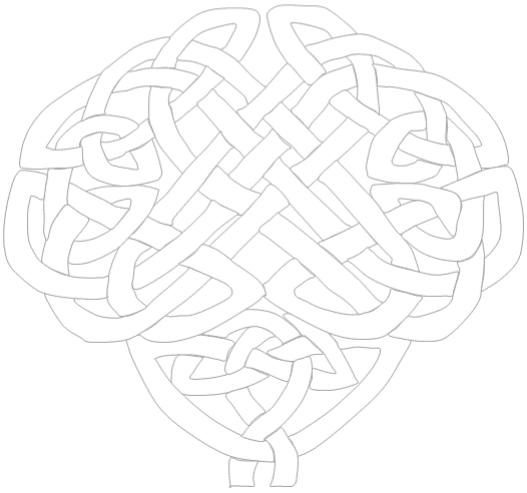
Modelling the dynamics in time-series of metabolomic and transcriptomic data

Valorisation



The human body is a highly complex system operating over several spatio-temporal scales. Advances is 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 poses 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 a 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 in 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 a 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 used of our MISI Calculator, applying the standardise 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 the 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 patters 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 burder 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, Alcala-Diaz 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 Agen 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 Agen 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.