimation, however the method for calculating the confidence intervals assumes the confidence interval is symmetric about the estimated parameter values. † terms that describe a fractional value were also bound above by 1 during parameter estimation.

4

Use of deep learning methods to translate drug-induced gene expression changes from rat to human primary hepatocytes exposed in vitro and in vivo

Shauna D. O'Donovan¹, Kurt Driessens², Daniel Lopatta², Florian Wimmenauer², Alexander Lukas², Jelmer Neeven², Tobias Stumm², Evgueni Smirnov², Michael Lenz^{1,3,4}, Gokhan Ertaylan^{1,5}, Danyel G.D. Jennen⁶, Natal A.W. van Riel^{1,7}, Rachel Cavill², Ralf L.M. Peeters^{1,2}, Theo M.C.M de Kok^{1,6}

¹Maastricht Center for Systems Biology (MaCSBio), Maastricht University. ²Dept. of Data Science & Knowledge Engineering, Maastricht University. ³Institute of Organismic and Molecular Evolution, Johannes Gutenberg University Mainz. ⁴Preventive Cardiology and Preventative Medicine - Center for Cardiology, University Medical Center of the Johannes Gutenberg University Mainz. ⁵Flemish Institute for Technological 1 Research (VITO), Mol, Belgium. ⁶Dept. of Toxicogenomics, GROW School of Oncology and Developmental Biology, Maastricht University. ⁸Dept. of Biomedical Engineering, Eindhoven University of Technology

PLoS ONE (2020) 15(8): e0236392.

Abstract

In clinical trials, animal and cell line models are often used to evaluate the potential toxic effects of a novel compound or candidate drug before progressing to human trials. However, relating the results of animal and *in vitro* model exposures to relevant clinical outcomes in the human *in vivo* system still proves challenging, relying on often putative orthologs. In recent years, multiple studies have demonstrated that the repeated dose rodent bioassay, the current gold standard in the field, lacks sufficient sensitivity and specificity in predicting toxic effects of pharmaceuticals in humans. In this study, we evaluate the potential of deep learning techniques to translate the pattern of gene expression measured following an exposure in rodents to humans, circumventing the current reliance on orthologs, and also from *in vitro* to *in vivo* experimental designs. Of the applied deep learning architectures applied in this study the convolutional neural network (CNN) and a deep artificial neural network with bottleneck architecture significantly outperform classical machine learning techniques in predicting the time series of gene expression in primary human hepatocytes given a measured time series of gene expression from primary rat hepatocytes following exposure *in vitro* to a previously unseen compound across multiple toxicologically relevant gene sets. With a reduction in average mean absolute error across 76 genes that have been shown to be predictive for identifying carcinogenicity from 0.0172 for a random regression forest to 0.0166 for the CNN model ($p < 0.05$). These deep learning architecture also perform well when applied to predict time series of *in vivo* gene expression given measured time series of *in vitro* gene expression for rats.

Introduction

The field of toxicology primarily concerns itself with the evaluation of compounds for potential adverse effects in humans. The liver is the primary organ for metabolism and detoxification of pharmaceuticals and other compounds, consequently adverse effects most often occur in the liver. However, the liver is not readily accessible for sampling. In the absence of relevant human data a plethora of alternative bioassays are used to predict adverse toxicological outcomes in humans. The current 'gold standard' for evaluating the toxicity of a compound is the two-year rodent bioassay [1]. However, this method is time-consuming, expensive, and requires the sacrifice of large numbers of animals in order to screen a single compound. Moreover, the predictive relevance of the rodent bioassay is increasingly being called into question, with a 2009 report by Hartung indicating that tests in rodents had correctly predicted just 43% of toxic effects of pharmaceuticals in humans [2]. Given the high attrition rates of novel candidate drugs in the final stages of research and drug development having alternatives to animal toxicity tests that are more reliable is of particular concern to the pharmaceutical industry [2]. Coupled with the increasingly restrictive legislation surrounding animal testing, this has motivated the development of alternatives, mostly based on human *in vitro* cellular models.

The intra-cellular response to an external stimulus, such as a potentially toxic exposure, is mediated by gene transcripts, also known as mRNA. The gene transcripts relay the genomic information stored within our DNA in the nucleus of each cell allowing proteins necessary to metabolise and detoxify the compound to be translated and activated, in accordance with the central dogma of molecular biology. With advancements in high-throughput omics platforms such as micro-array, and more recently RNA-Seq, it is now possible to measure the mRNA levels within a sample. Such transcriptomic profiles not only provide a particularly sensitive assessment of the cellular response to a potentially toxic external stimulus, but as they capture genome-wide alteration in mRNA levels following an exposure they may also facilitate mechanistic insight into modes of toxicological action or disease development. *in vitro* toxicogenomic assays have produced promising results in the discrimination between subclasses of carcinogenicity [3,4] with an overall accuracy of 80% for the prediction of *in vivo* rodent toxicity [5-10]. Nevertheless, *in vitro* assays are not without their limitations. These simple cell systems differ greatly in terms of functionality and signalling from the

organs they represent; lacking the systemic interplay with other tissues that exists *in vivo*. This is more so a problem with commonly used immortalized cell lines, such as tumour derived hepatocytes HepG2, where liver metabolic functions tend to vanish as culture time increases [11,12].

In addition, the analysis and interpretation of the high dimensional, highly correlated gene expression data presents many new challenges. While, multiple studies have reported positive results in predicting toxicity of a novel compound using genomic signatures of disease states [13-15], relating the genetic signatures identified from human disease states to *in vitro* or *in vivo* rodent toxicogenomics assays is achieved using, often putative, orthologs. Orthologs are genes in different species that share a similar function or role, often predicted based on similarities in genetic sequences. However, a study which compared the toxicogenomics responses for various compounds following 24 hour exposure for HepG2 cells, primary human, rat, and murine hepatocytes noted that while the biological replicates for a given compound and model were close together no single compound induced similar responses in all *in vitro* models [16]. The discrepancy between gene expression patterns between orthologs in different cellular models highlights the need for methods that can help us bridge not only the cross-species division but also resolve differences resulting from experimental techniques. Allowing for extrapolation of functional and mechanistic changes in cellular responses to the human *in vivo* disease state.

Artificial Neural Networks (ANNs) have emerged as a highly successful field of machine learning. ANNs consist of several layers of forward connected nodes, which transform the raw data into increasingly abstract feature representations for use in classification or prediction [17-19]. While neural networks themselves are not a novel idea, advances in computational power and training algorithms in more recent years, coupled with the availability of sufficiently large data sets, have allowed deeper architectures to be trained [20]. Deep ANNs have produced outstanding results on a range of applications including natural language [21] and image processing [22] and have been mooted as being a potential solution to the challenges faced with analysing high dimensional toxicogenomics data [23]. Deep learning methods have already been shown to outperform more classical techniques in predicting toxicity [24,25] and drug-induced liver injury [26,27] given information about the chemical structure and properties of a compound. In addition, multiple recent studies have had some success in applying deep learning architectures to the analysis of gene expression data, predicting genome-wide gene expres-

sion given information on histone modifications [28] or expression levels for a subset of, so-called, landmark genes [29]. Within the field of toxicology, deep learning models have also been successfully applied to predict the pharmacological properties of drugs given transcriptomic profiles, facilitating drug repurposing [30] and also predicting liver toxicity [31].

In this study, we evaluate the potential of deep learning methods to extrapolate transcriptomic profiles from rodent to human and from *in vitro* to *in vivo* in rats. Several deep learning network architectures are compared with more classical machine learning approaches, such as random regression forests and k-nearest neighbours, to determine the added value of the deep learning approaches. Firstly, we examine the cross-species prediction problem, applying the models to predict human *in vitro* gene expression given a time series of rat *in vitro* gene expression following exposure to a previously unseen compound. Subsequently, we look at predicting rat *in vivo* gene expression given a measured pattern of rat *in vitro* gene expression following an exposure using data from open TG-GATEs. In addition, a systematic analysis is conducted to determine the effect of the composition of the included genes on the accuracy of the model predictions.

Materials and methods

Open TG-GATEs

Open TG-GATEs is a large, publicly available toxicogenomics database containing gene expression profiles as well as traditional toxicological data from *in vitro* assays, in both primary rat and human hepatocytes, and *in vivo* rats following exposure to 170 compounds [32]. The data used in these analyses are summarised in Fig 4.1. For the *in vitro* human and rat experiments gene expression profiles were measured at three time points (2, 8, and 24 hours) following a single exposure to a given compound at up to three dosages (low, middle, and high) plus a control; with two biological replicates. The high dosage for each compound was determined as the maximally tolerated dose (yielding an 80-90% relative survival ratio). The ratio of concentrations for the low and middle doses were 1/25 and 1/5 of the high dose. For the rat *in vivo* experiments, gene expression profiles were measured at four time points (3, 6, 9, and 24 hours) following a single exposure. The high dosage for each compound was selected to match the level shown to induce the minimum toxic effect over the course of a four week toxicity study, with the ratio of concentrations for the low and

middle doses at 1/10 and 3/10 of the high dose. There are three biological replicates for each *in vivo* compound-dosage combination. Gene expression profiles were generated using the Affymetrix GeneChip, the Rat Genome 230 2.0 Array and the Human Genome U133 Plus 2.0 Array. Array data for all available compounds for rat and human *in vitro* and rat *in vivo* were downloaded in the form of CEL files from the Open TG-GATEs database (<https://toxico.nibiohn.go.jp>) and pre-processed using Affymetrix Power Tools using the robust multi-array average normalisation method, correcting the raw micro-array probe intensities for background and inter-array noise to produce interpretable gene expression data. Following normalisation, compounds with incomplete data, missing either time points or dosages, were removed, leaving 45 compounds with a complete set of data in all three domains which were included in this study (Supplementary Table S4.1).

Figure 4.1: Overview of toxicogenomics data available in open TG-GATEs.

Outline of toxicogenomics data available in Open TG-GATEs across three domains. In rat and human *in vitro* studies gene expression profiles are measured at three time points (2, 8, 24 hours) with two biological replicates at low, medium, and high dosages plus the control. The rat *in vivo* data is available at four time points (3, 6, 9, and 24 hours) at a low, medium, and high dosage plus a control with three biological replicates. No data is available for the human *in vivo* domain.

Experimental setup

The experimental setup of this study is visualised in Figure 4.2 and detailed below. In summary, gene sets are identified either from literature or randomly selected from a list of known rat-human orthologs. Once the relevant gene expression data has been extracted from the TG-GATEs data set machine learning examples can be generated by pairwise matching of compound-dose combinations from the source (rat *in vitro*) and target (human *in vitro* or rat *in vivo*) domains. The gene expression data is then re-encoded and reformatted into the preferred input for the models. In order to maximise the number of learning examples available for training each of the models are trained using a leave-one-compound-out cross validation approach. Model performance is assessed using the average mean absolute error in predicting gene expression over all 45 compounds. A detailed description of the three applied deep learning architectures and two machine learning models can also be found below.

Figure 4.2 : Schematic of experimental setup.

Overview of the design of this study. Firstly, gene sets are identified either from literature or by randomly selecting genes from a list of known rat-human orthologs. Once the relevant gene expression data has been extracted from the TG-GATEs data set and learning examples are generated by the pairwise matching of compound dose combinations. The models are then trained using a leave-one-compound-out cross validation loop, where by all instance for one compound are removed from the data set, the models are then trained on the data for the remaining 44 compounds and predictions assessed by predicting for this unseen compound.

Gene sets

Micro-array data measures the abundance of more than 20,000 gene transcripts. Consequently, our time series of gene expression profiles have a very high dimensional feature space. While Open TG-GATEs is a comparatively large toxicogenomics database, the 720 learning examples that could be generated by combining all available compounds and dosages are somewhat restrictive for effectively training a model to predict genome-wide gene expression. For example, a single fully connected layer of a neural network going from n nodes in the input layer to m nodes in the internal layer requires the training of $n*m$ weights. A genome-wide prediction for a single time-point would require an input layer consisting of 20,000 nodes resulting in $20,000*m$ parameters which would need to be optimised. Therefore, it was decided to select subsets of genes to evaluate the predictive ability of the neural networks. Four subsets of genes associated to relevant toxicological outcomes were identified from literature; namely steatosis (developed in-house using the KEGG pathway hsa04932), cholestasis [33-35], genotoxicity and carcinogenicity (GTX+C) [13,14], and NAFLD [15]. The published gene lists were filtered to contain only known rat-human orthologs. Gene sets can be found in Supplementary Tables S4.2-S4.5.

In order to evaluate the effect of gene set composition on model predictions several randomly selected gene sets of increasing size were generated. Initially, sets of 20 genes were randomly selected independently from both the source domain (rat *in vitro*) and the target domain (human *in vitro* or rat *in vivo*). Subsequently, sets of 15 randomly selected genes were iteratively added to the core set; generating nested gene sets consisting of 35, 50, 65, and 80 genes. Random sets of orthologs were also generated by repeating this nested gene set selection procedure within a reduced set of known rat-human orthologs. The expression of some genes was not altered following the exposure. Constant gene expression is easier to predict. Therefore, all generated gene sets were screened to ensure sufficient variation in gene expression. This threshold was set at 3% lower than the variance measured in the steatosis gene set.

Learning examples

Once the expression data for the selected gene sets have been extracted from the TG-GATEs data set. Machine learning examples are generated by pairing the time series of gene expression for a given

compound-dose combination in the source domain (rat *in vitro*) with the time series of gene expression for the same compound-dose combination in the target domain (human *in vitro* or rat *in vivo*). As each replicate for a given compound-dose combination in the source domain is a valid match for both replicates in the target domain four learning examples can be generated for each compound-dose combination (Figure 4.2, Supplementary Fig S4.1). In order to maintain the structure of the data in each domain one of the three replicates for the rat *in vivo* domain is randomly discarded. For the purpose of generating learning examples, the controls are treated as an additional dose. With 45 compounds, three dosages plus the control, and the pairwise matching of replicates it is possible to generate 720 learning examples for each source-target domain combination. For all models, with the exception of the convolutional neural network, gene expression data is arranged in a vector consisting of each time points of gene expression for consecutive genes for a given compound-dose combination, for example the rat *in vitro* data for the GTX+C gene set consists of gene expression values for 76 genes at three time points resulting in a (228 × 1) dimensional input vector. For the CNN the gene expression data is restructured into a two dimensional array with time points on one dimension and genes on the second, consequently the rat *in vitro* data for the GTX+C gene sets forms a (76 × 3) array.

Encoding the gene expression data

With time series of gene expression data, it is the changes in gene expression between consecutive time points, and not the levels of gene expression, which is of primary interest. Consequently, the input data was re-encoded, favouring the correct prediction of the pattern of gene expression. The first entry for each gene is the measured gene expression at time point one. The values have been re-scaled relative to maximum and minimum gene expression levels across all genes and compound-dose combinations such that the values are in the range [0,1], the preferred input to neural networks. The remaining entries encode the slopes between each consecutive pair of time points. The second entry is the difference in gene expression between the first time point and the second time point and so on. This re-encoding of the gene expression data produces a lower error for predictions with the correct gene expression pattern, even if the level of gene expression has been shifted, without the need to use a more complex error function when training (Supplementary Figures S4.2a and S4.2b).

Leave-one-compound-out cross validation

In order to maximise the number of learning examples available for training leave-one-compound-out cross validation is used. All sixteen instances for a given compound were removed from the training data, the models were then trained on the remaining 704 learning examples. The excluded instances were then used to validate the prediction accuracy of the model. This was repeated for all 45 compounds. The overall performance of each model was assessed using the average validation error over all 45 compounds. The validation error is the mean absolute error between the model predicted time series of gene expression and the measured time series of gene expression for each gene.

Models are trained using the mean absolute error across all time points and genes according to the following formula.

$$\text{MAE} = \frac{1}{G * T} \sum_{i=1}^G \sum_{j=1}^T |X_{ijc} - M_{ijc}| \quad (4.1)$$

Where T is the number of sampled gene expression time points and G is the number of genes in the gene set. X_{ijc} is the measured gene expression level and M_{ijc} is the model prediction of gene expression at time point j for gene i for a given compound-dose combination c . Overall model performance is assessed as the average mean absolute error over the 45 leave-one-compound-out loops.

Models

Artificial Neural Networks consist of many nodes arranged into multiple connected layers [17]. Each node processes the information received from the previous layer as a function of the inputs and their connection weights and responds accordingly. These responses, known as activations, become the input for the subsequent layer of nodes [36]. Through training the connection weights are tuned [20], prioritising the most informative features for the particular task. A bottleneck layer is a layer that contains fewer nodes compared to the previous layers, depicted in Figure 4.3, forcing the network to find a reduced dimensional representation of the data. The bottleneck forces learning by restricting the network to use only the most relevant features [37]. Consequently, the reduced dimensional latent space also helps avoid over-fitting, preventing learning of a one-to-one prediction.

Two bottleneck architectures were implemented in this study.

Naïve Encoder

The first model is a classical, fully connected ANN with a bottleneck structure (Figure 4.3). Through a grid search of the hyper-parameters a well-performing network structure was found to consist of two hidden layers containing 256 and 160 nodes respectively, a central layer containing 32 nodes, followed by two further layers with 96 and 256 nodes. All layers use rectified linear unit (ReLU) activation [38], except for the output layer which uses sigmoid activation. L1 activity regularisation was applied to all five hidden layers to enforce sparseness [39]. Initial connection weights were randomly assigned.

Figure 4.3 : Schematic diagram of the architecture of the Naïve Encoder model.

This model is a densely connected five-layer artificial neural network with a bottle neck architecture. The two encoding layers consisting of 256 and 160 nodes respectively, followed by a central bottleneck layer containing 32 nodes and two decoder layers containing 96 and 256 layers respectively. All layers use ReLU activation, other than the final output layer which uses sigmoid activation. L1 activity regulation is applied to all layers to enforce sparseness. Input and output layers are shown for the steatosis gene set in the rat *in vitro* to human *in vitro* prediction, where both the input and the output have 150 dimension (50 genes * 3 time points).

Modified Autoencoders

Autoencoders are a subclass of ANNs with a bottleneck design, used to learn an efficient encoding of data [40]. Rather than initialising our prediction model with random weights, it was decided to independently train separate autoencoders for each of our data domains. The trained weights for the encoder for our source domain were then concatenated with the trained weights for the decoder of the target domain. These weights became the initial weights for training our second bottleneck architecture, outlined in Figure 4.4. This network also consists of five hidden layers. The first two layers forming the encoder contain 70 nodes, the bottleneck layer contains 60 nodes, and the remaining two layers of the decoder have 70 nodes. To stimulate the learning of compact internal representation of the data L1 activity regularisation was applied to the three middle layers of the network. All layers use ReLU activation except for the output layer, which uses sigmoid activation.

Figure 4.4 : Schema for the structure of the Modified Autoencoder.

Initially, two autoencoders are trained independently one for the source domain (red) and a second for the target domain (blue), both networks have the same architecture as the final modified autoencoder consists of five hidden layers of 70, 70, 60, 70, and 70 nodes respectively.. The weights for the encoder portion of the source domain autoencoder (red) are concatenated with the weights for the decoder section of the target domain autoencoder (blue). These weights become the initialising weights for the prediction task. L1 activity regularisation is applied to the three middle hidden layers.

Convolutional Neural Network

The architecture of convolutional neural networks (CNNs) is analogous to the connectivity pattern in the visual cortex, with neurons responding to stimuli in only a limited region of the visual field [41]. As the filters applied in the convolutional layers look at local regions, CNNs greatly reduce the number of weights requiring training. For input to the CNN model, the gene expression data was reshaped into a 2D format, with genes on one axis and corresponding time points on the other. CNNs consist of alternating convolutional and pooling layers followed by fully connected layers [42]. In our network, Figure 4.5, the first convolutional layer consists of 16 filters with a 10×1 kernel, followed by MaxPooling [43] layer of pool size 2×1 . This is followed by an 8-filter convolution again with a 10×1 kernel, with a second MaxPooling layer. A third convolutional layer uses four

Figure 4.5 : Structure of the implemented convolutional neural network.

The convolutional neural network consists of three alternating convolutional and MaxPooling layers. The output from the final MaxPooling layer is then flattened into a vector which then passed through two fully connected layers to reconstruct the time series of gene expression pattern in the target domain. The above figure illustrates the rat *in vitro* to human *in vitro* prediction for the steatosis gene sets consisting of 50 genes. The gene expression data is reconstructed into a 2D format, with the genes on one axis and the respective time points on the other yielding an input of 50 genes by 3 time points as depicted in the diagram.

filters with a 2x1 kernel, followed by a third MaxPooling layer. The output of the final MaxPooling layer is then flattened to a vector which passed into a 20 node, fully connected layer with L1 activity regularisation. The output is then up-sampled using a 30 node fully connected layer, followed by a fully connected output layer.

All deep learning models are trained using Adam [44], an adaptive stochastic gradient descent algorithm.

Machine learning models

In the absence of a gold standard in the field, in order to evaluate the added benefit of using deep learning models to predict time series of gene expression, classical machine learning techniques were used as a benchmark. Random Regression Forest (RRF) [45] and k-nearest neighbours (KNN) [46] were selected as our bench-marking models. This bench-marking strategy is in line with the validation methods used in previous studies to assess the added value of deep learning approaches on genomics data [25, 28].

Data availability

All data analysed during the current study are available from the Open TG-GATes database (<https://toxico.nibiohn.go.jp>).

All model scripts are publicly available from the Maastricht Centre for Systems Biology GitHub repository (<https://github.com/macsbio/Deep-Learning-gene-expression>)

Results

Three deep learning models (Naïve Encoder, Modified Autoencoder and CNN) are evaluated for predicting time series of gene expression in a target domain (human *in vitro* or rat *in vivo*) given a measured time series of rat *in vitro* gene expression following exposure to a previously unseen compound.

Across-species prediction – rat *in vitro* to human *in vitro*.

The average mean absolute error for the model predictions of time series of human *in vitro* gene expression given rat *in vitro* gene

expression for the models included in this analysis are summarised in Table 4.1. The Naïve Encoder and CNN consistently outperformed both bench-marking methods (KNN and RRF) for all four toxicologically relevant gene sets included in this study. In fact, the CNN model has a significantly reduced average mean absolute error when compared to the RRF for each gene set, calculated using a paired two-tailed t-test with Benjamini-Hochberg correction for multiple testing within each gene set. The KNN model was consistently the poorest performing model across all gene sets. It is also noteworthy that the average mean absolute error for predicting time series of human *in vitro* gene expression decreases as the number of included genes increases for all methods. The average mean absolute error for the CNN model decreases from 0.0238 units for the cholestasis gene set (18 genes) to 0.01658 units for the GTX+C gene sets(76 genes). Suggesting the additional information added by including more genes improves the model predictions.

Figure 4.6 visualises the prediction of human *in vitro* gene expression for a subsets of genes from the GTX/C gene set using the CNN model for the low, medium, and high dosages of the validation compound hexachlorobenzene. In each row the measured time series of gene expression for *in vitro* rat (red) and both human *in vitro* replicates (blue) following an exposure are visualised. The rat *in vitro* and human *in vitro* gene expression often notably

Table 4.1 : Average mean absolute error from leave one out cross validation for each model predicting the four toxicologically relevant gene sets identified from literature

Gene set number of genes	ref.	Cholestasis 18	NAFLD 22	Steatosis 50	GTX + C 76
k-nearest-neighbour	[46]	0.02836	0.02303	0.02270	0.02076
random regression forest	[45]	0.02335	0.0183	0.01865	0.01718
naive encoder	[37]	0.02326	0.01819	0.01862	0.01667*
modified autoencoder	[40]	0.02507	0.01987	0.01972	0.01768
CNN	[41]	0.02138*	0.01742*	0.01841*	0.01658*

Table 4.1 shows the validation errors of the five models included in this analysis (rows) for predicting human *in vitro* gene expression given rat *in vitro* gene expression for each of the four toxicologically relevant orthologous gene sets identified from literature. The validation errors are the average of the mean absolute errors (mean prediction error across genes) calculated during the leave one out cross validation for the 45 compounds included in this study. Emboldened values indicate the best performing model for each gene set. * indicated models that have a significantly lower average mean absolute error than random regression forest for that gene set as calculated using a paired two-tailed t-test using Benjamini-Hochberg correction for multiple testing.

Figure 4.6 : Input data and model predictions for validation compound hexachlorobenzene for convolutional neural network trained on carcinogenicity/genotoxicity gene set.

The measured time series of rat *in vitro* (red) and human *in vitro* (blue) gene expression for a subset of genes from the GTX+C gene set for a low, medium, and high dosage of hexachlorobenzene are visualised along each row. A visual representation of the model predictions of a time series of human *in vitro* gene expression (yellow) relative to measured human *in vitro* gene expression in both biological replicates (blue) for the same subset of genes from the GTX/C gene set. The model prediction and human *in vitro* biological replicates for BEAN1 and MARCKS are visualised on a closer scale for the low dose exposure.

differ in pattern (CCNE1 and SGK11) demonstrating the potential issues of relying on orthologs to directly translate gene expression patterns between species. The model predicted time series of human *in vitro* gene expression are depicted in yellow. In instances where the measured human *in vitro* replicates have inconsistent gene expression patterns the

model predicts the mean gene expression pattern, as evident when zooming in on the model predictions for BEAN1 and MARCKS for a low dose of hexachlorobenzene in Figure 4.6. A visualisation of the CNN model prediction for the full GTX/C gene set following a low dose exposure to hexachlorobenzene is provided in Supplementary Figs S4.3.

Across experimental technique – rat *in vitro* to rat *in vivo*.

The above analysis was repeated to evaluate the potential for the deep learning models to also predict time series of *in vivo* gene expression given *in vitro* gene expression following an exposure in rats. The average mean absolute errors for the model predictions of time series of rat *in vivo* gene expression are summarised in Table 4.2. Again, the CNN and the Naïve Encoder consistently outperform the classical machine learning methods used for benchmarking, although the RRF does have a lower average mean absolute error than the Naïve Encoder for the cholestasis gene set. While the average mean absolute error values are greater than the human *in vitro* predictions, the error values again decrease as the size of the gene set increases for all models.

Table 4.2 : Average mean absolute error from leave one out cross validation for each model predicting rat *in vivo* gene expression from rat *in vitro* for the four toxicologically relevant gene sets.

Gene set number of genes	ref	Cholestasis 18	NAFLD 22	Steatosis 50	GTX + C 76
k-nearest-neighbour	[46]	0.04991	0.04027	0.03643	0.03140
random regression forest	[45]	0.03934	0.03188	0.02883	0.02479
naive encoder	[37]	0.03977	0.03102	0.02859	0.02437
modified autoencoder	[40]	0.04233	0.03438	0.03056	0.02626
CNN	[40]	0.03771*	0.03040*	0.02849	0.02462

Table 4.2 shows the validation errors of the models included in this analysis (rows) for predicting rat *in vivo* gene expression given rat *in vitro* gene expression for each of the four toxicologically relevant orthologous gene sets identified from literature. The validation errors are the average of the mean absolute errors (mean prediction error across genes) calculated during the leave one out cross validation for the 45 compounds included in this study. The number of genes included in each gene sets is indicated in the second row. Emboldened values indicate the best performing model for each gene set. * indicated models that have a significantly lower average mean absolute error than random regression forest for a given gene set as calculated using a paired two-tailed t-test with correction for multiple testing using Benjamini-Hochberg correction.

Figure 4.7 visualises the prediction of rat *in vivo* gene expression for a subset of genes from the GTX/C gene set using the CNN model for a low, medium, and high dose exposure the previously unseen compound azathioprine. As with the rat and human *in vitro* data, the measured time series of rat *in vitro* (red) and rat *in vivo* (blue) gene expression often differ in pattern (Figure 4.7). The model predicted time series of rat *in vivo* gene expression is visualised in yellow. There is a greater variance in gene expression patterns between the biological replicates for rat *in vivo* than observed in the human *in vitro* (Figure 4.7). As in the cross species predictions, the model often predicts the mean gene expression when the rat *in vivo* biological replicates differ in gene expression patterns (Figure 4.7) BEAN1 and APOM). A graphical representation of the CNN model prediction for all 76 genes in GTX/C gene set following a low dose exposure to azathioprine is available in Supplementary Figs S4.4.

Figure 4.7 : Input data and model predictions for validation compound azathioprine for Convolutional Neural Network trained on GTX+C gene set.

The measured time series of rat *in vitro* (red) and rat *in vivo* (blue) gene expression and the model predictions of rat *in vivo* gene expression (yellow) for a subset of genes from the GTX/C gene set for a low, medium, and high dose exposure of azathioprine. The model prediction and both measured rat *in vivo* biological replicates APOM are visualised on a closer scale for the low dose exposure to demonstrate the discrepancy in gene expression patterns that often exists for the rat *in vivo* biological replicates.

Effect of gene set composition on model predictions.

The toxicologically relevant gene sets identified from literature differed not only in gene composition but also in terms of the numbers of included genes. To evaluate the impact of gene set composition on the accuracy of the cross species prediction of human *in vitro* gene expression from rat *in vitro* gene expression each model was trained using leave-one-out cross validation for ten randomly selected gene sets of increasing size (20, 35, 50, 65, and 80 genes) reflecting the size of the toxicologically relevant gene sets identified from literature. The gene sets were screened to ensure minimal overlap between sets of equal size. There was a considerable variation in average mean absolute error across the ten randomly selected gene sets for all models (Figure 4.8). While the variation decreased, somewhat, as the number of genes increased, it was still marked. Training of each model ten separate times on the same gene set demonstrated that the large variation in prediction accuracy was not due to the randomisation in model initialisation (not shown). This establishes that the particular genes included in a gene set can have a profound effect on the resulting model prediction.

Figure 4.8 : Average mean absolute error for randomly generated gene sets of increasing size for Convolution Neural Network.

Average mean absolute error (AMAE) for CNN trained on ten randomly generated non-orthologous gene sets of increasing size (20, 35, 50, 65, and 80 genes) (blue). The average AMAE for each size of gene set is depicted in orange. Note the distribution of AMAE values for the ten gene sets of each size.

As each random gene set was generated independently, it was not possible to draw strong conclusions on the impact of the number of genes on the model predictions. To this end, 30 nested gene sets of increasing size were randomly generated, providing a more thorough analysis of the impact of added information (increasing number of genes) on the models' predictions. The mean validation error for all models decreases as more genes are added (Figure 4.9). This reflects the results obtained for the toxicologically relevant gene sets, with GTX+C, the largest gene set, yielding the lowest average mean absolute error for all five models. The above analysis was repeated for randomly selected orthologous gene sets from the rat *in vitro* to human *in vitro* prediction and also the *in vivo* predictions and gave comparable results (Supplementary material section 4.2). There appears to be no added benefit in using only orthologous genes.

Figure 4.9 : Average mean absolute error for each model trained on several nested sets of randomly selected genes of increasing size.

Each model included in the analysis (CNN, naïve encoder, modified autoencoder, and random regression forest) were trained on a population of randomly selected nested gene sets of increasing size (20, 35, 50, 60, 80 genes). The figure depicts the mean average mean absolute error for each model trained on a population of thirty randomly generated non-orthologous gene sets of each size. The error bars indicate the standard error of the mean.

Discussion

The advent of genomics technologies, coupled with the growing pressure to reduce the need for animal testing, has contributed to a surge in the development of *in vitro* toxicogenomics bio-assays. While the identification of genomic signatures of adverse toxicological outcomes have yielded positive results in the classification of novel compounds [13-15], relating changes in genomic profile across species or from *in vitro* to *in vivo*, particularly for biological interpretation of underlying mechanisms still proves challenging not only in toxicological research, but also in several other biological disciplines that utilise cellular and animal models. In this study, we demonstrate the potential of deep learning methods to translate time series of drug-induced changes in gene expression across species and experimental domains. The deep learning models evaluated in this study outperform classical machine learning algorithms in predicting time series of human *in vitro* and rat *in vivo* gene expression given a measured time series of rat *in vitro* gene expression following an exposure to a previously unseen compound.

Of the three deep learning models included in this analysis the Naïve Encoder, a fully connected artificial neural network with a bottleneck architecture, and the CNN have the lowest average mean absolute error in both the cross-species and *in vitro* to *in vivo* predictions, consistently outperforming the Random Regression Forest, the best performing traditional machine learning method. CNN are increasingly being investigated for their potential in the analysis of gene expression data [47,48]. Gene expression is a highly complex, co-regulated system, which is currently not fully understood. The kernels implemented in our CNN model have been restricted to filtering several genes at a single time point, identifying patterns of co-expression. By integrating further biological knowledge, it may be possible to apply a more logical structure to the genes in the input data, with neighbouring genes sharing some regulatory relationship. In this way, the local convolution may be more optimally utilised to filter for dependencies in neighbouring genes, improving the prediction accuracy.

Our Modified Autoencoder had consistently low prediction accuracies, outperforming only the KNN method. Use of the weights obtained by pre-training two independent autoencoders to initialise the prediction model may have hindered its training. A recent study applied domain adaptation to combine the pre-trained encoder and decoder weights [49], integration of this ap-

proach in the training of the network may improve the performance of our Modified Autoencoder.

As the number of genes included in the analysis increases, the accuracy of the predictions improves for all models. This is perhaps not surprising as the inclusion of additional genes increases the information from which the models can learn. While the performance of all models, including the classical machine learning approaches, improve with the addition of genes the reduction in average mean absolute error is more pronounced for the CNN and Naïve Encoder, suggesting the deep learning models have an advantage in prediction for larger gene sets. The improved performance is evident in both the toxicologically relevant gene sets identified from literature and the randomly selected gene sets. We have constrained our model predictions to relatively small gene sets consisting of approximately twenty to eighty genes. While, Open TG-GATEs was selected as a pilot data set as it is a comparatively large toxicogenomics data. The number of learning examples that could be generated is far less than the 20,000 genes measured by microarray analysis, necessitating the use of subsets of genes. This does not necessarily hinder our results; the use of subsets of genes, or genomic signatures, to classify potential adverse outcomes is commonplace in toxicological research. In addition, we also do not see any benefit in restricting the cross species prediction task to orthologous genes, with the deep-learning models performing equally as well on sets of randomly selected genes (Figure 4.9) as randomly selected sets of orthologs (Supplementary Fig S4.5).

In order to evaluate the impact of the number of genes on model predictions nested gene sets containing between twenty and eighty genes were generated. This range of values was selected to reflect the numbers of genes in our toxicologically relevant gene sets while also bearing in mind the 720 learning examples available for training. Generally, a ratio of 10 learning examples to number of features is advised, guiding the choice of gene set sizes included in these analyses. Looking at the average mean absolute error values for models trained on these randomly generated nested gene sets it is evident that model performance improves with the addition of more genes. However, we would expect this improvement in validation error to plateau at a certain gene set size as the dimensionality of the feature space would become too large for the limited number of learning examples. The models would then be expected to over-fit the training data and the validation error would increase. This point has not been identified in our analyses.

As this was a provisional study evaluating the potential of deep learning al-

gorithms for this cross-domain prediction task it was decided to restrict training to a single data set of gene expression measurements removing variation in the data due to differing experimental protocols. TG-GATEs was selected as it is a large toxicogenomics data set. Sprague-Dawley rats were used for both *in vitro* and *in vivo* experiments. The primary human hepatocytes used in the experiments included in these analyses were derived from a single human donor. Consequently, the subset of TG-GATEs used here provides a somewhat an idealised scenario for evaluating the performance of the deep learning models greatly reducing variation in data due to experimental technique. Despite this the average mean absolute error values for the *in vivo* prediction are greater than the human *in vitro* prediction (Tables 4.1 and 4.2), we believe this may be largely attributed to the greater variation in the gene expression pattern for the rat *in vivo* replicates when compared to the human *in vitro* replicates. It may be possible to increase the number of genes included in the analysis and improve the accuracy of the models predictions by combining time series of gene expression from other toxicological data bases, such as DrugMatrix [50], thereby increasing the number of learning examples. The variation in gene expression introduced by the differing experimental protocols, and more genetically diverse hepatocytes may further improve the generalisability of the models.

Within the context of this study, we were primarily concerned with the predictive ability of the models. We have not explored the connection weights or structure of the latent space of the trained models. Exploration of the latent spaces of neural networks applied in other classification tasks has provided insight into how the networks learn and make decisions [26]. Due to the bottleneck structure in the networks implemented within this study the models are forced to learn a compressed representation of the data. Accordingly, the time series of gene expression predicted in the target domain is formed by a nonlinear combination of the input gene expression patterns of multiple genes. Exploration of the most influential features in the final, fully trained models, may provide new biological insight into the relationship between genes. In addition, this reduced dimensional latent space may provide a novel method for the classification of potential adverse outcomes as it is anticipated that compounds that trigger similar responses in gene expression would cluster together in the compressed latent space [51,52].

We have demonstrated the ability of deep learning models to translate time series of gene expression across domains; predicting across species, from rat to human *in vitro*, and from *in vitro* to *in vivo* for rats. Our aim in future work would be to train a model to predict human *in vivo* gene expression. While

the simple deep learning architectures evaluated in this study show potential, they require vast quantities of data to train, which would be unavailable for the human *in vivo* domain. It may be possible to further adapt the *in vitro* to *in vivo* models trained on rat data to the human task by integrating some form of transfer learning as has been utilised by Herdon and Caragea to predict protein localisation and splice site locations in a given organism by leveraging large volumes of publicly available labelled data from well studied model organisms [53,54] .

We have demonstrated the potential of deep learning architectures, most notably Convolution Neural Networks and deep fully connected networks, in translating drug-induced changes in gene expression across the species and experimental domains. These deep learning models outperform classical machine learning algorithms in predicting time series of human *in vitro* and rat *in vivo* gene expression given a measured time series of rat *in vitro* gene expression following exposure to a novel compound. More importantly, we highlight the emerging potential of deep learning models in both the analysis of gene expression data and classification of toxicity.

Reference

1. Chhabra RS, Huff JE, Schwetz BS, Selkrk J. An overview of prechronic and chronic toxicity/carcinogenicity experimental study designs and criteria use by the National Toxicology Program. *Environ. Health Perspect.* 1990 86:313-21.
2. Hartung T. Toxicology for the twenty-first century. *Nature.* 2009 460:208-12.
3. van Delft JH, van Agen E, van Breda SG, Herwijnen MH, Staal YC, Kleinjans JC. Discrimination of genotoxic from non-genotoxic carcinogens by gene expression profiling. *Carcinogenesis.* 2004 25:1265-76.
4. van Delft JH, van Agen E, van Breda SG, Herwijnen MH, Staal YC, Kleinjans JC. Comparison of supervised clustering methods to discriminate genotoxic from non-genotoxic carcinogens by gene expression profiling. *Mutat. Res.* 2005 575:17-33.
5. Ellinger-Ziegelbauer H, Aubrecht J, Kleinjans JC, Ahr HJ. Application of toxicogenomics to study mechanisms of genotoxicity and carcinogenicity. *Toxicol. Lett.* 2009 186:36-44.
6. Hu T, Gibson DP, Carr GJ, Torontali SM, Tiesman JP, Chaney JG, Aardem MJ. Identification of a gene expression profile that discriminates indirect-acting genotoxins from direct-acting genotoxins. *Mutat. Res.* 2007 619:16-29.
7. Le Fevre AC, Boitier E, Marchandeu JP, Sarasin A, Thybaud V. Characterization of DNA reactive and non-DNA reactive anticancer drugs by gene expression profiling. *Mutat. Res.* 2007 619:16-29.
8. Li HH, Aubrecht J, Fornace AJ Jr. Toxicogenomics: overview and potential applications for the study of non-covalent DNA interacting chemicals. *Mutat. Res.* 2007 623:98-108.
9. Mathijs K, Brauers KJ, Jennen DG, Lizarraga D, Kleinjans JC, van Delft JH. Gene expression profiling in primary mouse hepatocytes discriminates true from false-positive genotoxic compounds. *Mutagenesis.* 2010 25:561-568.
10. Tsujimura K, Asamoto M, Suzuki S, Hokaiwado N, Ogawa K, Shirai T. Prediction of carcinogenic potential by a toxicogenomic approach using rat hepatoma cells. *Cancer Sci.* 2006 97:1002-1010.
11. Jover R, Bort R, Gomez-Lechon MJ, Castell JV. Re-expression of C/EBP alpha induces CYP2B6, CYP2C9 and CYP2D6 genes in HepG2 cells. *FEBS Lett.* 1998 431:227-30.
12. Rodriguez-Antona C, Donato MT, Boobis A, Edwards RJ, Watts PS, Castell JV, et al. Cytochrome P450 expression in human hepatocytes and hepatoma cell lines: molecular mechanisms that determine lower expression in cultured cells. *Xenobiotica.* 2002 80:580-604.
13. Magkoufopoulou C, Claessen SMH, Tsamou M, Jennen DGJ, Kleinjans JCS, van Delft JHM. A transcriptomics-based in vitro assay for predicting chemical genotoxicity in vivo. *Carcinogenesis.* 2012 33:1421-1429.
14. Caiment F, Tsamou M, Jennen D, Kleinjans J. Assessing compound carcinogenicity in vitro using connectivity mapping. *Carcinogenesis.* 2014 35:201-207.
15. Ryaboshapinka M, Hammar M. Human hepatic gene expression signature of non-alcoholic fatty liver disease progression, a meta-analysis. *Scientific Reports.* 2017 7:12361.
16. Van Delft J, Matijs K, Polman J, Coonen M, Szalowska E, et al. Toxicogenomics-Based Cellular Models - Chapter 5.1 - Hepatotoxicity Screening on In Vitro Models and the Role of 'Omics. *Academic Press.* 2014 193-212 ISBN 9780123978622.
17. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015 521:436-444.
18. Angermueller C, Parnamaa T, Parts L, Stegle O. Deep learning for computational biology. Molecular systems biology. *Mol. Sys. Bio.* 2016 12:878.
19. Lee H, Grosse R, Ranganath R, Ng AY. Unsupervised learning of hierarchical representa-

- tions with convolutional deep belief networks. *Commun. ACM*. 2011 54:95–103.
20. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagation errors. *Nature*. 1986 323:533–536.
 21. Socher R, Lin CCY, Ng AY, Manning CD. Parsing natural scenes and natural language with recursive neural networks. Proceedings of the 28th International Conference on Machine Learning (ICML-11) 2011 129–136.
 22. Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks. Advances in Neural Information Processing Systems 27 (NIPS 2014).
 23. Baskin I.I. Machine Learning Methods in Computational Toxicology. In: Nicolotti O. (eds) Computational Toxicology: Methods in Molecular Biology, vol 1800. (2018) Humana Press, New York, NY.
 24. Mayr A, Klambauer G, Unterthiner T, Hochreiter S. DeepTox: Toxicity prediction using Deep Learning. *Font. Environ. Sci.* 2016 3:80.
 25. Idakwo G, Thangapandian S, Luttrell IV J, Zhou Z, Zhang C, Gong P. Deep learning-Based Structure-Activity Relationship Modelling for Multi-Category Toxicity Classification: A case study of 10k Tox21 Chemicals with high-throughput cell-based androgen receptor bioassay data. *Front. Physiol.* 2019 10:1044.
 26. Xu Y, Dai Z, Chen F, Gao S, Pei J, Lai L. Deep learning for drug-induced liver injury. *J. Chem. Inf. Model.* 2015 55:2085–2093.
 27. Pu, L., Naderi, M., Liu, T. et al. eToxPred: a machine learning-based approach to estimate the toxicity of drug candidates. *BMC Pharmacol Toxicol.* 2019 20: 2.
 28. Singh R, Lanchantin J, Robins G, Qi Y. DeepChrome: deep-learning for predicting gene expression from histone modifications. *Bioinformatics*. 32, i639–i648, (2016).
 29. Chen Y, Li Y, Narayan R, Subramanian A, Xie X. Gene expression inference with deep learning. *Bioinformatics*. 2016 32:1832–9.
 30. Aliper A, Plis S, Artemov A, Ulloa A, Mamoshina P, Zhavoronkov A. Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. *Mol. Pharm.* 2016 13:2524–2530.
 31. Wang H, Liu R, Schyman P, Wallqvist A. Deep Neural Network Models for Predicting Chemically Induced Liver Toxicity Endpoints from Transcriptomics Responses. *Front. Pharmacol.* 2019 10:42.
 32. Igarashi Y, Nakatsu N, Yamashita O, Ono A, Urushidani T, Yamada H. Open TG-GATES: a large-scale toxicogenomics database. *Nucleic Acids Res.* 2015 43:21–7.
 33. Jung D, Elferink MGL, Stellaard F, Groothuis GMM. Analysis of bile acid-induced regulation of FXR target genes in human liver slices. *Liver Int.* 2007 27(1):137–144.
 34. Vinken M, Landesmann B, Goumenou M, Vinken s, Shah I, Jaeschke H, Willett C, Whelan M, Rogiers V. Development of an Adverse Outcome Pathway From Drug-Mediated Bile Salt Export Pump Inhibition to Cholestatic Liver Injury. *Toxicol. Sci.* 2013 136(1):97–106.
 35. Liu J, Lu H, Lu YF, Lei X, Cui JY, Ellis E, Strom SC, Klaassen CD. Potency of individual bile acids to regulate bile acid synthesis and transport genes in primary human hepatocyte cultures. *Toxicol. Sci.* 2014 141:538–46.
 36. Rosenblatt F. The Perceptron — A Perceiving and Recognizing Automaton. Tech. Rep. 85-460-1 (Cornell Aeronautical Laboratory, 1957).
 37. Kramer MA. Nonlinear principal component analysis using autoassociative neural networks. *AIChE Journal*. 1991 37:233–243.
 38. Nair V, Hinton GE. Rectified Linear Units Improve Restricted Boltzmann Machines. Proceedings of the 27th International Conference on International Conference on Machine Learning (ICML/10) 807–814 (2010)
 39. Glorot X, Bordes Y, Bengio J. Deep Sparse Rectifier Neural Networks. *PMLR* 2011 15:315–323.

-
40. Hinton GE, Salakhutdinov RR. Reducing the Dimensionality of Data with Neural Networks. *Science*. 2006 313:504-507.
 41. Hubel DH, Wiesel TN. Receptive fields of single neurons in the cat's striate cortex. *J Physiol*. 1954 148(3):574-59.
 42. Fukushima K. Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position. *Biological Cybernetics*. 1980 36(4):193-202.
 43. Weng J, Ahuja N, Huang TS. Learning recognition and segmentation of 3-D objects from 2-D images. Proc. 4th International Conf. Computer Vision: 121-128.
 44. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. *ICLR* (2014)
 45. Breiman L. Random forests. *Mach Learn*. 2001;45:5-32.
 46. Cover TM, Hart P. Nearest neighbor pattern classification. *IEEE Trans Inf Theory*. 1967 13:21-27
 47. Dutil F, Cohen JP, Weiss M, Derevyanko G, Bengio Y. Towards Gene Expression Convolutions using Gene Interaction Graphs. arXiv:1806.06975v1 (2018)
 48. Yuan Y, Bar-Joseph Z. Deep learning for inferring gene relationships from single-cell expression data, bioRxiv 2019, doi: <https://doi.org/10.1101/365007>
 49. Chai X, Wang Q, Zhao Y, Liu X, Bai O, Li Y. Unsupervised domain adaptation techniques based on auto-encoder for non-stationary EEG-based emotion recognition. *Comput Biol Med*. 2016 79:205-214.
 50. Ganter B, Tugendreich S, Pearson CI, Ayaboglu E, Baumhueter S, Bostian KA et al. Development of a large-scale chemogenomics database to improve drug candidate selection and to understand mechanisms of chemical toxicity and action. *Biotechnol*. 2005 119:219-244.
 51. Stein-O'Brien GL, Clark BS, Sherman T, Zibetti C, Hu Q et al. Decomposing cell identity for transfer learning across cellular measurements, platforms, tissues, and species. *Cell Syst*. 2019 8, 395-411.
 52. Taroni JN, Grayson PC, Hu Q, Eddy S, Kretzler M et al. Multi-PLIER: A transfer learning framework for transcriptomics reveals systemic features of rare disease. *Cell Syst*. 2019 8,380-394.
 53. Mourragui S, Loog M, Reinders MJT, Wessels LFA. PRECISE: A domain adaptation approach to transfer predictors of drug response from pre-clinical models to tumors. bioRxiv 536797; doi:<https://doi.org/10.1101/536797>
 54. Herndon N, Caragea D. Predicting protein localization using domain adaptation approach. CCIS 452, Springer-Verlag Berlin Heidelberg, 191-206 (2014).
 55. Herndon N, Caragea D. Empirical Study of Domain Adaptation with Naive Bayes on the Task of Splice Site Prediction. BIOINFORMATICS (BIOSPEC 2014) 2014 1:57-67.

Supplementary material

Supplemental Figure S4.1 : Pairing strategy for generating machine learning examples.

Machine learning examples are generated by pairing the time series of gene expression in the source and target domain for each compound-dose combination. Both biological replicates for a given compound-dose combination in the target domain are a valid pair for each biological replicate for that compound-dose combination in the source domain. Pairwise matching of the biological replicates generates four machine learning examples for each compound-dose combination (illustrated with the yellow arrows above). Treating the controls as an additional dose; with 45 compound * 3 dosages plus the control * pairwise matching of biological replicates creates 720 learning examples.

Supplemental Figure S4.2a : Error calculation using mean absolute error for raw normalised gene expression values

The above figure depicts two model predictions for a measured time series of human *in vitro* gene expression (blue). The gene expression values for model prediction 1 (green) are very close in magnitude to the measured time series of human *in vitro* gene expression, however, the gene expression pattern of Model prediction 1 differs greatly to the measure human *in vitro* gene expression pattern. Model prediction 2 (yellow) has the same gene expression pattern as the time series measured human *in vitro* gene expression, but it is shifted upwards. Evaluating the quality of the model predictions using a classical distance based cost function, such as sum of absolute errors, on the raw gene expression level values we see that model prediction 1 would have a much lower error value despite having such a different pattern in gene expression. Despite model prediction 2 predicting the correct gene expression pattern the penalty for incorrectly predicting the gene expression level at the first time point is also added to each subsequent time point.

Supplemental Figure S4.2b : Error calculation using mean absolute error for slope re-encoded gene expression values

The gene expression data is re-encoded, the first entry remains unchanged, subsequent entries now encode the slope, change in gene expression levels, between consecutive time points. Applying the same error calculation as before (mean absolute error) we see that while model prediction 2 is penalised for incorrectly predicting the gene expression value for time point one as it correctly predicts the changes in gene expression between time point one and two, and time point two and three, it receives no further penalty. While model prediction 1 is closer in predicting the level of gene expression for time point one as it fails to predict the correct changes in gene expression level for subsequent time point it has a greater error over all. In this way, model predictions are not unduly penalised for incorrectly predicting a single time point. By re-encoding the data in this manner the models are trained to correctly predict the gene expression pattern without the need to introduce a more complex error term.

Supplemental Figure S4.3 : Measured and model predicted time series of rat *in vitro* to human *in vitro* gene expression for the full GTX/C gene set for the validation compound hexachlorobenzene (low dose) for the CNN

Measured time series of rat *in vitro* gene expression (red) following an exposure to a low dose of the validation compound hexachlorobenzene for the full GTX/C gene set (76 genes, known rat-human orthologs). The CNN model predictions of human *in vitro* gene expression are in yellow. Both biological replicates for the time series of human *in vitro* gene expression are depicted in blue for reference.

Supplementary Material Section S4.2 - additional analyses

Nested sets of non-orthologous genes - rat *in vitro* to rat *in vivo*

As with the human *in vitro* predictions, as the models are trained on larger gene sets the average mean absolute error decreases for all models. The average mean absolute error values are greater for the rat *in vivo* predictions than for the human *in vitro* predictions. Higher average mean absolute error values for the rat *in vivo* predictions were also observed for the toxicologically relevant gene sets identified from literature.

Supplemental Figure S4.5 : Average mean absolute error in predicted rat *in vivo* gene expression for each model trained on several nested sets of randomly selected genes of increasing size

Each model included in the analysis (CNN, naïve encoder, modified autoencoder, and random regression forest) were trained to predict rat *in vivo* gene expression from rat *in vitro* gene expression on a population of randomly selected nested gene sets of increasing size (20, 35, 50, 60, 80 genes). The figure depicts the mean average mean absolute error for each model trained on a population of ten randomly generated non-orthologous gene sets of each size. The error bars indicate the standard error of the mean.

Nested sets of known rat-human orthologs – rat *in vitro* to human *in vitro*

Orthologs are two, or more, homologous gene sequences found in different species related by linear decent. Orthologs are commonly utilized to relate results from rodent *in vivo* and *in vitro* bio assays to the human system. In order to evaluate any added benefit of using known orthologs when predicting times series of human *in vitro* gene expression given a time series gene expression the models were also applied to randomly selected nested gene sets of increasing size of known rat human orthologs. The range of average mean absolute error values for predicting human *in vitro* gene expression from rat *in vitro* gene expression using known rat-human orthologs(Figure S6 below) shows no improvement over using randomly generated non-orthologous gene sets (Figure 8 in the text) This indicates there is no advantage in restricting gene predictions to known rat-human orthologs. This is unsurprising as all three deep learning models implemented in this system contain a bottleneck in their architecture. As a result, model predictions of a human gene expression pattern are made by a non-linear combination of the input rat gene expression patterns. As with the randomly selected non-orthologous gene sets, as more genes are included the average mean absolute error decreases for all models. Again, the convolutional neural network consistently out performs the random regression forest, our benchmarking classical machine learning method.

Supplemental Figure S4.6 : Average mean absolute error in predicted human *in vitro* gene expression for each model trained on several nested sets of randomly selected rat-human orthologs of increasing size

Each model included in the analysis (CNN, naïve encoder, modified autoencoder, and random regression forest) were trained to predict human *in vitro* gene expression from rat *in vitro* gene expression on a population of randomly selected nested gene sets of increasing size (20, 35, 50, 60, 80 genes). The figure depicts the mean average mean absolute error for each model trained on a population of thirty randomly generated gene sets of known rat-human orthologs of each size. The error bars indicate the standard error of the mean.

Supplementary material 4.3 : Data information

Supplemental Table S4.1 : Subset of compounds from TG-GATEs included in this analysis [1]

compound	abbreviation	Carcinogenicity	Genotoxicity
Acetaminophen	APAP	C	GTX
Adapin	ADP	unknown	unknown
Allopurinol	APL	unknown	GTX
Allyl alcohol	AA	unknown	unknown
Aspirin	ASA	NC	NGTX
Azathioprine	AZP	C	GTX
Benzbromarone	BBr	unknown	unknown
Bromobenzene	BBZ	unknown	NGTX
Carbamazepine	CBZ	unknown	unknown
Chlorpromazine	CPZ	unknown	NGTX
Cimetidine	CIM	NC	unknown
Clofibrate	CFB	C	GTX(in vitro), NGTX(in vivo)
Coumarin	CMA	C	GTX(in vitro), NGTX(in vivo)
Cyclophosphamide	CPA	C	GTX
Diazepam	DZP	NC	GTX
Diclofenac	DFNa	NC	NGTX
Ethionine	ET	unknown	unknown
Fluphenazine	FP	unknown	GTX
Flutamide	FT	unknown	unknown
Gemfibrozil	GFZ	unknown	unknown
Glibenclamide	GBC	unknown	unknown
Griseofulvin	GF	C	GTX(in vitro), NGTX(in vivo)
Haloperidol	HPL	unknown	NGTX
Hexachlorobenzene	HCB	C	GTX
Indomethacin	IM	unknown	unknown
Isoniazid	INAH	C	GTX
Ketoconazole	KC	unknown	unknown
Labetalol	LBT	unknown	NGTX
Lomustine	LS	C	GTX
Methapyrilene	MP	unknown	GTX
Methyltestosterone	MTS	unknown	unknown

compound	abbreviation	Carcinogenicity	Genotoxicity
Naphthyl isothiocyanate	ANIT	unknown	NGTX
Nitrofurantoin	NFT	C	GTX(in vitro), NGTX(in vivo)
Omeprazole	OPZ	unknown	NGTX
Perhexiline	PH	unknown	unknown
Phenobarbital	PB	C	GTX(in vitro), NGTX(in vivo)
Phenylbutazone	PhB	C	GTX(in vitro), NGTX(in vivo)
Phenytoin	PHE	C	NGTX
Propylthiouracil	PTU	C	NGTX
Rifampicin	RIF	unknown	unknown
Sulfasalazine	SS	C	NGTX
Tetracycline	TC	C	GTX
Thioacetamide	TAA	C	GTX
Thioridazine	TRZ	unknown	unknown
Valproic acid	VPA	unknown	NGTX
WY-14643	WY	C	NGTX

Overview of gene sets included in these analyses.

Supplemental Table S4.2 : Cholestasis gene set identified from literature [2-4]

HNF4A	SLC10A1	SLCO1B1	CYP7A1	CYP8B1
CYP27A1	CYP7B1	NR1H4	NR0B2	N21L2
NR1L3	FGF19	ABCB11	SLC51A	ABCC3
	UGT2B4	CYP3A4	SULT2A1	

Supplemental Table S4.3 : NAFLD gene set identified from literature [5]

PPARGC1A	IL6	SERPIN1	IL1B	STAT3	TCF7L2
PNPLA3	PEMT	TM6SF2	SREBF1	HFE	SAMM50
FDFT1	NR1I2	PPARA	PPP1R3B	CHDH	LYPLAL1
		SOD2	LEPR		

Supplemental Table S4.4 : Steatosis gene set; generated in house

FABP4	ACACA	AKT1	AKT2	AKT3	PRKAA1	PRKAA2	ADIPOR1
ADIPOR2	ADIPOQ	BCL2A1	CPT2	CPT1A	CPT1C	CASP8	MLXIPL
FABP5	ELOVL3	FAS	FOXO1	NR1H4	RXRA	FASLG	FABP3
FABP7	PMP2	GCKR	IL1A	IL10	IRS1	IRS2	MAPK10
NFKB1	NFKB2	RELA	RELB	PPARA	PPARG	PTEN	RXRΒ
RXRG	SCD	SOCS3	SREBF1	TGFB1	TGFB2	TGFB3	TLR4
		MTOR			PNPLA3		

The steatosis gene set was generated in house combining a literature search using the search terms “liver steatosis”, “Nonalcoholic fatty liver disease”, and “NAFLD”, the steatosis pathway from KEGG (hsa04932) and the steatosis adverse outcome pathway from Wikipathways. In addition genes were filtered to include only known human –rat orthologs measured by both the Rat Genome 230 2.0 Array and the Human Genome U133 Plus 2.0 Array used in this study. This gave rise to a seed gene set of 45 genes, these were then used an input for input for MetaCore (version 6.30, build 68780, accessed on 9th of May 2017) to generate a fully connected gene interaction network. Dijkstra’s shortest path algorithm is used to construct the network with allowing one node to be added if necessary. This analysis yielded the final set of 50 genes.

Supplemental Table S4.5: gene set reported as being a genomic signature of genotoxicity/carcinogenicity [6,7]

CEACAM1	CLCN4	EML1	PWWP2B	UBE2E2	USP13	GMFG
PROSC	TTR	NR0B2	NAT8	RBPMS	TBC1D9	SNX11
BCOR	ROBO2	DENND6B	APOM	NR1P3	PITHD1	AVEN
ZNRF3	BEAN1	SLC27A1	ANXA6	APOA4	BTD	EIF2D
AGFG1	NDUFA1-	NFATC3	PLAA	FAN1	SLC40A1	ANAPC5
MRPS5	GSTK1	HOGA1	FGA	SGK1	SLC6A4	SCRN2
CC2D1B	GPC3	MDK	COL5A2	TP53BP2	XPO1	AFP
CCNA2	CCNE1	COL1A1	COL4A1	CTNNB1	FBN1	FOXM1
STMN1	LGALS3BP	MARCKS	NME1	NRAS	PGK1	MAPK3
SMARCC1	COPS5	PEG10	HGFAC	IGFALS	LCAT	SLC22A1
ACADS	ACADVL	C9	DSG2	PLG	HAMP	

References

1. Igarashi, Y., Nakatsu, N., Yamashita, T., et al. Open TG-GATEs: a large-scale toxicogenomics database. *Nucleic Acids Res.* 2014;43(Database issue):D921-7.
2. Jung, D., Elferink, M. G. L., Stellaard, F., Groothuis, G. M. M.. "Analysis of bile acid-induced regulation of FXR target genes in human liver slices," *Liver Int.*, 27(1), 137–144 (2007).
3. Vinken, M., Landesmann, B., Goumenou, M., Vinken, S., Shah, I., Jaeschke, H. et al. "Development of an Adverse Outcome Pathway From Drug-Mediated Bile Salt Export Pump Inhibition to Cholestatic Liver Injury," *Toxicol. Sci.*, 136,. 97–106, (2013).
4. Liu, J., Lu, H., Lu, Y.-F., Lei, X., Cui, J., Ellis, J. Y E et al. "Potency of individual bile acids to regulate bile acid synthesis and transport genes in primary human hepatocyte cultures.," *Toxicol. Sci.*, 141, 538–46 (2014).
5. Ryaboshapinka, M., Hammar, M. "Human hepatic gene expression signature of non-alcoholic fatty liver disease progression, a meta-analysis." *Scientific Reports* 7:12361(2017)
6. Magkoufopoulou, C., Claessen, S.M.H., Tsamou, M., Jennen, D.G.J., Kleijnans, J.C.S., van Delft, J.H.M. "A transcriptomics-based in vitro assay for predicting chemical genotoxicity in vivo", *Carcinogenesis*, 33, 1421–1429,(2012)
7. Caiment, F., Tsamou, M., Jennen, D., Kleijnans J. Assessing compound carcinogenicity in vitro using connectivity mapping, *Carcinogenesis*, 35, 201–207, (2014)

5

Use of domain adaptation to predict human in vivo drug-induced gene expression changes using rat in vitro and in vivo gene expression data

Shauna D. O'Donovan¹, Rachel Cavill², Florian Wimmenauer², Alexander Lukas², Tobias Stumm², Evgueni Smirnov², Michael Lenz^{1,3,4}, Gokhan Ertaylan^{1,5}, Danyel G.D. Jennen⁶, Natal A.W. van Riel^{1,7}, Kurt Driessens², Ralf L.M. Peeters^{1,2}, Theo M.C.M de Kok^{1,6}

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

²Dept. of Data Science & Knowledge Engineering, Maastricht University, the Netherlands.

³Institute of Organismic and Molecular Evolution, Johannes Gutenberg University Mainz, Germany. ⁴ Preventive Cardiology and Preventative Medicine - Center for Cardiology, University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany.

⁵Flemish Institute for Technological 1 Research (VITO), Mol, Belgium. ⁶Dept. of Toxicogenomics, GROW School of Oncology and Developmental Biology, Maastricht University, the Netherlands. ⁸Dept. of Biomedical Engineering, Eindhoven University of Technology, the Netherlands.

In preparation

Abstract

The liver is the primary site for the metabolism and detoxification of pharmaceuticals and other compounds and , consequently, is also the primary location for any adverse outcomes. As the liver is not readily accessible for sampling in humans, rodents and human cell line models are often used to evaluate potential toxic effects of a novel compound or candidate drug. With the advent of toxicogenomics bioassays, it is now possible to measure genome wide alterations in gene expression in response to an exposure to a compound allowing for mechanistic insight into modes of toxicological action within the liver. However, relating the results of animal and *in vitro* studies to relevant clinical outcomes for the human *in vivo* situation still proves challenging. The comparative scarcity of relevant human *in vivo* exposure data further compounds the difficulty in effectively training computational a model to relate findings from rodent and *in vitro* data to the human *in vivo* state. In this study we integrate transfer learning techniques with deep artificial neural networks using the method proposed by Ganin et al., allowing us to leverage the relative abundance of rat *in vitro* and *in vivo* exposure data from the Open TG-GATEs data set to train a model that could be applied to translate changes in human *in vitro* gene expression measured following an exposure to the human *in vivo* state. We show that domain adaptation has been successfully achieved, with the rat and human *in vitro* data no longer being separable in the common latent spaceof the network. The network produces physiologically plausible predictions of the human *in vivo* gene expression pattern following an exposure to a previously unseen compound, however due to the lack of time series of human *in vivo* hepatic gene expression we were unable to validate our predictions. Integration of the human *in vitro* data when training the domain adaptation network significantly improves the accuracy of the model predictions of rat *in vivo* gene expression pattern following an exposure to a previously unseen compound. The reduced dimensional representation of the time series data was also evaluated as a novel method of classifying compound toxicity, however we did not achieve a good separation between the carcinogenic and non-carcinogenic compound.

Introduction

Drug induced liver injury (DILI) is a leading cause of the withdrawal of a drug from the market or the failure of a novel candidate drug at end stage clinical trials [1]. This is despite the compounds having already undergone a battery of costly and time-consuming tests prior to human testing. The current gold standard for evaluating a compound for potential adverse outcomes in humans is the two-year rodent *in vivo* bioassay. However, a survey published in 2000 that evaluated the concordance of toxicity observed in humans with the results of animal testing using data from twelve pharmaceutical companies reported that just 43% of toxic effects of pharmaceutical compounds in human were correctly predicted by tests in rodents [2]. Moreover, the two-year rodent *in vivo* bioassay is expensive, time consuming, and requires sacrificing of large numbers of animals to screen a single compound. The apparent lack of sensitivity and specificity in predicting potential adverse outcomes in humans, coupled with growing ethical concerns surrounding animal testing, has motivated the developments of alternatives to the traditional rodent *in vivo* bioassays, most notably *in vitro* cellular models [3]. These *in vitro* toxicogenomics approaches often make use of human cell lines, eliminating the need for animal testing while aiming to provide a more relevant prediction of adverse effects for the human system.

Indeed, multiple studies have reported promising results in differentiating between subclasses of carcinogenicity [4] and predicting hepatotoxicity of a novel compound using genomic signatures of human *in vivo* disease states [5-12]. Nevertheless, *in vitro* assays are not without their limitations. These cell systems lack the systemic interplay with other tissues that exists *in vivo*. Consequently, the *in vitro* models may differ greatly in functionality and metabolism from the tissues they represent. Immortalised human cell lines, such as hepatic HepaRG and HepG2 cells, are frequently used in toxicity testing as they continue to grow and divide indefinitely *in vitro* [13]. However, these cell lines are usually tumour derived and liver specific metabolic functions tend to vanish as culture time increases [14, 15]. Therefore, genomic signatures of obtained from cell lines following exposure may not necessarily reflect human *in vivo* disease states, particularly if we are trying to derive mechanistic insight into modes of action of toxicity and beyond simple classification of potential toxicants. Thus, there is a need for new methods that can relate the output from these *in vitro* exposure assays to potentially relevant human *in vivo* disease states.

In **Chapter 4** we had notable success in applying deep learning architectures to translate time series of hepatic gene expression following an exposure from one domain to another, predicting both human *in vitro* and rat *in vivo* gene expression patterns given a measured time series of rat *in vitro* gene expression following an exposure to a previously unseen compound. However, such deep learning models require large volumes of data upon which to train. While large data bases of both *in vitro* and rat *in vivo* hepatic gene expression following exposures to a vast array of compounds are available [16,17], human *in vivo* data is comparatively sparse as the liver is not a readily accessible tissue for sampling. Consequentially, there is insufficient human *in vivo* data available to effectively train such a deep learning model for human *in vivo* gene expression predictions.

In several biological applications, it may be laborious, costly, or even infeasible to obtain sufficient quantities of data to train a reliable predictor, as in the case of human hepatic gene expression. Transfer learning is a branch of machine learning in which knowledge gained from solving one problem is then utilised while solving another, similar, problem [18]. Multiple studies have applied a range of transfer learning and domain adaptation algorithms to the cross species prediction task [19,20,21] with mixed success. Transfer learning approaches have been applied to translate gene expression measured at a single time point from clinical *in vitro* models to primary human tumour profiles in order to better predict mutation status of the tumour [22] or predict responsiveness of a patient to a particular treatment [23].

In 2015, Ganin et al. proposed a method of integrating domain adaptation and deep feature prediction in the context of a generalisable neural network architecture [24]. This method allows for the training of a predictor model for an unlabelled target data set using a large, similarly distributed, labelled data set. This unsupervised domain adaptation (UDA) is achieved through the addition of a domain predictor branch from an internal hidden layer of the neural network trained to maximise the loss in classifying the domain of the input data. The introduction of a gradient reversal layer effectively coverts the maximisation of loss in domain classification to a minimisation problem, allowing the domain classifier to be trained in tandem with the predictor model, resulting in the emergence of features that are indiscriminate to the domain of the data but sufficiently discriminative for the prediction task. Ganin et al. applied their UDA algorithm to a range of deep neural network architectures, with their unsupervised domain adaptation method outperforming the state of the art in image processing and sentiment analysis of natural language [24].

In this study, we integrate domain adaptation within the deep learning architectures developed in **Chapter 4** using the method proposed by Ganin et al. We then apply the resulting domain adaptation network to leverage a large publicly available data set of measured rat *in vitro* and *in vivo* gene expression following an exposure to a range of compounds to facilitate the training of a predictor model of human *in vivo* gene expression from measured human *in vitro* gene expression. We evaluate the impact of incorporating the human *in vitro* data on the rat *in vitro* to rat *in vivo* prediction. Finally, we explore the potential of the reduced dimensional representation of the data generated by the bottleneck architecture of our deep neural network to classify compounds based on toxicity.

Materials & Method

Open TG-GATEs

Open TG-GATEs is a large, publicly available, toxicogenomics database containing gene expression profiles from *in vitro* assays in both primary rat and primary human hepatocytes and *in vivo* rats following exposure to 170 compounds collected over the course of ten years [16]. For the *in vitro* exposure studies gene expression profiles were measured at three time points (2, 8, and 24 hours) following a single exposure to a given compound at up to three dosages (low, medium, and high) plus a control, with two biological replicates for each compound-dose combination. The ratio of the compound dosages for the low, medium, and high dose levels were set at 1:3:10, where the high dosage for each compound is determined at the maximally tolerated dose (yielding a survival ratio of 80-90%). For the rat *in vivo* experiments gene expression profiles were measured at four time points (3, 6, 9, and 24 hours) following a single exposure to a compound. The ratio of the dose levels for low, medium, and high compound dosages for the *in vivo* studies are 1:5:25, with the high dose selected to match the level shown previously to induce the minimum toxic effects over the course of a four-week toxicity study. Gene expression profiles for the rat *in vitro* and *in vivo* samples were generated using the Affymetrix Rat Genome 230 2.0 Array, human *in vitro* gene expression profiles were measured using the Affymetrix Genome U133 Plus 2.0. Array data for all available compounds for rat and human *in vitro* and rat *in vivo* were downloaded in the form of CEL files from the Open TG GATEs database (<https://toxico.nibiohn.go.jp>) and pre-processed using Affymetrix Power Tools using the robust multi-array average normalisation method. Following normalisation, compounds with

incomplete data (missing either time points or dosages) were removed, leaving 45 compounds with a complete set of data used in this study.

Figure 5.1 : Overview of toxicogenomics data from TG-GATEs included in this study

Overview of toxicogenomics measurements available from the TG-GATEs data based used in this study. Gene expression profiles were measured for primary rat and primary human hepatocytes exposed *in vitro* at three time points (2, 8, and 24 hours) following exposure to 45 compounds at three dosages (low, medium, and high) plus a control. Two biological replicates were conducted for each compound-dosage combination. Gene expression profiles were measured at four time points (3, 6, 9, and 24 hours) for rat hepatic tissue exposed *in vivo* to 45 compounds at three dosages (low, medium, and high) plus a control. Three biological replicates are conducted for each rat *in vivo* exposure. The rat *in vitro* and *in vivo* gene expression data form the source domain. The rat data is used to train a model to predict human *in vivo* gene expression given a measured time series of human *in vitro* gene expression following exposure to a compound (the unlabelled target domain).

Learning examples

The rat *in vitro* and *in vivo* data form the source domain, will be used to train a predictor model for the unlabelled human *in vitro* target domain (Figure 5.1). Machine learning examples for the source domain are generated by pairing the time series of gene expression for a compound-dose combination for the rat *in vitro* data with a time series of gene expression for the same compound-dose combination for the rat *in vivo* data, as done in **Chapter 4**. As each rat *in vitro* replicate for a given compound-dose combination is a valid match for both rat *in vitro* biological replicates, four learning examples can be generated for each compound-dose combination (Chapter 4, Supplementary Figure S4.1). For the purposes of generating learning examples, the controls are treated as an additional dose. In order to maintain the structure of the data in each domain (human and rat *in vitro* and rat *in vivo*) one of the three replicates for the rat *in vivo* domain is randomly discarded. With 45 compounds, three dosages plus the control, and the pairwise matching of replicates it is possible to generate 720 learning examples for the source domain. The human *in vitro* data (the target domain) is processed in the same manner as the rat *in vitro* data generating 720 learning examples consisting of time series of unlabelled human *in vitro* gene expression.

Encoding the gene expression data

Micro-array data measures the expression of more than 20,000 gene transcripts, resulting in a high dimensional feature space. While Open TG-GATEs is a comparatively large database of toxicogenomics data, the 720 machine learning examples that can be generated from the data is somewhat restrictive for effectively training a model to predict genome-wide gene expression. Consequently, it was decided to restrict our analyses to toxicologically relevant subsets of genes that had previously been identified from literature in **Chapter 4**. Four subsets of genes with associations to relevant toxicological outcomes had been identified, namely steatosis (developed in-house from the KEGG pathway hsa04932, Supplementary Table S4.3), cholestasis [25-27], genotoxicity and carcinogenicity (GTX+C) [5,28,29], and non-alcoholic fatty liver disease (NAFLD) [30]. All gene lists have been filtered to contain only known rat-human orthologs. Complete gene sets are listed in Supplementary Tables S4.2-S4.5.

Model

Previous work, in **Chapter 4**, has demonstrated the ability of deep artificial neural networks (ANN) with a bottleneck architecture to out-perform classical machine learning techniques in translating time series of gene expression from rat to human and from *in vitro* to *in vivo* in rats. Here, Ganin et al.'s unsupervised domain adaptation [24] is applied to the previously validated ANN architecture to train a model to predict human *in vivo* gene expression from measured human *in vitro* gene expression (target domain) using a large measured data set of rat *in vitro* and *in vivo* gene expression following exposure to a variety of compounds (source domain). Domain adaptation is achieved through the introduction of a domain classification arm from a central hidden layer (Figure 5.2). The network is trained to maximise the loss in predicting the domain label of the input data (rat or human *in vitro*). In this way, the network constructs a common latent space for the rat and human *in vitro* data which is indiscriminate to the domain of the data. The inclusion of a gradient reversal layer, which leaves the input unchanged during forward propagation but reverses the gradient during back propagation by multiplying the gradient by a negative scalar, allows the domain classifier to be trained in tandem with the *in vivo* prediction using the standard backpropagation algorithm [31]. This results in the emergence of features that are invariant to the domain of the data but sufficiently discriminative for the prediction task, producing a predictor model of time series of both rat and human *in vivo* gene expression.

The structure of the deep neural network was optimised using a grid search. The final network consists of five hidden layers containing 135, 96, 64, 120, 190 nodes respectively (Figure 5.2). All layers use rectified linear unit (ReLU) activation [32], except for the output layer, which uses sigmoid activation producing output values are between 0 and 1. A single domain classification layer consisting of eight nodes is introduced at the central hidden layer (Figure 5.1). The gradient reversal parameter lambda increases at a logarithmic rate during training from zero to one, as in the original publication [24]. The domain classification error is calculated using soft max cross entropy, as in the original publication [24]. The prediction error is calculated as the sum of absolute errors between the measured rat *in vivo* gene expression and the model predicted gene expression pattern.

The model is trained using Momentum [33], a stochastic-gradient decent algorithm that accelerates convergence to an optimum solution by accumulating gradients from previous steps.

Figure 5.2 : Schema of the implemented unsupervised domain adaptation network

Domain labels are appended to the measured time series of rat (source domain) and human (target domain) *in vitro* gene expression to form the input for the model. The bottleneck architecture of the deep neural network finds a reduced dimension representation of the data. A simple, one layer, classifier is trained to predict the domain labels (rat or human) from the 64 dimensional latent space. The inclusion of a gradient reversal layer in this domain classifier multiplies the gradient by a negative scalar in the backpropagation step. Thereby, maximising the loss in predicting the domain label, resulting in domain adaptation as the model cannot discriminate between the source and target domains. The model is simultaneously trained to reconstruct *in vivo* gene expression patterns using the measured time series of rat *in vitro* gene expression. The network architecture was optimised using a grid search to find a well performing network. The final network, depicted above, was found to contain five layers consisting of 135, 90, 64, 120, and 190 nodes respectively with a domain classification layer containing 8 nodes. All layers use ReLU activation, except for the output layer which uses sigmoid activation.

Experimental setup

In order to assess how well the model would perform for a previously unseen compound, while still maximising the number of learning examples available for training, leave-one-compound-out cross validation is used. All sixteen instances for a given compound were removed from both the source and target domain and the model was trained on the remaining data. The excluded instances were then used to validate the prediction accuracy of the model. This was repeated for all 45 compounds. Training the network with domain adaptation, the error term is composed of the sum of the prediction error and the domain classification error, for each gene set. The network was also trained without adaptation, in which case the error term to be minimised consisted of just the prediction error. The overall performance of the model was assessed using the average validation error over all compound-dose instances. The validation error for each compound-dose combination is the mean absolute error between the model predicted time series of gene expression and the measured time series of rat *in vivo* gene expression for each gene in the gene set.

Latent space classification

The bottleneck structure of the neural network generates a reduced dimensional representation of the time series of both rat and human *in vitro* gene expression data. For example, the GTX+C gene set consists of 76 genes measured at three time points *in vitro*, resulting in a 228 dimensional feature space in the input layer which is compressed to 64 dimensions in the third hidden layer. In recent years, multiple studies have reported on the potential of the reduced dimensional latent spaces generated by methods such as variational autoencoders and non-negative matrix factorisation as novel methods for classification in high dimensional genomic data. With low dimension latent space representation having already being successfully utilised to identify subclasses of tumours from RNASeq data [34], predicting drug responses [35], and de-convoluting cell composition of samples [36]. As a result, it was decided to explore the 64 dimensional embedding of the rat and human *in vitro* gene expression data in common latent space as a potential method for the classification of toxicity of a novel compound under the assumption that compounds that trigger similar responses in gene expression would cluster together in the compressed latent space. Of the 45 compounds included in this study, the carcinogenicity status (yes/no) is known for 25 compounds. A

linear support-vector-machine (SVM), weighted to account for the unequal number of class labels (19 carcinogenic versus just 6 non-carcinogenic), was trained to discriminate between known carcinogenic and non-carcinogenic compounds using the 64 dimensional embedding of the rat and human *in vitro* gene expression for the low, medium, and high dosages for the labelled compounds during leave-out-out cross validation. The accuracy of the classification predictions was assessed using the prediction for the labelled leave-one-out compound. This process was repeated to train a predictor of genotoxicity using the 32 compounds for which genotoxicity labels are available. A list of carcinogenicity and genotoxicity labels used for the compounds can be found in the Supplementary Table S4.1.

Results

Domain adaptation

Figure 5.3 demonstrates the effect of Ganin et al.'s domain adaptation on the training of our network. Each row visualises the embedding of the rat (blue) and human (red) *in vitro* gene expression data along the first two principal components for the first three layers of the network. The network in Figure 5.3 was trained for the leave-one-out compound hexachlorobenzene; domain adaptation produced similar behaviour when the network was trained for the other 44 compounds. The first column depicts the encoding of the rat and human data for the network trained without domain adaptation.

For each layer, the first principal component explains at least 90% of the variance in the data. The rat and human data are clearly separable along the first principle component in each layer. The second column details the encoding of the rat and human data in the same network trained with domain adaptation. While, the rat and human data are still separable along the first principal component in the first layer, in this instance the first principal component explains just 36.9% of the total variance. In the second layer, the border between the rat and human data along the first and second principal components becomes less distinct, with the first component accounting for just 18.3% of the total variance in the data. In the third layer, it is no longer possible to discriminate between the rat and human *in vitro* gene expression instances in the 64 dimensional latent space. The domain classifier predicts the data domain using the encoding of the rat and human data from this third layer. Supplementary Figure S5.1 illustrates the

Figure 5.3 : Visualisation of the effect of domain adaptation on the embedding of rat and human data in the network.

Visualisation of the embedding of the rat (blue) and human (red) *in vitro* gene expression data in the first three hidden layers of the fully trained network. The first column depicts the embedding of the rat and human *in vitro* data for the network trained without domain adaptation. The second column shows the embedding of the rat and human data for the same network architecture trained with domain adaptation. The first row shows the embedding of each dose-compound combination for both the rat and human *in vitro* data in the first layer of the network, consisting of 135 nodes, projected along the first two principal components of the data. The second and third rows depict the embedding of the rat and human data in the

second and third layers of the network respectively. The rat and human data remain disjoint in the first three layers of the network trained without domain adaptation (column one). When the network was trained with domain adaptation the embedding of the rat and human data began to overlap in the second layer, and are no longer differentiable in the third layer.

progress of the domain adaptation during training, showing that at the mid-way point of training the domain adaptation has effectively merged the rat and human data.

Human *in vivo* predictions

The model utilises *in vitro* and *in vivo* gene expression data from rats to predict time series of human *in vivo* gene expression given a measured time series of human *in vitro* gene expression following an exposure. The network's human *in vivo* predictions differ in gene expression pattern from the human *in vitro* input and also from the corresponding measured rat *in vivo* gene expression pattern for a given dose-compound combination, as demonstrated in Figure 5.4. In addition, the model predicted gene expression pattern for a given gene changes for different dosages and exposures, as shown in Figure 5.4 for a sample gene NR0B2, a transcription regulator involved in the regulation of NPAS2-mediated hepatic lipid metabolism that has been associated with hepatic genotoxicity. The model predictions for all 76 genes from the GTX+C gene set following exposure to the previously unseen compounds hexachlorobenzene and omeprazole can be found in the supplementary material. (Figures S5.2 and S5.3). No human hepatic gene expression data following *in vivo* exposure to any of the compounds included in this study could be obtained to validate the model predictions.

Rat *in vivo* predictions

It is possible to assess the quality of the predictions of time series of rat *in vivo* gene expression following exposure to a previously unseen compound. Figure 5.5 depicts the model predictions for a selection of genes from the GTX+C gene set for a medium dosage of the previously unseen compound hexachlorobenzene. As before, the measured time series of rat *in vitro* gene expression is shown in red and both biological replicates of measured rat *in vivo* gene expression is in blue. The solid yellow line indicates the gene expression pattern predicted using the network trained with domain adaptation. The dashed yellow line shows the rat *in vivo* gene expression pattern predicted using the same network architecture trained

without domain adaptation. While, the network trained without domain adaptation predicts the general trend of the rat *in vivo* gene expression over the four time points, the network trained with domain adaptation captures more of the finer details in the gene expression pattern out-performing the model trained without domain adaptation (Figure 5.5).

Figure 5.4 : Input data and model predictions for time series of the gene NR0B2 for multiple doses of hexachlorobenzene and omeprazole

The first row depicts the measured time series of rat *in vitro* gene expression (red) and both measured biological replicates of rat *in vivo* gene expression for the gene NR0B2 following exposure to a low (columns 1 and 2) and medium (column 3) dose of hexachlorobenzene (HCB) and a medium (columns 4 and 5) and high dose (column 6) of omeprazole (OPZ). The model prediction of the time series of rat *in vivo* gene expression of NR0B2 for each exposure are shown in yellow. The second row displays the corresponding measured time series of NR0B2 in primary human hepatocytes exposed *in vitro* and the model predicted time series of NR0B2 for the human *in vivo* system.

Figure 5.5 : Input data and model predictions for a medium dosage of the validation compound hexachlorobenzene for the UDA network trained with domain adaptation for the GTX/C gene set.

The figure shows the time series of rat *in vitro* gene expression (red), the input to the model, and both rat *in vivo* biological replicates (blue) measured following an exposure to a medium dosage of hexachlorobenzene for a selection of gene from the GTX/C gene set. The model predictions of rat *in vivo* gene expression pattern for each gene is shown in yellow. The solid yellow line shows predictions of rat *in vivo* gene expression patterns for the network trained with domain adaptation. The dashed yellow lines indicate the predicted rat *in vivo* gene expression patterns for the same network trained without domain adaptation. The rat *in vivo* prediction for the model trained with domain adaptation outperforms those when the model is trained without domain adaptation. When the biological replicates have a similar gene expression pattern the model performs well (Ccna2, Nr0b2, Rbpms, Ccne1). Given the instance matching used to generate machine learning examples in this study, when the biological replicates have contradictory gene expression patterns the model predictions are inaccurate (Afp, Gstk1).

Comparing the average mean absolute error in predicting rat *in vivo* gene expression for the network model trained with and without domain adaptation indicates that the model trained with domain adaptation produces more accurate predictions of rat *in vivo* gene expression for each of the four toxicologically relevant genes sets (Table 5.1). In fact, the average mean absolute error is significantly lower for the predictions made using

Table 5.1 : Average mean absolute error from leave one out cross validation for the model trained with and without domain adaptation predicting rat *in vivo* gene expression the four toxicologically relevant gene sets identified from literature.

gene set	Number of genes	Network with domain adaptation	Network without domain adaptation	p-value
Cholestasis	18	0.0401 ± 0.0106	0.0409 ± 0.0109	0.0000915*
NAFLD	22	0.0369 ± 0.0071	0.0372 ± 0.007	0.0339*
Steatosis	50	0.0378 ± 0.0055	0.0379 ± 0.0064	0.6195
GTX/C	76	0.0337 ± 0.0087	0.0371 ± 0.0224	0.00003*

The table shows the average validation error ($\pm$ one standard deviation) for the network model trained with (column 3) and without domain adaptation (column 4) for each of the four toxicologically relevant genes sets identified from literature. The validation error is the mean absolute error between the measured time series of rat *in vivo* gene expression and the model predicted rat *in vivo* gene expression pattern using leave-one-out cross validation for the 45 compounds included in this study. The fifth column provides the p-value indicating if the difference in average mean absolute error between the network trained with domain adaptation and without domain adaptation are statistically significant, measured using a two-tailed paired t-test. The model trained with domain adaptation has a lower average mean absolute error than the model trained without domain adaptation for all gene sets included in these analyses, the error has significantly decreased for the Cholestasis, NAFLD, and GTX+C gene sets. Column 2 indicates the number of genes included in each gene set.

the UDA model than using the model without domain adaptation for the Cholestasis, NAFLD, and GTX+C gene sets (p-values < 0.05 , using a two-tailed paired t-test).

Latent space classification.

The training accuracy in predicting carcinogenicity using the latent space embedding of rat and human *in vitro* gene expression for genes in the GTX+C gene set of 85.2% indicates that a complete linear separation between carcinogenic and non-carcinogenic compounds within the 64 dimensional latent space of the network is not possible. The accuracy in predicting carcinogenicity for a previously unseen compound using the network trained with domain adaptation is just 67.3%. Moreover, the specificity and sensitivity of the method (31.9% and 78.5% respectively) indicate a high false positive rate (Table 5.2). This false short of existing methods such as connectivity mapping which has achieved accuracies of 80% in predicting carcinogenicity on the open TG-GATES data set [5].

Predicting genotoxicity labels for a compound using the latent space proves to be an even more challenging task. The training accuracy drops to 78.5% for the embedding trained with domain adaptation and 73.3% for the network trained without domain adaptation. The accuracy in predicting genotoxicity status for a previously unseen compound is just 51%.

Discussion

In recent years toxicogenomics assays have achieved notable success in predicting hepatotoxicity of a novel compound [4-12]. However, relating changes in gene expression profile from these rodent and cell line assays to relevant human *in vivo* outcomes still proves challenging. In **Chapter 4**, we successfully applied deep neural networks to translate

Table 5.2 : Training and validation accuracies in classifying carcinogenicity and genotoxicity of a compound from latent space embedding of rat and human *in vitro* gene expression data

GTX+C gene set	training accuracy	validation accuracy	sensitivity	specificity
Carcinogenicity (n=25)				
With domain adaptation	0.852	0.673	0.785	0.319
Without domain adaptation	0.807	0.590	0.741	0.111
Genotoxicity (n=32)				
With domain adaptation	0.785	0.510	0.299	0.638
Without domain adaptation	0.733	0.559	0.222	0.763

The first section of the above table lists the training and validation accuracies for a linear support vector machine trained to predict carcinogenicity status of a previously unseen compound using the 64 dimensional embedding of the rat and human *in vitro* gene expression from the third hidden layer of the neural network trained for the GTX+C gene set using leave-one-out cross validation. The linear SVM was trained to predict carcinogenicity for the latent space from the latent space embedding networks trained both with domain adaptation (row 1) and without domain adaptation (row 2). The training errors indicate that there is not a complete linear separation between carcinogenic and non-carcinogenic compounds in the 64 dimensional latent space. However, there is a reasonable prediction accuracy for the validation compounds (67.3% for the network trained with domain adaptation). The lower section presents the training and validation errors for the linear SVM trained to predicting genotoxicity from the latent space embeddings of the rat and human *in vitro* gene expression data. Both the training and validation error for the genotoxicity predictions are lower than those for carcinogenicity.

changes in gene expression patterns following exposure to a compound across species and from *in vitro* to *in vivo* in rats.

However, these deep learning networks require large volumes of data to train and relevant human *in vivo* hepatic gene expression data is scarce. Domain adaptation, a subclass of transfer learning, had previously been mooted as a potential solution to this problem [37]. Transfer learning allows knowledge gained solving one task to be reused to solve a similar task. In this study, we apply transfer learning to leverage a large, publicly available database of *in vitro* and *in vivo* gene expression in rats to train a deep learning model to predict human *in vivo* gene expression using the method proposed by Ganin et al. [24]. We demonstrate that this method successfully achieves domain adaptation, with the rat and human data being inseparable in the network latent space. Moreover, integration of the human *in vitro* data significantly improves the accuracy of the predictions of the rat *in vivo* gene expression.

The deep learning network produces physiologically plausible predictions of human *in vivo* gene expression given a measured time series of human *in vitro* gene expression following exposure to a previous unseen compound. However, given the invasive nature of sampling the liver, no relevant measured human *in vivo* gene expression data could be obtained to validate our model predictions. Nevertheless, the model predicted human *in vivo* gene expression patterns that differ from the time series of human *in vitro* gene expression. Demonstrating that the model is not simply reproducing the input it receives. Moreover, the human *in vivo* predictions also differ from the rat *in vivo* predictions for the same gene, indicating that the model is also not simply predicting the same output for both the rat and human. In addition, the predicted human *in vivo* gene expression pattern for a given gene alters for different exposure conditions (dosages and compounds), demonstrating that the model is not predicting the same human *in vivo* gene expression pattern in all instances for a given gene. While we do not have the data to validate the quality of our human *in vivo* predictions, Ganin et al.'s method for domain adaptation has been validated for unsupervised training and was shown to outperform the state-of-the-art algorithms in image recognition and natural language processing. Should relevant human *in vivo* data become available in the future our model predictions could be validated. Moreover, Ganin et al.'s method can be easily generalised to semi-supervised learning, allowing even sparse human *in vivo* data to be integrated into training to improve the human *in vivo* predictions [24].

The use of domain adaptation to integrate the rat and human *in vitro* gene

expression data significantly improves the accuracy of the rat *in vivo* gene expression predictions for three of the four toxicologically relevant gene sets included in this analysis. This is illustrated in Figure 5.5. While the network trained without domain adaptation predicts the general trend in the gene expression pattern (Figure 5.5, dashed yellow lines Ccena2, Nr0b2, Rbpms, and Afp) the improvement in prediction accuracy for the network trained with domain adaptation is evident with the solid yellow lines more closely predicting the measured time series of gene expression. The network trained without domain adaptation was optimised solely to predict the time series of rat *in vivo* gene expression. Therefore, it is somewhat unexpected that training the network to both predict rat *in vivo* gene expression and domain label would result in a significantly improved prediction of rat *in vivo* gene expression. As the training accuracy for the network trained with and without domain adaptation do not differ significantly we do not expect the higher error for the network trained without domain adaptation to be a result of overfitting the training data. We postulate that the integration of the human *in vitro* data introduces additional, relevant information for the rat *in vitro* to rat *in vivo* prediction task.

Recent studies have reported some success in separating subclasses of tumours from the latent space embedding of tumour derived RNA-Seq data [34] or identifying processes involved in rare diseases from a reduced dimension latent space trained on a large publicly available database of gene expression data [38]. While our network was trained to predict time series of rat *in vitro* gene expression and not carcinogenicity or genotoxicity status, it was hypothesised that compounds that induce a particular phenotype of interest, such as carcinogenic or genotoxic compounds, would trigger similar responses in gene expression pattern and would therefore cluster together in the reduced dimension common latent space. Consequently, the embedding of the rat and human *in vitro* gene expression data in the 64 dimensional common latent space was explored as a potential method for the classification of toxicity of a novel compound. Our analysis indicated it was not possible to get a complete linear separation between carcinogenic and non-carcinogenic compound in the common latent space generated by our model, with an accuracy of just 67.3% in predicting the carcinogenicity status for a previously unseen compound. The latent space classification falls short of existing classification methods which may achieve over 80% accuracy in classifying carcinogenicity in combination with the Ames mutagenicity assay [5]. Classifying genotoxicity proved to be even more difficult, with the SVM trained on the latent space representation achieving an accuracy of just 51% in predicting

genotoxicity status for a previously unseen compound. As it currently stands, the reduced dimensional common latent space generated by our models does not appear to be a viable novel method for the classification of compound toxicity. However, our network has been trained using just 720 learning examples of gene expression for just 76 genes (GTX+C gene set). Green and Way trained their variational autoencoder to reconstruct transcriptomic profiles of 5,000 genes using RNASeq data for 10,459 tumour and tumour adjacent normal samples from The Cancer Genome Atlas [34]. In future work it may be possible to improve the classification accuracy of the reduced dimensional latent space by integrating gene expression data from other data sets, thereby increasing the number of learning examples and, consequently, the number of genes that can be included.

The domain adaptation network was trained using Momentum, a stochastic gradient descent algorithm. The gradient reversal parameter (λ) increased from 0 to 1 at a logarithmic rate, as suggested by Ganin et al. [24]. When optimising the networks architecture it was observed that the domain classifier is highly susceptible to mode collapse, whereby the network maximises the loss in predicting the domain of the input data by simply predicting human for every case. This may in part be due to the limited number of learning examples as mode collapse was more prevalent when training the network for the larger gene sets. The frequency of mode collapse occurrence was reduced by modifying the respective error functions for the prediction task and the domain classification to ensure a lesser weight to the domain classifier in the early phase of training. In instances when mode collapse occurred in the domain classifier the network was retrained.

The open TG-GATEs dataset was selected for this study as it is a comparatively large dataset containing gene expression profiles for *in vitro* and *in vivo* rat as well as human *in vitro* following exposure to a range of compounds. Although supplementing the data set with data from other sources would have increased the number of learning, it was decided to restrict our analysis to just the TG-GATEs data set. This was done in order to minimise the variation between the rat and human *in vitro* data due to differing experimental protocols that may have negatively impacted the domain adaptation. Nevertheless, the addition of gene expression profiles from other toxicological databases, such as DrugMatrix [17], may improve the accuracy of the model predictions of gene expression and classification of compounds using the reduced dimensional common latent space. While, in this study, domain adaptation has been applied to transfer responses in gene expression levels across species, UDA could prove to be a valuable tool for the improved inte-

gration of data from multiple sources or data platforms, correcting for batch effects or variation in experimental protocol between the data sets. UDA may also facilitate the training of predictor models for small limited data sets of human *in vivo* data by leveraging larger, publicly available data sets from *in vitro* and animal studies.

To conclude, we have successfully applied domain adaptation, in the context of a deep neural network, merging rat and human gene expression data to facilitate the prediction of human *in vivo* gene expression using a large labelled set of rat data. Moreover, the incorporation of the human *in vitro* gene expression data when training the network significantly improves the accuracy of the predictions of the pattern of rat *in vivo* gene expression following exposure to a previously unseen compound. We also explored the reduced dimensional common latent space generated by our network as a novel method of classification of compound toxicity, as has been done in other studies. However, given the somewhat restrictive number of learning examples and limited number of genes included in this study, we did not get a good separation between the carcinogenic and non-carcinogenic compounds and the method did not outperform existing methods of compound classification.

References

1. Stevens, J. L. Future of toxicology—mechanisms of toxicity and drug safety: where do we go from here? *Chem. Res. Toxicol.* 19, 1393–1401 (2006).
2. Olson, H., Betton, G., Robinson, D., Thomas, K., Monro, A., Kolaja, G. et al. Concordance of the toxicity of pharmaceuticals in humans and in animals. *Regul. Toxicol. Pharmacol.* 32, 56–67 (2000).
3. Hartung, T. Toxicology for the twenty-first century. *Nature* 460, 208–12 (2009).
4. van Delft, J.H., van Aken, E., van Breda, S.G., Herwijnen, M.H., Staal, Y.C., Kleinjans, J.C. Discrimination of genotoxic from non-genotoxic carcinogens by gene expression profiling. *Carcinogenesis* 25, 1265–76 (2004).
5. Caiment, F., Tsamou, M., Jennen, D., Kleinjans, J. Assessing compound carcinogenicity *in vitro* using connectivity mapping. *Carcinogenesis* 35, 201–207 (2014).
6. Ryaboshapinka, M., Hammar, M. Human hepatic gene expression signature of non-alcoholic fatty liver disease progression, a meta-analysis. *Scientific Reports* 7, 12361 (2017).
7. Ellinger-Ziegelbauer, H., Aubrecht, J., Kleinjans, J.C., Ahr, H.J. Application of toxicogenomics to study mechanisms of genotoxicity and carcinogenicity. *Toxicol. Lett.* 186, 36–44 (2009).
8. Hu, T., Gibson, D.P., Carr, G.J., Torontali, S.M., Tiesman, J.P., Chaney, J.G., Aardem, M.J. Identification of a gene expression profile that discriminates indirect-acting genotoxins from direct-acting genotoxins. *Mutat. Res.* 549, 5–27 (2004).
9. Le Fevre, A.C., Boitier, E., Marchandeu, J.P., Sarasin, A., Thybaud, V. Characterization of DNA reactive and non-DNA reactive anticancer drugs by gene expression profiling. *Mutat. Res.* 619, 16–29 (2007).
10. Li, H.H., Aubrecht, J., Fornace, A.J. Jr. Toxicogenomics: overview and potential applications for the study of non-covalent DNA interacting chemicals. *Mutat. Res.* 623, 98–108 (2007).
11. Mathijs, K., Brauers, K.J., Jennen, D.G., Lizarraga, D., Kleinjans, J.C., van Delft, J.H. Gene expression profiling in primary mouse hepatocytes discriminates true from false-positive genotoxic compounds. *Mutagenesis* 25, 561–568 (2010).
12. Tsujimura, K., Asamoto, M., Suzuki, S., Hokaiwado, N., Ogawa, K., Shirai, T. Prediction of carcinogenic potential by a toxicogenomic approach using rat hepatoma cells. *Cancer Sci.* 97, 1002–1010 (2006).
13. Lundberg, A.S., Randell, S.H., Stewart, S.A., Elenbaas, B., Hartwell, K.A., Brooks, M.W. et al. Immortalization and transformation of primary human airway epithelial cells by gene transfer. *Oncogene* 21, 4577–86 (2002).
14. Jover, R., Bort, R., Gomez-Lechon, M.J., Castell, J.V. Re-expression of C/EBP alpha induces CYP2B6, CYP2C9 and CYP2D6 genes in HepG2 cells. *FEBS Lett.* 431, 227–30 (1998).
15. Rodriguez-Antona, C., Donato, M.T., Boobis, A., Edwards, R.J., Watts, P.S., Castell, J.V. et al. Cytochrome P450 expression in human hepatocytes and hepatoma cell lines: molecular mechanisms that determine lower expression in cultured cells. *Xenobiotica* 80, 580–604 (2002).
16. Igarashi, Y., Nakatsu, N., Yamashita, O., Ono, A., Urushidani, T., Yamada, H. Open TG-GATEs: a large-scale toxicogenomics database. *Nucleic Acids Res.* 43, 21–7 (2015).
17. Ganter, B., Tugendreich, S., Pearson, C.I., Ayaboglu, E., Baumhueter, S., Bostian, K.A. et al. Development of a large-scale chemogenomics database to improve drug candidate selection and to understand mechanisms of chemical toxicity and action. *J. Biotechnol.* 119, 219–244 (2005).

18. Pan, S. J., Yang, Q. A survey on transfer learning. *IEEE Trans. Know. Data Eng.* 22, 1345–1359 (2010).
19. Schweikert, G., Widmer, C., Scholkopf, B., Ratsch, G. An Empirical Analysis of Domain Adaptation Algorithms for Genomic Sequence Analysis. *NIPS'08* 1433-1440 (2009).
20. Herndon, N., Caragea, D. Predicting protein localization using domain adaptation approach. *CCIS 452*, Springer-Verlag Berlin Heidelberg, 191-206, (2014).
21. Herndon, N., Caragea, D., Empirical Study of Domain Adaptation with Naïve Bayes on the Task of Splice Site Prediction. In *Proceedings of the International Conference on Bioinformatics Models, Methods and Algorithms - Volume 1: BIOINFORMATICS*, (BIOSTEC 2014), 1,57-67 (2014).
22. Gonen, M., Margolin, A.A., Kernelized Bayesian Transfer Learning. In *AAAI*, 2014.
23. Mourragui, S., Loog, M., Reinders, M.J.T., Wessels, L.F.A. PRECISE: A domain adaptation approach to transfer predictors of drug response from pre-clinical models to tumors. *bioRxiv* 536797; doi: <https://doi.org/10.1101/536797>
24. Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., Lempitsky, V. Domain-adversarial training of neural networks. *J. Mach. Learn. Res.* 17, 1–35. (2016).
25. Jung, D. Elferink, M.G.L., Stellaard, F., Groothuis, G.M.M. Analysis of bile acid-induced regulation of FXR target genes in human liver slices, *Liver Int.* 27, 137–144 (2007).
26. Vinken, M., Landesmann B, Goumenou, M. S., Vinken, I., Shah, H., Jaeschke, C. Willett, M. Whelan et al. Development of an Adverse Outcome Pathway From Drug-Mediated Bile Salt Export Pump Inhibition to Cholestatic Liver Injury, *Toxicol. Sci.*, 136, 97–106, (2013).
27. J. Liu, H. Lu, Y.-F. Lu, X. Lei, J. Y. Cui, E. Ellis, S. C. Strom, C. D. Klaassen, Potency of individual bile acids to regulate bile acid synthesis and transport genes in primary human hepatocyte cultures. *Toxicol. Sci.*, 141, 538–46, (2014).
28. Magkoufopoulou, C., Claessen, S.M.H., Tsamou, M., Jennen, D.G.J., Kleinjans, J.C.S., van Delft, J.H.M. A transcriptomics-based *in vitro* assay for predicting chemical genotoxicity *in vivo*. *Carcinogenesis*, 33, 1421–1429 (2012).
29. Ryaboshapinka, M., Hammar, M. Human hepatic gene expression signature of non-alcoholic fatty liver disease progression, a meta-analysis. *Scientific Reports* 7, 12361(2017).
30. Van Delft, J., Matijs, K., Polman, J., Coonen, M., Szalowska, E., et al. Hepatotoxicity Screening on *in vitro* Models and the Role of 'Omics', *Book Toxicogenomics-based cellular models* (2014).
31. Rumelhart, D.E., Hinton, G.E., Williams, R.J. Learning representations by back-propagation errors. *Nature* 323, 533-536 (1986).
32. Nair, V., Hinton, G.E., Rectified Linear Units Improve Restricted Boltzmann Machines. *ICML'10 Proceedings of the 27th International Conference on International Conference on Machine Learning* 807-814 (2010).
33. Qian, N. On the momentum term in gradient descent learning algorithms. *Neural Networks* 12, 145–151(1999).
34. Way, G.P, Greene, C.S. Extracting biologically relevant latent space from cancer transcriptomics with variational autoencoders. *Pac. Symp. Biocomput.* 23,80-91 (2018).
35. Dincer, A.B., Celik, S., Hiranuma, N., Lee, S. DeepProfile: deep learning of cancer molecular profiles for precision medicine. *bioRxiv*:278739 (2018).
36. Stein-O'Brien, G.L., Clark, B.S., Sherman, T., Zibetti, C., Hu, Q. et al. Decomposing cell identity for transfer learning across cellular measurements, platforms, tissues, and species. *Cell Syst.* 8, 395-411 (2019).
37. Rhrissorakrai, K., Belcastro, V., Bilal, E., Norel, R., Poussin, C. et al. Understanding the

- limits of animal models as predictors of human biology: lessons learned from the IM-PROVER Species Translation Challenge, *Bioinformatics* 31(4), 471–483 (2015).
38. Taroni, J.N., Grayson, P.C., Hu, Q., Eddy, S., Kretzler, M. et al. MultiPLIER: A transfer learning framework for transcriptomics reveals systemic features of rare disease, *Cell Syst.* 8,380-394 (2019).

Supplementary Material - Chapter 5

Supplemental figure S5.1 : Impact of domain adaptation on the network during training

Visualisation of the effect of domain adaptation on the embedding of the source data (rat gene expression (blue)) and the target domain data (human gene expression (red)) during training. Each row shows a two dimensional projection of the embedding of the data along the first two principal components for the first three layers of the network. The first column shows the initial embedding of the rat and human data in the network before training. The rat and human data occupy disjoint regions of the latent space. The second column depicts the embedding of the rat and human data one quarter for the way through training. The difference in the distributions of the rat and human data becomes less distinct. The final column shows the embedding of the rat and human data mid-way through training. At this point the domain adaptation has effectively merged the rat and human data.

Supplemental Figure S5.2 : Measured and model predicted rat and human gene expression for GTX/C gene set for a low dose of hexachlorobenzene

The upper panel depicts the data from the source domain. The measured rat *in vitro* gene expression (model input) for the 76 gene in the GTX/C gene set are shown in red and both rat *in vivo* biological replicates in blue. The model rat *in vivo* predictions are shown in yellow. The lower panel shows corresponding target data. The measured human *in vitro* is shown in red and model human *in vivo* predictions in yellow.

Measured and model predicted rat and human gene expression for GTX/C gene set for a low dose of omeprazole.

The upper panel depicts the data from the source domain. The measured rat *in vitro* gene expression (input to the model) for the 76 gene in the GTX/C gene set for the previously unseen compound hexachlorobenzene are shown in red and both rat *in vivo* biological replicates are in blue. The model predictions of rat *in vitro* gene expression are shown in yellow. The lower panel shows corresponding target data. The measured human *in vitro* is shown in red and model predictions of human *in vitro* gene expression are in yellow.

6

General Discussion

Human physiology acts as a well-orchestrated machinery, adapting from the whole body to the cellular level to a continuously changing environment via a complex interplay between genes, proteins, and metabolites [1, 2]. With advancements in measuring platforms, coupled with a reduction in associated costs, it is now possible to generate extensive time series at multiple levels allowing the functional capacity of the system be observed at several spatiotemporal scales. However, while there has been a surge in the generation of multi-variate longitudinal data, the development of methods for effectively processing, integrating, and interpreting such data has not kept pace. Currently, there exists a patchwork of very often domain specific approaches used to tackle what is essentially a common challenge. In this thesis we explored a variety of computational methods which aim to capture dynamic behaviour of multivariate time series of biological data collected following a perturbation. While, in the previous chapters each approach has been evaluated in the context of their respective applications, in this chapter we will compare the methods, reflecting on their respective attributes in the broader Systems Biology domain.

Systems Biology is an emergent research discipline focusing on the use of computational and mathematical analysis to study the complex interactions between individual components of biological systems and how these interactions give rise to the function and behaviour of the system [3, 4]. Systems biology takes a more holistic approach to studying the biological system, connecting the various network levels from the molecular to the whole body response integrating targeted metabolomics measurements to genome-wide omics measurements. Systems Biology approaches

can be broken into two broad categories as depicted in Figure 6.1. The first category of methods, often termed 'bottom-up', build on existing biological knowledge to facilitate quantitative analysis of the system using computational and predictive models [4, 5]. The advent of high throughput omics technologies has heralded the big data era in biological research, giving rise to the development of more data driven, or 'top-down', Systems Biology approaches. These machine learning methods relying on pattern and inference to identify associations within the data and may provide novel functional insights into the underlying biological network [4, 5].

Figure 6.1 : Overview of Systems Biology approaches

Systems Biology utilises computational analysis to study the interactions between individual biological entities and how these interactions give rise to the functionality of the system as a whole. Systems Biology approaches can be divided into two broad categories; Bottom-up approaches which utilise computational models constructed using existing understanding of the underlying biological network to facilitate analysis of system behaviours. Data driven, or top-down approaches make use of statistical or machine learning methods to identify novel functional insights into the underlying biological network from system-wide omics data.

Bottom-up - Building from a strong foundation of existing knowledge

Energy homeostasis is maintained by the complex interplay between several tissues including the liver, skeletal muscle, adipose tissue, and brain [6]. An energy imbalance in overweight and obesity can induce disturbances in the metabolic crosstalk between these tissues giving rise to the development of diseases such as type 2 diabetes mellitus, non-alcoholic fatty liver disease (NAFLD), and cardiovascular disease [7]. Plasma excursions of metabolites in response to a nutritional challenge are regularly used to assess metabolic health, as impairments in glycaemic control are observable in the postprandial state long before any disturbance is detectable in the fasting state [8–11]. However, objectively quantifying relevant features of metabolic health from the post-load response curves remains challenging.

Simple surrogate measures for targeted questions.

In nutritional and epidemiological research, a plethora of targeted summary indices aiming to quantifying markers of metabolic health, such as insulin resistance or beta cell function from a standardised 75g oral glucose tolerance test (OGTT) have been constructed [12 – 16]. These measures are largely composed of combinations of fasting values, average response values, or area under the curve (AUC) for both the glucose and insulin response curves during the challenge test and have been validated by comparison with gold standard clamp measures of insulin sensitivity or β -cell capacity [14].

In **Chapter 2** we take a closer look at one of the measures, the Muscle Insulin Sensitivity Index (MISI) proposed by Abdul-Ghani et al. (2007) [13]. MISI aims to quantify skeletal muscle specific insulin sensitivity as the ratio of the rate of decay of the post-load plasma glucose excursion from peak to nadir and the mean postprandial plasma insulin concentration. While other OGTT based indices, such as the Matsuda index, have been modified to facilitate their application to differing sampling schedules [17], MISI had not. Exploration of the literature revealed there had been considerable variation in the calculation of MISI [18, 19]. Not only had there been variation in the units used to calculate the score, but also in the number of time points used, with some studies electing to omit certain time points and others using OGTTs

with differing sampling schedules. Despite finding a strong correlation in MISI values calculated on various sampling schedules ($\rho = (0.75, 0.95)$ **Chapter 2** Table 2.1) our analysis showed that even small alterations in MISI value due to differing sampling schedules can have a profound effect on the resulting insulin sensitivity classification (**Chapter 2**, Table 2.1) [20]. With the difference between a five and seven time point OGTT resulting in 15% of the population being reclassified into different insulin sensitivity classes. This discrepancy between sampling schedules may be attributed to miss-identification of the glucose peak, with 27% of the Maastricht Study population ($n=2631$) having a glucose peak at 45 mins, which is not measured in a typical five time point OGTT sampling schedule (**Chapter 2**).

We introduced a numerically more robust method for calculating MISI, utilising cubic splines to predict the glucose peak [20]. Our modified MISI approach appears to be particularly effective for less frequently sampled OGTTs or unequally sampled OGTTs with the correlation coefficient with the gold standard clamp measurement rising from $\rho = 0.48$ for the original method to $\rho = 0.66$ for our modified method when calculated on a four time point OGTT missing the 90 minute measurement.

The use of these surrogate indices of insulin resistance, such as MISI, is widespread throughout metabolic and epidemiological research as the scores provide a less invasive, more cost-effective alternative to time consuming clamp techniques. A review from 2015 lists three surrogate indices of whole-body insulin sensitivity (HOMA-IR [15], Matsuda [12], and QUICKI [16]) among the 100 most cited articles in diabetes research [21]. Searching the Scopus data base for papers with the keywords Diabetes, Type 2 Diabetes Mellitus and Type 2 Diabetes revealed that Matthew et al.'s 1985 paper introducing the Homeostatic Model Assessment index (HOMA-IR), a surrogate index of whole body insulin resistance based on fasting plasma glucose and insulin concentrations, was the most cited paper with these key words, with in excess of 28,000 citations as of December 2019 [15].

While these simple insulin resistance metrics provide an invaluable resource in epidemiological research, they have only been validated for the standardised 75g oral glucose load of an OGTT and are not generalisable for use on a generic meal. Technological advances in wearable devices for glucose monitoring in recent years mean that it is now possible to generate extensive time series of interstitial glucose levels over the course of several days [22]. Continuous glucose monitoring data provides the most physiologically relevant assessment of an individual's glycaemic control,

monitoring glucose excursions in response to their own diet over multiple days [23,24]. However, as the meals have not been standardised applying indices and cut-off values identified for OGTT data is uninformative. There is a need to develop more generalisable methods that not only account for the magnitude of the postprandial glucose excursion, but also the rate and the duration of glucose increase in response to an arbitrary meal.

Mechanistic insight from mathematical models

Mathematical models provide a particularly versatile method for studying the evolution of complex systems over space or time and are regularly employed across diverse range research fields from classical applications in physics and engineering to the social sciences and economics. Within biological research, physiology based mathematical models (PBMM) constructed using *a priori* knowledge allows researchers to investigate how complex regulatory processes are connected and how disruptions to these processes may contribute to disease development.

The use of PBMM to analyse challenge test data is in itself not a novel idea. In fact, the original glucose minimal model proposed by Bergman et al. as far back as 1969 was constructed to quantify insulin sensitivity using plasma glucose and insulin excursions during a frequently sampled intra-venous glucose tolerance test [25]. However, these initial models were quite rudimentary. In recent years, advances in sampling techniques, such as the incorporation of stable isotope tracers, have not only improved our understanding of postprandial glucose and lipid dynamics but also allow for the quantification of specific rates of reactions and turn-over of metabolites that were not previously possible, galvanising the construction and validation of more expansive PBMM [26-28]. In **Chapter 3** we utilise arteriovenous data collected across the abdominal subcutaneous adipose tissue following consumption of a high fat mixed meal [29] to construct a novel mechanistic model of adipose tissue dynamics of glucose, insulin, non-esterified fatty acids (NEFA), and triglyceride in the fasting and fed states [30]. Our model was shown to outperform three published models of postprandial adipose tissue metabolism.

While ODE based PBMM are commonly thought of as a method for simulating dynamic systems *in silico*, advances in numerical solvers and increased computational power mean that it is now feasible to rapidly resolve a model and estimate model parameters directly from data. The widespread use of computational models across several research domains has given rise to a rich array of computational resources and tools, from numerical ODE solvers

implemented in Matlab, Python, and Mathematica to specialist graphical editors such as Copasi, Simulink, and STELLA Modeler. Unlike purely data driven, or so called black-box, models where no model form is assumed, this grey-box approach to systems identification, whereby the model structure is defined accounting for known interactions between metabolic species, means that it is now possible to use a computational model as a scaffold of biological knowledge upon which multivariate time series of data can be analysed. In **Chapter 3** we utilised our adipose tissue model to quantify the effect of a diet intervention on adipose tissue metabolism. Parameter sets were estimated from the postprandial adipose tissue fluxes of glucose, glycerol, NEFA, and triglyceride collected at baseline and following a period of caloric restriction as part of the Yoyo study [30]. Comparison of parameter values indicated there was a significant increase in the transport of both glycerol and NEFA to and from the adipose tissue. The model attributed this to an increase in the rate of lipolysis of triglycerides stored within the adipose tissue, which is to be expected following such a period of caloric restriction, but not evident from the raw data. The model also predicted a large, but not statistically significant decrease in both insulin delay parameters. Decreases in these values mean that the response to plasma insulin is less damped post-intervention; indicative of improved insulin sensitivity. This finding is in line with current literature reports that periods of prolonged caloric restriction can yield improvements in whole-body insulin sensitivity. Further analysis using two surrogate indices of insulin resistance, namely HOMA-IR and ADIPO-IR, confirmed that there was indeed a significant improvement in both whole-body and adipose tissue specific insulin sensitivity following the diet intervention for the sixteen individuals included in this study.

The Yoyo study adipose tissue flux data had previously been analysed using incremental area under the curve (iAUC), and while the method detected a significant decrease in the adipose tissue triglyceride flux, no other significant changes could be identified using iAUC [29]. Here, we have demonstrated that PBMM are more sensitive to subtle changes in the dynamics, such as earlier/higher peaks in response to the meal than standard AUC approaches. In addition, PBMM permit the integrated analysis of multiple metabolites, for example, allowing for the interpretation of the reduced post-intervention triglyceride flux in the context of the plasma insulin to quantify insulin resistance. Moreover, as the structure of PBMM reflect the underlying biological network the estimated parameter values are physiologically interpretable, for example the model predicted decreases in insulin delay parameters can be directly related to a particular physiological feature, in this case insulin sen-

sitivity. This work highlights the mutually beneficial relationship between experimental and computational biology; while the adipose tissue flux data provides an unprecedented opportunity to evaluate, refine, and parameterise the computational model, the model in turn proves to be a powerful tool for the interpretation and quantification of the meal challenge test data.

In this study we applied the model to the average population response curves. However, there is considerable inter-individual heterogeneity in the meal response curves, as evident in the large values for standard error of the mean (Figure 3.4 **Chapter 3**). PBMM are particularly amenable to individualisation, allowing for personalised assessment of metabolic health from meal challenge test data. In future work, fitting of the model to individual data would allow for personalised analysis of the effect of the diet intervention, providing more informative results. Moreover, as there can be considerable inter-individual variation in the responsiveness to lifestyle interventions, such as the diet intervention implemented in the Yoyo study, personalised model simulation would allow for exploration of the observed heterogeneity in the baseline data and identification of features of metabolic health that are indicative of responders and non-responders moving towards the concept of personalised nutrition and precision medicine[31, 32]. In recent years initiatives, such as the Virtual Physiological Human Institute, have garnered increased recognition of the potential of computational models in facilitating the understanding and exploration of the human body as a single complex system [33]. It is envisioned that this Virtual Human could not only allow for *in silico* clinical trials, reducing the need for animal testing, but also through patient specific modelling facilitate personal health forecasting with real time simulations based on data collected by wearable devices such as activity and continuous glucose monitors.

Unlike the simple summary measures discussed in the previous section, as PBMM include phenomenological descriptions of rates of appearance of dietary derived nutrient in the plasma the models are readily generalisable to meals with variable nutrient composition. For example, the Eindhoven Diabetes Education Simulator [34] a physiology based mathematical model of glucose-insulin dynamics has been shown to be applicable for a diverse range of complex meals [35]. This feature of ODE based models has also been exploited in the field of toxicology, Dose-Time-Network-Identification (DTNI) incorporates ODEs to account for the evolution of gene expression over both time and dose when inferring gene regulatory networks for exposure data [36]. By integrating both the temporal and dose dependent information DTNI was shown to provide more in-depth information on

the mode-of-action of individual compounds than existing methods using the temporal information alone. However, PBMM are not without their disadvantages. As the models are constructed based on *a priori* knowledge, their scope may be limited by gaps in our current understanding of the system. In turn, mathematical models may also contribute to plugging these gap as they readily facilitate hypothesis testing. Coppack et al. utilised computational models to determine the most likely explanation for the observed discrepancies in measured adipose triglyceride, NEFA, and glycerol fluxes, predicting the postprandial uptake of glycerol into the interstitial adipose space [28].

Top-down Systems Biology; Learning from big data

The advent of high-throughput omics technologies has revolutionised research in several biological disciplines, however, the processing and interpretation of these high dimensional system-wide measurements, i.e. micro-array measure more than 20,000 human transcripts, has presented researchers with many new challenges. Machine learning approaches have emerged as a promising method for analysing complex high dimensional data. Unlike the previously described bottom-up approaches, machine learning methods are less susceptible to limitations in existing knowledge, primarily relying on patterns and inference to identify relationships in the data.

Gene expression can provide a highly sensitive measure of response of the system to external stimuli. In toxicology, transcriptomic based *in vitro* bio-assays have emerged as a viable alternative to the gold standard two-year rodent bio-assay [37,38]. Multiple studies have shown that *in vitro* toxicogenomics assays were capable of significantly discriminating between subclasses of carcinogenicity [39 – 42]. Moreover, as these assays capture genome-wide alterations in mRNA levels following an exposure, they can also provide insight into the mechanisms that underlie a toxic response [43], giving rise to the concept of Systems Toxicology. However, gene expression patterns in animal and cellular models do not necessarily reflect the human *in vivo* condition [44,45]. In Figure 4.5 (Chapter 4) we show that exposure to a low dose of hexachlorobenzene elicits different response trajectories in known rat-human orthologs for primary rat and primary human hepatocytes exposed *in vitro*. Furthermore, exposing the cells *in vitro* and *in vivo* often also yields differing responses, due to the interactions with other tissues and

cell types, demonstrated in Figure 4.6 (**Chapter 4**). Consequently, directly relating the gene expression patterns from these rodent and *in vitro* assays to the human *in vivo* state using orthologs may yield erroneous results.

Artificial neural networks (ANNs) have been shown to outperform machine learning approaches in classifying high dimensional, highly correlated data such as images [46] and natural language [47]. In **Chapter 4** we apply multiple deep learning architectures to translate gene expression from rat to human primary hepatocytes exposed *in vitro* exploiting temporal patterns in gene expression following an exposure to identify functional relationships in gene expression. As there is no objective measure for the quality of our human *in vitro* predictions we benchmark the predictive accuracy of the deep learning networks against more traditional machine learning methods, namely k-nearest-neighbours and random regression forests. Both the convolutional neural network and a modified autoencoder model consistently outperform the classical machine learning methods. Training the models on randomly selected gene sets of equal size we found considerable variation in the performance of all models, suggesting that the gene expression profile of certain genes may be more informative for the cross-species prediction than others. This is unsurprising, given the complex cross talk and regulatory relationships that exists in gene expression. Moreover, restriction of the analysis to known rat-human orthologs yielded no improvement in the cross-species prediction.

Machine learning methods, particularly deep learning approaches, require the training of many parameters. In order to prevent overfitting and produce a generalisable model larger networks require vast amounts of data on which to train. In **Chapter 4** and **Chapter 5** we used data from Open TG-GATEs [48], a comparatively large toxicogenomics data set containing gene expression profiles for rat and human primary hepatocytes exposed *in vitro* and rats exposed *in vivo* to a vast array of compounds. By leveraging technical replicates and controls it was possible to generate 720 learning examples, however, if we consider the size of the human genome, amounting to some 20,000 protein coding genes, it quickly becomes apparent that there is insufficient data to train a reliable predictor model. In an effort to circumvent the data bottleneck when training neural networks multiple studies have aggregated gene expression data from public repositories such as Array Express and Gene Expression Omnibus, generating sufficiently large data sets to train their models [49 – 52]. In our study, it was decided to restrict our analysis to a single large data set, open TG-GATEs, in order to minimise potential introduction of noise or batch effects due to differing

experimental protocols for data collection. Alternatively, the use of a simpler model structure can reduce the number of parameters needing to be estimated from the data. This approach was taken recently by Eastman and Pande, who make use of a shallow, single layer artificial neural network to process the TG-GATEs data [53]. As we hypothesise a compressed representation of the gene expression data would facilitate cross-species translation, we elected to reduce the number of predicted genes, thereby allowing the application of deeper networks with a bottleneck architecture. We utilised four toxicologically relevant subsets of genes reported in literature as being genomic signatures of cholestasis [54-56], steatosis [57], genotoxicity [58] and carcinogenicity [59] respectively.

Transferable knowledge

In many biological applications, it may be too laborious, costly, or even infeasible to obtain sufficient quantities of data to train a reliable predictor model. For example, human *in vivo* exposure data is scarce due to obvious ethical and practical concerns. These difficulties in collecting data are further compounded when we consider time series data, the focus of this thesis, which necessitates the collection of multiple samples. Transfer learning is a branch of machine learning in which knowledge gained from solving one problem is then utilised while solving another, similar, problem [60]. Multiple studies have applied a range of transfer learning and domain adaptation algorithms to the cross-species prediction task, utilising the abundance of data from well-studied model organisms [61-63]. In **Chapter 5** we integrated this principle of transfer learning with the deep learning architectures developed in **Chapter 4** using a method proposed by Ganin et al., leveraging a large paired data set of rat *in vitro* and *in vivo* gene expression profiles following an exposure to a range of compounds to predict human *in vivo* gene expression patterns. However, given the invasive nature of sampling the liver, no relevant measured human *in vivo* gene expression data could be obtained to validate our model predictions. Interestingly, the integration of human *in vitro* gene expression data resulted in an improved prediction of rat *in vivo* gene expression from the rat *in vitro* gene expression (**Chapter 5** Figure 5.5). While it could be suggested that the simultaneous training for the domain classifier and predictor model helps reduce the risk of over fitting the predictor model as the training errors show little difference between the model trained with and without domain adaptation is it more likely that the integration of the human *in vitro* gene expression data contributes additional information to the predictor.

Neural networks process the data by generating increasingly abstract feature representation of the input data, identifying common patterns across the data [64,65]. Exploring how the gene expression patterns are encoded within the network could produce a novel method for analysing and processing time series data. Several recently published studies have utilised similar bottle neck architectures to find reduced representations of their data, identifying relevant clusters that could be used to separate subclasses of tumours [66] or identify processes involved in rare disorders [51] trained on a large publicly available data bases of gene expression. In **Chapter 5** we explored the embedding of the rat and human *in vitro* gene expression data in the 64 dimensional latent space of the bottleneck architecture to predict genotoxicity and carcinogenicity for a previously unseen compound anticipating that harmful compounds would induce similar gene expression patterns and cluster together in this latent space. While the network had a reasonable performance, with an accuracy of 67.3% in predicting carcinogenicity status for a novel compound, it does not out-perform existing classification methods which rely on static gene expression measurements. Stein-O'Brein et al. used many data sets from the cancer genome atlas to train their model. However, our network has been trained for just 76 genes with 720 learning examples. In future work, inclusion of gene expression profiles from other toxicogenomic databases may yield an improved discrimination between carcinogenic and non-carcinogenic compounds using deep learning models.

Learning from each other

The availability of large publicly accessible online repositories of omics data, particularly gene expression measurements, has greatly accelerated the development of these top-down techniques. A range of pattern recognition methods such as STEM Clustering [67] and dynamic time warping [68] have been developed to relate time series of gene expression following an exposure to relevant disease phenotypes [69]. The application of machine learning methods, including clustering and pattern recognition approaches, to time series of omics measurements has led to the identification of novel associations between genes, proteins, and metabolites, providing insight into the functional capacity of the underlying biological network.

In 2018 Hall et al. applied spectral clustering using dynamic time warping as a method to identify relevant plasma glucose trajectories, termed

glucotypes, for assessing glycaemic control from continuous glucose monitoring data [24]. Spectral Clustering and Dynamic Time Warping have been applied to study longitudinal data in several other research disciplines, including toxicogenomics for many years [68]. The delay in the dissemination of these methods to metabolomics research may in part be attributed to the lack of sufficiently large datasets to effectively train such pattern recognition approaches. MetabolomeXchange [70] aggregates metabolomics data stored across four public metabolomics specific repositories (MetaboLights (EMBL-EBI) [71], the Metabolomics Workbench (NIH) [72], and Metabolomic Repository Bordeaux). As of December 2019 MetabolomeXchange contains just 1838 data sets, of which 130 contain time course data. This is dwarfed by gene expression repositories such as Gene-Expression Omnibus [73] and Array Express [74] which contain gene expression profiles from in excess of 75,000 experiments. The availability of publicly accessible data sets has stimulated the emergence of bioinformatics and systems biology as interdisciplinary fields of research in their own rights, developing a vast array of analysis tools.

Conclusions

Multivariate time-series measurements capture the complex interplay between genes, proteins and metabolites that underlie the system response, not only allowing for the identification of causative relationships between biological entities but also the assessment of the functional capacity of the system as a measure of fitness. However, without the methods to effectively process, analyse, and interpret this often very high dimensional data, much of the functional information can be lost.

Systems Biology is a rapidly emerging, interdisciplinary field of research integrating experimental and computational biology to uncover the complex interactions between individual components of biological systems and how these interactions give rise to the function and behaviour of the system.

Physiology-based mathematical models are regularly applied across a range of diverse biological disciplines to study the dynamic behaviour of biological systems *in silico*. Advances in sampling techniques and technologies in recent years have facilitated the construction of more expansive well validated models. In return, these refined mathematical models provide a scaffold of existing biological knowledge upon which to analyse and interpret this more complex longitudinal data, demonstrating a symbiosis that exists between data generation and model construction. Moreover, unlike many

black box machine learning algorithms the parameter values that are estimated for mechanistic models are physiologically interpretable. In this thesis we have shown that physiology based mathematical models are capable of capturing more subtle alterations in dynamic response trajectories, for example identifying an improvement in insulin sensitivity following a period of caloric restriction, that were not evident using more traditional methods. However, the construction of a mechanistic model can be highly laborious with many existing models describing only targeted sub-systems. As a result, computational models do not yet present a viable option for the analysis of system-wide temporal omics measurements. The expansion and connection of these model sub-systems is ongoing within the wider Systems Biology community.

With the generation of increasingly larger data sets, machine learning, and particularly deeper learning models, present an unprecedented opportunity to not only analyse and process multivariate longitudinal data, but also to identify novel relations in the data, providing new hypotheses for further research. These methods utilise pattern recognition and inference to identify common motifs in the temporal data. In this thesis we have demonstrated the ability of deep neural networks to learn informative patterns in gene expression follow an exposure, translating gene expression trajectories from primary rat to primary human hepatocytes. However, these deep learning approaches require large volumes of data in order to train a reliable and generalisable model.

In the ‘big data’ era one can easily be dismissive of simple summary measures of insulin resistance, such as MISI. However, well validated surrogate indices of health are an invaluable resource in nutritional and epidemiological research, providing a reliable alternative to invasive, time-consuming, and expensive clamp measures.

Despite the challenge of processing and interpreting multivariate time series data being experienced across many diverse research fields, to date the development of solutions has been highly field specific. Here, we have shown that many methods, particularly the data driven machine learning and pattern recognition approaches, are readily transferable between scientific domains. There is a need for greater crosstalk between research disciplines, accelerating the dissemination of computational techniques across various biological disciplines. Systems Biology is highly interdisciplinary field, bringing together people with a range of expertise from experimental biology to mathematics

and computer science, not only facilitating the spread of methods across biological disciplines but also the integration of methods from related research fields such as engineering and physics.

References

1. Guo, N.L. Network medicine: new paradigm in the Omics era. *Anat. Physiol.* 1(1):1000e106 (2011).
2. Barabási, A-L, Oltvai, Z. Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5:101-113 (2004).
3. Breitling, R. What is Systems Biology. *Front. Physiol.* 1:9 (2010).
4. Bruggeman, F.J., Hornberg, J.J., Booger, F.C., Westerhoff, H.V. Introduction to systems biology; Plant Systems Biology. Switzerland, Birkhauser Verlag (2007).
5. Shahzad, K., Loo, J.J. Application of Top-Down and Bottom-up Systems Approaches in Ruminant Physiology and Metabolism. *Curr. Genomics.* 13(5):379-394.
6. Hill, J.O., Wyatt, H.R., Peters, J.C. Energy Balance and Obesity. *Circulation.* 126(1):126-132.
7. Ford, E. S. Risk for all-cause mortality, cardiovascular disease, and diabetes associated with metabolic syndrome: a summary of the evidence. *Diabetes Care* 28, 1769-1778 (2005).
8. World Health Organization: Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia. Report of a WHO/IDF consultation. 2006, Geneva: World Health Organization
9. American Diabetes Association: Standards of Medical Care in Diabetes-2011. *Diabetes Care.* 34 (Suppl 1): S11-S61. (2011).
10. Vis, D.J., Westerhuis, J.A., Jacobs, D.M., van Dynhoven, J.P.M., Wopereis, S. et al. Analyzing metabolomics-based challenge tests. *Metabolomics.* 11(1):50-63 (2015).
11. van Ommen, B., Keijer, J., Kleemann, R., Elliott, R., Drevon, C. A., McArdle, H., et al. The challenges for molecular nutrition research 2: Quantification of the nutritional phenotype. *Genes Nutr.* 3, 51-59 (2008).
12. Matsuda, M. DeFronzo, R. A. Insulin sensitivity indices obtained from oral glucose tolerance testing: comparison with the euglycaemic insulin clamp. *Diabetes Care.* 22, 1462-1470 (1999).
13. Abdul-Ghani, M.A., Matsuda, M., Balas, B., DeFronzo, R.A., Muscle and liver insulin resistance indexes derived from the oral glucose tolerance test. *Diabetes Care.* 30(1):89-94 (2007).
14. Gutch M, Kumar S, Razi SM, Gupta KK, Gupta A. Assessment of insulin sensitivity/resistance. *Indian J Endocrinol Metab.* 19(1):160-164 (2015).
15. Matthews, D.R., Hosker, J.P., Rudenski, A.S., Naylor, B.A., Treacher, D.F., Turner, R.C. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia.* 28(7):412-9 (1985).
16. Katz, A, Nambi, SS, Mather, K., Baron, A.D, Follmann, D.A, et al. Quantitative insulin sensitivity check index: a simple, accurate method for assessing insulin sensitivity in humans. *J Clin Endocrinol Metab.* 85(7):2402-2410 (2000).
17. DeFronzo RA, Matsuda M: Reduced time points to calculate the composite index [letter] *Diabetes Care* 33:e93, (2010).
18. Croymans, D. M. et al. Resistance training improves indices of muscle insulin sensitivity and β -cell function in overweight/obese, sedentary young men. *J. Appl. Physiol.* 115, 1245-1253 (2013).
19. Muniyappa, R., Tella, S.H., Sortur, S., Mszar, R., Grewal, S. et al. Predictive Accuracy of Surrogate Indices for Hepatic and Skeletal Muscle Insulin Sensitivity. *J. Endocr. Soc.* 3(1):108-118 (2018).
20. O'Donovan, S.D., Lenz, M., Goossens, G.H. et al. Improved quantification of muscle insu-

- lin sensitivity using oral glucose tolerance test data: the MISI Calculator. *Sci Rep* 9, 9388 (2019)
21. Shuaib, W., Costa, J.L. Anatomy of success: 100 most cited articles in diabetes research. *Ther. Adv. Endocrinol. Metab.* 6(4):163-167 (2015).
22. Bailey, T.S., Ahmann, A., Brazg, R., Christiansen, M., Garg, S. et al. Accuracy and acceptability of the 6-day Enlite continuous subcutaneous glucose sensor. *Diabetes Technol. Ther.* 16:277-283 (2014).
23. Zeevi, D., Korem, T., Zmora, N., Israeli, D., Rothschild, D. et al. Personalized nutrition by prediction of glycemic responses. *Cell.* 163(5):1079-1094 (2015).
24. Hall, H., Perelman, D., Breschi, A., Limcaoco, P., Kellogg, R. et al. Glucotypes reveal new patterns of glucose dysregulation. *PLOS Biol.* 16(7):e2005143 (2018).
25. Bergman, R.N., Ider, Y.Z., Bowden, C.R., Cobelli, C. Quantitative estimation of insulin sensitivity. *Am. J. Physiol.* 236(6): E677-77 (1979)
26. Dalla Man, C., Rizza, R.A., Cobelli, C., Meal simulation model of the glucose-insulin system. *IEEE Trans. Biomed. Eng.* 54(10):1740-9 (2007).
27. Kim, J., Saidel, G.M., Kalhan, S.C. A computational model of adipose tissue metabolism: evidence for intercellular compartmentation and differential activation of lipases. *J.Theor. Biol.* 251(3):523-40 (2008).
28. Coppack, S.W., Chinkes, D.L., Miles, J.M., Patterson, B.W., Klein, S. A Multicompartmental Model of *in vivo* Adipose Tissues Glycerol Kinetics and Capillary Permeability in Lean and Obese Humans. *Diabetes* .54(7):1934–1941 (2005).
29. Vink, R.G., Roumans, N.J., van der Kolk, B.W., Fazelzadeh, P., Boekschooten, M.V., Mariman, E.C., van Baak, M.A. Adipose Tissue Meal-Derived Fatty Acid Uptake Before and After Diet-Induced Weight Loss in Adults with Overweight and Obesity. *Obesity*.25:1391–1399 (2017).
30. O'Donovan, S.D, Lenz, M., Vink, R.G., Roumans, N.J., de Kok, T.M.C.M., et al. A computational model of postprandial adipose tissue lipid metabolism derived using human arteriovenous stable isotope tracer data. *PLOS Comput. Bio.* 5(10): e1007400. (2019).
31. Ordovas, J.M., Ferguson, L.R., Shyong Tai, E., Mathers, J.C. Personalised nutrition and health. *BMJ* 361:bmj.k2173 (2018).
32. Fröhlich, H., Balling, R., Beerenwinkel, N. et al. From hype to reality: data science enabling personalized medicine. *BMC Med* 16, 150 (2018).
33. Ayache, N., Boissel, J.P, Brunal, S., Clapworthy, G., Fingberg, J., et al. Towards Virtual Physiological Human: multilevel modelling and simulation of the human anatomy and physiology. White pap. Eur. Comm., Barcelona. (2005).
34. Maas, A.H., Rozendaal, Y.J., van Pul, C., Hilbers, P.A., Cottaar, W.J., Haak H.R., van Riel, N.A. A physiology-based model describing heterogeneity in glucose metabolism: the core of the Eindhoven Diabetes Education Simulator (E-DES). *J. Diabetes Sci. Technol.* 9(2):282-92 (2015).
35. Maas, A.H., Rozendaal, Y.J., van Pul, C., Hilbers, P.A., Cottaar, W.J., Haak H.R., van Riel, N.A. A physiology-based model describing heterogeneity in glucose metabolism: the core of the Eindhoven Diabetes Education Simulator (E-DES). *J. Diabetes Sci. Technol.* 9(2):282-92 (2015).
36. Hendrix, D.M., Souza, T., Jennen, D.G.J., Kleinjans, J.C.S. DTNI: a novel toxicogenomics data analysis tool for identifying molecular mechanisms underlying the adverse effects of toxic compounds. *Arch. Toxicol.* 91(6) :2343-2352 (2017).
37. Kleinjans, J. Introduction to toxicogenomics based cellular models in Toxicogenomics-Based Cellular Models; Alternatives to animal testing for safety assessments. Academic Press (2014).
38. Alexander-Dann, B., Pruteanu, L.L., Oerton, E., Sharma, N., Berindan-Neagoe, I. et al. De-

-
- veloments in toxicogenomics: understanding and predicting compound induced toxicity from gene expression data. *Mol. Omics* 14(4):218-263 (2018).
39. van Delft, J.H., van Agen, E., van Breda, S.G., Herwijnen, M.H., Staal, Y.C., Kleinjans, J.C. Discrimination of genotoxic from non-genotoxic carcinogens by gene expression profiling. *Carcinogenesis*, 25: 1265-76 (2004).
 40. Ellinger-Ziegelbauer, H., Aubrecht, J., Kleinjans, J.C., Ahr, H.J., Application of toxicogenomics to study mechanisms of genotoxicity and carcinogenicity. *Toxicol. Lett.*, 186: 36-44 (2009).
 41. Hu, T., Gibson, D.P., Carr, G.J., Torontali, S.M., Tiesman, J.P., Chaney, J.G., Aardem, M.J., Identification of a gene expression profile that discriminates indirect-acting genotox-ins from direct-acting genotoxins. *Mutat. Res.*, 549: 5-27 (2004).
 42. Mathijs, K., Brauers, K.J., Jennen, D.G., Lizarraga, D., Kleinjans, J.C., van Delft, J.H., Gene expression profiling in primary mouse hepatocytes discriminates true from false-positive genotoxic compounds. *Mutagenesis*, 25, 561-568 (2010).
 43. Gautier, L., Taboureau, O., Audouze, K. The effect of network biology on drug toxicology. *Exp. Opin. Drug. Metab. Toxicol.* 9(11):1409-1418 (2013).
 44. Caillarec-Joly, A. et al. TOXsIgN: a cross-species repository for toxicogenomic signatures. *Informatics* 34(12):2116-2122 (2018).
 45. Rhrissorakrai, K., Belcastro, V., Bilal, E., Norel, R., Poussin, C. et al. Understanding the limits of animal models as predictors of human biology: lessons learned from the IM-PROVER Species Translation Challenge, *Bioinformatics* 31(4), 471-483 (2015).
 46. Sutskever, I., Vinyals, O., and Le, Q. V. Sequence to sequence learning with neural networks," in *Advances in Neural Information Processing Systems* 27 (NIPS 2014).
 47. Socher, R. Lin, C.C.Y., Ng, A.Y., Manning, C.D. Parsing natural scenes and natural language with recursive neural networks. In: *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, 129-136.
 48. Igarashi, Y., Nakatsu, N., Yamashita, O., Ono, A., Urushidani, T., Yamada, H., Open TG-GATES: a large-scale toxicogenomics database. *Nucleic Acids Res.* 43: 21-7.
 49. Singh, R., Lanchantin, J., Robins, G., Qi, Y., DeepChrome: deep-learning for predicting gene expression from histone modifications, *Bioinformatics*, 32, i639-i648, (2016).
 50. Aliper, A., Plis, S., Artemov, A., Ulloa, A., Mamoshina, P., Zhavoronkov, A., Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. *Mol Pharm* 13, 2524-2530. (2016).
 51. Taroni, J.N., Grayson, P.C., Hu, Q., Eddy, S., Kretzler, M. et al. MultiPLIER: A transfer learning framework for transcriptomics reveals systemic features of rare disease, *Cell Syst.* 8,380-394 (2019).
 52. Way, G.P., Greene, C.S. Extracting biologically relevant latent space from cancer transcriptomics with variational autoencoders. *Pac. Symp. Biocomput.* 23,80-91 (2018).
 53. Eastman, P., Pande, V.S. Predicting Gene Expression Between Species with Neural Networks. *arXiv:1907.03041* (2019).
 54. Jung, D., Elferink, M.G.L., Stellaard, F., Groothuis, G.M.M. Analysis of bile acid-induced regulation of FXR target genes in human liver slices. *Liver Int.* 2007 27(1):137-144.
 55. Vinken, M., Landesmann, B., Goumenou, M., Vinken, S., Shah, I., Jaeschke, H., et al. Development of an Adverse Outcome Pathway From Drug-Mediated Bile Salt Export Pump Inhibition to Cholestatic Liver Injury. *Toxicol. Sci.* 2013 136(1):97-106.
 56. Liu, J., Lu, H., Lu, YF., Lei, X., Cui, J.Y., Ellis, E., Strom, S.C., Klaassen, C.D. Potency of individual bile acids to regulate bile acid synthesis and transport genes in primary human hepatocyte cultures. *Toxicol. Sci.* 2014 141:538-46.
 57. Ryaboshapinka, M., Hammar, M. Human hepatic gene expression signature of non-alcoholic fatty liver disease progression, a meta-analysis. *Scientific Reports.* 2017

- 7:12361.
58. Magkoufopoulou, C., Claessen, S.M.H., Tsamou, M., Jennen, D.G.J., Kleinjans, J.C.S., van Delft, J.H.M. A transcriptomics-based *in vitro* assay for predicting chemical genotoxicity *in vivo*. . Carcinogenesis. 2012 33:1421–1429.
59. Caiment F, Tsamou M, Jennen D, Kleinjans J. Assessing compound carcinogenicity *in vitro* using connectivity mapping. Carcinogenesis. 2014 35:201–207.
60. Pan, S. J., Yang, Q. A survey on transfer learning. IEEE Trans. Know. Data Eng. 22, 1345–1359 (2010).
61. Schweikert, G., Widmer, C., Scholkopf, B., Ratsch, G. An Empirical Analysis of Domain Adaptation Algorithms for Genomic Sequence Analysis. NIPS'08 1433-1440 (2009).
62. Herndon, N., Caragea, D. Predicting protein localization using domain adaptation approach. CCIS 452, Springer-Verlag Berlin Heidelberg, 191-206, (2014).
63. Herndon, N., Caragea, D., Empirical Study of Domain Adaptation with Naïve Bayes on the Task of Splice Site Prediction. In Proceedings of the International Conference on Bioinformatics Models, Methods and Algorithms - Volume 1: BIOINFORMATICS, (BIOTEC 2014), 1,57-67 (2014).
64. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015 521:436-444.
65. Angermueller C, Parnamaa T, Parts L, Stegle O. Deep learning for computational biology. Molecular systems biology. Mol. Sys. Bio. 2016 12:878.
66. Stein-O'Brien, G.L., Clark, B.S., Sherman, T., Zibetti, C., Hu, Q. et al. Decomposing cell identity for transfer learning across cellular measurements, platforms, tissues, and species. Cell Syst. 8, 395-411 (2019).
67. Ernst, J. Bar-Joseph, Z. STEM: a tool for the analysis of short time series gene expression data. BMC bioinformatics, 7, 191. (2006)
68. Cavill, R., Kleinjans, J., Briede, J.J. DTW4Omics : Comparing patterns in biological timer series. PLOS One 8(8):e71823 (2013).
69. Hendrickx, D.M., Jennen, D.G., Briede, J.J., Cavill, R., de Kok, T.M., Kleinjans, J.C. Pattern recognition methods to relate time profiles of gene expression with phenotypic data: a comparative study. Bioinformatics. 31(13):2115-22 (2015).
70. Haug, K., Salek, R.M., Steinbeck, C. Global open data management in metabolomics. Curr Opin Chem Biol. 36:58–63 (2017)
71. Kale, N.S., Haug, K., Conesa, P., Jayseelan, K., Moreno, P., Rocca-Serra, P. et al. Metabo-Lights: an open-access database re-pository for Metabolomics data. Curr. Protoc. Bioinforma. 14.13.1–14.13.18 (2016).
72. Sud, M., Fahy, E., Cotter, D., Azam, K., Vadivelu, I., Burant, C., et al. Metabolomics workbench: an international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. Nucleic Acids Res. 44:D463–70 (2016).
73. Wang, Z., Lachmann, A., Ma'ayan, A. Mining data and metadata from the gene expression omnibus. Biophysical Reviews. 11(1); 103-110 (2019).
74. Parkinson, H., Kapushesky, M. Shojatalab, M., Abeygunawardena, N., Coulson, R., Farne, A. et al. ArrayExpress – a public database of microarray experiments and gene expression profiles. Nucleic. Acids Res. 35: D747-D750 (2007).

Valorisation

The human body is a highly complex system operating over several spatio-temporal scales. Advances in measurement techniques, particularly the advent of high through-put omics technologies, in recent decades has greatly improved our understanding of the human system, and how the biological entities in each sub-network from cells, to tissues and organs interact to give rise to systemic responses. However, how to effectively process, integrate, and analyse this often very high dimensional data presents one of the biggest challenges to biological researches today. In this thesis, “Modelling the Dynamics in Time Series of Metabolomics and Transcriptomics Data” we have focused on a range of computational approaches that allow multivariate time series of biological data to be integrated, analysed, and interpreted.

While the challenges of analysing time series data are common to several biological disciplines in this thesis we have elected to focus on two specific cases; the minimally invasive assessment of metabolic health in the obese system and alternatives to animal testing in toxicological screening. Nevertheless, the approaches used in this thesis can be generalised to several biological disciplines. Moreover, all chapters of this thesis re-use existing data, generating novel insights from previously published data.

Minimally invasive assessment of metabolic health

Our increasing sedentary lifestyle has given rise to an obesity epidemic, with approximately 2.1 billion adults worldwide being overweight or obese [1]. Overweight and obesity greatly increases the risk for the development of non-communicable diseases including type 2 diabetes mellitus (T2DM), cardiovascular diseases, and certain cancers [2-4]. These conditions not only adversely impact the life of the sufferers and their families, but also have a greater socio-economic impact; with the health care costs as well as loss of productivity associated with T2DM in the Netherlands alone being estimated at 4 billion euro annually [5]. Randomized controlled clinical trials into nutritional and lifestyle interventions are plagued by a considerable heterogeneity in response. Consequently, despite considerable research into lifestyle interventions over several decades, the best practice guidelines for nutritional intervention have not altered significantly.

The success of a nutritional or lifestyle intervention is typically assessed by over-all weight loss, or loss in fat mass. However, increasingly researchers are looking beyond simple weight loss and looking at improvements in metabolic health as the best metric for intervention success. Nevertheless, how we assess metabolic health still poses many challenges, with many

of the gold standard measurements, such as clamp-techniques, being costly and highly invasive. In **Chapter 3** we propose a physiology based computational model of adipose tissue metabolism. This model integrates postprandial response of several metabolites to quantify dynamic features of adipose tissue metabolism that are not directly measurable. Applying our model to data collected before and after a nutritional intervention reveals a large increase in adipose tissue specific insulin sensitivity coupled with significant increases in the mobilisation of free fatty acids and glycerol from the adipose tissue following caloric restriction. These alterations in adipose tissue metabolism were not evident using more traditional area under the curve techniques to assess the meal responses. In this way the computational models produced in this thesis provide researchers with a more sensitive measure of metabolic resilience that can be applied minimally invasive meal challenge test data. This greater understanding of how an individual responds to a particular diet intervention at a mechanistic level can support clinicians in designing more personalised lifestyle interventions with the potential to achieve greater and more sustainable weight loss and improvements in health for the individual, while also helping to alleviate the societal burden of the current obesity epidemic.

Recent studies have shown that certain phenotypical subgroups defined based on insulin resistance or glucose tolerance status may be more responsive to particular diet interventions, experiencing greater and more sustained weight loss [6,7]. However, the current gold standard measure for assessing insulin resistance is the hyperinsulinemic-euglycaemic clamp. Such clamp techniques are highly invasive and time-consuming, consequently it is prohibitive to apply these clamp techniques in clinics or in large populations. In **Chapter 2** we introduce the MISI Calculator, an application that allows researchers and clinicians to quantify skeletal muscle specific insulin resistance using oral glucose tolerance test data. The MISI Calculator is available as a stand-alone application with a user-friendly graphical user interface, allowing clinicians and biological researchers to calculate the index without need for programming experience or potentially expensive licensed software. In this way tissue specific insulin resistance indices can be quantified using minimally invasive clinical measurements. At the time of writing this thesis, multiple intervention studies are making use of our MISI Calculator to facilitate the standardised calculation of MISI. We also envision, that with increased use of our MISI Calculator, applying the standardised MISI score to a variety of study populations, clinically relevant insulin resistant

thresholds can be identified as has been done in that past with the well-established indices of whole body insulin resistance HOMA-IR and Matsuda.

Alternatives to animal testing in toxicological screening

The current gold standard for assessing novel compounds or candidate drugs for potential adverse outcomes in humans is the two-year rodent *in vivo* bio-assay. In recent years, there has been growing ethical concerns surrounding the number of animals being used in toxicological screening giving rise to the adoption of the 3Rs principle of replacement, refinement, and reduction of animal testing. The 2006 EU Regulation on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) legislation states

“The Commission, Member States, industry and other stakeholder should continue to contribute to the promotion of alternative test methods on an international and national level including computer supported methodologies, in vitro methodologies, as appropriate, those based on toxicogenomics, and other relevant methodologies.”

This desire for alternatives to animal testing has given rise to the development of a slew of *in vitro* assays primarily utilising genomics technologies. While these *in vitro* toxicogenomic assays have been shown to have good predictive power in discriminating between subclasses of carcinogenicity [8,9] they can differ greatly in terms of functionality and signalling from the tissues they represent limiting their use for mechanistic interpretation.

In **Chapters 4 and 5** we demonstrate that deep neural networks, particularly those with a bottleneck architectures, are capable not only of translating gene expression patterns from rat to human *in vitro* but also from *in vitro* to *in vivo* in rats. While this was an initial pilot study, using data from just a single toxicogenomics data base, our initial positive results show the potential of bottleneck deep learning architectures for translating *in vitro* gene expression patterns to the human *in vivo* system, in line with the REACH guidelines. Future research, aggregating gene expression data from several toxicology data bases as well as public gene expression repositories can be used to expand the scope of our model. Looking at genome wide changes in gene expression, allowing *in vitro* gene expression from human cell line models be directly translated to the human *in vivo* system could greatly reduce the need

for animal testing. Moreover, as well as being expensive and time consuming the current animal testing protocols have been shown to be a very poor predictor of adverse outcomes in the human system, with a recent study indicating that just 43% of the toxic effects of pharmaceutical compounds in human were correctly predicted by animal testing [10]. The availability of human *in vitro* assays combined with computational models that can reliably predict adverse outcomes in humans would greatly accelerate drug development both in terms of cost and quality. Beyond the field of toxicity assessment, these bottleneck neural networks are increasingly being applied to gene expression data sets in a range of biological disciplines as a powerful tool for dimensionality reduction

Transfer learning methods, as applied in **Chapter 5**, allow the wealth of data available for well studied model organisms or pathologies to be re-used to study other disease states. This not only, reduces the need for to generate large volumes of data for each instance, which may involve laborious and costly experiments, but is particularly beneficial in the case of rare diseases when it is not be possible to collect sufficient volumes of data to train a model. These transfer learning principles are readily being utilised in medical image analysis, where large, well labelled image data sets are being exploited to train models for tumor detection or tissue segmentation in medical images [11].

Open source Softwear

In order for computational models to be useable, it is essential that the source code underlying the model is made publicly accessible along with the published results. This is not only necessary to ensure reproducibility of the results, but also to allow other researchers to readily apply the models on their own data and also build upon and extend published models. In this thesis we evaluated several published models, however the underlying code was rarely supplied with the publication, forcing us to make several assumptions on parameter values and initial conditions when implementing the models ourselves. Each chapter in this these introduces a novel computational model. To ensure others can benefit from these models and approaches the source code for all chapters is publicly accessible from my own GitHub repository (<https://github.com/shauna-odonovan>) as well as in replication packages provided in the supplementary material each publication. The models and software described in this these are already being reused in multiple new

research projects.

Conclusions

In conclusion, I believe that greater cross discipline collaboration, where researchers can benefit from each other's expertise and work is essential to move scientific research forward faster. In particular, I believe computational models can provide a far more sensitive tool for biological researchers to integrate and analyse time series of omics measurements than current cruder approaches, providing predictive tools that can be more rapidly translated to the clinic. In addition, computational models can allow use to re-use available data, integrating it with existing knowledge, to infer information about interactions in the human body that may not be readily accessible for sampling.

References

1. Ng M, Fleming T, Robinson M et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systemic analysis for the Global Burden of Disease Study 2013. *Lancet* (2014); 384(9945):766-81.
2. Kopelman PG. Obesity as a medical problem. *Nature* (2000); 404(6778):635-43.
3. Van Gaal LF, Mertens IL, de Block CE. Mechanisms linking obesity with cardiovascular disease. *Nature* (2006); 444(7121):875-80.
4. Wilborn C, Beckham J, Campbell B et al. Obesity: prevalence, theories, medical consequences, management, and research directions. *J Int Soc Sports Nutr.* (2005); 2:4-31
5. Peters ML, Huisman EL, Schoonen M, Wolffenbuttel BHR. The current total economic burden of diabetes mellitus in the Netherlands. *Neth J Med.* (2017); 75(7):281-88.
6. Astrup A, Hjorth MF. Classification of obesity targeted personalized dietary weight loss management based on carbohydrate tolerance. *Eur J Nutr.* (2018) 72:1300-1304.
7. Blanco-Rojo R, Alcalá-Díaz JF, Wopereis S et al. The insulin resistance phenotype (muscle or liver) interacts with the type of diet to determine changes in disposition index after 2 years of intervention: the CORDIOPREV-DIAB randomized clinical trial. *Diabetologia.* (2016) 59:67-76.
8. van Delft JH, van Aken E, van Breda SG et al. Discrimination of genotoxic from non-genotoxic carcinogens by gene expression profiling. *Carcinogenesis*, (2004): 25: 1265-76.
9. van Delft JH, van Aken E, van Breda SGC. Comparison of supervised clustering methods to discriminate genotoxic from non-genotoxic carcinogens by gene expression profiling. *Mutat Res.* (2005); 575: 17-33.
10. Hartung T. Toxicology for the twenty-first century. *Nature.* (2009) 460: 208-12.
11. Cheplygina V, de Bruijne M, Pluim JPW. Not so-supervised: A survey of semi-supervised, multi-instance and transfer learning in medical image analysis. *Med Image Anal.* (2019) 54:280-296.

Summary

The human body is a complex machinery consisting of multiple tissues and organs interacting over several spatial and temporal domains in response to a continuously changing environment. Advances in high-throughput omics technologies in recent years now allow the identification and quantification of biological entities from each layer of the system. This reaches from from the underlying genetic sequence common to each cell, through gene expression and protein levels, to the resulting metabolite concentrations and how these biological entities may change as a result of an exposure or in certain disease states. However, knowing each component of a car in detail does not necessarily mean we understand how these individual components interact in order to move from A to B. Time-series of omics measurements in response to a stimulus not only facilitate the identification the full collection of genes, proteins, and metabolites that are changed under the new condition but also to untangle the complex interactions between individual components in the system. In this way we can go beyond mere associations and derive mechanistic insight into the behaviour of the underlying system, establishing causative relations. While there has been a surge in the generation of multi-variate time series of omics data in recent years, the developments of methods to effectively process, integrate, and interpret such data has experienced a more ad hoc evolution giving rise to a plethora of often domain specific approaches. In this thesis we explore a variety of computational methods that aim to capture the dynamic behaviour of time series of omics data following a perturbation. While the challenges of analysing time series data is common to many biological disciplines in this thesis we focus on two specific applications; the use of meal challenge tests to evaluate metabolic resilience and alternatives to animal testing in toxicological screening.

Building on a foundation of existing knowledge

Studying the behaviour of a system in response to an external stimulus can provide insight into the fitness of the underlying system, often referred to as resilience. This principle is utilised in nutritional and metabolic research, where meal challenge tests are regularly employed as a measure of metabolic resilience. Postprandial glucose and triglyceride levels have been shown to be more sensitive predictors of future risk of Type 2 Diabetes Mellitus (T2DM) and Cardiovascular Disease than their fasting values. The postprandial glucose trajectory is determined by the complex interplay between several tissues including the liver, skeletal muscle and adipose tissue. While several well-established cut-off values have been defined for

classifying glycaemic control from an oral glucose tolerance test (OGTT), more recent efforts have focused on deriving features of the metabolic health of underlying tissues, such as insulin resistance, from the shape of the postprandial glucose trajectory. In **Chapter 2** we take a closer look at one such metric, the muscle insulin sensitivity index (MISI), that aims to quantify skeletal muscle specific insulin resistance based on the rate of decay of the glucose response curve to a standard 75g OGTT. However, as this index relies on the identification of the glucose peak it is particularly sensitive to the discrete and often sparse sampling of an OGTT, typically consisting of five or fewer time points. We propose a numerically more robust method for calculating MISI that utilises cubic spline interpolation to reduce the index's sensitivity to the discrete sampling schedule. Our modified method produces a stronger correlation with the hyperinsulinemic-euglycaemic clamp, the gold standard for measuring peripheral insulin sensitivity, than the original formulation. Moreover, we provide a user-friendly MISI Calculator application, that allows for the standardised calculation of the MISI score, facilitation cross Study comparison of the index.

While such simple indices provide an invaluable alternative to invasive and time-consuming clamp techniques, currently only indices for quantify features of glycaemic control are available. With growing interest surrounding the role of lipids and amino acids in the etiology of T2DM giving rise to the use of more complex high fat and high protein meal challenge tests there is a need for more generalisable approaches. In **Chapter 3** we construct a novel physiology based mathematical model (PBMM) of postprandial adipose tissue metabolism. PBMMs are constructed based on the known interactions underlying the biological system. Such models are readily extendable to describe larger systems and generalisable to generic meal inputs. Application of our adipose tissue model to meal challenge test data collected before and after and nutritional intervention indicates there is an improvement in adipose tissue insulin resistance following a period of caloric restriction, as well as significant increases in the rates of free fatty acid and glycerol release by the adipose tissue. The computational model captures the more subtle changes in postprandial metabolite trajectories which were not evident using more traditional area under the curve approaches. These modelling approaches utilise our understanding of physiology underlying the postprandial response to facilitate the analysis and quantification of features of health from dynamic time series of plasma metabolites. The knowledge driven approaches are commonly referred to as bottom-up approaches.

Guided by the data

While the dynamic response of the system to a stimulus can give an indication of the health of the underlying system, in turn the trajectory of the response can also be used to infer information of how potentially harmful the exposure is. This principle is exploited in the field of toxicogenomics, where the gene expression pattern following exposure to a compound provides a particularly sensitive measure of the potential toxicity of that compound. While these genomics based animal and *in vitro* models have been shown to perform well in predicting the toxicity of compound, relating the gene expression patterns to the human *in vivo* system still proves challenging, limiting the mechanistic insight into modes of toxicological action that can be gained. In addition, the high dimensionality of gene expression data makes the construction of the knowledge based dynamic models used above prohibitive. Machine learning encompasses a range of methods that rely on inference and patterns to identify relationships in data. In **Chapter 4** we demonstrate that deep learning models, a subclass of machine learning, are capable of translating the post-exposure gene expression patterns from primary rat to primary human hepatocytes. The deep learning models consist of layers of artificial neural networks that transform the data into increasingly abstract features and as such are particularly suited to relating patterns in time series gene expression data. The deep learning models also significantly outperform more traditional machine learning techniques in translating gene expression patterns from the *in vitro* to *in vivo* situation in rats.

While deep neural networks can be a very powerful tool in data analysis, particularly in dimensionality reduction, they require very large volumes of data upon which to train. While large data bases of animal and cell line data are available, due to the challenges with sampling humans, especially tissues such as the liver, human *in vivo* data is comparatively scarce. In **Chapter 5** we build on the deep learning models applied in **Chapter 4** by integrating transfer learning within a neural network architecture in a method known as domain adaptation. In this way, a large set of paired rat *in vitro* and *in vivo* post-exposure gene expression patterns is leveraged to train a model to predict human *in vivo* gene expression given measured human *in vitro* gene expression data. Interestingly, we see that the addition of the human *in vitro* data also improves the model prediction of rat *in vivo* gene expression. These deep learning and machine learning methods rely on the patterns and

information in the data itself to find novel relations and do not require any *a priori* information on the underlying system. These data driven machine learning approaches are also known as top-down approaches.

Conclusion

Multivariate-time series measurements capture the complex interactions between genes, proteins, and metabolites that underly the systemic response to a stimulus. They not only allow for the identification of causative relationships between the biological entities but can also provide an assessment of the functional capacity of the system. However, currently there is a lack of methods to effectively process, integrate, and interpret this often high dimensional data without much of the functional information being lost. Systems Biology is an emergent research discipline focusing on the use of computational and mathematical analysis to study the complex interactions between individual components of biological systems and how these interactions give rise to the function and behaviour of the overall system. In this thesis, I explore a variety of both knowledge driven 'bottom-up' and data driven 'top-down' Systems Biology approaches to analyse and interpret multivariate time-series of omics measurements in two distinct research fields; the quantification of features of metabolic health from meal challenge test data and the assessment of toxicity of a compound using gene expression data in animal and cellular models.

Acknowledgements

First of all, I would like to thank my supervisors for their guidance over these last five years. Theo, Ralf, Natal, and Ilja your diverse perspectives gave me a true experience of interdisciplinary research, motivating me to develop knowledge and expertise beyond my computational background and growing into the researcher I am today. More than anything I would like to thank you for giving me the freedom and opportunity to explore a diverse range of research areas during this PhD project, allowing me to find a topic about which I am truly passionate. Perhaps one day I will finally find the time to do that time management course I agreed to do in first year.

Secondly to Michael, I am truly grateful for your guidance and support throughout this project. Without your help, especially in navigating what was a very new field for me in those first few months, I don't think this thesis would ever have been completed. You were always there with words of support, willing to help locate coding errors or proof read my typo filled documents, even on the multiple occasions when I forgot my own name,

Best,

Michael?

I wish you and your family every joy and happiness in the future.

I would like to thank all the people at MaCSBio for being so welcoming and creating such a pleasant work environment.

Bob, I would like to thank you for keeping me sane during this PhD and introducing me to the many cultural oddities of Maastricht. Over the last five years you have taught me many valuable things; that ants do not survive for very long in a freezer, the optimal recipe for long island ice tea (just a dash of cola for colour), and when in doubt just take the absolute

Acknowledgements

value. I want to thank you for the many great memories, and also the very many nights where there are no memories.

Isma you've been there for me from day one. I miss our daily coffee roulette, when it was good it was very very good, when it was bad it was probably the sister. You were always a great positive influence even when it seems like we would never those papers from that writing course published. Unfortunately the code fairies don't seem to have followed me to Wageningen.

Cristobal, I cannot thank you enough for taking pity on me when you saw my homemade dictionary of biological terms and patiently explaining glucose homeostasis to me, and even now lipoprotein dynamics. I have never laughed as hard as I did when sharing an office with you.

To the newer arrivals Bart, David, and Balazs, thank you for breathing new life into the office when things were becoming more tense. I would like to pay special thanks to Balazs for braving my resting bitch face and sitting next to me and listening to my random modelling rants and also taking on the mantle of the office cloud of doom.

I would like to thank my many co-authors, especially those that very generously shared their data with me. I would like to particularly thank Gijs and Roel for very patiently explaining in intricacies of clamps and challenge test data and Rachel and Kurt for taking the time to give me a crash course in field of deep learning, you were all very supportive.

I would like to thank my current employers Lydia, Natal, and Ilja, for being so patient and understanding over the last year as I finished my thesis and also for giving me the opportunity to further pursue my research interests.

A very big thank the various members of the Periodic Table Dancers down through the years especially the regulars Bob, Menica, Koen, Balazs, and Petra. We will never forget the Italian word for blood feud. Although we may need to actually brush up on the Periodic Table. I'm sure one day Peter's will add a small fries to the menu.

To my Irish PhD support group, Danny, Shane, Bryan, Liam, and Dan while we were often more like ships passing in the night it was always reassuring to know the struggle was real regardless of the country or topic. We all know

better than to mention the war. Is post istigh e.

I would like to explicitly thank John Daniel Riordan for his help with one line of Mathematica code that I subsequently did not use. I hope that you will accept this acknowledgement in lieu of the co-authorship you had demanded.

Dave I owe you a innumerable amount of pints for your stats advice, with a very vague descriptions of the data you always recommneded the correct test

Finally, I would like to thank my family, especially my parents, their encouragement and support throughout my studies. Even when things may not have gone the way I wanted you always reassured me what was ment for me would not pass me by. I'm also sorry for bringing a dampener on family gathering by reminding you all about the consequences of high blood sugars. The true test of this PhD will be whether I have learnt enough to stave off the family inheritance.

Publications

O'Donovan SD, Driessens K, Lopatta D, Wimmenauer F, Lukas A, Neveen J, Stumm T, Smirnov E, Lenz M, Ertaylan G, Jennen DGD, van Riel NAW, Cavill R, Peeters RLM, de Kok TMCM. "Use of deep learning methods to translate drug-induced gene expression changes from rat to human primary hepatocytes exposed in vitro and in vivo." PLOS ONE (2020).

O'Donovan SD, Lenz M, Vink RG, Roumans NJT, de Kok MCM, Mariman ECM, Peeters RLM, van Riel NAW, van Baak MA, Arts ICW. "A computational model of postprandial adipose tissue lipid metabolism derived using human arteriovenous stable isotope tracer data." PLOS Comp. Bio. (2019) 15(10):e1007400.

O'Donovan SD, Lenz M, Goossens GH, van der Kallen CJH, Eussen SJMP, Stehouwer CDA, van Greevenbroek MM, Schram MT, Sep SJ, Peeters RLM, Blaak EE, van Riel NAW, de Kok TMCM, Arts ICW. "Improved quantification of muscle insulin sensitivity using oral glucose tolerance test data: the MISI Calculator." Sci. Rep. (2019) 9(1):9388.

Erdos B, van Sloun B, Adriaens ME, **O'Donovan SD**, Langin D, Astrup A, Blaak EE, Arts ICW, van Riel NAW. Individualized computational model quantifies heterogeneity in postprandial responses to OGTT. (submitted)

About the author

Shauna O'Donovan was born in Cork, Ireland on the 13th of October 1990. In 2009 she completed her secondary school education in Christ King Girls Secondary School in Cork and went on to pursue a Bachelor's degree in Mathematical Sciences with a focus on mathematics and applied mathematics in University College Cork, Ireland, graduating with honors in October 2013.

In 2013, she undertook a Master's degree in Bioinformatics and Computational Biology in the Microbiology department of University College Cork. She conducted her Master's thesis research in The Laboratory of Post-Transcriptional control and bioinformatics in UCC on a project entitled "Why protein synthesis predominately initiates at AUG codons and terminates at UAR and UGA." She graduated with first class honors in February 2015.

In April 2015, Shauna moved to the Netherlands and joined the Maastricht Centre for Systems Biology as a PhD Candidate. Here, she worked on the application of a variety of computational modelling approaches to study dynamic time series of biological data with a special interest in the liver and the obese system.

In July 2019, Shauna moved to Wageningen to continue her research in the Division of Human Nutrition and Health in Wageningen University. Here, her work focuses on the application of personalisable computational models of metabolism to explore heterogeneity in the overweight and obese population and its relation to variation in diet response.