Genome-scale modelling of human adipocyte metabolism

Summary

Over the last few decades, there have been tremendous advances in technologies that can measure all the biomolecules in a cell. The suffix, 'omics', is added to describe these technologies depending on which type of molecule they examine, for example, DNA (genomics), RNA (transcriptomics) and proteins (proteomics). The ability to rapidly quantify a myriad of cellular molecules in a cost effective manner has revolutionised biomedical research. It has opened opportunities for investigating the changes occuring in our body during a disease (such as cancer, COVID) at cellular and molecular level. The molecular components of the altered processes serve as targets for developing drugs or novel therapies.

One of the commonly used approaches to interpreting such large-scale omics data is through simulating biological processes. Such simulations are achieved using mathematical models that combine information from multiple sources. Models have been used to simulate a wide range of biological phenomena such as spread of a disease, development of an embryo and propagation of electric signals in neurons. The studies conducted in this thesis used genome-scale models of cellular metabolism (metabolic models). Essentially, metabolic models are maps representing thousands of chemical reactions taking place within a cell. These models consist of numerous compounds or metabolites and enzymes participating in the reactions and are often very large. Combining metabolic models with omics data obtained from a disease of interest, provides an opportunity to simulate and investigate disease-specific cellular malfunctions. However, such an integration is not straightforward. In this thesis, two novel approaches were developed to incorporate omics data into human models of metabolism. These methods were then applied onto a model of adipocytes to study the metabolic changes occurring in obesity and during weight loss.

Visualisation is a crucial part of analytical pipelines as it provides a visual context via graphs and charts. It also facilitates identifying trends in large datasets. In the context of metabolic models, our understanding of the disease under study can be improved by methods that allow combining them with omics data through visual exploration. **Chapter 2** presents a computational method developed called EFMviz for visualising networks of cellular metabolism. EFMviz allows users to select reactions of interest. It then extracts components from the model that are connected to these reactions and visualises them as networks. These networks show the connections between genes, reactions and metabolites in the cell. In addition, EFMviz allows for viewing different types of omics data (from transcriptomics, proteomics and metabolomics) on the networks and can display the changes in the amount of molecules between normal and disease conditions. **Chapter 2** demonstrated the visualisation of subsets of reactions from microbial and human models. This workflow has been adapted in other chapters as well - **Chapters 3 and 4**, where it was instrumental in interpreting the analysis of data from the DiOGenes Study. To improve its usability, EFMviz has been integrated into COBRA Toolbox, a software suite that is widely used by researchers working with metabolic models.

Another ongoing challenge in the field of metabolic modelling is comparing different phenotypes in large models for identifying disease-specific metabolic signatures. **Chapter 4** addressed this challenge by developing an algorithm called ComMet (Comparing Metabolic states). Given two metabolic models of a cell, for instance, representing its healthy and diseased states, ComMet can extract the pathways having significant modifications in their reaction rate. **Chapter 4** illustrated application of ComMet in a model-based manner on two simulated

metabolic states of an adipocyte. In terms of usability, ComMet is applicable in both model-based and data-driven manner. In **Chapter 4**, we successfully predicted the metabolic pathways affected by the simulated perturbation. On the other hand, **Chapter 5** illustrated a data-driven application of ComMet. Transcriptomics data was used in place of simulated conditions to represent two states of adipocytes - before and after weight loss. ComMet successfully identified adipocyte-specific modules affected by the weight loss intervention.

Chapter 3 demonstrated the value of integrative analytical pipelines to study adipocyte metabolism in obesity. The integrative workflow developed in **Chapter 3** is based on metabolic network analysis and it successfully combined plasma metabolomics and transcriptomics data (from adipose tissue) related to human obesity. Using this workflow, we examined the relationship between dysregulated adipocyte metabolism on BCAA levels in circulation. Through network visualisation we clearly illustrated how different molecules interact and contribute towards the biological phenomenon under study. This enabled identifying several adipocyte-specific genes, metabolites and processes that are significantly associated with plasma BCAAs. In addition, sex disparities were uncovered in these molecular aspects of adipocyte metabolism.

Taken together, the findings of the studies presented in this thesis provide testable hypotheses for future research and several targets for validating the predictions. Thus, future studies can build upon our findings to improve our overall understanding of adipocyte metabolism in obesity.

In summary, mathematical models are instrumental to studying biological processes such as metabolism. This thesis showcased novel methods and workflows to overcome the challenges around integrating large-scale data with human metabolic models. Our approaches enabled integrating multiple types of omics data to investigate adipocyte metabolism in obesity. Moreover, we demonstrated several ways to visualise molecular data in the form of networks, which improved interpretability of our analyses. The methods developed here are building blocks for applications in human health as they can be used for studying several other human diseases. The data integration strategies presented here have great potential in using omics data from individuals to generate personalised models of diseased functions and thus, contribute to developing personalised medical treatments.