Human Cardiac Age-OME

A multi-facet approach to understanding healthy ageing of the heart

by

Giovanni Guglielmi

ORCID: https://orcid.org/0000-0002-0509-9666

A thesis submitted in total fulfilment for the
degree of Doctor of Philosophy

In the Department of Biomedical Engineering
University of Melbourne

In the School of Mathematics
University of Birmingham

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

The world's population is ageing, expecting individuals over 65 to represent 38% of the total population by 2050[705]. As lifestyle and medical practices have improved and one's life expectancy is longer, comorbidities associated with ageing are also expected to raise healthcare costs in the near future. This can limit access to appropriate medical treatments and introduce further disparities in healthcare accessibility due to one's socio-economic background[556]. Older age is a strong risk factor for cardiac morbidity and cardiovascular diseases, including heart failure, one of the leading causes of death. Furthermore, those who survive cardiomyopathies can often be left with disabilities, which would exert financial and socioeconomic impacts on themselves, family members and the community. The increase in cardiovascular diseases partially attributed to ageing has set a novel and challenging goal for the scientific community: understanding how cardiac tissues age. Although aged hearts might not develop cardiomyopathies, their functionality, shape, histology and molecular composition decrease over time. Frailty of the cardiac muscle is developed around the age of 50, where mortality for cardiovascular diseases arises for both genders. On the other hand, mortality due to non-congenital cardiovascular diseases is almost negligible for young individuals, usually aged between 12 and 25. Utilising a unique set of pre-mortem, cryopreserved, non-diseased human hearts, we analysed the molecular landscape (transcripts, proteins, metabolites and lipids) in young (≤ 25 years old) and old (≥ 50 years old) donors to characterise human cardiac ageing. Analysing all the molecules using multiple biostatistical and mathematical methods, including a novel framework for analysing biological networks and a newly engineered method for detecting dysregulated biological processes, we observed a downregulation of proteins involved in calcium signalling, such as SERCA, decreased abundance in key proteins of the contractile apparatus, like myosin heavy chain, in older hearts. In addition, we found a potential counteractive upregulation of central carbon metabolism, an increase of long-chain fatty acids, and possible dysregulation of the β-oxidation in older individuals. This is the first study on the molecular data set of normal human cardiac ageing, which, coupled with the innovative omics methods, enhances our understanding of the development of age-related heart diseases.

Declaration for The University of Melbourne

This is to certify that:

- The thesis comprises only my original work towards the Doctor of Philosophy.
- Due acknowledgement has been made in the text to all other material used.
- The thesis is fewer than 100,000 words in length exclusive of tables, bibliographies and appendices.

Declaration for the University of Birmingham

This is to certify that:

- The thesis comprises only my original work towards the Doctor of Philosophy.
- Due acknowledgement has been made in the text to all other material used.
- The thesis is fewer than 50,000 words in length exclusive of tables, bibliographies and appendices.

Acknowledgements

I would like to thank my supervisors, A/Prof Vijay Rajagopal and Prof Fabian Spill, for mentoring me throughout my programme. I am also thankful to the collaborators at the University of Sydney, who have patiently supported me through my academic growth. I am also grateful for the Priestly scholarship that allowed me to pursue my doctoral study. This thesis is dedicated first to my mother, Francesca, who has supported and cheered for me throughout my studies without hesitation. I would like to thank my younger sister, Nadia, who has patiently listened to me countless times during my struggles. I would also thank my dog, Jason, who passed away a few months before my PhD submission, for accompanying me throughout most of my studies, although I could not see him one last time. Finally, I would also like to dedicate this work to my partner, Nobuko, who has remained by my side on rainy and sunny days.

Contents

1 Introduction **10**

2 Heart, Metabolism, and Ageing **13**
2.1 Circulatory system, heart anatomy and its functions . . . 14
2.1.1 Overview of the circulatory system . . . 14
2.1.2 The function of the heart in the circulatory system . . . 14
2.1.3 Heart anatomy and its functions . . . 15
2.2 Cardiac cell types and extracellular matrix . . . 17
2.2.1 Cardiac cell types and cell counts . . . 17
2.2.2 Cardiomyocytes . . . 18
2.2.3 Cardiomyocyte's ultrastructure and excitation-contraction mechanism . 19
2.3 Lipids, glucose and cardiometabolism . . . 23
2.3.1 An overview of lipids . . . 23
2.3.2 Mitochondrial membranes and related lipids . . . 25
2.3.3 Lipid synthesis and trafficking in mitochondria . . . 26
2.3.4 Lipid droplets . . . 30
2.3.5 β-oxidation, GL/FFA cycle, and lipotoxicity . . . 32
2.3.6 Glucose metabolism . . . 36
2.3.7 Pyruvate trafficking and Acetyl-CoA synthesis . . . 38
2.3.8 Krebs cycle . . . 38
2.3.9 Oxidative phosphorylation . . . 39
2.3.10 Phosphocreatine energy system . . . 40
2.4 Signalling and metabolic changes in cardiac ageing . . . 40
2.4.1 Signalling pathways . . . 41
2.4.2 Metabolism . . . 44

3 Aims, Scopes and Contribution **46**
3.1 Molecular signature of ageing . . . 46
3.2 Omic as a biological network . . . 48
3.3 Omics SEA . . . 49
3.4 Other contributions . . . 50
3.5 Awards and funding . . . 51

4 Research Methods **52**
4.1 High-throughput proteomics . . . 53
4.2 High-throughput metabolomics . . . 54
4.3 High-throughput lipidomics . . . 56
4.4 High-throughput transcriptomics . . . 58
4.5 Transforming omics measurements . . . 60
4.6 Normalising omics measurements . . . 62
4.7 Handling missing measurements . . . 64
4.8 Identification of candidate biomarkers . . . 66
4.9 Co-expression network . . . 70
4.9.1 Construction of co-expression network . . . 73
4.10 Omics enrichment analysis . . . 74
4.10.1 Over-representation analyis . . . 76
4.10.2 Omics SEA . . . 77

5 Experimental Design and Data Exploration **80**
5.1 Five principles for good omics research . . . 81
5.2 Samples definition, selection and sample size calculation . . . 82
5.3 Proteomics protocol . . . 85
5.4 Metabolomics protocol . . . 86
5.5 Lipidomics protocol . . . 87
5.6 Transcriptomics protocol . . . 88
5.7 Computational protocol . . . 88
5.8 Explorative data analysis . . . 91

6 Molecular Signature of Cardiac Ageing **96**
6.1 Proteins . . . 97
6.2 Metabolites . . . 101
6.3 Lipids . . . 107
6.4 Transcripts . . . 114

7 Omic-inferred Biological Network **118**
7.1 Protein-Protein co-expression network . . . 119
7.2 Metabolite-Metabolite co-expression network . . . 122
7.3 Lipid-Lipid co-expression network . . . 124
7.4 Transcript-Transcript co-expression network . . . 126

8 Omics SEA: A Unified View of Biological Pathways **131**

9 Discussion of the Findings **138**
9.1 Discussion . . . 139
9.2 Summary . . . 144

10 Conclusion, Limitations and Future Directions **145**

11 Appendix - Background **220**
11.1 Endothelial cells 220
11.2 Leukocytes 222
11.3 Pericytes 222
11.4 Fibroblast and extracellular matrix 223

12 Appendix - Research Methods **226**
12.1 Transforming omics measurements - extended 226
12.2 Normalising omics measurements - extended 227
12.3 Handling missing measurements - Extended 230

13 Appendix - Transcriptomics Quality Assessment **232**

List of Figures

2.1 Simplified heart physiology (Created with Biorender.com). 15
2.2 Cardiomyocyte's ultrastructure. The figure shows the hierarchical organisation of the cardiomyocyte, with emphasis on the structure of myofibril, which comprises actin and myosin filaments. (Created with BioRender.com). 20
2.3 Overview of cardiac energy and contractility apparatus. (Created with BioRender.com). 22
2.4 A schematic of cardiac metabolism, fatty acid transport from capillaries, and lipid droplets. The figure also emphasises the interlink between metabolism and lipid peroxidation. (Created with BioRender.com). 24
2.5 Summary of fatty acid transport in the mitochondria and their oxidation (FAO or β-oxidation). The figure highlights the role of the acylcarnitine-carnitine shuttle that allows the transportation of the fatty acids in the mitochondrion membrane. The β-oxidation, or FAO, yields $FADH_2$ and NADH. (Created with BioRender.com). 33
2.6 Schematic of glycolysis. The glycolysis breaks down the glucose molecule into two pyruvate molecules, generating ATP and NADH. (Created with BioRender.com). . . . 37
2.7 Overview of the Krebs Cycle. Starting from Acetyl-CoA, the Krebs generates ATP, $FADH_2$ and NADH molecules. (Created with BioRender.com). 39
2.8 Summary of key molecular reactions of the oxidative phosphorylation from NADH. The interaction NADH-O_2 and $FADH_2$- O_2 yield electrons and energy availability, which will drive the synthesis of ATP from ADP. (Created with BioRender.com). 40

5.1 Schematic of the study design. Mathematical model details are not included. 81
5.2 Principal component analysis with 95% confidence intervals. **(a)** Application on proteomics. **(b)** Application on metabolomics. **(c)** Application on Lipidomics. **(d)** Application on transcriptomics 94

6.1 Proteomics. Heatmap. 98
6.2 Proteomics. **(a)** Volcano plot. **(b)** Bubble plot. 99
6.3 Metabolomics. Heatmap. 102
6.4 Metabolomics. **(a)** Volcano plot. **(b)** Bubble plot. 103
6.5 Metabolomics. **(a)** NADP/NADPH ratio. **(b)** NAD/NADH ratio. **(c)** ADP/ATP ratio. **(d)** GSSG/GSH ratio 106
6.6 Lipidomics. Heatmap 108
6.7 Lipidomics. Lipidomics. **(a)** Volcano plot. **(b)** Bubble plot. 109
6.8 Lipidomics - Cer. **(a)** Analysis of fatty acid tail. **(b)** Analysis of polyunsaturated fatty acids. 110

6.9 Lipidomics - DG. **(a)** Analysis of fatty acid tail. **(b)** Analysis of polyunsaturated fatty acids. . . . 111
6.10 Lipidomics - Cer. **(a)** Analysis of fatty acid tail. **(b)** Analysis of polyunsaturated fatty acids. . . . 112
6.11 Lipidomics. Analysis of fatty acids length . . . 112
6.12 Lipidomics. **(a)** Lipids' class $\log_2$ abundance. **(b)** Composition lipids dataset. **(c)** Lipids enrichment analysis. . . . 113
6.13 Transcriptomics. Heatmap. . . . 114
6.14 Transcriptomics. **(a)** Volcano plot. **(b)** Bubble plot. . . . 115
6.15 Transcriptomics. **(a)** Lipids' class $\log_2$ abundance. **(b)** Composition lipids dataset. **c** Lipids enrichment analysis. . . . 116

7.1 Proteomics. **(a)** Protein-protein network with largest community and hubs highlighted. **(b)** Overrepresentation analysis of the largest community of proteins. For each pathway, a sign analysis of the correlation is also presented. . . . 120
7.2 Metabolomics. **(a)** Metabolite-metabolite network with largest community and hubs highlighted. **(b)** Overrepresentation analysis of the largest community of metabolites. For each pathway, a sign analysis of the correlation is also presented. . . . 122
7.3 Lipidomics. **(a)** Lipid-lipid network with largest community and hubs highlighted. **(b)** Overrepresentation analysis of the largest community of lipids. For each pathway, a sign analysis of the correlation is also presented. . . . 125
7.4 Transcriptomics. **(a)** Transcript-transcript network with largest community and hubs highlighted. **(b)** Overrepresentation analysis of the largest community of transcripts. For each pathway, a sign analysis of the correlation is also presented. . . . 127

8.1 **(a)** Core metabolic and mechanical network interlinking the TCA cycle, aerobic glycolysis and oxidative phosphorylation with the contractile machinery and calcium signalling. The analytes are coloured based on the $\log_2$(Old/Young). The analytes with aquamarine borders are the leading edges of omics SEA. **(b)** Omics SEA. The enriched pathways with $p_{fdr} < 0.05$ are highlighted, and the bars are proportional to $-log_{10}p_{val}$ of each pathway. **(c)** PCr/ATP ratio. **(d)** Dysregulation of β-oxidation via omics SEA. . . . 134
8.2 Schematic of β-oxidation. . . . 135
8.3 Expression pattern in oxidative phosphorylation complexes. . . . 136

9.1 Summary of the findings. . . . 139

13.1 Transcriptomics. Quality control first strand of sample *Young 3* . . . 232
13.2 Transcriptomics. Quality control second strand of sample *Young 3* . . . 233
13.3 Transcriptomics. Quality control first strand of sample *Old 4* . . . 234
13.4 Transcriptomics. Quality control second strand of sample *Old 4* . . . 235

List of Tables

5.1 Profile of heart donors. . . . 85
5.2 Samples in each experiment. . . . 92

6.1 Top 10 dysregulated proteins. . . . 99
6.2 Top 10 dysregulated metabolites. . . . 103
6.3 Top 10 dysregulated lipids. . . . 109
6.4 Top 10 dysregulated lipid classes. . . . 113
6.5 Top 10 dysregulated transcripts. . . . 115

7.1 Top 10 proteins ranked by degree. . . . 120
7.2 Top 10 biological process associated with the largest community of proteins. . . . 121
7.3 Top 10 metabolites ranked by degree. . . . 123
7.4 Top 10 biological process associated with the largest community of metabolites. . . . 124
7.5 Top 10 lipids ranked by degree. . . . 125
7.6 Top 10 lipid classes associated with the largest community of lipids. . . . 126
7.7 Top 10 proteins ranked by degree. . . . 127
7.8 Top 10 biological process of the largest community of transcripts. . . . 128

8.1 Top 10 dysregulated pathways. . . . 133
8.2 Oxidative phosphorylation complexes. . . . 136
8.3 Top 10 age-related signalling pathways in the hearts. . . . 137

Chapter 1

Introduction

Cardiovascular diseases, such as heart failure, are one of the leading causes of death, and ageing is a well-established, albeit not widely known, risk factor for cardiac morbidity[499,559,825,936]. As a lifestyle and medical practices have improved, cardiovascular disease risks are expected to increase[564], which can contribute to the increase of healthcare costs and limit one's access to appropriate medical treatments[838]. Furthermore, those who survive cardiomyopathies can often be left with disabilities[369], which would exert financial and socioeconomic impacts on themselves, family members and the community[413,705]. The increase in cardiovascular diseases partially attributed to ageing has set a novel and challenging goal for the scientific community: understanding how cardiac tissues age. Although aged hearts might not develop cardiomyopathies, their functionality, shape, histology and molecular composition decrease over the course of time. Frailty of the cardiac muscle is developed around the age of 50[24,62,838], where mortality for cardiovascular diseases arise for both genders[564]. On the other hand, mortality due to non-congenital cardiovascular diseases is almost negligible for young individuals[564], usually aged between 12 and 25[240,541,542].

Studies have sought to characterise the molecular landscape by targeting specific molecular events[473] or conducting a scale characterisation of the whole cardiac tissue[299], and produced several insights and hypotheses[119]. Cardiac ageing is a biological process characterised by a constant accumulation of features that will eventually lead to the aberrant functioning of the biological processes in cardiac cell populations[304]. As the heart ages, an increase in the inflammatory reaction is observed[235], possibly due to poor diet, sedentary habits, accumulation of damaged proteins and cellular senescence[203,477]. Accumulation of damaged protein is also associated with cellular senescence in the heart[225], a cell state that propagates to surrounding cells and leads to the irreversible loss of cells[124] in the cardiac tissue, exposing it to the development of cardiac diseases[558]. The state of cellular senescence is induced by multiple occurrences, including elongation of the DNA damage. The senescence state brings neighbour cells to the same state by releasing multiple signalling molecules, such as growth factors, lipids and cytokines, and creates a domino effect that also affects the extracellular matrix[558]. Multiple signalling events exhibit age-related changes. Most notably, the NF-κB signalling pathway, a regulator of inflammatory reactions and fibrosis, suggests dysregulation with cardiac ageing, which is

generally induced by oxidative stress and sirtuin signalling[546]. Oxidative stress in cardiac cells is induced by HIF-1, a signalling molecule stimulated by the redox imbalance[587]. The activation of some signalling pathways can also affect the overall metabolism of the cardiac tissue; for instance, aberrant regulation of mTOR and AMPK signalling events can derail metabolic activities of the hearts, and both pathways show evidence of age-related changes in their activities[52,172]. The changes in metabolism are also well captured during ageing. Fatty acids, the preferred energy substrate in non-failing adult hearts, are oxidised through the β-oxidation and yield ATP, a highly energetic metabolite[775]. With ageing, the myocardial tissue derails the energy production machinery from β-oxidation to glycolysis, a metabolic chain of reaction that yields ATP from the breakdown of glucose[1,899]. While this particular pattern is not solely observed in ageing, it is well observed in heart failure[74].

As the heart appears to change its fuel preference and signalling mechanism with ageing[899], a thorough investigation of the enzymatic and non-enzymatic proteins, as well as a measure of fuel in the human cardiac tissue, such as lipids and metabolites, is required to better understand the human cardiac ageing and generate novel hypotheses on the relationship between ageing and cardiovascular diseases[473]. Detecting alteration in the molecular landscape in individuals who have reached the age that can be considered a risk factor for cardiovascular diseases can unveil a unique view of the driving factors for frailty in the aged heart. This novel perspective would enhance possible therapies[861] that can mitigate myocardial ageing and cardiac frailty to cardiovascular diseases[399] and eventually benefit society. Nevertheless, there are three main factors that have limited research into ageing[1]. The first factor concerns the biological models, as most of the studies have been conducted in non-human species, which affects the translation of the findings to humans[97,230]. Additionally, high-throughput experiments tend to utilise, albeit not always, a very small sample size[292]. The second limitation in the literature is the use of post-mortem human or animal models, which can restrict the validity of the findings in living humans[229]. The third limitation is the lack of simultaneous detection on the same biological sample of multiple types of molecules. This limits the understanding of interactions among different omics and identification of dysregulated biological functions and hinders a systematic exploration of cardiac ageing[394].

To overcome the shortcomings of previous studies on cardiac ageing, our project has used the unique collection of human cardiac tissues offered by the Sydney Heart Bank to examine and detect possible age-associated molecular and functional changes in the hearts[842]. The samples selected for our study are unique inasmuch as the left ventricular samples were excised from individuals whose hearts had been planned for transplantation due to irreversible brain death. Their myocardial tissues were obtained in the operating theatre before the living cardiac muscle. Overall, we obtained 25 myocardial tissues from healthy donors, 12 belonging to donors aged below 25, and 13 samples from donors over 50. To our knowledge, this is the first study conducted on pre-mortem myocardial tissues. The abundance of each sample has also allowed us to conduct multiple high-throughput investigations on the same samples and identify proteins, metabolites, lipidomics and transcriptomics. The vast molecular landscape mapped for young and old myocardial tissues was then employed to address the three main aims below.

This thesis delves into the mechanism of human cardiac ageing with three main aims:

- Characterise the age-related changes in the cardiac molecular landscape.
- Identify regulators and functional modules of cardiac ageing for each omic biological network.
- Identify dysregulated biological processes that drive the human cardiac ageing.

With the first aim, I identify molecules that exhibit the largest variation across the young and old ventricular tissues. The molecular landscape captured in the age-OME study includes proteins, metabolites, lipids and transcripts. This aim would greatly enhance the understanding of the ageing heart and further corroborate the findings in human plasma[742] presented by Shen et al. (2024). For the second aim, constructing a biological network allows the identification of master regulators of the cardiac ageing process in proteins, metabolites, lipids and transcripts. Interpreting each omic measurement as a biological network also provides an opportunity to investigate interactions among molecules and identify the functional modules in the network. To address this aim, I create a novel workflow that identifies modules/communities in any type of biological network and their functional roles. This novel framework has been applied in the age-OME study. The absence of appropriate tools to identify dysregulated biological processes to conduct an investigation and address the third aim has led me to develop a novel method. Utilising and empowering all the omics measurements characterised during the age-OME project, it detects aberrant biological processes. With the three aims, this thesis provides a comprehensive and unique overview of human cardiac ageing.

The thesis begins with a comprehensive literature review on the heart (Chapter 2). The first part introduces the circulatory systems and heart anatomy, followed by a review of different cardiac cell types and extracellular matrix, with an in-depth insight into cardiomyocytes. The chapter also touches on cardiac metabolism and the role of lipids in the heart and concludes with a detailed overview of cardiac age-associated changes detected by other studies. Chapter 3 elucidates in great detail the aims, scopes and research questions underlying this thesis and discusses my contributions outside the human age-OME project and relevant awards. In Chapter 4, I discuss and introduce the rationale of the research methods used to investigate human cardiac ageing. In particular, Section 4.9 discusses the novel framework developed to conduct the analysis of biological network (Aim 2), and Section 4.10 introduces the novel method to identify dysregulated pathways associated with the human cardiac ageing (Aim 3). Full disclosure of experimental and computational protocols, along with sample size calculations and explorative analysis for the age-OME project, can be found in Chapter 5. The following three chapters display the results. Chapter 6 discusses in detail the molecular signature of the cardiac human ageing identified in the pre-mortem myocardial tissues. In Chapter 7, I present the outcomes of the biological network analysis performed for each omic. Chapter 8 summarises the outcomes of the multiple biochemical network investigations I conducted during the age-OME study. In Chapter 9 I thoroughly discuss the findings and their implications, followed by the conclusive chapter 10.

Chapter 2

Heart, Metabolism, and Ageing

In this chapter, I expound on the background of this thesis by drawing from literature on the human cardiac system and its ageing process. The human cardiovascular system supplies cells, tissues and organs with oxygen and nutrients to maintain vitality and function. While cardiac function plays a pivotal role in an organism, the intrinsic ageing process inevitably leads to molecular and physiological changes within the cardiac tissues, increasing the risk of a range of cardiovascular diseases (CVD). At the cellular level, ageing involves dysregulation of biological processes, such as metabolism, signalling pathways, inflammatory responses and gene expressions[723], which increases the risk of CVDs[1]. Human cardiac ageing is known as the major risk factor of the ”frailty syndrome”, a clinical term that describes an individual at risk of disabling events[532]. A comprehensive clinical picture of a patient at risk of CVDs entails a range of factors and other age-associated comorbidities[709]. However, examining the specific effect of age-related processes on the heart holds significance in deepening our understanding of cardiovascular-related diseases and personalised medicine[87,621]. As such, I aim to dissect the comprehensive picture of human cardiac ageing and its causative role in cardiovascular diseases by focusing on the left ventricle[748]. This chapter critically reviews the existing research that has helped formulate the research hypotheses and identify the appropriate methods for this thesis.

This chapter consists of four sections. First, I provide a broad overview of the circulatory system, with a focus on the heart’s anatomy and functions. The following section introduces the histology of the cardiac wall, including cardiomyocytes and their ultrastructure, while other cell types and extracellular matrix composition of the cardiac wall are discussed in Chapter 11. Consequently, I review the main literature on cardiac metabolism, closely examining lipids and glucose, two essential substrates for energy production and storage in the heart[580]. In the final section, I discuss how ageing affects the heart, particularly the left ventricle[748], from geometrical to molecular changes, followed by the conclusion.

2.1 Circulatory system, heart anatomy and its functions

2.1.1 Overview of the circulatory system

For humans, the exchange of fluids, gases, metabolites and other substrates between the organism and the external environment is fulfilled by the double and closed circulatory system, where the heart plays the essential role. The cardiovascular system can be divided into pulmonary and systemic circuits, which involve blood vessels, namely, arteries and veins, blood, the lymphatic system and the heart. The heart comprises synchronised left (pressure) and right (volumetric) pumps; the pulmonary system originates with blood injection from the right pump into the pulmonary artery, which transports blood toward pulmonary capillaries embedded in the lungs. Once the blood is enriched with oxygen and nutrients, pulmonary capillaries converge to the pulmonary vein, letting blood flow into the left pump, concluding the pulmonary circuit. The systemic circuit begins with blood flowing from the left pump through the aorta and ends when the blood returns to the right pumps through the inferior or superior vena cava[23,410].

Specifically, the exchange of nutrients and cellular waste between interstitial fluid and blood occurs through diffusive processes. Containing blood rich in nutrients, arteries branch in each organ and become arterioles, which further branch as capillaries, small vessels with porous walls. The dense network of capillaries, known as capillary beds, spreads across different tissues and offers the exchange site. Cells trade nutrients for metabolic wastes, using extracellular fluid as a medium. Deoxygenated blood and cellular wastes collected in the systemic circuit are transported by veins and partially expelled through the lungs in the pulmonary circuit[108,410,426]. Whilst capillary walls are impermeable for macromolecules, small molecules can escape from capillaries and enter the cardiac interstitium. The cardiac lymph in excess is then collected by contractile lymphatic vessels and returned to the circulatory system through the subclavian vein after being filtered by lymph nodes. The main lymph nodes in the heart include pre-tracheal and cardiac lymph nodes, connected to each other close to the superior vena cava[209,569].

2.1.2 The function of the heart in the circulatory system

The heart's right and left pumps contain distinct chambers: the atrium and the ventricle (See Figure 2.1). The chambers are connected through the atrioventricular valves, called mitral for the left and tricuspid valve for the right pump. Both atrioventricular valves are anchored by the chordae tendineae, so-called "heartstrings", to the papillary muscle in the respective ventricular walls. The papillary muscles are embedded in the anterolateral and posteromedial locations in the left ventricle. The aortic valve connects the left ventricle to the systemic circulation, and the pulmonary valve connects the right ventricle to the pulmonary circulation. Their main function is to prevent regurgitation of blood. The aortic valve is also responsible for supplying blood to the whole organ by means of the left main coronary artery and the right coronary artery, which originate from the sinuses of Valsalva located in the right posterior cusp and the anterior cusp of the aortic valve. Furthermore, the left main coronary artery branches into the left anterior descending artery and circumflex artery, while

the right coronary artery gives rise to the posterior descending artery. Those main arteries supply blood to the entire cardiac walls through arterioles and perforating arteries[23,410,792].

Due to low systolic pressure and the amount of blood that can be stored, the right pump

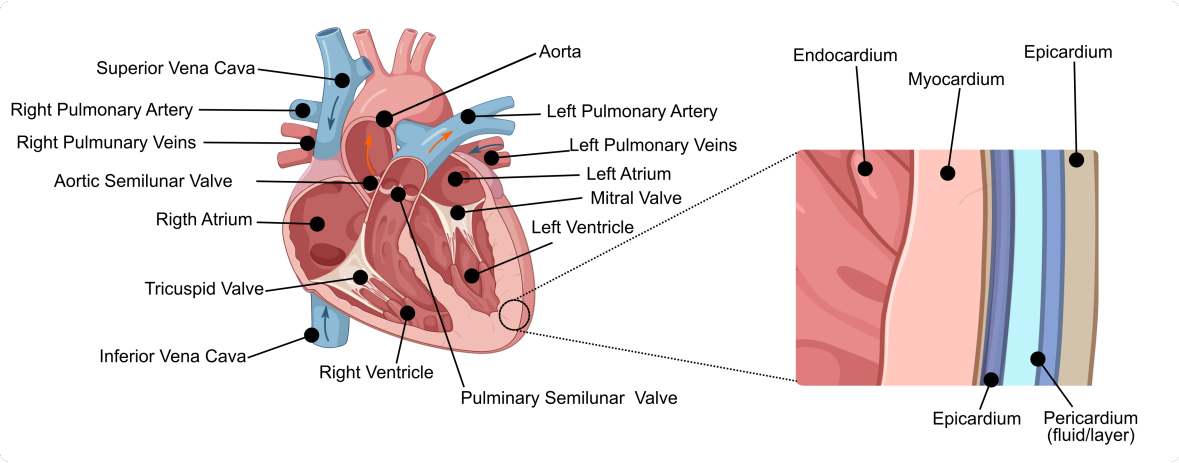


Figure 2.1: Simplified heart physiology (Created with Biorender.com).

can be seen as a volumetric pump[150,856]. The right atrium receives low-pressure blood, approximately $4mmHg$ ($1mmHg \approx 133.322Pa$), from the systemic circulation. Blood flows from the right atrium to the right ventricle via the tricuspid valve by atrial contractions. Similarly to the atrium, the right ventricle can store a considerable amount of blood with respect to the left, and the wall that wraps the right ventricle is thin but inelastic. The pulmonary valve links the right ventricle to the pulmonary artery, and the systolic pressure (a maximal pressure generated by the contraction of a ventricle) of the pulmonary artery ranges from $20mmHg$ to $30mmHg$. On the other hand, the left part of the heart is referred to as a pressure pump. With a pressure around $10mmHg$, oxygen-rich blood from pulmonary circulation enters the left atrium via the pulmonary vein. Blood in the atrium passively flows in the left ventricle through the mitral valve, while the atrial kick increases blood flow toward the ventricle only in a second moment. The left ventricle is smaller and surrounded by a thicker cardiac wall than the right one. The left ventricular contraction generates a high maximal pressure ranging from $100mmHg$ to $140mmHg$ in a healthy condition[426,905].

2.1.3 Heart anatomy and its functions

The heart is a highly contractile muscle with connective and endothelial tissues arranged in a helical shape[823]. The macroscopic structure of the heart comprises three main components. The first is a circumferential band, the basal loop, surrounding the left and right ventricles. The basal loop also forms the free wall of the right ventricle, which is responsible for heart compression. Next, the helix is embedded in the apical part of the cardiac muscle, surrounded by the basal loop. It branches into two oblique arms, inner and outer helical coils, which constitute the remaining main components of the heart. The inner helix ascends from the apex to the muscle base, constituting the septum and left ventricular free wall[956]. In the adult human left ventricle, the free wall is thicker (12-$15mm$ circa) at its obtuse angle

and thinner at its apex (1-$2mm$). The outer helix descends from the base to the apex of the cardiac muscle. Notably, while the septum and left ventricular walls are often described as different anatomical parts of the heart, they are made of the same muscle band[956]. The septum presents a smooth appearance except for a few scattered thick bundles of myofibers, and it is located between the two ventricles. In particular, the septum has an arch shape and is concave to the left ventricle[335]. The helices and circumferential band allow interactions between the two ventricles; the interactions become more vivid in the case of dilated cardiomyopathy when the stretched right ventricle affects the atrioventricular blood flow of the left ventricle during the diastolic phase[881].

The cardiac functions are closely related to the nervous system. The sinoatrial node is embedded in the septum near the right atrium. Cardiomyocytes are auto-rhythmic muscle cells and generate electrical signals that spell timing and rates for cardiac contractions[23,410]. The thin wall surrounding both atria contains the pectinate muscles, which deliver electrical signals from the sinoatrial node to the atrioventricular node. The pectinate muscles, known as the sinoatrial ring bundles, are made of Purkinje cells. The atrioventricular node is placed between the two pumps and functions as a transmitter-delayer, ensuring that atria contract ahead of the ventricles during systole[23,90,698,715]. The parasympathetic division is cross-linked to the heart by innervating the sinoatrial node with the right vagus nerve, while the left vagus nerve innervates the atrioventricular node; the activation of vagus nerves leads to a decrease in the heart rate[110]. The sympathetic system interacts mainly with atrial muscles via efferent nerves, which are present particularly in the atrial wall close to the sinoatrial node, although little efferent innervation is present in the ventricular walls. The sympathetic innervation can increase inotropy (contractility of cardiac muscle) and chronotropy (pace of heartbeats) by releasing norepinephrine to the cardiomyocytes[410,426].

The cardiac muscle can perform six mechanical actions - to shorten, narrow, increase its length, increase its width, twist and uncoil - while undergoing four main phases: isovolumic contraction, ejection phase, isovolumic relaxation and diastole. The first two phases are part of systole, during which cardiac torsion is brought about by the double helix model[796,828]. The inner and outer helix, respectively, give rise to a clockwise twist at the base and a counter-clockwise twist at the apex. The isovolumic contraction phase starts with the contraction of the myocardium, which decreases the length of the long and short axes of the heart while conserving the volume of the left ventricle[796]. Then, the ejection phase begins with the opening of the aortic valve. During the ejection phase, the muscle torsion peaks, and circumferential shortening exceeds longitudinal coiling. Studies also suggest the independence of recoil from torsion; uncoiling does not start unless the inner helix relaxes, as the ventricle cannot lengthen. The downward coiling of the cardiac muscle accounts for 60% of the total ejection fraction during the systolic phase, mainly dictated by the contraction of the two helices, which leads to the long axis strain[956]. The contraction of the circumferential wrap accounts for 30% of the total ejection fraction by straining the short axis of the heart; its aberrant function is usually observed in dilated hearts. The ejection phase ends with the closure of the aortic valve[681]. At the end of the systole, the mitral valve comprised of leaflets connected to an annulus through papillary muscles, opens and lets blood flow from the left atrium to the left ventricle. The opening is due to the clockwise twisting of the circumferential wrap, which increases the pressure of the left atrium above the

ventricular pressure. Simultaneously, the outer helix lengthens the cardiac muscle after the inner helix ends its contraction. Once blood ends in the ventricle, the circumferential wrap recoils, enabling the closure of the mitral valve leaflets. The time interval from the aortic valve closure to the mitral valve opening is known as isovolumic relaxation time. During this period, the ventricle begins its relaxation (diastole) while conserving its volume; the inner helix, however, displays contraction[956]. The synchronised torsion and coil are caused by the circumferential wrap and helices and play an essential role in healthy heart functions. In healthy individuals, the isovolumic relaxation time lasts, on average, for $90ms$. In contrast, it takes longer for those with several cardiac conditions, such as ischaemic cardiomyopathy. Finally, the diastole begins with the relaxation of the inner helix, which allows the blood to flow from the left atrium to the left ventricle through the mitral valve[828,956].

2.2 Cardiac cell types and extracellular matrix

2.2.1 Cardiac cell types and cell counts

The cardiac wall is an inelastic multi-layer fibrous structure that wraps the atria and ventricles, and multiple cell types populate it. The cardiac wall comprises three layers: endocardium, myocardium and epicardium. The endocardium is the innermost, thin, squamous layer of cells composed of a few endothelial cells supported by collagen and elastic fibres[923]. At the apex of the left ventricle, the endocardium contains interlaced bundles of muscle fibres[335]. As the thickest layer, the myocardium constitutes the main part of the cardiac wall. It lies between the endocardium and pericardium and consists of myocytes and connective tissue. The pericardium continuously extends until it becomes the epicardium, the external part of the cardiac sac. The outermost epicardium is a dense layer of connective tissue, elastic fibres and myocytes. The outermost part of the epicardium is covered by a layer of epithelial cells, which secrete lubricating fluids[410,923].

Studies have sought to reveal and better understand such cells in cardiac muscle, ranging from cardiomyocytes and nonmyocytes to non-cardiomyocytes. Cardiomyocytes, highly contractile striated muscle cells, occupy $70 - 85\%$ of the heart's volume[948]. Using genetic tracers combined with flow cytometry, Pinto et al. (2016) deduce that cardiomyocytes compose approximately $25 - 35\%$ of the cell count in cardiac muscle tissues. Amongst nonmyocytes, endothelial cells are the most predominant cell, accounting for more than 50% of the total cell population. Leukocytes, or white blood cells, account for about 2.8% while $1.4 - 1.9\%$ of the cell population are dedicated to interstitial cells, such as fibroblasts[948]. Yet, the remaining composition of the cell population is unclear[649]. Furthermore, the role of noncardiomyocytes in the cardiac muscle is still a matter of research, especially due to the difficulty of accessing human tissue samples[649].

Nonetheless, subsequent studies have presented evidence of different cell counts. Tucker et al. (2020) have shown that fibroblasts and cardiomyocytes have almost a $1:1$ ratio in healthy donors. The research was conducted using single nucleus RNAseq on 7 adult myocardia (4 female, 3 male) across four different sections of the heart, that are, right

ventricle, right atria, left ventricle and left atria[833]. Litviňuková et al. (2020), moreover, have examined the cell counting in the human heart and possible histological differences across six different cardiac regions, namely, left atrium, left ventricle, right atrium, right ventricle, septum, and apex. The study was conducted on 14 individuals (7 female, 7 male), where 8 hearts belonged to the United Kingdom Biobank, and 6 were from the Canada Biobank. The age of donors spanned from 40 to 75 years old. Six samples were from donors deceased after circulatory death, while eight were from donors deceased after brain death. They investigated the cell types in the region using single-nucleus RNA sequencing (snRNA-seq) on 14 samples, single-cell RNA sequencing (scRNA-seq) on 7 samples and multiplex smFISH imaging. Combining these three techniques, the authors determined different types of cardiomyocytes, endothelial cells and immune system cells and their spatial distribution. Cardiomyocytes accounted for 49.2% of the cells in the left ventricle, smooth muscle cells and pericytes 21.2%, fibroblasts 15.5%, while endothelial cells and immune cells accounted for 7.8% and 5.3% respectively. Cell distributions across ventricular cells tend not to change based on gender, although a qualitative increase of cardiomyocytes was observed in female donors compared to male[504].

2.2.2 Cardiomyocytes

Researchers have scrutinised molecular differences between atrial and ventricular cardiomyocytes and further examined cardiomyocytes[833]. Ventricular myocytes have approximately 2300 detected differential expressed genes, including high expression of myosin heavy chain 7 (MYH7) and hes-related family bHLH transcription factor with YRPW motif 2 (HEY2), two protein-encoding genes associated with cardiac dysfunction and diseases, in addition to myosin light chain 2 (MYL2) genes[685,846]. Ventricular cardiomyocytes also have high expression of transcription factors, such as Iroquois homeobox 3 (IRX3), Iroquois homeobox 5 (IRX5), Iroquois homeobox 6 (IRX6), MBL-associated serine protease 1 (MASP1), and PR/SET domain 16 (PRDM16). While highlighting the relatively higher percentage of cardiomyocytes in the left ventricle[649,948] as discussed in Section 2.2, Litviňuková et al. (2020) have also concluded that ventricular cardiomyocytes (vCMs) have five different types, ranging from vCMs1 to vCMs5.

vCMs1 are the most abundant cardiomyocytes in the left ventricle, accounting for 63.9% of ventricular myocytes. vCM2 comprises 9.1% of cardiomyocytes, showing high expression of the PRELI domain containing 2 (PRELID2), although its function is still unknown. vCMs2 also highly express myosin heavy chain 6 (MYH6) and cadherin 13 (CDH13) genes. vCMs5 comprise approximately 1% of cardiomyocytes, with a characteristic expression of DLC1, GTPase-activating protein regulating apoptosis[322], and EBF2, an EBF transcription factor that regulates brown adipocytes by changing the activities of peroxisome proliferator-activated receptor-γ (PPARγ)[667]. vCMs3 and vCMs4 account for the remaining cardiomyocytes. High abundances of MYH9, a nonmuscle myosin heavy chain[221], Nexilin F-actin binding protein (NEXN), a protein responsible for the development of transverse tubules (t-tubules)[773], and calponin 1 (CNN1), a protein that modulates contraction of smooth muscle cells[375], characterise the vCM3 genotype. vCM4 are characterised by high-level genes that encode subunits of complexes of oxidative phosphorylation, such as

NADH:ubiquinone oxidoreductase subunit B11 (NDUFB11), NDUFA4 mitochondrial complex associated, cytochrome c oxidase subunit 7C (COX7C), cytochrome c oxidase subunit 5B (COX5B), and crystallin alpha B (CRYAB), a small heat shock protein that protects the heart from various cellular stresses[941], in addition to phospholamban (PLN), an important inhibitor of ATP2A2 (or SERCA2)[122].

Cardiac myocytes can also be grouped based on functions and aspects in working myocytes, Purkinje, nodal and transition cells[410]. Working myocytes present a central nucleus and striation, and its striated aspect is conferred by the regular arrangement of actomyosin (actin and myosin) fibres. Working myocytes are rich in mitochondria and myofilament (or myosin filaments) and mainly rely on oxidative metabolisms, such as β-oxidation and aerobic glycolysis for producing adenosine triphosphate (ATP). Purkinje fibres, responsible for rapid electrical signals, are rich in glycogen yet lack actin-myosin filaments. Purkinje cells are oval-shaped and can have more than one nucleus, although they do not have regenerative potential after cardiac infarction[870]. Unlike working myocytes, Purkinje myocytes comprise most of the pectinate muscles and ventricular walls and can synthesise a high amount of ATP even in anaerobic conditions. Clusters of working myocytes and Purkinje myocytes are not separate since transition cells mediate between the two groups. Similarly to Purkinje fibres, nodal cells can anaerobically produce high energy and are placed in the sinoatrial and atrioventricular nodes. The main function of nodal cells is to regulate the heart rate and delay the electrical signal from the sinoatrial to the atrioventricular node, helping the atria to contract before the ventricles. Nodal cells are morphologically similar to Purkinje cells, whereas the former are smaller[410]. Despite different biological functions, cardiomyocytes have branches at both ends that interconnect adjacent cardiac cells through intercalated discs. The intercalated discs are positioned perpendicularly to the orientation axes of myocytes and facilitate the spread of contractile stimuli across different cardiac muscle fibres[117,426,923].

2.2.3 Cardiomyocyte's ultrastructure and excitation-contraction mechanism

Cardiomyocytes present a hierarchical ultra-structure (see Figure 2.2) with an average length of $150\mu m$[401], approximately $23.1 \pm 2.2\mu m$ diameter[826], and volume about $40,000\mu m^3$[401]. Using rat models, a study revealed that cardiomyocyte's volume comprises 36% mitochondria, 47% myofibrils, 2% nuclei and 3.5% sarcoplasmic reticulum, while the remaining 11.5% is mainly cytosol, yet standard deviation of the data was not reported numerically[410,624].

The cytoskeleton of cardiomyocytes is a dense network of protein filaments which guarantees the correct functioning of the cardiac cell[401]. The most common filaments in cardiac cells are microfilaments, microtubules and intermediate filaments. Microfilaments present an actin backbone and act as a cellular scaffold[410,502]. Vesicles and mitochondria can use a myosin motor to bind on the microfilament and quickly move within the cell[185]; they can also facilitate connections between cells and between the extracellular matrix and cells[73]. Microtubules are complexes mainly made of tubulin; like microfilaments, tubulin filaments can transport vesicles and organelles across the cell but are coupled to kinesin and dynein motor proteins[468]. Microtubules are also responsible for the movement of cilia

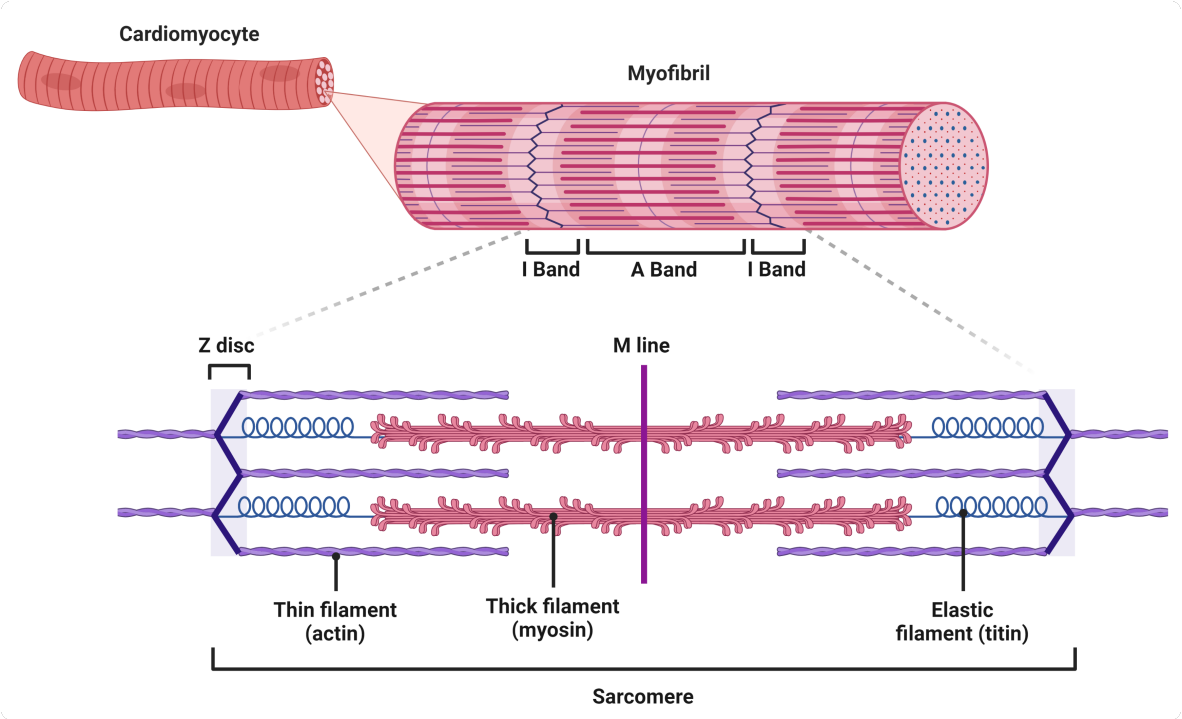


Figure 2.2: Cardiomyocyte's ultrastructure. The figure shows the hierarchical organisation of the cardiomyocyte, with emphasis on the structure of myofibril, which comprises actin and myosin filaments. (Created with BioRender.com).

and flagella[410]. Intermediate filaments are principally made of desmosomes, a polymer composed of several units of desmin. Desmosomes also contribute to linking a cell to adjacent cells or the extracellular matrix[111].

Cardiomyocytes are separated from the extracellular matrix by a phospholipid bilayer membrane, known as the sarcolemma (or plasma membrane). The sarcolemma contains two hydrophilic heads and a hydrophobic core. This particular structure allows it to act as a permeable barrier that partially blocks the exchange between the cytosol and the extracellular matrix. Receptors, ion channels and ion pumps are embedded in the plasma membrane. The difference between intracellular and extracellular matrix ion concentrations forms chemical gradients and confers electrical potential to the plasma membrane. The ions that greatly influence the resting membrane potential, approximately $-90mV$, are calcium (Ca^{2+}), potassium (K^{+}) and sodium (Na^{+}). While ion leaks can occur across the membrane, ionic channels and pumps are the main machinery that maintains ion gradients across the sarcolemma. The ion pumps require energy harvested through ATP hydrolysis, which is commonly referred to as ATP-dependent pumps[426]. Ion channels can be an alternative method for ion exchange across the membrane. Ion channels can be voltage-gated channels[127] or ligand-gated channels[291]. The voltage-gated channels operate when there is a perturbation in the resting membrane potential, while the ligand-gated channels work via chemical signalling[55].

Adjacent myocytes are connected by intercalated disks that contain desmosomes and gap junctions (or connexin channels) principally made of titin and connexin, and both structures are embedded in the sarcolemma. Intracellular invaginations of the sarcolemma are known as t-tubules. T-tubules facilitate communication between the cellular compartment and the extracellular matrix, which is essential for ionic currents. The dense network of sarcoplasmic reticulum, a specialised form of endoplasmic reticulum, surrounds the myofilaments and mitochondria, and it acts as a calcium warehouse, releasing and storing calcium as needed for cardiac contraction. The sarcoplasmic reticulum close to the t-tubules has a sac-like shape, named sarcoplasmic cisternae[697]. The sarcoplasmic reticulum is closely associated with t-tubules, as together, they regulate the excitation-contraction of the myocyte[747]. Specialised calcium pumps on the sarcoplasmic reticulum and cisternae regulate the calcium concentration of the cytosol; the calcium is pumped out of the cell during muscle relaxation while pumped in during contraction[502].

The genetic information of the cardiomyocyte is conserved in the nucleus or nuclei[567]. High energy requirements of the heart and its cardiomyocytes are reflected in the great number of mitochondria within the cells, small organelles that carry oxidative reactions for ATP production. The intercalated discs and myocytes are interconnected, creating a continuous sheet called functional syncytium. Within intercalated discs, gap junctions offer a low-resistance pathway for ionic current between adjacent myocytes so that the functional syncytium can act as a unit to different stimuli[410,426,731].

Structurally, a cardiomyocyte can be depicted as closely packed actomyosin fibrils (myofibrils) structured with bundled myosin filaments (myofilaments), each of which is made of several sarcomeres. The sarcomere is a unit of a myocyte and contains a thick (myosin backbone) and thin filament (actin). Its structure has a central A-band with a thick filament, wrapped by two half-H-bands that contain a thin filament. A complete H-band includes the Z-line, while two half-H-bands belong to two different sarcomeres. The H-band is a less organised part of sarcomeres and is predominantly composed of thin filaments and titin. Thin filaments are made of tropomyosin, troponin and actin proteins. Each helix-shaped tropomyosin can bind with several actin molecules, which can also be observed from the regularly spaced binding spot for the formation of the troponin complex. The troponin complex is comprised of troponin-T (TN-T). TN-T is a protein and can bind to tropomyosin, troponin-C (TN-C), which can bind to Ca^{2+}, and troponin-I (TN-I), a protein that inhibits the binding of actin-myosin or actomyosin, fibres formation. The molecular properties of the troponin complex are important during the excitation-contraction phase of the cardiac muscle, dictated by hormonal and neural stimulation. Hormonal and autonomic nervous stimuli, such as voltage- and calcium-dependent processes, modulate actomyosin fibre contractions and consequently regulate overall cardiomyocyte contractions[649,929]. The contraction of actomyosin fibres depends on the calcium and phosphorylation of the myosin. Indeed, calcium binds first to the calmodulin, activating the myosin's phosphorylation process. Then, the phosphorylated myosin binds to the actin, the phosphorylated form of the actomyosin, a process that permits the formation of cross-bridges. Restoring the basal level of calcium in the dyadic space or dephosphorylation of the actomyosin fibres leads to the return of the cardiomyocyte to its relaxed state[449]. On the other hand, the A-band is made of six thick filaments; the central thick filament is connected with hexagonally organised six thin

filaments through cross-bridges. There are also thin filaments within the hexagon[731]. The cross-bridge formation between thin and thick filaments relies on the interaction between ATP and myosin. The thick filament is made of titin and myosin polymers, and each thick filament contains approximately 300 molecules of myosin. The myosin strands present two heads as a binding site for myosin-ATPase, an enzyme that can hydrolyse ATP to obtain adenosine diphosphate (ADP) and energy[410,426].

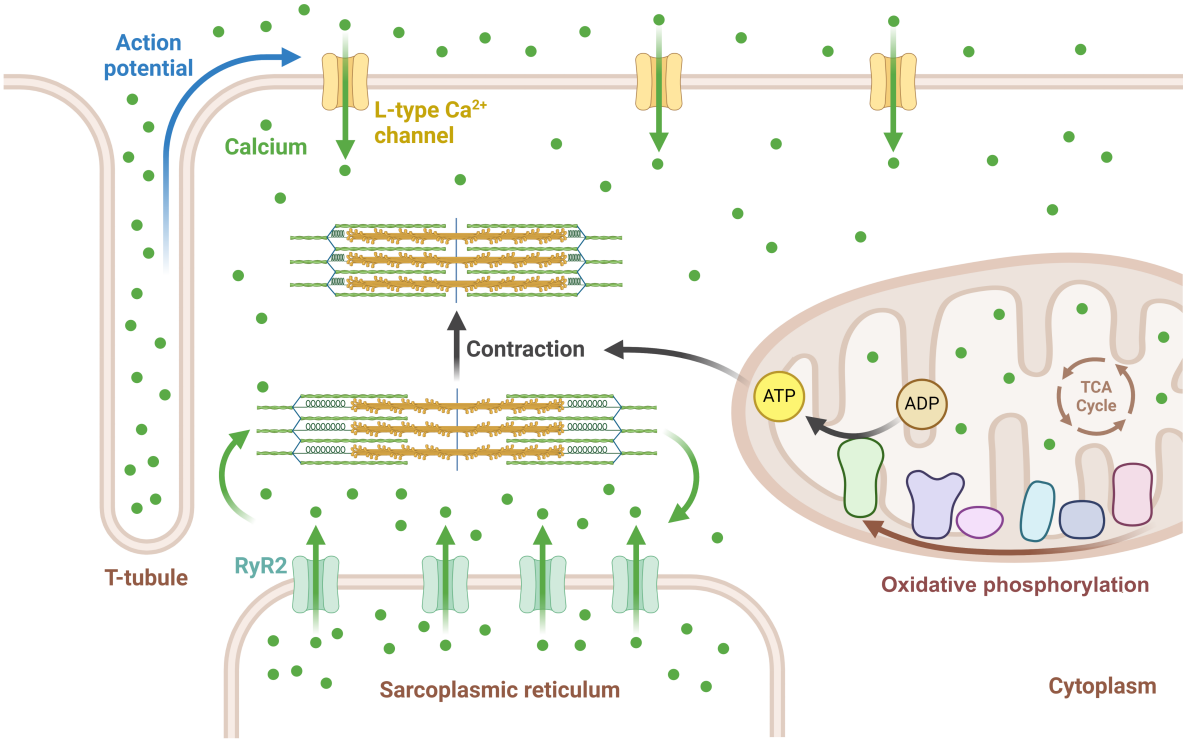


Figure 2.3: Overview of cardiac energy and contractility apparatus. (Created with BioRender.com).

The cross-talk amongst different functional cardiomyocytes is essential to guarantee the excitation-contraction of the cardiac muscle (see Figure 2.3). Depolarisation of the cardiac muscle is triggered by neural stimuli conveyed through pacemaker cells sited in the sinoatrial node. Spontaneous depolarisation of pacemaker cells causes potassium leakage from cardiomyocytes in exchange for sodium and calcium. Calcium, sodium and potassium gradients are responsible for the action potential of the cardiomyocyte. In the resting state, cardiomyocytes present a high concentration of K^+ but a low abundance of Na^+ and Ca^{2+}. The inflow of Na^+, combined with an outward leakage in K^+, triggers (fast) depolarisation of cardiomyocytes. The initial polarisation is then driven by the outflow of K^+ through specialised channels while NA^+ channels become inactive; concurrently, a slow uptake of Ca^{2+} from L-type channels is also observed. As a result, the muscle cell reaches the temporary potential plateau before the full repolarisation. The full repolarisation is characterised by an increase in K^+ uptake in exchange for Ca^{2+} and Na^+[449]. Once the

cardiomyocyte is depolarised, the excitation-contraction phase starts. The L-type calcium channels embedded in the membrane of the cell are responsible for bringing the (trigger) Ca^{2+} into the cytosol, consequently sensed by Ryanodine receptors (RyR). RyRs then release Ca^{2+} stored in the sarcoplasmic reticulum into the cytosol (dyadic space). The cytosolic Ca^{2+} binds with the troponin C, a protein that is part of the actin-troponin filament. Now, the actin filament is ready to bind with the myosin filament and cause the contraction of the cardiomyocyte and ATP hydrolysis. The SERCA ATPase pump meanwhile uptakes Ca^{2+} from the cytosol and places it into the sarcoplasmic reticulum. As the cytosolic abundance of Ca^{2+} decreases, the troponin C dissociates from Ca^{2+} and the cardiomyocyte returns to its relaxed state[433].

2.3 Lipids, glucose and cardiometabolism

As highlighted in Section 2.2, a quick and repetitive shift from contractile to relaxed state of the cardiomyocytes relies on the constant availability of ATP. Cardiomyocytes can survive for only about a minute in anaerobic conditions. In normal, healthy conditions, most of the energy is synthesised through aerobic respiration. A change in the cardiac metabolic pathway can influence the ability to generate ATP (see Figure 2.3) and lead to heart failure[694]. Due to high energy requirements, normal cardiac metabolism is mainly based on free fatty acids, although the heart can rely more on glucose metabolism to match the higher energy demands under stress. Fatty acids produce up to 90% of the total required energy, and the remaining is yielded mainly by glucose metabolism (see Figure 2.4)[409,591,631,775]. This section focuses on lipid trafficking, fatty acid metabolism and glucose metabolism, whilst other metabolic pathways with a marginal contribution to ATP production still guarantee the correct functioning of excitation-contraction in cardiomyocytes. After introducing lipids and their classification, I discuss lipids in the mitochondrial membrane, which are essential for maintaining healthy oxidative metabolism. I then move on to lipid synthesis and trafficking, followed by lipid droplets, lipid metabolism and lipotoxicity. Glucose metabolism, the Krebs cycle, and oxidative phosphorylation are examined subsequently. Finally, this section concludes by discussing the phosphocreatine-creatine pathway for energy production.

2.3.1 An overview of lipids

Lipids are organic compounds that play several essential roles in the body, such as constituting organelle and cell membranes and dictating their permeability to different molecules[152,844]. Lipidomics - the characterisation of lipids in tissues, cells and organelles - poses several challenges, not only technical ones due to the hydrophobic nature of the molecules but also practical ones, as the functions of lipids are highly dependent on intracellular abundance, cell locations as well as activity and regulation of biosynthetic and hydrolytic enzymes. The functions of lipids depend on their position in the cell, mainly due to the highly compartmentalised structure of the cell and the immobility of lipids in organelle and cell membranes without a proper transporter[844]. Recent developments in

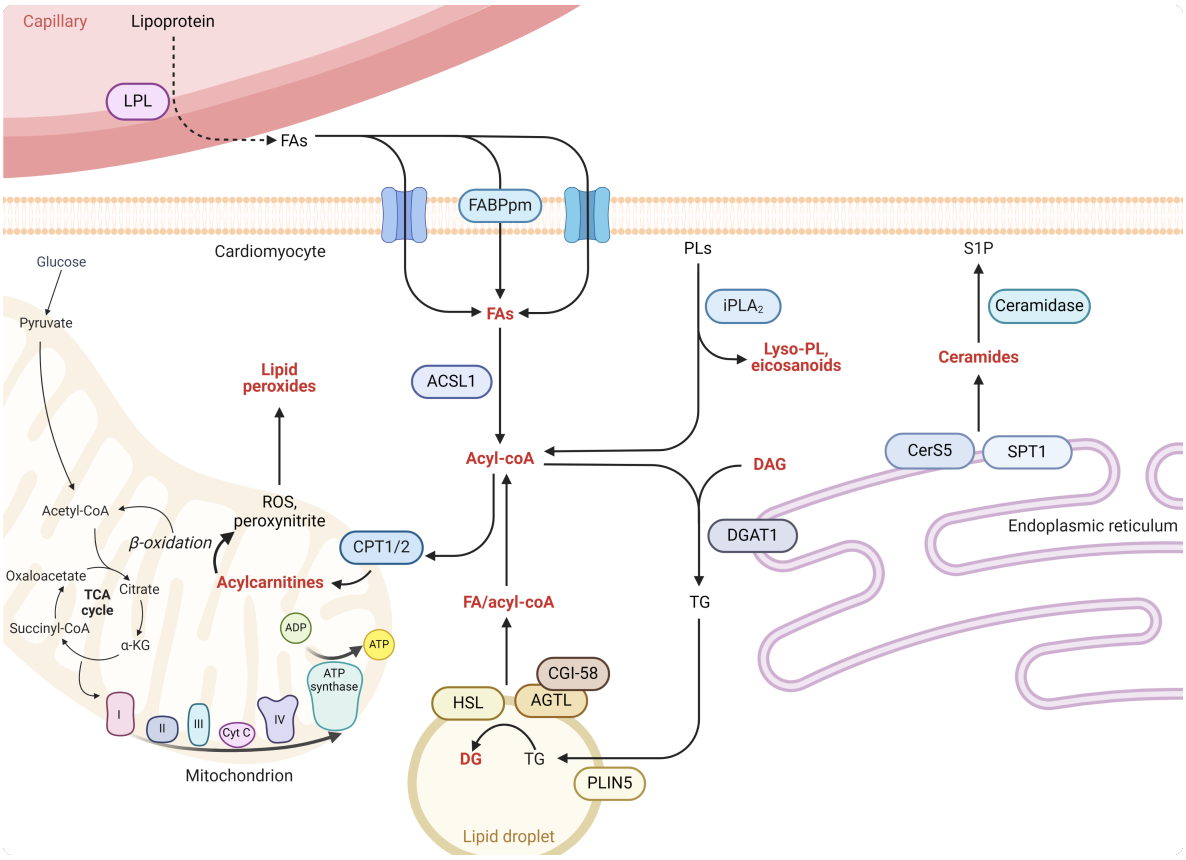


Figure 2.4: A schematic of cardiac metabolism, fatty acid transport from capillaries, and lipid droplets. The figure also emphasises the interlink between metabolism and lipid peroxidation. (Created with BioRender.com).

high-throughput technologies, such as mass spectrometry and nuclear magnetic resonance, have allowed researchers to partially quantify the lipidome, estimate the kinetics of lipid metabolism and unveil lipid-enzyme interactions[152,453,474]. This has enabled researchers to understand further the role of lipids in metabolic diseases, energy storage, production and mobilisation of lipids, as well as their impacts on heart functions[453,454,876,900,955]. Lipids can be classified into eight categories based on the chemical composition and functionality of the lipid backbone, which are: fatty acyls, glycerolipids (GL), glycerophospholipids (GP), sphingolipids (SP), sterols (ST), prenols (PR), saccharolipids (SL) and polyketides (PK)[226,227]. This classification has also been utilised by LIPID MAPS, one of the main databases for lipids[228]. The fatty acyls comprise all the lipids containing a fatty acid or its derivative. Fatty acids can be synthesised through lipid metabolism, such as triacylglycerides and phospholipids[908], and modified via multiple processes, for example, fatty acid elongation[421]. GL contain lipids with one or more fatty acids esterified to glycerol but no phosphate, while those with phosphate esterified to glycerol hydroxyl groups are identified as GP. SP are lipids with aliphatic amino alcohols and sphingosine (sphingoid bases) backbone. While ST and PR share a common biosynthetic pathway, the former contains four carbon rings, and the latter has a linear carbon chain backbone. SL contain lipid molecules with a sugar backbone, to which FA are directly bound. Finally, PK include several metabolites produced by microbes

and plants[226,227].

Based on the classification above, this section focuses on specific classes and related subclasses detected using the mass spectrometer (MS) in the ageing heart project, details of which will further be discussed in the following chapters. Regarding GP, our analyses have detected lysophosphatidylethanolamine (LPE), phosphatidylcholine (PC), phosphatidylglycerol & lysobisphosphatidic acid (PG & LBPA), cardiolipin (CL), lysophosphatidylcholine (LPC), phosphatidylinositol (PI) and phosphatidylethanolamine (PE) [i]. The MS analyses have detected triglyceride (TG) and diacylglycerol (DG) for GL and acylcarnitine (AcCA) for FA. Within SP, ceramide (Cer), hexosyl ceramide (HC), and sphingomyelin (SM) have been detected. This thesis follows lipid nomenclature listed by the International Union of Pure and Applied Chemists and the International Union of Biochemistry and Molecular Biology (IUPAC-IUBMB) Commission on Biochemical Nomenclature[496]. Briefly, the name of a lipid starts with an acronym of a subclass to which it belongs, followed by parentheses containing the length of the carbon chain of each fatty acid and the number of double carbon bonds in it. When double bonds occur, E/Z describes the cis/trans geometry of the lipid's chain fatty acid. The R/S can also be present in the parenthesis to indicate the molecule's stereochemistry, that is, α/β or D/L. Other information can be reported within the parenthesis, such as the stereospecific number, which denotes the position of the glycerol group if present[226]. An application of the lipid notation can be found in Chapter 6.3.

2.3.2 Mitochondrial membranes and related lipids

Mitochondria are the cell's power plant, separated by the cytosol from its double lipid-based membrane, and its complex composition appears to change during apoptosis[659]. The two membranes of mitochondria, the outer (OMM) and inner (IMM) mitochondrial membranes, further compartmentalise organelle functions, particularly the Krebs cycle, fatty acid β-oxidation and lipid synthesis. The space between IMM and OMM is also known as the intermembrane space, which entails different functions and molecular compositions. While the thin space between the membranes is known as the peripheral intermembrane space, the intermembrane space created by invagination of the IMM is called the intracristae space. The former tends to be a more reduced environment than the latter, as ROS and oxidised glutathione are less abundant, while the reduced form of glutathione is more present. With a high concentration of oxidised glutathione, reactive oxygen species (ROS), cytochrome c and protons, the intracristae space is clamped by specific complexes, which interact with lipids within the intermembrane space. Complexes I, II, III, and IV are embedded in the invaginated IMM, making the intracristae space an efficient environment for the electron transport chain. The alteration of IMM, composed mainly of CL, PE and PC lipid classes, can cause severe leakage of ROS from the intracristae space to the peripheral intermembrane space. ROS in the peripheral intermembrane can trigger ROS-induced lipid peroxidation, which converts unsaturated lipids to lipid peroxide, which can be further converted to

[i] PG and LBPA are associated since the MS cannot differentiate the molecule in each class when the m/z of these two subclasses overlap. They are not taken into account when interpreting findings in order to avoid any bias.

oxidatively truncated phospholipids. ROS-induced lipid peroxidation can have substantial implications for IMM and, consequently, for cellular respiration[659].

The mitochondrial membrane mainly comprises GP, while SP and ST are less abundant. Cer is not present in SP of the healthy mitochondria, unlike the plasma membrane. Additionally, GP are utilised as building blocks of mitochondrial double membranes, such as CL, PC, PE and phosphatidylserine (PS). Considering the dry lipid mass, the OMM includes up to 54% of PC, 29% of PE and 14% of PI, unlike the sarcolemma that is made of 50% of PC and 20% of PE. The membrane's cell or organelle features also depend on the distribution of lipids. With few amounts of complexes and the asymmetric distribution of lipids, the OMM embraces a smooth surface; on the other hand, the IMM displays a rough appearance due to the rich presence of complexes, the mix composition of PE and PC (75%) and CL (15-20%), and several invaginations. ST are sparsely present in the OMM, guaranteeing more fluidity to the membrane, while they are almost absent in the IMM[561,806,844,847]. The IMM is highly impermeable and preserves the compartmentalisation and biological function of the cell. Additionally, the similar geometry and function of PE and CL guarantee the cell's existence in case one of them is lost. On the other hand, the simultaneous deficiency of both lipids would compromise several chemical gradients within the intracristae space, which can trigger cell apoptosis[659].

2.3.3 Lipid synthesis and trafficking in mitochondria

Lipid synthase and endocytosis are the two main mechanisms that offer lipids in the cellular environment. Lipid synthase is an anabolic set of reactions that allows the cell to build lipids, and occurs primarily in the endoplasmic reticulum (ER) and Golgi apparatus[148]. Endocytosis refers to the process in which lipid molecules are imported into the cell environment through specialised vesicles[448]. Different categories of lipids play different roles in these processes.

ST are imported into the cellular environment through low-density lipoprotein (LDL) receptor-mediated endocytosis, while a part of ST can be newly synthesised through the ST biosynthesis pathway, involving multiple organelles such as the ER and Golgi apparatus[222]. The presence of ST plays a crucial role in the fluidity of OMM and IMM. Additionally, cholesterol is pivotal for steroid synthesis; the first step of steroidogenesis [ii] occurs within mitochondria[570]. The cholesterol transport in the mitochondrial matrix is mediated by the steroidogenic acute regulatory protein (StAR) family known as STARD1. Similarly, STARD4 also mediates transport of cholesterol into mitochondria but with lower efficiency with respect to STARD1. STARD4 has also been detected in the ER, where it activates acyl-CoA cholesterol acyl transferase 1 (ACAT1) and increases cholesterol esterification. The esterified cholesterol is then stored in lipid droplets, bonafide organelles[141,570]. Alternatively, cholesterol can enter the cytosol through endocytosis, mediated by the LDL receptor, or can be synthesised de novo from acetate in the ER[570].

[ii] Steroidogenesis is the steroid synthesis pathway.

SP also play an important role in signalling molecules for several biological processes, such as cellular and mitochondrial apoptosis[332]. Displaying the simple form of SP, Cer is synthesised in the ER, while more complex forms of SP are synthesised in the Golgi apparatus. Cer can be synthesised through four different enzymatic reactions, where SM, Cer-1-phosphate, sphingosine, or dihydroceramide are the substrates, and SMase, lipid phosphate phosphatase (LPP), ceramide synthase (CerS), and Dh-Cer desaturase are the respective enzymes[248]. Moreover, it has been shown that Cer may be able to create large OMM channels that allow the passage of cytochrome c[195], while it is yet unclear whether mitochondria can synthesise Cer[835]. Aberrant accumulation of Cer in mitochondria can trigger cellular apoptosis or disrupt the mitochondrial microenvironment[926]; the latter depends on the tails of Cer, as shorter tails can increase the IMM and OMM permeability while losing chemical gradients and polarity within the mitochondrion[659]. The Cer interaction, in particular, an increase of Cer18[iii] with dynamin-related protein 1 (Drp1), regulates mitophagy, a catabolic process that allows mitochondrial degradation in the lysosome. The study suggests that Drp1 knockdown can prevent mitophagy and change the Cer distribution in the OMM. Furthermore, the increase of Cer18 and Cer16 affects mitochondrial fission[248]. Cer also functions as the backbone of more complex forms of SL, such as sphingosine-1-phosphate (S1P). S1P is synthesised by the interaction between sphingosine and sphingosine kinase (SphK). The enzyme SphK can be found in two isoforms, namely SphK1 and SphK2. SphK2 in IMM, shown by recent studies, suggests that mitochondria can produce S1P[248]. The interaction of S1P and SphK2 also appears to be involved in assembling the electron transport chain[782]. The binding of S1P with S1PR2 and S1PR3 receptors triggers mitochondrial fission through a signalling network that activates the Ras homolog family member A (RhoA) protein and Rho-associated protein kinase (ROCK), which leads to Drp1 phosphorylation[248].

GP lipids are synthesised mainly in the ER and mitochondria-associated membranes (MAM)[237,844,847]. Three exceptions are PG, CL and PE, the only glycerophospholipids that are synthesised in mitochondria[847,851]. PG and CL share a part of the biosynthesis pathway, as PG are converted as a substrate to CL by cardiolipin synthase. PG biosynthesis involves two enzymatic processes. The first reaction converts phosphatidic acids, imported from the MAM, a membrane similar to the ER, to cytidine diphosphate diacylglycerol (CDP-DAG) by the TAMM41 enzyme. The second enzymatic reaction converts the CDP-DAG to PG by interacting with the PTPMT1 enzyme[851]. While PG and CL are produced only within mitochondria, PE synthesis can happen via two routes. Within the mitochondrial intermembrane space, the interaction between PS, imported from the MAM, and phosphatidylserine decarboxylase 1 (Psd1) converts PS into PE. Alternatively, PE can be synthesised through the CDP-ethanolamine pathway, which takes place in the ER and involves ethanolamine kinase and DG[851]. The study using the mouse model shows that a reduction of PE can perturb mitochondrial morphology, impact cell growth and oxygen consumption, and decrease the production of ATP in the oxidative phosphorylation pathway[806,847]. Generally, PC synthesis starts with converting choline into phosphocholine by choline kinase. Phosphocholine is then converted into CDP-choline, catalysed by CCT. CDP-choline and DAG are reaction substrates of CPT, and, as a result, PC are produced within the ER. The abundance of PS is partially related to the presence of PE and PC in the MAM. PE can be converted into PS

[iii] Cer with 18 carbon atoms on the fatty acid tail.

by interacting with phosphatidylserine synthase 2 (PSS2); similarly, PC can be transformed into PS by PSS1[582,851]. New evidence suggests that the PS shuttling occurs at the contact sites between the MAM and mitochondria[756]. Furthermore, the ER is also responsible for PI biosynthesis, which occurs in two steps; CDP1 and CDP2 convert phosphatidic acid into CDP-DAG, and the phosphatidylinositol synthase converts it into PI[851].

Lipids belonging to GP are transported back and forth to the ER through the MAM, which is rich in enzymes for phospholipid biosynthesis[847]. While the exact transport mechanism is still a matter of research for mammalian cells, it seems to vary by type of GP[806,851]. For instance, the transport of the PS between ER and mitochondria is modulated by three proteins associated with the MAM, namely oxysterol-binding protein-related proteins 5 (ORP5), 8 (ORP8), and VAMP-associated protein B and one protein embedded in the OMM, namely PTPIP5[255,582,920]. ORP5 and ORP8 work synergetically with mitofusin 2, which creates a PS-rich rigid domain that facilitates PS transport. Additionally, ORP5 and ORP8 are also responsible for PS transfer between the ER and the plasma membrane[255,582]. Similarly, PC and PE exhibit similar transport mechanisms as PS, although PC- and PE-rich domains are created by mitofusin 1, the homolog of the GTPase mitofusin 2. Both mitofusin homologs are embedded in the OMM[851]. The intermembrane space of mitochondria provides small regions, namely, contact sites, where two mitochondrial membranes are close to each other. Contact sites are claimed to be enriched with negatively charged GP, such as CL, and distribute lipids between the OMM and IMM. Experimentalists have detected lipid transfer activities at contact sites, which may happen spontaneously and/or through PRELI1, PRELI2, SLMO1 and SLMO2 proteins, a set of conserved protein families that exhibit evidence of lipid transfer activities. In particular, PRELI1-TRIAP1 complexes maintain CL and PE levels by regulating the flux of phosphatidic acids between the OMM and IMM. Additionally, protein PLSCR3 appears responsible for the CL transfer from the IMM to the OMM[658,806,851]. On the other hand, it is still unknown what types of lipid transport machinery allow the passage of phosphatidic acids from the MAM to the mitochondria.[851]

GL are often associated with FA, as GL can be converted to free fatty acids (FFA) via the GL/FFA cycle[iv], which contains lipolysis and is responsible for several signalling pathways, energy homeostasis and thermogenesis [v]. TG, DG, cholesterol and some phospholipids can be packed together in the cytosol by a layer of amphiphilic lipids, giving rise to lipid droplets. The transport and trafficking of lipid droplets appear very complex and not fully understood. However, it has been identified that lipid droplets contain GL, signalling lipids involved in FA oxidation and other critical biological processes. Lipid droplets are also the primary intracellular source of GL for the GL/FFA cycle; through the cycle, FA are esterified and GL are synthesised[660]. The GL/FFA cycle can be divided into two parts depending on whether the cycle performs anabolic and catabolic reactions. Anabolic reactions ensure energy storage by employing glycerol-3-phosphate and fatty acyl-coenzyme A as substrates to synthesise TG. Glycerol-3-phosphate is obtained through glycolysis, glycogenesis or from outside the cell. Some intermediates of the anabolic phase include lysophosphatidic acid, phosphatidic acid and DG. The subset of anabolic reactions, from glycerol-3-phosphate to

[iv] The GL/FFA cycle is often called the futile cycle due to its ATP consumption in return for heat. The same mechanism between TG and FFA is known as the TG/FFA cycle.

[v] Thermogenesis is the process that allows the organism to produce and maintain a specific temperature.

TG synthesis, is also known as the Kennedy pathway[918]. On the other hand, the catabolic phase (lipolysis) hydrolyses TG and generates energy for the cell. During this phase, glycerol is also generated; glycerol can be used during the anabolic phase and converted into glycerol-3-phosphate by glycerol kinase. DG and TG are intermediates of the chemical chain reactions of the GL/FFA cycle. Within this cycle, GL are synthesised mainly on the outer mitochondrial membrane and ER. During the catabolic phase, almost 40% of produced FFA is re-esterified into TG. This recycling ratio appears constant across different cell lines and metabolic conditions[660]. In cardiomyocytes, while TG represent almost 70% of energy source, they are mainly exogenous. Although TG are imported into the cellular environment, the efficient GL/FFA cycle facilitates the rapid oxidation of TG and the formation of FFA. This enables energy supply to the cell and a timely FFA re-esterification into GL for later consumption[424,660]. The study shows that the activation of adipose triglyceride lipase (ATGL), an enzyme involved in cytosolic lipolysis, could be pharmacologically utilised to cope with heart failure[424]. Additionally, aberrant TG accumulation is often associated with obesity and insulin resistance[918].

Besides the GL/FFA pathway, moreover, TG can be synthesised through the monoacylglycerol pathway. While the GL/FFA cycle is present in most cell types, the monoacylglycerol pathway tends to be restricted to more specific types, such as adipocytes and enterocytes[918]. Regardless of the cell line, both pathways share a common anabolic reaction that sees two diacylglycerol O-acyltransferases (DGAT1 and DGAT2) as enzymes and DG and fatty-acyl-CoA as substrates for TG production. The ER is enriched in both enzymes, indicating that most TG synthesis occurs there. Biochemically, DGAT1 and DGAT2 exhibit different affinities to substrates. DGAT2 seems more sensitive to the lower substrate concentration, suggesting that it has a higher Michaelis constant (K_m) than DGAT1. Further experiments on rat hepatoma models have revealed that overexpression of DGAT2 considerably increases the amount of TG accumulation in a cell[918]. TG accumulation in cardiomyocytes is often observed with ageing and diabetes. Research on mouse models has shown that a co-inhibition of DGAT1 and DGAT2 can prevent the aberrant TG accumulation without compromising cardiac functions and lead to the decreased activity of peroxisome proliferator-activated receptor α (PPARα). Additionally, it has also been pointed out that DGAT1 inhibition does not affect the TG turnover, contrary to DGAT2 inhibition[689].

AcCA, (L-)carnitine esterified to fatty acids, can be found across lipids associated with mitochondrial dynamics. They are responsible for transporting long-chain fatty acids into the mitochondrial matrix, where they are involved in the fatty acid β-oxidation, Krebs cycle, and branched-chain amino acids metabolism[479,551,592]. Long-chain fatty acids usually have 13 to 18 carbon atoms, regardless of their unsaturation. Fatty acids with more than 18 carbon atoms are known as very long-chain fatty acids. Medium-chain fatty acids contain between 6 and 12 carbon atoms, while short-chain fatty acids have less than 6 carbon atoms. AcCA synthesis occurs in two subcellular locations: peroxisome and mitochondria. Regarding mitochondria, AcCA synthesis is triggered by importing exogenous carnitine, a polar molecule made of lysine and methionine, into the cytosol by organic cation-carnitine transporter 2 (OCTN2). Carnitine and acyl-CoA interact with carnitine palmitoyltransferase 1 (CPT1), an enzyme integrated into the OMM, followed by AcCA production in the cytosol. The newly synthesised AcCA can also cross the OMM back via unspecific voltage-dependent

channels and flow in the cytosol. In exchange for carnitines, AcCA in the cytosol is imported into the mitochondrial matrix via carnitine-acylcarnitine translocase (CACT or SLC25A20). Interestingly, the SLC16 family of proteins also appears responsible for efflux of AcCA[553].

Once in the mitochondria, AcCA, along with substrate coenzyme A, is reconverted into acyl-CoA and carnitine by CTP2 in exchange for long-chain FA-CoA [vi] into the matrix. The described process of carnitine transport in and out of the mitochondria is called the carnitine shuttle. The free carnitine and coenzyme A in the mitochondrial matrix are successively employed in the β-oxidation process and Krebs cycle[551,821]. After several chemical reactions within the mitochondria, AcCA usually displays shorter fatty acid tails, which can be transported back into the cytosol by CACT or employed in the branched-chain amino acid metabolism[479]. Moreover, the study has shown that CPT2 can drive the reverse process, converting long-chain FA-CoA into long-chain AcCA when they cannot be efficiently metabolised[346,553]; as a consequence, the accumulation of long-chain AcCA can be observed in mitochondria and cytosol. The accumulation of long-chain AcCA can also occur due to the reduced CTP2 activity. Injection of long-chain AcCA (16 carbons) also activates the L-type Ca^{2+} channels, as observed in pig models[11]. The introduction of carnitine (16 carbons) in rat cardiomyocytes has suggested an increase in the uptake of mitochondrial calcium in cardiomyocytes[553]. Studies have shown that the length of fatty acid chains in AcCA is associated with metabolic perturbations in various diseases[551,553]. Failure to transport AcCA in the mitochondrial matrix profoundly impairs the β-oxidation pathway[821], and the ratio of acylcarnitine to carnitine concentration in blood and/or plasma can be used as potential biomarkers to detect perturbations in mitochondrial metabolism[479,551]. Further studies have revealed that the plasma level of acylcarnitine in healthy patients increases with age[380]. Additionally, significant changes in AcCA metabolism have been observed in heart failure[479]. The dysregulated fatty acid oxidation can be detected through the aberrant accumulation of long-chain fatty acids in a tissue or plasma, and long-chain AcCA can be employed as biomarkers[553].

2.3.4 Lipid droplets

Cardiomyocytes are one of the most energy-requiring cells, and human cardiomyocytes yield most of the energy from fatty acid oxidation while relying less on other energy-generating pathways. Mitochondria, cell power plants with a fundamental role in ageing and diseases, guarantee that cardiomyocytes never run out of energy[32,282]. Although fatty acid oxidation represents the primary energy source in cardiomyocytes, its excessive reliance on lipid oxidation could harm cells. Adult cardiomyocytes are characterised by their prompt rewiring of the metabolic network when changes in the energy homeostasis occur, matching their energy demands through pathways based on different substrates, such as glycolysis and Krebs and Randle cycles[711]. Metabolic rewiring is particularly explicit in myocardial ischaemia, where cardiomyocytes switch from the fatty acid- to glucose-based metabolic pathways to generate ATP, a highly energetic compound[610]. Metabolic rewiring

[vi] CTP2 and CACT are tightly associated, as both are part of an integral supercomplex of the IMM.

does not come to no harm since its over-reliance on glucose substrate can often lead to the accumulation of long-chain fatty acids in cardiac cells and, subsequently, an aberrant increase of lipids circulating in a cellular environment (lipotoxicity). Although accumulation of long-chain fatty acids can be observed through different classes of lipids, in particular TG, long-chain fatty acids on Cer and DG, which are two essential signalling lipids, can be far more threatening for the cellular environment[282]. While most lipids utilised by the cell are exogenous, lipid synthesis maintains cardiac functions in ageing[282], and a change in lipid composition can also affect cellular energy production by impairing mitochondrial functions and activating signalling networks like the unfolded protein response[32,610]. To mitigate the risk of lipotoxicity, the cell packs exceeding lipids into small lipid droplets, bona fide organelles rich in TG, or proceeds with the β-oxidation of fatty acids[32].

Lipid droplets are dynamic cytosolic organelles. Their size varies based on the cell's state and its needs for energy, and their biogenesis has not been fully characterised. In addition to precluding cellular lipotoxicity under different physiological states such as starvation, lipid droplets can also work as buffers to prevent mitochondrial intoxication from abnormally trafficking AcCA[610]. Researchers have identified the ER as the sole responsible organelle for building lipid droplets[924]. Lipid droplet biogenesis starts with the accumulation of neutral lipids, TG and ST, within the two lipid layers of the ER membrane, giving rise to a lens, an early stage droplet. No protein or enzyme appears to be involved at this stage, and it is unclear if the lens formation preferentially occurs in specific locations of the ER membrane. Lipid droplet formation from the lens is affected by the tension of the outermost layer of the ER membrane, which minimises the energy dissipation derived from the exposure of neutral lipids of the lenses to the aqueous cytosolic environment. The ER membrane phospholipids regulate the geometrical shape and surface tension of budding droplets. In vivo experiments on yeast cells have also suggested that DG and PI discourage droplet formation while LBPA prompts it[589]. Recent studies have also shown that the expression of fat storage-inducing transmembrane protein family (FIT), particularly FIT1 and FIT2, induces lipid droplet formation in mammalians. The absence of FIT2 in mammalian cells leads to the aberrant lens accumulation on the ER membrane and that of DG within the ER, suggesting that FIT2 actively contributes to droplet formation. In vitro experiments support that FIT2 can directly bind DG and TG. Ablation of FIT2 also leads to severe deregulation of metabolic functions[163,610]. Droplet formation also appears to be affected by seipin, a conserved protein in the ER membrane. The absence of seipin seems to decelerate the development of droplets, whilst mature droplets present different morphology and species compositions[610].

The perilipin (Plin) family is also associated with forming lipid droplets, especially in mammals, but it is yet unclear how Plin contributes. Studies have established that Plin3 is recruited from the cytosol at the very early stage of droplet formation, and its absence causes a delay in droplet development. Additionally, the activity of Plin3 may be related to the presence of TG in the lens, as Plin3 does not contribute to the droplet development when ST compose most of the lens[32,610]. At the protein level, lipid droplets of cardiomyocytes differ significantly from those of adipocytes, as the former presents a higher abundance of Plin2, Plin3, Plin4 and Plin5 proteins but almost none of Plin1 when compared to lipid droplets in adipocytes[282]. TG and ST are introduced in lenses through specialised bridges

or on-site lipid synthesis. As previously mentioned, fatty acids in lipid droplets are not stored in their free state, yet the triple FFA are esterified to a glycerol-3-phosphate molecule, giving rise to TG, which would be packed into the lens. Bridges also ensure the free diffusion of proteins typically associated with the ER. Other proteins in lipid droplets are recruited directly from the cytosol. After a lipid droplet reaches the mature state, it detaches from the ER membrane. It is still unclear as to how the detachment occurs and why only some droplets separate from the ER membrane[610].

The lipid droplet membrane is a monolayer derived from the ER membrane and mainly comprises phospholipids and Plin. It is suggested that lipid droplets and mitochondria have close interactions, as the two tend to be close to each other. The interactions between mitochondria and lipid droplets appear to be dynamic and change as the organism's physiological state changes; for instance, fasting and physical exercises could display very different interactions[32]. The spatial proximity between mitochondria and lipid droplets avoids FFA buffer in the remote part of the cell, preventing lipotoxicity and erroneous activation of signalling pathways. Communication between a mitochondrion and lipid droplets occurs via contact sites, although it is not fully understood how two organelles are tethered. In the heart cells, overexpression of Plin5 on the droplet membrane seems to promote the recruitment of mitochondria around the droplets, suggesting its potential involvement in tethering the droplet to the mitochondrion. Plin5 also promotes ATGL activity both directly and indirectly through its activator abhydrolase domain containing 5 (ABHD5). In brown adipocytes, the tethering between mitochondria and lipid droplets appears to be mediated by mitofusin 2, and Plin1[850]. This evidence suggests that docking between the two organelles might involve mitofusin and the Plin family, although mechanisms behind the formation of contact sites between the two organelles remain unclear[850]. Further studies on brown adipocytes have highlighted an increase of ATP and a decrease of β-oxidation in mitochondria close to lipid droplets, indicating a potential increase in TG synthesis, an anabolic set of reactions that requires energy[61]. Furthermore, it has also been suggested that mitochondria utilise lipid droplets as a warehouse of TG lipids, withdrawing TG for lipolysis and depositing them after lipogenesis[610].

2.3.5 β-oxidation, GL/FFA cycle, and lipotoxicity

Lipid droplets are tightly related to energy production, as fatty acid oxidation is crucial in the human heart. Cardiomyocytes use fatty acid oxidation (see Figure 2.5) to generate up to 90% of the energy needs. In contrast, the fetal heart yields energy mainly from glucose metabolism to minimise oxygen consumption. Glucose is also preferred as a substrate for energy production when the heart exhibits left ventricle hypertrophy, hypertension or myocardial ischaemia[631]. The catabolic series of reactions that oxidise fatty acids, highly energetic compounds with a long carbon backbone, and yield energy is known as β-oxidation[409,591]. Although enzymes involved in fatty acid oxidation are also present in peroxisomes, β-oxidation occurs mostly in mitochondria. As mentioned earlier, lipid oxidation requires the cell to import long-chain fatty acids into mitochondria. The uptake of long-chain fatty acids from the extracellular matrix is facilitated by several protein families

embedded in the cellular membrane, which are fatty acid translocases (FAT/CD36), fatty acid-binding proteins (FABP), and fatty acid transport proteins (FATP). In cardiomyocytes, FABP3 is the predominant homolog of FABP. In most human tissues, FATPs are found in all six homologous, namely FATP1-6 (or SCL27A1-6). In the human heart, FATP1 (SCL27A1) and FATP6 (SCL27A6) are overexpressed in the sarcolemma[345,551,721].

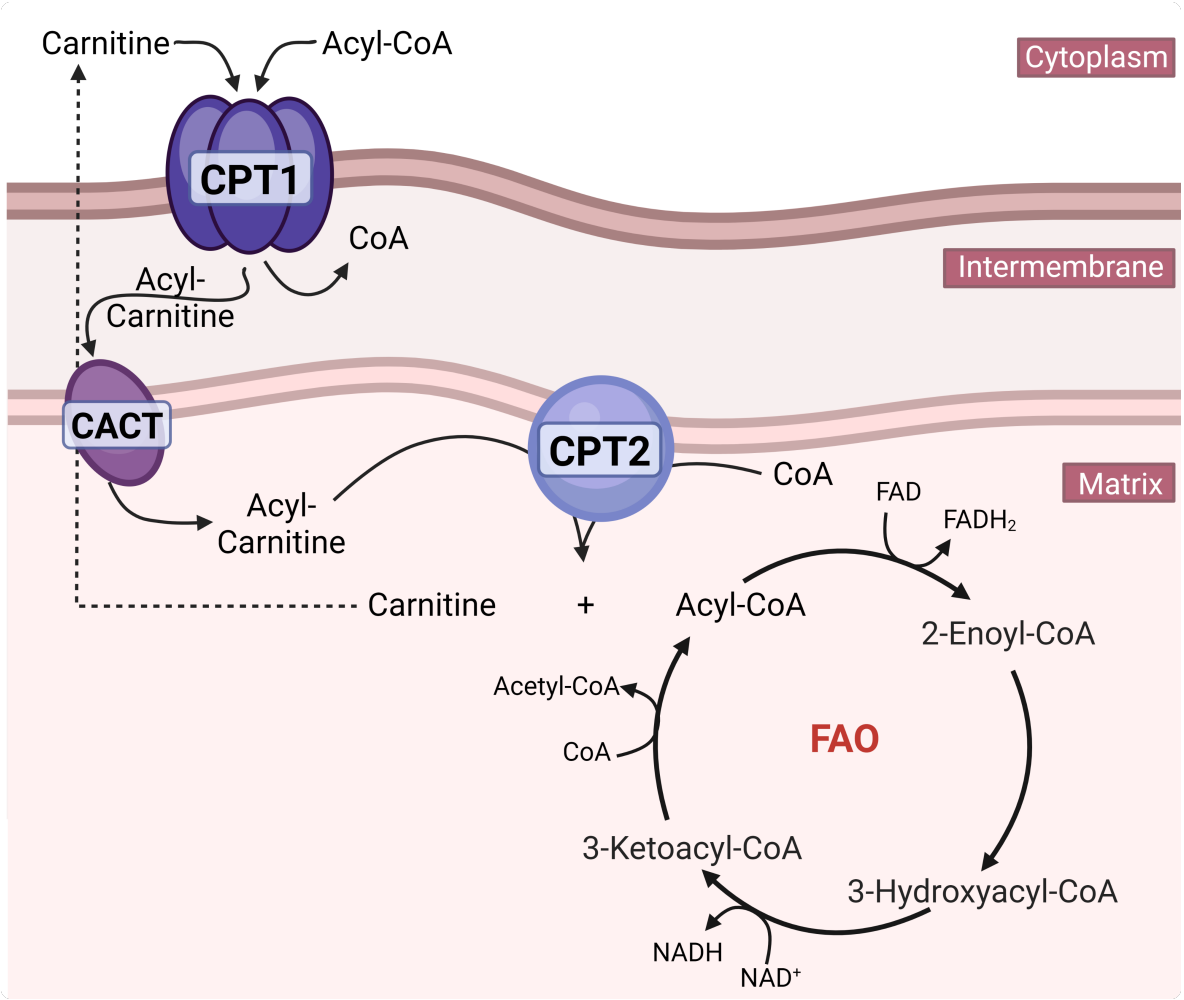


Figure 2.5: Summary of fatty acid transport in the mitochondria and their oxidation (FAO or β-oxidation). The figure highlights the role of the acylcarnitine-carnitine shuttle that allows the transportation of the fatty acids in the mitochondrion membrane. The β-oxidation, or FAO, yields $FADH_2$ and NADH. (Created with BioRender.com).

Interestingly, the FATP family exhibits evidence of acyl-CoA synthetase (ACSL), suggesting that the conversion of fatty acids to acyl-CoA occurs right after the cellular uptake of fatty acids. The cytosol also contains ACSL; ACSL is a family of proteins and takes fatty acid, CoA and ATP as substrates and returns acyl-CoA, pyrophosphate and adenosine monophosphate (AMP). As a result of ACSL and FATP activities, acyl-CoA compounds in the cytosol inherit the high energy content of exogenous fatty acids. Since the OMM is impermeable to the acyl-CoA, the carnitine shuttle machinery moves high-energy compounds that

comprise acyl-CoA into the mitochondrial matrix[551,721]. The carnitine shuttle requires carnitine as substrate. The acyl-CoAs and carnitines react with carnitine palmitoyltransferases (CPT), a family of proteins embedded in OMM, and AcCA are synthesised in the cytosol in return. In cardiomyocytes, CPT1B is the highest overexpressed isoform of CPT, while CPT1A is also present; both isoforms are inhibited by malonyl-CoA. Subsequently, AcCA is introduced in the mitochondrial matrix in exchange for carnitine from the mitochondrial matrix. CACT intermediates the exchange, which can happen in both directions. CPT2 in the IMM converts AcCA into acyl-CoA and carnitine. Carnitine O-acetyltransferase (CRAT) operates within the mitochondrial matrix by converting acetyl-CoA into AcCA and CoA. Acetyl-CoA in the mitochondrial matrix is also regulated by glycolysis. As soon as acyl-CoA is available in the mitochondrial matrix, β-oxidation can occur[345,346].

In general, β-oxidation is a cycle that starts with a long-chain acyl-CoA and ends with acyl-CoA short of two carboxyl groups[vii], which are released in the form of acetyl-CoA. In detail, β-oxidation begins by converting acyl-CoA, a compound with long-chain fatty acids, to enoyl-CoA by very long-chain acyl-CoA dehydrogenase (VLCAD, encoded by ACADVL). Enoyl-CoA is then used to generate 3-hydroxyacyl-CoA and 3-ketoacyl-CoA by interacting with mitochondrial trifunctional protein (MTP, encoded by HADHA and HADAHB). Then, the interaction between 3-ketoacyl-CoA and MTP yields acetyl-CoA, NADH, $FADH_2$, and acyl-CoA shorter of two carboxyl-terminals. β-oxidation occurs again if acyl-CoA still presents long-chain fatty acids. When acyl-CoA contains less than 14 carbon fatty acid, β-oxidation happens through a slightly different route; β-oxidation converts the medium-chain acyl-CoA into enoyl-CoA through the enzymatic activity of medium-chain acyl-CoA dehydrogenase (MCAD, encoded by ACADM). Then, 3-hydroxyacyl-CoA is metabolised by crotonase (ECHS1) using enoyl-CoA as substrate. Then, 3-ketoacyl-CoA is synthesised by medium-chain hydroxyacyl-CoA dehydrogenase (MCHAD) by consuming 3-hydroxyacyl-CoA. Finally, medium-chain 3-ketoacyl-CoA thiolase (MCKAT, encoded by ACAA2) converts 3-ketoacyl-CoA into acyl-CoA, acetyl-CoA, $FADH_2$ and NADH. When acyl-CoA contains a fatty acid with a chain with less than 6 carbons, short-chain acyl-CoA dehydrogenase (SCAD, encoded by ACADS) substitutes MCAD for converting acyl-CoA into enoyl-CoA. Similarly, short-chain hydroxyacyl-CoA dehydrogenase (SCHAD, encoded by HADH) substitutes MCHAD for producing 3-ketoacyl-CoA from 3-hydroxyacyl-CoA[345]. The oxidation processes described above assume that fatty acids are saturated. When a fatty acid is monounsaturated, 3,2-trans-enoyl-CoA isomerase (DCI) converts unsaturated 3-enoyl-CoA into 2-enoyl-CoA. The role of DCI is to shift the position of the double bond in the fatty acid. When acyl-CoA contains polyunsaturated fatty acids, an additional enzyme, 2,4 Dienoyl-CoA reductase (DECR1), is required to convert the compound into monounsaturated 3-enoyl-CoA[345].

Hormones and metabolites maintain a certain level of energy production and consequently regulate β-oxidation. In particular, adrenaline, norepinephrine, glucagon and adrenocorticotropin stimulate lipolysis, while insulin, a compound secreted by pancreatic β-cells, inhibits lipolysis by increasing glucose in the cytosol. FFAs are also signalling

[vii]The carboxyl group is defined as COOH-, where C stands for carbon atom, O for oxygen, H for hydrogen, and "-" for the remaining compound.

compounds; FFAs are ligands for the peroxisome proliferator-activated receptors (PPAR), such as PPARα and PPARβ. These two receptors are responsible for stimulating fatty acid oxidation. On the other hand, as previously mentioned, malonyl-CoA inhibits CPT1 and, consequently, β-oxidation. The concentration of malonyl-CoA depends on the ATP level. As ATP increases, a cascade of signals that include AMP, AMP-activated protein kinase (AMPK) and acetyl-CoA carboxylase increases the amount of malonyl-CoA, resulting in the inhibition of the CPT1 activity and the shutdown of β-oxidation[345]. AcCA and carnitines also have an active role in regulating fatty acid oxidation. Due to their close relationship with acetyl-CoA, AcCA and carnitines are associated with the activity of pyruvate dehydrogenase (PDH), a mitochondrial enzyme that converts pyruvate into acetyl-CoA. The presence of carnitine in the mitochondrial matrix stimulates the PDH activity in human muscles. At the same time, overexpression of acetyl-CoA seems to inhibit the PDH activity and consequently deregulates cellular respiration. Furthermore, a high ratio of NAD/NADH and acetyl-CoA/CoA appears to activate pyruvate dehydrogenase kinase (PDK), and, as a result, inhibit PDH[345,553]. Additionally, using canine and rabbit ventricular cardiomyocyte models, researchers have also revealed that, by introducing exogenous long-chain AcCA (16 carbons) and carnitine, ATPase activities involved in NA^+, K^+ and Ca^{2+} transport in the cytosol and sarcolemma reticulum are increased. In other words, the increase in carnitine and long-chain AcCA increases Ca^{2+} concentration, although the exact mechanism remains unclear[553]. Deregulated fatty acid oxidation is also associated with several cardiomyopathies, such as dilated and hypertrophic cardiomyopathy[345,631].

GL and the GL/FFA cycle contribute to energy production through lipolysis, increasing FFA availability. Since GL and FFA are closely related, malonyl-CoA is also central to GL/FFA regulation. The biosynthesis of malonyl-CoA [viii] is controlled by AMPK, which is activated in response to the high AMP/ATP ratio in the cellular membrane. The AMPK activity reduces the expression of glycerol-3-Phosphate acyltransferase (GPAT) and acetyl-CoA carboxylase (ACC). As a result, malonyl-CoA decreases while fatty acid oxidation is promoted, which leads to depletion of TG in the cell. The higher presence of glucose inhibits lipolysis and leads to TG accumulation, particularly in adipocytes[660]. The increase of TG in the left ventricle can also lead to a structural and functional adjustment of the heart, with potential impairment of myocardial contractility, causing ventricular and atrial hypertrophy and dilated cardiomyopathy[631].

Long-chain fatty acids can cause lipotoxicity, and elongation of fatty acids is generally reflected in SP. As a building block of SP, Cer contains fatty acids with lengths between 16 and 24 carbons, depending on the tissue[421]. The presence of Cer with long-chain fatty acids in OMM can cause mitochondrial dysfunction by compromising the membrane fluidity, complex formation and remodelling of cristae, and potential impact on the electron transport chain. Cer can also promote the activity of protein phosphatase 2 (PP2A), which is responsible for the phosphorylation of B-cell lymphoma 2 (BCL2) proteins that can pour cytochrome c into the cytosol. The increased amount of lipids can also induce stress in the ER. When the stress persists, the ER will increase the transport of Ca^{2+} to mitochondria,

[viii] Malonyl-CoA biosynthesis is regulated by ATP citrate lyase (ACLY), acetyl-CoA carboxylase (ACC), di- and tri-carboxylic acid transporters, and fatty acid synthase (FAS). The degradation of malonyl-CoA depends mainly on the activity of malonyl-CoA decarboxylase.

causing the opening of pores in its membrane, loss of microenvironment properties and leakage of chemicals into the cytosol[659]. The aberrant amount of long-chain fatty acids can indicate a dysfunction of fatty acid oxidation in mitochondria and peroxisomes, which, in cardiac tissue, can lead to cardiomyopathy[346,429]. Accumulation of SP is also associated with impaired glucose homeostasis and insulin sensitivity[459]. Recent evidence has shown that a higher presence of SP, particularly Cer with 22 and 18-carbon long fatty acids, is associated with increased fasting glucose level, while in vitro experiments have shown that Cer inhibits GLUT4, with a consequent reduction in glucose uptake[385,459]. Meanwhile, changes in the length and saturation of fatty acid tails in GP can greatly affect the fluidity of mitochondrial membranes. Similarly, changes in phospholipids in the mitochondrial membrane, such as CL, can alter the permeability of the OMM and IMM. The reduction of CL can be observed during ageing, associated with decreased production of ATP and increased ROS species in cardiac tissues[631].

2.3.6 Glucose metabolism

Glycolysis is triggered by glucose in the cell and can be aerobic or anaerobic (see Figure 2.4), depending on oxygen availability[128]. Glucose homeostasis ensures the survival of cardiomyocytes, and cardiomyopathies such as ischaemia and hypertrophy highlight the significance of glucose for cardiac contractility[186]. Glucose is driven inside the cell by a chemical gradient, yet, due to the semi-permeability of the sarcolemma, it requires a transporter to enter the cytosol. The glucose transporter (GLUT) is a specific membrane protein that allows glucose trafficking between the extracellular matrix and cytosol. GLUT1 (SLC2A1) and GLUT4 (SLC2A4) are the two specific isoforms for cardiomyocytes, and their expression in adult cardiomyocytes is dictated during the developmental phase of the cell. The two isoforms have slightly different behaviours. GLUT1 is a transporter that controls basal glucose uptake, while GLUT4 activity is regulated by insulin[738]. Once in the cell, glucose undergoes glycolysis. Hexokinase (HK) catalyses the glycolytic process by phosphorylating glucose, yielding glucose 6-phosphate. HK1 and HK2 are predominant in the heart; HK1 activity is higher in fetal and newborn hearts, while HK2 is insulin-regulated[186]. At this stage, glucose 6-phosphate can be employed for glycogen synthesis or proceeds through glycolysis, where it is converted into fructose 6-phosphate by interacting with phosphoglucose isomerase. Subsequently, most of fructose 6-phosphate is phosphorylated into fructose-1,6-bisphosphate by interacting with the enzyme phosphofructokinase 1 (PFK1), whereas some fructose 6-phosphate fractions are phosphorylated by phosphofructokinase 2 (PFK2) to obtain fructose 2,6-bisphosphate (F2,6BP). The presence of F2,6BP and AMP in the cardiomyocyte promotes the PFK1 activity, while ATP, citrate and protons are allosteric inhibitors of the reaction[775]. By the mediation of aldolase, fructose-1,6-bisphosphate is cleaved into two molecules of glyceraldehyde 3-phosphate, each with three carbon atoms. glyceraldehyde 3-phosphate can be converted into its isomer dihydroxyacetone phosphate by triose phosphate isomerase. The enzyme glyceraldehyde 3-phosphate dehydrogenase (GAPDH) then converts Glycerol-3-phosphate into glyceraldehyde 1,3-bisphosphate, which is further dephosphorylated by phosphoglycerate kinase (PGK1), yielding 3-phosphoglycerate. The enzyme phosphoglycerate mutase 2 (PGAM2) mediates

the conversion of 3-phosphoglycerate into 2-phosphoglycerate, an isomer of the substrate. The interaction between 2-phosphoglycerate and enolase forms phosphoenolpyruvate, which is then dephosphorylated by pyruvate kinase into pyruvate, determining the end of the aerobic glycolytic pathway[744]. At the end of aerobic glycolysis, there is a net gain of two ATP and two NADH molecules, where the latter is the reduced form of an electron carrier nicotinamide adenine dinucleotide (NAD^+), a coenzyme of GAPDH[554,591]. Under limited oxygen availability, lactate dehydrogenase A (LDHA) can convert pyruvate into lactate, concluding anaerobic glycolysis. LDHA has been identified as a possible therapeutic target to facilitate cardiac repair after myocardial infarction; lactate production and LDHA activity promote cardiomyocyte proliferation while alleviating ROS synthesis[129].

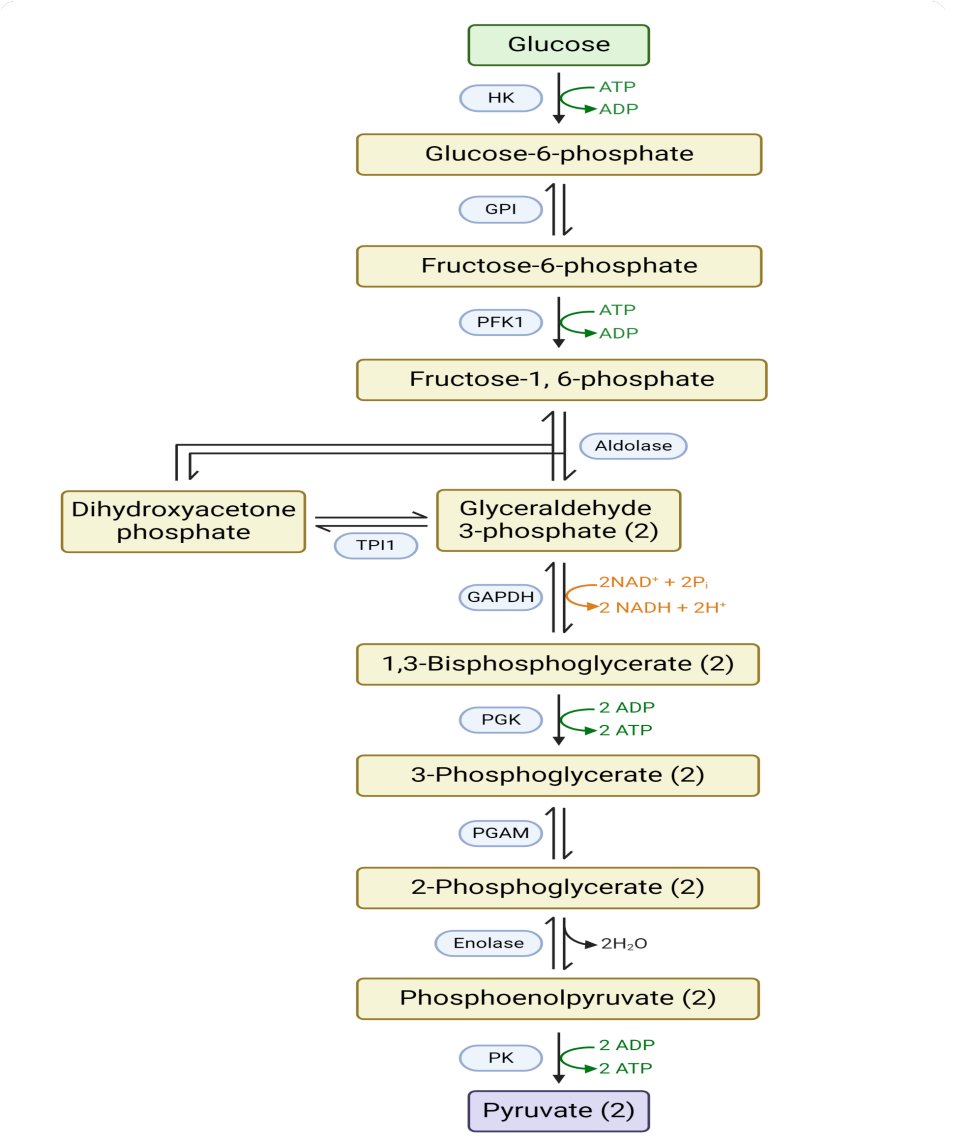


Figure 2.6: Schematic of glycolysis. The glycolysis breaks down the glucose molecule into two pyruvate molecules, generating ATP and NADH. (Created with BioRender.com).

2.3.7 Pyruvate trafficking and Acetyl-CoA synthesis

While still a matter of research, the voltage-dependent anion channel (VDAC) is understood to be responsible for transporting pyruvate across the OMM and releasing it in the mitochondria intermembrane space. The mitochondrial pyruvate carrier (MPC) carries pyruvate from the mitochondrial intermembrane space into the IMM. The composition of MPC is yet unknown. Bioinformatics and genetic methods have enabled the discovery of two complexes of MPC, namely MPC1 and MPC2. Pharmacological and biological experiments suggest that both complexes are sufficient and necessary for pyruvate transport into the IMM, as the absence of even one complex heavily affects its transport from the intermediate space to the inner matrix[554]. Next, pyruvate dehydrogenase (PDH) strips pyruvate from its carboxyl group (-$COOH^-$). The resulting decarboxylated molecule is oxidised and bound to coenzyme A (CoA), giving rise to acetyl-CoA (A-CoA), an essential molecule for central carbon metabolism[894]. The PDH expression is activated by PDH-phosphatase and inhibited by pyruvate dehydrogenase kinase (PDH_k). The production of PDH-phosphatase is stimulated by positive changes of $Ca^{2}+$ in the mitochondrion. The enzyme PDH is inhibited by pyruvate and a decrease in A-Coa/CoA and NADH/NAD^+ ratios. The oxidation of carbon atoms in A-CoA occurs in the Krebs cycle[554,591,894].

2.3.8 Krebs cycle

Regardless of the pathway that leads to A-CoA synthesis, the Krebs cycle (see Figure 2.7), also known as the tricarboxylic acid (TCA) cycle, metabolises A-CoA to produce high-energy molecules as well as electron carriers that will be utilised in the later stage of the metabolism[514]. The Krebs cycle begins when the reaction between A-CoA and oxaloacetate returns a molecule of citrate (6 carbons), which is then converted into isocitrate (the isomer of citrate). Isocitrate is then decarboxylated while NAD^+ is oxidised. As a result, NADH and α-ketoglutarate (5 carbons) are attained. Consequently, by losing carbon atoms, the molecule α-ketoglutarate becomes succinyl-CoA that is further transformed into succinate (4 carbons), whereas guanosine triphosphate (GTP) is released. The electron carrier $FADH_2$ is then produced when succinate is converted into fumarate (4 carbons) that is further modified into malate. The additional NADH release occurs when malate interfaces with H_2O and leads to oxaloacetate production that concludes the Krebs cycle. As mentioned above, the TCA cycle aims to generate electron carriers like NADH, $FADH_2$, and a high energy particle like GTP, which is successively employed for ATP production. The TCA cycle relies on oxygen (O_2) availability and oxidising agents NAD^+ and FAD^+ in the mitochondrion, and its disruption is associated with several cardiac diseases, such as ischaemia[171,324,409,930].

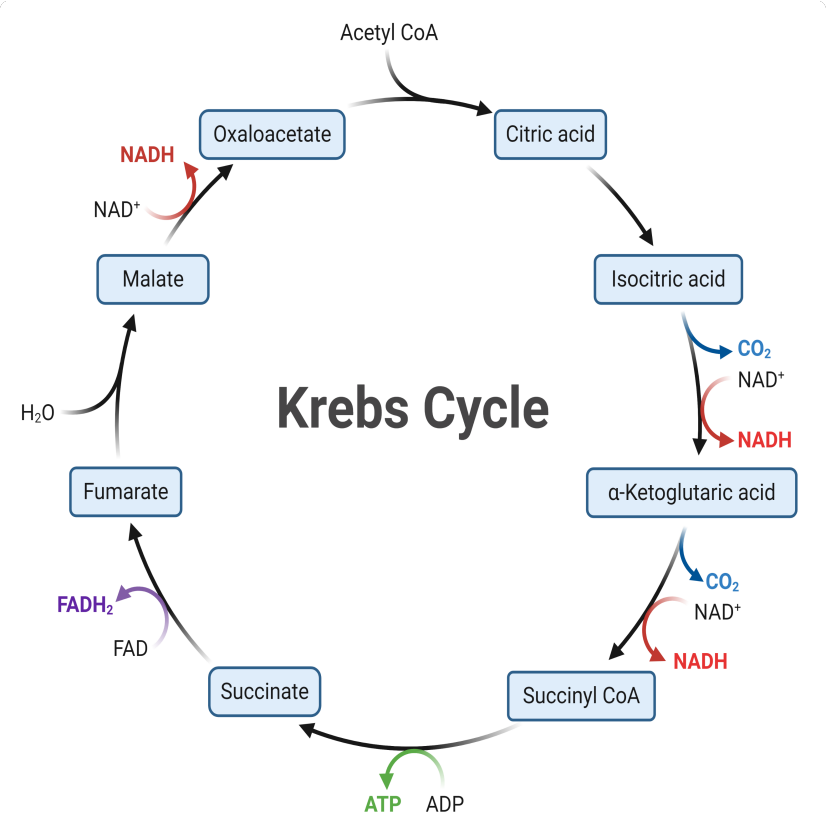


Figure 2.7: Overview of the Krebs Cycle. Starting from Acetyl-CoA, the Krebs generates ATP, $FADH_2$ and NADH molecules. (Created with BioRender.com).

2.3.9 Oxidative phosphorylation

The electron carriers retain most energy harvested by glucose or fatty acids. The main role of oxidative phosphorylation is to extract the energy for ATP synthesis by oxidising electron carriers. As a first step, NADH is oxidised through interaction with flavoprotein (FMN), a protein that comprises a part of complex I of the electron transport chain. Next, the electron moves towards iron-sulfur (FeS) proteins while $FADH_2$ is oxidised by complex II. Complex I donates the electron to ubiquinone (coenzyme Q), as does complex II. The electron collected by coenzyme Q is transferred to complex III and subsequently to complex IV by a sequence of cytochromes, a class of electron-carrying proteins. Similarly, the electron reaches ATP synthase, which produces ATP and H_2O in the cytosol[883]. The electron transport chain complexes comprise several proteins with different domains, and their formation and function remain unclear[663,753]. A visual summary of the oxidative phosphorylation is presented in Figure 2.8

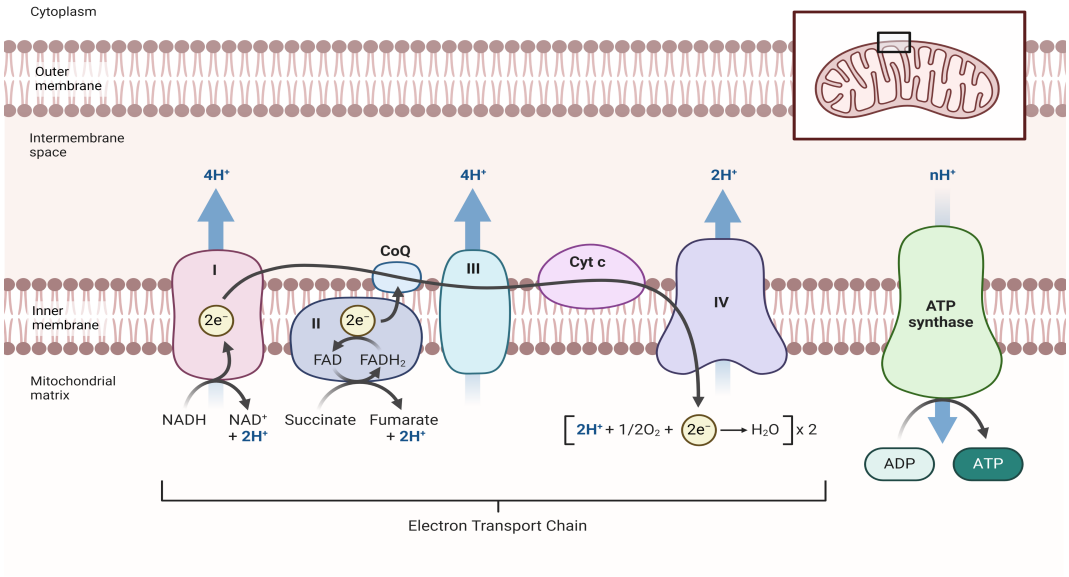


Figure 2.8: Summary of key molecular reactions of the oxidative phosphorylation from NADH. The interaction NADH-O_2 and $FADH_2$- O_2 yield electrons and energy availability, which will drive the synthesis of ATP from ADP. (Created with BioRender.com).

2.3.10 Phosphocreatine energy system

The role of the creatine-phosphocreatine energy system is determinant for systolic and diastolic functions of the heart, and it offers an efficient re-synthesis of ATP in relation to oxidative phosphorylation[49]. In healthy hearts, the ATP fluxes from oxidative phosphorylation are estimated to be slower than the ATP fluxes generated through the creatine-phosphocreatine system[875]. A study on humans has shown that in mild heart failure individuals, the heart mechanical work decreases proportionally to the ATP fluxes generated by the phosphocreatine energy apparatus[251], possibly suggesting that supplementation of creatine could ameliorate the cardiac mechanical work[49,251]. The energy production through the phosphocreatine energy systems occurs when Phosphocreatine (PCr) reacts with ADP to form ATP and creatine (Cr). The reaction is catalysed by creatine kinase (CK) and can occur in the cytosol or mitochondria[182,301,409].

2.4 Signalling and metabolic changes in cardiac ageing

Cardiac ageing is a multidimensional phenomenon encompassing changes in the heart's cell composition, increased heart frailty to cardiovascular diseases and cardiac remodelling, and changes in biochemical networks[35,300,528]. Among several definitions of ageing, one can understand it as a decline in the efficiency of the cell in terms of energy production, with a consequently increasing risk of developing diseases. With ageing, cardiomyocytes show

changes in structures and biochemical reactions, possibly underlying reduced capability to respond to stress and damage. The age-dependent biochemical alterations are possibly due to a mismatch between the supply and demand of energy, prompting the aged cardiomyocytes to mimic the behaviour of failing myocytes[88]. Due to limited oxygen availability, failing cardiomyocytes rely more on aerobic glycolysis than fatty acids for energy production, a pattern that is also observed in heart failure[469,908]. This section focuses mainly on the changes occurring in the molecular landscape of cardiac ageing, particularly revising two facets of the age-associated changes occurring in the biochemical pathways: signalling and metabolic pathways.

2.4.1 Signalling pathways

Changes and synthesis of molecules can lead to a cascade of molecular reactions that can result in the overall alteration of functions and processes of cardiac tissues[88]. Harman (1972) have hypothesised that mitochondria dictate cellular ageing through the production and accumulation of ROS from central carbon metabolism and oxidative phosphorylation reaction networks[21,314,822]; this theory is usually known as the mitochondrial free radical theory of ageing[88]. ROS are molecules with unpaired electrons on the outermost orbit[886] and comprise an important class of molecules that are essential for maintaining cellular homeostasis of cardiomyocytes[26]. With ageing, cardiomyocytes lose their efficiency, and homeostasis imbalance affects activities in different receptors, which can trigger several cellular responses. The aberrant accumulation of ROS, usually associated with ageing and cellular senescence[822], can damage the mitochondrion[207] and cause mitochondrial swelling and rupture of its outer membrane[88]. For instance, the rupture of the mitochondrial membrane can promote the release of cytochrome c and the second mitochondria-derived activator of caspase (SMAC) in the cytosol. Cytochrome c and SMAC bring about apoptosis[176]. ROS are also linked to cellular necrosis; the upregulation of phosphoglycerate mutase 5 (PGAM5) induces higher ROS production and mitochondrial dysfunction, which consequently activates cellular necroptosis. Using rat models to study cardiac ischaemia, She et al. (2019) have found that the inhibition of PGAM5 reduces ROS, membrane potential and lactate dehydrogenase (LDH) and prevents the cell from activating the necrotic pathways. ROS synthesis is constant in mitochondria and is catalysed through several enzymes, such as NADPH oxidase. Since up to 90% of ROS are produced in mitochondria, the decline of mitochondrial functions prompted by ageing leads to the increased ROS generation[47] and the imbalance between antioxidants and ROS, which causes oxidative stress[207]. ROS and oxidative stress can activate several signalling pathways, leading to various complex responses in the cell, whose mechanisms require further investigation[88].

ROS synthesis and oxidative stress are central to the cardiac ageing process. Recent studies have identified that mice lacking p66shc, an adaptor protein encoded by the Shc locus, have lower ROS levels and improved cellular resistance to oxidative stress. p66shc is localised in the mitochondrial intermembrane space and maintains the reduced oxygen level high while oxidising cytochrome c[157]. PKA-β phosphorylates p66shc and prolyl isomerase Pin1, and, when accumulated, the phosphorylated p66shc can trigger the Ca^{2+} response

followed by cellular apoptosis. The deletion of both alleles in diabetic mice has proven to be effective in reducing left ventricular hypertrophy as well as apoptosis in cardiac myocytes of diabetic mice[678]. In the study conducted on ischemic and reperfusion injuries in mice, p66shc inhibition has shown its efficacy in reducing the damage with oxidative stress and decreasing ROS in cardiomyocytes[113].

Evidence indicates that ROS and oxidative stress trigger nuclear factor-kappa B (NF-kB), a regulator of the inflammatory response and the immune system[509]. As the cell undergoes oxidative stress, the NF-kB transcription complex, comprised of P65, P55 and IkB, is separated into the P65-P50 complex and IkB. The two dimers P65-P50 successively enter the nucleus and promote the activity of several genes involved in apoptosis, morphogenesis, inflammation, cell proliferation and differentiation, such as Fas, Bcl2, MCP-1, IL-18, ICAM-1 and VCAM-1[662]. Meanwhile, oxidative stress activates specificity protein-1 (Sp1), a transcription factor, leading to higher Sp1 expression. Sp1 is an important promoter of SERCA2A, a protein pivotal for excitation and contraction, and studies point to its mediation of the β-adrenergic pathway[432,538]. Sp1 overexpression is also a promising biomarker of cardiac hypertrophy[44]. Oxidative stress provokes the activation of the Sp1 transcription factor via its phosphorylation, which regulates several gene expressions, including p53, p73, and YAP. Sp1 activation has a repressive effect on COX-2 inhibitors, and consequently, COX-2 is not repressed and stimulates inflammatory reactions[104]. In cardiac cells, oxidative stress promotes Sp1 and Sp3 expressions. Sp1 and Sp3 interact with vascular endothelial growth factor (VEGF), promoting ROS synthesis[267]. Recent research shows that Sp1 inhibition has a silencing effect on poly(ADP-ribose) polymerase-1 (PARP1), an enzyme that contributes to the progression of several cardiovascular diseases. Sp1 inhibition can be a promising therapeutic target for mitigating the effects of myocardial ischemia[903]. Interestingly, in the ageing process, the Sp1 activity is expected to decrease while impacting different other proteins[478].

Hypoxia-inducible factor 1 (HIF-1) is involved in several biological processes, such as angiogenesis and metabolism and promotes angiogenesis by regulating VEGF[662]. HIF-1 is a transcription factor composed of two units, one stable, known as HIFβ, and one unstable, known as HIFα, which becomes stable by interacting with oxygen. HIF-1 can be present in several isomeres, such as HIF-2 and HIF-3. HIFα is lowly expressed in normal conditions, yet its abundance increases in oxygen constraints environments, as HIFα cannot be stabilised and the complex HIF-1 cannot bind to its promoter region with a limited oxygen availability[769]. Based on the free radical theory of ageing, ROS production increases as the cardiac tissue ages, eventually inducing damage from oxidative stress and impaired mitochondrial biogenesis. SIRT1 is part of the sirtuin family and is pivotal in several biological processes, including HIF-1 regulation[666]. It has also been pointed out that Sp1 regulates SIRT1 expression, as they both decrease during the ageing process[478]. The decrease of SIRT1 can negatively impact DNA repair machinery[938] and signalling pathways associated with ageing and longevity, such as AMPK, insulin/IGF-1, FOXO and mTOR. The decrease of SIRT1 also reduces HIFα stabilisation, while later enhancing HIFα expression. The increased HIFα level restrains mitochondrial biogenesis due to the Mxi1 gene product that constrains the interaction between c-Myc and mitochondrial transcription factor A

(TFAM)[919]. The role of SIRT1 and HIFα also extends to modulating energy production in cardiomyocytes. Indeed, HIFα expression inhibits activities of FATP, CPTI and MCAD and reduces oxygen and fatty acids demands for energy production while increasing activities of aerobic glycolysis and glucose demands[769,938].

The forkhead box (FOX) transcription factor holds significance in cardiac ageing. The forkhead box O (FOXO) family comprises six isoforms, denoted as FOXO1-6[103]. Dysregulation of FOXO is associated with age-related diseases as they can alter cellular processes, such as oxidative stress and metabolism[612]. With regard to metabolism, FOXO1 demonstrates metabolic regulation by augmenting PDHK transcription, followed by inhibition of glucose metabolism[288]. The FOXO3 variant in cardiac human tissues is also associated with one's longevity, as evidence suggests that it might have a cardioprotective role against metabolic diseases[78,125], while suppression of FOXO1 and FOXO3 makes the cardiac tissue more susceptible to myocardial injuries[728]. The cross-talk between FOXO and NF-kB regulates inflammatory reactions. In particular, an increase in FOXO1 expression shows an increment in inflammatory activities[868].

As mentioned earlier in this chapter, fibroblasts can greatly impact the composition of the extracellular matrix and, consequently, the mechanical functioning of the cardiac pump (see Section 2.2), and Klotho protein can regulate several fibroblast functions. Klotho can be found in two isoforms, α-Klotho and β-Klotho. In mice, the partial inhibition of one of the alleles impairs the synthesis of endothelial nitric oxide and reduces SIRT1 expression while increasing the stiffness of cardiac vasculatures[256]. The presence of α-Klotho has an inhibitory effect on transforming growth factor β 1 (TGFβ) and insulin-like growth factor 1 (IGF1)[450]. With ageing, TGFβ appears to increase in the plasma, although the circulating TGFβ and gene expression do not correlate[820]. Overexpression of TGFβ strongly stimulates cardiac fibrosis, a condition that can occur with ageing or after cardiac injuries[749]; evidence shows that the activation of TGFβ signalling pathways in mice is regulated by the ASK1-signalosome. Indeed, ASK1 knockout inhibits TGFβ expression with a consequent reduction of the fibrotic process[627]. Regarding IGF1, experiments in the mouse model show that IGF1 inhibition increases the life span of the mouse, although IGF1 inhibition for younger mice appears to compromise their cardiac health[3]. The impairment of Klotho and its cascade effects appear to be associated with a risk of coronary artery diseases[458].

Low sterile inflammation characterises the ageing process, and the mammalian target of rapamycin (mTOR) is a pathway modulating inflammatory reactions in the cardiac pump[902]. mTOR is a serine/threonine kinase and is composed of two functional enzymes, mTORC1 and mTORC2, although the interaction between those two complexes and the functionality of mTORC2 is still a matter of research[247,392]. mTOR1 activation can occur through several growth factors, such as IGF1. Substrates, like amino acids, can also regulate mTORC1. On the other hand, mTORC2 is less sensitive to rapamycin and other substrates, which makes it a difficult target for therapeutic applications[75,774]. mTORC1 and mTORC2 have different downstream effects on cellular processes, and their roles in age-associated cardiac diseases are well recognised[392,725]. mTORC1 inhibition displays several beneficial effects, such as decreasing the progression of cardiac hypertrophy and enhancing the regression in heart failure. Furthermore, mTORC2 activity benefits the

heart by promoting cell survival[724]. Although mTORC1 appears to be a suitable target for age-related cardiovascular diseases such as dilation and inflammation of the cardiac muscle[488], treatments that target it have not been shown to have not shown any relevant beneficial effects[154].

2.4.2 Metabolism

Cardiac metabolism plays a pivotal role in cardiac ageing, and, as discussed in Section 2.3, fatty acids and glucose guarantee energy homeostasis in the heart. Mitochondria are at the central of the ageing research[501], and the study has shown that in healthy individuals aged 90 and over, 95% of mitochondria in muscles were damaged, while no mitochondrial damages were detectable in those aged 5[500]; similar findings have been reported in mice cardiac tissues[167]. Cardiomyocytes, the predominant cell type in the left ventricle, require constant energy to fulfil the excitation and contraction and allow the cardiac pump to work properly. Consequently, they exhibit a very flexible metabolic network that can switch from one substrate to another to keep the energy production machinery up and working. Cardiomyocytes in a fetus yield energy mainly from carbon central metabolism and oxidative phosphorylation, and glucose is driven from the maternal-fetal gradient[259]. Even though cardiomyocytes may have reached maturation[302], glucose-based metabolism remains as the main energy supplier during the first days of the newborn[259]; nonetheless, the in vitro experiment on pluripotent stem cell-derived cardiomyocytes has displayed that the metabolic switch from glucose toward fatty acids oxidation occurs upon cell maturation. In newborns, after initiation of breastfeeding that is rich in fatty acids, their glucose level increases to a similar level observed in adults[815]. Experiments in human pluripotent-derived cardiomyocytes have also shown that cultures grown in a high-fatty acids medium are intoxicated by lipids[341], while high glucose levels can impair cardiomyocyte differentiation[913].

The mature cardiomyocyte maintains the energy production relying mainly on fatty acid oxidation. Evidence has highlighted that further metabolic reprogramming of cardiomyocytes occurs with ageing through switching ATP production from fatty acid oxidation to carbohydrate metabolism[469,529]. The metabolic rewiring observed in ageing also occurs as a response to different energy demands from young and aged cardiac myocytes[916]. Changes in the metabolic processes are also noted when taking into account the ATP/PCr ratio, a well-established metric summarising the energy utilisation in cells and tissues[179]. The ATP/PCr ratio decreases in the older hearts compared to the younger ones[338,916], which supports the hypothesis that ageing is associated with a decreased efficiency in cardiac contraction[179,231]. Although the literature supports the association between ageing and metabolic rewiring, metabolic activity is also strongly influenced by other factors, such as a healthy lifestyle, regular physical exercises and balanced nutrition, which can contribute to delaying the changes induced by ageing[300,910]. The switch of substrates for energy production has drawn the attention of researchers, who have examined the crucial roles of energy maintenance and ATP in ageing, which has led to the novel hypothesis: the Energy Maintenance Theory of Aging[119]. This theory embraces the hypothesis that higher ATP expression in mature cells is necessary for prolonging one's life span, as the energy allows

several biological processes to take place, such as mitochondrial biogenesis, mitochondrial fusion, lipid damage repair, proteostasis and DNA repair. The Energy Maintenance Theory of Ageing is also compatible with the Mitochondria Free Radical Theory, as a higher abundance of ROS would undermine mitochondrial and cellular health[119]. In support of the Emerging Maintenance Theory of Ageing, experiments on C. elegans, Drosophila melanogaster and mice have confirmed the association between a longer lifespan and a higher ATP level in ageing[60]. The experiment in mice has also demonstrated that the cohort with the highest metabolic activity exhibits a longer lifespan despite the increased food intake[771].

Conclusion

The ageing of the human heart is a process that needs to be addressed from multiple aspects[1,158,595], and it can be studied at the anatomy, tissue or cell level. This chapter has discussed these approaches through an in-depth literature review on the human heart, including examples of related changes observed in cardiovascular diseases[145]. Although the heart anatomy was discussed based on the four chambers, I have also introduced more recent explanations anatomical description of the heart and how the contraction occurs[956]. Successively, I have described the ultrastructure of the heart, presenting the latest evidence on the cell population[504,649,833,948], with a focus on the left ventricle[504], as it undergoes through multiple changes in accordance with ageing-associated cardiovascular diseases, such as heart failure[748,781]. Studies suggest that the cardiac ageing process and age-related diseases compromise the metabolic[92,158] and signalling events[627]. I have also touched upon the key metabolic pathways utilised by cardiac cells and molecular changes relevant to cardiac ageing investigations. These approaches to understanding cardiac ageing are indeed interconnected to the underlying molecular events, which has led to significant discoveries in the past decades not only in molecular biology but also in anatomy[384], physiology[674] and cellular characterisation[504].

Chapter 3

Aims, Scopes and Contribution

In this chapter, I mainly discuss the aims and scope of this study and highlight my contribution to the project as well as to the field to establish the rationale for this study. As previous chapters have delved into, human cardiac ageing is a prominent area that needs further examination from several perspectives, and the in-depth insights this study offers bring several benefits.

The first three sections respectively introduce the three aims and scope for the characterisation of human cardiac ageing as well as the contributions each objective can bring about. In the following section, I briefly refer to the contributions this study makes to other studies in the field outside the context of human cardiac ageing and the role I played in them. One example of such contribution is the interactive web platform of data regarding human cardiac experiments and measurements with an interactive interface that allows researchers and practitioners to retrieve valuable information and conduct a statistical and mathematical investigation (`https://humancardiac.shinyapps.io/ageing/`). The later part of this chapter also mentions prizes and awards received during this project that involved my contribution, followed by conclusion.

3.1 Molecular signature of ageing

The first aim of this thesis is to advance the current knowledge about healthy human cardiac ageing. This is done by examining human cardiac tissues at the molecular level across young and old individuals. The molecular examination is conducted across four different categories: proteins (age-proteome), metabolites (age-metabolome), lipids (age-lipidome) and transcripts (age-transcriptome). To identify changes that human cardiac tissues undergo in ageing, our study is conducted on unique pre-mortem, cryopreserved, left ventricular tissues excised from healthy donor hearts held in the Sydney Heart Bank[199], the largest cryopreserved human heart biobank in the world[842]. This study is one of the largest and extensive studies on human cardiac tissues; especially as a cardiac age-ome study, it is the first study ever conducted on pre-mortem human cardiac tissues (see Chapter 5. Capturing the molecular landscape of human cardiac ageing will benefit the research community in testing outstanding theories on the ageing heart[258] and generating novel hypotheses on how ageing sets fertile conditions for the development of cardiovascular

diseases and increases the frailty of the cardiac tissue[376]. While the cardiac age-OME study provides a unique resource by elucidating the molecular landscape of cardiac ageing in left ventricular tissues, other studies have examined the role of metabolic fuels used by the heart and how molecular landscape in blood plasma changes with age[584,742]. As previously mentioned (see Chapter 1) Shen et al. (2024) have conducted a multi-omics analysis on human blood plasma in 108 participants for a median period of 1.7 years and have provided multiple insights including the non-linearity of the molecular abundance across ages of the individuals[742]. Similarly, Murashige et al. (2020) provided a comprehensive analysis of the metabolites and lipids detected in the transcardiac gradients in failing and non-failing human hearts. Although Murashige et al. (2020) does not investigate the effect of ageing, it gives valuable information on the fuel the non-failing human hearts consume instantly. The study found that, on average, the 87 non-failing heart individuals with 63.95 ± 12.32 years old mainly consume free fatty acids, from where it obtains almost 70% of the carbon, while 15% of the carbon is yielded by ketone bodies. The transcardiac analysis also revealed that the heart mainly released amino acids, with alanine and glutamine being the most abundant. Interestingly, non-failing hearts show evidence of high reliance on the TCA cycle; however, the heart also abundantly releases intermediates, probably due to the excess of TCA intermediates accumulated in the heart[584].

Using the left ventricle, this study offers a possibility of extending the results on cardiomyocytes, the predominant cell type of the left ventricle (see Chapter 2.2). Furthermore, it provides a deeper understanding of the role of the left ventricle in cardiovascular diseases[19,832,921]. Based on sample size calculations, sufficient samples were available in the Sydney Heart Bank for the originally planned investigations on proteomics, metabolomics and lipidomics. We were further able to characterise tissue transcriptome with the remaining tissues, with possible consequences on the power of statistical analysis. As a result, the transcript-wise examination was conducted to explore whether the protein landscape is coupled with the gene expression of the tissues. Identifying molecular signatures of the healthy ageing process, this additional examination offers insights into healthy ageing dynamics of the heart. Characterisation of healthy ageing will allow further research by comparing healthy ageing molecular patterns with molecular changes due to cardiovascular diseases, although the findings are explorative and further studies and meta-analyses are required to validate them.

The results of this study (see Chapters 5 and 6) provide some notable implications. The cardiac age-ome study detected and analysed 2938 proteins, 150 metabolites, 341 lipids and 15447 transcripts; this marks the most extensive quantification on pre-mortem cardiac tissues. Computational analysis of the cardiac age-ome unveiled novel features regarding age-related molecular changes that have not been reported in previous studies. Our analysis has further captured the dysregulation of pivotal proteins, such as SERCA2 and MYH6, both required for heart contractility. The protein findings are not observed in the RNA counts, suggesting a phenotypic variation between younger and older cohorts of human hearts rather than changes in the gene transcription[56,210]. At the transcription level, the sodium leak channel NALCN, a highly conserved protein in the heart, increases its RNA counts in the aged cohort compared to the younger one[727]. The NALCN is established in literature to be associated with coronary diseases[579,911], and pharmacological studies have shown that

it can impact the pacemaker activity of cardiomyocytes[523]. The quantification of metabolites offers a substrate-oriented perspective of the age-ome. The metabolites analysis identified aberrant carbon metabolism, given that several analytes, such as fructose-1-6-bisphosphate and glyceraldehyde 3-phosphate, present a higher abundance in the elder cohort. Our findings also corroborate a higher central carbon metabolic activity in the healthy aged cohort reported in the literature (see Chapter 2.4). Analysis of lipids in the cardiac tissues identified a higher abundance of phosphatidylglycerol (PG) with long and very long fatty acid tails in the aged cohort, and similar results are reported in a study on human plasma[395]. A technical contribution is also introduced in this part. The changes observed in each class of lipids have been linked to the length and polyunsuturation of the fatty acids tail of the compound using a weighted least square approach. To our knowledge, no method is available in the literature that allows such investigation[576]. When lipids are analysed based on their class rather than on individual abundance, the increase of triglycerides is observed together with a generalised increment in the abundance of acylcarnitine. This indicates that metabolic activities, such as β-oxidation, are altered as the cardiac tissue ages. Dysregulation of metabolic activities during ageing is also reported in animal models[60,771] and supports the Energy Maintenance Theory of Ageing[119].

Furthermore, the molecular signature's outcomes were then used in a novel manner to drive three biologically informed mathematical models. In detail, Dr Michael Pan and I created a Monte Carlo simulation using the averages estimated for each analyte and the observed biological variability across the young and old population. The generated data were then used to feed the model developed by Dr Michael Pan.

3.2 Omic as a biological network

Although identifying molecular signatures is the well-established approach to exploring differences across cohorts, it does not consider common trends and interactions amongst analytes of the same type and consequently neglects some effects that, for example, a metabolite has on other associated metabolites[377]. The exclusive focus on molecular signatures would not only limit the study's findings to molecules that exhibit the largest variations but also prevent us from gaining a better understanding of preserved and rewired interactions amongst analytes occurring in the ageing process[944]. The proteins, metabolites, lipids and transcripts system investigations have been performed via weighted correlation network analysis (WCNA), a well-established method in systems biology[702] that enables us to identify highly connected vertices, known as hubs, potential regulators of the healthy cardiac ageing process[71,265]. The systems-level can also unveil functional modules that contribute to ageing. A functional module is identified as a set of highly connected nodes which elicits the discovery of biological functions[265] that operate to ensure the healthy ageing of the heart.

As a second aim, this thesis extends the age-ome study beyond the molecular signature approach by analysing transcripts, proteins, metabolites and lipids as whole systems rather than as single molecular entities. The analysis focuses on identifying hubs and functional modules[136] within each omics layer crucial for maintaining the healthy ageing of the cardiac tissue. The results can be utilised to identify possible prognostic factors[803]

of human cardiac ageing with potential application in the development of pharmacological solutions[43]. WCNA also offers the opportunity to employ the estimated association network from healthy cardiac ageing data as a baseline when comparing disease state tissue networks[71]. My contribution to the omics analysis on the system-level spaces is from novel biological observations obtained from applying the network approach to the human cardiac ageing data (see Chapter 7) and the novelty of workflow introduced in analyses of the network (see Chapter 4.9).

As a novel technical contribution, the network for each omic was inferred by maintaining the sign of the correlations across different analytes. Consequently, summary statistics of the network can capture a wider variety of behaviours with respect to the one offered from currently available frameworks[455,927]. Furthermore, our novel workflow also employs a different method for community (or module) detection in the co-expression network. In fact, our work ensures that communities in the network are identified by taking into account the strength of analytes-analyte association[455]. The functional annotations of communities in a network have been limited to gene and protein networks[455]. As part of the technical contribution, our workflow allows the functional annotation of the community in the network independently from the type of omic network (see Chapter 4.9 for further details of our technical contribution).

Our workflow was then utilised to derive biological insights into the cardiac ageing process. The application of our WCNA on protein abundances has led to the identification of three key proteins responsible for the organisation of the biological ageing heart protein network[884], SOD2, MB, and NDUFAB1, as important players, or hubs, for the ageing process. The largest community detected in the protein network hints that several pathways, such as oxidative phosphorylation and cellular proteostasis, contribute to maintaining healthy ageing. The weighted correlation analysis on the transcripts data identifies ARPC4, MAZ and MAD as network hubs, while the largest module in the network suggests that biological functions such as IL-5 signalling as well as metabolic processes of ceramides and sphingolipids are also involved in the ageing process. When applied to the metabolite's measurements, WCNA ascertains Purine, Homocysteine, Fructose and 1-methylhistamine as key metabolites, together with several metabolic processes, such as one-carbon and NAD^+ metabolism, associated with the largest community in the network. Analysis of the age-lipidome using the weighted correlation network indicates that phosphatidylcholine and lipids with long and very fatty acid tails are the pivotal lipids in the network. Additional network module analysis shows PC lipids have a key role in ageing, as the module contains only PC lipids (see Chapter 7 for discussion of the findings).

3.3 Omics SEA

The systems approach and the identification of molecular signatures that determine the phenotypic differences between younger and older tissues give invaluable insights into possible molecular functions, molecular association and master regulators of the ageing process. Those indisputable essential investigations confine the insights and

discoveries we can draw on the healthy cardiac ageing process only within a single molecular type, namely proteins, transcripts, metabolites and lipids. A more integrative approach could let us gain greater insight into possible dysregulated cellular functions[574,784].

The third aim of this thesis is to explore aberrant biological pathways in the context of human cardiac ageing. The scope of this study is to utilise all the evidence collected in the differentially abundant analysis and inspect annotated pathways across databases to see if any biological functions are affected by the ageing of the cardiac tissue and cardiomyocytes. Although we had initially aimed to apply existing methods, no tool in the literature allowed us to investigate the pathway activities at the multi-omics level, especially when more than 3-omics are considered[109]. This is because it is not possible to use transcripts, proteins, metabolites and lipids to perform the pathway analysis[937]. To push the understanding of the ageing process, we have created a novel enrichment method, namely omics SEA, capable of integrating any omics measurements to infer aberrant molecular pathways across different cohorts. Although omics SEA was created during the age-OME project, this method can be used in any multi-omics study. In fact, omics SEA is a general method that infers up-or down-regulation of a specific pathway and couples dysregulation with a cohort in the study, allowing researchers to identify the cohort in which dysregulation occurs. Details of the method can be found in Chapter 4.10.2.

The omics SEA application to the age-ome has shed light on alterations of metabolic activities during cardiac ageing. The most relevant metabolic changes observed in the aged cohort is the higher reliance on central carbon metabolism for energy production rather than fatty acid oxidation. Metabolic dysregulation and changes in energy utilisation were also captured by the phosphocreatine to ATP ratio[179]. The application of omics SEA to the age-OME data is presented in Chapter 8.

3.4 Other contributions

While this study has made direct contributions to research into cardiac ageing and bioinformatics, I have applied my skills and approaches to other projects within and beyond the characterisation of cardiac ageing, which I list in this section. The human cardiac ageing project is part of a large project that establishes the first web platform of data from diseased and non-diseased hearts. I have conceptualised, designed and developed the platform. It will serve to drive future research in the cardiovascular field and be available to the public at: `https://humancardiac.shinyapps.io/ageing/`. Furthermore, I have contributed to the study on the RNA Polymerase II and DNA replication led by Rojas et al. (2024) by designing and implementing a test for examining if there was evidence of a clash between the transcription and replication machinery. Details can be found on the published manuscript[691]. In a study on the impacts of the cellular ultrastructure on the metabolism in cardiomyopathy conducted by Ghosh et al. (2022), I have applied my knowledge of linear mixed models to test if any difference occurred across multiple conditions investigated in the study. Details are available on the published manuscript[269]. In addition, I have contributed to a study *Diabetic ischaemic cardiomyopathy: Characterising the left ventricular myocardial*

molecular profile in advanced human heart failure by Dr Benjamin Hunter by performing quality control of the RNA sequencing reads and subsequent alignment and counting of the genes. The manuscript is currently in review for *Cell*. Lastly, in a study conducted by O'Sullivan et al. (2024), I have applied my expertise in lipidomics and pathway analysis to investigate consumed and accumulated pathways in HFpEF patients. The work has been published recently, and the article was featured by a dedicated editorial comment[20].

3.5 Awards and funding

The unique approach and findings of this study have indeed offered possibilities to extend our knowledge beyond the characterisation of cardiac ageing and explore medical applications. I have been part of several grant applications that aimed to advance the field and inform practice. At the time of writing, I have been awarded two funding schemes, which demonstrate the significant possibility this study provides. The first funding awarded is the Accelerator Award from the British Heart Foundation [AA/18/2/34218], with Professor Fabian Spill as principal investigator. Fabian and I jointly developed the idea and wrote the grant proposal. A second source of funding has been awarded through the EPSRC Impact Acceleration Account (IAA) - Doctoral Prize for Innovation. This award provided funding to translate my research into a prognostic medical app that allowed medical practitioners to estimate an individual's metabolic status and monitor possible divergence from its normal status. The idea was awarded 50,818.91 GBP. I solely developed the idea and proposal, and Prof Fabian Spill was listed as a potential line manager for the project.

Conclusion

The human cardiac ageing study, the age-OME project, is the first study that examines the pre-mortem human cardiac tissue across ageing, which is currently under review at *Cell - Stem Cell*. The manuscript provides great insights into basic science that will advance the research into cardiac ageing. Additionally, the project also contributes to the field of bioinformatics through the development of new frameworks for analysing biological networks and multi-omics biological functional annotation. I have made further contributions through the interactive web platform for human cardiac data and my dedication to other studies. These contributions highlight the significant theoretical and practical insights this study – its approach, employed methods, findings and their applications – provides to scholarship.

Chapter 4

Research Methods

In this chapter, I introduce analytical instruments and computational methods I utilised for the age-OME investigation. It opens up with Section 4.1 to briefly discuss proteomics investigations and research methods adopted for protein screening in the age-OME study. Section 4.2 is devoted to applications of metabolomics to medical research and our experimental setting for the metabolome characterisation in the ageing cardiac tissue. The following Section 4.3 touches on high-throughput technologies for lipids, a prominent area of research in the omics sphere, providing insights into the methods utilised for studying human cardiac ageing. Then, Section 4.4 concisely reviews transcriptomics and details the methods used for the age-OME project. In Sections 4.5 - 4.7, I introduce the workflow for analysing omics measurements by highlighting the relevant steps of data pretreatment. In particular, I discuss possible transformations that can be applied to the omics measurements to match the assumption underlying the follow-up statistical and mathematical methods to investigate the human cardiac ageing process. Then, I examine normalisation methods that mitigate the amount of unwanted variation in the data. In the following section, I succinctly review some methods that can be utilised to handle missing measurements for omics experiments. This chapter then moves onto expounding on methods applied to data analysis. Section 4.8 shortly reviews statistical methods that can be employed to identify the potential biomarkers, followed by a detailed description of limma, the framework employed in our project.

In Section 4.9, I discuss our novel co-expression network framework used in our project to investigate common trends in the omics data. I introduce the novelty that I brought into network analysis by performing the pathways enrichment analysis to the largest identified community for each omic network. The section also explains the algorithm used for reconstructing the network. Section 4.10 covers the review of the methods applied to perform enrichment analysis. Then, it extensively describes over-representation analysis used to attribute the functional role of the largest community detected in co-expression network analysis as well as the new framework that I developed to perform pathway analysis in multi-omics studies, followed by the conclusion.

4.1 High-throughput proteomics

Proteomics refers to the full characterisation of proteins in a certain system[635]. Proteomics as a study encompasses protein detection in a system, the understanding of functions of proteins, and their regulation, structure and possible interactions with other proteins and small molecules[66,131]. The technological advancement in high-throughput methods has profoundly influenced clinical research and allowed researchers to identify diseases and healthy phenotypes across individuals by profiling proteins in biological fluids, tissues or cells[560]. For instance, proteomic analysis amongst clinically obese patients has revealed differences between individuals affected by heart failure with preserved ejection fraction and individuals without cardiac issues[443]. Combined with medical research, the application of proteomics in clinical screening has improved personalised medicinal practices and early detection of diseases[843]. However, one of the main intrinsic challenges in proteomics is one-to-many relationships that a gene or a set of genes is responsible for protein-encoding[99]. The one-to-many relationship between a protein and genes encoding proteins is due to several post-translational events, such as post-translational modifications, RNA splicing and translation errors[284]; the total amount of proteins in the human body is still unknown to this day[6].

Among several protein investigation technologies, mass spectrometry (MS) may be the most prevalent due to its sensitivity and accuracy in detecting the presence of different analytes in a biological sample[164,295]. Two possible approaches can be taken in proteomics research: top-down and bottom-up[164]. The top-down approach aims to calculate the exact molecular mass of a few pre-selected proteins and mainly relies on selected or multiple reaction monitoring coupled to triple quadrupole MS[447]. As an advantage, this approach quantifies the molecular mass of analytes in biological samples, but only for a few targeted proteins[447]. Targeted proteomics is a valuable tool for confirmatory[877] and mechanistic studies[560], but not suitable for protein-wide investigations and identification of candidate biomarkers[766], which this age-OME study aims. On the other hand, bottom-up proteomics has the opposite objective since it aims to identify all possible spectra of proteins in the injected samples.

Proteins are quantified via their intensities or ion counts, a measure proportional to the quantity of a protein in the biological system[717]. Before samples are injected into the MS plates, enzymes are added to the sample in order to cleave polypeptides at a specific site and create simpler peptides[239]. A common enzyme utilised for digestion of proteins is trypsin[848], which cleaves the protein at the C-terminus of lysine and arginine, two amino acids that build proteins[537]. Despite some available enzymes[270], trypsin is still understood as the most widely used protease for proteomic analysis[205]. Trypsinised samples are injected into plates and ready to be analysed. Liquid Chromatography (LC) and Gas Chromatography (GC) are the most employed platforms for untargeted MS due to their capability of separating the samples in an online fashion, reducing sample preparation in the lab[406]. The difference between GC and LC relies on how peptides are delivered (mobile phase) to MS; in GC, the mobile phase is in a gaseous state, while in LC, the mobile state is in a liquid state[604]. Once the spectrometer concludes the run, which can comprise several plates, it returns the peaks associated with the mass-over-charge. Using

a database, the protein is associated with the peptide that differs from the background noise[406]. Higher-quality examination of peptides can be conducted using tandem MS (MS/MS or MS^2)[447]. MS/MS enhances peptide identification by further fragmenting it in a collision chamber; the fragments are then analysed with the second MS[676]. MS/MS employs two MS and can be applied with chromatographic and ionisation methods[676].

For the age-OME project, we have taken a bottom-up proteomics approach. As outlined earlier, the objectives of this study are to characterise as much protein as possible in the young and old cohorts and to identify candidate biomarkers that can have clinical relevance in detecting the ageing state of the cardiac tissues. Trypsin has been used for digesting peptides. Moreover, using the MS/MS technique has boosted the sensitivity of the measurements and protein identification[447], coupled with liquid chromatography columns[349]. Further details regarding the analytical platform, sample preparation and data acquisition are discussed in Chapter 5.

4.2 High-throughput metabolomics

Identification and quantification of all the metabolites (metabolome), small molecules with molecular weight less than $1500Da$, can offer a snapshot of energy status, active or inactive signalling events[175,757,867], immune responses and production of endogenous toxins[885] in the system. Metabolites in an organism are both endogenous and exogenous and can be derived from micro-organisms that live in symbiosis with the individual[165]. Recent studies have also shown that metabolites can regulate gene expressions through allosteric interactions with transcription factors[466,614], posing metabolic processes under a new light as not merely involved in energy management but rather being an integral part of the system that can reflect its effects on proteins and gene expression[169,466,497]. For instance, increase in reactive oxygen species has been linked to post-translational modification of proteins, protein activity, lipid peroxidation and changes in DNA expressions[168,754]. Application of metabolomics in clinical practice has successfully identified several biomarkers in cardiovascular diseases[555]; for example, quantification of phosphocreatine (PCr) and adenosine triphosphate (ATP) allows medical practitioners to evaluate the PCr/ATP ratio, a prognostic biomarker employed to assess cardiac metabolic status[179]. A decreased PCr/ATP ratio at resting state is associated with several cardiovascular disorders such as left ventricular dysfunction[598] and diabetes[101]. Recent evidence has also shown that the PCr/ATP ratio has better prognostic biomarker performance for heart failure than left ventricular ejection fraction[644]. Characterisation of the metabolome also enhances understanding of metabolic cardiac disease; examples include circulating branched-chain amino acids such as valine, leucine and isoleucine that are prognostic biomarkers to identify subjects at risk of developing cardiovascular disorders[600,817]. While metabolomics has found efficient applications in biomarker discovery, mapping of metabolites in several biological systems has also enabled researchers to identify key dysregulated biological functions unique to a specific phenotype and characterise physiological conditions. This approach attempts to determine differences at the systems level rather than identifying a specific molecular signature[390,391]. For example, phenotype profiling of young individuals has allowed researchers to identify

those at risk of cardiovascular diseases by taking into account all the molecular abundances in the systems rather than comparing single molecules between cohorts; the results have been validated in two independent cohorts[588].

There are several analytical methods to screen metabolites[713]. MS coupled with chromatography techniques and nuclear magnetic resonance (NMR) are amongst the most popular platforms for metabolome-wide investigations[220]. NMR exposes biosamples to a strong magnetic field and records the frequency of electromagnetic radiation emitted during the relaxation of certain atomic nuclei, such as hydrogen. Since the electromagnetic radiation of the nucleus is heavily affected by surrounding atoms, it can be exploited to identify the compound[220]. NMR allows the identification of metabolites not only in biological samples but also in living cells in a non-invasive and non-destructive manner with minimal sample preparations[57]. However, the main shortcoming of NMR is its low sensitivity to biological samples with complex matrix samples such as left ventricle tissues[162]. As such, for human tissues, MS is more widely employed than NMR due to its higher sensitivity.

Approaches to metabolome studies involving MS can be categorised into three types: untargeted, targeted, and semi-targeted. The untargeted approach aims to quantify as many metabolites as possible in a biological sample, where metabolites are unknown until the spectra are collected. This approach allows the researcher to pursue semi-quantitative measurements across a wide spectrum of metabolites and has led to the discovery of novel compounds[263]. The main limitation of this approach, nonetheless, lies in the difficulties of identifying the true unknown metabolites, as multiple compounds can have the same m/z signals when used for clinical or medical studies[106]. Considering this, this approach is unsuitable for the age-OME investigation, as occurrences of unidentified biomarkers might increase the complexity of the study. Secondly, the targeted approach in metabolomics determines the precise abundances of known molecules before acquiring the spectra. The targeted approach enables the researcher to precisely identify and quantify metabolites of interest by using isotopically labelled internal standards that help construct calibration curves to estimate the absolute quantification for each analyte. This procedure is extremely precise despite being labour-intensive to scan large-scale molecular[714]. Although the targeted approach would give us a precise identification and measure of the metabolites, it does not appear suitable for the age-OME project, as we aim to conduct a metabolome-wide characterisation of the left ventricular tissue.

The third option, the semi-targeted approach, allows researchers to combine the strengths of targeted and untargeted approaches while mitigating their respective limitations. Using this framework, hundreds of metabolites are targeted before data acquisition, but the quantification is not absolute; it is estimated by a calibration curve and internal standards for more than one metabolite[123]. The sample preparation for the semi-targeted metabolomics investigation requires standards for each metabolite the study aims to detect. A few standards are also injected in the MS plate to monitor the quality of the reads. Further assessment of the quality of the peaks and evaluation of within-batch and between-batches is conducted using pooled samples as quality control to match the complexity of the matrix in the biological samples[18]. Then, the samples are injected and run in MS. In semi-targeted metabolomics, MS can be coupled to column chromatography to enhance separation, identification and

purification of biological samples, such as gas or liquid chromatography (GC/LC-MS)[510]. GC-MS presents the advantage of good resolution and stable fragmentation of compounds but is limited to molecules that can be converted into volatile analytes, such as sugars and alcohols[236,701]. LC-MS is a valid alternative to metabolomics experiments, as it requires a simple sample size preparation, superior separation and detection of the metabolites, and a broader coverage of metabolites[510]. Further molecular structure can be unveiled using MS/MS connected to the liquid chromatography column (LC-MS/MS). Regarding metabolites, LC-MS/MS provides additional information on the structure of selected ions to better characterise the molecular structure of metabolites and differentiate possible isotopes[638].

As the aim of the age-OME project is to characterise as many metabolites as possible in the left ventricle tissue, the semi-targeted approach seems appropriate due to its reliable semi-quantitative measure of the abundance of metabolites and a large number of metabolites that can be unambiguously identified. The semi-targeted metabolomics can identify hundreds of metabolites with relative semi-quantitative measures that allow the estimation of fold changes across cohorts[273]. LC-MS/MS has been selected as the analytical platform for metabolite quantification. For the study, two different runs were performed for each sample; one run fulfilled the quantified amino acids, such as valine, while the second run detected the metabolites belonging to the central carbon metabolism (for further details, see Chapter 5).

4.3 High-throughput lipidomics

As Section 2.3 has highlighted, lipid molecules are involved in signalling, energy production, and building blocks of cellular and subcellular membranes, and further research is required to deepen our understanding[590]. While lipidomics centres around the characterisation of all lipids in a biological system, the complete characterisation of each lipid still remains an ambitious goal mainly due to the diverse nature of lipids in biological samples[745]. Lipidomics has further flourished owing to recent technological advancements, such as electrospray ionisation MS[98], that identified different phenotypic signatures in diseases and in aged individuals at the class level[343,866] or the single molecule level[472]. Furthermore, examining the length of the carbon chain and its saturation or (poly-)unsaturation attached to a lipid of a specific class can lead to a better characterisation of different diseases. For instance, ceramides with long or very long fatty acid tails can impair mitochondrial functions and develop cardiovascular diseases[137,835]. In addition, lipids are building blocks of several (intra-)cellular structures, such as membranes, and they can also act as energy storage and direct regulators of gene expression, duplication and transcription. For instance, a direct mechanism of gene regulation occurs when lipids bind to transcriptional receptors[864]. Lipids have also found their way into clinical applications. A recent study presents that fatty acid levels in the plasma of rodents are comparable to the level of fatty acids identified in their cardiac muscles, indicating that the heart is the leading consumer of fatty acids[551,584]. The study also highlights that changes in plasma acetylcarnitine can be observed under different conditions, such as a fasted state, which suggests that acetylcarnitine reflects changes in

metabolic activities. This finding points to the use of acetylcarnitine as a possible biomarker for several heart conditions[551]. An investigation on heart failure individuals also showed that tracking changes in the long chain of acetylcarnitine detected in plasma can help to diagnose heart failure and predict possible clinical outcome[9].

Two possible approaches can be employed in lipidomics: targeted and untargeted. The targeted approach can accurately quantify and identify a specific set of lipid species and find its natural application in confirmatory studies and personalised medicine[812]. While the targeted approach does not match the aim of the age-OME study, the untargeted approach, however, allows us to screen hundreds of lipid molecules simultaneously and generate hypotheses from the cardiac ageing data. Yet, the untargeted approach comes with drawbacks, as it only provides the approximation of each molecule's abundance rather than the exact quantification[107,352]. It is therefore necessary for us to balance the strengths and limitations of the untargeted approach for the age-OME project.

To conduct untargeted lipidomics analysis, MS is currently understood as the gold standard[953]. Lipidome investigation starts with sample homogenisation. Samples are generally kept frozen to halt any enzymatic reactions while homogenisation increases the efficacy of solvent[837]. Then, extraction methods are applied to the samples to reduce the complexity of the sample matrix and increase the signal-to-noise ratio of lipids. Several extraction protocols have been proposed in literature[837]. For instance, a protocol for lipids extraction involves methyl tert-butyl ether, which simplifies the sample preparation and improves the lipid profiling in untargeted lipidomics[547] with respect to the well-established Folch or Bligh-Dyer protocol[79,461]. Literature also discusses multi-phase extraction protocols, such as two-[547] or three-phase protocols[839], which increase the spectrum of lipids identified by MS and the complexity of sample preparation. Whilst the sample would be ready for injection for direct infusion electrospray ionisation MS, direct infusion would require a higher abundance of samples with respect to other frameworks[503]. Ion suppression that derives from competitive analytes reaching the detector at the same time and the presence of isomers are other shortcomings of direct infusion electrospray ionisation MS. Indeed, this MS can lead to complex spectra where some analytes are almost indistinguishable and can be wrongly attributed, increasing the complexity of the spectra and consequently reducing the lipid detection power of the platform. Despite new methods, this still remains one of the main challenges[350]. For this project, due to the limited amount of tissues available, the direct infusion would not match the parsimony principle of biobanks[373]; liquid chromatographic techniques, such as HILIC or HPLC, would reduce the number of samples necessary for lipid-wide detection and increase the identifiability of lipids[503]. It would require internal standards for each lipid class that we aim to investigate, which add complexity to the sample preprocessing phase[503,953].

To sum up, the untargeted approach aligns with the aim of the age-OME study. Direct infusion electrospray ionisation MS was not utilised due to the higher amount of biosamples required with respect to when MS is coupled with liquid chromatographic columns despite the increase of sample preprocessing and experimental costs (see Chapter5 for the sample preprocessing protocol and analytical platform).

4.4 High-throughput transcriptomics

DNA contains a vast amount of information about a human, from hereditary traits to environmental influences[939], and genomics is the discipline devoted to studying and characterising genes in biological organisms. Amongst research efforts, the Human Genome Project, a world-scale project, has established modern genomics by sequencing most of the human genes[271]; the mapping of the human genome has been completed recently[605]. This project has revolutionised research approaches toward genetics research, and advances in high-throughput technologies have allowed large-scale genetic investigation and amplified accessibility to sequencing devices by reducing the cost of analysis and sequencing platforms[811], which have consequently benefited genome-wise research[776].

Meanwhile, transcriptomics, the study of RNA molecules, has garnered researchers' attention[562]. It aims to characterise coding and non-coding RNA molecules[177] and their role in biological samples. RNA molecules originate during the process of transcription led by the RNA polymerase II machinery[232] and are made of four nucleotides: cystine, guanine, thymine and uracil[571]. The ability of cells to keep their genetic lineage is pivotal for all cells in a tissue, as it can alter their functional role. Transcriptomics explores changes at the transcript level, reflecting genes accessible to RNA polymerase II and consequently characterising the active/expressed genes in the biological system[434]. For example, a recent study has analysed RNA expression and revealed that the inability of cardiomyocytes to maintain the platelet isoform of phosphofructokinase gene expression is a possible cause of cardiac hypertrophy[852], conditions that can lead to heart failure[607]. Since gene expression is reflected in transcript abundance[161,791], transcriptomics has been predominantly employed to study differentially expressed genes across different cohorts[776] and characterise related genotypical and/or phenotypical profiles of biological systems[594]. Researchers have also found applications beyond the differential analysis of genes, including alternative splicing[127] and single nucleotide polymorphism detection[257]. Furthermore, transcriptomics has been widely applied in medical[261] and clinical research[789]. For example, in medical research, transcriptomics has been employed to understand the link between gene expressions and genotypical and phenotypical profiles of patients affected by Type 2 diabetes[383]. In clinical research, it has helped researchers understand the effect of beta-blockers, common drugs for patients affected by heart failure, on their gene expressions to predict and improve their therapy outcomes[327].

There are two main approaches to studying RNA expressions: targeted and untargeted. Targeted transcriptomics places hundreds of complementary DNA or cDNA of genes robotically on a grid. Reference and test biological samples are prepared, dyed with different colours and applied on different grids. Subsequently, confocal microscopy is used to quantify the colour intensity in each cDNA microarray, and the intensities across tests and references can be used to compare gene expression across two types of samples[204]. High-throughput sequencing platforms have replaced microarrays due to their higher coverage of transcriptome and sensitivity[776]; however, microarrays cost less than sequencing technologies and can still be a valuable option for possible screening tests on large-scale populations[790]. Targeted RNA-seq is the state-of-the-art technology for studies that aim to target a specific set of genes and serves for confirmatory studies as well[334,805]. Although the targeted

analysis represents a valuable option for the age-OME project, limiting the sequencing to only a few genes would exclude the opportunity to generate a hypothesis on the ageing process.

Untargeted transcriptomics aims to fully investigate RNA molecules in a biological system[805]. In summary, a possible approach to RNA sequencing starts with sample preparation, including purification, isolation and extraction of RNA molecules[720]. Once the RNA is extracted, ribosomal RNA (rRNA) depletion follows and eliminates RNA molecules that constitute the majority of RNA molecules in a biological sample and are integral to ribosomes[441]. The remaining RNA molecules are successfully fragmented and used to construct RNA libraries; alternatively, the RNA fragments can be reversely transcribed in cDNA via enzymatic reactions and then use the cDNA for building the library[653]. Once the library is built, one of the two strands of cDNA is degraded for a stranded protocol. Stranded protocols reduce artefacts when gene expression are quantified[942]. cDNA molecules are then processed in the sequencer, which generates reads of various lengths. The sequence can occur from one or both ends of a cDNA molecule. The study has shown that there is little difference in gene quantification for reads with more than 25 base pairs of length whereas pair-end sequencing tends to outperform the single-ended counterpart when it aims to detect splicing[133]. Notably, not all the sequencing platforms require the reverse transcription from RNA to cDNA or can work directly with RNA molecules[190,619]. The reads from the sequencer are converted into FASTQ, a format that retains metadata regarding the analytical platform as well as quality measures for each base of each RNA sequence[147].

The quality measure for each base is a Phred quality score, defined as $Q_{ri} = -10 \cdot \log_{10} p_{ri}$, where p_{ri} is the probability that the i^{th} base or cycle in the r^{th} read is incorrectly attributed. The p_{ri} can be estimated using several methods[932]. The average quality score calculated per cycle is commonly used to evaluate the overall quality of the nucleotide detection of the sequencer. A common practice is to use the cycle-wise average of a Phred quality score ($\bar{Q}_i$) to assess the quality of each cycle; $\bar{Q}_i \geq 20$ are usually considered as cycles with good quality[81]. If the quality of some reads is below the cutoff, the researcher can evaluate to improve the quality of the data by trimming the reads after a certain sequencing cycle[297]. Other visual inspections involving a Phred quality score have been used during the age-OME study to further explore the data quality. For instance, boxplots of the read-wise average Phred quality score ($\bar{Q}_r$) were used to assess the overall samples' quality. Additionally, a histogram of all Phred quality scores was also plotted to ensure that they exhibit a left-skewed empirical density function. Possible bias can also be introduced during the library preparation[442]. To investigate the redundancy of reads, a visual inspection of a number of reads against occurrences can reveal a possible bias when reads have an anomalous abundance frequency. A $\log_{10} - \log_{10}$ scale of the plot can help to visualise this relationship. An extension of this analysis is to look for redundancy in sequences of nucleotides or k-mers. The k-mer analysis can reveal the nucleotide sequence amongst all reads that are particularly abundant, hinting at possible bias in the library preparation[147,442]. A sequence of nucleotides is over-represented when it occurs more than 0.1% of the total amount of reads. The over-representation of a sequence could be an artefact introduced by the adapter[946] or bias due to cDNA amplification techniques, leading to serious bias in following analyses[151]. A complementary metric to the over-represented sequence

relies on the frequency distribution of guanine-cytosine combined (G-C), also known as G-C-content bias analysis[64,249]. The frequency of G-C across all the reads is expected to be approximately normally distributed, and any deviation from the Gaussian distribution can suggest a possible bias in the sequenced data. Amplification of cDNA can be a possible source of G-C bias[64]. Additional exploration of frequencies for each nucleotide can further help to assess bias in the samples[249]. For the first cycles, a strong random fluctuation in each nucleotide frequency is expected due to the randomness of hexamer priming used for the reverse transcription before each frequency stabilises at around 0.25[313]. During the age-OME project, we also kept track of null values (N) returned by the sequencer when no nucleotide could be attributed in a read cycle. N frequencies are expected to be close to 0 across all the cycles[147,244]. Further summary statistics were also reported for each sequencing conducted in each cohort. Additional summary statistics on Phred quality scores and other relevant metrics were also calculated and reported during this project. Once the quality of the reads is confirmed and relevant data pre-processing has been carried out, the resulting sequences are mapped to the human genome, and a gene count is performed. Amongst several publicly available human genomes, hg38 appears to be the most indicated for aligning the sequences and identifying genes[626].

In light of the aims of the age-OME project, the untargeted transcriptome-wide investigation would offer an unbiased insight into gene expression of the pre-mortem human left ventricle tissue. Transcriptomics analysis can be conducted using sequencing technologies (see Chapter 5 for a brief description of the experimental design).

4.5 Transforming omics measurements

As mentioned earlier, untargeted or semi-targeted MS measurements and RNA sequencing quantify ion or gene counts, which are proportional to the analyte or gene abundance in the tissue. Consequently, the omics measurements are non-negative and usually right-skewed distributed[907]. Mapping omics measurements onto a new domain can enhance its interpretability[462], help to match the assumptions of statistical methods[76], and appropriately tackle possible heteroscedasticity in the data[34]. However, a lack of care in transforming the data can lead to bias in the statistical and mathematical results[457]. Additionally, the choice of function for remapping omics measurements should not be determined based on the significance level of the statistical test, a practice known as p-hacking[325]. In high-throughput biomedical studies, data transformation represents an essential step in preceding any statistical and mathematical analysis[687,841]. A data transformation is defined as a function f that maps the i-th analyte's measure detected in the j-th sample, $\tilde{\mathbf{Z}}_{i,j} \geq 0$, to a new value, $\mathbf{Z}_{i,j} \in \mathbf{Z}$, $\forall i, j$. The data transformation can be linear if $\mathbf{Z}_{i,j} = f(\tilde{\mathbf{Z}}_{i,j}) = a + b \cdot \tilde{\mathbf{Z}}_{i,j}$ with $a, b \in \mathbb{R}$ or a non-linear function[238]. This section, together with Chapter 12.1, briefly introduces some of the transformations that can be used for MS and sequencing data, their applications in omics analysis and the rationale for $\log_2$ for the age-OME project.

Logarithm transformation The logarithm transformation belongs to the family of non-linear transformations, and $\log_2$ is most commonly applied for biomedical[726] and omics

data[494,498]. The choice of the base 2 emphasises the interest in detecting when a molecule or gene doubles or quadruplicates its abundance[928]. The $\log_2$ maps the non-negative omics measurements onto the whole real line. Additionally, any follow-up statistical method that relies on averages is applied to the geometric average rather than the arithmetic average, and consequently, the estimate is less susceptible to extreme values due to its mathematical properties[726]. The transformed $\log_2$ omics measurements read

$$\mathbf{Z}_{i,j} = \log_2\left(\tilde{\mathbf{Z}}_{i,j}\right), \tag{4.1}$$

for each i-th omics measurements in j-th sample. The $\log_2$ transformation has some compelling properties. When two averages are compared, their difference is the ratio of the two averages, also known as the log fold change. The log fold change is a simple quantity to interpret; since the logarithm is in base 2, a log fold change equal to 1 indicates that the average of the omics measurement in the first average doubles with respect to the second average. Additionally, the logarithm transformation works element-wise, which handles well any missing values returned by the MS analysis.

There are some points to be considered when dealing with sequencing data. RNA libraries usually present several zeros, and applying the logarithm transformation without taking them into account might fall into an inappropriate zero-inflation[755]. The size of libraries should also be considered when dealing with RNA sequencing, inasmuch as gene counts can differ significantly for each sample, from hundreds to thousands[653]. Consequently, the count of transcripts attributed to a specific gene is highly related to the size of the library[653]. One possible approach is to divide each gene count by the total size of the library (expressed in millions), which yields the count per million for each gene[464] and allows comparisons among libraries. Specifically, the vector of rescaling factors $\mathbf{r}$ for the j-th sample (or library) is $\mathbf{r}_j = \sum_i \tilde{\mathbf{Z}}_{i,j}$. Successively, the $\log_2$ count per million of the i-th analyte in the j-th sample reads

$$\mathbf{Z}_{i,j} = \log_2\left(\frac{\tilde{\mathbf{Z}}_{i,j} + 1/2}{\mathbf{r}_j + 1} \cdot 10^6\right). \tag{4.2}$$

The constant 1 in the calculation of the library ensures that normalised counts are positive and less than the unit. The 10^6 allows the count to be expressed in millions. The logarithm-transformed omics measurements also help to map the non-negative onto the whole real line, reduce possible skewness and heteroscedasticity in the data and help the sample-wise and element-wise distribution to look bell-shaped, which would increase the precision of statistical methods that require data to be normally distributed[238,726]. A limitation of the logarithm transformation is its incapacity of handling the value 0, which can occur with MS when the amount of ion received from the detector is below the detection level. Yet, in MS, those values are treated as missing values[808], which are further discussed in Section 4.7. For example, a simple adjusted logarithm transformation $\mathbf{Z}_{i,j} = \log_2(\tilde{\mathbf{Z}}_{i,j} + 0.5)$ would allow the transformation to handle zero values in high-throughput data, but differences in library sizes would not be accounted in RNA sequencing data, as discussed previously[909]. While extensions of logarithm transformations, such as variance-stabilizing transformation[206,363], have been introduced in the literature, the interpretation of the results can be compromised due to the assumptions that the method imposes.

The $\log_2$ was used in the age-OME process in order to facilitate interpretation of the follow-up results of the logarithm transformation and mitigate heteroscedasticity and skewness of the data without concerning whether to apply the sample- or analyte-wise transformation. Details on the experimental design for each omics are summarised in Chapter 5.

4.6 Normalising omics measurements

Sample storage, sample processing and analytical platforms can introduce different sources of variations in omics measurements, known as systematic bias or unwanted variation[254]. Although a strict sample processing protocol can mitigate systematic bias, it cannot completely remove it, which undermines reproducibility[641]. Different normalisation methods have been proposed to reduce unwanted variation, each with its advantages and disadvantages. This section, together with Chapter 12.1, briefly introduces some of the main normalisation methods in the omics field, followed by the rationale behind the choice of EigenMS normalisation for the age-OME study. While it was created originally for proteomics[407], it was later adapted for metabolomics[405] and lipidomics studies[463]. Normalisation methods reviewed in this section can be applied to all omics[404] although, for protein, I rely on the normalisation integrated into the workflow offered by the DIA-NN software suite[183]. Based on the previous section, I denote with $\mathbf{Z}_{i,j} \in \mathbb{R}$ the transformed i-th omics measurement in the j-th sample throughout this section. In the last part of this section, I introduce the normalisation of RNA sequencing gene counts.

EigenMS normalisation The approaches discussed above assume that each sample's measurement is affected by systematic bias that can be deflated by multiplying omics measurements by a normalisation constant. An alternative approach, initially introduced in genomics studies, addresses this issue of systematic bias[465] by using residuals of generalised linear models to estimate and detract systematic bias from the transformed omics measurements $\tilde{\mathbf{Y}}$. EigenMS normalisation[405] applies a similar idea to spectrometry experiments, including metabolomics and lipidomics[405,953]. EigenMS starts with constructing the design matrix, $\mathbf{X}$, which carries the pairwise comparison between two or more cohorts[326] (see Section 4.8). The design $\mathbf{X}$ has a row for each sample in the omics data, and one column for each cohort that is to be estimated. For example, the age-OME study requires two columns, one for the young and one for the old cohort. Then, the following linear model would describe the relation between the analyte expression and each i-th omics measurement,

$$\tilde{\mathbf{Y}}_{i,\cdot}^T = \mathbf{X}\beta + \epsilon_i, \tag{4.3}$$

where ϵ_i is the error term associated with the i-th analyte. The coefficients $\hat{\beta}$ are estimated by solving

$$\hat{\beta} = \arg\min_{\beta} \|\tilde{\mathbf{Y}}_{i,\cdot}^T - \mathbf{X}\beta\|_2^2, \tag{4.4}$$

and then used to obtain fitted values $\hat{\mathbf{Y}}_{i,\cdot}^T = \mathbf{X}\hat{\beta}, \forall i$. In the equation above, $\|\cdot\|$ indicates Euclidean norm [i]. The fitted values are used to calculate the residual matrix $\mathbf{R}$, where its i-th row reads $\mathbf{R}_{i,\cdot} = \tilde{\mathbf{Y}}_{i,\cdot} - \hat{\mathbf{Y}}_{i,\cdot}$. In other words, the i-th row of the residual matrix contains the residual of the i-th omics measurements. Next, singular value decomposition[13] is applied to the residual matrix, leading to $\mathbf{R} = \mathbf{U}\mathbf{D}\mathbf{V}^T$. Hereafter, only a few eigenvectors are kept for the follow-up analysis and they are contained in the matrix $\mathbf{V}^*$. The eigenvector selection is based on significance analysis conducted on the associated eigenvalues. The analysis involves the bootstrap-based statistical test[405,544]. Subsequently, the following multivariate linear model represents the analyte-specific bias,

$$\mathbf{R} = \mathbf{B}\mathbf{V}^* + \mathbf{E}, \tag{4.5}$$

where $\mathbf{B}$ is the matrix of unknown coefficients and $\mathbf{E}$ is the error matrix. The coefficients are then estimated by solving

$$\hat{\mathbf{B}} = \arg\max_{\mathbf{B}} \|\mathbf{R} - \mathbf{B}\mathbf{V}^*\|_F^2, \tag{4.6}$$

where $\|\cdot\|_F$ is the Frobenius norm [ii]. The estimated coefficient $\hat{\mathbf{B}}$ is used to compute the fitted residual values $\hat{\mathbf{R}} = \hat{\mathbf{B}}\mathbf{V}^*$. The normalised omics measurements are computed as

$$\mathbf{Y} = \tilde{\mathbf{Y}} - \hat{\mathbf{R}}. \tag{4.7}$$

EigenMS proposes a distinct and effective approach to removing unnecessary variations. The method relies on singular value decomposition of the residual matrix to first construct surrogate variables[544] and then select a few based on significance analysis[405]. EigenMS has been widely employed in omics data, and computational evidence has shown its competitive performance over other normalisation methods[840]. EigenMS was included in the experimental protocol to mitigate the systematic variation in metabolomics and lipidomics.

Variance modelling normalisation Variance modelling normalisation, also known as voom, is a particular framework that has been developed for transcriptomics data[460]. In Section 4.5, the $\log_2$ count per million ($\mathbf{Z}_{i,j}$) of the count data ($\tilde{\mathbf{Z}}_{i,j}$) was defined in Equation 4.2. The model voom functions based on the understanding that there is a trend between $\log_2$ transformed raw count ($\tilde{\mathbf{Z}}_{i,j}$) and residual standard deviation after a linear model is used to conduct a statistical study. In detail, let $\mathbf{X}$ be the design matrix for the following statistical model that describes the i-th $\log_2$ gene count per million across all the samples. The statistical model reads

$$\mathbf{Z}_{i,\cdot}^T = \mathbf{X}\beta_i + \epsilon_i, \tag{4.8}$$

where ϵ_i is the error term. The unknown coefficients β_i are estimated by solving

$$\hat{\beta}_i = \arg\min_{\beta_\mathbf{i}} \|\mathbf{Z}_{i,\cdot}^T - \mathbf{X}\beta_i\|_2^2, \tag{4.9}$$

and then the fitted values are calculated as $\hat{\mathbf{Z}}_{i,\cdot}^T = \mathbf{X}\hat{\beta}_i$. Then, the residual variance for the i-th gene is computed as $\sigma_i^2 = \|\hat{\mathbf{Z}}_{i,\cdot} - \mathbf{Z}_{i,\cdot}\|_2^2/(N-2)$. Equation 4.2 is then employed to

[i] Euclidean norm for a vector $\mathbf{a}$ is $\|\mathbf{a}\|_2 = \sqrt{\sum_i \mathbf{a}_i^2}$.

[ii] The Frobenius norm of a matrix $\mathbf{A}$ is defined as $\|\mathbf{A}\|_F = \sqrt{\sum_{i,j} A_{i,j}^2}$.

calculate the predicted $\log_2$ raw count $\log_2(\tilde{\mathbf{Z}}_{i,j})$, which can be approximated as

$$\log_2(\hat{\tilde{\mathbf{Z}}}_{i,j}) = \hat{\mathbf{Z}_{i,j}} + \log_2(\mathbf{r}_j + 1) + 6\log_2(10). \tag{4.10}$$

In a similar way, the average of the predicted $\log_2$ raw counts for each i-th gene denoted by $\log_2\left(\bar{\hat{\tilde{\mathbf{Z}}}}_i\right)$ is obtained by substituting $\mathbf{Z}_{i,j}$ in Equation 4.2 with the average of the predicted values $\hat{\mathbf{Z}}_{i,\cdot}$ for each i-th gene, denoted by $\bar{\hat{\mathbf{Z}}}_i$, and $\mathbf{r}_j$ with the geometric average of the J library size, that is $\bar{r} = (\prod_j \mathbf{r}_j)^{1/J}$. The equation reads

$$\log_2\left(\bar{\hat{\tilde{\mathbf{Z}}}}_{i,j}\right) = \hat{\mathbf{Z}}_{i,j} + \log_2(\bar{r} + 1) + 6\log_2(10). \tag{4.11}$$

Similarly to Equation 12.6, loess function is applied to approximate the relation between the square root of the residual variance $\sqrt{\sigma_i}$ of each i-th gene and the $\log_2$ average of predicted raw gene count $\log_2\left(\bar{\hat{\tilde{\mathbf{Z}}}}_{i,j}\right)$. The square root transformation is applied to residual standard deviations to promote symmetricity and generate a distribution that is close to Gaussian. Loess function is trained to the $\log_2(\hat{\tilde{\mathbf{Z}}}_{i,j})$ so as to predict the residual standard deviations $\sqrt{\sigma_i}$. It is subsequently applied on the prediction from Equation 4.10 to estimate the weight for each $\log_2$ count per million, $\mathbf{Z}^T_{i,\cdot}$. The rescalled $\log_2$ count per million by its weight is denoted as $\mathbf{Y}_{i,j}$ (Further details of voom can be found in Law et al. (2014)). This age-OME project has adopted voom as a normalisation method for RNA sequencing data due to recent critics raised on other competitive techniques[320]. Additionally, voom has allowed us to use a very similar framework for identifying the molecular signature and simplify the interpretation of the results.

In this section, I have briefly discussed normalisation methods that help to reduce the unwanted variation in omics data, which supports our rationale for applying EigenMS in the metabolomics[405] and lipidomics[953] analyses and voom for the transcript reads. Regarding the proteomics analyses, normalisation was performed directly by Dr Mark Lawrence's lab, which regularly uses DIA-NN suite[183].

4.7 Handling missing measurements

Whereas high-throughput analytical techniques allow the screening of multiple analytes, some measurements can be missing. In RNA sequencing data, missing data are generally attributed to the low abundance of the specific transcript, and the sequencer returns a zero when the transcript could be attributed to a specific gene. Certainly, many zeros in the gene count matrix can cause issues such as zero-inflation and impact the follow-up mathematical and statistical analysis[389,755]. Zero-inflation can be mitigated by simply removing the genes that exhibit a high number of zeros across samples[189,755]. In bottom-up proteomics, missing data can occur for several reasons, including the peptide abundance below the detection threshold and the ion suppression between two competitive precursors[755]. Like proteomics, missing data can occur in untargeted and semi-targeted metabolomics due to ion suppression, gradual contamination of the chromatographic platform, ions below the sensitivity of

the spectrometer, matrix effect, and incorrect processing of the spectra[118,673,872]. Lipidomics analyses also produce missingness in lipids measurements. Missing values in lipids data can be due to similar factors listed above for proteomics and metabolomics; indeed, possible mechanics that generate missingness in the lipids abundances are ion suppression, contamination or poor quality of the peak in the spectra[609]. This section introduces the generative mechanism of missing values and common, albeit not exhaustive, methods to handle missing data, including the k-nearest neighbour algorithm that has been used for the age-OME omics. I also explain how missing data have been handled for the follow-up analyses in the ageing heart project.

Correctly handling missing data can enhance the interpretation of the data and unlock the use of several multivariate methods that cannot process data with missing data[438]. As generative causes of missing values can vary, scholarship has identified three possible mechanisms to explain and distinguish the missing measurements[211,704]. First, missing completely at random (MCAR) values occur when an analyte intensity is missing due to unknown and unpredictable reasons, such as a glitch of the instrument or poor sample ionisation. The second generative mechanism is missing at random (MAR); in the context of MS, MAR values occur when an analyte in a sample is not detected under specific conditions. MAR assumes that missing values are present due to a specific set of experimental variables, and if known, missing values can be predicted[404,874]. The third generative mechanism of missing data is missing not at random (MNAR). MNAR depends on the experimental protocol or the analytical platform employed; for instance, MNAR values are brought about when a certain molecule is below the sensitivity threshold of the MS and cannot be detected[404]. Although the MNAR generative mechanism is the least desirable option, computational experiments have shown that a small portion of MNAR does not compromise follow-up statistical analyses[285]. The omics measurements can contain multiple mechanisms that generate missing data due to the complexity of the sample matrix, sample processing protocols and limitations in the analytical platforms[286,476]. MCAR and MAR generative mechanisms of missing values are most prevalent in MS datasets[874].

Depending on the mechanism that generates missing values, different methods can be employed to predict and impute missing intensities in the MS[94]. With the increasing use of high-throughput technologies, researchers and practitioners have explored missing values and created new data-driven algorithms for imputation purposes of MAR and MCAR values. Despite such new statistical learning algorithms, such as random forest, support vector machine and neural network, it seems that traditional and less-performing methods, including median imputation, mean imputation and zero imputation, are yet favoured in bioinformatics due to their simplicity and low computational cost; however, these methods increase risk of creating bias in follow-up analyses as they fail to take into account the uncertainty behind missing values[91,198,495,873]. For more robust analyses, multiple imputation algorithms, such as Multiple Imputation by Chained Equations (MICE), should be applied to generate several sets of values for missing data points, which return multiple parallel datasets[58]. Consequently, one can analyse each generated dataset and combine the results. The combination of parallel analyses takes into consideration the uncertainty of the unknown value and ensures more consistent results than the pointwise imputation[198,779]. Another common approach to inferring missing data in computational biology is the k-nearest

neighbours (kNN) algorithm. A large body of literature has shown the reliability of the kNN for inferring missing data in MS[67,130,196,347]. A brief description of each method and its advantages and pitfalls that can derive from their applications are presented in Chapter 12.3, while a brief description the KNN is presented here. The notation is similar to the one employed in Section 4.6; the normalised data are presented in a matrix form $\mathbf{Y}_{i,j}$, where samples are listed on the columns and analytes on rows.

kNN imputation kNN imputation is a popular technique for inferring missing data in omics experiments. It was originally introduced to handle missing values in microarray[830] and later applied to MS data[347]. kNN allows the researcher to obtain a point estimate for the missing value present in the transformed and normalised data matrix $\mathbf{Y}$. Assuming that the i-th analyte in the j-th sample has a missing value, the imputed value of kNN reads

$$\mathbf{Y}_{i,j} = \sum_{i' \in I_k} \mathbf{Y}_{i',j} \cdot w_{i'} \tag{4.12}$$

where I_k is the set of the k closest analytes' measurements in sample j. The closest analytes' measurements are identified by calculating Euclidean distance among all analytes in all samples, except for j. The weight $w_{i'}$ is the rescaled Euclidean distance between the analyte $i' \in I_k$ and the analyte i-th across all the samples except j[347,830]. Despite being a point-estimate method, kNN has shown robust estimation of MCAR and MAR values in MS data[67,196,347]. Further experiments have also suggested that a neighbour equal to 10 is the best choice for data imputation[196]. kNN has been used for the cardiac ageing project.

In the cardiac ageing project, the first step to mitigating missing values is to reduce the number of analytes that exhibit too much sparsity. This was achieved by excluding samples that had more than 25% of missing values in the younger and older cohorts. In other words, only analytes measured in at least 75% of omics measurements in each cohort were considered for the analyses. The cut-off for the missing observations was obtained by the sample size calculations for each omics (see Chapter 5). The imputed data matrix, $\mathbf{Y}_{i,j}^{imp}$ was then used only for conducting principal component analysis (see Chapter 5) and the heat map (see Chapters 5.7 and 6).

4.8 Identification of candidate biomarkers

Identifying biomarkers is a pivotal step for clinical and medical research in general and for the age-OME project[677]. The indication of biomarkers is also known as differential abundance/expression analysis[362], which aims to identify key molecules amongst multiple cohorts that are up or down-regulated with respect to a baseline. For the human cardiac ageing project, the young cohort was set as a baseline, to which the older cohort of donors was compared (see Chapter 5). Since this project focuses on two cohorts, this section introduces some methods that help identify omics biomarkers by contrasting abundances of analytes detected in two sample groups. The section further introduces linear models and empirical Bayes methods, also known as limma or moderated t-test[764], employed in the age-OME project for detecting biomarkers. Throughout the section, I assume that the data matrix $\mathbf{Y}$ is log_2 transformed and appropriately normalised if not stated differently (see

Sections 4.5 and 4.6).

Given two groups of observation, namely $\mathcal{G}_A$ and $\mathcal{G}_B$, differential expression analysis aims to test the null hypothesis that there is no tendency for the observation of $\mathcal{G}_A$ to be different from the observation of $\mathcal{G}_B$[677]. The differences between two sample cohorts can be summarised with several metrics, such as average difference, which would eventually make the hypothesis test well-defined[116]. Several methods have been developed to conduct differential expression analysis between two cohorts[200,586]. One basic approach is to calculate the average difference between the two cohorts in the data matrix $\mathbf{Y}$ to calculate the $\log_2$ fold change ($\log_2 FC$) of each analyte and identify it as differentially expressed analytes exhibiting the absolute value of the $\log_2 FC$ value greater or equal to 2. This method allows straightforward implementation and interpretation and has been extensively applied in biosciences. Nevertheless, it lacks a solid probabilistic ground, and its cut-off value does not take into account that some analytes cannot exhibit such high variation, as it could be incompatible with the living conditions of the model used for the experiment[170,586].

Alternatively to using the $\log_2 FC$ for the biomarker identification, approaches grounded onprobability theory offer more powerful tools to investigate differentially expressed analytes between two conditions[17]. The t-test, similar to ANOVA with two groups, is designed to test whether the mean difference between two samples is statistically relevant or not[572]. Despite its wide use in literature, nevertheless, t-statistics and ANOVA exhibit poor sensitivity when applied to small sample sizes or unbalanced experimental design[586], common issues in omics experiments[340,418]. A permutation test can be a useful alternative, although the appropriate analytical framework should always be preferred when possible[418]. Moreover, ANOVA and t-tests rely on the assumption that the two compared groups exhibit a common pooled standard error. This assumption can be matched by applying proper data transformation and normalisation methods (see Sections 4.5 and 4.6). This assumption can be circumnavigated by employing Welch's t-test instead, a statistical test that allows two mean sampling distributions to have different variances[116,381]. Like t-test, Welch's t-test has low sensitivity in small sample sizes, which may not be suitable for omics experiments[381]. Extensions of the t-test have been proposed to overtake its limitations. For example, reproducibility-optimised sample test statistic (ROTS) decreases the pooled variability by rescaling the standard error of the test[217,218,788]. Evidence suggests that ROTS addresses the limitation of the t-test and shows appealing performances compared to the class of moderated t-test[28], although recent study has confirmed that moderated t-test, such as limma, exhibits better sensitivity and specificity[200].

A non-parametric alternative to Welch's and t-test is represented by the Mann-Whitney-Wilcoxon U-test[120]. Designed for testing ordinal data, the U-test has been applied to quantitative data and offers a distribution-free approach for testing whether there is a difference between two samples. The idea behind the U-test lies in ranking the observation across two groups, $\mathcal{G}_A$ and $\mathcal{G}_B$, and then summing up the number of observations in group $\mathcal{G}_B$ that proceed the one in group $\mathcal{G}_A$. Once the observed statistic U is computed, the p-value is obtained by permutating observations. In literature, the U-test has also been routinely utilised for testing the median difference between two samples[316]. In omics analyses, the U-test has shown equivalent performance to the parametric t-test[120] without providing

any additional advantage over it[381] but offered better reliability in biomarker detection than negative binomial models[518,686] that are popular in transcriptomics analyses[483]. While some extensions on the U-test have also been proposed[308], further studies are needed to explore their sensitivity in omics data. A competitive test to the U-test and t-test, which aims to test the median difference rather than the average, is offered by local pooled error z-test[379], but it does not demonstrate better sensitivity than the t-test for biomarker discovery[586]. Significance analysis of microarray (SAM) can be a non-parametric option to the t-test[836]. SAM framework assumes that a random fluctuation exists across genes, which should account for the standard error when the mean difference of an analyte is tested. SAM estimates the p-value through a permutation test and displays a low false discovery rate and low misidentification of the true biomarker. This method can be impractical for large omics investigations due to the large computational efforts it requires[836,888].

To overcome the limitation of computational costs required by non-parametric tests, semi-parametric approaches can be applied for differential expression analysis. For example, semi-parametric differential abundance analysis can be regarded as a three-stage method that handles normalised MS data without any transformations. The first stage applies a logistic regression for zero values in the data matrix, while the second stage applies regression for non-zero analytes in the data. Consequently, a kernel-smoothed likelihood is applied to estimate the model coefficients and related standard errors, which enables the detection of possible differentially abundant analytes. However, this framework is only limited to MS data[482].

Another class of methods relies on mixture model frameworks yet shows concerns when distributional assumptions are violated[278]. Extensions of the mixture model have also been introduced in literature to address the instability of the mixture models[278] by using a shrinkage variance method to obtain a robust estimation of the variability of each metabolite[361]. Yet, further computational studies are required to ensure the reliability of identified biomarkers. A geometrical approach to revealing differentially expressed analytes is offered by the characteristic direction. Based on naive Bayesian classifiers, the characteristic direction takes into account not only a single expression level but also linear interactions across analytes. This method seems to outperform other available methods, although its interpretation is not straightforward[142].

To overcome the issue with the test performance in small sample size, a new class of estimators has been introduced by fusing the Bayesian statistical framework with the frequentist one. The approach estimates the mean difference based on the frequentist approach, but the pooled variance is partially estimated through a hierarchical model[513]. The moderated t-test was introduced by Lönnstedt and Speed (2002), who utilised a hierarchical model to obtain the log posterior odds statistics and conduct significance analysis on the observed odds[513]. Several extensions have been proposed based on this framework, such as random variance models known as the BRB t-test. BRB t-test uses the linear regression framework with random errors with a zero mean and variance distributed as the inverse gamma. This particular method appears to be robust in both small and large samples[887]. CyberT t-test is another example of moderated t-tests. CyberT t-test aims to regularise the pooled variance by performing a weighted average between a

background variance estimated by the Bayesian framework and the observed variance[48,412]. In terms of moderated t-test, Smyth (2004) have established the golden standard with the limma framework[764]. The limma t-test extends the work of[513] ([513]) on a framework that works closely to generalised linear models and enables researchers to obtain a t-test with higher statistical power than the standard t-test. We have utilised the limma for the age-OME project, inasmuch as it has displayed outstanding performance in small and large sample sizes across different systematic reviews as described later in this section[200,381,586]. Extensions of the limma framework have also been elaborated[358,734,951]; for instance, when investigating differential expression analysis with a magnitude of $\log_2 FC$ above a certain threshold, TREAT method can be employed to test hypothesis appropriately[552]. Furthermore, a moderated Welch's t-test has been proposed, whereas it is less versatile than the moderated t-test, as it cannot accommodate a more complex experimental design[184,616].

Limma t-test In this section, I briefly introduce the limma t-test, a moderated t-test employed in the human ageing heart project to detect the molecular signatures of ageing in human cardiac tissues (see Smyth (2004) for more details). For the age-OME project, the data matrix, $\mathbf{Y}$, contains on the i-th row the abundance of the i-th analyte across all the samples (columns of the matrix). The samples can belong to the young or the old cohort of donors, conveniently denoted with $\mathcal{G}_y$ and $\mathcal{G}_o$. To derive the limma t-test, we assume that each i-th analyte, $\mathbf{Y}_{i,\cdot}^T$ expression has a linear relation with a design matrix $\mathbf{X}$, which contains two columns; the first one containing all ones and the second columns with the ones in correspondence of $\mathbf{Y}_{i,j} \in \mathcal{G}_o$ and zero otherwise. Then linear relationship between $\mathbf{Y}_{i,\cdot}$ and $\mathbf{X}$ is quantified through the following link:

$$\mathbb{E}[\mathbf{Y}_i] = \mathbf{X}\beta^i, \tag{4.13}$$

where β^i is the vector containing two unknown coefficients and $\mathbb{E}[\mathbf{Y}_i]$ denotes the expected value of random variable $\mathbf{Y}_i$ that describes measurements of the i-th analyte. Additionally, the key point of the limma framework relies on the assumption related to the variance of $\mathbf{Y_i}$, here indicated by $\mathbb{V}[\mathbf{Y}_i]$, which reads

$$\mathbb{V}[\mathbf{Y}_i] = \sigma_i^2 / \mathbf{W}_{i,j}. \tag{4.14}$$

The $\mathbf{W}_{i,j}$ are weights introduced in order to address heteroscedasticity within each observation of the i-th analyte, and σ_i^2 is the unknown variance of the i-th analyte. In general, for MS data, weights are set to the unit and usually used for RNA sequencing data, as mentioned earlier in this chapter (see variance modelling normalisation in Section 4.6). Coefficients $\beta^{\mathbf{i}}$ can be estimated by solving the weighted least square problem[692]:

$$\hat{\beta}^i = (\mathbf{X}^T \mathbf{D}^i \mathbf{X})^{-1} \mathbf{X}^T \mathbf{D}^i \mathbf{Y}_{i,\cdot}^T, \tag{4.15}$$

where $\mathbf{D}^i$ is a diagonal matrix, where each element of the diagonal contains the weight of the i-th measurements corresponding to the j-th sample $\mathbf{D}^i_{j,j} = \mathbf{W}_{i,j}$. Without generality, we assume that the i-th analyte is measured for all the J samples; consequently, the residual variance of the linear model for the i-th analyte (see Equations 4.13 and 4.14) is estimated as

$$\hat{s}_i^2 = \frac{1}{J-2} \Big(\mathbf{Y}_{i,\cdot}^T - \mathbf{X}\hat{\beta}^i \Big)^T \Big(\mathbf{Y}_{i,\cdot}^T - \mathbf{X}\hat{\beta}^i \Big), \tag{4.16}$$

where $J-2$ is also the degree of freedom d_i of the Student's t-test applied so as to test whether the slope in Equation 4.13, namely $\hat{\beta}_2^i$, is different from zero. Here, since we are interested in detecting if the analyte changes in the older cohort $\mathcal{G}_o$, we focus on the slope. The limma framework assumes that the conditional distribution of $d_i s_i^2 | \sigma_i^2 \sim \sigma_i^2 \chi_{d_i}^2$, where s_i^2 represents the estimator of s_i^2 and $\chi_{d_i}^2$ indicates the chi-squared distribution with d_i degree of freedom[116]. Additionally, the limma sets the prior distribution for σ_i^{-2} to be a $\chi_{d_0}^2$ distribution with d_0 degree of freedom multiplied by $s_0^2 d_0$; the prior distribution reads $\sigma_i^{-2} \sim \frac{\chi_{d_0}}{s_0^2 d_0}$. The application of the Bayes theorem[116] leads to the posterior distribution:

$$\sigma_i^{-2} | \tilde{s}_i^2 \sim \frac{\chi_{d_0+d_i}^2}{s_0^2 d_0 + d_i s_i^2}. \tag{4.17}$$

Using the property of linearity of the expected value[116], the posterior expectation of $\sigma_i^{-2} | \tilde{s}_i^2$ is

$$\begin{aligned} \mathbb{E}\left[\sigma_i^{-2} | \tilde{s}_i^2\right] &= \frac{d_0 + d_i}{(s_0^2 d_0 + d_i s_i^2)} \\ &=: \frac{1}{\tilde{s}^2}, \end{aligned} \tag{4.18}$$

where empirical Bayes is used to estimate d_0 and s_0^2[115]. Consequently, the observed limma t-statistic for the i-th analyte reads

$$t_i = \frac{\hat{\beta}_2^i}{\sqrt{\tilde{s}_i^2 v_2}}, \tag{4.19}$$

where v_2 is the second element on the diagonal of projection matrix $(\mathbf{X}^T \mathbf{D}^i \mathbf{X})^{-1}$ (see Equation 4.15). Under the null hypothesis that $\beta_2 = 0$, the statistic distributes as a Student's t with $d_0 + d_i$ degree of freedom[764]. Finally, the $\log_2 FC$ of the i-th analyte is calculated by summing up the coefficients $\hat{\beta}^i$, that is $\log_2 FC_i = \hat{\beta}_1^i + \hat{\beta}_2^i$.

This section has discussed the role of the molecular signature in the omics data and different statistical frameworks to tackle this to provide a rationale for employing the limma in the computational protocol for the age-OME project [iii]. While the limma t-test is a hybrid approach between frequentist and Bayesian statistics, multiple testing leads to exponential growth of the type I error. To mitigate the false rate of discovery, the p-values have been adjusted with the Benjamini-Hochberg procedure[63].

4.9 Co-expression network

Identifying molecular biomarkers elucidates possible genotypes and phenotype differences between two groups or conditions. While the molecular signature conducted for each omic layer can offer important insights, it cannot measure if the two changes occur independently or coordinated. Enrichment analysis can be conducted to address this issue (see Section 4.10), yet its framework relies on prior knowledge that can be incomplete or scattered across

[iii] Further details of the limma such as the calculation of odds can be found in Smyth (2004).

different databases[29,42]. Unlike molecular signature and enrichment analysis, co-expression analysis can offer a pure data-driven view of each omic layer and how analytes change in a coordinated manner and inform researchers in generating mechanistic hypotheses of the underlying phenomenon[671]. This section expounds on co-expression analysis and co-expression network in the context of the omics network. A brief description of the method applied in the age-OME project is discussed later in Section 4.9.1.

Co-expression analysis can be conveniently arranged as a network $\mathcal{G} = (\mathbf{V}, \mathbf{E})$, where $\mathbf{V}$ is the vector of nodes (analytes) and $\mathbf{E}$ is the vector of edges representing co-expression in a pair of analytes[824,927]. Co-expression network $\mathcal{G}$ is a convenient representation of the biological systems under study, where each node represents a molecule and each edge contains the strength and the direction of the co-variation between two molecules[550]. Co-expression network analysis helps to understand regulatory dynamics across omics measurements and has been extensively used in systems biology[71,816]. For instance, co-expression network on proteomics and transcriptomics data has been applied for inferring gene regulatory networks[250,661]. It has also been applied to other omics, such as metabolomics and lipidomics data, in order to detect changes across the metabolite-metabolite and lipid-lipid interactions in patients with acute ischaemic stroke[489].

Constructing a co-expression network begins with estimating the relation and the strength of the relation across each pair of analytes. Amongst several metrics that have been proposed for this scope, the most common metric used to estimate the edges between analytes is the Pearson correlation coefficient, which aims to detect linear relationships across a pair of analytes[455]. Partial correlation coefficient is an extension of the Pearson correlation coefficient, yet the conditioning variable needs to be specified for each pair of analytes[38]. There are some solutions derived from a penalised linear model to estimate partial correlation coefficient without prior knowledge of conditioning variables, such as elastic net[952], lasso[268] and Gaussian graphical models[444,716,879,880]. However, these approaches can be impractical for estimating large networks[386], inasmuch as model coefficients and hyperparameters are required to be estimated for each pair of analytes measured in the omic data. Alternatively, the mutual information[934] or biweight midcorrelation[768] can characterise non-linear changes across the analytes. Jahagirdar and Saccenti (2020) have investigated how different metrics affect the estimation of the metabolite-metabolite network inference and concluded that the Pearson correlation coefficient shows robust performance compared to other metrics[377]. In another study, they have employed in-silico models to generate a known network and revealed a superior performance of correlation metrics amongst several metrics[378]. This supports our choice of the Pearson's correlation coefficient for inferring the pairwise relation between analytes in the age-OME project.

A common issue of network reconstruction on pairwise correlation is the large number of inferred edges. Indeed, the initial network is a complete weighted network in which all the analytes are interconnected without a loop or multiple edges[333]. The strength of the connection is the estimated correlation; the complete co-expression network contains several spurious relations between analytes that need to be removed[93]. Several methods have been proposed in the literature to tackle this issue. The simplest method is to reduce the number of edges that exhibit strength less or equal to a chosen cut-off[475]. This practice,

albeit effective, relies on a cut-off that is chosen by the user, with subsequent introduction of subjectivity. A more rigorous approach can involve using statistical tests. For example, if the Pearson's correlation coefficient is used, it can be conveniently transformed using Fisher's z transformation. Fisher's z transformation is distributed as a Gaussian random variable under the null hypothesis that it is null[116]. Then, the network retains only the edge with the adjusted p-value below a threshold[71]. Some studies have also introduced a cut-off on the correlation simultaneously to the adjusted p-value in order to exclude edges with small magnitudes[603]. An alternative approach that does not require a cut-off on the coefficient and avoids the overuse of statistical tests is known as soft threshold[455,927]. It relies on the assumption that there is a linear relation between the degree of nodes [iv] and their probability in log_{10}-log_{10} scale. While this assumption was originally introduced to estimate the power-law coefficients, studies have pointed out that the estimates resulting from this procedure are not reliable[143] and it is extremely susceptible to model miss-specification as the linear regression can fit a wider range of degree distributions[493]. Lima-Mendez and Van Helden (2009) have also demonstrated that biological networks are not necessarily scale-free but exhibit different behaviours, although the relation between degree probability and degree in log_{10}-log_{10} scale might resemble a straight line[493]. Non-scale free distribution in biological networks was also confirmed in follow-up studies[95]. Although the soft threshold had initially intended to assume a scale-free distribution upon the network, studies[493] have concluded that this assumption is indeed less restrictive[455,927] and networks built with the soft threshold can exhibit a wider variety of behaviours than a scale-free network[493]. This ageing heart project has utilised the soft threshold. A brief description of the network reconstruction via soft threshold is presented in Section 4.9.1.

The reconstruction of a biological network for each omic measured in the age-OME study aimed to detect the largest community[242,277] of analytes as well as possible hubs[549], most connected nodes that the network presented. With the non-stationary biological network[289], identifying the largest community in each omic network allows us to explore analytes that exhibit simultaneous changes during the ageing process. This approach differs from molecular signature analysis since they are connected to each other in the largest community. For the ageing heart study, the most populated community was identified using walktrap[86,656], a random walk-based algorithm that takes into account the strength of connection across the analytes, summarised by the absolute value of pairwise correlations. Walktrap has been selected amongst other conventional methods in biological networks, such as hierarchical clustering[455], due to its well-rounded performance demonstrated in simulation studies[260,617]. Furthermore, functional enrichment analysis is performed on the largest community using annotated databases, such as WikiPathways[543]. Enrichment analysis is computed using the over-representation framework (see Section 4.10.1 for further details). Although it was originally available in the framework provided by WCNA software[455], enrichment analysis was limited to gene or protein data. For this project, we extended it to all the omics layers, including metabolomics and lipidomics.

[iv] In the undirected network, the degree of a node (or analyte) is the number of edges incident to it.

4.9.1 Construction of co-expression network

Co-expression analysis with a soft threshold enables a pure data-driven reconstruction of biological networks relying solely on the assumption that the relation between degrees of nodes and their probabilities, considered in $\log_{10}$-$\log_{10}$ scale, have a linear relation[143,455,927]. This reconstruction of the network is based on the algorithm introduced by Zhang and Horvath (2005). The first step of the network reconstruction with the soft threshold is calculating the correlation matrix across each pairwise analyte. The correlation is calculated using transformed and normalised omic measurements (see Sections 4.5 and 4.6), denoted by $\mathbf{Y}$. The pairwise correlation is calculated on complete observations. The correlation between i-th and the i'-th analyte for $i \neq i'$ is computed as

$$\mathbf{C}_{i,i'} = \frac{\sum_j (\mathbf{Y}_{i,j} - \mu_{i,\cdot})(\mathbf{Y}_{i',j} - \mu_{i',\cdot})}{\sqrt{\sum_k (\mathbf{Y}_{i,k} - \mu_{i,\cdot})^2 \sum_l (\mathbf{Y}_{i',l} - \mu_{i',\cdot})^2}}, \tag{4.20}$$

where $\mu_{i,\cdot}$ is the arithmetic average of the i-th analyte across all the samples. Once the correlation matrix $\mathbf{C}$ is finalised for all the analytes, the degree of i-th analyte ($\mathbf{d_i}$) is calculated. Subsequently, the ordered vector of degrees, $\mathbf{k} = [1, \ldots, \max(\mathbf{d})]^T$, is calculated with their probabilities $\mathbf{p} = [\mathbf{p}_1, \ldots, \mathbf{p}_{\max(\mathbf{d})}]^T$[550]. With the first network complete, all the nodes have the same degree. To save computational time, our algorithm introduces an initial cut-off τ on the absolute correlation.

In the second step of the algorithm, all the elements in the correlation matrix $\mathbf{C}$ that have an absolute value above$|\mathbf{C}_{i,i'}| > \tau$ are assigned the value 0, which stands for no connection between the two analytes.

In the third step, the degree vector $\mathbf{k}$ and related probability vector are updated on the new correlation matrix obtained in the previous stage. Then, the regression model

$$\log_{10} \mathbf{p} = \alpha + \beta \log_{10} \mathbf{k} + \epsilon \tag{4.21}$$

is employed to describe the relation between degrees and their probabilities, where ϵ is the error term. Denoting with $\hat{\alpha}$ and $\hat{\beta}$ the estimated α and β obtained by solving the least square, the coefficient of determination of the model, denoted with R^2, is calculated as

$$R^2 = 1 - \frac{\sum_i (\log_{10} \mathbf{p}_i - \hat{\alpha} + \hat{\beta} \mathbf{k}_i)^2}{\sum_j (\log_{10} \mathbf{p}_j - \log_{10} \bar{\mathbf{p}}))^2}, \tag{4.22}$$

where $\bar{\mathbf{p}}$ is the arithmetic average of $\log_{10}$ transformed probabilities.

Since the soft threshold procedure aims to identify a network with degrees and probabilities that exhibit linear behaviour in $\log_1 0$-$\log_{10}$ scale, the algorithm is terminated when the coefficient of determination R^2 is above 0.9[135]. When R^2 is below 0.9, the algorithm assigns $\tau = \min \mathbf{C}_{i,i'}$ and repeats the second and third steps. These steps are reiterated but with τ corresponding to the lowest correlation present in $\mathbf{C}$.

In this section, I have reviewed different strategies for constructing a pure data-driven omic network. The networks for the age-OME study have three aims; (i) to identify the

most connected analytes or hubs, potential regulators of the ageing process[71]; (ii) to find the largest community of analytes that co-vary during the ageing process[242]; and (iii) to infer the functional role of the community[455]. The biological network was constructed using the soft threshold method to avoid over-reliance on multiple testing and use of cut-off on the correlation metric[455,927]. The correlation metric was adopted to estimate the strength of connection of the pair of analytes with their effectiveness in reconstructing biological network[250,377,378,671]. Specifically, we used the Pearson's correlation. Over-representation analysis was used to investigate the functional role of the largest community (see Section 4.10.1). Results of this method for each omics are presented in Chapter 7.

4.10 Omics enrichment analysis

Differential expression analysis (see Section 4.8) provides insights into the molecular signature of a group compared to a control group and informs possible biological functions. By only looking at the molecular signature, the identification of dysregulation occurring in biological processes is left to the researcher, which inevitably introduces a certain degree of subjectivity. To avoid this pitfall in interpreting omics measurements, novel statistical tests, known as enrichment analysis, are designed to identify dysregulated pathways, cellular functions and molecular functions[814]. This section presents a succinct review of the most commonly applied enrichment methods and their extension for multi-omics measurements. I then present two frameworks applied to the cardiac ageing project in detail. The first method belongs to over-representation analysis (ORA) and attributes the most likely biological functions to a given set of molecules. As Chapter 7 explains, ORA is applied to detect the most likely biological functions that the largest connected community of proteins, metabolites, lipids and transcripts represent. The second method is a novel framework mapped during the age-OME project to detect dysregulated biological functions by using all the omics measurements available in the multi-omics investigation. The proposed method is referred to as "omics set enrichment analysis (omics SEA) in the thesis. Application of omics SEA to the multi-omics data collected for the human cardiac ageing study is presented in Chapter 8. Further application of omics SEA for detecting possible dysregulated lipids classes can be found in Section 6.

Pathway enrichment analysis helps to unveil the potential dysregulation that occurs at the biological systems level. Dysregulated biological functions are identified by integrating data into knowledge-based pathways. Biological pathways can be interpreted as a holistic characterisation of molecular reactions that fulfil a certain scope. For example, glycolysis is a metabolic pathway containing all the reactions to break down a molecule of glucose into pyruvate and/or lactate (see Section 2.3). Public databases, such as WikiPathways[543,762] and KEGG[400], provide access to a detailed list of biological pathways and related chemical reactions.

Biological pathways collected in databases can be used to test statistical hypotheses at the systems level. One of the common classes of algorithms used for identifying possible dysregulation at the pathway level is over-representation analysis (ORA). ORA, a statistical

method, tests if a hypothesis that a certain set of molecules belonging to a pathway occurs by chance[807]. ORA requires two kinds of information: the set of analytes in a pathway and the set of analytes that should be linked to biological functions. Since the set of analytes that are usually processed through ORA are the ones identified as differentially expressed (see Section 4.8), ORA can be regarded as an extension of differential expression analysis to the systems level. The ORA framework has been prevalent in research, and several methods have been created upon it[356,357,655]. While ORA provides a solid framework to examine whether a set of analytes belongs to a biological function, it heavily relies on the molecular signature of a group when compared to a baseline. The imposition of a cut-off on a metric, such as p-value and $\log_2 FC$ (see Section 4.8), always introduces a certain degree of subjectivity[51,585,869] that can affect the outcome of ORA and follow-up analysis.

Meanwhile, gene set enrichment analysis (GSEA), a class of pathway enrichment methods, does not require any prior selection of molecules as input and, therefore, overcomes the limitations of ORA. GSEA utilises all the available information in the data to examine changes that occur in biological processes[783,849]. The underlying statistics of GSEA can be calculated using different metrics, such as observed moderated t-statistics or log fold changes[898,954]. The GSEA framework has shown no evidence of false prediction and has indeed been promoted for identifying pivotal biological functions based on proteins and gene data[601]. However, since the GSEA test statistics lack analytical distribution, the sampling distribution for each pathway is calculated by permutating the measurements[642], which requires an extensive run time if multiple pathways are tested simultaneously. To reduce this computational burden, several variations of GSEA have been proposed, such as PAGE[422], GAGE[527] and Camera[889], for which analytical test statistics derives from more restrictive assumptions compared to the original GSEA formulation. Despite the analytical approximation of GSEA test statistics, the original GSEA has shown superior statistical performance[601]. To avoid imposing assumptions on GSEA, researchers have proposed more efficient sampling schemes. For example, the scheme based ondynamic programming algorithm has decreased the GSEA runtime[415]. Recent implementation based on an adaptive multi-level split Monte Carlo scheme, known as FGSEA, has decreased the computational complexity even further[437].

ORA and GSEA methods have been extensively applied in literature, but their use has been mainly confined to detecting possible pathway dysregulations by looking at only one omic layer per time, such as transcriptomics and proteomics[367]. The application of enrichment analysis to every single omics layer can be helpful, as is the case for O'Sullivan et al. (2024). They successfully applied pathway analysis to investigate changes of biological processes at lipids and metabolites levels in heart failure patients[20,620]. But, multi-omics investigation would generate unified insights on changes at the systems level. Some options have been proposed to link omics measurements to the analysis of biological processes[565,804]. MultiGSEA is one of the first methods that tackled multi-omics integration in the context of pathway enrichment[109]. This approach was then generalised to other methods, including ORA[382,387]. MultiGSEA independently tests a pathway on each omic layer, namely transcriptomics, proteomics and metabolomics, and combines all the p-values obtained from each test[109,922].

Whilst this novel approach benefits the integration of multi-omics measurements[492], it also entails some drawbacks[387]. The first shortcoming of multiGSEA is its exclusion of lipidomics data as it only allows proteins, transcripts and metabolites data[109]. The second limitation arises from a close inspection of the source code; multiGSEA does not consider the whole set of molecules in the pathways. In fact, the method only extrapolates proteins that interact with other proteins and metabolites that interact with other metabolites, leaving out proteins and metabolites that interact with each other from the list of analytes in the pathway[109]. Moreover, multiGSEA cannot calculate unique observed test statistics known as enrichment scores; the methods only return the p-values related to every single test and the combined p-value[857,922] and hide the magnitude of enrichment scores and its directionality in each omic[109,783]. The multiple testing for each omic layer becomes overwhelming when proteomics and transcriptomics are simultaneously included in multiGSEA. The method approaches enrichment analysis by testing the same set of analytes twice but neglects three aspects: (i) the central dogma of molecular biology that proteins are translated by transcripts[370]; (ii) proteins are mainly responsible for enzymatic reactions in biological pathways[139]; and (iii) the combined p-value might hide evidence of dysregulation in one of the omics layers. To tackle the drawbacks, we have introduced a novel enrichment analysis method named omics SEA. Omics SEA removes the multiple testing for each layer by choosing an appropriate metric that addresses the heterogeneity of different omics data and enables the calculation of a single test statistic for each pathway. Additionally, omics SEA also regards proteins as responsible for enzymatic reactions; consequently, when a protein and a transcript that belong to the same pathway are present in the study, the protein expression is preferred to drive enrichment analysis, according to the central dogma of molecular biology[370] (see Section 4.10.2 for further details of omics SEA).

4.10.1 Over-representation analyis

Over-representation analysis (ORA) aims to identify the most likely biological functions that a given set of molecules can fulfil. It hypothesises that the proportion of molecules belonging to a pathway is no higher than that of molecules that do not belong. In order to derive the test statistic of Fisher's exact test in the context of the pathway analysis, let m be the number of analytes belonging to the pathway that has been detected in omic measurements as well, and n be the molecules detected in our measurements but not belonging to the pathway. Additionally, we indicate with x the analytes that belong to the pathway and the list of analytes of interest, and with $k - x$ the analytes that belong to the pathway but not the list of analytes that we aim to attribute a biological process. Then, the observed test statistics reads

$$q = \frac{\binom{m}{x}\binom{n}{k-x}}{\binom{x+k}{k}}. \tag{4.23}$$

Under the null hypothesis, q_x has a hypergeometric distribution with $n + m$, n and k parameters. The hypothesis is tested on the right tail of the hypergeometric distribution. Since the hypergeometric test requires quite challenging calculations when multiple tests are performed, it can be stressed that, when $n + m$ is large, the hypergeometric distribution can be approximated with a binomial distribution[202]. Consequently, the observed test statistics can

be simplified as

$$q = \binom{k}{x} p^x (1-p)^{k-x}, \tag{4.24}$$

where $p = \frac{m}{n+m}$. Similarly to the hypergeometric test, this test is performed on the right tail of the binomial distribution. We adopted it as the underlying distribution for Fisher's exact test for the age-OME project since the binomial distribution requires fewer computational resources (see Chapter 7).

4.10.2 Omics SEA

Omics SEA is a non-parametric test[783], whose sampling distribution is built on the hypothesis that a given pathway does not show any evidence of enrichment and is approximated computationally by using the multilevel-split Monte Carlo scheme; this permutation scheme is also known as FGSEA[437]. While the multiGSEA[109] framework independently tests every single omics and combines the p-values[857] obtained by each test, our framework performs only one test for each pathway and obtains a metric for each pathway that describes the strength of its dysregulation. In order to compute omics SEA, the first step includes selecting a certain metric that appropriately addresses the issue of heterogeneity across the analytes measures if needed. For example, when this framework is applied to the sole lipid data, $\log_2 FC$ of each lipid molecule can be used, as large differences in variation of different lipid molecules are not expected (see Chapter 2.3)[388]. However, when dealing with molecules of different natures, the variability across them is expected to differ largely; for instance, metabolites can be expected to be more volatile with respect to proteins[638], and using an appropriate transformation of the omics measurements can address this issue (see Section 4.5). The observed moderated t-statistics, obtained by dividing $\log_2 FC$ of each analyte by its standard error, represents a suitable choice[954]. Indeed, the standard error of each $\log_2 FC$ creates a metric $\mathbf{m}$ that takes into account the biological variability as well as the sample size (see Section 4.8).

In order to construct the statistic for omics ESA, let us first denote with $\mathcal{S}$ the set of $N_{\mathcal{S}}$ analytes belonging to the pathway of interest, and with $\mathcal{A}$ the set of all omics' measurements arranged in decreasing order with respect to the metric $\mathbf{m}$. The number of analytes detected is denoted with $N_{\mathcal{A}}$[v].

Now, $N_{\mathcal{A}}$ subsets of $\mathcal{A}$ are generated; the i-th subset is denoted with $\mathcal{A}_i \subseteq \mathcal{A}$ and contains the first i analytes in $\mathcal{A}$. Once all the subsets $\{\mathcal{A}_i\}_{i=1}^{N_{\mathcal{A}}}$ are constructed, the enrichment score for the i-th subset is calculated as

$$E_i = \sum_{a \in \mathcal{A}_i} \frac{\delta_{\{a \in \mathcal{S} \cap \mathcal{A}_i\}}}{\sum_{a \in \mathcal{A}_i \cap S} m_a} \cdot |m_a| - \frac{\delta_{\{a \in \mathcal{A}_i - \mathcal{S}\}}}{N_{\mathcal{A}} - N_{\mathcal{S}}}, \tag{4.25}$$

where the $\delta_{\{\cdot\}}$ is the Kronecker delta and $|\cdot|$ denotes the absolute value. Once E_i is calculated for each $A_i \subset A$, two quantities are extrapolated. The first quantity represents the maximum

[v] More precisely, the pathway set ($\mathcal{S}$) is constructed by taking the intersection between the set of analytes that belong to the pathway and the analytes measured in the study. As a result, $N_{\mathcal{A}} \geq N_{\mathcal{S}}$.

value of the sequence of enrichment scores, namely,

$$E^+ = \max_{1 \leq i \leq N_{\mathcal{A}}} E_i, \tag{4.26}$$

which occurs in correspondence of $i^+ = \arg\max_{1 \leq i \leq N_{\mathcal{A}}} E_i$,. The second quantity is the lowest value of the enrichment scores, which reads

$$E^- = \min_{1 \leq i \leq N_{\mathcal{A}}} E_i, \tag{4.27}$$

occurring at position $i^- = \arg\min_{1 \leq i \leq N_{\mathcal{A}}} E_i$. Successively, the magnitude of E^+ and E^- are compared by taking their absolute values to determine the directionality of the pathway enrichment[437,783]. Without losing generality, we can assume that $|E^+| > |E^-|$. Consequently, the multilevel-split Monte Carlo scheme is employed to create permutation across the genes in order to create a sampling distribution for the enrichment score E^+. Once the sampling distribution is obtained, the p-value can be computed and interpreted. If the p-value suggests a dysregulation of the pathway, all the analytes in $\mathcal{A}_i^+$ are the cause of the dysregulation, which are usually called leading edges. A normalised enrichment score can be obtained by dividing E^+ by the average of the sampling distribution[783]. A similar procedure can be applied when $|E^-| > |E^+|$.

ORA and omics SEA allow researchers to connect omics measurements at the systems level. While the former requires a pre-selected set of molecules to perform enrichment analysis[403], the latter considers all the evidence in the omics and conducts enrichment analysis[783]. For the age-OME project, we have utilised both frameworks with different scopes. ORA was applied to identify the possible functions that the largest community in the omic network could represent (see Chapter 7). On the other hand, omics SEA was used to identify possible dysregulation of biological function and enrichment of lipids (see Chapter 8 and Chapter 6.3).

Conclusion

This chapter has introduced relevant methods applied to this study and provided a rationale for the choice of such methods. I have offered an overview of relevant technologies for proteomics, metabolomics, lipidomics and transcriptomics and discussed common methods for omics pretreatment. I have also described different data transformations and their impacts on data, as well as the normalisation of the omics measurements that can mitigate the unwanted variance in data. The later part of this chapter has reviewed missing data and related imputation methods in addition to statistical frameworks for omics analysis and biomarker identification methods. Building on the extensive methodological reviews, I have also introduced the co-expression network and the novel framework we employed to identify the communities in the network and associate it plausible biological functions. This co-expression network and the novel framework were then applied in the context of the cardiac age-OME project (see Chapter 7). Finally, after a brief literature review, I expounded on the

newly engineered method, omics SEA, that permits the detection of dysregulated pathways when multiple omics data are available. The omics SEA was then applied in the context of the human cardiac ageing project (see Chapter 8). Based on the methods described in this chapter, the following chapter provides in detail the study design. This work could be extended in multiple directions. For instance, we used the combined omics SEA to express data to understand how changes observed in carbon metabolic intermediates affect older individuals more than younger ones (see Chapter 3.1). Future investigations could also employ kinetic-based mathematical models combined with constraint-based models to infer metabolic fluxes in key metabolic fluxes using observed metabolomics measurements[155,156].

Chapter 5

Experimental Design and Data Exploration

High-throughput technologies have generated a vast amount of knowledge by simultaneously measuring the abundance of multiple molecules across several samples[639] and unlocked new biomedical discoveries[732]. Omics-driven discoveries and the increased accessibility to high-throughput technologies[39] have paved the way for personalised medicine[10,317,351,799] along with statistical[68] and mathematical frameworks[166]. Despite such assets omics measurements offer, however, the risk of generating non-reproducible results should not be overlooked[758]. Researchers have advocated the adoption of minimal standards in experimental and computational protocols to avoid a possible reproducibility pitfall in omics analyses[188,431,772]. A thorough experimental design can enhance the reproducibility of research findings in the omics era[241]. Yamada et al. (2021) have tackled this issue by presenting a unified framework for any omics with five pivotal principles that should be followed during an experiment involving omics measurements[907]. Our study of the human cardiac ageing heart has addressed these five principles for omics experiments[907], and this chapter discusses in-depth details of experimental design. This includes sample definition, storage, selection, processing and pre-treatment, and laboratory and computational protocols to increase the transparency of the study and its repeatability. Codes and data have been deposited in public repositories, and results are publicly available on the interactive web platform (`https://humancardiac.shinyapps.io/ageing/`).

I first review and discuss the five principles introduced by Yamada et al. (2021) to establish a solid framework in omics experiments[907]. The following section explains the collection, description and storage of the samples involved in the study and the criteria for sample selection and study approval. Then, I move onto presenting the laboratory protocol and pre-treatment of the measures applied in proteomics and summarising laboratory details for metabolomics. The details of lipidomics and transcriptomics are respectively presented. Furthermore, I explain the computational protocol with details on sample size calculations and summarise methods utilised to address each aim of this study (see Section 3). Each section also reports the data and code repositories used for the project [i]. Consequently, I expound on the part of explorative data analysis performed for the age-OME study,

[i] The data and code for this project will be accessible upon publication of the manuscripts.

taking into consideration its large dataset size [ii]. Finally, a succinct visual summary of the experimental design can be found in Figure 5.1.

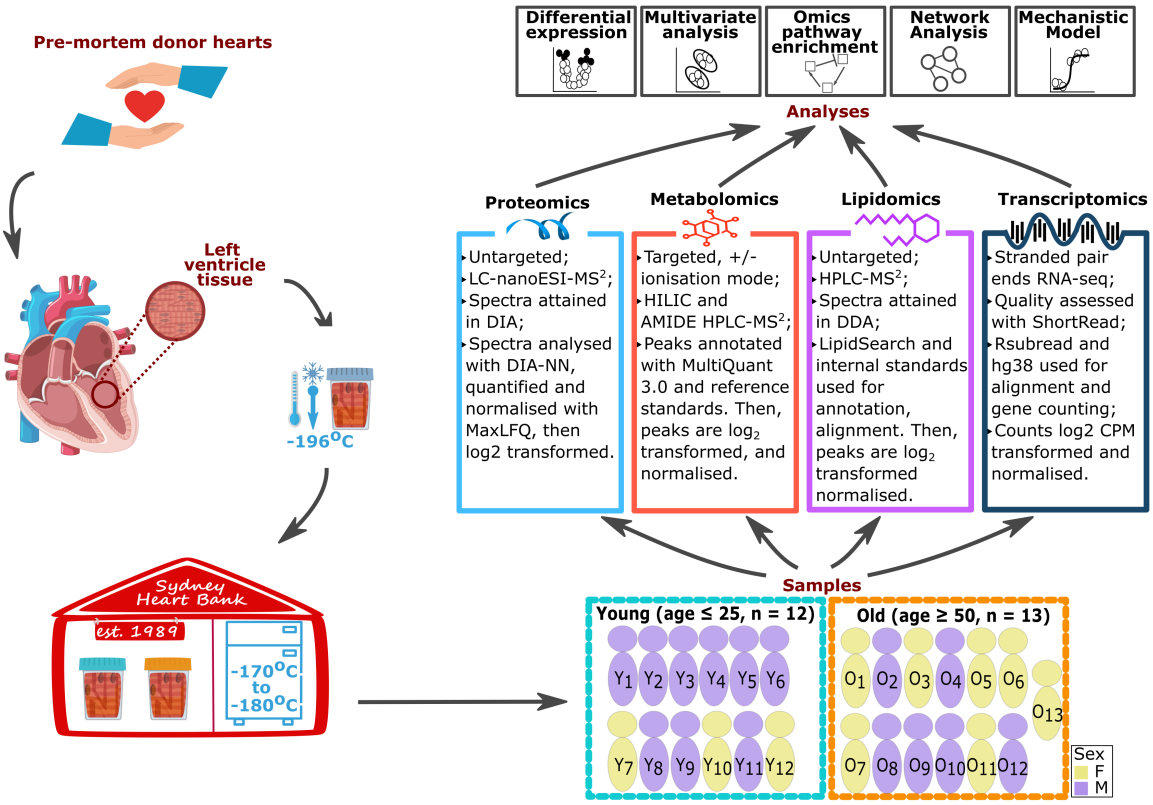


Figure 5.1: Schematic of the study design. Mathematical model details are not included.

5.1 Five principles for good omics research

Yamada et al.s' first recommendation refers to the management of samples and omics measurements. Researchers in the omics field should accurately describe the samples without clashing with the anonymity preservation of human data[215]. This point also requires researchers to disclose the locations and storage conditions of samples utilised for the analysis. Researchers are also recommended to report where the omics measurements are archived.

The second recommendation orbits around laboratory and computational protocols[907]. Researchers involved in wet labs should accurately describe methods applied during the experiment and each sample processing step. The same can be said for dry-lab experiments. Computational protocols with a detailed description of statistical and mathematical methods should be clearly reported. Creating the workflow before conducting analyses helps researchers to avoid poor practices such as p-hacking and b-hacking[778]. In addition,

[ii] Full details of explorative data analyses are available on a repository created for this project (`https://humancardiac.shinyapps.io/ageing/`).

for this study, it was helpful to include an additional standard for the data volume collected for the experiment. Establishing sample size calculations beforehand ensures that outcomes calculated from the experimental data have the desired level of statistical significance, power and effect size while addressing the variability between samples[72,787].

The third recommendation involves validating the software employed[907] for statistical and mathematical analyses. The applied software must be error-free and return reliable computations. The validation of software also relies on a researcher; it is important that the researcher is fully trained on how to perform analyses, fully comprehends the inputs of the software and carries out correct interpretation of the computational results that the software returns[201,398,530].

The fourth recommendation focuses on the reproducibility of experimental and computational results[374]. The repeatability of the results allows replication of the same results from independent studies on the same or similar data. The reproducibility also points to the integrity of a researcher, inasmuch as poor reproducibility of the results could derive from fabrication of results in some cases[675]. Although it is almost impossible to obtain a perfect reproducibility of a study, overall outcomes drawn from a repeated experiment should resemble the original findings[192]. A strict experimental design and careful description of computational methods along with a strong research integrity[16,59,331] establish the reproducibility[675,736].

The final recommendation proposed by Yamada et al. (2021) calls for researchers to gain a deep insight into the data. Explorative data analysis is a crucial step for understanding omics measurements[859], which can facilitate analyses and derive hypotheses beyond the ones initially reported in the experimental protocol. Exploratory data analysis, moreover, is a useful tool for assessing data quality by calculating relevant metrics and arranging the data in graphs, which helps to further the researcher's understanding of the phenomenon and improve the reproducibility of the results[4,633,859].

5.2 Samples definition, selection and sample size calculation

The samples utilised for this study are control donor hearts. Although they were suitable and ready for transplantation, they did not go through a transplant procedure due to various reasons, such as the immune incompatibility of a recipient and logistics issues. The biobank of donor hearts stores human hearts from 1990 onward[452], and the samples for this study range from 1990 to early 2000 due to the rarity of obtaining donor hearts with the current technological and medical advancements. The donors were hosted at the hospital, where several examinations were performed before proceeding with the organ collection; all the hearts used for our study underwent pathophysiological and histological examinations. Echocardiography was also performed, which observed normal cardiac functions. To be eligible for donation, the donors were required to have a body mass index below 30 and no comorbidities associated with cardiovascular diseases, such as diabetes, hypertension and smoking. Additionally, they were not affected by any chronic diseases. The donors, who had

been declared brain dead due to events such as accidents and suicide, were placed on life support. Consequently, all the donors' hearts were collected pre-mortem, and the cause of death was declared non-cardiac (further details on the donors' profiles can be found in 5.1). Surgical excision of several tissues for each donor's heart, including the left ventricle tissue, was then performed in the operating theatre. Tissues were immediately frozen in liquid nitrogen at a temperature of 196^oC. The samples were then transported and stored in the range of 170-180^oC in the Sydney Heart Bank located at the University of Sydney.

This study was primarily concerned with the changes associated with ageing at the molecular level in the myocardial tissue, as mentioned in Chapter 3. The study used left ventricle tissues surgically excised from the donors' hearts. The left ventricle was identified as a suitable model firstly because of its frailty due to age-related cardiovascular diseases (see Chapter 2.4). Secondly, the left ventricle tissue can shed light on cardiomyocyte molecular changes since cardiac cells are predominant in left ventricular tissues according to histological analyses[504,649,948]. Thirdly, the left ventricle does not exhibit changes in the number of cardiac cells observed when male and female adult hearts were compared[504], which excludes sex from a potential confounder in the findings[330] (see Chapter 2.2).

The human cardiac ageing investigation was conducted by stratifying the individuals into two groups, *young* and *old* [iii]. The left ventricular tissue in the young cohort belonged to individuals aged between 14 and 25 since systematic studies have shown that hearts share similar parameters from the age of 12 onward[240], while many studies identify the beginning of adulthood from the age of 25 of an individual[541,542]. From a clinical perspective, the transition from paediatrician cardiology to cardiology of an individual occurs at 12[339,364]. The samples for the old cohort included the left ventricular tissues of individuals aged between 50 and 65. The threshold of 50 years of age was determined given the increased risk of cardiovascular diseases above 50[24,62]. In total, this study utilised 25 samples. The younger cohort included 12 individuals, with an average age (standard deviation) of 19 (± 3.4), and the older cohort included 13 individuals with an average age (standard deviation) of 57 (± 4.7). The younger cohort comprised 9 left ventricular tissues drawn from male and 3 from female donors while the older cohort contained 6 male samples and 7 female samples. Although, qualitatively speaking, this indicates an asymmetry between the distribution of biological sex of the samples across the two cohorts, further statistical examination showed that there is little or no evidence between the two cohorts. In fact, applying the test of equal proportions[599], knowing that $9/12 = 0.75$ and $6/13 \approx 0.46154$ are the proportions of male respectively in the young and old cohort, the estimated difference is 0.28846, with 95% confidence interval equals to $CI_{95\%} = [-0.15699, 0.73391]$ and p-value equal to $p_{val} = 0.2881$.

While omics experiments require both material resources and manpower, it is imperative that the sample size is adequate in order to increase the reproducibility of the results[907]. After identifying an appropriate biological model to characterise the molecular landscape of the myocardium and its changes with ageing and selecting the candidate samples, we performed sample size calculations to ensure that we had enough power in the statistical

[iii] The terms young and old are relative, and they should not be interpreted in a demographic sense. The two arms have been identified based on medical criteria rather than demographic.

tests. The primary research question of this study involves identifying candidate biomarkers that drive the ageing process, as discussed in Chapter 3. In statistical terms, this can be addressed by testing if a change can be detected across the two cohorts in each molecule. This hypothesis can be tested using the tools discussed in Chapter 4.8. For this study, we used the moderated t-test computed with limma[764] (see Section 5.7). In order to determine the sample size calculation, we first defined the effect size for the i-th analyte, which reads

$$\Delta_i = \frac{\delta_i}{SD_i(2/n)^{1/2}}, \tag{5.1}$$

where n is the sample size, δ_i is the $\log_2 FC_i$, and SD_i is the standard deviation of the i-th analyte estimated using limma (see Section 3.1). Here, it is notable that, although the definition of effect size in Equation 5.1 is very similar to Cohen's effect size[787], it would be inappropriate to regard it as such since the standard deviation is estimated using limma rather than the pooled standard deviation[764]. Our minimum sample size is then computed focusing on the effect size, namely $\delta_i/SD_i = 3$, for each i-th analyte[787]. To finalise the sample size calculation, a few more considerations are necessary. First, we set the probability of type II error, β, equal to 0.1. Consequently, the power of the test is set to $1 - \beta = 0.9$. Then, we set α, probability of type I error, equal to 0.05[730]. The α here does not represent the type I error for the single test but rather the type I error corrected by a false discovery rate. This is processed by assuming that only 95% are truly dysregulated amongst all analytes[507]. Further mathematical details on the sample size calculation with multiple testing corrections are available in Liu and Hwang (2007). The minimal sample size for the study leads to a minimum of

$$n = 7 \tag{5.2}$$

samples. As mentioned in Section 4.7, measurements detected using high-throughput technologies can present a degree of missingness. To avoid missing values from affecting the power of the sample, I created a simple solution to maintain the desired power of the test. Assuming that we test analytes that are present up to 75% of the measurements in each group, the minimum sample size, n^* is then recomputed by solving the following equation

$$1 : 0.75 = n^* : n, \tag{5.3}$$

where $n = 7$, from Equation 5.2. Solving Equation 5.3

$$\begin{aligned} n^* &= \frac{n}{0.75} \\ &= 9.333, \end{aligned} \tag{5.4}$$

we can conclude that we need

$$n^* \approx 10 \tag{5.5}$$

samples in each cohort to retain the desired power of the test $(1 - \beta) = 0.9$ and type I error $\alpha = 0.05$. In light of all the reported details, the human cardiac ageing study was approved by the Ethics Committee of the University of Sydney (USYD # 2021/122).

Sample	Age	Gender	Cause of Death
Young 1	14	M	Hypoxia from hanging
Young 2	14	M	Head Trauma
Young 3	16	M	Hypoxia from hanging
Young 4	17	M	Head Trauma
Young 5	18	M	Hypoxic brain injury secondary to alcohol toxicity
Young 6	19	M	Motor vehicle accident
Young 7	19	F	Motor vehicle accident
Young 8	19	M	Hypoxia from hanging
Young 9	21	M	Motor vehicle accident
Young 10	23	F	Motor vehicle accident
Young 11	23	M	Hypoxic brain injury secondary to self-strangulation
Young 12	25	F	Subarachnoid haemorrhage
Old 1	51	F	Subarachnoid haemorrhage
Old 2	53	M	Infrarenal aortic aneurysm
Old 3	53	F	Cerebrovascular accident
Old 4	54	M	Crush injury involving the right ribs with a resulting embolism
Old 5	54	F	Subarachnoid haemorrhage
Old 6	54	F	Subarachnoid haemorrhage
Old 7	55	F	Cerebral haemorrhage
Old 8	56	M	Cerebral haemorrhage
Old 9	60	M	Motor vehicle accident
Old 10	62	M	Middle cerebral aneurysm + haemorrhage and brain herniation
Old 11	62	F	Hypoxic brain injury
Old 12	65	M	Motor vehicle accident
Old 13	65	F	Kidney tumour

Table 5.1: Profile of heart donors.

5.3 Proteomics protocol

Bottom-up proteomics analyses were conducted by Dr Mark Lawrence and his team at the University of Sydney. The frozen cardiac tissues were powdered liquid nitrogen. Then, about $10mg$ of the mixture was homogenised in 4% sodium deoxycholate and $100mM$ Tris-HCl pH 7.5; successively, the samples were heated at 95^oC. The mixture was exposed to 70% amplitude of ultrasonic vibration using QSonica R2 (Q Sonica). Centrifugation of the samples occurred at the maximum relative centrifugal field of $18000 \times g$. The BCA assay (Pierce) was used to quantify the protein concentration, and then each $20\mu g$ of protein was trypsinised and held overnight at 37^oC.

Peptides were prepared[315] and injected in a fused silica column ($30cm \times 70\mu m$ C18, $1.9\mu m$ particle size) with a $10\mu m$ pulled tip[349], coupled online to a nanospray ESI source at $2.3Kv$; the peptides were further resolved over a gradient of acetonitrile (5%-40%) for over 2 hours with a flow rate of $30nL \cdot min^{-1}$. Q-Exactive Fusion Lumos (ThermoFisher) was used for the tandem MS analysis with 27% normalised higher energy collisional dissociation for

fragmentation. The spectra were obtained in data-independent acquisition (DIA) using 20 variable isolation windows[290] and saved in RAW format[187]. All the runs were performed on the same day on the same batch. The RAW files, comprising of high pH fractions[536] data, were then integrated through DIA-Neural Network (DIA-NN) 1.7v, known as DIA-NN 1.7v[183]. The neural network classifier was set in double-pass mode[517]. MaxLFQ was integrated into DIA-NN for the untargeted quantification of peptides[159]. Peptide identification was conducted on the human database (downloaded on 5th May 2020), imposing the false discovery rate cut-off at 1% on the precursor ions[160,348,831], and maximum occurrence of missed cleavages equal to 2[608,751]. The search parameters also included possible modifications. For our research, cysteine carbamidomethylation was set as a fixed modification, while methionine oxidation was set as a variable modification[102]. The search also included the retention time to increase the accuracy of the protein hits[428]. The normalised counts were then $\log_2$ transformed (see Chapter 4.5). The MS proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier: PXD050610 [iv].

5.4 Metabolomics protocol

The semi-targeted metabolomics study was conducted by Professor John O'Sullivan and his team at the University of Sydney. About $50mg$ each of left ventricle tissue was withdrawn from the frozen hearts and then homogenised. After embedding the tissue in the extraction medium of methanol, water and chloroform, the mixture was centrifuged at $14300rpm$, at 4^oC for about 25 minutes. The aqueous layer was then extracted and injected into another microcentrifuge tube. The tubes were subsequently placed in the concentrator Thermo Fisher Scientific SpeedVac SPD120, where nitrogen steam was used to dry the samples. The evaporated samples were then reconstructed using two solvents. The first solvent comprises acetonitrile/methanol/formic acid in $75 : 25 : 0.2(v/v/v)$ proportion. The resulting mixture was then used to detect metabolites in positive ionised mode through the HILIC column (Waters Atlantis HILIC column). The second solvent was composed of acetonitrile/methanol in $25 : 25(v/v)$ proportion and used the AMIDE column (Waters XBridge Amide column) for negative mode investigation[430]. Extraction buffers were added to the samples with four standards: phenylalanine-d8, valine-d8, thymine-d4, and citrate-d4. The added standards allowed the technicians to measure the reliability of the MS reads. Similarly, blanks and pooled samples (quality control) were used for quality evaluation and assurance[18].

The MS runs were performed in one batch on the same day. The analytical platform used for the semi-targeted metabolomics study was LC-MS/MS. Agilent 1260 Infinity Liquid Chromatography was coupled to AB Sciex QTRAP 5500. Multiple reaction monitoring Q1/Q3 analyses were conducted using MultiQuant 3.0v and Analyst Software 1.6.2v for peak integration[430,854]. The raw ion counts were $\log_2$ transformed (see Chapter 4.5), and

[iv] The data will be publicly available upon publication of the manuscript.

normalisation of the spectra was conducted using EigenMS[405] (see Chapter 4.6) [v].

5.5 Lipidomics protocol

Untargeted lipidomics analysis for the age-OME study was coordinated by Prof. Anthony Don and conducted in facilities at the University of Sydney. For each heart, about $20mg$ of the left ventricle was used for the analysis. The tissue was homogenised using $250\mu L$ of methanol containing 0.01% (weight/volume %) of butylhydroxytoluene. Additionally, the following internal standards were introduced: 2 nmol of PC(19:0/19:0); 2 nmol of SM(d18:1/12:0); 1 nmol of GluCer(d18:1/12:0), 1 nmol of Cer(d18:1/17:0), 1 nmol of PS(17:0/17:0), 1 nmol of PE(17:0/17:0), 1 nmol of 1 nmol of PA(17:0/17:0), 1 nmol of PI(d7-18:1/15:0), 1 nmol of PG(17:0/17:0), 1 nmol of CL(14:0/14:0/14:0/14:0), and 1 nmol of TG(17:0/17:0/17:0); 0.5 nmol of DG(d7-18:1/15:0); 0.5 nmol of CholE(17:0); 0.5 nmol of LPC (17:0); 0.5 nmol of LPE(17:1); and 0.5 nmol of AcCa(d3-16:0); 0.2 nmol of Sph(d17:1); 0.2 nmol of S1P(d17:1), 0.2 nmol of LacCer(d18:1/12:0), and 0.2 nmol of MG(d7-18:1). The lipid extraction followed a two-phase protocol, organic and acqueous, using methyl tert-butyl ether and water, respectively for each phase. Consequently, $850\mu L$ methyl tert-butyl ether was added to the mixture and then sonicated in 4^oC water for 30 minutes.

Successively, the composite was introduced in the centrifuge set at $2000g$ and kept for 5 minutes. After centrifugation, $212\mu L$ of water (MS graded) was introduced into the mixture. A $5mL$ glass tube was employed for the organic phase. For the aqueous phase, $5000\mu L$ of methyl tert-butyl ether and $150\mu L$ of methanol were used. At the end of the aqueous phase, the mixture was exposed to ultrasound or sonicated for extra 15 minutes and then underwent phase separation for $125\mu L$ of water. Aqueous and organic phases were combined and dried using ThermoFisher Scientific Savant SC210 SpeedVac. Then, $400\mu L$ liquid comprising of 80% methanol/20% water/0.1% formic acid with 0.01% (weight/volume) butylated hydroxytoluene was used for the lipid reconstruction. The resulting mixture was stored at the temperature of -80^oC. The LC-MS/MS was used for lipidomics, and the sample was injected in a randomised fashion and acquired in positive and negative modes.

Vanquish High-Performance Liquid Chromatography was used with $2.1 \times 100mm$ Waters C18 HPLC column with 1.7 μm pores size for analysing the samples. The MS employed was ThermoFisher Q-Exactive HF-X. For the liquid phase, two solvents were used. The first solvent was $10mM$ of ammonium formate with 0.1% of formic acid in acetonitrile/water ($60\%/40\%$). The second solvent was $10mM$ of ammonium formate in 0.1% formic acid in isopropanol/acetonitrile ($90\%/10\%$). In the liquid phase, different gradient settings were then used, alternating the two solvents. Specifically, for a total of 27 minutes gradient, $0.28mL/min$ of solvent mixture was applied for the gradient as follows: $80\%/20\%$ A/B as a start; $80\%/20\%$ A/B, after 3 minutes; $55\%/45\%$ A/B, after 5 minutes 30 seconds; $35\%/65\%$ A/B, after 8 minutes; $15\%/85\%$ A/B, after 13 minutes; $0\%/100\%$ A/B, after 14 minutes; $0\%/100\%$ A/B, after 20 minutes; $70\%/30\%$ A/B, after 20 minutes and 12 seconds;

[v] The MS metabolomics will be available on Metabolomics Workbench at `http://dx.doi.org/10.21228/M8N42X` upon publication of the manuscript.

and $70\%/30\%$ A/B at the end[834]. The full scan and data-dependent mode were used to acquire the spectra from the tandem MS[706,834]. Some ions were removed based on their signals found in the blanks. Thermo Fisher LipidSearch software 4.2v was used for lipid annotation[893] with tolerance set at $5ppm$ and diagnostic product ions tolerance set at $8ppm$. For each class, the abundance of each lipid was expressed as a ratio of its internal standard and then multiplied by the amount of the internal standard injected in the beginning to approximate the lipid abundance. Then the peaks were first $\log_2$ transformed (see Chapter 4.5) and normalised using EigenMS[953] (see Chapter 4.6) [vi].

5.6 Transcriptomics protocol

Dr Cassandra Malecki and I jointly coordinated transcriptomics investigations for the age-OME study. The sequencing was performed at the Ramaciotti Centre of Genomics, University of New South Wales, Australia. For each sequencing, $50mg$ of left cardiac tissue was used. The sample was injected in a $2ml$ Qiagen tube with a stainless steel bead of $5mm$. The tissue was then homogenised with the TRIzol reagent manufactured by Invitrogen in a Qiagen TissueLyser LT. The purification/extraction of RNA was performed following TRIzol's protocol[684]. The sample was then DNAse treated with the Qiagen RNAse Set. The purification involved the Qiagen RNAse Mini Kit based on the manufacturer's instructions. The quality and purity of the RNA were assessed by Nanodrop spectrophotometer, and then the $260/230nm$ and $260/280nm$ absorbance ratios were assessed[181].

The RNA Nano Chip on Agilent Bioanalyzer was used to evaluate the integrity of RNA. The Illumina Stranded Total RNA Prep, Ligation with Ribo-Zero Plus was used to prepare the RNA library with paired-end $250bp$ kit on the S4 flow cell of the Illumina NovaSeq 6000[470]. The quality of the RNA sequence was assessed using ShortRead[581], and the genes were identified using human genome assembly HG38[626]. The alignment of the RNA sequencing to the genome and the gene count was computed using subread[487] without any trimming[486,878]. RNA counts showed $\log_2$ counts per million (see Chapter 4.5) and then normalised the counts using variance modelling (see Chapter 4.6) [vii].

5.7 Computational protocol

According to Yamada et al.'s (2021) principles of good practices for omics research, the lack of computational protocol could impact the reproducibility of the research. This section extensively reports the main computational methods used in the context of the ageing heart project in order to avoid such a pitfall. According to the sample size calculations and literature[770] (see Section 5.2), data filtering was performed, and only analytes present in 75% of the cases in each cohort were considered for follow-up statistical and mathematical

[vi] The lipidomics data are hosted on Metabolomics Workbench at `http://dx.doi.org/10.21228/M8N42X` and will be publicly available upon acceptance of the manuscript

[vii] Transcriptomics source data will be available on the GEO public repository at `https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE263566`. upon manuscript publication.

analyses[770], ensuring that the power of the test $(1 - \beta)$ for the molecular signature was not compromised in any analyte.

Principal component analysis (PCA) was conducted as part of explorative data analyses[293,393]. PCA was mainly used to qualitatively understand if any clustering occurred within the data[475] and to quantitatively detect possible outliers[336,475]. PCA is applied on the covariance matrix calculated on the kNN ($k = 10$) imputed dataset of each omic's measurements[393,818]. Briefly, let $\mathbf{D}^*$ be the transposed of the imputed omics measurements $\mathbf{Y}^{imp}$ centred to zero. The matrix $\mathbf{D}^*$ contains the samples on the columns and molecules on the rows, and applying the singular value decomposition on it, it can be denoted as

$$\mathbf{D}^* = \mathbf{ULA}^T, \tag{5.6}$$

where $\mathbf{U}$ and $\mathbf{A}$ are the metric of left and singular component, and $\mathbf{L}$ is a diagonal matrix containing the singular values. The rank approximation of matrix $\mathbf{D}^*$ reads

$$\mathbf{D}^* \approx \mathbf{U}_2\mathbf{L}_2\mathbf{A}_2^T, \tag{5.7}$$

where $\mathbf{L}_2$ contains the two largest singular values of $\mathbf{L}$, while $\mathbf{U}_2$ and $\mathbf{A}_2$ are the left and right singular components associated with the two largest singular values in $\mathbf{L}$. The two principal components utilised to inspect the omic measurements are calculated as

$$\mathbf{P} = \mathbf{D}^*\mathbf{A}_2. \tag{5.8}$$

In other words, the matrix $\mathbf{P}$ has as many rows as samples in the omic measurements, with two columns representing two principal components. The total variation captured by each component is calculated as the squared singular value divided by the sum of all the squared singular values. The 95% confidence intervals are created using bivariate normal distribution fitted on the two principal components contained in $\mathbf{P}$.

The characterisation of the molecular signature in each omic is performed using the moderated t-test calculated through limma[764] (see Section 4.8). For each molecular signature conducted within the limma framework, we report the $\log_2$ fold change ($\log_2 FC$), fold change (FC), standard deviation (SD), observed statistic (t), the degree of freedom if the sampling distribution (df), the lower and upper bound of the 95% confidence interval (CI_l and CI_u), p-value (p_{val}), and B-statistics (B)[764]. The odds ratio for each analyte is then calculated as $od = e^B$. The corrected p-value (p_{fdr}) for multiple testing is performed by using the Benjamini-Hochberg procedure[63]. Based on the sample size calculations, an analyte is differentially expressed when $p_{fdr} < 0.05$.

On each omic level, the cut-off at $p_{fdr} < 0.05$ is significant in identifying molecules with strong evidence of dysregulation between the two cohorts. Yet, the p_{fdr} similar to the p_{val} should be considered a draw from a continuous uniform random distribution, and its value should be interpreted as a continuous value[22,417,557,585,787]. To facilitate the interpretation of the data, we subscribed to the recommendations of Bland (2015) and extended our exploration to patterns of all the analytes that exhibited $p_{fdr} < 0.1$[77]. The expression of those analytes was then extracted from the imputed dataset of each omic. The missing

data were imputed using the kNN with $k = 10$[830]. Next, the expression of the analytes with associated $p_{val} < 0.1$ was centred to zero and then used to construct a heat map with two independent clustering procedures. A heat map construction benefited the graphical visualisation of the expression patterns. The sample- and analyte-wise agglomerative hierarchical clustering procedures investigated potential groups in the data. The clustering procedure was performed on the Euclidean distance matrix calculated on the selected analytes' expressions and visually summarised in a dendrogram. Then, Ward's minimum variance linkage was employed to identify clusters and calculate the distance amongst them at each step of the clustering procedure[262,935].

In terms of metabolomics, the ratio of analytes can play an important role in interpreting the results. For example, detecting a difference between the oxidised/reduced state of a metabolite can unveil if there is a general accumulation of the analyte or a higher activity of its reductase or oxidase enzyme. One can detect if the change in expression occurs due to higher enzymatic activity by testing if there is any change in a ratio of the metabolites, such as the oxidised/reduced state of an analyte[897]. The Welch's t-test was employed to detect if the ratio of metabolites changed with ageing (see Section 4.8). Each ratio was tested by considering the differences in the analytes' abundances in the respective cohort, as those are expressed in $\log_2$ scale (see Sections 5.3 - 5.6).

In lipidomics analyses, the length of fatty acid tails was investigated by extracting the number of carbon molecules present on each lipid molecule and then dividing the length into classes. Based on the lipids detected in the MS analysis, three classes were identified in the dataset (see Chapter 2.3). The first class comprises medium-chain fatty acids, which present carbon atoms between 6 and 12. The second class is the long-chain fatty acids, where tails exhibit a number of carbon atoms between 13 and 18. The third class represents the very long fatty acids that contain all the lipids with tails with more than 18 carbon atoms. For each fatty acid, the $\log_2$ fold change of the corresponding lipid was extracted and then divided for the standard deviation estimated using limma[764], creating weighted averages. To test if a class of fatty acids tends to be more abundant with ageing, the t-test is applied to the weighted averages. For this study, the weighted averages were used to help the data match the assumption of equal distribution required by the t-test[423].

Further analyses were conducted on lipids measurements to identify and understand any evident linear patterns between the number of carbon double bonds in a fatty acids chain of lipids and their $\log_2 FC$, representative of the metric of the lipid trend in the ageing process. This analysis aimed to detect if there was an accumulation of saturated or polyunsaturated lipids in their respective classes, such as Ceramides (see Chapter 2.3). The analyses were conducted by fitting a linear model, where the independent variable was $\log_2 FC$ of the lipid and the independent variable was the number of carbon bonds on its tail. The coefficients of the linear model were estimated using the weighted least square procedure to address possible heteroscedasticity across the $\log_2 FC$s[692]. Regression weights were the inverse standard deviation estimated with limma[764]. A similar approach was employed to examine if there was a linear trend between the $\log_2 FC$s of each lipid class and the length of the fatty acids tail.

Consequently, the co-expression network was then applied to each omic level and assessed protein-protein, metabolite-metabolite, lipid-lipid and transcript-transcript associations. The network inference of each omic interaction was based on the soft threshold on the edge strength. The soft threshold was obtained by setting $R^2 = 0.9$. The edge's strength was estimated through Pearson's correlation coefficient estimated on the dataset using only completed observations between analytes. Statistical metrics were then calculated on the inferred network[636]. The strength of each network conserved the sign of the correlation coefficient. The community in the network were identified using walktrap, a random walk-based algorithm[656] (see Chapter 4.9 for further details on the network). The functional annotation of the largest community in protein, metabolite and transcript networks was conducted using ORA on WikiPathways database[762] (see Chapter 4.10.1). The functional annotation for lipids was conducted using LIPID MAPS[228].

Omics SEA was applied to address two hypotheses whose statistical formulations resembled (see Section 4.10.2 for the mathematical details of omics SEA). The first hypothesis focused on the pathways that exhibit a behaviour divergent from a simple random fluctuation and regarded that its dysregulation was associated with the ageing process. To address this, we considered all the omics data at once; when proteins and transcript measurements were available for the same gene, we ensured to favour the protein expression to reflect the central dogma of molecular biology and the fact that proteins are mainly responsible for enzymatic reactions[139,370]. The omics analysis was conducted with the WikiPathways database due to its wide coverage of lipid annotation in biological pathways along with LIPID MAPS[645,646,762,767]. Using omics SEA, the second hypothesis was tested to detect if there was any change associated with ageing in the lipid class rather than at a single molecule level. The lipid classes were defined based on LIPID MAPS annotation[228] (see chapter 2.3) [viii].

5.8 Explorative data analysis

The samples utilised for this study are non-diseased cryopreserved human hearts across the lifespan, which makes it hardly possible to repeat and replicate the study. The hearts used for our study were deemed pathologically normal following formal histological evaluation by anatomical pathology at the respective hospital and echocardiography. The donors had no cardiac disease risk factors. As previously discussed in Section 5.2, these samples were collected in the past when there was only one heart transplantation site in Australia and transplantation methods and technologies were not yet advanced. Since modern logistics and patient matching techniques have enabled smooth matching and transplant of most hearts, especially young ones, healthy human hearts are rarely available for research purposes. Young donors' hearts have not been collected for over ten years, while older hearts stored in other biobanks tend to have been regarded as insufficient for transplant and, therefore, may not suffice the criteria of this study.

[viii] The code will be hosted and publicly available at `https://github.com/gguglielmi92/Human-Cardiac-Ageing.git`. upon publication of the manuscript

This unique nature of our sample collection has enabled this omics study, and the measurements we have taken are made available to the public in the online repository to advance the field and our understanding of mechanisms behind cardiac ageing[759]. The online repository (available at: `https://humancardiac.shinyapps.io/ageing/`) aims to present information related to the samples, source data, explorative analyses and univariate and multivariate analyses without bias. The online repository contains full data explorations that cannot be entirely described in this thesis[639,759]. Furthermore, the interactivity of explorative and statistical analyses in the online repository allows researchers to interface with data and results without requiring any programming skills. This section focuses on exploratory data analyses that are pivotal to understanding the data and presents the composition of each omic layer and principal component analysis (see Section 5.7).

Applying high-throughput technologies to cardiac tissues (see Section 5.2) has enabled the characterisation of four different molecular layers, namely proteins, metabolites, lipids and transcripts (see Sections 5.3-5.6). Based on the criteria reported in Section 5.2, this study includes only molecules that are contained in 75% of the dataset. The summary of the samples used for each molecular characterisation and the number of analytes identified in each omic layer can be found in Table 5.2. According to the sample size calculations presented in Section 5.2, a minimum of 10 samples were required. Considering the order of high-throughput investigations described above, left ventricular tissues that were left to maintain the required sample size for the analyses of the transcripts were insufficient; indeed, they were inadequate from lipidomics analyses.

Omics	**Gender**	**N. young**	**N. old**	**Identified molecules**
Proteins	Female	3 samples	7 samples	
	Male	9 samples	6 samples	
	Total	12 samples	13 samples	2938 identified proteins
Metabolites	Female	3 samples	7 samples	
	Male	9 samples	6 samples	
	Total	12 samples	13 samples	150 identified metabolites
Lipids	Female	3 samples	5 samples	
	Male	9 samples	6 samples	
	Total	12 samples	11 samples	341 identified lipids
Transcripts	Female	2 samples	4 samples	
	Male	6 samples	4 samples	
	Total	8 samples	8 samples	15447 identified genes

Table 5.2: Samples in each experiment.

As shown in Table 5.2, lipidomics analyses were performed on a total of 23 samples, 12 from young and 11 from old cohort. The left ventricular tissue of the young cohort was adequate to guarantee the lipid investigation on each sample, whereas 2 samples from the old cohort were insufficient; the missing samples were Old 3 and Old 6, respectively (see Table 5.1). While the sample size still matched the desired power of the test for analysing the molecular signature, the biological sex composition of the groups was affected and required further

tests. The test of equal proportion[599] was applied to see if possible unbalance occurs. The young cohort contained $9/12$ males, and the old cohort was $6/11$ males. The estimated difference of male proportion between the young and old cohorts was 0.2045455, and the 95% confidence interval was $CI_{95\%} = [-0.26540.67456]$, with associated $p_{val} = 0.5548$. The test of equal proportion did not exhibit any evidence of unbalanced samples.

Though it had not been planned originally, the transcriptomics investigation was conducted to clarify if changes detected at the protein level could also be observed in the transcripts. The sample available for this investigation was insufficient to guarantee the desired power (see Section 5.2), yet their analyses could still unveil changes occurring with ageing. As also reported in Table 5.2, the transcriptomics analysis was performed on 16 samples, 8 from the young and 8 from the older cohort. In the younger cohort, Young 1, Young 2 and Young 6 could not be utilised due to insufficient biological samples for the analyses. Similarly, Old 1, Old 2, Old 3, Old 5 and Old 12 were not used for the transcriptomics analyses (see Table 5.1). The male ratio in the young and old cohorts were respectively $6/8$ and $4/8$, making the difference between ratios 0.25. The 95% confidence interval was $CI_{95\%} = [-0.3333, 0.8333]$, and the p-value returned by the test was $p_{val} = 0.6056$. The test for equal proportion did not exhibit any evidence against the hypothesis that there was no difference between the composition of the two samples. As such, the initial exploration of data compositions for each omics level helped us to gain insights into sample size, the size of each omic library and possible demographic unbalance that ventricular tissues could present. Principal component analysis was performed on each dataset to better understand omics measurements and further evaluate whether any clustering occurs in the data. Amongst all the omics on which principal component analysis was performed, the first two principal components were presented, following the common empirical approach in the field[475] (see Section 5.7 for the mathematical details of principal component analyses).

Principal component analysis conducted on all the detected protein measurements is presented in Figure 5.2a. The principal component analysis has enabled the representation of the samples used for proteomics in two dimensions. The first dimension is summarised by the first principal component (PC_1). The PC_1 explains 15.18% of the total variability. The second dimension is summarised by the second principal component (PC_2), which explains 14.23% of the total variability. Drawing the 95% confidence interval on each group, no indication of any clustering within the young and the old cohorts was observed due to gender in accordance with the test of equal proportion. Within each cohort, no outliers could be detected across the samples. Furthermore, samples belonging to each cohort displayed a marginal separation, albeit not perfectly.

Principal component analysis of the metabolomics data, displayed in Figure 5.2b, represents the rank-2 approximation of the samples by considering all the 150 metabolites detected in our investigation. The first principal component (PC_1) is representative of the 15.65% of the total variability in the data, while the second principal component (PC_2) captures 13.23% of the total variability in the data. Conducting a visual inspection of the outcome of the principal component analysis did not identify any clustering due to gender. Furthermore, no observation lay outside confidence intervals, indicating that no sample could represent a potential outlier. Principal component analysis displayed a perfect separation between the two groups, suggesting that metabolites may play a pivotal role in

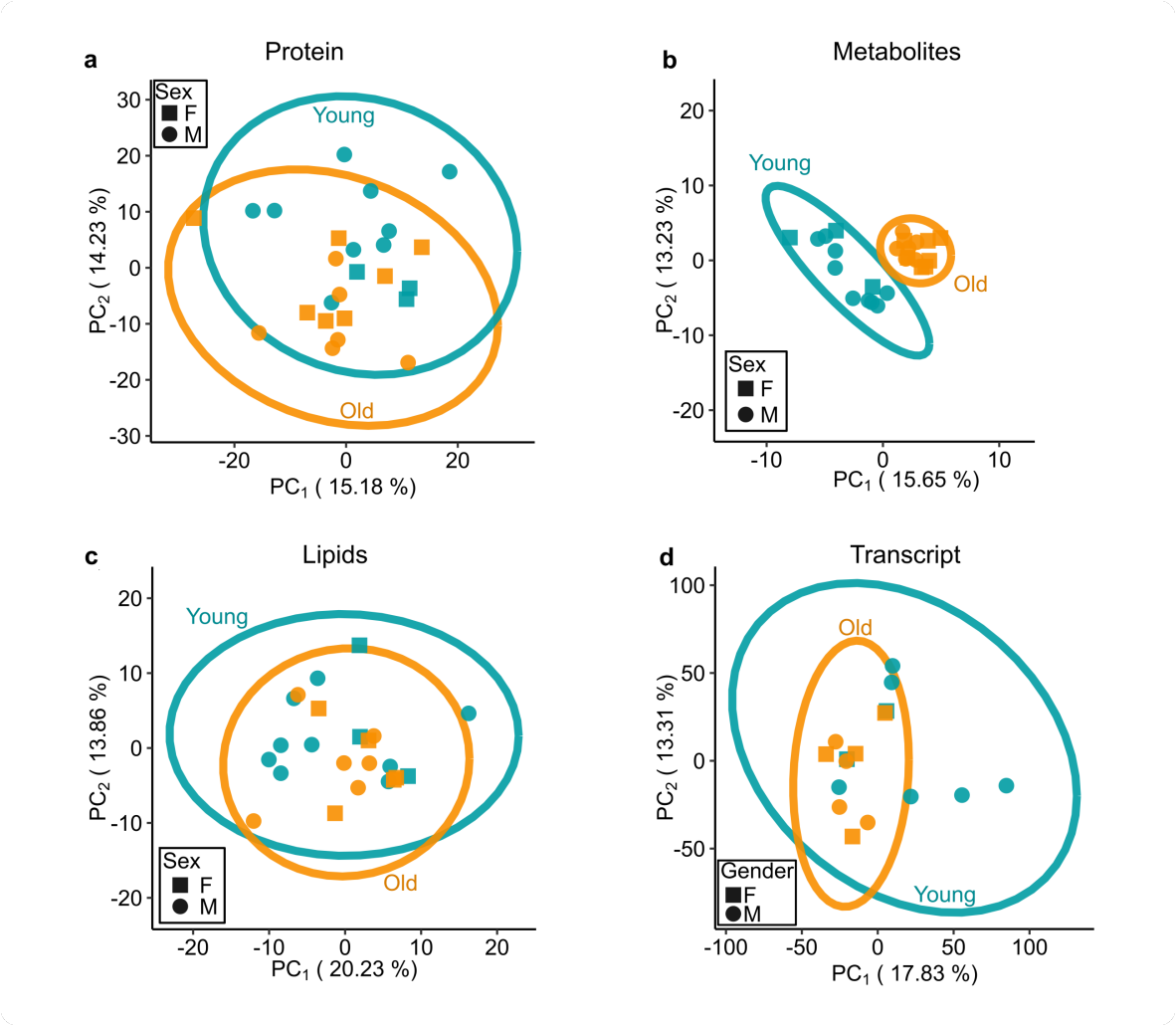


Figure 5.2: Principal component analysis with 95% confidence intervals. **(a)** Application on proteomics. **(b)** Application on metabolomics. **(c)** Application on Lipidomics. **(d)** Application on transcriptomics

the ageing process.

Principal component analysis was then applied for lipids in order to assess possible groups hidden in the data. The visual display of principal component analysis for lipids is shown in Figure 5.2c, where the first principal component (PC_1) explains 20.23% of the total variation and the second principal component (PC_2) addresses 13.86% of the variability in the data. The samples exhibited in each cohort appeared to cluster together and indicated that no sample had outstanding behaviour. The data did not suggest any clustering due to gender between and within the cohort. A visual inspection of principal component analysis further suggested that lipids did not exhibit a separation between the two cohorts.

Lastly, principal component analysis was utilised to explore the transcript measurements, as shown in Figure 5.2d. Principal components used to reduce the dimensionality

of the data, namely PC_1 and PC_2, respectively, explain 17.83% and 13.31% of the total variability of the data. The samples did not show any particular cluster due to gender. Similarly, the observation within each group appeared to not fall out of their respective 95% confidence intervals. Notably, the younger cohort exhibited a larger variation across its samples than the older one. Although observations between the two groups exhibited a degree of separation on the first principal component, it was not possible to assess it clearly due to the large variability intrinsic to the samples in the younger cohort.

Conclusion

Omics experiments have enabled researchers to quantify large quantities of molecules simultaneously and led to thousands of biomedical discoveries in the past two decades; nonetheless, some studies cannot be reproduced[219]. During this study of human cardiac ageing, we have embraced the importance of adopting appropriate experimental designs and protocols to increase the reproducibility of omics research, as emphasised by Yamada et al. (2021). Following their five principles of good omics research, this chapter has fully disclosed key details of our study design, ranging from information on the donor heart selection, sample storage, sample size calculation and laboratory and computational protocols. A complete set of information that this thesis does not cover, including data and code, is available online to enhance its reproducibility and transparency. Furthermore, the last part of this chapter has presented relevant explorative data analyses, which provide valuable insights into the omics data and, therefore, strengthen the credibility and reproducibility of the downstream analyses[759,907]. Built on the methods described in Chapter 4 and the study design in this chapter, I will present findings in the next three chapters.

Chapter 6

Molecular Signature of Cardiac Ageing

This chapter describes the outcome of molecular signature analysis (differential expression) obtained from molecular investigations on the tissues[173]. Molecular investigations provide in-depth insights into the dynamics of ageing and the interplay of molecular changes with those at the higher level, such as changes in cellular function or in the heart anatomy[384]. The application of modern high-throughput technologies has enabled us to explore multiple dimensions of the cardiac ageing process using biomarkers across multiple types of molecules. As discussed in the previous chapter, to explore the age-related changes occurring across the variety of molecules in the heart, we used our unique collection of pre-mortem human cardiac tissues for proteomics, metabolomics, lipidomics and transcriptomics analyses (see Sections 5.3-5.6). Differential expression analysis was then applied to every set of omic measurements, using moderated t-test computed via limma[764], a method presented in Chapter 4.8.

This chapter opens up with the results regarding the changes observed in proteins between the young and old cohorts. The following section discusses the results from metabolites, which indicate a possible dysregulation in small molecules. I then extend the analysis to include ratios of metabolites; ratio analyses benefit the interpretation of the findings and the outcomes of biological pathways analysis, as discussed in Chapter 8. The characterisation of lipid changes associated with ageing is introduced in the following section. The analysis of lipids also includes the examination of a possible accumulation or depletion of lipids during ageing at the lipid class level. In addition, the length and unsaturation of fatty acids are also analysed and linked to age-related changes. This chapter concludes by presenting the age-related changes observed in RNA molecules and their connection to protein expression 6.4, followed by a conclusion. Throughout the chapter, the young cohort represents the baseline, and the results are interpreted accordingly.

6.1 Proteins

The proteomics investigation via MS was conducted in an untargeted fashion and identified 2938 proteins (see Chapter 4.1). The analysis was performed on 12 myocardial samples belonging to the young cohort and 13 belonging to the old cohort. Based on explorative data analysis discussed in Chapter 5.8, samples did not exhibit excessive gender imbalance, and principal component analysis did not indicate any specific clustering in the samples used for proteome analysis (details on samples, sample size, experimental design and explorative data analysis can be found in Chapter 5). Comparing the proteome between the old and young cohorts (baseline), we identified 56 molecules showing evidence of dysregulation ($p_{fdr} < 0.1$) associated with the ageing process[77], 23 of which were considered age-related changes ($p_{fdr} < 0.05$) and potential biomarkers of cardiac ageing (see Chapter 4.8). The candidate biomarkers were about 0.78% of the total proteins measured in the study. Among the 23 candidate biomarkers representing ageing process, I display the first 10 in Table 6.1 and Figure 6.2. I expound on each of these biomarkers below.

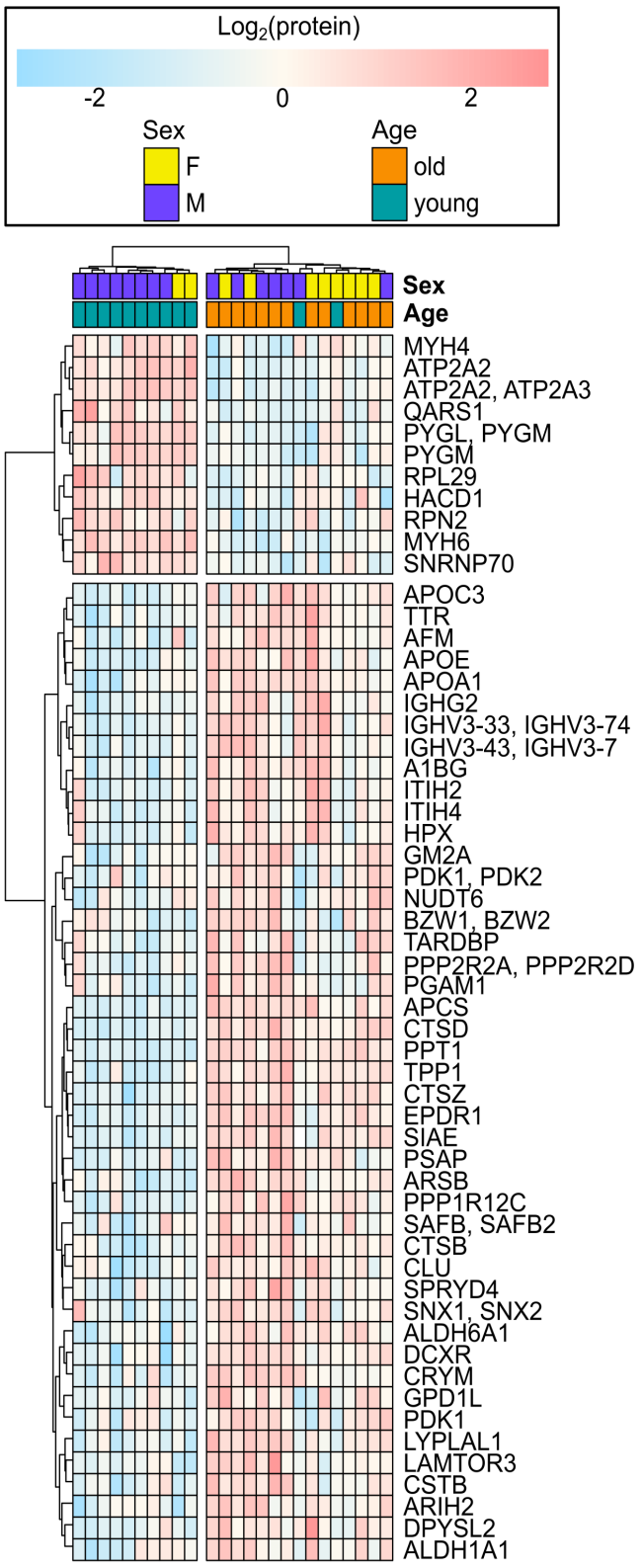


Figure 6.1: Proteomics. Heatmap.

The 56 molecules that indicated dysregulation ($p_{fdr} < 0.1$) associated with the ageing process[77] were utilised for constructing a heat map to explore possible patterns across samples and analytes accompanied by two unsupervised clustering procedures, one on the sample and the other on protein space (see Chapter 5.7). The heat map is shown in Figure 6.1. Exploring the outcome of the sample-wise clustering procedure, it can be understood that the cardiac tissues separate neatly into two groups. The first group, identified by the unsupervised clustering, contains all tissues belonging to the young donors except for two young samples that fall in the second cluster. Performing the analysis on all the proteins we could see a clear separation in two groups of the 56 proteins. The first group comprises proteins that have generally higher abundance in the older cohort with respect to the younger. Analysing the pattern of the second group suggests that it contains proteins with lower abundance in the older samples when compared to the younger ones.

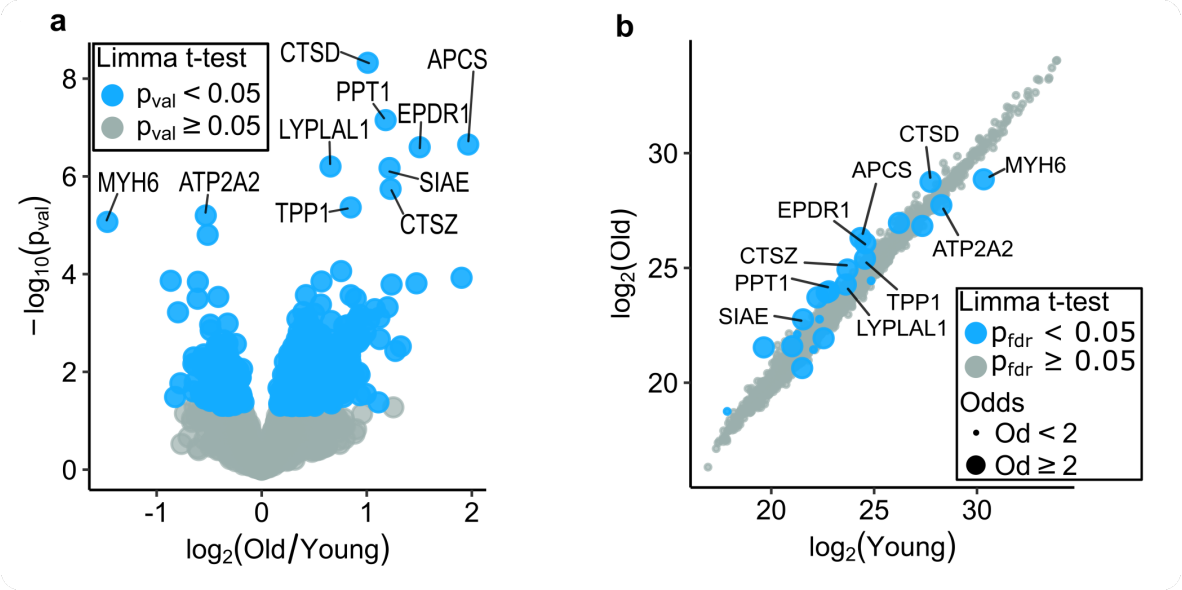


Figure 6.2: Proteomics. **(a)** Volcano plot. **(b)** Bubble plot.

Proteins	$\log_2$ **FC**	**t**	$[\mathbf{CI_l}, \mathbf{CI_u}]$	$\mathbf{p_{val}}$	$\mathbf{p_{fdr}}$
CTSD	1.0096	8.4980	[0.7656, 1.2536]	4.74E-09	1.39E-05
PPT1	1.1796	7.3655	[0.8507, 1.5084]	7.08E-08	0.000104
APCS	1.9656	6.9045	[1.3810 , 2.5503]	2.23E-07	0.000186
EPDR1	1.5039	6.8537	[1.0533, 1.9545]	2.53E-07	0.000186
LYPLAL1	0.6547	6.4964	[0.4478, 0.8617]	6.28E-07	0.00033
SIAE	1.2170	6.4683	[0.8306, 1.6034]	6.75E-07	0.00033
CTSZ	1.2265	6.0883	[0.8128, 1.6402]	1.80E-06	0.000756
TPP1	0.8469	5.7536	[0.5446, 1.1492]	4.32E-06	0.001587
ATP2A2	-0.5323	-5.609	[-0.7272, -0.3374]	6.32E-06	0.002063
MYH6	-1.4691	-5.4955	[-2.0181, -0.9202]	8.54E-06	0.002508

Table 6.1: Top 10 dysregulated proteins.

According to differential expression analysis, three components of cathepsin protein family were upregulated in the older hearts, namely cathepsin D (CTSD) ($FC = 2.01$, $p_{fdr} = 1.39x10^{-5}$, cathepsin Z (CTSZ) ($FC = 2.34$, $p_{fdr} = 7.56 \cdot 10{-}4$) and cathepsin B (CTSB) ($FC = 1.68$, $p_{fdr} = 2.1259 \cdot 10{-}3$). Cathepsin proteins are major components of the lysosomal proteolytic system[615], and their dysregulation has been linked to failing hearts and cardiac remodelling[132,578]. Palmitoyl-protein thioesterase 1 (PPT1) ($FC = 2.26$, $p_{fdr} = 1.04 \cdot 10{-}4$) removes long-chain fatty acids (LCFAs) from specific proteins[520] and found upregulated in the older heart. A recent study has suggested PPT1 as a possible biomarker for cardiovascular diseases[397]. Tripeptidyl peptidase 1 (TPP1) ($FC = 1.79$, $p_{fdr} = 1.587 \cdot 10{-}3$) and arylsulfatase B (ARSB) ($FC = 1.79$, $p_{fdr} = 3.97 \cdot 10{-}2$), two proteins involved in the breakdown of glycosaminoglycans[654], were detected in the old cohort. Two other proteins related to the lipids were found to be dysregulated in the old cohort: lysophospholipase-like 1 (LYPLAL1) ($FC = 1.57$, $p_{fdr} = 3.3 \cdot 10{-}4$) upregulated and glycogen phosphorylase L;glycogen phosphorylase, muscle associated (PYGL;PYGM) ($FC = 0.65$, $p_{fdr} = 2.6329 \cdot 10{-}2$) downregulated in the older cohort[223,298]. LYPLAL1 has recently been associated with cardiac hypertrophy[485], and PYGL;PYGM with glycogen storage disorder, which should not affect the heart, inasmuch as it is not the predominant isoform in the cardiac tissue[8]. Ganglioside GM2 activator (GM2A) ($FC = 3.73$, $p_{fdr} = 2.6329 \cdot 10{-}2$), generally associated with sphingolipidoses dysfunctions[735], was found upregulated with ageing. It is suggested that GM2A can be a possible biomarker in blood plasma for acute heart failure[809]. Ependymin-related 1 (EPDR1) ($FC = 2.83$, $p_{fdr} = 1.86 \cdot 10{-}4$) was upregulated in the cardiac tissue. EPDR1 is an adipokine secreted by brown adipose cells, which can regulate heart metabolism via endocrine. Researchers have been studying the potential of adipokines to reduce oxidative stress caused by their cardioprotective functions[688].

Sialic acid acetylesterase (SIAE) ($FC = 2.32$, $p_{fdr} = 3.3 \cdot 10{-}4$), an upregulated protein in the old heart tissues, regulates the stability of glycoprotein in cells[855]. While the role of sialic acids in heart failure has drawn the attention of researchers, its relation with cardiovascular diseases is not yet fully known[740]. Two apolipoproteins, Apolipoprotein C3 (APOC3) ($FC = 2.77$, $p_{fdr} = 2.6615 \cdot 10{-}2$) and Apolipoprotein E (APOE) ($FC = 2.35$, $p_{fdr} = 2.6615 \cdot 10{-}2$), were also found upregulated in the older hearts. Apolipoproteins can be synthesised or taken up by different types of cells and is associated with the regulation of cardiometabolism[126,491,526] and cardiovascular diseases[524,539]. Amyloid P component, serum (APCS) ($FC = 3.90$, $p_{fdr} = 1.86 \cdot 10{-}4$) was also more abundant in older donor hearts. APCS can be found in the cardiac extracellular matrix and is critical for forming amyloid plaques, which can lead to cardiac amyloidosis[440].

Amongst the downregulated proteins, two belonged to the myosin heavy chains family, myosin heavy chain 6 (MYH6) ($FC = 0.36$, $p_{fdr} = 2.508 \cdot 10{-}3$) and myosin heavy chain 4 (MYH4) ($FC = 0.75$, $p_{fdr} = 4.0683 \cdot 10{-}2$), which are pivotal for cardiac muscle contraction[471]. Furthermore, two ATPases, ATPase sarcoplasmic/endoplasmic reticulum Ca2+ transporting 2 (ATP2A2 or SERCA2) ($FC = 0.69$, $p_{fdr} = 2.063 \cdot 10{-}3$) and ATPase sarcoplasmic/endoplasmic reticulum Ca2+ transporting 2; 3 (ATP2A2; ATP2A3 or SERCA2; SERCA3) ($FC = 0.70$, $p_{fdr} = 4.189 \cdot 10{-}3$), indicated downregulation in the older cohort. They are both responsible for transporting calcium from the cytoplasm to the sarcoplasmic

reticulum[522]. Literature suggests that a higher abundance of SERCA2 restores cardiac contractility to baseline and can be cardioprotective in heart failure[411], while the function of SERCA3 is still unclear due to its low affinity to calcium[411,665,945].

Ribosomal protein L29 (RPL29) ($FC = 0.54$, $p_{fdr} = 2.6329 \cdot 10{-}2$) showed evidence of downregulation in the older ventricular tissues. It is responsible for ribosomal biogenesis and is recently linked to congenital cardiac diseases, although further research is required[679]. Overexpression of SPRY domain containing 4 (SPRYD4) ($FC = 1.48$, $p_{fdr} = 2.6329 \cdot 10{-}2$) was observed in the older cohort. SPRYD4 is a nuclear protein recently found overexpressed in blood cells of patients with end-stage heart failure[795]. Sorting nexins 1; 2 (SNX1; SNX2) ($FC = 1.33$, $p_{fdr} = 3.97 \cdot 10{-}2$) also displayed evidence of upregulation in older hearts; its role is essential for regulating protein trafficking in the cytosol, and their dysregulation has been linked to cardiovascular diseases[912]. Ariadne RBR E3 ubiquitin protein ligase 2 (ARIH2) ($FC = 1.85$, $p_{fdr} = 4.0683 \cdot 10{-}2$) was found in higher abundance in the older cohort. ARIH2 is associated with the immune response of the cardiac tissue. When statin is applied to reduce cholesterol in individuals with type 2 diabetes, the gene site of ARIH2 undergoes higher methylation[722], suggesting that ARIH2 might play an active role in diabetes. Finally, we found dysregulation of small nuclear ribonucleoprotein U1 subunit 70 (SNRNP70) ($FC = 0.65$, $p_{fdr} = 4.0683 \cdot 10{-}2$). Recent evidence points to the upregulation of SNRNP in patients with type 2 diabetes[253], although further research is needed[307].

6.2 Metabolites

The characterisation of metabolites in old and young cardiac tissues was performed using a semi-targeted MS approach (see Chapters 4.2 and 5.4). The analysis identified 150 metabolites on 12 young and 13 old donor heart tissues. No evidence of gender grouping was detected. The analysis of the molecular signature in metabolites in cardiac ageing identified 55 metabolites with evidence of dysregulation ($p_{fdr} < 0.1$), of which 47 were ($p_{fdr} < 0.05$) potential biomarkers. Table 6.2 presents the top 10 dysregulated metabolites, which I further explain below.

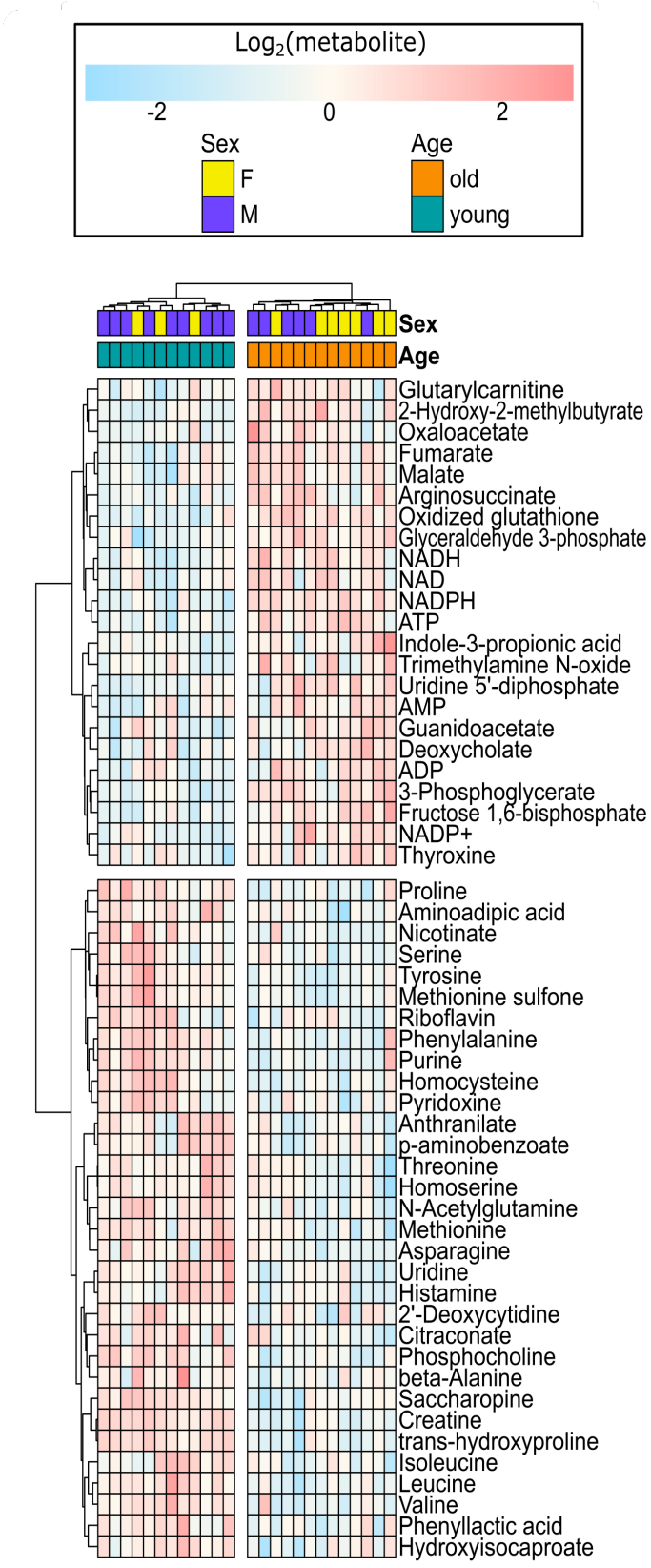


Figure 6.3: Metabolomics. Heatmap.

In order to visually evaluate the expression patterns of the 55 metabolites that exhibit evidence of dysregulation ($p_{fdr} < 0.1$), we constructed a heat map on their detected measurements, which is presented in Figure 6.3. The sample-wise clustering procedure unveiled two main clusters in the data, one for the old cohort and one for the young cohort. Likewise, the application of unsupervised clustering on the metabolites space revealed two distinct groups. The first group comprised analytes with higher expression in the older cohort, while the second group comprised analytes with higher measured abundance in the young donors.

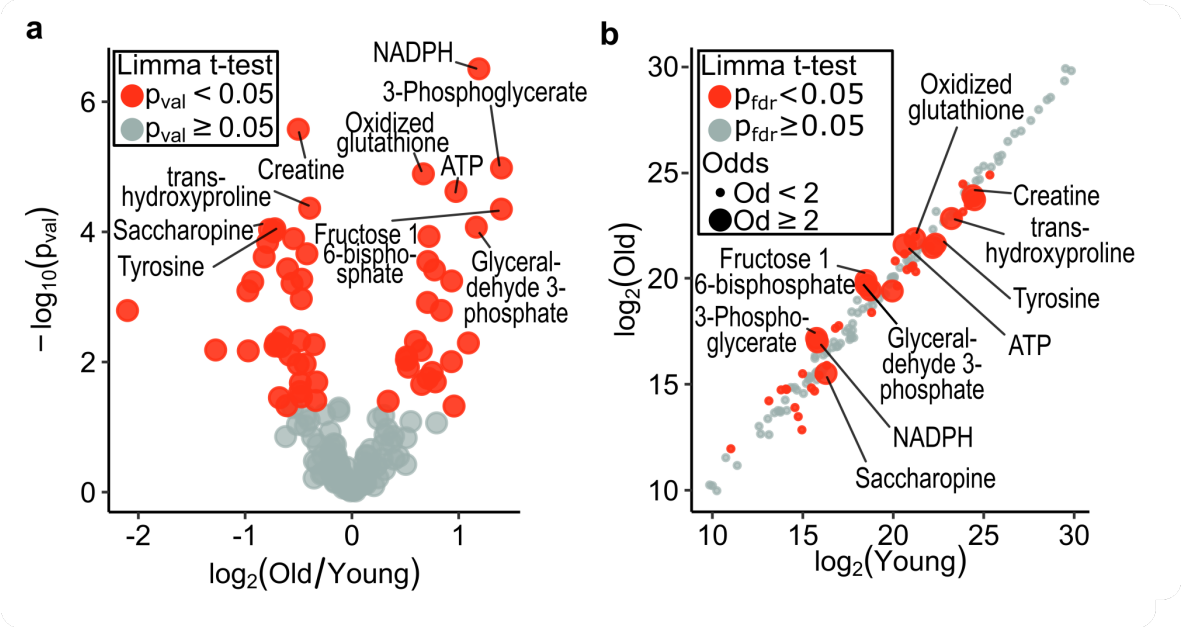


Figure 6.4: Metabolomics. **(a)** Volcano plot. **(b)** Bubble plot.

Metabolites	$\log_2$ **FC**	**t**	$[\mathbf{CI_l}, \mathbf{CI_u}]$	$\mathbf{p_{val}}$	$\mathbf{p_{fdr}}$
NADPH	1.1889	6.8450	[0.8316, 1.5462]	3.14E-07	4.71E-05
Creatine	-0.5024	-5.9881	[-0.6749, -0.3298]	2.70E-06	0.000203
3 Phosphoglycerate	1.4006	5.4631	[0.8732, 1.9281]	1.05E-05	0.000486
Oxidized glutathione	0.6697	5.3811	[0.4137, 0.9257]	1.29E-05	0.000486
Adenosine triphosphate	0.9742	5.1442	[0.5846, 1.3638]	2.40E-05	0.000721
Trans-hydroxyproline	-0.3958	-4.9120	[-0.5616, -0.2300]	4.41E-05	0.000961
Fructose 1,6 bisphosphate	1.3996	4.9057	[0.8127, 1.9866]	4.48E-05	0.000961
Glyceraldehyde 3 phosphate	1.1651	4.6623	[0.6510, 1.6792]	8.48E-05	0.00137
Tyrosine	-0.7218	-4.6392	[-1.0419, -0.4017]	9.01E-05	0.00137
Saccharopine	-0.7715	-4.6294	[-1.1143, -0.4287]	9.25E-05	0.00137

Table 6.2: Top 10 dysregulated metabolites.

Amongst these age-associated biomarkers, multiple amino acids[114] changed their abundance during the ageing process. Multiple proteinogenic amino acids were downregulated: tyrosine ($FC = 0.60$, $p_{fdr} = 1.37 \cdot 10^{-3}$); leucine ($FC = 0.68$, $p_{fdr} = 1.48 \cdot 10^{-3}$); valine ($FC = 0.65$, $p_{fdr} = 3.048 \cdot 10^{-3}$); phenylalanine ($FC = 0.72$, $p_{fdr} = 3.994 \cdot 10^{-3}$); methionine

($FC = 0.52$, $p_{fdr} = 3.994 \cdot 10^{-3}$); isoleucine ($FC = 0.66$, $p_{fdr} = 2.9347 \cdot 10^{-2}$), proline ($FC = 0.50$, $p_{fdr} = 5.043 \cdot 10^{-3}$); threonine ($FC = 0.61$, $p_{fdr} = 2.3322 \cdot 10^{-2}$); asparagine ($FC = 0.78$, $p_{fdr} = 2.3947 \cdot 10^{-2}$); and serine ($FC = 0.70$, $p_{fdr} = 3.6239 \cdot 10^{-2}$). On the other hand, thyroxine ($FC = 1.56$ $p_{fdr} = 2.5938 \cdot 10^{-2}$) was found in higher abundance in aged hearts. Hydroxyisocaproate ($FC = 0.51$, $p_{fdr} = 2.5938 \cdot 10^{-2}$), a metabolite involved in the leucine and valine metabolism was lower in older hearts, in accordance with valine and leucine. In addition, arginosuccinate ($FC = 1.78$, $p_{fdr} = 8.663 \cdot 10^{-3}$), an intermediate of arginine and proline[306], was found upregulated in older cardiac tissues. Similarly, histamine ($FC = 0.71$ $p_{fdr} = 2.3322 \cdot 10^{-2}$), a key intermediate of histidine metabolism, was found upregulated in older hearts; it is reported that heart failure patients have displayed higher concentrations of histamine[425]. N-acetylglutamine ($FC = 0.56$, $p_{fdr} = 2.283 \cdot 10^{-3}$), related to glutamine and glutamate metabolism, was downregulated in older hearts. Homoserine ($FC = 0.60$, $p_{fdr} = 2.4227 \cdot 10^{-2}$) was lower in the older cohort. It plays an important role in serine, glycine and threonine metabolism, and its low abundance has been detected in multiple metabolic disorders. It is therefore suggested that homoserine can be a possible biomarker for cardiac metabolic diseases[632].

Multiple downregulations were also recorded across non-proteinogenic amino acids: trans-hydroxyproline ($FC = 0.76$, $p_{fdr} = 9.61 \cdot 10^{-4}$), a metabolite part of collage, and homocysteine ($FC = 0.67$, $p_{fdr} = 4.062 \cdot 10^{-3}$), a well-established biomarker for cardiovascular diseases[303]. Links that non-proteinogenic amino-acid intake with cardiovascular diseases have been reported[813]. Additionally, phenyllactic acid ($FC = 0.41$, $p_{fdr} = 2.5938 \cdot 10^{-2}$), an acid related to phenylalanine metabolism, was downregulated in the older cohort.

Across the age-associated changes, a clear decrease of saccharopine ($FC = 0.58$, $p_{fdr} = 1.37 \cdot 10^{-3}$) was observed in the older cohort of tissue. Saccharopine is a key intermediate in lysine degradation and methionine sulfone, where the latter is an oxidised derivative of methionine[947]. Similarly to saccharopine, methionine sulfone ($FC = 0.60$ $p_{fdr} = 1.37 \cdot 10^{-3}$) was also identified downregulated in the old cohort. Glutarylcarnitine ($FC = 1.90$, $p_{fdr} = 3.4462 \cdot 10^{-2}$) was found more abundant, while aminoadipic acid ($FC = 0.73$, $p_{fdr} = 3.6239 \cdot 10^{-2}$) was less abundant in the older cohort. Phosphocholine ($FC = 0.74$, $p_{fdr} = 2.195 \cdot 10^{-3}$), a metabolite involved in the synthesis of glycerophosphocholine, was also downregulated in the old donors. Higher levels of glycerophosphocholine were associated with early development of cardiovascular diseases[794]. Moreover, an increase of 2-hydroxy-2-methylbutyrate ($FC = 1.70$ $p_{fdr} = 3.048 \cdot 10^{-3}$) was found in the older cohort. The p-aminobenzoate ($FC = 0.63$, $p_{fdr} = 2.4227 \cdot 10^{-2}$) was downregulated in the older cohort.

Antioxidants and redox homeostasis are essential for maintaining healthy ageing (see Chapter 2.4), and multiple analytes involved in maintaining a reduced-oxidised environment were dysregulated in the aged cohort. For instance, reduced nicotinamide adenine dinucleotide phosphate (NADPH) ($FC = 2.27$ $p_{fdr} = 4.71 \cdot 10^{-5}$) was more abundant in the older myocardial tissues. Moreover, its oxidised form (NADP), appeared upregulated ($FC = 1.63$ $p_{fdr} = 2.525 \cdot 10^{-3}$). To assess if there was any imbalance between NADP and NADPH with ageing, their ratio was compared across the two cohorts (see 5.7). The NADP/NADPH ratio exhibited evidence of a decrease with ageing ($p_{val} = 0.03247$), suggesting that the old hearts

contained more NADPH than NADP. The NADP/NADPH is displayed in Figure 6.5a. The accumulation of NADPH plays a pivotal role in ROS defence[233,897]. Oxidised glutathione (GSSG) was found upregulated ($FC = 1.59$ $p_{fdr} = 4.86 \cdot 10^{-4}$) in older donors. Comparing GSSG with its reduced form (see Figure 6.5d) reveals a high imbalance in their ratio toward the oxidised form ($p_{val} = 0.00025$). Accumulation of GSSG can lead to oxidative stress, a well-known age-related phenomenon[897] (see Chapter 2.4).

Amongst the several age-related changes observed in our data, multiple molecules were central to energy production and maintenance of the heart (see Chapter 2.3). Indeed, creatine ($FC = 0.70$, $p_{fdr} = 2.03 \cdot 10^{-4}$) was less abundant in the older group, while guanidoacetate ($FC = 1.42$ $p_{fdr} = 3.3407 \cdot 10^{-2}$), precursor of creatine, was more abundant in the older hearts, suggesting a possible aberrant functioning of creatine synthesis machinery. Similar patterns were observed in heart failure[182]. Phosphocreatine (PCr) did not exhibit any age-related changes. Notably, however, adenosine triphosphate (ATP) ($FC = 1.96$ $p_{fdr} = 7.21 \cdot 10^{-4}$) and adenosine diphosphate (ADP) ($FC = 1.44$ $p_{fdr} = 3.9228 \cdot 10^{-2}$) showed an increase in abundance in the older cohort. The ADP/ATP ratio was analysed in order to interpret the results and determine if this increase was an accumulation rather than a change in energy supply. As presented in Figure 6.5c, the results suggest that ATP production/consumption did not increase since the ADP/ATP ratio did not show any change ($p_{val} = 0.1289$) and that a constant accumulation occurred with ageing.

Furthermore, nicotinamide adenine dinucleotide (NAD) ($FC = 1.42$ $p_{fdr} = 3.0709 \cdot 10^{-2}$) showed evidence of upregulation in the older donors, and so did its reduced form (NADH) ($FC = 1.64$ $p_{fdr} = 1.477 \cdot 10^{-3}$). Nicotinate ($FC = 0.23$ $p_{fdr} = 8.663 \cdot 10^{-3}$), a closely related acid to NAD, was found instead low in the aged cohort. From the NAD/NADH ratio analysis (see Figure 6.5b), strong evidence of imbalance was observed in the older cohort ($p_{val} = 0.0129$), with the ratio lower in the old myocardial tissues. Accumulation of NADH with ageing can lead to inhibition of enzymes that produce NADH from NAD, with subsequent increase of reactive oxygen species, which promotes oxidative stress[436]; this is usually observed in diabetic patients[890]. To further assess if any changes occurred in the balance of consumption/supply machinery, the PCr/ATP ratio was compared between the young and old[179]. A decrease in the ratio was observed in the older cohort, which suggests a possible dysregulation of a ready supply of energy via the creatine-kinase phosphagen shuttling mechanism[916].

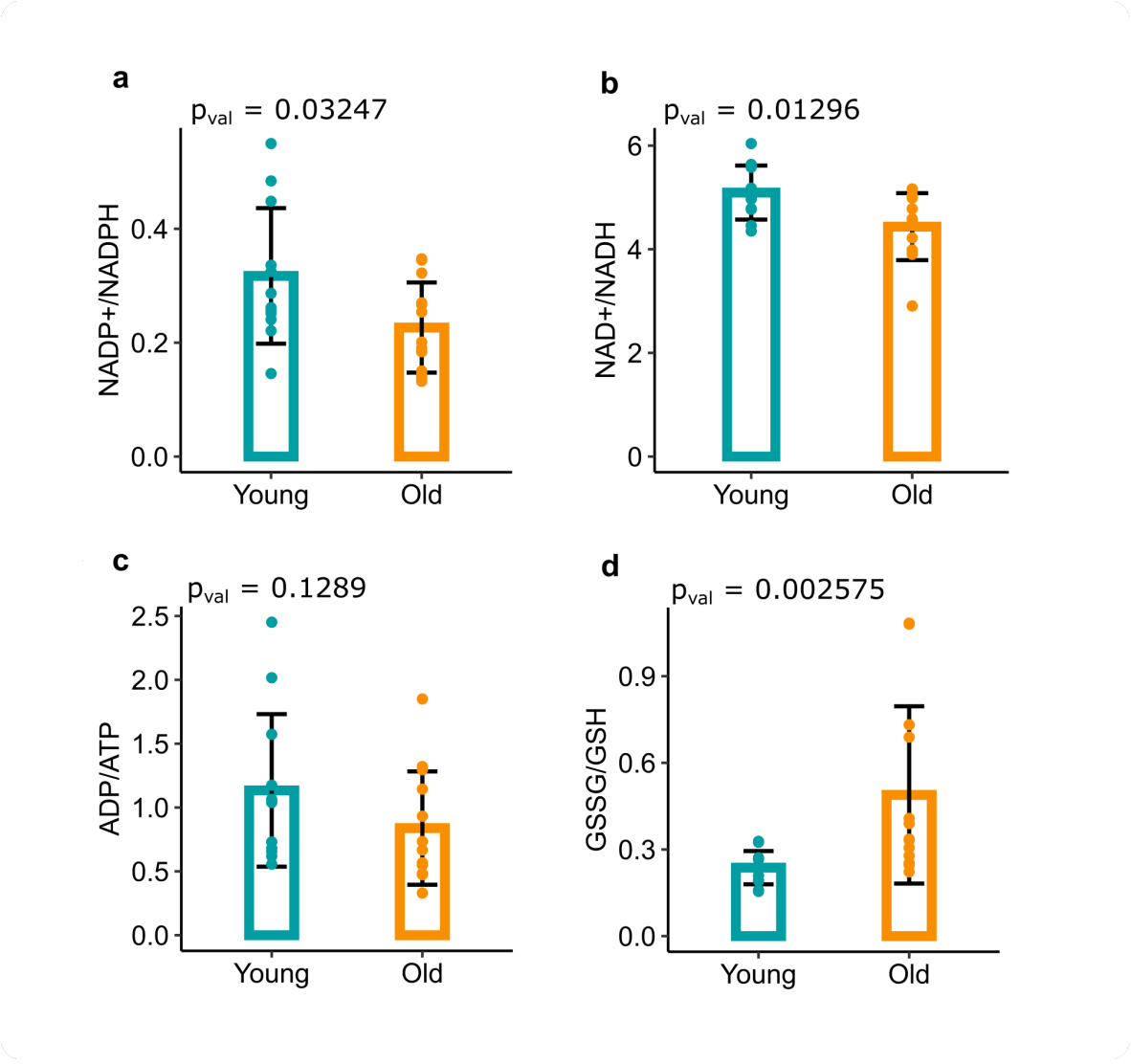


Figure 6.5: Metabolomics. **(a)** NADP/NADPH ratio. **(b)** NAD/NADH ratio. **(c)** ADP/ATP ratio. **(d)** GSSG/GSH ratio

Some intermediates of the central carbon metabolism were altered in the older cohort of donors. Fructose 1,6-bisphosphate (F1,6BP) ($FC = 2.63$ $p_{fdr} = 9.61 \cdot 10^{-4}$), glyceraldehyde 3-phosphate (GAP) ($FC = 2.24$ $p_{fdr} = 1.37 \cdot 10^{-3}$) and 3-Phosphoglycerate (3PG) ($FC = 2.64$ $p_{fdr} = 4.86 \cdot 10^{-4}$) are substrates present in glycolysis and exhibited a higher abundance in the older hearts. This coordinated increase indicates a generalised higher activity in the aged cohort, as already discussed in the literature (see Chapter 2.4). An increase in glycolysis is also observed in heart failure[515]. Some intermediates of the TCA cycle (see Chapter 2.3) were recorded as overabundant in the aged cohort. In fact, fumarate ($FC = 1.63$, $p_{fdr} = 7.025 \cdot 10^{-3}$) and malate ($FC = 1.51$, $p_{fdr} = 2.3322 \cdot 10^{-2}$) were both upregulated in the older cohort. An increase of TCA intermediate fluxes [i] points to longevity

[i] A flux can be informally defined as a substrate in - substrate out.

observed in human and rat models[85], while TCA intermediates are downregulated in heart failure[515].

Uridine 5'-diphosphate ($FC = 1.91$, $p_{fdr} = 3.994 \cdot 10^{-3}$) and purine ($FC = 0.72$, $p_{fdr} = 6.449 \cdot 10^{-3}$) were found upregulated, while uridine ($FC = 0.57$, $p_{fdr} = 1.559 \cdot 10^{-3}$) downregulated in the older cohort. The three intermediates belong to purine and pyrimidine metabolism. A recent study has associated perturbation of purine and pyrimidine metabolism with cardiac hypertrophy[863].

Moreover, perhaps reflecting an ageing-related change to the gut microbiome, there was an increase in metabolites created by the gut microbiota in the old cohort. Indole-3-propionic acid (IPA) ($FC = 2.12$, $p_{fdr} = 2.3322 \cdot 10^{-2}$) significantly increased in older hearts as well. IPA is a microbial tryptophan derivative and has been recognised as a key protective factor for various cardiovascular diseases, including atherosclerosis and heart failure, with preserved ejection fraction[904]. A sharp increase was also observed for trimethylamine N-oxide (TNO) ($FC = 1.68$, $p_{fdr} = 4.6236 \cdot 10^{-2}$), an atherogenic metabolite, in older hearts[65]. The aged cohort witnessed downregulation of anthranilate ($FC = 0.63$, $p_{fdr} = 2.1445 \cdot 10^{-2}$). Being part of tryptophan metabolism, the decreased activity of anthranilate is associated with heart failure, while dietary intervention has shown promising results to prevent such risk[672].

6.3 Lipids

Lipidomics analysis was performed using untargeted liquid chromatography-MS on 12 young and 11 old myocardial samples. We identified a total of 341 lipids, 11 of which showed evidence of possible dysregulation ($p_{fdr} < 0.1$) with ageing. Among the 11 lipids, 7 lipids were identified as candidate biomarkers of human cardiac ageing ($p_{fdr} < 0.05$). The initial explorative analysis did not highlight any cluster due to biological sex.

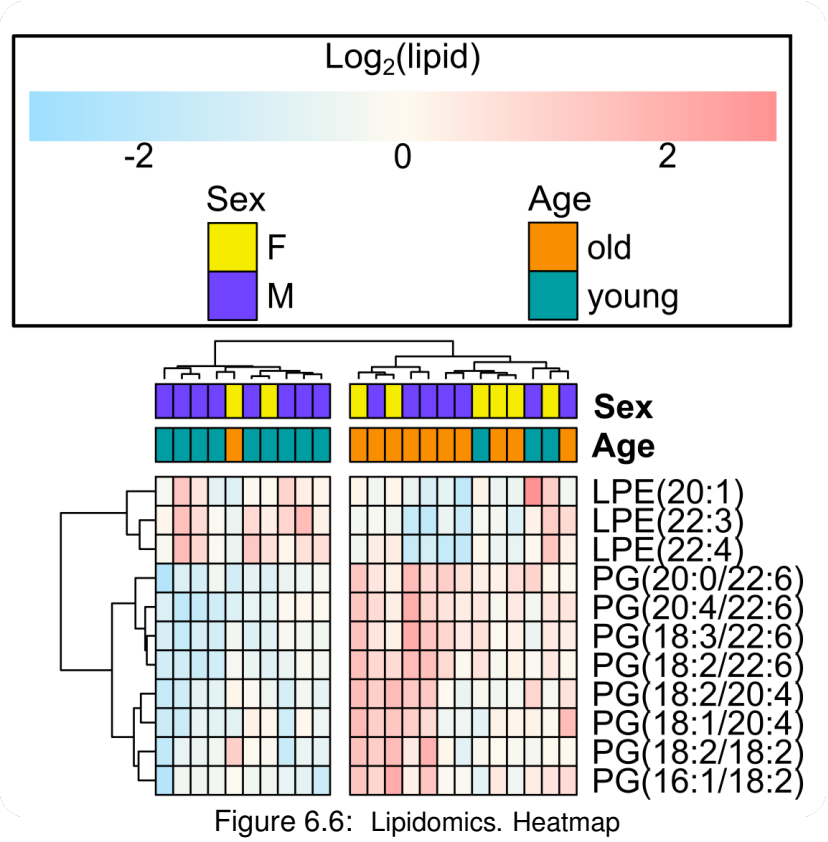


Figure 6.6: Lipidomics. Heatmap

Amongst the most age-related changes recorded at the lipid level, we observed a contradistinction pattern in the data, as shown in the heat map (see Figure 6.6). Applying the unsupervised clustering procedure on the samples, we detected the sample classified into two main groups: the first group with all younger donors except for the 1 sample, and the second group mainly with old donors, except for the 3 young samples. The unsupervised clustering performed on the lipids with $p_{val} < 0.1$ identifies two distinct groups of lipids. The first group contained lipids that were less abundant in the older cohort than in the younger, while the second group comprised the lipids more abundant in the older cohort than in the younger. Notably, the first group contained all lysophosphatidylethanolamines (LPE), while the second group comprised all phosphatidylcholine lipids (see Chapter 2.3).

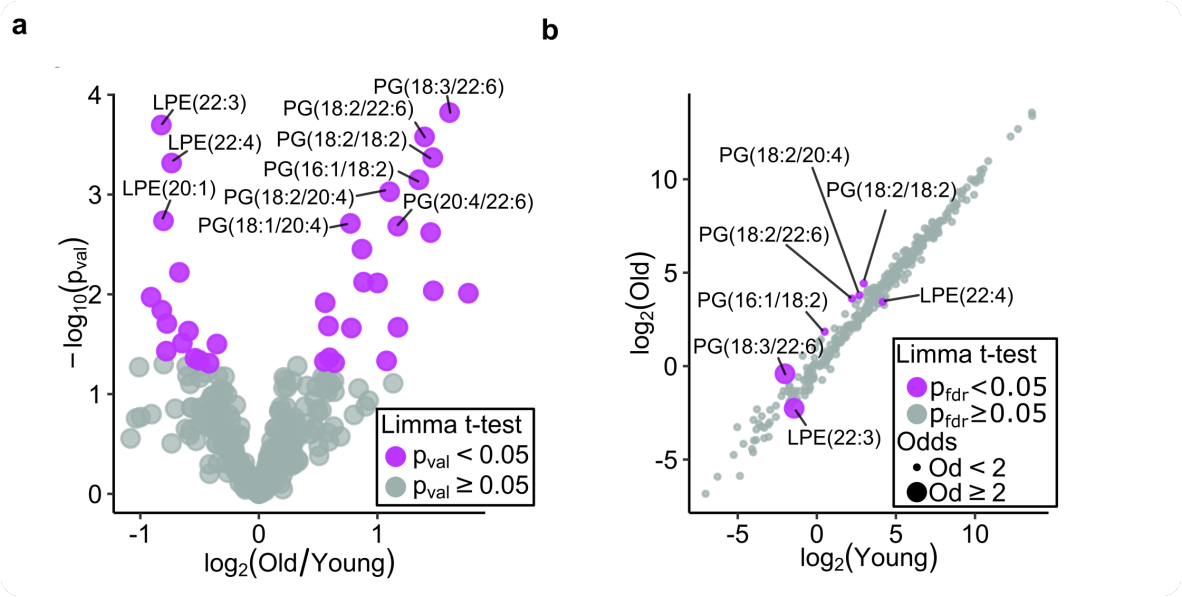


Figure 6.7: Lipidomics. Lipidomics. **(a)** Volcano plot. **(b)** Bubble plot.

Lipids	$\log_2$ **FC**	**t**	$[\mathbf{CI_l}, \mathbf{CI_u}]$	$\mathbf{p_{val}}$	$\mathbf{p_{fdr}}$
PG(18:3/22:6)	1.6085	4.4985	[0.8703, 2.3467]	0.000151	0.030033
LPE(22:3)	-0.8231	-4.3847	[-1.2108, -0.4355]	0.000201	0.030033
PG(18:2/22:6)	1.3978	4.2770	[0.7230, 2.0726]	0.000264	0.030033
PG(18:2/18:2)	1.4665	4.0881	[0.7259, 2.2072]	0.000426	0.032958
LPE(22:4)-H	-0.7362	-4.0382	[-1.1126, -0.3598]	0.000483	0.032958
PG(16:1/18:2)	1.3492	3.8853	[0.6322, 2.0662]	0.00071	0.040358
PG(18:2/20:4)	1.1033	3.7745	[0.4998, 1.7068]	0.000938	0.045672
LPE(20:1)	-0.8047	-3.5053	[-1.2787, -0.3307]	0.001831	0.070709
PG(18:1/20:4)	0.7722	3.4805	[0.31414, 1.2302]	0.001946	0.070709
PG(20:4/22:6)	1.1717	3.4548	[0.4715, 1.8720]	0.002074	0.070709

Table 6.3: Top 10 dysregulated lipids.

Several lipids with $p_{fdr} < 0.05$ were identified as age-related biomarkers: 5 upregulated and 2 downregulated in the older cohort. From the results summarised in Table 6.3 and Figure 6.7, a clear pattern can be observed. The 5 upregulated lipids are phosphatidylglycerol (PG/LPG), namely: PG(18:3/22:6) ($FC = 3.04$, $p_{fdr} = 3.0 \cdot 10^{-2}$); PG(18:2/22:6) ($FC = 2.63$, $p_{fdr} = 3.0 \cdot 10^{-2}$); PG(18:2/18:2) ($FC = 2.76$, $p_{fdr} = 3.29 \cdot 10^{-2}$); PG(16:1/18:2) ($FC = 2.54$, $p_{fdr} = 4.03 \cdot 10^{-2}$); and PG(18:2/20:4) ($FC = 2.14$, $p_{fdr} = 4.56 \cdot 10^{-2}$).

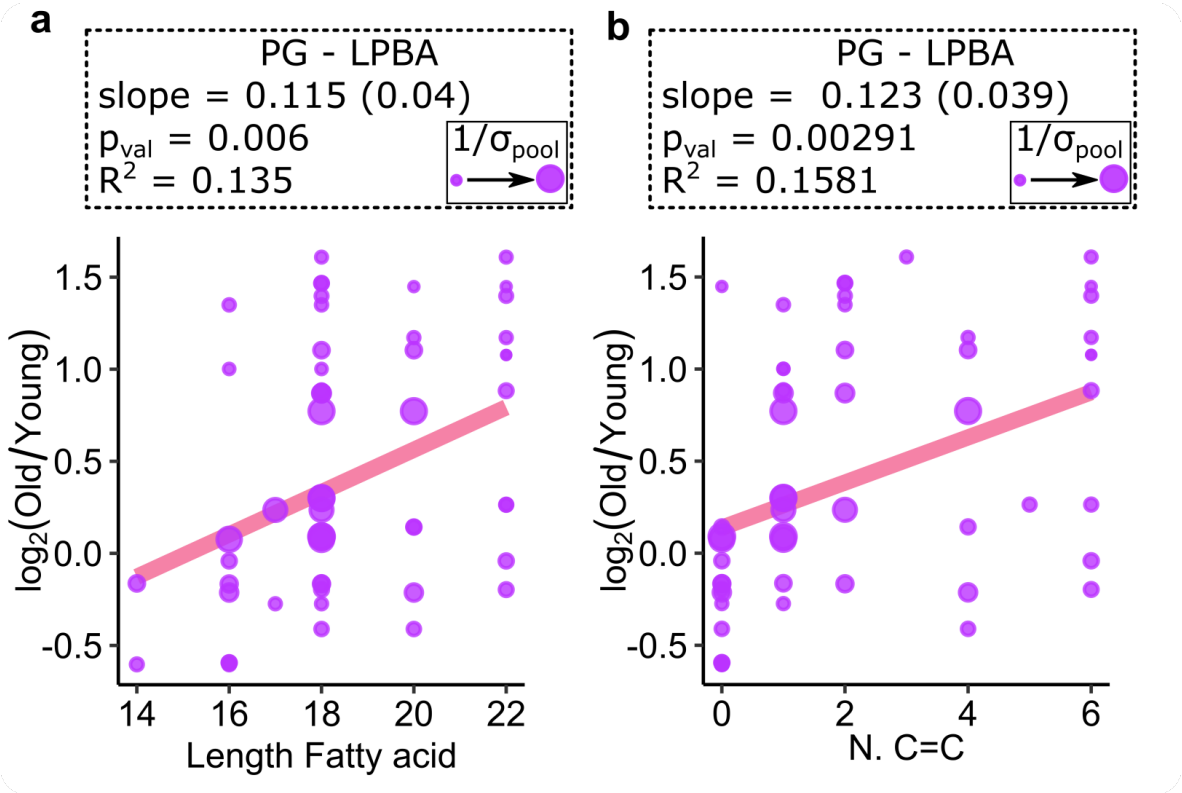


Figure 6.8: Lipidomics - Cer. **(a)** Analysis of fatty acid tail. **(b)** Analysis of polyunsaturated fatty acids.

Although there is no direct relation between the higher abundance of PG and cardiovascular diseases[548], it is known that heart failure can impair the cardiolipin abundance in cardiac tissue, which requires PG for the synthesis process[651]. A deeper look at the upregulated lipids also unveils that all the PGs comprise long and very long fatty acid tails (see Chapter 2.3). This hypothesis was tested with regression analysis (see Chapter 5.7), which indicated the increase of expression in PG with longer chained fatty acids as the cardiac tissue ages ($p_{val=0.006}$). The generalised increase of PG expression in the older cohort was also noted in polyunsaturated PG ($p_{val} = 0.0029$). Both results are displayed in Figure 6.8. With regard to the downregulated lipids in the aged hearts, all of them belonged to the class of LPE, namely: LPE(22:3) ($FC = 0.56$, $p_{fdr} = 3.0 \cdot 10^{-2}$) and LPE(22:4) ($FC = 0.60$, $p_{fdr} = 3.29 \cdot 10^{-2}$). Recent evidence has suggested that a higher LPE abundance is associated with heart failure[786].

Further investigation between lipid classes, length and polyunsaturation of the fatty acid chains were conducted (see Chapter 2.3 for their importance for the cardiac tissue and Chapter 5.7 for mathematical details). Diacylglycerols (DG) exhibited evidence of a decrease in their abundance in the older cohort as the fatty acid tails became longer ($p_{val} = 0.0323$). Moreover, abundance in polyunsaturated DGs decreased in the older donors ($p_{val} = 0.02514$), as presented in Figure 6.9

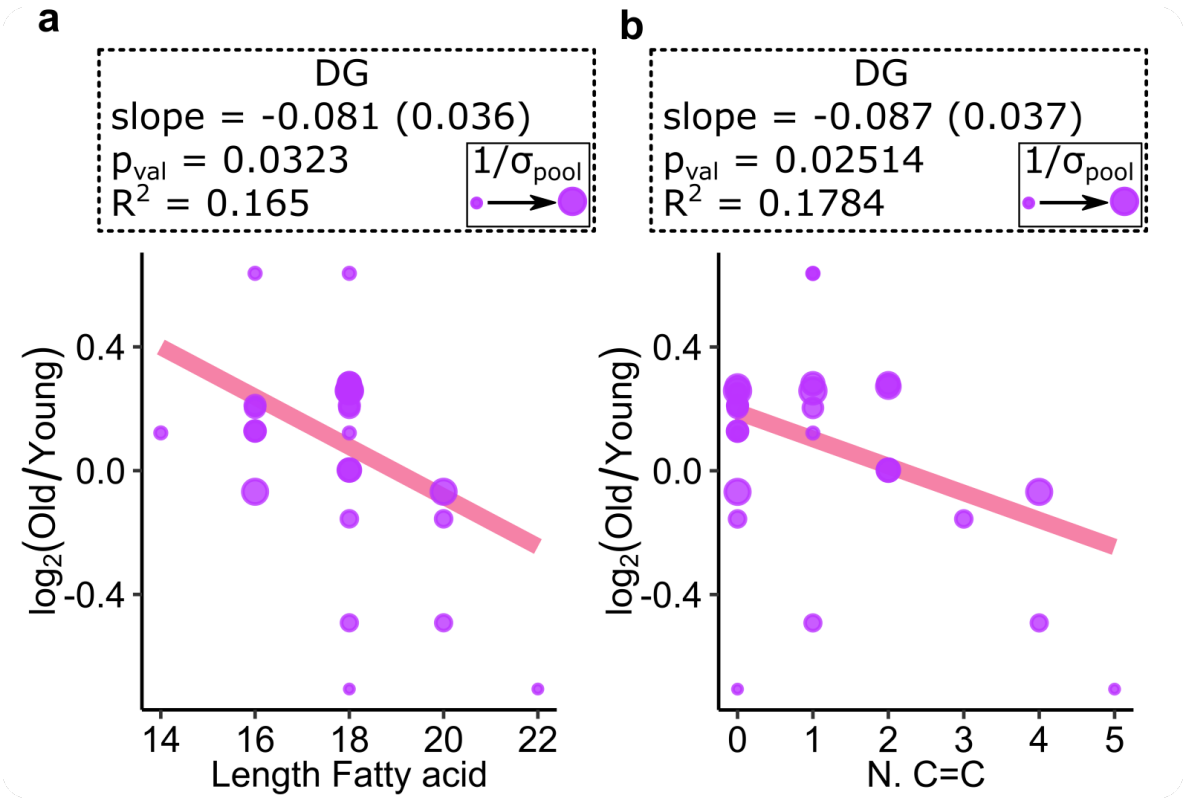


Figure 6.9: Lipidomics - DG. **(a)** Analysis of fatty acid tail. **(b)** Analysis of polyunsaturated fatty acids.

In order to understand whether the DG dynamics were inversely correlated to triglycerides (TGs) and acylcarnitines (AcCAs), a similar analysis for the latter lipid classes was performed. As a result, both AcCAs and TGs did not exhibit any evidence of a linear trend between their change in abundance and the length of the fatty acid tails in the old cohort. On the other hand, $\log_2 FC$ of AcCAs showed a strong correlation with the amount of polyunsaturation ($p_{val} = 2.02 \cdot 10^{-6}$) present on their tails. A similar pattern was also observed in the TGs when changes between young and old hearts were coupled to the polyunsaturations of AcCAs fatty acids ($p_{val} = 5.46 \cdot 10^{-3}$).

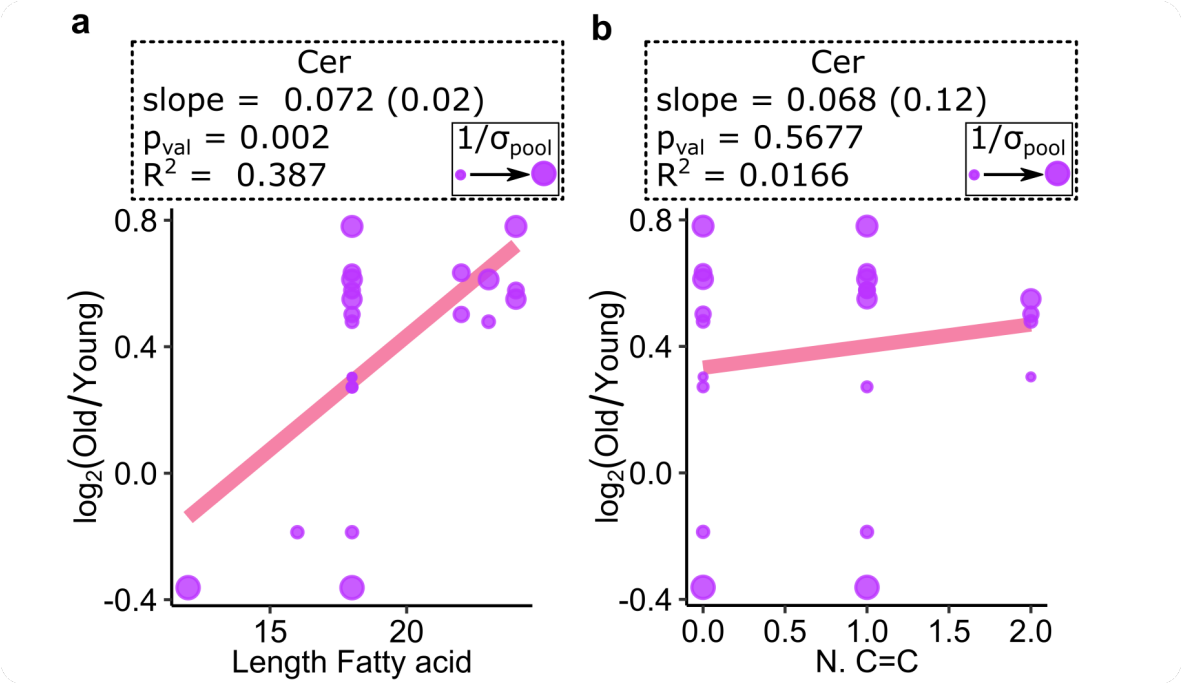


Figure 6.10: Lipidomics - Cer. **(a)** Analysis of fatty acid tail. **(b)** Analysis of polyunsaturated fatty acids.

Furthermore, extending the analysis for ceramides (Cer), lipids that affect lipotoxicity (see Chapter 2.3), we observed Cers' $\log_2 FC$ showing a positive linear trend with the length of the fatty acids, which suggests that Cer with longer fatty acids chains are more abundant in the older cohort rather than in the younger one ($p_{val} = 0.002$), but no change in fatty acids unsaturation (see Figure 6.10).

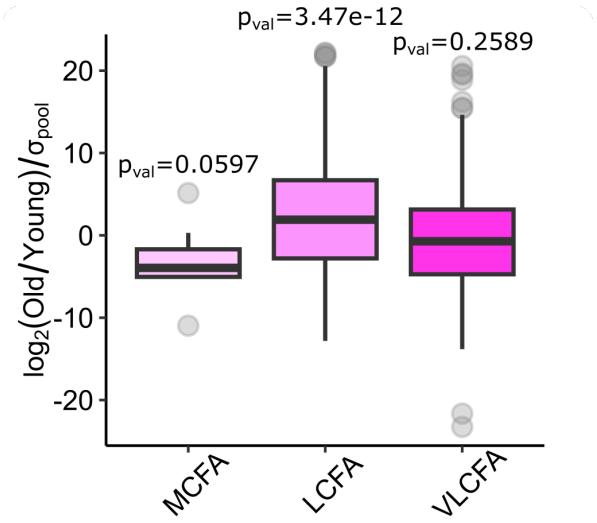


Figure 6.11: Lipidomics. Analysis of fatty acids length

Further analysis of the overall length of fatty acid tails in ageing was also conducted to understand the overall impact that lipid metabolism can have in the older cohorts. The

analysis showed that long-chain fatty acids tended to be more abundant in the older cohort ($p_{val} = 3.47 \cdot 10^{-12}$), while medium-chain fatty acids showed weak evidence of decreasing in their abundance in the older cohort ($p_{val} = 0.0597$) (see Chapter 2.3).

While single molecular investigation can be helpful for identifying biomarkers, lipid analysis based on the class can shed light on the possible accumulation of specific lipids during the ageing process, which can have a strong impact on one's health (see Chapter 2.3). Enrichment analysis identified 6 lipid classes that exhibited overall changes across the younger and older cohorts, and the results are shown in Figure 6.12 and Table 6.4. Consistent with the individual lipid analysis, phosphatidylglycerol & lysobisphosphatidic acid (PG & LBPA), with LBPA being a late endosome and lysosome-specific phospholipid[366], were increased in the old cohort. AcCAs, which are essential in shuttling fatty acids into the mitochondria for β-oxidation[11], were also statistically enriched in the older cohort. Cers, predictive biomarkers of cardiometabolic complications[137], and TGs, which are stored in the heart under metabolic stress, including obesity and T2D[281], and demonstrate a clear association with atherogenic CVD[800], were also increased in the old hearts.

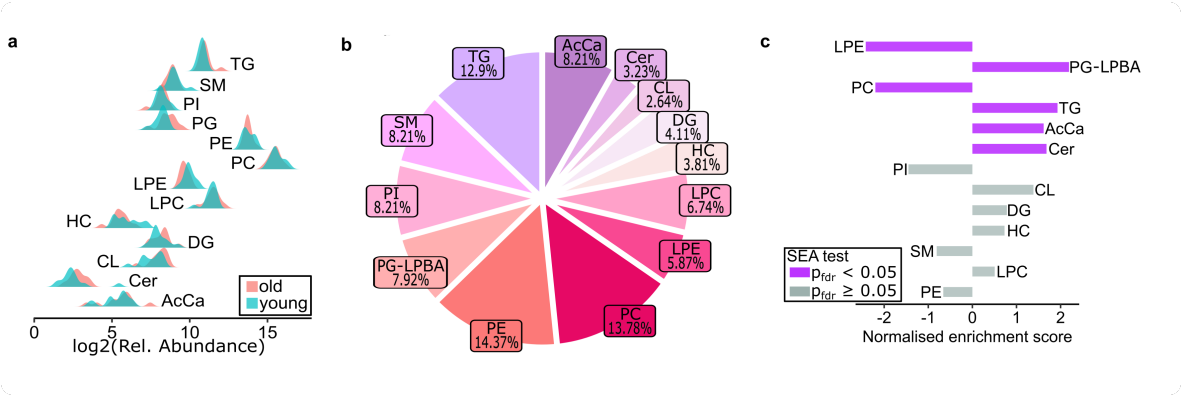


Figure 6.12: Lipidomics. **(a)** Lipids' class $\log_2$ abundance. **(b)** Composition lipids dataset. **(c)** Lipids enrichment analysis.

Lipid class	ES	$\mathbf{p_{val}}$	$\mathbf{p_{adj}}$	Set size
LPE	-0.8263	4.71E-08	6.12E-07	20
PG-LBPA	0.760883	7.53E-07	4.89E-06	27
PC	-0.61561	1.51E-06	6.53E-06	47
TG	0.60671	0.000107	0.000347	44
AcCA	0.555874	0.013349	0.028924	28
Cer	0.716934	0.011954	0.028924	11
PI	-0.45434	0.03563	0.06617	28
CL	0.623662	0.11413	0.185462	9
DG	0.313723	0.751712	0.974799	14
HC	0.297163	0.82483	0.974799	13

Table 6.4: Top 10 dysregulated lipid classes.

6.4 Transcripts

Transcriptomics analysis of the myocardial tissue was conducted on 8 young individuals and 8 old individuals. Although the analysis did not reach the specific sample size (see Section 5.2), its analysis and downstream results, along with the high quality of the reading (see Chapter 13 for an example of Quality Assessment), still provide valuable insights for cardiology and bioinformatics. From the myocardial samples, a total of 15447 transcripts were measured. The samples and possible clustering have been assessed in the explorative phase (see Chapter 5.8) through principal component analysis, and no specific clustering across samples was identified. Additionally, samples within each group showed good agreement. The top 10 dysregulated transcripts are summarised in Table 6.5.

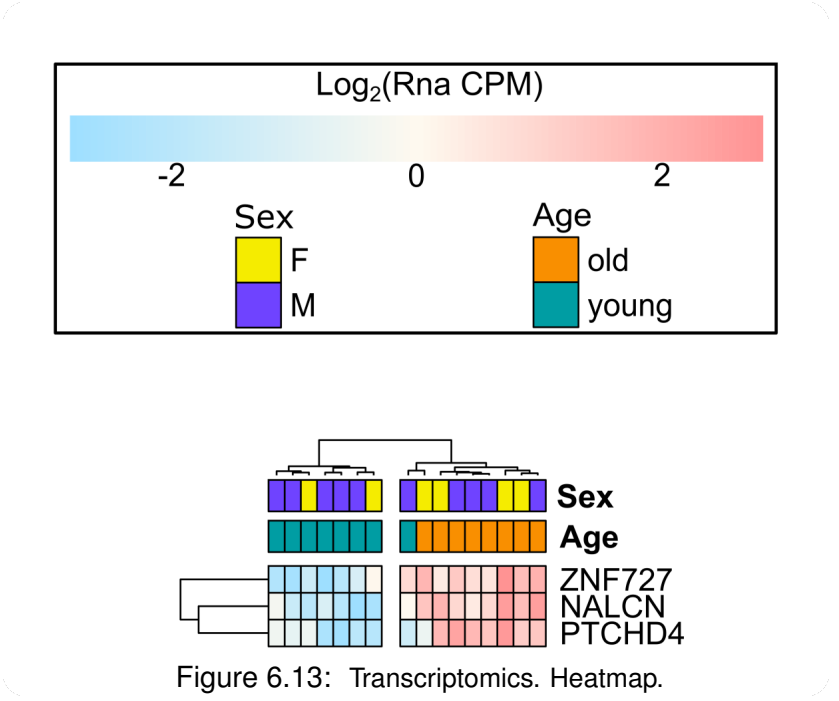


Figure 6.13: Transcriptomics. Heatmap.

Applying differential expression analysis to transcriptomics measurements, 3 transcripts exhibited evidence of possible age-associated dysregulation $p_{fdr} < 0.1$, while only one transcript was a possible age-related biomarker. The 3 transcripts with $p_{fdr} < 0.1$ were visualised using the heat map, as displayed in Figure 6.13. Applying the unsupervised clustering procedure on the samples (see Chapter 5.7) identified two distinct groups. The first group contained all young donors, and the second was all old donors except for one. Unsupervised clustering analysis applied to the transcript space did not highlight any group across the concerned transcripts; indeed, all the transcripts showed good agreement and were more abundantly expressed in the older cohort than the younger one.

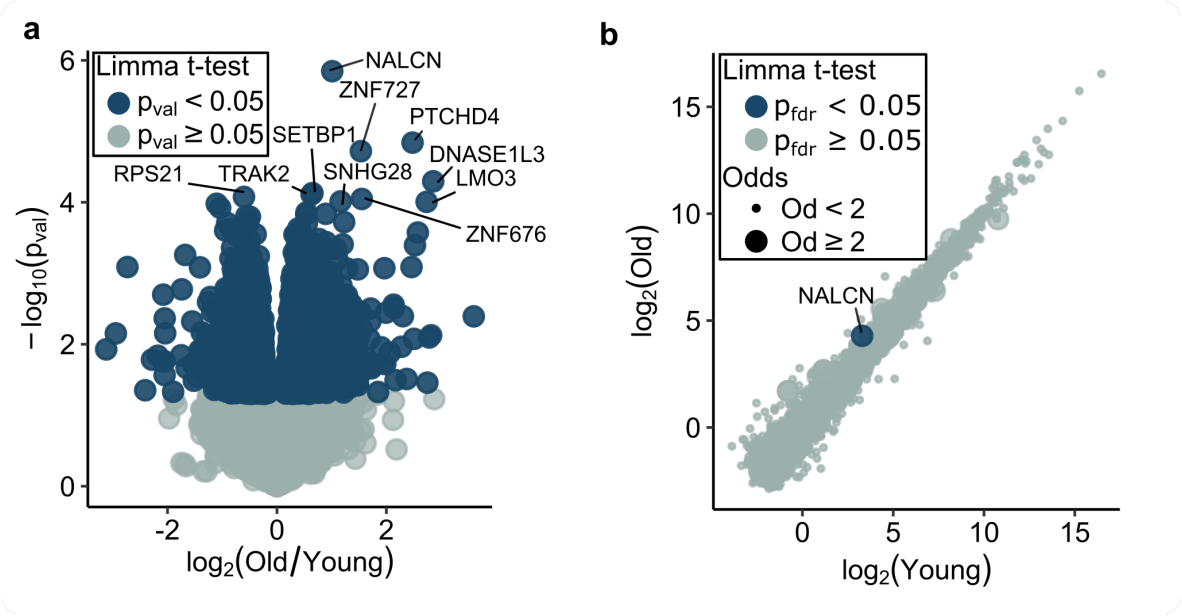


Figure 6.14: Transcriptomics. **(a)** Volcano plot. **(b)** Bubble plot.

Transcript	$\log_2$ **FC**	**t**	$[\mathbf{CI_l}, \mathbf{CI_u}]$	$\mathbf{p_{val}}$	$\mathbf{p_{fdr}}$
NALCN	1.010004	7.129849	[0.71179, 1.308217]	1.42E-06	0.021988
PTCHD4	2.478532	5.934998	[1.599391, 3.357673]	1.44E-05	0.098016
ZNF727	1.533324	5.798573	[0.976654, 2.089993]	1.90E-05	0.098016
DNASE1L3	2.856299	5.324621	[1.727024, 3.985574]	5.07E-05	0.12411
TRAK2	0.648435	5.143355	[0.383033, 0.913838]	7.42E-05	0.12411
SETBP1	0.632811	5.125093	[0.372881, 0.892742]	7.71E-05	0.12411
RPS21	-0.60343	-5.08494	[-0.85325, -0.35361]	8.40E-05	0.12411
ZNF676	1.550735	5.053652	[0.904758, 2.196711]	8.97E-05	0.12411
SNHG28	1.163442	5.015451	[0.675105, 1.651779]	9.73E-05	0.12411
LMO3	2.735965	5.009776	[1.586285, 3.885644]	9.85E-05	0.12411

Table 6.5: Top 10 dysregulated transcripts.

As Figure 6.14 and Table 6.5 show, the sodium leak channel, non-selective (NALCN) ($FC = 2.01$, $p_{fdr} = 0.021988$) was the only age-associated transcript identified as a potential biomarker. The sodium leak channel allows a flux of calcium, potassium and sodium. A patched domain containing 4 (PTCHD4) acts as a repressor of hedgehog signalling, while zinc finger protein (ZNF727) is a protein that may be involved in transcriptional regulation. Additional evidence has linked NALCN with a possible regulator of the metabolism, and its inhibition can affect the pacemaker activity of the heart[146,891].

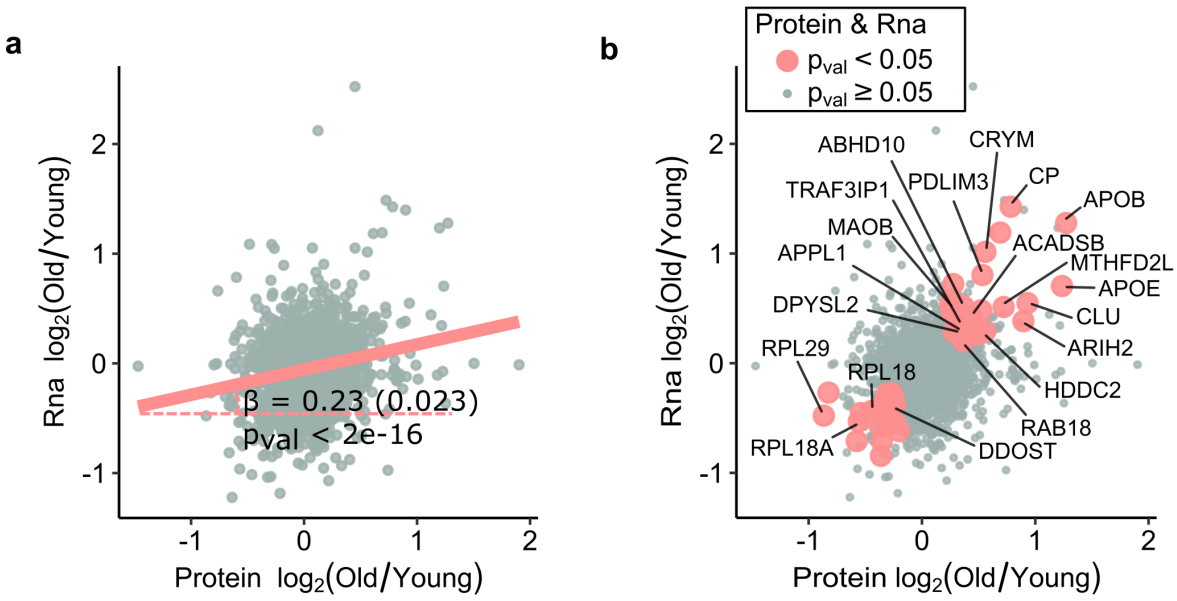


Figure 6.15: Transcriptomics. **(a)** Lipids' class $\log_2$ abundance. **(b)** Composition lipids dataset. **c** Lipids enrichment analysis.

Regression analysis was conducted to assess if any correlation occurred between protein and transcripts. The results showed a positive correlation between proteins and transcripts changes in abundance ($p_{val} = 2 \cdot 10^{-16}$), as displayed in Figure 6.15a. The lack of a strong correlation between proteins and transcripts has been pointed out in literature[690], and further examination was conducted using the raw p-value, p_val, obtained by differential expression analysis conducted on each respective omic. Figure 6.15b reports molecules that exhibited $p_{val} < 0.05$ simultaneously in proteins and transcripts.

Conclusion

In this chapter, I have presented the outcomes of the analysis for molecular signatures conducted on proteomics, metabolomics, lipidomics and transcriptomics. For each candidate biomarker detected during this part, I have presented possible connections with age-associated cardiovascular diseases. Indeed, with regard to proteomics, I have highlighted a possible change in contractility with multiple myosin heavy chains isoforms downregulated. The impairment also appeared in calcium delivery machinery, with SERCA2 and SERCA3 less abundant in the older cohort. At the metabolite level, multiple intermediates of the central carbon metabolism were upregulated, such as F1,6BP, fumarate and malate. In addition, higher concentrations of ATP and ADP in the older cohort were observed; since ADP/ATP indicate no change in supply/consumption of ATP occurred, it could be the accumulation of the two molecules in the tissue during the ageing process. Similar dynamics have been observed in early heart failure individuals[197]. It is notable that the change in the overall balance of energy consumption/supply machinery was noted when the PCr/ATP ratio was analysed and exhibited a strong decrease in the aged cohort. The imbalance of the energy machinery

was also suggested by creatine being downregulated in the older cohort. Further analysis of cardiac lipidome demonstrated a strong accumulation of AcCAs and TGs, which signals possible interruption in fatty acid oxidation and a gradual TG accumulation in the heart. Furthermore, transcriptomics analysis showed a strong dysregulation of NALCN.

Chapter 7

Omic-inferred Biological Network

This chapter presents findings from co-expression network analysis. Co-expression network is a powerful and widely used approach in systems biology to infer and study biological networks[455,927]. The investigation of human cardiac ageing can be conducted by recording the changes for each molecule identified and reconstructing the underlying molecular signature. Although this approach is essential to unravel and hypothesise on the potential mechanisms that regulate ageing, some of the patterns can be daunting to explain, as scattered and sparse identified dysregulated molecules are used to infer the whole ageing process. However, this does not mean that proteins, metabolites, lipids and transcripts work alone; their interactions highly affect each other's expression[153]. To address the lack of a holistic view of the analysis of molecular signatures, each omic layer can be studied using networks. The characterisation of omic measurements as a network helps researchers to infer protein-protein, metabolite-metabolite, lipid-lipid and transcript-transcript interactions and better understand interactions across analytes[895]. Co-expression network not only examines trends in the data but also identifies key molecules that play a pivotal role in human cardiac ageing or other complex diseases by allowing researchers to identify highly connected nodes (hubs or regulators)[12] essential for maintaining the structure of the network[884,895]. Exhibiting a high degree of connectivity with other analytes, pivotal molecules in co-expression networks have been referred to as master regulators of the complex phenomenon[71]. Their deletion can severely impact the healthy cardiac ageing process. Highly connected molecules are usually referred to as hubs. The co-expression network inferred from healthy human left ventricle omics measurements represents a resource for the scientific community, as it can be employed to identify differences in the network's topology compared to a network inferred on cardiovascular disease data[178,377].

Co-expression network has brought some benefits to this study. Representing data as a network allows researchers to calculate important metrics for nodes and edges and better understand the human cardiac ageing phenomenon[456]. Furthermore, it enables the possible identification of the modules (or communities); the module with more nodes inside a network represents the largest agglomerate of analytes that exhibit a positive or a negative linear trend in their expression patterns. Given the advantages that co-expression network analysis offers, we merged the existing framework[455] with a novel way to detect modules/communities and their functional annotation, independently from the type of omic

we are using for constructing the network (see Chapter 4.9). Although co-expression network analysis is widely applied in multi-omics data, we decided to apply it to each omic singularly, as bias across each omic level is introduced by different experimental conditions to which samples are subject in different omic experiments (see Chapter 5)[280]. This chapter presents findings from co-expression analysis of proteomics, metabolites, lipidomics and transcriptomics in order.

7.1 Protein-Protein co-expression network

Co-expression protein network was constructed on the whole proteome measured in the myocardial tissues. To mitigate the issues related to multiple testing, the soft threshold was applied to the correlation matrix till the degree of nodes and their frequencies did not exhibit a linear trend in $\log_{10}$-$\log_{10}$ scale with goodness of fitness at least $R^2 = 0.9$. The co-expression network is a simple network, and consequently, there is at most one edge that connects two nodes. The co-expression network is also a weighted network, and each edge's weight is determined by Pearson's correlation coefficient between two nodes. Furthermore, the inferred co-expression protein network is not fully connected. The final network is shown in Figure 7.1a. The protein-protein network exhibited an average degree (<degree>) of 13.01, indicating that proteins in the network did not have more than 13 neighbours on average. The mean strength of the edges (<strength>) of the network was 7.73, and each protein had incident edges with the sum of the weights equal to 7.73. The density of the network was $7 \cdot 10^{-3}$; the density represents the ratio between the actual number of edges over edges that a complete graph would have, and its low density suggets that the protein network is far from resembling a complete graph. The greatest geodesic distance between two proteins, also known as diameter, was approximately 14, indicating that 14 edges needed to be crossed to connect the farthest proteins in the network. Furthermore, the average of the shortest paths between proteins (<Path length>) was 4.86; 2 proteins were approximately 5 edges away from each other on average. The mean clustering coefficient (<Clustering coeff.>) is a metric that summarises how well a node is connected to its neighbourhood and can be obtained by calculating the ratio of theoretical triangles and actual triangles observed in the network. For the protein-protein network, it was about 0.43, signifying that proteins were not strongly connected to their neighbour proteins[333] (see Chapters 4.9 and 5.7).

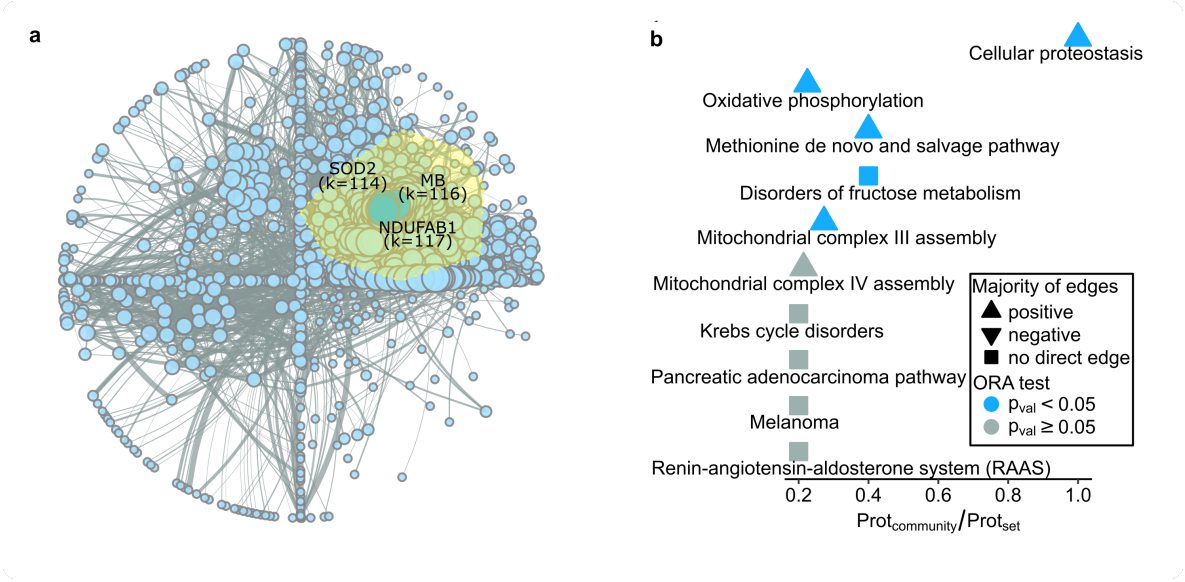


Figure 7.1: Proteomics. **(a)** Protein-protein network with largest community and hubs highlighted. **(b)** Overrepresentation analysis of the largest community of proteins. For each pathway, a sign analysis of the correlation is also presented.

Next, we explored the network by computing summary statistics, such as degree, betweenness and clustering coefficient. The nodes with the highest degree and their relative statistics are reported in Table 7.1. Amongst the nodes with the highest degree, we defined the top 3 as the main hubs of the network, as they are pivotal in maintaining the healthy ageing process and can be potential biomarkers for diagnosis and prognosis[310]. The node

Protein	Degree	Betweenness	Clustering Coeff
NDUFAB1	117	8071	0.34025
MB	116	7361	0.33538
SOD2	114	18494	0.35429
UQCRB	108	13880	0.33073
NDUFS6	103	13980	0.32172
TPM1;TPM3	101	7276	0.39762
ATP5PF	97	7647	0.35760
SOD1	96	14109	0.37785
FXN	95	4584	0.40828
PFDN6	94	11681	0.40242

Table 7.1: Top 10 proteins ranked by degree.

with the largest NADH:ubiquinone oxidoreductase subunit AB1 (NDUFAB1) is a subunit of NADH dehydrogenase or mitochondrial complex I of the mitochondrial electron transport chain and plays a role in oxidative phosphorylation and NAD synthesis[739]. In addition, NDUFAB1 coordinates complexes II and III and supercomplexes by regulating the iron-sulfur biosynthesis. Suppression of NDUFAB1 in mice elevates reactive oxygen species, impairs the electron transport chain and develops dilated cardiomyopathy[344]. Myoglobin (MB) is an

oxygen-binding protein in striated muscle and essential for oxygen storage and delivery[611]. The high amount of MB in the blood can lead to the early detection of myocarditis[439]. Lower levels of oxidised MB have also been observed in patients with myocardial infarction post-reperfusion[41]. Another protein that exhibited the highest degree is superoxide dismutase 2 (SOD2), a mitochondrial antioxidant enzyme pivotal in clearing ROS and protecting against oxidative stress[625]. A recent study highlighted a lower concentration of SOD2 in subjects with abnormal left ventricular geometry (see Chapter 2.1) and heart failure compared to one in the non-disease group[480].

Applying the clustering procedure allowed us to identify multiple communities in the network (see Chapter 4.9). The largest community, or module, represents the largest set of analytes exhibiting co-tendency during the cardiac ageing process. The functional analysis of the community can help to reveal the biological processes that have increased or decreased their activities during ageing. Unlike the one on differentially expressed analytes, overrepresentation analysis on the largest module takes into account the analytes that exhibit a trend together regardless of whether they are differentially expressed or not. The functional annotation of the largest community, which comprised 199 proteins, was performed using WikiPathways[543], and results are summarised in Figure 7.1b and Table 7.2. Overrepresentation analysis unveiled proteins belonging to cellular proteostasis

Pathway	$\mathbf{N_{path}}$	$\mathbf{N_{com}}$	$\mathbf{p_{val}}$
Cellular proteostasis	6	6	4.68E-06
Oxidative phosphorylation	40	9	0.0116
Methionine de novo and salvage pathway	5	2	0.0170
Disorders of fructose metabolism	5	2	0.0170
Mitochondrial complex III assembly	11	3	0.0313
Mitochondrial complex IV assembly	14	3	0.0647
Krebs cycle disorders	5	1	0.1008
Pancreatic adenocarcinoma pathway	5	1	0.1008
Melanoma	5	1	0.1008
Renin-angiotensin-aldosterone system (RAAS)	5	1	0.1008

Table 7.2: Top 10 biological process associated with the largest community of proteins.

($p_{val} = 4.68 \cdot 10^{-6}$). Protein homeostasis and proteostasis are fundamental to the correct functioning of cardiac cells, and their dysregulation can occur at the cytosolic or endoplasmic reticulum level[33]. Aberrant regulation of proteostasis can lead to the gradual derailment of cardiac functions and is associated with multiple heart conditions[329]. The functional role of the community also extended to oxidative phosphorylation ($p_{val} = 0.0116$) and mitochondrial complex III ($p_{val} = 0.0313$). Impairment of the oxidative phosphorylation is usually observed in heart failure[696].

7.2 Metabolite-Metabolite co-expression network

The metabolite-metabolite network was inferred using the co-expression network framework on the whole set of metabolites identified through semi-target MS. So as not to rely on statistical tests, the network was inferred using the soft threshold, which requested that degrees and their related probability had a linear trend when analysed in $\log_{10}$-$\log_{10}$ scale. The goodness of fitness of the linear model was set at $R^2 = 0.9$. Each edge was accompanied by a weight, representing Pearson's correlation coefficient between the two incident nodes, representing the metabolites. The outcome of the inferred network was simple and weighted but not fully connected. Exploring the network characteristics, the average degree of the metabolite-metabolite network, <degree>, was approximately 3.77, indicating that each metabolite had about 4 metabolite interactions. The average strength, <strength>, of the metabolite interactions was 1.13. The <Clustering Coeff.>, representing the connectivity of nodes with their neighbours, was approximately 0.44. The average path length, which indicated the number of edges needed to be crossed to connect two proteins, was 4.10. Additional investigation on the network's diameter revealed that the largest geodesic distance between two metabolites was 11, while the network's density was about 0.034.

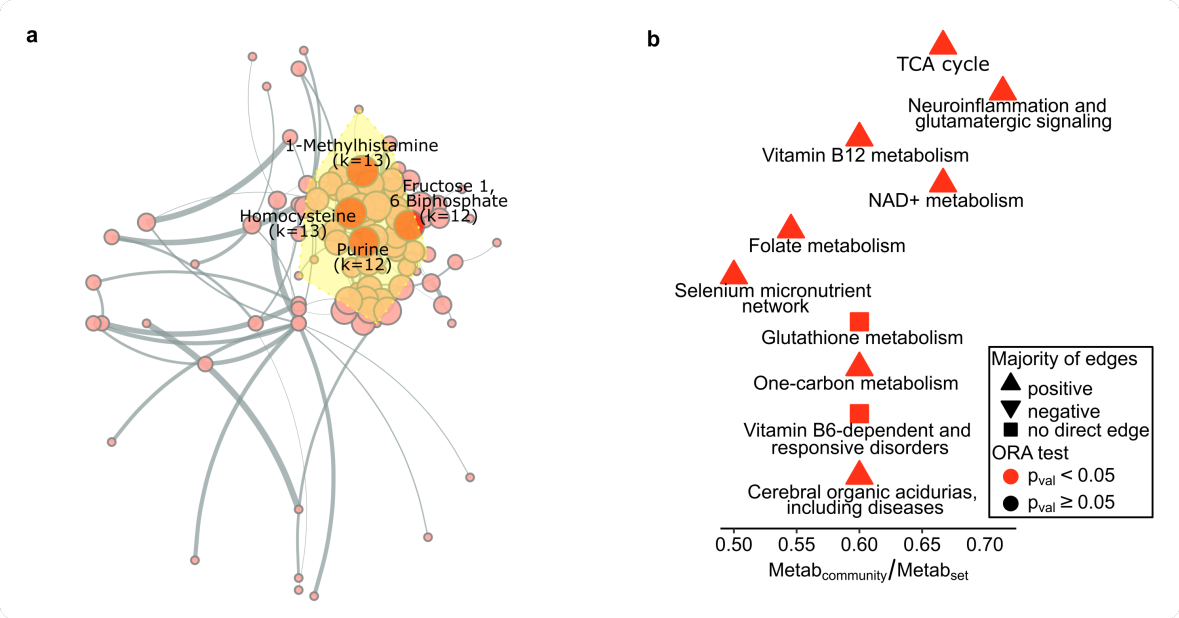


Figure 7.2: Metabolomics. **(a)** Metabolite-metabolite network with largest community and hubs highlighted. **(b)** Overrepresentation analysis of the largest community of metabolites. For each pathway, a sign analysis of the correlation is also presented.

Properties of the metabolites in the network were then further explored by computing summary statistics for each of them. When sorting the nodes based on the highest degree, a total of 6 hubs were identified, potential biomarkers in metabolic processes during cardiac ageing. Results are summarised in Table 7.3. The first identified hub was 1-methylhistamine, a part of histidine metabolism. While 1-methylhistamine is found in higher abundance in the

Metabolite	Degree	Betweenness	Clustering Coeff
1-Methylhistamine	13	618	0.24359
Homocysteine	13	453	0.320513
Fructose 1,6-bisphosphate	12	1272	0.19697
Purine	12	136	0.318182
Saccharopine	11	539	0.327273
Creatine	11	573	0.327273
Acetoacetic acid	9	6	0.722222
Acetylcholine	9	751	0.694444
Phenylalanine	9	99	0.416667
Trans-hydroxyproline	9	679	0.388889

Table 7.3: Top 10 metabolites ranked by degree.

plasma of patients in the pre-heart failure stage[933], its role in cardiac ageing and heart failure needs further investigation. The second analyte was homocystine. It is a well-established biomarker for cardiovascular disease[303], and was identified as downregulated in older donor hearts (see Chapter 6.2. Interestingly, F1,6BP was also identified as a candidate hub for healthy ageing, showing strong upregulation in the older cohort. F1,6BP is an important regulator of oxidative stress and cardiac contraction and plays a well-established cardioprotective role in the heart (see Chapter 2.3)[319,682].

Purine was also detected as a hub in the metabolite-metabolite network. Regulating ageing, purine indicated downregulation in the older cohort in the analysis of molecular signatures of the cardiac metabolome (see chapter 6.2). Purine dysregulation has also been observed in recent studies on human cardiac tissue in patients with dilated cardiomyopathy[860] and cardiac hypertrophy[863]. Another candidate hub of cardiac ageing was saccharopine; this metabolite also displayed downregulation in the aged cohort in the analysis of molecular signature (see Chapter 6.2). Saccharpione is crucial for maintaining mitochondrial homeostasis[947]. Lastly, creatine was identified as the hub. Differential expression analysis presented in Chapter 6.2 exhibited a change in its abundance, leading to downregulation in the older cohort. Rapidly supplying ATP to the myofibrils and facilitating cardiac contraction, creatine plays an important role in energy supply (see Chapter 2.3). Cardiac remodelling occurs in individuals with hypertrophic hearts and is tightly associated with the downregulation of creatine kinase, an enzyme that permits the ATP production from phosphocreatine[414].

The largest community identified in the metabolite-metabolite network counted 29 analytes, including the hubs, except for F1,6BP. To detect possible functional roles that the most interconnected community of metabolites can represent, we performed overrepresentation analysis. We utilised the WikiPathways database for the functional annotation[543]. Overrepresentation analysis showed multiple key metabolic pathways that exhibited a possible regulatory role in the ageing of the cardiac tissue. The most enriched pathway was the TCA cycle, which is crucial for cardiac ageing and diseases[899] (see Chapter 2.4). Metabolisms of vitamin B12, B6 and folate were also represented by the largest community of metabolites. Deficiency in metabolisms of vitamin B and folic acid has been observed

Pathway	$\mathbf{N_{path}}$	$\mathbf{N_{com}}$	$\mathbf{p_{val}}$
TCA cycle (aka Krebs or citric acid cycle)	9	6	0.008037
Neuroinflammation and glutamatergic signaling	7	5	0.009238
Vitamin B12 metabolism	10	6	0.013971
NAD+ metabolism	6	4	0.01983
Folate metabolism	11	6	0.02262
Selenium micronutrient network	12	6	0.034535
Glutathione metabolism	5	3	0.042101
One-carbon metabolism	5	3	0.042101
Vitamin B6-dependent and responsive disorders	5	3	0.042101
Cerebral organic acidurias, including diseases	5	3	0.042101

Table 7.4: Top 10 biological process associated with the largest community of metabolites.

among heart failure patients, and possible nutritional supplementation can be beneficial if integrated with current practices and treatments[45,650]. Another metabolic pathway that was associated with the largest community was nicotinamide adenine dinucleotide (NAD) metabolism. NAD metabolism is central to cardiac diseases and ageing, as it regulates redox homeostasis, energy maintenance and DNA repairs[2]. Consistent with the findings in Chapter 6.2, glutathione metabolism was identified as a possible key process that regulated cardiac ageing; it helps to regulate oxidative stress in the cardiac tissue[40,871]. Furthermore, overrepresentation analysis showed that one-carbon metabolism contributed to the cardiac ageing process. One-carbon is a non-oxidative metabolic pathway and assumes a possible cardioprotective role according to the study that shows one-carbon mitigating the remodelling of the left ventricle during heart failure when pyruvate metabolism is suppressed[515] (see Chapter 2.1).

7.3 Lipid-Lipid co-expression network

Co-expression network framework discussed in Chapter 4.9 was applied to infer lipid-lipid interactions from the lipids measured through untargeted MS (see Chapter 5.5). The soft threshold was applied to avoid overreliance on multiple statistical tests. The soft threshold aimed to identify the edges by filtering out interaction across lipids till there was a linear trend between degrees and probabilities of nodes in the $\log_{10} - \log_{10}$ scale; the goodness of fitness of the linear model was imposed at $R^2 = 0.9$. The resulting graph was weighted and not fully connected. The weight of an edge represented Pearson's correlation coefficient. Network summary statistics were then computed to further understand the properties of the network. The average degree in the lipid-lipid interaction network was $<$ degree $>= 3.89$, and each node of the network had approximately 4 interactions. The average strength of lipid interactions, $<$strength$>$, was about 3.44. The average clustering coefficient, denoted by $<$Clustering Coeff.$>$, was 0.67. The $<$Clustering Coeff.$>$ summarises how close a lipid is connected to its neighbour lipids in the network. Additionally, the average path length, $<$ Path length $>$, which summarises the average length of the shortest paths between pairs of lipids in the network, was 2.48, while the network diameter showed 7. Finally, the density of

the network was 0.018.

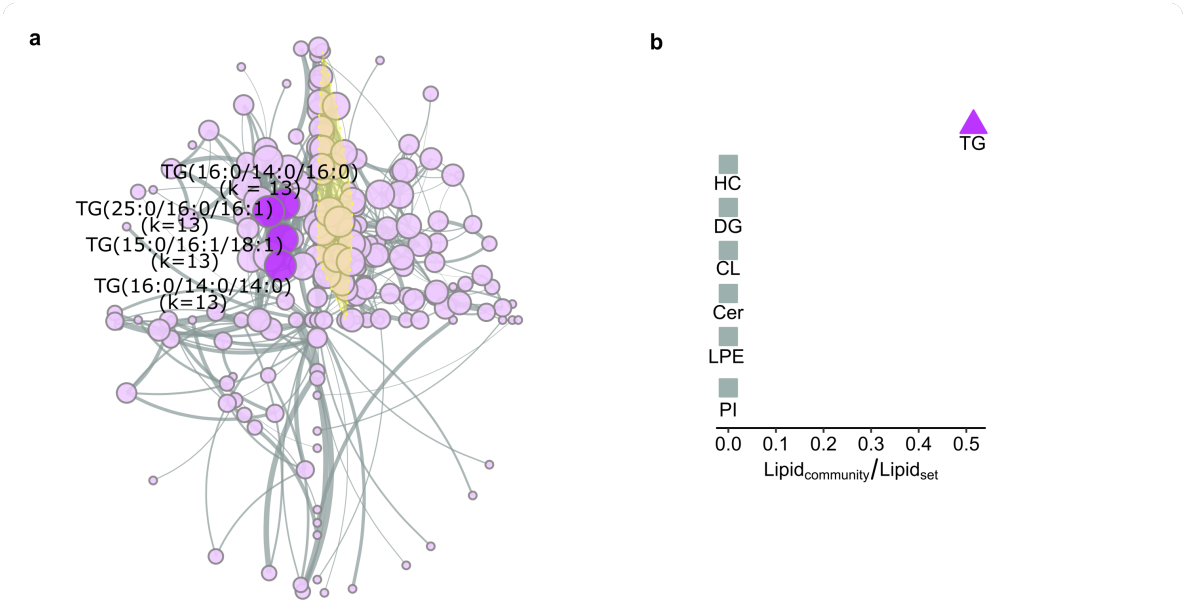


Figure 7.3: Lipidomics. **(a)** Lipid-lipid network with largest community and hubs highlighted. **(b)** Overrepresentation analysis of the largest community of lipids. For each pathway, a sign analysis of the correlation is also presented.

Summary statistics were calculated for each node, and the 10 nodes with the highest degree are presented in Table 7.5. Considering the nodes with the highest 3 degrees as hubs, we identified 12 potential lipids that can be associated with healthy ageing. All the 12 potential

Lipid	Degree	Betweenness	Clustering Coeff
TG(15:0/16:1/18:1)	13	0	0.74359
TG(16:0/14:0/14:0)	13	33	0.74359
TG(16:0/14:0/16:0)	13	36	0.717949
TG(25:0/16:0/16:1)	13	24	0.717949
TG(15:0/14:0/16:0)	12	0	0.80303
TG(16:0/16:1/18:1)	12	21	0.636364
TG(16:0/16:1/24:0)	12	10	0.757576
TG(15:0/14:0/16:1)	11	33	0.690909
TG(15:0/16:0/16:1)	11	72	0.836364
TG(15:0/16:1/16:1)	11	5	0.818182

Table 7.5: Top 10 lipids ranked by degree.

lipids that regulate the ageing process belonged to the TG class. The significance of TG was also highlighted by the identification of the largest community, which comprised 17 lipids, all of which belonged to the TG class, as shown in Table 7.6. Fatty acid tails on the TG hubs were long or very long (see Chapter 2.3). During the normal activity of the heart, TG cytosolic hydrolysis can be performed by adipose triglyceride lipase, a class of enzymes that

extract fatty acids from TGs. Fatty acids are then oxidised in order to supply energy to the cardiac muscle. As mentioned in Chapter 2.3, fatty acid oxidation is the main energy supplier of the heart, and dysregulation of TG hydrolysis can lead to a change in the energy supply/consumption of the heart. The link between dysregulation of TG synthesis and hydrolysis with heart failure has been studied by researchers since the aberrant TG accumulation can often be observed[424] (see Chapter 6.3 for further discussion on the role of TGs in cardiac ageing) .

Pathway	$\mathbf{N_{path}}$	$\mathbf{N_{com}}$	$\mathbf{p_{val}}$
TG	33	17	0
HC	3	0	0.213375
DG	6	0	0.383345
CL	7	0	0.431852
Cer	10	0	0.556719
LPE	11	0	0.592246
PI	11	0	0.592246
LPC	13	0	0.655413
AcCa	19	0	0.794148
PG	19	0	0.794148

Table 7.6: Top 10 lipid classes associated with the largest community of lipids.

7.4 Transcript-Transcript co-expression network

The network of transcripts and their interaction was inferred using the co-expression network framework (see Section 4.9), and the edges of the network were obtained through the soft threshold procedure, which required the degrees of nodes and their associated probabilities to exhibit a linear relationship when evaluated in the $\log_{10}$-$\log_{10}$ space. The resulting network was weighted and not necessarily connected. Analysing some summary metrics of the network showed 39.92 for the average degree $<$degree$>$ of the nodes, indicating that each transcript was connected to 40 other transcripts on average. The average strength of the network indicated by $<$strength$>$ was 5.52. The longest path in the network, the diameter, was 14, while the average of the shortest paths, $<$ Path length$>$, was 4.26. The mean clustering coefficient of the network was 0.28 and showed a little connection between a node and its neighbours. The density of the network was 0.002.

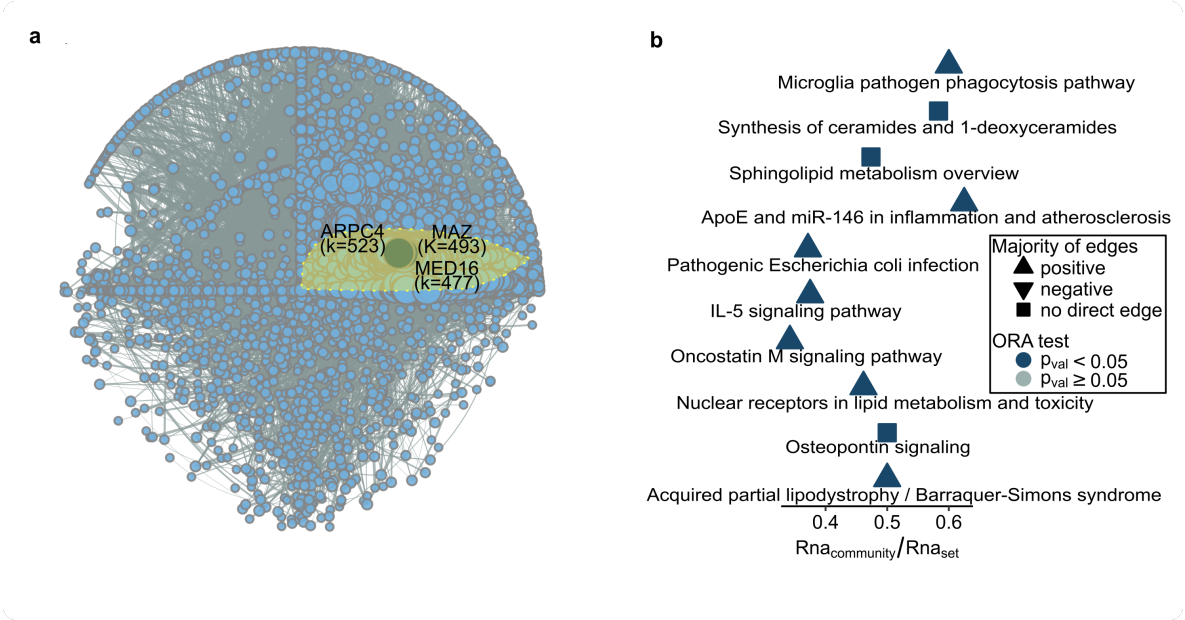


Figure 7.4: Transcriptomics. **(a)** Transcript-transcript network with largest community and hubs highlighted. **(b)** Overrepresentation analysis of the largest community of transcripts. For each pathway, a sign analysis of the correlation is also presented.

Further summary statistics were elaborated for each node, which are shown in Table 7.7. Actin-related protein 2/3 complex subunit 4 (ARPC4 or 10093) was the transcript with the

Transcript	Degree	Betweenness	Clustering Coeff
10093	523	105246	0.258265
4150	493	123761	0.25289
10025	477	93057	0.252559
396	475	168944	0.237086
1192	464	117606	0.250047
10095	455	94985	0.261171
823	452	63301	0.270235
391	450	88054	0.256936
80012	448	121998	0.24984
4650	445	107085	0.255491

Table 7.7: Top 10 proteins ranked by degree.

highest degree in the network, identified as a hub. As part of the actin-related protein 2/3, ARPC4 mediates the stability and regeneration of the branch junction in the actin network[266] (see Chapter 2.2.3). If the actin network of a cardiomyocyte is compromised, cardiac cells may not reorganise the cytoskeleton when stimuli are received, although further research is required to understand its impact on overall cardiac functioning[246]. The second hub was MYC-associated zinc finger protein (MAZ or 4150), a transcription factor that is usually associated with inflammatory diseases. Chronic inflammations are one of the hallmarks of cardiac ageing[899] (see Chapter 2.4), and MAZ could represent an interesting biomarker to

track during the cardiac ageing process in future experiments[829]. The last hub was mediator complex subunit 16 (MED16 or 10025), a component essential to connect multiple modules of the mediator complex. Although the specific role of MED16 in cardiac tissue requires further investigation, its ablation is known to decrease MED23 and MED24[140]. The overall dysregulation of the mediator complex activity is well associated with heart conditions and dysregulation of the overall energy balance[718].

The final step of this analysis involved detecting the largest community of interconnected transcripts and determining possible functions that it could represent. The functional annotation of the community was performed using overrepresentation analysis on the WikiPathways database[543]. The largest interconnected community contained 2576 transcripts, and the top 10 associated biological functions are presented in Table 7.8. The first

Pathway	$\mathbf{N_{path}}$	$\mathbf{N_{com}}$	$\mathbf{p_{val}}$
Microglia pathogen phagocytosis pathway	20	12	6.72E-05
Synthesis of ceramides and 1-deoxyceramides	12	7	0.001382
Sphingolipid metabolism overview	19	9	0.002127
ApoE and miR-146 in inflammation and atherosclerosis	8	5	0.003028
Pathogenic Escherichia coli infection	35	13	0.003901
IL-5 signaling pathway	32	12	0.00467
Oncostatin M signaling pathway	38	13	0.00775
Nuclear receptors in lipid metabolism and toxicity	13	6	0.008351
Osteopontin signaling	10	5	0.008669
Acquired partial lipodystrophy / Barraquer-Simons syndrome	10	5	0.008669

Table 7.8: Top 10 biological process of the largest community of transcripts.

biological function associated with the largest community in the transcript-transcript network is the microglia pathogen phagocytosis pathway, an inflammatory regulator with a novel role in cardiovascular diseases and cardiac remodelling[862]. Another pathway that represented the largest community is proteasome degradation ($p_{val} = 0.040$), a pathway that monitors the protein quality and degrades them to ensure the correct functioning of the heart[481,623]. Multiple pathways associated with the lipids were also detected ($p_{val} < 0.05$), particularly sphingolipids (SP) and Cer. SP and Cer regulations are essential, as both lipids are involved in the sarcolemma and cellular signalling[629] (see Chapter 2.3). Additionally, ApoE and miR-146 in inflammation and atherosclerosis was another pathway enriched in the community of transcripts; it has a key role in protecting against inflammation in the cardiac tissue[89]. The community was also likely to represent multiple interleukin signalling pathways, such as interleukin signalling (1/3/5/11). The interleukin 1 ($p_{val} = 0.0267$) signalling pathway is well associated with inflammatory reactions during myocardial injury[845], while interleukin 3 regulates the formation of endothelial cells to mitigate further damages provoked by myocardial injury[577]. Interleukin 5 appears to increase the healing process after heart infarction occurs, although further research is required[901]. Interleukin 11 is associated with cardioprotective activities during ischaemia[801].

Further signalling pathways associated with the largest community were MAPK cas-

cade ($p_{val} = 0.0174$), TSH signalling pathway ($p_{val} = 0.0164$) and ATM signalling pathways ($p_{val} = 0.0449$). The MAPK signalling pathway covers a pivotal role in regulating various biological processes, including the transformation of growth factor-β, which has been linked to the development of cardiac hypertrophy[323] (see Chapter 2.4). Thyroid hormone also controls cardiac hypertrophy by regulating SERCA2 and myosin heavy chains expressions[194]. The ATM pathway regulates cardiac apoptosis by interacting with the β-adrenergic receptor[243]. Kit receptor signalling pathway ($p_{val} = 0.0281$) was also enriched in the community, suggesting its influence over cardiac ageing; indeed, experiments in mice have highlighted the cardioprotective role of Kit signalling pathway for cardiac injury[191]. Wnt/B-catenin signalling by small molecules ($p_{val} = 0.0306$) is another pathway represented by the largest community of transcripts, and its activation decreases the size of the cardiac injury[533]. Likewise, the RANKL/RANK signalling pathway ($p_{val} = 0.0256$) was also enriched; it appears to maintain the cardiac function post ischaemia through the interaction with interleukin 1 when inhibited[761]. The oncostatin M signalling pathway is another pathway represented by the largest community, and the study has suggested its active role in de-differentiating cardiomyocytes and cardiac remodelling following myocardial infarction[446]. The osteopontin signalling, which is essential to regulate cell-matrix interactions in heart failure[535], was also enriched. Another biological process represented by the largest community in the network was the inhibitors of DNA binding (ID) signalling pathway ($p_{val} = 0.0347$), which has recently been studied in the context of multiple heart diseases despite the need for further research[353]. The last signalling pathway associated with the community is purinergic signalling ($p_{val} = 0.0324$), which mediates multiple intracellular signal transduction, including metabolism and oxidative stress during myocardial ischaemia[949].

Furthermore, the community identified in the network was also represented by some metabolic pathways. The first is the pyrimidine metabolism ($p_{val} = 0.0193$) and, as briefly mentioned in Chapter 6.2, its perturbation is associated with cardiac hypertrophy[863]. The creatine pathway ($p_{0.0347}$) was also enriched, with a clear role in cardiac metabolism and energy supply/consumption (see Chapter 2.3).

Conclusion

This chapter has demonstrated how co-expression network has informed our investigation of potential biomarkers and regulators of the cardiac ageing process across different omic levels. Hubs and functional annotation of the largest community in each network revealed interesting features of the network, which provided further insights into the potential pathways involved in healthy ageing. From the protein-protein co-expression network, we observed NDUFAB1 is a hub in the network, a node essential for maintaining the overall network structure of healthy ageing. Overrepresentation analysis has further highlighted the role of mitochondria, as the largest community of interconnected proteins appeared to represent multiple mitochondrial complexes and oxidative phosphorylation. Moreover, overrepresentation analysis performed on the largest community of metabolites placed mitochondria and energy machinery in the spotlight of the cardiac ageing process; indeed, the largest community of metabolites was represented by two key processes associated with the mitochondria, namely TCA cycle, which takes place in the mitochondria, and glutathione metabolism, which

contributes to redox metabolism in the mitochondria[467]. The lipid-lipid interaction network revealed that TGs not only regulated the ageing process but also represented the largest community in the lipidome interactions. TGs are important lipids for the heart, and their regulation and hydrolysis are essential to supplying fatty acids to the β-oxidation, a process in mitochondria. Meanwhile, transcriptome interactions provided another perspective of the ageing process. At the transcript level, ARPC4, an actin-related complex, had a key role in the ageing process. Overrepresentation analysis suggested the role of multiple signalling pathways in the ageing heart, especially interleukin signalling and lipids-related pathways.

Chapter 8

Omics SEA: A Unified View of Biological Pathways

In this chapter, I discuss the omics SEA application for 4 different instances to understand better how biological processes possibly derail with ageing. While co-expression network can effectively help to identify key regulators and pathways driving the human cardiac ageing process, it entails a limitation; each omic has been evaluated individually and not as part of the same biological process. To investigate potential pathways that derail during the cardiac ageing process, pathway enrichment analysis was employed to consider all the analytes[622]. Pathway enrichment analysis helps researchers unveil the potential dysregulated biological function by integrating data into knowledge-based pathways. While previous approaches required results from differential expression analysis of respective data, gene set enrichment analysis (GSEA), a class of pathways enrichment methods, does not require any pre-selection on the analytes. In fact, GSEA uses all the available information in the data to examine changes occurring in biological systems[783] regardless of whether they belong to differentially expressed analytes (see Chapter 4.8). GSEA summarises the expression pattern by calculating an appropriate metric for each analyte, which is used to subsequently rank the analytes[954]. The ranked analytes and their associated metrics are then used to produce an adequate statistical test to identify possible dysregulated biological functions when two cohorts are compared[783]. While GSEA has facilitated the identification process of pivotal biological functions based on proteins and gene data, a few alternatives have been proposed to compute pathway enrichment analysis in the presence of multi-omics data. MultiGSEA addresses omics integration in the context of pathway enrichment[109]; however, it has several drawbacks, including its inability to integrate lipid measurements for enrichment analysis and to calculate a metric for each pathway, such as the enrichment score that allows researchers to better interpret the outcome of pathways analysis (see Chapter 4.10).

During the age-OME project, I engineered a new method that allowed a full multi-omics pathway enrichment analysis that integrates any omics layer and effectively overcomes the limitations of multiGSEA. This new method, defined as omics SEA, builds a set of analytes from the whole pathways and facilitates the integration process of multiple omics, including lipidomics data. Omics SEA returns 1 enrichment score (ES) and p-value for each tested pathway, and the ES statistics and testing were performed using the fast GSEA algorithm[437]

(see Chapter 4.10 for further details of omics SEA). The uniquely calculated ES and adjusted p-value (p_{fdr}) enhance the interpretation of the directionality of the pathway alteration. In fact, assuming that p_{fdr} is below the significance level, in our case $p_{fdr} < 0.05$, when the ES is positive, the pathway can be interpreted as up-regulated, while a negative ES hints toward the down-regulation of the pathway. The interpretation should always be accompanied by a graphical representation, such as Figure 8.2, as it helps to interpret the directionality of the pathway alteration further. Additionally, Omics SEA, similar to any pathway enrichment analysis, can indicate possible altered biological function, as the analysis relies solely on omics measurements[783]. This chapter first introduces the application of omics SEA to explore possible dysregulation of biological systems using WikiPathways databases[543]. Although molecular pathways on WikiPathways comprise several lipids, we found that β-oxidation was not detailed enough to allow the pathway evaluation. Consequently, I discuss the omics SEA application to the manually curated β-oxidation, which takes into account lipids that are essential for the fatty acids oxidation: acetylcarnitine (AcCA). Finally, this chapter evaluates potential signalling pathways associated with ageing, employing omics SEA. Although WikiPathways already contains multiple signalling pathways, the Reactome database was selected for this task due to its detailed curation of cellular signalling events[568]. From Reactome, the signalling pathways associated with cardiac ageing were selected based on a meticulous literature review (see Chapter 2.4). The final section summarises the findings and concludes the chapter.

Omics SEA application on human cardiac ageing

The application of omics SEA on the omics measurements obtained from myocardial tissue excised from pre-mortem donors aimed to identify potential dysregulated pathways associated with ageing. For this purpose, the database WikiPathways was used to retrieve all the molecular pathways[543]. As explained in Chapters 4.10.2 and 5.7, we computed the moderated t statistic for each omic measurement, a dimensional metric that summarises the generic trend of an analyte; the baseline was represented by the young cohort of donors. The results are summarised in Table 8.1 and Figure 8.1b. Omics SEA performed on the proteomics, transcriptomics, metabolomics and lipidomics data identified 8 dysregulated pathways in older hearts compared to the younger ones, among which 5 pathways belonged to the cell metabolism. Among the metabolic pathways, glycolysis (set size = 20, $p_{fdr} = 1.99 \cdot 10 - 2$) and TCA cycle (set size = 32, $p_{fdr} = 9.71 \cdot 10^{-4}$), which are pivotal components of central carbon metabolism, presented substantial evidence of dysregulation. The Cori cycle (set size = 31, $p_{fdr} = 3.39 \cdot 10^{-2}$), a chemical pathway responsible for lactate metabolism, also exhibited significant age-related changes. Evidence of aberrant regulation was also found for glutathione metabolism (set size = 24, $p_{fdr} = 3.68 \cdot 10^{-2}$), with possible consequences for cellular oxidative stress. GABA metabolism, or GABAergic system, was also dysregulated (set size = 20, $p_{fdr} = 1.93 \cdot 10^{-2}$), possibly affecting Ca^{2+} cardiac dynamics in cardiomyocytes. Age-related metabolic changes were further indicated by the Phosphocreatine (PCr)/ATP ratio (see Figure 8.1c and Chapter 6.2), a prognostic biomarker for cardiac tissue frailty[916].

Given that pathways are characterised by multiple feedback and feedforward loops[216],

Pathway	ES	$\mathbf{p_{val}}$	$\mathbf{p_{fdr}}$	Set Size
TCA cycle	0.6566	2.26E-06	0.0009	32
Inducation pathway by cyclosporin A	-0.7516	1.30E-05	0.0027	16
11p11.2 copy number variation syndrome	-0.5163	0.0001	0.0127	44
miR-509-3p alteration of YAP1/ECM axis	-0.6959	0.0001	0.0127	17
GABA metabolism (aka GHB)	0.6586	0.0002	0.0193	20
Aerobic glycolysis	0.6532	0.0002	0.0199	20
Cori cycle	0.5623	0.0005	0.0339	31
Glutathione metabolism	0.5866	0.0006	0.0368	24
Cholesterol metabolism	-0.4411	0.0032	0.1003	44
Complement and coagulation cascades	0.4572	0.0023	0.1003	45
Hippo-Merlin signaling dysregulation	-0.416	0.0032	0.1003	50

Table 8.1: Top 10 dysregulated pathways.

Figure 8.1 illustrates a schematic capturing the comprehensive changes and subcellular localisation of the metabolic enzyme. It highlights transcript, protein, metabolite and lipid changes in some cardiac bioenergetics pathways that were found dysregulated, such as glycolysis and TCA cycle. The schematic also includes the oxidative phosphorylation and sarcoplasmic reticulum Ca^{2+} pump to further display how cardiac bioenergetics is interconnected with contractility machinery[827,899]. Despite the lack of evidence of the activities of these two pathways that limit the mechanistic explanation, there are several key results that pointed to possible cross-talk between bioenergetic and contractility machinery. Amongst these results, we identified decreased MYH6, SERCA2, creatine and PCr/ATP ratio in old hearts (see Chapter 6). Oxidative phosphorylation also appeared as a key regulator when the protein-protein interactions network was analysed, as shown in Chapter 7.1. This was relevant to the understanding of the changes in ATP observed in Chapter 6.2, as myosin ATPase and SERCA are the two largest ATP consumers, while oxidative phosphorylation is one of the main ATP suppliers of ATP in the heart[827] (see Chapter 2.3). The change in these essential functional/structural proteins may lead to the increase in ATP indicative of either increased ATP production or decreased ATP utilisation.

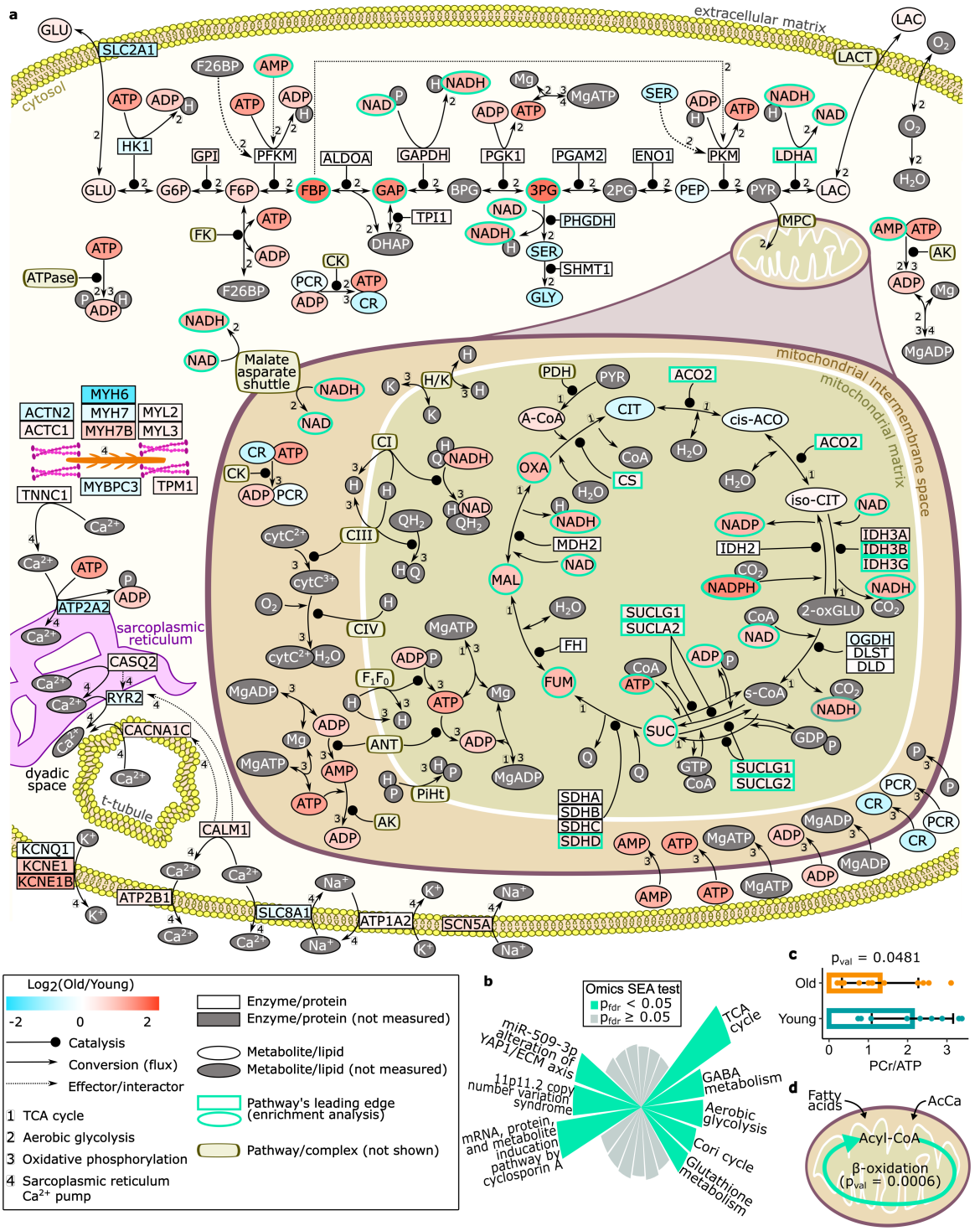


Figure 8.1: **(a)** Core metabolic and mechanical network interlinking the TCA cycle, aerobic glycolysis and oxidative phosphorylation with the contractile machinery and calcium signalling. The analytes are coloured based on the $\log_2$ (Old/Young). The analytes with aquamarine borders are the leading edges of omics SEA. **(b)** Omics SEA. The enriched pathways with $p_{fdr} < 0.05$ are highlighted, and the bars are proportional to $-log_{10}p_{val}$ of each pathway. **(c)** PCr/ATP ratio. **(d)** Dysregulation of β-oxidation via omics SEA.

Taking into account the evidence in Chapter 6.3, AcCA accumulation can be observed in older hearts. AcCAs are pivotal substrates for the β-oxidation and contain fatty acid tails that are oxidised. Despite β-oxidation being in WikiPathways, fatty acids and AcCA were not sufficiently annotated. Hence, we proceeded with the analysis by manually reconstructing the β-oxidation pathway[345,346,683]. Omics SEA was then applied to test if the manually reconstructed β-oxidation exhibited age-related dysregulation. The oxidation of fatty acids displayed evidence of alteration (set size = 54, p_{fdr} = $5.84 \times 10 - 4$) with ageing. The results are shown in Figure 8.1c, and a schematic of the expression pattern is available in Figure 8.2.

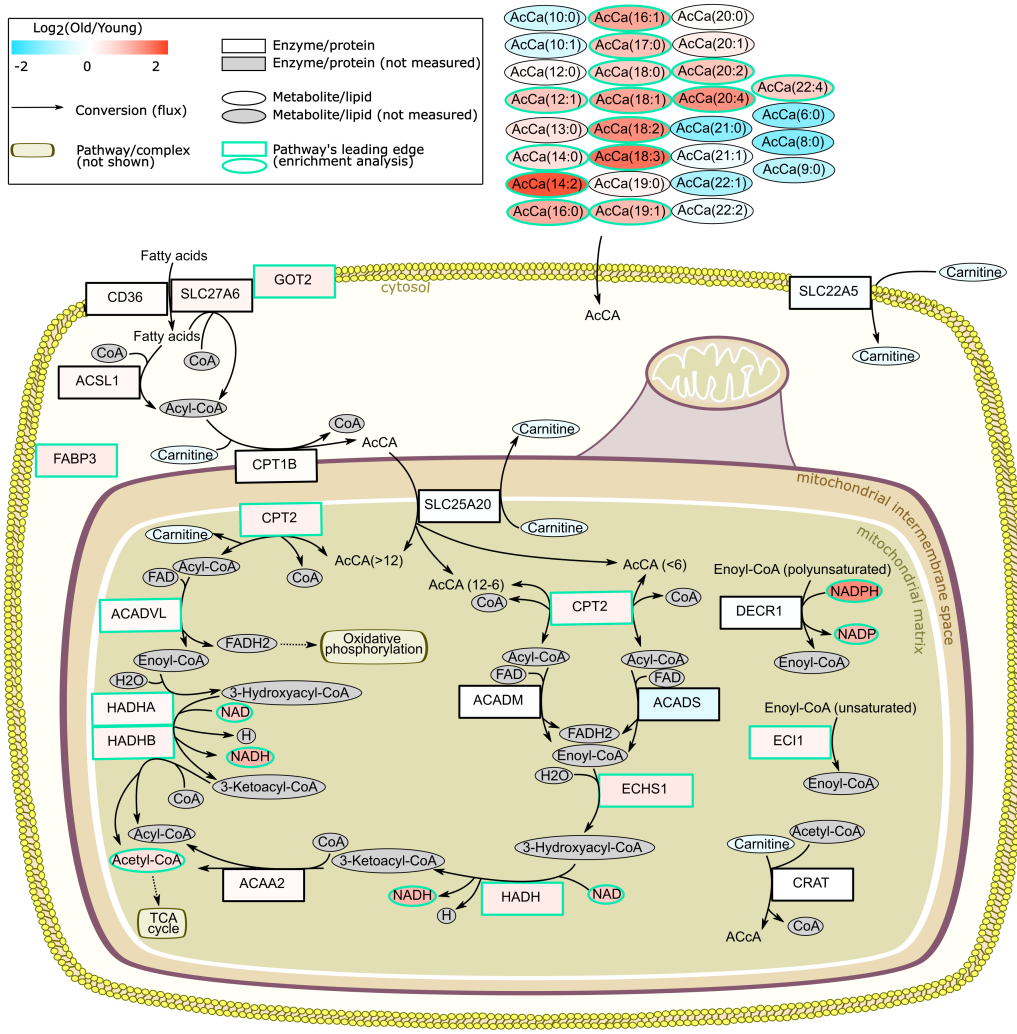


Figure 8.2: Schematic of β-oxidation.

To further explore if any of the 5 complexes belonging to oxidative phosphorylation showed age-related changes, I manually reconstructed each complex based on literature[663,753,883], followed by omics SEA on each of complex. The results are shown in Table 8.2. Omics

Complex	ES	p_{val}	p_{fdr}	Set Size
C I	-0.2034	0.6746	0.6767	51
C II	0.6414	0.2044	0.6767	5
C III	-0.2896	0.6155	0.6767	13
C IV	-0.3094	0.3633	0.6767	20
C V	0.2881	0.6767	0.6767	18

Table 8.2: Oxidative phosphorylation complexes.

SEA did not detect any age-related changes in the complexes belonging to oxidative phosphorylation. The expression pattern of complexes in the electron transport chain is displayed in Figure 8.3.

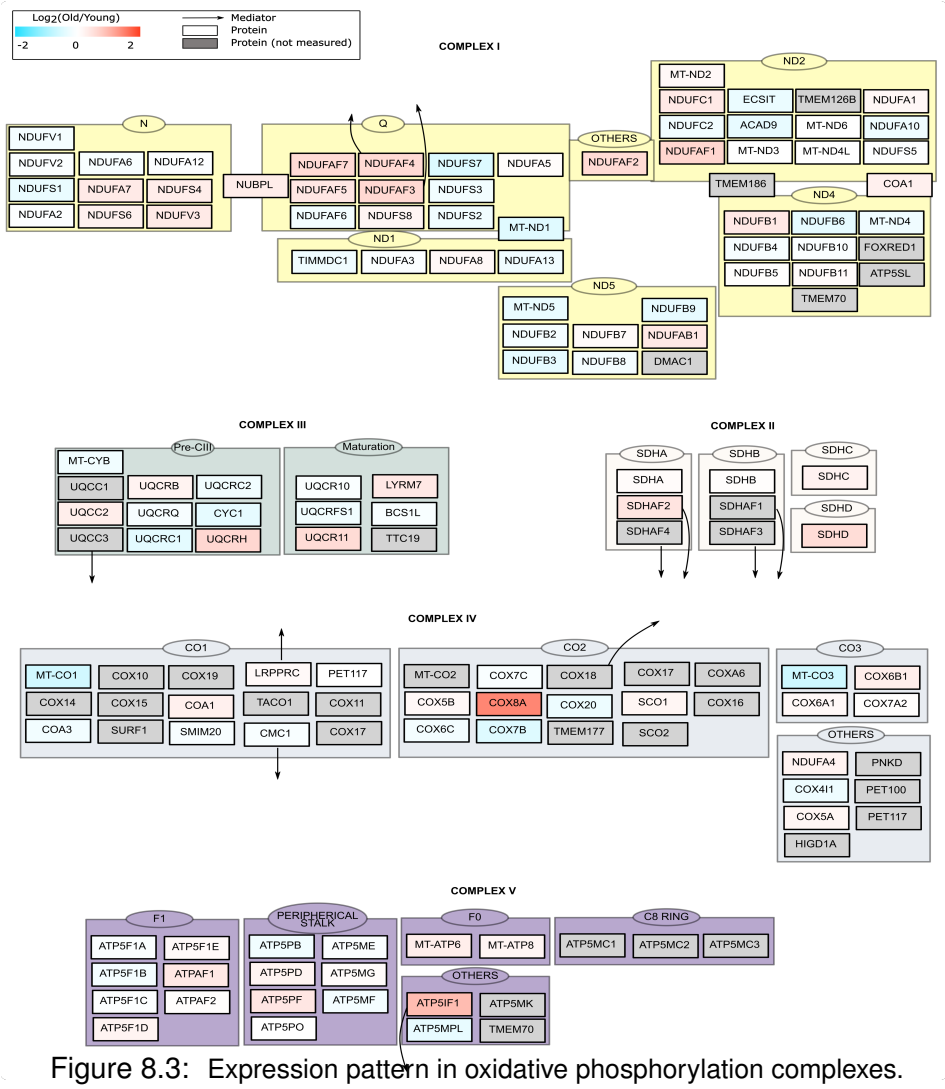


Figure 8.3: Expression pattern in oxidative phosphorylation complexes.

Further assessment of possible dysregulation of signalling pathways associated with cardiac ageing was conducted using Reactome[568]. The Reactome signalling pathways involved with ageing were selected based on literature[88,176,469,508,528,627,643,662,664,710,917] (see Chapter 2.4). Table 8.3 shows the results of the computational investigation on signalling pathways.

Signalling pathways	**ES**	$\mathbf{p_{val}}$	$\mathbf{p_{fdr}}$	**Set Size**
Regulation of IGF by IGFBP	0.3546	0.0061	0.2588	97
Signaling by VEGF	-0.2997	0.0267	0.4263	105
Regulation of FOXO by acetylation	0.6230	0.0304	0.42634	10
Advanced glycosylation receptor signaling	0.6489	0.0491	0.4630	8
NF-kB and signals survival	-0.5172	0.0584	0.4630	13
Regulation of NF-kappa B signaling	-0.4616	0.0744	0.4630	18
PI5P, PP2A and IER3 Regulate PI3K/AKT	-0.2713	0.0960	0.4630	90
JNK phosphorylation mediated by TAK1	-0.4356	0.1015	0.4630	18
PTK6 Regulates RTKs	-0.5436	0.1037	0.4630	8
IRS-related events triggered by IGF1R	0.3387	0.1154	0.4630	42

Table 8.3: Top 10 age-related signalling pathways in the hearts.

Conclusion

To sum up, proteins that are essential for efficient contraction mostly decreased in the older heart, including SERCA2 (or ATP2A2) and MYH6. The central cardiac energetic pathways (glycolysis, TCA cycle, beta-oxidation) were impaired in older hearts, which was also corroborated by the decrease in the PCr/ATP ratio (see Chapter 6.2). No dysregulation of oxidative phosphorylation was found at the complex level, and results were also supported by the non-changing ATP/ADP ratio when compared between young and old donors[534]. No significant dysregulation in signalling pathways was detected. Results presented throughout the chapter suggest a relatively inefficient contraction in older hearts accompanied by a, perhaps compensatory, reroute in myocardial energy production from β-oxidation to aerobic glycolysis. This hypothesis is also supported by further evidence, such as the accumulation of AcCAs and TGs, presented in Chapter 6.3. Taking into account the findings presented in this chapter and the past two chapters, I will discuss their implications for the cardiac ageing process in the next chapter.

Chapter 9

Discussion of the Findings

Throughout this thesis, I have provided the most comprehensive molecular characterisation of human cardiac ageing ever conducted on pre-mortem donor tissues, which certainly contributes to the entire scholarship. The molecular characterisation of cardiac ageing was conducted using high-throughput technologies (see Chapters 4.1 - 4.4) in order to capture molecular measurements in the proteome, metabolome, lipidome and transcriptome space. In this chapter, I discuss and interpret the findings and their implications while raising analogies with cardiovascular diseases. The visual interpretation of key findings, albeit not comprehensive, can be found in Figure 9.1.

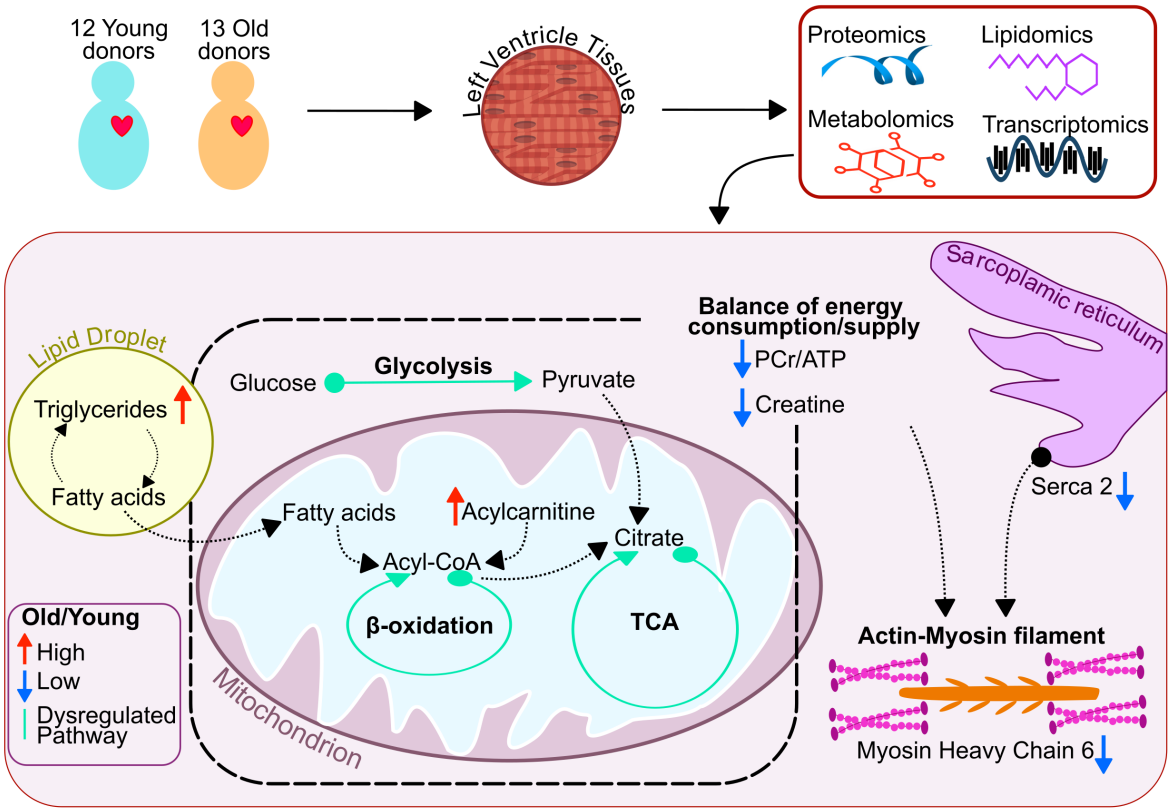


Figure 9.1: Summary of the findings.

9.1 Discussion

First of all, this study has been one of the first to thoroughly investigate and capture human cardiac ageing, and its peculiar samples should be acknowledged. Sample tissues used in our study were acquired from 25 pre-mortem hearts. Among the 25 donors, 12 donors were aged below 25, and 13 donors were over 50 (see Chapter 5.2). Furthermore, using the left ventricular tissue also allowed us to identify age-related changes in cardiomyocytes since histological studies have shown that they take up 85% of the volume of the tissue (see Chapter 2.2)[504,649,948]. A similar study on cardiac ageing was conducted using human heart left ventricular tissues excised from patients with cardiac pathologies who underwent bypass surgery. However, the study confined the investigation to subjects above 45 and below 85 years of age. Moreover, tissues were analysed using Western blotting and immunochemistry[264], technologies with lower sensitivity compared with MS[7,274]. Although further studies on cardiac ageing were also conducted on non-diseased cohorts, the tissues were excised post-mortem[138,180]. Even though studies on pre-mortem human tissue have been scarce, larger studies have been conducted on human ageing or on the development of cardiovascular diseases have been conducted on human plasma. For instance, Ramonfaur et al. (2024) have conducted a longitudinal observational study to link ageing with cardiac frailty, as well as heart failure with and without reperfusion[669]. In particular, Ramonfaur et al. (2024) show a causal association of EFEMP1, FSTL3, TREM, WFDC1, and GDF15

with cardiac frailty and heart failure, as well as alteration of fibrosis and immune pathways. Murashige et al. (2020) have also conducted a study on a large cohort of non-failing individuals, a study that aimed to characterise the fuel of non-failing cardiac hearts by analysing plasma samples across the transcardiac gradient[584]. The study revealed that non-failing hearts heavily depend on fuel retrieved from fatty acids, as well as amino acids, including glutamine[584]. In light of recent research efforts, our age-OME study can further our understanding of human cardiac ageing and become substantial to scholarship.

The molecular signature at the protein level presented in Chapter 6.1 included a reduction in the abundance of SERCA2 in the older cohort. This can suggest an initial stage of impaired diastolic filling that begins to develop as part of the normal ageing process[27,596]. In fact, SERCA2, an intracellular pump responsible for Ca^{2+} handling, clears the Ca^{2+} in the cytoplasm after the cardiac contraction by transferring it into the sarcoplasmic reticulum[213,630]. A reduction in SERCA2 can affect the uptake rate of Ca^{2+} by the sarcolemma reticulum from the cytosol, which delays cardiomyocyte relaxation and impacts cardiac contractility and functioning[945]. Experiments conducted on mice showed that SERCA2 knockout mice[640] and heterozygous SERCA2 mice had a significant decrease in diastolic function, supporting its importance in modulating cardiac relaxation. The age-associated decrease in SERCA2 expression was also observed in rats and mice models[80,309,798]. Notably, SERCA2 abundance and activity also decrease in heart failure both for rodents and humans; this can point to a possible therapeutic intervention for improving cardiac function in patients affected by cardiac failure[214,294,318,693].

Another molecular signature of cardiac ageing, presented in Chapter 6.1, is the decrease in expression of α-MYH6 in the aged cohort. A reduced quantity of MYH6 indicates a decline in cardiac contractility. Similar results were also reported in rodents' left ventricular tissue[112,511]; for instance, MYH6 abundance in rat hearts declined as rats grew older[511]. In a similar experiment, MYH6 appears to be a more prevalent isoform of myosin heavy chain in younger rats, but it decreases in older rats. On the one hand, the opposite trend is observed for β-myosin heavy chain, which becomes the prevalent isoform in older hearts[112]. In human cardiomyocytes, β-myosin heavy chain is the most common isoform, while MYH6 accounts for only the 7% of the total myosin heavy chain isoforms[575]. Despite its relatively low abundance, MYH6 requires higher ATPase activity, resulting in greater shortening velocity and power generation[112,708]. Additionally, levels of α-myosin heavy chains decrease to less than 1% of the total myosin heavy chain in failing human hearts[575]. Although they are not the most predominant isoform, a reduction in α-myosin heavy chains can lead to a less efficient contractility apparatus as humans age[519].

Multiple cathepsin proteins showed higher abundance in the older cohort in this study. Cathepsin proteins are a major component of lysosomal proteolysis[670] and are also present in cardiomyocytes. They are key in apoptosis, stress response and cardiometabolic stressors mediation[355]. Evidence shows that suppressing cathepsins-related genes in mice can alleviate cardiac dysfunction during ageing[354,593]. In more detail, the findings showed cathepsin D (CTSD) as upregulated in aged hearts, and a similar trend has been observed in circulating CSTD of patients with heart failure and diabetes[566]. In vitro experiments on human stem-cell cardiomyocytes have elucidated that silencing of CTSD increases

stress-induced metabolism, autophagy and necrosis[337], suggesting its role in a coping mechanism during cellular stress. Similarly, cathepsin B (CSTB) was upregulated in older donors for this study; its overabundance is generally observed during cardiac fibrosis, hypertrophy[615,892] and ageing[451,521]. Supporting findings regarding lysosome dysregulation, lysobisphosphatidic acid (LBPA), a lysosomal lipid, also increased overall in the aged cohort of donors, possibly responding to the accumulation of multiple lipids in the tissue and intermediates of glycolysis[208,371]. Dysregulation of lysosome homeostasis is usually associated with metabolic and lipid disorders[105] and can lead to impairments of apoptosis, cholesterol metabolisms, inflammation and protein homeostasis[505]. The impairment of lysosome homeostasis and lipid disorders has been recently linked to a reduction in the mitophagy process (mitochondrial autophagy) in rodents' cardiomyocytes[30]. Possible age-associated dysregulation of proteostasis in the heart was supported by our network analysis of the proteome, which detected positive co-tendency across the proteins involved in cellular proteostasis 7.1. Indeed, Klaips et al. (2018) observed aberrant functioning of protein homeostasis and accumulation of damaged proteins in the tissue with ageing, which supports our findings. One of the explanations can be the aberrant functioning of the lysosomal-mediated protein degradation machinery[853,882]. In fact, a failure to remove damaged protein can lead to cellular damage and cellular apoptosis[427], events that can cause the increase of damaged tissue and the loss of cells in the myocardial tissue with ageing[283]. In other words, proteostasis degrades damaged proteins, and the lysosome is the site where the degradation occurs; dysregulation of one inevitably affects the other, eventually leading to defective cardiac repair[96]. Defective cardiac repairment machinery stimulates the production of lysosomes to remove damaged cellular structures, which can explain the increased abundance of molecules related to lysosomes in older donors[234,810]. However, our data also suggest the possible consequences of aberrant protein homeostasis and lysosome; three proteins, closely related to each other, were found upregulated in the older cohort. Two of them are apolipoproteins, APOE and APOC3, which can be synthesised or uptake from the extracellular matrix by multiple cell types[53,491], while the other is the amyloid component, APCS, a well-established biomarker for amyloid deposition[321], and transported through plasma in the extracellular matrix[54]. Apolipoproteins play a cardioprotective role against heart failure[490] but also stimulate cardiac amyloidogenesis together with amyloid components[606], caused by aberrant accumulation of damaged proteins due to the dysregulated proteostasis and lysosomes[329,733]. The formation of nonhereditary amyloids in cardiac tissue is well associated with cardiomyopathies and cardiac ageing[634,865].

Oxidative stress emerges when there is an imbalance between the production and consumption of reactive oxygen species, and an increase in reactive oxygen species concomitant with a decrease in antioxidant defence is often observed during ageing[276,899]. Oxidative stress is extremely harmful to the cellular environment, as it can affect cellular structure and function and damage proteins[652] (see Chapters 2.3 and 2.4). Accumulation of ceramides, a class of lipids composed of sphingosine and fatty acids, is well associated with the insurgence of oxidative stress and inflammation as well as multiple cardiometabolic diseases, and can consequently become cardiotoxic[137,252]. Our findings suggest the age-related accumulation of ceramides in cardiac tissues, especially amongst ceramides with long and very long chain fatty acids, as reported in Chapter 6.3. Similar changes in ceramides were also reported in a cardiac ageing study performed on rats[750]. Furthermore,

oxidised glutathione, a metabolite promoting oxidative stress, increased in our older hearts (see Chapter 6.2). The reduced (GSH) and oxidised (GSSG) forms of glutathione, coupled with glutathione reductase and peroxidase, constitute one of the major antioxidant systems of the organism[31,296]. Further evidence of possible impairment of glutathione and its metabolism was also observed in our network analysis of the metabolome. A further assessment of the reduced/oxidised glutathione ratio, an indicator of oxidative stress[950], also pointed to a higher oxidative stress state in the older cohort by showing a significant increase and a possible decrease in glutathione reductase activity[896]. One of the possible causes of the reduction in glutathione reductase activity could be a reduction of klotho protein that can occur during ageing. Klotho is known as an anti-ageing protein, and its absence in mice is associated with accelerated ageing and impaired glutathione reductase[121,484]. While our omics measurements did not capture the klotho expression, our study indicated NADPH more abundant in the older cohort, a metabolite that removes part of the ROS produced by glutathione[896]. Yet, at the same time, its oxidised form, NADP, which contributes to oxidative stress, was elevated in the older donor[897]. Assessing the NADP/NADPH ratio between the young and old cohorts, we identified a significant decrease in the older one. A greater abundance of NADPH with respect to its oxidised form signifies a possible compensatory mechanism caused by oxidised glutathione, which copes with oxidative stress.

While NAD has often been expected to decrease in older age[2,906], our findings suggest the opposite; NAD and its reduced form, NADH, were detected in higher abundance in the older cohort of donors, which was further confirmed by detecting a positive association across the metabolites involved in NAD metabolism in the metabolite-metabolite network. The increase of NAD can indeed seem contradictory to the findings from previous studies, which assumed and supported the age-related decrease[2]. However, this assumption is not necessarily inferred from evidence[637], and the results from a recent study on human plasma partially support our findings[144]. Extending the analysis to the NAD/NADH ratio, we observed the ratio decreasing in the aged donors, which indicates that NADH increased in higher abundance than its oxidised form with ageing. Impairing cardiac contractility and relaxation and promoting dilated cardiomyopathies, the imbalance of the ratio toward the reduced state of NAD is associated with the initial stage of diabetic stress[134]. It has also been linked to the metabolic response to oxidative stress[896], induced likely by the elevated levels of oxidised glutathione. Failure to maintain sufficient NAD to balance its reduced state can also affect the cardiac energy state[2]. An imbalance in the NAD/NADH ratio is also associated with a change in sirtuins, a protein that modulates metabolism and its response to oxidative stress[14,372].

Possible dysregulation of the cardiac metabolic state can be observed from the nearly simultaneous decrease of multiple proteogenic amino acids, possibly due to the observed proteostasis impairment. Indeed, as damaged proteins are not efficiently cleared with ageing, fewer amino acids are produced, subsequently leading to lower activity of their metabolism[329]. Among the downregulated amino acids detected in older hearts in our study, 3 branched-chain amino acids (BCAA), isoleucine, valine and leucine, involved in cardiac energy supply. Although BCAA address a minor part of the energy supply of the cardiac muscle, its decrease is also observed in studies on heart failure[408] and cardiac ageing[174], suggesting its possible cardioprotective role and possibility as a therapeutic target[174].

Age-related changes in cardiac energy metabolism also involve lipids and their oxidation[899]. In the analysis of the lipidome, we observed the age-associated accumulation of triacylglycerol (or triglyceride), a lipid responsible for efficiently storing fatty acids. Accumulation of triglycerides was also observed in the study on mice cardiac tissue[224]. The increase in triacylglycerol with ageing appears to be a characteristic signature of heart and muscle tissues, while the lipid synthase does not exhibit any age-associated changes in cardiac and muscle tissues[940]. A higher abundance of triacylglycerol is also observed in mice with cardiometabolic conditions, for instance, mice affected by obesity[563]. Interestingly, a study in human plasma and mice skeletal muscles shows an increase in expression of apolipoprotein E (APOE) promoting the accumulation of triacylglycerol[360,365], and our findings corroborate this pattern. In the skeletal muscle of mice, the expression of APOE3, the most common allele of APOE, is also associated with lower fatty acid oxidation compared to a less common variant, APOE4, regardless of the type of diet that mice underwent[365]. Consequently, although APOE has a cardioprotective role against heart failure[490], its higher abundance can impair fatty acid oxidation. Despite the need for further study, our findings suggest impairment of fatty acid oxidation with regard to energy metabolism (see Chapter 8). To take place, β-oxidation requires the availability of fatty acids in the mitochondria matrix. While short and medium-chain fatty acids can freely diffuse through the mitochondrial membrane, long-chain fatty acids require a specialised transport system performed by the carnitine shuttle[695]. An increase in long-chain fatty acids was also supported by our data. As long-chain fatty acids increase with ageing[628], the proper functioning of the carnitine shuttle becomes important. The carnitine shuttle necessitates acetylcarnitine, a lipid that acts as a fatty acid carrier[695]. We observed an accumulation of acetylcarnitine in older donors simultaneously with a decrease in carnitine, which can indicate possible operative issues in the carnitine shuttle and reduced activity of the β-oxidation. This finding supports studies on the reduced activity of CPT1 in rat cardiac tissues[287] and age-related increase of acetylcarnitine in human plasma[272].

The reduced supply of acetyl-CoA through the β-oxidation combined with a state of oxidative stress can lead to an increase in activity of the glycolytic pathway without hypoxic condition[746,899]. Derangement of cardiac metabolism has also been proposed as a possible mechanism that triggers cardiac remodelling and develops heart failure[69]. Dysregulation in fatty acids oxidation was also confirmed by a reduction in PCr/ATP ratio, a metric inversely correlated with the abundance of circulating fatty acids[69]. The overall TCA intermediates in the older cohort were also observed in our metabolomics analysis, which was reinforced by our analysis of the metabolite-metabolite network and multi-omics analysis. A recent review has also highlighted that a higher abundance of TCA intermediates increases the frailty of the heart to develop heart failure[858]. The increase in TCA cycle activity was also supported by the decrease in the NAD/NADH ratio in the older cohort, whose quantity tightly correlated to the increase in the acetyl-CoA/CoA ratio[70]. This suggets that the citrate cycle has access to a larger pool of acetyl-CoA, likely produced by a high activity of aerobic glycolysis in order to compensate for the reduced activity of β-oxidation and sustain the optimal contraction. In oxidative phosphorylation, a potential co-tendency of the analyte shows its potential regulator role in the proteome ageing process, although this was not confirmed by our omics enrichment analysis of each complex in accord with the study that observed no change in

the electron transport chain in rat model[940].

Perturbation of the cardiac metabolism and cardiac contractility with ageing can also lead to a possible ATP accumulation, which might not be used as efficiently as in younger heart. Indeed, our findings support that there are no higher conversions of ATP from ADP, as the ATP/ADP ratio did not exhibit any age-related change. Additionally, a higher ATP abundance has also been associated with cardioprotection against diastolic dysfunction[777]. An impaired ATP utilisation is also supported by a decrease in myosin ATPase and SERCA2, two of the main consumers of ATP in the heart[827]. The age-related decrease of SERCA2 and SERCA3 can also be linked to an impairment of the contractility apparatus in the old cohort, a hypothesis also sustained by age-related decrease in myosin heavy chains 4 and 6. Further decrease in contractility is also suggested by a significant decline of the creatine availability in the aged cohort[49], coupled with a significant decrease in the PCr/ATP ratio; similar trends have also been found in heart failure[182]. A possible mechanistic explanation for the decrease in creatine can be the increase in guanidinoacetate detected in older hearts in our study. Guanidinoacetate is a precursor of creatine, and its increase along with the decrease of creatine can point to possible dysregulation of the creatine synthesis pathway[729].

9.2 Summary

Our analysis has drawn a unique insight into human cardiac ageing, and, in this chapter, I have discussed and contextualised the findings presented through Chapters 6-8. In particular, a reduction in SERCA proteins was observed in the older cohort together with myosin-heavy chains, suggesting a possible dysregulation of the contractility apparatus. At the same time, multiple changes in key molecules related to lysosomes and proteostasis were also detected, which are reflected in the overall reduction of amino acid availability in older cohorts with possible impairments of BCAA metabolism. Central carbon metabolism was upregulated, while β-oxidation appeared less active in older donors likely engulfed by the impaired carnitine shuttle. Furthermore, as the decrease in levels of mechanical proteins did not seem to occur in tandem with a decrease in energy provision, it can be presumed that decreased mechanical efficiency led to a ”backup” or accumulation of energetic substrates. However, according to Balaban (2002), there is a decrease both in the mechanical and energetic apparatus in heart failure, and conservation of energy generation may prevent the ageing heart from becoming the failing aged heart, inasmuch as the heart failure is an outcome of failed compensation for cardiac injury. Further mechanistic studies can be drawn from our findings to fully elucidate the dynamics that led to a generalised ATP increase in the old donor tissues.

Chapter 10

Conclusion, Limitations and Future Directions

Human cardiac ageing is a multifaceted phenomenon that requires high-throughput technologies and mathematical creativity to capture changes in the cardiac molecular landscape. While analytical and computational analysis provides invaluable insights into the mechanism of ageing in the biological system, biological intuition is also essential to link pieces of evidence together and explore possible mechanistic processes that drive the ageing process. This will inform further studies into the heart and ageing-related cardiovascular diseases such as heart failure, one of the leading causes of death in many countries[499,559,825,936]. In this thesis, I have dissected the mechanism of cardiac ageing based on three aims:

- Characterise the age-related changes in the cardiac molecular landscape.
- Identify regulators and functional modules of cardiac ageing for each omic biological network.
- Identify dysregulated biological processes that drive the human cardiac ageing.

This study extensively explores human cardiac ageing by characterising proteins, lipids, transcripts, and metabolites in pre-mortem human donor cardiac tissue. The characterisation of the ageing process in cardiac tissue has been carried out in multiple animal models and post-mortem human cardiac tissue[1,899], yet no study, up to our knowledge, has been conducted on pre-mortem cardiac tissues excised from donor's hearts. As Chapter 5 explained, our sample included 25 cryopreserved donors' left ventricular tissues stored in the facility of the Sydney Heart Bank[842], among which 12 were from young donor hearts (age ≤ 25), and 13 from old donors (age ≥ 50). The donors had a BMI < 30 with no cardiovascular diseases or comorbidities due to chronic diseases. We examined the molecular landscape of the two cohorts by measuring four main classes of molecules, proteins, metabolites, lipids and transcripts.

To address the first aim, I applied differential expression analysis to all the measured molecules and identified the molecular signature (see Chapter 4.8 and Chapter 6). This has enhanced our understanding of the molecular landscape of the human heart and how it evolves during ageing. Ratio analysis of selected metabolites was also conducted to

contextualise findings in the metabolomics measurements and subsequently evaluate the redox and cardiac energetic age-related changes. With regard to lipids' measurements, I further explored the association between ageing and the length of fatty acids tails as well as polyunsaturation in each lipid class. The overall length of the fatty acids tails was also evaluated. To further assess if the myocardial tissue has a preferential accumulation of a certain class of lipids with ageing, I applied lipid set enrichment analysis and identified if any accumulation or diminishing of specific lipid classes occurred. When analysing the transcripts, possible protein-transcript connections were also evaluated to find simultaneous trends between the transcription and translation machinery of analytes.

For the second aim, to understand the possible association between the molecules in the same omic layer and capture possible biological processes that arise from this holistic view of the omics, we constructed the coexpression network for proteins, metabolites, lipids and transcripts measurements (see Chapter 4.9 and Chapter 7). Analytes in each omics investigation derived from our donor tissue do not work alone but interact with each other both directly and indirectly[895]. For each inferred association network, we identified the largest interconnected community of analytes and linked it to potential biological processes that drive cardiac ageing. Additionally, my analysis also identified possible hubs that act as regulators/keepers of the ageing process.

I addressed the third aim by developing the novel method, omics SEA, and identifying the biological functions that were dysregulated (see Chapter 4.10.2 and Chapter 8). To enhance our understanding of human cardiac ageing, I used all available molecules' measurements to analyse age-related functional changes occurring in the myocardial tissue. I applied omics SEA to publicly available, manually curated biological pathways to explore age-related aberrant functioning in metabolic and signalling events.

The multifaceted exploration of cardiac ageing has provided some insights into the age-related changes in myocardial tissues, as discussed in detail in Chapter 9. First, my analysis has unveiled the age-related change in the cardiac contractility apparatus and calcium handling. The tissues from the old cohort also exhibited an accumulation of cathepsins accompanied by an increase in the accumulation of lysosome-related lipids. The findings further indicate the increased effort of the cardiac tissue in cleansing damaged proteins accumulated in the heart, which was also confirmed by the protein-protein association network and downregulation of amino acids. Supporting evidence of aberrant proteostatis was also emphasised by the presence of apolipoproteins and amyloid-related proteins. Further analysis also suggested that older donor hearts were on higher oxidative stress, as oxidised glutathione and ceramides were found in higher abundance in the aged donors. It can be considered that the anti-oxidant activities compensated for the alternated redox balance. Furthermore, the dysregulation at the metabolic level suggests that older hearts prefer central carbon metabolism rather than fatty acid oxidation and creatine for energy production and maintenance.

This study has contributed in multiple ways to the scholarship. Firstly, we analysed the ageing process in myocardial samples excited from pre-mortem donor hearts and applied multiple high-throughput technologies to characterise the molecular landscape in

young and old hearts. This allowed us to confirm multiple aspects already observed from animal and post-mortem studies conducted in the context of the ageing heart. Secondly, this study has identified that switching of the preferred substrate for the supply and maintenance of cardiac energy from fatty acids to glucose[899] already occurs in donors aged between 50 and 65. Showing some overlapping patterns with heart failure[858], this result supports that individuals over 50 are more susceptible to developing cardiovascular conditions[24,62] and metabolism is a driver of the cardiac ageing process. Additionally, we also observed a decrement of simultaneous downregulation of three branched amino acids, which aligns with the evidence of dysregulation of proteostasis. On a technical side, this thesis has advanced the field of omics analysis by introducing a novel workflow for co-expression analysis and lipids analysis, as well as a novel method to investigate the aberrantly regulated biological pathways. This study makes a further contribution by creating a publicly available repository where researchers can view our results and inspire future research.

The main limitation of this study is the sample size, as a larger number of samples could have allowed us to increase the number of findings in human pre-mortem cardiac ageing and extend our investigation across a wider age range. However, obtaining younger donor hearts in the modern era is almost impossible due to improvements in logistic facilities, immunosuppression techniques and surgical techniques, the increased number of transplantation centres and the priority of young donor hearts for transplantation. Although older donor hearts are still collected in other biobanks, the sample quality would not match, as our pre-mortem samples were obtained from individuals qualified for heart donation in the operative theatre. Intrinsic confounding factors regarding the lifestyle of the individuals can also be another limitation of the study. It is impossible to track the lifestyle of each donor, and donors lost their lives in unrelated and independent events. Therefore, no bias could be introduced in the selection of the samples. Furthermore, the healthy donor hearts utilised in this study were collected from patients who had been placed on life support following a fatal event. Therefore, we were unable to record information regarding their lifestyle and diet, as they were not chosen as samples before being declared brain dead to match possible confounders between the two cohorts. Despite such limitations, our sample collection method was the only ethical way to collect healthy human heart tissue. It is also important to note that we mitigated other important confounders. All the donor hearts were harvested from patients who did not have any comorbidities related to any chronic illnesses and cardiovascular diseases, such as diabetes, high BMI, hypertension and smoking, given the fact that they were ready to be transplanted. Beyond the randomness of deadly events, myocardial samples were randomised in the MS and sequencer to mitigate possible batch effects (see Chapter 5). Another limitation of our study is that only the whole myocardial tissue was analysed, and analysis at single cell or cellular compartments, such as mitochondria, would have helped us to better understand as to where the molecular changes can originate. Nonetheless, we applied the most reliable approach to examining the pre-mortem cryopreserved normal human myocardium of different ages with a limited sample availability. Although this approach does not distinguish if the changes are causative of cardiac ageing, detecting changes associated with ageing has already made a significant contribution that drives future research in the field. Indeed, the uniqueness of the samples and omics analyses allowed us to construct a dedicated public repository that benefits further studies on ageing (the repository will be updated and expanded accordingly at

https://humancardiac.shinyapps.io/ageing/).

The age-OME project has provided several points of entry for researchers to explore the mechanism of cardiac ageing and beyond. Further study is needed to confirm and replicate our results, yet it might not be possible due to the limited availability of pre-mortem tissues. One of the most suitable alternatives that can be employed to replicate and confirm the findings could be pigs; the porcine heart shows similar physiology, size, morphology and immunology to the human heart[525], although ethical issues might rise[583]. The use of pigs would allow researchers to extract, through biopsies[305], left ventricular tissues across different time points and perform a detailed map of molecular changes within pigs and across ages. The study's longitudinal design allows researchers to investigate the causality of the molecular modifications further and corroborate our findings[416,648]. For instance, one of the most exciting results that arise across literature and in our Age-OME study is the alteration of central carbon metabolism and beta-oxidation across ageing (see Chapter 2.4 and Chapter 9). To investigate this particular aspect, a longitudinal study on pigs can help us understand not only the causality of this change but also how the plasma and diet affect the alteration of metabolic processes during ageing by tracking blood plasma and food intakes. A study design as such can potentially unlock the inter-play food, metabolism and metabolic changes across ageing[512,737]. Another approach that does not involve animals and can be employed to replicate part of our findings could be utilising in-vitro induced pluripotent stem cells (iPSC) differentiated into cardiomyocytes[618]. The iPSC cardiomyocytes can emulate young and old cardiomyocyte phenotypes in control individuals and permit a targeted investigation of how young and old cells differ in molecular profile. The iPSC approach allows a researcher to investigate the effect of ageing on cardiomyocytes rather than tissue level, enhancing our understanding of how ageing affects the muscle cells and the role of cardiomyocytes in the cardiac frailty associated with ageing[5]. Additionally, iPSC omics measurements could be employed to estimate a co-expression network and corroborate the findings and insights obtained by the inferred network on human cardiac tissue (see Chapter 7). Once the network results are corroborated in iPSC, CRISP-based tools could elucidate the impact of knockdown or knock-out of NADUFAB1, ARPC4, and other hubs on the ageing network[311,540]. IPSC-derived cardiomyocytes could also provide a feasible model to monitor age-related changes in fluxes of the key substrates and intermediates of the central carbon metabolism as well as energy molecules such as ATP[743] and confirm our findings on the ATP and ADP accumulations with age and changes in metabolism. Further extension of our work can be linked to the molecular changes observed in the contractility apparatus and cardiac energy to the clinical indexes such as cardiac output, stroke ejection fraction and heartbeat, although in vivo experiments are yet required[100,785].

Bibliography

1 Abdellatif, M., Rainer, P. P., Sedej, S. and Kroemer, G. 2023, 'Hallmarks of cardiovascular ageing', *Nature Reviews Cardiology* **20**(11), 754–777.

2 Abdellatif, M., Sedej, S. and Kroemer, G. 2021, 'Nad+ metabolism in cardiac health, aging, and disease', *Circulation* **144**(22), 1795–1817.

3 Abdellatif, M., Trummer-Herbst, V., Heberle, A. M., Humnig, A., Pendl, T., Durand, S., Cerrato, G., Hofer, S. J., Islam, M., Voglhuber, J., Ramos Pittol, J. M., Kepp, O., Hoefler, G., Schmidt, A., Rainer, P. P., Scherr, D., von Lewinski, D., Bisping, E., McMullen, J. R. and Diwan, A. 2022, 'Fine-tuning cardiac insulin-like growth factor 1 receptor signaling to promote health and longevity', *Circulation* **145**(25).

4 Abzalov, M. and Abzalov, M. 2016, 'Exploratory data analysis', *Applied Mining Geology* pp. 207–219.

5 Acun, A., Nguyen, T. D. and Zorlutuna, P. 2019, 'In vitro aged, hipsc-origin engineered heart tissue models with age-dependent functional deterioration to study myocardial infarction', *Acta biomaterialia* **94**, 372–391.

6 Aebersold, R., Agar, J. N., Amster, I. J., Baker, M. S., Bertozzi, C. R., Boja, E. S., Costello, C. E., Cravatt, B. F., Fenselau, C., Garcia, B. A. et al. 2018, 'How many human proteoforms are there?', *Nature chemical biology* **14**(3), 206–214.

7 Aebersold, R., Burlingame, A. L. and Bradshaw, R. A. 2013, 'Western blots versus selected reaction monitoring assays: time to turn the tables?', *Molecular & Cellular Proteomics* **12**(9), 2381–2382.

8 Aeppli, T. R., Rymen, D., Allegri, G., Bode, P. K. and Häberle, J. 2020, 'Glycogen storage disease type vi: clinical course and molecular background', *European journal of pediatrics* **179**, 405–413.

9 Ahmad, T., Kelly, J. P., McGarrah, R. W., Hellkamp, A. S., Fiuzat, M., Testani, J. M., Wang, T. S., Verma, A., Samsky, M. D., Donahue, M. P. et al. 2016, 'Prognostic implications of long-chain acylcarnitines in heart failure and reversibility with mechanical circulatory support', *Journal of the American college of cardiology* **67**(3), 291–299.

10 Ahmed, Z. 2020, 'Practicing precision medicine with intelligently integrative clinical and multi-omics data analysis', *Human genomics* **14**(1), 35.

11 Aitken-Buck, H. M., Krause, J., Zeller, T., Jones, P. P. and Lamberts, R. R. 2020, 'Long-chain acylcarnitines and cardiac excitation-contraction coupling: links to arrhythmias', *Frontiers in Physiology* **11**, 577856.

12 Åkesson, J., Lubovac-Pilav, Z., Magnusson, R. and Gustafsson, M. 2021, 'Comhub: Community predictions of hubs in gene regulatory networks', *BMC bioinformatics* **22**, 1–12.

13 Akritas, A. G. and Malaschonok, G. I. 2004, 'Applications of singular-value decomposition (svd)', *Mathematics and computers in simulation* **67**(1-2), 15–31.

14 Alcendor, R. R., Gao, S., Zhai, P., Zablocki, D., Holle, E., Yu, X., Tian, B., Wagner, T., Vatner, S. F. and Sadoshima, J. 2007, 'Sirt1 regulates aging and resistance to oxidative stress in the heart', *Circulation research* **100**(10), 1512–1521.

15 Alex, L., Tuleta, I., Harikrishnan, V. and Frangogiannis, N. G. 2022, 'Validation of specific and reliable genetic tools to identify, label, and target cardiac pericytes in mice', *Journal of the American Heart Association* **11**(1).

16 Aliukonis, V., Poškutė, M. and Gefenas, E. 2020, 'Perish or publish dilemma: challenges to responsible authorship', *Medicina* **56**(3), 123.

17 Allison, D. B., Cui, X., Page, G. P. and Sabripour, M. 2006, 'Microarray data analysis: from disarray to consolidation and consensus', *Nature reviews genetics* **7**(1), 55–65.

18 Alseekh, S., Aharoni, A., Brotman, Y., Contrepois, K., D'Auria, J., Ewald, J., C. Ewald, J., Fraser, P. D., Giavalisco, P., Hall, R. D. et al. 2021, 'Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices', *Nature methods* **18**(7), 747–756.

19 Ambale-Venkatesh, B., Yoneyama, K., Sharma, R., Ohyama, Y., Wu, C., Burke, G., Shea, S., Gomes, A., Young, A., Bluemke, D. and Lima, J. 2016, 'Left ventricular shape predicts different types of cardiovascular events in the general population', *Heart* **103**(7).

20 Ambardekar, A. V. and Kopecky, B. J. 2024, 'The farm to table hfpef kitchen: Selecting the right ingredients for the discerning palate'.

21 Amorim, J. A., Coppotelli, G., Rolo, A. P., Palmeira, C. M., Ross, J. M. and Sinclair, D. A. 2022, 'Mitochondrial and metabolic dysfunction in ageing and age-related diseases', *Nature reviews. Endocrinology* **18**(4).

22 Amrhein, V., Greenland, S. and McShane, B. 2019, 'Scientists rise up against statistical significance', *Nature* **567**(7748), 305–307.

23 Anderson, R. H., Razavi, R. and Taylor, A. M. 2004, 'Cardiac anatomy revisited', *Journal of Anatomy* **205**(3).

24 Andersson, C. and Vasan, R. S. 2018, 'Epidemiology of cardiovascular disease in young individuals', *Nature Reviews Cardiology* **15**(4), 230–240.

25 Ando, K., Tong, L., Liu, Y., Vazquez-Liebanas, E., Chiyoda, H., He, L., Liu, J., Kawakami, K., Mochizuki, N., Fukuhara, S., Grutzendler, J. and Betsholtz, C. 2022, 'Kcnj8/abcc9-containing k-atp channel modulates brain vascular smooth muscle development and neurovascular coupling', *Developmental Cell* **57**(11).

26 Andreadou, I., Schulz, R., Papapetropoulos, A., Turan, B., Ytrehus, K., Ferdinandy, P., Daiber, A. and Di Lisa, F. 2020, 'The role of mitochondrial reactive oxygen species, no and h2s in ischaemia/reperfusion injury and cardioprotection. journal of cellular and molecular medicine', *Journal of cellular and molecular medicine* **24**(12).

27 Angeja, B. G. and Grossman, W. 2003, 'Evaluation and management of diastolic heart failure', *Circulation* **107**(5), 659–663.

28 Anna, P., Saira, A., Tomi, S., Thaman, C., Leena, S., Matti, P., Anne, R. et al. 2015, 'Optimization of statistical methods impact on quantitative proteomics data'.

29 Antezana, E., Kuiper, M. and Mironov, V. 2009, 'Biological knowledge management: the emerging role of the semantic web technologies', *Briefings in bioinformatics* **10**(4), 392–407.

30 Aon, M. A., Cortassa, S., Juhaszova, M., González-Reyes, J. A., Calvo-Rubio, M., Villalba, J. M., Lachance, A. D., Ziman, B. D., Mitchell, S. J., Murt, K. N. et al. 2021, 'Mitochondrial health is enhanced in rats with higher vs. lower intrinsic exercise capacity and extended lifespan', *npj Aging and Mechanisms of Disease* **7**(1), 1.

31 Aon, M. A., Cortassa, S., Maack, C. and O'Rourke, B. 2007, 'Sequential opening of mitochondrial ion channels as a function of glutathione redox thiol status', *Journal of Biological Chemistry* **282**(30), 21889–21900.

32 Aon, M., Bhatt, N. and Cortassa, S. 2014, 'Mitochondrial and cellular mechanisms for managing lipid excess', *Frontiers in Physiology* **5**.

33 Arrieta, A., Blackwood, E. A., Stauffer, W. T. and Glembotski, C. C. 2020, 'Integrating er and mitochondrial proteostasis in the healthy and diseased heart', *Frontiers in Cardiovascular Medicine* **6**, 193.

34 Atkinson, A. C., Riani, M. and Torti, F. 2016, 'Robust methods for heteroskedastic regression', *Computational Statistics & Data Analysis* **104**, 209–222.

35 Aunan, J. R., Watson, M. M., Hagland, H. R. and Soreide, K. 2016, 'Molecular and biological hallmarks of ageing', *British Journal of Surgery* **103**(2).

36 Austin, P. C., White, I. R., Lee, D. S. and van Buuren, S. 2021, 'Missing data in clinical research: a tutorial on multiple imputation', *Canadian Journal of Cardiology* **37**(9), 1322–1331.

37 Awwad, K., Hu, J., Shi, L., Mangels, N., Abdel Malik, R., Zippel, N., Fisslthaler, B., Eble, J. A., Pfeilschifter, J., Popp, R. and Fleming, I. 2015, 'Role of secreted modular calcium-binding protein 1 (smoc1) in transforming growth factor β signalling and angiogenesis', *Cardiovascular Research* **106**(2).

38 Baba, K., Shibata, R. and Sibuya, M. 2004, 'Partial correlation and conditional correlation as measures of conditional independence', *Australian & New Zealand Journal of Statistics* **46**(4), 657–664.

39 Babu, M. and Snyder, M. 2023, 'Multi-omics profiling for health', *Molecular & Cellular Proteomics* **22**(6).

40 Bachhawat, A. K., Yadav, S., Jainarayanan, A. K. and Dubey, P. 2020, 'Heart failure and the glutathione cycle: An integrated view', *Biochemical Journal* **477**(17), 3123–3130.

41 Badawi, S., Leboullenger, C., Chourrout, M., Gouriou, Y., Paccalet, A., Pillot, B., Augeul, L., Bolbos, R., Bongiovani, A., Mewton, N. et al. 2024, 'Oxidation–reduction imaging of myoglobin reveals two-phase oxidation in the reperfused myocardium', *Basic Research in Cardiology* pp. 1–17.

42 Bader, G. D., Cary, M. P. and Sander, C. 2006, 'Pathguide: a pathway resource list', *Nucleic acids research* **34**(suppl_1), D504–D506.

43 Bai, J., Melas, I., Hur, J. and Guo, E. 2017, 'Advances in omics for informed pharmaceutical research and development in the era of systems medicine', *Expert Opinion on Drug Discovery* **13**(1).

44 Bai, L., Kee, H. J., Han, X., Zhao, T., Kee, S. and Jeong, M. H. 2021, 'Protocatechuic acid attenuates isoproterenol-induced cardiac hypertrophy via downregulation of rock1–sp1–pkcγ axis', *Scientific Reports* **11**(1).

45 Bajic, Z., Sobot, T., Skrbic, R., Stojiljkovic, M. P., Ponorac, N., Matavulj, A. and Djuric, D. M. 2022, 'Homocysteine, vitamins b6 and folic acid in experimental models of myocardial infarction and heart failure—how strong is that link?', *Biomolecules* **12**(4), 536.

46 Balaban, R. S. 2002, 'Cardiac energy metabolism homeostasis: role of cytosolic calcium', *Journal of molecular and cellular cardiology* **34**(10), 1259–1271.

47 Balaban, R. S., Nemoto, S. and Finkel, T. 2005, 'Mitochondria, oxidants, and aging', *Cell* **120**(4).

48 Baldi, P. and Long, A. D. 2001, 'A bayesian framework for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes', *Bioinformatics* **17**(6), 509–519.

49 Balestrino, M. 2021, 'Role of creatine in the heart: Health and disease', *Nutrients* **13**(4), 1215.

50 Ballman, K. V., Grill, D. E., Oberg, A. L. and Therneau, T. M. 2004, 'Faster cyclic loess: normalizing rna arrays via linear models', *Bioinformatics* **20**(16), 2778–2786.

51 Banerjee, A., Chitnis, U., Jadhav, S., Bhawalkar, J. and Chaudhury, S. 2009, 'Hypothesis testing, type i and type ii errors', *Industrial psychiatry journal* **18**(2), 127–131.

52 Barcena, M. L., Tonini, G., Haritonow, N., Breiter, P., Milting, H., Baczko, I., Müller-Werdan, U., Ladilov, Y. and Regitz-Zagrosek, V. 2023, 'Sex and age differences in ampk phosphorylation, mitochondrial homeostasis, and inflammation in hearts from inflammatory cardiomyopathy patients', *Aging Cell* **22**(8), e13894.

53 Baroukh, N., Lopez, C. E., Saleh, M.-C., Recalde, D., Vergnes, L., Ostos, M. A., Fiette, L., Fruchart, J.-C., Castro, G., Zakin, M. M. et al. 2004, 'Expression and secretion of human apolipoprotein ai in the heart', *FEBS letters* **557**(1-3), 39–44.

54 Bart, N. K., Thomas, L., Korczyk, D., Atherton, J. J., Stewart, G. J. and Fatkin, D. 2020, 'Amyloid cardiomyopathy', *Heart, Lung and Circulation* **29**(4), 575–583.

55 Bartos, D. C., Grandi, E. and Ripplinger, C. M. 2015, 'Ion channels in the heart', *Comprehensive Physiology* **5**(3), 1423.

56 Battle, A., Khan, Z., Wang, S. H., Mitrano, A., Ford, M. J., Pritchard, J. K. and Gilad, Y. 2015, 'Genomic variation. impact of regulatory variation from rna to protein', *Science* **347**(6222).

57 Beckonert, O., Keun, H. C., Ebbels, T. M., Bundy, J., Holmes, E., Lindon, J. C. and Nicholson, J. K. 2007, 'Metabolic profiling, metabolomic and metabonomic procedures for nmr spectroscopy of urine, plasma, serum and tissue extracts', *Nature protocols* **2**(11), 2692.

58 Beesley, L. J. and Taylor, J. M. 2021, 'A stacked approach for chained equations multiple imputation incorporating the substantive model', *Biometrics* **77**(4), 1342–1354.

59 Begley, C. G. and Ioannidis, J. P. 2015, 'Reproducibility in science: improving the standard for basic and preclinical research', *Circulation research* **116**(1), 116–126.

60 Begum, R., Calaza, K., Kam, J. H., Salt, T. E., Hogg, C. and Jeffery, G. 2015, 'Near-infrared light increases atp, extends lifespan and improves mobility in aged drosophila melanogaster', *Biology Letters* **1**(3).

61 Benador, I. Y., Veliova, M., Mahdaviani, K., Petcherski, A., Wikstrom, J. D., Assali, E. A., Acín-Pérez, R., Shum, M., Oliveira, M. F., Cinti, S. et al. 2018, 'Mitochondria bound to lipid droplets have unique bioenergetics, composition, and dynamics that support lipid droplet expansion', *Cell metabolism* **27**(4), 869–885.

62 Benjamin, E. J., Blaha, M. J., Chiuve, S. E., Cushman, M., Das, S. R., Deo, R., De Ferranti, S. D., Floyd, J., Fornage, M., Gillespie, C. et al. 2017, 'Heart disease and stroke statistics—2017 update: a report from the american heart association', *circulation* **135**(10), e146–e603.

63 Benjamini, Y. and Hochberg, Y. 1995, 'Controlling the false discovery rate: a practical and powerful approach to multiple testing', *Journal of the Royal statistical society: series B (Methodological)* **57**(1), 289–300.

64 Benjamini, Y. and Speed, T. P. 2012, 'Summarizing and correcting the gc content bias in high-throughput sequencing', *Nucleic acids research* **40**(10), e72–e72.

65 Bennett, B. J., de Aguiar Vallim, T. Q., Wang, Z., Shih, D. M., Meng, Y., Gregory, J., Allayee, H., Lee, R., Graham, M., Crooke, R. et al. 2013, 'Trimethylamine-n-oxide, a metabolite associated with atherosclerosis, exhibits complex genetic and dietary regulation', *Cell metabolism* **17**(1), 49–60.

66 Bennett, J. L., Nguyen, G. T. and Donald, W. A. 2021, 'Protein–small molecule interactions in native mass spectrometry', *Chemical Reviews* **122**(8), 7327–7385.

67 Beretta, L. and Santaniello, A. 2016, 'Nearest neighbor imputation algorithms: a critical evaluation', *BMC medical informatics and decision making* **16**, 197–208.

68 Bersanelli, M., Mosca, E., Remondini, D., Giampieri, E., Sala, C., Castellani, G. and Milanesi, L. 2016, 'Methods for the integration of multi-omics data: mathematical aspects', *BMC bioinformatics* **17**, 167–177.

69 Bertero, E. and Maack, C. 2018, 'Metabolic remodelling in heart failure', *Nature Reviews Cardiology* **15**(8), 457–470.

70 Berthiaume, J. M., Kurdys, J. G., Muntean, D. M. and Rosca, M. G. 2019, 'Mitochondrial nad+/nadh redox state and diabetic cardiomyopathy', *Antioxidants & redox signaling* **30**(3), 375–398.

71 Bhuva, D. D., Cursons, J., Smyth, G. K. and Davis, M. J. 2019, 'Differential co-expression-based detection of conditional relationships in transcriptional data: comparative analysis and application to breast cancer', *Genome biology* **20**, 1–21.

72 Biau, D. J., Kernéis, S. and Porcher, R. 2008, 'Statistics in brief: the importance of sample size in the planning and interpretation of medical research', *Clinical Orthopaedics and Related Research®* **466**(9), 2282–2288.

73 Bildyug, N. 2019, 'Extracellular matrix in regulation of contractile system in cardiomyocytes', *International journal of molecular sciences* **20**(20), 5054.

74 Birkenfeld, A. L., Jordan, J., Dworak, M., Merkel, T. and Burnstock, G. 2019, 'Myocardial metabolism in heart failure: purinergic signalling and other metabolic concepts', *Pharmacology & therapeutics* **194**, 132–144.

75 Blagosklonny, M. V. 2019, 'Rapamycin for longevity: opinion article', *Aging* **11**(19).

76 Bland, J. M. and Altman, D. G. 1996, 'Statistics notes: Transforming data', *Bmj* **312**(7033), 770.

77 Bland, M. 2015, *An introduction to medical statistics*, Oxford university press.

78 Blice-Baum, A. C., Zambon, A. C., Kaushik, G., Viswanathan, M. C., Engler, A. J., Bodmer, R. and Cammarato, A. 2017, 'Modest overexpression offoxomaintains cardiac proteostasis and ameliorates age-associated functional decline', *Aging Cell* **16**(1).

79 Bligh, E. G. and Dyer, W. J. 1959, 'A rapid method of total lipid extraction and purification', *Canadian journal of biochemistry and physiology* **37**(8), 911–917.

80 Bodyak, N., Kang, P. M., Hiromura, M., Sulijoadikusumo, I., Horikoshi, N., Khrapko, K. and Usheva, A. 2002, 'Gene expression profiling of the aging mouse cardiac myocytes', *Nucleic acids research* **30**(17), 3788–3794.

81 Bokulich, N. A., Subramanian, S., Faith, J. J., Gevers, D., Gordon, J. I., Knight, R., Mills, D. A. and Caporaso, J. G. 2013, 'Quality-filtering vastly improves diversity estimates from illumina amplicon sequencing', *Nature methods* **10**(1), 57–59.

82 Boland, E., Quondamatteo, F. and Van Agtmael, T. 2021, 'The role of basement membranes in cardiac biology and disease', *Bioscience Reports* **41**(8), BSR20204185.

83 Bolstad, B. M., Irizarry, R. A., Åstrand, M. and Speed, T. P. 2003, 'A comparison of normalization methods for high density oligonucleotide array data based on variance and bias', *Bioinformatics* **19**(2), 185–193.

84 Bonnans, C., Chou, J. and Werb, Z. 2014, 'Remodelling the extracellular matrix in development and disease', *Nature Reviews Molecular Cell Biology* **15**(12).
URL: *doi:10.1038/nrm3904*

85 Borkum, J. M. 2023, 'The tricarboxylic acid cycle as a central regulator of the rate of aging: implications for metabolic interventions', *Advanced Biology* **7**(7), 2300095.

86 Borsboom, D., Deserno, M. K., Rhemtulla, M., Epskamp, S., Fried, E. I., McNally, R. J., Robinaugh, D. J., Perugini, M., Dalege, J., Costantini, G. et al. 2021, 'Network analysis of multivariate data in psychological science', *Nature Reviews Methods Primers* **1**(1), 58.

87 Bosch, L., Assmann, P., de Grauw, W. J. C., Schalk, B. W. M. and Biermans, M. C. J. 2019, 'Heart failure in primary care: prevalence related to age and comorbidity', *Primary Health Care Research & Development* **20**.

88 Bou-Teen, D., Kaludercic, N., Weissman, D., Turan, B., Maack, C., Di Lisa, F. and Ruiz-Meana, M. 2021, 'Mitochondrial ros and mitochondria-targeted antioxidants in the aged heart', *Free radical biology and medicine* **167**.

89 Bouchareychas, L. and Raffai, R. L. 2018, 'Apolipoprotein e and atherosclerosis: from lipoprotein metabolism to microrna control of inflammation', *Journal of cardiovascular development and disease* **5**(2), 30.

90 Bove, A. A. and Santamore, W. P. 1981, 'Ventricular interdependence', *Progress in Cardiovascular Diseases* **23**(5).

91 Bramer, L. M., Irvahn, J., Piehowski, P. D., Rodland, K. D. and Webb-Robertson, B.-J. M. 2020, 'A review of imputation strategies for isobaric labeling-based shotgun proteomics', *Journal of proteome research* **20**(1), 1–13.

92 Bratic, I. and Trifunovic, A. 2010, 'Mitochondrial energy metabolism and ageing', *Biochimica et Biophysica Acta (BBA)-Bioenergetics* **1797**(6-7), 961–967.

93 Breda, J., Zavolan, M. and van Nimwegen, E. 2021, 'Bayesian inference of gene expression states from single-cell rna-seq data', *Nature Biotechnology* **39**(8), 1008–1016.

94 Briggs, A., Clark, T., Wolstenholme, J. and Clarke, P. 2003, 'Missing.... presumed at random: cost-analysis of incomplete data', *Health economics* **12**(5), 377–392.

95 Broido, A. D. and Clauset, A. 2019, 'Scale-free networks are rare', *Nature communications* **10**(1), 1017.

96 Broughton, K. M., Wang, B. J., Firouzi, F., Khalafalla, F., Dimmeler, S., Fernandez-Aviles, F. and Sussman, M. A. 2018, 'Mechanisms of cardiac repair and regeneration', *Circulation research* **122**(8), 1151–1163.

97 Brubaker, D. K. and Lauffenburger, D. A. 2020, 'Translating preclinical models to humans', *Science* **367**(6479), 742–743.

98 Brügger, B. 2014, 'Lipidomics: analysis of the lipid composition of cells and subcellular organelles by electrospray ionization mass spectrometry', *Annual review of biochemistry* **83**, 79–98.

99 Buccitelli, C. and Selbach, M. 2020, 'mrnas, proteins and the emerging principles of gene expression control', *Nature Reviews Genetics* **21**(10), 630–644.

100 Budde, H., Hassoun, R., Mügge, A., Kovács, Á. and Hamdani, N. 2022, 'Current understanding of molecular pathophysiology of heart failure with preserved ejection fraction', *Frontiers in Physiology* **13**, 928232.

101 Bugger, H. and Abel, E. D. 2010, 'Mitochondria in the diabetic heart', *Cardiovascular research* **88**(2), 229–240.

102 Bugyi, F., Szabó, D., Szabo, G., Revesz, A., Pape, V. F., Soltész-Katona, E., Tóth, E., Kovács, O., Langó, T., Vékey, K. et al. 2021, 'Influence of post-translational modifications on protein identification in database searches', *ACS omega* **6**(11), 7469–7477.

103 Burgering., B. M. 2008, 'A brief introduction to foxology', *Oncogene* **27**(16).

104 C., P. 2016, 'Cardiovascular effects of cyclooxygenase-2 inhibitors: a mechanistic and clinical perspective', *British journal of clinical pharmacology* **82**(4).

105 Cabrera-Reyes, F., Parra-Ruiz, C., Yuseff, M. I. and Zanlungo, S. 2021, 'Alterations in lysosome homeostasis in lipid-related disorders: Impact on metabolic tissues and immune cells', *Frontiers in Cell and Developmental Biology* **9**, 790568.

106 Cai, Y., Zhou, Z. and Zhu, Z.-J. 2023, 'Advanced analytical and informatic strategies for metabolite annotation in untargeted metabolomics', *TrAC Trends in Analytical Chemistry* **158**, 116903.

107 Cajka, T., Smilowitz, J. T. and Fiehn, O. 2017, 'Validating quantitative untargeted lipidomics across nine liquid chromatography–high-resolution mass spectrometry platforms', *Analytical chemistry* **89**(22), 12360–12368.

108 Campbell, N. A., Urry, L. A., Cain, M. L., Wasserman, S. A. and Minorsky, P. V. 2018, *Biology: A Global Approach, Global Edition, 11th Edition*, Pearson, London, UK.

109 Canzler, S. and Hackermüller, J. 2020, 'multigsea: a gsea-based pathway enrichment analysis for multi-omics data', *BMC Bioinformatics* **21**(561).

110 Capilupi, M. J., Kerath, S. M. and Becker, L. B. 2020, 'Vagus nerve stimulation and the cardiovascular system', *Cold Spring Harbor perspectives in medicine* **10**(2), a034173.

111 Caporizzo, M. A., Chen, C. Y. and Prosser, B. L. 2019, 'Cardiac microtubules in health and heart disease', *Experimental Biology and Medicine* **244**(15), 1255–1272.

112 Carnes, C. A., Geisbuhler, T. P. and Reiser, P. J. 2004, 'Age-dependent changes in contraction and regional myocardial myosin heavy chain isoform expression in rats', *Journal of applied physiology* **97**(1), 446–453.

113 Carpi, A., Menabò, R., Kaludercic, N., Pelicci, P., Di Lisa, F. and Giorgio, M. 2009, 'The cardioprotective effects elicited by p66shc ablation demonstrate the crucial role of mitochondrial ros formation in ischemia/reperfusion injury', *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1787**(7).

114 Carubelli, V., Castrini, A. I., Lazzarini, V., Gheorghiade, M., Metra, M. and Lombardi, C. 2015, 'Amino acids and derivatives, a new treatment of chronic heart failure?', *Heart failure reviews* **20**, 39–51.

115 Casella, G. 1985, 'An introduction to empirical bayes data analysis', *The American Statistician* **39**(2), 83–87.

116 Casella, G. and Berger, R. 2002, *Statistical Inference*, Duxbury Resource Center, Warsaw, Poland.

117 Caulfield, J. and Borg, T. K. 1979, 'The collagen network of the heart', *Laboratory investigation; a journal of technical methods and pathology* **40**(3).

118 Chamberlain, C. A., Rubio, V. Y. and Garrett, T. J. 2019, 'Impact of matrix effects and ionization efficiency in non-quantitative untargeted metabolomics', *Metabolomics* **15**, 1–9.

119 Chaudhari, S. N. and Kipreos, E. T. 2018, 'The energy maintenance theory of aging: Maintaining energy metabolism to allow longevity', *BioEssays: news and reviews in molecular, cellular and developmental biology* **40**(8).

120 Chen, J., Wang, S., Tsai, C. and Lin, C. 2007, 'Selection of differentially expressed genes in microarray data analysis', *The pharmacogenomics journal* **7**(3), 212–220.

121 Chen, K., Wang, S., Sun, Q. W., Zhang, B., Ullah, M. and Sun, Z. 2021, 'Klotho deficiency causes heart aging via impairing the nrf2-gr pathway', *Circulation research* **128**(4), 492–507.

122 Chen, L. T. L., Yao, Q., Soares, T. A., Squier, T. C. and Bigelow, D. J. 2009, 'Phospholamban modulates the functional coupling between nucleotide domains in ca-atpase oligomeric complexes in cardiac sarcoplasmic reticulum', *Biochemistry* **48**(11).

123 Chen, L., Zhong, F. and Zhu, J. 2020, 'Bridging targeted and untargeted mass spectrometry-based metabolomics via hybrid approaches', *Metabolites* **10**(9), 348.

124 Chen, M. S., Lee, R. T. and Garbern, J. C. 2022, 'Senescence mechanisms and targets in the heart', *Cardiovascular Research* **118**(5), 1173–1187.

125 Chen, R., Morris, B. J., Donlon, T. A., Masaki, K. H., Willcox, D. C., Davy, P. M. C., Allsopp, R. C. and Willcox, B. J. 2020, 'Foxo3 longevity genotype mitigates the increased mortality risk in men with a cardiometabolic disease', *Aging* **12**(23).

126 Chen, W.-Y., Chen, Y.-F., Chan, H.-C., Chung, C.-H., Peng, H.-Y., Ho, Y.-C., Chen, C.-H., Chang, K.-C., Tang, C.-H. and Lee, A.-S. 2020, 'Role of apolipoprotein e in electronegative low-density lipoprotein-induced mitochondrial dysfunction in cardiomyocytes', *Metabolism* **107**, 154227.

127 Chen, X., Liu, Y., Xu, C., Ba, L., Liu, Z., Li, X., Huang, J., Simpson, E., Gao, H., Cao, D. et al. 2021, 'Qki is a critical pre-mrna alternative splicing regulator of cardiac myofibrillogenesis and contractile function', *Nature communications* **12**(1), 89.

128 Chen, Y., Fry, B. C. and Layton, A. T. 2017, 'Modeling glucose metabolism and lactate production in the kidney', *Mathematical Biosciences* .

129 Chen, Y., Wu, G., Li, M., Hesse, M., Ma, Y., Chen, W., Huang, H., Liu, Y., Xu, W., Tang, Y., Zheng, H., Li, C., Lin, Z., Chen, G., Liao, W., Liao, Y., Bin, J. and Chen, Y. 2022, 'Ldha-mediated metabolic reprogramming promoted cardiomyocyte proliferation by alleviating ros and inducing m2 macrophage polarization', *Redox biology* **56**.

130 Cheng, D., Zhang, S., Deng, Z., Zhu, Y. and Zong, M. 2014, k nn algorithm with data-driven k value, *in* 'Advanced Data Mining and Applications: 10th International Conference, ADMA 2014, Guilin, China, December 19-21, 2014. Proceedings 10', Springer, pp. 499–512.

131 Cheng, F., Zhao, J., Wang, Y., Lu, W., Liu, Z., Zhou, Y., Martin, W. R., Wang, R., Huang, J., Hao, T. et al. 2021, 'Comprehensive characterization of protein–protein interactions perturbed by disease mutations', *Nature genetics* **53**(3), 342–353.

132 Cheng, X. W., Shi, G.-P., Kuzuya, M., Sasaki, T., Okumura, K. and Murohara, T. 2012, 'Role for cysteine protease cathepsins in heart disease: focus on biology and mechanisms with clinical implication', *Circulation* **125**(12), 1551–1562.

133 Chhangawala, S., Rudy, G., Mason, C. E. and Rosenfeld, J. A. 2015, 'The impact of read length on quantification of differentially expressed genes and splice junction detection', *Genome biology* **16**, 1–10.

134 Chiao, Y. A., Chakraborty, A. D., Light, C. M., Tian, R., Sadoshima, J., Shi, X., Gu, H. and Lee, C. F. 2021, 'Nad+ redox imbalance in the heart exacerbates diabetic cardiomyopathy', *Circulation: Heart Failure* **14**(8), e008170.

135 Chicco, D., Warrens, M. J. and Jurman, G. 2021, 'The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation', *Peerj computer science* **7**, e623.

136 Chodary Khameneh, S., Razi, S., Shamdani, S., Uzan, G. and Naserian, S. 2022, 'Weighted correlation network analysis revealed novel long non-coding rnas for colorectal cancer', *Scientific reports* **12**(1).

137 Choi, R. H., Tatum, S. M., Symons, J. D., Summers, S. A. and Holland, W. L. 2021, 'Ceramides and other sphingolipids as drivers of cardiovascular disease', *Nature reviews Cardiology* **18**(10), 701–711.

138 Choudhury, S., Huang, A. Y., Kim, J., Zhou, Z., Morillo, K., Maury, E. A., Tsai, J. W., Miller, M. B., Lodato, M. A., Araten, S. et al. 2022, 'Somatic mutations in single human cardiomyocytes reveal age-associated dna damage and widespread oxidative genotoxicity', *Nature aging* **2**(8), 714–725.

139 Christopher, J. A., Geladaki, A., Dawson, C. S., Vennard, O. L. and Lilley, K. S. 2022, 'Subcellular transcriptomics and proteomics: a comparative methods review', *Molecular & Cellular Proteomics* **21**(2).

140 Chu, Y., Rosso, L. G., Huang, P., Wang, Z., Xu, Y., Yao, X., Bao, M., Yan, J., Song, H. and Wang, G. 2014, 'Liver med23 ablation improves glucose and lipid metabolism through modulating foxo1 activity', *Cell research* **24**(10), 1250–1265.

141 Clark, B. 2011, 'The mammalian start domain protein family in lipid transport in health and disease', *Journal of Endocrinology* **212**(3).

142 Clark, N. R., Hu, K. S., Feldmann, A. S., Kou, Y., Chen, E. Y., Duan, Q. and Ma'ayan, A. 2014, 'The characteristic direction: a geometrical approach to identify differentially expressed genes', *BMC bioinformatics* **15**, 1–16.

143 Clauset, A., Shalizi, C. R. and Newman, M. E. 2009, 'Power-law distributions in empirical data', *SIAM review* **51**(4), 661–703.

144 Clement, J., Wong, M., Poljak, A., Sachdev, P. and Braidy, N. 2019, 'The plasma nad+ metabolome is dysregulated in "normal" aging', *Rejuvenation research* **22**(2), 121–130.

145 Coats, A. J. S. 2019, 'Ageing, demographics, and heart failure', *European Heart Journal Supplements* **21**(Supplement_L), L4–L7.

146 Cochet-Bissuel, M., Lory, P. and Monteil, A. 2014, 'The sodium leak channel, nalcn, in health and disease', *Frontiers in cellular neuroscience* **8**, 132.

147 Cock, P. J., Fields, C. J., Goto, N., Heuer, M. L. and Rice, P. M. 2010, 'The sanger fastq file format for sequences with quality scores, and the solexa/illumina fastq variants', *Nucleic acids research* **38**(6), 1767–1771.

148 Cockcroft, S. 2021, 'Mammalian lipids: structure, synthesis and function', *Essays in biochemistry* **65**(5).

149 Colliva, A., Braga, L., Giacca, M. and Zacchigna, S. 2019, 'Endothelial cell–cardiomyocyte crosstalk in heart development and disease', *The Journal of Physiology* **598**(14).

150 Comunale, G., Peruzzo, P., Castaldi, B., Razzolini, R., Di Salvo, G., Padalino, M. A. and Susin, F. M. 2021, 'Understanding and recognition of the right ventricular function and dysfunction via a numerical study', *Scientific reports* **11**(1), 3709.

151 Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X. et al. 2016, 'A survey of best practices for rna-seq data analysis', *Genome biology* **17**, 1–19.

152 Consortium, L. S. I. 2019, 'Lipidomics needs more standardization', *Nature Metabolism* **1**(8).

153 Cordell, H. J. 2009, 'Detecting gene–gene interactions that underlie human diseases', *Nature Reviews Genetics* **10**(6), 392–404.

154 Correia-Melo, C., Birch, J., Fielder, E., Rahmatika, D., Taylor, J., Chapman, J., Lagnado, A. B., Carroll, B., Miwa, S., Richardson, G. D., Jurk, D., Oakley, F., Mann, J., Mann, D. A., Korolchuk, V. I. and Passos, J. F. 2018, 'Rapamycin improves healthspan but not inflammaging in nf κ b1-/- mice', *Aging Cell* **18**(1).

155 Cortassa, S., Caceres, V., Bell, L. N., O'Rourke, B., Paolocci, N. and Aon, M. A. 2015, 'From metabolomics to fluxomics: a computational procedure to translate metabolite profiles into metabolic fluxes', *Biophysical Journal* **108**(1), 163–172.

156 Cortassa, S., Caceres, V., Tocchetti, C. G., Bernier, M., de Cabo, R., Paolocci, N., Sollott, S. J. and Aon, M. A. 2020, 'Metabolic remodelling of glucose, fatty acid and redox pathways in the heart of type 2 diabetic mice', *The Journal of physiology* **598**(7), 1393–1415.

157 Cosentino, F., Francia, P., Camici, G. G., Pelicci, P. G., Volpe, M. and F., L. 2008, 'Final common molecular pathways of aging and cardiovascular disease', *Arteriosclerosis, Thrombosis, and Vascular Biology* **28**(4).

158 Costantino, S., Paneni, F. and Cosentino, F. 2016, 'Ageing, metabolism and cardiovascular disease', *The Journal of physiology* **594**(8), 2061–2073.

159 Cox, J. and Mann, M. 2008, 'Maxquant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification', *Nature biotechnology* **26**(12), 1367–1372.

160 Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V. and Mann, M. 2011, 'Andromeda: a peptide search engine integrated into the maxquant environment', *Journal of proteome research* **10**(4), 1794–1805.

161 Cramer, P. 2019, 'Organization and regulation of gene transcription', *Nature* **573**(7772), 45–54.

162 Crook, A. A. and Powers, R. 2020, 'Quantitative nmr-based biomedical metabolomics: Current status and applications', *Molecules* **25**(21), 5128.

163 Cruz, A. L. S., Barreto, E. d. A., Fazolini, N. P. B., Viola, J. P. B. and Bozza, P. T. 2020, 'Lipid droplets: platforms with multiple functions in cancer hallmarks', *Cell Death & Disease* **11**(2).

164 Cui, M., Cheng, C. and Zhang, L. 2022, 'High-throughput proteomics: a methodological mini-review', *Laboratory Investigation* **102**(11), 1170–1181.

165 Curtis, M. M. and Sperandio, V. 2011, 'A complex relationship: the interaction among symbiotic microbes, invading pathogens, and their mammalian host', *Mucosal immunology* **4**(2), 133–138.

166 Dahal, S., Yurkovich, J. T., Xu, H., Palsson, B. O. and Yang, L. 2020, 'Synthesizing systems biology knowledge from omics using genome-scale models', *Proteomics* **20**(17-18), 1900282.

167 Dai, D. F., Chiao, Y., Marcinek, D. J., Szeto, H. H. and Rabinovitch, P. S. 2014, 'Mitochondrial oxidative stress in aging and healthspan', *Longevity & Healthspan* **3**(1).

168 Dai, Y., Chen, D. and Xu, T. 2022, 'Dna methylation aberrant in atherosclerosis', *Frontiers in Pharmacology* **13**, 815977.

169 Dai, Z., Mentch, S. J., Gao, X., Nichenametla, S. N. and Locasale, J. W. 2018, 'Methionine metabolism influences genomic architecture and gene expression through h3k4me3 peak width', *Nature communications* **9**(1), 1955.

170 Dalman, M. R., Deeter, A., Nimishakavi, G. and Duan, Z.-H. 2012, Fold change and p-value cutoffs significantly alter microarray interpretations, *in* 'BMC bioinformatics', Vol. 13, Springer, pp. 1–4.

171 Dambrova, M., Zuurbier, C. J., Borutaite, V., Liepinsh, E. and Makrecka-Kuka, M. 2021, 'Energy substrate metabolism and mitochondrial oxidative stress in cardiac ischemia/reperfusion injury', *Free Radical Biology and Medicine* **165**.

172 Daneshgar, N., Rabinovitch, P. S. and Dai, D.-F. 2021, 'Tor signaling pathway in cardiac aging and heart failure', *Biomolecules* **11**(2), 168.

173 Dann, E., Henderson, N. C., Teichmann, S. A., Morgan, M. D. and Marioni, J. C. 2022, 'Differential abundance testing on single-cell data using k-nearest neighbor graphs', *Nature biotechnology* **40**(2), 245–253.

174 D'Antona, G., Ragni, M., Cardile, A., Tedesco, L., Dossena, M., Bruttini, F., Caliaro, F., Corsetti, G., Bottinelli, R., Carruba, M. O. et al. 2010, 'Branched-chain amino acid supplementation promotes survival and supports cardiac and skeletal muscle mitochondrial biogenesis in middle-aged mice', *Cell metabolism* **12**(4), 362–372.

175 Danzi, F., Pacchiana, R., Mafficini, A., Scupoli, M. T., Scarpa, A., Donadelli, M. and Fiore, A. 2023, 'To metabolomics and beyond: a technological portfolio to investigate cancer metabolism', *Signal Transduction and Targeted Therapy* **8**(1), 137.

176 Davidson, S. M., Adameova, A., Barile, L., Cabrera Fuentes, H. A., Lazou, A., Pagliaro, P., Stenslokken, K., Garcia-Dorado, D. and on behalf of the EU-CARDIOPROTECTION COST Action (CA16225) 2020, 'Mitochondrial and mitochondrial-independent pathways of myocardial cell death during ischaemia and reperfusion injury', *Journal of Cellular and Molecular Medicine* **24**.

177 de Goede, O. M., Nachun, D. C., Ferraro, N. M., Gloudemans, M. J., Rao, A. S., Smail, C., Eulalio, T. Y., Aguet, F., Ng, B., Xu, J. et al. 2021, 'Population-scale tissue transcriptomics maps long non-coding rnas to complex disease', *Cell* **184**(10), 2633–2648.

178 de la Fuente, A. 2010, 'From 'differential expression'to 'differential networking'–identification of dysfunctional regulatory networks in diseases', *Trends in genetics* **26**(7), 326–333.

179 de Wit-Verheggen, V. H. W., Schrauwen-Hinderling, V. B., Brouwers, K., Jorgensen, J. A., Schaart, G., Gemmink, A., Nascimento, E. B. M., Hesselink, M. K. C., Wildberger, J. E., Segers, P., Montaigne, D., Staels, B., Schrauwen, P., Lindeboom, L., Hoeks, J. and van de Weijer, T. 2023, 'Pcr/atp ratios and mitochondrial function in the heart. a comparative study in humans', *Scientific reports* **13**(1).

180 Debessa, C. R. G., Maifrino, L. B. M. and de Souza, R. R. 2001, 'Age related changes of the collagen network of the human heart', *Mechanisms of ageing and development* **122**(10), 1049–1058.

181 Decruyenaere, P., Verniers, K., Poma-Soto, F., Van Dorpe, J., Offner, F. and Vandesompele, J. 2023, 'Rna extraction method impacts quality metrics and sequencing results in formalin-fixed, paraffin-embedded tissue samples', *Laboratory Investigation* **103**(2), 100027.

182 Del Franco, A., Ambrosio, G., Baroncelli, L., Pizzorusso, T., Barison, A., Olivotto, I., Recchia, F. A., Lombardi, C. M., Metra, M., Ferrari Chen, Y. F. et al. 2022, 'Creatine deficiency and heart failure', *Heart Failure Reviews* pp. 1–12.

183 Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. and Ralser, M. 2020, 'Diann: neural networks and interference correction enable deep proteome coverage in high throughput', *Nature methods* **17**(1), 41–44.

184 Demissie, M., Mascialino, B., Calza, S. and Pawitan, Y. 2008, 'Unequal group variances in microarray data analyses', *Bioinformatics* **24**(9), 1168–1174.

185 DePina, A. S. and Langford, G. M. 1999, 'Vesicle transport: the role of actin filaments and myosin motors', *Microscopy research and technique* **47**(2).

186 Depre, C., Vanoverschelde, J. L., J. and Taegtmeyer, H. 1999, 'Glucose for the heart', *Circulation* **99**(4).

187 Deutsch, E. W. 2012, 'File formats commonly used in mass spectrometry proteomics', *Molecular & cellular proteomics* **11**(12), 1612–1621.

188 Deutsch, E. W., Vizcaíno, J. A., Jones, A. R., Binz, P.-A., Lam, H., Klein, J., Bittremieux, W., Perez-Riverol, Y., Tabb, D. L., Walzer, M. et al. 2023, 'Proteomics standards initiative at twenty years: current activities and future work', *Journal of proteome research* **22**(2), 287–301.

189 Deyneko, I. V., Mustafaev, O. N., Tyurin, A., Zhukova, K. V., Varzari, A. and Goldenkova Pavlova, I. V. 2022, 'Modeling and cleaning rna-seq data significantly improve detection of differentially expressed genes', *BMC bioinformatics* **23**(1), 488.

190 Di, L., Fu, Y., Sun, Y., Li, J., Liu, L., Yao, J., Wang, G., Wu, Y., Lao, K., Lee, R. W. et al. 2020, 'Rna sequencing by direct tagmentation of rna/dna hybrids', *Proceedings of the National Academy of Sciences* **117**(6), 2886–2893.

191 Di Siena, S., Gimmelli, R., Nori, S. L., Barbagallo, F., Campolo, F., Dolci, S., Rossi, P., Venneri, M., Giannetta, E., Gianfrilli, D. et al. 2016, 'Activated c-kit receptor in the heart promotes cardiac repair and regeneration after injury', *Cell death & disease* **7**(7), e2317–e2317.

192 Diaba-Nuhoho, P. and Amponsah-Offeh, M. 2021, 'Reproducibility and research integrity: the role of scientists and institutions', *BMC Research Notes* **14**, 1–4.

193 Dieterle, F., Ross, A., Schlotterbeck, G. and Senn, H. 2006, 'Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. application in 1h nmr metabonomics', *Analytical chemistry* **78**(13), 4281–4290.

194 Dillmann, W. 2010, 'Cardiac hypertrophy and thyroid hormone signaling', *Heart failure reviews* **15**, 125–132.

195 Ding, S., Li, G., Fu, T., Zhang, T., Lu, X., Li, N. and Geng, Q. 2024, 'Ceramides and mitochondrial homeostasis', *Cellular Signalling* p. 111099.

196 Do, K. T., Wahl, S., Raffler, J., Molnos, S., Laimighofer, M., Adamski, J., Suhre, K., Strauch, K., Peters, A., Gieger, C. et al. 2018, 'Characterization of missing values in untargeted ms-based metabolomics data and evaluation of missing data handling strategies', *Metabolomics* **14**, 1–18.

197 Doenst, T., Nguyen, T. D. and Abel, E. D. 2013, 'Cardiac metabolism in heart failure: implications beyond atp production', *Circulation research* **113**(6), 709–724.

198 Donders, A. R. T., Van Der Heijden, G. J., Stijnen, T. and Moons, K. G. 2006, 'A gentle introduction to imputation of missing values', *Journal of clinical epidemiology* **59**(10), 1087–1091.

199 dos Remedios C., Lal, S., Li, A., McNamara, J., Keogh, A., Macdonald, P., Cooke, R., Ehler, E., R., K., Marston, S., Stelzer, J., Granzier, H., Bezzina, C., van Dijk, S., De Man, F., Stienen, G., Odeberg, J., Pontén, F., Linke, W. and van der Velden, J. 2017, 'The sydney heart bank: improving translational research while eliminating or reducing the use of animal models of human heart disease.', *Biophysical Reviews* **9**(4).

200 Dowell, J. A., Wright, L. J., Armstrong, E. A. and Denu, J. M. 2021, 'Benchmarking quantitative performance in label-free proteomics', *ACS omega* **6**(4), 2494–2504.

201 Drew, J., Morgan, W., Galindo, S., Kleinschmit, A. J., McWilliams, M., Pauley, M., Triplett, E. W., Williams, J., Murdoch, B. and Rosenwald, A. 2023, Revisiting barriers to implementation of bioinformatics into life sciences education, *in* 'Frontiers in Education', Vol. 8.

202 Drăghici, S., Khatri, P., Martins, R. P., Ostermeier, G. C. and Krawetz, S. A. 2003, 'Global functional profiling of gene expression', *Genomics* **81**(2), 98–104.

203 Dugan, B., Conway, J. and Duggal, N. A. 2023, 'Inflammaging as a target for healthy ageing', *Age and Ageing* **52**(2), afac328.

204 Duggan, D. J., Bittner, M., Chen, Y., Meltzer, P. and Trent, J. M. 1999, 'Expression profiling using cdna microarrays', *Nature genetics* **21**(1), 10–14.

205 Dupree, E. J., Jayathirtha, M., Yorkey, H., Mihasan, M., Petre, B. A. and Darie, C. C. 2020, 'A critical review of bottom-up proteomics: the good, the bad, and the future of this field', *Proteomes* **8**(3), 14.

206 Durbin, B., Hardin, J., Hawkins, D. and Rocke, D. 2002, 'A variance-stabilizing transformation for gene-expression microarray data', *Bioinformatics* **18**.

207 Dutta, D., Calvani, R., Bernabei, R., Leeuwenburgh, C. and Marzetti, E. 2012, 'Contribution of impaired mitochondrial autophagy to cardiac aging', *Circulation Research* **110**(8).

208 D'Souza, K., Paramel, G. V. and Kienesberger, P. C. 2018, 'Lysophosphatidic acid signaling in obesity and insulin resistance', *Nutrients* **10**(4), 399.

209 E., S. E. 1978, 'Structural basis for some permeability properties of the air–blood barrier', *Federation proceedings* **37**(11).

210 Edfors, F., Danielsson, F., Hallstrom, B. M., Kall, L., Lundberg, E., Ponten, F., Forsstrom, B. and Uhlen, M. 2016, 'Gene-specific correlation of rna and protein levels in human cells and tissues', *Molecular systems biology* **12**(10).

211 Eekhout, I., de Boer, R. M., Twisk, J. W., de Vet, H. C. and Heymans, M. W. 2012, 'Missing data: a systematic review of how they are reported and handled', *Epidemiology* **23**(5), 729–732.

212 Eelen, G., de Zeeuw, P., Treps, L., Harjes, U., Wong, B. W. and Carmeliet, P. 2018, 'Endothelial cell metabolism', *Physiological Reviews* **98**(1).

213 Eisner, D. A., Caldwell, J. L., Trafford, A. W. and Hutchings, D. C. 2020, 'The control of diastolic calcium in the heart: basic mechanisms and functional implications', *Circulation research* **126**(3), 395–412.

214 Eisner, D., Caldwell, J. and Trafford, A. 2013, 'Sarcoplasmic reticulum ca-atpase and heart failure 20 years later', *Circulation research* **113**(8), 958–961.

215 El Emam, K., Rodgers, S. and Malin, B. 2015, 'Anonymising and sharing individual patient data', *bmj* **350**.

216 El-Samad, H. 2021, 'Biological feedback control—respect the loops', *Cell Systems* **12**(6), 477–487.

217 Elo, L. L., Filén, S., Lahesmaa, R. and Aittokallio, T. 2008, 'Reproducibility-optimized test statistic for ranking genes in microarray studies', *IEEE/ACM transactions on computational biology and bioinformatics* **5**(3), 423–431.

218 Elo, L. L., Hiissa, J., Tuimala, J., Kallio, A., Korpelainen, E. and Aittokallio, T. 2009, 'Optimized detection of differential expression in global profiling experiments: case studies in clinical transcriptomic and quantitative proteomic datasets', *Briefings in bioinformatics* **10**(5), 547–555.

219 Emmert-Streib, F. 2022, 'Severe testing with high-dimensional omics data for enhancing biomedical scientific discovery', *npj Systems Biology and Applications* **8**(1), 40.

220 Emwas, A.-H., Roy, R., McKay, R. T., Tenori, L., Saccenti, E., Gowda, G. N., Raftery, D., Alahmari, F., Jaremko, L., Jaremko, M. et al. 2019, 'Nmr spectroscopy for metabolomics research', *Metabolites* **9**(7), 123.

221 England, J. and Loughna, S. 2012, 'Heavy and light roles: myosin in the morphogenesis of the heart', *Cellular and Molecular Life Sciences* **70**(7).

222 Espenshade, P. and Hughes 2007, 'Regulation of sterol synthesis in eukaryotes', *Annual Review of Genetics* **41**(1).

223 Essandoh, K., Teuber, J. P. and Brody, M. J. 2024, 'Regulation of cardiomyocyte intracellular trafficking and signal transduction by protein palmitoylation', *Biochemical Society Transactions* p. BST20221296.

224 Eum, J. Y., Lee, J. C., Yi, S. S., Kim, I. Y., Seong, J. K. and Moon, M. H. 2020, 'Aging-related lipidomic changes in mouse serum, kidney, and heart by nanoflow ultrahigh-performance liquid chromatography-tandem mass spectrometry', *Journal of Chromatography A* **1618**, 460849.

225 Evangelou, K., Vasileiou, P. V., Papaspyropoulos, A., Hazapis, O., Petty, R., Demaria, M. and Gorgoulis, V. G. 2023, 'Cellular senescence and cardiovascular diseases: moving to the "heart" of the problem', *Physiological reviews* **103**(1), 609–647.

226 Fahy, E., Subramaniam, S., Brown, H., Glass, C., Merrill, A., Murphy, R., Raetz, C., Russell, D., Seyama, Y., Shaw, W., Shimizu, T., Spener, F., van Meer, G., VanNieuwenhze, M., White, S., Witztum, J. and Dennis, E. 2005, 'A comprehensive classification system for lipids', *Journal of Lipid Research* **46**(5).

227 Fahy, E., Subramaniam, S., Murphy, R., Nishijima, M., Raetz, C., Shimizu, T., Spener, F., van Meer, G., Wakelam, M. and Dennis, E. 2009, 'Update of the lipid maps comprehensive classification system for lipids', *Journal of Lipid Research* **50**.

228 Fahy, E., Sud, M., Cotter, D. and Subramaniam, S. 2007, 'Lipid maps online tools for lipid research', *Nucleic Acids Research* **35**.

229 Fang, S., Dai, X., Shi, X., Xiao, L., Ye, Y. and Liao, L. 2023, 'A pilot study investigating early postmortem interval of rats based on ambient temperature and postmortem interval-related metabolites in blood', *Forensic Science, Medicine and Pathology* pp. 1–9.

230 Fanjul, V., Jorge, I., Camafeita, E., Macías, Á., González-Gómez, C., Barettino, A., Dorado, B., Andrés-Manzano, M. J., Rivera-Torres, J., Vázquez, J. et al. 2020, 'Identification of common cardiometabolic alterations and deregulated pathways in mouse and pig models of aging', *Aging Cell* **19**(9), e13203.

231 Fares, E. and Howlett, S. E. 2010, 'Effect of age on cardiac excitation-contraction coupling', *Clinical and Experimental Pharmacology and Physiology* **37**(1).

232 Fenstermaker, T. K., Petruk, S., Kovermann, S. K., Brock, H. W. and Mazo, A. 2023, 'Rna polymerase ii associates with active genes during dna replication', *Nature* **620**(7973), 426–433.

233 Fernandez-Marcos, P. J. and Nóbrega-Pereira, S. 2016, 'Nadph: new oxygen for the ros theory of aging', *Oncotarget* **7**(32), 50814.

234 Fernandez-Sanz, C., Ruiz-Meana, M., Miro-Casas, E., Nunez, E., Castellano, J., Loureiro, M., Barba, I., Poncelas, M., Rodriguez-Sinovas, A., Vazquez, J. et al. 2014, 'Defective sarcoplasmic reticulum–mitochondria calcium exchange in aged mouse myocardium', *Cell Death & Disease* **5**(12), e1573–e1573.

235 Ferrucci, L. and Fabbri, E. 2018, 'Inflammageing: chronic inflammation in ageing, cardiovascular disease, and frailty', *Nature Reviews Cardiology* **15**(9).

236 Fiehn, O. 2016, 'Metabolomics by gas chromatography–mass spectrometry: Combined targeted and untargeted profiling', *Current protocols in molecular biology* **114**(1), 30–4.

237 Filadi, R., Theurey, P. and Pizzo, P. 2017, 'The endoplasmic reticulum-mitochondria coupling in health and disease: Molecules, functions and significance', *Cell Calcium* **62**.

238 Fink, E. L. 2009, 'The faqs on data transformation', *Communication Monographs* **76**(4), 379–397.

239 Finoulst, I., Pinkse, M., Van Dongen, W., Verhaert, P. et al. 2011, 'Sample preparation techniques for the untargeted lc-ms-based discovery of peptides in complex biological matrices', *BioMed Research International* **2011**.

240 Fleming, S., Thompson, M., Stevens, R., Heneghan, C., Plüddemann, A., Maconochie, I., Tarassenko, L. and Mant, D. 2011, 'Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: a systematic review of observational studies', *The Lancet* **377**(9770), 1011–1018.

241 Forshed, J. 2017, 'Experimental design in clinical 'omics biomarker discovery', *Journal of proteome research* **16**(11), 3954–3960.

242 Fortunato, S. and Newman, M. E. 2022, '20 years of network community detection', *Nature Physics* **18**(8), 848–850.

243 Foster, C. R., Zha, Q., Daniel, L. L., Singh, M. and Singh, K. 2012, 'Lack of atm induces structural and functional changes in the heart: Role in β-adrenergic receptor-stimulated apoptosis', *Experimental Physiology* **97**(4), 506.

244 Frampton, M. and Houlston, R. 2012, 'Generation of artificial fastq files to evaluate the performance of next-generation sequencing pipelines', *PLoS One* **7**(11), e49110.

245 Frangogiannis, N. G. 2019, 'The extracellular matrix in ischemic and nonischemic heart failure', *Circulation Research* **125**(1).

246 Fregoso, F. E. 2024, 'Biochemical and structural investigation of arp2/3 complex regulation by mammalian coronin-7', *Biophysical Journal* **123**(3), 536a–537a.

247 Fu, W. and Hall, M. N. 2020, 'Regulation of mtorc2 signaling', *Genes* **11**(9), 1045.

248 Fugio, L., Coeli-Lacchini, F. and Leopoldino, A. 2020, 'Sphingolipids and mitochondrial dynamic', *Cells* **9**(3).

249 Fukasawa, Y., Ermini, L., Wang, H., Carty, K. and Cheung, M.-S. 2020, 'Longqc: a quality control tool for third generation sequencing long read data', *G3: Genes, Genomes, Genetics* **10**(4), 1193–1196.

250 Fuller, T. F., Ghazalpour, A., Aten, J. E., Drake, T. A., Lusis, A. J. and Horvath, S. 2007, 'Weighted gene coexpression network analysis strategies applied to mouse weight', *Mammalian Genome* **18**, 463–472.

251 Gabr, R. E., El-Sharkawy, A.-M. M., Schär, M., Panjrath, G. S., Gerstenblith, G., Weiss, R. G. and Bottomley, P. A. 2018, 'Cardiac work is related to creatine kinase energy supply in human heart failure: a cardiovascular magnetic resonance spectroscopy study', *Journal of Cardiovascular Magnetic Resonance* **20**(1), 81.

252 Gaggini, M., Ndreu, R., Michelucci, E., Rocchiccioli, S. and Vassalle, C. 2022, 'Ceramides as mediators of oxidative stress and inflammation in cardiometabolic disease', *International Journal of Molecular Sciences* **23**(5), 2719.

253 Gahete, M. D., del Rio-Moreno, M., Camargo, A., Alcala-Diaz, J. F., Alors-Perez, E., Delgado-Lista, J., Reyes, O., Ventura, S., Perez-Martínez, P., Castaño, J. P. et al. 2018, 'Changes in splicing machinery components influence, precede, and early predict the development of type 2 diabetes: from the cordioprev study', *EBioMedicine* **37**, 356–365.

254 Gallego-Paüls, M., Hernández-Ferrer, C., Bustamante, M., Basagaña, X., Barrera-Gómez, J., Lau, C.-H. E., Siskos, A. P., Vives-Usano, M., Ruiz-Arenas, C., Wright, J. et al. 2021, 'Variability of multi-omics profiles in a population-based child cohort', *BMC medicine* **19**, 1–16.

255 Galmes, R., Houcine, A., Vliet, A., Agostinis, P., Jackson, C. and Giordano, F. 2016, 'Orp5/orp8 localize to endoplasmic reticulum–mitochondria contacts and are involved in mitochondrial function', *EMBO reports* **17**(6).

256 Gao, D., Zuo, Z., Tian, J., Ali, Q., Lin, Y., Lei, H. and Sun, Z. 2016, 'Activation of sirt1 attenuates klotho deficiency–induced arterial stiffness and hypertension by enhancing amp-activated protein kinase activity', *Hypertension* **68**(5).

257 Gao, J., Collyer, J., Wang, M., Sun, F. and Xu, F. 2020, 'Genetic dissection of hypertrophic cardiomyopathy with myocardial rna-seq', *International journal of molecular sciences* **21**(9), 3040.

258 Gao, X., Jakovljevic, D. and Beard, D. 2019, 'Cardiac metabolic limitations contribute to diminished performance of the heart in aging', *Biophysical Journal* **117**(12).

259 Garbern, J. C. and Lee, R. T. 2021, 'Mitochondria and metabolic transitions in cardiomyocytes: lessons from development for stem cell-derived cardiomyocytes', *Stem Cell Research & Therapy* **12**(1).

260 Gates, K. M., Henry, T., Steinley, D. and Fair, D. A. 2016, 'A monte carlo evaluation of weighted community detection algorithms', *Frontiers in neuroinformatics* **10**, 45.

261 Gatsiou, A. and Stellos, K. 2023, 'Rna modifications in cardiovascular health and disease', *Nature Reviews Cardiology* **20**(5), 325–346.

262 Gehlenborg, N. and Wong, B. 2012, 'Heat maps', *Nature Methods* **9**(3), 213.

263 Gentry, E. C., Collins, S. L., Panitchpakdi, M., Belda-Ferre, P., Stewart, A. K., Carrillo Terrazas, M., Lu, H.-h., Zuffa, S., Yan, T., Avila-Pacheco, J. et al. 2024, 'Reverse metabolomics for the discovery of chemical structures from humans', *Nature* **626**(7998), 419–426.

264 Gergs, U., Mangold, W., Langguth, F., Hatzfeld, M., Hauptmann, S., Bushnaq, H., Simm, A., Silber, R.-E. and Neumann, J. 2019, 'Alterations of protein expression of phospholamban, zasp and plakoglobin in human atria in subgroups of seniors', *Scientific Reports* **9**(1), 5610.

265 Ghafouri-Fard, S., Safarzadeh, A., Taheri, M. and Jamali, E. 2023, 'Identification of diagnostic biomarkers via weighted correlation network analysis in colorectal cancer using a system biology approach', *Scientific Reports* **13**(13637).

266 Ghasemi, F., Cao, L., Jegou, A. and Romet-Lemonne, G. 2023, 'The arp2/3 complex contributes to the regeneration of branched actin filament networks', *Biophysical Journal* **122**(3), 166a–167a.

267 Ghosh, A. and Shcherbik, N. 2020, 'Effects of oxidative stress on protein translation: Implications for cardiovascular diseases', *International Journal of Molecular Sciences* **21**(8).

268 Ghosh Roy, G., Geard, N., Verspoor, K. and He, S. 2020, 'Polobag: Polynomial lasso bagging for signed gene regulatory network inference from expression data', *Bioinformatics* **36**(21), 5187–5193.

269 Ghosh, S., Guglielmi, G., Orfanidis, I., Spill, F., Hickey, A., Hanssen, E. and Rajagopal, V. 2022, 'Effects of altered cellular ultrastructure on energy metabolism in diabetic cardiomyopathy: an in silico study', *Philosophical Transactions of the Royal Society B* **377**(1864), 20210323.

270 Giansanti, P., Tsiatsiani, L., Low, T. Y. and Heck, A. J. 2016, 'Six alternative proteases for mass spectrometry–based proteomics beyond trypsin', *Nature protocols* **11**(5), 993–1006.

271 Gibbs, R. A. 2020, 'The human genome project changed everything', *Nature Reviews Genetics* **21**(10), 575–576.

272 Giesbertz, P., Brandl, B., Volkert, D., Hauner, H. and Skurk, T. 2023, 'Age-related metabolite profiles and their relation to clinical outcomes in young adults, middle-aged individuals, and older people', *The FASEB Journal* **37**(6), e22968.

273 Gika, H. G., Zisi, C., Theodoridis, G. and Wilson, I. D. 2016, 'Protocol for quality control in metabolic profiling of biological fluids by u (h) plc-ms', *Journal of chromatography B* **1008**, 15–25.

274 Gilbertson, J. A., Theis, J. D., Vrana, J. A., Lachmann, H., Wechalekar, A., Whelan, C., Hawkins, P. N., Dogan, A. and Gillmore, J. D. 2015, 'A comparison of immunohistochemistry and mass spectrometry for determining the amyloid fibril protein from formalin-fixed biopsy tissue', *Journal of clinical pathology* **68**(4), 314–317.

275 Giordano, F. J., Gerber, H.-P., Williams, S.-P., VanBruggen, N., Bunting, S., Ruiz-Lozano, P., Gu, Y., Nath, A. K., Huang, Y., Hickey, R., Dalton, N., Peterson, K. L., Ross, J., Chien, K. R. and Ferrara, N. 2001, 'A cardiac myocyte vascular endothelial growth factor paracrine pathway is required to maintain cardiac function', *Proceedings of the National Academy of Sciences* **98**(10).

276 Giorgi, C., Marchi, S., Simoes, I. C., Ren, Z., Morciano, G., Perrone, M., Patalas-Krawczyk, P., Borchard, S., Jedrak, P., Pierzynowska, K. et al. 2018, 'Mitochondria and reactive oxygen species in aging and age-related diseases', *International review of cell and molecular biology* **340**, 209–344.

277 Girvan, M. and Newman, M. E. 2002, 'Community structure in social and biological networks', *Proceedings of the national academy of sciences* **99**(12), 7821–7826.

278 Gleiss, A., Dakna, M., Mischak, H. and Heinze, G. 2015, 'Two-group comparisons of zero-inflated intensity values: the choice of test statistic matters', *Bioinformatics* **31**(14), 2310–2317.

279 Gogiraju, R., Bochenek, M. L. and Schäfer, K. 2019, 'Angiogenic endothelial cell signaling in cardiac hypertrophy and heart failure', *Frontiers in Cardiovascular Medicine* **6**.

280 Goh, W. W. B., Wang, W. and Wong, L. 2017, 'Why batch effects matter in omics data, and how to avoid them', *Trends in biotechnology* **35**(6), 498–507.

281 Goldberg, I. J. 2018, '2017 george lyman duff memorial lecture: fat in the blood, fat in the artery, fat in the heart: triglyceride in physiology and disease', *Arteriosclerosis, thrombosis, and vascular biology* **38**(4), 700–706.

282 Goldberg, I., Trent, C. and Schulze, P. 2012, 'Lipid metabolism and toxicity in the heart', *Cell Metabolism* **15**(6).

283 Goldspink, D. F., Burniston, J. G. and Tan, L.-B. 2003, 'Cardiomyocyte death and the ageing and failing heart', *Experimental physiology* **88**(3), 447–458.

284 Goldtzvik, Y., Sen, N., Lam, S. D. and Orengo, C. 2023, 'Protein diversification through post-translational modifications, alternative splicing, and gene duplication', *Current Opinion in Structural Biology* **81**, 102640.

285 Gomer, B. 2019, 'Mcar, mar, and mnar values in the same dataset: a realistic evaluation of methods for handling missing data', *Multivariate Behavioral Research* **54**(1), 153–153.

286 Gomer, B. and Yuan, K.-H. 2023, 'A realistic evaluation of methods for handling missing data when there is a mixture of mcar, mar, and mnar mechanisms in the same dataset', *Multivariate Behavioral Research* **58**(5), 988–1013.

287 Gómez, L. A., Heath, S.-H. D. and Hagen, T. M. 2012, 'Acetyl-l-carnitine supplementation reverses the age-related decline in carnitine palmitoyltransferase 1 (cpt1) activity in interfibrillar mitochondria without changing the l-carnitine content in the rat heart', *Mechanisms of ageing and development* **133**(2-3), 99–106.

288 Gopal, K., Saleme, B., Al Batran, R., Aburasayn, H., Eshreif, A., Ho, K. L., Ma, W. K., Almutairi, M., Eaton, F., Gandhi, M., Park, E. A., Sutendra, G. and Ussher, J. R. 2017, 'Foxo1 regulates myocardial glucose oxidation rates via transcriptional control of pyruvate dehydrogenase kinase 4 expression', *American journal of physiology. Heart and circulatory physiology* **313**(3).

289 Gosak, M., Markovič, R., Dolenšek, J., Rupnik, M. S., Marhl, M., Stožer, A. and Perc, M. 2018, 'Network science of biological systems at different scales: A review', *Physics of life reviews* **24**, 118–135.

290 Gotti, C., Roux-Dalvai, F., Joly-Beauparlant, C., Mangnier, L., Leclercq, M. and Droit, A. 2021, 'Extensive and accurate benchmarking of dia acquisition methods and software tools using a complex proteomic standard', *Journal of Proteome Research* **20**(10), 4801–4814.

291 Grant, A. O. 2009, 'Cardiac ion channels', *Circulation: Arrhythmia and Electrophysiology* **2**(2), 185–194.

292 Graw, S., Chappell, K., Washam, C. L., Gies, A., Bird, J., Robeson, M. S. and Byrum, S. D. 2021, 'Multi-omics data integration considerations and study design for biological systems and disease', *Molecular omics* **17**(2), 170–185.

293 Greenacre, M., Groenen, P. J., Hastie, T., d'Enza, A. I., Markos, A. and Tuzhilina, E. 2022, 'Principal component analysis', *Nature Reviews Methods Primers* **2**(1), 100.

294 Greenberg, B., Butler, J., Felker, G. M., Ponikowski, P., Voors, A. A., Desai, A. S., Barnard, D., Bouchard, A., Jaski, B., Lyon, A. R. et al. 2016, 'Calcium upregulation by percutaneous administration of gene therapy in patients with cardiac disease (cupid 2): a randomised, multinational, double-blind, placebo-controlled, phase 2b trial', *The Lancet* **387**(10024), 1178–1186.

295 Griffiths, J. 2008, 'A brief history of mass spectrometry', *Analytical Chemistry* **80**(15). **URL:** *doi:10.1021/ac8013065*

296 Grivennikova, V. G., Kareyeva, A. V. and Vinogradov, A. D. 2010, 'What are the sources of hydrogen peroxide production by heart mitochondria?', *Biochimica et Biophysica Acta (BBA)-Bioenergetics* **1797**(6-7), 939–944.

297 Grünberger, F., Ferreira-Cerca, S. and Grohmann, D. 2022, 'Nanopore sequencing of rna and cdna molecules in escherichia coli', *Rna* **28**(3), 400–417.

298 Grünert, S. C., Hannibal, L. and Spiekerkoetter, U. 2021, 'The phenotypic and genetic spectrum of glycogen storage disease type vi', *Genes* **12**(8), 1205.

299 Gu, Y., Zhou, Y., Ju, S., Liu, X., Zhang, Z., Guo, J., Gao, J., Zang, J., Sun, H., Chen, Q. et al. 2022, 'Multi-omics profiling visualizes dynamics of cardiac development and functions', *Cell Reports* **41**(13).

300 Gude, N. A., Broughton, K. M., Firouzi, F. and Sussman, M. A. 2018, 'Cardiac ageing: extrinsic and intrinsic factors in cellular renewal and senescence', *Nature Reviews Cardiology* **15**(9).

301 Guimaraes Ferreira, L. 2014, 'Role of the phosphocreatine system on energetic homeostasis in skeletal and cardiac muscles.', *Einstein (Sao Paulo)* **12**(1).

302 Guo, Y. and Pu, W. T. 2020, 'Cardiomyocyte maturation', *Circulation Research* **126**(8).

303 Habib, S., Al-Khlaiwi, T., Almushawah, A., Alsomali, A. and Habib, S. 2023, 'Homocysteine as a predictor and prognostic marker of atherosclerotic cardiovascular disease: a systematic review and meta-analysis.', *European Review for Medical & Pharmacological Sciences* **27**(18).

304 Hägg, S. and Jylhävä, J. 2021, 'Sex differences in biological aging with a focus on human studies', *Elife* **10**, e63425.

305 Hahn, V. S., Knutsdottir, H., Luo, X., Bedi, K., Margulies, K. B., Haldar, S. M., Stolina, M., Yin, J., Khakoo, A. Y., Vaishnav, J. et al. 2021, 'Myocardial gene expression signatures in human heart failure with preserved ejection fraction', *Circulation* **143**(2), 120–134.

306 Haines, R. J., Pendleton, L. C. and Eichler, D. C. 2011, 'Argininosuccinate synthase: at the center of arginine metabolism', *International journal of biochemistry and molecular biology* **2**(1), 8.

307 Håkansson, K. E., Goossens, E. A., Trompet, S., van Ingen, E., de Vries, M. R., van der Kwast, R. V., Ripa, R. S., Kastrup, J., Hohensinner, P. J., Kaun, C. et al. 2019, 'Genetic associations and regulation of expression indicate an independent role for 14q32 snornas in human cardiovascular disease', *Cardiovascular research* **115**(10), 1519–1532.

308 Hallstrom, A. P. 2010, 'A modified wilcoxon test for non-negative distributions with a clump of zeros', *Statistics in Medicine* **29**(3), 391–400.

309 Hamilton, S. and Terentyev, D. 2019, 'Altered intracellular calcium homeostasis and arrhythmogenesis in the aged heart', *International journal of molecular sciences* **20**(10), 2386.

310 Han, J.-D. J., Bertin, N., Hao, T., Goldberg, D. S., Berriz, G. F., Zhang, L. V., Dupuy, D., Walhout, A. J., Cusick, M. E., Roth, F. P. et al. 2004, 'Evidence for dynamically organized modularity in the yeast protein–protein interaction network', *Nature* **430**(6995), 88–93.

311 Han, J. L. and Entcheva, E. 2023, 'Gene modulation with crispr-based tools in human ipsc-cardiomyocytes', *Stem Cell Reviews and Reports* **19**(4), 886–905.

312 Han, M. and Zhou, B. 2022, 'Role of cardiac fibroblasts in cardiac injury and repair', *Current Cardiology Reports* **24**(3).

313 Hansen, K. D., Brenner, S. E. and Dudoit, S. 2010, 'Biases in illumina transcriptome sequencing caused by random hexamer priming', *Nucleic acids research* **38**(12), e131–e131.

314 Harman, D. 1972, 'The biologic clock: The mitochondria?', *Journal of the American Geriatrics Society* **20**(4).

315 Harney, D. J., Hutchison, A. T., Su, Z., Hatchwell, L., Heilbronn, L. K., Hocking, S., James, D. E. and Larance, M. 2019, 'Small-protein enrichment assay enables the rapid, unbiased analysis of over 100 low abundance factors from human plasma*[s]', *Molecular & Cellular Proteomics* **18**(9), 1899–1915.

316 Hart, A. 2001, 'Mann-whitney test is not just a test of medians: differences in spread can be important', *Bmj* **323**(7309), 391–393.

317 Hasanzad, M., Sarhangi, N., Ehsani Chimeh, S., Ayati, N., Afzali, M., Khatami, F., Nikfar, S. and Aghaei Meybodi, H. R. 2022, 'Precision medicine journey through omics approach', *Journal of Diabetes & Metabolic Disorders* **21**(1), 881–888.

318 Hasenfuss, G., Reinecke, H., Studer, R., Meyer, M., Pieske, B., Holtz, J., Holubarsch, C., Posival, H., Just, H. and Drexler, H. 1994, 'Relation between myocardial function and expression of sarcoplasmic reticulum ca (2+)-atpase in failing and nonfailing human myocardium.', *Circulation research* **75**(3), 434–442.

319 Hassinen, I. E., Nuutinen, E. M., Ito, K., Nioka, S., Lazzarino, G., Giardina, B. and Chance, B. 1991, 'Mechanism of the effect of exogenous fructose 1, 6-bisphosphate on myocardial energy metabolism.', *Circulation* **83**(2), 584–593.

320 Hawinkel, S., Rayner, J., Bijnens, L. and Thas, O. 2020, 'Sequence count data are poorly fit by the negative binomial distribution', *PloS one* **15**(4), e0224909.

321 Hazenberg, B. P., van Rijswijk, M. H., Piers, D. A., Lub-de Hooge, M. N., Vellenga, E., Haagsma, E. B., Hawkins, P. N. and Jager, P. L. 2006, 'Diagnostic performance of 123i-labeled serum amyloid p component scintigraphy in patients with amyloidosis', *The American journal of medicine* **119**(4), 355–e15.

322 He, H., Huang, J., Wu, S., Jiang, S., Liang, L., Liu, Y., Liu, W., Xie, L., Tao, Y., Jiang, Y. and Cong, L. 2021, 'The roles of gtpase-activating proteins in regulated cell death and tumor immunity', *Journal of Hematology & Oncology* **14**(1).

323 He, X., Du, T., Long, T., Liao, X., Dong, Y. and Huang, Z.-P. 2022, 'Signaling cascades in the failing heart and emerging therapeutic strategies', *Signal transduction and targeted therapy* **7**(1), 134.

324 He, Y., Huang, W., Zhang, C., Chen, L., Xu, R., Li, N., Wang, F., Han, L., Yang, M. and Zhang, D. 2020, 'Energy metabolism disorders and potential therapeutic drugs in heart failure', *Acta Pharmaceutica Sinica B* **11**(5).

325 Head, M. L., Holman, L., Lanfear, R., Kahn, A. T. and Jennions, M. D. 2015, 'The extent and consequences of p-hacking in science', *PLoS biology* **13**(3), e1002106.

326 Hector, A., Von Felten, S. and Schmid, B. 2010, 'Analysis of variance with unbalanced data: an update for ecology & evolution', *Journal of animal ecology* **79**(2), 308–316.

327 Heidecker, B., Kittleson, M. M., Kasper, E. K., Wittstein, I. S., Champion, H. C., Russell, S. D., Baughman, K. L. and Hare, J. M. 2016, 'Transcriptomic analysis identifies the effect of beta-blocking agents on a molecular pathway of contraction in the heart and predicts response to therapy', *JACC: Basic to Translational Science* **1**(3), 107–121.

328 Hellström, M., Phng, L.-K., Hofmann, J. J., Wallgard, E., Coultas, L., Lindblom, P., Alva, J., Nilsson, A.-K., Karlsson, L., Gaiano, N., Yoon, K., Rossant, J., Iruela-Arispe, M. L., Kalén, M., Gerhardt, H. and Betsholtz, C. 2007, 'Dll4 signalling through notch1 regulates formation of tip cells during angiogenesis', *Nature* **445**(7129).

329 Henning, R. H. and Brundel, B. J. 2017, 'Proteostasis in cardiac health and disease', *Nature Reviews Cardiology* **14**(11), 637–653.

330 Hernán, M. A., Hernández-Díaz, S., Werler, M. M. and Mitchell, A. A. 2002, 'Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology', *American journal of epidemiology* **155**(2), 176–184.

331 Herndon, N. C. 2016, 'Research fraud and the publish or perish world of academia'.

332 Hernández-Corbacho, M., Salama, M., Canals, D., Senkal, C. and Obeid, L. 2017, 'Sphingolipids in mitochondria', *Biochimica et Biophysica Acta (BBA)* **1862**(1).

333 Hevey, D. 2018, 'Network analysis: a brief overview and tutorial', *Health Psychology and Behavioral Medicine* **6**(1), 301–328.

334 Heyer, E. E., Deveson, I. W., Wooi, D., Selinger, C. I., Lyons, R. J., Hayes, V. M., O'Toole, S. A., Ballinger, M. L., Gill, D., Thomas, D. M. et al. 2019, 'Diagnosis of fusion genes using targeted rna sequencing', *Nature communications* **10**(1), 1–12.

335 Ho, S. Y. 2009, 'Anatomy and myoarchitecture of the left ventricular wall in normal and in disease', *European Journal of Echocardiography* **10**(8).

336 Hodge, V. and Austin, J. 2004, 'A survey of outlier detection methodologies', *Artificial intelligence review* **22**, 85–126.

337 Hoes, M. F., Tromp, J., Ouwerkerk, W., Bomer, N., Oberdorf-Maass, S. U., Samani, N. J., Ng, L. L., Lang, C. C., van der Harst, P., Hillege, H. et al. 2020, 'The role of cathepsin d in the pathophysiology of heart failure and its potentially beneficial properties: a translational approach', *European journal of heart failure* **22**(11), 2102–2111.

338 Hollingsworth, K. G., Blamire, A. M., Keavney, B. and MacGowan, G. A. 2012, 'Left ventricular torsion, energetics, and diastolic function in normal human aging', *American Journal of Physiology-Heart and Circulatory Physiology* **302**(4).

339 Hollon, H. 2023, 'Preventive cardiology in pediatrics. a fellow's voice', *American Journal of Preventive Cardiology* **15**.

340 Hoover, D. R. 2002, 'Power for t-test comparisons of unbalanced cluster exposure studies', *Journal of Urban Health* **79**, 278–294.

341 Horikoshi, Y., Yan, Y., Terashvili, M., Wells, C., Horikoshi, H., Fujita, S., Bosnjak, Z. and Bai, X. 2019, 'Fatty acid-treated induced pluripotent stem cell-derived human cardiomyocytes exhibit adult cardiomyocyte-like energy metabolism phenotypes', *Cells* **8**(9).

342 Horn, M. A. and Trafford, A. W. 2016, 'Aging and the cardiac collagen matrix: Novel mediators of fibrotic remodelling', *Journal of Molecular and Cellular Cardiology* **93**.

343 Hornburg, D., Wu, S., Moqri, M., Zhou, X., Contrepois, K., Bararpour, N., Traber, G. M., Su, B., Metwally, A. A., Avina, M. et al. 2023, 'Dynamic lipidome alterations associated with human health, disease and ageing', *Nature Metabolism* **5**(9), 1578–1594.

344 Hou, T., Zhang, R., Jian, C., Ding, W., Wang, Y., Ling, S., Ma, Q., Hu, X., Cheng, H. and Wang, X. 2019, 'Ndufab1 confers cardio-protection by enhancing mitochondrial bioenergetics through coordination of respiratory complex and supercomplex assembly', *Cell research* **29**(9), 754–766.

345 Houten, S. and Wanders, R. 2010, 'A general introduction to the biochemistry of mitochondrial fatty acid β-oxidation', *Journal of Inherited Metabolic Disease* **33**(5).

346 Houten, S., Wanders, R. and Ranea-Robles, P. 2020, 'Metabolic interactions between peroxisomes and mitochondria with a special focus on acylcarnitine metabolism', *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1866**(5).

347 Hrydziuszko, O. and Viant, M. R. 2012, 'Missing values in mass spectrometry based metabolomics: an undervalued step in the data processing pipeline', *Metabolomics* **8**, 161–174.

348 Hsieh, E. J., Hoopmann, M. R., MacLean, B. and MacCoss, M. J. 2010, 'Comparison of database search strategies for high precursor mass accuracy ms/ms data', *Journal of proteome research* **9**(2), 1138–1143.

349 Hsieh, Y., Duncan, C. J. and Brisson, J.-M. 2007, 'Fused-core silica column high-performance liquid chromatography/tandem mass spectrometric determination of rimonabant in mouse plasma', *Analytical chemistry* **79**(15), 5668–5673.

350 Hu, C., Duan, Q. and Han, X. 2020, 'Strategies to improve/eliminate the limitations in shotgun lipidomics', *Proteomics* **20**(11), 1900070.

351 Hu, C. and Jia, W. 2021, 'Multi-omics profiling: the way toward precision medicine in metabolic diseases', *Journal of molecular cell biology* **13**(8), 576–593.

352 Hu, T. and Zhang, J.-L. 2018, 'Mass-spectrometry-based lipidomics', *Journal of separation science* **41**(1), 351–372.

353 Hu, W., Xin, Y., Hu, J., Sun, Y. and Zhao, Y. 2019, 'Inhibitor of dna binding in heart development and cardiovascular diseases', *Cell Communication and Signaling* **17**(1), 51.

354 Hua, Y., Robinson, T. J., Cao, Y., Shi, G.-P., Ren, J. and Nair, S. 2015, 'Cathepsin k knockout alleviates aging-induced cardiac dysfunction', *Aging cell* **14**(3), 345–351.

355 Hua, Y., Zhang, Y., Dolence, J., Shi, G.-P., Ren, J. and Nair, S. 2013, 'Cathepsin k knockout mitigates high-fat diet–induced cardiac hypertrophy and contractile dysfunction', *Diabetes* **62**(2), 498–509.

356 Huang, D. W., Sherman, B. T. and Lempicki, R. A. 2009*a*, 'Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists', *Nucleic acids research* **37**(1), 1–13.

357 Huang, D. W., Sherman, B. T. and Lempicki, R. A. 2009*b*, 'Systematic and integrative analysis of large gene lists using david bioinformatics resources', *Nature protocols* **4**(1), 44–57.

358 Huang, T., Staniak, M., Veiga Leprevost, F. d., Figueroa-Navedo, A. M., Ivanov, A. R., Nesvizhskii, A. I., Choi, M. and Vitek, O. 2023, 'Statistical detection of differentially abundant proteins in experiments with repeated measures designs and isobaric labeling', *Journal of Proteome Research* **22**(8), 2641–2659.

359 Huang, X., Feng, T., Jiang, Z., Meng, J., Kou, S., Lu, Z., Chen, W., Lin, C.-P., Zhou, B. and Zhang, H. 2019, 'Dual lineage tracing identifies intermediate mesenchymal stage for endocardial contribution to fibroblasts, coronary mural cells, and adipocytes', *Journal of Biological Chemistry* **294**(22).

360 Huang, Y., Liu, X. Q., Rall, S. C., Taylor, J. M., von Eckardstein, A., Assmann, G. and Mahley, R. W. 1998, 'Overexpression and accumulation of apolipoprotein e as a cause of hypertriglyceridemia', *Journal of Biological Chemistry* **273**(41), 26388–26393.

361 Huang, Z., Lane, A. N., Fan, T. W., Higashi, R. M., Weiss, H. L., Yin, X. and Wang, C. 2020, 'Differential abundance analysis with bayes shrinkage estimation of variance (dasev) for zero-inflated proteomic and metabolomic data', *Scientific Reports* **10**(1), 876.

362 Huang, Z. and Wang, C. 2022, 'A review on differential abundance analysis methods for mass spectrometry-based metabolomic data', *Metabolites* **12**(4), 305.

363 Huber, W., Von Heydebreck, A., Sültmann, H., Poustka, A. and Vingron, M. 2002, 'Variance stabilization applied to microarray data calibration and to the quantification of differential expression', *Bioinformatics* **18**(suppl_1), S96–S104.

364 Hudsmith, L. and Thorne, S. 2007, 'Transition of care from paediatric to adult services in cardiology', *Archives of Disease in Childhood* **92**(10), 927–930.

365 Huebbe, P., Dose, J., Schloesser, A., Campbell, G., Glüer, C.-C., Gupta, Y., Ibrahim, S., Minihane, A.-M., Baines, J. F., Nebel, A. et al. 2015, 'Apolipoprotein e (apoe) genotype regulates body weight and fatty acid utilization—studies in gene-targeted replacement mice', *Molecular nutrition & food research* **59**(2), 334–343.

366 Hullin-Matsuda, F., Luquain-Costaz, C., Bouvier, J. and Delton-Vandenbroucke, I. 2009, 'Bis (monoacylglycero) phosphate, a peculiar phospholipid to control the fate of cholesterol: Implications in pathology', *Prostaglandins, Leukotrienes and Essential Fatty Acids* **81**(5-6), 313–324.

367 Hung, J.-H., Yang, T.-H., Hu, Z., Weng, Z. and DeLisi, C. 2012, 'Gene set enrichment analysis: performance evaluation and usage guidelines', *Briefings in bioinformatics* **13**(3), 281–291.

368 Hunt, T. K., Aslam, R. S., Beckert, S., Wagner, S., Ghani, Q. P., Hussain, M. Z., Roy, S. and Sen, C. K. 2007, 'Aerobically derived lactate stimulates revascularization and tissue repair via redox mechanisms', *Antioxidants & Redox Signaling* **9**(8).

369 Ijaz, N., Buta, B., Xue, Q.-L., Mohess, D. T., Bushan, A., Tran, H., Batchelor, W., DeFilippi, C. R., Walston, J. D., Bandeen-Roche, K. et al. 2022, 'Interventions for frailty among older adults with cardiovascular disease: Jacc state-of-the-art review', *Journal of the American College of Cardiology* **79**(5), 482–503.

370 Ille, A. M., Lamont, H. and Mathews, M. B. 2022, 'The central dogma revisited: Insights from protein synthesis, crispr, and beyond', *Wiley Interdisciplinary Reviews: RNA* **13**(5), e1718.

371 Ilnytska, O., Jeziorek, M., Lai, K., Altan-Bonnet, N., Dobrowolski, R. and Storch, J. 2021, 'Lysobisphosphatidic acid (lbpa) enrichment promotes cholesterol egress via exosomes in niemann pick type c1 deficient cells', *Biochimica et biophysica acta. Molecular and cell biology of lipids* **1866**(6), 158916.

372 Imai, S.-i. and Guarente, L. 2014, 'Nad+ and sirtuins in aging and disease', *Trends in cell biology* **24**(8), 464–471.

373 Imboden, M. and Probst-Hensch, N. M. 2013, 'Biobanking across the phenome-at the center of chronic disease research', *BMC public health* **13**, 1–13.

374 Ioannidis, J. P. and Khoury, M. J. 2011, 'Improving validation practices in "omics" research', *Science* **334**(6060), 1230–1232.

375 Iyer, D., Gambardella, L., Bernard, W. G., Serrano, F., Mascetti, V. L., Pedersen, R. A., Talasila, A. and Sinha, S. 2015, 'Robust derivation of epicardium and its differentiated smooth muscle cell progeny from human pluripotent stem cells', *Development* .

376 J., R., B., F., M., R.-G. and A., G.-B. 2022, 'Ageing, age-related cardiovascular risk and the beneficial role of natural components intake', *nternational Journal of Molecular Sciences* **23**(1).

377 Jahagirdar, S. and Saccenti, E. 2020, 'On the use of correlation and mi as a measure of metabolite—metabolite association for network differential connectivity analysis', *Metabolites* **10**(4), 171.

378 Jahagirdar, S., Suarez-Diez, M. and Saccenti, E. 2019, 'Simulation and reconstruction of metabolite–metabolite association networks using a metabolic dynamic model and correlation based algorithms', *Journal of proteome research* **18**(3), 1099–1113.

379 Jain, N., Thatte, J., Braciale, T., Ley, K., O'Connell, M. and Lee, J. K. 2003, 'Local-pooled-error test for identifying differentially expressed genes with a small number of replicated microarrays', *Bioinformatics* **19**(15), 1945–1951.

380 Jarrell, Z., Smith, M., Hu, X., Orr, M., Liu, K., Quyyumi, A., Jones, D. and Go, Y. 2020, 'Plasma acylcarnitine levels increase with healthy aging', *Aging* **12**(13).

381 Jeanmougin, M., De Reynies, A., Marisa, L., Paccard, C., Nuel, G. and Guedj, M. 2010, 'Should we abandon the t-test in the analysis of gene expression microarray data: a comparison of variance modeling strategies', *PloS one* **5**(9), e12336.

382 Jeanneret, F. and Gazut, S. 2021, 'Towards an integrative multi-omics workflow', *bioRxiv* pp. 2021–07.

383 Jenkinson, C. P., Göring, H. H., Arya, R., Blangero, J., Duggirala, R. and DeFronzo, R. A. 2016, 'Transcriptomics in type 2 diabetes: Bridging the gap between genotype and phenotype', *Genomics Data* **8**, 25–36.

384 Jensen, B., Christoffels, V. M. and Moorman, A. F. 2020, 'An appreciation of anatomy in the molecular world', *Journal of cardiovascular development and disease* **7**(4), 44.

385 Jensen, P., Fretts, A., Yu, C., Hoofnagle, A., Umans, J., Howard, B., Sitlani, C., Siscovick, D., King, I., Sotoodehnia, N., McKnight, B. and Lemaitre, R. 2019, 'Circulating sphingolipids, fasting glucose, and impaired fasting glucose: The strong heart family study', *EBioMedicine* **41**.

386 Jenul, A., Schrunner, S., Liland, K. H., Indahl, U. G., Futsæther, C. M. and Tomic, O. 2021, 'Rent—repeated elastic net technique for feature selection', *IEEE Access* **9**, 152333–152346.

387 Jeon, J., Han, E. Y. and Jung, I. 2023, 'Mopa: An integrative multi-omics pathway analysis method for measuring omics activity', *Plos one* **18**(3), e0278272.

388 Jha, P., McDevitt, M. T., Halilbasic, E., Williams, E. G., Quiros, P. M., Gariani, K., Sleiman, M. B., Gupta, R., Ulbrich, A., Jochem, A. et al. 2018, 'Genetic regulation of plasma lipid species and their association with metabolic phenotypes', *Cell systems* **6**(6), 709–721.

389 Jiang, R., Sun, T., Song, D. and Li, J. J. 2022, 'Statistics or biology: the zero-inflation controversy about scrna-seq data', *Genome biology* **23**(1), 31.

390 Jobard, E., Trédan, O., Postoly, D., André, F., Martin, A.-L., Elena-Herrmann, B. and Boyault, S. 2016, 'A systematic evaluation of blood serum and plasma pre-analytics for metabolomics cohort studies', *International Journal of Molecular Sciences* **17**(12), 2035.

391 Johnson, C. H., Ivanisevic, J. and Siuzdak, G. 2016, 'Metabolomics: beyond biomarkers and towards mechanisms', *Nature reviews Molecular cell biology* **17**(7), 451–459.

392 Johnson, S. C., Rabinovitch, P. S. and Kaeberlein, M. 2013, 'mtor is a key modulator of ageing and age-related disease', *Nature* **493**(7432).

393 Jolliffe, I. T. and Cadima, J. 2016, 'Principal component analysis: a review and recent developments', *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences* **374**(2065), 20150202.

394 Joshi, A., Rienks, M., Theofilatos, K. and Mayr, M. 2021, 'Systems biology in cardiovascular disease: a multiomics approach', *Nature Reviews Cardiology* **18**(5), 313–330.

395 Jové, M., Naudi, A., Gambini, J., Borras, C., Cabre, R., Portero-Otin, M., J., V. and Pamplona, R. 2016, 'A stress-resistant lipidomic signature confers extreme longevity to humans', *The Journals of Gerontology: Series A* **72**(1).

396 Jridi, I., Canté-Barrett, K., Pike-Overzet, K. and Staal, F. J. T. 2021, 'Inflammation and wnt signaling: Target for immunomodulatory therapy?', *Frontiers in cell and developmental biology* **8**(615131).

397 Kalamohan, K., Tayubi, I. A., Madar, I. H. and Chen, Z. 2022, 'Integrative coexpression network detects the epicardial adipose tissue specific exosome vesicles as key predictor biomarker of cardiovascular disease', *Human Gene* **34**, 201088.

398 Kamali, A. H., Giannoulatou, E., Chen, T. Y., Charleston, M. A., McEwan, A. L. and Ho, J. W. 2015, 'How to test bioinformatics software?', *Biophysical reviews* **7**, 343–352.

399 Kane, A. E., Bisset, E. S., Keller, K. M., Ghimire, A., Pyle, W. G. and Howlett, S. E. 2020, 'Age, sex and overall health, measured as frailty, modify myofilament proteins in hearts from naturally aging mice', *Scientific Reports* **10**(1), 10052.

400 Kanehisa, M. and Goto, S. 2000, 'Kegg: kyoto encyclopedia of genes and genomes', *Nucleic acids research* **28**(1), 27–30.

401 Karbassi, E., Fenix, A., Marchiano, S., Muraoka, N., Nakamura, K., Yang, X. and Murry, C. E. 2020, 'Cardiomyocyte maturation: advances in knowledge and implications for regenerative medicine', *Nature Reviews Cardiology* **17**(6), 341–359.

402 Karkkainen, M. J., Haiko, P., Sainio, K., Partanen, J., Taipale, J., Petrova, T. V., Jeltsch, M., Jackson, D. G., Talikka, M., Rauvala, H., Betsholtz, C. and Alitalo, K. 2003, 'Vascular endothelial growth factor c is required for sprouting of the first lymphatic vessels from embryonic veins', *Nature Immunology* **5**(1).

403 Karp, P. D., Midford, P. E., Caspi, R. and Khodursky, A. 2021, 'Pathway size matters: the influence of pathway granularity on over-representation (enrichment analysis) statistics', *BMC genomics* **22**, 1–11.

404 Karpievitch, Y. V., Dabney, A. R. and Smith, R. D. 2012, 'Normalization and missing value imputation for label-free lc-ms analysis', *BMC bioinformatics* **13**, 1–9.

405 Karpievitch, Y. V., Nikolic, S. B., Wilson, R., Sharman, J. E. and Edwards, L. M. 2014, 'Metabolomics data normalization with eigenms', *PloS one* **9**(12), e116221.

406 Karpievitch, Y. V., Polpitiya, A. D., Anderson, G. A., Smith, R. D. and Dabney, A. R. 2010, 'Liquid chromatography mass spectrometry-based proteomics: biological and technological aspects', *The annals of applied statistics* **4**(4), 1797.

407 Karpievitch, Y. V., Taverner, T., Adkins, J. N., Callister, S. J., Anderson, G. A., Smith, R. D. and Dabney, A. R. 2009, 'Normalization of peak intensities in bottom-up ms-based proteomics using singular value decomposition', *Bioinformatics* **25**(19), 2573–2580.

408 Karwi, Q. G. and Lopaschuk, G. D. 2023, 'Branched-chain amino acid metabolism in the failing heart', *Cardiovascular Drugs and Therapy* **37**(2), 413–420.

409 Karwi, Q. G., Uddin, G. M., Ho, K. L. and Lopaschuk, G. D. 2018, 'Loss of metabolic flexibility in the failing heart', *Frontiers in Cardiovascular Medicine* **5**.
URL: *https://doi.org/10.3389/fcvm.2018.00068*

410 Katz, A. M. 2010, *Physiology of the Heart*, Lippincott Williams and Wilkins, Philadelphia, USA.

411 Kawase, Y. and Hajjar, R. J. 2008, 'The cardiac sarcoplasmic/endoplasmic reticulum calcium atpase: a potent target for cardiovascular diseases', *Nature Clinical Practice Cardiovascular Medicine* **5**(9), 554–565.

412 Kayala, M. A. and Baldi, P. 2012, 'Cyber-t web server: differential analysis of high-throughput data', *Nucleic acids research* **40**(W1), W553–W559.

413 Keating, N. 2022, 'A research framework for the united nations decade of healthy ageing (2021–2030)', *European Journal of Ageing* **19**(3), 775–787.

414 Keceli, G., Gupta, A., Sourdon, J., Gabr, R., Schär, M., Dey, S., Tocchetti, C. G., Stuber, A., Agrimi, J., Zhang, Y. et al. 2022, 'Mitochondrial creatine kinase attenuates pathologic remodeling in heart failure', *Circulation research* **130**(5), 741–759.

415 Keller, A., Backes, C. and Lenhof, H.-P. 2007, 'Computation of significance scores of unweighted gene set enrichment analyses', *BMC bioinformatics* **8**, 1–7.

416 Kelly, J., Berzuini, C., Keavney, B., Tomaszewski, M. and Guo, H. 2022, 'A review of causal discovery methods for molecular network analysis', *Molecular genetics & genomic medicine* **10**(10), e2055.

417 Kennedy-Shaffer, L. 2019, 'Before p¡ 0.05 to beyond p¡ 0.05: using history to contextualize p-values and significance testing', *The American Statistician* **73**(sup1), 82–90.

418 Kerr, K. F. 2009, 'Comments on the analysis of unbalanced microarray data', *Bioinformatics* **25**(16), 2035–2041.

419 Khalil, H., Kanisicak, O., Prasad, V., Correll, R. N., Fu, X., Schips, T., Vagnozzi, R. J., Liu, R., Huynh, T., Lee, S.-J., Karch, J. and Molkentin, J. D. 2017, 'Fibroblast-specific tgf-β–smad2/3 signaling underlies cardiac fibrosis', *The Journal of Clinical Investigation* **127**(10).

420 Khalilimeybodi, A., Riaz, M., Campbell, S. G., Omens, J. H., McCulloch, A. D., Qyang, Y. and Saucerman, J. J. 2023, 'Signaling network model of cardiomyocyte morphological changes in familial cardiomyopathy', *Journal of Molecular and Cellular Cardiology* **174**, 1–14.

421 Kihara, A. 2012, 'Very long-chain fatty acids: elongation, physiology and related disorders', *Journal of Biochemistry* **152**(5).

422 Kim, S.-Y. and Volsky, D. J. 2005, 'Page: parametric analysis of gene set enrichment', *BMC bioinformatics* **6**, 1–12.

423 Kim, T. K. and Park, J. H. 2019, 'More about the basic assumptions of t-test: normality and sample size', *Korean journal of anesthesiology* **72**(4), 331.

424 Kintscher, U., Foryst-Ludwig, A., Haemmerle, G. and Zechner, R. 2020, 'The role of adipose triglyceride lipase and cytosolic lipolysis in cardiac function and heart failure', *Cell Reports Medicine* **1**(1).

425 Kitakaze, M. 2016, 'Clinical evidence of the role of histamine in heart failure'.

426 Klabunde, R. E. 2012, *CARDIOVASCULAR PHYSIOLOGY CONCEPTS*, Lippincott Williams and Wilkins, Philadelphia, USA.

427 Klaips, C. L., Jayaraj, G. G. and Hartl, F. U. 2018, 'Pathways of cellular proteostasis in aging and disease', *Journal of Cell Biology* **217**(1), 51–63.

428 Klammer, A. A., Yi, X., MacCoss, M. J. and Noble, W. S. 2007, 'Improving tandem mass spectrum identification using peptide retention time prediction across diverse chromatography conditions', *Analytical chemistry* **79**(16), 6111–6118.

429 Knottnerus, S., Bleeker, J., Wüst, R., Ferdinandusse, S., IJlst, L., Wijburg, F., Wanders, R., Visser, G. and Houtkooper, R. 2018, 'Disorders of mitochondrial long-chain fatty acid oxidation and the carnitine shuttle', *Reviews in Endocrine and Metabolic Disorders* **19**(1).

430 Koay, Y. C., Wali, J. A., Luk, A. W., Macia, L., Cogger, V. C., Pulpitel, T. J., Wahl, D., Solon-Biet, S. M., Holmes, A., Simpson, S. J. et al. 2019, 'Ingestion of resistant starch by mice markedly increases microbiome-derived metabolites', *The FASEB Journal* **33**(7), 8033–8042.

431 Köfeler, H. C., Ahrends, R., Baker, E. S., Ekroos, K., Han, X., Hoffmann, N., Holčapek, M., Wenk, M. R. and Liebisch, G. 2021, 'Recommendations for good practice in ms-based lipidomics', *Journal of lipid research* **62**.

432 Koitabashi, N., Arai, M., Tomaru, K., Takizawa, T., Watanabe, A., Niwano, K., Yokoyama, T., Wuytack, F., Periasamy, M., Nagai, R. and Kurabayashi, M. 2005, 'Carvedilol effectively blocks oxidative stress-mediated downregulation of sarcoplasmic reticulum ca2+-atpase 2 gene transcription through modification of sp1 binding', *Biochemical and Biophysical Research Communications* **328**(1).

433 Koivumaki, J. T., Takalo, J., Korhonen, T., Tavi, P. and Weckstrom, M. 2009, 'Modelling sarcoplasmic reticulum calcium atpase and its regulation in cardiac myocytes', *Philosophical Transactions of the Royal Society A* **1896**(367).

434 Kolobkov, D. S., Sviridova, D. A., Abilev, S. K., Kuzovlev, A. N. and Salnikova, L. E. 2022, 'Genes and diseases: insights from transcriptomics studies', *Genes* **13**(7), 1168.

435 Korff, T., Braun, J., Pfaff, D., Augustin, H. G. and Hecker, M. 2008, 'Role of ephrinb2 expression in endothelial cells during arteriogenesis: impact on smooth muscle cell migration and monocyte recruitment', *Blood* **112**(1).

436 Korge, P., Calmettes, G. and Weiss, J. N. 2016, 'Reactive oxygen species production in cardiac mitochondria after complex i inhibition: modulation by substrate-dependent regulation of the nadh/nad+ ratio', *Free Radical Biology and Medicine* **96**, 22–33.

437 Korotkevich, G., Sukhov, V., Budin, N., Shpak, B., Artyomov, M. N. and Sergushichev, A. 2016, 'Fast gene set enrichment analysis', *BioRxiv* p. 060012.

438 Köse, T., Özgür, S., Coşgun, E., Keskinoğlu, A., Keskinoğlu, P. et al. 2020, 'Effect of missing data imputation on deep learning prediction performance for vesicoureteral reflux and recurrent urinary tract infection clinical study', *BioMed Research International* **2020**.

439 Kottwitz, J., Bruno, K. A., Berg, J., Salomon, G. R., Fairweather, D., Elhassan, M., Baltensperger, N., Kissel, C. K., Lovrinovic, M., Baltensweiler, A. et al. 2020, 'Myoglobin for detection of high-risk patients with acute myocarditis', *Journal of cardiovascular translational research* **13**, 853–863.

440 Kourelis, T. V., Dasari, S. S., Dispenzieri, A., Maleszewski, J. J., Redfield, M. M., Fayyaz, A. U., Grogan, M., Ramirez-Alvarado, M., Abou Ezzeddine, O. F. and McPhail, E. D. 2020, 'A proteomic atlas of cardiac amyloid plaques', *Cardio Oncology* **2**(4), 632–643.

441 Kraus, A. J., Brink, B. G. and Siegel, T. N. 2019, 'Efficient and specific oligo-based depletion of rrna', *Scientific reports* **9**(1), 12281.

442 Krehenwinkel, H., Wolf, M., Lim, J. Y., Rominger, A. J., Simison, W. B. and Gillespie, R. G. 2017, 'Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding', *Scientific reports* **7**(1), 17668.

443 Kresoja, K.-P., Rommel, K.-P., Wachter, R., Henger, S., Besler, C., Klöting, N., Schnelle, M., Hoffmann, A., Büttner, P., Ceglarek, U. et al. 2021, 'Proteomics to improve phenotyping in obese patients with heart failure with preserved ejection fraction', *European Journal of Heart Failure* **23**(10), 1633–1644.

444 Krumsiek, J., Suhre, K., Illig, T., Adamski, J. and Theis, F. J. 2011, 'Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data', *BMC systems biology* **5**, 1–16.

445 Krützfeldt, A. 1990, 'Metabolism of exogenous substrates by coronary endothelial cells in culture', *Journal of Molecular and Cellular Cardiology* **22**(12).

446 Kubin, T., Pöling, J., Kostin, S., Gajawada, P., Hein, S., Rees, W., Wietelmann, A., Tanaka, M., Lörchner, H., Schimanski, S. et al. 2011, 'Oncostatin m is a major mediator of cardiomyocyte dedifferentiation and remodeling', *Cell stem cell* **9**(5), 420–432.

447 Kulyyassov, A., Fresnais, M. and Longuespée, R. 2021, 'Targeted liquid chromatography-tandem mass spectrometry analysis of proteins: basic principles, applications, and perspectives', *Proteomics* **21**(23-24), 2100153.

448 Kumari, S., MG, S. and Mayor, S. 2010, 'Endocytosis unplugged: multiple ways to enter the cell', *Cell Research* **20**(3).

449 Kuo, I. Y. and Ehrlich, B. E. 2015, 'Signaling in muscle contraction.', *Cold Spring Harbor Perspectives in Biology* **7**(2).

450 Kuro-o, M. 2018, 'The klotho proteins in health and disease', *Nature Reviews Nephrology* **15**(1).

451 Lakatta, E. G. 2003, 'Arterial and cardiac aging: major shareholders in cardiovascular disease enterprises: Part iii: cellular and molecular clues to heart and arterial aging', *Circulation* **107**(3), 490–497.

452 Lal, S., Li, A., Allen, D., Allen, P. D., Bannon, P., Cartmill, T., Cooke, R., Farnsworth, A., Keogh, A. and Dos Remedios, C. 2015, 'Best practice biobanking of human heart tissue', *Biophysical reviews* **7**, 399–406.

453 Lam, S. and Shui, G. 2013, 'Lipidomics as a principal tool for advancing biomedical research', *Journal of Genetics and Genomics* **40**(8).

454 Lam, S., Wang, Z., Li, B. and Shui, G. 2021, 'High-coverage lipidomics for functional lipid and pathway analyses', *Analytica Chimica Acta* **1147**.

455 Langfelder, P. and Horvath, S. 2008, 'Wgcna: an r package for weighted correlation network analysis', *BMC Bioinformatics* **9**(1).

456 Langfelder, P., Mischel, P. S. and Horvath, S. 2013, 'When is hub gene selection better than standard meta-analysis?', *PloS one* **8**(4), e61505.

457 Lantz, B. 2013, 'The impact of sample non-normality on anova and alternative methods', *British Journal of Mathematical and Statistical Psychology* **66**(2), 224–244.

458 Lanzani, C., Citterio, L. and Vezzoli, G. 2020, 'Klotho: a link between cardiovascular and non-cardiovascular mortality', *Clinical kidney journal* **13**(6).

459 Larsen, P. and Tennagels, N. 2014, 'On ceramides, other sphingolipids and impaired glucose homeostasis', *Molecular Metabolism* **3**(3).

460 Law, C. W., Chen, Y., Shi, W. and Smyth, G. K. 2014, 'voom: Precision weights unlock linear model analysis tools for rna-seq read counts', *Genome biology* **15**, 1–17.

461 Lebaron, F. and Folch, J. 1959, 'The effect of ph and salt concentration on aqueous extraction of brain proteins and lipoproteins', *Journal of neurochemistry* **4**(1), 1–8.

462 Lee, D. K. 2020, 'Data transformation: a focus on the interpretation', *Korean journal of anesthesiology* **73**(6), 503.

463 Lee, I. C., Tumanov, S., Wong, J. W., Stocker, R. and Ho, J. W. 2023, 'Integrative processing of untargeted metabolomic and lipidomic data using multiabler', *iScience* **26**(6).

464 Lee, S., Zhang, A. Y., Su, S., Ng, A. P., Holik, A. Z., Asselin-Labat, M.-L., Ritchie, M. E. and Law, C. W. 2020, 'Covering all your bases: incorporating intron signal from rna-seq data', *NAR Genomics and Bioinformatics* **2**(3), lqaa073.

465 Leek, J. T. and Storey, J. D. 2007, 'Capturing heterogeneity in gene expression studies by surrogate variable analysis', *PLoS genetics* **3**(9), e161.

466 Lempp, M., Farke, N., Kuntz, M., Freibert, S. A., Lill, R. and Link, H. 2019, 'Systematic identification of metabolites controlling gene expression in e. coli', *Nature communications* **10**(1), 4463.

467 Lennicke, C. and Cochemé, H. M. 2021, 'Redox metabolism: Ros as specific molecular regulators of cell signaling and function', *Molecular Cell* **81**(18), 3691–3707.

468 Leone, M., Musa, G. and Engel, F. B. 2018, 'Cardiomyocyte binucleation is associated with aberrant mitotic microtubule distribution, mislocalization of rhoa and iqgap3, as well as defective actomyosin ring anchorage and cleavage furrow ingression', *Cardiovascular research* **114**(8), 1115–1131.

469 Lesnefsky, E. J., Chen, Q. and Hoppel, C. L. 2016, 'Mitochondrial metabolism in aging heart', *Circulation Research* **118**(10).

470 Levin, J. Z., Yassour, M., Adiconis, X., Nusbaum, C., Thompson, D. A., Friedman, N., Gnirke, A. and Regev, A. 2010, 'Comprehensive comparative analysis of strand-specific rna sequencing methods', *Nature methods* **7**(9), 709–715.

471 Lewis, C. T. and Ochala, J. 2023, 'Myosin heavy chain as a novel key modulator of striated muscle resting state', *Physiology* **38**(1), 3–9.

472 Li, D. K., Smith, L. E., Rookyard, A. W., Lingam, S. J., Koay, Y. C., McEwen, H. P., Twigg, S. M., Don, A. S., O'Sullivan, J. F., Cordwell, S. J. et al. 2022, 'Multi-omics of a pre-clinical model of diabetic cardiomyopathy reveals increased fatty acid supply impacts mitochondrial metabolic selectivity', *Journal of Molecular and Cellular Cardiology* **164**, 92–109.

473 Li, H., Hastings, M. H., Rhee, J., Trager, L. E., Roh, J. D. and Rosenzweig, A. 2020, 'Targeting age-related pathways in heart failure', *Circulation research* **126**(4), 533–551.

474 Li, J., Vosegaard, T. and Guo, Z. 2017, 'Applications of nuclear magnetic resonance in lipid analyses: An emerging powerful tool for lipidomics studies', *Progress In Lipid Research* **68**.

475 Li, M., Parker, B. L., Pearson, E., Hunter, B., Cao, J., Koay, Y. C., Guneratne, O., James, D. E., Yang, J., Lal, S. et al. 2020, 'Core functional nodes and sex-specific pathways in human ischaemic and dilated cardiomyopathy', *Nature communications* **11**(1), 2843.

476 Li, M. and Smyth, G. K. 2023, 'Neither random nor censored: estimating intensity-dependent probabilities for missing values in label-free proteomics', *Bioinformatics* **39**(5), btad200.

477 Li, N. and Brundel, B. J. 2020, 'Inflammasomes and proteostasis novel molecular mechanisms associated with atrial fibrillation', *Circulation research* **127**(1), 73–90.

478 Li, N., Muthusamy, S., Liang, R., Sarojini, H. and Wang, E. 2011, 'Increased expression of mir-34a and mir-93 in rat liver during aging, and their impact on the expression of mgst1 and sirt1', *Mechanisms of ageing and development* **3**(132).

479 Li, S., Gao, D. and Jiang, Y. 2019, 'Function, detection and alteration of acylcarnitine metabolism in hepatocellular carcinoma', *Metabolites* **9**(2).

480 Li, X., Lin, Y., Wang, S., Zhou, S., Ju, J., Wang, X., Chen, Y. and Xia, M. 2020, 'Extracellular superoxide dismutase is associated with left ventricular geometry and heart failure in patients with cardiovascular disease', *Journal of the American Heart Association* **9**(15), e016862.

481 Li, Y.-F. and Wang, X. 2011, 'The role of the proteasome in heart disease', *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* **1809**(2), 141–149.

482 Li, Y., Fan, T. W., Lane, A. N., Kang, W.-Y., Arnold, S. M., Stromberg, A. J., Wang, C. and Chen, L. 2019, 'Sda: A semi-parametric differential abundance analysis method for metabolomics and proteomics data', *BMC bioinformatics* **20**, 1–10.

483 Li, Y., Ge, X., Peng, F., Li, W. and Li, J. J. 2022, 'Exaggerated false positives by popular differential expression methods when analyzing human population samples', *Genome biology* **23**(1), 79.

484 Li-Zhen, L., Chen, Z.-c., Wang, S.-S., Liu, W.-b. and Zhuang, X.-d. 2021, 'Klotho deficiency causes cardiac ageing by impairing autophagic and activating apoptotic activity', *European Journal of Pharmacology* **911**, 174559.

485 Liao, X., Zhou, S., Zeng, D., Ying, W., Lian, D., Zhang, M., Ge, J., Chen, M., Liu, Y. and Lin, Y. 2023, 'Roles of the crucial mitochondrial dna in hypertrophic cardiomyopathy prognosis and diagnosis: A review', *Medicine* **102**(48), e36368.

486 Liao, Y. and Shi, W. 2020, 'Read trimming is not required for mapping and quantification of rna-seq reads at the gene level', *NAR genomics and bioinformatics* **2**(3), lqaa068.

487 Liao, Y., Smyth, G. K. and Shi, W. 2019, 'The r package rsubread is easier, faster, cheaper and better for alignment and quantification of rna sequencing reads', *Nucleic acids research* **47**(8), e47–e47.

488 Liberale, L., Kraler, S., Camici, G. G. and Lüscher, T. F. 2020, 'Ageing and longevity genes in cardiovascular diseases', *Basic & clinical pharmacology & toxicology* **127**(2), 120–131.

489 Licari, C., Tenori, L., Giusti, B., Sticchi, E., Kura, A., De Cario, R., Inzitari, D., Piccardi, B., Nesi, M., Sarti, C. et al. 2021, 'Analysis of metabolite and lipid association networks reveals molecular mechanisms associated with 3-month mortality and poor functional outcomes in patients with acute ischemic stroke after thrombolytic treatment with recombinant tissue plasminogen activator', *Journal of Proteome Research* **20**(10), 4758–4770.

490 Liehn, E. A., Lupan, A.-M., Diaconu, R., Ioana, M., Streata, I., Manole, C. and Burlacu, A. 2021, 'Heart function assessment during aging in apolipoprotein e knock-out mice', *Discoveries* **9**(3).

491 Liehn, E. A., Ponomariov, V., Diaconu, R., Streata, I., Ioana, M., Crespo-Avilan, G. E., Hernández-Reséndiz, S. and Cabrera-Fuentes, H. A. 2018, 'Apolipoprotein e in cardiovascular diseases: novel aspects of an old-fashioned enigma', *Archives of medical research* **49**(8), 522–529.

492 Lim, J., Park, C., Kim, M., Kim, H., Kim, J. and Lee, D.-S. 2024, 'Advances in single-cell omics and multiomics for high-resolution molecular profiling', *Experimental & Molecular Medicine* pp. 1–12.

493 Lima-Mendez, G. and Van Helden, J. 2009, 'The powerful law of the power law and other myths in network biology', *Molecular BioSystems* **5**(12), 1482–1493.

494 Lin, S. M., Du, P., Huber, W. and Kibbe, W. A. 2008, 'Model-based variance-stabilizing transformation for illumina microarray data', *Nucleic acids research* **36**(2), e11–e11.

495 Lin, W.-C. and Tsai, C.-F. 2020, 'Missing value imputation: a review and analysis of the literature (2006–2017)', *Artificial Intelligence Review* **53**, 1487–1509.

496 Lin, Z., Long, F., Kang, R., Klionsky, D. J., Yang, M. and Tang, D. 2024, 'The lipid basis of cell death and autophagy', *Autophagy* **20**(3), 469–488.

497 Lindqvist, L. M., Tandoc, K., Topisirovic, I. and Furic, L. 2018, 'Cross-talk between protein synthesis, energy metabolism and autophagy in cancer', *Current opinion in genetics & development* **48**, 104–111.

498 Lindskrog, S. V., Prip, F., Lamy, P., Taber, A., Groeneveld, C. S., Birkenkamp-Demtröder, K., Jensen, J. B., Strandgaard, T., Nordentoft, I., Christensen, E. et al. 2021, 'An integrated multi-omics analysis identifies prognostic molecular subtypes of non-muscle-invasive bladder cancer', *Nature communications* **12**(1), 2301.

499 Lindstrom, M., DeCleene, N., Dorsey, H., Fuster, V., Johnson, C. O., LeGrand, K. E., Mensah, G. A., Razo, C., Stark, B., Varieur Turco, J. et al. 2022, 'Global burden of cardiovascular diseases and risks collaboration, 1990-2021', *Journal of the American College of Cardiology* **80**(25), 2372–2425.

500 Linnane, A. W., Kovalenko, S. and Gingold, E. B. 1998, 'The universality of bioenergetic disease: Age-associated cellular bioenergetic degradation and amelioration therapy', *Annals of the New York Academy of Sciences* **854**(1).

501 Linton, P., Gurney, M., Sengstock, D., Mentzer, R. M. and Gottlieb, R. A. 2015, 'This old heart: Cardiac aging and autophagy', *Journal of Molecular and Cellular Cardiology* **83**.

502 Lipsett, D. B., Frisk, M., Aronsen, J. M., Nordén, E. S., Buonarati, O. R., Cataliotti, A., Hell, J. W., Sjaastad, I., Christensen, G. and Louch, W. E. 2019, 'Cardiomyocyte substructure reverts to an immature phenotype during heart failure', *The Journal of Physiology* **597**(7).

503 Lísa, M., Cífková, E., Khalikova, M., Ovčačíková, M. and Holčapek, M. 2017, 'Lipidomic analysis of biological samples: comparison of liquid chromatography, supercritical fluid chromatography and direct infusion mass spectrometry methods', *Journal of Chromatography A* **1525**, 96–108.

504 Litviňuková, M., Talavera-Lopez, C., Maatz, H., Reichart, D., Worth, C. L., Lindberg, E. L., Kanda, M., Polanski, K., Heinig, M., Lee, M., Nadelmann, E. R., Roberts, K., Tuck, L., Fasouli, E. S., DeLaughter, D. M., McDonough, B., Wakimoto, H., Gorham, J. M., Samari, S., Mahbubani, K. T., Saeb-Parsy, K., Patone, G., Boyle, J. J., Zhang, H., Zhang, H., Viveiros, A., Oudit, G. Y., Bayraktar, O. A., Seidman, J. G., E., S. C., Noseda, M., Hubner, N. and Teichmann, S. A. 2020, 'Cells of the adult human heart', *Nature* **588**(7838).

505 Liu, C.-L., Guo, J., Zhang, X., Sukhova, G. K., Libby, P. and Shi, G.-P. 2018, 'Cysteine protease cathepsins in cardiovascular disease: from basic research to clinical trials', *Nature Reviews Cardiology* **15**(6), 351–370.

506 Liu, M. and Gomez, D. 2019, 'Smooth muscle cell phenotypic diversity', *Arteriosclerosis, Thrombosis, and Vascular Biology* **39**(9).

507 Liu, P. and Hwang, J. G. 2007, 'Quick calculation for sample size while controlling false discovery rate with application to microarray analysis', *Bioinformatics* **23**(6), 739–746.

508 Liu, R., Bian, Y., Liu, L., Liu, L., Liu, X. and Ma, S. 2022, 'Molecular pathways associated with oxidative stress and their potential applications in radiotherapy', *International Journal of Molecular Medicine* **49**(5), 1–11.

509 Liu, T., Zhang, L., Joo, D. and Sun, S. C. 2017, 'Nf-κb signaling in inflammation', *Signal Transduction and Targeted Therapy* **2**(17023).

510 Liu, X. and Locasale, J. W. 2017, 'Metabolomics: a primer', *Trends in biochemical sciences* **42**(4), 274–284.

511 Long, X., Boluyt, M. O., O'Neill, L., Zheng, J.-S., Wu, G., Nitta, Y. K., Crow, M. T. and Lakatta, E. G. 1999, 'Myocardial retinoid x receptor, thyroid hormone receptor, and myosin heavy chain gene expression in the rat during adult aging', *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences* **54**(1), B23–B27.

512 Longo, V. D. and Anderson, R. M. 2022, 'Nutrition, longevity and disease: From molecular mechanisms to interventions', *Cell* **185**(9), 1455–1470.

513 Lönnstedt, I. and Speed, T. 2002, 'Replicated microarray data', *Statistica sinica* pp. 31–46.

514 Lopaschuk, G. D. and Jaswal, J. S. 2010, 'Energy metabolic phenotype of the cardiomyocyte during development, differentiation, and postnatal maturation', *Journal of Cardiovascular Pharmacology* **56**(2).

515 Lopaschuk, G. D., Karwi, Q. G., Tian, R., Wende, A. R. and Abel, E. D. 2021, 'Cardiac energy metabolism in heart failure', *Circulation research* **128**(10), 1487–1513.

516 Lother, A., Bergemann, S., Deng, L., Moser, M., Bode, C. and Hein, L. 2018, 'Cardiac endothelial cell transcriptome', *Arteriosclerosis, Thrombosis, and Vascular Biology* **38**(3).

517 Lou, R., Cao, Y., Li, S., Lang, X., Li, Y., Zhang, Y. and Shui, W. 2023, 'Benchmarking commonly used software suites and analysis workflows for dia proteomics and phosphoproteomics', *Nature Communications* **14**(1), 94.

518 Love, M. I., Huber, W. and Anders, S. 2014, 'Moderated estimation of fold change and dispersion for rna-seq data with deseq2', *Genome biology* **15**, 1–21.

519 Lowes, B. D., Minobe, W., Abraham, W. T., Rizeq, M. N., Bohlmeyer, T. J., Quaife, R. A., Roden, R. L., Dutcher, D. L., Robertson, A. D., Voelkel, N. F. et al. 1997, 'Changes in gene expression in the intact human heart. downregulation of alpha-myosin heavy chain in hypertrophied, failing ventricular myocardium.', *The Journal of clinical investigation* **100**(9), 2315–2324.

520 Lu, J.-Y., Verkruyse, L. A. and Hofmann, S. L. 2002, 'The effects of lysosomotropic agents on normal and incl cells provide further evidence for the lysosomal nature of palmitoyl-protein thioesterase function', *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids* **1583**(1), 35–44.

521 Lu, L., Guo, J., Hua, Y., Huang, K., Magaye, R., Cornell, J., Kelly, D. J., Reid, C., Liew, D., Zhou, Y. et al. 2017, 'Cardiac fibrosis in the ageing heart: Contributors and mechanisms', *Clinical and Experimental Pharmacology and Physiology* **44**, 55–63.

522 Lu, Q., Pan, B., Bai, H., Zhao, W., Liu, L., Li, G., Liu, R., Lv, T., Huang, X., Li, X. et al. 2022, 'Intranuclear cardiac troponin i plays a functional role in regulating atp2a2 expression in cardiomyocytes', *Genes & diseases* **9**(6), 1689–1700.

523 Lu, T. and Feng, Z. 2012, 'Nalcn: A regulator of pacemaker activity', *Molecular Neurobiology* **45**(3).

524 Lumsden, A. L., Mulugeta, A., Zhou, A. and Hyppönen, E. 2020, 'Apolipoprotein e (apoe) genotype-associated disease risks: a phenome-wide, registry-based, case-control study utilising the uk biobank', *EBioMedicine* **59**.

525 Lunney, J. K., Van Goor, A., Walker, K. E., Hailstock, T., Franklin, J. and Dai, C. 2021, 'Importance of the pig as a human biomedical model', *Science translational medicine* **13**(621), eabd5758.

526 Luo, J., Xu, L., Li, J. and Zhao, S. 2018, 'Effects and mechanisms of apolipoprotein av on the regulation of lipid accumulation in cardiomyocytes', *Lipids in Health and Disease* **17**, 1–13.

527 Luo, W., Friedman, M. S., Shedden, K., Hankenson, K. D. and Woolf, P. J. 2009, 'Gage: generally applicable gene set enrichment for pathway analysis', *BMC bioinformatics* **10**, 1–17.

528 López Otín, C., Blasco, M. A., Partridge, L., Serrano, M. and Kroemer, G. 2023, 'Hallmarks of aging: An expanding universe', *Cell* **186**(2).

529 Ma, Y. and Li, J. 2015, 'Metabolic shifts during aging and pathology', *Comprehensive Physiology* **5**(2).

530 Macleod, M. R., Michie, S., Roberts, I., Dirnagl, U., Chalmers, I., Ioannidis, J. P., Salman, R. A.-S., Chan, A.-W. and Glasziou, P. 2014, 'Biomedical research: increasing value, reducing waste', *The Lancet* **383**(9912), 101–104.

531 Maejima, Y., Zablocki, D., Nah, J. and Sadoshima, J. 2022, 'The role of the hippo pathway in autophagy in the heart', *Cardiovascular Research* **118**(17).

532 Maggiorani, D., Manzella, N., Edmondson, D. E., Mattevi, A., Parini, A., Binda, C. and Mialet-Perez, J. 2017, 'Monoamine oxidases, oxidative stress, and altered mitochondrial dynamics in cardiac ageing', *Oxidative Medicine and Cellular Longevity* **2017**.

533 Majidinia, M., Aghazadeh, J., Jahanban-Esfahlani, R. and Yousefi, B. 2018, 'The roles of wnt/β-catenin pathway in tissue development and regenerative medicine', *Journal of cellular physiology* **233**(8), 5598–5612.

534 Maldonado, E. N. and Lemasters, J. J. 2014, 'Atp/adp ratio, the missed connection between mitochondria and the warburg effect', *Mitochondrion* **19**, 78–84.

535 Mamazhakypov, A., Sartmyrzaeva, M., Sarybaev, A. S., Schermuly, R. and Sydykov, A. 2022, 'Clinical and molecular implications of osteopontin in heart failure', *Current Issues in Molecular Biology* **44**(8), 3573–3597.

536 Manadas, B., Mendes, V. M., English, J. and Dunn, M. J. 2010, 'Peptide fractionation in proteomics approaches', *Expert review of proteomics* **7**(5), 655–663.

537 Manea, M., Mező, G., Hudecz, F. and Przybylski, M. 2007, 'Mass spectrometric identification of the trypsin cleavage pathway in lysyl-proline containing oligotuftsin peptides', *Journal of peptide science: an official publication of the European Peptide Society* **13**(4), 227–236.

538 Maning, J., Desimine, V. L., Pollard, C. M., Ghandour, J. and Lymperopoulos, A. 2022, 'Carvedilol selectively stimulates β arrestin2-dependent serca2a activity in cardiomyocytes to augment contractility', *International Journal of Molecular Sciences* **23**(19).

539 Marais, A. D. 2019, 'Apolipoprotein e in lipoprotein metabolism, health and cardiovascular disease', *Pathology* **51**(2), 165–176.

540 Marchiano, S., Nakamura, K., Reinecke, H., Neidig, L., Lai, M., Kadota, S., Perbellini, F., Yang, X., Klaiman, J. M., Blakely, L. P. et al. 2023, 'Gene editing to prevent ventricular arrhythmias associated with cardiomyocyte cell therapy', *Cell Stem Cell* **30**(4), 396–414.

541 Maron, B. J., Friedman, R. A., Kligfield, P., Levine, B. D., Viskin, S., Chaitman, B. R., Okin, P. M., Saul, J. P., Salberg, L., Van Hare, G. F. et al. 2014*a*, 'Assessment of the 12-lead ecg as a screening test for detection of cardiovascular disease in healthy general populations of young people (12–25 years of age) a scientific statement from the american heart association and the american college of cardiology', *Circulation* **130**(15), 1303–1334.

542 Maron, B. J., Friedman, R. A., Kligfield, P., Levine, B. D., Viskin, S., Chaitman, B. R., Okin, P. M., Saul, J. P., Salberg, L., Van Hare, G. F. et al. 2014*b*, 'Assessment of the 12-lead electrocardiogram as a screening test for detection of cardiovascular disease in healthy general populations of young people (12–25 years of age) a scientific statement from the american heart association and the american college of cardiology', *Journal of the American College of Cardiology* **64**(14), 1479–1514.

543 Martens, M., Ammar, A., Riutta, A., Waagmeester, A., Slenter, D. N., Hanspers, K., A. Miller, R., Digles, D., Lopes, E. N., Ehrhart, F. et al. 2021, 'Wikipathways: connecting communities', *Nucleic acids research* **49**(D1), D613–D621.

544 Martin, M. A. 2007, 'Bootstrap hypothesis testing for some common statistical problems: A critical evaluation of size and power properties', *Computational Statistics & Data Analysis* **51**(12), 6321–6342.

545 Martucciello, S., Turturo, M. G., Bilio, M., Cioffi, S. M. G., Chen, L., Baldini, A. and Illingworth, E. 2020, 'A dual role for tbx1 in cardiac lymphangiogenesis through genetic interaction with vegfr3', *The FASEB Journal* **34**(11).

546 Matsumori, A. 2023, 'Nuclear factor-κb is a prime candidate for the diagnosis and control of inflammatory cardiovascular disease', *European Cardiology Review* **18**.

547 Matyash, V., Liebisch, G., Kurzchalia, T. V., Shevchenko, A. and Schwudke, D. 2008, 'Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics', *Journal of lipid research* **49**(5), 1137–1146.

548 Maulik, N., Kagan, V. E., Tyurin, V. A. and Das, D. K. 1998, 'Redistribution of phosphatidylethanolamine and phosphatidylserine precedes reperfusion-induced apoptosis', *American Journal of Physiology-Heart and Circulatory Physiology* **274**(1), H242–H248.

549 May, R. M. 2006, 'Network structure and the biology of populations', *Trends in Ecology & Evolution* **21**(7), 394–399.

550 Ma'ayan, A. 2011, 'Introduction to network analysis in systems biology', *Science signaling* **4**(190), tr5–tr5.

551 McCann, M. R., George De la Rosa, M. V., Rosania, G. R. and Stringer, K. A. 2021, 'L-carnitine and acylcarnitines: mitochondrial biomarkers for precision medicine', *Metabolites* **11**(1), 51.

552 McCarthy, D. J. and Smyth, G. K. 2009, 'Testing significance relative to a fold-change threshold is a treat', *Bioinformatics* **25**(6), 765–771.

553 McCoin, C., Knotts, T. and Adams, S. 2015, 'Acylcarnitines-old actors auditioning for new roles in metabolic physiology', *Nature Reviews Endocrinology* **11**(10).

554 McCommis, K. S. and Finck, B. N. 2015, 'Mitochondrial pyruvate transport: a historical perspective and future research directions', *Biochemical Journal* **466**(3).

555 McGranaghan, P., Saxena, A., Rubens, M., Radenkovic, J., Bach, D., Schleussner, L., Pieske, B., Edelmann, F. and Trippel, T. D. 2020, 'Predictive value of metabolomic biomarkers for cardiovascular disease risk: A systematic review and meta-analysis', *Biomarkers* **25**(2), 101–111.

556 McMaughan, D. J., Oloruntoba, O. and Smith, M. L. 2020, 'Socioeconomic status and access to healthcare: interrelated drivers for healthy aging', *Frontiers in public health* **8**, 512143.

557 McShane, B. B., Gal, D., Gelman, A., Robert, C. and Tackett, J. L. 2019, 'Abandon statistical significance', *The American Statistician* **73**(sup1), 235–245.

558 Mehdizadeh, M., Aguilar, M., Thorin, E., Ferbeyre, G. and Nattel, S. 2022, 'The role of cellular senescence in cardiac disease: basic biology and clinical relevance', *Nature Reviews Cardiology* **19**(4), 250–264.

559 Mehta, N. K., Abrams, L. R. and Myrskylä, M. 2020, 'Us life expectancy stalls due to cardiovascular disease, not drug deaths', *Proceedings of the National Academy of Sciences* **117**(13), 6998–7000.

560 Meissner, F., Geddes-McAlister, J., Mann, M. and Bantscheff, M. 2022, 'The emerging role of mass spectrometry-based proteomics in drug discovery', *Nature Reviews Drug Discovery* **21**(9), 637–654.

561 Mejia, E. and Hatch, G. 2015, 'Mitochondrial phospholipids: role in mitochondrial function', *Journal of Bioenergetics and Biomembranes* **48**(2).

562 Melé, M., Ferreira, P. G., Reverter, F., DeLuca, D. S., Monlong, J., Sammeth, M., Young, T. R., Goldmann, J. M., Pervouchine, D. D., Sullivan, T. J. et al. 2015, 'The human transcriptome across tissues and individuals', *Science* **348**(6235), 660–665.

563 Menahan, L. 1983, 'Age-related changes in lipid and carbohydrate metabolism of the genetically obese mouse', *Metabolism* **32**(2), 172–178.

564 Mensah, G. A., Fuster, V., Murray, C. J., Roth, G. A., of Cardiovascular Diseases, G. B. and Collaborators, R. 2023, 'Global burden of cardiovascular diseases and risks, 1990-2022', *Journal of the American College of Cardiology* **82**(25), 2350–2473.

565 Miao, Z., Humphreys, B. D., McMahon, A. P. and Kim, J. 2021, 'Multi-omics integration in the age of million single-cell data', *Nature Reviews Nephrology* **17**(11), 710–724.

566 Mijanovic, O., Petushkova, A. I., Brankovic, A., Turk, B., Solovieva, A. B., Nikitkina, A. I., Bolevich, S., Timashev, P. S., Parodi, A. and Zamyatnin Jr, A. A. 2021, 'Cathepsin d—managing the delicate balance', *Pharmaceutics* **13**(6), 837.

567 Miko, M., Kyselovic, J., Danisovic, L., Barczi, T., Polak, S. and Varga, I. 2017, 'Two nuclei inside a single cardiac muscle cell. more questions than answers about the binucleation of cardiomyocytes', *Biologia* **72**(8).

568 Milacic, M., Beavers, D., Conley, P., Gong, C., Gillespie, M., Griss, J., Haw, R., Jassal, B., Matthews, L., May, B. et al. 2024, 'The reactome pathway knowledgebase 2024', *Nucleic acids research* **52**(D1), D672–D678.

569 Miller, A. J. 1963, 'The lymphatics of the heart', *Archives of Internal Medicine* **112**(4).

570 Miller, W. and Auchus, R. 2011, 'The molecular biology, biochemistry, and physiology of human steroidogenesis and its disorders', *Endocrine Reviews* **3**(1).

571 Minchin, S. and Lodge, J. 2019, 'Understanding biochemistry: structure and function of nucleic acids', *Essays in biochemistry* **63**(4), 433–456.

572 Mishra, P., Singh, U., Pandey, C. M., Mishra, P. and Pandey, G. 2019, 'Application of student's t-test, analysis of variance, and covariance', *Annals of cardiac anaesthesia* **22**(4), 407–411.

573 Misra, B. B. 2020, 'Data normalization strategies in metabolomics: Current challenges, approaches, and tools', *European Journal of Mass Spectrometry* **26**(3), 165–174.

574 Misra, B., Langefeld, C., Olivier, M. and Cox, L. 2019, 'Integrated omics: tools, advances and future approaches', *Journal of Molecular Endocrinology* **62**(1).

575 Miyata, S., Minobe, W., Bristow, M. R. and Leinwand, L. A. 2000, 'Myosin heavy chain isoform expression in the failing and nonfailing human heart', *Circulation research* **86**(4), 386–390.

576 Mohamed, A., Molendijk, J. and Hill, M. M. 2020, 'Lipidr: a software tool for data mining and analysis of lipidomics datasets', *Journal of proteome research* **19**(7), 2890–2897.

577 Moldenhauer, L. M., Cockshell, M. P., Frost, L., Parham, K. A., Tvorogov, D., Tan, L. Y., Ebert, L. M., Tooley, K., Worthley, S., Lopez, A. F. et al. 2015, 'Interleukin-3 greatly expands non-adherent endothelial forming cells with pro-angiogenic properties', *Stem Cell Research* **14**(3), 380–395.

578 Molvin, J., Jujić, A., Melander, O., Pareek, M., Råstam, L., Lindblad, U., Daka, B., Leósdóttir, M., Nilsson, P. M., Olsen, M. H. et al. 2020, 'Proteomic exploration of common pathophysiological pathways in diabetes and cardiovascular disease', *ESC Heart Failure* **7**(6), 4151–4158.

579 Monteil, A., Senatore, A., Gil-Nagel, A., Parra-Díaz, P., Lory, P. and Guérineau, N. 2024, 'New insights into the physiology and pathophysiology of the atypical sodium leak channel nalcn', *Physiological Reviews* **104**(1).

580 Moreira, J. B., Wohlwend, M. and Wisløff, U. 2020, 'Exercise and cardiac health: physiological and molecular insights', *Nature Metabolism* **2**(9), 829–839.

581 Morgan, M., Anders, S., Lawrence, M., Aboyoun, P., Pagès, H. and Gentleman, R. 2009, 'Shortread: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data', *Bioinformatics* **25**(19), 2607–2608.

582 Muallem, S., Chung, W., Jha, A. and Ahuja, M. 2017, 'Lipids at membrane contact sites: cell signaling and ion transport', *EMBO reports* **18**(11).

583 Mukherjee, P., Roy, S., Ghosh, D. and Nandi, S. 2022, 'Role of animal models in biomedical research: a review', *Laboratory Animal Research* **38**(1), 18.

584 Murashige, D., Jang, C., Neinast, M., Edwards, J. J., Cowan, A., Hyman, M. C., Rabinowitz, J. D., Frankel, D. S. and Arany, Z. 2020, 'Comprehensive quantification of fuel use by the failing and nonfailing human heart', *Science* **370**(6514), 364–368.

585 Murdoch, D. J., Tsai, Y.-L. and Adcock, J. 2008, 'P-values are random variables', *The American Statistician* **62**(3), 242–245.

586 Murie, C., Woody, O., Lee, A. Y. and Nadon, R. 2009, 'Comparison of small n statistical tests of differential expression applied to microarrays', *BMC bioinformatics* **10**, 1–18.

587 Murray, T. V., Ahmad, A. and Brewer, A. C. 2014, 'Reactive oxygen at the heart of metabolism', *Trends in cardiovascular medicine* **24**(3), 113–120.

588 Murthy, V. L., Reis, J. P., Pico, A. R., Kitchen, R., Lima, J. A., Lloyd-Jones, D., Allen, N. B., Carnethon, M., Lewis, G. D., Nayor, M. et al. 2020, 'Comprehensive metabolic phenotyping refines cardiovascular risk in young adults', *Circulation* **142**(22), 2110–2127.

589 Mutijima, F., Basiony, M. and Fujita, M. 2020, 'Lipid droplets formation and their effects on phosphatidylinositol level in yeasts', *Novel Research in Microbiology Journal* **4**(4), 856–867.

590 Mutlu, A. S., Duffy, J. and Wang, M. C. 2021, 'Lipid metabolism and lipid signals in aging and longevity', *Developmental cell* **56**(10), 1394–1407.

591 Nagoshi, T., Yoshimura, M., Rosano, G. M. C., Lopaschuk, G., D. and Mochizuki, S. 2011, 'Optimization of cardiac metabolism in heart failure', *Current Pharmaceutical Design* **17**(35).

592 Nakamura, M., Yudell, B. and Loor, J. 2014, 'Regulation of energy metabolism by long-chain fatty acids', *Progress in Lipid Research* **53**.

593 Nakayama, H., Nishida, K. and Otsu, K. 2016, 'Macromolecular degradation systems and cardiovascular aging', *Circulation research* **118**(10), 1577–1592.

594 Nam, A. S., Kim, K.-T., Chaligne, R., Izzo, F., Ang, C., Taylor, J., Myers, R. M., Abu-Zeinah, G., Brand, R., Omans, N. D. et al. 2019, 'Genotyping of transcriptomes links somatic mutations and cell identity', *Nature* **571**(7765), 355.

595 Nanayakkara, S., Marwick, T. H. and Kaye, D. M. 2018, 'The ageing heart: the systemic and coronary circulation', *Heart* **104**(5), 370–376.

596 Nayor, M., Cooper, L. L., Enserro, D. M., Xanthakis, V., Larson, M. G., Benjamin, E. J., Aragam, J., Mitchell, G. F. and Vasan, R. S. 2018, 'Left ventricular diastolic dysfunction in the community: impact of diagnostic criteria on the burden, correlates, and prognosis', *Journal of the American Heart Association* **7**(11), e008291.

597 Neff, L. S. and Bradshaw, A. D. 2021, 'Cross your heart? collagen cross-links in cardiac health and disease', *Cellular Signalling* **79**.

598 Neubauer, S. 2007, 'The failing heart—an engine out of fuel', *New England Journal of Medicine* **356**(11), 1140–1151.

599 Newcombe, R. G. 1998, 'Interval estimation for the difference between independent proportions: comparison of eleven methods', *Statistics in medicine* **17**(8), 873–890.

600 Newgard, C. B., An, J., Bain, J. R., Muehlbauer, M. J., Stevens, R. D., Lien, L. F., Haqq, A. M., Shah, S. H., Arlotto, M., Slentz, C. A. et al. 2009, 'A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance', *Cell metabolism* **9**(4), 311–326.

601 Nguyen, T.-M., Shafi, A., Nguyen, T. and Draghici, S. 2019, 'Identifying significantly impacted pathways: a comprehensive review and assessment', *Genome biology* **20**, 1–15.

602 Nicin, L., Wagner, J. U. G., Luxán, G. and Dimmeler, S. 2021, 'Fibroblast-mediated intercellular crosstalk in the healthy and diseased heart', *FEBS Letters* **596**.

603 Nishimura, T., Végvári, Á., Nakamura, H., Fujii, K., Sakai, H., Naruki, S., Furuya, N. and Saji, H. 2024, 'Cancer cell immunity-related protein co-expression networks are associated with early-stage solid-predominant lung adenocarcinoma', *Frontiers in Oncology* **14**, 1273780.

604 Noor, Z., Ahn, S. B., Baker, M. S., Ranganathan, S. and Mohamedali, A. 2021, 'Mass spectrometry–based protein identification in proteomics—a review', *Briefings in bioinformatics* **22**(2), 1620–1638.

605 Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A. V., Mikheenko, A., Vollger, M. R., Altemose, N., Uralsky, L., Gershman, A. et al. 2022, 'The complete sequence of a human genome', *Science* **376**(6588), 44–53.

606 Obici, L., Franceschini, G., Calabresi, L., Giorgetti, S., Stoppini, M., Merlini, G. and Bellotti, V. 2006, 'Structure, function and amyloidogenic propensity of apolipoprotein ai', *Amyloid* **13**(4), 191–205.

607 Oldfield, C. J., Duhamel, T. A. and Dhalla, N. S. 2020, 'Mechanisms for the transition from physiological to pathological cardiac hypertrophy', *Canadian Journal of Physiology and Pharmacology* **98**(2), 74–84.

608 Olsen, J. V., Ong, S.-E. and Mann, M. 2004, 'Trypsin cleaves exclusively c-terminal to arginine and lysine residues', *Molecular & cellular proteomics* **3**(6), 608–614.

609 Olshansky, G., Giles, C., Salim, A. and Meikle, P. J. 2022, 'Challenges and opportunities for prevention and removal of unwanted variation in lipidomic studies', *Progress in Lipid Research* **87**, 101177.

610 Olzmann, J. and Carvalho, P. 2018, 'Dynamics and functions of lipid droplets', *Nature Reviews Molecular Cell Biology* **20**(3).

611 Ordway, G. A. and Garry, D. J. 2004, 'Myoglobin: an essential hemoprotein in striated muscle', *Journal of Experimental Biology* **207**(20), 3441–3446.

612 Orea Soufi, A., Paik, J., Braganca, J., Donlon, T. A., Willcox, B. J. and Link, W. 2022, 'Foxo transcription factors as therapeutic targets in human diseases', *Trends in Pharmacological Sciences* **43**(12).

613 Ortiz Genga, M. F., Cuenca, S., Dal Ferro, M., Zorio, E., Salgado Aranda, R., Climent, V., Padron-Barthe, L., Duro Aguado, I., Jimenez Jaimez, J., Hidalgo Olivares, V. M., Garcia-Campo, E., Lanzillo, C., Suarez Mier, M. P., Yonath, H., Marcos Alonso, S., Ochoa, J. P., Santome, J. L., Garcia Giustiniani, D., Rodríguez-Garrido, J. L. and et al. 2016, 'Truncating flnc mutations are associated with high-risk dilated and arrhythmogenic cardiomyopathies', *Journal of the American College of Cardiology* **68**(22).

614 Ortmayr, K., Dubuis, S. and Zampieri, M. 2019, 'Metabolic profiling of cancer cells reveals genome-wide crosstalk between transcriptional regulators and metabolism', *Nature communications* **10**(1), 1841.

615 O'Toole, D., Zaeri, A. A. I., Nicklin, S. A., French, A. T., Loughrey, C. M. and Martin, T. P. 2020, 'Signalling pathways linking cysteine cathepsins to adverse cardiac remodelling', *Cellular Signalling* **76**, 109770.

616 Ouyang, W., An, Q., Zhao, J. and Qin, H. 2016, 'Integrating mean and variance heterogeneities to identify differentially expressed genes', *BMC bioinformatics* **17**, 1–14.

617 Oza, P. R., Agrawal, S., Ravaliya, D. and Kakkar, R. 2023, 'Evaluating the igraph community detection algorithms on different real networks', *Scalable Computing: Practice and Experience* **24**(2), 173–180.

618 Ozcebe, S. G., Bahcecioglu, G., Yue, X. S. and Zorlutuna, P. 2021, 'Effect of cellular and ecm aging on human ipsc-derived cardiomyocyte performance, maturity and senescence', *Biomaterials* **268**, 120554.

619 Ozsolak, F. and Milos, P. M. 2011, 'Single-molecule direct rna sequencing without cdna synthesis', *Wiley Interdisciplinary Reviews: RNA* **2**(4), 565–570.

620 O'Sullivan, J. F., Li, M., Koay, Y. C., Wang, X. S., Guglielmi, G., Marques, F. Z., Nanayakkara, S., Mariani, J., Slaughter, E. and Kaye, D. M. 2024, 'Cardiac substrate utilization and relationship to invasive exercise hemodynamic parameters in hfpef', *JACC: Basic to Translational Science* .

621 Packer, M. 2020, 'Longevity genes, cardiac ageing, and the pathogenesis of cardiomyopathy: implications for understanding the effects of current and future treatments for heart failure', *European Heart Journal* **41**(39).

622 Paczkowska, M., Barenboim, J., Sintupisut, N., Fox, N. S., Zhu, H., Abd-Rabbo, D., Mee, M. W., Boutros, P. C., Reimand, J. et al. 2020, 'Integrative pathway enrichment analysis of multivariate omics data', *Nature communications* **11**(1), 735.

623 Pagan, J., Seto, T., Pagano, M. and Cittadini, A. 2013, 'Role of the ubiquitin proteasome system in the heart', *Circulation research* **112**(7), 1046–1058.

624 Page, E. 1978, 'Quantitative ultrastructural analysis in cardiac membrane physiology', *American Journal of Physiology-Cell Physiology* **235**(5), C147–C158.

625 Palma, F. R., He, C., Danes, J. M., Paviani, V., Coelho, D. R., Gantner, B. N. and Bonini, M. G. 2020, 'Mitochondrial superoxide dismutase: what the established, the intriguing, and the novel reveal about a key cellular redox switch', *Antioxidants & redox signaling* .

626 Pan, B., Kusko, R., Xiao, W., Zheng, Y., Liu, Z., Xiao, C., Sakkiah, S., Guo, W., Gong, P., Zhang, C. et al. 2019, 'Similarities and differences between variants called with human reference genome hg19 or hg38', *BMC bioinformatics* **20**, 17–29.

627 Papaconstantinou, J. 2019, 'The role of signaling pathways of inflammation and oxidative stress in development of senescence and aging phenotypes in cardiovascular disease', *Cells* **8**(11).

628 Pararasa, C., Ikwuobe, J., Shigdar, S., Boukouvalas, A., Nabney, I. T., Brown, J. E., Devitt, A., Bailey, C. J., Bennett, S. J. and Griffiths, H. R. 2016, 'Age-associated changes in long-chain fatty acid profile during healthy aging promote pro-inflammatory monocyte polarization via ppar γ', *Aging cell* **15**(1), 128–139.

629 Park, T.-S. and Goldberg, I. J. 2012, 'Sphingolipids, lipotoxic cardiomyopathy, and cardiac failure', *Heart failure clinics* **8**(4), 633–641.

630 Park, W. J. and Oh, J. G. 2013, 'Serca2a: a prime target for modulation of cardiac contractility during heart failure', *BMB reports* **46**(5), 237.

631 Pascual, F. and Coleman, R. 2016, 'Fuel availability and fate in cardiac metabolism: A tale of two substrates', *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **1861**(10).

632 Patel, A., Abdelmalek, L., Thompson, A. and Jialal, I. 2020, 'Decreased homoserine levels in metabolic syndrome', *Diabetes & Metabolic Syndrome: Clinical Research & Reviews* **14**(4), 555–559.

633 Patel, H., Guttula, S., Mittal, R. S., Manwani, N., Berti-Equille, L. and Manatkar, A. 2022, Advances in exploratory data analysis, visualisation and quality for data centric ai systems, *in* 'Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining', pp. 4814–4815.

634 Patel, K. and Hawkins, P. 2015, 'Cardiac amyloidosis: where are we today?', *Journal of internal medicine* **278**(2), 126–144.

635 Patterson, S. D. and Aebersold, R. H. 2003, 'Proteomics: the first decade and beyond', *Nature genetics* **33**(3), 311–323.

636 Pavlopoulos, G. A., Secrier, M., Moschopoulos, C. N., Soldatos, T. G., Kossida, S., Aerts, J., Schneider, R. and Bagos, P. G. 2011, 'Using graph theory to analyze biological networks', *BioData mining* **4**, 1–27.

637 Peluso, A., Damgaard, M. V., Mori, M. A. and Treebak, J. T. 2021, 'Age-dependent decline of nad+—universal truth or confounded consensus?', *Nutrients* **14**(1), 101.

638 Perez de Souza, L., Alseekh, S., Scossa, F. and Fernie, A. R. 2021, 'Ultra-high-performance liquid chromatography high-resolution mass spectrometry variants for metabolomics research', *Nature Methods* **18**(7), 733–746.

639 Perez-Riverol, Y., Zorin, A., Dass, G., Vu, M.-T., Xu, P., Glont, M., Vizcaíno, J. A., Jarnuczak, A. F., Petryszak, R., Ping, P. et al. 2019, 'Quantifying the impact of public omics data', *Nature communications* **10**(1), 3512.

640 Periasamy, M., Reed, T. D., Liu, L. H., Ji, Y., Loukianov, E., Paul, R. J., Nieman, M. L., Riddle, T., Duffy, J. J., Doetschman, T. et al. 1999, 'Impaired cardiac performance in heterozygous mice with a null mutation in the sarco (endo) plasmic reticulum ca2+-atpase isoform 2 (serca2) gene', *Journal of Biological Chemistry* **274**(4), 2556–2562.

641 Perng, W. and Aslibekyan, S. 2020, 'Find the needle in the haystack, then find it again: Replication and validation in the 'omics era', *Metabolites* **10**(7), 286.

642 Pesarin, F. and Salmaso, L. 2010, 'The permutation testing approach: a review', *Statistica* **70**(4), 481–509.

643 Picca, A., Mankowski, R. T., Burman, J. L., Donisi, L., Kim, J.-S., Marzetti, E. and Leeuwenburgh, C. 2018, 'Mitochondrial quality control mechanisms as molecular targets in cardiac ageing', *Nature Reviews Cardiology* **15**(9), 543–554.

644 Piché, M.-E., Tchernof, A. and Després, J.-P. 2020, 'Obesity phenotypes, diabetes, and cardiovascular diseases', *Circulation research* **126**(11), 1477–1500.

645 Pico, A. 2010, 'Wikipathways: Community curation of biological pathways', *Nature Precedings* pp. 1–1.

646 Pico, A. R., Kelder, T., Van Iersel, M. P., Hanspers, K., Conklin, B. R. and Evelo, C. 2008, 'Wikipathways: pathway editing for the people', *PLoS biology* **6**(7), e184.

647 Pils, V., Terlecki-Zaniewicz, L., Schosserer, M., Grillari, J. and Lämmermann, I. 2021, 'The role of lipid-based signalling in wound healing and senescence', *Mechanisms of Ageing and Development* **198**.

648 Pingault, J.-B., O'reilly, P. F., Schoeler, T., Ploubidis, G. B., Rijsdijk, F. and Dudbridge, F. 2018, 'Using genetic data to strengthen causal inference in observational research', *Nature Reviews Genetics* **19**(9), 566–580.

649 Pinto, A. R., Ilinykh, A., Ivey, M. J., Kuwabara, J. T., D'Antoni, M. L., Debuque, R., Chandran, A., Wang, L., Arora, K., Rosenthal, N. A. and Tallquist, M. D. 2016, 'Revisiting cardiac cellular composition', *Circulation Research* **118**(3).

650 Piquereau, J., Boitard, S. E., Ventura-Clapier, R. and Mericskay, M. 2021, 'Metabolic therapy of heart failure: Is there a future for b vitamins?', *International Journal of Molecular Sciences* **23**(1), 30.

651 Pistritu, D.-V., Vasiliniuc, A.-C., Vasiliu, A., Visinescu, E.-F., Visoiu, I.-E., Vizdei, S., Martínez Anghel, P., Tanca, A., Bucur, O. and Liehn, E. A. 2023, 'Phospholipids, the masters in the shadows during healing after acute myocardial infarction', *International Journal of Molecular Sciences* **24**(9), 8360.

652 Pizzino, G., Irrera, N., Cucinotta, M., Pallio, G., Mannino, F., Arcoraci, V., Squadrito, F., Altavilla, D., Bitto, A. et al. 2017, 'Oxidative stress: harms and benefits for human health', *Oxidative medicine and cellular longevity* **2017**.

653 Podnar, J., Deiderick, H., Huerta, G. and Hunicke-Smith, S. 2014, 'Next-generation sequencing rna-seq library construction', *Current protocols in molecular biology* **106**(1), 4–21.

654 Pohl, S., Angermann, A., Jeschke, A., Hendrickx, G., Yorgan, T. A., Makrypidi-Fraune, G., Steigert, A., Kuehn, S. C., Rolvien, T., Schweizer, M. et al. 2018, 'The lysosomal protein arylsulfatase b is a key enzyme involved in skeletal turnover', *Journal of Bone and Mineral Research* **33**(12), 2186–2201.

655 Pomyen, Y., Segura, M., Ebbels, T. M. and Keun, H. C. 2015, 'Over-representation of correlation analysis (orca): a method for identifying associations between variable sets', *Bioinformatics* **31**(1), 102–108.

656 Pons, P. and Latapy, M. 2006, 'Computing communities in large networks using random walks.', *J. Graph Algorithms Appl.* **10**(2), 191–218.

657 Porter, K. E. and Turner, N. A. 2009, 'Cardiac fibroblasts: At the heart of myocardial remodeling', *Pharmacology & Therapeutics* **123**(2).

658 Potting, C., Tatsuta, T., König, T., Haag, M., Wai, T., Aaltonen, M. and Langer, T. 2013, 'Triap1/preli complexes prevent apoptosis by mediating intramitochondrial transport of phosphatidic acid', *Cell Metabolism* **18**(2).

659 Poulaki, A. and Giannouli, S. 2022, 'Mitochondrial lipids: From membrane organization to apoptotic facilitation', *International Journal of Molecular Sciences* **23**(7).

660 Prentki, M. and Madiraju, S. 2008, 'Glycerolipid metabolism and signaling in health and disease', *Endocrine Reviews* **29**(6).

661 Prill, R. J., Marbach, D., Saez-Rodriguez, J., Sorger, P. K., Alexopoulos, L. G., Xue, X., Clarke, N. D., Altan-Bonnet, G. and Stolovitzky, G. 2010, 'Towards a rigorous assessment of systems biology models: the dream3 challenges', *PloS one* **5**(2), e9202.

662 Priya Dharshini, L. C., Vishnupriya, S., Sakthivel, K. M. and Rasmi, R. R. 2020, 'Oxidative stress responsive transcription factors in cellular signalling transduction mechanisms', *Cellular Signalling* **72**.

663 Protasoni, M. and Zeviani, M. 2021, 'Mitochondrial structure and bioenergetics in normal and disease conditions', *International Journal of Molecular Sciences* **22**(2).

664 Prud'homme, G. J., Kurt, M. and Wang, Q. 2022, 'Pathobiology of the klotho antiaging protein and therapeutic considerations', *Frontiers in Aging* **3**, 931331.

665 Rahate, K., Bhatt, L. K. and Prabhavalkar, K. S. 2020, 'Serca stimulation: A potential approach in therapeutics', *Chemical Biology & Drug Design* **95**(1), 5–15.

666 Rahman, S. and Islam, R. 2011, 'Mammalian sirt1: insights on its biological functions', *Cell communication and signaling : CCS* **9**(11).

667 Rajakumari, S., Wu, J., Ishibashi, J., Lim, H.-W., Giang, A.-H., Won, K.-J., Reed, R. R. and Seale, P. 2013, 'Ebf2 determines and maintains brown adipocyte identity. cell metabolism', *Cell Metabolism* **17**(4).

668 Rakusan, K. and Nagai, J. 1994, 'Morphometry of arterioles and capillaries in hearts of senescent mice', *Cardiovascular Research* **28**(7).

669 Ramonfaur, D., Buckley, L. F., Arthur, V., Yang, Y., Claggett, B. L., Ndumele, C. E., Walker, K. A., Austin, T., Odden, M. C., Floyd, J. S. et al. 2024, 'High throughput plasma proteomics and risk of heart failure and frailty in late life', *JAMA cardiology* .

670 Rawlings, N. D., Barrett, A. J., Thomas, P. D., Huang, X., Bateman, A. and Finn, R. D. 2018, 'The merops database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the panther database', *Nucleic acids research* **46**(D1), D624–D632.

671 Rayaprolu, S., Higginbotham, L., Bagchi, P., Watson, C. M., Zhang, T., Levey, A. I., Rangaraju, S. and Seyfried, N. T. 2021, 'Systems-based proteomics to resolve the biology of alzheimer's disease beyond amyloid and tau', *Neuropsychopharmacology* **46**(1), 98–115.

672 Razquin, C., Ruiz-Canela, M., Toledo, E., Hernández-Alonso, P., Clish, C. B., Guasch-Ferré, M., Li, J., Wittenbecher, C., Dennis, C., Alonso-Gómez, A. et al. 2021, 'Metabolomics of the tryptophan–kynurenine degradation pathway and risk of atrial fibrillation and heart failure: potential modification effect of mediterranean diet', *The American journal of clinical nutrition* **114**(5), 1646–1654.

673 Redestig, H., Kobayashi, M., Saito, K. and Kusano, M. 2011, 'Exploring matrix effects and quantification performance in metabolomics experiments using artificial biological gradients', *Analytical chemistry* **83**(14), 5645–5651.

674 Reitz, C. J., Kuzmanov, U. and Gramolini, A. O. 2023, 'Multi-omic analyses and network biology in cardiovascular disease', *Proteomics* **23**(21-22), 2200289.

675 Resnik, D. B. and Shamoo, A. E. 2017, 'Reproducibility and research integrity', *Accountability in research* **24**(2), 116–123.

676 Révész, Á., Hevér, H., Steckel, A., Schlosser, G., Szabó, D., Vékey, K. and Drahos, L. 2023, 'Collision energies: Optimization strategies for bottom-up proteomics', *Mass spectrometry reviews* **42**(4), 1261–1299.

677 Rhee, J.-W., Ky, B., Armenian, S. H., Yancy, C. W. and Wu, J. C. 2020, 'Primer on biomarker discovery in cardio-oncology: application of omics technologies', *Cardio Oncology* **2**(3), 379–384.

678 Ribeiro, A. S. F., Zerolo, B. E., Lopez-Espuela, F., Sanchez, R. and Fernandes, V. S. 2023, 'Cardiac system during the aging process', *Aging and disease* **14**(4).

679 Richter, F., Morton, S. U., Kim, S. W., Kitaygorodsky, A., Wasson, L. K., Chen, K. M., Zhou, J., Qi, H., Patel, N., DePalma, S. R. et al. 2020, 'Genomic analyses implicate noncoding de novo variants in congenital heart disease', *Nature genetics* **52**(8), 769–777.

680 Ricketts, S. N. and Qian, L. 2022, 'The heart of cardiac reprogramming: The cardiac fibroblasts', *Journal of molecular and cellular cardiology* **172**.

681 Riffel, J. H., Keller, M. G. P., Rost, F., Arenja, N., Andre, F., aus dem Siepen, F., Fritz, T., Ehlermann, P., Taeger, T., Frankenstein, L., Meder, B., Katus, H. A. and Buss, S. J. 2016, 'Left ventricular long axis strain: a new prognosticator in non-ischemic dilated cardiomyopathy?', *Journal of Cardiovascular Magnetic Resonance* **18**(1).

682 Rigobello, M. P., Galzigna, L. and Bindoli, A. 1994, 'Fructose 1, 6-bisphosphate prevents oxidative stress in the isolated and perfused rat heart', *Cell Biochemistry and Function: Cellular biochemistry and its modulation by active agents or disease* **12**(1), 69–75.

683 Rinaldo, P., Matern, D. and Bennett, M. J. 2002, 'Fatty acid oxidation disorders', *Annual review of physiology* **64**(1), 477–502.

684 Rio, D. C., Ares, M., Hannon, G. J. and Nilsen, T. W. 2010, 'Purification of rna using trizol (tri reagent)', *Cold Spring Harbor Protocols* **2010**(6), pdb–prot5439.

685 Ritter, A., Leonard, J., Gray, C. S., Izumi, K., Levinson, K., Nair, D., O'Connor, M. J., Rossano, J. W., Shankar, V., Chowns, J. L., Marzolf, A., Owens, A. T. and Ahrens-Nicklas, R. C. 2022, 'Myh7 variants cause complex congenital heart disease', *American Journal of Medical Genetics* **188**(9).

686 Robinson, M. D., McCarthy, D. J. and Smyth, G. K. 2010, 'edger: a bioconductor package for differential expression analysis of digital gene expression data', *bioinformatics* **26**(1), 139–140.

687 Rocca-Serra, P., Salek, R. M., Arita, M., Correa, E., Dayalan, S., Gonzalez-Beltran, A., Ebbels, T., Goodacre, R., Hastings, J., Haug, K. et al. 2016, 'Data standards can boost metabolomics research, and if there is a will, there is a way', *Metabolomics* **12**, 1–13.

688 Rodríguez, A., Becerril, S., Hernández-Pardos, A. W. and Frühbeck, G. 2020, 'Adipose tissue depot differences in adipokines and effects on skeletal and cardiac muscle', *Current opinion in pharmacology* **52**, 1–8.

689 Roe, N., Handzlik, M., Li, T. and Tian, R. 2018, 'The role of diacylglycerol acyltransferase (dgat) 1 and 2 in cardiac metabolism and function', *Scientific Reports* **8**(1).

690 Rogers, S., Girolami, M., Kolch, W., Waters, K. M., Liu, T., Thrall, B. and Wiley, H. S. 2008, 'Investigating the correspondence between transcriptomic and proteomic expression profiles using coupled cluster models', *Bioinformatics* **24**(24), 2894–2900.

691 Rojas, P., Wang, J., Guglielmi, G., Sadurnì, M. M., Pavlou, L., Leung, G. H. D., Rajagopal, V., Spill, F. and Saponaro, M. 2024, 'Genome-wide identification of replication fork stalling/pausing sites and the interplay between rna pol ii transcription and dna replication progression', *Genome Biology* **25**.

692 Romano, J. P. and Wolf, M. 2017, 'Resurrecting weighted least squares', *Journal of Econometrics* **197**(1), 1–19.

693 Romeo, F. J., Mavropoulos, S. A. and Ishikawa, K. 2023, 'Progress in clinical gene therapy for cardiac disorders', *Molecular diagnosis & therapy* **27**(2), 179–191.

694 Rosano, G., Fini, M., Caminiti, G. and Barbaro, G. 2008, 'Cardiac metabolism in myocardial ischemia', *Current Pharmaceutical Design* **14**(25).

695 Rosca, M. G., Lemieux, H. and Hoppel, C. L. 2009, 'Mitochondria in the elderly: is acetylcarnitine a rejuvenator?', *Advanced drug delivery reviews* **61**(14), 1332–1342.

696 Rosca, M. G., Vazquez, E. J., Kerner, J., Parland, W., Chandler, M. P., Stanley, W., Sabbah, H. N. and Hoppel, C. L. 2008, 'Cardiac mitochondria in heart failure: decrease in respirasomes and oxidative phosphorylation', *Cardiovascular research* **80**(1), 30–39.

697 Rossi, D., Pierantozzi, E., Amadsun, D. O., Buonocore, S., Rubino, E. M. and Sorrentino, V. 2022, 'The sarcoplasmic reticulum of skeletal muscle cells: A labyrinth of membrane contact sites', *Biomolecules* **12**(4), 488.

698 Rossi, M. A., Abreu, M. A. and Santoro, L. B. 1998, 'Connective tissue skeleton of the human heart', *Circulation* **97**(9).

699 Rousseeuw, P. J. 1991, 'Tutorial to robust statistics', *Journal of chemometrics* **5**(1), 1–20.

700 Rousseeuw, P. J. and Hubert, M. 2011, 'Robust statistics for outlier detection', *Wiley interdisciplinary reviews: Data mining and knowledge discovery* **1**(1), 73–79.

701 Rowan, D. D. 2011, 'Volatile metabolites', *Metabolites* **1**(1), 41–63.

702 Ruan, J., Dean, A. and Zhang, W. 2010, 'A general co-expression network-based approach to gene expression analysis: comparison and applications', *BMC Systems Biology* **4**(1).

703 Rubattu, S., Forte, M. and Raffa, S. 2019, 'Circulating leukocytes and oxidative stress in cardiovascular diseases: A state of the art', *Oxidative Medicine and Cellular Longevity* **2019**.

704 Rubin, D. B. 1976, 'Inference and missing data', *Biometrika* **63**(3), 581–592.

705 Rudnicka, E., Napierala, P., Podfigurna, A., Mkeczekalski, B., Smolarczyk, R. and Grymowicz, M. 2020, 'The world health organization (who) approach to healthy ageing', *Maturitas* **139**, 6–11.

706 Rudt, E., Feldhaus, M., Margraf, C., Schlehuber, S., Schubert, A., Heuckeroth, S., Karst, U., Jeck, V., Meyer, S., Korf, A. et al. 2023, 'Comparison of data-dependent acquisition, data-independent acquisition, and parallel reaction monitoring in trapped ion mobility spectrometry–time-of-flight tandem mass spectrometry-based lipidomics', *Analytical Chemistry* **95**(25), 9488–9496.

707 Rudziak, P., Ellis, C. G. and Kowalewska, P. M. 2019, 'Role and molecular mechanisms of pericytes in regulation of leukocyte diapedesis in inflamed tissues', *Mediators of Inflammation* **2019**.

708 Rundell, V. L., Manaves, V., Martin, A. F. and de Tombe, P. P. 2005, 'Impact of β-myosin heavy chain isoform expression on cross-bridge cycling kinetics', *American Journal of Physiology-Heart and Circulatory Physiology* **288**(2), H896–H903.

709 Rysz, J., Franczyk, B., Rysz-Górzyńska, M. and Gluba-Brzózka, A. 2021, 'Ageing, age-related cardiovascular risk and the beneficial role of natural components intake', *International Journal of Molecular Sciences* **23**(1), 183.

710 Sahoo, S., Meijles, D. N. and Pagano, P. J. 2016, 'Nadph oxidases: key modulators in aging and age-related cardiovascular diseases?', *Clinical Science* **130**(5), 317–335.

711 Sakaguchi, A. and Kimura, W. 2021, 'Metabolic regulation of cardiac regeneration: roles of hypoxia, energy homeostasis, and mitochondrial dynamics', *Current opinion in genetics & development* **70**, 54–60.

712 Salazar, N. and Zabel, B. A. 2019, 'Support of tumor endothelial cells by chemokine receptors', *Frontiers in Immunology* **10**.

713 Saoi, M. and Britz-McKibbin, P. 2021, 'New advances in tissue metabolomics: A review', *Metabolites* **11**(10), 672.

714 Sarmad, S., Viant, M. R., Dunn, W. B., Goodacre, R., Wilson, I. D., Chappell, K. E., Griffin, J. L., O'Donnell, V. B., Naicker, B., Lewis, M. R. et al. 2023, 'A proposed framework to evaluate the quality and reliability of targeted metabolomics assays from the uk consortium on metabolic phenotyping (map/uk)', *Nature protocols* **18**(4), 1017–1027.

715 Savinova, O. V. and Gerdes, A. M. 2012, 'Myocyte changes in heart failure', *Heart Failure Clinics* **8**(1).

716 Schäfer, J. and Strimmer, K. 2005, 'An empirical bayes approach to inferring large-scale gene association networks', *Bioinformatics* **21**(6), 754–764.

717 Schessner, J. P., Voytik, E. and Bludau, I. 2022, 'A practical guide to interpreting and generating bottom-up proteomics data visualizations', *Proteomics* **22**(8), 2100103.

718 Schiano, C., Casamassimi, A., Vietri, M. T., Rienzo, M. and Napoli, C. 2014, 'The roles of mediator complex in cardiovascular diseases', *Biochimica et Biophysica Acta (BBA)-Gene Regulatory mechanisms* **1839**(6), 444–451.

719 Schielzeth, H. 2010, 'Simple means to improve the interpretability of regression coefficients', *Methods in Ecology and Evolution* **1**(2), 103–113.

720 Scholes, A. N. and Lewis, J. A. 2020, 'Comparison of rna isolation methods on rna-seq: implications for differential expression and meta-analyses', *BMC genomics* **21**, 1–9.

721 Scholtes, C. and Giguère, V. 2022, 'Transcriptional control of energy metabolism by nuclear receptors', *Nature Reviews Molecular Cell Biology* .

722 Schrader, S., Perfilyev, A., Martinell, M., García-Calzón, S. and Ling, C. 2021, 'Statin therapy is associated with epigenetic modifications in individuals with type 2 diabetes', *Epigenomics* **13**(12), 919–925.

723 Schumacher, B., Pothof, J., Vijg, J. and Hoeijmakers, J. H. 2021, 'The central role of dna damage in the ageing process', *Nature* **592**(7856), 695–703.

724 Sciarretta, S., Forte, M., Frati, G. and Sadoshima, J. 2018, 'New insights into the role of mtor signaling in the cardiovascular system', *Circulation research* **122**(3).

725 Sciarretta, S., Forte, M., Frati, G. and Sadoshima, J. 2022, 'The complex network of mtor signalling in the heart', *Cardiovascular research* **118**(2), 424–439.

726 Sedgwick, P. 2012, 'Log transformation of data', *BMJ* **345**.

727 Senatore, A., Monteil, A., van Minnen, J., Smit, A. and Spafford, J. 2013, 'Nalcn ion channels have alternative selectivity filters resembling calcium channels or sodium channels', *PLoS ONE* **8**(1).

728 Sengupta, A., Molkentin, J. D., Paik, J. H., DePinho, R. A. and Yutzey, K. E. 2011, 'Foxo transcription factors promote cardiomyocyte survival upon induction of oxidative stress', *Journal of Biological Chemistry* **286**(9).

729 Seper, V., Korovljev, D., Todorovic, N., Stajer, V., Ostojic, J., Nesic, N. and Ostojic, S. M. 2021, 'Guanidinoacetate-creatine supplementation improves functional performance and muscle and brain bioenergetics in the elderly: a pilot study', *Annals of Nutrition and Metabolism* **77**(4), 244–247.

730 Serdar, C. C., Cihan, M., Yücel, D. and Serdar, M. A. 2021, 'Sample size, power and effect size revisited: simplified and practical approaches in pre-clinical, clinical and laboratory studies', *Biochemia medica* **31**(1), 27–53.

731 Severs, N. J. 2000, 'The cardiac muscle cell', *Bioessays* **22**(2).

732 Seyhan, A. A. and Carini, C. 2019, 'Are innovation and new technologies in precision medicine paving a new era in patients centric care?', *Journal of translational medicine* **17**(1), 114.

733 Shah, K. B., Inoue, Y. and Mehra, M. R. 2006, 'Amyloidosis and the heart: a comprehensive review', *Archives of Internal Medicine* **166**(17), 1805–1813.

734 Shahjaman, M., Mollah, M. M. H., Rahman, M. R., Islam, S. S. and Mollah, M. N. H. 2020, 'Robust identification of differentially expressed genes from rna-seq data', *Genomics* **112**(2), 2000–2010.

735 Shaimardanova, A. A., Solovyeva, V. V., Issa, S. S. and Rizvanov, A. A. 2023, 'Gene therapy of sphingolipid metabolic disorders', *International Journal of Molecular Sciences* **24**(4), 3627.

736 Shamoo, A. 1987, 'Ensuring scientific integrity', *Nature* **327**(6123), 550.

737 Shannon, O., Ashor, A., Scialo, F., Saretzki, G., Martin-Ruiz, C., Lara, J., Matu, J., Griffiths, A., Robinson, N., Lillà, L., Stevenson, E., Stephan, B., Minihane, A., Siervo, M. and Mathers, J. 2021, 'Mediterranean diet and the hallmarks of ageing', *European Journal of Clinical Nutrition* **75**(8).

738 Shao, D. and Tian, R. 2015, 'Glucose transporters in cardiac metabolism and hypertrophy', *Comprehensive Physiology* **6**(1).

739 Sharma, L. K., Lu, J. and Bai, Y. 2009, 'Mitochondrial respiratory complex i: structure, function and implication in human diseases', *Current medicinal chemistry* **16**(10), 1266–1277.

740 Sharma, M. and Bhatt, L. K. 2022, 'Emerging therapeutic targets for heart failure', *Current Cardiology Reports* **24**(11), 1737–1754.

741 She, L., Tu, H., Zhang, Y. Z., Tang, L., Li, N. S., Lin, Q., Liu, B., Li, Q., Luo, X. and Peng, J. 2019, 'Inhibition of phosphoglycerate mutase 5 reduces necroptosis in rat hearts following ischemia/reperfusion through suppression of dynamin-related protein 1', *Cardiovascular Drugs and Therapy* **33**(1).

742 Shen, X., Wang, C., Zhou, X., Zhou, W., Hornburg, D., Wu, S. and Snyder, M. P. 2024, 'Nonlinear dynamics of multi-omics profiles during human aging', *Nature aging* pp. 1–16.

743 Shen, Y., Dinh, H. V., Cruz, E. R., Chen, Z., Bartman, C. R., Xiao, T., Call, C. M., Ryseck, R.-P., Pratas, J., Weilandt, D. et al. 2024, 'Mitochondrial atp generation is more proteome efficient than glycolysis', *Nature Chemical Biology* pp. 1–10.

744 Shestov, A. A., Liu, X., Ser, Z., Cluntun, A. A., Hung, Y. P., Huang, L., Kim, D., Le, A., Yellen, G., Albeck, J. G. and Locasale, J. W. 2014, 'Quantitative determinants of aerobic glycolysis identify flux through the enzyme gapdh as a limiting step', *ELife* **3**.

745 Shevchenko, A. and Simons, K. 2010, 'Lipidomics: coming to grips with lipid diversity', *Nature reviews Molecular cell biology* **11**(8), 593–598.

746 Shi, D.-y., Xie, F.-z., Zhai, C., Stern, J. S., Liu, Y. and Liu, S.-l. 2009, 'The role of cellular oxidative stress in regulating glycolysis energy metabolism in hepatoma cells', *Molecular cancer* **8**, 1–15.

747 Shiels, H. A. and Galli, G. L. 2014, 'The sarcoplasmic reticulum and the evolution of the vertebrate heart', *Physiology* **29**(6), 456–469.

748 Shih, H., Lee, B., Lee, R. J. and Boyle, A. J. 2011, 'The aging heart and post-infarction left ventricular remodeling', *Journal of the American College of Cardiology* **57**(1), 9–17.

749 Shimizu, N., Yoshiyama, M., Omura, T., Hanatani, A., Kim, S., Takeuchi, K., Iwao, H. and Yoshikawa, J. 1998, 'Activation of mitogen-activated protein kinases and activator protein-1 in myocardial infarction in rats', *Cardiovascular Research* **38**.

750 Shu, H., Peng, Y., Hang, W., Li, N., Zhou, N. and Wang, D. W. 2022, 'Emerging roles of ceramide in cardiovascular diseases', *Aging and disease* **13**(1), 232.

751 Siepen, J. A., Keevil, E.-J., Knight, D. and Hubbard, S. J. 2007, 'Prediction of missed cleavage sites in tryptic peptides aids protein identification in proteomics', *Journal of proteome research* **6**(1), 399–408.

752 Sies, H., Belousov, V. V., Chandel, N. S., Davies, M. J., Jones, D. P., Mann, G. E., Murphy, M. P., Yamamoto, M. and Winterbourn, C. 2022, 'Defining roles of specific reactive oxygen species (ros) in cell biology and physiology', *Nature Reviews Molecular Cell Biology* **23**(7).

753 Signes, A. and Fernandez-Vizarra, E. 2018, 'Assembly of mammalian oxidative phosphorylation complexes i–v and supercomplexes', *Essays in Biochemistry* **62**(3).

754 Silveira Rossi, J. L., Barbalho, S. M., Reverete de Araujo, R., Bechara, M. D., Sloan, K. P. and Sloan, L. A. 2022, 'Metabolic syndrome and cardiovascular diseases: Going beyond traditional risk factors', *Diabetes/Metabolism Research and Reviews* **38**(3), e3502.

755 Silverman, J. D., Roche, K., Mukherjee, S. and David, L. A. 2020, 'Naught all zeros in sequence count data are the same', *Computational and structural biotechnology journal* **18**, 2789–2798.

756 Simoes, I. C., Morciano, G., Lebiedzinska-Arciszewska, M., Aguiari, G., Pinton, P., Potes, Y. and Wieckowski, M. R. 2020, 'The mystery of mitochondria-er contact sites in physiology and pathology: A cancer perspective', *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease* **1866**(10), 165834.

757 Sindelar, M. and Patti, G. J. 2020, 'Chemical discovery in the era of metabolomics', *Journal of the American Chemical Society* **142**(20), 9097–9105.

758 Singh, P. P. and Benayoun, B. A. 2023, 'Considerations for reproducible omics in aging research', *Nature Aging* **3**(8), 921–930.

759 Singh, U., Hur, M., Dorman, K. and Wurtele, E. S. 2020, 'Metaomgraph: a workbench for interactive exploratory data analysis of large expression datasets', *Nucleic acids research* **48**(4), e23–e23.

760 Sissaoui, S., Yu, J., Yan, A., Li, R., Yukselen, O., Kucukural, A., Zhu, L. and Lawson, N. D. 2020, 'Genomic characterization of endothelial enhancers reveals a multifunctional role for nr2f2 in regulation of arteriovenous gene expression', *Circulation Research* **126**(7).

761 Slavic, S., Andrukhova, O., Ford, K., Handschuh, S., Latic, N., Reichart, U., Sasgary, S., Bergow, C., Hofbauer, L. C., Kostenuik, P. J. et al. 2018, 'Selective inhibition of receptor activator of nf-κb ligand (rankl) in hematopoietic cells improves outcome after experimental myocardial infarction', *Journal of Molecular Medicine* **96**, 559–573.

762 Slenter, D. N., Kutmon, M., Hanspers, K., Riutta, A., Windsor, J., Nunes, N., Mélius, J., Cirillo, E., Coort, S. L., Digles, D. et al. 2018, 'Wikipathways: a multifaceted pathway database bridging metabolomics to other omics research', *Nucleic acids research* **46**(D1), D661–D667.

763 Smilde, A. K., van der Werf, M. J., Bijlsma, S., van der Werff-van der Vat, B. J. and Jellema, R. H. 2005, 'Fusion of mass spectrometry-based metabolomics data', *Analytical chemistry* **77**(20), 6729–6736.

764 Smyth, G. K. 2004, 'Linear models and empirical bayes methods for assessing differential expression in microarray experiments', *Statistical applications in genetics and molecular biology* **3**(1).

765 Smyth, G. K. and Speed, T. 2003, 'Normalization of cdna microarray data', *Methods* **31**(4), 265–273.

766 Sobsey, C. A., Ibrahim, S., Richard, V. R., Gaspar, V., Mitsa, G., Lacasse, V., Zahedi, R. P., Batist, G. and Borchers, C. H. 2020, 'Targeted and untargeted proteomics approaches in biomarker development', *Proteomics* **20**(9), 1900029.

767 Soh, D., Dong, D., Guo, Y. and Wong, L. 2010, 'Consistency, comprehensiveness, and compatibility of pathway databases', *BMC bioinformatics* **11**, 1–16.

768 Song, L., Langfelder, P. and Horvath, S. 2012, 'Comparison of co-expression measures: mutual information, correlation, and model based indices', *BMC bioinformatics* **13**, 1–21.

769 Sousa Fialho, M. d. L., Abd Jamil, A. H., Stannard, G. A. and Heather, L. C. 2019, 'Hypoxia-inducible factor 1 signalling, metabolism and its therapeutic potential in cardiovascular disease', *Biochimica et Biophysica Acta (BBA)* **1865**(4).

770 Southam, A. D., Weber, R. J., Engel, J., Jones, M. R. and Viant, M. R. 2017, 'A complete workflow for high-resolution spectral-stitching nanoelectrospray direct-infusion mass-spectrometry-based metabolomics and lipidomics', *Nature protocols* **12**(2), 310–328.

771 Speakman, J. R., Talbot, D. A., Selman, C., Snart, S., McLaren, J. S., Redman, P., Krol, E., Jackson, D. M., Johnson, M. S. and Brand, M. D. 2004, 'Uncoupled and surviving: individual mice with high metabolism have greater mitochondrial uncoupling and live longer', *Aging Cell* **3**(3).

772 Spicer, R. A., Salek, R. and Steinbeck, C. 2017, 'A decade after the metabolomics standards initiative it's time for a revision', *Scientific Data* **4**(1), 1–3.

773 Spinozzi, S., Liu, C., Chen, Z., Feng, W., Zhang, L., Ouyang, K., Evans, S. M. and Chen, J. 2020, 'Nexilin is necessary for maintaining the transverse-axial tubular system in adult cardiomyocytes', *Circulation: Heart Failure* **13**(7).

774 Stallone, G., Infante, B., Prisciandaro, C. and Grandaliano, G. 2019, 'mtor and aging: An old fashioned dress', *International Journal of Molecular Sciences* **20**(11).

775 Stanley, W. C., Recchia, F. A. and Lopaschuk, G. D. 2005, 'Myocardial substrate metabolism in the normal and failing heart', *Physiological Reviews* **85**(3).

776 Stark, R., Grzelak, M. and Hadfield, J. 2019, 'Rna sequencing: the teenage years', *Nature Reviews Genetics* **20**(11), 631–656.

777 Starling, R. C., Hammer, D. F. and Altschuld, R. A. 1998, 'Human myocardial atp content and in vivo contractile function', *Molecular and cellular biochemistry* **180**, 171–177.

778 Stefan, A. M. and Schönbrodt, F. D. 2023, 'Big little lies: A compendium and simulation of p-hacking strategies', *Royal Society Open Science* **10**(2), 220346.

779 Sterne, J. A., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., Wood, A. M. and Carpenter, J. R. 2009, 'Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls', *Bmj* **338**.

780 Stobdan, T., Zhou, D., Williams, A. T., Cabrales, P. and Haddad, G. G. 2018, 'Cardiac-specific knockout and pharmacological inhibition of endothelin receptor type b lead to cardiac resistance to extreme hypoxia', *Journal of Molecular Medicine* **96**(9).

781 Strait, J. B. and Lakatta, E. G. 2012, 'Aging-associated cardiovascular changes and their relationship to heart failure', *Heart failure clinics* **8**(1), 143–164.

782 Strub, G., Paillard, M., Liang, J., Gomez, L., Allegood, J., Hait, N., Maceyka, M., Price, M., Chen, Q., Simpson, D., Kordula, T., Milstien, S., Lesnefsky, E. and Spiegel, S. 2010, 'Sphingosine-1-phosphate produced by sphingosine kinase 2 in mitochondria interacts with prohibitin 2 to regulate complex iv assembly and respiration', *The FASEB Journal* **25**(2).

783 Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S. et al. 2005, 'Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles', *Proceedings of the National Academy of Sciences* **102**(43), 15545–15550.

784 Subramanian, I., Verma, S., Kumar, S., Jere, A. and Anamika, K. 2020, 'Multi-omics data integration, interpretation, and its application', *Bioinformatics and Biology Insights* **14**.

785 Suciu-Petrescu, M., Truta, A., Suciu, M. D., Trifa, A. P., Petrescu, D., Rosianu, H. S., Sabin, O., Popa, D. E., Macarie, A. E., Vesa, S. C. et al. 2021, 'Clinical impact of echocardiography parameters and molecular biomarkers in heart failure: Correlation of ace2 and mcp-1 polymorphisms with echocardiography parameters: A comparative study', *Experimental and Therapeutic Medicine* **22**(1), 1–9.

786 Sugihara, R., Taneike, M., Murakawa, T., Tamai, T., Ueda, H., Kitazume-Taneike, R., Oka, T., Akazawa, Y., Nishida, H., Mine, K. et al. 2023, 'Lysophosphatidylserine induces necrosis in pressure overloaded male mouse hearts via g protein coupled receptor 34', *Nature communications* **14**(1), 4494.

787 Sullivan, G. M. and Feinn, R. 2012, 'Using effect size—or why the p value is not enough', *Journal of graduate medical education* **4**(3), 279–282.

788 Suomi, T. and Elo, L. L. 2017, 'Enhanced differential expression statistics for data-independent acquisition proteomics', *Scientific reports* **7**(1), 5869.

789 Supplitt, S., Karpinski, P., Sasiadek, M. and Laczmanska, I. 2021, 'Current achievements and applications of transcriptomics in personalized cancer medicine', *International Journal of Molecular Sciences* **22**(3), 1422.

790 Suratannon, N., Van Wijck, R. T., Broer, L., Xue, L., Barendregt, B. H., van der Burg, M., Dik, W. A., Langerak, A. W., Dalm, V. A., Suphapeetiporn, K. et al. 2020, 'Rapid low-cost microarray-based genotyping for genetic screening in primary immunodeficiency', *Frontiers in immunology* **11**, 497722.

791 Sutherland, H. and Bickmore, W. A. 2009, 'Transcription factories: gene expression in unions?', *Nature Reviews Genetics* **10**(7), 457–466.

792 Sutton, J. P., Yen Ho, S. and Anderson, R. H. 1995, 'The forgotten interleaflet triangles: A review of the surgical anatomy of the aortic valve', *The Annals of Thoracic Surgery* **59**(2).

793 Swirski, F. K. and Nahrendorf, M. 2013, 'Leukocyte behavior in atherosclerosis, myocardial infarction, and heart failure', *Science (New York, N.Y.)* **339**(6116).

794 Syme, C., Czajkowski, S., Shin, J., Abrahamowicz, M., Leonard, G., Perron, M., Richer, L., Veillette, S., Gaudet, D., Strug, L. et al. 2016, 'Glycerophosphocholine metabolites and cardiovascular disease risk factors in adolescents: a cohort study', *Circulation* **134**(21), 1629–1636.

795 Szmit, S., Jank, M., Maciejewski, H., Grabowski, M., Glowczynska, R., Majewska, A., Filipiak, K., Motyl, T. and Opolski, G. 2010, 'Gene expression profiling in peripheral blood nuclear cells in patients with refractory ischaemic end-stage heart failure', *Journal of applied genetics* **51**, 353–368.

796 Taber, L. A., Yang, M. and Podszus, W. W. 1996, 'Mechanics of ventricular torsion', *Journal of Biomechanics* **26**(6).

797 Tacconi, C., He, Y., Ducoli, L. and Detmar, M. 2020, 'Epigenetic regulation of the lineage specificity of primary human dermal lymphatic and blood vascular endothelial cells', *Angiogenesis* **24**.

798 Taffet, G. E. and Tate, C. A. 1993, 'Caatpase content is lower in cardiac sarcoplasmic reticulum isolated from old rats', *American Journal of Physiology-Heart and Circulatory Physiology* **264**(5), H1609–H1614.

799 Tahir, U. A. and Gerszten, R. E. 2020, 'Omics and cardiometabolic disease risk prediction', *Annual review of medicine* **71**, 163–175.

800 Talayero, B. G. and Sacks, F. M. 2011, 'The role of triglycerides in atherosclerosis', *Current cardiology reports* **13**, 544–552.

801 Tamura, Y., Kohno, H., Mohri, T., Fujio, Y. and Matsumiya, G. 2018, 'The cardioprotective effect of interleukin-11 against ischemia-reperfusion injury in a heart donor model', *Annals of Cardiothoracic Surgery* **7**(1), 99.

802 Tang, H.-Y., Chen, A.-Q., Zhang, H., Gao, X.-F., Kong, X.-Q. and Zhang, J.-J. 2022, 'Vascular smooth muscle cells phenotypic switching in cardiovascular diseases', *Cells* **11**(24), 4060.

803 Tang, J., Yang, Q., Cui, Q., Zhang, D., Kong, D., Liao, X., Ren, J., Gong, Y. and Wu, G. 2019, 'Weighted gene correlation network analysis identifies rsad2, herc5, and ccl8 as prognostic candidates for breast cancer', *Journal of Cellular Physiology* **235**(1).

804 Tarazona, S., Arzalluz-Luque, A. and Conesa, A. 2021, 'Undisclosed, unmet and neglected challenges in multi-omics studies', *Nature Computational Science* **1**(6), 395–402.

805 Tarca, A. L., Romero, R., Xu, Z., Gomez-Lopez, N., Erez, O., Hsu, C.-D., Hassan, S. S. and Carey, V. J. 2019, 'Targeted expression profiling by rna-seq improves detection of cellular dynamics during pregnancy and identifies a role for t cells in term parturition', *Scientific reports* **9**(1), 848.

806 Tatsuta, T., Scharwey, M. and Langer, T. 2014, 'Mitochondrial lipid trafficking', *Trends in Cell Biology* **24**(1).

807 Tavazoie, S., Hughes, J. D., Campbell, M. J., Cho, R. J. and Church, G. M. 1999, 'Systematic determination of genetic network architecture', *Nature genetics* **22**(3), 281–285.

808 Taylor, S. L., Leiserowitz, G. S. and Kim, K. 2013, 'Accounting for undetected compounds in statistical analyses of mass spectrometry 'omic studies', *Statistical applications in genetics and molecular biology* **12**(6), 703–722.

809 Templeton, E., Lassé, M., Cameron, V., Pickering, J., Kleffmann, T., Hertzberg, D., Frampton, C., Ellmers, L., Palmer, S., Davidson, T. et al. 2022, 'Harnessing the power of "omics": Discovering novel microrna and protein biomarkers of acute kidney injury in acute heart failure', *Heart, Lung and Circulation* **31**, S23–S24.

810 Terman, A., Kurz, T., Gustafsson, B. and Brunk, U. T. 2008, 'The involvement of lysosomes in myocardial aging and disease', *Current cardiology reviews* **4**(2), 107–115.

811 Teufel, M. and Sobetzko, P. 2022, 'Reducing costs for dna and rna sequencing by sample pooling using a metagenomic approach', *BMC genomics* **23**(1), 613.

812 Tham, Y. K., Bernardo, B. C., Huynh, K., Ooi, J. Y., Gao, X. M., Kiriazis, H., Giles, C., Meikle, P. J. and McMullen, J. R. 2018, 'Lipidomic profiles of the heart and circulation in response to exercise versus cardiac pathology: a resource of potential biomarkers and drug targets', *Cell reports* **24**(10), 2757–2772.

813 Tharrey, M., Mariotti, F., Mashchak, A., Barbillon, P., Delattre, M., Huneau, J.-F. and Fraser, G. E. 2020, 'Patterns of amino acid intake are strongly associated with cardiovascular mortality, independently of the sources of protein', *International Journal of Epidemiology* **49**(1), 312–321.

814 Thomas, P. D. 2017, 'The gene ontology and the meaning of biological function', *The gene ontology handbook* pp. 15–24.

815 Thompson-Branch, A. and Havranek, T. 2017, 'Neonatal hypoglycemia', *Pediatrics in Review* **38**(4).

816 Ting, C. F., Harun, S., Daud, K. M., Sulaiman, S. and Muhammad, N. A. N. 2024, 'Exploring gene regulation and biological processes in insects: Insights from omics data using gene regulatory network models', *Progress in Biophysics and Molecular Biology* .

817 Tobias, D. K., Lawler, P. R., Harada, P. H., Demler, O. V., Ridker, P. M., Manson, J. E., Cheng, S. and Mora, S. 2018, 'Circulating branched-chain amino acids and incident cardiovascular disease in a prospective cohort of us women', *Circulation: Genomic and Precision Medicine* **11**(4), e002157.

818 Todorov, V. and Filzmoser, P. 2013, Comparing classical and robust sparse pca, *in* 'Synergies of soft computing and statistics for intelligent data analysis', Springer, pp. 283–291.

819 Tokareva, A. O., Chagovets, V. V., Kononikhin, A. S., Starodubtseva, N. L., Nikolaev, E. N. and Frankevich, V. E. 2021, 'Normalization methods for reducing interbatch effect without quality control samples in liquid chromatography-mass spectrometry-based studies', *Analytical and Bioanalytical Chemistry* **413**, 3479–3486.

820 Tominaga, K. and Suzuki, H. I. 2019, 'Tgf-β signaling in cellular senescence and aging-related pathology', *International Journal of Molecular Sciences* **20**(20).

821 Tonazzi, A., Giangregorio, N., Console, L., Palmieri, F. and Indiveri, C. 2021, 'The mitochondrial carnitine acyl-carnitine carrier (slc25a20): Molecular mechanisms of transport, role in redox sensing and interaction with drugs', *Biomolecules* **11**(4).

822 Torday, J. S. 2015, 'Homeostasis as the mechanism of evolution', *Biology* **4**(3).

823 Torrent-Guasp, F., Buckberg, G. D., Clemente, C., Cox, J. L., Coghlan, H. C. and Gharib, M. 2001, 'The structure and function of the helical heart and its buttress wrapping. i. the normal macroscopic structure of the heart', *Seminars in Thoracic and Cardiovascular Surgery* **13**(4).

824 Toubiana, D., Puzis, R., Sadka, A. and Blumwald, E. 2019, 'A genetic algorithm to optimize weighted gene co-expression network analysis', *Journal of Computational Biology* **26**(12), 1349–1366.

825 Townsend, N., Kazakiewicz, D., Lucy Wright, F., Timmis, A., Huculeci, R., Torbica, A., Gale, C. P., Achenbach, S., Weidinger, F. and Vardas, P. 2022, 'Epidemiology of cardiovascular disease in europe', *Nature Reviews Cardiology* **19**(2), 133–143.

826 Tracy, R. E. 2012, 'Cardiomyocyte size estimated from noninvasive measurements of left ventricular wall thickness and chamber diameter', *Journal of the American Society of Hypertension* **6**(3), 185–192.

827 Tran, K., Loiselle, D. S. and Crampin, E. J. 2015, 'Regulation of cardiac cellular bioenergetics: mechanisms and consequences', *Physiological reports* **3**(7), e12464.

828 Trifunović-Zamaklar, D., Jovanović, I., Vratonjić, J., Petrović, O., Paunović, I., Tešić, M., Boričić-Kostić, M. and Ivanović, B. 2022, 'The basic heart anatomy and physiology from the cardiologist's perspective: Toward a better understanding of left ventricular mechanics, systolic, and diastolic function', *Journal of clinical ultrasound : JCU* **50**(8).

829 Triner, D., Castillo, C., Hakim, J. B., Xue, X., Greenson, J. K., Nuñez, G., Chen, G. Y., Colacino, J. A. and Shah, Y. M. 2018, 'Myc-associated zinc finger protein regulates the proinflammatory response in colitis and colon cancer via stat3 signaling', *Molecular and cellular biology* .

830 Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D. and Altman, R. B. 2001, 'Missing value estimation methods for dna microarrays', *Bioinformatics* **17**(6), 520–525.

831 Tsai, Y. S., Scherl, A., Shaw, J. L., MacKay, C. L., Shaffer, S. A., Langridge-Smith, P. R. and Goodlett, D. R. 2009, 'Precursor ion independent algorithm for top-down shotgun proteomics', *Journal of the American Society for Mass Spectrometry* **20**, 2154–2166.

832 Tsao, C., Gona, P., Salton, C., Chuang, M., Levy, D., Manning, W. and O'Donnell, C. 2015, 'Left ventricular structure and risk of cardiovascular events: A framingham heart study cardiac magnetic resonance study', *Journal of the American Heart Association* **4**(9).

833 Tucker, N. R., Chaffin, M., Fleming, S. J., Hall, A. W., Parsons, V. A., Bedi, K. C., Akkad, A. D., Herndon, C. N., Arduini, A., Papangeli, I., Roselli, C., Aguet, F., Choi, S. H., Ardlie, K. G., Babadi, M., Margulies, K. B., Stegmann, C. M. and Ellinor, P. T. 2020, 'Transcriptional and cellular diversity of the human heart', *Circulation* **5**(142).

834 Turner, N., Lim, X. Y., Toop, H. D., Osborne, B., Brandon, A. E., Taylor, E. N., Fiveash, C. E., Govindaraju, H., Teo, J. D., McEwen, H. P. et al. 2018, 'A selective inhibitor of ceramide synthase 1 reveals a novel role in fat metabolism', *Nature communications* **9**(1), 3165.

835 Turpin-Nolan, S. M. and Brüning, J. C. 2020, 'The role of ceramides in metabolic disorders: when size and localization matters', *Nature Reviews Endocrinology* **16**(4), 224–233.

836 Tusher, V. G., Tibshirani, R. and Chu, G. 2001, 'Significance analysis of microarrays applied to the ionizing radiation response', *Proceedings of the National Academy of Sciences* **98**(9), 5116–5121.

837 Ulmer, C. Z., Koelmel, J. P., Jones, C. M., Garrett, T. J., Aristizabal-Henao, J. J., Vesper, H. W. and Bowden, J. A. 2021, 'A review of efforts to improve lipid stability during sample preparation and standardization efforts to ensure accuracy in the reporting of lipid measurements', *Lipids* **56**(1), 3–16.

838 Vaduganathan, M., Mensah, G. A., Turco, J. V., Fuster, V. and Roth, G. A. 2022, 'The global burden of cardiovascular diseases and risk: a compass for future health'.

839 Vale, G., Martin, S. A., Mitsche, M. A., Thompson, B. M., Eckert, K. M. and McDonald, J. G. 2019, 'Three-phase liquid extraction: a simple and fast method for lipidomic workflows [s]', *Journal of lipid research* **60**(3), 694–706.

840 Välikangas, T., Suomi, T. and Elo, L. L. 2018, 'A systematic evaluation of normalization methods in quantitative label-free proteomics', *Briefings in bioinformatics* **19**(1), 1–11.

841 Van den Berg, R. A., Hoefsloot, H. C., Westerhuis, J. A., Smilde, A. K. and Van der Werf, M. J. 2006, 'Centering, scaling, and transformations: improving the biological information content of metabolomics data', *BMC genomics* **7**, 1–15.

842 van Dijk, S. 2020, 'The sydney heart bank: a core facility that goes the extra mile to advance cardiac research', *Biophysical Reviews* **12**(4).

843 Van Eyk, J. E. and Snyder, M. P. 2018, 'Precision medicine: role of proteomics in changing clinical management and care', *Journal of proteome research* **18**(1), 1–6.

844 van Meer, G. 2005, 'Cellular lipidomics', *The EMBO Journal* **24**(18).

845 Van Tassell, B. W., Toldo, S., Mezzaroma, E. and Abbate, A. 2013, 'Targeting interleukin-1 in heart disease', *Circulation* **128**(17), 1910–1923.

846 Van Walree, E. S., Dombrowsky, G., Jansen, I. E., Mirkov, M. U., Zwart, R., Ilgun, A., Guo, D., Clur, S. B., Amin, A. S., Savage, J. E., Van der Wal, A. C., Waisfisz, Q., Maugeri, A., Wilsdon, A., Bu'Lock, F. A., Hurles, M. E., Dittrich, S., Berger, F., Audain Martinez, E., Christoffels, V. M., Marc-Philip Hitz, M. P., Milewicz, D. M., Posthuma, D., Meijers-Heijboer, H. and Mathijssen, I. B. 2021, 'Germline variants in hey2 functional domains lead to congenital heart defects and thoracic aortic aneurysms', *Genetics in Medicine* **23**(1).

847 Vance, J. 2014, 'Mam (mitochondria-associated membranes) in mammalian cells: Lipids and beyond', *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **1841**(4).

848 Vandermarliere, E., Mueller, M. and Martens, L. 2013, 'Getting intimate with trypsin, the leading protease in proteomics', *Mass spectrometry reviews* **32**(6), 453–465.

849 Väremo, L., Nielsen, J. and Nookaew, I. 2013, 'Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods', *Nucleic acids research* **41**(8), 4378–4391.

850 Veliova, M., Petcherski, A., Liesa, M. and Shirihai, O. S. 2020, The biology of lipid droplet-bound mitochondria, *in* 'Seminars in cell & developmental biology', Vol. 108, Elsevier, pp. 55–64.

851 Ventura, R., Martínez-Ruiz, I. and Hernández-Alvarez, M. 2022, 'Phospholipid membrane transport and associated diseases', *Biomedicines* **10**(5).

852 Vigil-Garcia, M., Demkes, C. J., Eding, J. E., Versteeg, D., de Ruiter, H., Perini, I., Kooijman, L., Gladka, M. M., Asselbergs, F. W., Vink, A. et al. 2021, 'Gene expression profiling of hypertrophic cardiomyocytes identifies new players in pathological remodelling', *Cardiovascular research* **117**(6), 1532–1545.

853 Vilchez, D., Saez, I. and Dillin, A. 2014, 'The role of protein clearance mechanisms in organismal ageing and age-related diseases', *Nature communications* **5**(1), 5659.

854 Virgiliou, C., Gika, H. G. and Theodoridis, G. A. 2018, 'Hilic-ms/ms multi-targeted method for metabolomics applications', *Metabolic Profiling: Methods and Protocols* pp. 65–81.

855 Visser, E. A., Moons, S. J., Timmermans, S. B., de Jong, H., Boltje, T. J. and Büll, C. 2021, 'Sialic acid o-acetylation: From biosynthesis to roles in health and disease', *Journal of Biological Chemistry* **297**(2).

856 Vonk-Noordegraaf, A. and Westerhof, N. 2013, 'Describing right ventricular function', *European Respiratory Journal* **41**(6), 1419–1423.

857 Vovk, V. and Wang, R. 2020, 'Combining p-values via averaging', *Biometrika* **107**(4), 791–808.

858 Vujic, A., Koo, A. N., Prag, H. A. and Krieg, T. 2021, 'Mitochondrial redox and tca cycle metabolite signaling in the heart', *Free Radical Biology and Medicine* **166**, 287–296.

859 Wang, A. Y., Epperson, W., DeLine, R. A. and Drucker, S. M. 2022, Diff in the loop: Supporting data comparison in exploratory data analysis, *in* 'Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems', pp. 1–10.

860 Wang, G., Zou, R., Liu, L., Wang, Z., Zou, Z., Tan, S., Xu, W. and Fan, X. 2022, 'A circular network of purine metabolism as coregulators of dilated cardiomyopathy', *Journal of translational medicine* **20**(1), 532.

861 Wang, L., Chen, P. and Xiao, W. 2021, 'β-hydroxybutyrate as an anti-aging metabolite', *Nutrients* **13**(10), 3420.

862 Wang, M., Pan, W., Xu, Y., Zhang, J., Wan, J. and Jiang, H. 2022, 'Microglia-mediated neuroinflammation: a potential target for the treatment of cardiovascular diseases', *Journal of inflammation research* pp. 3083–3094.

863 Wang, W., Wang, J., Yao, K., Wang, S., Nie, M., Zhao, Y., Wang, B., Pang, H., Xu, J., Wu, G. et al. 2022, 'Metabolic characterization of hypertrophic cardiomyopathy in human heart', *Nature Cardiovascular Research* **1**(5), 445–461.

864 Wang, X., Han, X. and Powell, C. A. 2022, 'Lipids and genes: Regulatory roles of lipids in rna expression', *Clinical and translational medicine* **12**(5).

865 Wang, X. and Robbins, J. 2006, 'Heart failure and protein quality control', *Circulation research* **99**(12), 1315–1328.

866 Wang, Y., Liu, W., Sun, L., Zhang, Y., Wang, B., Yuan, Y., Li, T., Yao, R., Han, H., Qian, Q. et al. 2020, 'A novel indicator, childhood lipid accumulation product, is associated with hypertension in chinese children and adolescents', *Hypertension Research* **43**(4), 305–312.

867 Wang, Y.-P. and Lei, Q.-Y. 2018, 'Metabolite sensing and signaling in cell metabolism', *Signal transduction and targeted therapy* **3**(1), 30.

868 Wang, Y., Zhou, Y. and Graves, D. T. 2014, 'Foxo transcription factors: Their clinical significance and regulation', **2014**.

869 Wasserstein, R. L., Schirm, A. L. and Lazar, N. A. 2019, 'Moving to a world beyond "p¡ 0.05"'.

870 Watanabe, H., Tao, G., Gan, P., Westbury, B. C., Cox, K. D., Tjen, K., Song, R., Fishman, G. I., Makita, T. and Sucov, H. M. 2023, 'Purkinje cardiomyocytes of the adult ventricular conduction system are highly diploid but not uniquely regenerative', *Journal of cardiovascular development and disease* **10**(4).

871 Watanabe, K., Nagao, M., Toh, R., Irino, Y., Shinohara, M., Iino, T., Yoshikawa, S., Tanaka, H., Satomi-Kobayashi, S., Ishida, T. et al. 2021, 'Critical role of glutamine metabolism in cardiomyocytes under oxidative stress', *Biochemical and Biophysical Research Communications* **534**, 687–693.

872 Watrous, J. D., Henglin, M., Claggett, B., Lehmann, K. A., Larson, M. G., Cheng, S. and Jain, M. 2017, 'Visualization, quantification, and alignment of spectral drift in population scale untargeted metabolomics data', *Analytical chemistry* **89**(3), 1399–1404.

873 Webb-Robertson, B.-J. M., Wiberg, H. K., Matzke, M. M., Brown, J. N., Wang, J., McDermott, J. E., Smith, R. D., Rodland, K. D., Metz, T. O., Pounds, J. G. et al. 2015, 'Review, evaluation, and discussion of the challenges of missing value imputation for mass spectrometry-based label-free global proteomics', *Journal of proteome research* **14**(5), 1993–2001.

874 Wei, R., Wang, J., Su, M., Jia, E., Chen, S., Chen, T. and Ni, Y. 2018, 'Missing value imputation approach for mass spectrometry-based metabolomics data', *Scientific reports* **8**(1), 663.

875 Weiss, R. G., Gerstenblith, G. and Bottomley, P. A. 2005, 'Atp flux through creatine kinase in the normal, stressed, and failing human heart', *Proceedings of the National Academy of Sciences* **102**(3), 808–813.

876 Wenk, M. 2005, 'The emerging field of lipidomics', *Nature Reviews Drug Discovery* **4**(7).

877 Whiteaker, J. R., Lin, C., Kennedy, J., Hou, L., Trute, M., Sokal, I., Yan, P., Schoenherr, R. M., Zhao, L., Voytovich, U. J. et al. 2011, 'A targeted proteomics–based pipeline for verification of biomarkers in plasma', *Nature biotechnology* **29**(7), 625–634.

878 Williams, C. R., Baccarella, A., Parrish, J. Z. and Kim, C. C. 2016, 'Trimming of sequence reads alters rna-seq gene expression estimates', *BMC bioinformatics* **17**, 1–13.

879 Williams, D. R. and Mulder, J. 2020, 'Bayesian hypothesis testing for gaussian graphical models: Conditional independence and order constraints', *Journal of Mathematical Psychology* **99**, 102441.

880 Williams, D. R. and Rast, P. 2020, 'Back to the basics: Rethinking partial correlation network methodology', *British Journal of Mathematical and Statistical Psychology* **73**(2), 187–212.

881 Williams, L. and Frenneaux, M. 2006, 'Diastolic ventricular interaction: From physiology to clinical practice', *Nature Clinical Practice Cardiovascular Medicine* **3**(7).

882 Willis, M. S., Schisler, J. C., Portbury, A. L. and Patterson, C. 2009, 'Build it up–tear it down: protein quality control in the cardiac sarcomere', *Cardiovascular research* **81**(3), 439–448.

883 Wilson, J. H. and Hunt, T. 2015, *Molecular biology of the cell*, Garland Science, New York, USA.

884 Winterbach, W., Mieghem, P. V., Reinders, M., Wang, H. and Ridder, D. d. 2013, 'Topology of molecular interaction networks', *BMC systems biology* **7**, 1–15.

885 Wishart, D. S. 2019, 'Metabolomics for investigating physiological and pathophysiological processes', *Physiological reviews* **99**(4), 1819–1875.

886 Woods, J. R., Plessinger, M. A. and Fantel, A. 1998, 'An introduction to reactive oxygen species and their possible roles in substance abuse', *Obstetrics and Gynecology Clinics of North America* **25**(1).

887 Wright, G. W. and Simon, R. M. 2003, 'A random variance model for detection of differential gene expression in small microarray experiments', *Bioinformatics* **19**(18), 2448–2455.

888 Wu, B. 2005, 'Differential gene expression detection using penalized linear regression models: the improved sam statistics', *Bioinformatics* **21**(8), 1565–1571.

889 Wu, D. and Smyth, G. K. 2012, 'Camera: a competitive gene set test accounting for inter-gene correlation', *Nucleic acids research* **40**(17), e133–e133.

890 Wu, J., Jin, Z., Zheng, H. and Yan, L.-J. 2016, 'Sources and implications of nadh/nad+ redox imbalance in diabetes and its complications', *Diabetes, metabolic syndrome and obesity: targets and therapy* pp. 145–153.

891 Wu, Q., Burley, G., Li, L.-C., Lin, S. and Shi, Y.-C. 2023, 'The role of dietary salt in metabolism and energy balance: Insights beyond cardiovascular disease', *Diabetes, Obesity and Metabolism* **25**(5), 1147–1161.

892 Wu, Q.-Q., Xu, M., Yuan, Y., Li, F.-F., Yang, Z., Liu, Y., Zhou, M.-Q., Bian, Z.-Y., Deng, W., Gao, L. et al. 2015, 'Cathepsin b deficiency attenuates cardiac remodeling in response to pressure overload via tnf-α/ask1/jnk pathway', *American Journal of Physiology-Heart and Circulatory Physiology* **308**(9), H1143–H1154.

893 Wu, T., Guo, H., Lu, Z., Zhang, T., Zhao, R., Tao, N., Wang, X. and Zhong, J. 2020, 'Reliability of lipidsearch software identification and its application to assess the effect of dry salting on the long-chain free fatty acid profile of tilapia muscles', *Food Research International* **138**, 109791.

894 Wu, Z., Liang, X., Li, M., Ma, M., Zheng, Q., Li, D., An, T. and Wang, G. 2023, 'Advances in the optimization of central carbon metabolism in metabolic engineering', *Microbial Cell Factories* **22**(1).

895 Xiao, J., Li, F., Yang, Q., Zeng, X.-F. and Ke, Z.-P. 2020, 'Co-expression analysis provides important module and pathways of human dilated cardiomyopathy', *Journal of Cellular Physiology* **235**(1), 494–503.

896 Xiao, W. and Loscalzo, J. 2020, 'Metabolic responses to reductive stress', *Antioxidants & redox signaling* **32**(18), 1330–1347.

897 Xiao, W., Wang, R.-S., Handy, D. E. and Loscalzo, J. 2018, 'Nad (h) and nadp (h) redox couples and cellular energy metabolism', *Antioxidants & redox signaling* **28**(3), 251–272.

898 Xiao, Y., Hsiao, T.-H., Suresh, U., Chen, H.-I. H., Wu, X., Wolf, S. E. and Chen, Y. 2014, 'A novel significance score for gene selection and ranking', *Bioinformatics* **30**(6), 801–807.

899 Xie, S., Xu, S.-C., Deng, W. and Tang, Q. 2023, 'Metabolic landscape in cardiac aging: insights into molecular biology and therapeutic implications', *Signal Transduction and Targeted Therapy* **8**(1), 114.

900 Xu, C., Song, D., Holck, A., Zhou, Y. and Liu, R. 2020, 'Identifying lipid metabolites influenced by oleic acid administration using high-performance liquid chromatography–mass spectrometry-based lipidomics', *ACS Omega* **5**(20).

901 Xu, J. Y., Xiong, Y. Y., Tang, R. J., Jiang, W. Y., Ning, Y., Gong, Z. T., Huang, P. S., Chen, G. H., Xu, J., Wu, C. X. et al. 2022, 'Interleukin-5-induced eosinophil population improves cardiac function after myocardial infarction', *Cardiovascular research* **118**(9), 2165–2178.

902 Xu, L. and Brink, M. 2016, 'mtor, cardiomyocytes and inflammation in cardiac hypertrophy', *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1863**(7, Part B).

903 Xu, Y., Wang, B., Liu, H., Deng, Y., Zhu, Y., Zhu, F., Liang, Y. and Liu, H. 2021, 'Sp1 targeted parp1 inhibition protects cardiomyocytes from myocardial ischemia–reperfusion injury via downregulation of autophagy', *Frontiers in Cell and Developmental Biology* **9**.

904 Xue, H., Chen, X., Yu, C., Deng, Y., Zhang, Y., Chen, S., Chen, X., Chen, K., Yang, Y. and Ling, W. 2022, 'Gut microbially produced indole-3-propionic acid inhibits atherosclerosis by promoting reverse cholesterol transport and its deficiency is causally related to atherosclerotic cardiovascular disease', *Circulation research* **131**(5), 404–420.

905 Yacoub, M. H. 1995, 'Two hearts that beat as one', *Circulation* **92**(2).

906 Yaku, K., Okabe, K. and Nakagawa, T. 2018, 'Nad metabolism: Implications in aging and longevity', *Ageing research reviews* **47**, 1–17.

907 Yamada, R., Okada, D., Wang, J., Basak, T. and Koyama, S. 2021, 'Interpretation of omics data analyses', *Journal of human genetics* **66**(1), 93–102.

908 Yamamoto, T. and Sano, M. 2022, 'Deranged myocardial fatty acid metabolism in heart failure', *International Journal of Molecular Sciences* **23**(2).

909 Yamamura, K. 1999, 'Transformation using (x+ 0.5) to stabilize the variance of populations', *Population Ecology* **41**(3), 229–234.

910 Yan, H. and Liu, W. 2023, 'Metabolic changes in cardiac aging', *Reviews in Cardiovascular Medicine* **24**(3).

911 Yan, X., Huang, Y. and Wu, J. 2018, 'Identify cross talk between circadian rhythm and coronary heart disease by multiple correlation analysis', *Journal of computational biology* **25**(12).

912 Yang, J., Villar, V. A. M., Rozyyev, S., Jose, P. A. and Zeng, C. 2019, 'The emerging role of sorting nexins in cardiovascular diseases', *Clinical science* **133**(5), 723–737.

913 Yang, P., Chen, X., Kaushal, S., Reece, E. A. and Yang, P. 2016, 'High glucose suppresses embryonic stem cell differentiation into cardiomyocytes: High glucose inhibits es cell cardiogenesis', *Stem cell research & therapy* **7**(1).

914 Yang, Y. H., Dudoit, S., Luu, P., Lin, D. M., Peng, V., Ngai, J. and Speed, T. P. 2002, 'Normalization for cdna microarray data: a robust composite method addressing single and multiple slide systematic variation', *Nucleic acids research* **30**(4), e15–e15.

915 Yang, Y. H., Dudoit, S., Luu, P. and Speed, T. P. 2001, Normalization for cdna microarry data, *in* 'Microarrays: optical technologies and informatics', Vol. 4266, SPIE, pp. 141–152.

916 Yaniv, Y., Juhaszova, M. and Sollott, S. J. 2013, 'Age-related changes of myocardial atp supply and demand mechanisms', *Trends in endocrinology and metabolism: TEM* **24**(10).

917 Yeh, C.-H., Chou, Y.-J., Kao, C.-H. and Tsai, T.-F. 2020, 'Mitochondria and calcium homeostasis: Cisd2 as a big player in cardiac ageing', *International Journal of Molecular Sciences* **21**(23), 9238.

918 Yen, C., Stone, S., Koliwad, S., Harris, C. and Farese, R. 2008, 'Thematic review series: Glycerolipids. dgat enzymes and triacylglycerol biosynthesis', *Journal of Lipid Research* **49**(11).

919 Yeo, E. J. 2019, 'Hypoxia and aging', *Experimental & Molecular Medicine* **51**(6).

920 Yeo, H., Park, T., Kim, H., Jang, H., Lee, J., Hwang, G., Ryu, S., Park, S., Song, H., Ban, H., Yoon, H. and Lee, B. 2021, 'Phospholipid transfer function of ptpip51 at mitochondria-associated er membranes', *EMBO reports* **22**(6).

921 Yildiz, M., Oktay, A., Stewart, M., Milani, R., Ventura, H. and Lavie, C. 2020, 'Left ventricular hypertrophy and hypertension', *Progress in Cardiovascular Diseases* **63**(1).

922 Yoon, S., Baik, B., Park, T. and Nam, D. 2021, 'Powerful p-value combination methods to detect incomplete association', *Scientific reports* **11**(1), 6980.

923 Young, B., O'Dowd, G., Young, B. and Wheater, P. R. 2014, *Wheater's functional histology: A text and colour atlas.*, Churchill Livingstone Elsevier, Philadelphia, USA.

924 Zadoorian, A., Du, X. and Yang, H. 2023, 'Lipid droplet biogenesis and functions in health and disease', *Nature Reviews Endocrinology* **19**(8), 443–459.

925 Zeigler, A., Richardson, W., Holmes, J. and Saucerman, J. 2016, 'A computational model of cardiac fibroblast signaling predicts context-dependent drivers of myofibroblast differentiation', *Journal of molecular and cellular cardiology* **94**, 72–81.

926 Zelnik, I., Ventura, A., Kim, J., Silva, L. and Futerman, A. 2020, 'The role of ceramide in regulating endoplasmic reticulum function', *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **1865**(1).

927 Zhang, B. and Horvath, S. 2005, 'A general framework for weighted gene co-expression network analysis', *Statistical applications in genetics and molecular biology* **4**(1).

928 Zhang, H., Du, Y., Wang, Z., Lou, R., Wu, J., Feng, J. et al. 2019, 'Integrated analysis of oncogenic networks in colorectal cancer identifies guca2a as a molecular marker', *Biochemistry research international* **2019**.

929 Zhang, J., DeFelice, A. F., Hanig, J. P. and Colatsky, T. 2010, 'Biomarkers of endothelial cell activation serve as potential surrogate markers for drug-induced vascular injury', *Toxicologic Pathology* **38**(6).

930 Zhang, J., Wang, Y. T., Miller, J. H., Day, M. M., Munger, J. C. and Brookes, P. S. 2018, 'Accumulation of succinate in cardiac ischemia primarily occurs via canonical krebs cycle activity', *Cell Reports* **23**(8).

931 Zhang, L., Shi, X., Gu, C., Chen, B., Wang, M., Yu, Y., Sun, K. and Zhang, R. 2020, 'Identification of cell-to-cell interactions by ligand-receptor pairs in human fetal heart', *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1866**(12).

932 Zhang, S., Wang, B., Wan, L. and Li, L. M. 2017, 'Estimating phred scores of illumina base calls by logistic regression and sparse modeling', *BMC bioinformatics* **18**, 1–14.

933 Zhang, X., Liu, H., Gao, J., Zhu, M., Wang, Y., Jiang, C. and Xu, M. 2019, 'Metabolic disorder in the progression of heart failure', *Science China Life Sciences* **62**, 1153–1167.

934 Zhang, X., Zhao, X.-M., He, K., Lu, L., Cao, Y., Liu, J., Hao, J.-K., Liu, Z.-P. and Chen, L. 2012, 'Inferring gene regulatory networks from gene expression data by path consistency algorithm based on conditional mutual information', *Bioinformatics* **28**(1), 98–104.

935 Zhang, Z., Murtagh, F., Van Poucke, S., Lin, S. and Lan, P. 2017, 'Hierarchical cluster analysis in clinical research with heterogeneous study population: highlighting its visualization with r', *Annals of translational medicine* **5**(4).

936 Zhao, D. 2021, 'Epidemiological features of cardiovascular disease in asia', *JACC: Asia* **1**(1), 1–13.

937 Zhao, K. and Rhee, S. 2023, 'Interpreting omics data with pathway enrichment analysis', *Trends in Genetics* **39**(4).

938 Zhao, L., Cao, J., Hu, K., He, X., Yun, D., Tong, T. and Han, L. 2020, 'Sirtuins and their biological relevance in aging and age-related diseases', *Aging and disease* **11**(4).

939 Zhao, L.-Y., Song, J., Liu, Y., Song, C.-X. and Yi, C. 2020, 'Mapping the epigenetic modifications of dna and rna', *Protein & cell* **11**(11), 792–808.

940 Zhao, L., Zou, X., Feng, Z., Luo, C., Liu, J., Li, H., Chang, L., Wang, H., Li, Y., Long, J. et al. 2014, 'Evidence for association of mitochondrial metabolism alteration with lipid accumulation in aging rats', *Experimental gerontology* **56**, 3–12.

941 Zhao, R., Jiang, S., Zhang, L. and Yu, Z. 2019, 'Mitochondrial electron transport chain, ros generation and uncoupling (review)', *International Journal of Molecular Medicine* **44**(1).

942 Zhao, S., Zhang, Y., Gordon, W., Quan, J., Xi, H., Du, S., von Schack, D. and Zhang, B. 2015, 'Comparison of stranded and non-stranded rna-seq transcriptome profiling and investigation of gene overlap', *BMC genomics* **16**, 1–14.

943 Zhao, Y., Wong, L. and Goh, W. W. B. 2020, 'How to do quantile normalization correctly for gene expression data analyses', *Scientific reports* **10**(1), 15534.

944 Zheng, P., Chen, L., Guan, Y.-Z. and Liu, P. 2021, 'Weighted gene co-expression network analysis identifies specific modules and hub genes related to coronary artery disease', *Scientific Reports* **11**(1).

945 Zhihao, L., Jingyu, N., Lan, L., Michael, S., Rui, G., Xiyun, B., Xiaozhi, L. and Guanwei, F. 2020, 'Serca2a: a key protein in the ca 2+ cycle of the heart failure', *Heart failure reviews* **25**, 523–535.

946 Zhong, X., Heinicke, F., Lie, B. A. and Rayner, S. 2019, 'Accurate adapter information is crucial for reproducibility and reusability in small rna seq studies', *Non-coding RNA* **5**(4), 49.

947 Zhou, J., Wang, X., Wang, M., Chang, Y., Zhang, F., Ban, Z., Tang, R., Gan, Q., Wu, S., Guo, Y. et al. 2019, 'The lysine catabolite saccharopine impairs development by disrupting mitochondrial homeostasis', *Journal of Cell Biology* **218**(2), 580–597.

948 Zhou, P. and Pu, W. T. 2016, 'Recounting cardiac cellular composition', *Circulation Research* **118**(3).

949 Zhuang, Y., Yu, M.-l. and Lu, S.-f. 2023, 'Purinergic signaling in myocardial ischemia–reperfusion injury', *Purinergic Signalling* **19**(1), 229–243.

950 Zitka, O., Skalickova, S., Gumulec, J., Masarik, M., Adam, V., Hubalek, J., Trnkova, L., Kruseova, J., Eckschlager, T. and Kizek, R. 2012, 'Redox status expressed as gsh: Gssg ratio as a marker for oxidative stress in paediatric tumour patients', *Oncology letters* **4**(6), 1247–1253.

951 Zolotareva, O., Nasirigerdeh, R., Matschinske, J., Torkzadehmahani, R., Bakhtiari, M., Frisch, T., Späth, J., Blumenthal, D. B., Abbasinejad, A., Tieri, P. et al. 2021, 'Flimma: a federated and privacy-aware tool for differential gene expression analysis', *Genome biology* **22**(1), 338.

952 Zou, H. and Hastie, T. 2005, 'Regularization and variable selection via the elastic net', *Journal of the Royal Statistical Society Series B: Statistical Methodology* **67**(2), 301–320.

953 Züllig, T. and Köfeler, H. C. 2021, 'High resolution mass spectrometry in lipidomics', *Mass spectrometry reviews* **40**(3), 162–176.

954 Zyla, J., Marczyk, M., Weiner, J. and Polanska, J. 2017, 'Ranking metrics in gene set enrichment analysis: do they matter?', *BMC bioinformatics* **18**, 1–12.

955 Züllig, T., Trötzmüller, M. and Köfeler, H. 2019, 'Lipidomics from sample preparation to data analysis: a primer', *Analytical And Bioanalytical Chemistry* **412**(10).
Buckberg et al.

956 Buckberg, G., Nanda, N., Nguyen, C. and Kocica, M. 2018, 'What is the heart? anatomy, function, pathophysiology, and misconceptions', *Journal of Cardiovascular Development and Disease* **5**(2).

Chapter 11

Appendix - Background

11.1 Endothelial cells

Endothelial cells play an active role in cardiac physiology and disease progression[516] although the recent study suggests they may not be predominant in the left ventricle[504]. Endothelial cells are also responsible for regulating coronary angiogenesis, an essential biological process that enables the development and maintenance of cardiac vascularisation[212]. In the human heart, histological analysis has highlighted the presence of at least one capillary close to each cardiac myofibril. On average, endothelial cells are no farther than $3\mu m$ from cardiomyocytes, and the density of coronary capillaries increases as cardiomyocytes enlarge their volume. A decreased density in the vasculature close to Purkinje fibres has also been observed[279]. Moreover, analysing the effect of ageing on the cardiac microvasculature in mouse hearts, Rakusan and Nagai (1994) revealed rarefaction of vascular beds in the cardiac wall in senescence mice[668]. Additionally, a lack of myocardial blood flow in the left ventricle of canines leads to a perfusion deficit and eventually a possible ischaemic attack due to the augmented cardiac stress[279].

The close spatial distance between cardiomyocytes and endothelial cells indicates that both cells embark on cross-talks. One example can be angiogenesis, a process carried out entirely by endothelial cells but regulated by cardiomyocytes through vascular endothelial growth factor-A (VEGF-A) synthesis. Giordano et al. (2001) have demonstrated that deleting the encoding gene of VEGF-A results in the activation of hypoxia-responsive genes, which consequentially changes myocardial metabolism toward glucose-based energy production by over-expressing glucose transporter 1 (GLUT1). The expression of vascular endothelial growth factor receptors, VEGFR1 and VEGFR2, affects the responsiveness of endothelial cells to the related growth factor. A high VEGFR1/VEGFR2 ratio tends to be found in endothelial cells stalked in the newer vessels. In contrast, a lower VEGFR1/VEGFR2 ratio is associated with the phenotypes of endothelial cells on the tip of the newly born vessel. Stalked and tip endothelial cells do not stay still but shuttle continuously[212]. Metabolism of endothelial cells greatly impacts the sprouting of the new vessel. As for every cell type, the main energy sources of endothelial cells are glucose, amino acid and fatty acid. Endothelial cells rely on anaerobic glycolysis, as most of the glucose goes through the whole chain of reactions and produces lactate. While anaerobic glycolysis is suboptimal in producing ATP

per one mole of glucose with respect to the Krebs cycle and oxidative phosphorylation, it can benefit endothelial cells indirectly. For instance, reactive oxygen species, prominent signalling metabolites associated with fibrosis, ischaemia-reperfusion and tissue ageing[752], are produced less during the metabolic process[88]. Using rats, Krützfeldt (1990) have shown that endothelial cells metabolise 98% of glucose into lactate[445]. Acting as a signalling molecule, lactate also stimulates angiogenesis[368].

Furthermore, recent studies have identified different types of endothelial cells across different heart regions. Litviňuková et al. (2020) have identified ten different cell types of endothelial cells. Three endothelial cells, EC1, EC2, and EC3, are specific for capillaries; they exhibit higher expression of RGCC and CA4 genes and comprise 57.4% of the total population of endothelial cells in the heart. While resembling the first three cells, EC4 cells are likely to have immune and inflammatory regulatory processes. Genes such as chemokines CX3CL1 and CCL2, cytokine protein IL-6 and protein ICAM1 are expressed[504,929]. Arterial endothelial cells, EC5, display enrichment of SEMA3G, a protein capable of modulating several functions of the cardiovascular system, and EFNB2 and DLL4 proteins, regulators of angiogenesis[328,435,504]. Associated with veins, endothelial cells are enriched with NR2F2, a transcription factor with multiple roles in determining the cell fate[760], and ACKR1, a chemokine receptor that inhibits angiogenesis[712]. EC7 is found in the atria and enriched with SMOC1, a negative regulator of cell proliferation, migration and angiogenesis[37]. It also expresses NPR3, a receptor for a potential role in cardiomyocytes-endothelial cells cross-talk[931]. The phenotype EC8 is a lymphatic endothelial cell and accounts for 1% of the endothelial cells in cardiac tissues. EC8 expresses PROX1, a transcriptor factor regulating the cell fate and lineage differentiation[402], TBX1, a transcription factor involved in lymphangiogenesis[545], and PDPN, a transmembrane protein typical of the lymphatic lineage of endothelial cells[797].

The remaining two types of endothelial cells are fibroblast-like EC9 and cardiomyocyte-like EC10[504]. The contractility of cardiomyocytes is also a result of cross-talks with endothelial cells. Nitric oxide, synthesised by nitric oxide synthase 3 (NOS3) in endothelial cells, controls the relaxation of ventricles through cyclic guanosine monophosphate (cGMP) and protein kinase G mediation of troponin I phosphorylation. When there is a low nitric oxide concentration in the cardiac wall, ventricles are relaxed through different routes that produce peroxynitrite, a metabolite associated with oxidative stress. The ET_B receptor in the endothelial cell facilitates the release of nitric oxide, stimulated by endothelin-1 (ET-1), a metabolite synthesised within the endothelial cell. The ET_B receptor can also be found in the nuclear membrane of cardiomyocytes and regulates calcium signalling in nuclei during extreme hypoxia[780]. Endothelial cells offer protection against oxidative stress by acting on several buffers, such as ATP-binding cassette transporter G2 (ABCG2), NgR1, BH4, and thioredoxin, while the ET_B on endothelial cell membranes stimulate the release of nitric oxide[149].

11.2 Leukocytes

Leukocytes are known as keepers of the immune system, and their role in cardiac tissues has been explored by Rubattu et al. (2019), who utilised circulating leukocytes. While their approach concerned leukocytes in the bloodstream, their molecular functions and biological processes can mirror those in cardiac tissues without any invasive sampling methods[793]. It is notable that leukocytes in cardiac tissues increase in number after an episode of myocardial infarction and ischaemia. In other words, lesions are quickly overwhelmed by neutrophils, the first defence of the immune system, and $CD16^{-}CD14^{+}$, inflammatory monocytes. The recruitment of neutrophils and $CD16^{-}CD14^{+}$ occurs after cardiac endothelial cells increase the production of monocyte chemoattractant protein-1 (MCP-1). Several days after the injury, $D16^{-}CD14^{dim}$ replaces the $CD16^{-}CD14^{+}$ monocytes in cardiac tissues. Although the replacement occurs due to CX3CL1 accumulation, the reason for the switch is unclear. The $CD16^{-}CD14^{dim}$ cells promote angiogenesis and synthesis of the extracellular matrix by supplementing vascular endothelial cells and transforming growth factors, namely VEGF and TGF$-\beta$. Furthermore, monocytes invade non-damaged parts of the tissues adjacent to the infarction area, potentially activating fibroblasts, which promote fibrosis[793]. After reaching the injured area, monocytes respond to stimuli such as lipid accumulation and differentiate into macrophages that detect and destroy bacteria. Observing the matching profile of lipids in senescent cells and wound healing in skin, Pils et al. (2021) have conjectured that cellular senescence and senescence-associated phenotype contribute to wound healing changes seen with ageing[647]. Similar dynamics might occur in aged hearts after cardiac injury, yet further research is needed to elucidate this conjecture[558].

11.3 Pericytes

Inflammatory responses also involve a cross-talk between leukocytes and pericytes, contractile interstitial cells that wrap vessels[707]. Recent studies have provided the tools to trace the lineage of pericytes in the myocardium by targeting CSPG4 (or NG2), an antigen highly expressed in pericytes but not in fibroblasts[15]. Pericytes mediate diapedesis, leukocyte infiltration into the vascular wall, by expressing proinflammatory molecules, such as MHCII that boost the phagocytic activity of neutrophils. Ando et al.'s (2022) research on brain tissues, furthermore, has suggested that their inhibition reduces the differentiation of vascular smooth muscle cells. Litviňuková et al. (2020) have highlighted that cardiac pericytes express ABCC9 and KNIJ8, two encoding genes of ATP-sensitive potassium channels. They also found that cardiac pericytes can be divided into four different clusters of pericytes. Type-1 pericytes are generally found in the ventricle and have higher expression of adhesive proteins NCAM2, CD38 and CSPG4[15]. While type-2 pericytes are usually found in atria, type-3 is characterised by a low expression of endothelial cell transcripts, suggesting that type-3 pericytes are a possible transition of pericytes toward endothelial cell phenotype. Type-4 pericytes are similar to cardiomyocytes[504].

Vascular smooth muscle cells are differentiated cells usually found in the medial layer of blood vessels and play an important role in blood supply to the cardiac wall. Smooth

muscle cells have a unique expression profile, being enriched with myosin heavy chain 11 (MYH11), α-actin2 (ACTA2), transgelin (TAGLN), and myocardin (MYOCD). Vascular smooth muscle cells can also switch their phenotype from contractile cells to dedifferentiated synthetic cells. When vascular muscle cells are dedifferentiated into synthetic cells, they lose their characteristic gene markers while increasing the synthesis of extracellular components[506,802]. Vascular smooth muscle cells' phenotypic flexibility makes their tracing quite difficult. An increase of smooth muscle synthetic cells is observed after vascular trauma[506]. There are two possible phenotypes within Vascular smooth muscle cells with MYH11 expression, namely, SMC1 and SMC2[504]. SMC1 has a lower level of CNN1 and ACTA2 than SMC2 and matches the profile of vein-derived smooth muscle cells, while SMC2 matches smooth muscle cells derived from arteries[504].

11.4 Fibroblast and extracellular matrix

Fibroblasts are the guardians of microenvironments surrounding the left ventricle and the second most abundant population of cells in human ventricles[504]. Cardiac fibroblasts are situated parallel to cardiomyocytes across different layers of the myocardium. The number of fibroblasts can fluctuate with time as their counts reflect the pathophysiology of the ventricular tissue; for instance, the number of cardiac fibroblasts increases in the area affected by myocardial infarction[657]. Cardiac fibroblasts originate from the epicardium, where epithelial cells transit to mesenchymal cells, which further differentiate into fibroblasts after exposure to growth factors. While a smaller portion of fibroblasts derives from the endothelial cells, how the difference in derivation impacts fibroblast physiology remains unclear[312]. Fibroblasts can transmit electrophysiological signals between cardiomyocytes through gap junctions, characterised by connexin and cadherin proteins[657]. Fibroblasts have a contradistinctive expression of epicardial transcription factor 21 (Tcf21) and platelet-derived growth factor receptor alpha (PDGFRα)[602]. Although fibroblasts are mature cells, they can further differentiate. In a cardiac injury, endocrine and paracrine signalling induces the proliferation and migration of fibroblasts into the affected area and consequently differentiates into activated fibroblasts[312]. Activated fibroblasts are enriched with periostin (Postn), and matrix metalloproteinases (MMPs). The ablation of Postn leads to the reduction of fibrotic scarring. MMPs are activators of TGFβ, cytokines associated with the increase of phosphorylation of SMAD2, SMAD3 and SMAD4, with subsequent stimulation of the differentiation process of fibroblasts into activated fibroblasts[312]. Activated fibroblasts can further differentiate into myofibroblasts; notably, the resident-activated fibroblasts are the ones that undergo differentiation[312,359]. Myofibroblasts express ACTA2, which confers the ability to exercise contractile forces[680]. Myofibroblasts are responsive to several molecules, including vasoactive protein angiogenesis II, aldosterone, and endothelin-1. Myofibroblasts also respond to proinflammatory cytokines such as interleukin-1 (IL-1), interleukin-6 (IL-6) and TNFα, conferring them a prominent role when a cardiac injury occurs[657]. The final stage of differentiation turns myofibroblasts into matrifibrocytes, a type of cell with limited proliferation capacity and low ACTA2 expression[312,680].

One of the main functions of cardiac fibroblasts is to modulate the extracellular matrix

turnover both directly and indirectly. Their direct modulation involves secreting and degrading the matrix components. Fibroblasts can synthesise and deposit collagen types I and III, proteins that transmit contractile forces across the cardiac left ventricle. Collagen fibrils build a supporting mesh in the cardiac muscle; as discussed later, in the extracellular matrix, collagen type I accounts for 85% of the total collagen and collagen type III approximately 10% in non-diseased hearts. The composition of the extracellular matrix can vary within different locations in the heart, and an overall increase of collagens is associated with the progression of cardiovascular diseases, such as fibrosis and hypertrophy. Adaptive responses of fibroblasts due to changes in the cardiac environment can lead to dysfunction in the regulation of the extracellular matrix, followed by changes in the propagation of electrical and transduction signals and mechanical forces within the cardiac tissues. On the other hand, the indirect modulation of the extracellular matrix occurs through TGF-β, stimulated by angiogenesis II[657].

Fibroblasts become senescent with ageing or a tissue injury, producing inflammatory molecules and activating several receptors. For instance, TGF-β and TGF-β receptors I/II cause phosphorylation of Smad2/3/4 receptors[312], and silencing of Smad3 appears to reduce fibrosis when cardiac injuries occur[419]. Inflammation also stimulates the production of Wnt1, signalling molecules for the Wnt pathways involved in cell fate processes[396]. Activating the Wnt pathway increases the abundance of β-catenin, a regulator of cardiac hypertrophy and fibrosis. Recent evidence has shown that cross-talk between Wnt and Hippo pathways increases cardiac dilation and autophagy, whereas its mechanism has not been fully understood[531]. Furthermore, fibroblast-cardiomyocyte interactions also contribute to several biological functions within the left ventricle and the heart. For instance, the paracrine signalling pathway regulates the swelling of cardiomyocytes through Ang II and TGFβ, signalling molecule secretion by fibroblasts. Cardiomyocyte swelling induces impaired cell-cell communication. The IL-6 pathway is not only responsible for regulating angiogenesis but also for modulating apoptosis of cardiomyocytes[602]. Indirect interactions between cardiocytes and fibroblasts occur through the extracellular matrix; cardiomyocytes increase their cell-cell interaction proteins, such as cadherin, in the stiff extracellular matrix, tightening the interactions with the surrounding cells[602].
The non-myocytes account for $60\% - 70\%$ of the total cells within the heart, although they occupy only the $20\% - 30\%$ of the cardiac volume. Despite the early assumption in scholarship that fibroblasts were the most predominant non-myocytes, it was later proven that endothelial cells account for 60% of non-muscle cells along with leukocytes (white blood cells) that constitute 10% while fibroblasts are only 15% of the non-myocytes. The endothelial cells have a polygonal shape and are connected through junctions. The endothelial cells have several metabolic functions such as synthesis of collagen, proteoglycan and formation of a permeable barrier when clustered together. The leukocytes are a class of cells that belongs to the immune system and their role is to tackle infection and external agents. The significant amount of white blood cells reflects the high density of blood, and lymphatic vessels are present throughout the heart. Among the non-myocytes, fibroblasts secrete the extracellular matrix components such as collagen proteins and elastin and are therefore essential for maintaining biophysical properties of the tissue. The fibroblasts are in general small-sized cells with an elongated shape and can host more than one nucleus[649,923,948].

The extracellular matrix offers mechanical support to the organs by preserving the integrity of tissues and coordinates cellular responses in case of external stimuli by transmitting signal transduction across nearby cells. The extracellular matrix also guarantees cardiac muscle homeostasis; structural and molecular changes of the extracellular matrix are often observed in cardiac diseases, such as dilated cardiomyopathy[84,597,613]. The extracellular matrix holds cardiomyocytes interconnected via Z-band integrin connections, which are essential for transmitting electrochemical signals across cells[342]. The extracellular network can be classified into two main layers: the basement membrane (or external lamina) and the interstitial matrix. The basement membrane is a thin sheet of the specialised extracellular matrix and is generally interposed between the different cells in the hearts and the underlying interstitial matrix[82]. The basement membrane offers a selectively permeable barrier for biomolecules such as metabolites, helping chemicals pass from and to epithelial cells. It also modulates the growth and differentiation of epithelial cells and isolates cardiomyocytes from the interstitial matrix. The basement membrane contains abundant type IV collagen, laminins, fibronectin, nidogen-1 (or entactin) and heparan sulphate. The nidogen-1 and laminins complexes connect two basement membrane layers. The interstitial matrix, the predominant layer of the extracellular matrix, principally contains collagen proteins and other components, including proteases, growth factors, glycoproteins and fibronectin. The interstitial matrix further comprises three sheets: endomysium, perimysium and epimysium. While endomysium is the innermost part of the muscle tissue and rich in type III collagen fibres, which draw a fine network surrounding cardiac cells, epimysium and perimysium are the outer-most and intermediate layers that mainly contain type I collagen[245,923]. As such, collagen synthesis and post-translational modification are heavily regulated[342]. The extracellular matrix structure varies by several factors, such as positions of the analysed tissue, age and gender. The study suggests that an anomalous accumulation of collagen in the extracellular matrix may increase the stiffness of the cardiac wall (cardiac fibrosis) with consequent valvular dysfunction[597].

Chapter 12

Appendix - Research Methods

12.1 Transforming omics measurements - extended

Standardisation Standardisation, or z-score, belongs to a family of linear functions and expresses the centre and dispersion of the data distribution in order to determine coefficients a and b of the linear transformation $f(\tilde{\mathbf{Z}}_{i,j}) = a + b \cdot \tilde{\mathbf{Z}}_{i,j}$. Standardisation maps omics measurements onto the real space and can be applied for analytes or samples, which could lead to different results in the follow-up analyses[238]. For example, the best option for preserving a part of the analyte variability would be to perform the sample-wise standardisation by imposing a common variability and centre of mass for each sample. Alternatively, if the researcher is not interested in the data dispersion, which can be affected by the dimension attached to each variable, the analyte-wise standardisation can be appropriate, as it assigns equal weight to each analyte in the data set[719]. Since this study aims to preserve the variability of a molecule, consequently, sample-wise standardisation is defined as

$$\mathbf{Z}_{i,j} = \frac{\tilde{\mathbf{Z}}_{i,j} - \mu_{i,\cdot}}{\sigma_{i,\cdot}},$$

where $\mu_{i,\cdot}$ is the sample-wise mean, $\mu_{i,\cdot} = \frac{1}{J}\sum_j \tilde{\mathbf{Z}}_{i,j}$ and $\sigma_{i,\cdot}$ is the sample-wise standard deviation $\sigma_{i,\cdot} = \sqrt{\frac{1}{J-1}\sum_j \left(\tilde{\mathbf{Z}}_{i,j} - \mu_{i,\cdot}\right)^2}$, where J denotes the total amount of samples in the study. An extension of standardisation relies on the use of the median and median absolute deviation instead of mean and standard deviation, which are robust measures for the centre of mass and dispersion of the data[700]. Standardisation, nonetheless, entails some limitations. Although standardisation aligns moments of the distribution of samples or analytes, standardisation would not help to match the assumption of statistical methods if the distribution does not exhibit normality. Additionally, for the sample-wise performance, standardisation complicates the interpretation of the follow-up results. Another drawback is the possible inflation of measurement errors intrinsic in any high-throughput technology. While standardisation represents a valuable transformation, it adds complexity to the interpretation of the statistical analysis.

Min-max transformation Min-max transformation aims to express the omics data in a predefined range of interest $[l, u]$ for any lower-bound l and upper-bound u. While min-max

transformation can be applied sample- or analyte-wise, this thesis presents the sample-wise min-max transformation only. Additionally, the analyte-wise min-max transformation would always lead to a transformed analyte abundance with a minimum of 0 and a maximum equal to 1, augmenting the difficulty in interpreting the data. The sample-wise min-max transformation finds its primary application in mathematical models, such as continuous logic models, where the omics measurements can be expressed $[0, 1]$ to improve curve fittings[420]. The sample-wise min-max transformation reads

$$\mathbf{Z}_{i,j} = l + \frac{\tilde{\mathbf{Z}}_{i,j} - \min_j(\mathbf{Z}_{i,j})}{\max_j(\mathbf{Z}_{i,j}) - \min_j(\mathbf{Z}_{i,j})} \cdot (u - l), \tag{12.1}$$

where $\min_j(\mathbf{Z}_{i,j})$ and $\max_j(\mathbf{Z}_{i,j})$ are respectively the minimum and maximum across all the analyte measurements in the i-th sample. A variation of the min-max is known as range scaling[763], where $\mathbf{Z}_{i,j} = \tilde{\mathbf{Z}}_{i,j} - \mu_{i\cdot}(\mathbf{Z}_{i,j}) / \max_j(\mathbf{Z}_{i,j}) - \min_j(\mathbf{Z}_{i,j})$. Min-max transformation can be a valuable transformation in the context of mathematical modelling for signalling pathways, as it allows the data expression in the same range of mathematical functions employed to estimate coefficients or validate the proposed model on the observed experimental data[925]. Albeit its easy implementation, the sample-wise min-max transformation can inflate measurement errors and is highly susceptible to outliers. Additionally, it does not address possible divergence of the measurement data from the Gaussian distribution, making it unsuitable for the age-OME project, where some statistical methods require that analytes are approximately Gaussian.

12.2 Normalising omics measurements - extended

Scale normalisation Scale normalisation is a method that was first introduced in the context of microarray and later applied in MS data[765]. The aim of scale normalisation is to reduce the unwanted variation by dividing all the analytes across samples for a common factor so that all the samples have a common median absolute deviation[765]. Scale normalisation assumes that an analyte should not increase its difference across samples as its average expression increases. The first step of the scale normalisation is to calculate minus components for each analyte. For all the minus components for the i-th analyte, $\mathbf{M}_i$ reads

$$\mathbf{M}_{i,\cdot} = \mathbf{Z}_{i,j} - \mathbf{Z}_{i,j'} \tag{12.2}$$

for each pair of samples $j \neq j'$. The second step of the scale normalisation is to compute the average components for each analyte. The average components for the i-th analyte are calculated as

$$\mathbf{A}_{i\cdot} = \frac{1}{2} \cdot \left(\mathbf{Z}_{i,j} + \mathbf{Z}_{i,j'}\right). \tag{12.3}$$

This results in two matrices: the $\mathbf{A}$ matrix and the minus matrix $\mathbf{M}$. In both matrices, the element in the (i, j) position identifies the average and the minus values for the i-th analyte in the j-th sample. The normalising factor for the j-th sample is then calculated as

$$c_j = \frac{MAD_j}{\left(\prod_{j=1}^{J} MAD_j\right)^{1/J}}, \tag{12.4}$$

where $MAD_j = median_i(|\mathbf{M}_{i,j} - median_i(\mathbf{M}_{i,j})|)$ [i] and J stands for the total number of samples[915]. The corrected minus components for each j-th sample are then recalculated as

$$\mathbf{M}'_{\cdot,j} = \mathbf{M}_{\cdot,j} - c_j. \tag{12.5}$$

The normalised transformation is then calculated by substituting $\mathbf{M}$ with $\mathbf{M}'$ in Equation 12.2 and solve it simultaneously with Equation 12.2 in order to retrieve the normalised $\mathbf{Z}_{i,j}$ and $\mathbf{Z}_{i,j'}$ omics measurements for samples j and j' with $j \neq j'$. In other words, the normalised omics measurements are $\mathbf{Y}_{i,j} = \mathbf{A}_{i,\cdot} + 0.5 \cdot \mathbf{M}'_{\cdot,j}$ for the j-th sample and $\mathbf{Y}_{i,j'} = \mathbf{A}_{i,\cdot} + 0.5 \cdot \mathbf{M}'_{i,\cdot}$ for the j'-th sample.

This particular framework offers some variety[914]. One is known as the median normalisation. The median normalisation procedure substitutes the median absolute deviation in Equation 12.4 with the median absolute value. Additionally, the median normalisation uses corrected average components $\mathbf{A}'$ to calculate the normalised omics measurements, where $\mathbf{A}'$ are calculated in a similar fashion to $\mathbf{M}'$ (see Equations 12.5 and 12.4).

Cyclic loess normalisation is an extension of scale normalisation that relaxes the linearity assumption of systematic bias[50] (see Equation 12.4). Similar to scale normalisation, cyclic loess normalisation starts with calculating the minus ($\mathbf{M}$) and average ($\mathbf{A}$) components for two different samples across the J samples available in the omics measurements. Let $neigh([\mathbf{A}_{\cdot,j})$ be the neighbourhood of size $J \cdot 40\%$ of the element $\mathbf{A}_{i,j}$, then the loess polynomial regression problem reads

$$\min_{d,\mathbf{b}} \frac{1}{2} \sum_i w_i \left(\mathbf{M}_{i,j} - \sum_{k=0}^{d} b_k \mathbf{A}_{i,j}^k \right)^2 = \min_{d,\mathbf{b}} \frac{1}{2} \sum_i w_i \left(\mathbf{M}_{i,j} - f(k, \mathbf{b}, \mathbf{A}_{i,j}) \right)^2 \tag{12.6}$$

where $d \in \mathbb{N}$ represents the degree of the polynomial function, $\mathbf{b}^T = [b_0, \ldots, b_d] \in \mathbb{R}^{d+1}$ regression coefficients, and w_i the weight defined as

$$w_i = \begin{cases} (1 - |\mathbf{A}_{i,j}|^3)^3 & \text{if } \mathbf{A}_{i,j} \in neigh(\mathbf{A}_{\cdot,j}) \\ 0 & \text{otherwise} \end{cases} \tag{12.7}$$

for each j. Cyclic loess normalisation $d = 1$ reduces the complexity of the problem while regarding local systematic bias as linear[765]. Once Equation 12.6 is solved for each point in the j-th sample, the estimated vector of values is subtracted to $\mathbf{M}_{\cdot,j}$. The normalised values are obtained in a similar approach as the scale normalisation. Then, the algorithm is reiterated until the normalised values converge[50]. While those types of normalisation – scale, median and loess – are widely applied[573], they assume the minus components close to zero, and the minus component does not vary with the average components, which indicates that most samples should not differ in molecular abundances. This assumption, together with the overreliance on the minus-average plot for detecting systematic bias, leads to the conclusion that the scale normalisation and its variations are not appropriate for the age-OME project.

[i] The median is defined as $median_i = \{m : \sum_{\mathbf{x}_i \leq m} i \geq N/2 \ \wedge \sum_{\mathbf{x}_i \geq m} i \geq N/2\}$, where N is the length of a numerical vector $\mathbf{x}$

Quantile normalisation Quantile normalisation is a well-established framework for microarray data[83], and its use has been extended for RNA sequencing and MS measurements[943]. Its aim is to map omics measurements across samples on common distributions in order to remove systematic bias[943]. In the original formulation[83], quantile normalisation requires the transformed omics measurements $\mathbf{Z}_{\cdot,j}$ in each j-th sample to be ranked from the lowest to the highest abundance. The ranked samples are denoted as $\mathbf{Z}^{rank}_{\cdot,j}$. Once all the samples are ranked and arranged in the matrix $\mathbf{Z}^{rank}$, the i-th row corresponds to the analyte with i-th lowest abundance level in each sample. Notably, this normalisation keeps track of the original position of omics measurements. The values from the common distribution, $\mathbf{q} \in \mathbb{R}^J$, are generated by averaging the element on each row of $\mathbf{Z}^{rank}$; in other words, the i-th element of the vector $\mathbf{q}$ is calculated as

$$\mathbf{q}_i = \frac{1}{J} \sum_J \mathbf{Z}^{rank}_{i,j}, \tag{12.8}$$

by definition $\mathbf{q}_j \leq \mathbf{q}'_j$ for any $j < j'$. Successively, each ranked row in $\mathbf{Z}^{rank}$ is substituted with the vector $\mathbf{q}$, generating the matrix $\mathbf{Y}^{rank}$, where $\mathbf{Y}^{rank}_{\cdot,j} = \mathbf{Y}^{rank}_{\cdot,j'} = \mathbf{q}^T$, for any $j \neq j'$. As a final step of the quantile normalisation, data $\tilde{\mathbf{Y}}$ are re-arranged in the same order as the original transformed measurements $\mathbf{Z}$, generating the matrix of quantile normalised measures, $\mathbf{Y}$ where each sample (column) exhibits the same median. Recent study has suggested quantile normalisation within each cohort if samples are collected across different cohorts. This allows researchers not to remove the characteristic variability that is due to phenotypical or genotypical traits of the cohort[943]. Although quantile normalisation is widely applied and can effectively reduce systematic variability, the convergence of measurements of the analytes in each sample to a unique distribution can compromise the biological variation of the samples.

Probabilistic quotient normalisation Probabilistic quantile normalisation was originally proposed for normalising metabolite measurements obtained through the nuclear magnetic resonance[193] and adapted successively for MS data for metabolomics and lipidomics[770]. It aims to reduce the systematic error by multiplying each metabolite in a sample by a common constant. Unlike other types of normalisation, probabilistic quantile normalisation is applied before the transformation of omics measurements (see Section 4.5). Using the notation from Section 4.5, let $\tilde{\mathbf{Z}}$ be the matrix containing omics measurements, where each column j is a sample and each row i is a detected molecule. The first step of probabilistic quantile normalisation is to divide each spectrum by the total ion count received by the detector and multiply it by 100, after which the results are stored in matrix $\tilde{Y}^{step1}$. The j-th column of the matrix consists in

$$\tilde{\mathbf{Y}}^{step1}_{\cdot,j} = \frac{\tilde{\mathbf{Z}}_{\cdot,j}}{\sum_i \tilde{\mathbf{Z}}_{i,j}} \cdot 100. \tag{12.9}$$

Once the spectra are normalised for the total ion count, the reference spectrum, $\mathbf{s}$, is calculated by taking the median of each column; the j-th element of the reference spectrum is

$$\mathbf{s}_j = median(\tilde{\mathbf{Z}}_{\cdot,j}). \tag{12.10}$$

The reference spectrum plays a role in expressing each sample on a common base. The next step of probabilistic quotient normalisation is to compute the ratio matrix, $\tilde{\mathbf{Y}}^{step2}$, where each sample is divided by the reference spectrum. The element in the (i, j) position of the ratio matrix $\mathbf{Y}^{\tilde{step2}}$ is defined as

$$\tilde{\mathbf{Y}}_{i,j}^{step2} = \frac{\tilde{\mathbf{Y}}_{i,j}^{step1}}{\mathbf{s}_j}. \tag{12.11}$$

The ratio matrix is then employed to compute normalising quotients for each sample j. The vector of normalising quotients is denoted by $\mathbf{q}$, and its j-th element is calculated as

$$\mathbf{q}_j = median(\mathbf{Y}_{\cdot,j}^{step2}). \tag{12.12}$$

Then, normalising quotients are employed to calculate the normalised but non-transformed omics measurements by dividing each original sample by the corresponding quotient. Denoting normalised omics measurements by $\tilde{\mathbf{Y}}^{norm}$, the j-th column of the matrix reads

$$\tilde{\mathbf{Y}}_{\cdot,j}^{norm} = \frac{\tilde{\mathbf{Y}}_{\cdot,j}^{step2}}{\mathbf{q}_j}. \tag{12.13}$$

The transformed and normalised omics measurements $\mathbf{Y}$ are then obtained by $\log_2$, transforming the element in $\tilde{\mathbf{Y}}^{norm}$. Probabilistic quantile normalisation is frequently applied in metabolomics and lipidomics studies[770]. It overcomes the overreliance of the normalisation on the minus-average relation and does not impose a common distribution on all samples. Probabilistic quantile normalisation assumes that systematic bias in each sample can be addressed by rescalling each sample measure for a certain constant although this may not always happen, as discussed earlier for cyclic loess normalisation. A recent study has exhibited that probabilistic quotient normalisation does not appear to reduce systematic bias in data, as its performance is comparable to the data transformation such as analyte-wise standardisation[819] (see Section 4.5), which rejected probabilistic quantile normalisation for the age-OME project.

12.3 Handling missing measurements - Extended

Median imputation Assuming that the missing value in the transformed and normalised matrix of omics measurements occurs in correspondence of the i-th analyte and j-th sample, the missing value is then substituted by the median value calculated analyte-wise. In other words,

$$\mathbf{Y}_{i,j} = median_{j \in J'}(\mathbf{Y}_{i,j}), \tag{12.14}$$

where J' is the set of samples with the i-th analyte's measurement recorded. The median imputation is relatively straightforward and simple and a preferred metric when dealing with a small number of samples since the mean value can be biased by the presence of a few extreme values[699]. As a limitation, the median imputation relies on substituting the missing value with the median of the analyte. Although the imputation of missing values with the median is computationally efficient and easy to implement, literature has shown that it can also introduce greater bias in the data when compared to other imputation methods[196,874].

Mean imputation A similar alternative to the median imputation is mean imputation. Assuming that the missing value occurs in position (i, j) in the transformed and normalised data matrix $\mathbf{Y}$, the missing values is then calculated as

$$\mathbf{Y}_{i,j} = \frac{1}{n(J')} \sum_{j \in J'} \mathbf{Y}_{i,j} \tag{12.15}$$

where J' is the set of samples without missing values in correspondence with the i-th analyte. The $n(J')$ represents the cardinality of the set. Similar to median imputation, mean imputation is a pointwise imputation method with simplicity. Nonetheless, it assumes that the missing value is equal to the average of the analyte, which leads to the underestimation of the variability of the phenomenon[779]. Computational experiments have proven that mean imputation returns less reliable imputed values compared to MICE and kNN.

Zero imputation While mean and median normalisation methods are applied on transformed and normalised dataset $\mathbf{Y}$, zero imputation is applied on raw ion counts data $\tilde{\mathbf{Z}}$. The zero imputation presumes that values are missing since they are absent in the samples. Consequently, assuming that $\mathbf{Z}_{i,j}$ has a missing entry, the method substitutes the missing value with 0, that is

$$\tilde{\mathbf{Z}}_{i,j} = 0. \tag{12.16}$$

The close-zero imputation is a variant of zero imputation attribute to the missing value a near zero value, such as 10^{-3}. The choice of using the close-zero instead of the zero imputation depends on the choice of data transformation on the imputed data (see Section 4.5). While the zero imputation is very easy to implement, it regards the i-th analyte as absent in the j-th sample. Additionally, the procedure ignores other possible generative mechanisms of missing values in the data, making it a suboptimal option for imputing missing values[874].

MICE imputation The MICE strategy gives competitive results compared to the kNN imputation method and addresses the point-estimate limitation by providing a series of likely values for the same missing data. The MICE algorithm is a recursive procedure and generates multiple data matrices, each containing a different value in correspondence with the missing observation. MICE relies on the point estimate of the missing values first and then on a regression model to estimate the missing value in the analyte i by using analyte i' for all $i' \neq i$. The predicted missing values are then randomly perturbed by a random error. The procedure returns multiple estimates for each missing value in the data matrix[36]. While the MICE technique allows to address the uncertainty behind each missing value, this method is based on multiple regression models and is computationally infeasible when multiple large data matrices are processed. Additionally, it also increases the computational burden in the follow-up analyses; each data matrix generated by MICE has to undergo the same set of analyses to combine the results. Furthermore, computational results point to kNN as a simpler and faster alternative to MICE for avoiding bias in the data[196].

Chapter 13

Appendix - Transcriptomics Quality Assessment

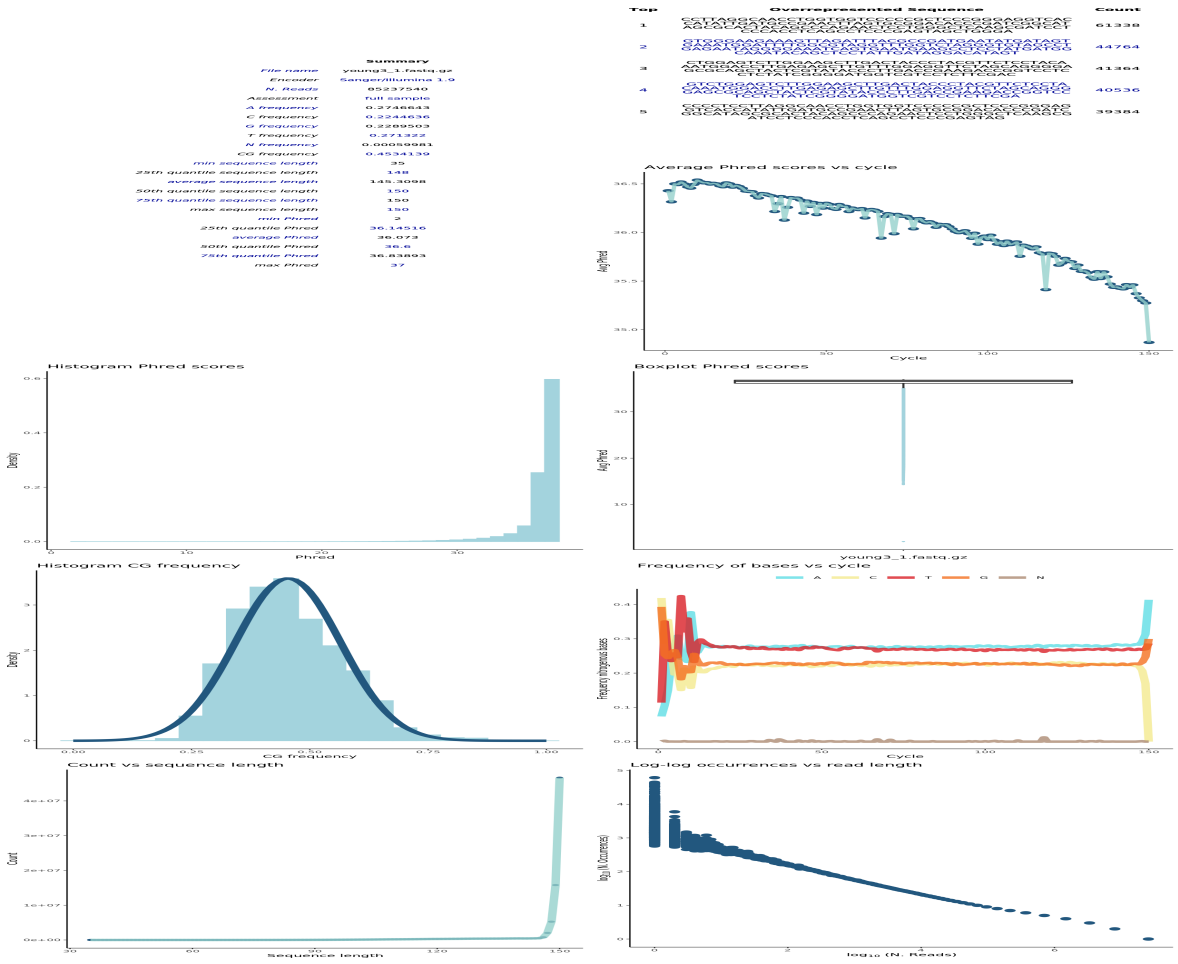


Figure 13.1: Transcriptomics. Quality control first strand of sample *Young 3*

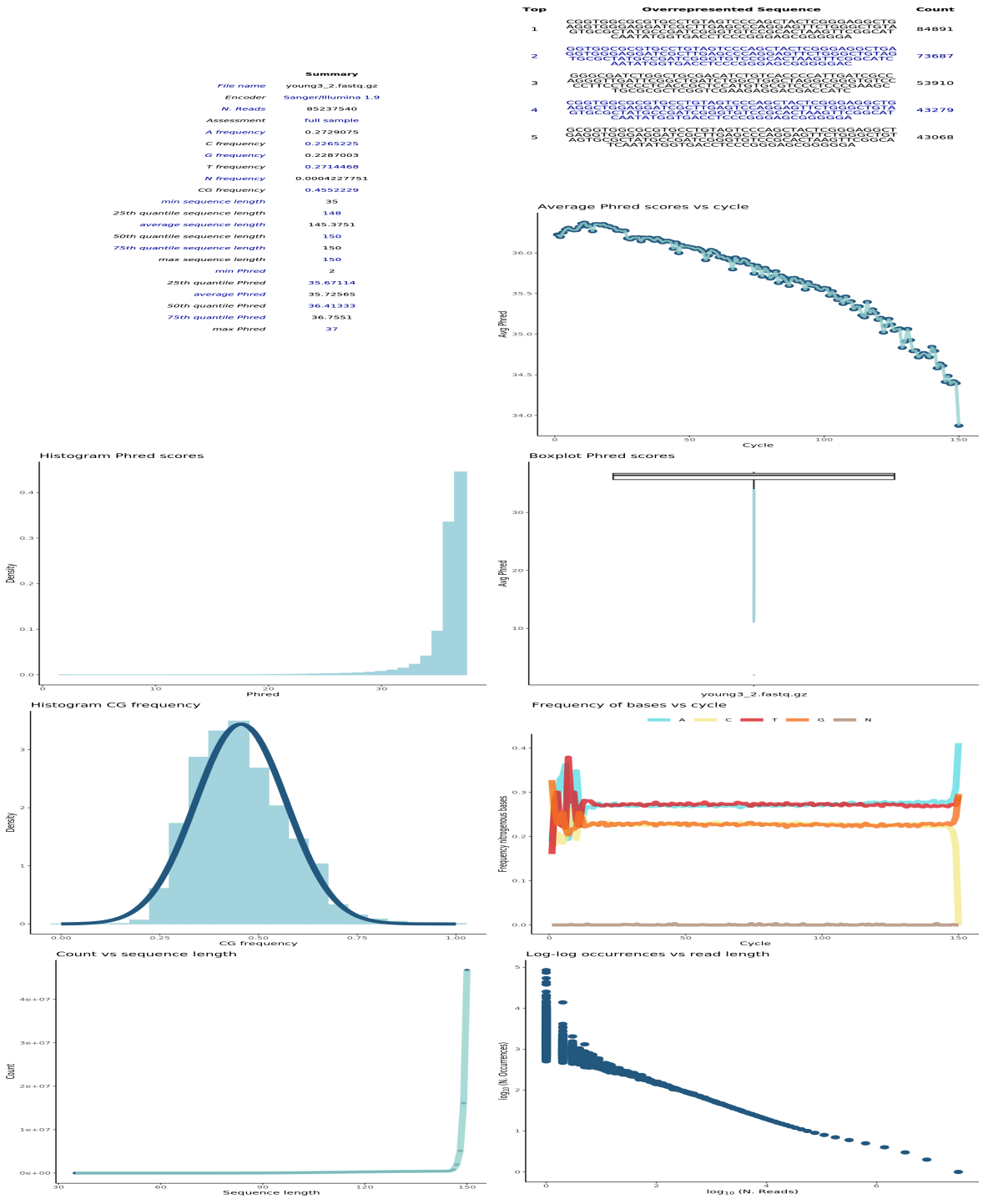


Figure 13.2: Transcriptomics. Quality control second strand of sample *Young 3*

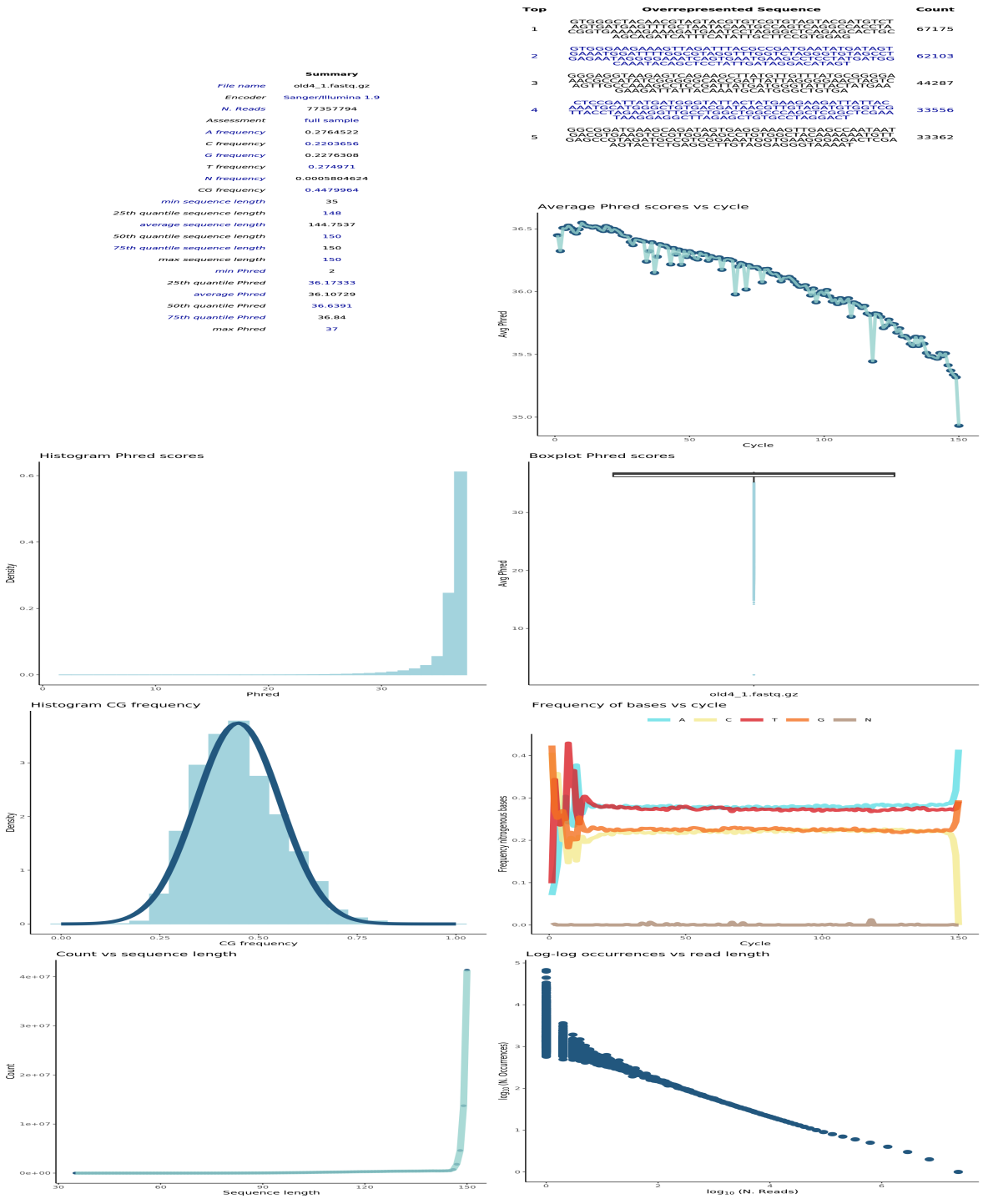


Figure 13.3: Transcriptomics. Quality control first strand of sample *Old 4*

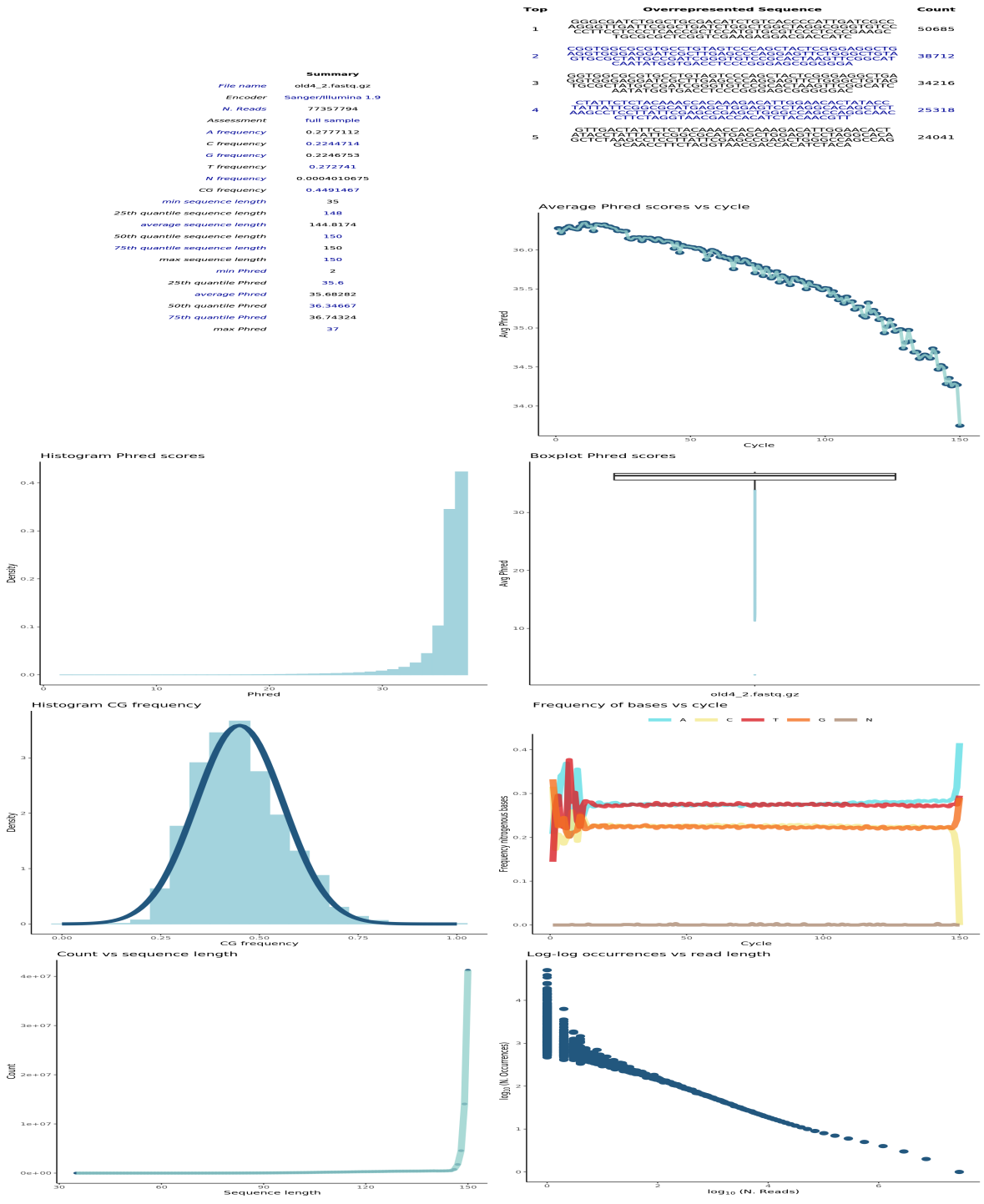


Figure 13.4: Transcriptomics. Quality control second strand of sample *Old 4*