
University of Glasgow

# Identifying Early Genetic Determinants of Adverse Cardiac Remodelling in Complex Genetic Models of Human Cardiovascular Disease

Cara Trivett MRes BSc (Hons)

Submitted in fulfilment of the requirements for the Degree of Doctor of Philosophy (PhD)

School of Cardiovascular and Metabolic Health

College of Medical, Veterinary and Life Sciences
University of Glasgow

May 2024

# Abstract

Globally, cardiovascular diseases (CVD) are the leading cause of morbidity and mortality. High blood pressure, or hypertension, is the primary modifiable risk factor contributing to the development of CVD. An independent risk factor of cardiac and all cause mortality is an enlargement of the heart, specifically the left ventricle (LV), clinically defined as left ventricle hypertrophy (LVH). Accounting for known biological and environmental factors, a large part of variability in LV mass is unexplained, and is thought to come under genetic control. Studies of heritability estimate structural measures of the heart are highly heritable, ranging between 15–60%. Translation of genetic variation into mechanisms underlying progression of human disease can be methodologically difficult, and thus far understanding of the genetic architecture of LV mass is not well-developed. Genetically determined differences are required to alter molecular phenotypes to result in measurable differences in cardiac structure. Unravelling connections between genetic variation and pathways of pathology are therefore key to defining genetic contribution to LVM index (LVMI). The rat is an effective small rodent model to recapitulate human disease phenotypes without surgical or dietary intervention. Genetic models of hypertension include the Stroke Prone Spontaneously Hypertensive Rat (SHRSP) and its closest genetic comparator, Wistar Kyoto (WKY).

To determine potential genetic factors underlying the development of LV hypertrophy, a linkage study in WKY and SHRSP rats identified a region of chromosome 14 independently associated with LVMI. Following this, the chromosome 14 region associated with LVMI was the target of congenic strain generation, where a segment of chromosome 14 was introgressed from donor WKY and SHRSP strains into the opposing strain (SP.WKYGla14a and WKY.SPGla14a respectively). Initial phenotyping, in male animals, show cardiac phenotypes of chromosome 14 congenic strains diverged from their background strain, with measurable differences detectable from as early as 5-weeks. The WKY.SPGla14a developed increased LVMI without an excessively increased blood pressure, whilst the SP.WKYGla14a developed increased blood pressure without significant remodelling of the LV. Following the development and establishment of the chromosome 14 congenic strains on both normotensive and hypertensive backgrounds, work undertaken in this thesis comprised a series of investigations *in-vivo*,

*ex-vivo*, *in-vitro*, and *in-silico* to determine potential key genetic factors and molecular effectors predicting and resulting in excessively increased LVMI in rat models of human hypertension.

Short-read DNA sequences, from WKY and SHRSP colonies housed in Glasgow, were aligned to the recently published SHRSP/BbbUtx reference quality genome. High quality variant calling analysis identified a number of key sites of variance between SHRSP and WKY strains, which could be contributing to differences in cardiac gene expression and LV structure. These include localised clusters of high impact variants within previously identified QTLs for cardiac mass on chromosomes 1 & 3. Prior to this work, secreted phosphoprotein 1 (*Spp1*) was identified as a positional and functional candidate gene governing increased LVMI associated with the SHRSP chromosome 14 region. Variants within the *Spp1* gene and upstream of its transcriptional start site were fine mapped between the WKY and SHRSP strains for future investigations determining causal variants underlying observed differences in expression of *Spp1* in cardiac tissues.

Expanding on initial phenotyping, *in-vivo* studies were conducted to determine the effect of introgression of SHRSP 'risk' haplotypes into the 'protected' WKY background (and vice versa), in both female and male congenic strains. Phenotype assessment in neonate, 5-, and 16-week animals included transthoracic ultrasound echocardiography (5- & 16-week) and tail cuff plethysmography (16-weeks). Male and female SHRSP animals developed a significantly increased LVMI compared to sex- and age-matched WKYs by 5-weeks of age. This was accompanied by a persistent upregulation of *Spp1* mRNA in the LV, which was more pronounced in females at 16-weeks of age. The chromosome 14 congenic strains did not develop a cardiac phenotype which differed significantly from their respective parental strain. Despite this, *Spp1* mRNA expression was increased in the LV of strains harbouring SHRSP genome at the chromosome 14 locus. The increase in *Spp1* expression in SHRSP and WKY.SPGla14a hearts persisted across all measured time-points; gestational day 18.5, 1–3 day neonate, 5-, and 16-weeks.

To better understand biological pathways altered by genetic variation in parental and congenic strains, short read RNA sequencing was performed in the gestational day 18.5 heart. Although prior to the development of divergent cardiac phenotypes, introgression of the SHRSP genome significantly and dramatically altered gene expression and transcript usage during cardiac development. Alignment of the RNA short read sequences to the SHRSP/BbbUtx, revealed variance within genes that also displayed evidence of significant differential expression and transcript usage. The hearts of SHRSP and WKY.SPGla14a strains are potentially primed for future dysfunction through dysregulation of genes to increase mitochondrial dysfunction and decrease oxidative phosphorylation. In addition, key receptors and signalling factors involved

in the maintenance and production of the extracellular matrix were dysregulated in SHRSP and WKY.SPGla14a hearts during development.

Molecular investigations focused on *Spp1*, an extracellular matrix protein and a known biomarker of established heart disease. Adapting a H9c2 cell model used in Angiotensin-II (Ang-II) studies, transfection of H9c2 cells with vectors containing *Spp1* mRNA increased cell size 48-hours post-transfection compared to pcDNA controls. The H9c2 cell model was further expanded to investigate the role of small extracellular vesicles (sEV). These membrane-bound vesicles (<200nm) are produced by cardiac and other cells of the heart for cell-to-cell communication and signal transduction, carrying nucleotide species and/or protein messengers to mediate physiological and pathological processes. sEVs derived from cells transfected with *Spp1* were applied to naïve H9c2 cells resulting in an equal increase in cell size to direct transfection of *Spp1*. This increase in cell size following incubation with sEV was blocked when cells were treated with 60µM of EV uptake inhibitor, dynasore.

Finally, characterisation of a CRISPR/Cas9 *Spp1* knock-out on the SHRSP strain suggested alternative transcripts of *Spp1* are produced in the rat cardiac transcriptome. The alternative transcript identified resembled human short forms of *SPP1* known as osteopontin-c. Alternative *Spp1* transcripts produced by the knock-out removes the exon containing point mutations and may rescue osteopontin protein expression in these animals. In the H9c2 model, overexpression of *Spp1* transcripts produced by the knock-out have equal function to canonical *Spp1*.

Taken together, these data have implicated a role for sEV in the pathway of *Spp1* overexpression that results in H9c2 cell size to increase. Genomic and transcriptomic studies support a role of *Spp1* dysregulation underpinning differences in cardiac phenotypes associated with the SHRSP chromosome 14 region. These studies have generated a basis to further validate and translate findings in described experimental models. The wealth of genomic data provided by large consortia such as GTEx, the GWAS catalogue, and the UK BioBank represent rich resources to combine and correlate findings from genetic rat models to targets in human disease.

# Contents

# List of Tables

# List of Figures

# Acknowledgement

The first and biggest acknowledgement of this thesis must go to my supervisors, Martin McBride and Delyth Graham, who have been endlessly supportive of my ambitions, both in my PhD project and in other aspects of science. I have them to thank for my personal and professional development in academia; their patience, understanding, and depth of knowledge was indispensable. I hope, in them, I have found mentorship and friendship that will remain long after the conclusion of this period of study.

I would also like to acknowledge the British Heart Foundation for their funding of the PhD programme, and the incredible opportunity the 4-year scholarship has afforded me. Without the mentorship of fellow students, their friendship, support, and laughter through the high and lows of science, much of the work herein could not have been possible. To name everyone in the School who has helped me in some way would take far too long, so thank you to everyone who made me feel part of a dynamic and supportive community. A special thanks should go to my lab mates past and present– Simon, Kayley, Amrita, and Manshi, I wish you all every success in the world. The last of my 'professional' thanks must go to the staff, and colleagues, in the Cardiovascular Research Unit, Glasgow Tissue Research Facility, and across the pond at Medical College of Wisconsin. I learnt so much over the course of my PhD and their patience in teaching me is greatly appreciated.

To my Mum and Dad, who never fail to keep my feet firmly on the ground, and reassure me that the world is not falling, thank you for never letting me feel like there was anything I couldn't do. To my sister, Cassandra, whose bravery inspires me every day, and whose love of science sparked my own, thank you. I came to Glasgow on my own, and found friends who became my family; Gemma, Rob, Kieran, and Jane, I could not have come through the pandemic or the final years of my PhD without you. Also friends who were far from me physically but never failed to make me cry with laughter and sing with joy, Hannah, Jen, Kiki, Alex, and Phil, thank you all for your support.

The final thank-you is for my partner Jamie and I don't know where I should start. Your dedication and passion for your PhD helps to remind me of my own. You've shared in the stress of it all and celebrated the successes. As my emotional, technical, and IT support, this thesis really would not have been possible without you.

# Authors Declaration

I declare that this thesis has been written by myself and is a record of research performed myself. For certain experiments, other students were involved in practical aspects of the work and were present during these processes. Calum McGeoch and Abby Shaw (BSci(Hons)) were present during RNA and protein extraction from SHRSP-*Spp1* knock-out models and performed cDNA preparation and end-point PCR under my supervision. Nicola Gilroy (MRes) carried out functional assays of alternative *Spp1* transcripts in H9c2 cells with my supervision. Roberta Munro (BSci (Hons)) analysed a selection of ultrasound echocardiography images as the secondary investigator to improve reliability of measurements. All data obtained from work carried out by students was analysed and prepared for this thesis by myself.

Glasgow Polyomics facility performed all library preparations and all next-generation sequencing reactions within this thesis, but did not process samples prior to library preparations. Glasgow polyomics performed the initial analyses of RNA sequencing experiments, but not the downstream analysis or visualisation.

Dr Hannah Morgan and Dr Jennifer Hay from the Glasgow Tissue Research Facility (GTRF) performed tissue sectioning for immunofluorescence immunohistochemistry assays. The team at GTRF also provided technical support for single and multiplex assays. Multiplex analysis was prepared with the help of Molly McKenzie, at the Beatson Institute.

Genomic analyses were done in collaboration with the Rat Genome Database at the Medical College of Wisconsin, with support from MVLS Skills Training Award and Graham Wilson Scholarship.

Contribution from others has been clearly referenced and reproduced with permission. This work has not been submitted previously for a higher degree and was supervised by Dr Martin W. McBride and Dr Delyth Graham.

# COVID-19 Impact Statement

Experimental work contained with this thesis was begun in October 2021, during the coronavirus-19 (COVID-19) pandemic. The pandemic altered the research environment globally, the extent of which was determined by the nature of the research, and the specific geographical location it took place within. The *in-vivo* aspects of data collection were affected by both the impact of local restrictions and the requirement of close contact isolation which continued into the middle of 2022.

Although outside the period of UK wide lockdown in March– April 2021, significant restrictions remained in place for the first 6–12 months of experimental work. These resulted in restricted access to training in key *in vivo* methods employed in Chapter 3 and Chapter 6, as well as *in vitro* methods included in Chapter 5. Access to the laboratory space and animal unit was severely restricted and progress of experimental work was slow as a result. Mistakes, which would not have occurred if the laboratory had been more populated, were more frequent, and engagement in the scientific discussions which drive creative thinking in experimental research was reduced. Cumulatively, these reduced research output during the first year of experimental work.

The impact of these restrictions was partially mitigated by access to previously generated RNA microarray data from the genetic rat models employed within this thesis. Although not directly included in results chapters presented, the bioinformatic skills developed during this period became an integral part of RNA-sequencing and variant calling analyses in Chapter 7 and Chapter 8.

# Abbreviations

**ACE** Angiotensin-converting enzyme.

**ADAM10** A Disintegrin And Metalloproteinase Domain-containing protein 10.

**Ang-II** Angiotensin-II.

**ANOVA** Analysis of Variance.

***B2m*** beta-2-microglobulin.

**BCA** Bicinchoninic acid assay.

**CI** Confidence interval.

**CO** Cardiac Output.

**CRISPR** Clustered regularly interspaced short palindromic repeats.

**DAB** 3,3' Diaminobenzidine.

**DGE** Differntial gene expression.

***Dmp1*** Dentin matrix protein 1.

**DMSO** Dimethylsulfoxide.

***Dspp*** Dentin sialophosphoprotein.

**DTU** Differntial transcript usage.

**ECDF** Empirical Cumulative Distribution Function.

**EDTA** Ethylenediaminetetraacetic acid.

**EDV** End Diastolic Volume.

**EF** Ejection Fraction.

***Egfr*** Epidermal growth factor.

**Egr-1** Early growth response factor 1.

**ELISA** Enzyme-linked immunosorbent assay.

**ERK** Extracellular signal-regulated kinase.

**ESM-1** endothelial cell-specific molecule-1.

**ESV** End Systolic Volume.

**FFPE** Formalin-fixed paraffin embedded.

**FS** Fractional Shortening.

**GATA4** GATA binding protein 4.

**GSEA** Gene Set Enrichment Analysis.

***Gstm1*** Glutathione S-transferase µ-type 1.

**GWAS** Genome Wide Association Study.

**HHR** Hypertrophic Heart Rat.

**HRP** Horseradish peroxidase.

***Ibsp*** Integrin binding bone sialoprotein.

**IGF** Insulin-like growth factor.

**IGFR-1** Insulin-like growth factor receptor 1.

**LH** Lithium heparin.

**LVH** Left Ventricle Hypertrophy.

**LVM** Left Ventricle Mass.

**LVMI** Left Ventricle Mass Index.

**MAPK** Mitogen-activated protein kinase.

**MCSF-1** macrophage colony-stimulating factor-1.

**MEF2A** Myocyte enhancer factor 2A.

***Mepe*** matrix extracellular phosphoglycoprotein.

**MES** 2-(N-morpholino)ethanesulfonic acid.

**miRNA** mirco-RNA.

**mRNA** messenger RNA.

**mRNA** messenger RNA.

**MTT** 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide.

**NAD+** Nicotinamide adenine dinucleotide (+).

**NADH** Nicotinamide adenine dinucleotide.

**ncRNA** non-coding RNA.

**NFAT** nuclear factor of activated T cells.

**NHR** Normal Heart Rat.

***NPPA*** Natriuretic Peptide A.

***NPPB*** Natriuretic Peptide B.

**NT-proBNP** N-terminal pro B-type natriuretic peptide.

**NTA** Nanoparticle Tracking Analysis.

**OPN** osteopontin.

**ORA** Over Representation Analysis.

**PBS** Phosphate-buffered Saline.

**PCA** Principal Component Analysis.

**PCR** Polymerase chain reaction.

**PDGF** Platelet derieved growth factor.

**PFA** paraformaldehyde.

**PI3K** Phosphoinositide 3-kinases.

**PPAR** Peroxisome proliferator-activated receptors.

**RBM20** RNA binding motif protein 20.

**RIPA** Radioimmunoprecipitation assay buffer.

**RNA** Ribonucleic acid.

**rRNA** ribosomal RNA.

**RWT** Relative Wall Thickness.

**s-UC** Sucrose cushion ultracentrifugation.

**sEV** Small extracellular vesicles.

**sgRNA** single-guide RNA.

**SHR** Spontaneously Hypertensive Rat.

**SHRSP** Stroke Prone Spontaneously Hypertensive Rat.

***Spp1*** secreted phosphoprotein 1.

**STAT** Signal transducer and activator of transcription.

**SV** Stoke Volume.

**TALEN** Transcription activator-like effector nuclease.

**TBE** Tris-borate EDTA.

**TBS** Tris-buffered Saline.

**TBS-T** Tris-buffered Saline-Tween.

**TGF-β** Transforming growth factor-β.

**TSG101** Tumor susceptibility gene 101.

**UC** Differential ultracentrifugation.

**VEGF** Vascular endothelial growth factor.

**WKY** Wistar Kyoto.

**ZFN** Zinc-finger nuclease.

# List of Publications, Conferences and Awards

**Trivett, C.**, Lees, Z.J. & Freeman, D.J. Adipose tissue function in healthy pregnancy, gestational diabetes mellitus and pre-eclampsia. *Eur J Clin Nutr* 75, 1745-1756 (2021). https://doi.org/10.1038/s41430-021-00948-9

Fullerton, J.L., Thomas, J.M., Gonzalez-Trueba, L., **Trivett, C.**, Van Kralingen, J.C., Allan, S.M., Quinn, T.J. and Work, L.M., Systematic review: Association between circulating microRNA expression & stroke. *Journal of Cerebral Blood Flow & Metabolism*, 0(0) 1-17 (2022) https://doi.org/10.1177/0271678X221085090

### Oral

- Complex Trait Consortium 2023, Hybrid. Quantifying the Early Cardiac Transcriptome in Rat Models of Left Ventricular Hypertrophy with or without Predisposition to Hypertension. **Trivett, C**, Graham, D., McBride, M.W.

- Scottish Cardiovascular Forum 2023 Aberdeen, UK. Functional characterisation of alternative osteopontin transcripts from the stroke-prone spontaneously hypertensive rat heart in H9c2 cells. *Heart Apr 2023, 109 (Suppl 2) A1; DOI: 10.1136/heartjnl-SCF-2023.3.* **Trivett, C.**, Gilroy, N., Graham, D., McBride, M.W.

- Complex Trait Consortium 2022, Online. Alternative Splicing of Osteopontin in CRISPR/Cas9 Knock Out Rats. **Trivett, C.**, Graham, D., McBride, M.W.

- British and Irish Hypertension Society 2022, York UK. O-12 Investigating the Effects of CRISPR/Cas9 Gene Editing on Cardiac Structure and Function in a Rat Model of Cardiovascular Disease. **Trivett, C.**, Asirvatham, A., Graham, D., McBride, M.W.

### Poster

- GRC Extracellular Vesicles 2022, Boston, USA. The Role of Small Extracellular Vesicles in the Functional Effects of Spp1 Overexpression in Cardiac Cells. **Trivett, C.**, Asirvatham, A., Graham, D., McBride, M.W. July 2022

- UK Extracellular Vesicles Forum 2021, Online. Comparison of Isolation Methods and Effect of Small Extracellular Vesicles Derived from Rat Cardiac H9c2 Cells Overexpressing Spp1. **Trivett, C.**, Asirvatham, A., Graham, D., McBride, M.

December 2021

- Scottish Cardiovascular Forum 2021, Online. Comparing the Uterine Artery Long Non-Coding RNA Profile in WKY and SHRSP Rats During Early Pregnancy. **Trivett, C.**, Scott K., Morgan H., McBride, M., Graham, D.

Skills Training Award and Graham Wilson Travel Scholarship (University of Glasgow, School of Cardiovascular and Metabolic Health). **May 2023**. Awarded £2,350 to enable a collaboration with the Rat Genome Database at the Medical College of Wisconsin for -omic data skills training (NGS data processing and analysis).

ECR Pump Priming Grant (funded by Wellcome Institutional Strategic Support Fund (ISSF), University of Glasgow). **August 2022**. £800 awarded for multiplex IHC in collaboration with Glasgow Tissue Research Facility

Feasibility Award (funded by Wellcome Institutional Strategic Support Fund - University of Glasgow). **June 2022**. Awarded £7,943 awarded to fund an RNA sequencing project with Glasgow Polyomics.

# Chapter 1

# Introduction

Cardiovascular disease (CVD) is an umbrella term describing a collection of pathological conditions involving the heart and/or the vasculature. In broad terms, CVD can be divided into coronary heart disease, myocardial disease, cerebrovascular disease, and peripheral artery disease (Figure 1.1). The loss of health, cost to the health system, and mortality caused by CVDs remains one of the biggest threats to healthcare both in the UK and around the world (Vaduganathan et al., 2022).

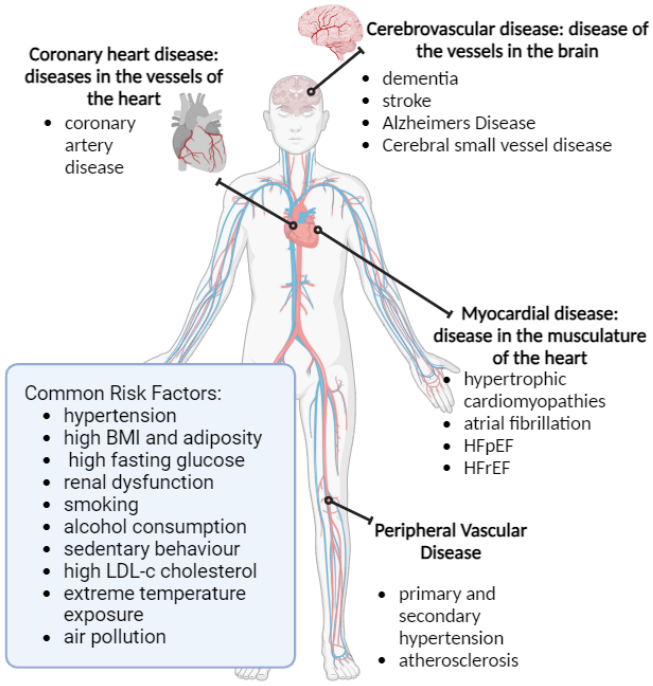



**Figure 1.1: Overview of Cardiovascular Diseases** CVD is an umbrella term encompassing a number of diseases broadly categorised into cerebrovascular, myocardial, peripheral, and coronary heart disease based on their primary site of injury. Examples of specific clinical diagnoses included on the infographic are not extensive. BMI; body mass index, LDL-c; low density lipoprotein cholesterol, HFpEF; heart failure with preserved ejection fraction, HFrEF; heart failure with reduced ejection fraction. Figure created in BioRender.com

The first Global Burden of Disease (GBD) was published in 1990. By 1993, 'cardiovascular diseases' had become the largest single contributor to GBD and continues to contribute to the loss of over 300 million Disability Adjusted Life Years (DALYs) every year (Roser, Ritchie, and Spooner, 2021) (Figure 1.2). Along with cancer, neonatal disorders, and respiratory infections, CVD contributed to the largest % of DALYs, globally in 2019 (Figure 1.2). In the UK, CVD contributed to 14.9% of DALYs and was the second largest contributor to disease burden over cancer (19.1%) (Roser, Ritchie, and Spooner, 2021).

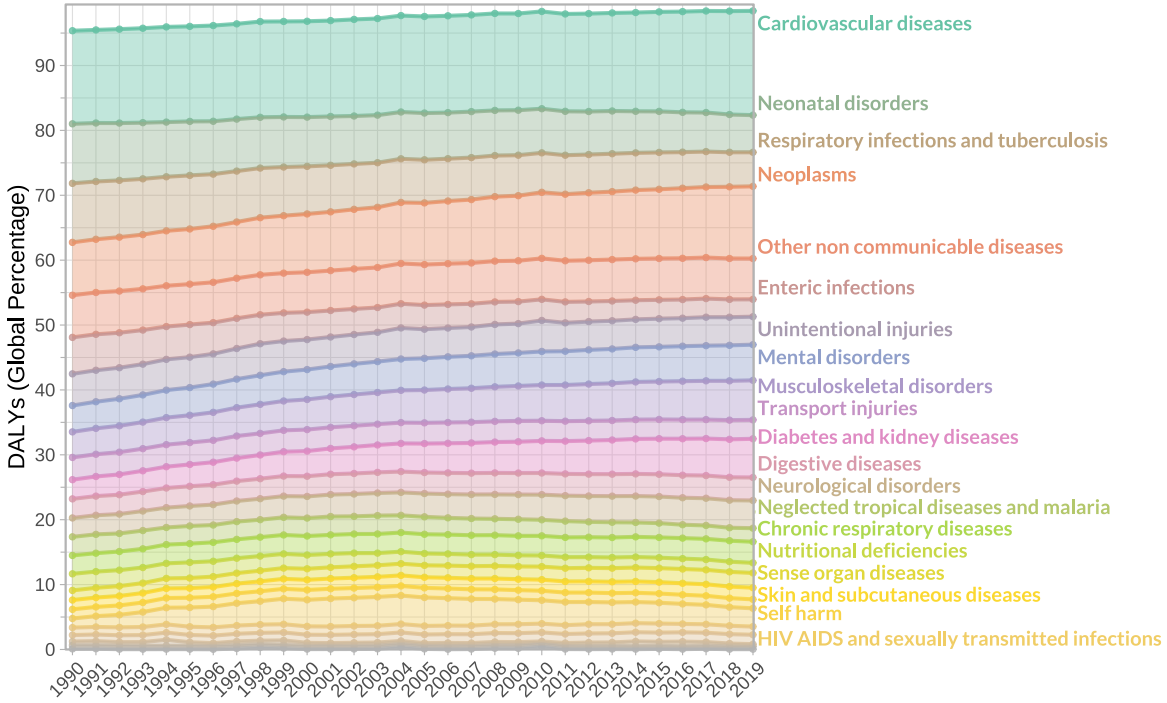



**Figure 1.2: Global Burden of Cardiovascular Disease** Global burden of disease measured in Disability Adjusted Life Years (DALYs). Reproduced using custom R script and data downloaded from Our World in Data `https://ourworldindata.org/burden-of-disease`

## 1.1 Blood Pressure as a Modifiable Risk Factor for CVD

CVDs affecting different organ systems share common risk factors and etiologies (Figure 1.1). The leading environmental, metabolic, and behavioural risks include elevated blood pressure (hypertension), high body mass index and adiposity, poor glucose control, renal dysfunction, tobacco smoking, alcohol use, physical inactivity and sedentary behaviour, high LDL-c cholesterol, extreme temperature exposure, and air pollution (ambient particle matter or household / solid fuel exposure) (Vaduganathan et al., 2022). High systolic blood pressure (SBP) is the top modifiable risk factor for premature death, globally, accounting for 10.8 million CVD deaths, and 11.3 million total deaths in 2021 (Vaduganathan et al., 2022). High blood pressure represents a particular risk for ischaemic heart disease (Razo et al., 2022) and according to the British Heart Foundation, 28% of adults in the UK are living with high blood pressure (BHF,

2023). Estimates published by the research charity predict at least half of individuals with raised blood pressure are not receiving antihypertensive treatment.

'Healthy' blood pressure is considered to be a SBP and diastolic blood pressure (DBP) of 120mmHg and 80mmHg respectively and hypertension is defined as a blood pressure >130/80mmHg, on multiple readings (Unger et al., 2020). Clinically speaking, primary hypertension is defined as a sustained elevation of SBP and DBP not secondary to an underlying disease. Conversely, hypertension that can be attributed to an underlying but separate disease is defined as 'secondary hypertension'. Hypertension is the clinically defined threshold of high blood pressure, where higher blood pressures act as a modifiable risk factor for CVD. To date, over 1,500 single nucleotide polymorphisms and ∼30 monogenic rare variants are known to comprise the genetic architecture of hypertension (Garimella et al., 2023). Mendelian Randomisation studies have also demonstrated genetically predicted blood pressure is causally-associated with incidence of CV conditions and CVD mortality (Wan et al., 2021). Data from the Framingham Heart Study indicated risk of developing CVD was almost triple in participants with a BP >130/85mmHg compared to those with good pressure control (<120/80mmHg) (Vasan, M.G. Larson, Leip, et al., 2001). Genetic risk scores, generated using >1,000 variants, were linearly associated with risk of incident CVD, myocardial infarction and stroke (E. Evangelou et al., 2018). As well as the increased risk of CVD, hypertension is also associated with 'hypertension-mediated end organ damage' (HMOD) (B. Williams et al., 2018). This damage is gradual, begins asymptomatically, and can include left ventricle hypertrophy (LVH) and micro-albuminuria. Instances of HMOD can themselves be classified as pre-clinical markers of CVD.

Hypertensive heart disease is a term applied to pathology of the heart, usually concerning the left ventricle, left atrium, or coronary arteries, which occurs as a result of sustained increases in blood pressure (Masenga and Kirabo, 2023). Common complications of sustained hypertension include diastolic or systolic heart failure, risk of atrial fibrillation, peri-operative ischaemia, coronary artery disease, heart valve disease, and aortic dissection (Masenga and Kirabo, 2023). High blood pressure increases afterload in the left ventricle and increases peripheral vascular resistance to the pumping of the heart. As a result, pressure- and volume-mediated structural remodelling of the left ventricle is often associated with hypertrophy of the left ventricle (Drazner, 2011; Oh and H.J. Cho, 2020; Masenga and Kirabo, 2023). This compensatory mechanism of LVH underpins almost all manifestations of hypertensive heart disease, contributing to risk of myocardial infarction, impaired contractility, reduced filling, and risk of sudden cardiac death due to arrhythmia development (Schmieder, 2010). Remodelling of the heart resulting in dysfunction is often defined as heart failure with preserved ejection fraction (HFpEF), or heart failure with reduced ejection fraction (HFrEF), and is a common CVD driven by hypertension, underlying approximately 17–22%, and 31% of

cases respectively (Savarese et al., 2023).

### 1.1.1 Current Treatments and Challenges in the Management of Hypertension and CVD

Medications which lower cholesterol, blood pressure, and/or reduce inflammation to improve CVD mortality are widely available, however, in >50% of patients, hospitalisation due to acute cardiac events is the first symptom of CVD (Cicha, 2021). It is estimated that 6–8 million people in the UK have undiagnosed or untreated high blood pressure (BHF, 2023). Other risk factors for CVD, including poor diet, low physical activity, smoking, and high adiposity are also common in the UK (BHF, 2023), increasing the potentially untreated risk of CVD mortality.

CVD risk management using risk score prediction to inform treatment is currently recommended by the European Society of Hypertension (Mancia, Kreutz, et al., 2023). Alongside recommendations to reduce blood pressure and CVD risk by changing lifestyle factors, pharmacological intervention to reduce blood pressure is begun immediately upon a diagnosis of hypertension (Mancia, Kreutz, et al., 2023). Currently, antihypertensive medications include, but are not limited to, angiotensin-converting enzyme inhibitors (ACE-I), angiotensin II receptor blockers (ARBs), thiazide-type diuretics, calcium channel blockers, mineralocorticoid receptor antagonists (MRA), α-blockers, and β-blockers. Blood pressure lowering is associated with decreasing asymptomatic CVD and CV mortality across a range of blood pressures, down to 120mmHg (Ettehad et al., 2016; Blood Pressure Lowering Treatment Trialists Collaboration, 2021; Mancia, Kreutz, et al., 2023).

The most recent ESH guidelines suggest 'controlled' management of hypertension is effective lowering of SBP to <130-139 mmHg (Figure 1.3). Even with effective antihypertensive treatment, CVD risk is not normalised in hypertensive patients (Mancia, Kreutz, et al., 2023). This 'residual risk' is of importance to understand in improving outcomes for hypertensive patients, highlighting the need for a potentially combined and holistic approach to hypertension and CVD management (Suzuki et al., 2021).

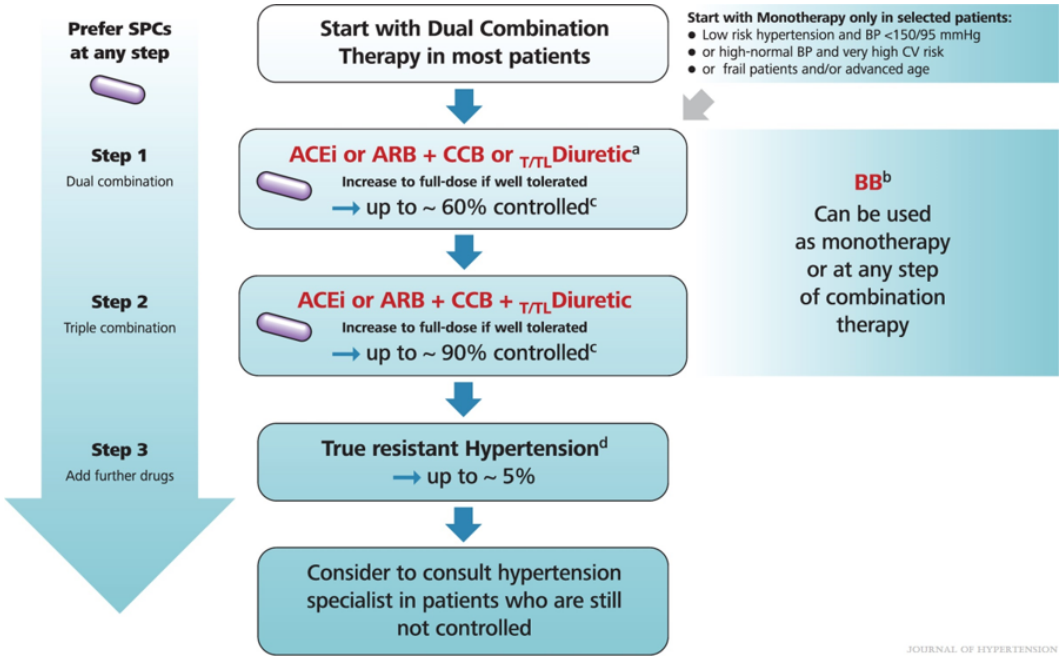



**Figure 1.3: ESH Guidelines for Hypertension Management** Taken from (Mancia, Kreutz, et al., 2023); General BP-lowering strategy in patients with hypertension. [a] Consider loop diuretic if eGFR is between 30–45 ml/min/1.73m$^2$. [b] BB should be used as guideline directed medical therapy in respective indications or considered in several other conditions. [c] BP controlled below 140/90mmHg. [d] when SBP is >140mmHg or DBP is >90 mmHg provided that maximum recommended and tolerated doses of a three-drug combination has been confirmed and various causes of pseudo-resistant hypertension (especially poor medication adherence) and secondary hypertension have been excluded. (Journal of Hypertension 41(12):1874-2071, December 2023.)

## 1.2 Left Ventricular Hypertrophy

Over the life-course, the heart will undergo many structural and functional alterations in response to changes in the internal and external environment. Left ventricular mass (LVM) is often indexed to body size, body surface area, or height (Bella and Goring, 2012; Taylor et al., 2023). Left ventricular hypertrophy (LVH) is defined as a pathological or physiological increase in LVM index (LVMI) in response to a stimulus (Aung, Vargas, et al., 2019), and can occur in response to a range of stimuli. Increases in LVMI can be stimulated by an increase in haemodynamic load as a consequence of an increased training load, pregnancy, or chronically elevated ambulatory blood pressures (Nakamura and Sadoshima, 2018). LVMI increases over the life-course, with population-based studies estimating LVMI increases ∼19–22% in the general population over a 10-year follow-up period (Bombelli et al., 2023). Age-associated increase in LVMI is potentially dependent on the presence or absence of additional risk factors, and stimuli for physiological remodelling (Dannenberg, Levy, and R.J. Garrison, 1989). Importantly, physiological adaptations to increased exercise, or uncomplicated pregnancies will result in hypertrophy associated with a concomitant enhancement of cardiac function (Nakamura and Sadoshima, 2018). Physiological improvements in-

clude an increased filling capacity of the heart and/or improved ability of contractile units to generate force (increased contractility). Increased diastolic filling and cardiac contractility result in increased cardiac output (volume of blood pushed out of the heart with each beat) and ejection fraction (proportion of blood returned to the LV removed by a single contraction), effectively increasing the ability of the heart to perform its function. Physiological LVH is also a reversible adaptation, which regresses after de-training or post-partum (Maillet, Berlo, and Molkentin, 2013). Conversely, pathological cardiac hypertrophy is associated with cardiomyocyte loss, and reduction in systolic and diastolic function, often progressing to heart failure (Shimizu and Minamino, 2016). Pathological LVH is also accompanied by an excessive increase in components of the extracellular matrix (ECM), impaired electrical conduction, reduced calcium uptake and release, and reduced myofibrillar and myosin ATPase activity (Oldfield, Duhamel, and Dhalla, 2020).

LV geometry can be altered through an increase in ventricle wall thickness and/or chamber dilation. The relationship between the chamber wall size and chamber volume is simply defined as the ratio between wall thickness and diastolic diameter known as relative wall thickness (RWT). Eccentric hypertrophy occurs when there is an increase in LVMI without a change to RWT (Figure 1.4). Concentric hypertrophy is described as an increase in LVMI, with a concomitant increase in RWT, normally as a result of reduced diastolic diameter (Figure 1.4). Before progression to LVH, an increase in cardiac RWT, without a signifiant increase in LVMI, is termed ‘concentric remodelling’. Both physiological and pathological hypertrophic growth can be eccentric, where both wall thickness and diastolic diameter are increased concomitantly. However, concentric growth tends to be pathological. The increased wall thickness acts initially to preserve cardiac contractility but impairs diastolic function and is likely to decompensate into either HFpEF or HFrEF (Drazner, 2011).

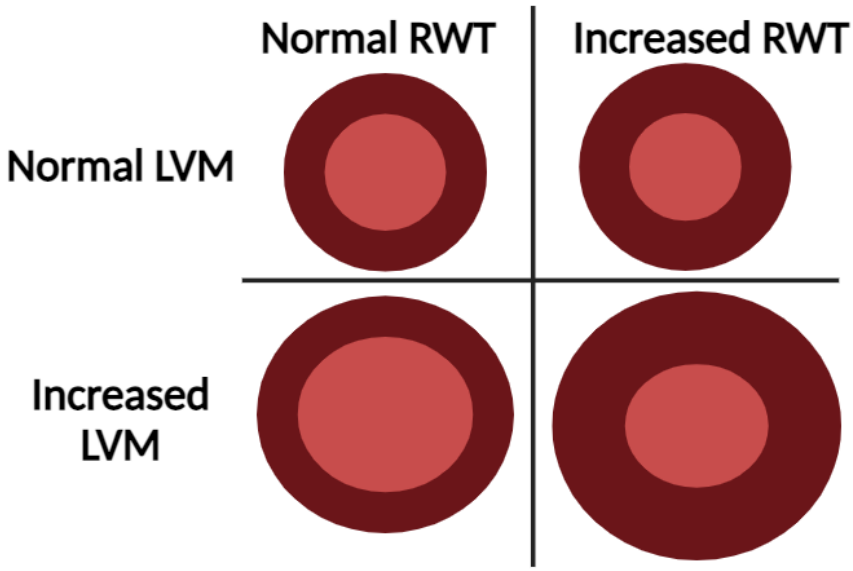



**Figure 1.4: Types of Cardiac Remodelling** Concentric vs eccentric remodelling adapted from (Drazner, 2011). 'Walled' circles represent cross-section of the left ventricle with dark shade representing wall thickness and lighter shade representing LV lumen (LV volume). Eccentric hypertrophy is defined as the increase in LVMI without significant change to the ratio between wall thickness and diastolic diameter (Relative Wall Thickness (RWT)). Concentric remodelling (increase RWT and normal LVMI) and concentric hypertrophy occur when RWT is increased, reducing the diastolic filling capacity of the heart.

### 1.2.1 Identification and Prevalence of LVH

Patients are defined as having 'LVH' when LVMI increases over a clinically defined threshold depending on the method of measurement. The American Society of Echocardiography and European Association of Cardiovascular Imaging define LVH as an increase in LVMI >95g/m in women and 115g/m in men (Bornstein, Rao, and Marwaha, 2023). Independent of additional risk factors, LVH is an independent predictor of morbidity and mortality, even after adjustment for BP and other confounding factors. There is considerable variability in the extent of cardiac remodelling in response to sustained high blood pressure (Drazner, 2011). Factors including severity and duration of pressure load, age, ethnicity, gender, multi-morbidities, and genetic factors all influence the extent to which LVMI increases during the life-course (Schmieder, 2010; Drazner, 2011). Despite being taken from a general population (mean 24-hour BP </85mmHg), over a 10-year follow up ∼1/4 of patients experienced an age-associated increase in LVMI that became classified as LVH (Bombelli et al., 2023). Greater changes in 10-year LVMI was associated with increased all-cause and CV mortality in the general population (Bombelli et al., 2023). In addition to the association between an increased pressure-load and progression to LVH as part of HMOD (B. Williams et al., 2018), there is a risk of developing LVH which is independent of poor blood pressure control. Clinically defined LVH also develops non-uniformly in hypertensive patients, and given that increased LVMI is associated with poor outcomes in the general population (Krumholz, M. Larson, and Levy, 1995; Bombelli et al., 2023), the pressure-independent factors

contributing to LV phenotypes are of importance to dissect.

It can be methodologically difficult to assess the presence of LVH, where estimates can vary depending on study population, location, and method of normalisation (indexing factor) to name but a few. Measurement modality can also affect LVH prevalence estimates; LVH can be identified by electrocardiography, echocardiography, and cardiac magnetic resonance imaging (CMR). Different measurements have varying precision, but their application is limited by availability (or unavailability), cost, functional parameters estimated, and requirements for skilled operators or interpretations (Arnett, Fuentes, and Broeckel, 2004). Estimates of LVH prevalence from primary care data ranges between 6–25% (Schirmer, Lunde, and Rasmussen, 1999; Ching, Chia, and Wan Azman, 2012; Cuspidi, Sala, et al., 2012; Viwatrangkul et al., 2021; Mancia, Facchetti, et al., 2022; X. Wang, Hao, et al., 2022; Bombelli et al., 2023; Taki et al., 2023). These estimates are higher in hypertensive populations and those with LVH tend to be of male gender, older age, have increased body mass index, and higher SBP. The presence of LVH is also associated with valvular heart disease, existing CVD, and antihypertensive medications (Schirmer, Lunde, and Rasmussen, 1999; Taki et al., 2023). Longitudinal cohort studies have suggested those who developed LVH have greater, but non-clinically raised, office and ambulatory BP measurements (SBP <130mmHg, DBP <80mmHg), and worse baseline metabolic profiles (Bombelli et al., 2023). Other factors reported to increase LVMI and predict LVH include dietary sodium intake, volume overload, diabetes, arterial hypertrophy, vascular stiffness, insulin resistance, and neurohormonal factors (i.e. adrenergic factors, renin-angiotensin-aldosterone system (RAAS)) (Arnett, Fuentes, and Broeckel, 2004).

Given the association of hypertension with factors that can increase LVH risk, it is therefore unsurprising that the prevalence of LVH is increased to approximately 30–40% in hypertensive populations (Cuspidi, Sala, et al., 2012; Mancia, Facchetti, et al., 2022; X. Wang, Hao, et al., 2022; Heimark et al., 2023). A Chinese cohort study, published in 2022, reported the weighted prevalence of LVH was almost double in a hypertensive population versus a normotensive cohort, and was higher in 'uncontrolled' hypertensives versus those with good pressure control (X. Wang, Hao, et al., 2022). This study highlights that whilst LVH occurs at a higher rate in hypertensives, LVH develops in both non-hypertensive populations and hypertensive populations. The interaction between pressure control and LVH is also of clinical importance as a number of antihypertensive medications reduce LVH (Ferreira Filho et al., 2010). Recently published data suggest blood pressure reduction to <130 mmHg in a hypertensive population with LVH resulted in increased risk of both cardiac and all-cause mortality (Heimark et al., 2023). Adverse effects of more aggressive blood pressure lowering in this cohort contrasts findings that more aggressive BP lowering in hypertension is beneficial compared to the 140mmHg target used in previous guidelines (Heimark

et al., 2023). Reliably differentiating the presence or absence of risk of developing LVH is a potentially critical clinical variable, whereby predicting the onset of LVH within patient populations could better inform clinical care, reducing poorer outcomes associated with LVH (H. Zhang, Hu, and Wei, 2020).

### 1.2.2 Heritable Cardiomyopathies

There are a number of cardiomyopathies characterised by an increase in LVMI. Hypertrophic Cardiomyopathy (HCM) is a common inherited CVD and is defined as hypertrophic remodelling of the LV not solely explained by abnormal loading conditions (B.J. Maron and M.S. Maron, 2013; Mazzarotto et al., 2020). HCM is the leading cause of sudden cardiac death in young people (B.J. Maron and M.S. Maron, 2013; Mazzarotto et al., 2020), and it is estimated ∼70% of cases of HCM are caused by mutations in sarcomere proteins (McKenna and Judge, 2021). Despite robust genetic association with autosomal dominant inheritance, HCM is a clinically diverse condition often not explained by single gene mutations (Aung, Vargas, et al., 2019; Mazzarotto et al., 2020; Harper et al., 2021). Dilated Cardiomyopathy (DCM) is characterised by LV dilation and reduced systolic function, which is often asymptomatic, and associated with >60 different genes (McKenna and Judge, 2021). Similar to DCM, Peripartum Cardiomyopathy is a reduction in LV systolic function at the end of pregnancy or following delivery, without another identifiable cause. PCM is an important cause of CVD mortality which affects only women (Vogel et al., 2021), but shares risk factors of LVH and DCM, including African-ancestry, hypertension, and increased age (Davis et al., 2020). The genetic basis and heritability of PCM is heterogeneous, although several shared genetic loci have been reported between DCM and PCM, namely within TTN and BAG3 genes (Davis et al., 2020).

Data from the UK BioBank estimated the prevalence of HCM was ∼0.11–0.22% in the adult population (Lopes et al., 2021) and genome wide association studies have recently reported common genetic variants can influence both DCM and HCM phenotypes (Harper et al., 2021; Tadros et al., 2021). Using the UK BioBank and a HCM registry cohort, 21 loci in the human genome demonstrated significant association with LV wall thickness in both the general population and HCM specific cohorts (Aung, Lopes, et al., 2023), indicating a shared genetic background between LVH in the general population, and genetically influenced cardiomyopathies characterised by increased LVMI.

### 1.2.3 Sex Differences in LVH

In 2021, the Lancet Women and Cardiovascular Disease commission described the problems of under-recognition, under-diagnosis, under-treatment, and lack of good quality studies in women with CVD (Vogel et al., 2021). Both males and females in

the general population exhibit increases in LVMI during ageing (Bombelli et al., 2023) however the prevalence of LVH is potentially higher in women than in men (Krumholz, M. Larson, and Levy, 1995; Leibowitz et al., 2010). Specifically, women may experience an increased rate of concentric remodelling and concentric hypertrophy with respect to men (Leibowitz et al., 2010). A study of an Indo-Asian population found women were at a higher risk of developing LVH than men, although this study was potentially confounded by a greater incidence of hypertension in the female population, as well as additional risk factors including increased adiposity, and low physical activity (Jafary and Jafar, 2008). In contrast, a Scandinavian population study found no difference in the prevalence of LVH between men and women (Schirmer, Lunde, and Rasmussen, 1999). Regardless of differences in prevalence, adjusting for age, hypertension, and % ejection fraction, the relative risk of all-cause mortality was double in women with LVH compared to men with LVH (Liao et al., 1995). In a population of patients with hypertrophic cardiomyopathy (HCM), women had worse survival than men (Geske et al., 2017). More strikingly, women with HCM also had a greater reduction in 10-year survival compared to an age and sex-matched population than males (-8.1% in females vs -0.7% in males) (Geske et al., 2017). In a non-hypertensive population, the risk of a 1 standard deviation (SD) increase in LVM on CVD mortality was significant only in the male population, once adjusted for additional confounders (Bombelli et al., 2023). This was likely due to the lower number of CV outcomes in the female population rather than a true difference in risk of raised LVMI. Hazard ratios for Major Adverse Cardiac Events (MACE) was lower in women with hypertension than men only when LVH was not present (Gerdts, Izzo, et al., 2018). As a consequence of hypertension, women more often develop LVH than men (Gerdts, Okin, et al., 2008; Drazner, 2011; Gerdts, Izzo, et al., 2018; X. Wang, Hao, et al., 2022), with the risk of LVH reportedly up to 70% higher in women than in men (X. Wang, Hao, et al., 2022; Cuspidi, Faggiano, and Tadic, 2023). Another study of hypertension and CVD risk demonstrated women were at increased risk of CVD at a lower SBP than men (Elfassy et al., 2023), although these authors did not report the presence or absence of LVH and other end-organ damage.

The mechanisms underlying sex-based differences in LVH and CVD risk are yet to be fully elucidated. Epidemiological studies report retrospective analysis of risks associated with LVH without investigating potentially dimorphic responses to stressors accumulating over the lifetime (Bairey Merz, Nelson, et al., 2020). Women appear to disproportionately suffer from coronary microvascular dysfunction (Bairey Merz, Pepine, et al., 2017), providing an enhanced ischaemic signal potentially perpetuating pathological adaptations of the heart. LVM is not different between boys and girls during infancy and childhood, and the initial number of cardiomyocytes is not reportedly sexually dimorphic (Burke et al., 1987; Simone, Devereux, Daniels, et al., 1995). Sex-based differences in systolic stiffness (LV elastance), resting heart rate, and SBP could potentially underlie the divergence of LVMI which becomes measurable during

prepubescent and pubescent years (Burke et al., 1987; Goble et al., 1992; Simone, Devereux, Daniels, et al., 1995; Bairey Merz, Nelson, et al., 2020). How these differences in cardiac structure affect the development of adverse cardiac remodelling is yet to be determined. In a rat model of physiological hypertrophy, the hypertrophic growth conferred by swim-training was more pronounced, and resulted in a greater increase in stroke volume and cardiac contractility in female rats than in male rats (Olah et al., 2019). Hearts from aged male rats also demonstrate more fibrosis and larger cardiomyocytes than female hearts however this is coupled with a reduced RWT when normalised to differences in body weight (Forman et al., 1997). Differences in energy metabolism between females and males in response to exercise training have also been reported (Soto et al., 2008; Foryst-Ludwig et al., 2011). Again, how these sex-specific responses to physiological stimuli translate into pathological cardiac remodelling is largely unknown, although some investigations have focused on the potential protective role of estrogen and estrogen receptors in females (Skavdahl et al., 2005). The contribution of genetic factors to sex-based differences are likely to be tissue specific, where gene expression profiles cluster by tissue, and then by sex in human, mouse, rat, dog, and macaque tissues (Naqvi et al., 2019).

### 1.2.4 Pathological vs Physiological Processes in Hypertrophic Growth

Pathways associated with cardiac remodelling that are physiological and pathological can often somewhat overlap, becoming 'pathological' depending on the context within which remodelling is occurring. In physiological hypertrophy, increased LV wall thickness is achieved through an increase in individual cardiomyocyte length and width, with preserved or enhanced myocyte contractility (Nakamura and Sadoshima, 2018). The molecular stimuli associated with physiological growth also activate mitochondrial biogenesis, improved calcium handling, and angiogenesis, including the release of growth factors such as insulin, vascular endothelial-, platelet derived-, and insulin like-growth factors (VEGF, PDGF and IGF respectively) (Nakamura and Sadoshima, 2018). These physiological processes are represented in Figure 1.5. Physiological hypertrophy is accompanied by the upregulation of genes involved in blood vessel morphogenesis, vascular development and muscle development (Farrell et al., 2018) suggesting increases in cardiomyocyte size is accompanied expansion of the surrounding vasculature to ensure physiological functioning can be maintained with an adequate blood supply. Contrasting pro-growth and chemotaxis pathways specific to physiological hypertrophy, genes associated with pathological growth were involved in extracellular matrix organisation and development (Farrell et al., 2018). This is consistent with the association of pathological growth with fibrosis, resulting from an excessive expansion of the extracellular matrix (Farrell et al., 2018).

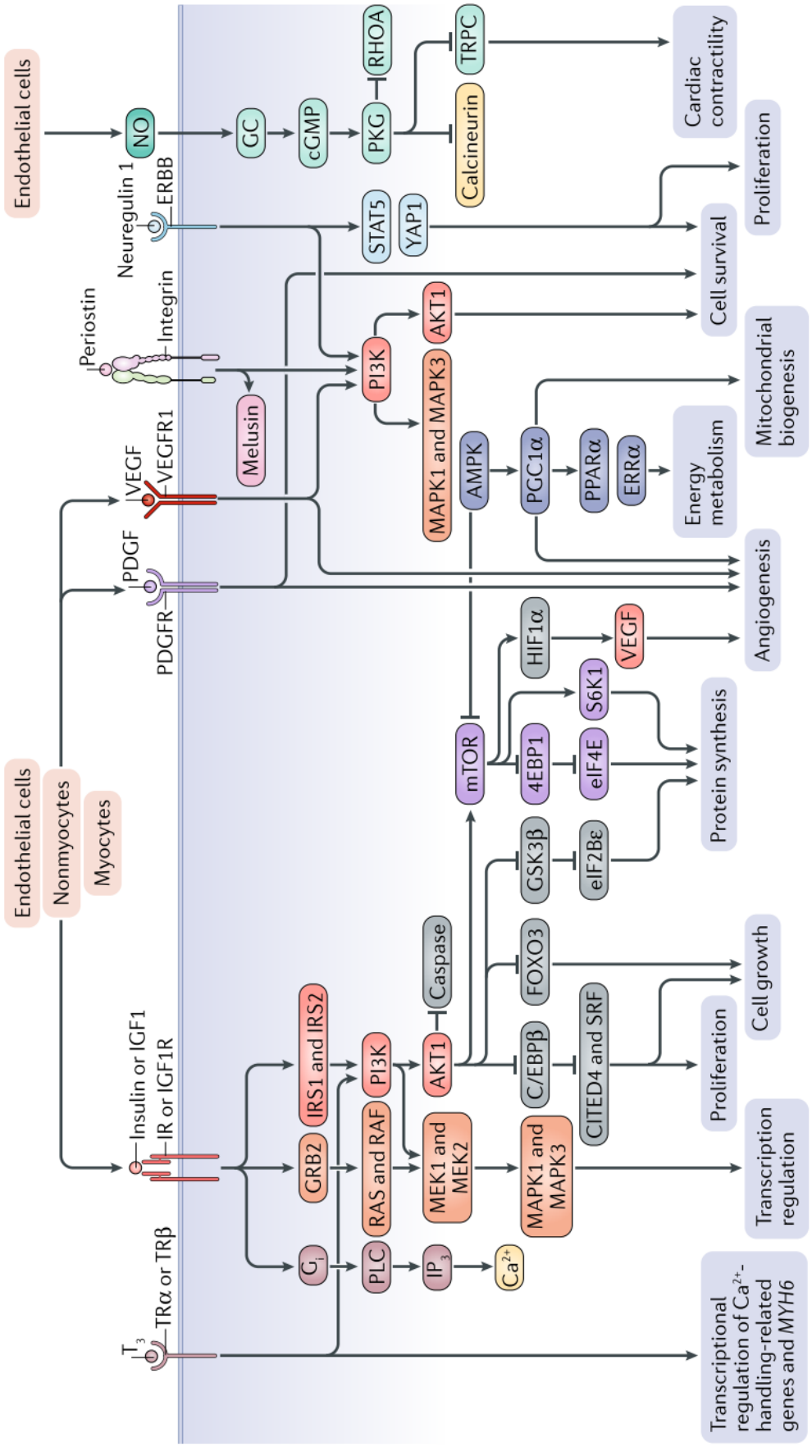



**Figure 1.5: Pathways in Physiological Cardiac Hypertrophy** Image taken from Nakamura and Sadoshima (2018) detailing pathways involved in physiological cardiac remodelling.

Activation of fibroblast populations and the expansion of ECM components forms an important part of pathological hypertrophy, distinguishing it from physiological adaptations to external stimuli. Myocytes and fibroblasts constitute the two primary cell types of the LV. Paracrine regulation of both myocytes by fibroblasts, and the reverse, form an important part of their regulation. Cross-talk between myocytes and fibroblasts is produced through paracrine signalling, mechanical stimuli, vesicle-mediated crosstalk, and electrical coupling (Nicin et al., 2022). Cardiac fibrosis is mediated by a number of fibroblast growth factors and can be triggered by increased cardiomyocyte death (Fujiu and Nagai, 2014). Stimulation of cardiomyocytes with pro-hypertrophic signals, resulted in increased cross-signalling between myocytes and fibroblasts, directly activating fibroblasts to increase ECM deposition (Manabe, Shindo, and Nagai, 2002). Signal transduction of external signalling molecules interacting with cell-surface receptors on fibroblast cells results in the activation of transcription factors including AP-1, NF-$\varkappa$B, early growth response factor-1 (Egr-1), STATs (signal transducer and activator of transcription) and Smads (Manabe, Shindo, and Nagai, 2002). Egr-1 and NF-$\varkappa$B are likely involved in the pro-inflammatory response of fibroblasts (Silverman and T. Collins, 1999; Valen, Z.Q. Yan, and Hansson, 2001). Fibroblast cells under prolonged stress will begin to alter their phenotype, becoming 'activated' myofibroblasts, expressing several smooth muscle markers including $\alpha$-actin and tropomyosin. As such, activated myofibroblasts have increased production of growth factors, cytokines, chemokines, ECM proteins, and proteases (Manabe, Shindo, and Nagai, 2002). One of the best characterised effects of myofibroblast transdifferentiation is the release of TGF-$\beta$ which both promotes proliferation and ECM production of fibroblasts, and stimulates hypertrophic growth in cardiomyocytes (Nicin et al., 2022).

At the cellular level, pathological growth is regulated by cardiomyocytes, fibroblasts, vascular, and resident immune cells of the heart, as well as infiltrating immune cells (Nakamura and Sadoshima, 2018). Angiotensin-II (Ang-II), endothelin-1, naturetic peptides (atrial and B-type; *NPPA* & *NPPB* respectively), and catecholamines have all been shown to stimulate pathological remodelling through their promotion of mitochondrial dysfunction, fibrosis, cell death, and contractile dysfunction (Figure 1.6). Calcineurin, a serine/threonine phosphatase, is activated downstream of angiotensin and endothelin receptors in the LV. In a pathway specific to pathological remodelling, calcineurin dephosphorylates nuclear factor of activated T cells (NFAT) to promote its nuclear localisation which, in-turn, interacts with transcription factors GATA4 and MEF2A to upregulate hypertrophy associated genes (Molkentin et al., 1998; Nakamura and Sadoshima, 2018). During foetal development, the cardiac muscle relies primarily on glycolytic energy processes. In the postnatal period the heart undergoes extensive remodelling to increase mitochondrial content and increase fatty acid oxidation (Porrello et al., 2009; Prestes, Marques, Lopez-Campos, Lewandowski, et al., 2018). NFAT is known to regulate the expression of foetal genes, forming part of the 'foetal gene

program', that becomes re-activated during pathological hypertrophy (Rajabi et al., 2007). Whilst physiological hypertrophy is associated with increased cardiac oxidative capacity and expansion of mitochondrial content and function (Shimizu and Minamino, 2016), in the pathologically hypertrophied heart there is reversion to a foetal-like gene expression program, resulting in an increased reliance on glycolysis and other forms of metabolism (Neubauer, 2007).

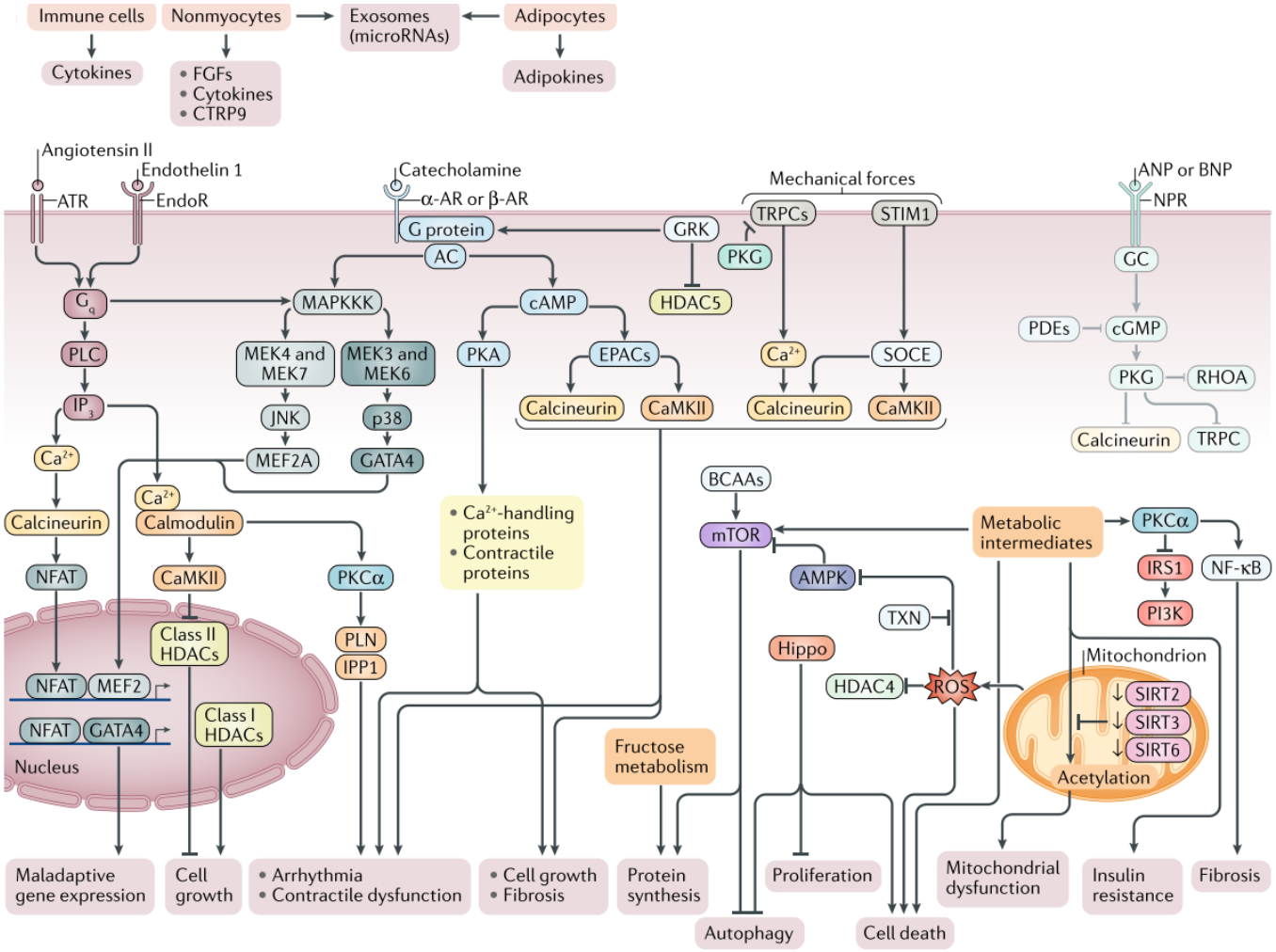



**Figure 1.6: Pathways in Pathological Cardiac Hypertrophy** Taken from Nakamura and Sadoshima (2018), overview of signalling pathways involved in pathological cardiac remodelling.

The oxygen rich environment in post-natal development and consequent upregulation of oxidative metabolism has been identified as an upstream signal to trigger cell-cycle arrest and terminal differentiation of cardiomyocytes shortly after birth (Puente et al., 2014). During pathological hypertrophy, mitochondrial proteins become hyperacetylated, reducing NAD+ levels, and inactivating mitochondria-associated sirtuin proteins (Nakamura and Sadoshima, 2018). Sirtuins 3 and 5, which exhibit cardioprotective effects, become downregulated during cardiac hypertrophy and contribute to mitochondrial dysfunction (Tang et al., 2017). A simultaneous increase in sirtuin-4 promotes accumulation of ROS and further activates the MAPK/ERK pathway to increase pro-hypertrophic signalling (Tang et al., 2017). During post-natal development, in response to increased production of reactive oxygen species (ROS), the DNA damage response (DDR) is induced to cause cell-cycle arrest in cardiomyocytes (Puente et al., 2014). In CVD pathophysiology, the DDR, accumulation of DNA damage, and incomplete DNA repair is associated with simultaneously inducing cardiomyocyte hypertrophy and cell death, fibroblast activation and fibrosis, endothelial cell dysfunction and a pro-thrombotic phenotype, and relaxation of vascular smooth muscle cells (Nikfarjam and K.K. Singh, 2023). Persistent DDR activation can result in the accumulation of DNA damage, triggering cell death and senescence (Hoeijmakers, 2009). In animals models of cardiac hypertrophy, induced by Ang-II infusion or surgical induction of pressure overload, an increase in DNA double-strand breaks was observed (Nakada et al., 2019), a response consistent with increased cardiomyocyte apoptosis during pathological hypertrophy.

### 1.2.5 Current Therapeutic Strategies in LVH

European guidelines for the management of hypertension currently recommend cardiac assessment by electrocardiography, 2D/3D echocardiography, or CMR (Diez and Butler, 2023; Mancia, Kreutz, et al., 2023), for the early detection of changes in LVMI and development of LVH. Data from the Framingham Heart Study suggested use of effective antihypertensive therapy reduces the prevalence of LVH (Mosterd et al., 1999). Some antihypertensive drugs have been shown to reverse LVH or reduce LVM in meta-analyses of clinical trials, with reduction in LVM reportedly greatest in ARBs and worst in diuretics and β-blockers (Klingbeil et al., 2003; Fagard et al., 2009). Both diuretics and β-blockers as hypertensive treatments conferred a reduction in LVMI (Klingbeil et al., 2003; Fagard et al., 2009). In 2022, analysis of almost 2,000 patients showed a positive association between the reduction in SBP and LVMI conferred by antihypertensive medications, with the strongest reduction in patients with LVH at baseline (H.M. Kim et al., 2022). That being said, Simone, Devereux, Izzo, et al. (2013) reported no significant reduction in the regression of LVH in a 5-year follow-up, although women experienced a greater regression of LVH than men. In contrast, a recent evaluation of LVH regression with antihypertensive medication suggested the association with decreased SBP and LVMI was stronger in males than in females (H.M. Kim et

al., 2022). The higher proportions of LVH in the female population at baseline and follow-up likely contributed to the described differences (H.M. Kim et al., 2022).

Treatment of LVH is dependent on the underlying cause, particularly as LVH is common to a number of diseases. As a consequence treatment of LVH is often reported as secondary to the reduction of the phenotype itself, as described above in hypertension. The protective effect of ARBs on LVH has also been reported in dialysis patients (L.Y. Yang et al., 2013) and diabetes (Raff et al., 2014). Currently, ability to effectively determine those at risk of LVH relies on cardiac assessment by electrocardiography, echocardiography, or CMR. An insertion/deletion mutation in the angiotensin-converting enzyme (ACE) gene was identified as marker of LVH in an untreated cohort of hypertensives, which might synergistically predict LVH alongside a polymorphism in the angiotensin-I receptor gene ($AT_1$) (Diez and Frohlich, 2010). The circulating levels of cardiotrophin 1 (CT-1) is increased in hypertensive patients with LVH, and HCM patients with more severe LVH (Diez and Frohlich, 2010; Monserrat et al., 2011).

Identifying genes, molecular biomarkers or specific pathways influencing LVMI is an important step in improving the identification of patients requiring treatment to prevent development of LVH (Diez and Frohlich, 2010). Persistence of LVH, despite treatment of hypertension, was an independent predictor of adverse outcomes, highlighting both the importance of LVMI reduction during hypertensive treatment, and the relative independence of CVD risk between LVH and hypertension (H.M. Kim et al., 2022).

## 1.3 -Omics Technologies to Unveil Markers and Mechanisms of Complex Polygenic Phenotypes

'Omics' technologies are high-throughput biochemical assays that measure molecules of the same type (e.g. DNA, RNA, proteins, metabolites) from the same biological sample (Conesa and Beck, 2019). Since the completion of the first human, mouse, and rat genome sequences in the early 2000s, the 'post-genomic era' saw rapid advancements in sequencing technologies, with Next Generation Sequencing (NGS) now replacing traditional DNA and RNA microarrays (Parikh and Ashley, 2017; Joshi et al., 2021; R.S. Wang, B.A. Maron, and Loscalzo, 2023).

Genomic data comprises collection and sequencing of DNA from populations of interest and can be utilised in genome-wide association studies (GWAS), Mendelian randomisation (MR), and quantitative trait loci (QTL) studies, all of which aim to elucidate relationships between genetic factors and specific phenotypes and/or diseases. Genomics is the most mature of the -omics technologies and GWAS studies have been

widely applied across almost all human phenotypes. As genomic studies have advanced, it has inspired the development of other 'omics', including transcriptomics (study of the transcriptome), proteomics (study of whole proteomes), metabolomics (study of the metabolome), lipidomics (study of the lipidome) and epigenomics (study of the complete set of epigenetic modifications) (R.S. Wang, B.A. Maron, and Loscalzo, 2023). Whilst genomics represents the stable genetic influence on phenotypes, transcriptomics, proteomics, lipidomics, metabolomics, and epigenomics represent state-specific landscapes within tissues or cells, and can be explored to determine not only relationships between the genetic landscape and phenotype development, but the pathways and mechanisms by which molecules influence phenotype development. The non-selective and high-throughput approaches taken by -omics studies oppose a traditional, targeted or 'reductionist' approach, and makes an unbiased assessment of specific populations (R.S. Wang, B.A. Maron, and Loscalzo, 2023). Both approaches are integral to the modern research landscape, and it is likely that a combined approach is required in complex diseases like CVD.

### 1.3.1 Improving NGS Technologies: Short to Long Read Sequencing

Short read sequencing techniques require the genome or transcriptome to be broken into small fragments that are enriched by PCR during a 'library preparation' step. These libraries of RNA or DNA fragments are then sequenced by massive parallel sequencing to generate reads 50–300bp long. The short reads are then checked for quality and aligned to a reference genome. The exact method of sequencing depends on the NGS platform, for example, Illumina applies an 'sequencing-by-synthesis' approach, using fluorescently labelled bases incorporated into the growing DNA strands. Other technologies include the SOLiD platform, which uses clonal amplification, where DNA templates bound to microbeads are flanked by adapters and hybridised to a growing cDNA strand by DNA ligase.

Short read sequencing technologies are well established, thought to be highly accurate, and increasingly widely available in clinical and research-based laboratories. The requirement to fragment data for short read sequencing reduces genome continuity in alignments, and repetitive regions that cannot be uniquely mapped will be discarded. Regions of high GC content, complex genomic regions and structural variation can also limit practical applications of short read sequencing (M.O. Pollard et al., 2018). Conversely, long read sequencers can sequence reads in excess of 10kb (Amarasinghe et al., 2020). Although traditionally less accurate than short read technologies, 'third-generation' sequencing from Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT) have begun to overcome these limitations and broadened application of long read sequencing in genomics (Amarasinghe et al., 2020). In 2017, the Vertebrate Genomes Project (VGP) was established to generate high-quality complete reference genomes for 70,000 vertebrae species, using IsoSeq long reads to improve gene

annotation structure (Rhie et al., 2021). The first rat genome produced by the VGP was the mRatBN7.2, which become the first rat reference to be adopted by the Genome Reference Consortium. In 2024, the GRCr8 (GCA_036323735.1) rat genome was published, generated from NHGRI-funded 'Inbred Rat Genomes Project' using PacBio HiFi sequencing. The first human telomere to telomere human genome sequencing has recently published a gap-less human genome assembly, allowing investigation into regions of the genome previously inaccessible (Miga et al., 2020; Mao and G. Zhang, 2022).

### 1.3.2 Genetic Contribution to LVH

Known clinical and haemodynamic stimuli of LVH account for only one half to two thirds of inter-individual variability in LVMI (Bella and Goring, 2012). Robust evidence from familial studies and studies of monozygotic and dizygotic twins show electrocardiographic and echocardiographic derived indices of LVM have significant genetic basis, with estimates of heritability ratios ranging from 0.12 to 0.84 (Bella and Goring, 2012; Aung, Vargas, et al., 2019; H. Lin et al., 2021). These estimates of heritability can be influenced by modality of cardiac measurement where measurements derived from electrocardiography result in higher estimates of heritability (<60%) than cardiac MRI-derived (CMR) indices (<44%) (Nethononda et al., 2019). LVH is more common in those of African ancestry (Arnett, Fuentes, and Broeckel, 2004), where prevalence is reportedly 3-fold higher in black men and women from multi-ethnic cohorts (A.A. Lewis et al., 2020). Both candidate gene studies, where a gene of interest in a relevant pathway is investigated for its role in pathological processes, and GWAS studies, have explored the link between genetic variance and LVM (Arnett, Fuentes, and Broeckel, 2004; Bella and Goring, 2012). Polygenic risk scores (PRS) combine genomic and clinical information to synergistically and additively contribute to risk prediction in patient populations, but require a highly specific set of casual variants to be more effective than traditional clinical risk scores. In the context of CVD, GWAS studies have been applied in both as a method of causal variant discovery, and for the generation of such polygenic 'risk scores'.

Structural measures of the heart tend to show the greatest heritability scores, making LVMI, End Diastolic Volume (EDV) and End Systolic Volume (ESV) good candidates for genomic studies (Aung, Vargas, et al., 2019). That being said, relatively few genomic studies have been undertaken to determine genetic contribution to cardiac structure in human populations. One of the earliest GWAS studies for LVMI was performed in the HyperGEN cohort study using 'cases' and 'controls' from the extremes of LVMI distribution to increase power (Arnett, N. Li, et al., 2009). The authors found 12 SNPs associated with echocardiographic-determined LVMI, 2 of which were within 2Mb of SNPs previously associated with EDV in the Framingham Heart Study (Vasan, M.G. Larson, Aragam, et al., 2007; Arnett, N. Li, et al., 2009). An intragenic SNP

on the *KCNB1* gene was associated with LVMI in both the Caucasian discovery, and African-American validation cohorts (Arnett, N. Li, et al., 2009). Using data from the UK BioBank, where genomic and CMR data was available, sentinel variants annotated at the *TTN, SH2B3, BAG3, MTSS1, CLCNKA, CDKN1A, ZNF592*, and *DERL3* gene loci were associated with either EDV, ESV, EF or LVM (Aung, Vargas, et al., 2019). These replicated findings in a European meta-analysis of echocardiographic traits, where *MTSS1*, and *CDKN1A* were associated with EDV and fractional shortening respectively (Wild et al., 2017). The *TTN* and *BAG3* loci on human chromosomes 2 and 10, were associated with EDV, SEV EF, and LVM (*TTN* only). Apart from *MTSS1*, most other variants were only nominally associated with structural parameters in multi-ethnic validation cohort and sex-stratified analyses were not reported (Aung, Vargas, et al., 2019).

A more recent GWAS of CMR-derived measures of LVMI replicated the association with the *TTN* locus and found 11 additional lead SNPs significantly associated with LVMI (Khurshid et al., 2023). Despite replication of a few significant associations across non-European ancestries, the UK Biobank is known to be enriched for health and socioeconomic status compared to the general population (Fry et al., 2017). The BioBank Japan project is one of the largest, non-European, single-descent biobanks containing detailed genome and phenome information (Kanai et al., 2018). A GWAS of 58 traits, including 9 echocardiographic parameters found only 4 distinct loci associated with EDV, ESV, RWT, Fractional Shortening (FS) and EF, all of which were shared with at least one other echocardiographic trait (Kanai et al., 2018). Two of these loci, on chromosome 8 (lead variant *TRIB1*) and on chromosome 2 (lead variant*EML6*), were also shared with at least one other blood pressure, haematological, liver, or metabolic trait indicating the potential pleiotropic effects of variants affecting cardiac structure or function (Kanai et al., 2018). Echocardiographic measures of LVMI showed a significant association with SNPs in *MYOM1, MYOCD* and *CHD13* in a large American case-control cohort of normotensive and hypertensive patients (Barve et al., 2016).

NGS technologies have allowed the expansion of GWAS studies from common variants on DNA microarrays, to full genome (whole exome) screens examining the relationship between genetic variants and measures of LV structure and function. Similar to Arnett, N. Li, et al. (2009), NGS sequencing was conducted in 'cases' and 'controls' taken from the extreme values of LVMI (<5th percentile, >95th percentile) in a Taiwanese Community cohort (H.Y. Fan et al., 2022). Of the 16 independent SNPs identified, none overlapped previously reported SNPs, but were mostly nearby genes associated with cardiovascular traits including *CASQ2, COL51*, and *FXN* (H.Y. Fan et al., 2022), although collation of published associated SNPs can be complicated by the availability of data. The GWAS catalogue, a collaborative project between the

EMBL-EBI and NHGRI, provides GWAS data for variant-trait associations which can be assessed by API or through an online graphical user interface (GUI) (Sollis et al., 2023). Human SNPs associated with cardiac structural traits are presented in Table 1.1. Despite strong heritability in cardiac structural and functional traits, many efforts to associate these with variants in the human genome have resulted in few significant findings. Whilst some replicate findings from previous studies, some of the 'missing heritability' is in part due to physiological differences in posture, hydration, heart rate or pharmacologic agents and methodological differences in trait measurement (Heineke and Molkentin, 2006; Nethononda et al., 2019).

An alternative to GWAS analyses are so-called 'PheWAS' analyses, which test associations between one or many variants (usually from GWAS studies in a phenotype of interest) and a range of defined characteristics. PheWAS analyses can identify clinical diagnoses or phenotypes that are linked with the genetic variability associated with a phenotype of interest (Mosley et al., 2020). In a clinical population of healthy and diseased individuals, obesity, hypertension, coronary artery disease, atrial fibrillation, and type II diabetes were significantly associated with genetically predicted LVMI (Mosley et al., 2020). Interestingly, in the same study, Mendelian Randomisation found that whilst obesity is causally linked to LVMI in both males and females, only in males were causal associations of hypertension and coronary artery disease with LVMI significant (Mosley et al., 2020). Higher scores in a developed polygenic risk score for LVMI has also been associated with incidence of implantable cardioverter-defibrillator (ICD) implant, HCM and DCM (Khurshid et al., 2023).

**Table 1.1:** GWAS Catalogue SNPs Associated with LVMI

| Trait | Variant ID | Gene | Risk Allele | Risk Freq. | *p*-value | Effect & Direction | Study |
|---|---|---|---|---|---|---|---|
| Left ventricular mass | rs112555002 | | T | | 0.00 | 1.02 unit decrease | GCST90244711 36944631 |
| Left ventricular mass | rs1421085 | | T | | 0.00 | 1.03 unit decrease | GCST90244711 36944631 |
| Left ventricular mass to end-diastolic volume ratio | rs2070458 | SMARCB1 | T | 0.81 | 0.00 | 0.0067 unit decrease | GCST011208 33495596 |
| Left ventricular mass indexed by body surface area | rs28552516 | | C | | 0.00 | 0.58 unit decrease | GCST90244710 36944631 |
| Left ventricular mass indexed by body surface area | rs34163229 | | G | | 0.00 | 0.6 unit decrease | GCST90244710 36944631 |
| Left ventricular mass | rs3729989 | | T | | 0.00 | 1.32 unit decrease | GCST90244711 36944631 |
| Left ventricular mass indexed by body surface area | rs3729989 | | T | | 0.00 | 0.61 unit decrease | GCST90244710 36944631 |
| Left ventricular mass | rs6503451 | | T | | 0.00 | 1.01 unit decrease | GCST90244711 36944631 |
| Left ventricular mass indexed by body surface area | rs6503451 | | T | | 0.00 | 0.52 unit decrease | GCST90244710 36944631 |

. . . **continued next page**

| Trait | Variant ID | Gene | Risk Allele | Risk Freq. | *p*-value | Effect & Direction | Study |
|---|---|---|---|---|---|---|---|
| Left ventricular mass | rs9388498 | | G | | 0.00 | 1.43 unit decrease | GCST90244711 36944631 |
| Left ventricular mass indexed by body surface area | rs9388498 | | G | | 0.00 | 0.55 unit decrease | GCST90244710 36944631 |
| Left ventricular mass | chr5:133066736 | | CTT | | 0.00 | 1.02 unit increase | GCST90244711 36944631 |
| Left ventricular mass indexed by body surface area | chr5:133066736 | | CTT | | 0.00 | 0.5 unit increase | GCST90244710 36944631 |
| Left ventricular mass indexed by body surface area | rs10497529 | | G | | 0.00 | 1.28 unit increase | GCST90244710 36944631 |
| Left ventricular mass | rs10878349 | | A | | 0.00 | 0.99 unit increase | GCST90244711 36944631 |
| Left ventricular mass to end-diastolic volume ratio | rs116904997 | PXN | G | 0.98 | 0.00 | 0.0155 unit increase | GCST011208 33495596 |
| Left ventricular mass indexed by body surface area | rs143800963 | | C | | 0.00 | 0.95 unit increase | GCST90244710 36944631 |
| Left ventricular mass to end-diastolic volume ratio | rs185178768 | FAF1 | G | 0.99 | 0.00 | 0.0205 unit increase | GCST011208 33495596 |

. . . **continued next page**

| Trait | Variant ID | Gene | Risk Allele | Risk Freq. | *p*-value | Effect & Direction | Study |
|---|---|---|---|---|---|---|---|
| Left ventricular mass | rs2255167 | | T | | 0.00 | 1.83 unit increase | GCST90244711 36944631 |
| Left ventricular mass indexed by body surface area | rs2255167 | | T | | 0.00 | 0.97 unit increase | GCST90244710 36944631 |
| Left ventricular mass | rs4985155 | | A | | 0.00 | 1 unit increase | GCST90244711 36944631 |
| Left ventricular mass indexed by body surface area | rs56252725 | | G | | 0.00 | 0.54 unit increase | GCST90244710 36944631 |
| Left ventricular mass | rs62621197 | | C | | 0.00 | 2.85 unit increase | GCST90244711 36944631 |
| Left ventricular mass indexed by body surface area | rs62621197 | | C | | 0.00 | 1.11 unit increase | GCST90244710 36944631 |
| Left ventricular mass | rs6755784 | TTN | G | 0.82 | 0.00 | 1.094 unit increase | GCST011207 33495596 |
| Left ventricular mass to end-diastolic volume ratio | rs8039472 | ALPK3 | G | 0.52 | 0.00 | 0.0052 unit increase | GCST011208 33495596 |
| Left ventricular end diastolic mass end diastolic volume ratio | rs11350493 | | TG | | 0.00 | | GCST90268131 37126556 |

. . . **continued next page**

| Trait | Variant ID | Gene | Risk Allele | Risk Freq. | *p*-value | Effect & Direction | Study |
|---|---|---|---|---|---|---|---|
| Left ventricular end diastolic mass end diastolic volume ratio | rs12452627 | | G | | 0.00 | | GCST90268131 37126556 |
| Cardiac hypertrophy | rs1320448 | COL17A1 | | | 0.00 | | GCST000960 21348951 |
| Left ventricular end diastolic mass end diastolic volume ratio | rs133890 | | C | | 0.00 | | GCST90268131 37126556 |
| Cardiac muscle measurement | rs1436109 | NCAM1 | | | 0.00 | | GCST000936 21212386 |
| Left ventricular end diastolic mass end diastolic volume ratio | rs17692129 | | C | | 0.00 | | GCST90268131 37126556 |
| Left ventricular end diastolic mass end diastolic volume ratio | rs2503715 | | A | | 0.00 | | GCST90268131 37126556 |
| Left ventricular end diastolic mass end diastolic volume ratio | rs2696421 | | A | | 0.00 | | GCST90268131 37126556 |
| Left ventricular end diastolic mass end diastolic volume ratio | rs34504162 | | A | | 0.00 | | GCST90268131 37126556 |

. . . **continued next page**

| Trait | Variant ID | Gene | Risk Allele | Risk Freq. | *p*-value | Effect & Direction | Study |
|---|---|---|---|---|---|---|---|
| Left ventricular end diastolic mass end diastolic volume ratio | rs35009641 | | G | | 0.00 | | GCST90268131 37126556 |
| Left ventricular end diastolic mass end diastolic volume ratio | rs56864281 | | C | | 0.00 | | GCST90268131 37126556 |
| Left ventricular end diastolic mass end diastolic volume ratio | rs5760054 | | C | | 0.00 | | GCST90268131 37126556 |
| Left ventricular end diastolic mass end diastolic volume ratio | rs7166287 | | C | | 0.00 | | GCST90268131 37126556 |
| Left ventricular end diastolic mass end diastolic volume ratio | rs730506 | | G | | 0.00 | | GCST90268131 37126556 |
| Left ventricular end diastolic mass end diastolic volume ratio | rs74640693 | | A | | 0.00 | | GCST90268131 37126556 |

GWAS Catalog is a service provided by the EMBL-EBI and NHGRI that offers a manually curated and freely available database of published genome-wide association studies (GWAS). The R-package gwasrapidd (Magno and Maia, 2019) was used to retrieve all variants associated with the trait IDs; EFO_0009289, EFO_0009290 & EFO_0010556. EMBL-EBI; European Molecular Biology Laboratory-European Bioinformatics Institute (EBI), NHGRI; National Human Genome Research Institute.

### 1.3.3 Assessment of the Transcriptome During Physiological and Pathological Hypertrophy

Genomic studies require translation of function from associated variants to active pathways in state-specific contexts. This requires high-throughput analysis of the 'active'-genome, that is the parts of DNA which are actively transcribed in specific cells or tissue types. The dynamic expression landscape contained within the transcriptome is specific to the sample it has been extracted from, and relies on mRNA-based quantification by RNA microarray or RNA-sequencing. Compared to traditional microarray, short read RNA sequencing offers increased accuracy for low abundance transcripts, and higher resolution for tissue-specific transcripts, increasing the ability to distinguish between expression profiles of closely related paralogues and different splice variants (Dam et al., 2018).

In 2013, the Genotype Tissue Expression Project (GTEx) was developed, which aimed to determine tissue specific variation in gene expression in response to genotype (Lonsdale et al., 2013). The current release (version 8) contains RNA-sequencing in over 15,201 samples from 49 tissues of 838 postmortem donors (GTEx Consortium, 2020). Importantly, genotype information is also available such that loci associated with specific phenotypes can both be assessed not only at the intra-individual level, to determine relative tissue specificity of genes, but also at the inter-individual level; determining differences in tissue specific gene expression based on genetic background (Lonsdale et al., 2013; GTEx Consortium, 2020). Aside from the potential of catalogues of diversity such as GTEx, transcriptomics can be performed in samples which do not have diverse genetic landscapes, and is useful for the comparison of control and case tissue. Transcriptomics can therefore be applied where GWAS can not, including model systems, cells, tissues derived from inbred animal models, and clinical samples from distinct populations.

Reconstruction of the transcriptome from short read sequence data represents a great technological advancement from the reliance on expressed sequence tags (ESTs), protein sequences and *in silico* prediction (Mardis, 2017; X. Wang, You, et al., 2019). The median length of transcripts is far greater than the short read length, such that alternative transcript start sites, alternative exon splicing, and gene regions containing tandem repeats are difficult to characterise from short read sequencing (Mardis, 2017; X. Wang, You, et al., 2019). Although less widely applied than short read sequencing in transcriptomics, application of the PacBio long read RNA sequencing platform in the rat hippocampus demonstrated the number of isoforms per gene could be increased by more than twofold during *de novo* transcript assembly (X. Wang, You, et al., 2019), these technologies represent the opportunity to improve genome annotation as well as increase the ability to determine state-specific differences in transcript usage, define alternative start sites, and infer functionality in previous unaccessible regions of the

genome.

#### 1.3.3.1 Bulk RNA Sequencing

Bulk or whole tissue RNA-sequencing refers to any approach where mRNA is extracted and transcripts from a population of cells are aggregated and sequenced using short or long read sequencing. RNA is isolated and reverse transcribed in a non-specific manner to generate cDNA of the complete transcriptome present within the sample, which is then sequenced using short or long read RNA sequencing. Once sequenced, aligned to a reference genome, and quantified, the expression of coding and non-coding transcripts can then be compared across samples and conditions. An important element of transcriptomics is the ability to perform pathway analysis, where linkage of expressed genes to annotated pathways and functions can be used to attribute functionality to gene-sets in a phenotype of interest. Due to the high tissue and environment specificity of gene expression, attribution of gene functionality can be inferred by exploration of the transcriptome across models of defined phenotypes. The close connection between tissue expression specificity, and gene function, was revealed in a recent meta-analysis of over 300 sequencing experiments in the rat. Combining data from 7,700 RNA-sequencing datasets, only 0.85% of genes were detected in the active transcriptome of all 585 rat tissues studied (Summers et al., 2022). RNA-sequencing experiments between genetic models relevant tissues, at time points preceding or following the development of divergences in phenotype can be utilised to determine pathways causing or resulting from phenotypes of interest. Rather than implicating specific single genes, RNA sequencing experiments can thus be used to determine the functional pathways suitable for disruption in order to validate mechanistic studies, or target for reversal and treatment of pathological phenotypes.

As the number of published microarray and RNA-sequencing experiments increased, the NCBI developed the 'Gene Expression Omnibus' (GEO), a public functional genomics data repository where published transcriptomics data can be queried and downloaded. This has allowed expression 'meta-analyses' to be performed, particularly useful in conditions when the number of cases or controls may be small. In a network comparison of 6 microarray datasets of physiological and pathological LVH, genes contained within pathological hypertrophy-specific co-expression networks were enriched for phenotypes such as abnormal cardiovascular system morphology (Drozdov et al., 2013). Co-expression networks of genes with increased expression in physiological hypertrophy contained genes encoding ECM and cytoskeletal proteins, with a concomitant reduction in contribution of networks encoding mitochondrial translation, metabolic pathways, and oxidative phosphorylation (Boileau et al., 2020). Analysis of genes only in models of pathological or physiological hypertrophy indicated genes expressed in pathological remodelling were associated with metabolic, apoptotic, and energy production pathways, whilst expression profiles in physiological remodelling were angiogenic, and cell

cycle specific (Drozdov et al., 2013). This is complemented by a similar comparison of the LV transcriptome following either surgically-induced pressure-overload or exercise training, where distinct patterns of expression were associated with physiological and pathological hypertrophy (H.K. Song et al., 2012). In response to pathological hypertrophy, cardiac expression of genes within immune response and cell cycle regulation pathways was increased, whilst also demonstrating a downregulation of genes involved in autoimmunity and cell adhesion, muscle contraction, and metabolism (H.K. Song et al., 2012). Pressure-overload induced hypertrophy was also associated with the upregulation of p53 signalling, ECM-receptor interactions and adhesion, leukocyte migration, and cardiomyopathies (H.K. Song et al., 2012). Combined analysis of 3 of hypertrophy datasets showed differentially expressed genes were related to HIF-1α signalling in the hypoxia pathway, and other known hypertrophic signalling pathways including MAPK and PI3K-signalling (Z. Zhang and C. Wang, 2022).

Targeted gene analysis and microarray data have implicated a role of the 'foetal-gene' programme in the development of LVH in pathological contexts (Fazilaty and Basler, 2023), controlled through activation of transcription factors such as NFAT, MEF2A and GATA4 (Dirkx, Costa Martins, and De Windt, 2013). A comparison of gene expression during foetal development and pressure-overload induced hypertrophy showed a shared upregulation of genes including *Glut1, Nppa, Nppb, Acta1*, and *Camk2a*, genes involved in a number of hypertrophy-associated pathways including response to reactive oxygen species and glucose metabolism (Ames et al., 2013). RNA-sequencing provides an unbiased assessment of the transcriptome, which lends itself to enrichment analysis of transcriptional regulators controlling differentially expressed genes. Shared upstream regulators between foetal gene expression and LVH models included IGF receptor-1 (IGFR-1), GATA4, and growth hormone (GH)-1 (Ames et al., 2013). In pathological hypertrophy, motifs for the *Foxm1* and *PU.1* transcription factors were enriched in 417 genes upregulated from a low basal expression in control, sedentary, or physiological hypertrophy conditions (H.K. Song et al., 2012). Targets of the developmental *Hox5a* transcription factor were commonly enriched between foetal and failing myocardia (Dewey et al., 2011). Given that the number of genes coordinately regulated in foetal hearts and hearts with LVH is higher than would be expected in the absence of shared mechanisms (Ames et al., 2013), genetic variants not overtly pathological in development could become pathological if reactivated under hypertrophic stimuli in the context of remodelling. Full understanding of pathways and mechanisms contributing to LVH could lie in understanding developmental processes of the heart.

#### 1.3.3.2 Single Cell and Single Nuclei RNA Sequencing

Single cell (sc-) or single nuclei (sn)RNA sequencing is a technological enhancement of bulk RNA sequencing. Rather than averaging expression of transcripts over the tissue,

sc/sn-RNA seq obtains the gene expression profile of single, individual cells. This allows identification of cell types which have different phenotypes between healthy and diseased models, rare cell populations, cell-cell communication networks, and lineage trajectory reconstitution (Gladka, 2021) (Figure 1.7). Individual cells are captured in wells or droplets with a single bead containing oligonucleotide primers, a cell specific barcode, and unique molecular identifiers (UMI). Reverse transcription into barcoded cDNA (tagged with UMI) occurs within individual beads used for library preparation and deep-sequencing. Sequenced reads are subject to quality control, differential gene expression and 2D visualisation using t-distributed stochastic neighbour embedding (tSNE) and unsupervised clustering analyses.

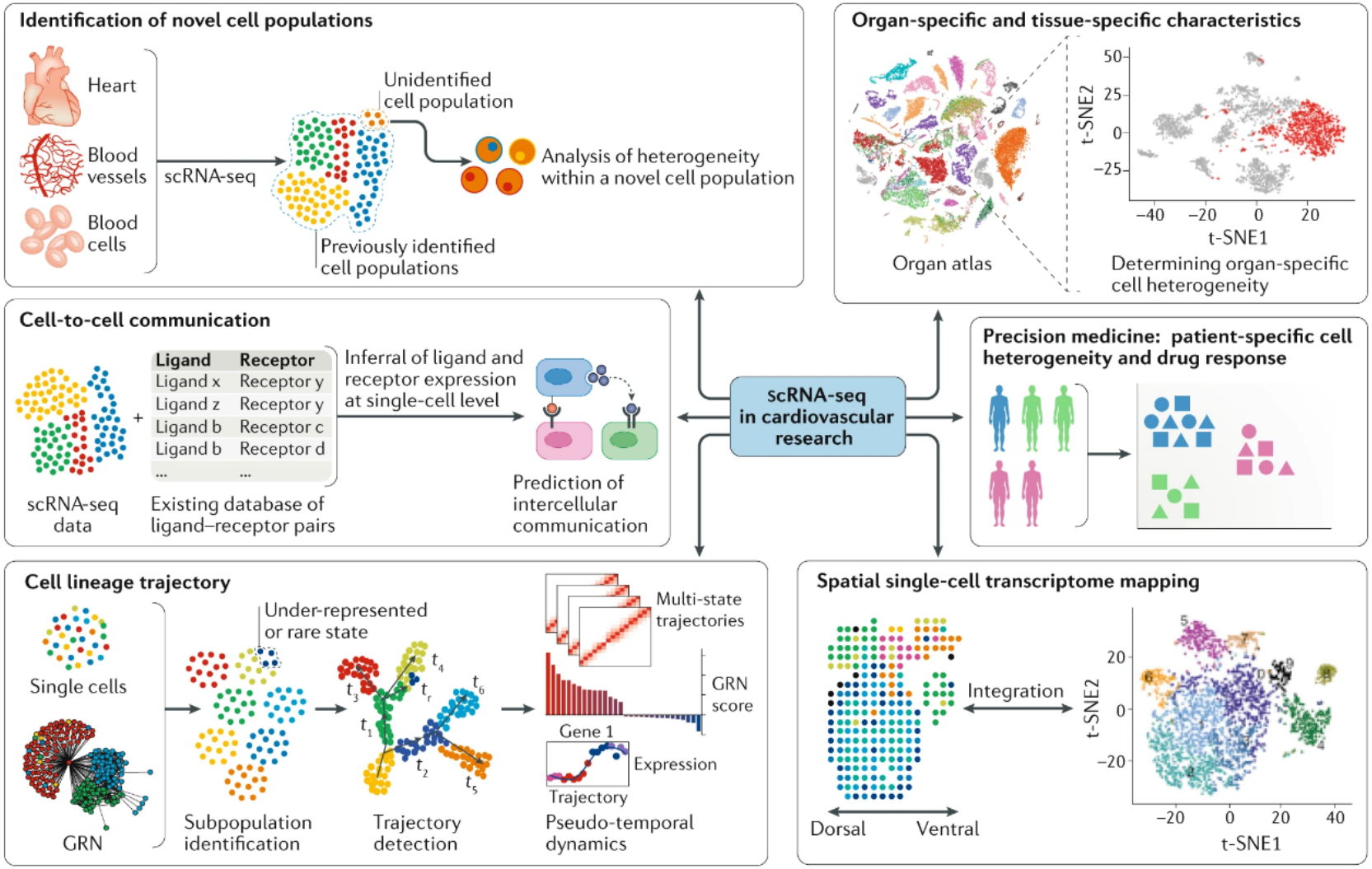



**Figure 1.7: Applications of Single Cell RNA sequencing** Taken from (Paik et al., 2020), graphical representation of key applications of single cell ang single nuclear RNA sequencing in cardiovascular disease.

Single nuclei (sn)RNA sequencing is an alternative approach to scRNA-seq which overcomes two problems in the heart; firstly, the heart is difficult to digest due to strong ECM interactions, and cell viability after enzymatic preparations from fresh tissue can be low. Second, cardiomyocytes are large, elongated cells, which are too big for most scRNA-seq platforms unless isolated from small models (zebrafish, mice) (Pimpalwar et al., 2020). snRNA sequencing is also suitable for flash frozen or archived tissues, from which single cell preparations can not be isolated (Slyper et al., 2020). Nuclei can be isolated from fresh or frozen tissue for snRNA sequencing analysis in the same way as individual cells, and pipelines are the same for isolated single nuclei suspensions as isolated single cells. In the heart, comparison of sc- and snRNA sequencing confirmed nuclear and cellular transcriptomes were highly correlated and display high concordance (Pimpalwar et al., 2020).

The output of sc/sn-RNA sequencing experiments is a cluster-based analysis of transcriptionally concordant units which represent individual cell types. Both genes and pathways highly enriched within clusters can be used to assign function and/or cellular identity to specific clusters. In the case of snRNA sequencing, proportions of cells reported from cardiomyocyte-enriched samples should be interpreted with caution as ∼25% of cardiomyocytes are reportedly multinucleated (Pimpalwar et al., 2020). The heart is made up of a complex and changing milieu of cells that are altered during growth and development. Cardiac fibroblasts and cardiomyocytes are the predominant cell types of the heart, but other cell populations include endothelial and smooth muscle cells associated with the extensive vasculature, adipocytes, and invading/resident immune-related cells (Litvinukova et al., 2020). As sc- and snRNA-sequencing has become more accessible, mapping cell types of the heart in greater depth has allowed the characterisation of more than 20 cell types across the 4 chambers of the human heart (Tucker et al., 2020). In the adult heart, ∼25–35% of cells in the adult heart are cardiomyocytes (Pinto et al., 2016) however despite their relatively low number, cardiomyocytes are thought to occupy ∼70% of adult myocardial volume (Anto Michel et al., 2022; Dewing et al., 2022). In the healthy adult heart, populations of 'activated-' (termed myofibroblasts) and 'non-activated' fibroblasts (Litvinukova et al., 2020; Tucker et al., 2020), equal in number the cardiomyocytes (Anto Michel et al., 2022). However, in diseased hearts, the transcriptional and proteomic landscape of these cells change to become pro-inflammatory and pro-fibrotic (Tallquist and Molkentin, 2017; Garvin et al., 2021).

Following induction of ischaemic injury in mice, scRNA-sequencing in control and ischaemic injured hearts indicated the appearance of 'disease-associated' cell clusters where an additional fibroblast cluster was present in injured myocardia, characterised by expression of genes associated with fibroblast activation (Gladka et al., 2018). The *Myoz2* sarcomere protein has been shown to tether α-actinin to calcineurin, reduc-

ing the availability of calcineurin to dephosphorylate NFAT and effectively inhibiting pathological hypertrophy in cardiomyocytes. Disease-specific clustering of cells has also been demonstrated in HCM, where UMAP projections generated HCM-specific clusters and a global enrichment of *Nppa, Nppb, Acta1* and *Myh7* (Wehrens et al., 2022). Increased cell-to-cell heterogeneity has been reported in models of ageing and senescence and was also observed following pressure-overload induced hypertrophy (Nomura et al., 2018; Aghagolzadeh et al., 2023). Increased transcriptional heterogeneity might underlie emergence of disease specific clusters, as gene regulation becomes disrupted. Time-course analysis of the failing heart, in response to pressure-overload, concluded clusters of cells activating mitochondrial biogenesis were the source of oxidative stress development leading to DNA damage and p53 signalling to drive hypertrophic responses (Nomura et al., 2018).

sc & snRNA sequencing analyses often includes exploration of ligand-receptor partnerships. Maps of ligand-receptor pairings in sc/sn-RNA sequencing have shown high inter- and intra-connectivity between cell types of the heart revealing fibroblasts have the highest number of receptor-ligand connections, and represent a highly regulatory cell type in the myocardium (Skelly et al., 2018). In response to myocardial infarction, scRNA sequencing showed localisation of signals to specific cell types, where pro-fibrotic TGF-β signalling preferentially targeted activated myofibroblasts (Aghagolzadeh et al., 2023). In the developing heart, scRNA sequencing identified Notch signalling pathways that controlled epicardial cell differentiation and the enrichment of Peroxisome proliferator-activated receptors (PPAR) in ventricular compared to atrial cardiomyocytes (Cui et al., 2019).

### 1.3.4 Multi-Omic Analysis in Cardiovascular Disease

Association between genetic variants, transcriptional activation, or differences in proteomes, metabolome, or lipidome require integrated investigation of potential function to attribute cause or effect to individual molecules or pathways. As the field has developed, the integration of multiple -omics datasets, from the same or functionally similar samples have the potential to unravel pathways of effect from the appearance of divergence in molecular phenotype (genetic mutation, dysregulated expression, changes in proteome, lipidome or metabolome) back to its genetic cause, or to the measurable phenotype. In a HCM cohort, samples underwent RNA-sequencing, mass spectrometry, and phosphoproteomic profiling, revealing a downregulation in hypertrophy associated pathways in the transcriptome associated with a proteome and phosphoproteome concurrently supporting enhanced RAS-MAPK signalling cascades (Garmany et al., 2023).

Correlation, network formation and machine learning are currently utilised in multi-omic experiments and a number of online tools exist to facilitate this (C. Chen et al., 2023). In the context of LVH, an integration of GWAS signals, RNA sequencing and

epigenome wide mapping of DNA-methylation identified associations between methylation status and expression of natriuretic peptides and DCM. In this case, hypomethylation of the 5'- CpG was associated with increased mRNA expression of *NPPA* and *NPPB* (Meder et al., 2017). Integrated analysis of the transcriptome and proteome in mouse models of LVH indicated an enrichment of cell adhesion, glycolytic process, actin filament organisation, translation, and sodium ion transport (Lau et al., 2018; C. Chen et al., 2023). Combining in-depth analyses has been aided by the development of repositories and large 'atlases', such that integration of the knowledge base can be inferred without requiring additional costly experiments. The *MTSS1* locus was identified as a lead variant for LVMI traits in both non-european and european cohorts and was found to have differential expression in the LV based on genotype in the GTEx database, where the risk allele was associated an increase in LV *MTSS1* expression (Wild et al., 2017). Similarly, variants in *MYOZ1* and *TNNT3*, associated with CMR traits, were also supported by eQTL or transcription wide association (TWAS) (Khurshid et al., 2023). Integrating -omics analyses is challenging both methodologically, statistically and computationally. Experiments differ in their sample collection time, collection conditions, and sample preparation, meaning careful design is required for integrating such data.

## 1.4 Animal Models of Human CVD and Related Disorders

Despite some understanding of the epidemiology of hypertension and LVH, scientific understanding of the mechanisms causing the heart to adversely remodel, and clinical ability to stop or even reverse such remodelling is currently lacking. Animal models used in biomedical research include sheep, pig, hamster, rats, mice, and zerbafish. Each model requires extensive characterisation to determine its potential similarities and limitations in the context of human disease (Leong, C.Y. Ng, and Jaarin, 2015). Rats were one of the first mammalian species to be used in CVD research, preferred due to their larger size and greater physiological similarity to humans (Homberg, Wohr, and Alenina, 2017). The larger size of the laboratory rat over the mouse affords the opportunity for serial blood collections, more complicated surgery, and radiotelemetry of aortic vessels (Homberg, Wohr, and Alenina, 2017; Szpirer, 2020). The rat also makes a good choice in pre-clinical models for drug development as it shares a similar pathway for eradicating toxins as humans (Szpirer, 2020). In addition, rats have a relatively short breeding cycle (21-days) and large litters (6-15) (Homberg, Wohr, and Alenina, 2017).

Since the first appearance of brown rats in research in the early 1800s, over 4000 rat strains have been created and documented on the Rat Genome Database (J.R. Smith et al., 2020). Rat strain ontology reflects the breeding history, genetic manipulation and genetic background of rat strains used in research (Nigam et al., 2013) (Table 1.2). As

in a human population, rat strains have different susceptibility to complex diseases, and in many cases, have been selectively inbred to maintain specific, relevant phenotypes (Hermsen et al., 2015). Inbred since 1958, the Brown Norway rat was used to generate the first rat reference genome, which has undergone continual updates since its first release in 2003 (Gibbs et al., 2004; J.R. Smith et al., 2020). According to the most recent update of the Rat Genome Database, CVD is linked to 4,455 rat genes, 704 quantitative trait loci (QTL) and is modelled by ∼175 strains (Vedi et al., 2023).

**Table 1.2:** Rat Strain Ontology Definitions

| Strain Type | Definition |
|---|---|
| Inbred | Defined genetic background. Maintained by brother sister mating for 20 or more consecutive generations. Inbred rat genome is understood to have on average only 0.01 residual heterozygosity (excluding any genetic drift) and regarded as genetically identical. |
| Mutant | Structural change in the DNA. This could be an insertion, deletion or chromosomal rearrangement generated spontaneously or by artificial methods |
| Transgenic | DNA fragment that has been stably introduced into the germline which can be generated by a random insertion into the genome |
| Coisogenic | Inbred strains that differ at only a single locus through mutation occurring in that strain |
| Congenic | Produced by repeated backcrosses to an inbred strain with selection for a particular marker from the donor strain |
| Conplastic | Strains in which the nuclear genome from one strain has been crossed onto the cytoplasm of the other during the backcrossing program. The mitochondrial donor is always the female parent. |
| Consomic | Produced by repeated backcrossing of a whole chromosome onto an inbred strain |
| Hybrid | Progeny of two inbred strains, crossed in the same direction, and are genetically identical |
| Segregating Inbred | Inbred strains in which a particular allele or mutation is maintained in a heterozygous state |
| Recombinant Inbred | Formed by crossing two inbred strains, followed by 20 or more consecutive generations of brother and sister mating |
| Outbred | Genetically undefined, these are bred with animals that are unrelated to each other, and not bred with siblings or close relatives |
| Wild | Rats Captured in the wild |

Strain definitions as described by the Rat Genome Database

Inbreeding creates a stable, fixed genetic background which can be risk-related or offer relative protection. A converse approach is to create a controlled yet genetically diverse 'strain'. Rather than displaying phenotypes which comprise the two extremes of possible values, a genetically diverse strain will have a diverse range of phenotypes which can be associated with the presence or absence of specific alleles. Commonly used in research utilising mouse models, recombinant inbred (RI, Table 1.2) strains,

can be applied in genetic research for linkage and association studies. A collaborative effort using the HXB-BXH and FXLE-LEXF rat RI panels, mapped over 16,000 single nucleotide polymorphisms (SNP), and identified QTLs associated with 74 phenotype parameters (Star Consortium et al., 2008), improving the rat SNP and genome variation catalogues. Since, the Hybrid Rat Diversity Panel (HRDP) have generated a strain-specific variant catalogue for the most recent rat reference genome which is available via the Rat Genome Database (Tabakoff et al., 2019; Vedi et al., 2023). As individual RI strains are inbred, they have the benefit of being useful in follow-up studies with genetically identical animals. An alternative method of increasing genetic diversity in rat models is the generation of 'heterogenous stock' (HS) animals. In order to generate a colony which has complex genetic variability mimicking human populations, 8 'founder' inbred strains were extensively interbred, resulting in an 'outbred' colony where each rat is a unique mosaic of the founding strains (Hansen and Spuhler, 1984). HS animals are suited to GWAS analyses, and even QTL mapping in the genetically diverse HS could improve resolution up to 30x (Mott et al., 2000). CVD-associated phenotypes have not yet been fully explored by GWAS in these populations, however exploration of renal phenotypes has shown a diversity of renal traits suitable for fine-mapping of associated genetic regions (Solberg Woods et al., 2010).

A large GWAS in HS stock animals is yet to investigate cardiac parameters, but adiposity related SNPs have been reported in *Rnf213, Xbp, Prlhr, Adcy3* and *Ngp* genes (Chitre et al., 2020). A large QTL-based analysis of HS rats mapped 122 phenotypes, and found association with the *Shank2* gene on rat chromosome 1 with heart weight (Rat Genome Sequencing and Mapping Consortium et al., 2013). Blood pressure and heart weight were also among 195 traits reported in a population of 2,006 genotyped HS rats (Baud et al., 2014). How one translates findings from rat GWAS or QTL is also difficult, as it is unlikely identical SNPs exist between human, mouse, and rat genomes. In addition, the large number of animals required, and the relatively higher cost of phenotyping, has limited its application. Using a network co-localisation framework, variants associated with BMI in human and rat GWAS studies were found to converge on a conserved set of 'BMI-network' genes (S.N. Wright et al., 2023). The molecular pathways of the 'BMI-associated' genes from each species showed significant convergence in gene function. Whilst genes containing variants did not themselves significantly overlap, the cross species functionally-associated networks could be annotated with more specificity than individual networks alone (S.N. Wright et al., 2023). The study of genetic variance in the rat is therefore of use to study in relation to human phenotypes. Finding variants contributing to phenotypic variability cross-species is methodologically challenging and requires consideration of both functional networks and phylogenetic conservation of genes (Hardison, 2016).

### 1.4.1 Genetic Models of Cardiovascular Disease

All models of disease utilised in research can be broadly categorised as inducible, superimposed or genetic. Inducible models are those where a genotype becomes active when particular diet or chemical stimulus is applied, such as tamoxifen-induced Cre-recombinase activity to induce gene editing in mice (Doetschman and Azhar, 2012). Genetic models are those which develop a phenotype without additional external stimuli, normally because they have been created by selective breeding of animals expressing desired phenotypes to generate genetic models of complex, polygenic traits. Finally, a model is superimposed when surgical or dietary interventions are given to cause specific phenotypes to develop. Interventions are often applied in specific, carefully selected relevant strains such that these categories are not mutually exclusive. Models selected for research are often a combination, dependent on specific phenotypes being recapitulated, availability of relevant animal models, and the context or aims of experimental questions.

Unlike in mice, where CVD phenotypes are often induced or superimposed, models of polygenetic CVD exist in the rat. This makes rat genetic models particularly suited for use in research as they have chronic and stable disease states coupled with predictable and controllable symptoms, to allow measurement of relevant cardiac, haemodynamic, and biochemical parameters (Doggrell and L. Brown, 1998). There are a number of genetic models of hypertension in the rat which also develop relevant end-organ disease, comparable to HMOD reported in human populations (Doggrell and L. Brown, 1998; Dornas and M.E. Silva, 2011). The Dahl Salt-Sensitive (SS) and Dahl Salt-Resistant (SR) strains were developed as new genetically inbred strains from the Sprague-Dawley rat, where the SS has a relative hypersensitivity to sodium intake (Jasinska-Stroschein, 2022). This genetic and superimposed model recapitulates the two extremes of human salt-sensitivity such that SS animals develop systemic arterial hypertension after ingesting a high-salt diet, while the SR animals can maintain BP even with the same diet (Dornas and M.E. Silva, 2011).

One of the most commonly used genetic model of human hypertension is the Wistar-derived Spontaneously Hypertensive Rat (SHR). The SHR develops systemic hypertension from 4-6 weeks of age without pharmacological or surgical intervention (Doggrell and L. Brown, 1998; Dornas and M.E. Silva, 2011; Jasinska-Stroschein, 2022). The closest genetic control to the SHR is the Wistar-Kyoto (WKY) rats, that are normotensive but can experience hypertension when superimposed by surgical, pharmacological, or dietary intervention (Doggrell and L. Brown, 1998). Other genetically hypertensive strains include the Milan, New Zealand, Prague, Lyon, and San Juan strains (Doggrell and L. Brown, 1998). The SHR strain was divided into substrains from which the SHR-Stroke Prone (SHRSP) was generated, selected for higher incidence of stochastic stroke, for which hypertension is a major risk factor in humans. Hypertension and

associated end organ diseases (stroke, LVH, renal damage) are more severe in male SHR and SHRSP versus females (Doggrell and L. Brown, 1998; Olivera and Graham, 2023), however females of the SHR-derived strains still develop more severe phenotypes than 'control' WKYs. The SHR, SHRSP, and WKY strains were distributed to several laboratories before they were established as fully inbred strains. It is therefore important to determine the exact sub-strain or colony of the WKY, SHR, and SHRSP utilised in experimental designs (Figure 1.8). In 1991, colonies of the WKY and SHRSP were received from the NIH by the University of Glasgow. The resulting WKY/Gla and SHRSP/Gla colonies are the result of strain specific brother-sister mating of 13 WKY/SHRSP (6 males and 7 females of each) animals that were obtained from Dr D.F. Bohr (Department of Physiology, University of Michigan, Ann Arbor), where they had been maintained as inbred colonies for more than 15 years.

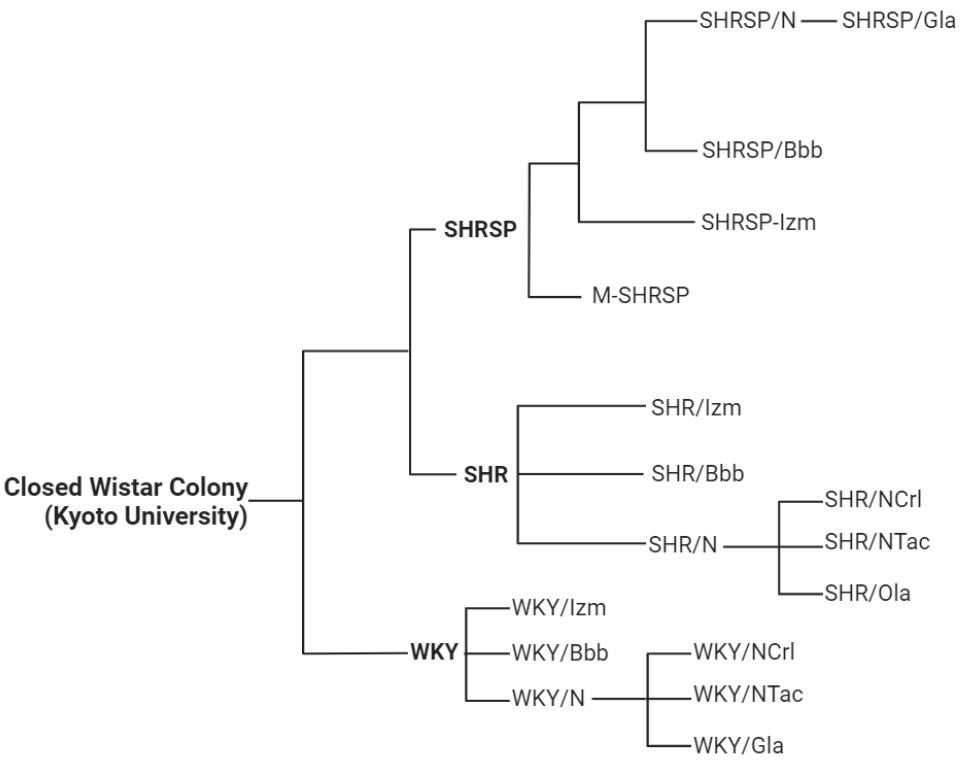



**Figure 1.8: Common Wistar derived WKY, SHR, and SHRSP sub-strains in CVD Research** Adapted from Nabika et al. (2012) shows the substrains of WKY, SHR, and SHRSP commonly used in research. The SHRSP/Gla and WKY/Gla have been maintained as fully inbred colonies at the University of Glasgow since 1991, when they were received from the NIH breeding stocks (WKY/N & SHRSP/N). Gla; also known as gcrc, Glasgow Cardiovascular Research Centre, Bbb; University of Heidelberg, Heidelburg, Germany, Izm; Izumo colony National BioResource Project for the Rat in Japan, M-SHRSP (malignant or precocious SHRSP); also known as SHRSP-A3, National BioResource Project for the Rat in Japan, N; NIH Autoimmune Rat Model Repository and Development Center, NCrl; Charles River Laboratories, received in Charles River in 1974 from NIH at F11 (not fully inbred), NTac; Taconic, received from the NIH Animal Genetic Resource in 1974 at F10 (not fully inbred), Ola; Czech Academy of Sciences, Prague, Czech Republic

The SHRSP is a unique model of severe hypertension and stroke, mimicking a

number of important phenotypes developed in human hypertension. Depending on the precise sub-strain, SHRSP animals will develop impaired endothelial function, vascular remodelling, decreased renal function, cardiac hypertrophy, reduced cerebral blood flow, and in some cases, progression to heart failure (Dornas and M.E. Silva, 2011). Compared to inducible or superimposed models, the WKY and SHRSP are useful to determine causative genetic mutations and genetically-influenced pathways associated with pathological processes which result in disease development. Unlike inducible or superimposed models, the SHRSP is not a single gene model, recapitulating polygenetic hypertension phenotypes observed in humans.

### 1.4.2 Models of Left Ventricular Hypertrophy

According to the RGD, including the SHRSP/Gla, there are 24 substrains of the SHRSP distributed around research institutes across the world. In addition, there are 8 consomic strains, and a further 120 congenic strains which use the SHRSP as either the background or regional donor strain. There are 4 mutant strains catalogued by the RGD; SHRSP-*Spp1*$^{em1Mcwi}$, SHRSP-*Spp1*$^{em2Mcwi}$, SHRSP-*Stim1*$^{em1Izm}$, SHRSP/Ta, and a single transgenic strain (SHRSP-Tg(Tagln-ACE2)6918Bdr). Each strain will have a specific cardiac, vascular and renal phenotype depending on their specific genetic background. An additional number of locally generated, unpublished strains will exist. Whilst hypertension and stroke susceptability is the major phenotype of the SHRSP, increased LVMI compared to the WKY is observable in the SHRSP/Gla strain from approximately 5-week of age. This precedes the development of increased blood pressure in the SHRSP such that these strains specifically represent hypertensive patients at higher genetic risk of LVH, stroke and other CVD (Nabika et al., 2012).

Akin to human LVH which can be caused by environmental or lifestyle factors, LVH can be superimposed by high fat diet (HFD) feeding, resulting in obesity and systolic dysfunction of the heart (Jasinska-Stroschein, 2022). Primarily utilised as model of type 2 diabetes, the Zucker Diabetic Fatty rat have a missense mutation in the leptin receptor and develop cardiac phenotypes including increased LVMI, diastolic dysfunction, and fibrosis, without reducing systolic function (Marsh et al., 2007; Jasinska-Stroschein, 2022). The deoxycorticosterone acetate (DOCA)-salt model was first used in the 1970s as a superimposed model of hypertension, specifically focused on the role of the sympathetic nervous and renin-angiotensin systems (Basting and Lazartigues, 2017). Administration of 20-50 mg/kg of DOCA results in hypertension as well as an increase in LVMI and cardiomyocyte cross-sectional area (Basting and Lazartigues, 2017). Systolic function is preserved, but diastolic function tends to be decreased in this model of superimposed hypertension and LVH (Basting and Lazartigues, 2017; Jasinska-Stroschein, 2022). Administration of AngII or aldosterone via injection, or implation of intraperitoneal minipump, is commonly used to superimpose LVH as a result of hypertension in animal models. Following a period of AngII or al-

dosterone administration, ejection fraction tends to be preserved despite development of increased LVMI and excess fibrosis (Jasinska-Stroschein, 2022), effectively reducing diastolic function (Jasinska-Stroschein, 2022). Finally, cardiac hypertrophy can also be induced surgically through transverse aortic constriction (TAC), whereby constriction of the ascending aorta causes pressure-overload to stimulate cardiac remodelling as a consequence of chronically elevated peripheral pressure (Gs et al., 2014). Genetic background of models can alter responses to hypertrophic stimuli and can be an important factor in study design and interpretation.

Depsite the robust increase in LVH prevalence among hypertensive patients, LVH is not uniformly developed. Due to the strong genetic influence on LVMI, models of genetically determined LVH might offer improved sensitivity to determine causal mechanisms of LVH than superimposed models. In part, the benefit of genetically determined LVH overcomes confounding factors introduced by the fact that LVH is not always the major phenotype of superimposed models, and is often a consequence of elevated blood pressure, renal dysfunction, or increased adiposity (Gs et al., 2014). The Dahl-SS hypertension model often experiences reduced systolic and diastolic function, increased LVMI, fibrosis, and exercise intolerance on exposition to 2-8% NaCl, the severity of which depends on induction period and intensity of salt loading (Jasinska-Stroschein, 2022). Thus, the Dahl-SS does not represent a true genetic model of LVH that is independent of blood pressure. Harrap et al. (2002) developed a normotensive model of LVH from interbreeding the F344 & SHR strains and selectively breeding littermates with the lowest BP combined with either high or low heart weight. The resultant Hypertrophic Heart Rat (HHR) strain develop increased LVMI compared to both age and weight matched Normal Heart Rats (NHR) despite similar, 'normotensive' blood pressure profiles (Harrap et al., 2002). The HHR and NHR are locally developed substrains by a group in Australia and are not currently documented by the RGD. The SHRSP, whilst not a pressure independent model, develops measurable differences in LVMI before the onset of high blood pressure, and therefore represents the hypertensive populations at greater genetic risk of LVH. In contrast, the HHR represents a non-hypertensive population at a greater genetic risk of developing LVH, and might be more representative of polygenic cardiomyopathies such as DCM and/or HCM.

### 1.4.3 QTL Mapping in Rat Models of Left Ventricular hypertrophy

Quantitative Trait Loci (QTL) mapping is a hugely popular statistical method which links phenotypic data (trait measurements) and to genotypic data to explain genetic basis of variation in complex traits (Grisel and Crabbe, 1995). Unlike GWAS, which tends to be performed on unrelated individuals, QTL mapping is based on linkage of traits in a bi-parental population. In addition, QTL infers linkage of traits to genetic regions by association of molecular markers to phenotype of interest, these markers can be genetic markers defined along chromosomal regions, or markers associated with

gene or protein expression (termed eQTL and pQTL respectively). In contrast to QTL studies, GWAS also tends to be performed on a higher density of genetic markers. QTL studies in the rat use F2 hybrid strains (Table 1.2), generated by mating two opposing inbred strains which differ phenotypically and genetically. The F2 progeny (resulting from 2 generations of brother-sister mating) are then phenotyped and genotyped and loci linked to the trait/s of interest are identified.

The RGD contains a catalogue of QTLs associated with CVD traits, including increased cardiac and left ventricle mass. There is an estimated 704 loci associated with CVD traits catalogued (Vedi et al., 2023), however, the true number of loci in laboratory rat populations which modify CV physiology and pathology is still unknown. Using the F344 as a normotensive, low heart weight control, and the SHRSP, QTL mapping of the F2 population identified a large region on rat chromosome 1 was independently associated with LVMI (Grabowski, Koplin, et al., 2013). Generation of a chromosome 1 consomic strain (Table 1.2) on the SHRSP background exhibited an intermediate phenotype compared to parental SHRSP and F344 chomosome 1 donor strains (Grabowski, Koplin, et al., 2013). Another study using an F2 population derived from same strains reported an association of LVMI with a similar region of chromosome 1, along with an additional region of chromosome 2 (Innes et al., 1998). The loci was then finemapped to generate a number of candidate genes including *Fgf2, Trpc3, Trpc4, Sclt1* and *Kcnmb2* (Di Nicolantonio et al., 2006). An earlier study used the SHR, SHRSP, and WKY strains to generate two F2 populations, and reported significant association of LVMI across 4 chromosomes (Inomata et al., 2005). Chromosome 3 harbored a sex-independent QTL for LVMI in the F2 populations, in addition to a number of sex-specific loci including a male-specific linkage on chromosome 10 (proximal to the *Ace* gene), and a female-specific loci on chromosome 6 (Inomata et al., 2005). An F2 hybrid generation from a Dahl-SS and SHR cross found 4 significant QTLs for SBP and LVH, supporting an association between rat chromosome 3 and LVMI (Siegel et al., 2003). QTLs are a useful tool in rat genetics. However, QTL analyses implicate large regions of the genome in an associative manner, and determination of mechanistic pathways or identification of specific causal gene/s requires further validation using consomic, congenic or mutant strains.

## 1.5 *Spp1*: A positional and functional candidate gene

The WKY/Gla and SHRSP/Gla were crossed to generate a hybrid strain suitable for QTL mapping with and without salt loading of the F2 progeny (J.S. Clark et al., 1996). A QTL on rat chromosome 14 for LVMI was detected, localised between markers D14Mgh3 and R58 (∼12.3cM apart), which accounted for 32.3% and 12.3% of variance in LVMI in males and females respectively (J.S. Clark et al., 1996). Congenic strains represent a more precise transfer of genetic material compared to consomic

strains, and to confirm the associated region of chromosome 14 as a positional, 'susceptibility' locus for pathological cardiac remodelling, reciprocal congenic strains were developed on the SHRSP/Gla and WKY/Gla backgrounds. Congenic strains are generated by transferring a chromosome segment into a background, 'recipient' strain from a 'donor' strain by serial backcrossing and marker based selection. Selection is made at microsatellite marker locus (M) through a series of backcrosses such that the allele of a donor ($M_1$) replaces the allele of a recipient ($M_2$) in a method origanlly developed by Snell (1948). The classic method of congenic generation employs selection of progeny for heterozygosity at the desired region, requiring 8–12 backcrosses according to Mendelian principles, before brother-sister mating to fix the desired chromosomal region in a homozygous state for donor alleles (Jeffs et al., 2000) (Figure 1.9).

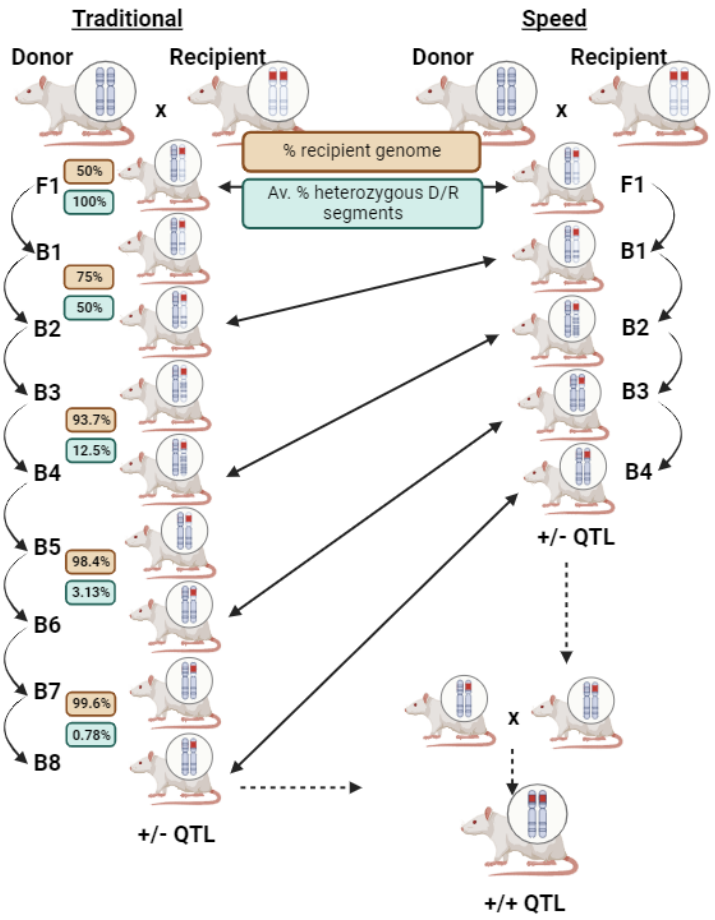



**Figure 1.9: Speed Congenic Generation** Illustration created in BioRender.com as an adaptation from (McBride, Charchar, et al., 2004) to demonstrate the difference between traditional and marker-assisted speed congenic approaches. As a graphical representation, the QTL region of interested is colored red in the donor strain. Chromosomal segments become gradually darker to represent the serial dilution of transferred genetical material from the donor strain in the recipient with each round of backcrossing. Congenic becomes fixed after >99% of background recipient genome is present by brother sister mating of littermates heterozygous for QTL, 25% of resulting litter homozygous for QTL are brother sister mated to maintain congenic strain, (D; donor strain alleles, R; recipient strain alleles, QTL; quantiative trait loci, F1; first filial generation, B; backcross).

The chromosome 14 congenic strains generated on the WKY/Gla and SHRSP/Gla background (termed WKY.SPGla14a and SP.WKYGla14a respectively) employed a 'speed' congenic strategy (Figure 1.9) such that screening of polymorphic marker loci distributed throughout the genetic background allowed selection of males with the least amount of donor alleles remaining from the donor strain (Jeffs et al., 2000). In contrast to the traditional method, the number of backcrosses to eliminate donor genetic material from the recipient is dramatically reduced (Jeffs et al., 2000). This 'speed'-congenic strategy is also referred to as a marker assisted strategy (MAS) and was first applied in mice (Morel et al., 1996; Estill and J.A. Garcia, 2000). Prior to the development of chromosome 14 congenic strains, the strategy was successfully employed in the WKY/Gla and SHRSP/Gla to transfer two independent segments of chromosome 2 (containing blood pressure associated QTLs) into both backgrounds, generating 4 congenic strains. The SHRSP background congenic strains showed decreased baseline and salt-loaded BP and there was also a small but significant increase in the BP of one of the WKY background congenic strains (Jeffs et al., 2000). Analysis of the developed strains identified Glutathione S-transferase µ-type 1 (*Gstm1*) as a positional and functional candidate gene in hypertension, where reduced expression of *Gstm1* was associated with the SHRSP/Gla genome (McBride, Carr, et al., 2003; McBride, Brosnan, et al., 2005). Sequencing of the WKY/Gla and SHRSP/Gla *Gstm1* gene also identified multiple promoter region variants potentially responsible for the differences in renal *Gstm1* expression (McBride, Brosnan, et al., 2005).

The 'speed' congenic protocol was applied to generate two chromosome 14 congenic strains on the WKY and SHRSP backgrounds (generated the WKY.SPGla14a and SP.WKYGla14a respectively). Congenic generation was performed independently on both backgrounds such that the region of interest in the resulting strains is comparable, but not identical (Figure 1.10). Some level of residual, contaminating, donor genome will become fixed in the process of congenic generation and is unavoidable, even using the traditional method (Jeffs et al., 2000). The exact level and chromosomal locations of these sites of residual genetic transfer from the donor will also differ between the WKY.SPGla14a and SP.WKYGla14a.

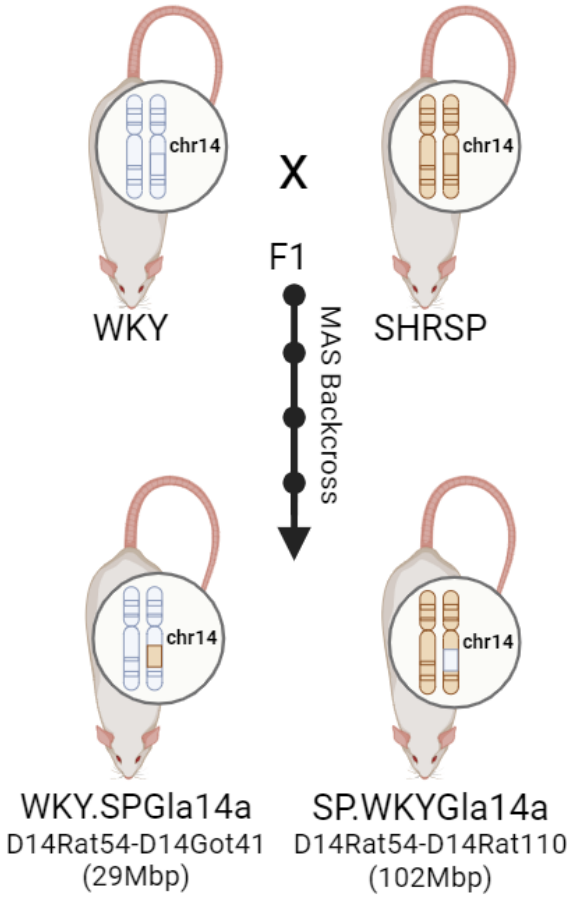



**Figure 1.10: SHRSP and WKY Chromosome 14 Congenic Generation** Schematic representation of WKY and SHRSP chromosome 14 congenic strains harbouring donor chromosome 14 region identified as a QTL for LVMI. Micosatellite genotyping confirmed transfer of WKY chromosome 14 region from D14Rat54 (5116492bp)–D14Rat100 (107119581bp) into the SHRSP. Transfer of 29Mbp region from D14Rat54 (5116492bp)–D14Got41 (34375297bp) from the SHRSP into the WKY recipient.

Based on the significant association of the congenic chromosome 14 region with LVMI, the congenic strains; WKY.SPGla14a and SP.WKYGla14a, were assessed for cardiac structure and function following their establishment as an inbred congenic strains (Figure 1.10). Initial phenotyping focused on male rats, and using a combination of echocardiography, pressure-volume (PV) loop, and histological assessment, significant divergence of cardiac phenotypes in both congenic strains from respective backgrounds was determined. Whilst SBP was slightly raised in WKY.SPGla14a compared to the background WKY, the SP.WKYGla14a did not have significantly decreased SBP compared to the SHRSP. Conversely, LVMI and RWT were reduced in the SP.WKYGla14a congenic, compared to its SHRSP background strain. LVMI and RWT was simultaneously increased in WKY.SPGla14a congenic compared to its WKY background strain (Douglas et al., 2010). Once divergence of phenotype was established in the male chromosome 14 congenic strains, gene expression microarray was carried out in neonate, 5-week, and 16-week male congenic animals to determine potential causal genes and pathways associated with the development of LVH in normoten-

sive (WKY.SPGla14a) and hypertensive (SHRSP) backgrounds. Located within the chromosome 14 congenic region, secreted phosphoprotein 1 (*Spp1*) was significantly upregulated in the heart of SHRSP and WKY.SPGla14a 1-day neonates, compared to the SP.WKYGla14a and WKY strains. Four other genes were identified in microarray analysis including hydroxy-steroid dehydrogenase 17-B13 (*Hsd17b13*), Zinc finger protein 644 (*Znf644*) and Ras-related glycolysis inhibitor and calcium channel regulator (*Rrad*) (Asirvatham, 2022). Only *Spp1* mapped to the congenic region and was subsequently validated in cardiac tissue and isolated primary cardiomyocytes from neonatal SHRSP and WKY strains (Crawford, 2010).

### 1.5.1 *Spp1* Structure and Physiological Function

*Spp1* encodes for the osteopontin (OPN) protein, a multi-functional, non-collagenous matrix glycophosphoprotein expressed in a number of cells and tissues (Du et al., 2022). The *Spp1* gene is a member of the small integrin binding N-linked glycoprotein (SIBLING) family, along with integrin binding bone sialoprotein (*Ibsp*), dentin matrix protein 1 (*Dmp1*), dentin sialophosphoprotein (*Dspp*), and matrix extracellular phosphoglycoprotein (*Mepe*) (Bellahcene et al., 2008). On the human and mouse genomes, these genes cluster in a forward orientation on chromosomes 4 and 5 respectively (hg38 and mm39). Although on the opposing strand (reverse orientation), the *Dspp, Dmp1, Ibsp, Mepe* and *Spp1* genes cluster along a 2Kb region of the rat chromosome 14 (Figure 1.11). SIBLING proteins interact with cell surface integrins and CD44, acting as signal transducers to stimulate intracellular kinase cascades promoting cell adhesion, motility, and survival (Bellahcene et al., 2008). *Spp1* is translated into the osteopontin (OPN) protein, a negatively charged protein that is highly acidic. The OPN protein contains an arginine-glycine-aspartic (RGD) binding domain which becomes exposed following thrombin or matrix-metalloprotease (MMP-) cleavage (Icer and Gezmen-Karadag, 2018; Du et al., 2022). In addition to the highly conserved RGD binding site, OPN also contains a conserved Ser-Val-Val-Tyr-Gly-Leu-Arg (SVVYGLR) recognition site which binds the intergin $\alpha\beta1$ (Yokosaki et al., 1999). As is standard in the field, reference to the *Spp1* gene and osteopontin (OPN) protein will henceforth refer to study at the gene/transcript and protein level respectively.

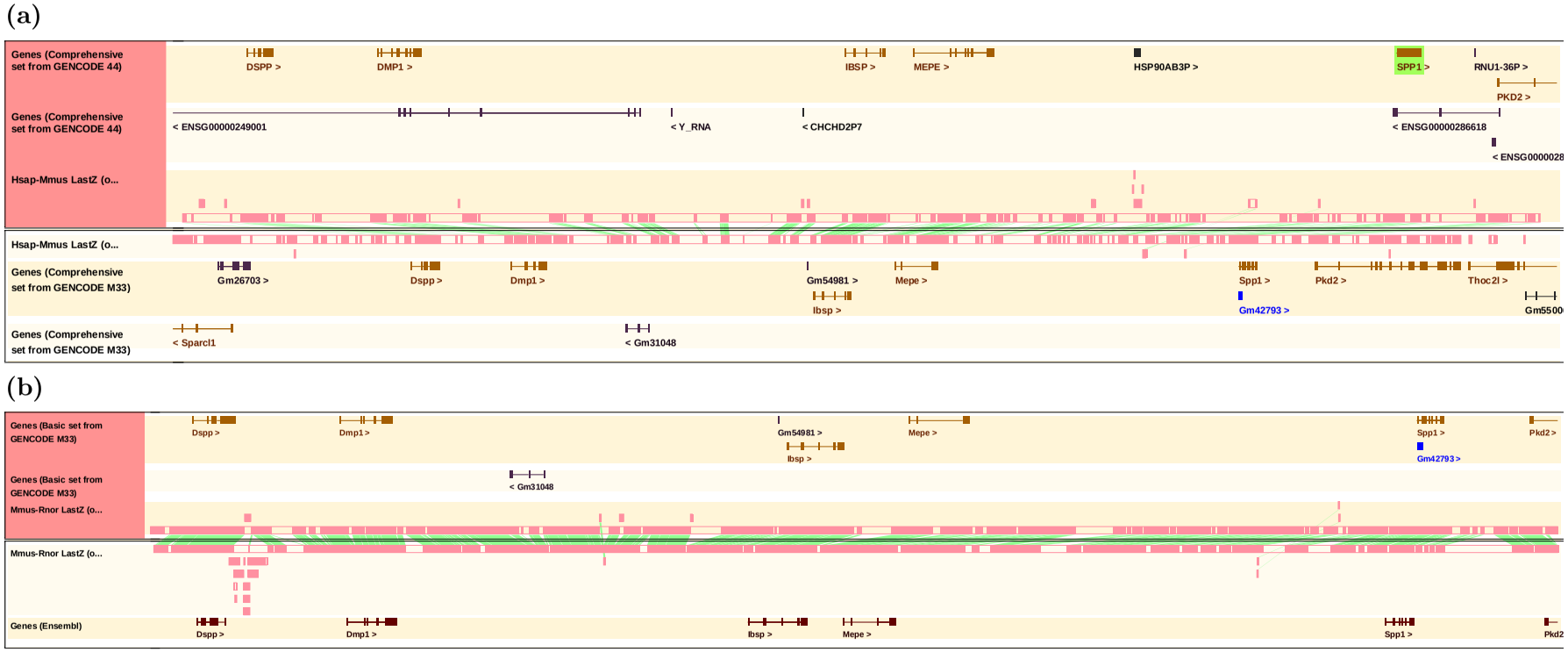



**Figure 1.11: SIBLING Family Genes in Human, Mouse, and Rat Genomes** Ensembl region comparison of (a) human (top) to mouse (bottom) and (b) mouse (top) to rat (bottom), to show the positional conservation of SIBLING genes along human chromosome 4, mouse chromosome 5, and rat chromosome 14

The canonical *Spp1* transcript in the human, mouse, and rat contains 7 exons where exon 1 is untranslated, and exons 2-7 contain the coding sequences. Exons 2-5 are relatively short and 80% of the coding region is encoded by exons 6 & 7 (Bastos et al., 2023), containing the RGD and SVVGLR domains. In the human and mouse, multiple splice variants of *Spp1* exist whereas the most recent rat annotation (mRatBN7.2) contains only two variants. In human dendritic cells, initiation of transcription from an alternative non-AUG start site produced a shortened OPN protein which localised within the cytoplasm rather than secretory vesicles (Shinohara, H.J. Kim, et al., 2008a). This intracellular OPN is predicted to localise to mitochondria and/or the nucleus (Bastos et al., 2023), however data on its physiological function is limited. In a murine model of innate anti-viral responses, intracellular OPN positively regulated IFN-β production to potentiate the anti-viral response of host cells (K. Zhao et al., 2016). Intracellular forms of OPN were also reported during all stages of cell differentiation in a rat osteogenic cell culture model (Zohar et al., 1997). This work preceded publication of the rat genome, and did not determine potential transcriptional regulation of intracellular OPN forms. Nuclear localisation of OPN was observed in embryonic kidney 293 (HEK293) cells, co-localising with mitosis regulator *Plk-1*, and was correlated with chromatin condensation during cell division (Junaid et al., 2007). The same authors showed OPN increased in the 'cell growth' ($G_2M$) phase and nuclear localisation coincided with DNA-synthesis during interphase/S-phase (Junaid et al., 2007).

GTEx single cell RNA-sequencing data predicts *Spp1* is most abundant in dendritic cells, macrophages, and Schwann cells of the left ventricle. *Spp1* is also detectable in fibroblasts, cardiomyocytes, vascular cells (endothelial and vascular smooth muscle), and adipocytes (GTEx Consortium, 2020). OPN interacts with the integrins αVβ3, αVβ5, αVβ1, α4β1, α8β1, and α9β1. Integrins have ubiquitous functions (adhesion, formation of extracellular matrix cyctoskeletal junctions, and signal transduction), and cardiomyocyte specific functions including; modification of ion channel function, stem-cell growth, hypertrophic growth, mechanotransduction, and ischaemic protection (Israeli-Rosenberg et al., 2014). Through integrin-mediated activation of NFϰB, OPN is thought to promote survival in dopaminergic neurons, endothelial cells, and dendritic cells (Kawamura et al., 2005; Iczkiewicz et al., 2006; J. Rice et al., 2006). OPN null mice show deficient wound healing, where OPN is important in the activation of macrophages for wound repair (Liaw et al., 1998). In response to sciatic nerve injury, *Spp1* regulated cytokines, c-Fos, PKCα, and the p-ERK/ERK pathway, to act as an anti-apoptotic factor preventing non-programmed cell death (X. Liu, Y. Sun, et al., 2017). Global knockout of *Spp1* does not appear to be lethal or even deleterious to embryogenesis in mice (Liaw et al., 1998; S. Li and Jakobs, 2022).

### 1.5.2 *Spp1* & Cardiovascular Disease

Whilst it is clear that *Spp1* is likely to be involved in a number of physiological pathways, it has been extensively investigated in association with a number of cancers, aging disorders of the brain (Parkinsons and ALS), and cardiovascular diseases (Du et al., 2022). As indicated by data from the GTEx project, expression of *Spp1* in the normal heart is relatively low and tightly controlled. Models of LVH in the rat (Giachelli et al., 1995; Ashizawa et al., 1996; Graf et al., 1997) and mouse (A.R. Collins et al., 2004; Lorenzen et al., 2015) indicate expression of *Spp1* is increased during pathological remodelling of the heart. Surgical induction of pressure overload by clipping of the renal artery, or aortic banding, increased OPN expression in the LV of Sprague-Dawley rats (Graf et al., 1997). Meta-analysis of heart failure in Dahl rats also showed *Spp1* was significantly increased in heart failure versus controls (Yim, H. Cho, and Rabkin, 2018). Increased expression of *Spp1* in human disorders characterised by LVH has been reported in patients undergoing heart transplant with idiopathic cardiomoyopathy, ischaemic heart disease, atrial fibrillation, and coronary heart disease (Graf et al., 1997; Irion et al., 2020; Hulsmans et al., 2023). Increased expression of *Spp1* is associated with increased LV filling pressure, LV stiffness, and increased expression of collagens (Okamoto and Imanaka-Yoshida, 2012; Lopez et al., 2013). LVH is associated with increased fibrosis, potentially mediated by the *Sox9* transcription factor, which colocalised with *Spp1* in regions of increased liver fibrosis (Pritchett et al., 2012; Schauer et al., 2019). Reducing expression of *Sox9* also reduced expression of *Spp1* mRNA and protein in hepatic stellate cells (Pritchett et al., 2012). Conversely, in embroyonic and post-natal heart valves, *Sox9* negatively regluated *Spp1* through binding to an SRY-binding site in the 5'- enhancer region of *Spp1* (Peacock et al., 2011).

In a proteomic study of patients across different stages of heart failure, OPN and 5 other proteins were upregulated at all disease stages (Michelhaugh et al., 2020). The upregulated proteins were enriched in pathways associated with cellular growth and signalling, post-translational phosphorylation, and regulation of the extracellular matrix (Michelhaugh et al., 2020). Comparison of myocardial expression of *Spp1* protein between failing and non-failing hearts from SHR and WKY rats, showed an increase in OPN within failing hearts, which was localised to the interstitial space and perivascular non-myocytes (K. Singh, Sirokman, et al., 1999). Endomyocardial biopsies of patients with hypertensive heart diease showed increased OPN expression that was localised to areas of fibrosis where fibroblasts and myofibroblasts reside (Lopez et al., 2013). Aortic banding of Sprawgue Dawley rats suggested superimposed hypertrophy induced OPN expression that was localised to myocytes and vascular smooth muscle cells but not interstitial spaces (Graf et al., 1997). Immunohistochemistry and in-situ hybridisation in samples from patients with DCM, also showed *Spp1* was primarily located in the cytoplasm of cardiomyocytes and determined cardiomyocytes were the primary source of increased *Spp1* in these patients (Satoh et al., 2005). Single-cell

RNA sequencing of fibroblasts in the SHR characterised 7 populations of fibroblasts in hypertensive and hypertrophied hearts (Garvin et al., 2021). *Spp1* was highly expressed in 'mildly-fibrobotic' gateway pool of fibroblasts, which served as precursor pool for proinflammtory/activated fibroblasts (Garvin et al., 2021). The cluster was also enriched for pathways associated with 'response to wound-healing', 'cell/fibroblast proliferation', 'regulation of cell-cycle', and 'response to cytokine' (Garvin et al., 2021). In contrast, following to myocardial infarction, *Spp1* was detected in cardiomyocytes, fibroblasts and immune cell populations at temporally distant timepoints, however *Spp1* was thought to originate from macrophages recruited to the infact zone (Komatsubara et al., 2003). In the diseased myocardium, the precise subcellular localisation and mechanism of action of *Spp1* is therefore unclear, and the precise location is likely state-specific.

The expression and function of the *Spp1* gene is potentially affected by genetic variants. Putative binding sits for the SP1, MYT zinc finger, and RUNX2 transcription factors were reportedly altered by 2 SNP mutations, and a small insertion (respectively) found in the promoter region of the human *Spp1* gene (Giacopelli et al., 2004). The SP1 binding site is highly conversed between species (Giacopelli et al., 2004). Investigation of the T>G SNP at the -66 position (rs28357094) showed T alleles conferred increased transcription factor binding in humans (Giacopelli et al., 2004), and might represent a 'risk' allele in human populations. This is consistent with a DCM cohort of patients where T/T genotype at rs28357094 was associated with lower median age of DCM onset than T/G and G/G genotypes (Barp et al., 2015). In another DCM cohort, plasma OPN was increased versus controls, however in the same study, circulating OPN levels in HCM patients did not differ from controls (Podzimkova et al., 2019). This study did not investigate the effect of *Spp1* genotype. In skeletal muscle isolated from patients with Duchenne Muscular Dystrophy (DMD), T alleles at the -66 position were associated with increased OPN expression and more rapid disease progression (Pegoraro et al., 2011; Vianello et al., 2017). Patients with DMD often develop DCM, with no correlation to disease progression (Barp et al., 2015), similar to the non-uniform cardiac remodelling across hypertensive cohorts. Increased plasma OPN was associated with higher LVMI and poorer E'/e' ratio, but not with reduced % ejection fraction, in a hypertensive Chinese cohort (Y. Yang, Y. Wang, and P.J. Gao, 2020). The association of higher *Spp1* expression with increased diastolic dysfunction is consistent with a conditional transgenic mice model of cardiomyocyte specific *Spp1* overexpression developed by Renault et al. (2010). The transgenic mice experience development of a DCM-like phenotype with ventricular dilation and systolic dysfunction resulting in premature death by 12-weeks of age (Renault et al., 2010).

The small insertion polymoprhism at -156del/G poisition of the *Spp1* promoter was predicted to generate an additional RUNX2 binding site (Giacopelli et al., 2004),

and was associated with reduced diastolic function in a Japanese cohort of hypertensives (Nakayama et al., 2011). Interestingly, in the Japanese hypertensive cohort, the association of the -156 G allele with diastolic dysfunction was independent of cardiac hypertrophy (Nakayama et al., 2011). A T>C polymoprhism at position -442 in the *Spp1* promotor showed differential binding of a nuclear transcription factor, suspected to be the MYT zinc-finger (Giacopelli et al., 2004), and has also been associated with a loss of a STAT6 binding site (Deepti et al., 2022). The C/C haplotype was associated with increased *Spp1* expression in HPV positive cervical cancer and in plasma of breast carcinoma cases and controls (Liang et al., 2019; Deepti et al., 2022). A GWAS of serum OPN levels, found significant association between serum OPN and the rs10011284 SNP which is located upstream of *Spp1* (position -63412) and downstream of SIBLING family gene *Mepe* (Y. Cheng et al., 2022b).

Differences in *Spp1* expression are potentially influenced by common variants in promoter and upstream regions, however, a clear mechanism through which increased *Spp1* might contribute to LVH is unclear. The pathological role of *Spp1* was recently demonstrated in the context of atrial fibrillation, where treatment of wild-type mice with bone-marrow from OPN knockouts reduced *Spp1* in macrophages and improved resilience against atrial fibrillation (Hulsmans et al., 2023). Treatment of mouse neonatal cardiomyocytes with recombinant OPN significantly reduced ATP-linked oxygen consumption and suggested overexpression of OPN might induce mitochondrial dysfunction (Yousefi et al., 2019). The deleterious effect of *Spp1* overexpression was also demonstrated in the conditional transgenic mouse model developed by Renault et al. (2010). In addition to development of ventricular dilation and systolic dysfunction, cardiomyocyte-specific overexpression of *Spp1* induced accumulation of fibrosis and increased cell apoptotic activity (Renault et al., 2010). Conversely, OPN knockout mice treated with Angiotensin II (AngII) experienced reduced fibrosis and an attenuated hypertrophic response versus controls (A.R. Collins et al., 2004). At the cellular level, AngII treatment of neonatal rat cardiomyocytes did not increase OPN expression whereas both endothelin-1 and norepinephrine treatment induced a more than threefold increase in *Spp1* expression (Graf et al., 1997). An earlier study suggested AngII treatment increased *Spp1* expression in rat cardiac-fibroblasts and thus demonstrates potential cell-specific control of expression (Ashizawa et al., 1996). A mechanism by which *Spp1* mediates AngII induced hypertrophy has been proposed through activation of the AP-1 transcription factor to induce expression of the mircoRNA-21 (mir-21) and culminate in enhanced fibroblast survival and excess ECM production (Lorenzen et al., 2015). The same authors suggested secreted OPN could directly induce PI3-kinase activity (Lorenzen et al., 2015) presumably through interactions with integrins (Bellahcene et al., 2008). Similarly, OPN knock-out mice show reduced phosphorylation of c-Jun N-terminal kinases, p38 kinase, and Akt in response to pressure induced cardiac hypertrophy, supporting an ability of OPN to activate adverse remodelling through the

Akt, ERK-, and PI3-kinase pathways (Xie, M. Singh, and K. Singh, 2004b). Stimulation of the sodium ($Na^{2+}$)/ hydrogen$^{+}$ transporter (NHE1), which regulates intracellular pH and $Na^{2+}$, has been implicated in pathological cardiac hypertrophy in a pathway dependent on *Spp1* (Mohamed and Mraiche, 2015). In rat primary cardiomyocytes, and the H9c2 cardiac cell line, NEH1 adenovirus treatment resulted in upregulation of *Spp1* mRNA and OPN protein, mediated by the hypertrophic calcineurin/nuclear factor of activated T-cells (NFAT) pathway (Mohamed, Gadeau, et al., 2015). Double transgenic mouse combining overexpression of NHE1 with OPN knock-out showed improved cardiac phenotypes compared to NEH1 overexpression alone (Abdulrahman et al., 2018).

*Spp1* is tightly controlled and plays a physiological role in normal functioning however overexpression of *Spp1* has been described as potential biomarker predicting poorer outcomes and reduced survival. The OPN protein was upregluated throughout all stages of heart failure development (Michelhaugh et al., 2020) and promoter polymorphisms increased *Spp1* expression in the plasma of both control and breast carinoma patients (Liang et al., 2019). The heart failure drug Sacubitril/Valsartan was shown to confer its cardioprotective effects through reduction of *Spp1* (Shen et al., 2022) and *Spp1* overexpression is a key biomarker in a number of pathologies where excess fibrosis and inflammation are common (M. Singh, Dalal, and K. Singh, 2014). Whether *Spp1* is a marker of disease or a casual mediator of pathological pathways is less clear.

## 1.6 Thesis Hypothesis

We hypothesise that *Spp1* is a key positional and functional candidate gene contributing to LVH and fibrosis in our congenic model of LVH. The parental and congenic models represent a complex genetic model of human CVD and LVH, both on hypertensive and normotensive backgrounds. Genetically influenced changes in the expression of *Spp1*, prior to the onset of hypertension and increased LVMI, are thus hypothesised to initiate the cardiac phenotype of the SHRSP/Gla. To this end, we aim to systematically characterise functional effects of *Spp1* overexpression, as a basic biological pathway and in the context of a complex model of hypertension and CVD.

### 1.6.1 Thesis Aims

The aims of the experimental work conducted in this thesis are;

- Characterise the effect of the chromomsome 14 congenic region in male and female animals during early cardiac development into adulthood. To determine genetic influences on LVH, differences in cardiac phenotypes will be assessed without applying additional external pressures.

- Examine the functional pathway through which increased expression of *Spp1* influences hypertrophic growth within the H9c2 cardiac cell model.

- Determine the functional effect of reducing *Spp1* in the SHRSP rat model of hypertension and CVD through CRISPR/Cas9-induced genetic mutation.

- Use genomic variant analyses, bulk-RNA, and single nuclear RNA sequencing techniques in the WKY, SHRSP, and congenic rat strains to investigate specific genetic determinants of cardiac phenotypes, and the potential causal pathways influencing increased risk of LVH in adulthood.

# Chapter 2

# General Materials and Methods

This section will describe common methods within this body of work. Specific methods are outlined in each chapter with reference to methods detailed below where appropriate.

## 2.1 *In-Vivo* Methods

Animal procedures were approved by the Home Office according to the Animals Scientific Procedures Act (1986) and subject to local ethics approval (project licence: PP0895181, Glasgow Experimental Request Form (GERF): 037). WKY, SHRSP, chromosome 14 congenic and SHRSP-*Spp1*$^{em1Mcwi}$ strains were housed in the Cardiovascular Research Unit (CVRU) at the University of Glasgow. Animals were kept at 21°C ambient temperature, on a 12-hour light/dark cycle with *ad libitum* access to chow diet and water. Littermates were group housed, separated by sex, following weaning at approximately 28 days.

### 2.1.1 Tail Cuff Plethysmography

Animals underwent systolic blood pressure assessment by tail cuff plethysmography, using a validated system built in-house for the purpose of blood pressure measurement in conscious rats (A.L. Evans et al., 1994). To facilitate blood pressure measurements, and encourage dilation of peripheral vessels, rats were warmed for 10–15 minutes by placement in a polystyrene incubation box containing bedding to minimise stress. The animals were wrapped in a towel and gently restrained on a warmed mat with their tail exposed. An inflation cuff was placed around the base of the tail and a piezoelectric pressure transducer cuff was fitted to the tail-tip. This non-invasive method utilises an occluding tail cuff placed at the base of the tail which is inflated to occlude blood flow. Upon inflation sensors placed distal to the occlusion cuff detect changes in pulse-waves to measure blood pressure. To make each measurement, the occlusion cuff was inflated

by 1 mmHg in steps up to 250 mmHg to impede blood flow to the tail. Upward pressure is measured from the disappearance of pulses upon inflation of the cuff, the cuff is then gradually deflated, and downward pressure is recorded from the reappearance of pulses while deflating. The piezoelectric ceramic crystals in the transducer cuff record the pulse signal detected from the tail vein during occlusion and deflation.

### 2.1.2 Ultrasound Echocardiography

Echocardiographic assessment of cardiac geometry and function was performed using an Acuson Sequoia c512 ultrasound system with a linear array transducer set at a frequency of 15MHz. Presently, only 5- and 16-week old rats were assessed using echocardiography. Rats were anaesthetised and placed in a supine position. Light anaesthetic was maintained at 1.5% isoflurane in 1.5L/min medical $O_2$ during imaging. Hair was removed from the chest using an electric shaver and animals were positioned in a supine position with slight left lateral rotation. Imaging the heart along the short-axis plane, M-mode images were taken at the level of the papillary muscles. Images were adjusted during examination so that the LV cavity appeared black and myocardial walls a mid-grey to white. Simultaneous 2D imaging allowed correct positioning on the plane of interest during M-mode recordings. Examinations lasted until 5–10 images had been recorded with a regular cardiac cycle, and good edge definition between epicardial and endocardial surfaces of the ventricle walls.

Images analyses were conducted using the ImageJ software. Three images with good edge definition were selected and distance between wave peaks and troughs were measured to derive values of anterior and posterior wall thickness (AWT and PWT respectively) and, end diastolic and systolic diameter (EDD and ESD respectively, Figure 2.1). Heart rate was determined by counting the number of peaks in each analysed image.

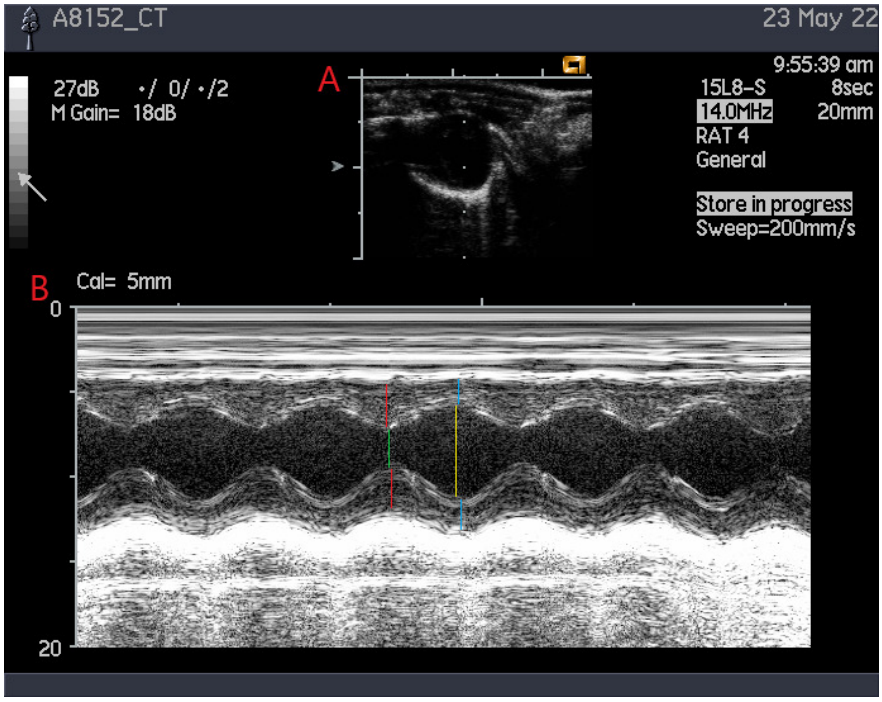



**Figure 2.1: Example of Echocardiograph Image Obtained by Ultrasound.** (A) 2D imaging of the heart along the long-axis plane to visualise the left atrium (LA) and left ventricle (LV). M-Mode imaging measurements are made along the dotted line. (B) Waveforms in M-mode used to measure anterior wall thickness (AWT) and posterior wall thickness (PWT) in systole (red lines) and diastole (blue lines). Finally, measures of end diastolic and systolic diameter (EDD and ESD) are made along yellow and green lines respectively.

The data across three images were averaged and used to calculate an estimate of LVM using a cubed model (Devereux and Reichek, 1977). To calculate other measures of cardiac function, end diastolic and systolic volumes (EDV and ESV respectively) were calculated and used to derive estimates of Stoke Volume (SV), Cardiac Output (CO), Ejection Fraction (EF), Fractional Shortening (FS), and Relative Wall Thickness (RWT). Described measures were derived using the formulae:

$$ASE\ cube = 1.04 \times ((AWTd + EDD + PWTd) \times 10)^3 - (EDD \times 10)^3$$

$$LVM = (0.8 \times ASE\ cube + 0.6)/1000$$

$$FS = (EDD - ESD)/EDD \times 10$$

$$RWT = (AWTd + PWTd) \times 10/(EDD \times 10)$$

$$EDV = EDD^3 \times 1.047$$

$$ESV = ESD^3 \times 1.047$$

$$SV = EDV - ESV$$

$$CO = SV \times bpm$$

$$EF = (SV/EDV) \times 100$$

### 2.1.3 Sacrifice Procedure and Tissue Collection

Method of euthanisation depended on the age of the animal. Prior to cull, all animals were weighed and anaesthetised with 5% isoflurane in oxygen mix by placement in an induction chamber until loss of righting response. Complete anaesthetisation was confirmed by non-response to pressure applied to the tail-tip. Animals were maintained under anaesthesia by attaching a face mask whilst in the supine position. Blunt forceps raised the skin above the diaphragm allowing fur and skin covering the chest to be removed. An incision was made into the peritoneum and cut along either side of the xiphoid process of the sternum. A second incision was made to the diaphragm, and the rib cage was cut to expose the organs of the chest. Care was taken to avoid large arteries surrounding the ribs to minimise blood loss prior to cardiac puncture.

Cardiac puncture was performed by inserting a 23G needle fitted to a 5mL syringe (Sigma-Aldrich, Manchester, UK) into the lumen of the left ventricle via the apex of the heart. Blood collected from the cardiac puncture was split evenly between vacuum-sealed EDTA, and LH-coated collection tubes. Blood samples were stored on ice during the rest of the procedure. Following exsanguination, connective tissue holding the aorta to the posterior wall of the thoracic cavity was cut using the sharp end of curve-tip scissors. The aorta was then held at the base of the spine and cut away to the aortic arch. The aortic section was dried of excess blood and split such that approximately 2/3 was placed directly into a 1.5mL collection tube and snap frozen in liquid nitrogen. The remaining 1/3 was placed in 10% formalin solution. The heart was then removed, cleared of connective tissue and weighed. Any remaining aorta and atrial tissue was removed from the heart and directly snap frozen in liquid nitrogen. The right ventricle was carefully cut away and snap frozen immediately.

The left ventricle was weighed to directly assess left ventricle mass. If possible, the apex was removed and fixed in 10% formalin solution. The remaining LV tissue was divided into two equally sized pieces and snap frozen in liquid nitrogen. Once the heart and aorta had been removed and stored, the abdomen was cut to expose the organs below the diaphragm. The left and right kidneys were removed, cleaned of connective tissue and weighed individually. One half of one kidney was fixed by placement in a bijou tube containing 10% formalin solution. The remaining tissue was divided into quarters and snap frozen in liquid nitrogen. Finally, the brain was removed and divided along the corpus callosum to separate the two hemispheres. Half was placed immediately in liquid nitrogen and half was fixed in 10% formalin solution.

Gestational day 18.5 (GD18.5) and neonate animals were euthanised by decapitation. Following decapitation, the chest was opened to expose the organs of the chest and the heart was removed, cleaned of connective tissue, and weighed. The whole heart was immediately snap frozen in liquid nitrogen without further dissection. The

abdomen was then further opened to collect, clean, and weigh both left and right kidneys. No other organs were collected from GD18.5 and neonate animals, with the exception of a small piece of liver tissue extracted from SHRSP-*Spp1*$^{em1Mcwi}$ litters for use in an assay to confirm genotype.

All snap frozen tissues were stored in -80°C until further use. After 24 hours at room temperature, tissues in 10% formalin solution were cleaned of residual formalin solution, washed with PBS 3x and stored in 70% ethanol at 4°C. Blood samples stored in tubes containing anti-coagulants were centrifuged at 3000x$g$ at 4°C for 15 minutes (Mega Star 3.0R Centrifuge, VWR, Leicestershire, UK). The plasma supernatant was removed by pipetting, placed into a DNase/RNase free eppendorf, and stored in -80°C for future use.

### 2.1.4 Generation of Gestational Day 18.5 Tissue

Adult (12+ weeks) female WKY, WKY.SPGla14a, SHRSP, and SP.WKYGla14a were placed in timed mating cages with males of the same strain for 3 days. Presence of copulation plug was used as evidence of successful mating and determined gestational day 0.5 (GD0.5). Pregnant dams were euthanised on GD18.5 by exsanguination under anaesthesia, followed by removal of the heart. The uterus was removed and individual foetal-placental units were collected. Each foetus was weighed, measured and euthanised by decapitation. Head weight was assessed, and foetal heart and kidneys were extracted, cleaned, and immediately snap frozen in liquid nitrogen.

## 2.2 General H9c2 Cell Culture

All cell experiments were conducted under aseptic technique in Class II biological safety cabinets (MSC-Advantage$^{TM}$, ThermoFisher, Paisley, UK). The primary H9c2(2-1) cell line was obtained from (ATCC CTRL-1446, Virginia, USA) and once expanded, stored in liquid nitrogen until required.

### 2.2.1 Media Preparation and Cell Passage

H9c2 cells were cultured in Dulbeccos Modified Eagle Medium (DMEM)-high glucose with L-glutamine and sodium pyruvate (Gibco, Invitrogen 41966-029), supplemented with 10% heat inactivated fetal bovine serum (HI-FBS) and 1% penicillin-streptomycin. Cells were maintained at 37°C in a humidified incubator, containing 5% CO2, and passaged once they reached 70% confluence. Media was replaced on growing cells every 48-72 hours. All solutions were heated to 37°C in a water bath prior to application onto cells in culture.

Cells were cultured and expanded in T75 (15mL) or T150 (25mL) Corning Flasks

(Corning, Arizona, USA). Once approximately 70% confluent, cells were split and passaged by removing media, washing twice with PBS, and adding 10% trypsin solution (prepared in PBS). Flasks were then returned to the incubator for 5 minutes to allow cells to detach from the flask surface. Following incubation with trypsin, reaction was stopped by adding 3x the reaction volume of full culture media. The resulting solution was collected and centrifuged to form a pellet. Once the supernatant was discarded, 1mL per flask, of full culture medium was added to the pellet and mixed gently to disperse the cells in the media. To new flasks, 500µL of cell mixture was added, splitting cells in a 1:2 ratio (i.e. equivalent to 1 flask into 2 flasks).

### 2.2.2 Plating Cells for Experimental Assays

For use in experiments, cells were passaged as above and once resuspended, 10µL of cell resuspension was pipetted into a hemocytometer (Sigma, Dorset, UK). Cell counting was performed under light microscopy where each quadrant on the hemocytometer was used to count cells, and average count was calculated over the four quadrants. In experiments using 10cm cell culture dishes, H9c2 cells were seeded at 10.0E+4 and left to adhere for 24-hours. In experiments using 6-well plates, H9c2 cells were seeded at 3.0E+4 and left to adhere for 24 hours.

## 2.3 Nucleic Acid Isolation and Polymerase Chain Reaction

Prior to work with nucleic acids, worktops and fume-hoods were cleaned with RNAzap (Sigma-Aldricj, Dorset, UK) and 70% ethanol. To minimise contamination, filter pipette tips were used (Greiner Bio-One, Gloucestershire, UK). All thermal cycling was conducted on PCRmax Alpha Thermal Cycler 4 (Cole-Parmer Instrument Company, St. Neots, UK).

### 2.3.1 DNA

Samples were processed using DNeasy Blood and Tissue Kits (Qiagen, Manchester, UK) according to the best practices recommended by the manufacturers. Buffer AL was warmed prior to use to ensure any precipitates were dissolved. To prevent overloading of the columns, large frozen samples were fragmented using a pestle and mortar pre-cooled by liquid nitrogen, and sections of tissues of ∼25mg were collected. Ear notches were cut to ∼0.5cm. Samples were lysed in a protein kinase K:buffer ATL solution prepared at a ratio of 1:10. Lysis was performed by incubation in protein kinase solution at 56°C for 15-30 minutes until completely lysed. Samples were vortexed every 10 minutes to aid lysis. Once fully lysed, 200µL of buffer AL followed by 200µL of 100% ethanol was added and mixed by vortex. Resulting solution was transferred to a DNeasy mini-spin column and centrifuged at 8000x$g$ for 1 minute. Flow through was discarded and DNA bound to the column membrane was washed by two rounds

of buffer application and centrifugation (Buffer AW1 and AW2 respectively) to remove any contaminating proteins, divalent cations and enzyme inhibitors. The membrane was dried by centrifugation to remove residual ethanol and reduce interference with downstream reactions. The fully dried column was added to a DNase/RNase free 1.5mL eppendorf tube and without touching the membrane, 50µL of nuclease free water was applied directly to the membrane. The membrane was incubated for 1 minute at room temperature and DNA was eluted by centrifugation for 1 minute at 6000x$g$. To maximise DNA yield, the elution step was repeated by reapplication of flow-through to membrane. DNA concentration was quantified using NanoDrop ND-1000 Spectrophotometer (ThermoFisher, Paisley, UK) and resulting DNA was stored at 4°C until use.

### 2.3.2 RNA

RNA was isolated from whole tissue or cell samples using a phenol/guanidine-based lysis solution combined with column filter purification as provided by miRNeasy Mini kits (Qiagen, Manchester, UK). Tissue samples were fragmented in liquid nitrogen and pre-cooled pestle and mortar to obtain fragments <30mg. Whole hearts from 1-3 day neonates were split to allow matched protein and RNA extractions. Tissue fragments were disrupted by the addition of single stainless steel ball (Qiagen, Manchester, UK) to 2mL DNase/RNase free sample collection tube containing sample and 700µL QIAzol lysis reagent. Samples were homogenised using bench-top TissueLyser at 25Hz for 30 seconds, repeated 2-3 times, until tissue was completely lysed. Cells at a concentration of no more than 3.0-4.0E+6 were lysed directly in 700µL QIAzol Lysis Reagent. Cell lysates were collected into DNase/RNase free eppendorf tubes and homogenised by vortex for 1 minute.

Homogenates were incubated at room temperature for 5 minutes to promote dissociation of nucleoprotein complexes. In fume-hood, 140µL of chloroform was added to tube and the solution was mixed by vigorous shaking for 15 seconds. Prior to phase separation, the homogenate was incubated for 3 minutes at room temperature. Phase separation was achieved by centrifugation for 15 minutes at 12,000x$g$ at 4°C. This separated the sample into an upper aqueous phase containing RNA, a white interphase and lower organic phase. The upper aqueous phase was transferred into a new collection tube and mixed with 1.5×100% ethanol. In most cases, ∼525µL ethanol was added to the collected aqueous phase (usually 350µL), but was dependent on the size of tissue fragment or number of lysed cells. Up to 700µL of the resulting solution was transferred to miRNeasy columns and centrifuged at ≥ 8000x$g$ for 15 seconds at room temperature. The flow through was discarded, and any remaining lysis product was applied to the column and subjected to the same method until all the sample had been processed.

To reduce DNA contamination of RNA samples, on-column DNase digest was performed as per manufacturer instructions. First, 350µL buffer RWT was applied to wash the column. Columns were centrifuged ($\geq$ 8000x$g$, 15 seconds) and flow through discarded. A mix containing DNase-I (Buffer RDD:DNase-I 70µL:10µL) was added to the column and tubes were incubated for 15 minutes at room temperature. Buffer RWT (350µL) was then added to wash the column, followed by centrifugation at $\geq$ 8000x$g$ for 15 seconds. Flow through was discarded and 500µL of a second wash buffer (RPE) was applied and columns were centrifuged for 2 minutes. Flow through and collection tube was discarded, and an additional centrifugation ($\geq$ 8000x$g$, 1 minute) in a fresh collection tube was performed to ensure membrane was dried of residual RPE buffer.

Columns were transferred to a final 1.5mL eppendorf tube and 30µL of nuclease free water was pipetted directly onto the membrane. The columns were rested for 1 minute at room temperature before elution of RNA by centrifugation at $\geq$ 8000x$g$ for 1 minute. The flow through was then reapplied to the membrane and re-centrifuged to maximise RNA yield. RNA concentration was quantified using NanoDrop ND-1000 Spectrophotometer (ThermoFisher, Paisley, UK) and was stored at -80°C until use.

### 2.3.3 cDNA Generation

Double-stranded complimentary DNA (cDNA) was prepared from isolated RNA using reverse transcription using the SuperScript$^{TM}$ II enzyme (ThermoFisher, Paisley, UK). Non-specific primers in the form of random hexamers (short oligodeoxyribonucleotides of random sequence) were used to amplify all RNA forms. Reactions were prepared to a 20µL volume on a 96-well plate containing 1ng-5µg of total RNA as detailed in Table 2.1. Random hexamers, dNTPs, RNA and nunclease free water were first added to wells and the plate was heated to 65°C for 5 minutes. Following this, 6µL of first-strand buffer and DTT (dithiothreitol) were added to solution and the plate was incubated at room temperature for 2 minutes. SuperScriptII RT enzyme was added to each well and the plate was then placed on a thermocycler. Samples were heated to 25°C for 10 minutes to anneal primers, then temparature was increased to 42°C for 50 minuts to allow transcription and extension of RNA into cDNA. Finally the reaction was inactivated by heating to 70°C for 15 minutes. The resulting cDNA product was collected and stored at -20°C until further use.

**Table 2.1:** cDNA Reverse Transcription Reaction Reagents

| Reagent | Volume |
|---|---|
| Template RNA | XµL containing 1ng-5µg RNA |
| dNTP (10mM) | 1µL |
| Random Hexamers (50µM) | 1µL |
| 5x first strand buffer | 4µL |
| DTT (0.1M) | 2µL |
| SuperScript II RT | 1µL |
| Nuclease free water | YµL added to 20µL total volume |

### 2.3.4 End-Point Polymerase Chain Reaction

End-point PCR was carried out on either genomic-DNA (Section 2.3.1) from transgenic animals to confirm genotype of litters from heterozygous mating, or cDNA generated from cardiac tissue RNA (Section 2.3.2), as part of alternative transcription assays.

Depending on the specific experiment, exact amount of cDNA/genomic DNA (gDNA) added to PCR reaction varied according to requirements. HotStarTaq DNA Polymerase (Qiagen, Manchester, UK) was used with specifically designed primers in PCR reactions and amplification of regions of interest was achieved by PCR using 96-well plates (Starlab, Milton Keynes, UK). Reactions were prepared as per Table 2.2 with nuclease free water used to bring reaction volume to a total of 20µL. Thermal cycle conditions are detailed in Table 2.3

**Table 2.2:** End Point Polymerase Chain Reaction Reagents

| Reagent | Volume |
|---|---|
| 10x buffer | 2.5µL |
| dNTP (10mM) | 0.5µL |
| Forward Oligonucleotide Primers (10µM) | 1µL |
| Reverse Oligonucleotide Primers (10µM) | 1µL |
| Taq Polymerase | 0.25µL |
| gDNA/cDNA | 1ng-5µg |

**Table 2.3:** HotStarTaq Polymerase Chain Reaction Thermal Cycling Conditions

| Temperature | Time |
|---|---|
| **Enzyme Activation** | |
| 95 °C | 15 min |
| **3-step cycles (n=35)** | |
| 94 °C | 30 seconds |
| 50-68 °C | 30 seconds |
| 72 °C | 1 min |
| **Final Extension** | |
| 72 °C | 10 min |
| 12 °C | Hold |

### 2.3.5 Agarose Gel Electrophoresis

To visualise products from end-point PCR, 1–1.5% agarose gels were prepared by adding UltraPure Agarose (Invitrogen, Carlsbad, USA) to 100–150mL Tris-borate EDTA (TBE) buffer and heating in a microwave to fully dissolve agarose in TBE buffer. Upon cooling 1.5–3µL of GelRed Nucleic Acid Stain (ThermoFisher, Paisley, UK) was added and mixed. Mixture was poured into a gel template with a 10, 15, or 20-well comb and left to cool and solidify for at least 15 minutes.

The gel was transferred into a BioRad electrophoresis tank (BioRad, Hampstead, UK) filled with TBE buffer and comb was carefully removed. To prepare samples for loading, 8µL PCR product was mixed with 2µL 6x loading dye (New England Biolabs, Ipswich, USA) and 10µL of solution was added to each well. A 100bp DNA ladder (New England Biolabs, Ipswich, USA) was added to lane 1 and electrophoresis of samples was performed at 90V for 1.5 hours, using a BioRad Powerpac-300 direct current system. Gels were visualised using a BioRad Molecular Imager XRS+ transillumination system (BioRad, Hampstead, UK) and images were saved for qualitative analysis. Band intensity was adjusted to ensure the best representation of products and ladder.

### 2.3.6 Quantitative Real Time PCR

Quantitative real time PCR (qRT-PCR) was performed using Applied Biosystems TaqMan$^{TM}$ Assays. TaqMan$^{TM}$ qRT-PCR probes are single DNA oligos consisting of a 5' reporter dye and a 3' quencher which produce a fluorescent signal when bound to DNA. The amount of fluorescence detected is linearly proportional to the amount of cDNA in the original sample.

TaqMan probes were applied in a duplex assay where the gene of interest and housekeeper probes were simultaneously quantified in the same reaction. FAM and VIC reporter-dyes which fluoresce at different wavelengths, were obtained for a suitable house-keeping and target gene respectively. FAM-conjugated target gene and VIC-conjugated house-keeping gene probes were added to the TaqMan MasterMix (ThermoFisher, Paisley, UK) solution to prepare a reaction master mix for N samples ±10% as described in Table 2.4. On a 384-well plate, 2µL of cDNA was mixed with 3µL master mix. Reactions were performed in triplicate and contained a no-RT and no-cDNA/water negative control. To generate a no-RT control, a cDNA reaction without the addition of SuperScript II enzyme was performed during cDNA preparation such that no reverse transcription of RNA would have occurred.

**Table 2.4:** Master Mix for TaqMan qRT-PCR

| **Reagent** | **Volume** |
|---|---|
| TaqMan Master Mix (no UNG) | 2.5µL |
| FAM-flurophore qRT-PCR probe | 0.5µL |
| VIC-flurophore qRT-PCR probe | 0.5µL |

Plates were amplified using a QuantStudio 12K Flex Real-Time PCR System (ThermoFisher, Paisley, UK) and reactions were quantified using the double-delta-Ct method (ddCt) for relative quantification (RQ). Each biological replicate was performed in triplicate such that missing Ct values could be removed if it was apparent there was a user error which would affect downstream calculations. Triplicates were summarised as mean Ct values for each housekeeper and target gene. The mean Ct value of the housekeeping gene was subtracted from the mean Ct value of the gene of interest to produce a ΔCt value per biological replicate. Statistical evaluation was carried out on ΔCt values. Group mean ΔCt values were normalised to the mean ΔCt of the control group giving ΔΔCt. The ΔΔCt was converted into fold change gene expression (RQ) by calculating the second exponent of the negative ΔΔ Ct. RQ min and max was calculated as the second exponent of ΔΔCt ± group ΔCt standard error.

$$\Delta Ct = Ct(target) - Ct(housekeeper)$$
$$\Delta\Delta Ct = \Delta Ct - \Delta Ct(control)$$
$$RQ = 2^{-\Delta\Delta Ct}$$
$$RQ\ min = 2^{-\Delta\Delta Ct - SE\Delta Ct}$$
$$RQ\ max = 2^{-\Delta\Delta Ct + SE\Delta Ct}$$

For visualisation of results, RQ (fold change) per group was plotted with calculated

RQ min and RQ max as error. Statistically significant comparisons of ΔCt values were then added to the plot area, above the relevant groups. Probes used in TaqMan experiments are listed in Table 2.5.

**Table 2.5:** TaqMan qRT-PCR Probes

| **Probe Target** | **Probe ID** | **Exon Boundary/Assay Location** |
|---|---|---|
| *Spp1* | Rn00681031_m1 FAM-MGB | 5–6/327 (NM_012881.2) |
| *B2m* | Rn00560865_m1 VIC-MGB | 1–2/86 (NM_012512.2) |

## 2.4 Plasmid Generation

Plasmid generation was performed from glycerol stocks that were generated previously and stored in -80°C. New glycerol stocks of pTarget vectors containing alternative *Spp1* transcripts were created and stored as glycerol stocks for long term use.

### 2.4.1 LB Medium and Agar Plates

All transformed and stock bacteria were cultured using sterile Luria Bertani (LB) medium supplemented with 100µg/mL ampicillin as per Table 2.6. Agar plates for growth of transformed competent cells and single colonies were prepared from sterile LB-Agar, supplemented with ampicillin, no less than 24-hours prior to use (Table 2.6). LB-Agar was poured into 10-cm dishes, to fill approximately 1/3 of the total volume. Once solidified, 40µL of X-Gal (5-bromo-4-chloro-3-indolyl β-D-galactopyranoside) was applied to the top of the Agar and spread over the surface. Plates were incubated at 37°C to dry and wiped clean of condensation.

**Table 2.6:** Media and Stock Recipes

| **Component** | **Mass or Volume / 1L** |
|---|---|
| **Luria Bertani Medium** | |
| Miller's LB Broth Base | 25g |
| Ampicillin | 100mg (100µg/mL) |
| $ddH_2O$ | 1L |
| **LB-Agar Plate** | |
| LB medium | 1L |
| Agar Power | 1g |
| Ampicillin | 100mg (100µg/mL) |

### 2.4.2 Plasmid Production and Transformation into Competent Cells

New plasmids containing alternative transcripts of *Spp1* were generated using the pTarget expression system. Full length *Spp1* RNA was amplified and reverse transcribed into cDNA (Section 2.3.3) and products resolved by end-point PCR (Section 2.3.4). PCR products were cloned into the pTarget Mammalian Expression Vector System (Promega, Wisconsin, USA) at a 3:1 ratio (vector:insert). Ligation reactions in 1.5mL eppendorfs were prepared (Table 2.7) and incubated overnight at 4°C to facilitate ligation of products into the vector.

**Table 2.7:** pTarget Ligation Reaction Reagents

| **Reagent** | **Volume µL** |
|---|---|
| T4 DNA Ligase | 1 |
| pTarget Vector | 1 |
| PCR product or control insert | X |
| Nuclease free water | to 10µL |

All transformations were performed using aseptic technique to remove potential contaminants. LB-Agar plates, coated with X-Gal, were allowed to equilibrate to room temperature. JM109 High Efficiency Competent Cells were removed from storage in -80°C and thawed on ice. Once thawed, 50µL of JM109 cells were combined with 2µL of ligation product in a 1.5mL eppendorf which was gently flicked to disperse the mixture. The eppendorf was incubated on ice for 20 minutes before the cells were subjected to heat shock for 50-seconds in a 42°C water bath. Cells were then immediately returned to ice for 2 minutes. Finally, 950µL of LB media was added to each reaction and this mixture was incubated for 1.5 hours at 37°C with shaking ( 150RPM) before 100µL of cells were spread onto LB-Agar plates and incubated at 37°C overnight.

Colonies were selected by blue/white screening where white colonies indicate recombinant colonies and blue indicate unsuccessful recombination. Successful colonies were picked with a sterile pipette tip and spread onto fresh LB-Agar plates. A small amount of the colony was left on the tip and was added to a PCR amplification reaction prepared using a T7-promoter forward primer and *Spp1* reverse primer to confirm forward orientation of the plasmid within the vector.

Single colony forming units were used to prepare an overnight starter culture to generate glycerol stocks. Chosen single colonies were selected with a sterile pipette tip which was placed in 10mL LB-media and incubated overnight at 37°C with shaking. The following day, 1mL of starter culture was added to 1mL 50% glycerol stock and stored at -80°C for future use.

### 2.4.3 DNA Plasmid Purification

Starter cultures were expanded from glycerol stocks by collection of a small amount of frozen glycerol stock onto a sterile pipette tip, adding tip to 10mL LB-media, and incubating overnight at 37°C with shaking. These starter cultures were then expanded again by adding 1mL of starter culture into 500mL LB-broth (supplemented with ampicillin) in a large canonical flask. The colony was expanded overnight with gentle shaking at 37°C. Purification of plasmids from overnight colonies was performed using Qiagen Plasmid Maxi Kit (Qiagen, Manchester, UK) as per manufacturers instructions. Overnight culture was purified by centrifugation at 6000x$g$ (Beckman Coulter, Avanti J26XP) for 15 minutes at 4°C. All following centrifugations were also performed at 4°C. The supernatant was discarded, and the pelleted bacterial cells were resuspended in 10mL of chilled buffer P1, and moved into sterilised polypropylene centrifuge tubes. To this, 10mL of P2 buffer was added, the solution was vigorously shaken by hand and incubated at room temperature for 5 minutes. To neutralise the lysate, 10mL of pre-chilled buffer P3 was added and the solution was mixed until it became colourless. Tubes were incubated on ice for 20 minutes and then returned to the centrifuge for 30 minutes at, 20,000x$g$, to collect the precipitates. The supernatant containing the plasmid DNA was then collected and retained.

During centrifugation steps above, QIAGEN-tip 500 was equilibrated by applying 10mL of buffer QBT to the column, which perfused into the tip and emptied by gravity flow. Following complete flow through of Buffer QBT, the supernatant was added to the column and allowed to empty by gravity flow. The column tip was washed by applying 30mL buffer QC twice. Following flow through of the wash buffer, collection tube was changed, and plasmid DNA was eluted from the column by applying 15mL of buffer QF. Care was taken throughout to ensure the tip did not fully dry between applications of buffers or supernatant, whilst ensuring all the previous buffer or supernatant had fully passed through.

Plasmid DNA was precipitated from the collected solution by the addition of 10.5mL molecular grade isopropanol at room temperature and subjecting the solution to centrifugation at 15,000x$g$ for 15 minutes. The supernatant was carefully discarded and pellet was washed with 5mL 70% ethanol. The DNA and alcohol solution was centrifuged at 15,000x$g$ for 15 minutes and supernatant was carefully decanted, taking care not to disrupt the DNA pellet. Pellets were air-dried for 10–20 minutes before re-suspending DNA plasmid in 450µL of nuclease free water. The concentration and yield of the resuspended DNA solution was determined by Nanodrop Spectrophotometer.

## 2.5 Protein Extraction

Protein was extracted from whole tissue using T-PER (Pierce) in a 1:20 weight to volume ratio supplemented with 1x protease and phosphatase inhibitor cocktail (PPC1010, Sigma-Aldrich, St Louis, USA). Approximately 30mg of tissue was added to chilled T-PER:Protease/phosphatase inhibitor solution, along with a 5mm stainless steel bead (Qiagen, Manchester, UK) and disrupted using 4× 30s pulses at 30kHz (TissueLyser, Qiagen, Manchester, UK). Tissue was further homogenised using freeze-thaw cycles. Simply, samples were placed on dry ice until completely frozen and then moved to a hot plate at 37°C until thawed. This cycle was repeated 3 times to ensure complete lysis of proteins. Between cycles samples were briefly vortexed. Samples were centrifuged in a bench-top centrifuge for 5min at 10,000x$g$ and protein containing supernatant was collected. A 10x dilution of protein sample was prepared for use in BCA assay (Pierce) and the remaining sample was stored in -80°C or -20°C until further use.

To isolate protein from cell culture, media was removed and cells washed with PBS 3 times. Pre-chilled RIPA buffer (ThermoFisher, Paisley, UK), supplemented with 1x protease and phosphatase inhibitor cocktail, was applied directly onto cells in a volume appropriate for the culture dish in use. Dishes were incubated in lysis solution, on ice, for 5 minutes. Lysis was further aided by scratching the dish was a sterile cell scraper. The supernatant was collected and stored at -20°C for future use.

Protein concentration was determined using Pierce$^{TM}$ BCA Protein Assay Kit (ThermoFisher, Paisley, USA) according to the manufacturers' protocol. Protein concentration was estimated from a standard concentration curve fitted to a 4-parameter logistic regression (4PL) model. Bovine Serum Albumin (BSA) standards were prepared by diluting 2000µg/mL BSA standards in working solution (50:1 ratio reagent A:reagent B). Serial dilutions were prepared to concentrations of 2000, 1500, 1000, 750, 500, 250, 125, 25, and 0µg/mL, and labelled A-F respectively. In a flat-bottom 96-well plate, 200µL of working solution was added each reaction well. Plates were prepared such that all samples and standards were repeated in triplicate and included a blank control (no standard or sample added to working solution). Equal volume of sample and standard was then added to each well and plate was incubated at 37°C for 30 minutes. The plate was analysed using the 480nm wavelength filter in a PerkinElmer VICTOR plate reader (PerkinElmer, USA). Raw absorbance values were collected from the plate reader and analysed using R.

From all wells, absorbance was corrected to the blank control. Mean values for each sample and standard were calculated by generating an average of the triplicate wells. The absorbance values for the BSA standards (0–2000µg/mL) were then fitted to a 4PL model (as shown in Equation 2.1), from which unknown sample concentrations were calculated (Figure 2.2).

$$absorbance = raw\ sample\ absorbance - blank\ absorbance$$

$$y = d + (a - d)/1 + (x/c)^{\mathrm{b}} \tag{2.1}$$

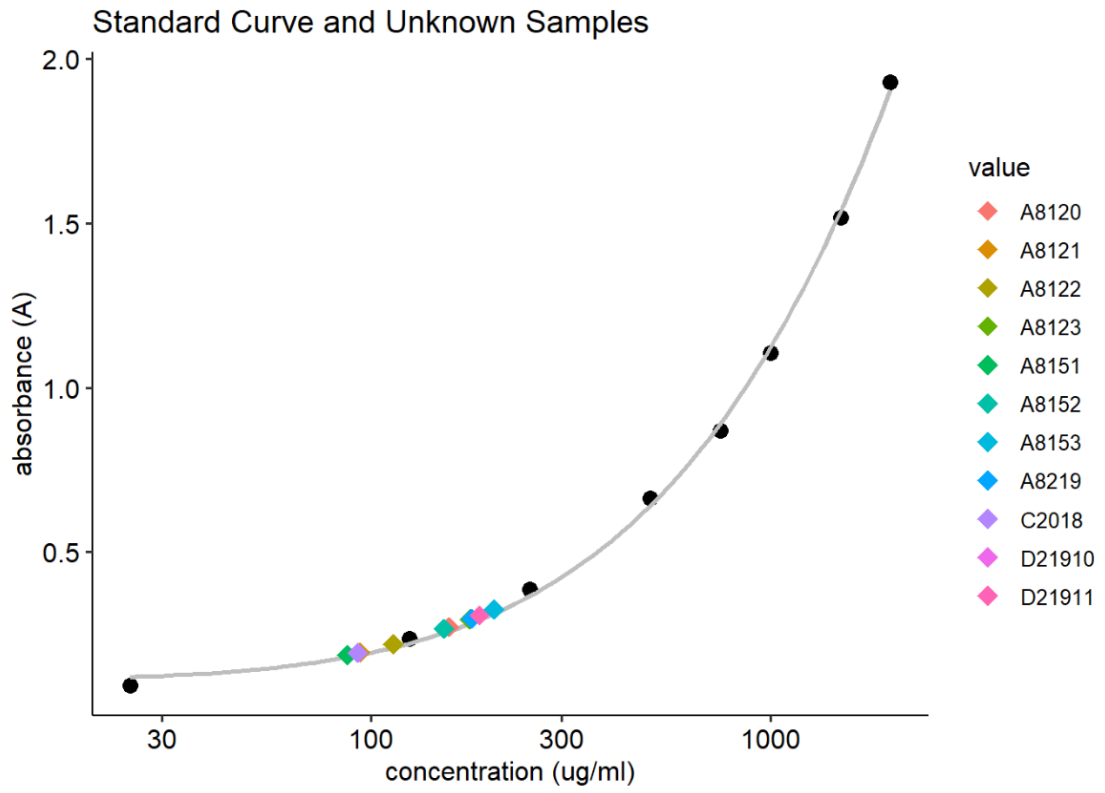



**Figure 2.2: BCA Protein Assay Standard Curve** An example of fitting unknown samples to standard curve modelled with 4-parameter logistic regression (4PL).

## 2.6 SDS-PAGE and Immunoblotting

Protein samples isolated from cells and tissues were diluted to 0.5µg/µL. Samples were mixed with 4x SDS buffer containing reducing agent and boiled at 95°C for 5 minutes. Lysed samples were loaded in 10µL volume to achieve a total of 5µg protein per lane. Proteins were separated by gel electrophoresis (4–12% gradient gels, ThermoFisher Scientific, UK) using Bolt™ MES running buffer (1x solution). To each gel, 3µL of PageRuler Pre-stained Protein Ladder (ThermoFisher, Paisley, UK) was added to the leftmost well. Proteins were electrophoresed using a BioRad power-pac 300 direct-current system at 200V for 25–35 minutes, or until the dye had migrated to the bottom of the gel.

Prior to transfer, 1x transfer buffer was prepared by combining 850mL MilliQ water with 50mL of 20x Bolt™ transfer buffer and 100mL of 100% methanol. This solution was pre-cooled and used to soak components of the transfer; nitrocellulose membrane,

filter paper, and sponges utilised during transfer. Components were soaked for approximately 10 minutes. Once sample and ladder had migrated sufficiently along the gel, the gel was excised from its case and paired with pre-soaked, activated nitrocellulose membrane, in a Mini Gel tank module. Elements were layered as followed; 3x pre-soaked sponges, 2x pre-soaked filter papers, gel, nitrocellulose membrane, 2x pre-soaked filter papers, 3x pre-soaked sponges. The module was submerged in transfer buffer and transfer of products from the gel onto the membrane was performed in a cold room at 4°C by applying a current for 90 minutes at 100V.

Normalisation of target protein was achieved using Revert 700 Total Protein Kit (LI-COR, Nebraska, USA). Membranes were dried, rehydrated in MilliQ water and then submerged in REVERT$^{TM}$ total protein stain for 5 minutes with gentle rocking. Membranes were washed 2x30 seconds with wash solution (LI-COR, 6.7% (v/v) glacial acetic acid, 30% (v/v) methanol, in water) and imaged immediately using the 700nm channel (Odessey CLx, LI-COR).

The membranes were blocked in TBS blocking buffer (Intercept (TBS) LI-COR, 927-60000) for 1 hour at room temperature with gentle rocking. Primary antibodies were diluted in blocking buffer and applied to membranes overnight at 4°C. Following overnight incubation with primary antibody solution, membranes were washed with TBS-T (50mM Tris-HCL at pH=7.4, 150mM NaCl and 0.1% Tween-20) and incubated in donkey anti-rabbit IgG secondary antibody solution (LI-COR IRDye 800CW, 1:20000) for 1 hour at room temperature. Membranes were visualised using 800nm channel (Odessey CLx, LI-COR) and band intensities quantified by ImageStudio software. As per the manufacturers' protocol, relative quantification of target protein was achieved by normalising target to a given lane normalisation factor (LNF), calculated from total protein signal.

## 2.7 Histology and Immunohistochemistry

All tissue processing and some sectioning was carried out by technical support staff at the Glasgow Cardiovascular Research Centre and Queen Elizabeth University Hospital (Glasgow Tissue Research Facility). All downstream histology was performed by the author. Chromogenic and immunofluorescent immunohistochemistry was carried out in LV apex tissues collected from adult WKY, SHRSP and chromosome 14 congenic strains. Apical tissue was orientated such that sectioning began at the base of the apex.

### 2.7.1 Tissue Processing and Sectioning

Tissues fixed in 10% formalin were dehydrated and placed into paraffin wax at 60°C (Excelsior AS and Histostar Embedding Workstation respectively, Epredia, Michigan,

USA). Tissue processing was performed by Ms. Nicola Britton and individual steps of tissue processing are detailed in Table 2.8. In brief, tissues were exposed to a timed sequence of alcohol and xylene before being embedded in warmed paraffin wax (≤60°C) and cooled. Once embedded, blocks were kept at room temperature. Prior to sectioning, tissue blocks were placed in -20°C. Sections were cut using Leica RM2235 Microtome (Leica Biosystems, Milton Keynes, UK), at a thickness of 4µm, and placed in a 38°C water bath. Once the wax had softened, 2-3 sections were transferred onto 76×26 1.0mm silane treated microscope slide (Clarity, Smith Scientific, Kent, UK). Slides were baked dry by placement in a bench-top oven at 50°C overnight.

**Table 2.8:** Processing Conditions for Formalin Fixed Tissues

| **Solution** | **Incubation time (mins)** |
|---|---|
| 70% Ethanol | 15 |
| 85% Ethanol | 15 |
| 90% Ethanol | 25 |
| 95% Ethanol | 25 |
| 100% Ethanol | 15 |
| 100% Ethanol | 15 |
| 100% Ethanol | 15 |
| Xylene | 30 |
| Xylene | 30 |
| Paraffin Wax | 30 |
| Paraffin Wax | 30 |

### 2.7.2 Picrosirius Red Solution

Prior to staining, sections were dewaxed and rehydrated at room temperature by serial placement of slides into Histo-Clear (Fisher Scientific, UK) twice, followed by an ethanol gradient (100%, 90%, 70%) and finally water. Slides were then transferred and submerged in Weigarts haematoxylin (Sigma, Dorset, UK) for 10 minutes. The slides were transferred into a fresh container and washed in running water tap water for 5–10 minutes until the water ran clear. A 0.1% picrosirius red solution was prepared using 0.1% w/v Sirius Red F3B (Sigma Aldrich, Dorset, UK) in dd$H_2O$. Slides were incubated in 0.1% picrosirius red solution for 90 minutes in a light protected/dark box. Excess stain solution was removed by transferring slides into a fresh container of acidified water (0.01N (0.1 mEq/mL) hydrochloric acid) for 5 minutes, twice. Excess moisture was removed by blotting with paper towel and sections were dehydrated and cleared for

mounting. Slides were dehydrated following staining by placing them sequentially into 70%, 90% and 100% ethanol for 5 minutes. Final clearing step was performed by placing slides into Histo-Clear for 5 minutes, twice (Table 2.9). A coverslip was mounted over the sections using DPX mounting medium applied with a pasteur pipette.

**Table 2.9:** Summary of Steps Taken During Staining

| Stage | Solution | Incubation time (mins) |
|---|---|---|
| 1. Deparafinisation | Histo-clear | 2x5 |
| 2. Rehydration | 100% Ethanol | 5 |
| | 90% Ethanol | 5 |
| | 70% Ethanol | 5 |
| 3. Staining | **Weigarts haematoxylin, picrosirius red** | |
| 4. Dehydration | 70% Ethanol | 5 |
| | 90% Ethanol | 5 |
| | 100% Ethanol | 5 |
| 5. Deparafinisation | Histo-clear | 2x5 |

### 2.7.3 Immunofluorescent Immunohistochemistry

Immunohistochemistry (IHC) was performed on LV sections for osteopontin alone (single immunofluorescence), or in combination with the cell type markers α-SMA, vimentin and myosin heavy chain (MYH) in a multiplexed analysis.

Antibody sensitivity and specificity, as well as optimal concentration, was first evaluated using chromogenic IHC. Dewaxing and antigen retrieval was carried out at pH 6 and pH 9 and antibodies were titrated at concentrations above, below and at the manufacturers recommended dilution. Appropriate negative controls were employed alongside positive control tissues described in manufacturer product data sheets (tonsil, heart or liver). DAB (3,3' Diaminobenzidine) chromogen staining was assessed visually using a light microscope to determine optimal concentrations.

To transition to immunofluorescence (IF), each optimal concentration was then diluted further and compared to chromogenic 'optimal' to determine concentration appropriate for immunofluorescent IHC. Immunofluorescent stains use flurophores, which absorb photons at a certain wavelength of light and emit them at a higher wavelength. Flurophores can be directly bound to primary antibodies, or can be added to any primary antibody as a secondary reagent, following an amplification step usually using horseradish peroxidase (HRP). OPAL secondary antibodies (Akoya Biosciences, Marlborough, USA) utilise tyramide signal amplification (TSA) chemistry to label pri-

mary antibodies with fluorescent markers. TSA uses the HRP conjugated to primary antibodies to catalyse the activation of dye tyramides, resulting in the binding of tyramides to nearby tyrosine residues. Once labelled in this way, primary antibodies and secondary-HRP antibodies are removed by microwave treatment whilst the fluorescent signal is retained (dye bound to tyrosine residues). This allows subsequent targets to be detected without antibody cross-reactivity, allowing multiple antibodies raised in the same species to used one after the other applying different coloured dye tyramides in each round of staining. Both TSA and DAB oxidation are peroxidase-mediated reactions, as such, primary antibody conditions could first be determined using DAB detection and then directly applied to the fluorescent assays, using Opal reagents.

After determination of antibody concentration, single or multiplex experiments were undertaken as per the pipeline depicted in Figure 2.3. Single immunofluorescent protocol followed the same workflow without requiring additional microwave and repeated rounds of primary/secondary antibody staining. Following sectioning, slides were placed in Dewax and Heat Induced Epitope Retrieval (HIER) Buffer (Expredia$^{TM}$, ThermoFisher, Paisley, UK) to simultaneously perform deparaffinisation, rehydration and epitope retrieval. Slides were placed in HIER solution and brought to boil using a domestic microwave at full power. Power was reduced to ∼50% and maintained at boiling (∼98°C) for 20 minutes. Solution was then left to cool for 10-15 minutes and slides were rinsed in cold water.

Antigen retrieval was automated using a Pre-Treatment (PT) Module$^{TM}$ (ThermoFisher, Paisley, UK). Tanks were filled with 1.5L of HIER buffer and pre-heated to 60°C, slides were then added into the solution and module was heated to 98°C for 20 minutes. Module was then cooled to 60°C. Once cooled, slides were removed and washed with water. The PT module automated treatment for up to 24 slides per tank.

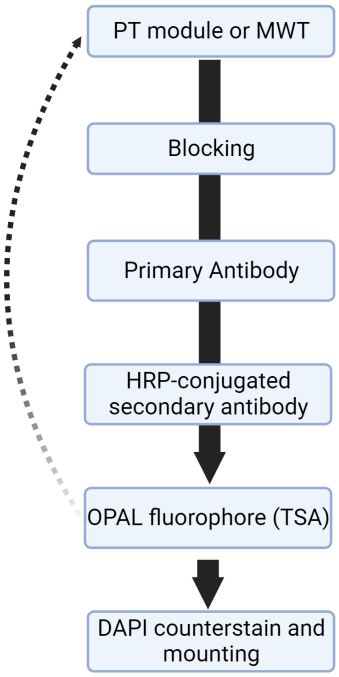



**Figure 2.3: Multiplex Immunofluorescence Workflow** Simplified workflow utilised in multiplex IHC experiments using the OPAL TSA system. Slides undergo deparaffinisation, rehydration and epitope retrieval in Dewax and Heat Induced Epitope Retrieval (HIER) Buffer by heating in PT module or domestic microwave. Slides are then serially blocked for endogenous peroxidase and non-specific antibody staining before incubation with primary antibody directed to protein target. Following incubation with primary antibody, amplification buffer, and HRP-conjugated secondary antibody are applied sequentially with washing steps in between. Secondary OPAL reagent with coloured dye tyramides containing selected fluorophore is added. Slides are then washed and cycle is repeated until all primary antibody and OPAL fluorophore pairs have been added. Slides are then counter-stained with the nuclear fluorophore 4',6-diamidino-2-phenylindole (DAPI) and stored in water. Coverslip was applied to slides using antifade mountant (ProLong$^{TM}$ Diamond, Invitrogen, Paisley, UK).

Following cooling, slides were processed using the UltraVision$^{TM}$ Quanto Detection System (Expredia$^{TM}$). Slides were washed with milliQ water and endogenous peroxidase activity was quenched by incubation with Hydrogen Peroxide block for 10 minutes. TSB-T buffer was used to wash slides before application of Ultra V Block (Expredia$^{TM}$) was applied for 5 minutes, to block non-specific background antibody staining. Slides were washed again before primary antibody was added directly to the tissue and incubated at room temperature for 30 minutes. Primary antibody was collected and slides were washed with TBS-T buffer. Primary antibody incubation was followed by an amplification step using Primary Antibody Amplifier Quanto, which was applied to the slides for 10 minutes and washed using TBS-T buffer. Slides were then incubated in HRP-polymer Quanto for 10 minutes and then washed with TBS-T followed by water. Blocked, stained and amplified slides were then incubated in OPAL fluorophore-TSA solution (1:200 OPAL dilution buffer) corresponding to the specific primary-fluorophore pair. OPAL TSA solution was applied for 10 minutes, collected and slides were washed with TBS-T. Figure 2.3 details the major steps described. Following incubation with OPAL TSA flurophores, slides were recycled into the start of

the process and primary antibody was removed using microwave treatment. If slides were being processed for single IF or, if the final antibody-fluorophore pair had been applied, slides were washed and then incubated in DAPI counter-stain for 10 minutes. Following counter-stain, slides were washed with TBS-T and then placed in water. Slides were mounted and coverslip applied using 50µL antifade mountant (ProLong™ Diamond, Invitrogen, Paisley, UK). Slides were dried and stored at -4°C until imaging.

### 2.7.4 Image Acquisition and Digital Image Analysis

All sections of chromogenic stains (picrosirius red and DAB) were digitized by scanning at 20x using the Hamamatsu Nanozoomer 2.0-HT slide scanner (Hamamatsu Photonics, UK), and images were viewed via the NDP.view 2.0 viewing software (Hamamatsu Photonics, UK). Immunofluorescent-stained slides were scanned by NanoZoomer S60 digital slide scanner (Hamamatsu Photonics, UK) with 20x magnification. Image acquisition was performed by Dr. Hannah Morgan at the Glasgow Tissue Research Facility.

Picrosirius red, single IF and multiplexed IF images were analysed using the open-source software QuPath (version 0.4.2). QuPath employs groovy scripting as well as a comprehensive GUI for image analysis. It contains built-in algorithms for tissue detection and utilises machine learning for object and pixel classification. Specialised analysis pipelines were developed for chromogenic and immunofluorescence-based images. Outputs were exported analysed in R.

## 2.8 Data Handling and Statistical Methods

Data was analysed R.4.3 (or equivalent) using tidyverse (Wickham et al., 2019) and rstatix (Kassambara, 2023) libraries. Appropriate statistical methods were determined by number of experimental groups or conditions, where 3+ groups were analysed using one-way ANOVA or non-parametric equivalent followed by Tukey *post-hoc* testing. Comparisons between 2 groups were performed using t-tests. For all data, a $p$-value of $\leq 0.05$ was considered statistically significant. Adjustments for multiple testing were applied where required and specific statistical methods are detailed within relevant chapters. Source codes developed for analyses are publicly available at `https://github.com/c-triv97/Source-Codes`.

# Chapter 3

# Cardiac Structure and Function in Chromosome 14 Congenic Rats

## 3.1 Introduction

Early studies suggested in human populations of hypertensives, the prevalence of LVH only differed significantly from a non-hypertensive population in older age groups (Tingleff et al., 1996). Estimates of LVH prevalence can vary considerably, with some reports suggesting prevalence of LVH is almost doubled in a hypertensive population versus normotensive populations (X. Wang, Hao, et al., 2022). Despite a potentially greater risk of LVH in hypertensive patients, population-based longitudinal study predict age-associated increase in LVMI will become classified as LVH in around 1/4 of adults (Bombelli et al., 2023). Given the poorer outcomes experienced in patients with LVH, determining pressure independent mechanisms which contribute to increased risk of elevated LVMI is therefore of importance. A relatively large proportion of LVMI is thought to be under genetic control and using an F2 hybrid derived from SHRSP/Gla and WKY/Gla strains, a QTL for LVMI was identified on rat chromosome 14 (J.S. Clark et al., 1996).

Using a speed congenic protocol, a chromosome 14 congenic strain on both the WKY and SHRSP backgrounds was established (Jeffs et al., 2000). Outlined in Section 1.5, initial phenotyping experiments of male WKY.SPGla14a and SP.WKYGla14a animals demonstrated significant differences in LVMI versus parental strains from as early as 5 weeks. Unbiased transcriptional analyses using microarray identified significant upregulation of secreted phosphoprotein 1 (*Spp1*) in the strains which developed increased LVMI (WKY.SPGla14a and SHRSP) versus strains which did not (SP.WKYGla14a and WKY). However, whilst cardiac phenotypes mirrored chromosome 14 donor strains, blood pressure profiles reflected parental strains (Figure 3.1).

As variance explained by the loci was higher in males (32.3%, LOD score 3.7) than females (12.3%, LOD score 3.1) J.S. Clark et al. (1996), initial phenotyping focused only on male animals. That being said, the peak LOD score for was $\geq 3.0$ in both sexes and thus could be considered a 'suggestive' linkage (J.S. Clark et al., 1996) affecting both males and females in the developed chromosome 14 congenic strains.

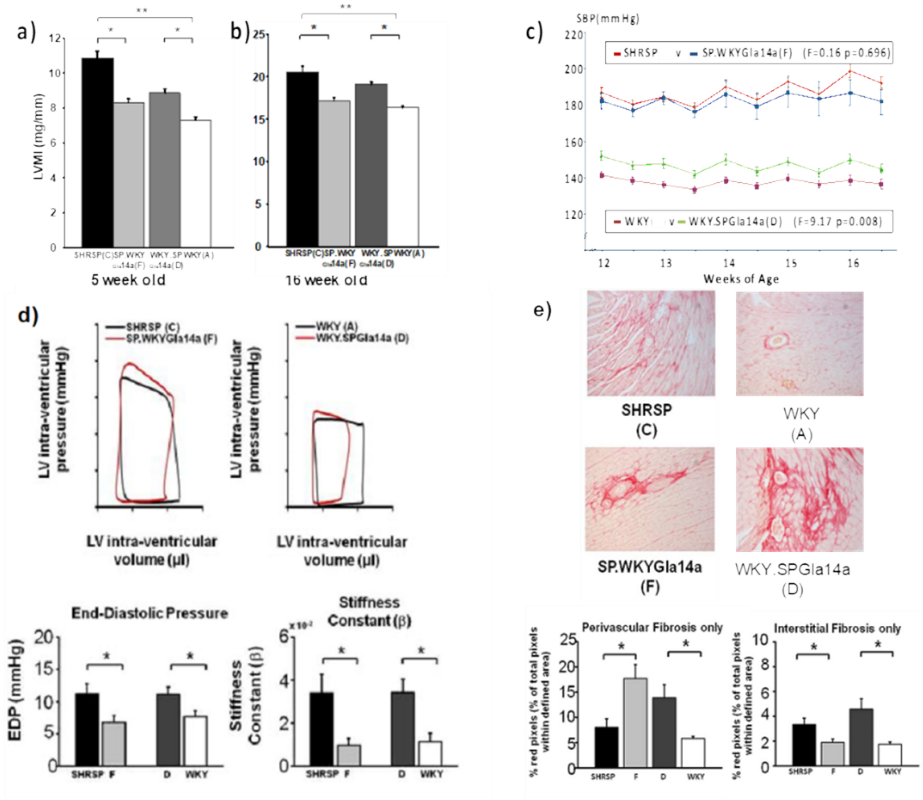



**Figure 3.1: Phenotype of Male Chromosome 14 Congenic Strains** Characterisation and functional analyses of WKY, SHRSP, WKY.SPGla14a and SP.WKYGla14a congenic strains. LVMI index at (a) 5- and (b) 16-weeks old. (c) Radiotelemetry SBP recording from 12- to 16-weeks of age. (d) Pressure-Volume (PV) loops and mean diastolic parameters assessed using PV catheters and (e) picrosirus read staining highlighting increased cardiac fibrosis (unpublished data, courtesy of Dr Martin McBride and Dr Delyth Graham.)

Increased expression of *Spp1* was first implicated by microarray across WKY, SHRSP, and chromosome 14 congenic strains in 1 day neonate hearts. This finding was further validated by qRT-PCR in males, as well as isolated primary cardiomyocytes from neonatal pups. *Spp1* encodes the osteopontin (OPN) protein, which has been frequently associated with polygenic and monogenic cardiovascular and related diseases (K. Singh, Sirokman, et al., 1999; Rysa et al., 2005; Brooks et al., 2010; Yim, H. Cho, and Rabkin, 2018; Y. Yang, Y. Wang, and P.J. Gao, 2020). Comparison of failing and non-failing hearts from SHR and WKY rats showed increased OPN protein expression in failing hearts when failure was surgically induced by aortic banding (K. Singh, Sirokman, et al., 1999). In a hypertensive Chinese cohort, increased OPN protein was associated with LVMI and diastolic dysfunction, but not with systolic function, in patients grouped by high or low plasma OPN (Y. Yang, Y. Wang, and P.J. Gao, 2020).

In a proteomic study of patients across different stages of heart failure, OPN was one of five proteins upregulated in all heart failure stages, along with N-terminal pro B-type natriuretic peptide (NT-proBNP), endothelial cell-specific molecule-1 (ESM-1), cathepsin L1, and macrophage colony-stimulating factor-1 (MCSF-1) (Michelhaugh et al., 2020). Crucially, in these settings where LVH is already established, increased expression of *Spp1* is a marker of dysfunction, inflammation, and expansion of the extracellular matrix.

The role of the *Spp1* gene as causal effector rather than biomarker in driving the development of cardiac phenotypes is not fully understood. From studies in another rat model of polygenic cardiac hypertrophy, the Hypertrophic Heart Rat (HHR), it has been hypothesised that genetic contribution to LVMI originates in dysregulated gene expression during cardiac development and neonatal growth (Porrello et al., 2009). GWAS of LV phenotypes, and other related cardiovascular diseases, demonstrate strong associations between lead variants and so-called 'developmental' genes (Nielsen et al., 2018; Aung, Vargas, et al., 2019), following hypotheses that the stressed heart undergoes a reversion to the foetal gene program. Contrasting low basal expression in the healthy adult myocardium, in cardiovascular diseases, there is an increased expression of genes normally expressed during cardiac development (such as *NPPA*/*NPPB*, and α/β myosin heavy chain) (Yim, H. Cho, and Rabkin, 2018). *Spp1* displays a similar developmental expression pattern, with a reduced basal expression in adulthood and a pathological increase during LVH, heart failure and acute phase recovery from myocardial infarction (M. Singh, Dalal, and K. Singh, 2014). The role of genes contained within the chromosome 14 region during cardiac development before adverse cardiac phenotypes become measurable is therefore important to understand.

### 3.1.1 Hypothesis and Aims

Previous work supports the hypothesis that introgression of a region of chromosome 14 into the WKY.SPGla14a and SP.WKYGla14a will have deleterious or protective effects, respectively, on the cardiac phenotypes of both male and female animals from approximately 5-weeks of age.

A secondary hypothesis to this work is an increased cardiac expression of *Spp1*, is associated with the SHRSP chromosome 14 region. In particular, the SHRSP and WKY.SPGla14a will have an increase in cardiac *Spp1* expression.

Specifically, we aim to:

- Determine genetic influence of the chromosome 14 region in both males and females through assessment of cardiac phenotype in WKY, WKY.SPGla14a, SHRSP and SP.WKYGla14a strains at three developmental time-points (1-3 day

neonatal, 5-weeks and 16-weeks).

- Characterise the extent of whole LV, interstitial and perivascular fibrosis in all 4 strains at 16-weeks and correlate with *Spp1* expression determined by qRT-PCR.

- Assess *Spp1* expression at both RNA and protein level in the hearts of male and female parental, congenic and transgenic strains during the life-course.

## 3.2 Specific Methods

The cardiac structure and function of parental WKY and SHRSP strains (officially WKY/Gla and SHRSP/Gla), and chromosome 14 congenic strains, WKY.SPGla14a and SP.WKYGla14a, was assessed using *in vivo* and *ex vivo* methods at time-points throughout development. Complimenting previous studies across the aforementioned strains, animals were selected for phenotypic assessment and tissue collection at 1-3 days (neonate), 5-weeks, and 16-weeks (Table 3.1). Both males and females were studied in all groups. The sacrifice procedure varied depending on the age of the rat. Neonatal animals were euthanised by decapitation. At 5- and 16- weeks rats were anaesthetised using sustained 5% isoflurane in 1.5L of medical oxygen and euthanised by exsanguination via cardiac puncture and aortic severance (Section 2.1.3).

**Table 3.1:** N Number of Parental and Congenic Strains Observational Developmental Study

| **Age** | **Parental Strain** | **Number (M/F)** | **Congenic Strain** | **Number (M/F)** |
|---|---|---|---|---|
| 1-3 days | WKY/Gla | 6/6 | WKY.SPGla14a | 8 |
| | SHRSP/Gla | 6/6 | SP.WKYGla14a | 8 |
| 5-weeks | WKY/Gla | 6/6 | WKY.SPGla14a | 8/8 |
| | SHRSP/Gla | 6/6 | SP.WKYGla14a | 8/8 |
| 16-weeks | WKY/Gla | 6/6 | WKY.SPGla14a | 6/6 |
| | SHRSP/Gla | 6/7 | SP.WKYGla14a | 8/8 |

Whole blood samples were collected during cardiac puncture. Between 2-5mL of blood per animal was withdrawn from the still beating left ventricle, transferred to LH or EDTA-treated vacuum tubes and stored on ice. Whole blood was fractionated by centrifugation at 3000x$g$ at 4°C for 15 minutes. The plasma supernatant was collected and stored at -80°C.

Following euthanisation, some or all of; aorta, heart, kidneys, mesenteric arteries and brain, were dissected and cleared of connective tissue. Extent of tissue collection depended on the age of the animal. Whole organ weight was recorded for heart and kidneys to the nearest milligram (mg). Atria and right ventricle were then removed

from 5- and 16-week whole hearts and the left ventricle cleared and weighed. Due to the small size of neonatal hearts, atria and right ventricles were not dissected. All tissues were immediately snap frozen in liquid nitrogen and stored in -80°C until use.

The larger tissue size of 16-week animals allowed the lower apex of LV, a section of aorta and a quarter of kidney to be fixed in 10% neutral buffered formalin for 24 hours. Following fixation, tissues were washed with PBS and stored in 70% ethanol at 4°C. The mesentery of 16-week animals was gently pulled away from the intestines and stored immediately in RNAlater (Invitrogen). The arteries were microscopically dissected by removing the surrounding adipose tissue. Isolated arteries were placed back into RNAlater, in a fresh 1.5mL eppendorf and stored at 4°C for RNA extraction. RNA was extracted within 2–3 weeks and stored at -80°C.

### 3.2.1 Tail Cuff Plethysmography

Animals at 16-weeks underwent systolic blood pressure assessment by tail cuff plethysmography as described in Section 2.1.1. A total of 8 measurements were made per blood pressure measurement session. The highest and lowest pressures were removed and an average of the final 6 measurements was recorded. If standard deviation was above 20, additional measurements were made to reduce noise and improve accuracy of individual recordings.

### 3.2.2 Transthoracic Echocardiography

Echocardiographic assessment of cardiac geometry and function was performed using ultrasound echocardiography (Section 2.1.2). During the final stages of collection in this study the Acuson Sequoia c512 ultrasound system failed and was replaced with the VisualSonics Vevo F2 (FujiFilm). Equivalent M-mode recordings were taken using this system (n=6 animals).

Image analysis was conducted using the ImageJ software as detailed in Section 2.1.2. The functional parameters; Stoke Volume (SV), Cardiac Output (CO), Ejection Fraction (EF), Fractional Shortening (FS) were determined as well as structural measures of Relative Wall Thickness (RWT) and LVM. Heart, LV and kidney mass indices were calculated by normalising organ weight (g) measured at cull or estimated from echocardiographic assessment to body weight at cull (kg).

### 3.2.3 Assessment of Inter-investigator Agreement

For a subset of the study, the ultrasound images were analysed by two separate investigators. Bland-Altman (BA) analysis was conducted to determine the extent to which estimates of cardiac parameters from echocardiographic analyses agreed between multiple investigators. In a BA analysis, the mean bias (difference) between two inves-

tigators is plotted against the mean of measures for each individual. As exemplified in Figure 3.2, the plot includes a shaded region containing the 95% confidence interval of the mean bias. If this interval contains 0, the investigator bias is likely not significant.

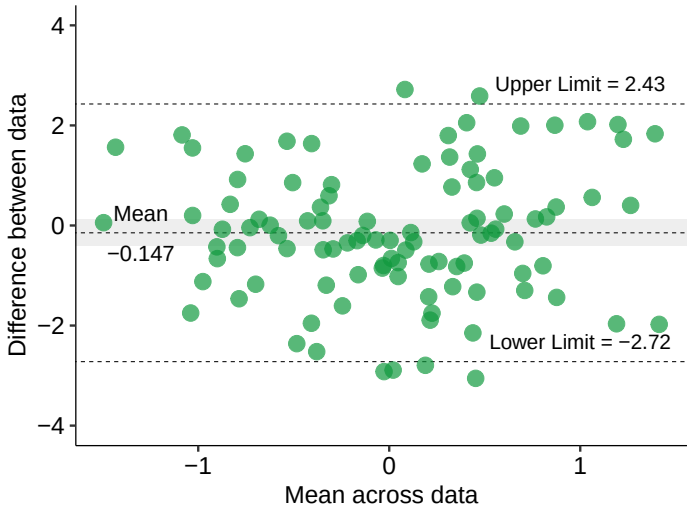



**Figure 3.2: Example of Bland-Altman Plot** Example plot generated from n=100 random points

The upper and lower dashed lines (Figure 3.2) represent the upper and lower limits of agreement (LOA) respectively. These are calculated as the mean bias ±1.96 * standard deviation of the mean. The LOA will capture 95% of all points, indicating the range of agreement that measurements will lie within, 95% of the time. Thus, the assessment of inter-investigator agreement must therefore be interpreted from these intervals keeping in mind suitable biological interpretations.

Investigator bias can be introduced at two points in echo analysis; firstly in the selection of images to analyse, which can affect estimation of heart rate (HR). A second bias can be introduced in the physical measurement of wall thickness. All echocardiographic parameters are derived from wall thickness alone or a combination of wall thickness and heart rate. Relative wall thickness (RWT) is dependant solely on measures of wall thickness and has few additional mathematical adjustments. Assessment of investigator bias and LOA in length measurements during analysis is therefore best assessed by this measure. Left ventricular mass (LVM) is of high biological significance and is similarly dependent on the length measurements taken during analysis. Finally, as HR significantly influences cardiac output (CO), an assessment of agreement in this measure can indicate how image selection and HR estimation might affect differences between investigators.

### 3.2.4 Histological Assessment of Cardiac Fibrosis

In 16-week animals, cardiac fibrosis was assessed using picrosirius red staining from sections of the apical LV (Section 2.7). Images were analysed using QuPath (Bankhead

et al., 2017), an open source software for bio-image analysis. Positive and negative stain area from whole slide images were measured using the Pixel Classifier tool. The tool was trained using a random selection of tiles from representative images to classify pixels as 'positive' or 'negative' using assessment of intensity across the RBG channels. A simple thresholder was used to define tissue area. Total positive tissue area per section was divided by total tissue area per section and this ratio was used a measure of cardiac fibrosis, where higher ratios represented higher total fibrosis content.

### 3.2.5 qRT-PCR and Immunoblotting

Whole hearts collected from neonate animals were split in liquid nitrogen for use in RNA and protein extraction procedures. Flash frozen pieces of LV were used for RNA isolation from 5-week and 16-week animals. cDNA was prepared using superscript II reverse transcriptase enzymes as described in Section 2.3.3. TaqMan qRT-PCR was performed to assess *Spp1* expression normalised to *B2m* housekeeper gene and analysed using the relative quantification method described in Section 2.3.6.

Protein was extracted from frozen nenonate hearts according to protocol described in Section 2.5. Expression of *Spp1* protein was analysed using standard western blotting techniques (Section 2.6). Protein isolated from neonate hearts was mixed with 4x SDS buffer containing reducing agent and boiled at 95°C for 5 minutes. Lysed samples were loaded in 10µL volume to achieve a total of 5µg protein per lane. Proteins were separated by gel electrophoresis and transferred to a nitrocellulose membrane. Osteopontin protein was normalised to total protein (Revert 700 Total Protein Kit, LI-COR). Membranes were blocked in TBS blocking buffer before incubation with primary antibody solution (OPN, Abcam, ab3856, 1:1000) overnight at 4°C. Following overnight incubation with primary antibody solution, membranes were washed with TBS-T (50mM Tris-HCL at pH=7.4, 150mM NaCl and 0.1% Tween-20) and incubated in secondary antibody solution (LI-COR IRDye 800CW, 1:20000) for 1 hour at room temperature. Membranes were visualised using 800 nm channel (Odessey CLx, LI-COR) and band intensities quantified by ImageStudio software. As per the manufacturers' protocol, relative quantification of target protein was achieved by normalising target to a given lane normalisation factor (LNF), calculated from total protein signal.

### 3.2.6 Data Handling and Statistical Methods

Data was analysed R.4.3 (or equivalent) using tidyverse (Wickham et al., 2019) and rstatix (Kassambara, 2023) libraries. An analysis pipline was developed utilising custon scripts (`https://github.com/c-triv97/Source-Codes`). Individual data were analysed using one-way ANOVA followed by Tukey *post-hoc* testing. A $p$-value of $\leq 0.05$ was considered statistically significant. Adjustments for multiple testing were applied during *post-hoc* testing and significant pairwise comparisons are displayed in the plot

area. At neonate and 5-week timepoints, one-way ANOVA was carried out on sex-aggregated data to determine group differences based on genotype. At the 16-week timepoint, ANOVA was carried out separately in male and female groups.

## 3.3 Results

### 3.3.1 Neonate Parental and Congenic Strains

Body weight was significantly, positively associated with heart and kidney weight in neonate animals (Figure 3.3). All parameters, including heart and kidney mass indexed to body weight, were significantly, positively correlated with age (in days). Body weight was not significantly correlated with either heart or kidney weights normalised to body weight however correlation with age remained significant (Figure 3.3). As shown in Table 3.2 the day (age) at sacrifice was varied between groups. In the WKY and WKY.SPGla14a strains, multiple litters were used and thus age was varied within group only in these strains.

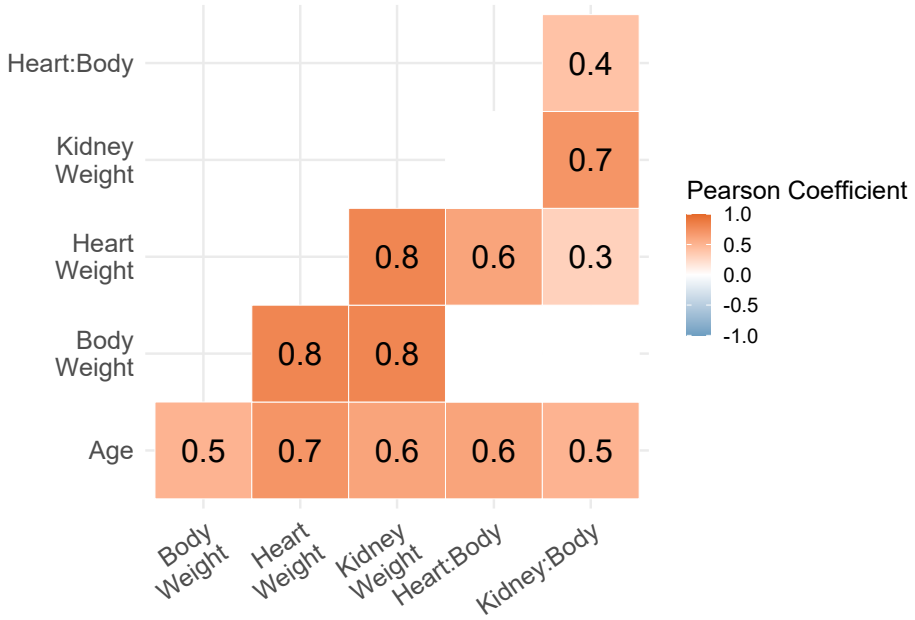



**Figure 3.3: Correlation Matrix of Measured Variables in Neonates** Pearson correlation matrix across all individuals, dark color represents higher pearson correlation coefficient. Non-significant correlations ($p$>0.05) are blank. Replicate correlations on the top half of the matrix were removed for readability.

**Table 3.2:** Neonate Litters Utilised for Study

| **Strain** | **Day 1** | **Day 2** | **Day 3** | **Num. Litters** |
|---|---|---|---|---|
| WKY | 0 | 8 | 4 | 4 |
| SHRSP | 0 | 12 | 0 | 1 |
| WKY.SPGla14a | 4 | 0 | 4 | 2 |
| SP.WKYGla14a | 0 | 0 | 8 | 1 |

Num. litters describes the number of different litters required to reach N numbers in experimental neonate groups. In most cases, the number of pups taken was less than the number of pups in the litter, to reduce stress and improve subsequent mating success.

Absolute body weight, heart weight and kidney weight was significantly different between genotypes (Figure 3.4). Compared to WKY and WKY.SPGla14a, SHRSP neonates had significantly reduced heart and body weight. In contrast, compared to the WKY, WKY.SPGla14a, and SHRSP neonates, the SP.WKYGla14a day 1–3 neonates had significantly increased heart and body weight (Figure 3.4, all $p<.001$). There was no significant difference between heart and body weights of WKY and WKY.SPGla14a neonates.

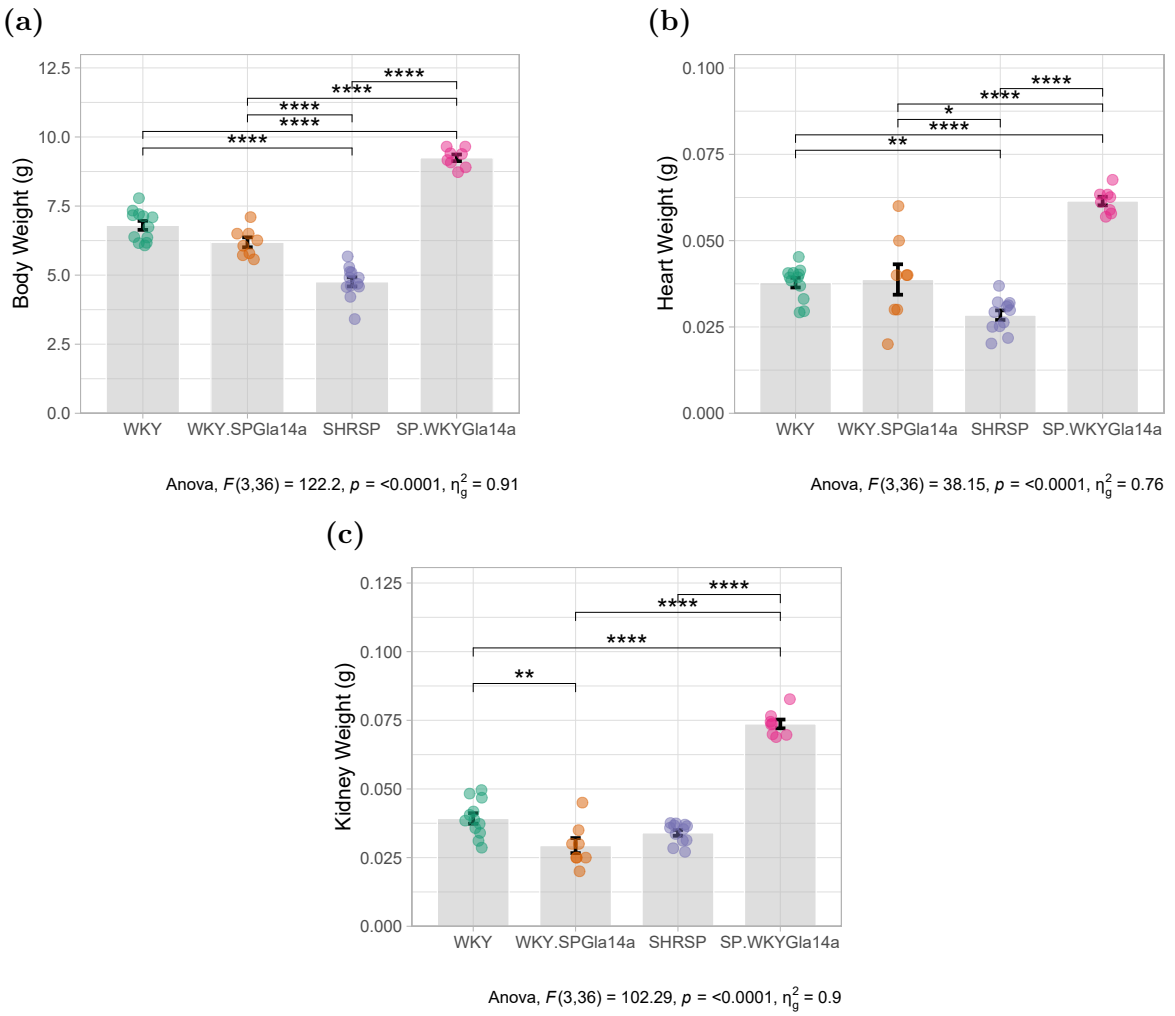



**Figure 3.4: Absolute Organ Weight of 1–3 Day Neonates.** (a) Body, (b) heart, and (c) kidney weights of 1–3 day neonates across WKY, WKY.SPGla14a, SHRSP and SP.WKYGla14a strains. Group means were compared using one-way ANOVA and significant *post-hoc* comparisons are displayed (*$p$<0.05, **$p$<0.01, ***$p$<0.001, **** $p$<0.0001, n=8–12 per group)

When normalised to body weight, only kidney weight was significantly different between neonate animals from different genetic backgrounds (Figure 3.5). There were no significant differences in kidney to body weight ratio between the chromosome 14 congenic strains and their relative background strains. Compared to WKY and WKY.SPGla14a neonates, SHRSP and SP.WKYGla14a neonates have significantly increased kidney weight when corrected for body weight (p<0.05).

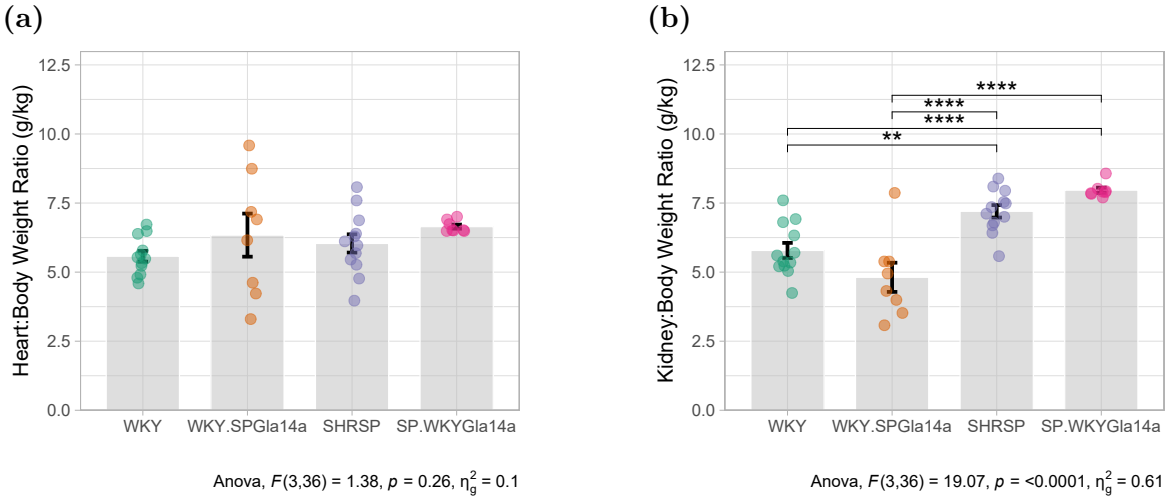



**Figure 3.5: Relative Heart and Kidney Weight of 1–3 Day Neonates** (a) Heart, and (b) kidney mass indexed to whole body weight of 1-3-day neonates from WKY, WKY.SPGla14a, SHRSP and SP.WKYGla14 strains. ANOVA results and significant *post-hoc* comparisons are displayed with the plot areas (*$p$<0.05, **$p$<0.01, ***$p$<0.001, **** $p$<0.0001, n=8-12 per group)

### 3.3.2 Inter-Investigator Validity of Echocardiographic Measurements

Two independent investigators analysed 5- and 16-week echocardiographic images. Agreement and bias was assessed using Bland-Altman (BA) plots. In all measures assessed, data points are spread evenly around the mean bias at each point along the x-axis, indicating investigator bias was not more severely affected at the extremes of measurements (high or low mean value, Figure 3.6). Interpretation of BA plots is dependent on the biological relevance of the mean bias and limits of agreement (LOA). Analysis was conducted across the measures of RWT, LVM and CO. The mean bias for RWT and LVM was estimated to be -0.005mm and -0.008g, respectively (Figure 3.6). In both cases the 95% CI contained the null value, confirming inter-investigator measures derived from echocardiography images are comparable. The range of values derived for RWT were within 0.316 mm and 0.262 g for LVM between investigators (Figure 3.6).

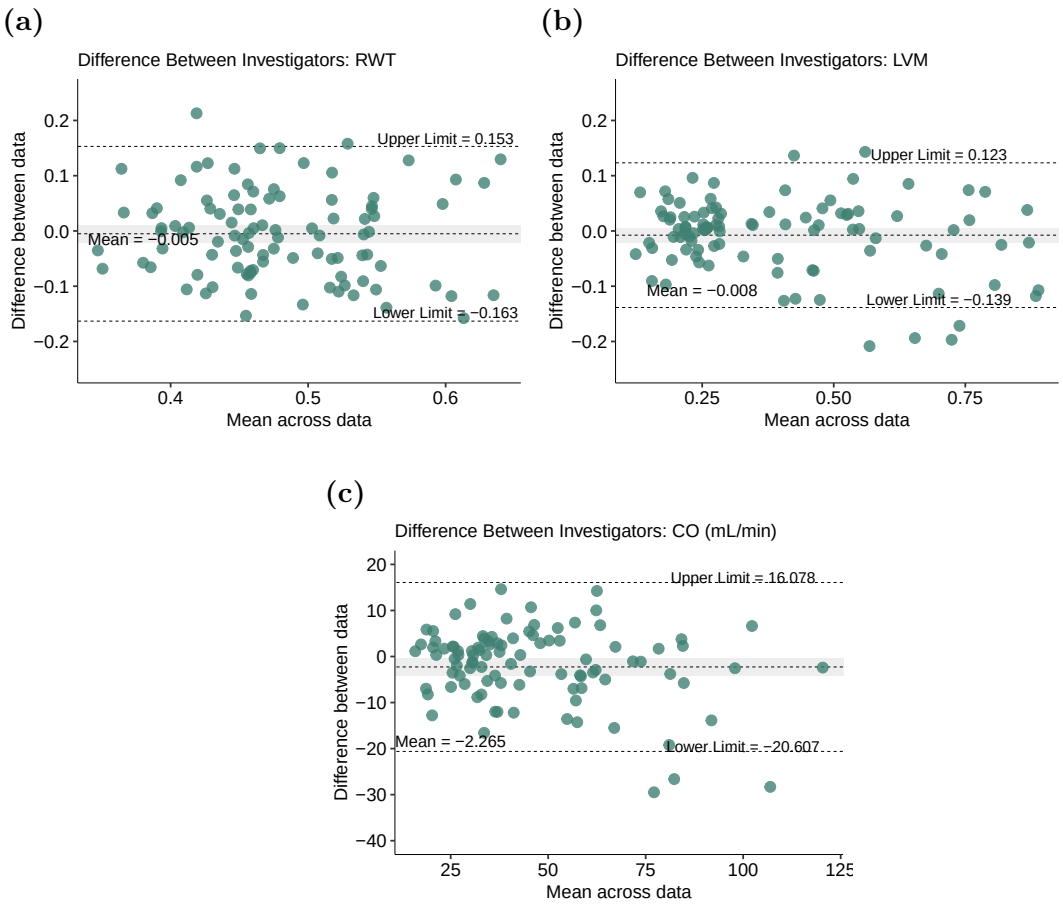



**Figure 3.6: Bland-Altman Analysis of Echocardiographic Measurements** Bland-Altman plots showing mean bias and limits of agreement for (a) relative wall thickness (RWT), (b) left ventricle mass (LVM) and (c) cardiac output (CO). 95% confidence interval of the mean bias is shaded light grey and limits of agreement are represented by dashed lines on upper and lower limits.

The estimates of cardiac output were more variable, with the mean bias centered around -2.265 mL/min. Unlike RWT and LVM which rely purely on length measurements, estimates of cardiac output also depend on the estimation of HR. This increased opportunity for bias to be introduced is reflected in the wider 95% CI (-4.1410859, -0.3885681) which did not contain the null value, and the larger range of LOA (∼35mL/min range).

### 3.3.3 Cardiac Structure and Function in 5-week Parental and Congenic Strains

At 5-weeks of age, body weight was significantly different by genotype according to one-way ANOVA (Figure 3.7a). In post-hoc testing, the WKY.SPGla14a were significantly smaller than their donor SHRSP strains ($p$<0.05). When normalised to body weight, there were no significant differences in heart weight between parental or congenic strains (Figure 3.7b). However, after dissection of the atria and RV from the LV, there was a significant effect of genotype on LVM relative to body weight, where both SHRSP and

SP.WKYGla14a had an increased LVMI compared to the WKY animals (Figure 3.7c).

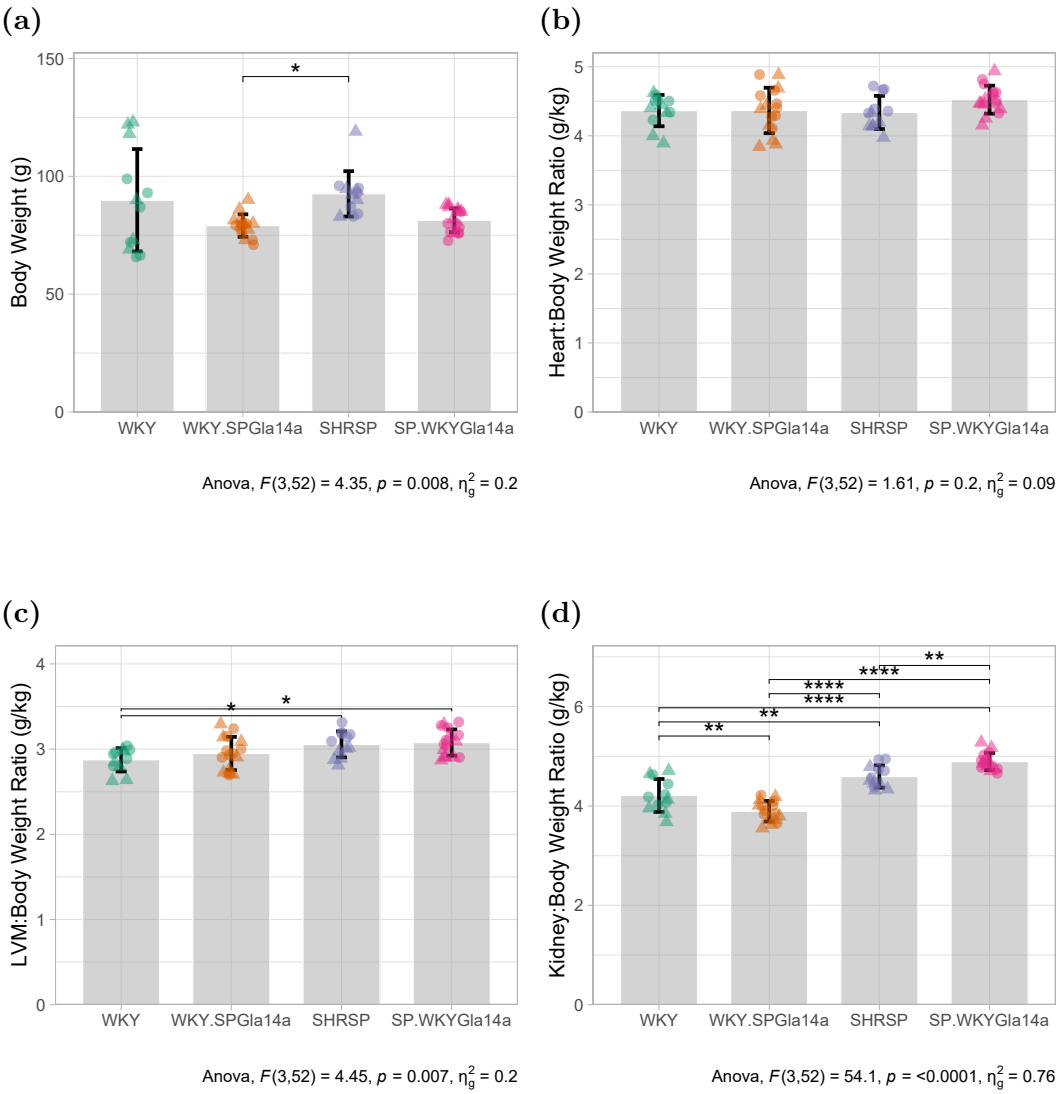



**Figure 3.7: Structural Measures in 5-week Parental and Congenic strains** (a) Body, (b) heart, (c) LV, and (d) kidney mass relative to body weight at cull in WKY, WKY.SPGla14a, SHRSP and SP.WKYGla14a animals. Individual animals are represented by points, closed circles represent females and closed triangles represent males. One-way ANOVA was performed and significant post-hoc comparisons are displayed in plot area (*$p$<0.05, **$p$<0.01, ***$p$<0.001, **** $p$<0.0001. N=6-8 male & female per group.)

There was a significant effect of genotype on normalised kidney weight (g/kg body weight) where both the SHRSP and SP.WKYGla14a had significantly increased relative kidney weights compared to both the WKY and WKY.SPGla14a. Compared to the WKY, the WKY.SPGla14a had a significantly reduced kidney weight whilst the SP.WKYGla14 had an increased kidney weight compared to its parental SHRSP strain (Figure 3.7d).

When LVM was estimated from echocardiographic measurements, post-hoc comparisons showed estimated LVMI in SP.WKYGla14a animals was significantly increased

compared to both WKY animals and SHRSP animals (Figure 3.8a). In functional measures derived from echocardiography, one-way ANOVA showed no significant effect of genotype on stroke volume, ejection fraction or cardiac output by 5-weeks of age (Figure 3.8 & Figure 3.9). Fractional shortening showed a significant main effect of genotype in ANOVA, although no comparisons reached significance after adjustment for multiple corrections, WKY.SPGla14a and SHRSP strains appeared to have greater fractional shortening than WKY and SP.WKYGla14a (Figure 3.8).

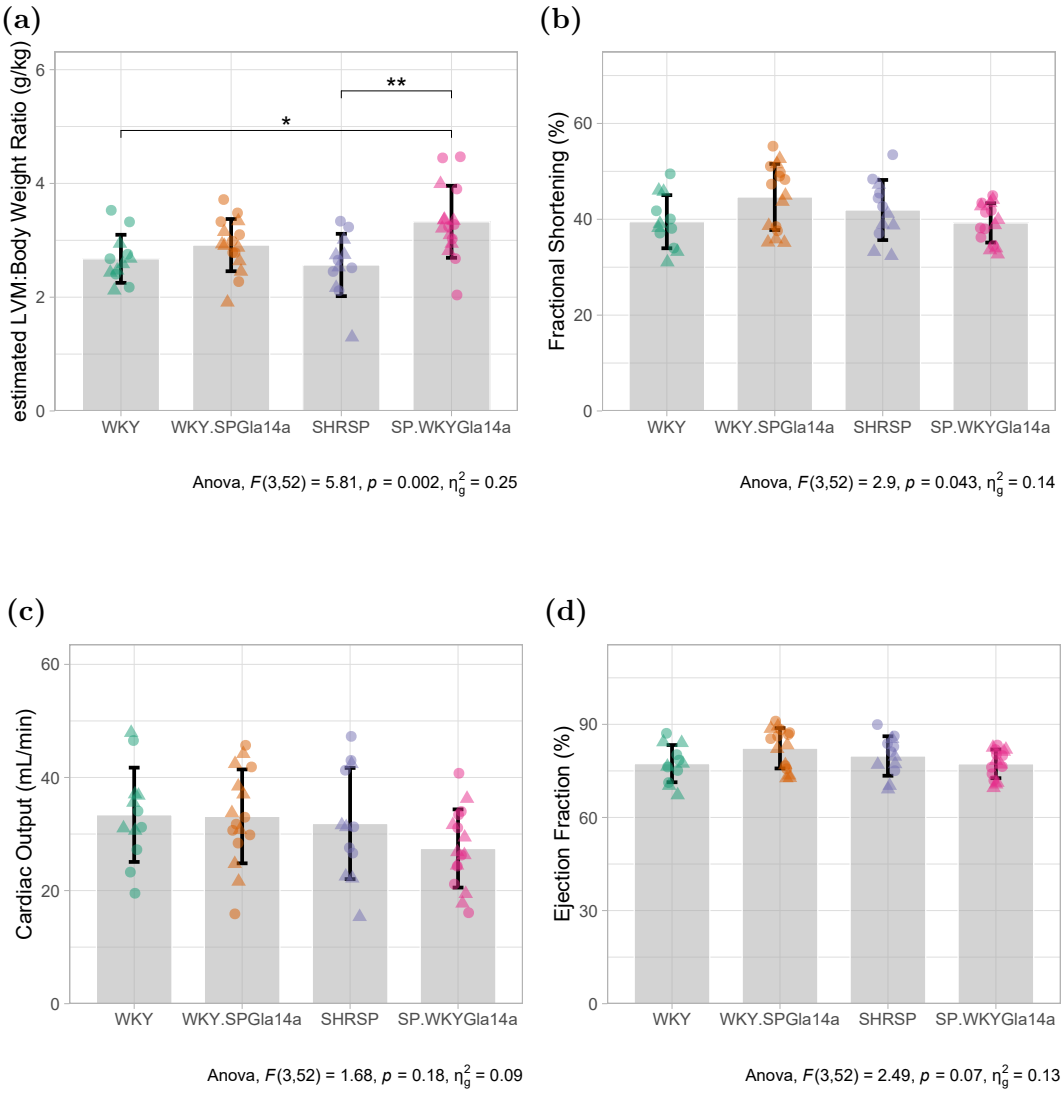



**Figure 3.8: Measures of Cardiac Function in 5-week Parental and Congenic Strains** Echocardiography derived measures of (a) LVMI estimated by echocardiography, (b) fractional shortening (%), (c) cardiac output (mL/min), and (d) ejection fraction (EF) in WKY, WKY.SPGla14a, SHRSP and SP.WKYGla14a animals. Individual animals are represented by points, closed circles represent females and closed triangles represent males. One-way ANOVA was performed and significant post-hoc comparisons are displayed in plot area (*$p$<0.05, **$p$<0.01, ***$p$<0.001, **** $p$<0.0001. N=6-8 males & females per group).

RWT, a measure of cardiac hypertrophy sensitive to concentric remodelling of the left ventricle, showed there was a significant effect of genotype. Compared to all other strains, SP.WKYGla14 had significantly increased RWT (Figure 3.9a).

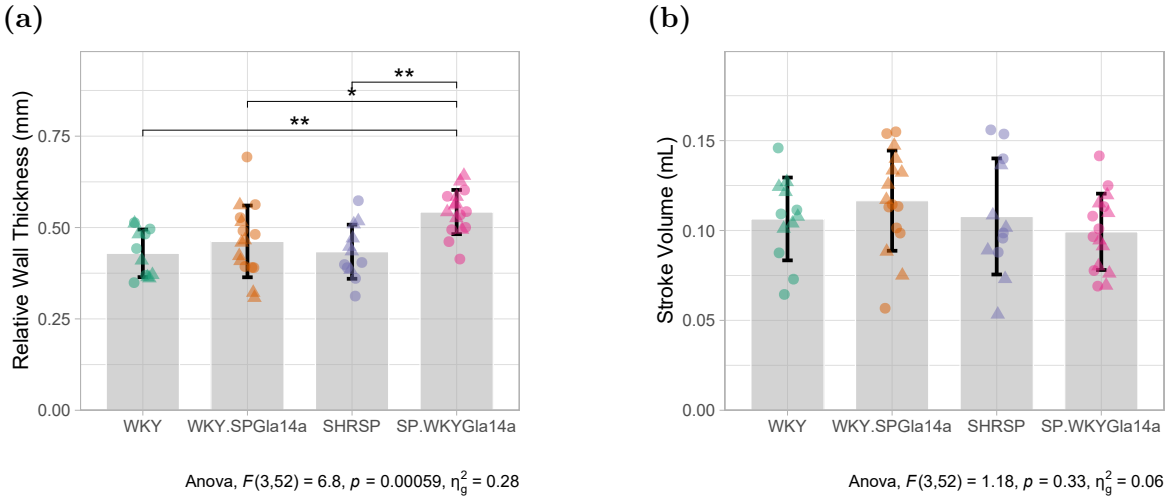



**Figure 3.9: Relative Wall Thickness at 5-weeks in Parental and Congenic Strains** (a) Relative wall thickness (RWT) and (b) stroke volume (mL) estimated from echocardiography in WKY, WKY.SPGla14a, SHRSP, and SP.WKYGla14a. Individual animals are represented by points, closed circles represent females and closed triangles represent males. One-way ANOVA was performed and significant post-hoc comparisons are displayed in plot area (*$p$<0.05, **$p$<0.01, ***$p$<0.001, **** $p$<0.0001. N=6-8 males & females per group)

### 3.3.4 Blood Pressure and Cardiac Phenotype in 16-week Parental and Congenic Strains

By 16-weeks of age, rats have undergone sexual maturity and show significant sex effects in all measures, not corrected for when increased male body size is accounted for by indexing measures to body weight. Additional sex-determined factors might also play a causal role in determining blood pressure and cardiac structure and function. Thus, at 16-weeks, all genotype comparisons were carried out in male and female groups separately.

In the SHRSP, onset of high blood pressure occurs at approximately 8–12 weeks of age. By 16-weeks, as shown in Figure 3.10, blood pressure is significantly increased in 16-week male SHRSP and SP.WKYGla14a animals compared to WKY. Similarly, female WKY.SPGla14a animals have a significantly lower blood pressure than both the SHRSP and SP.WKYGl14a animals. The downward trend in blood pressure between SHRSP and SP.WKYGla14a and WKY females is significant only in the congenic versus donor comparison (Figure 3.10).

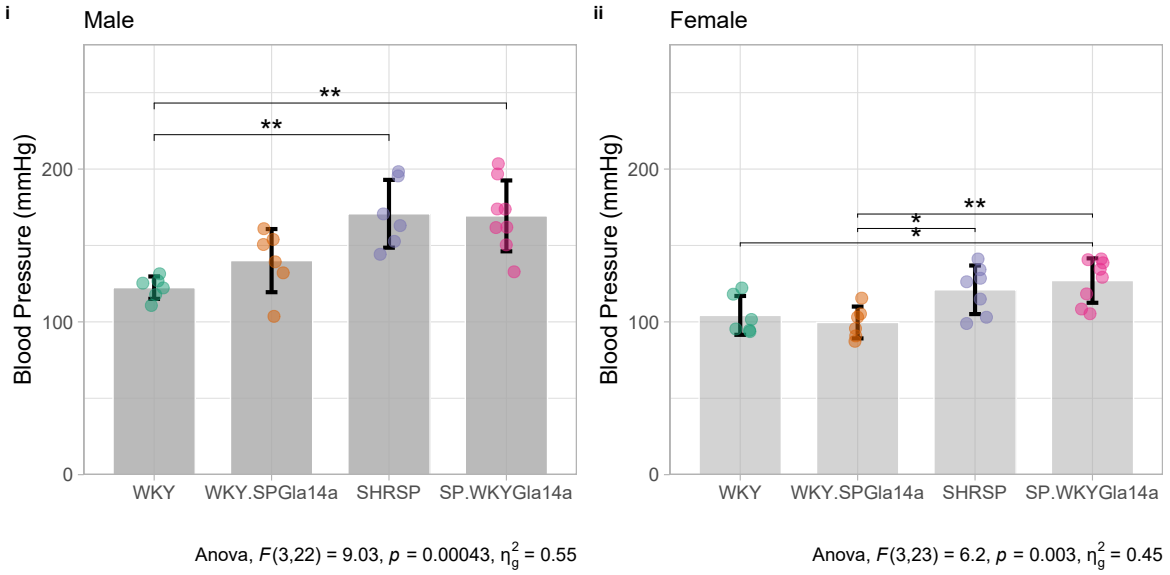



**Figure 3.10: Blood Pressure in 16-week Rats** Blood pressure recorded using tail cuff plethsmography at 16-weeks of age in male and female parental and chromosome 14 congenic strains. One-way ANOVA was performed and significant post-hoc comparisons are displayed in plot area (*$p$<0.05, **$p$<0.01, ***$p$<0.001, **** $p$<0.0001, N=6–8 male & female per group)

In male animals only, there was a significant difference in body weight between genotypes (Figure 3.11), where WKY males were significantly larger than SHRSP and SP.WKYGla14a males. As expected, males had larger body weights than female animals (Figure 3.11, F(1,51)=706.9, $p$<.001 (not shown on graph)).

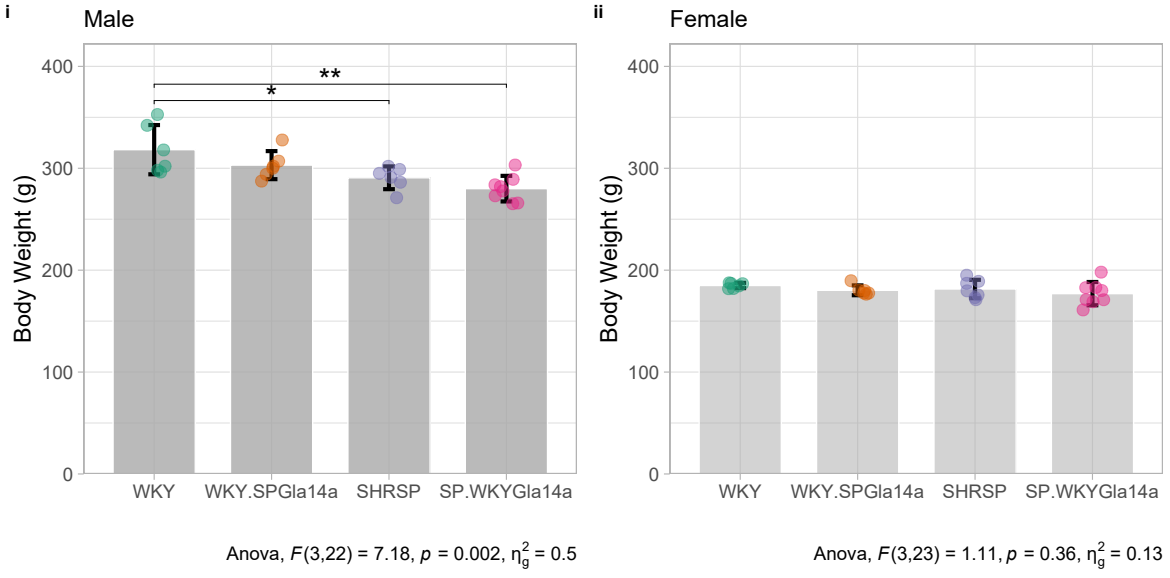



**Figure 3.11: Body Weight of 16-week Parental and Congenic Strains** Body weight measured at cull in male and female WKY, SHRSP, WKY.SPGla14a and SP.WKYGla14a strains. One-way ANOVA was performed and significant post-hoc comparisons are displayed in plot area (*$p$<0.05, **$p$<0.01, ***$p$<0.001, **** $p$<0.0001. N=6–8 male & female per group)

The effect of genotype is the same across sexes in indices of whole heart and LV mass. Both the SHRSP and SP.WKYGla14a had an increased heart and LVMI compared to both WKY and WKY.SPGla14a strains (Figure 3.12). In the female comparison, the SHRSP had a significantly reduced heart and LV indices compared to its background congenic SP.WKYGla14a.

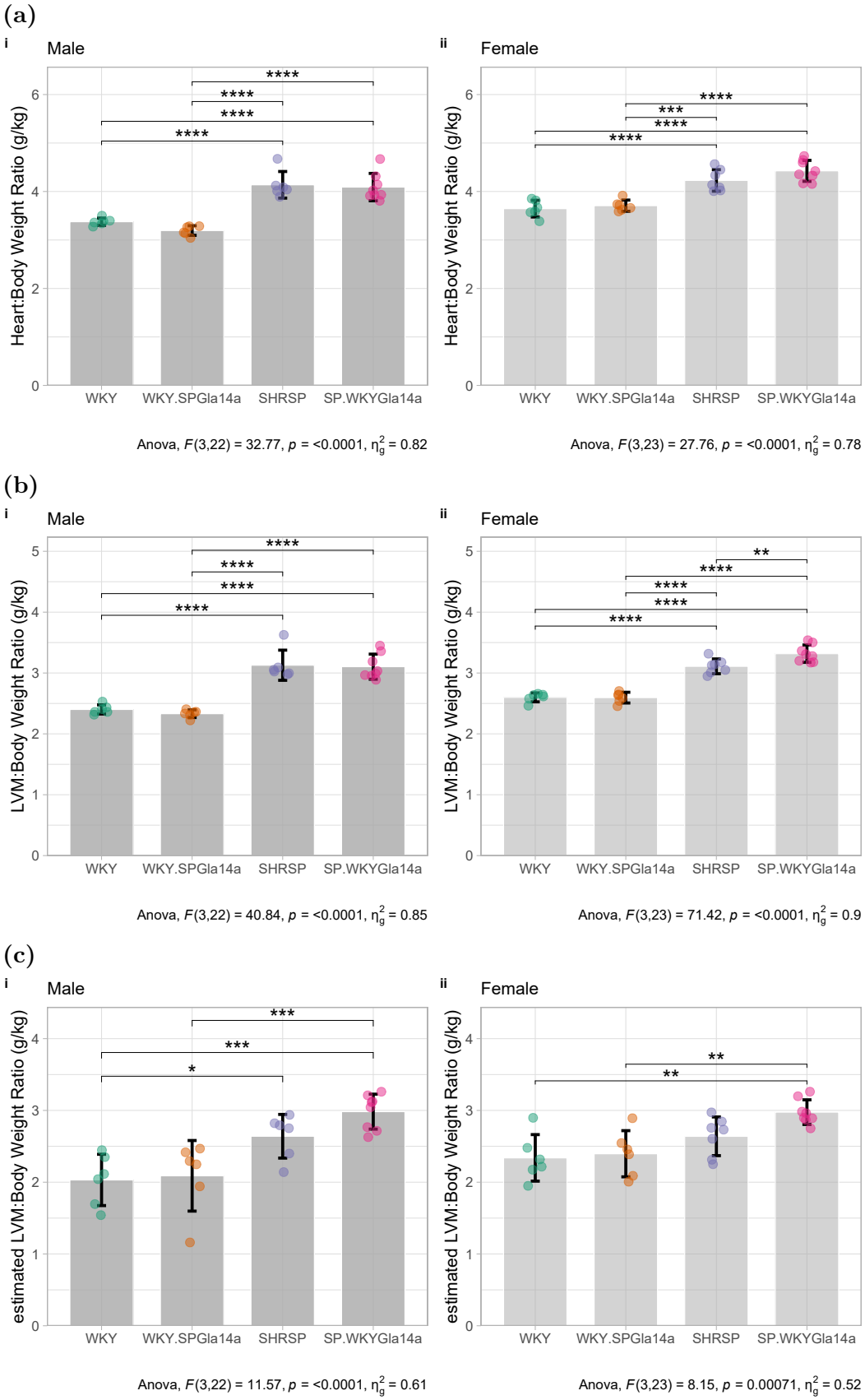



**Figure 3.12: Cardiac Structure in 16-week Animals** (a) Heart, and (b) LV relative to body weight (LVMI) at cull and (c) estimated by echocardiography in male and female WKY, WKY.SPGla14a, SHRSP and SP.WKYGla14a animals. One-way ANOVA was performed and significant post-hoc comparisons are displayed in plot area (*$p$<0.05, **$p$<0.01, ***$p$<0.001, **** $p$<0.0001. N=6-8 male & female per group)

With higher variability in the echocardiographic data, fewer comparisons reached significance in post-hoc testing. In females and males, compared to the WKY, estimated LVMI is greater in SHRSP and SP.WKYGla14a strains (comparison of SHRSP to WKY in females is not significant, Figure 3.12c). SP.WKYGla14a strains also show an increased LVMI measured by echocardiography compared to the WKY.SPGla14a (Figure 3.12c).

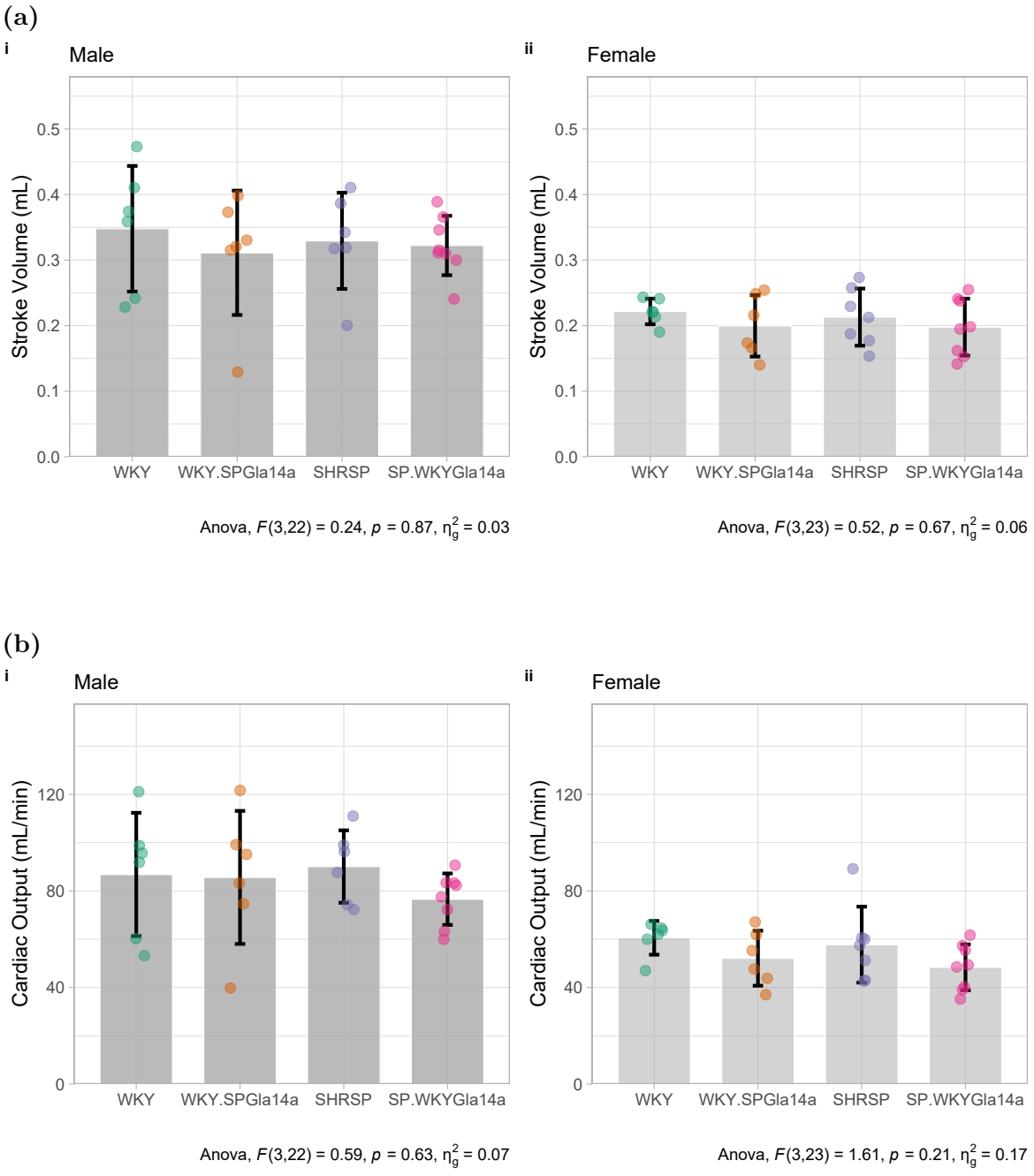



**Figure 3.13: Stroke Volume and Cardiac Output at 16-weeks in Parental and Congenic Strains** Echocardiography derived measures of (a) stroke volume (mL) and (b) cardiac output (mL/min) in male and female parental and congenic strains at 16-weeks of age. One-way ANOVA was performed in both male and female groups and significant post-hoc comparisons. One-way ANOVA was performed and significant post-hoc comparisons are displayed in plot area (*$p$<0.05, **$p$<0.01, ***$p$<0.001, **** $p$<0.0001). N=6-8 male & female per group

As in the 5-week animals, there were no significant differences between parental and congenic strains in functional measures of stroke volume, cardiac output (Figure 3.13), ejection fraction and fractional shortening (Figure 3.14). By 16 weeks, RWT differed between strains only in females, where compared to its chromosome 14 donor strain (WKY), SP.WKYGla14 had significantly increased RWT (Figure 3.15). In both sexes, there is a non-significant trend toward increasing RWT in WKY.SPGla14a, SHRSP and SP.WKYGla14a versus the WKY.

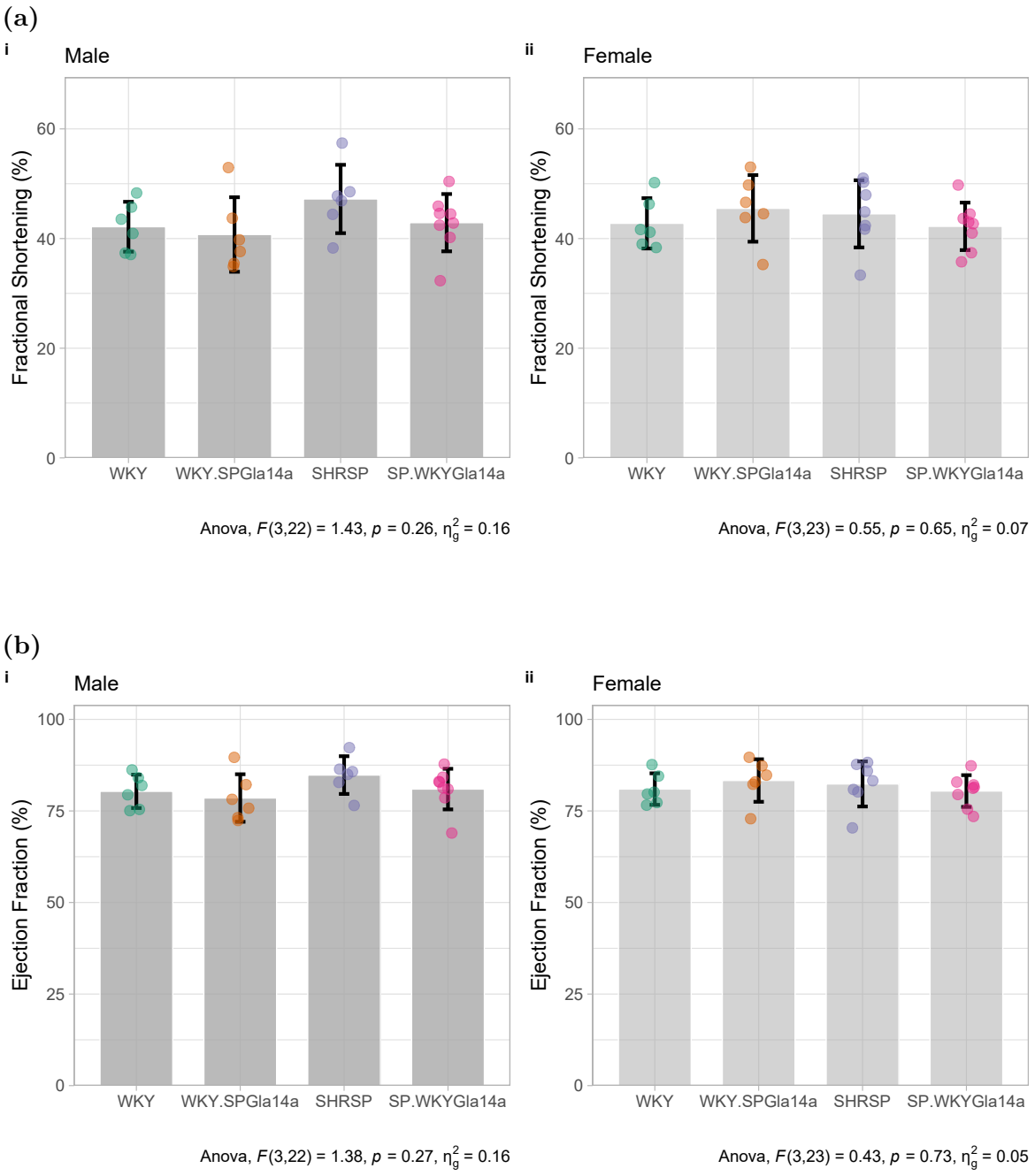



**Figure 3.14: Ejection Fraction and Fractional Shortening at 16-weeks in Parental and Congenic Strains** Echocardiography derived measures of (a) fractional shortening (FS), and (b) ejection fraction (%) in male and female parental and congenic strains at 16-weeks of age. One-way ANOVA was performed in both male and female groups and significant post-hoc comparisons are displayed in plot area (*$p$<0.05, **$p$<0.01, ***$p$<0.001, **** $p$<0.0001. N=6-8 male & female per group)

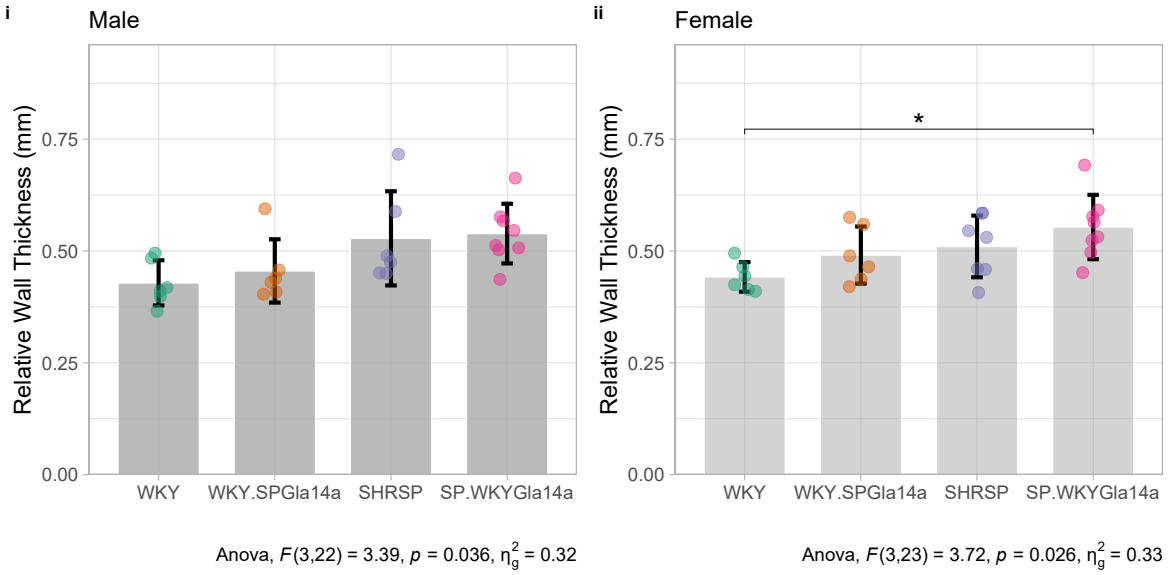



**Figure 3.15: Relative Wall Thickness at 16-weeks in Parental and Congenic Strains** Relative wall thickness (RWT) estimated from echocardiography in WKY, WKY.SPGla14a, SHRSP, and SP.WKYGla14a. One-way ANOVA was performed in both male and female groups and significant post-hoc comparisons are displayed in plot area (*$p$<0.05, **$p$<0.01, ***$p$<0.001, **** $p$<0.0001. N=6-8 male & female per group)

There was no significant effect of genotype in percentage positive staining by picrosirus red, to assess fibrosis in the apex of the LV (Figure 3.16). Data was highly variable within group, with some biological replicates displaying coefficients of variation >0.6 (60%) between technical replicates.

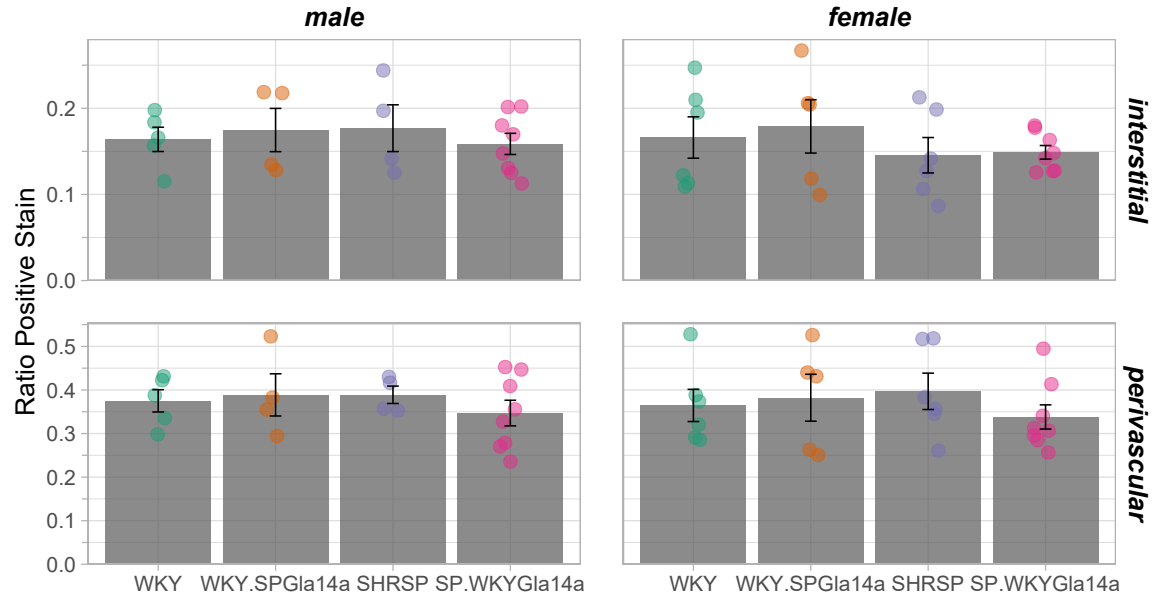



**Figure 3.16: Assessment of LV Fibrosis at 16-weeks** Ratio of area positive stain to total area in (top) interstitial and (bottom) perivascular regions of the LV. Assessment of fibrosis was assessed using chromogenic staining of LV sections with picrosirus red. N=4–8 males & females per group.

Kidney weight was significantly altered by genotype in both male and female comparisons. Both SHRSP and SP.WKYGla14a strains had significantly increased kidney mass compared to both WKY and WKY.SPGla14a strains (Figure 3.17).

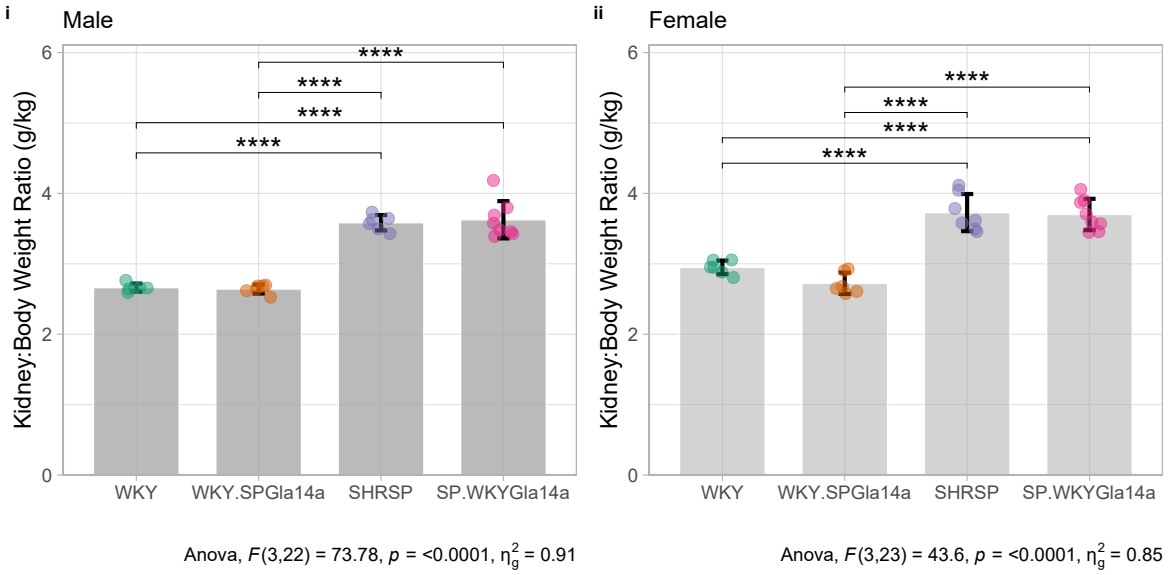



**Figure 3.17: Kidney Weight at 16-weeks in Parental and Congenic Strains** Kidney weight at cull in 16-week WKY, WKY.SPGla14a, SHRSP, and SP.WKYGla14a males and females. One-way ANOVA was performed in both male and female groups and significant post-hoc comparisons are displayed in plot area (*$p$<0.05, **$p$<0.01, ***$p$<0.001, **** $p$<0.0001. N=6-8 male & female per group)

### 3.3.5 qRT-PCR of *Spp1* Expression Over the Life Course

Quantitative real-time PCR was performed in whole heart (neonate) or LV tissue (5- and 16-weeks) in all strains. Expression was quantified in relation to the WKY, such that an RQ of 1 is equal to expression in the WKY, <1 indicates a reduced expression and >1 indicates increased expression.

In the neonates, WKY.SPGla14a and SHRSP strains had significantly higher *Spp1* expression (group RQ $\geq$ 5) compared to the WKY. The effect was signifiant in post-hoc comparison comparisons (Figure 3.18a). Compared to its background SHRSP and the WKY.SPGla14a congenic, the SP.WKYGla14a had significantly reduced *Spp1* expression. SP.WKYGla14a *Spp1* expression was not significantly different compared to the WKY (Figure 3.18a).

In 5-week animals, there was a similar significant effect of genotype such that WKY.SPGla14a and SHRSP had increased *Spp1* expression versus the WKY, whilst cardiac expression of *Spp1* was relatively decreased in the SP.WKYGla14a compared to both its background SHRSP, and its chromosome 14 donor, WKY (Figure 3.18b). In post-hoc comparisons, only the increase in expression in the WKY.SPGla14a versus the WKY and SP.WKYGla14a comparisons were significant at the $\alpha$=0.05 level

(Figure 3.18b).

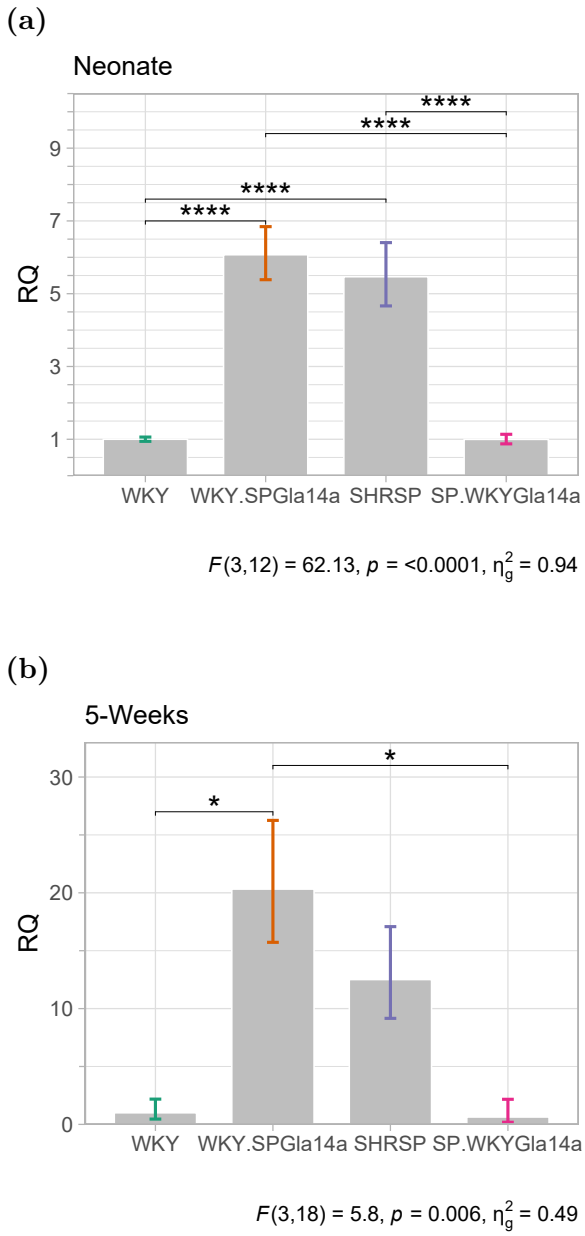



**Figure 3.18: *Spp1* Expression in Neonate and 5-week Hearts** TaqMan qRT-PCR of *Spp1* expression in (a) neonate hearts and (b) LV from 5-week parental and chromosome 14 congenic strains. Expression was normalised in each replicate to housekeeper *B2m*. One-way ANOVA was performed on delta-Ct values averaged across 3 technical replicates (where possible). Relative quantification values in relation to the WKY are displayed on plots. N=3 per group

At 16-weeks, compared to female WKY, expression of *Spp1* is increased ∼100-fold in WKY.SPGla14a and SHRSP strains. Due to high variability in biological samples (n=3/group) no significance was found in one-way ANOVA. In males, there was a more modest increase in *Spp1* expression, in WKY.SPGla14a and SHRSP compared to both the WKY and SP.WKYGla14a strains (Figure 3.19). In post-hoc comparisons, only the increase in male WKY.SPGla14a was significant compared to the WKY.

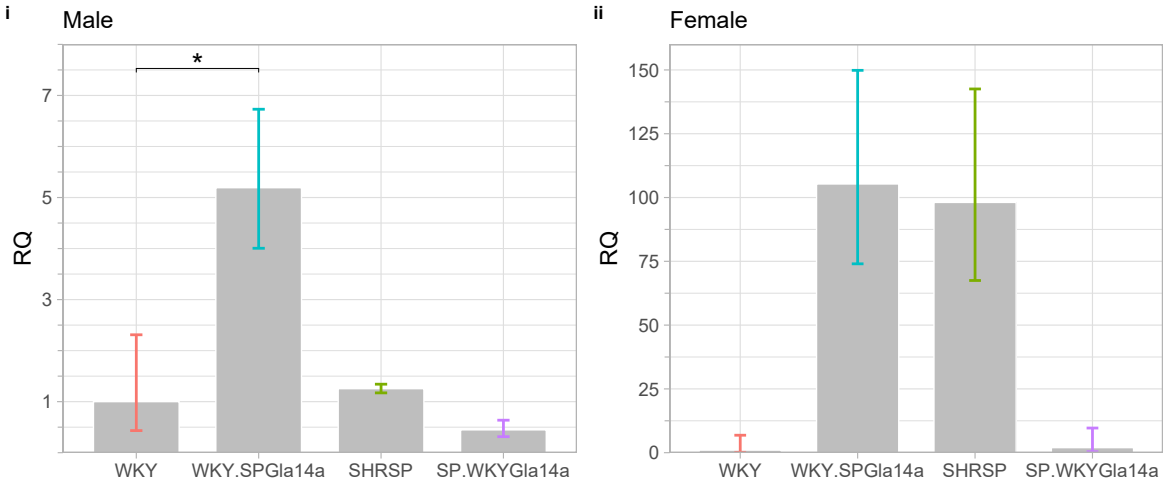



**Figure 3.19: *Spp1* Expression in 16-week Left Ventricle** qRT-PCR of *Spp1* expression in the LV of parental and chromosome 14 congenic strains. Expression was normalised in each replicate to housekeeper *B2m*. One-way ANOVA was performed on delta-Ct values averaged across 3 technical replicates (where possible). Relative quantification values in relation to the WKY are displayed on plots. N=3 per group

At the protein level, there was a significant difference in *Spp1* protein expression, assessed by western blot in neonate hearts (Figure 3.20). In post-hoc comparisons, the expression of *Spp1* was reduced in SP.WKYGla14a compared to its parental SHRSP strain. Protein expression of *Spp1* was not quantified in 5- and 16-week animals by western blot. An assessment of *Spp1* protein expression was undertaken through floursence based immunohistochemistry in 16-week animals, presented in Chapter 4.

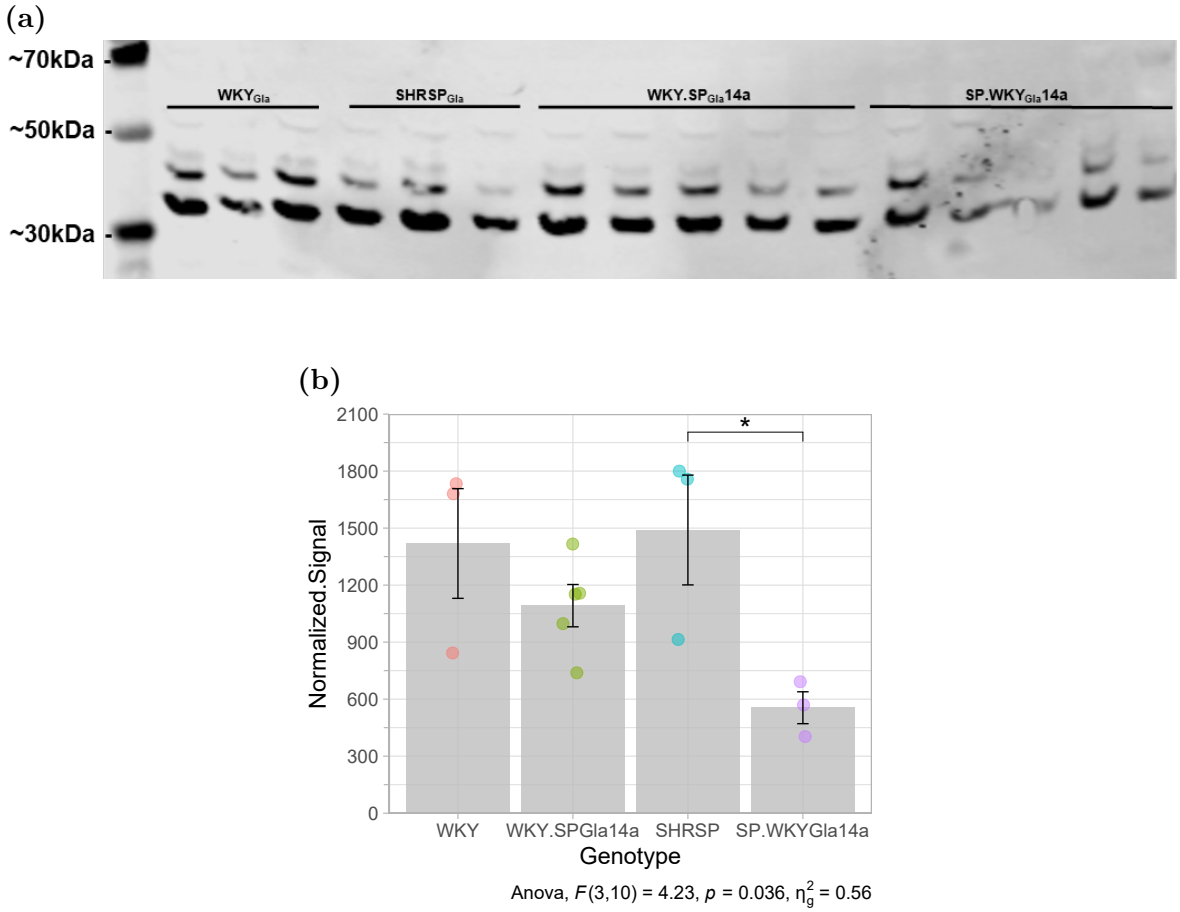



**Figure 3.20: *Spp1* Protein Expression in Neonate Hearts** Western blot analysis of parental and congenic *Spp1* expression in neonate hearts, quantified by normalisation to total protein. One-way ANOVA and significant post-hoc test displayed on graph. Lanes 13 and 14 were excluded from analysis. N=3 WKY, SHRSP and SP.WKYGla14a and N=6 WKY.SPGla14a. (*$p$<0.05, **$p$<0.01, ***$p$<0.001, **** $p$<0.0001).

## 3.4 Discussion

The SHRSP/Gla strain is known to develop increased blood pressure in adulthood, between 8–12 weeks of age, following puberty and the divergence of phenotype in male and female animals. Blood pressure exhibits an effect on heart geometry and function however, differences in cardiac structure is measurable between SHRSP and WKY strains from 5-weeks. Here we have replicated these common findings, where organ weight data at cull shows a significant increase in LVM indexed to body weight (LVMI) in SHRSPs compared to WKYs by 5-weeks of age. This is maintained following the development of significantly increased blood pressure in SHRSP animals by 16-weeks. Consistent with the expected phenotype, there were no measurable differences in measures of systolic function (CO, FS, EF, and SV) at 16-weeks between WKY and SHRSP animals. In addition to the expected differences in LV geometry and blood pressure, kidney to body weight ratio was consistently increased (neonate, 5- and 16-weeks) in SHRSP animals compared to the WKY. These results support the development of subclinical, but measurable changes to the heart and renal systems

of the SHRSP, which are subject to hypertension-mediated end organ damage as the animals age (X. Wang, Hao, et al., 2022).

At the neonatal time-point, in line with studies in WKY and SHRSP strains, there were no significant differences in relative cardiac weight between chromosome 14 congenic strains and their background or donor strains. The post-natal heart rapidly grows following the increase in haemodynamic load in the first 10–12 days after birth (F. Li et al., 1996), and is significantly affected by conditions such as inter-uterine growth restriction (Wadley et al., 2013). A study of post-natal growth in Sprague Dawley rats suggested heart to body weight ratio was not significantly increased from day 1 until day 4 (F. Li et al., 1996). The present study was designed such that we collected neonate tissue between 1 and 3 days post birth, however even after adjusting for differences in body weight, age remained significantly correlated with all organ weight measures. WKY and WKY.SPGla14a pups were collected from multiple litters however SHRSP and SP.WKYGla14a pups were collected from single litters such that we cannot assess the effect of post-natal age on cardiac structure. In the HHR model, differences in cardiomyocyte number and size is present at post-natal day 2 in the HHR versus the NHR, which underlies a significantly reduced cardiac weight index (Porrello et al., 2009). This post-natal restriction in cardiac growth is replaced by cardiac hypertrophy by 12-weeks of age (Harrap et al., 2002; Porrello et al., 2009; Bell et al., 2016; Marques et al., 2016). In the SHRSP model, reduced cardiac weight at neonatal time points is not commonly reported (versus WKY) and, although differences in litter size and number is a known consequence of hypertensive pregnancy in the SHRSP (Small et al., 2016), results presented are as expected, and are potentially affected by differences in post-natal age (1–3 days).

Contrary to experimental hypotheses, by 5- and 16-weeks, chromosome 14 congenic strains do not differ in cardiac structure or systolic function when compared to their background strain. This represents a potential drift in the phenotype of male chromosome 14 congenic animals from the initial phenotyping efforts. Following establishment of the congenic strains as inbred, a worsening of cardiac structure and diastolic function in the WKY.SPGla14a compared to its background WKY strain was found. Since these initial studies, a number of external and environmental factors have changed. Although no effect should be seen in established genetic models of CVD, changes in diet, housing location, and animal husbandry practices in the >10 years since congenic strain establishment, could have increased environmental pressures resulting in a drift in phenotype. At 5-weeks, the WKY.SPGla14a has a non-significant increase in LVMI compared to the WKY. Cardiac phenotypes also consistently opposed kidney phenotype, where kidney weight ratios indicated worse phenotypes in the SHRSP and SP.WKYGla14a. Although no significance between congenic strains and their parental and donor strains was found in the heart, the kidney data suggests some divergence in

cardiac phenotype has occurred in the congenic strains.

Molecular study of cardiac tissue through qRT-PCR showed a distinct and persistent upregulation of the *Spp1* gene in the SHRSP and WKY.SPGla14a, which is consistent with the microarray studies performed during the initial phenotyping. Contained within the chromosome 14 congenic region, cardiac *Spp1* is upregulated in SHRSP and WKY.SPGla14a across both sexes. Increased expression of *Spp1* relative to the WKY and SP.WKYGla14a was evident in post-natal cardiac development and continued into adulthood (5- and 16-weeks). The absence of increased *Spp1* in the SP.WKYGla14a demonstrates a degree of local control within the congenic region of *Spp1* expression. In a Japanese population, a polymorphism in the *Spp1* promoter region was associated with diastolic dysfunction in hypertensive patients without an observable difference in LVMI (Nakayama et al., 2011). Indeed, initial phenotyping in chromosome 14 congenic models used Pressure-Volume loop (PV loop) technology to demonstrate differences in diastolic function between chromosome 14 congenic strains and their background strain. These experiments were not possible in the relatively high numbers used herein. The Acuson Sequoia c512 ultrasound system did not have sufficient resolution to perform accurate measurements of mitral flow for an estimation of diastolic function using E/e' ratio. RWT, a measure sensitive to concentric modelling was variable, particularly at the 5-week time point, but did show a trend towards increased RWT in male and female WKY.SPGla14a vs the WKY. The VisualSonics Vevo F2 (FujiFilm) provided an improvement in ability to perform accurate ultrasound echocardiography, including mitral flow for the estimation of E/e' ratio. However, only n=6 animals were assessed using the improved ultrasound machine and could not power a more accurate study of diastolic function.

Remodelling of the myocardium during pathological hypertrophy includes the expansion of the extracellular matrix, increasing myocardial fibrosis (Mohamed and Mraiche, 2015) through fibroblast activation and increased collagen deposition (Mihl, Dassen, and Kuipers, 2008). On a macro level, differences in cardiac fibrosis in the left ventricle were assessed through chromogenic histological assessment, and despite differences in LVMI, there was no difference in left ventricle fibrosis between the WKY and SHRSP strains at 16-weeks. Although 16-weeks represents an 'adult' time point, it is relatively soon after maturity. Considering increased fibrosis often develops later in pathology, differences in parental and congenic strains at this macro-level might not become apparent until >20 weeks. Ang-II and surgical induction of pressure overload are used in a range of studies to induce cardiac remodelling, and in these cases increased expression of *Spp1* is marker of fibroblast activation (McLellan et al., 2020). Despite this more severe model of cardiac remodelling, authors reported a high variability in chromogenic straining which resulted in borderline significance of histological assessment of fibrosis (McLellan et al., 2020). Similarly, in a model of superimposed

hypertension and cardiac remodelling by aortic banding, *Spp1* deletion attenuated cardiac dysfunction and hypertrophy, without affecting quantitative analysis of fibrosis by trichrome-staining (Xie, M. Singh, and K. Singh, 2004b). Differences in *Spp1* expression at 16-weeks support a model of fibroblast activation in SHRSP hearts, however variability in picrosirus red staining may mask expected and potentially subtle increases in fibrosis at this time point.

One of the difficulties determining and defining the onset of LVH is the index to which LV mass is normalised to, such that LVMI is correctly adjusting for differences in body size both within and between sexes. In the clinic, there has been growing interest in the sex modified associations between LV mass and height, lean mass, and body surface area (Taylor et al., 2023). In a cohort of adolescent humans, lean mass explained much more of the variance in LV mass than height (Taylor et al., 2023). The use of tibial length (comparable to height in humans) or body weight for indexing LV mass in rats has been subject of much less investigation. In cases where body weight is altered by intervention, independent of potential cardiac remodelling, the use of tibia length might be more appropriate (Yin et al., 1982). The use of lean mass or approximations of lean mass derived from raising measures of height to a specific power, might better provide means of normalisation to more effectively understand differences in cardiac structure (Simone, Daniels, et al., 1992). Such measures have not been explored in animal models to date. In practice the ability to measure tibial length can be less accurate than that of body weight and as such, body weight was determined a more appropriate means of normalisation in this observational study.

A myriad of interacting genetic, internal, and external factors will result in measurable increases in the SHRSP LVMI compared to WKY. The genetic differences between the congenic strains and their backgrounds is greatly reduced compared to the difference between the two parental strains, which is a major benefit of congenic strain development. However, the introgressed region might not be sufficient to cause phenotypic change, without additional stressors to confer the increase in LVMI observed between parental strains. The observational focus on developmental time points aimed to reproduce the independent effect of the chromosome 14 region on cardiac structure but did not assess the possibility that the region potentially compounds responses to hypertrophic stimuli. Risk polymorphisms in *Spp1* promoter region exacerbated the reduced diastolic dysfunction observed in patients with diabetes (Nakayama et al., 2011). Plasma or tissue levels of *Spp1* mRNA or protein were not assessed in the study by Nakayama et al. (2011), however genetically determined increases in *Spp1* expression conferred by SHRSP chromosome 14 genome might contribute to a similar effect. Cardiac structure is a dynamic and continuous trait which adapts to stimuli according to the genetic background. A basal or background over-expression of *Spp1* could contribute to adverse responses to stimuli, such as an increase in blood pressure,

which occurs only in the SHRSP and not the WKY.SPGla14a. Similarly, higher *Spp1* expression is often associated with poorer prognosis and outcomes in cancer (Icer and Gezmen-Karadag, 2018; M. Zhao et al., 2020). It is not possible in these cases to determine basal expression of *Spp1*, that is, before the 'stressor' event (tumour development, diabetes etc.) occurs (Nakayama et al., 2011; Icer and Gezmen-Karadag, 2018; M. Zhao et al., 2020). Despite initial indications that congenic animals developed an increase in LVMI as observed in the SHRSP, it is possible that the region provides a genetic susceptibility to adverse remodelling, that does not become apparent until stressed by external stimuli such as high fat diet, or salt loading.

Genetic drift in inbred strains may develop from residual heterozygosity or fixation of *de novo* spontaneous mutations (Benavides et al., 2020). The QTL identified on chromosome 14 was associated with cardiac mass (J.S. Clark et al., 1996). In an unexpected finding, it has been observed that the SP.WKYGla14a strain experiences an increase in spontaneous stroke compared to the SHRSP. Measures of blood pressure and cardiac parameters at 16-weeks were often most extreme in these animals, compared to SHRSP, WKY, and WKY.SPGla14a strains. Such anecdotal evidence of an increased stroke susceptibility, particularly in male animals, has not been formally investigated, but the introduction of the chromosome 14 region from the WKY into these animals is potentially changing responses to genetically induced development of higher blood pressure. In the short term, this may reduce ability of SP.WKYGla14a animals to adapt to increased blood pressure. These hypotheses are yet to be investigated formally, however during this period of study, mesenteric arteries were isolated for RNA extraction from a subset of all strains at 16-weeks. The complex genetic models generated by congenic strains are a useful translational model for human disease, when specific phenotypes are discernable. Any protective spontaneous mutations which may have arisen in the WKY.SPGla14a should be investigated through genomic sequencing analyses. Alternatively, studies applying external stressors to the model might exploit a genetic predisposition to maladaptive remodelling which mirrors human development of LVH in non-hypertensive and hypertensive human populations. The UK biobank and other similar repositories of rich population data could also be explored to determine risks associated with *Spp1* over-expression, before clinical presentation of pathological phenotypes (Backman et al., 2021).

Changes in external environment and methods of analysis, as well as genetic drift through spontaneous mutations may all have contributed to differences in chromosome 14 congenic strains between their establishment and current study. That being said, their usefulness as a model of genetic risk of LVH remains. Despite no discernable differences in cardiac phenotypes of chromosome 14 congenic and background strains, molecular differences in *Spp1* represents an interesting target for future investigations. *Spp1* is heavily researched in the field of cancer and end-stage cardiac disease and

any differences in expression seemingly driven by the chromosome 14 region warrant investigation as a mechanistic target in the development of LVH.

# Chapter 4

# Multiplex Immunohistochemistry and Single Nuclei RNA-Sequencing to Localise Cardiac *Spp1* Expression

## 4.1 Introduction

Complex tissues like the heart are made up of a number of different and interacting cell types, each performing a specific set of functions. The cellular composition of the foetal and developing heart are different to that of the adult heart in both the defined cell-'marker' genes and the compositional make-up (Anto Michel et al., 2022). Cardiomyocytes form a greater proportion of the total cell population in foetal hearts, and potentially exist in proliferative and non-proliferative states (Suryawanshi et al., 2020). Cardiomyocytes of the foetal heart do not cluster in chamber specific subtypes as they do in the adult heart (Litvinukova et al., 2020; Suryawanshi et al., 2020) and compared to adult hearts, foetal hearts express genes involved in cell cycle regulation, chromatin organisation, and matrix organisation at a higher level (Berg et al., 2015). During cardiac development *in utero*, cardiomyocyte populations are the predominant cell type. During early post-natal growth (days 1–3), the number of cardiomyocytes and fibroblasts increases through hyperplastic growth, rapidly increasing cell numbers. Expansion of fibroblasts during post-natal growth supports the need for greater structural integrity within the myocardium. By 3-weeks of age, the relative mitochondrial content and volume of cardiomyocytes resembles that of the adult myocardium, all hyperplastic growth of cardiomyocytes is replaced with hypertrophic growth, and the number of binucleated cells reaches around 90% (of cardiomyocytes). After reaching sexual maturity (∼8.5 weeks) up to 96% of cardiomyocytes are binucleated and relative cell populations stabilises at approximately 64% fibroblast, 30% myocyte and 6% 'other' (Banerjee et al., 2007).

Just as bulk RNA sequencing has revealed the tissue-specific nature of gene expression, single cell and spatial transcriptomics have characterised the fine-tuned specificity of transcriptional activity in complex biological systems. Spatial transcriptomics, immunohistochemistry assays, and single-cell RNA sequencing rely on 'gold-standard' or commonly reported marker genes and proteins to annotate cell types or differentiate between adult and foetal-like cells. Markers identified through immunohistochemistry or following unsupervised clustering of transcriptomic data have demonstrated that although exact proportions of cell types and markers differ between human and rodents, there is a high degree of functional overlap defining cardiac cell populations (Figure 4.1).

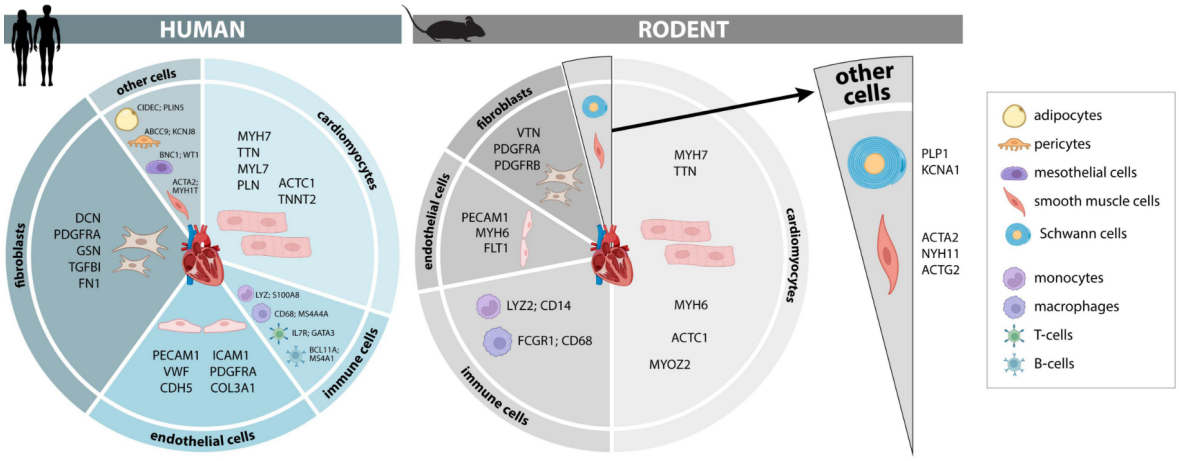



**Figure 4.1: Cells of the Adult Rodent and Human Heart** Image taken from (Anto Michel et al., 2022). Major cell lineages identified in the human and rodent adult heart including consistently identified markers used for identification of cell types in cardiac tissue. MYL; myosin light chain, MHC; myosin heavy chain, TNNT; troponins, PLN; phospholamban, PECAM1; platelet and endothelial cell adhesion molecule 1

*Spp1* expression is inducible in cardiomyocytes by a number of factors including; aldosterone, dexamethasone, endothelin-1, norepinephrine, and reactive oxygen species (K. Singh, Balligand, et al., 1995; Graf et al., 1997; Jiang et al., 2017; C.M. Pollard et al., 2019). Increased *Spp1* is not observed in response to stimulation of cardiomyocytes with Ang-II, interleukin-1β, and interferon-γ (K. Singh, Balligand, et al., 1995; Xie, Pimental, et al., 2001; Xie, M. Singh, and K. Singh, 2004a), all of which have been shown to increase *Spp1* expression in other cell types. Treatment with purified OPN protein is associated with increased apoptosis and endoplasmic reticulum (ER) stress in primary rat ventricular cardiomyocytes (Dalal et al., 2014) and *Spp1* is a marker of transition towards a pro-inflammatory state in fibroblasts (Garvin et al., 2021). Aged *Spp1* deficient mice demonstrate a lack of fibroblast activation (Sawaki et al., 2018) and cardiac fibroblasts isolated from *Spp1* knock-out mice also show reduced adhesion to ECM substrates and reduced differentiation into myofibroblasts following TGF-β stimulation (A.R. Collins et al., 2004; Lenga et al., 2008). Available evidence supports that *Spp1* enhances survival and proliferation of fibroblasts, promoting a

pro-pro-inflammatory and pro-fibrotic phenotype.

Some growth factors and inflammatory mediators have also been shown to increase *Spp1* expression in endothelial cells (K. Singh, Balligand, et al., 1995; Xie, Pimental, et al., 2001; Xie, M. Singh, and K. Singh, 2004a). Contrasting the apoptotic and pro-fibrotic effects *Spp1* produces in myocytes and fibroblasts, *Spp1* activation of the PI3K/Akt pathway in endothelial cells induced angiogenesis and stimulated production of VEGF to support angiogenesis (Dai et al., 2009). The potentially beneficial effects of *Spp1* in the vasculature is further supported by the reduced angiogenic response in *Spp1* knock-out mice following myocardial infarction (X. Zhao et al., 2007). *Spp1* is also thought to play an important role in the immune response (Mamazhakypov et al., 2022), with specific expression in activated and differentiated macrophages (Krause et al., 1996). The role of *Spp1* in macrophages appears to be as an important mediator of appropriate response to cardiac injury (Shirakawa et al., 2018). Following myocardial infarction, *Spp1* is produced by macrophages in the infarct peripheral zone during the inflammatory response to ischaemia-reperfusion (Komatsubara et al., 2003). The importance of immune cell expression of *Spp1* in hypertrophic response not following ischaemia-reperfusion injury is less well characterised (Mamazhakypov et al., 2022).

Given the contrasting roles of *Spp1* when expressed in different cell types, and the ability of various stimuli to induce *Spp1* expression, localising *Spp1* expression in the cardiac tissue of genetic models discussed herein could help to define activated pathological pathways in genetically-mediated hypertrophy. Tracking the expression of *Spp1* through developmental stages, from foetal and neonatal development to adulthood, could determine differences in where *Spp1* is produced between genetic models. Multiplex immunohistochemistry (mIHC) is a method of detecting multiple markers of interest on a single slide/sample using the principles of antibody detection, amplification, and visualisation using fluorescence. The basic principle of a multiplex assay relies on ability of different fluorophores to absorb and emit photons at specific wavelengths of light. Spectrally distinct fluorophores can be attached to antigens directly or indirectly, either by attaching them to primary antibodies (direct labelling) or as a secondary reagent following signal amplification and polymer-horseradish peroxidase (HRP) detection (indirect labelling). Traditional approaches to multiple antigen detection was limited by the need to ensure each antibody is raised in a different species, constraining analyses to those in which suitable antibodies could be found in multiple, distinct species. The OPAL system (Akoya Biosciences, Marlborough, USA) utilises tyramide signal amplification (TSA) to allow sequential rounds of antibody incubation and detection with different fluorophores as an indirect labelling technique. TSA antibodies carrying HRP as a secondary agent catalyse the activation of dye tyramides to covalently attach to nearby tyrosine residues. After labelling is complete, primary and secondary antibodies are removed by thermal or chemical treatment leaving the tyro-

sine bound fluorophores intact, and allowing for the next target to be detected without antibody cross-reactivity (Figure 4.2). This also means that multiple antibodies raised in the same species can be applied one after the other, with different coloured dye tyramides. The main advantage of mIHC is added dimensionality of acquired data. Not only can the presence and abundance of target proteins be quantified, but the spatial distribution can be characterised in relation to markers of cell type or damage, as well as co-expression or co-localisation (Stack et al., 2014).

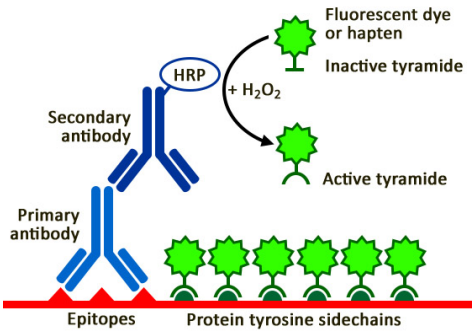



**Figure 4.2: Tyramide Signal Amplification** HRP-conjugated antibodies convert inactive tyramide to active tyramide, which then covalently attaches to nearby tyrosine residues. Image taken from Biotium `https://biotium.com/blog/seeing-more-and-brighter-colors-with-tyramide-signal-amplification/`

### 4.1.1 Specific Hypothesis and Aims

As our ability to dissect the cellular milieu of complex tissues has evolved, the importance of understanding contribution of specific cell populations has become apparent. We hypothesised the role of *Spp1* in adverse cardiac development is driven by its precise cellular location in the cardiac muscle. To explore this hypothesis, we used a hypothesis-free single nuclei RNA-sequencing experiment alongside mIHC to determine the location and quantity of *Spp1* expression in rat cardiac tissue at gestational day 18.5 (*in utero*), neonate (days 1–3), and 16-weeks (adult).

Specifically, the work in this Chapter aimed to:

- Optimise a mIHC assay in Formalin-fixed paraffin embedded (FFPE) LV tissue from WKY, SHRSP, and chromosome 14 congenic rats
- Use multiplex immunofluorescence to localise *Spp1* in relation to vascular, endothelial, fibroblast, and cardiac myocyte cells
- Isolate single nuclei from GD18.5 and adult flash frozen cardiac tissue for single-nuclei RNA-sequencing

## 4.2 Specific Methods

Immunohistochemistry experiments were undertaken in collaboration with the Glasgow Tissue Research Facility (GTRF) housed at the Queen Elizabeth University Hospital in Glasgow. Tissue sectioning was performed by Dr Jennifer Hay and Dr Hannah Morgan. Freshly isolated nuclei were sent to Glasgow Polyomics for single nuclei RNA sequencing.

### 4.2.1 Tissue Collection and Sample Preparation

The apex of the LV was collected during cull of male and female 16-week old rats as described (Section 2.1.3). Following removal and dissection of the heart, the apex of the LV was cut away using a sharp scalpel and fixed for 24h in 10% formalin. The tissue was washed and transferred to 70%EtOH and stored in -4°C until use. Collection of foetal tissue was performed primarily for RNA sequencing experiments described in Chapter 7 and is described in Section 2.1.4. Foetal hearts were extracted, cleaned, and immediately snap frozen in liquid nitrogen.

### 4.2.2 Single Nuclei RNA Sequencing

The age, strain, and tissue used in nuclei isolation is shown in Table 4.1. Nuclei were prepared using the 10x Chromium Nuclei Isolation kit (10x Genomics, California) with RNase inhibitor and prepared following manufacturers recommendations. Flash frozen tissue was removed from storage in -80°C, transferred into a fresh pre-chilled eppendorf tube, and weighed. Whole GD18.5 hearts could be processed however adult LV tissue required splitting in liquid nitrogen so that 3-50mg of tissue was added to 200µL of lysis buffer. Tissue was disrupted in 2mL eppendorfs containing lysis buffer using a sterile pestle. An additional 300µL of lysis buffer was added and samples were incubated on ice for 10 minutes. Dissociated tissue was pipetted onto the Chromium nuclei isolation column (10x Genomics, California) and centrifuged at 16,000x$g$ for 20 seconds at 4°C. The nuclei suspension in the collection tube was retained for further purification and the column was discarded. Nuclear suspension was vortexed briefly and centrifuged at 500x$g$, for 3 minutes (4°C) to pellet the nuclei. Supernatant was removed and the nuclei were resuspended in debris removal buffer. The suspension was centrifuged for 10 minutes at 700x$g$ at 4°C and the supernatant was removed. The nuclei pellet was washed with 1mL of wash and resuspension buffer (containing 10% ultrapure BSA) and centrifuged at 500x$g$ for 5 min at 4°C to re-form the pellet of nuclei. The wash step was repeated, and the final pellet was resuspended in 100µL of resuspension buffer. To determine concentration, 10µL of nuclei suspension was mixed with trypan blue and added to a standard hemocytometer. Nuclei were counted using a 20x light microscope, where intact nuclei were distinguished from debris by their spherical shape. Suspension volumes were adjusted to achieve a final concentration of

~300 nuclei/µL and transported to Glasgow Polyomics to be loaded on a Chromium Controller (10x Genomics) with a targeted nuclei recovery of 5,000 nuclei per reaction.

**Table 4.1:** Tissue Utilised in Single-Nuclear RNA-Sequencing

| Age | Strain | Sex | Tissue |
|---|---|---|---|
| GD18.5 | WKY/Gla | Unknown | Whole heart |
| GD18.5 | SHRSP/Gla | Unknown | Whole heart |
| 16-weeks | WKY/Gla | Male | Left ventricle |
| 16-weeks | WKY.SPGla14a | Male | Left ventricle |

### 4.2.3 Tissue Preparation for Immunohistochemistry

Prior to processing, tissues were de-identified to ensure investigator blinding during histological processing and analysis. Fixed tissues were embedded in paraffin and labelled with a randomly generated, unidentifiable ID number by lab technicians at the University of Glasgow. For use in development and optimisation of the multiplex assay, hearts were cut in 4µm thick sections and mounted on slides by collaborators at the GTRF. Sections were baked overnight at 60°C and then stored in 4°C until use.

### 4.2.4 Immunofluorescence

For each antibody, a cycle of antigen retrieval (AR) and removal, primary antibody incubation, and tyramide signal amplification fluorescent staining was performed as described in Section 2.7.3. Osteopontin was probed in a round of single immunofluorescence for an N=3 males and females per strain. The same steps were followed as for multiplex IHC, and Opal520 was used as the secondary reagent.

For rounds performed in an automated fashion, slides, and reagents were loaded onto the Lab Vision S480S (Thermo Scientific, UK). To reduce and block non-specific staining, slides were incubated for 10 minutes in UltraVision hydrogen peroxide ($H_2O_2$, ThermoScientific) followed by UltraVision protein block (ThermoScientific). Slides were washed with TBS-T and then incubated in primary antibody for 30 minutes at room temperature. After washing, signal was amplified (UltraVision Amplifier) and secondary antibody HRP (UltraVision Polymer HRP) was applied. Finally, the OPAL fluorophore was added and slides were incubated for a further 10 minutes. Final washing steps were performed before AR for subsequent antibody cycles continued (if required). After the final round of staining, slides were counter-stained with nuclear DAPI and mounted with 50µL antifade mountant (ProLong$^{TM}$ Diamond, Invitrogen, Paisley, UK). An overview of the final AR conditions, antibody concentration, and corresponding OPAL fluorophore is displayed in Table 4.2.

**Table 4.2:** Antibody Conditions in Multiplex IHC

| **Order & Opal Fluorophore** | **Antibody** | **Antigen Retrival pH** | **Working Concentration** | **Company** |
|---|---|---|---|---|
| 1: Opal480 | $\alpha$-SMA | pH6 | 1:6000 | Abcam (ab124964) |
| 2: Opal570 | MYH7 | ph9 | 1µg/mL | Developmental Studies Hybridoma Bank (MF-20) |
| 3: Opal650 | OPN | ph6 | 1:1600 | Abcam (ab63856) |
| 4: Opal780 | VIM | ph6 | 1:1600 | Leica (NCL-L-VIM-V9) |

### 4.2.5 Optimisation of Antibodies for Multiplex Immunofluorescence

A number of factors need to be considered in construction of multiplex assays. First optimal AR conditions (buffer pH6 or pH9) and concentration of each individual antibody should be determined through antibody titration. To reduce cost, initial optimisation experiments were performed using DAB (3,3'-diaminobenzidine). DAB immunohistochemistry works on the principal that DAB is oxidised in the presence of hydrogen peroxidase (HRP) and forms a brown, alcohol insoluble precipitate at the site of enzymatic activity. The benefit of DAB staining over fluorescence during optimisation is that is can be viewed under a bright-field microscope for easy visual assessment. Once optimal conditions were determined, single immunofluorescence assay was performed at optimal and reduced concentrations to account for the increased sensitivity of fluorescent assays. Finally, to maintain sensitivity and specificity of subsequent antibodies in multiplex assays, the ability of AR steps to effectively remove bound primary antibodies both unconjugated and conjugated to secondary HRP was validated in a series of antigen 'stripping' tests. Workflow of multiplex optimisation is represented in Figure 4.3.

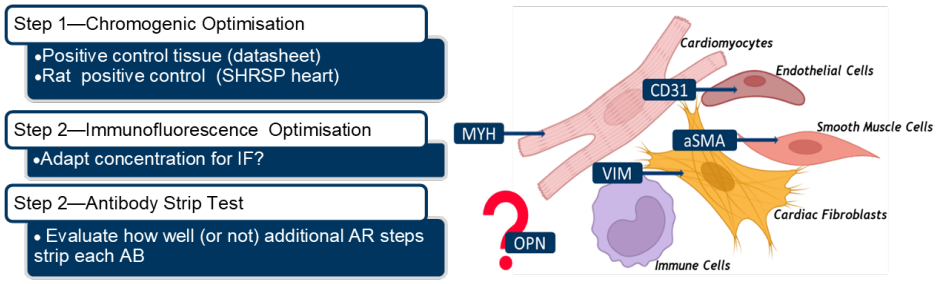



**Figure 4.3: Multiplex Optimisation** (left) Steps of multiplex immunofluorescence optimisation and (right) schematic depiction of multiplex experiment where cell types and respective markers are defined. Cellular location of osteopontin is unknown in rat heart and co-localisation of signal with cellular markers will be used infer osteopontin deposition. MYH; myosin heavy chain, VIM; vimentin, $\alpha$-SMA; alpha smooth muscle actin, CD31; cluster of differentiation 31 also known as platelet endothelial adhesion molecule (PECAM-1/CD31), OPN; osteopontin

Antibody titration was performed using DAB chromogenic assay. Each antibody was tested at 4 concentrations following AR in pH6 and pH9. Slides were visually assessed and optimal AR condition was selected based on DAB intensity and specificity. Titration was performed on positive control tissue provided by GTRF based on antibody product data-sheet. Selected concentration and AR condition was then performed in rat positive control LV tissue (SHRSP dam obtained in a previous study). Steps followed those detailed in Section 2.7.3 up to and including incubation with HRP-secondary reagent. Following amplification and application of secondary-HRP reagent, slides were incubated in UltraVision DAB reagent and counter-stained for nuclei using haematoxylin.

Each antibody was then tested at determined optimal concentration and 1 factor dilution in a single immunofluorescence assay using an OPAL reactive in the FITC channel (475 to 650 nm, peak 525nm). Following determination of optimal IF conditions, strip test was performed using an additional 2, or additional 4 AR steps. Steps of IHC were followed up to the application of OPAL reagent, where reagent was not applied until additional AR steps were completed. Secondary reagent was then applied under normal conditions (10 minutes incubation) and slides were counter-stained with DAPI, mounted with a coverslip, and imaged.

Multiplex order was determined based on ability of AR steps to effectively strip primary AB. Finally, a test was carried out where the order of primary antibodies was rotated to determine masking effects of primary antibodies before final multiplex order was defined.

### 4.2.6 Microscopy, Image Analysis and Quantification

All fluorescent slides were scanned by collaborators at the Glasgow Tissue Research Facility (Section 2.7.4). Single and multiplex images were analysed using the open-

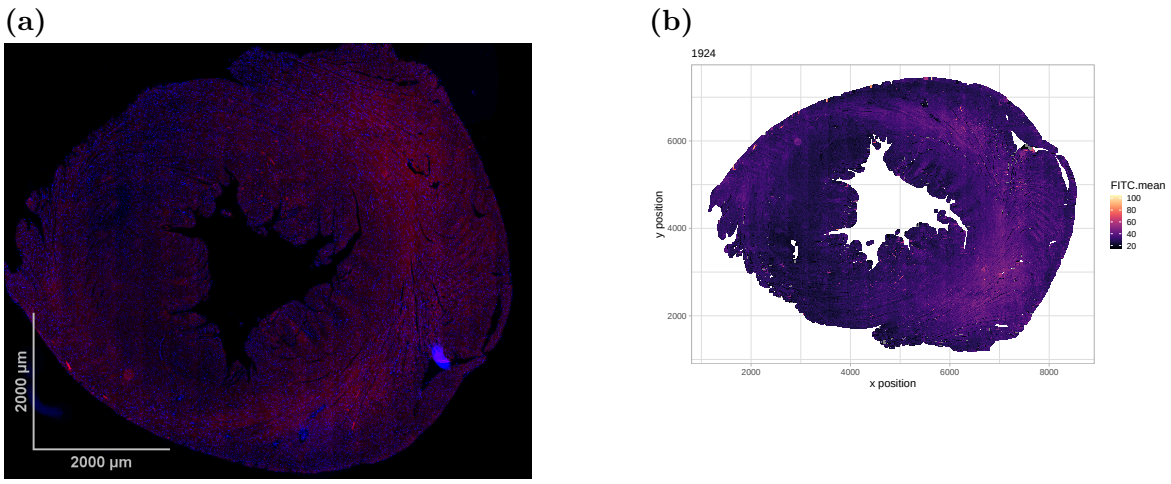



**Figure 4.4: Representative Image of QuPath and R Analysis** (a) Full scan of LV stained with OPN and DAPI (FITC and DAPI channels) (b) graphical representation of mean FITC intensity across tissue detected by QuPath using pixel threshold, retraction of edges, and division into tiles.

source software QuPath (version 0.4.2). Images were processes using '.groovy' scripts written for the purpose of these analyses (Appendix B). Tissue detection was performed using a pixel classification method based on DAPI and fluorescent channels. The thresholder tool was applied at a resolution of 3.54 pixels/µm using gaussian pre-filter and smoothing sigma of 3. Threshold-based algorithm was applied to define 'tissue' from 'background' using a threshold value of 12. Detected tissue regions were then eroded -20µm to retract boundaries away from the edges of tissue to both remove edge effects and areas where tissue was ripped. Tissue region (ROI; region of interest) was then divided into 10µm tiles which were trimmed to fit ROI.

In single-plex analyses, from all tiles, measurement of mean, standard deviation, minium, amd maximum intensity was recorded from FITC (OPAL 520) and DAPI channels at a resolution of 0.5µm per pixel. Resolution was chosen to match scanning resolution and X and Y coordinates of each tile was extracted along with intensity features. Data was exported from QuPath as a .csv file and analysed in R. All detections (tiles) were classified as 'background', 'positive' or 'artefact' based on mean FITC signal. Thresholds were set as $\leq 23$, $>23 \leq 90$ and $>90$ respectively, based on assessment of blinded images with clear examples of background, positive stain and/or artefacts caused by tissue folding, tearing, or hyper-intensities.

Images were reconstructed based on mean FITC intensity per tile using X and Y values in R, example of image and reconstructions is displayed in Figure 4.4. Due to the strong linear relationship between mean FITC intensity and mean DAPI intensity, FITC was normalised to DAPI using a simple ratio (FITC/DAPI) to generate arbitrary units of OPN fluorescence.

After optimisation at the GTRF using an auto-stainer, the pump system broke

down and full multiplex run was performed manually. This batch was affected by an edge effect where staining did not occur in the extremes of tissues. DAPI is used by the NanoZoomer S60 digital slide scanner (Hamamatsu Photonics, UK) to properly focus during scanning, and lack of uniform DAPI staining reduced scan quality. As a result analysis herein is preliminary. In multiplex analyses, following tissue detection and division into tiles, mean intensity of all 4 colour channels and DAPI was measured, and a threshold based classification of tiles was performed using cut-offs chosen by manual annotation of instances of 'true' positives. As a quality control measure to avoid detections of hyper-intensities from other channels, and account for the issues with DAPI as described, tiles with a mean DAPI below a threshold of detection was removed from analysis. Tiles were then classified as positive or negative for each marker, and data was exported as a .csv for downstream analyses.

## 4.3 Results

### 4.3.1 Troubleshooting Single Nuclei RNA Sequencing

CellRanger reported ∼10x more nuclei per sample than the expected 5,000 nuclei per sample (Figure 4.5). The barcode rank plot was also atypical, an ideal Barcode Rank plot has a distinctive shape, which is referred to as a 'cliff and knee'. This was not demonstrated, indicative of poor separation between the cell-associated barcodes and the barcodes associated with empty wells.

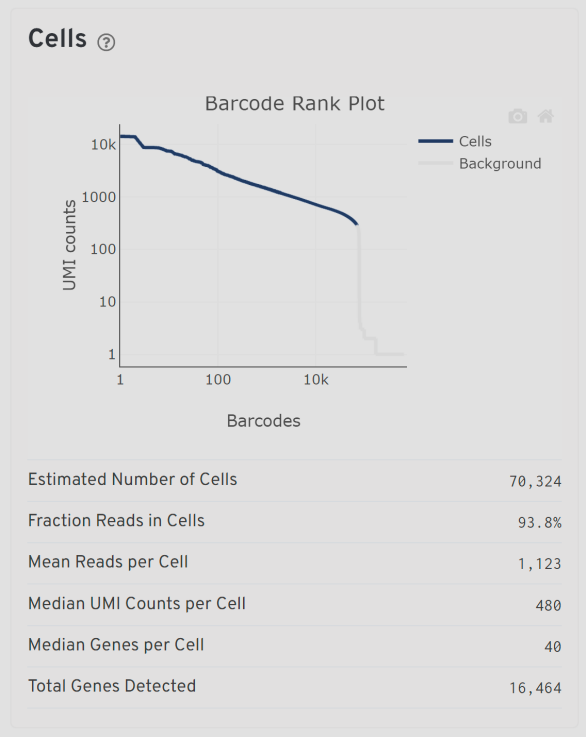



**Figure 4.5: Example of CellRanger Output in snRNA Sequencing** The output of CellRanger for one sample, Barcode Rank Plot; The plot shows the count of filtered UMIs mapped to each barcode. Barcodes are not determined to be cell-associated strictly based on their UMI count, the colour of the graph in these regions is based on the local density of barcodes that are cell-associated. Estimated Number of Cells; The number of barcodes associated with at least one cell. Fraction Reads in Cells; The fraction of valid-barcode, confidently-mapped reads with cell-associated barcodes. Mean Reads per Cell; The total number of sequenced reads divided by the number of barcodes associated with cell-containing partitions. Median UMI Counts per Cell; The median number of UMI counts per %s cell-associated barcode. Median Genes per Cell; The median number of genes detected per cell-associated barcode. Total Genes Detected; The number of genes with at least one UMI count in any cell.

Discussion with Glasgow Polyomics and 10x Genomics (California) established this could be due to either; wetting failure during library preparation and sequencing, high RNA background, or poor quality nuclei. A wetting failure occurs when there is improper emulsion formation and loss of single nuclei partitioning, causing multiple nuclei to be assigned the same UMI. This was ruled out as a cause after follow up with technicians confirmed that there were no indications at the QC level that wetting event had occurred. Single nuclei RNA sequencing is commonly subject to contamination by high amounts of ambient RNA which can be reduced using the CellBender software to assign single cell identity to UMIs. However, CellBender is a complex and resource intensive software. Data was not made available for inclusion herein.

### 4.3.2 Optimisation of IHC Workflow

Chromogenic staining confirmed the sensitivity and specificity of antibodies directed to myosin heavy chain (Myh), α-smooth muscle actin (α-SMA), vimentin (VIM) and

osteopontin (OPN) (Figure 4.6). CD31, a marker of endothelial cell origin, did not demonstrate specific staining in rat tissue and was removed from multiplex panel. Single immunofluorescence assay was performed to determine suitability of individual antibodies for multiplex analysis.

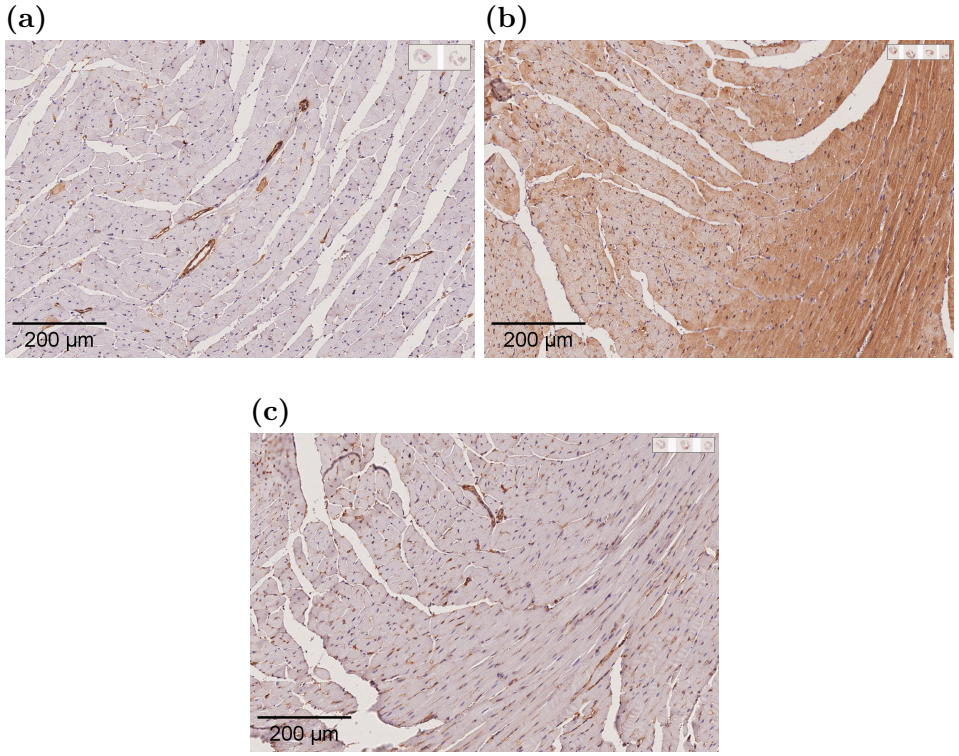



**Figure 4.6: Exemplar Images of Chromogenic IHC Optimisation** Examples of chromogenic DAB staining in; (a) α-SMA, (b) osteopontin and (c) vimentin antibodies in rat LV tissue sections. Magnification; 10x, scale bar; 200µm.

Strip testing was then performed to determine efficiency of antigen retrieval to remove primary and secondary reagents between rounds of staining. Figure 4.7 shows an example in α-SMA and VIM before and after 4 rounds of antigen retrieval prior to incubation of tissue with OPAL fluorescent reagent. Figure 4.7 also demonstrates the autofluorescence of erythrocytes and cells of myocardium, where high background is created by auto-fluorescent properties of the cardiac tissue.

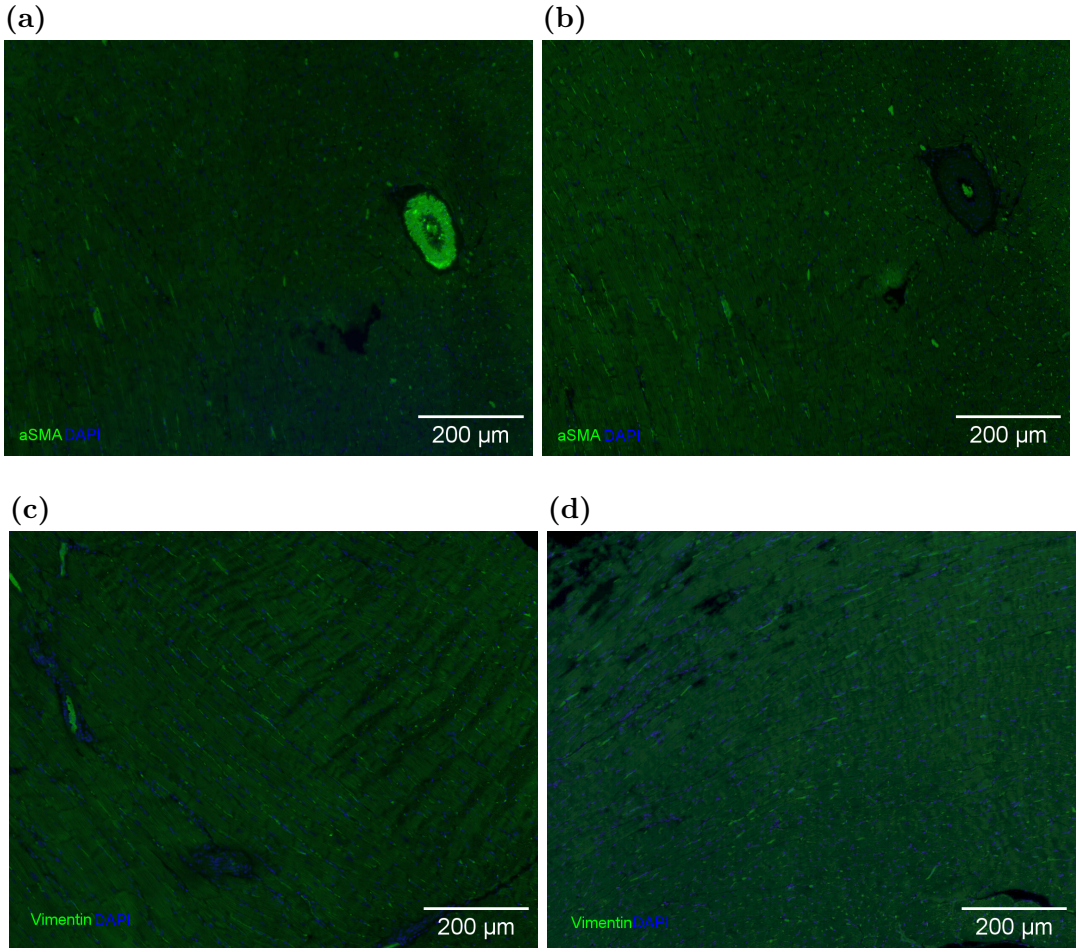



**Figure 4.7: Exemplar Images of Fluorescent IHC Optimisation** Examples of fluorescent IHC antibody test strip test where primary antibody was added to tissue, followed by amplifier and HRP secondary reagents. OPAL fluorophore was added either directly after HRP secondary reagent (a&c) or following 4 rounds of antigen retrieval and cooling (b&d) to replicate the conditions of multiplex IHC. Primary antibody directed against α-SMA (a&b) and Vimentin (c&d) are used in example to demonstrate an antibody which AR effectively stripped (a&b) and an antibody which was less effectively removed by AR cycles (c&d). Magnification; 10x, scale bar; 200µm.

### 4.3.3 Single Immunofluorescence of Osteopontin in the Adult Myocardium

Representative image of whole LV stained with OPN from WKY, SHRSP, WKY.SPGla14a and SP.WKYGla14a is shown in Figure 4.8.

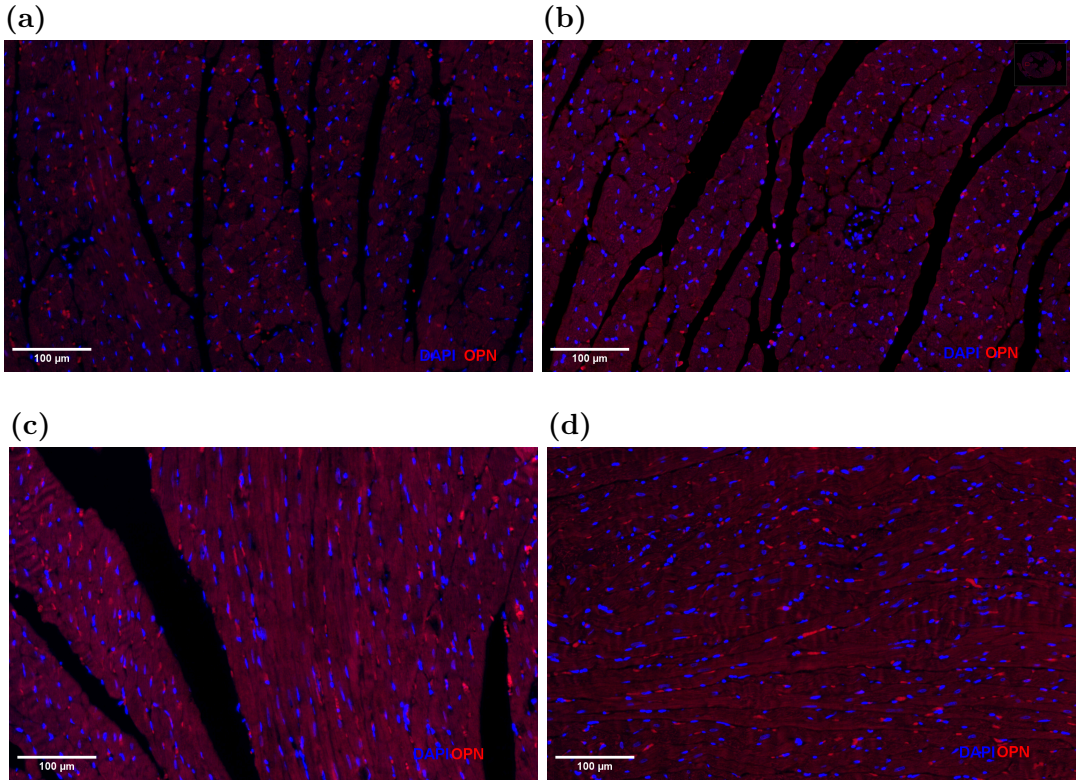



**Figure 4.8: Representative Image of Osteopontin Expression in 16-week Parental and Congenic Strains** Representative OPAL540 & DAPI stained sections using osteopontin primary antibody in LV of (a) WKY, (b) WKY.SPGla14a, (c) SHRSP, and (d) SP.WKYGla14a. Magnification; 20x, scale bar; 100µm

Female WKY and WKY.SPGla14a, and male SP.WKYGla14a did not have complete N numbers (N=3) and OPN fluorescence was compared statistically on sex-aggregated data (Figure 4.9). Quantification of mean intensity across the LV showed mean fluorescence was equal across male and female groups, and one-way ANOVA of mean OPN fluorescence by genotype indicated there was not a significant effect of strain. In both the male and female groups, there was a trend towards increased OPN intensity in WKY.SPGla14a and SHRSP LV compared to both the WKY and SP.WKYGla14a (Figure 4.9).

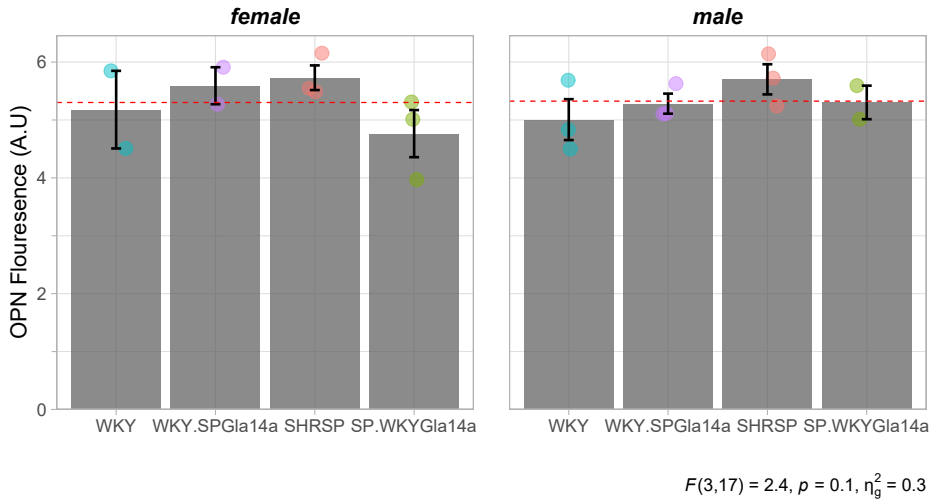



**Figure 4.9: Osteopontin Expression in LV of Parental and Congenic Strains** OPN intensity normalised to DAPI mean intensity across the whole of LV in WKY, SHRSP and chr14 congenic strains WKY.SPGla14a and SP.WKYGla14a. Red dotted line indicates female (left) and male (right) mean OPN fluorescence (A.U.). One-way ANOVA on sex aggregated data showed no significant effect of strain.

### 4.3.4 Multiplex Immunofluorescence in Adult and Neonate Myocardium

As visually demonstrated in Figure 4.10, an issue with DAPI staining was present in the final multiplex analysis meant outer region of tissue was unaccessible for analysis. DAPI staining is utilised by the microscope scanning system to appropriately focus and obtain fluorescent images. For analysis, any tiles in which DAPI could not be detected above a threshold were excluded from analyses.

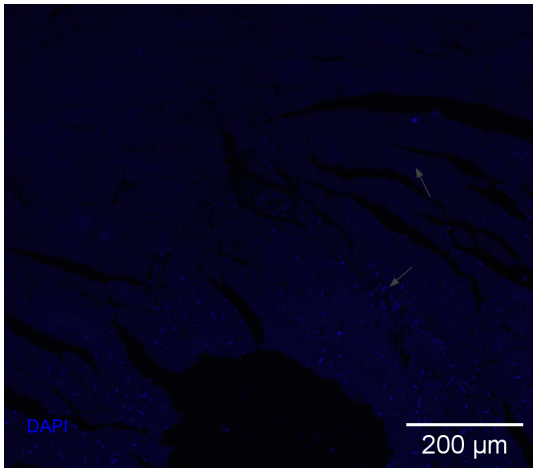



**Figure 4.10: Exemplar of Issue with DAPI Staining in LV Sections** Example of issue with DAPI staining, demonstrating a boundary where grey arrows indicate areas of relative good and poor DAPI resolution

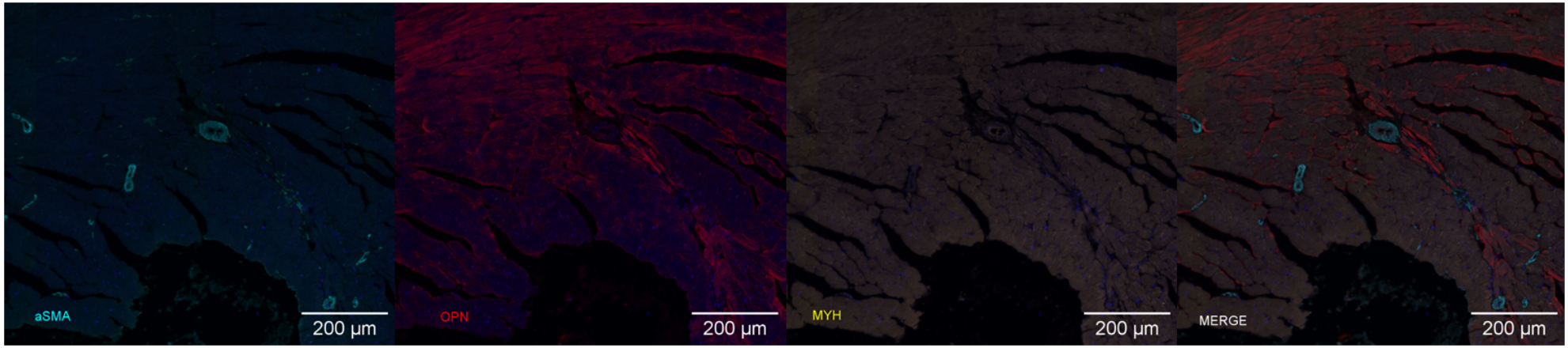



**Figure 4.11: Representative image of Multiplex IHC** Representative merged image of 3 channels in multiplex IHC (α-SMA, OPN and Myh).

An example of a merged, multiplex image showing positive OPN, α-SMA and Myh is shown in Figure 4.11. As there were 4 defined markers, and all combinations of markers were considered possible, there were 18 'classes' which tiles with adequate DAPI staining could be classified as. The relative proportions of each classification, for each strain is detailed in Table 4.3. $\chi^2$ test of association indicated a signifiant difference in relative proportion of each class between genotypes. As expected, few tiles were positive for all 4 markers, <1% in all genotypes, and most tiles were Myh positive with or without additional positive staining.

**Table 4.3:** Proportion (%) Tiles in Parental and Congenic LV Tissue Across All Potential Classifications

| **Class** | **Genotype (% tiles in class)** | | | |
|---|---|---|---|---|
| | WKY | WKY.SPGla14a | SHRSP | SP.WKYGla14a |
| CFP-aSMA | 1.20 | 1.10 | 1.90 | 1.80 |
| **CFP-aSMA: CY5-OPN** | <0.01 | 0.10 | 0.10 | 0.10 |
| CFP-aSMA: TRITC-Myh | 4.30 | 12.70 | 1.70 | 1.40 |
| **CFP-aSMA: TRITC-Myh: CY5-OPN** | 1.10 | 4.00 | 0.40 | 0.40 |
| **CY5-OPN** | 2.40 | 14.40 | 7.90 | 21.70 |
| CY7-VIM | 32.50 | 15.00 | 51.80 | 7.70 |
| CY7-VIM: CFP-aSMA | 0.90 | 0.10 | 0.40 | <0.01 |
| **CY7-VIM: CFP-aSMA: CY5-OPN** | <0.01 | <0.01 | <0.01 | <0.01 |
| CY7-VIM: CFP-aSMA: TRITC-Myh | 1.00 | 0.10 | 0.90 | 0.50 |
| **CY7-VIM: CFP-aSMA: TRITC-Myh: CY5-OPN** | 0.20 | 0.10 | 0.30 | 0.20 |
| **CY7-VIM: CY5-OPN** | 2.10 | <0.01 | 1.50 | 0.10 |
| CY7-VIM: TRITC-Myh | 3.70 | 0.30 | 6.30 | 0.90 |

| | | | | |
|---|---|---|---|---|
| **CY7-VIM: TRITC-Myh: CY5-OPN** | 0.90 | 0.70 | 0.50 | 0.20 |
| TRITC-Myh | 44.60 | 33.90 | 19.30 | 60.70 |
| **TRITC-Myh: CY5-OPN** | 5.20 | 17.50 | 6.90 | 4.40 |
| $\chi^2$ = 275382.7, d.f. = 42, p=2e-16 | | | | |

Bold typeface indicates classifications containing osteopontin (OPN).

In order to answer the specific research question, classifications positive for osteopontin were isolated, and proportion of classifications were compared between genotypes (Table 4.4). In WKY and WKY.SPGla14a, the highest proportion of OPN positive tiles were also Myh positive. SHRSP and SP.WKYGla14a tissues contained OPN positive tiles that were mostly not positive for another stain. In the WKY, OPN was also more often expressed with VIM than in SHRSP, WKY.SPGla14a and SP.WKYGla14a.

**Table 4.4:** Co-localisation of Osteopontin in Parental and Chromosome 14 Congenic Strains

| **Class** | **Genotype (% tiles in class)** | | | |
|---|---|---|---|---|
| | WKY | WKY.SPGla14a | SHRSP | SP.WKYGla14a |
| CY5-OPN | 20.20 | 39.30 | 44.80 | 80.50 |
| TRITC-Myh: CY5-OPN | 43.70 | 47.60 | 39.20 | 16.20 |
| CY7-VIM: CY5-OPN | 17.40 | <0.01 | 8.20 | 0.30 |
| CY7-VIM: TRITC-Myh: CY5-OPN | 7.60 | 1.80 | 2.80 | 0.70 |
| CFP-aSMA: TRITC-Myh: CY5-OPN | 9.00 | 10.80 | 2.50 | 1.40 |
| CY7-VIM: CFP-aSMA: TRITC-Myh: CY5-OPN | 1.60 | 0.20 | 1.50 | 0.70 |
| CFP-aSMA: CY5-OPN | 0.30 | 0.20 | 0.80 | 0.30 |
| CY7-VIM: CFP-aSMA: CY5-OPN | 0.30 | <0.01 | <0.01 | <0.01 |

## 4.4 Discussion

In a single-plex immunohistochemistry assay, LV tissue from male and female animals showed no signifiant change in OPN expression across the LV of parental and congenic strains. There was a non-significant trend toward increased OPN intensity in the WKY.SPGla14a and SHRSP compared to the WKY and SP.WKYGla14a. Limited time resulted in the study being underpowered to detect differences, either in sex aggregated or disaggregated images, and increased N numbers are required to determine if the observed trend was an artefact of data visualisation or representative of a true difference in OPN expression in the LV at the gross tissue level.

Analysis was affected by a high level of background fluorescence in the LV tissue. Erythrocytes (red blood cells) exhibit a broad spectrum of autofluorescence due to the polyphyrin ring in the heme group (Whittington and Wray, 2017). Nicotinamide adenine dinucleotide (NADH) and collagen, which are key components of most mammalian tissues and cells, also have emission spectra ∼450 and ∼ 300-450nm respectively (Georgakoudi et al., 2002). Other sources of autofluorescence include flavoproteins, lipofuscin, and elastin (Z. Zhang, H. Fan, et al., 2023). In highly metabolically active tissues, like the myocardium, there is an increase in NADH, which is an enzyme essential to metabolism. As a result, FFPE LV samples are subject to autofluorescence at 450-520nm wavelengths (Baschong, Suetterlin, and Laeng, 2001; Z. Zhang, H. Fan, et al., 2023).

The emission spectra of utilised antibodies in multiplex analysis is illustrated in Figure 4.12. FITC was channel was not used in multiplex analysis to avoid spectral overlap with CFP and TRITC, however its spectral peaks overlap with the autofluorescence spectra of NADH and erythrocytes (Bioquest, 2023; Z. Zhang, H. Fan, et al., 2023). Rather than an intensity based assessment of staining, employing the deep-learning capabilities of QuPath (Humphries, Maxwell, and Salto-Tellez, 2021) to determine a % positive tissue area, similar to picrosirius red analysis, might better be able to distinguish true positive areas of fluorescence from false positive areas caused by red blood cells or the natural fluorescence of cardiac tissues. The need for training images and time-cost of generating such a model was beyond the limits of this project however represents an opportunity to improve upon presented analysis strategy using current data. IHC images may also be improved by the addition of a dye to quench autofluorescence from the tissue. The application of Sudan Black B was attempted herein, without success and the data is not presented.

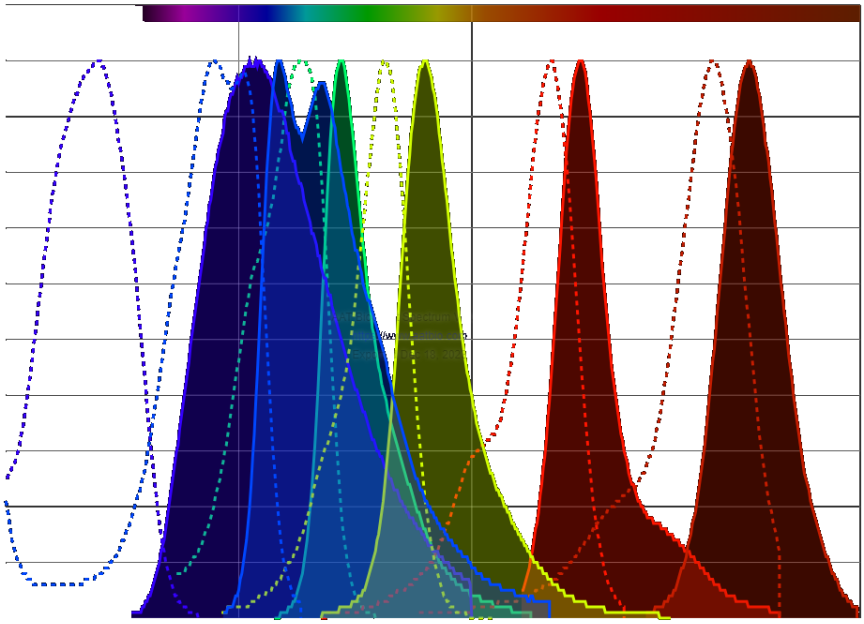


**Figure 4.12: Excitation-Emission Spectra of OPAL Dyes** Excitation spectra are shown by dotted lines, and emission spectra by solid lines and filled area plot. From left most to right most pair, the plot displays; DAPI, CFP, FITC, TRITC, Cy5, Cy7.

Despite strong foundation for multiplex assay, problems arising at the staining facility at GTRF impacted the ability to perform the multiplex to the same standards as the optimisation. The College of Medical, Veterinary, and Life Sciences at the University of Glasgow is currently undergoing centralisation of facilities which would allow the optimised multiplex assay to be performed in-house. Future study should aim to repeat multiplex IHC to confirm subcellular localisation of OPN within the LV. Initial analyses have suggested OPN primarily localises within myocytes in the 16-week heart.

Analysis of snRNA-sequencing will compliment multiplex IHC assay, and allow prediction of whether the source of *Spp1* transcription is the same as its destination once translated into its protein form. Due to the loss of single nuclei behaviour, library preparation and sequencing will be undertaken from freshly isolated cardiac nuclei from the tissues obtained during these investigations. Optimisation of nuclei characterisation prior to sequencing should include DAPI staining and fluorescence imaging of nuclei suspensions, as well as fluorescence-activated cell sorting (FACS) for nuclei recovery (Santos et al., 2021).

# Chapter 5

# *In Vitro* H9c2 Cell Assay of *Spp1* Function

## 5.1 Introduction

Whilst animal models provide a good genetic model of human diseases, identifying *de novo* pathways and mechanisms of function can be difficult in complex biological systems. This is coupled with an increased need to reduce the number of animals used in research by finding relevant cell and/or computational models. The use of functional assays in 2D and 3D cell cultures from immortalised cell lines both provides an opportunity to simplify the experimental system and reduce the need to sacrifice animals for primary cell isolations. These allow specific pathways to be implicated as targets in future investigations.

Given *Spp1* is consistently overexpressed in strains susceptible to cardiac hypertrophy and fibrosis (WKY.SPGla14a and SHRSP), investigating its molecular actions through a relevant cell system will help elucidate biological mechanisms that can be targeted later in translational or pre-clinical studies. Compared to other tissues, there are relatively few immortalised cell lines derived from the cells of the myocardium. Cardiomyocytes undergo terminal differentiation during neonatal development (Puente et al., 2014), a characteristic opposed by immortalised cell lines whose proliferative capacity is enhanced (Jimenez-Tellez and Greenway, 2019). Whilst immortalised types of fibroblast, smooth muscle, endothelial, and immune cell lines exist, none originate from the heart (Jimenez-Tellez and Greenway, 2019). Cardiac fibroblasts have a unique function due to their mechanical and electrical coupling to cardiomyocytes which is difficult to replicate in culture (Nicin et al., 2022). HL-1 cells are derived from mouse atrial cardiomyocytes and are phenotypically similar to embryonic atrial cardiac cells in culture (Claycomb et al., 1998). Alternatively, H9c2 cells are derived from embryonic ventricle tissue harvested from the BDIX/CrCrl rat strain (Kimes and Brandt, 1976). The morphology of H9c2 cells is akin to immature cardiomyocytes and are commonly

used in cardiac cytotoxicity assays (Witek et al., 2016). Crucially however, the ability of both HL-1 and H9c2 cells to divide distinguishes them from the myocardial cells they are employed to model, which have undergone terminal differentiation (Claycomb et al., 1998; Witek et al., 2016). That being said, when maintained suitable conditions, H9c2 cells retain some ability to be differentiated to resemble mature cardiac muscle cells (Witek et al., 2016).

Cells isolated from patient or animal tissue samples are also commonly applied in research. Primary cells arguably most faithfully represent the underlying biology of disease conditions as they are isolated from the tissue of interest. However, primary cells are difficult isolate and maintain, and have a finite lifespan, requiring replenishment from freshly obtained tissue (Jimenez-Tellez and Greenway, 2019). Additionally, as patient derived cells represent an already developed pathological state, mechanisms through which disease is caused might not be possible to elucidate from these cells. Induced pluripotent stem cells (iPSC) can be differentiated into 'mature' cardiomyocytes, however these cells are expensive, require additional, terminal differentiation steps, and do not fully resemble adult cardiomyocytes (Jimenez-Tellez and Greenway, 2019; Vuckovic et al., 2022). Given that cardiac hypertrophy primarily takes places in the left ventricle, and rat models have been employed in this research, the H9c2 line was determined the most physiologically relevant model, despite aforementioned differences compared to isolated primary cardiomyocytes.

Studies have suggested cardiac overexpression of *Spp1* protein dose-dependently increases fibroblast migration (Lorenzen et al., 2015) and represses ATP-linked oxygen consumption (Yousefi et al., 2019), indicating the diverse range of biological pathways through which increased *Spp1* expression could pathologically affect cardiac structure and function. Rat ventricular tissues isolated from renovascular hypertension and aortic banding showed an increase in *Spp1* expression in hypertrophied hearts, and treatment of primary cardiomyocytes with both endothelin-1 and norepinephrine induced a more than three-fold increase in *Spp1* expression (Graf et al., 1997). Another known stimulant of cardiac hypertrophy, AngII, increased expression of *Spp1* in primary rat cardiac fibroblasts (Ashizawa et al., 1996). In contrast, using the H9c2 model in this lab, stimulation with AngII increased cell size without increasing *Spp1* expression (data not published) suggesting *Spp1* has an ability to function independently of AngII in these cells. Stimulation of H9c2 cells with hypertrophic mineralocorticoid aldosterone also resulted in increased *Spp1* expression (C.M. Pollard et al., 2019). Patient samples derived from atrial septal defect demonstrated the *SPP1-CD44* ligand-receptor gene pair was significantly activated in patient versus control samples, which was prominent in immune cell populations (Z. Wang et al., 2021). In H9c2 cells, the pro-inflammatory role of endogenous H9c2 *Spp1* production was demonstrated through the ability of *Spp1* to inhibit the anti-fibrotic effects of catecholamine activation of the $\beta_2$-adrenergic re-

ceptor (C.M. Pollard et al., 2019). The secretome of fibrin activated bone marrow cells is highly enriched for the OPN protein and application of conditioned media induced H9c2 cell growth but not their proliferation (Borrego et al., 2022).

Hypertrophy of the left ventricle results from cardiomyocyte hypertrophy in response to specific external or internal stimuli. There has been growing interest in extracellular vesicles (EVs) as paracrine or autocrine messengers to modulate cell behaviour. These lipid-membrane bound vesicles contain cargoes of various bioactive molecules, including micro-RNAs, mRNA, protein, DNA, and lipids (Pathan et al., 2019). EVs can be released from cardiac cells including cardiomyocytes, fibroblasts, endothelial cells and resident immune cells (Fu et al., 2020). Depending on their size and formation, EVs are classified into three major subtypes (Figure 5.1) (Niel et al., 2022). The small extracellular vesicles, sometimes referred to as exosomes, can transfer proteins, nucleic acid and metabolites between cells (Claridge et al., 2021). In left ventricular hypertrophy and progression to heart failure, EVs have been investigated as disease biomarkers and diagnostic tools (Fu et al., 2020). Recently, studies using small extracellular vesicles (sEV) derived from obese and hypertensive patients demonstrated sEV induced a pro-hypertrophic, pro-fibrotic response in human induced pluripotent stem cell-derived cardiomyocytes (hiPSC-CMs) (Fandl et al., 2023). However, these sEV were plasma derived, and there is currently relatively little focus on sEV produced by the cells of the myocardium. In response to ischaemia in cell culture conditions, the size and number of EVs released by H9c2 and primary cardiomyocytes was altered, as well as the cargo they contained (Ribeiro-Rodrigues et al., 2017). Given their role in mediating a wealth of physiological and pathological processes (Fu et al., 2020), sEV produced by cardiac cells have the potential to mediate cardiac phenotypes, and it is therefore important to understand their role in more detail.

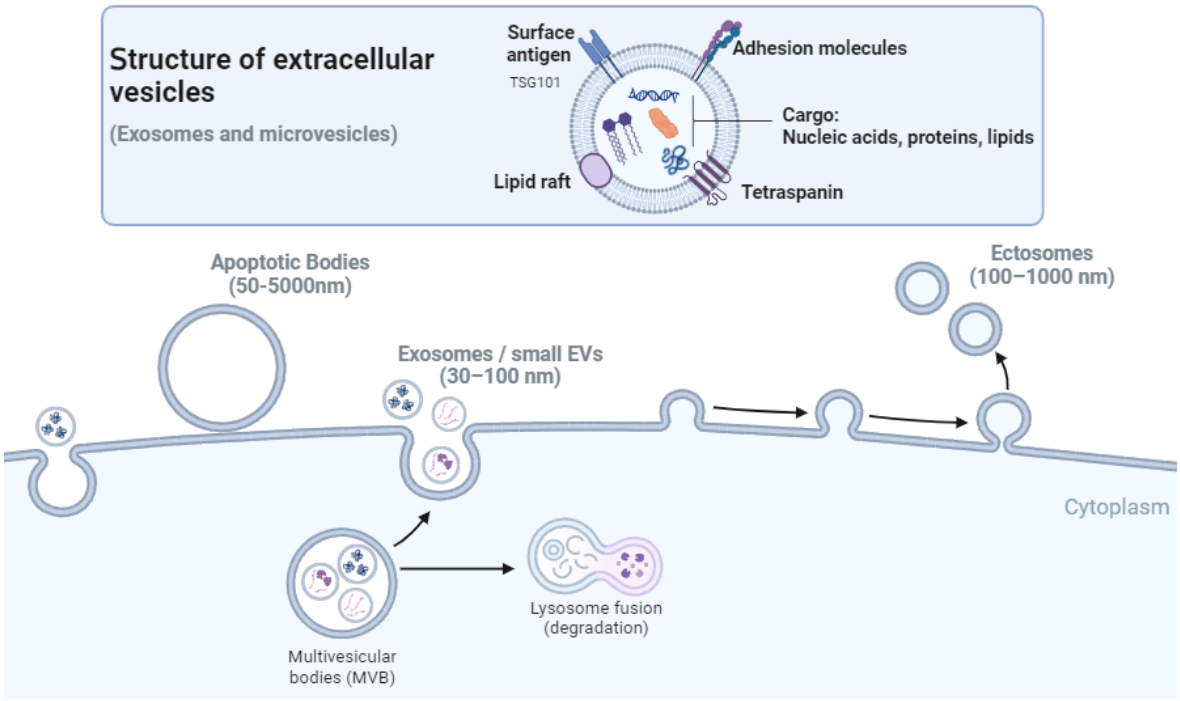



**Figure 5.1: Extracellular Vesicle Formation and Categorisation** Extracellular vesicles are distinguished by their size and formation. Apoptotic bodies are released by cells undergoing apoptosis formed by membrane blebbing into the extracellular space. Ectosomes (including micro-vesicles and oncosomes) are formed by plasma membrane budding and have an intermediary size to apoptotic bodies and small extracellular vesicles (sEV) or exosomes. The smallest vesicles can be formed through ESCRT (endosomal sorting complex required for transport)-dependent or ESCRT-independent pathways. ESCRT-pathways form multivesicular bodies (MVB) within the cell. Small EVs are released when MVBs fuse with the plasma membrane. The membrane of small EVs contains a number of specific proteins including tetraspanins (CD63, CD9, CD81), flotillin, TSG101, Alix and heat shock proteins (HSP70, HSP60, HSP5A). Microvesicle specific markers include integrins, selectins and CD40 (Konoshenko et al., 2018). Figure created with BioRender.com.

Previously, this lab used a traditional method of differential ultracentrifugation to isolate sEV from conditioned cell medium (Asirvatham, 2022). However, this method is limited by both the volume of medium that can be collected from cells and/or loaded in centrifugation tubes (max volume 17mL). These experiments cannot be scaled up economically. Other methods of isolation, such as size exclusion chromatography (SEC), isolate different populations of EVs and it was important to find a method of improving sEV yield, without changing the population of vesicles isolated. Density gradient isolation is the gold standard for sEV isolation by ultracentrifugation but is a time-consuming protocol requiring relatively higher skill and precision than other methodologies. As a more time efficient method of isolation, it has been reported that the quantity and quality of EVs isolated can be improved by applying a 30% sucrose 'cushion' to the bottom of ultracentrifugation tubes (Gupta et al., 2018; Duong et al., 2019).

## 5.2 Hypothesis and Aims

Over-expression of *Spp1* mRNA has a hypertrophic effect on H9c2 cells. Small extracellular vesicles mediate this effect through increased trafficking of excess *Spp1* to surrounding cells.

The studies presented aim to:

- Validate data showing over-expression of *Spp1* induces an increase in H9c2 cell size.
- Improve sEV isolation using an adapted method of isolation without changing isolated population characteristics
- Characterise sEV released by H9c2 cells overexpressing *Spp1* and determine their functional effect in the presence or absence of EV uptake inhibitor dynasore.

## 5.3 Specific Methods

### 5.3.1 Cellular Model of Over-expression

Cells were used in functional experiments between passage 5–19 to maintain a consistent phenotype (Witek et al., 2016). Biological replicates were repeated at various passages, as such that any variation induced by passage number will be consistent between groups.

H9c2 cells were seeded into 10cm tissue culture dishes or 6-well plates at a density of 10E+4 or 3–3.4E+4, respectively. Cells were plated with full maintenance media and allowed to adhere for 24 hours. Plasmids containing SHRSP-*Spp1* and control pcDNA were isolated from bacterial glycerol stocks as described in Section 2.4.3. Adherent, sub-confluent cells were transfected with plasmid DNA using Xfect$^{TM}$ Transfection Reagent (Takara Bio, Japan) at concentrations scaled as in Table 5.1.

**Table 5.1:** Concentrations to Scale Transfections

| Culture Dish | Growth Medium (mL) | DNA (ug) |
|---|---|---|
| 6-well plate | 2 | 5 |
| 10cm dish | 10 | 20 |

All reagents were stored on ice and prepared in a sterile environment. Plasmid DNA was mixed with Xfect reaction buffer to a final volume of 100µL. For every 1µg DNA plasmid, 0.3µL of Xfect Polymer was added to reaction buffer to a final volume

of 100µL. Both polymer and plasmid solutions were prepared in volumes suitable for 3 technical replicates. The polymer solution was then added to the plasmid solution and thoroughly mixed by vortex. The solution was incubated for 10 minutes at room temperature and 200µL was added to each well/plate. Following transfection, cells were returned to incubators at 37°C for 4 hours. Post-transfection, H9c2 cells were washed with PBS and fresh medium was added for 48 hours.

FBS is known to contain extracellular vesicles. To ensure collected EVs were produced by H9c2 cells, in experiments involving EVs, EV-depleted FBS was applied. EVs were depleted from FBS by subjecting HI-FBS to an 18-hour ultra-centrifugation at 120,000x*g*. The resulting supernatant was used to supplement cell culture medium and is referred to as EV-free medium in experiments where transfected cells were used for EV isolation.

### 5.3.2 Isolation of Small Extracellular Vesicles

H9c2 cells were transfected as described in Section 5.3.1. Four hours post-transfection, plates were washed with PBS and EV-free medium was added for 48-hours. sEV were isolated from conditioned cell medium collected from transfected H9c2 cells using a one-step Sucrose cushion ultracentrifugation (s-UC) method optimised from (Gupta et al., 2018). In brief, 20-30mL of media was collected and centrifuged 2000x*g* for 20-min to eliminate cellular debris and large contaminating particles. The supernatant was filtered using a 0.22µm filter to generate 'Clarified Conditioned Medium' (CCM) containing sEV.

CCM was subjected to centrifugation at 10,000x*g* for 30-min to pellet and remove microvesicles. The supernatant was carefully loaded over 2mL of a 30% sucrose solution and centrifuged at 100,000x*g* for 90 mins at 4°C using a SwTi32 rotor (Beckman-Coulter Optima L-80 XP). The sEV containing sucrose layer (approximately 3mL) was needle aspirated, re-suspended in PBS and washed by a final ultra-centrifugation at 100,000x*g* for 1 hour. The resulting sEV pellet was re-suspended in either PBS or RIPA buffer and stored in -80°C.

### 5.3.3 Optimisation of Small Extracellular Vesicle Isolation

The sucrose-cushion method was compared to standard methods of EV isolation using Differential ultracentrifugation (UC), as this method was previously used to implicate sEV in the cellular response to *Spp1* over-expression (Figure 5.2). Isolation of sEV by UC from CCM was achieved by three subsequent centrifugations at 100,000x*g* for 90-mins at 4°C using a SwTi32 rotor (Beckman-Coulter Optima L-80 XP). After each ultracentrifugation, the supernatant was removed and replaced with additional CCM or PBS. Following the final ultra-centrifugation step, the pellet was resuspended in

PBS and subjected to an additional 100,000x$g$ centrifugation for 60-mins to further concentrate and wash sEV. sEV were stored in -80°C until further use and characterisation.

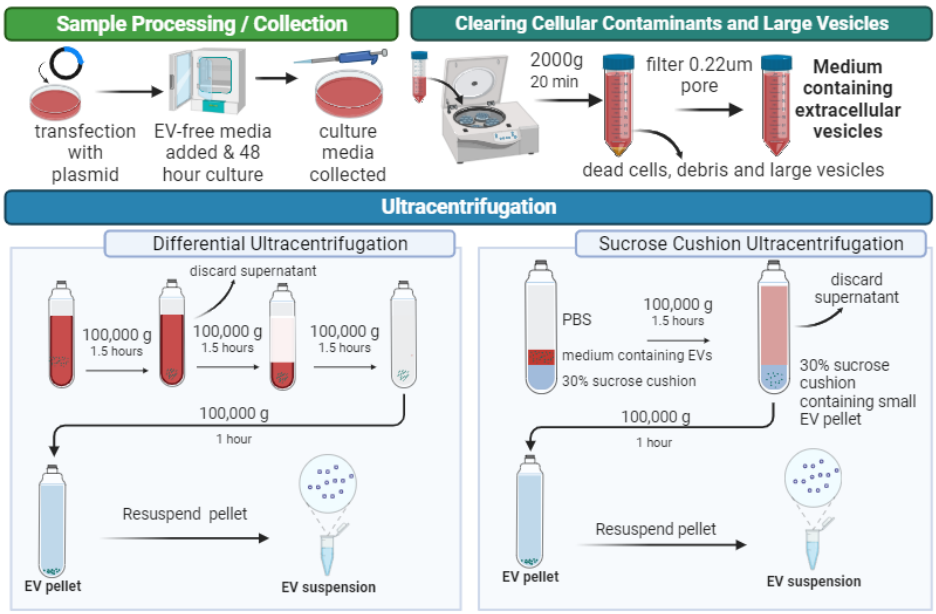



**Figure 5.2: Methods of EV Isolation** Infographic created in BioRender.com to demonstrate the two methods of ultracentrifugation utilised to isolate small extracellular vesicles released by H9c2 cells 48-hours following transient *Spp1* overexpression.

To compare the methods, 30mL of media from each transfection was split evenly (15mL) and sEV were isolated using either UC or s-UC. sEV size, concentration, and presence of marker proteins was assessed and compared using independent t-test.

### 5.3.4 Characterisation of Small Extracellular Vesicles

The size and concentration of isolated small EVs were characterised using Nanoparticle Tracking Analysis (NTA, NanoSight LM10, Malvern Panalytical). NTA visualises and measures nanoparticles in suspension using laser-illuminated microscopy. Suspensions containing nanoparticles are passed into a chamber and illuminated using a laser. The Brownian motion of particles is collected by a camera recording the movement of particles within the chamber. The tracking of particles is used to calculate a diffusion constant applied in the Stokes-Einstein equation to calculate hydrodynamic diameters.

Isolating a subpopulation of vesicles creates a negatively skewed distribution of particle size in NTA measurements (Figure 5.3). In such cases, the average particle size in a given sample is therefore better represented in measures of central tendency such as median (given as D50) and mode values. As an estimate of average particle size, D50 and mode values from each sample were extracted for between group comparisons. The particle concentration and distribution were extracted and corrected for the dilution factor applied to NTA measurements.

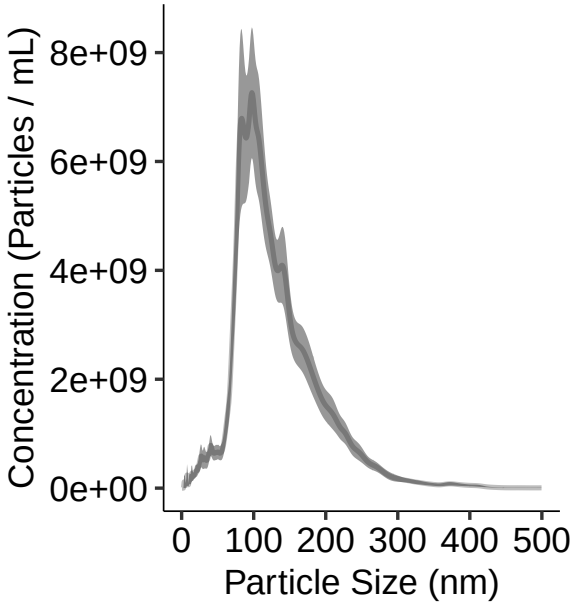



**Figure 5.3: Example of a Particle Distribution Curve Generated from Nanoparticle Tracking Analysis** Text data randomly generated to mimic particle distribution expected in EV samples isolated by ultracentrifugation

Expression of *Spp1* and EV marker proteins were analysed using standard western blotting techniques (Section 2.6). Equal volumes of sEV resuspended directly into RIPA buffer were thawed on ice, mixed with 4x SDS buffer containing reducing agent, and boiled at 95°C for 5 minutes. Proteins were separated by gel electrophoresis and then transferred to a nitrocellulose membrane. Normalisation of target protein was achieved using Revert 700 Total Protein Kit (LI-COR Biosciences, UK). Membranes were dried, rehydrated in MilliQ water, and submerged in REVERT$^{TM}$ stain for 5 minutes with gentle rocking. Membranes were washed and imaged immediately using the 700nm channel (Odessey CLx, LI-COR Biosciences, UK). Membranes were blocked in TBS blocking buffer for 1 hour at room temperature, with gentle rocking. sEV marker proteins, Tumor susceptibility gene 101 (TSG101) and A Disintegrin And Metalloproteinase Domain-containing protein 10 (ADAM10) were utilised as a primary antibody in EV experiments (Anti-TSG101, T5701, Sigma, Dorset, UK & ADAM10, 25900-1-AP, ProteinTech, Manchester, UK). Primary antibodies were diluted in blocking buffer and applied to membranes overnight at 4°C with gentle rocking. Following incubation with primary antibody solution, membranes were washed with TBS-T, incubated in secondary antibody solution (LI-COR IRDye 800CW, 1:20000) for 1 hour, and visualised using 800nm channel (Odessey CLx, LI-COR). Band intensities were quantified by ImageStudio software (LI-COR Biosciences, UK).

### 5.3.5 H9c2 Cell Sizing Assay

Following plating of H9c2 in 6-well plates, cells were assigned to either direct transfection or co-incubation conditions. H9c2 cells undergoing direct transfection were trans-

fected with plasmid containing pcDNA or SHRSP-*Spp1* as described (Section 5.3.1).

Alternatively, adherent H9c2 cells were co-incubated with one of; clarified conditioned media from transfected cells, or sEV from transfected H9c2 cells, for 48 hours. In experiments using sEV, EV-free maintenance media was added with equal volumes of relevant sEV suspensions.

After 48 hours, H9c2 cells were fixed in 4% paraformaldehyde (PFA) and stained for cell sizing using 0.5-1% crystal violet stain. Plates were visualised using an EVOS$^{TM}$ XL Core Cell Imaging System (ThermoFisher Scientific, UK) and cell sizing was carried out in ImageJ. Cell length was measured using the 'straight' line tool in ImageJ, along the long axis of each cell. 'Cell size' is used herein to describe this measure of H9c2 cell elongation. To avoid bias in sampling, a random sample of measurements was generated from the population of measured cells per biological replicate.

### 5.3.6 Inhibition of sEV Uptake by Dynasore

To reduce the ability of cells to uptake sEV, H9c2 cells were treated with dynasore and co-incubated with sEV as described in Section 5.3.5. Dynasore (60µM) was delivered in 1% DMSO-vehicle to H9c2 cells during co-incubation with sEV. Control cells were treated with 1% DMSO (vehicle) during co-incubation with sEV. Following 48 hours of co-incubation with sEV, in the presence or absence of dynasore, cells were fixed, stained, and imaged for cell sizing analysis (Section 5.3.5). One-well per biological replicate was used for RNA extraction (Section 2.3.2), cDNA generation and TaqMan qRT-PCR (Section 2.3.3, & Section 2.3.6). TaqMan qRT-PCR was performed to assess *Spp1* expression normalised to *B2m* housekeeper gene and analysed using the relative quantification method described in Section 2.3.6.

#### 5.3.6.1 Determination of Working Dynasore Concentration

Dynasore is a non-competitive, cell-permeable semi-carbazone compound inhibitor of Dynamin I and Dynamin II (sc-202592, Santa-Cruz Biotechnology, USA) utilised to block dynamin dependant, clatherin mediated endocytosis of sEV.

Cytotoxicity of dynasore was assessed using the MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) toxicity assay (Sigma, UK). The assay is based on the conversion of MTT to formazan crystals by living cells and is broadly assumed to represent number of metabolising cells. H9c2 cells were seeded into 6-well plates at a concentration of 3.0E+4. Serial dilutions of dynasore in 1% DMSO were generated from 0µM (DMSO vehicle control) to 200µM. All experiments were carried out in using EV-free media to mimic experimental conditions of functional assays and each concentration was tested in triplicate.

After 24 or 48 hours incubation with dynasore cells were treated with 20µL of MTT labelling reagent with gentle shaking for 5 minutes. Plates were then returned to a 37°C incubator for 2 hours. MTT reagent was then removed and crystals dissolved in DMSO by shaking for 10 minutes at room temperature. Absorbance was read at 550nm wavelength using VICTOR plate reader (PerkinElmer, USA).

Absorbance values were averaged at each concentration and absorbance of a cell-free control was subtracted to give absorbance values normalised to background. As the amount of absorbance is proportional to live cell number, cytotoxicity was assessed by cell survival as a percentage absorbance from vehicle control (Equation 5.1).

$$Relative\ Cell\ Survival(\%) = 100 * \frac{sample}{control} \tag{5.1}$$

## 5.4 Results

Adoption of the Sucrose cushion ultracentrifugation (s-UC) method of isolation was compared to Differential ultracentrifugation (UC). Following optimisation, s-UC was used to isolate vesicles for functional assays.

### 5.4.1 Sucrose Cushion Improves Small Extracellular Vesicle Recovery

Compared to UC, the addition of a 30% sucrose cushion did not change average particle size ($p$<0.05, Figure 5.4), with both methods isolating vesicles between 100–150nm.

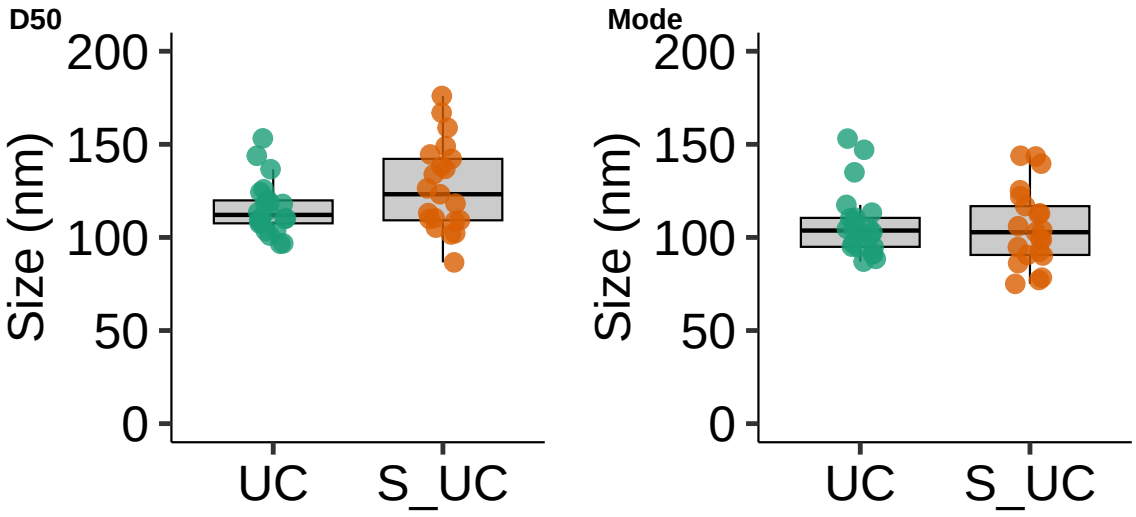



**Figure 5.4: Analysis of H9c2-derived sEV Size** D50 (median) and mode particle size estimated by nanoparticle tracking analysis in sEV samples isolated by standard differential ultracentrifugation (UC) and adapted 30% sucrose cushion ultracentrifugation (s-UC).

When compared to UC, s-UC significantly increased the number of nanoparticles isolated from the same volume of cell culture medium (Figure 5.5). Concentration was increased ($p<.001$) without changing the distribution of particle size within samples.

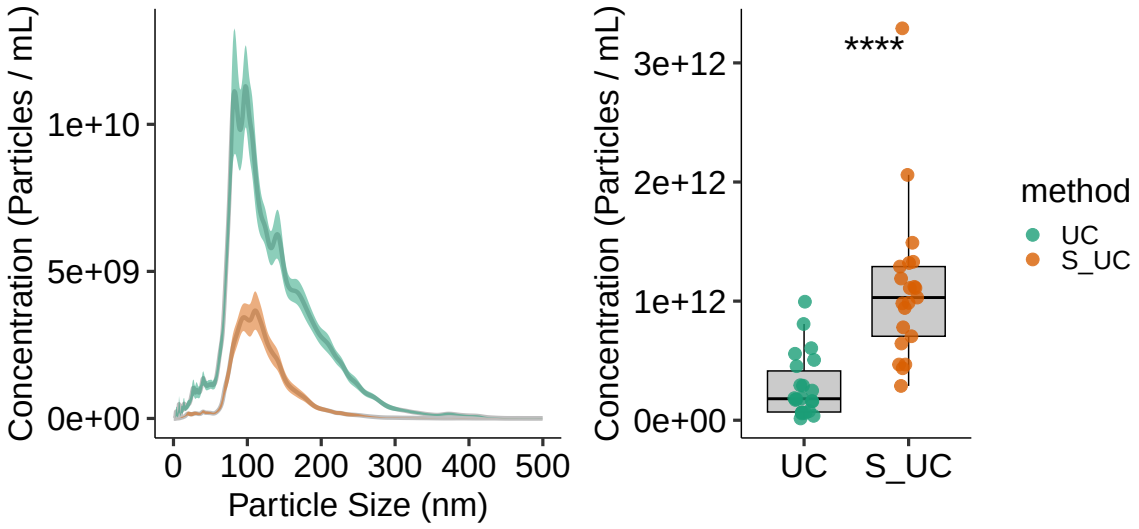



**Figure 5.5: Analysis of H9c2-derived sEV Concentration** Estimation of particle concentration in sEV preparations isolated from cell conditioned medium using nanoparticle tracking analysis. Samples were isolated from an equal amount of conditioned media using either UC or s-UC methods. *$p\leq0.05$, **$p\leq0.01$, ***$p\leq0.001$, ****$p\leq0.0001$.

### 5.4.2 Characterisation of Small Extracellular Vesicles

sEV isolated from H9c2 cells were around 100–150nm in size. *Spp1* transfection did not significantly alter sEV size compared to untransfected H9c2 cells, or cells transfected with pcDNA plasmids (Figure 5.6a). One-way ANOVA shows no difference in average EV size between transfection conditions as measured by D50 ($F(18,2)=0.445$, $p=0.648$) or by mode ($F(18,2)=0.383$, $p=0.687$). Transfection condition did not significantly influence particle concentration ($F(18,2)=2.414$, $p=0.118$, Figure 5.6b).

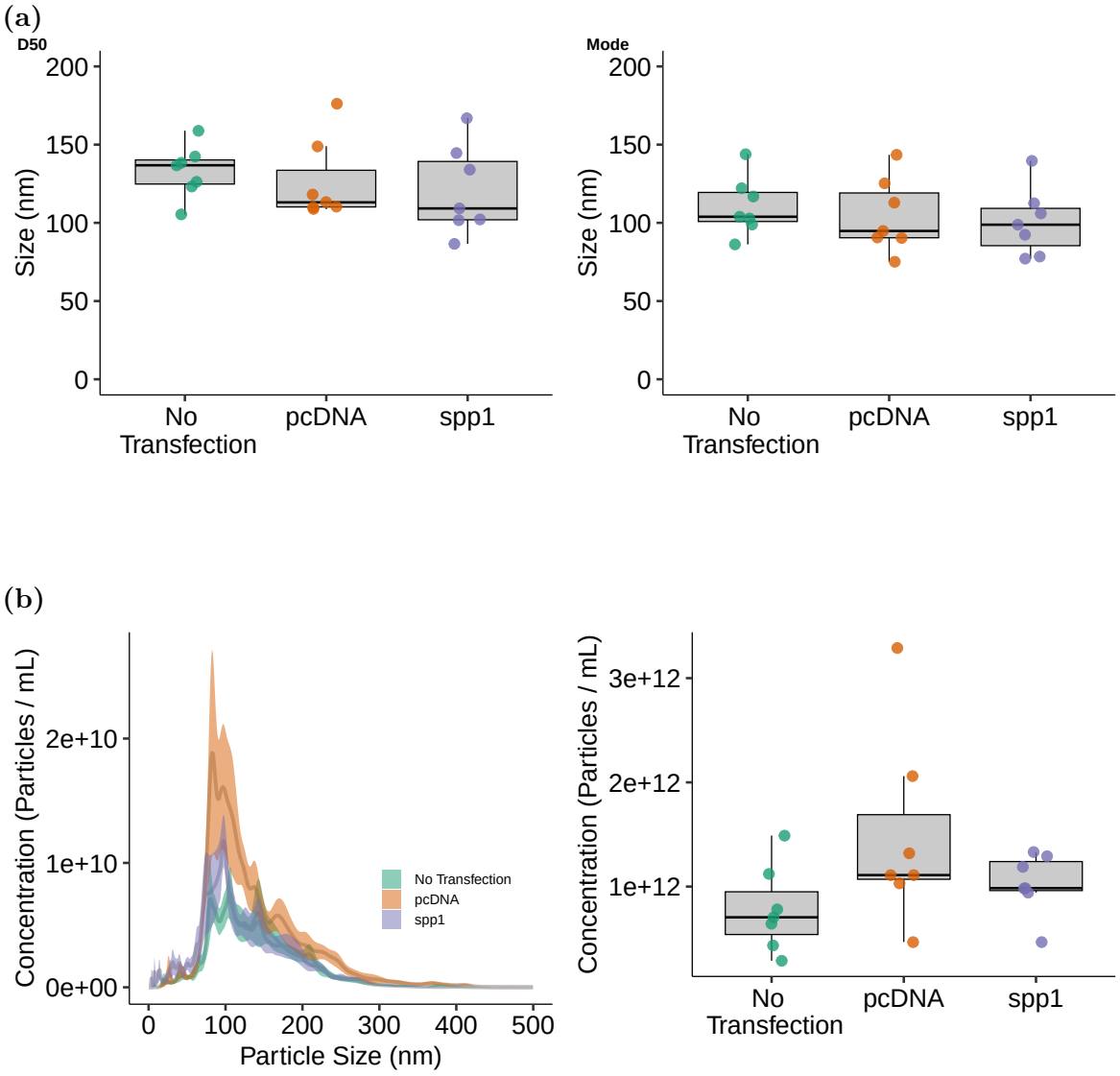



**Figure 5.6: Characterisation of H9c2-derived sEV following *Spp1* Transfection** Size and concentration of isolated sEV from H9c2 cells after 48 hours of no transfection, pcDNA or *Spp1* transfection. (a) Average particle size (D50 and Mode) is shown per transfection experiment. (b) Concentration is displayed in 0.5nm-increment bins (left) and averaged (right) across samples

Western blotting analysis showed sEV isolated from cells transfected with pcDNA or *Spp1* were positive for sEV marker protein TSG101. When normalised to total protein content, EVs contained higher amounts of TSG101 protein than cell lysates (Figure 5.7). Protein concentration was too low for assessment by BCA, so equal volume of isolated proteins were loaded into wedge well gels (∼60µL) and protein expression was quantified relative to total protein. Enough biological replicates could not be obtained for statistical analyses between transfection conditions and cell lysates. Where 3 successful *Spp1* and pcDNA replicates (n=1 isolation of 3 replicates) were performed (Figure 5.7), independent two-sided T-test suggested *Spp1* isolated EV contained more TSG101 protein than pcDNA (t=4.45, $p$=0.015). This pattern was not present in a repeated western and may represent differences in EV isolation success

than true difference in TSG101 or EV content.

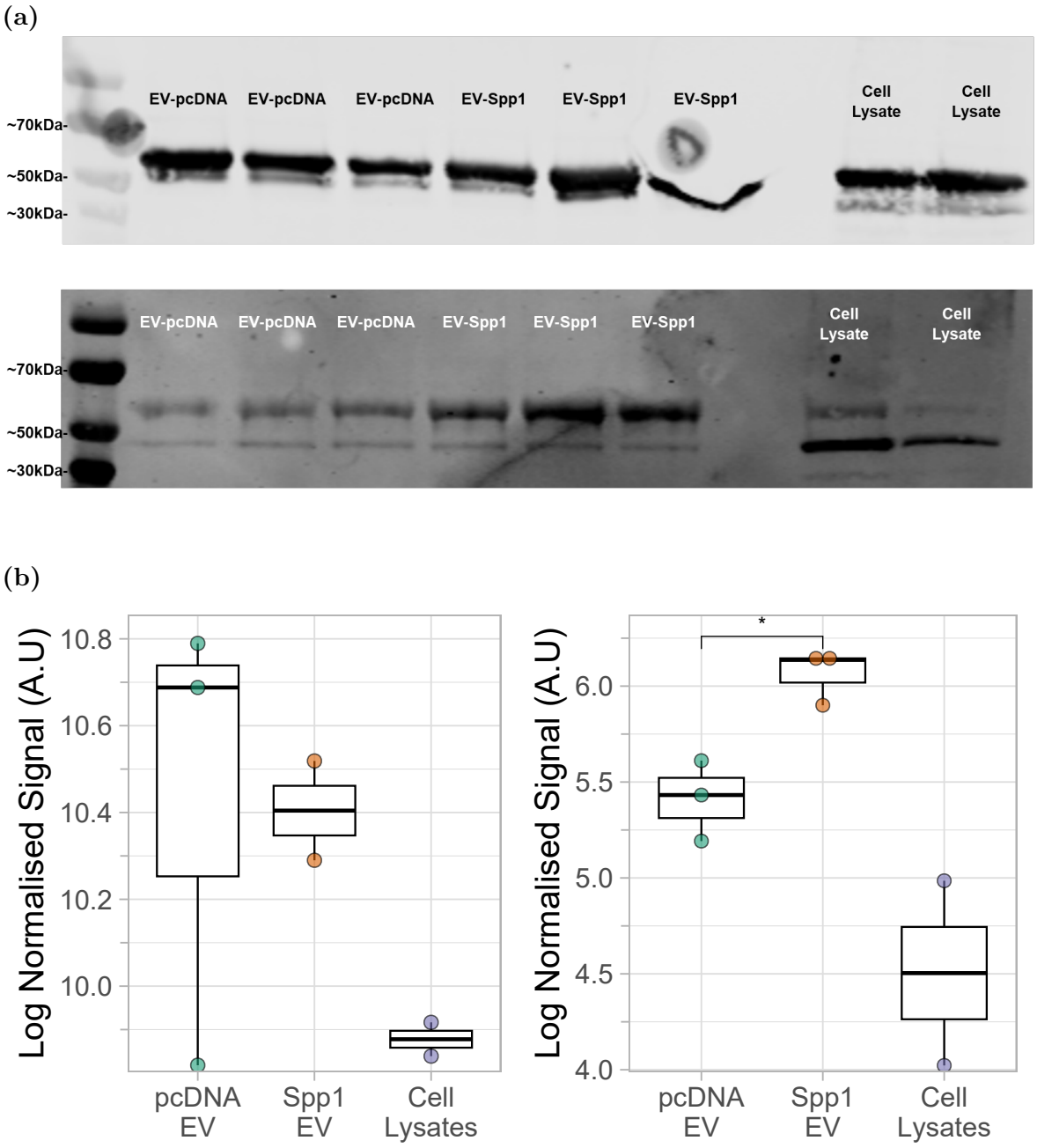



**Figure 5.7: Western Analysis of TSG101 Marker Protein Expression** (a) and (b) show two experiments with positive staining for TSG101 at ∼50-60kDa in pcDNA, *Spp1* and cell lysate protein fractions. After removal of cell lysates (n=2), independent t-test was performed on pcDNA vs *Spp1* samples. *$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$, ****$p \leq 0.0001$.

Two western analyses for ADAM10 sEV marker protein was performed using western blotting with an n=1 per condition. In both cases, sEV isolated from H9c2 cells were positive for ADAM10 (Figure 5.8).

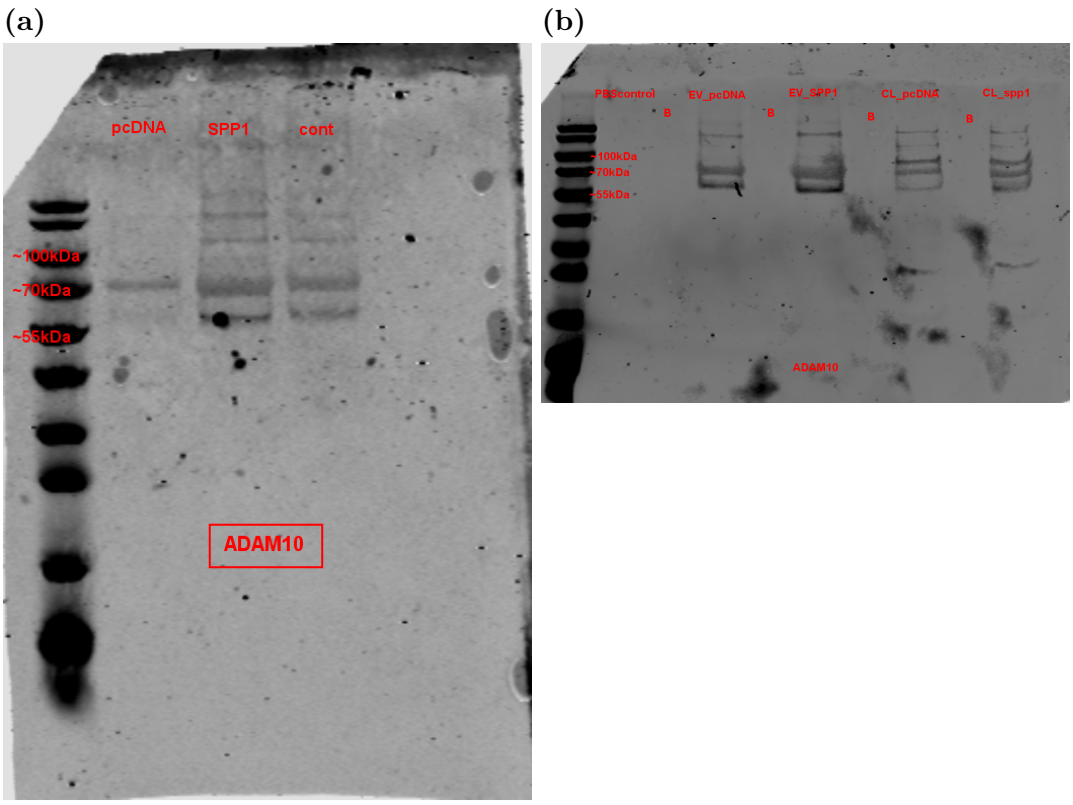



**Figure 5.8: Western Analysis of ADAM10 Protein Expression** (a) and (b) show two experiments with positive staining for ADAM at ~70kDa in pcDNA, *Spp1* and cell lysate protein fractions. Western blot was not quantified due to n=1 biological replicate.

### 5.4.3 Toxicity of Dynasore Compound

Dynasore concentrations up to 80µM resulted in >50% cell survival at both 24 and 48 hour incubation periods (Figure 5.9). As 60µM has been used in other experiments to effectively reduce EV uptake, this concentration was used in functional assays as being sufficient to effect EV uptake without causing severe stress to the cells.

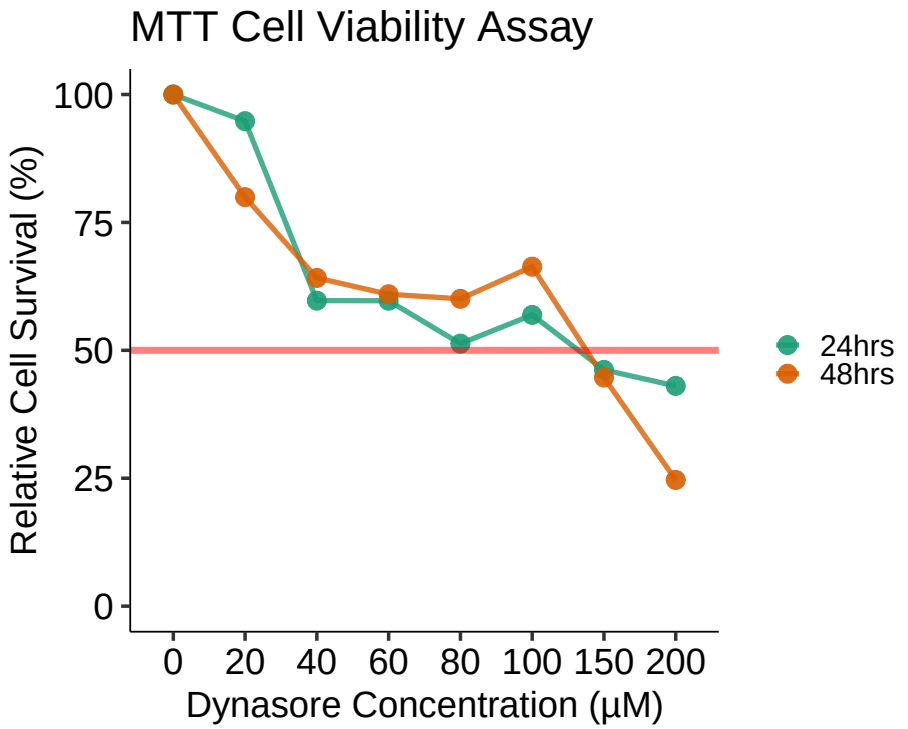



**Figure 5.9: Dynasore Cytotoxicity Assay** MTT cytotoxicity assay at a range of working concentrations of dynasore applied to H9c2 cells for 24- or 48-hours. Red line indicates greater than 50% cell death

### 5.4.4 Effects of *Spp1* Over-Expression in H9c2 Cells

Transient transfection of eukaryotic expression vector containing SHRSP-*Spp1* cDNA caused H9c2 cell size to increase 48 hours post-transfection. Cells transfected with plasmids containing SHRSP-*Spp1* were elongated compared to both pcDNA and untreated controls (Figure 5.10).

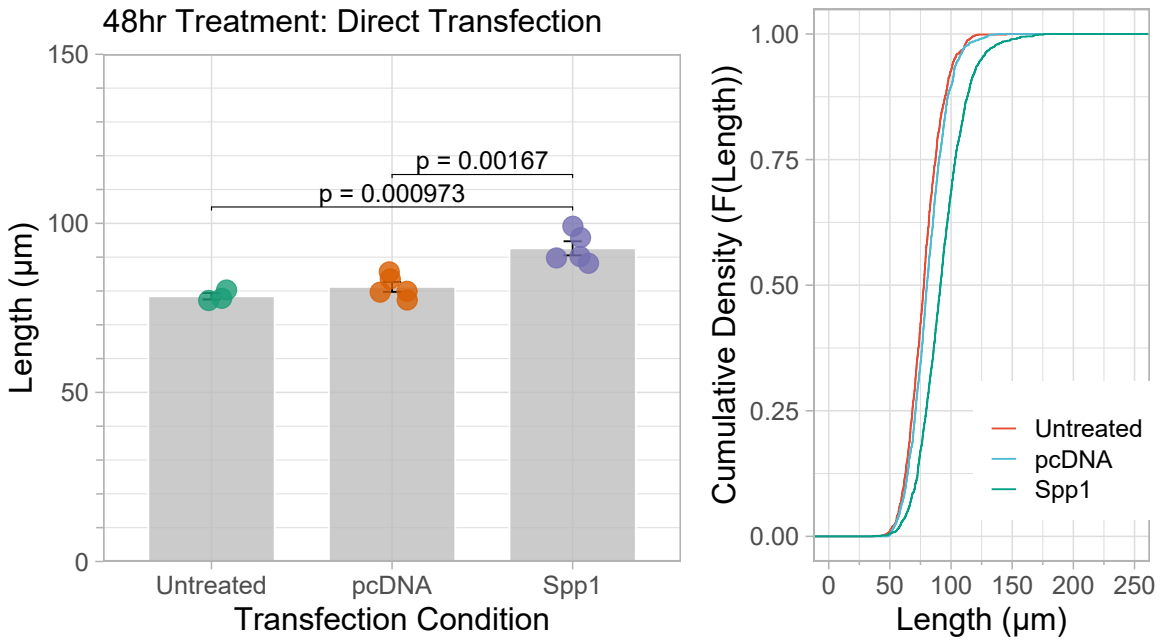



**Figure 5.10: Effect of Direct Transfection in H9c2 Cells** (left) mean cell length of H9c2 cells and (right) Empirical Cumulative Distribution Function of H9c2 cell populations 48 hours following transient pcDNA, *Spp1* or no plasmid transfection (untreated).

Conditioned culture media and small EVs derived from transfected cells were then applied to fresh H9c2 cells. Culture of naïve H9c2 cells with clarified conditioned media from H9c2 cells overexpressing SHRSP-*Spp1* led to an increase in cell size equal to that of direct transfection with SHRSP-*Spp1* containing plasmid (Figure 5.11).

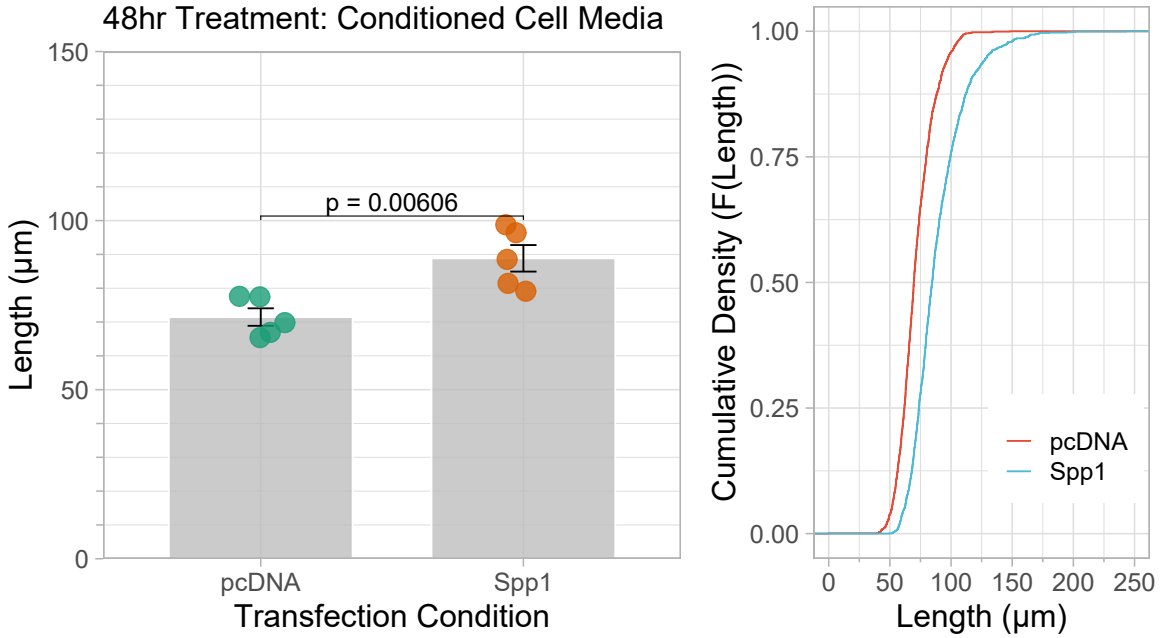



**Figure 5.11: Effect of Co-culture with Clarified Conditioned Medium** (left) mean cell length of naïve H9c2 cells and (right) Empirical Cumulative Distribution Function of cell populations 48 hours following co-incubation with conditioned cell culture medium from pcDNA, or *Spp1*.

This effect was retained when naïve H9c2 cells were co-cultured with sEV isolated from H9c2 cells 48 hours following SHRSP-*Spp1* transfection (Figure 5.12). The effect of co-culture was equal to direct transfection (∼30µm increase *Spp1* vs pcDNA).

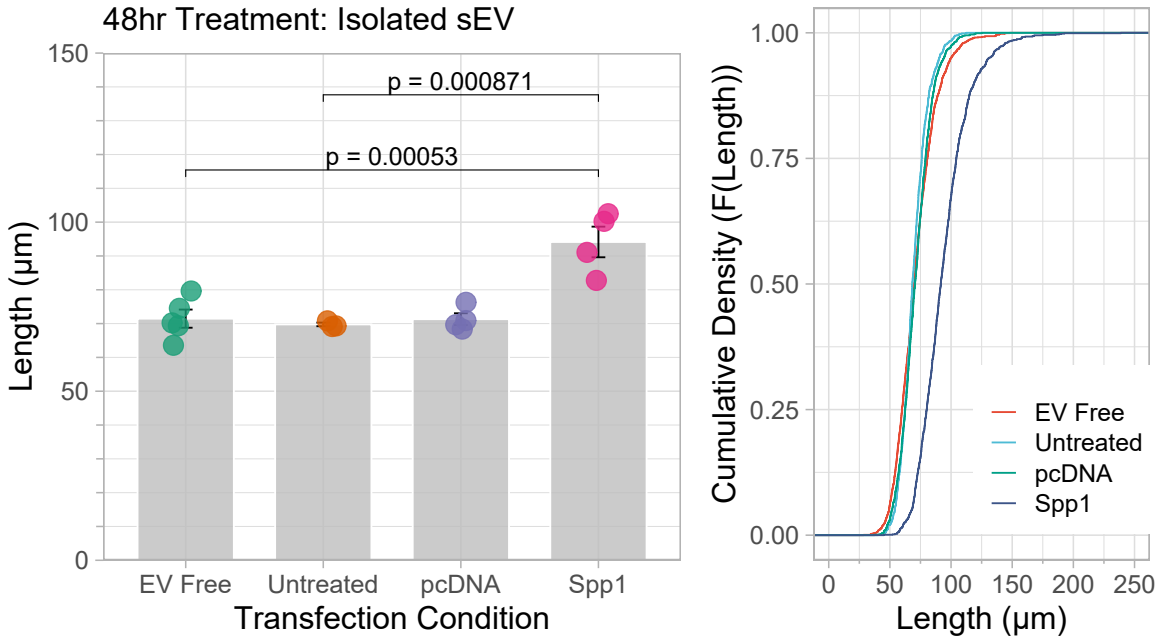



**Figure 5.12: Effect of Treatment with sEV from Transfected H9c2 Cells** (left) mean cell length of naïve H9c2 cells and (right) Empirical Cumulative Distribution Function of cell populations 48 hours following co-incubation with isolated small extracellular vesicles from pcDNA, *Spp1* or untreated control H9c2 cells

### 5.4.5 Blocking sEV Uptake Reduces the Functional Effect of *Spp1* Containing small Extracellular Vesicles

Co-incubation of naïve H9c2 cells in the presence of sEV uptake inhibitor dynasore significantly altered the response of H9c2 cells to co-incubation with sEV derived from SHRSP-*Spp1* over-expressing cells ($p<0.05$). There was a significant increase in cell size when cells were incubated with sEV isolated from SHRSP-*Spp1* transfected cells, compared to cells co-incubated with sEV derived from pcDNA and DMSO vehicle control (Figure 5.13).

Following Tukey HSD corrections, comparisons of cell size following co-incubation with sEV obtained from pcDNA and SHRSP-*Spp1* transfected cells were non-significant between dynasore treatment condition. When dynasore was added to co-culture of H9c2 cells with sEV derived from SHRSP-*Spp1* transfected cells, there was no significant increase in cell size compared to co-culture of H9c2 cells with sEV derived from pcDNA transfected cells (Figure 5.13). Cell size was not significantly smaller than the vehicle control co-culture of H9c2 cells with sEV derived from cells transfected with SHRSP-*Spp1*. Empirical Cumulative Distribution Function (ECDF) plot showed there was a clear left shift in cell size distribution when uptake of SHRSP-*Spp1*-derived EVs

was blocked by dynasore inhibition (Figure 5.13).

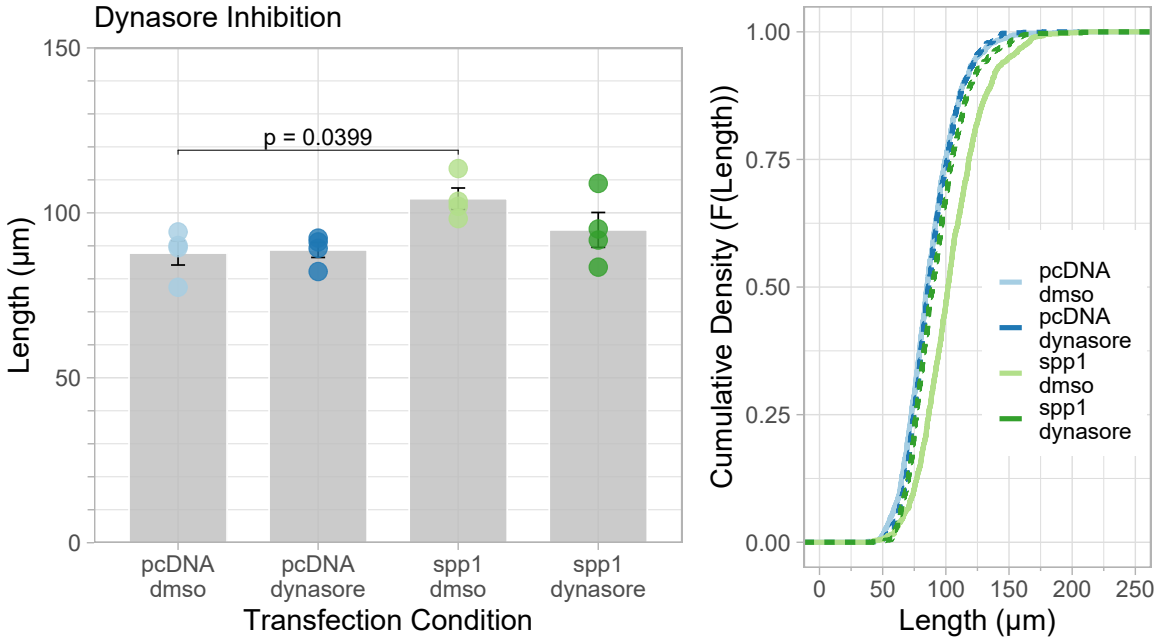



**Figure 5.13: Effect of Inhibition of sEV Uptake in H9c2 Cells** (left) mean cell length of naïve H9c2 cells and (right) Empirical Cumulative Distribution Function (ECDF) of cell populations 48 hours following co-incubation with vehicle control (DMSO, solid line) or Dynasore (dashed line) and isolated small extracellular vesicles from pcDNA or *Spp1*.

Expression of *Spp1* in cells co-incubated with small EVs was assessed using TaqMan RT-qPCR. There was no difference in the expression of *Spp1* following co-incubation with small EVs (Figure 5.14). However, due to technical difficulties in TaqMan qRT-PCR assays, the number of biological replicates was limited, and no formal testing could be carried out. Only pcDNA+DMSO and *Spp1*+dynasore conditions produced 3 biological replicates and thus the other groups did not reach requirements of statistical testing (n=±3).

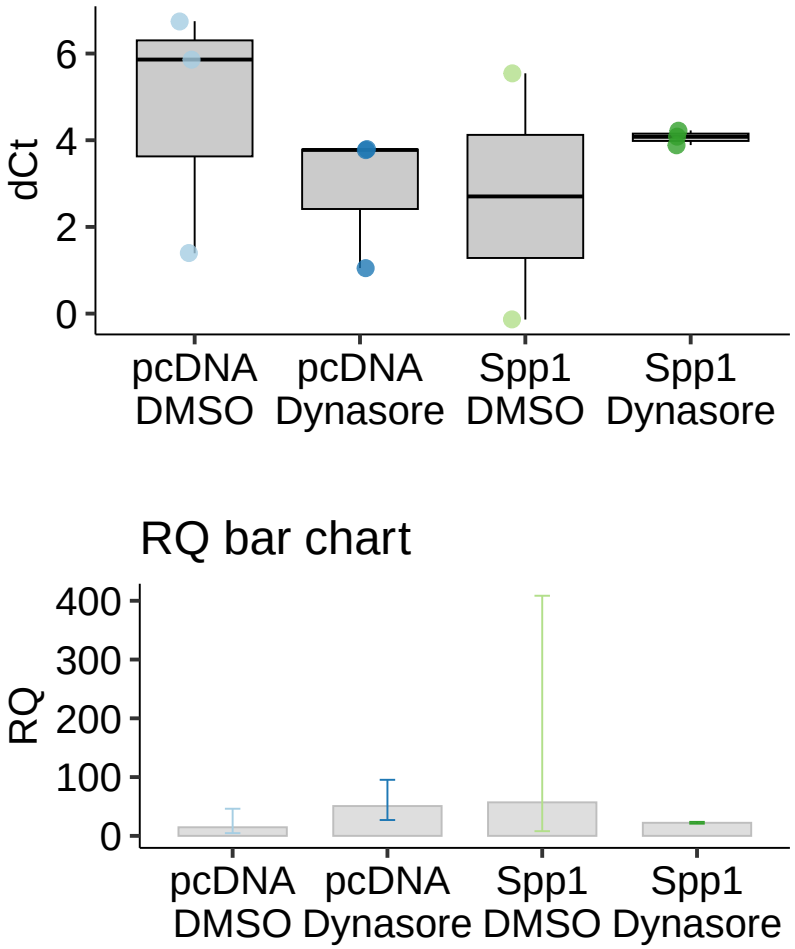



**Figure 5.14: Relative Quantification of *Spp1* in H9c2 cells Treated with sEVs and Dynasore** (top) delta cycle threshold (dCt) values for *Spp1* compared to housekeeper *B2m* used in statistical analyses. (Bottom) Relative Quantification values were calculated with pcDNA-DMSO as reference control (RQ = 1). Error bars were calculated as logarithmic standard error of ddCt.

## 5.5 Discussion

Cellular models provide an opportunity to interrogate the function of specific biological molecules within known or unknown pathways, providing the correct assay of function is selected. Functionality can be determined through overexpression, knock-down, and/or interference of the molecule or pathway of interest. Observation of effects on chosen outcome measures will implicate specific pathways of action. Overexpression of *Spp1* mRNA using transient transfection caused H9c2 cell size to increase over 48 hours following treatment. This increase in cell size was retained when naïve cells were treated directly with plasmids or with sEV derived from transfected cells. The study of sEV can be difficult methodologically as methods of isolation and characterisation are rapidly changing. Attributing functionality in this evolving landscape can therefore be difficult (Mateescu et al., 2017). We successfully improved concentrations of sEV

isolations without changing characteristics.

H9c2 cells act as a simplified model of the ventricular cardiomyocytes which undergo hypertrophic growth during pathophysiological remodelling of the heart. Genes often function within networks and can act within multiple pathways depending on the specific molecules they directly or indirectly interact with. In other cell-based experiments using Schwann cells (sciatic nerve) and fibroblast cells from the lung and foreskin (HFL-1, HFF-1), reducing *Spp1* expression caused a reduction in the expression of cytokines, *PKC*-α, ERK, and phosphorylated ERK (p-ERK) (X. Liu, Y. Sun, et al., 2017; Q. Cheng et al., 2022a), suggesting these signalling cascades are activated in the presence of *Spp1*. Exosomes released by cardiac myocytes treated with AngII showed a significant increase in the inflammatory IL-6 and Hsp90 proteins (Datta et al., 2017), supporting the role of inflammatory cytokines in the regulation of cardiac hypertrophy (Mutlak and Kehat, 2015). Cardiac fibroblast derived EVs activated MAPKs including ERK and p38 when applied to primary cardiomyocytes (Lyu et al., 2015). Increased activation of these pathways by the overexpression of *Spp1* is potentially contributing to the observed increase in H9c2 cell size.

The interplay between cardiomyocytes and cardiac fibroblasts can be difficult to determine in monocultures of single cells. Cultures of single cell populations do not reflect the cell-cell communication system mediated by a combination of EV populations, the ECM, and mechanical/electrical coupling of cells (Nicin et al., 2022). It has been shown that cardiomyocyte-derived sEV can be uptaken by fibroblasts *in-vivo*, using EVs labelled with phospholipid membrane dye (J. Yang et al., 2018). Fibroblasts release sEV in response to AngII treatment, altering the exosomal proteome to increase exosomal OPN and other proteins associated with MAPK and PI3K/Akt pathways (Lyu et al., 2015). Activation of cardiac fibroblasts resulting in changes to their signalling is thought to contribute to inducing cardiomyocyte growth. Although AngII stimulation does not affect expression of *Spp1* in H9c2 cells, directly increasing *Spp1* expression by transient transfection reliably increased H9c2 cell size. These findings support hypotheses that *Spp1* is a functional candidate gene from the chromosome 14 congenic region and increased *Spp1* is a sufficient stimulus to begin to induce a hypertrophic response.

Signalling molecules and effectors like *Spp1* are potentially trafficked within and between cells by EVs, as a method of intra- and intercellular communication and signal propagation. Proteomic analyses of serum-derived sEV from both lymphoma patients and rat osteoclasts showed OPN protein was present in sEV, and was increased in sEV from lymphoma patients versus controls (Lovisa et al., 2021; Faqeer et al., 2023). In our H9c2 model, concentration and size of sEV produced by H9c2 cells did not change depending on transfection condition, thus it is likely that it is the cargo which is changing to cause their effect. Co-culture with EVs containing OPN

protein resulted in quicker and greater activation of the TGFβ/SMAD3 pathway in mesenchymal stem cells (MSC), and treatment with a small interfering (si)RNA against *Spp1* reduced their effect (Faqeer et al., 2023). In H9c2 cells, isolated sEV from SHRSP-*Spp1* transfected cells affected naïve cells in a manner equivalent to that produced by direct transfection with SHRSP-*Spp1*. Dynasore experiments indicate the uptake of sEV cargo is required to propagate a growth response equal to that of direct transfection as H9c2 cell elongation was partially blocked by adding the EV uptake inhibitor into EV co-culture conditions. Although presence of OPN protein in EVs was not confirmed, previous RNA-sequencing analysis of sEV produced by *Spp1*-transfected H9c2 cells supports the enrichment of *Spp1* mRNA in these vesicles (Asirvatham, 2022). Together, results support a functional role of *Spp1* transported within small extracellular vesicles.

Understanding whether excess *Spp1* mRNA or protein is encapsulated into sEV is important to determine their function. Indeed, it may be products downstream of *Spp1* which are produced and exported by H9c2 cells. Evidence from plasma, cardiac fibroblast, and osteoclast-derived EVs support the hypothesis that *Spp1* itself is transported in this way (Lyu et al., 2015; Lovisa et al., 2021; Faqeer et al., 2023). A functional role of *Spp1* over-expression in cardiac cell lines, mediated through the export of sEV has been demonstrated, validating experiments previously performed, using an improved method of EV isolation. Future investigations should therefore focus on determining the molecular pathway activated by increasing *Spp1* mRNA, through assessment of activation of downstream signalling cascades, and characterisation of sEV cargo. Non-biased proteomic and transcriptomic approaches in cells post-transfection, and in sEV populations would provide a robust means of answering these questions.

# Chapter 6

# Assessing *Spp1* Function in CRISPR-Cas9 *Spp1* Mutant Rat Strain

## 6.1 Introduction

Transgenic and mutant knock-out rat models have been used in research since 1990, however difficulties introducing selective genetic modifications into the rat restricted the use of gene editing compared to mice models (Homberg, Wohr, and Alenina, 2017). Upon the introduction of Zinc-finger nuclease (ZFN), Transcription activator-like effector nuclease (TALEN), and most recently, Clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 technologies (Homberg, Wohr, and Alenina, 2017), the ability to generate effective and viable transgenic and mutant rat strains improved. The CRISPR/Cas9 technology, involves two elements; a single-guide RNA (sgRNA) and a Cas9 endonuclease, that act to guide the system to the target DNA, and induces a double-stranded break, respectively (Redman et al., 2016). The break is repaired by non-homologous end joining (NHEJ), leading to the introduction of a random insertion or deletion mutation. Alternatively, the break can be repaired by homology directed repair (HDR) where a homologous section of DNA is used as a template for repair (Lu et al., 2021). Despite the potentially higher specificity and reduced off-target effects of ZFN and TALENs, the comparative simplicity and efficiency of CRISPR/Cas9 has lead to its rapid adoption across basic and translational research (Lu et al., 2021).

A common application of CRISPR/Cas9 is the generation of a genetic knock out (KO) by Cas9-mediated deletion and NHEJ which results in the generation of a frameshift mutation (Smits et al., 2019). According to the Rat Genome Database (Vedi et al., 2023), there are over 200 different transgenic and mutant rat strains generated using this technology, however, only 3 are on the SHRSP background (data correct as of October 2023) and the number of publications utilising mouse models applying

CRISPR/Cas9 gene editing technology still outstrips that of the rat (Figure 6.1). Despite their reduced popularity compared to mice, rats have a number of characteristics which make them a superior genetic and physiological model (Chenouard et al., 2021). Although there have been many advances in the application of CRISPR/Cas9 and other targeted gene editing in human cell systems and mice, the use of this technology in the rat is progressing at a slower rate (Chenouard et al., 2021).

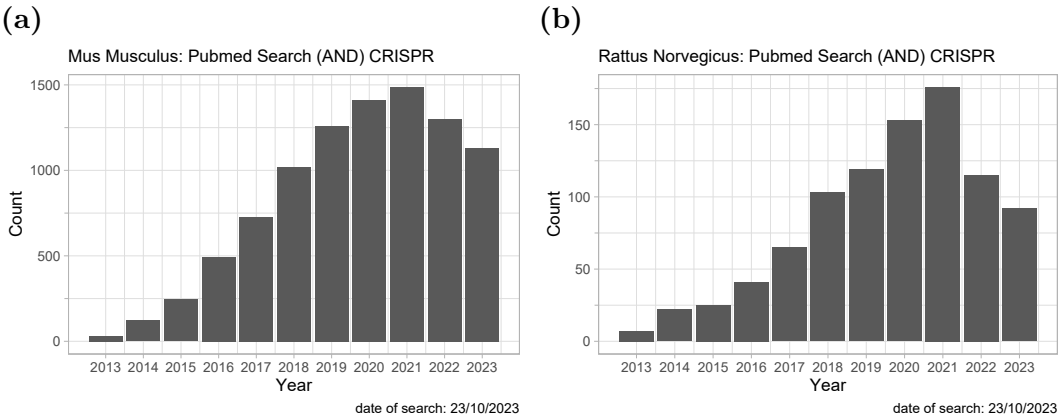



**Figure 6.1: PubMed Database Search of CRISPR/Cas9 Models** Graphs illustrating the number of search results by year of publication as retrieved from PubMed, plotted on different y-axis scales to maintain good visualisation. Advanced search was constructed using the keyword 'CRISPR' and the boolean 'AND' in combination with (a) 'mouse OR mice' and (b) 'rat'.

A number of mutant and transgenic mouse models exist whereby *Spp1* is expression is increased or reduced both globally and in specific cell types. An endothelial cell (EC) specific *Spp1* knock out on the 129/SvJ background demonstrated reduced EC proliferation during wound healing which was oestrogen-dependant (Leen et al., 2008). Global knock-out of *Spp1* on the same genetic background altered wound healing and reduced macrophage function (Liaw et al., 1998), and in mdxB10 mice TGFβ was reduced in the muscle of a global *Spp1* knock-out (Kramerova et al., 2019). Transgenic FVB mice overexpressing *Spp1* in mammary epithelial cells caused cells to become hyperproliferative in culture, potentially due to an increase in expression of the receptors OPN interacts with (α and β integrins, and CD44 receptors) (Hubbard et al., 2013).

Experimental work in Chapter 5 suggested *Spp1* overexpression increases H9c2 cell size. In these studies, the *Spp1* cDNA incorporated into plasmids was derived from the SHRSP/Gla. Preceding the development of increased LVMI by 5-weeks of age, and continuing into adulthood, SHRSP/Gla animals overexpress *Spp1* in cardiac tissue compared to age-matched WKY/Gla controls. In a transgenic mouse model, specifically over-expressing *Spp1* in cardiomyocytes led to increased cardiac fibrosis and premature death from dilated cardiomyopathy by 12 weeks of age (Renault et al., 2010). In another mouse model, interbreeding 129X1/SvJ($Col4a3^{-/-}$) with C57Bl/6 ($Spp1^{-/-}$) mice, to generate a double knock-out, improved cardiac function and decreased the

myocardial fibrosis developed as a result of *Col4a3* knock out (Yousefi et al., 2019). An *Spp1* knock-out mouse model constructed on a 129/SvJ background showed reduced fibrosis and cardiac cell apoptosis compared to wild type mice in response to 4-weeks aldosterone infusion (Sam et al., 2004).

There are a number of rat strains which develop a variety of human disease phenotypes, without intervention and with increased similarity versus mouse models (Chenouard et al., 2021). That being said, modulation of *Spp1* expression by target gene editing has focused on mice models. Following initial investigations indicating an increase in *Spp1* affected LVMI in male chromosome 14 congenic strains, the CRISPR/Cas9 system was used to introduce a mutation into the *Spp1* gene of SHRSP/A3NCrl rat embryos, resulting a 5-bp deletion mutation in exon 4 of the *Spp1* gene (Figure 6.2). The mutant SHRSP-$Spp1^{em1Mcwi}$ strain, initially received from Prof. Aron Geurts, at the Medical College of Wisconsin (MCW), is maintained as a heterozygous colony at the University of Glasgow. Male litters of homozygous wild type and knock out rats have been investigated in as yet unpublished work, which focused on cardiac and pressure responses to aortic banding in adulthood. The effect of reduced *Spp1* expression throughout the life-course, effectively reducing genetic predisposition to increased LVMI in the SHRSP, is yet to be explored.

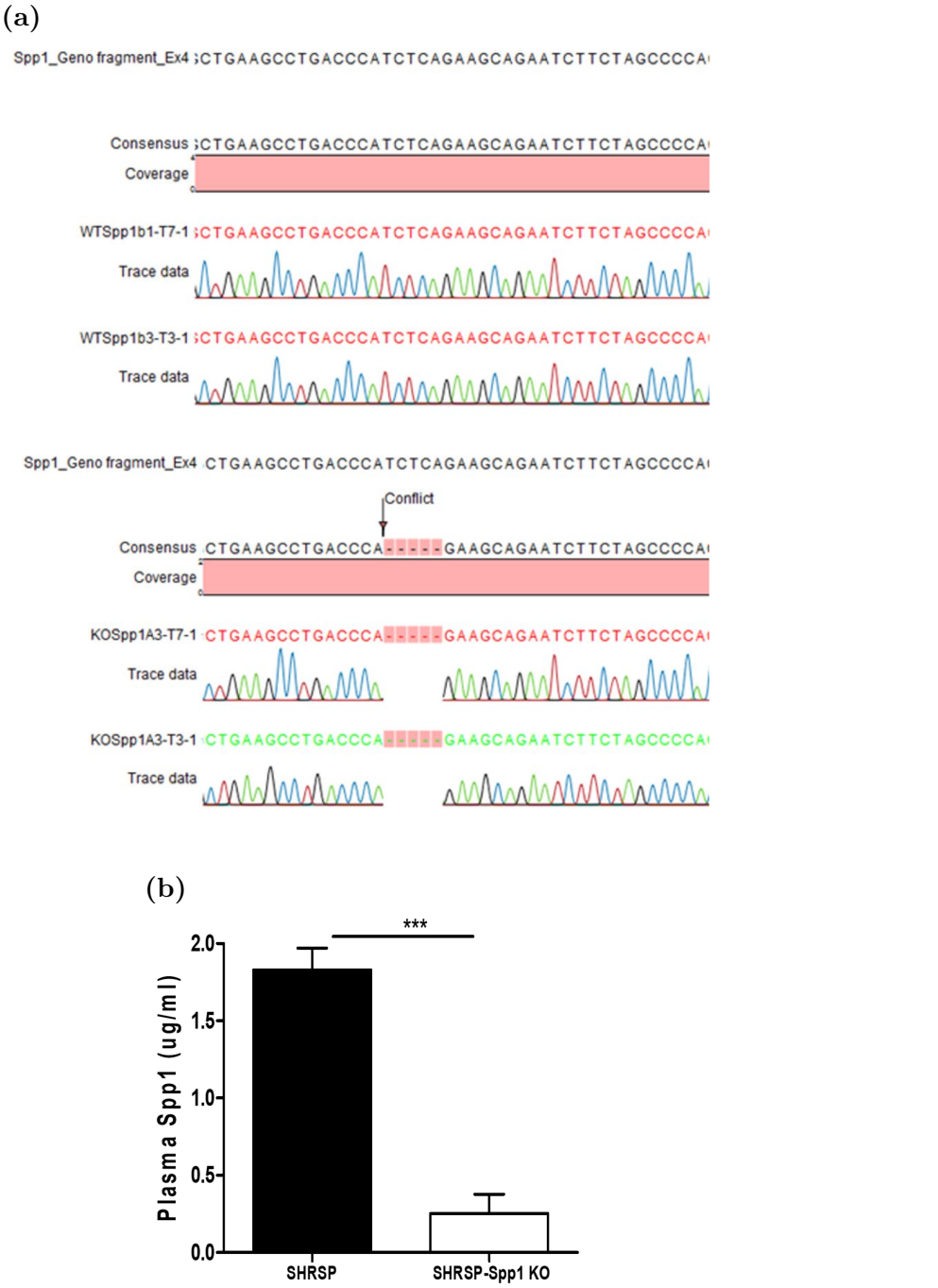



**Figure 6.2: SHRSP-*Spp1*$^{em1Mcwi}$ CRISPR-Cas9 Model** (A) Sequencing results showing the 5-bp deletion into exon 4 of the *Spp1* gene. (B) Plasma levels of osteopontin (OPN) as assessed by ELISA in SHRSP and homozygous *Spp1* knock-out animals. Data in collaboration with Prof. Aron Geurts, courtesy of Dr. Tamara Martin, University of Glasgow and reproduced with permission.

In the human transcriptome, there are many more annotated isoforms of *Spp1* than in the rat or even the mouse. Exploration of *Spp1* in studies of the human transcriptome have demonstrated expression of *Spp1* is achieved through a combination of a number of differentially spliced isoforms, which may have different functional outcomes (Shinohara, H.J. Kim, et al., 2008a). The human transcriptome contains more than 20 splice variants (Table 6.1). Of the protein coding isoforms, exons 4, 5 or both can be spliced out to produce proteins of 287, 300 and 273 amino acids

(compared to 314 in canonical 7-exon protein). Alternative isoforms can regulate the localisation of translated proteins which can often be cell or tissue specific (Xu, Modrek, and C. Lee, 2002). Given the potentially diverse functions of different *Spp1* isoforms, the possibility of functional, alternative transcripts in the rat warrants investigation. Nuclear localisation of *Spp1* is increased in dilated cardiomyopathy (Irion et al., 2020), and in immune cells, the nuclear form of *Spp1* promotes interactions with T-cells by activation of dendritic cells (Shinohara, H.J. Kim, et al., 2008a; Shinohara, J.H. Kim, et al., 2008b). In H9c2 cells, overexpression of the canonical rat *Spp1* transcript increases H9c2 cell size. Whether additional *Spp1* transcripts retain this functional effect is unknown. Targeted knock-out in the SHRSP-*Spp1*$^{em1Mcwi}$ therefore provides a unique opportunity to study potential additional, functional, *Spp1* transcripts that may exist in the rat cardiac transcriptome.

**Table 6.1:** Spp1 Isoforms from Human Ensembl Database

| **Transcript ID** | **UniProt ID** | **Length** |
|---|---|---|
| ENST00000237623.11 (SPP1-201) | P10451-5 | 1581bp/300aa |
| ENST00000360804.4 (SPP1-202) | P10451-3 | 1196bp/287aa |
| ENST00000395080.8 (SPP1-203) | P10451-1 | 1561bp/314aa |
| ENST00000508233.6 (SPP1-207) | Q3LGB0 | 923bp/273aa |
| ENST00000614857.5 (SPP1-211) | P10451-1 | 1728bp/314aa |
| ENST00000682655.1 (SPP1-218) | A0A804HLB3 | 834bp/192aa |
| SPP1-204–206, SPP1-208–210, SPP1-212–217 and SPP1-219–226 are all retained introns of varying length with no protein or defined CDS | | |

### 6.1.1 Hypothesis and Aims

CRISPR/Cas9 induced mutation into the *Spp1* gene of an SHRSP strain will alter cardiac expression of *Spp1* mRNA and protein, reducing differences in LVMI developed before the onset of high blood pressure (by 5-weeks of age).

The specific aims of the works presented are:

- Investigate potentially protective effect of reducing *Spp1* expression by assessment of cardiac phenotypes in SHRSP-*Spp1*$^{em1Mcwi}$ knock out, wild-type and heterozygous littermates at neonate and 5-week time-points.

- Characterise the functional effect of alternative *Spp1* isoforms expressed in the SHRSP-*Spp1*$^{em1Mcwi}$-/- and SHRSP-*Spp1*$^{em1Mcwi}$+/+ heart.

## 6.2 Specific Methods

SHRSP-*Spp1*$^{em1Mcwi}$ litters were assessed at 1-3 days (neonate) and 5-weeks of age. Additional tissue was collected during sacrifice to confirm genotype by PCR followed by restriction assay. Genomic DNA was extracted from flash frozen liver or ear notch tissue (Section 2.3.1). After assessing DNA quality and concentration using NanoDrop ND-1000 Spectrophotometer (ThermoFisher, Paisley, UK), 20ng was added into a PCR reaction with forward and reverse primers targeting regions upstream and downstream (respectively) of the CRISPR/Cas9 induced deletion (Figure 6.3). Following confirmation of successful PCR by agarose gel electrophoresis (Section 2.3.4), 10µL of PCR product was added to a reaction vessel containing the Dde1 restriction endonuclease enzyme (1:10, Dde1 (10 units/µL):buffer, New England BioLabs, Ipswich, USA) and incubated for 1.5 hours at 37°C. Dde1 is a restriction endonuclease that recognizes the sequence CTNAG and cuts between the C and T bases. There are two such locations in the wild type PCR product which results in three DNA products of 78, 38 and 130bp respectively. In contrast, there is only one cut site for the Dde1 enzyme in the PCR product containing 5bp deletion. This results in two DNA products of 78 and 163bp. Gel electrophoresis of restriction products allows resolution of wild type (78 + 130 bp band), knock out (78 + 168 bp bands) and heterozygous (78, 130 and 168bp bands) genotypes (Figure 6.3).

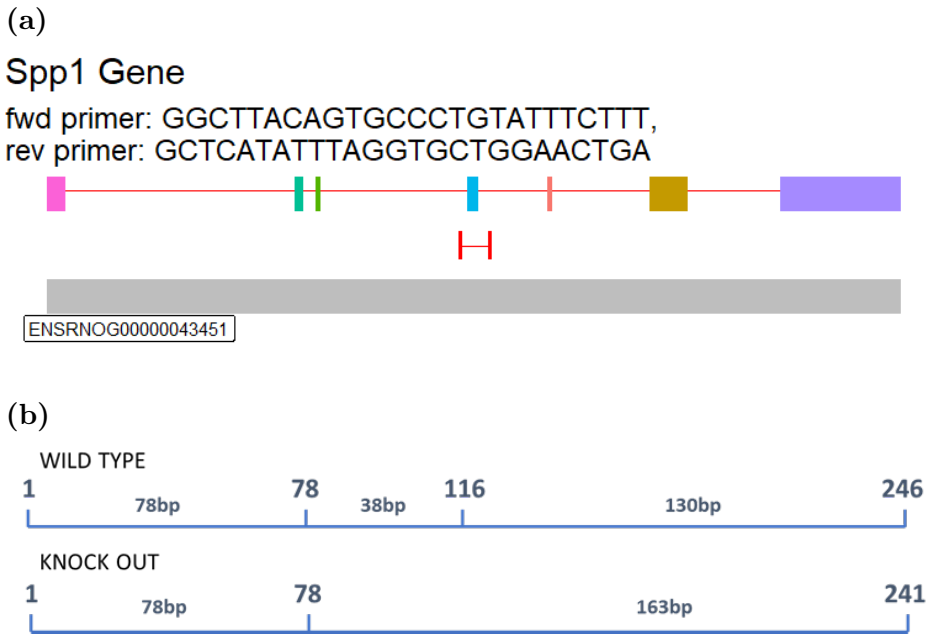



**Figure 6.3: Genotype Determination of SHRSP-*Spp1*$^{em1Mcwi}$ Litters** Graphical representation of (a) the genomic coordinates targeted by *Spp1* primers and (b) restriction products produced by incubation with Dde1 endonuclease.

### 6.2.1 *In-vivo* Assessment of *Spp1* Mutant Rat Models

SHRSP-*Spp1*$^{em1Mcwi}$ are maintained by brother-sister mating of heterozygous littermates. This produces heterozygous, homozygous knockout and homozygous wild type animals in 50:25:25 ratio according to Mendelian ratios. To reduce stress in dams, only 50% of litters were removed in the first 1-3 days for collection of neonate tissue and the remaining were left until weaning. After weaning, the litters were genotyped and cardiac structure and function was assessed by ultrasound echocardiography (Section 2.1.2). Neonate and 5-week heart and kidney tissues were collected and weighed as described in Section 2.1.3. Neonate animals were too small for echocardiographic assessment using the Acuson Sequoia c512 ultrasound system.

Prior to extraction of RNA or protein, neonate hearts were split in liquid nitrogen, following which approximately half was used to isolate RNA and the remainder used to isolate total protein. mRNA was extracted from WKY, SHRSP and SHRSP-*Spp1*$^{em1Mcwi}$-/- neonate hearts as described in Section 2.3.2. RNA was reverse transcribed into cDNA and stored at -20°C until use in TaqMan qRT-PCR assays (Section 2.3.6). The remaining neonate heart tissue was processed for protein extraction as per Section 2.5, quantified by BCA assay, and utilised in SDS-PAGE. The quantification of OPN protein, produced by *Spp1*, was assessed by immunoblotting (Section 2.6). Following transfer of separated proteins onto a nitrocellulose membrane, total protein was quantified using total protein stain (REVERT$^{TM}$, LI-COR, Nebraska, USA). After blocking, membranes were immersed in 5mL of *Spp1* primary antibody solution (1:1000, Ab63856, Abcam, Cambridge, UK) and incubated overnight at 4°C with gentle shaking. The following morning, membranes were washed with TBS-T and incubated with secondary antibody (1:20,000 LICOR IRDye 800CW) for 1 hour at room temperature. Membranes were visualised using LI-COR Odyssey system (LI-COR Biosciences Ltd) and total OPN was normalised to total protein.

### 6.2.2 Primer Design and End Point PCR

To investigate the expression of different *Spp1* isoforms, primers were designed to target the coding sequence of *Spp1*, surrounding the region affected by CRISPR/Cas9-induced deletion mutation (Figure 6.4). As described in Section 2.3.4, cDNA generated from neonate hearts was added to PCR reactions. PCR products were mixed with 6x DNA Gel loading dye (ThermoFisher, Paisley, UK) and loaded into a 1.5% agarose gel. Results were visualised following gel electrophoresis as described in Section 2.3.4. Full sequences and predicted product lengths are displayed in Table 6.2, predictions were based on the canonical *Spp1* transcript annotated in the mRatBN7.2 assembly.

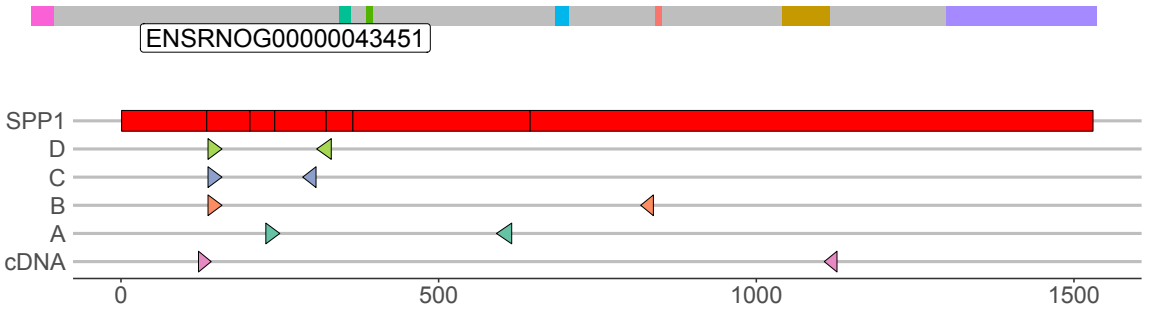



**Figure 6.4: Specific Primer Pairs Designed to Target *Spp1* mRNA Transcripts** Primer Pairs utilised in end point PCR using cDNA generated from mRNA isolated from whole neonate hearts. The forward primer in primer pair A crossed the exon 3–4 boundary. Primer pairs B-D used the same forward primer initiating amplification from within exon 2. The reverse primer of primer pair C was within the 4th exon whereas the reverse primer of primer pair D was contained in exon 5, close to the exon 4–5 boundary. cDNA primer was designed to encompass the open reading frame of *Spp1*, from the ATG canonical start codon to TAA stop codon.

**Table 6.2:** Primer Pair Design for End Point PCR Reactions

| Pair | Primer ID | Sequence | Fragment Length (bp) |
|---|---|---|---|
| A | Exon3 FwdA | CAGAGGAGAAGGCGCATTACAG | 411 |
| | Exon6 Rev1 | CATCAGAAACAGGGAAACTCCTGG | 411 |
| B | Exon2 FwdB | CCAACTACAACCATGAGACTGG | 721 |
| | Exon7 Rev2 | GGCTGGCCCTCTGCTTATAC | 721 |
| C | Exon2 FwdB | CCAACTACAACCATGAGACTGG | 191 |
| | Exon4 Rev3 | GAATTCTGTGGGGCTAGAAGA | 191 |
| D | Exon2 FwdB | CCAACTACAACCATGAGACTGG | 217 |
| | Exon5 Rev4 | AGTCATCCGTTTCTTCAGAGGAC | 217 |
| cDNA | cDNA Fwd | GAAGCCAGCCAAGGACCAACT | 1006 |
| | cDNA Rev | AAGGAACTGTGGTTTTGCCTC | 1006 |

### 6.2.3 Sanger Sequencing of PCR Products

PCR products with multiple bands were excised from the agarose gel using a blade and added to an appropriate volume of Monarch gel extraction buffer (New England Biolabs). Excised gel fragments were heated at 50°C until gel was dissolved. PCR products producing only one band, and excised gel fragments were purified using Monarch PCR purification kit (New England Biolabs). Concentration of PCR products was assessed using Nanodrop spectrophotometer and appropriate concentrations were sent to Eurofins/GATC for SupremeRun$^{TM}$ Sanger Sequencing (Eurofins Genomics, Ebersberg, Germany). Alignments of the sequences were visualised in RStudio and NCBI nucleotide Basic Local Alignment Search Tool (BLASTnt). The R-package MSA (Bodenhofer et al., 2015) was used for multiple sequence alignment with output into LaTeX.

### 6.2.4 Functional Assay of Alternative *Spp1* Transcripts

Products from end-point PCR reactions with primers targeting full length *Spp1* cDNA (start to stop codon) were amplified from cDNA generated from SHRSP-*Spp1*$^{em1Mcwi}$ wild type and knock out neonate hearts.

PCR products were purified using Monarch PCR purification kit (New England Biolabs) and ligated into pTarget mammalian expression vector (Figure 6.5, Section 2.4). Selected vectors were transformed into JM109 cells (Section 2.4.2) and screened using blue/white colony selection. Successfully transformed colonies were checked for correct orientation of insert cDNA using a T7 forward primer in PCR reactions. Following confirmation of sequence insertion by sanger sequencing, purified DNA plasmids were isolated and stored in -20°C until use in functional experiments. Glycerol stocks of plasmids were created (as per Section 2.4) and stored in -80°C.

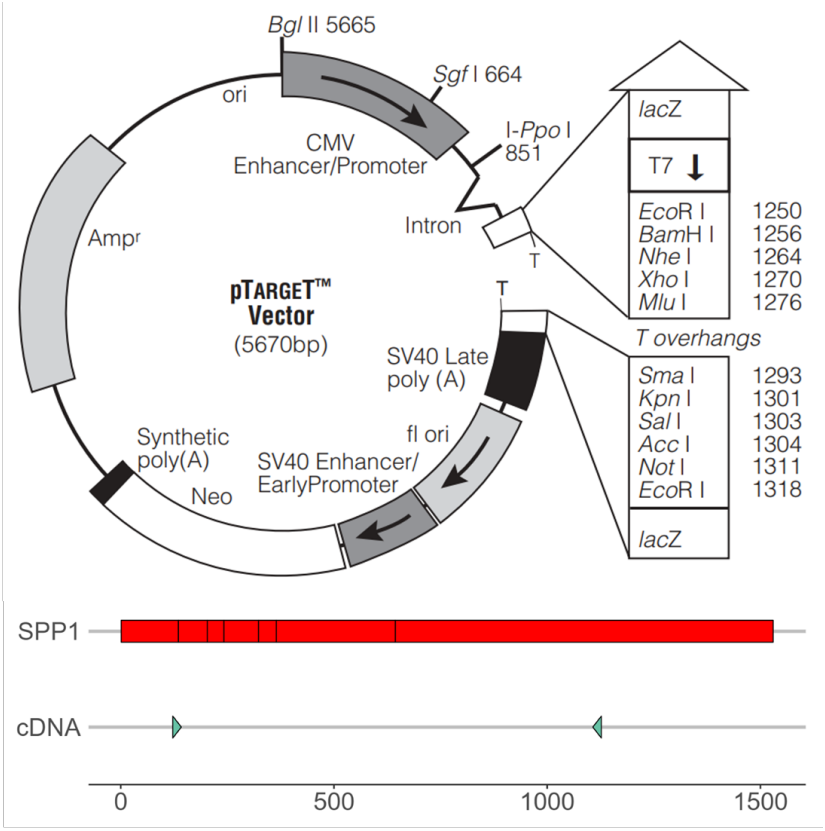



**Figure 6.5: pTarget Mammalian Expression Vector Design** Rat *Spp1* isoforms isolated using primers designed to capture full coding transcript from start to stop codon were ligated into the T-overhangs. Figure taken from Promega Technical Manual.

Following seeding into 6-well plates, H9c2 cells were transfected with no plasmid, plasmid containing pTarget control sequence (542bp sequence containing multiple stop codons in 6 reading frames), plasmid containing cDNA isolated from SHRSP-*Spp1*$^{em1Mcwi}$-wild type (*Spp1*-WT), or SHRSP-*Spp1*$^{em1Mcwi}$-knock out (*Spp1*-KO) hearts. These are referred to as control plasmid, *Spp1*-WT, and *Spp1*-KO within this chapter of work. After 48 hours, H9c2 cells were fixed in 4% PFA and stained for cell sizing using 0.5–1% crystal violet stain. Plates were visualised using an EVOS$^{TM}$ XL Core Cell Imaging System (ThermoFisher Scientific, UK) and cell sizing was carried out in ImageJ. To avoid bias in sampling, a random sample of measurements was generated from the population of measured cells.

RNA was collected from cell lysates (Section 2.3.2) and cDNA was prepared using superscript II reverse transcriptase enzymes (Section 2.3.3). TaqMan qRT-PCR was performed to assess *Spp1* expression normalised to *B2m* housekeeper gene and analysed using the relative quantification method, with untransfected condition acting as the reference control (Section 2.3.6).

Protein was extracted from adherent cells by the addition of RIPA buffer to the culture dish, supplemented with a protease and phosphatase inhibitor cocktail. Culture dishes were incubated on ice, and lysis was aided utilising sterile cell scrapers

(Section 2.5). H9c2 cell expression of *Spp1* protein, OPN, was analysed using western blot (Section 2.6) using OPN primary antibody (Abcam, ab3856, 1:1000), and normalisation to total protein. Membranes were visualised on the 700nm (total protein) and 800nm (Odessey CLx, LI-COR) and band intensities quantified by ImageStudio software (LI-COR Biosciences, UK).

### 6.2.5 Statistical Methods

Data was analysed R.4.3 (or equivalent) using code developed for *in-vivo* and *in-vitro* studies described. Assessment of *Spp1*-mutant heterozygous, and homozygous knock-out and wild type animals were compared to WKY and SHRSP strains using one-way ANOVA. Similarly, knock-out and wild type *Spp1* transcripts were compared to control plasmid and untransfected control using one-way ANOVA. ANOVA was followed by Tukey *post-hoc* testing and a $p$-value of $\leq 0.05$ was considered statistically significant. Adjustments for multiple testing were applied during *post-hoc* testing and significant pairwise comparisons are displayed in the plot area.

## 6.3 Results

### 6.3.1 Preliminary Study of Developmental Phenotype in *Spp1*-Mutant Rats

After initiation of the developmental study, preliminary results from this and aortic banding studies determined a closer inspection of the efficiency of *Spp1* knock out was required. In accordance with the 3R's (Diaz et al., 2020), the study of neonate and 5-week animals was suspended, and preliminary results are displayed. Final number of animals in each group is displayed in Table 6.3. WKY and SHRSP litters collected for Chapter 3 were used as control strains. The SHRSP/Gla and SHRSP-*Spp1*$^{\text{em1Mcwi}}$ wild type are sub-strains of the original SHRSP strains, such that the SHRSP/Gla makes a suitable, additional control strain.

**Table 6.3:** Numbers of Transgenic Strains by Genotype in Study

| **Strain** | **Neonate** | **5-weeks** |
|---|---|---|
| WKY | 16 | 12 |
| SHRSP | 12 | 12 |
| SHRSP-*Spp1*$^{em1Mcwi}$+/- | 4 | 6 |
| SHRSP-*Spp1*$^{em1Mcwi}$+/+ | 2 | 3 |
| SHRSP-*Spp1*$^{em1Mcwi}$-/- | 3 | 2 |

SHRSP-*Spp1*$^{em1Mcwi}$+/-; heterozygous wild-type/knock-out,
SHRSP-*Spp1*$^{em1Mcwi}$+/+; homozygous wild-type, SHRSP-*Spp1*$^{em1Mcwi}$-/-;
homozygous knock-out.

At the neonatal time-point, study was underpowered to statistically evaluate wild type (SHRSP-*Spp1*$^{em1Mcwi}$ +/+) neonates. After removing wild-type groups from analysis, one-way ANOVA show significant effect of genotype on body weight and kidney to body weight ratio, but not heart to weight ratio (Figure 6.6).

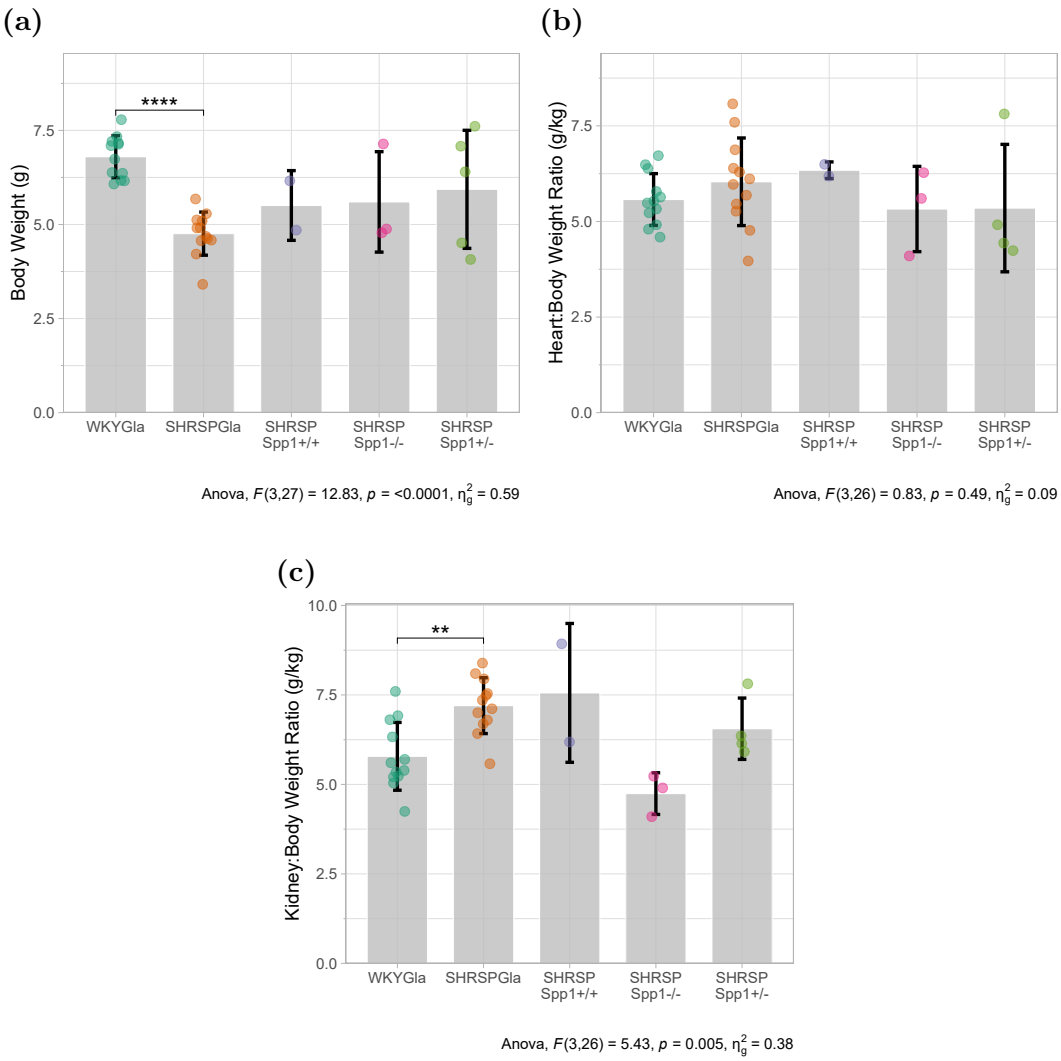



**Figure 6.6: Comparison of 1–3 day neonate *Spp1*-mutant littermates** (a) Body mass, (b) heart, and (c) kidney to body mass ratios measured at cull in WKY, SHRSP and SHRSP-*Spp1*$^{em1Mcwi}$ strains. Homozygous wild type (SHRSP-*Spp1*$^{em1Mcwi}$ +/+), knock out (SHRSP-*Spp1*$^{em1Mcwi}$ -/-) and heterozygous (SHRSP-*Spp1*+/-) animals were assigned a genotype following analysis of liver tissue collected at cull. Individual animals are represented by points. After removing wild-type animals, one-way ANOVA was performed and significant post-hoc comparisons are displayed in plot area (*$p$<0.05, **$p$<0.01, ***$p$<0.001, **** $p$<0.0001).

Differences between WKY and SHRSP Glasgow strains were as reported in Chapter 3, where SHRSP/Gla neonates were smaller than WKY/Gla. After correction for this difference, post-hoc testing showed no difference in heart weight between WKY and SHRSP strains. However, SHRSP neonates had larger kidney:body weight ratio than the WKY (Figure 6.6).

*Post-hoc* testing showed body weight at cull was not significantly different between knock-out (SHRSP-*Spp1*$^{em1Mcwi}$-/-) and heterozygous (SHRSP-*Spp1*$^{em1Mcwi}$+/-) littermates. There were also no significant differences in kidney or heart weight ratios following post-hoc tests between SHRSP-*Spp1*$^{em1Mcwi}$-/- and SHRSP-*Spp1*+/- littermates, or between *Spp1*-mutant and Glasgow WKY or SHRSP strains. There was

a non-significant trend for SHRSP-*Spp1*$^{em1Mcwi}$-/- kidney mass index to be reduced compared to the SHRSP/Gla and SHRSP-*Spp1*+/- littermates (Figure 6.6).

At 5-weeks of age, study was underpowered to statistically evaluate the SHRSP-*Spp1*$^{em1Mcwi}$-/-. Whilst body weight did not differ significantly between the strains ($p$>0.05 in ANOVA), kidney mass index was significantly altered by genotype (Figure 6.7). Kidney to body weight ratio was greater in the SHRSP and heterozygous SHRSP-*Spp1*+/- compared to the WKY (Figure 6.7). Comparison between the wild type SHRSP-*Spp1*$^{em1Mcwi}$+/+ and WKY did not reach significance in post-hoc testing, despite a comparable mean value to SHRSP/Gla. The SHRSP-*Spp1*$^{em1Mcwi}$ and SHRSP strains did not significantly differ from one another (Figure 6.7).

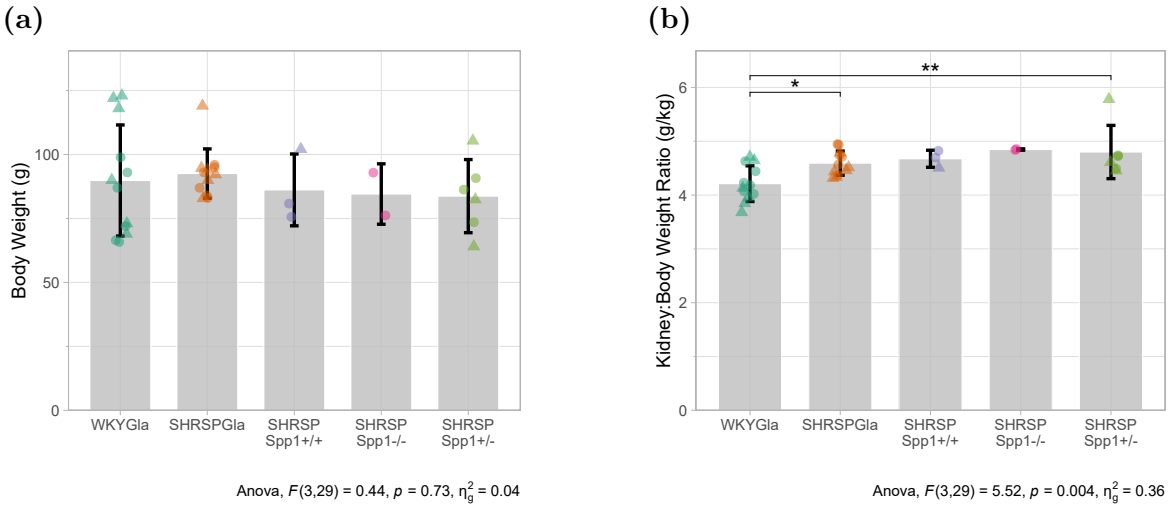



**Figure 6.7: Body and Kidney Weight at 5-weeks in *Spp1* Mutant Littermates** Comparison of (a) body weight, and (b) kidney to body weight ratio at cull in WKY, SHRSP and SHRSP-*Spp1*$^{em1Mcwi}$ strains. Homozygous wild type (SHRSP-*Spp1*$^{em1Mcwi}$ +/+), knock out (SHRSP-*Spp1*$^{em1Mcwi}$ -/-) and heterozygous (SHRSP-*Spp1*+/-) animals are displayed. After removing knock-out animals, one-way ANOVA was performed and significant post-hoc comparisons are displayed in plot area (*$p$<0.05, **$p$<0.01, ***$p$<0.001, **** $p$<0.0001). Males are represented by triangles and females by circles, to maintain power, ANOVA was performed on sex aggregated data. N=3-6 per group.

Heart to body weight ratio was not significantly affected by genotype in one-way ANOVA. LVMI measured at cull, but not by echocardiography, showed a significant main effect of genotype in one-way ANOVA (Figure 6.8, $p$=0.047). No pairwise comparisons between the strains were significant after adjustment for multiple testing (Figure 6.8). At cull, there was a trend for SHRSP-background strains to have increased LVMI compared to the WKY/Gla, however this was lost in the echocardiography-derived measurements which had increased variability.

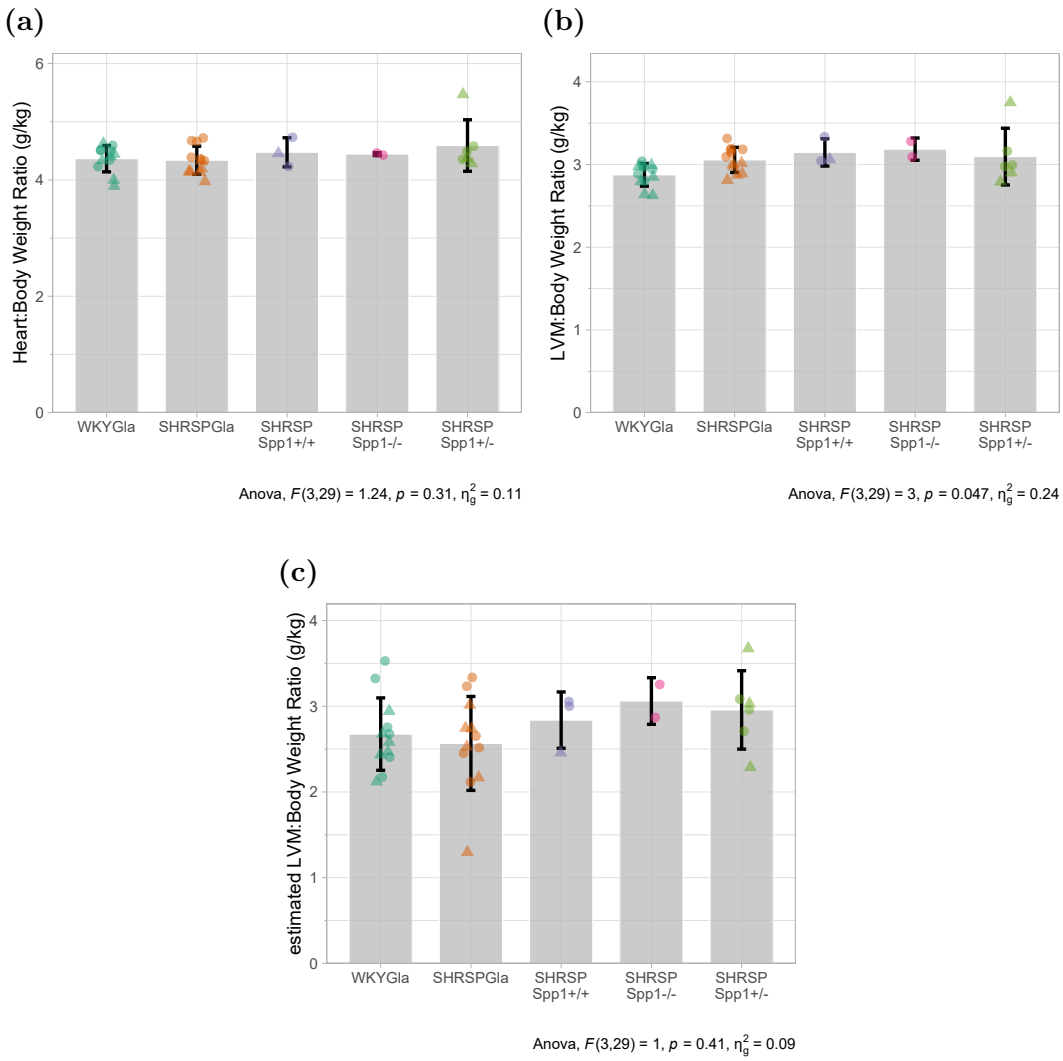



**Figure 6.8: Cardiac Measures at 5-weeks in *Spp1* Mutant Littermates** Indexed to body weight, (a) Heart, and (b) LV mass at cull is shown in WKY, SHRSP and SHRSP-*Spp1*$^{em1Mcwi}$ strains. (c) LVMI was also estimated via ultrasound echocardiography. After removing knock-out animals, one-way ANOVA was performed and significant post-hoc comparisons are displayed in plot area (*$p$<0.05, **$p$<0.01, ***$p$<0.001, **** $p$<0.0001). Males are represented by triangles and females by circles, to maintain power, ANOVA was performed on sex aggregated data.

There was no significant main effect of genotype in measures of cardiac function (SV, CO, EF, FS and RWT) assessed by echocardiography (Figure 6.9).

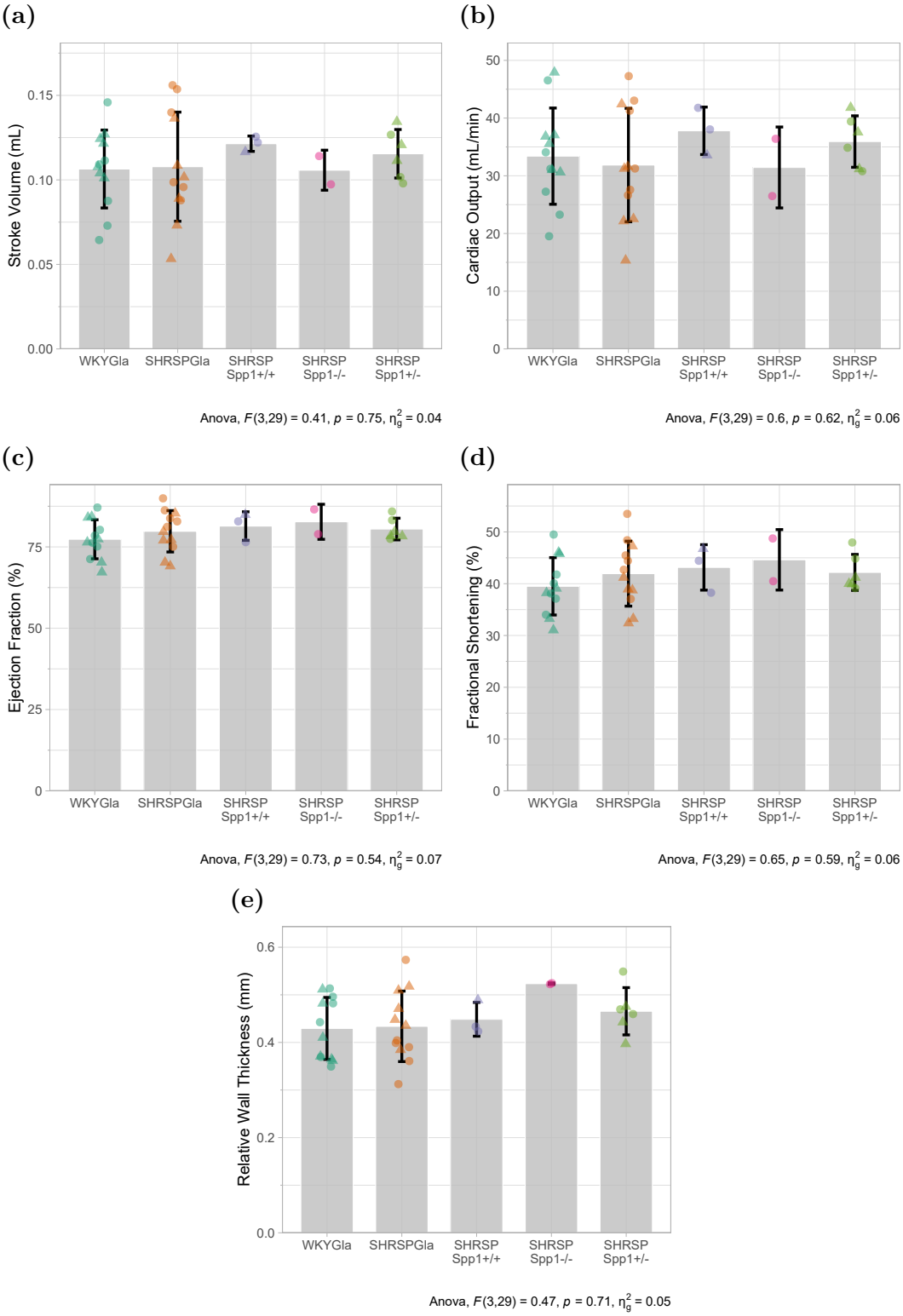



**Figure 6.9: Echocardiography of 5-week *Spp1* Mutant Littermates** Echocardiography derived measures of (a) stroke volume (mL), (b) cardiac output (Q; mL/min), (c) ejection fracion (%), (d) fractional shortening (%), and (e) relative wall thickness (mm) in WKY, SHRSP and SHRSP-*Spp1*$^{em1Mcwi}$ strains. After removing knock-out animals, one-way ANOVA was performed and significant post-hoc comparisons are displayed in plot area (*$p$<0.05, **$p$<0.01, ***$p$<0.001, **** $p$<0.0001). Males are represented by triangles and females by circles, to maintain power, ANOVA was performed on sex aggregated data.

### 6.3.2 Assessment of *Spp1* mRNA and Protein in *Spp1*-Mutant Rats

In the WKY, SHRSP, and SHRSP-*Spp1*$^{em1Mcwi}$-/- neonate heart, expression of both *Spp1* mRNA and protein content was determined. As reported in Chapter 3, *Spp1* expression was significantly increased in SHRSP neonate hearts compared to WKY neonate hearts. The SHRSP-*Spp1*$^{em1Mcwi}$-/- had significantly reduced cardiac *Spp1* mRNA compared to both the WKY/Gla and SHRSP/Gla when assessed by TaqMan qRT-PCR (Figure 6.10).

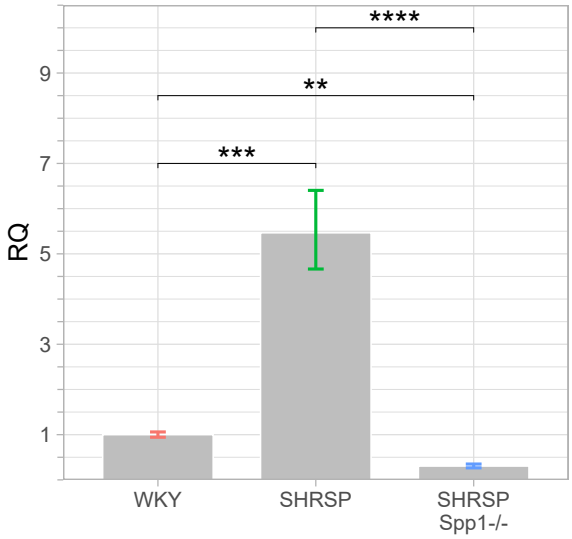



**Figure 6.10: TaqMan qRT-PCR in SHRSP-*Spp1*$^{em1Mcwi}$-/- Neonate Hearts** *Spp1* expressed was assessed using mRNA extracted from neonate SHRSP-*Spp1*$^{em1Mcwi}$-/-, SHRSP/Gla, and WKY/Gla hearts. Fold change relative to the WKY was estimated by the ddCt method. One-way ANOVA was performed on dCt values and significant post-hoc comparisons are displayed in plot area (*$p$<0.05, **$p$<0.01, ***$p$<0.001, **** $p$<0.0001)

At the protein level, one-way ANOVA showed a main effect of genotype on *Spp1* expression normalised to total protein ($F(2,6)=17.17$, $p=.003$). There was an increase in the *Spp1* protein, OPN, when assessed by western blot in the SHRSP-*Spp1*$^{em1Mcwi}$-/-, compared to both the WKY/Gla and SHRSP/Gla (Figure 6.11).

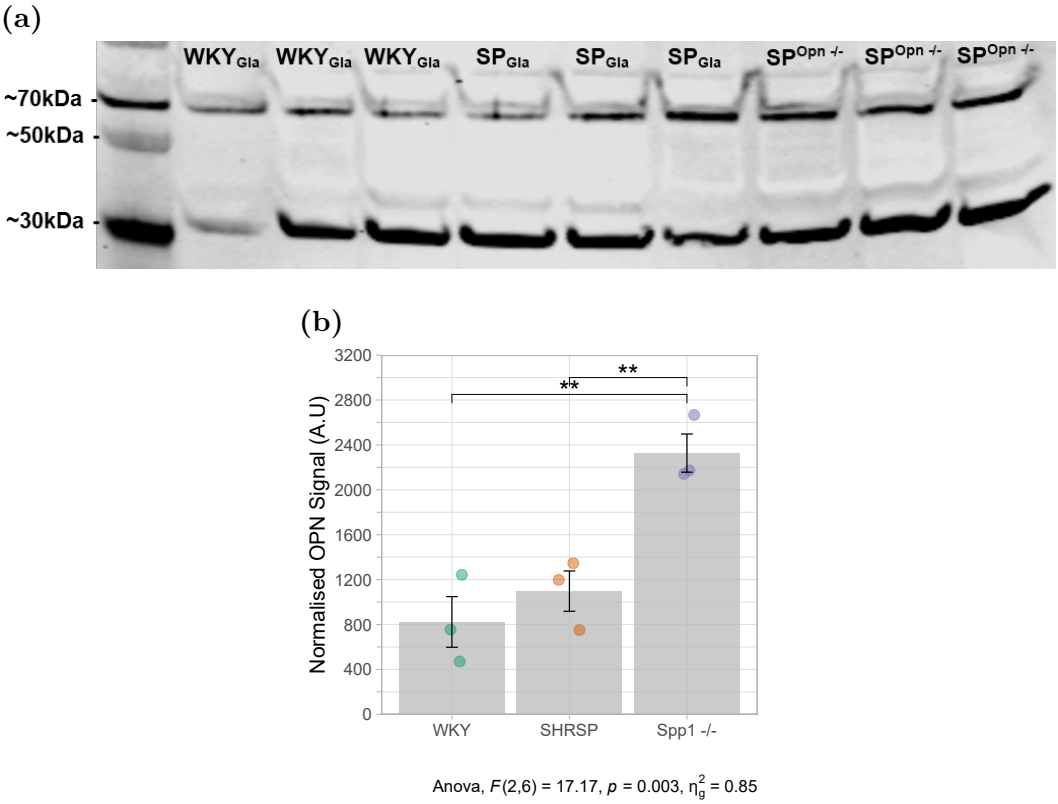



**Figure 6.11: *Spp1* Protein in SHRSP-*Spp1*$^{em1Mcwi}$-/- Hearts** Immunoblotting confirmed the presence of *Spp1* protein, OPN, in the hearts of SHRSP-*Spp1*$^{em1Mcwi}$-/-, SHRSP/Gla and WKY/Gla 1–3 day neonates. OPN signal was normalised to total protein. One-way ANOVA was performed on normalised OPN values and significant post-hoc comparisons are displayed in plot area (*$p$<0.05, **$p$<0.01, ***$p$<0.001, **** $p$<0.0001)

### 6.3.3 Alternative *Spp1* Transcripts in the SHRSP-*Spp1*$^{em1Mcwi}$

Endpoint PCR was unsuccessful in all strains using primer pair C and results are not presented. Results of successful PCR using the primer pairs A, B, D, and full length cDNA are shown in Figure 6.12. Primer pair A was designed crossing the exon 3–4 boundary in the forward direction, to capture the canonical transcript containing the 5bp deletion in the SHRSP-*Spp1*$^{em1Mcwi}$. WKY/Gla, SHRSP/Gla, wild-type (SHRSP-*Spp1*$^{em1Mcwi}$+/+), and knock-out (SHRSP-*Spp1*$^{em1Mcwi}$-/-) hearts showed evidence of transcription of the full length *Spp1* (Figure 6.12a). The 5bp deletion cannot be resolved by agarose gel electrophoresis.

Primer pairs B and D produced additional transcripts in SHRSP-*Spp1*$^{em1Mcwi}$-/- neonate hearts which were not visibly present in SHRSP-*Spp1*$^{em1Mcwi}$+/+, WKY/Gla, or SHRSP/Gla neonates. Exon 4, containing the 5bp deletion is 81bp in length, which corresponds to the approximate difference between the two transcripts produced in the knock-out (SHRSP-*Spp1*$^{em1Mcwi}$-/-) hearts (Figure 6.12b&c).

Primers designed to target full length *Spp1* cDNA (start to stop codon) produced small amounts of three potential transcripts in the SHRSP-*Spp1*$^{em1Mcwi}$-/-, two of which

reflected the calculated sizes for transcripts with and without exon 4 (1006 & 925bp respectively, Figure 6.12d). The minus exon 4 product, which removes 5bp deletion induced by CRISPR/Cas9, was the target of subsequent sequencing, cloning, and functional experiments. One of the SHRSP-*Spp1*$^{em1Mcwi}$+/+ neonate hearts also showed evidence of a shorter *Spp1* transcript, although the band was very faint and not present in the 2nd replicate (Figure 6.12d).

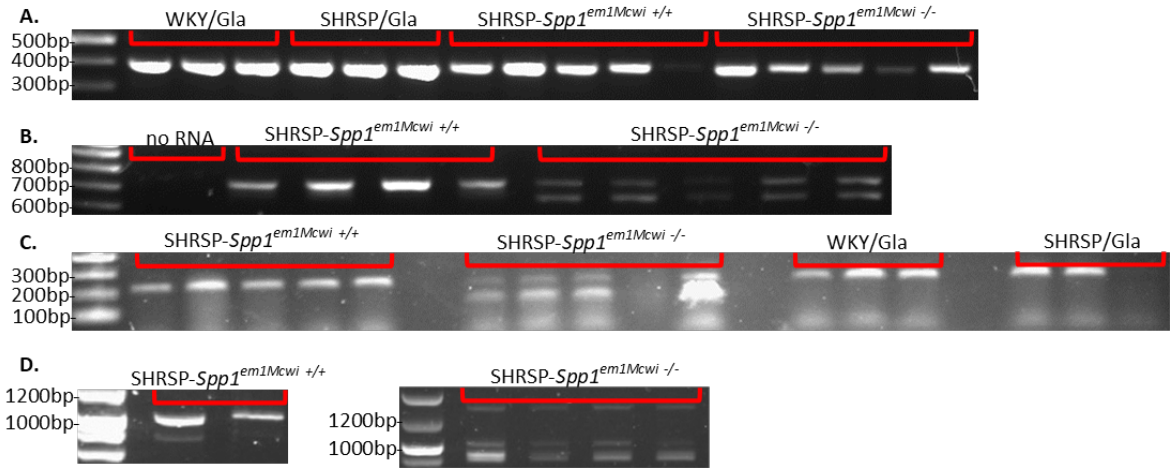



**Figure 6.12: Endpoint PCR Investigating Alternative Forms of *Spp1* in Mutant Neonates** cDNA produced from neonate hearts in WKY, SHRSP and SHRSP-*Spp1*$^{em1Mcwi}$ strains was amplified by PCR using different primer pairs designed around the CRISPR/Cas9-induced 5bp deletion, in exon 4 of the canonical *Spp1* gene. Results from (a) primer pair A, (b) primer pair B, (d) primer pair D and (d) full length cDNA are shown. Unsuccessful PCR using primer pair C is not shown.

Products from end-point PCR targeting full length cDNA (start to stop codon, Figure 6.12d) were transformed into JM109 *e-coli* cells using the pTarget mammalian vector. Transformations of transcripts isolated from SHRSP-*Spp1*$^{em1Mcwi}$-/- and SHRSP-*Spp1*$^{em1Mcwi}$+/+ neonate hearts are referred to as 'KO' and 'WT' respectively. Blue/white selection of 49 KO and 21 WT colony forming units (CFU) generated 70 single colonies to be tested for forward orientation of inserts. Colonies were chosen following PCR amplification of expanded single colonies against the T7 forward primer and *Spp1* cDNA reverse primer. Successful PCR amplification from the T7 promoter indicated correct orientation of the *Spp1* insert and reduced the number of viable colonies to 18 KO and 4 WT (data not shown). Colonies for plasmid isolation were selected by size for sanger sequencing. Following alignment of sequences to the canonical rat *Spp1* transcript, 3 KO colonies and 1 WT colony were stored in 50% glycerol (glycerol stock solution) and maintained in -80°C.

Sequencing and alignment of the WT5 CFU and KO8 CFU show *Spp1* was successfully cloned in both its canonical and minus-exon 4 containing forms (Figure 6.13). Plasmids were isolated from these colonies and used for downstream functional analyses.

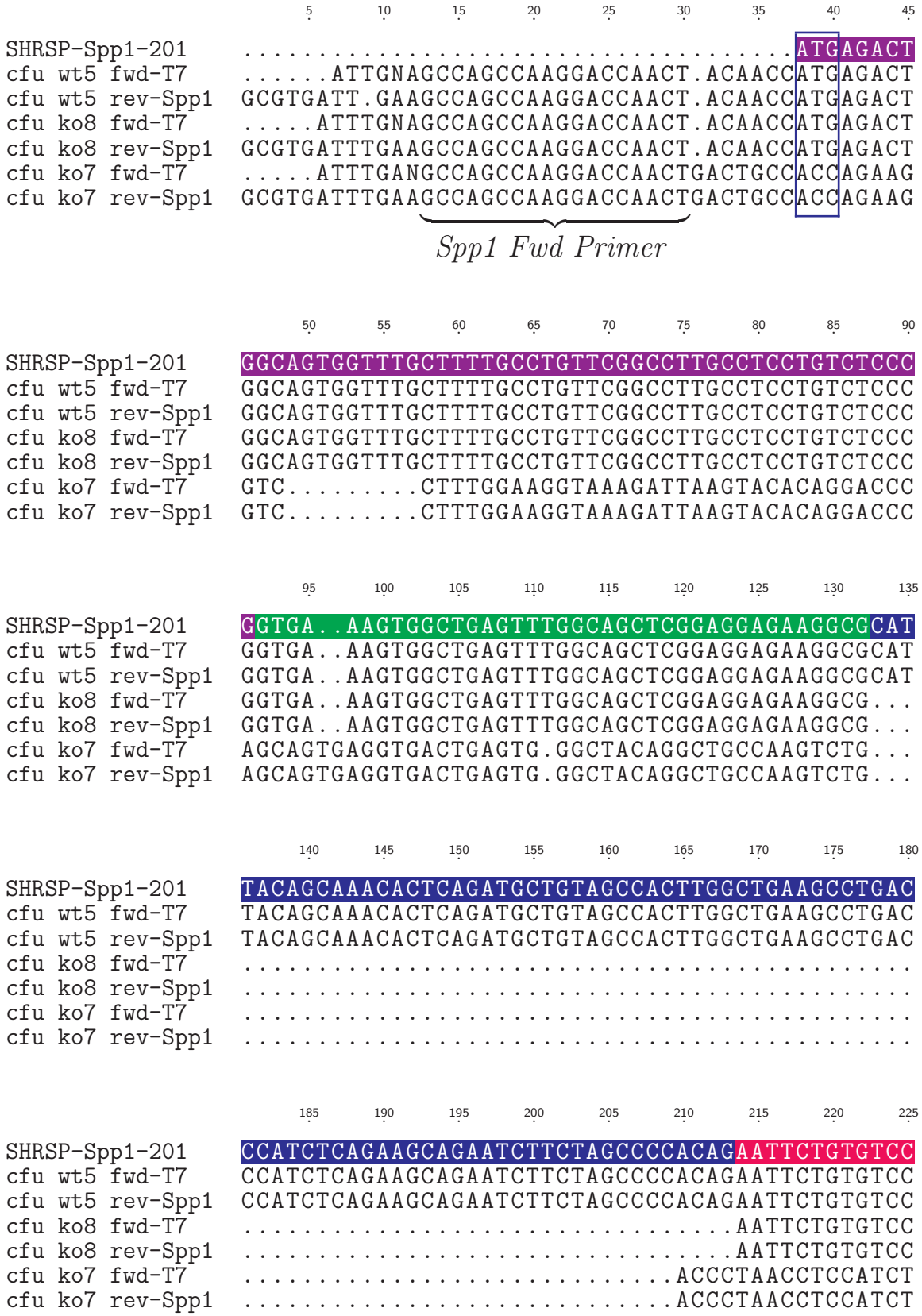



**Figure 6.13: Multiple Sequence Alignment of WT and KO Transcripts Transformed into pTarget Vectors** Trimmed sequences obtained from sanger sequencing (-25 nucleotides from start and end of sequence) were aligned to the canonical coding sequence (CDS) transcript of *Spp1* in the SHRSP/BbbUtx (Ensembl Spp1-201). Exons are shaded with alternative colours and blue box indicates transcription start sequence ATG, contained within exon 2. Full alignment is included in appendices. The original *Spp1* cDNA forward primer, used in initial PCR can be seen in all CFU sequenced.

### 6.3.4 Functional Assessment of Alternative *Spp1* Transcripts in H9c2 Cells

H9c2 cell size was significantly affected by transfection condition ($F(20, 3)=11.50$, $p<.001$). Transfection of plasmids containing SHRSP-*Spp1*$^{em1Mcwi}$+/+ (WT) increased H9c2 cell size after 48-hours, compared to untreated and control plasmids (Figure 6.14). The alternative form of *Spp1*, expressed in SHRSP-*Spp1*$^{em1Mcwi}$-/- (KO), also increased H9c2 cell size compared to untreated and control plasmid (Figure 6.14). There was no difference in H9c2 cell size between the two *Spp1* isoforms or between control and untreated control conditions.

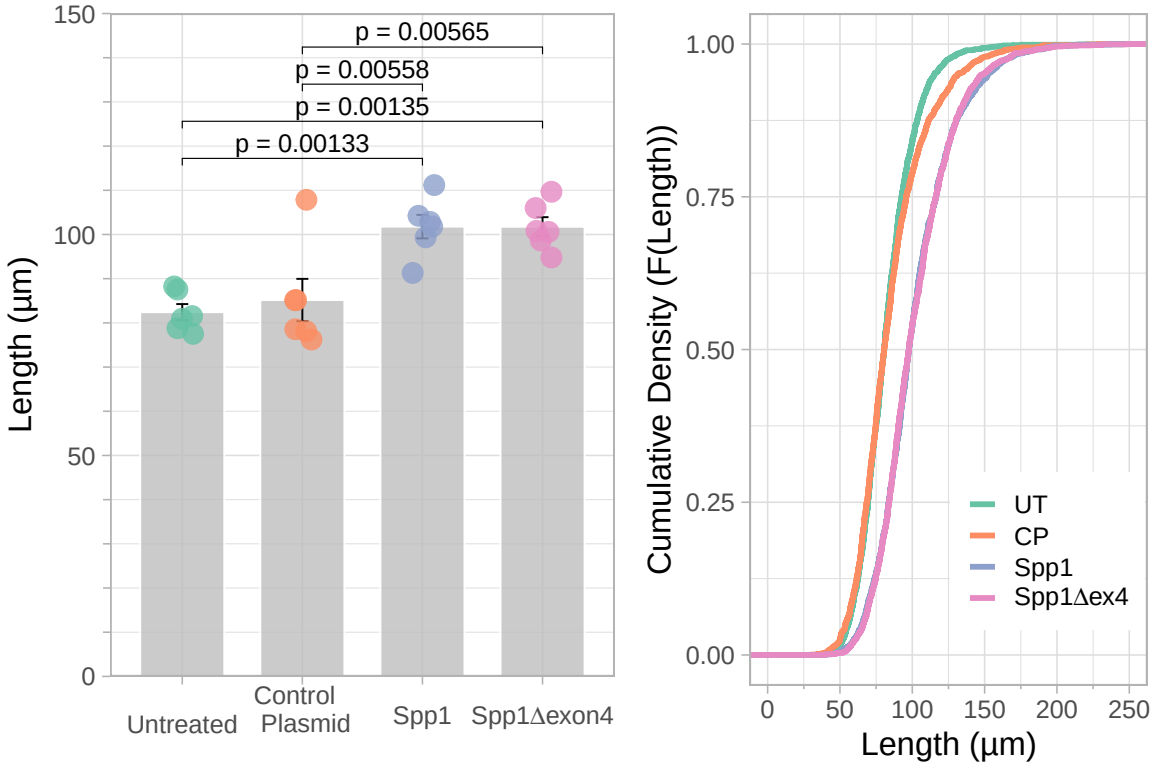



**Figure 6.14: H9c2 Cell Sizing Assay using Alternative *Spp1* Transcripts** H9c2 cells were transfected with plasmids containing pTarget control plasmid (CP), SHRSP-*Spp1*$^{em1Mcwi}$ +/+ derived (*Spp1*) or a short, alternative form produced by SHRSP-*Spp1*$^{em1Mcwi}$ -/- (Spp1Δex4) animals. One-way ANOVA was performed on mean cell size across biological replicates (n=6 per group) and significant pairwise comparisons are shown on the left. (right) Empirical cumulative distribution frequency (ECDF) curves show the distribution of all cells measured across all biological replicates, per group. NT; no transfection, CP; control plasmid, WT; wild type, KO; knock-out.

*Spp1* expression was assessed 48-hours post transfection using TaqMan qRT-PCR. One-way ANOVA of dCt values showed a significant main effect of transfection condition on *Spp1* expression in H9c2 cells ($F(3,18)=40.843$, $p<.001$). TaqMan qRT-PCR confirmed transfection of both *Spp1*-WT and *Spp1*-KO containing plasmids significantly increased mRNA expression in H9c2 cells 48-hours post transient transfection, compared to both untreated and control plasmid (Figure 6.15).

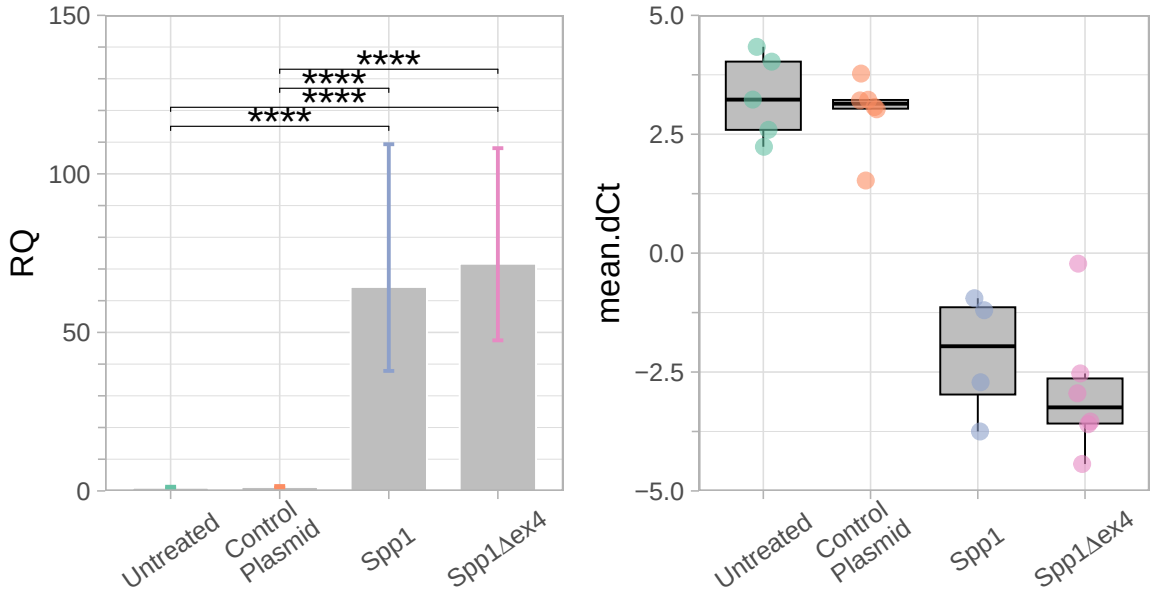



**Figure 6.15: TaqMan qRT-PCR in H9c2 Cells 48-hours Post Transfection** H9c2 cells were transfected with plasmids containing pTarget control plasmid (CP), wild type *Spp1* cloned from SHRSP-*Spp1*$^{em1Mcwi}$ +/+ (Spp1), or a short, alternative form produced in SHRSP-*Spp1*$^{em1Mcwi}$ -/- (Spp1Δex4) hearts. *Spp1* expression was normalised to rat *B2m* and quantified as RQ values against untreated control. *$p$<0.05, **$p$<0.01, ***$p$<0.001, **** $p$<0.0001. N=6 per group included in triplicate, only biological replicates with at least successful 1 technical replicate were included in analyses.

Analysis of *Spp1* at the protein level by immunoblotting, confirmed *Spp1* protein was produced by H9c2 cells 48-hours after transient transfection. However, unlike mRNA levels, which remained significantly increased in *Spp1*-containing plasmid conditions, protein levels were not elevated beyond endogenous H9c2 production (Figure 6.16). There was no significant difference in *Spp1* protein assessed by one-way ANOVA. The additional bands of potentially cleaved and modified OPN (∼30kDa) in lysates from H9c2 cells transfected with plasmids containing *Spp1*+/+ (WT) were not quantified due to the low signal-to-noise ratio, which was below the reliable threshold for distinguishing signal from background noise.

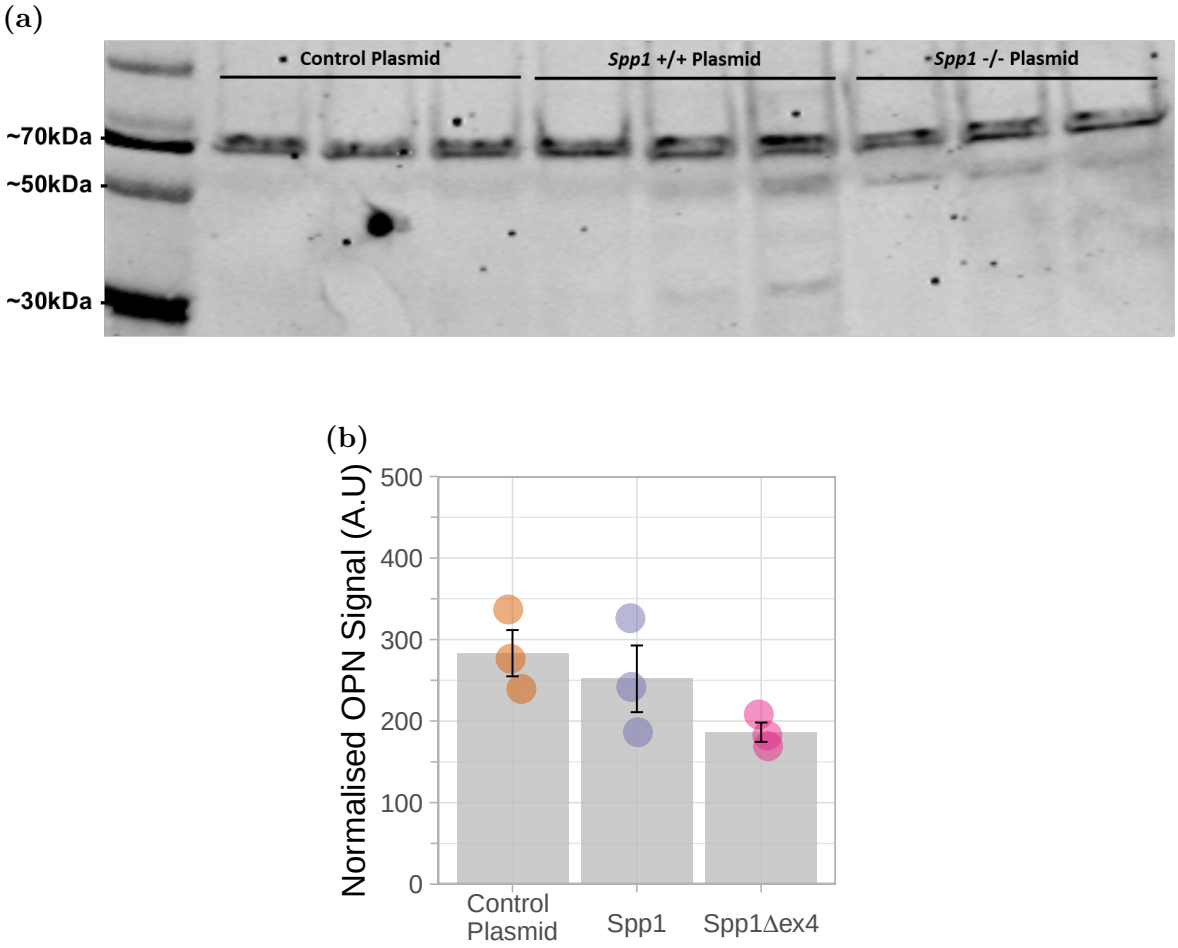



**Figure 6.16: *Spp1* Protein Assay in Transfected H9c2 Cells** Immunoblotting confirmed the presence of *Spp1* protein, OPN, in H9c2 cells following 48-hours transfection with control plasmid, plasmid containing *Spp1* from wild-type (Spp1) and knock-out (Spp1Δex4) SHRSP-*Spp1*$^{em1Mcwi}$ neonates. OPN signal was normalised to total protein. One-way ANOVA was performed on normalised OPN values and significant post-hoc comparisons are displayed in plot area (*$p$<0.05, **$p$<0.01, ***$p$<0.001, **** $p$<0.0001)

## 6.4 Discussion

Phenotypic differences present at neonate, 5-, and 16-weeks between Glasgow WKY and SHRSP strains were discussed in Chapter 3. As the SHRSP-*Spp1*$^{em1Mcwi}$ is maintained as a heterozygous colony, only 25% of each litter (in a perfect Mendelian ratio) are homozygote wild type (+/+) or knock-out (-/-), increasing the number of litters required to meet suitable power. Preliminary results indicating *Spp1* protein expression was retained in neonate and male SHRSP-*Spp1*$^{em1Mcwi}$-/- heart and kidneys required closer inspection of the efficiency of *Spp1* knock out. Evidence of *Spp1* protein expression in neonate and adult tissues warranted further investigation of CRISPR/Cas9 knock-out before increasing animal numbers. The study was therefore stopped prematurely to determine the efficiency of expected reduction in functional *Spp1*, and in accordance with the 3R's (Diaz et al., 2020). Whilst targeted genetic disruption in

animal models or cell lines are an effective model to assess gene function, is not always successful (Smits et al., 2019). Residual mRNA and protein expression of target genes can be maintained due to the presence of existing and/unknown weakly expressed splice variants. Different gene isoforms produced following the induction of genetic mutations may or may not retain gene function (Smits et al., 2019; Hosur et al., 2020), affecting the efficiency of intended disruption.

When tissue collection was stopped, data collected at 5-weeks was underpowered in SHRSP-Spp1$^{em1Mcwi}$-/- animals to determine differences in cardiac size or function. There was no clear trend toward improved LVMI in N=2 SHRSP-*Spp1*$^{em1Mcwi}$-/- compared to their wild-type and heterozygous littermates. The SHRSP sub-strain utilised for genetic mutation, SHRSP/A3NCrl, did show adverse cardiac remodelling relative to the WKY/Gla and is confirmed as a good genetic model to test the effect of reducing *Spp1* expression in the context of genetically influenced hypertension.

There was decreased expression of *Spp1* mRNA, assessed by TaqMan qRT-PCR, in the SHRSP-*Spp1*$^{em1Mcwi}$-/- neonate heart, compared to SHRSP/Gla and WKY. This was contrasted with a strong protein signal in immunoblotting, where SHRSP-*Spp1*$^{em1Mcwi}$-/- showed the strongest relative expression when normalised to total protein. The use of genetic disruption by CRISPR/Cas9 relied on the deletion mutation in *Spp1* exon 4, causing a frame-shift, premature stop, and eventual nonsense-mediated decay (NMD) of produced mRNA transcripts. From studies in cell and mouse models of CRISPR/Cas9 induced gene disruption, as many as 50% report rescue of protein expression (Mou et al., 2017; Smits et al., 2019; Tekath and Dugas, 2021a). A rescue of protein expression by skipping exon containing mutations may not always retain function, as truncated proteins may misfold or lose active sites (Mou et al., 2017). As with transfection of SHRSP-*Spp1* containing plasmids discussed in Chapter 5, both the short form of *Spp1* from the SHRSP-*Spp1*-/- and wild-type containing plasmids produced a functional effect in the H9c2 cell assay. Although not quantitative, band intensity in end-point PCR supports reduced expression of *Spp1* full-length transcript and smaller amounts of alternative transcripts in the SHRSP-*Spp1*$^{em1Mcwi}$-/-. This is supported by the reduced *Spp1* expression in SHRSP-*Spp1*$^{em1Mcwi}$, assessed by TaqMan qRT-PCR. Similarly, in human muscle biopsies from Duchenne muscular dystrophy (DMD), the non-canonical, alternatively spliced *Spp1* isoforms were expressed at a much lower level than full length *Spp1* (Many et al., 2016). Protein level of *Spp1* was not quantified in this case (Many et al., 2016), however results herein suggest the mRNA produced was enough to maintain equal OPN protein expression. Whether endogenous levels of *Spp1* transcripts in SHRSP-*Spp1*$^{em1Mcwi}$-/- animals is sufficient to stimulate adverse cardiac remodelling cannot be concluded. The H9c2 cell assay suggested alternative isoforms produced in the SHRSP-*Spp1*$^{em1Mcwi}$-/- heart, likely retain the ability to stimulate an increase in cell size. That being said, the over-expression induced in cell based assays

assumes equal quantities of *Spp1* isoforms are produced in *Spp1* knock-out animals, which was not supported by TaqMan qRT-PCR data from neonate cardiac tissue.

Results presented in previous chapters did not show differences in *Spp1* protein despite robust differences in *Spp1* mRNA expression. There are a number of processes that can regulate the translation of mRNA into protein, including mRNA stability, localisation, relative transcription and translation rates, and protein-turnover (Buccitelli and Selbach, 2020). *Spp1* has been attributed a number of potentially important functions, including regulation of muscle development during post-natal growth and development, and is tightly regulated in response to muscle injury (E.P. Hoffman et al., 2013). In studies of skeletal muscle from Duchenne Muscular Dystrophy (DMD) samples, *Spp1* mRNA expression did not predict OPN protein expression levels (Vianello et al., 2017). The OPN protein is subject to post-translational modifications (PTMs) by glycosylation and phosphorylation, as well as cross-linking and proteolysis (Pagel et al., 2014). These PTMs will be dependent on the biological system and specific stimuli, and may affect congruency mRNA and protein quantification.

In addition to differences at the post-translational level, gene expression is also controlled at the transcriptional level by alternative splicing, where multiple isoforms of a gene can exist, with distinct subcellular localisation and function (Saitoh et al., 1995). The human, mouse, and rat full length *Spp1* isoform contains 7 exons, with the annotation of the human transcriptome determining existence of two alternative protein coding forms of *SPP1* with exons 5 and 4 spliced out to generate OPN-a (*SPP1-203*), OPN-b (*SPP1-201*), OPN-c (*SPP1-202*), respectively. The annotation of the rat does not contain additional isoforms. End point PCR determined CRISPR/Cas9-induced mutation increased production of alternative *Spp1* transcripts in the neonate heart of SHRSP-*Spp1*$^{em1Mcwi}$-/-. The protein form of *Spp1*, OPN, has two highly conserved integrin binding sites (RGD and SVVYGLR), which can be affected by phosphorylation and other post-translational modifications (Schytte et al., 2020). In all human transcripts (OPN-a, OPN-b, and OPN-c), these active sites are conserved within the translated OPN protein. Through these binding sites, OPN interacts with integrin $\alpha$ and $\beta$ receptors, which have been shown to be a potent regulator of embryonic cardiomyocyte proliferation independent of cell attachment (Ieda et al., 2009). Interaction of OPN with integrins expressed on terminally differentiated myocytes, that cannot undergo hyperplastic growth, might be a potential pathway through which OPN overexpression directly induces cell growth in myocytes. In our *Spp1* mutant model, we showed additional transcripts retain these functional domains within an open reading frame not containing exon 4. Transfection with this alternative form of *Spp1* also produced an increase in H9c2 cell equal to that of the wild type.

In humans, the OPN-c isoform (*SPP1-202*) is a 6-exon transcript, missing residues 31-57 from exon 4 compared to the canonical 7-exon transcript (OPN-a). This short-

form of OPN aligned well to the transcript produced in SHRSP-*Spp1*$^{em1Mcwi}$-/- cardiac tissue. The SHRSP genetic background is associated with increased *Spp1* mRNA expression. Potentially analogous to the human OPN-c isoform of OPN, end point PCR identified an additional *Spp1* isoform produced in the rat heart when placed under the selection pressure of reduced canonical *Spp1*. In human breast carcinoma, OPN-c was expressed in 75-80% of cancer tissues but not in normal tissues, and was associated with tumour grade (severity) (Mirza et al., 2008). Transfection of 293T and MCF-7 cells with OPN-a and OPN-c mRNA produced an increase in OPN protein expression in both cell lysates and supernatant (Mirza et al., 2008) which was not the case in H9c2 cells transfected with wild-type and knock-out forms of *Spp1*. A few important methodological differences should be considered, Mirza et al. (2008) performed SDS-PAGE under non-reducing conditions, which was not the case herein. Additionally, immunoblotting and qRT-PCR rely on specific probes targeting regions of interest magnitudes smaller than the whole mRNA or protein molecule, such that differences in affinities of utilised antibodies for various mRNA or protein isoforms can also affect results. In immunoblotting experiments, the C-terminus from cleaved OPN, as well as full length OPN were detectable using the antibody applied herein. Differences in molecular weight conferred by alternative splicing around exon 4 in both human and rat transcriptome occur upstream of the thrombin and MMP cleavage site, in the chemoattractive N-terminus (E.P. Hoffman et al., 2013). A specific antibody directed to the C- and N- terminals of *Spp1* in the rat would be required to elucidate differences in expression of OPN isoforms and confirm the existence of an 'OPN-c'-like transcript in the rat.

In protein lysates from neonate cardiac tissue, bands at 70kDa and 30kDa, representing both the full length and MMP/thrombin cleaved C-terminal, were visible in WKY, SHRSP and SHRSP-*Spp1*$^{em1Mcwi}$-/- strains. However, in H9c2 cells only the full length 70kDa OPN was detectable. Following transfection with plasmids containing *Spp1* cDNA, there was no increase in the amount of full length OPN produced. There was evidence of a cleaved form only in cells transfected with wild-type derived *Spp1* plasmids, although this was not reliably distinguishable from noise. OPN protein is actively secreted from immune cells and remodelling tissues, where it is likely to be cleaved by MMPs into its biologically active isoforms (Yokosaki et al., 1999; E.P. Hoffman et al., 2013). The isoform produced in SHRSP-*Spp1*$^{em1Mcwi}$-/- neonate hearts, which removes exon 4, retains both the active and cleavage sites. Where tissue lysis can encapsulate both cell retained and secreted forms of *Spp1* protein, the collection of protein lysates from adherent cells captures only cellular proteins and 'misses' proteins secreted into the media, either directly or bound in extracellular vesicles. All human OPN isoforms increase inflammatory cytokine expression in isolated skeletal muscle cells, where OPN-a caused a greater increase in cytokine production than OPN-c (Many et al., 2016). Similarly, in carcinoma cells OPN-a induced the

strongest macrophage response (Mirza et al., 2008) and differences in the inflammatory profile of OPN isoforms, likely arises due to the loss of exon 4 or 5 in OPN-c and OPN-b affecting the interaction between OPN active site (SVVGLR, in exon 6) and integrins. Interestingly, promotion of inflammatory response by increasing cytokine production was RGD-dependant and differed between isoforms, whereas treatment of human monocytes with OPN isoforms induced an equal phagocytosis response which was independent of the RGD-binding site (Many et al., 2016).

The multiple isoforms of *Spp1* might fall under different regulatory mechanisms or be differentially affected by cis-acting variants. Over the past 10 years, a number of studies have investigated the role of polymorphisms in *Spp1* promoter regions in association with DMD (Pegoraro et al., 2011; Barp et al., 2015; M. Chen et al., 2020). Barp et al. (2015) report the protective G allele in *SPP1* rs28357094 was associated with fewer cases of dilated cardiomyopathy. These studies did not determine whether polymorphisms were differentially associated with OPN isoforms. Following the identification of SNPs in the *Spp1* promoter region in association with DMD phenotypes, studies have found implicated transcription factors SP1 and TGFβ in the control of *Spp1* expression (Vianello et al., 2017). In the initial phases of muscle regeneration, OPN forms an important part of the inflammatory response and is likely to be beneficial during repair. It has been suggested that targeting OPN-a, but not OPN-c, isoforms in humans might be a useful therapeutic strategy in DMD (Many et al., 2016). Investigation of differences in the pro-inflammatory effects of *Spp1* in the wild type and knock-out SHRSP-$Spp1^{em1Mcwi}$ is a good candidate animal model for determining the effect of different *Spp1* isoforms in the heart. Recent improvements in the rat reference genome provide a unique opportunity to explore effects of polymorphisms that exist between inbred, congenic, and/or mutant strains, providing a more meaningful evaluation of *Spp1* function during the development of complex cardiac phenotypes.

# Chapter 7

# Analysis of Cardiac Transcriptome during Early Development

## 7.1 Introduction

The heart is one of the first organs to develop during embryogenesis. The cardiac loop is developed from Carnegie stage 13–17 (∼28 days) of human foetal development and becomes a 'mature' heart (which undergoes progressive compaction, and atrioventricular and semilunar valve refinement until birth) by Carnegie stage 23 (∼56 days) (Krishnan et al., 2014). In mice and rats, cardiac embryogenesis begins at gestational day 8 and 9 is complete by gestational day 14.5 and 16 respectively (Marcela et al., 2012). In response to an increase in haemodynamic load and changes to cardiac strain, the heart grows rapidly in the first few days of life (Kerckhoffs, 2012). As cardiomyocytes terminally differentiate, the heart transitions from hyperplastic to hypertrophic growth of cardiomyocytes. During neonatal development, the cellular composition of the heart is altered as other cells continue to divide and differentiate, reducing the number of myocytes relative to other cell types. The Hypertrophic Heart Rat (HHR), a polygenetic model of cardiac hypertrophy, shows pathological growth in adulthood is preceded by growth suppression during gestation, followed by increased post-natal 'catch-up' growth which culminates in excessive cardiac hypertrophy (Porrello et al., 2009). Birth is accompanied by a metabolic switch, as cardiomyocytes shift from a reliance on lactate and glucose oxidation to fatty acid oxidation, with a concurrent switch in the production of enzymes involved in these metabolic pathways. In addition to severe perturbations causing congenital heart defects, monogenic cardiomyopathy genes have also been implicated in contributing to polygenetic models of cardiac hypertrophy (Prestes, Marques, Lopez-Campos, Lewandowski, et al., 2018).

Genetic, epigenetic, and external factors influence phenotypes through changing

the transcriptome, proteome and/or metabolome. Assessment of the transcriptome can be used to determine the effect of genetic variance on phenotype as individual or network gene functionality is often tissue or cell specific, and is potentially highly correlated between species (Sollner et al., 2017). Although relatively easier to obtain, studies of the transcriptome in early neonatal life could be confounded by the strong growth stimulus governing foetal adaptation during the transition from the *in-utero* environment into early life. During cardiac development *in-utero*, gene expression drives organisation, growth, and development of the heart, as cells proliferate and differentiate (Pervolaraki et al., 2018; Cui et al., 2019). Reactivation of several developmental genes has been explored as part of the pathological process occurring during cardiac hypertrophy (Ghatpande et al., 1999; Barry, S.M. Davidson, and Townsend, 2008; Ames et al., 2013; Fazilaty and Basler, 2023). In LVH, low oxygen conditions cause increased reliance on glucose metabolism, as a protective mechanism for the heart, mimicking conditions *in utero* (Rajabi et al., 2007). An earlier study using RNA microarray compared the cardiac transcriptome of SHRSP and F344 rat strains at gestational day 20 and showed differences in cardiac gene expression, despite no measurable differences in relative heart weight (Grabowski, Riemenschneider, et al., 2015). This study was limited by the annotation of the RNA microarray platform utilised, where only 40 differentially expressed genes could be reliably annotated (Grabowski, Riemenschneider, et al., 2015). In addition, the use of hypertensive SHRSP and normotensive, but genetically distant F344, could not allow separation of genes increasing susceptibility to LVH independent of susceptibility to increased blood pressure. Differences in the transcriptome occurring as a result of the phylogenetic distance between F344 and SHRSP strains would also have likely masked functional differences causal in the pathogenesis of increased cardiac volume.

There are a number of approaches to bulk RNA-sequencing which are employed depending upon the aims of individual settings (Hegenbarth et al., 2022). Adapted from microarray studies, the first, and most common approach, is analogous to microarray analysis. Once mapped to a reference genome, reads are counted at each gene locus to measure gene expression in each sample. The relative abundance of genes across samples, conditions and/or time-points is then explored using methods such as Limma, EdgeR, or DEseq2. Whilst this approach explores differences in the expression of genes between experimental conditions, it does not explore the potential differences in individual transcript expression and assumes all important genes are annotated in the current assembly. Alternative approaches to quantifying reads obtained from an RNA-sequencing experiment involve additional 'assembly' steps following alignment. The short read RNA fragments are assembled into full length transcripts and expression is quantified by measuring fragments per kilobase of exon per million fragments mapped (FPKM) (Trapnell et al., 2010). Newer installments of software replace FPKM with transcripts per million (TPM), which eliminates some statistical biases inherent to

the measures of FPKM (G.P. Wagner, Kin, and Lynch, 2012). Transcript assembly allows differences in individual transcript expression to be quantified. Using *de-novo* transcript assembly step for transcript-level quantification both improves sensitivity to low expression (X. Liu, J. Zhao, et al., 2022), and allows discovery of potentially unannotated transcripts or genes. Finally, RNA-analysis can be conducted at the 'exon level', whereby the assembly step is removed and differences are inferred between experimental groups using mapping of raw data fragments at the exon level. This type of analysis is less frequently used and is described in Anders, Reyes, and Huber (2012). The aim of transcriptomic experiments should therefore determine the analysis methods employed.

Transcript assembly of short or long read sequences obtained from RNA-sequencing requires alignment to the whole genome, which is computationally expensive and often the time-limiting step in RNA-sequencing pipelines. 'Alignment-free' method such as Kallisto and Salmon, allow transcript-level quantification on standard CPUs offering reduced computational and time requirements. These software programs do not require separate alignment, transcript assembly, and quantification steps, instead relying on 'pseudo-alignment'. Kallisto, introduced in 2015 and widely adopted by the scientific community, compares short-reads to reference transcriptome sequences and directly quantifies abundance using a 'De Bruijin'-graphs constructed from k-mers present. As such Kallisto greatly reduces the time and computational cost of RNA-sequencing, combining alignment, transcript-assembly and quantification steps (Bray et al., 2016).

Implicating potential disease pathways and specific genes relies on accurate quantification of differences in gene/transcript expression, but is dependent on high quality reference genomes, with good functional annotations. In general, an RNAseq workflow consists of; sample and library preparation, single or paired-end sequencing, alignment, transcript assembly, quantification, normalisation, and finally between-group analyses (Table 7.1). Whilst sample and library preparation can affect sequence quality and impact experimental outcomes, general best practices and standardised methods have been adopted by the community (Hegenbarth et al., 2022). However, quantification method used following alignment and assembly steps is more likely to directly affect statistical models and downstream pathway analysis (Corchete et al., 2020; X. Liu, J. Zhao, et al., 2022). The effect of different quantification tools can affect genes with particularly high or low expression more significantly than those with a moderate expression (X. Liu, J. Zhao, et al., 2022). Simple, count-based estimations show reduced correlation with 'true' TPM versus quantification by Salmon (Soneson, Love, and M.D. Robinson, 2015). That being said, quantifying differences between samples at the gene-level is both more statistically robust, and biologically relevant. Compared to gene-level counts by HTseq FeatureCounts, generation of gene-level counts following transcript quantification by Salmon improves the accuracy of gene expression quantifi-

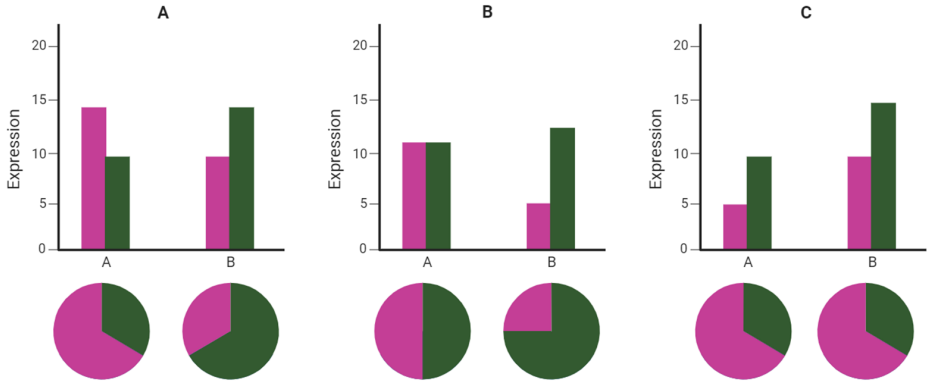
cation in the human and mouse (Kanitz et al., 2015; C. Zhang et al., 2017; X. Liu, J. Zhao, et al., 2022). The increased numbers of transcripts compared to genes may also reduce statistical power and reduce ability to correctly identify biologically relevant differences in expression. To improve integration of RNAseq pipelines, the R-package *Tximport* was developed to import transcript length and abundance estimates from a range of quantification packages, and export estimated count matrices that can be used as inputs to common statistical engines, such as DESeq2.

**Table 7.1:** Overview of RNA Sequencing Workflows

| **Step** | **Step in Pipeline** | **Select Available Software/Programs** |
|---|---|---|
| 1 | RNA extraction and Library Preparation | *NA* |
| 2 | Single or Paired end sequencing | Illumina, PacBio |
| 2a | Alignment | HISAT2, Kallisto, Salmon |
| 2b | Transcript Assembly | StringTie, Kallisto, Salmon |
| 2c | Quantification & Normalisation | HTSeq, Kallisto, Salmon |
| 3a | DGE | DESeq2, EdgeR, Limma |
| 3b | DTU | IsoformAnalyseR, DTUrtle |
| 4 | Biological interpretation | GSEA, ORA, IPA |

DGE: differential gene expression, DTU: differential transcript usage, GSEA: gene set enrichment analysis, ORA: over-representation analysis, IPA: Ingenuity Pathway Analysis™

Traditional analyses of RNA sequencing data focuses solely on differential gene expression (DGE) and do not take advantage of the wealth of data obtained from a single RNA-sequencing experiment. The raw paired-end reads represent the transcriptional landscape of its source and allow exploration of differences between samples in transcript usage. Approaches applying differential transcript usage (DTU) analysis provides information regarding isoform switching and alternative splicing events which may be tissue- or condition specific. DTU analyses work under the null hypothesis that the transcripts contributing to gene expression are proportionally equal between comparison groups. It is possible that potential differences in transcript usage can be masked in differential expression analyses, particularly where overall expression is approximately equal (Figure 7.1).




| Analysis Gene Identified By: | A | B | C |
|---|---|---|---|
| DGE | NO | YES | YES |
| DTU | YES | YES | NO |

**Figure 7.1: Transcript Usage in Differential Expression Analyses** Adapted from (Tekath and Dugas, 2021a), example of cases where differential transcript usage analysis can identify an additional set of genes with significantly different regulation which may be missed from traditional differential gene expression analysis (DGE: differential gene expression, DTU: differential transcript usage).

A recent study combining a number of RNA-sequencing datasets from Alzheimer's and control brains showed a small degree of overlap between genes identified as differentially expressed (Marques-Coelho et al., 2021). Genes that showed evidence of differential transcript usage were enriched for vesicle-mediated transport and synapse-related terms, pathways not observed in enrichment analysis of genes with differential expression only (Marques-Coelho et al., 2021). In a study of early-onset atrial fibrillation, DTU identified alternative splicing in the sarcomere proteins *TPM1* and *LDB3*, as well as the *MYOM1* gene which increased expression of isoforms associated with reduced stability of sarcomere M-bands (Vad et al., 2023). Analysis also found loss of function mutations in splice-regulator *RBM20* driving the increased expression of non-cardiac *TPM1* and *LDB3* isoforms (Vad et al., 2023).

The chromosome 14 congenic and parental strains present a unique opportunity to assess genetically encoded mechanisms of hypertrophy, with and without the confounding susceptibility to high blood pressure. Transcriptomic assessment during late gestation, where the heart is fully formed but not yet under a full haemodynamic load, might provide insights into the foetal gene program promoting LVH. The gestational time period also reduces the influence of the growth stimulus present during the first week of post-natal development. To this end, experimental work in this chapter was supported by funding from the Wellcome Trust Institutional Strategic Support Fund (ISSF) to conduct a transcriptomic comparison between the WKY, SHRSP, and chromosome 14 congenic strains (WKY.SPGla14a and SP.WKYGla14a) at gestational day 18.5 (GD18.5), following the completion of cardiac embryogenesis but prior to full term delivery at day 21.

### 7.1.1 Hypothesis and Aims

RNA-sequencing is an unbiased, hypothesis-free approach to genome sequencing that allows researchers to generate detailed hypotheses that underpin phenotypes. The exploration of early differences in cardiac gene expression was designed to implicate genes and pathways contributing to increased susceptibility to adverse cardiac remodelling developed post-natally.

This chapter explores the secondary hypothesis that transcript-level quantification of short-read RNA-sequencing plays an important role in deriving appropriate gene-level expression counts, and will offer improved quantification over simply counting feature overlap.

RNA-sequencing aimed to:

- Compare methods of alignment and quantification between commonly utilised in RNA-sequencing, namely, HiSAT2 alignment followed by either; FeatureCounts, StringTie *de novo* transcript alignment, or alignment-free transcript assembly and quantification by Kallisto.

- Implicate relevant genes and pathways in CVD developed that are altered in chromosome 14 congenic strains using RNA-sequencing of the heart during cardiac development *in-utero* (GD18.5).

- Explore the role of alternative splicing in cardiac remodelling by employing differential transcript usage (DTU) analysis of RNA-sequencing data.

## 7.2 Specific Methods

Timed mating of WKY, WKY.SPGla14a, SHRSP, and SP.WKYGla14a males and females was performed as per Section 2.1.4. Pregnant dams were euthanised on GD18.5 and the uterus was removed. Individual foetal-placental units were collected, dissected and each foetus was weighed and measured. Head weight was assessed and the heart and kidneys were removed, cleaned, and immediately snap frozen in liquid nitrogen. Whole heart tissue was homogenised and RNA extracted using miRNeasy spin columns (Section 2.3.2). RNA concentration and quality was checked using nanodrop spectrophotometer and sent to Glasgow Polyomics for library preparation, sequencing, and alignment.

Sequencing libraries were prepared from total RNA using the Illumina TruSeq Stranded mRNA Sample Preparation Kit (Illumina, San Diego, USA) which includes ribosomal RNA depletion using Ribo-Zero rRNA reduction chemistry. All remaining RNA species (mRNA, ncRNA, miRNA) were retained for sequencing. Libraries were

sequenced in 75 base, paired-end mode on the Illumina NextSeq 500 platform. Raw sequence reads were trimmed for contaminating sequence adapters and poor quality bases using the programme FastP (S. Chen et al., 2018). Bases with an average Phred score lower than 15 were trimmed. Reads that were trimmed to less than 54 bases were discarded. The quality of the reads were checked using Fastqc (Wingett and Andrews, 2018) before and after trimming. Sequenced reads were checked for contamination using the Fastq-Screen2 programme, which selects a sample of the reads and aligns them to a selection of reference genomes routinely sequenced at Glasgow Polyomics. Following QC, trimmed reads were aligned to the mRatBN7.2 reference genome using HISAT2 (D. Kim et al., 2019). Quality control and alignment reports were combined for all samples and presented in one report by MultiQC (Ewels et al., 2016).

### 7.2.1 Workflows of Quantification

Gene and transcript level quantification was determined and compared across 3 pipelines (Figure 7.2). Pipeline 3 combines the traditional 2-step alignment and transcript assembly step to generate transcript level expression data using pseudoalignment software Kallisto (Bray et al., 2016). Rather than perform a computationally expensive alignment to the whole genome, kallisto determines which transcripts each read could have originated from and determines identity using transcriptome de Bruijn graphs (Compeau, Pevzner, and Tesler, 2011).

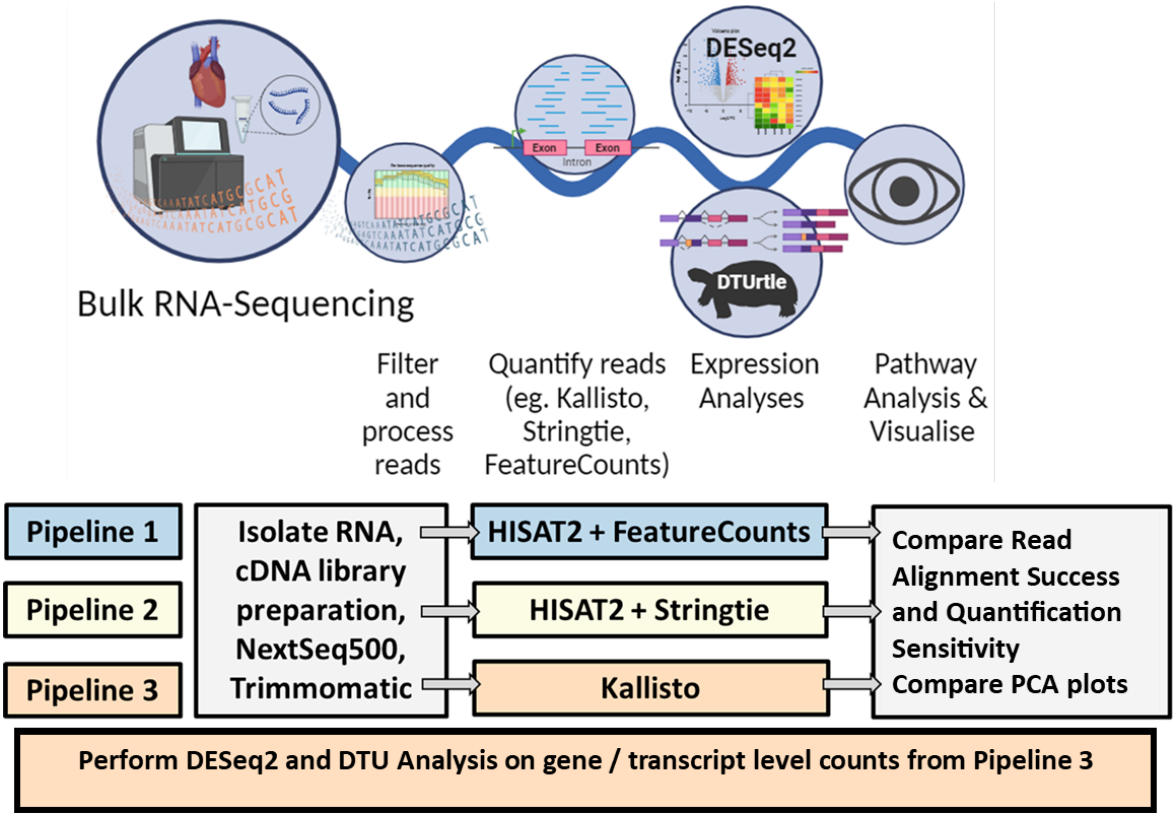



**Figure 7.2: RNA Sequencing Analysis Pipeline** Workflow followed for RNA sequencing. RNA extracted from whole heart tissue of GD18.5 pups (n=12) was sequenced using Illumina NextSeq 500 Platform. Reads trimmed of adaptor sequences were assessed for quality and quantified using 3 different methods. Pipelines 1, 2 & 3 were compared for alignment efficiency (time, success rate) and ability to quantify the expression of genes within the mRatBN7.2 transcriptome. Differential gene expression (DGE) and differential transcript usage (DTU) analyses were carried out in R, using the data derived from Pipeline 3. Figure was created using BioRender.com

Pipeline 1 was performed entirely by Glasgow Polyomics. Alignment of short reads to the mRatBN7.2 genome using splice sensitive aligner HISAT2 was followed by quantification by simple gene counts or *de-novo* transcript assembly (Figure 7.2). In pipeline 1, gene abundance per sample was generated using HTSeq Count (Anders, Pyl, and Huber, 2015), where for each given input file, aligned reads were compared to the reference gene feature file and the number of reads overlapping each gene (or feature) were counted, with a minimum overlap of 1 bp. Reads overlapping multiple features were not counted. In pipeline 2, quantification was performed at the transcript level using StringTie (M. Pertea et al., 2016). StringTie assembles short-reads into transcripts on a per sample basis. Sample-level transcript assemblies were then merged, so counts can be extracted against the same assembly. Assembled transcripts are compared to the rat reference and transcripts assembled without a corresponding reference were assigned 'MSTRG' prefix.

Pipelines 2 & 3 produced transcript level counts. The R-package Tximport was used to compute gene level counts from StringTie and Kallisto output files for use in

DESeq2.

### 7.2.2 Comparison of *de novo* Transcript Assembly to mRatBN7.2

GffCompare (G. Pertea and M. Pertea, 2020) was used to compare the final merged StringTie generated assembly to the mRatBN7.2 reference. Given a query input file, GffCompare will compare and estimate accuracy of the input to a given reference annotation, assigning StringTie assembled 'query' transcripts, a class code which describes the type of relationship between query and reference transcripts as described in Table 7.2. Annotated transcripts which match known isoforms are described by class code '='. Potential novel isoforms are described class codes 'c', 'k', 'j', 'm', 'n', or 'o'. Novel loci are denoted by class codes 'i', 'u', 'y', or 'x'. Potential artefacts and/or mapping errors are denoted by class codes 'e', 's', 'r', or 'p'.

**Table 7.2:** Class Code Descriptions as Reported in GffCompare Documentation

| Class Code | Description |
|---|---|
| = | Complete, exact match of intron chain |
| c | Contained in reference (intron compatible) |
| k | Containment of reference (reverse containment) |
| m | Retained introns, all introns matched or retained |
| n | Retained introns, not all introns matched/covered |
| j | Multi-exon with at lease one junction match |
| o | Other same strand overlap with reference exon |
| i | Fully contained within a reference intron |
| u | None of the above, unknown/intergenic |
| x | Exonic overlap on the opposite strand |
| y | contains a reference within its intron |
| e | Single exon transfrag partially covering an intron, possible pre-mRNA fragment |
| s | Intron match on the opposite strand (likely mapping error) |
| p | Possible polymerase run-on (no actual overlap) |
| r | Repeat (at least 50% bases soft-masked) |

Definitions as described in G. Pertea and M. Pertea (2020)

A measure of the accuracy of StringTie assembled transcripts (query) vs the ref-

erence is also provided. The accuracy, or measure of agreement, is calculated on the input vs the reference and expressed as sensitivity and specificity (precision) using the calculations:

$$Sensitivity = TP/(TP + FN)$$
$$Specificity = TP/(TP + FP)$$

A true positive (TP) was defined as agreement between the query and reference. A false positive (FP) was defined as features from the experimental 'query' that are not in the reference and false negatives (FN) were defined as features from the reference not found in the 'query'. As such, the sensitivity can be decreased in samples where there are many genes which are not expressed within the samples, increasing the number of false negatives. The precision of the assembly can be decreased when there are many unannotated features and is therefore heavily influenced by the quality and completeness of the reference.

### 7.2.3 DESeq2 Differential Gene Expression Analysis

A DESeq2 model was constructed from counts obtained by all three pipelines of analysis. The DESeq2 pipeline contains 5 steps; the first step is normalisation of the data by calculating size factors for each sample using the median of ratios method. This is distinct from the transcript per million (TPM) approach where the between sample variance is used as a normalisation method rather than within sample variance. DESeq2 calculates a gene-level mean ($\mu$) and a size factor estimate ($s$) as the median of ratios to the $\mu$. The second step involves performing gene-wise dispersion estimates using maximum likelihood estimation. In the third step, a curve is fit to gene-dispersion estimates which determines an estimate for the expected dispersion value for genes of a given expression value. In step 4, gene-wise dispersion estimates are shrank towards expected dispersion values. The size of the shrinkage depends on how close gene dispersions are from the curve and the size of the sample. The final step is to fit a generalised linear model (GLM) to each gene and perform a Wald test for significance. Specifically, DESeq2 models individual gene counts ($K_i$) following a negative binomial distribution ($\gamma$-Poisson distribution) with normalised mean; $\mu_{\mathrm{i}}$ and dispersion; $\alpha_{\mathrm{i}}$.

Contrasts are provided to the model to perform differential expression testing using the Wald test. For each contrast, shrunken estimates of log2 fold change (LFC) account for higher dispersion in low count genes by shrinking estimates toward 0 when genes have either; low counts, or high dispersion. It is worth noting that shrinking LFC does not change the number of genes identified as significantly differentially expressed. The Wald test is used to compare observed LFC to the null hypothesis LFC = 0. The $p$-value returned per gene, is corrected for the effects multiple testing using the

Benjamini-Hochberg (B-H) false discovery rate (FDR) adjustment (Love, Huber, and Anders, 2014).

Three contrasts were applied to the DESeq2 model constructed. Simply put, first the parental strains were compared to one another (SHRSP vs WKY), and each chromosome 14 congenic was compared to its background strain. Contrasts were defined as A vs B, where B reflects the control genotype in each comparison. Both SHRSP and WKY.SPGla14a animals were compared to WKY animals, where WKY animals reflect the natural control. Conversely, SP.WKYGla14a animals were compared to SHRSP animals such that the SHRSP acts as background genetic control for the chromosome 14 congenic strain Table 7.3. Genes were considered to have significantly different expression with a false discovery rate ($p$-adjusted) <0.05.

**Table 7.3:** Comparisons for Differential Expression and Differential Transcript Use Analyses

| | **A** | **B** |
|---|---|---|
| i. | SHRSP | WKY |
| ii. | WKY.SPGla14a | WKY |
| ii. | SP.WKYGla14a | SHRSP |

### 7.2.4 DTUrtle Differential Transcript Usage

Complementary to DGE analysis, DTU defines the proportional changes in transcript composition contributing to individual gene expression. StringTie and Kallisto abundance tables were normalised for DTU analysis using Tximport 'dtuScaledTPM', which re-scales counts by median transcript length. DTUrtle (Tekath and Dugas, 2021b) applies the Dirichlet-multinomial model to estimate the precision parameter from DRIMSeq (Nowicka and M.D. Robinson, 2016). The precision parameter is combined with the group-wise maximum-likelihood estimates of transcript proportions ($\pi_A$ and $\pi_B$) in a likelihood ratio test that $\pi_A = \pi_B$. Following initial DTU calling, a two-stage statistical correction with StageR and post-hoc filtering is applied (Tekath and Dugas, 2021b).

### 7.2.5 Pathway Analysis

Results from DGE and DTU analyses were aggregated and visualised to find targets with potential biological significance. A range of open-source and proprietary software programs exist to perform downstream analysis of complex -omic datasets. Termed 'pathway-analysis', biological significance can be attributed to -omic analyses using functional gene-sets, where genes with shared functions, cellular locations, or protein domains are grouped together into a 'set'. The most basic analysis consists of counting

the number of genes in an input gene set (usually list of genes with DGE or DTU with FDR (padj) $\leq$0.05) that overlap annotated gene sets (Over Representation Analysis (ORA)). The number of overlapping genes from the input set is compared to expected overlap achieved from comparing all expressed genes (termed 'background') to each gene-set. A hyper-geometric test with adjustment for multiple testing is then applied (Huang da, Sherman, and Lempicki, 2009). Genes with FDR (padj) $\leq$0.05 in DGE and DTU analyses were uploaded to SRINGdb, which computes functional enrichment analysis on gene sets from databases including; Gene Ontology (divided into 'biological processes', 'cellular compartment' and 'molecular function'), KEGG, Reactome, and WikiPathways (Szklarczyk et al., 2023). STRINGdb also determines potential enrichment of protein domains and features using InterPro and SMART protein domains (Szklarczyk et al., 2023). The 'Rattus Norvegicus' whole genome was used as background for ORA.

As an extension of ORA, Gene Set Enrichment Analysis (GSEA) was proposed. GSEA does not require an *ad-hoc* cutoff for the inclusion/exclusion of genes, and allows directionality of comparative gene expression to be encompassed in enrichment analysis (Subramanian et al., 2005). Input gene lists for GSEA, therefore, need to be ordered, usually by log fold change. For a given a set of genes (e.g. genes involved in glucose metabolism), GSEA determines whether the members of the gene-set are randomly distributed in the input gene list, or primarily found at the top (positive enrichment score) or bottom (negative enrichment score). There are 3 steps to GSEA; the first is calculation of an enrichment score describing the degree to which a gene-set is overrepresented at the extremes of input/query gene-list. The Enrichment Score (ES) is normalised (NES) to account the size of each gene-set. An adjusted $p$-value, correcting for multiple testing (FDR/padj), is then calculated to determine the significance of each the over-representation score (Subramanian et al., 2005). The open-source R-package, clusterProfiler v.4.0, was utilised to perform GSEA in Gene-Ontology and KEGG databases and the most up-to-date genome annotation for rat was loaded (org.Rn.eg.db). As GSEA relies on a directional, ordered input gene-list, DTU enrichment analyses were confined to ORA. Full list of expressed genes, organised by log2 fold change were used as input for GSEA in clusterProfiler.

For each comparison, differential expression data were uploaded to Ingenuity Pathway Analysis (Qiagen, Manchester, UK) and filtered to include genes with an FDR (padj) $\leq$0.05. IPA consist of independently curated gene sets which are distinct from publicly curated gene sets utilised in open-source GSEA or ORA analysis. IPA employs a custom algorithm to perform GSEA combined with a Fishers Exact Test to detect differences in over-represented pathways. The software has an additional suite of algorithms and tools which employ the Ingenuity Knowledge Base, including the prediction of 'upstream regulators'. This analysis determines likely upstream regulators of gene-

sets through a set of direct or indirect relationships (Kramer et al., 2014). The software uses 2 scores to address the inference problem providing both an 'enrichment score' assessed by Fisher's exact test, and a Z-score. These two scores measure the overlap of observed and predicted regulated gene sets, and assess the match of observed and predicted up or down regulatory patterns respectively (Kramer et al., 2014). 'Upstream regulators' analysis was performed as part of the core analysis pipeline of IPA and data was extracted for plotting using custom R scripting.

### 7.2.6 Data Handling and Analysis

Kallisto, StringTie, and GffCompare were carried out using a Windows Subsystem for Linux (Ubuntu 20). Following transcript assembly, DGE and DTU analyses were carried out using R software, attaching user developed, open source packages as described. IPA analysis was conducted in IPA version 107193442.

## 7.3 Results

RNA sequencing data and generative code was disseminated from Glasgow Polyomics. In addition to the analysis of data obtained through the standard pipeline employed by Polyomics, aligned BAM files were utilised in 2 additional pipelines to quantify the effect on downstream analysis. The pipelines were constructed such that two methods of alignment (HISAT2; whole genome alignment, and Kallisto; a pseudoalignment to the transcriptome) were used in combination with three methods of quantification (FeatureCounts, StringTie and Kallisto).

### 7.3.1 Phenotype at Gestational Day 18.5

The first round of tissue collection for GD18.5 pups was affected by the COVID-19 pandemic. As such, measures of body and head weight, as well as body length, were not recorded in WKY.SPGla14a pups. Table 7.4 details the number of litters and pups collected per litter per genotype.

**Table 7.4:** Pups Collected for GD18.5 Study

| **Genotype** | **Mother ID** | **N Pups Collected** | **N in RNA-seq** |
|---|---|---|---|
| WKY | A8232 | 2 | 0 |
| | A8301 | 10 | 3 |
| WKY.SPGla14a | D1854 | 8 | 1 |
| | D1854 | 4 | 1 |
| | D1854 | 8 | 1 |
| SHRSP | C1884 | 9 | 3 |
| SP.WKYGla14a | F2817 | 12 | 3 |

The litter of 2 pups from A8232 was classified as an unusual pregnancy and the tissue obtained from was not used in RNA-sequencing experiments. Consequently, RNA extraction and sequencing was performed in foetuses' from independent litters only in the WKY.SPGla14a.

Body weight was significantly different between genotypes as analysed by one-way ANOVA. Post-hoc comparisons show WKY pups were significantly heavier than SHRSP and SP.WKYGla14a pups at GD18.5 (Figure 7.3a). When head weight and body length were corrected for body weight, there was a significant effect of genotype where SHRSP and SP.WKYGla14a pups had an increased head:body weight ratios versus the WKY. There was no significant difference in head:body weight ratio of SHRSP and background congenic strain, SP.WKYGla14a. Similarly, body length:body weight ratios were significantly different in one-way ANOVA, where post-hoc comparisons indicated that SHRSP and SP.WKYGla14a pups had a significantly increased body length ratio compared to WKY pups. As in head:body weight ratio, there was no significant difference between SHRSP and its congenic background SP.WKYGla14a strain.

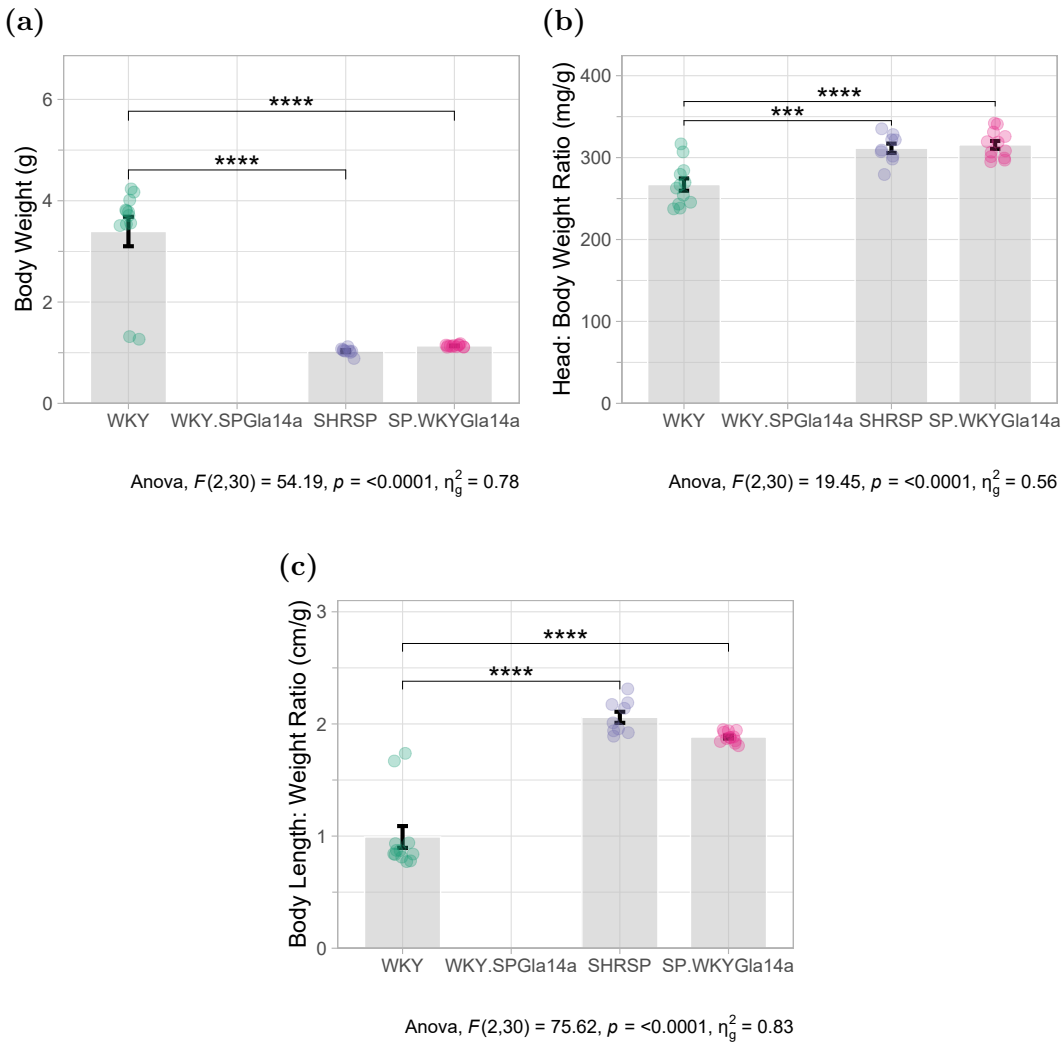



**Figure 7.3: Characterisation of Gestational Day 18.5 Pups** Individual data points and group mean ±SEM are displayed for (a) body weight (g), (b) head:body weight ratio (mg/g), and (c) body length:weight ratio (cm/g) in GD18.5 pups. WKY, SHRSP and SP.WKYGla14a. Strain means were compared by one way ANOVA. Significant Tukey Pairwise comparisons are displayed in plot area (*$p$<0.05, **$p$<0.01, ***$p$<0.001, **** $p$<0.0001). N = 9-12 per group.

### 7.3.2 Comparison of Alignment Methods

MultiQC report indicated HISAT2 successfully aligned ∼90% of paired-end reads to the mRatBN7.2 reference genome in all samples, except C1884-5, which had a slightly lower alignment (around 85%, Figure 7.4a). Only 40% of total paired-end reads were aligned to the transcriptome (as assessed by picard), where the percentage of alignments overlapping untranslated regions (UTRs) and coding mRNA transcripts is quantified by comparing aligned BAM files to input reference genome and transcriptome files. More detailed analysis of read assignment shows a high number of 'intronic', 'intergenic', and 'not aligned' reads contaminated samples (Figure 7.4c). As a consequence of ribosomal depletion from RNA before sequencing, very little rRNA was detected ($\leq$ 1%, Figure 7.4c).

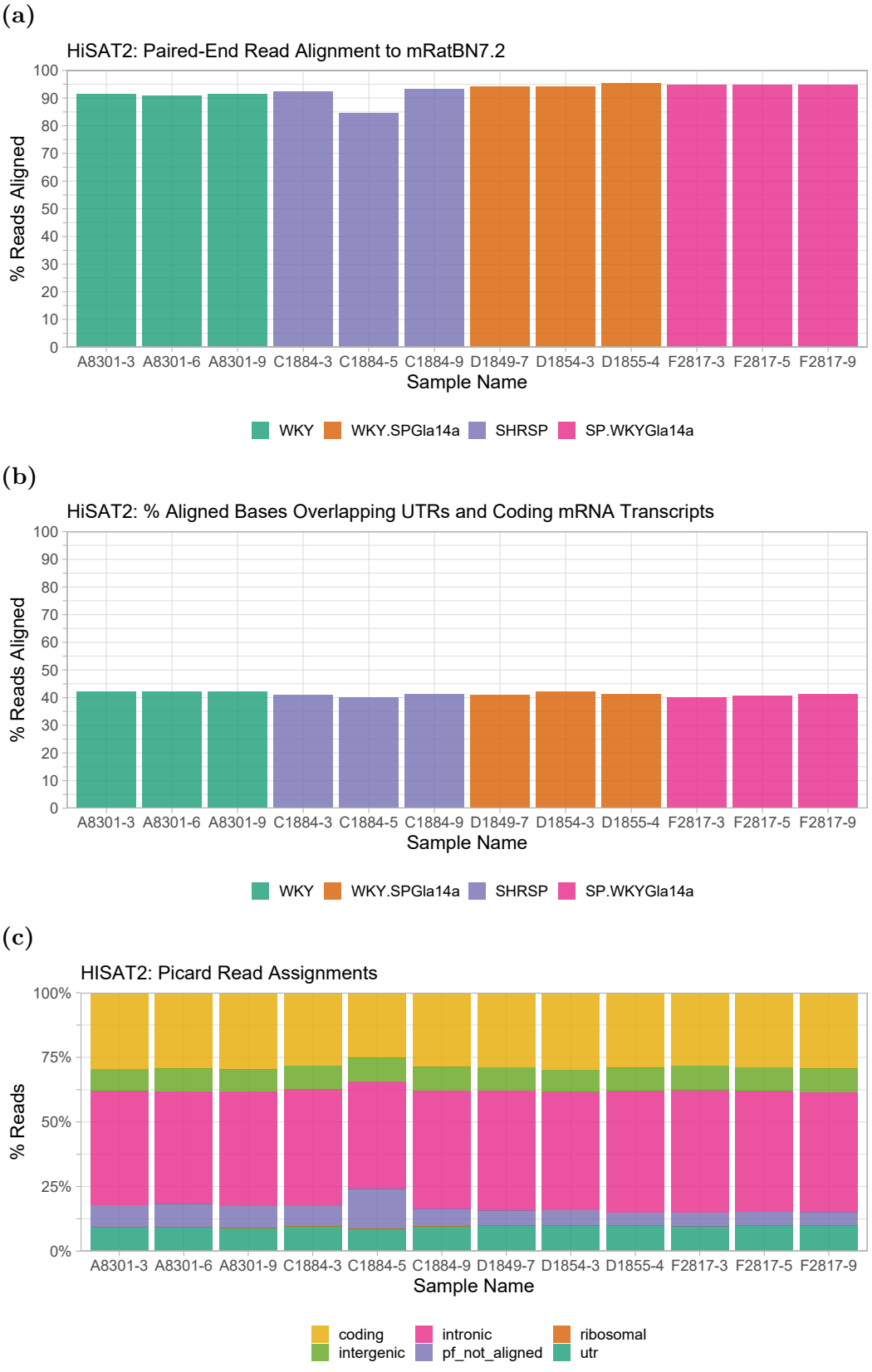



**Figure 7.4: HISAT2 Alignment Quality Metrics** (a) Overall alignment rate of paired end reads to the mRatBN7.2 reference genome (Ensembl Version 110), per sample. (b) percentage of paired end reads aligned to mRNA (including UTRs) as assessed by picard. (c) Picard RnaSeqMetrics toolkit compared aligned BAM files to reference genome and transcriptome to annotate percentage of reads aligned to coding, UTR, intergenic and intronic regions. RnaSeqMetrics also annotated the percentage of non-aligned fragments and the number of ribosomal RNAs present in the sequence files. Animal IDs are coded such that A; WKY, C; SHRSP, D; WKY.SPGla14a and, F; SP.WKYGla14a.

Whilst HISAT2 does not discard reads that do not align to untranscribed regions of the genome, Kallisto utilises a so-called *pseudoalignment* to determine the compatibility of reads to a reference transcriptome. Any reads that do not come from transcripts annotated within the reference transcriptome, are discarded. Kallisto achieved an alignment rate of just under 50% (Figure 7.5), which was similar to that of HISAT2 when considering only mRNA and UTRs as a 'successful' alignment (Figure 7.4b).

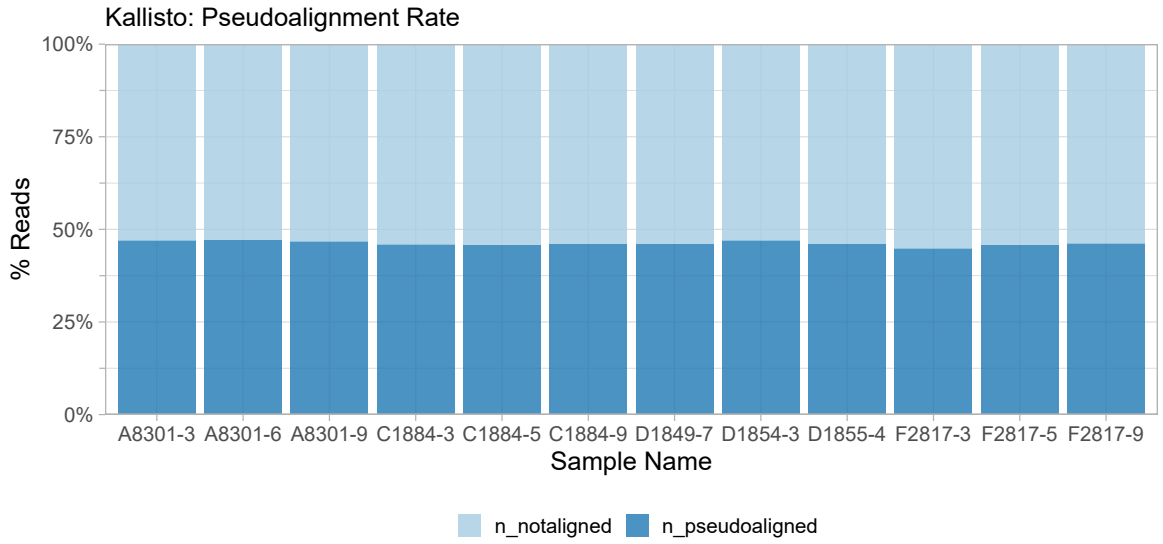



**Figure 7.5: Kallisto Metrics of Alignment and Quantification** Percentage of reads successfully aligned to the mRatBN7.2 transcriptome by the pseudoalignment software Kallisto by sample. Animal IDs are coded such that A; WKY, C; SHRSP, D; WKY.SPGla14a and, F; SP.WKYGla14a.

### 7.3.3 Comparison of Quantification Methods

Kallisto combines both alignment and quantification steps, which traditionally require two specific programmes. Reads aligned by HISAT2 were quantified into gene-level counts by HT-Seq FeatureCounts or assembled into a transcriptome and quantified by StringTie. Gene-level counts were generated and imported into R by Tximport. Before filtering, Kallisto contained count information for all 30,562 genes annotated in the mRatBN7.2 transcriptome, whereas FeatureCounts and StringTie had count information on 24,057 and 30,552 genes respectively. As Kallisto included gene count information for all annotated genes, there were no unique genes identified by StringTie or FeatureCounts (Figure 7.6a). All three methods determined expression values for 24,049 genes in the rat GD18.5 heart. There was only 8 genes FeatureCounts quantified which StringTie did not, and there were relatively more genes (6503) quantified only by Kallisto and StringTie. StringTie failed to count only 10 genes from the annotated transcriptome. The raw mean count distribution (across all 12 samples) was heavily skewed toward 0, particularly in StringTie and Kallisto datasets (Figure 7.6b).

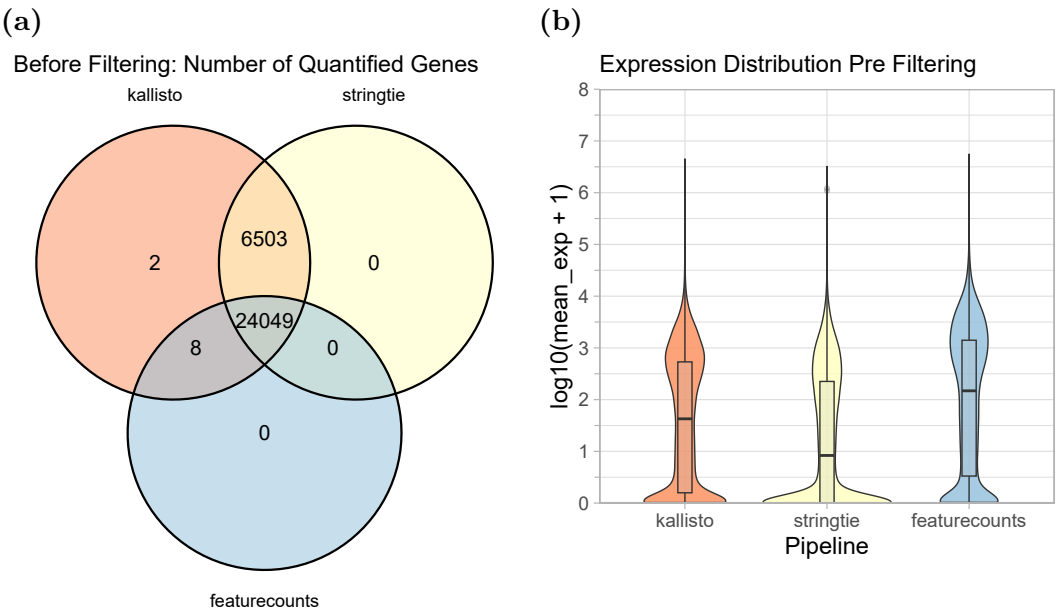



**Figure 7.6: Comparison of Gene Level Quantification** (a) Comparison of number of genes included in raw expression matrices following quantification by Kallisto, StringTie or FeatureCounts. (b) Violin plot of mean expression collapsed across all 12 samples prior to any filtering in the 3 pipelines of analysis, a value of 1 was added to mean values before log10 transformation to avoid generation non-finite value by the log transformation.

To reduce noise and increase power, filtering to remove very low or unexpressed genes (expression count $\leq 1$) and genes with very low expression, is often applied to RNA-seq data. Standard practice in DESeq2 is to apply a filter based on the sum expression of a single gene across all samples, in a minimum of $N_{group}$-1 samples. First a filter was applied which retained genes with a count value $\geq$1 in at least 2 samples. A second filter was then applied such that any genes with a row sum (i.e. across 12 samples) of $\leq$10 were removed. Following filtering, Kallisto quantified the expression of 19,300 genes. As Kallisto initially contained information for the entire transcriptome, approximately 65% of genes were expressed in GD18.5 hearts (Figure 7.7a).

Almost 50% of genes in raw count tables from StringTie quantification were removed by filtering, which left 15,997 genes for downstream analyses. Filtering normalised the distribution of mean counts in StringTie, however both the range of gene counts and median gene counts remained lower than both FeatureCounts and Kallisto (Figure 7.7b). FeatureCounts retained the greatest proportion of genes after filtering (Figure 7.7a). Despite more genes passing expression filters, FeatureCounts still quantified fewer genes than Kallisto, where 17,048 genes were included in expression tables. More than 13,000 genes were commonly quantified by all three pipelines (Figure 7.7c).

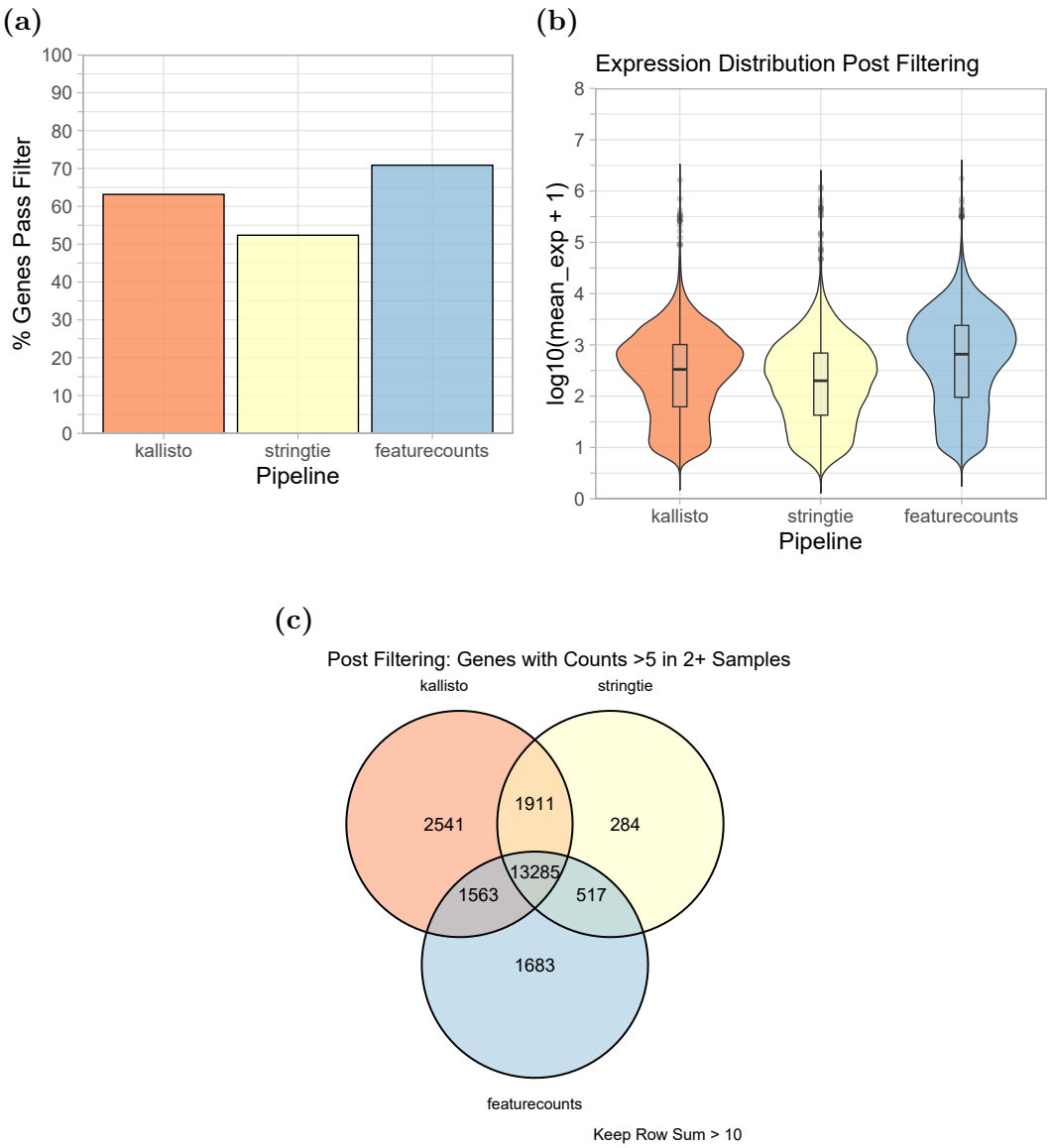



**Figure 7.7: Gene Level Quantification Post Expression Filtering** After removing non- and low expression genes, defined as genes with an expression value <10 across all samples in <2 samples, comparison of gene counts across the 3 quantification methods was repeated. (a) Percentage of genes which passed the filtering thresholds to be included in expression analysis in all three pipelines. (b) Violin plot of mean gene expression across all 12 samples (log10(mean+1)) in all three pipelines. Expression counts were averaged across all samples and log10 transformed for normalisation and visualisation purposes. (c) Venn diagram comparing numbers of genes counted by each pipeline.

Following regularised log (rlog) transformation, principal component analysis (PCA) was conducted on gene level expression data obtained from all 3 pipelines. PCA is a method of reducing high dimensionality data to explore high level patterns of similarity across samples. From an expression data matrix, PCA creates principal components which are ordered according to how much of the variation present in the data they contain.

The 12 samples tend to cluster in space by genotype (Figure 7.8). Gene expression profiles in WKY hearts at GD18.5 is distinct from all strains containing SHRSP

genome, split along the first principal component in all three PCA plots. Due to the higher range of expression values in the Kallisto and FeatureCounts expression matrices, the first 2 principal components account for a higher % of total variance compared to StringTie. In the PCA from StringTie, PC1 accounts for <20% of total variance compared to ∼30% in FeatureCounts and Kallisto pipelines. In contrast, PC2 accounts for ∼15% of total variance in all 3 analyses (Figure 7.8).

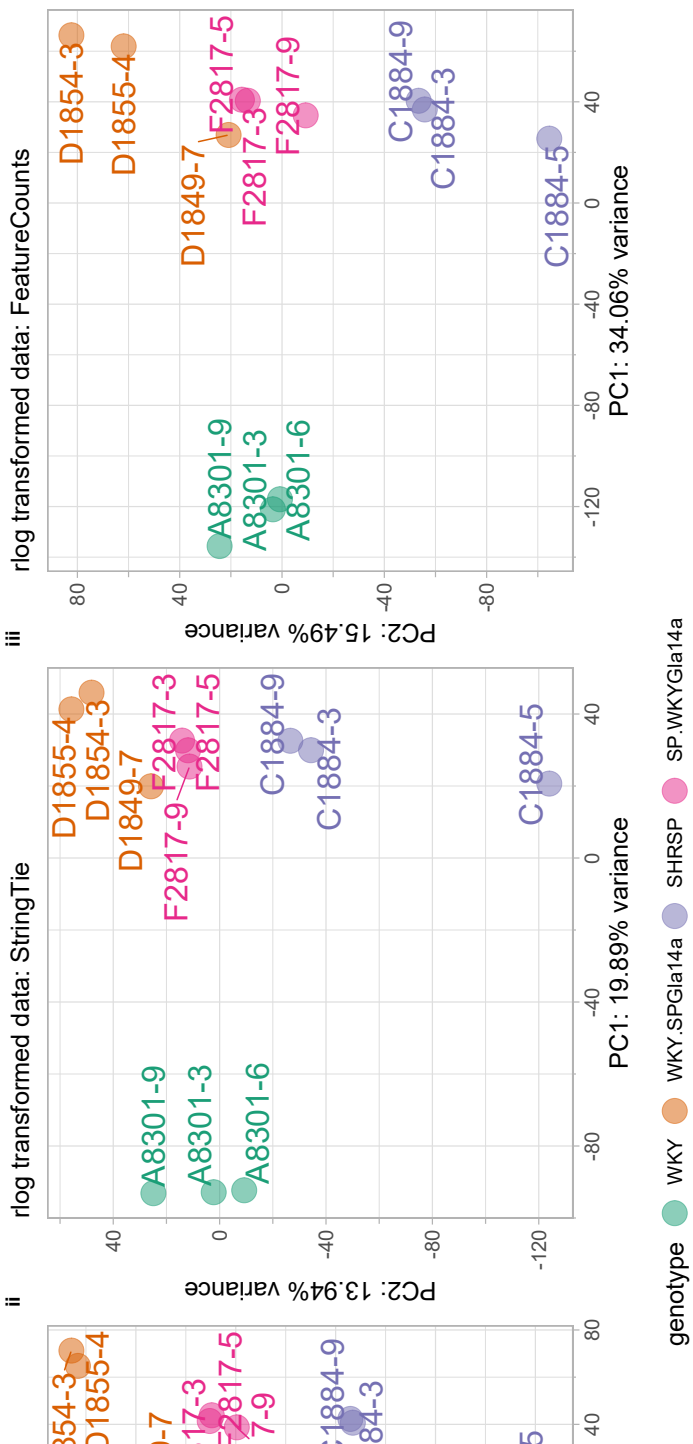



**Figure 7.8: Principal Component Analysis** Graphical summary of principal component analysis shows clustering of samples by genetic background along the first 2 principal components using expression quantification from (i) Kallisto, (ii) StringTie and (iii) FeatureCounts

All genes used as variable inputs to PCA have a 'loading' value which describes their relative contribution to each PC-axis. Loading values range from -1 to 1, with higher absolute loading values indicating a greater influence on the specific principal component. The 10 genes with the highest absolute loading value in PC1 and PC2 are displayed in Figure 7.9. There were only a few genes that appeared across all three quantification workflows. Those in common include *Gata3*, *Carhsp1* and *AABR07071935.1* driving PC2 and *Hspb6*, *Ndrg4*, *Zbtb16* and *Dpysl2* driving PC1. In all cases, the loading values are $<0.03$ and thus individually contribute relatively little to variance in each direction.

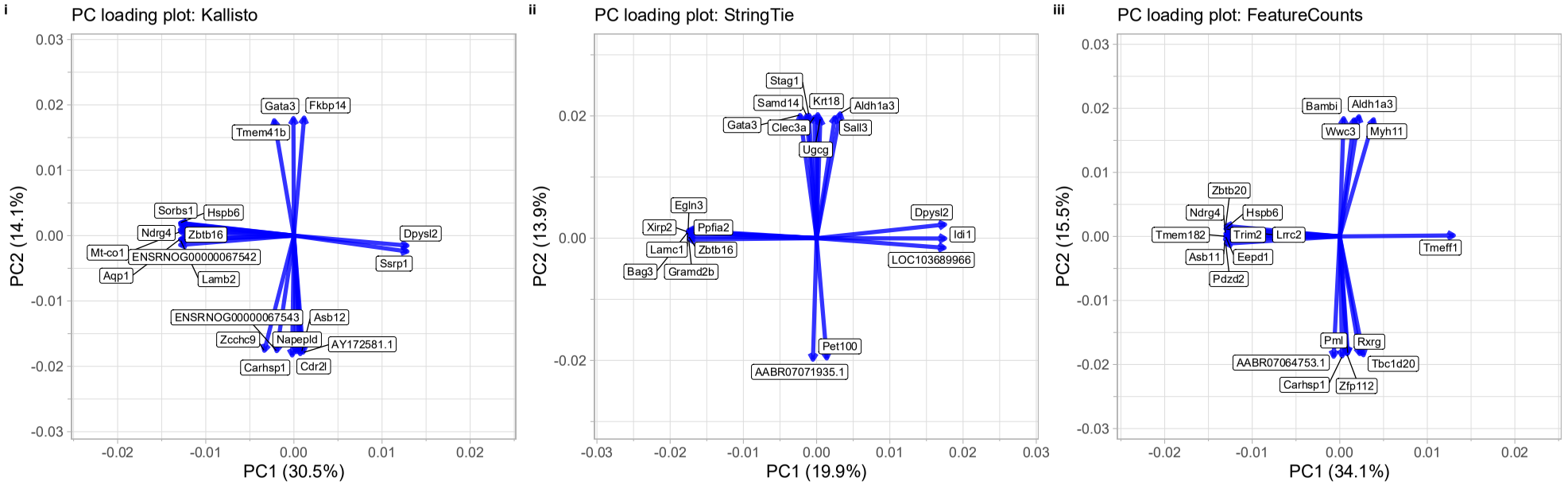



**Figure 7.9: Loading Plots Corresponding to Principal Component Analyses** Loading plot shows n=10 genes with the highest loading scores from PC1 and PC2. Loading values range from -1, +1 such that loadings close to 0 have a weak influence on the component whilst those at the limits (±1) have a strong influence on the component. Arrow head shows space along PC1 and PC2 each gene projects too. Expression matrices obtained using (i) Kallisto, (ii) StringTie and (iii) FeatureCounts were used for PCA analysis.

### 7.3.4 StringTie *de-novo* Transcript Assembly

The rat reference (mRatBN7.2), at the time of writing, has 54,993 reference mRNAs (47,971 multi-exon) at 29,983 loci, comprising 23,141 coding genes (+ 7,421 non-coding or pseudogenes) according to Ensembl Version 110. StringTie assembled 99,370 transcripts (83,556 multi-exon) at 43,466 loci. Assembly of 54,482 annotated transcripts resulted in a sensitivity of 99.1% at the transcript level. However, the high number of unannotated transcripts reduced the precision to just 54.8% at the transcript level. Precision was higher at the exon level; 77.6%.

All query transcripts from the StringTie assembly are given a class code by the GffCompare function (Figure 7.10a). The majority of assembled transcripts were complete, exact matches of intron chains (i.e. annotated transcripts). The next largest category 'j', represents multi-exon transcripts with at least one junction match of an annotated transcript, which are likely alternative transcripts of genes. Taking class codes 'c', 'k', 'j', 'm', 'n', or 'o' as potential novel transcripts, 26.05% of transcripts (n=25,886) represented alternative transcripts of genes already annotated. Taking class codes 'i', 'u', 'y' or 'x' as novel gene loci, 18.59% (n=18,477) of transcripts are potential unannotated rat genes. There was a low error rate as determined by class codes 'e', 's', 'r' or 'p' which represent only 0.53% of assembled transcripts (n=525, (Table 7.5)). Transcript length of annotated transcripts, potential novel transcripts and novel gene loci demonstrate comparable distribution of transcript length, that is relatively longer than likely errors (Figure 7.10b).

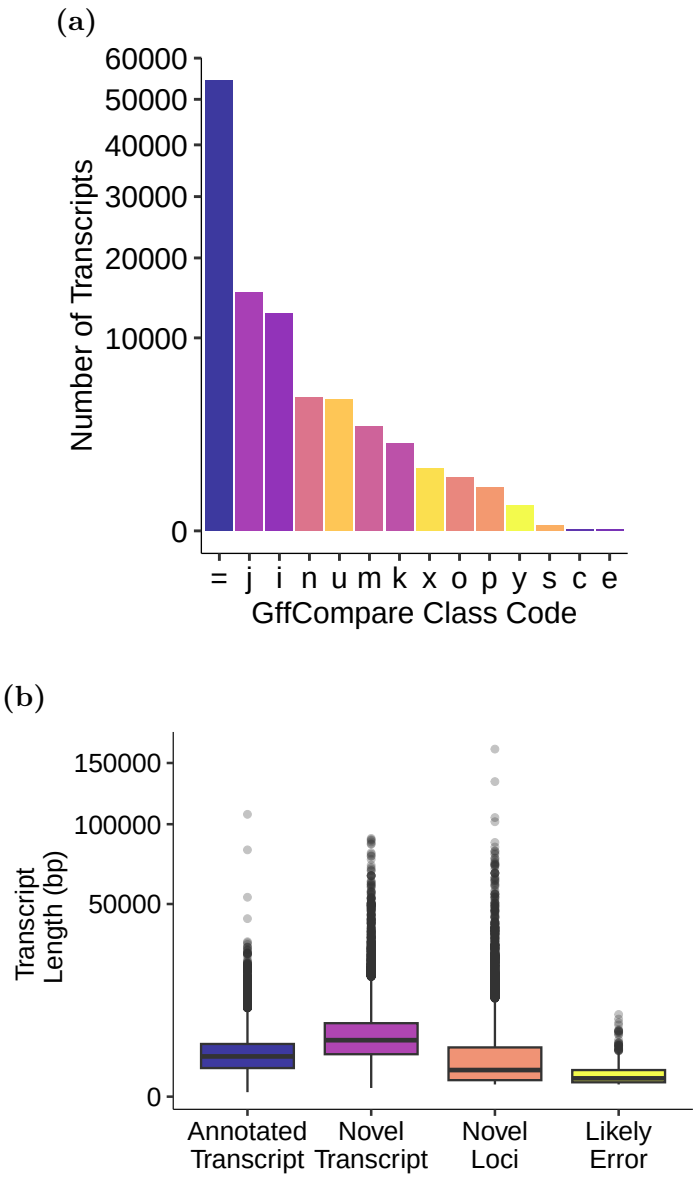



**Figure 7.10: StringTie Assembly Statistics** (a) Class codes assigned to GD18.5 rat *de novo* transcript assembly as constructed by StringTie, compared to the reference mRatBN7.2 obtained from Ensembl. (b) Box-plot summarising length (bp) of StringTie assembled transcripts based on class codes. Annotated transcript; class code '=', Novel Transcript; class code 'c', 'k', 'j', 'm', 'n', or 'o', Novel Loci; 'i', 'u', 'y' or 'x', Likely Error; 'e', 's', 'r' or 'p'.

**Table 7.5:** Results Obtained from StringTie *de novo* Transcript Assembly

| Class Code | Number Transcripts | Percent of Total (%) |
|---|---|---|
| = | 54,482 | 54.83 |
| j | 15,296 | 15.39 |
| i | 12,643 | 12.72 |
| n | 4,801 | 4.83 |
| u | 4,624 | 4.65 |
| m | 2,955 | 2.97 |
| k | 2,056 | 2.07 |
| x | 1,041 | 1.05 |
| o | 777 | 0.78 |
| p | 515 | 0.52 |
| y | 169 | 0.17 |
| s | 9 | 0.01 |
| c | 1 | <0.01 |
| e | 1 | <0.01 |

### 7.3.5 Differential Gene Expression Analysis

A DESeq2 model was constructed for all three quantification workflows. Inspection of model dispersion-estimate plots suggested an improved model fit in the Kallisto quantified data compared to StringTie and FeatureCounts. Combined with the improved diversity of expressed genes in the Kallisto pipeline, contrasts with LFC shrinkage were applied to the quantification data from Kallisto workflow for downstream analyses.

There were 19,300 genes with non-zero read counts which were included in the construction of the DESeq2 model. Contrasts were applied with log fold change shrinkage and Cooks cutoff to produce log fold change (LFC) and Wald test statistics for three comparisons; SHRSP vs WKY, WKY.SPGla14a vs WKY, and SP.WKYGla14a vs SHRSP. In the parental comparison (SHRSP vs WKY), there were 4,638 genes differentially expressed at a significance threshold $\leq$0.05, 2,359 of these were upregulated (LFC >0) and 2,279 were downregulated (LFC <0). LFC values range up to $\pm$15, which is greater than range of LFC seen in comparisons between chromosome 14 congenic and parental strains (Figure 7.11a).

The effect of the chromosome 14 congenic region on the WKY background was greater than on the SHRSP background. The WKY background chromosome 14 con-

genic, WKY.SPGla14a, had evidence of differential expression in 5,300 genes; 2,585 upregulated and 2,715 downregulated (Figure 7.11b). Compared to the WKY, the WKY.SPGla14a showed significantly altered regulation of genes within the congenic region; *Spp1*, *Slit2*, *Nipsnap1*, *Ccdc85a*, *Slc10a6*, and *LOC688553*. Over 3,000 of these changes were shared between the comparison of the SHRSP to the WKY (Figure 7.11d), and included chromosome 14 genes *Spp1*, *Nipsnap1* and *Slc10a6*.

The reciprocal comparison between the SP.WKYGla14a and its background SHRSP strain produced more than 10x fewer significant differences (Figure 7.11c). The SHRSP background congenic showed differential expression of 266 genes. Of those, 146 were upregulated and 120 were downregulated in the SP.WKYGla14a compared to the SHRSP. There were 76 differentially expressed genes that were shared across all 3 comparisons (Figure 7.11d). Volcano plots show *Spp1* is differentially expressed in all three comparisons (Figure 7.11). The number of genes changing in each direction at FDR (padj) $\leq$0.05 and $\leq$0.01 are shown in Table 7.6.

**Table 7.6:** Results of Differential Expression Analyses

| FDR ($p$-adjust) | SHRSP vs WKY | WKY.SPGla14a vs WKY | SP.WKYGla14a vs SHRSP |
|---|---|---|---|
| 0.05 | 4,638 | 5,300 | 266 |
| 0.01 | 3,276 | 3,725 | 87 |

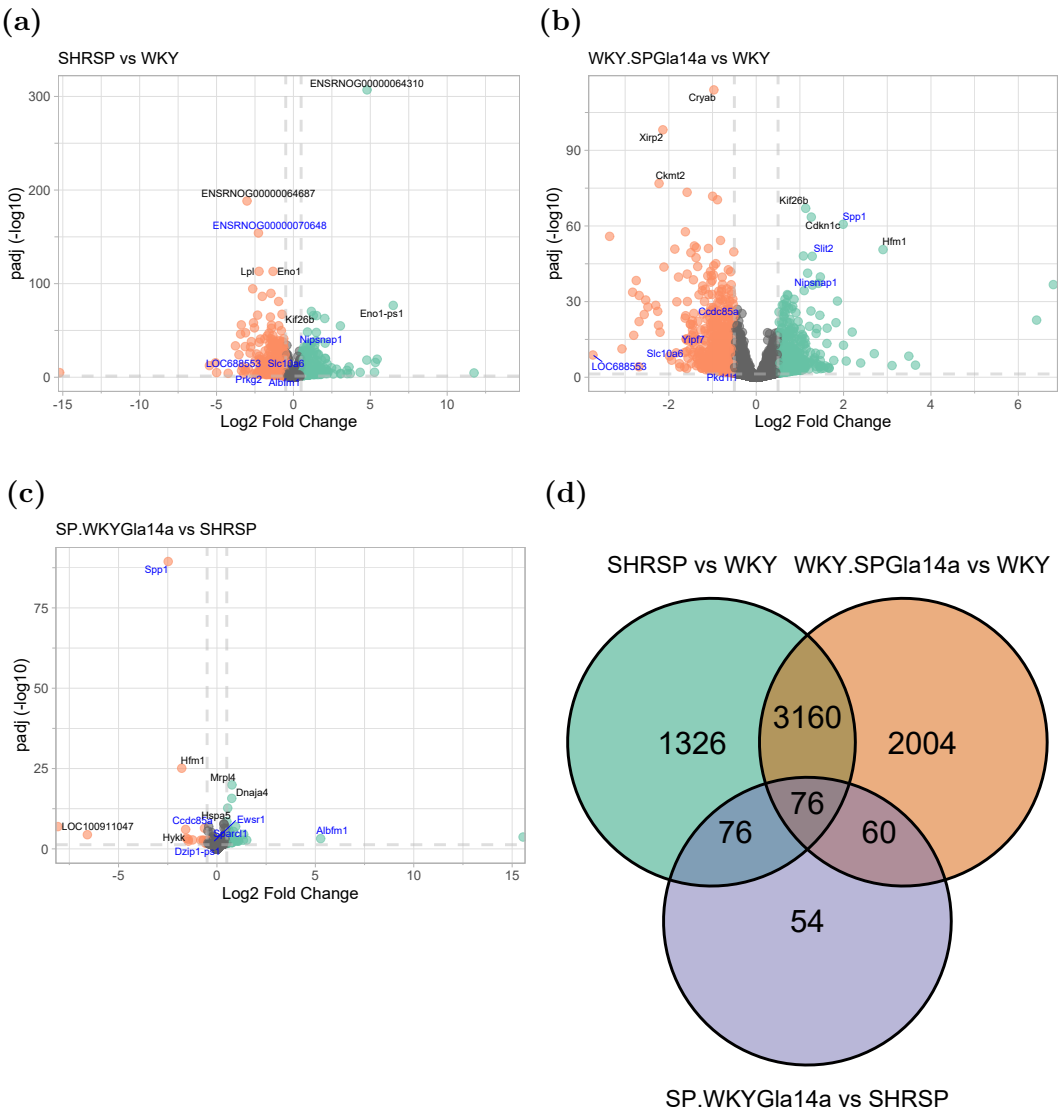



**Figure 7.11: Differential Gene Expression in Congenic and Parental Strains** (a-c) Volcano plots showing genes included in DESeq2 model. Genes with a B-H adjusted $p$-value and LFC $\geq \pm 0.5$ are colored green and orange to show upregulation and downregulation respectively. The 5 genes with greatest LFC in both directions within the chromosome 14 congenic region are labelled in blue. The top most significant 3 genes with LFC $\geq \pm 0.5$ are labelled in black. In plot area, the maximum allowed overlap was set to 5, such that gene labels that excessively overlap are removed. Horizontal grey dashed lines represent -log10 $p$-adjusted value $\leq 0.05$, vertical grey dashed lines represent LFC $\geq \pm 0.5$. (d) Venn diagram comparing number of differentially expressed genes shared or unique to each comparison group.

Over-representation analysis (ORA) of the 76 genes shared across all three comparisons showed significant over representation of GO Cellular Compartment terms 'smooth endoplasmic reticulum' and 'ER chaperone complex' suggesting shared differences in genes which are involved the main functions of the ER, including protein synthesis and transport, protein folding, lipid and steroid synthesis, carbohydrate metabolism, and calcium storage (Figure 7.12a). Although below the level of significance after adjustment, the 'PI3K-Akt'-, 'NF-$\varkappa$B'- and 'TGF$\beta$-' signalling pathways

are also over-represented in shared differentially expressed genes (Figure 7.12a).

TaqMan qRT-PCR was used to validate the upregulation of *Spp1* at GD18.5. Relative quantification of *Spp1* showed a more than 6-fold increase in the heart of SHRSP and WKY.SPGla14a as compared to the WKY and SP.WKYGla14a. The increase was significant between both the WKY and SP.WKYGla14a with both the SHRSP and WKY.SPGla14a. Expression of *Spp1* was not different between the chromosome 14 congenic strains and their chromosome 14 donor strain (Figure 7.12b).

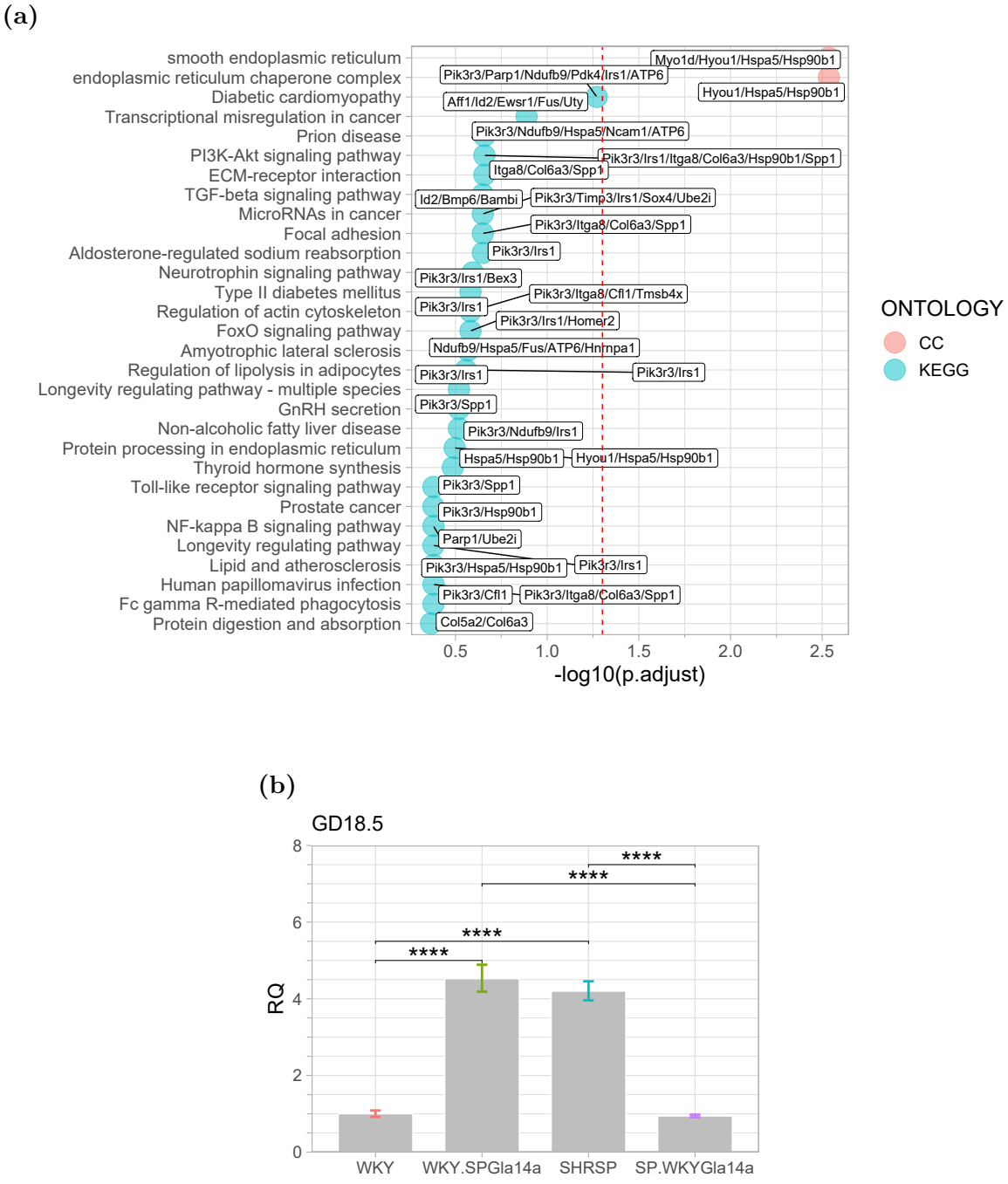



**Figure 7.12: Over-Representation Analysis in Shared Genes and Validation of *Spp1* Overexpression** (a) Over Representation Analysis (ORA) of Gene Ontology and KEGG terms of the 76 genes in common included as gene list, and all expressed genes as background. Negative log10 transformation was applied to adjusted $p$-values for visualisation along the x-axis. Genes in pathway are labelled in the plot. Gene Ontology terms are divided into 'Molecular Function (MF)', 'Cellular Compartment (CC)' and 'Biological Process (BP)'. $p$-values were adjusted using B-H correction and n=25 terms with the lowest $p$-adjusted values are displayed on the plot. Red line is equivalent to $p$-adjusted $\leq 0.05$. (b) qRT-PCR using TaqMan probes for target gene *Spp1* and housekeeper gene *B2m* using RNA from GD18.5 whole hearts utilised in RNA-sequencing experiment. RQ was calculated relative to WKY. *$p<0.05$, **$p<0.01$, ***$p<0.001$, **** $p<0.0001$.

When FDR (padj) was reduced to $\leq$0.01, there were 28 genes with shared differen-

tial expression shared across in all 3 comparisons (Figure 7.13a). Genes with shared differential regulation tend to be regulated in the same direction in the WKY-background congenic (WKY.SPGla14a) and the SHRSP compared to the WKY (Figure 7.13b). The contrasting comparison, comparing the SHRSP-background chromosome 14 congenic to the SHRSP tends to have the opposite directional effect (Figure 7.13b).

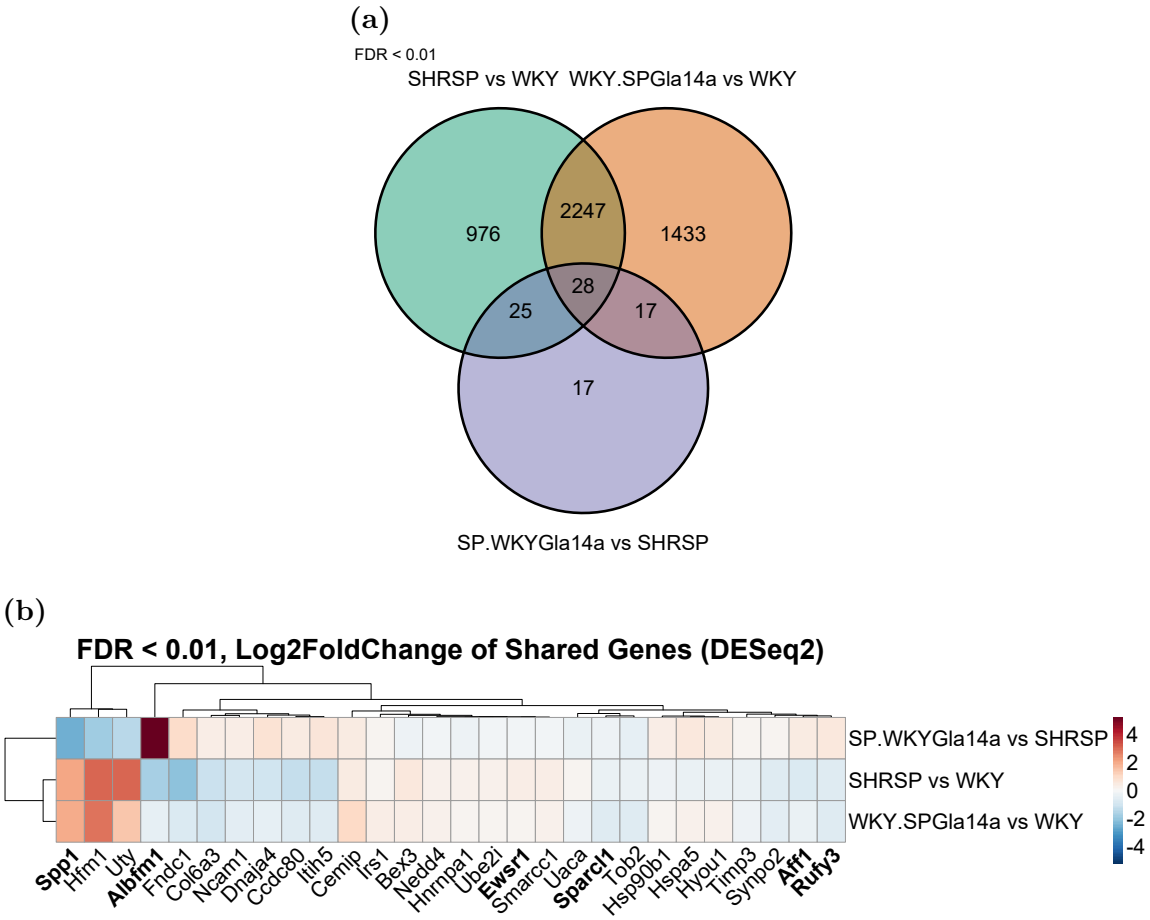



**Figure 7.13: Chromosome 14 Congenic Region in Differential Expression Analyses** (a) Venn diagram of overlapping and unique differentially expressed genes at the FDR ≤0.01. (b) Heatmap of log2 fold change value in each of the 3 comparisons in 28 genes with FDR ≤0.01 shared by all 3 comparison groups. Genes that are located within the congenic region on chromosome 14 introgressed into the WKY.SPGla14a are highlighted by bold type face. Genes are labelled with external gene name, unless no external gene name exists, when ensembl stable ID is used.

To focus on the effective genes within the congenic region, genes within the smaller chromosome 14 congenic region on the WKY.SPGla14a congenic (approximately defined as chromosome 14:5000000-29000000) was used as a filter in pathway analyses. In the SHRSP vs WKY comparison, there were 133 genes differentially expressed at a significance level ≤0.05 within this region. There were 172 differentially expressed genes within the congenic region in the WKY.SPGla14a comparison to the WKY. In the SP.WKYGla14a there were only 13 genes within the defined chromosome 14 region differentially expressed compared to the SHRSP. As with the commonly regulated genes, these genes tended to be regulated in the same direction in the comparison

against the WKY and the opposite in the comparison against the SHRSP. The greatest LFC was in the *Spp1* and *Albfm1* genes (Figure 7.13b).

### 7.3.6 Gene Set Enrichment Analysis of Differentially Expressed Genes

The open-source ClusterProfiler software was used to conduct gene set enrichment analysis (GSEA) in all three comparison groups. GSEA found no significantly enriched GO-terms or KEGG pathways in the SP.WKYGla14a vs SHRSP comparison. Before adjustment for multiple testing, there were 79 and 449 significantly enriched terms in SHRSP and WKY.SPGla14a comparisons to the WKY respectively.

To more effectively filter the enriched terms, terms with $p$-adjust $\leq$0.05 and containing at least one gene from within the chromosome 14 congenic region were considered. GSEA showed a number of terms were similarly enriched in the same direction relative to the SHRSP genome in the parental and WKY-background chromosome 14 congenic. Most shared pathways were in the 'Biological Processes' ontology and tended to be positively enriched (Figure 7.14a). Pathways with positive enrichment included 'DNA replication'/'DNA-templated DNA replication', 'DNA recombination', and a number of processes regulating chromosome conformation (Figure 7.14a). Genes within pathways associated with DNA/RNA processing demonstrated similar patterns of regulation in the SHRSP and WKY.SPGla14a compared to the WKY (Figure 7.14b). These were contrastingly regulated in the SP.WKYGla14a comparison, and contained a number of genes with the approximate congenic region. The genes with the strongest differences in expression were *Hfm1*, *Ooep*, and *Tent5d* (Figure 7.14b).

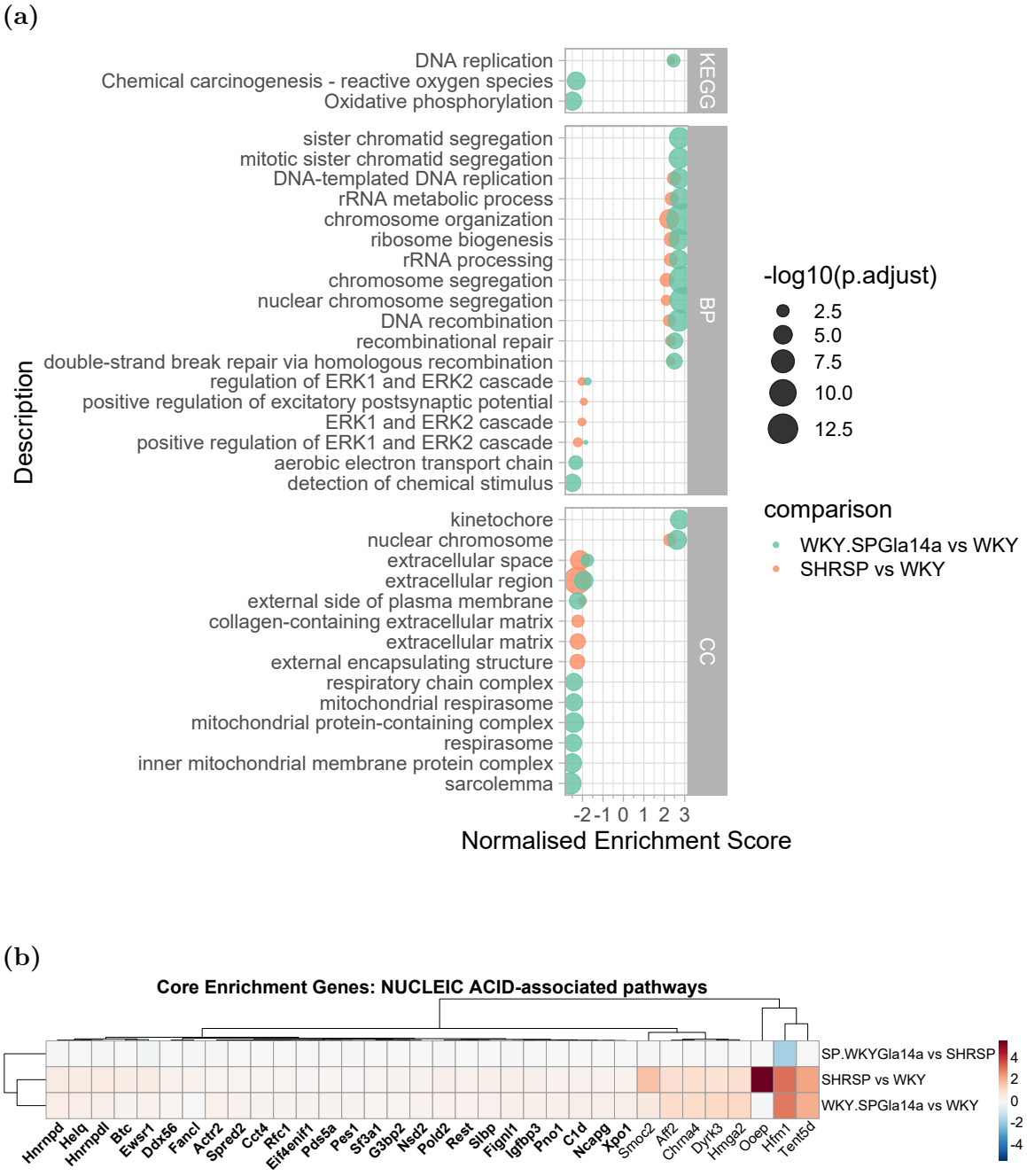



**Figure 7.14: GSEA of Differential Gene Expression Analyses** Gene set enrichment analysis of GO and KEGG terms produced no enrichment in SP.WKYGla14a vs SHRSP comparison. Bubble plot of Normalised Enrichment Score (NES) in terms containing at least one gene within the congenic region with a $p$-adjusted value $\leq 0.05$. (b) Heatmap of LFC in all 3 comparison groups for core enrichment genes involved in pathways containing 'DNA' or 'RNA'. Genes in bold typeface are contained within the WKY.SPGla14a chromosome 14 congenic region.

The most significant, negatively enriched terms in the WKY.SPGla14a comparison to the WKY were 'Oxidative Phosphorylation' and 'Reactive Oxygen Species', supported by a negative enrichment of mitochondria-related cellular compartments and the 'aerobic electron transport chain' (Figure 7.14a). Differentially expressed genes from mitochondrial pathways include several cytochrome-c oxidase subunits (Cox) and NADH-ubiquinone oxidoreductases (Ndufs), which are reduced in the SHRSP and

WKY.SPGla14a compared to the WKY (Figure 7.15a). Changes are in the opposite direction or close to 0 in the SP.WKYGla14a vs SHRSP comparison.

Gene expression profiles from the WKY.SPGla14a and SHRSP compared to the WKY had shared negative enrichment of 'extracellular'-space related cellular compartments (Figure 7.14a). Key genes associated with the 'extracellular'-compartment were mostly downregulated in WKY.SPGla14a or SHRSP comparisons to the WKY (Figure 7.15b). In contrast, there is an upregulation of the chromosome 14 genes in the SP.WKYGla14a vs SHRSP comparison, with chromosome 14 genes *Albfm1* and *Efemp1* showing the strong upregulation in the SP.WKYGla14a compared to the SHRSP (Figure 7.15b). There was a negative enrichment of 'Erk1 and Erk2 signalling' in the WKY.SPGla14a and SHRSP vs WKY, which included the *Egfr* gene in the chromosome 14 congenic region (Figure 7.15c).

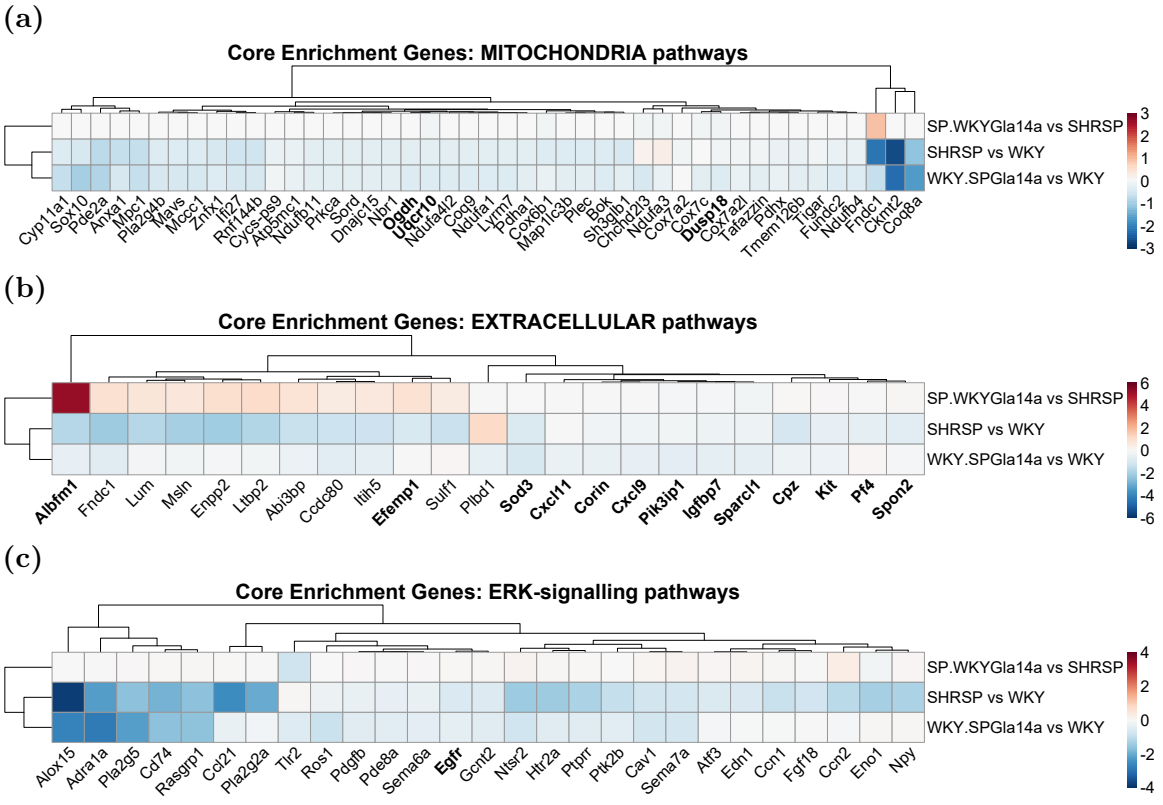



**Figure 7.15: Heatmap of Core Enrichment Genes in GSEA Pathway Analysis** Heatmap of LFC in all 3 comparison groups for core enrichment genes involved in pathways from GSEA analysis. Each heatmap is labelled with core enrichment genes taken from pathways containing (a) 'mitochondria', (b) 'extracellular' (c) 'ERK'-signalling pathways. Genes in bold-face are contained within the approximated WKY.SPGla14a chromosome 14 congenic region.

GD18.5 follows complete cardiac organogenesis in the rat, as such, filtering of terms containing 'development' OR 'cardio' OR 'cardiac' was performed to determine pathways with relevance to the time point investigated (Figure 7.16a). Most of the pathways were only enriched in the WKY.SPGla14a comparison, likely due to the increased

number of differentially expressed genes that reached significance in this comparison. GO 'Biological Processes' of 'muscle tissue development' are negatively enriched in both SHRSP and WKY.SPGla14a comparisons to the WKY. KEGG pathways of 'cardiomyopathy' (dilated, hypertrophic and diabetic) were also negatively enriched in the WKY.SPGla14a vs WKY comparison.

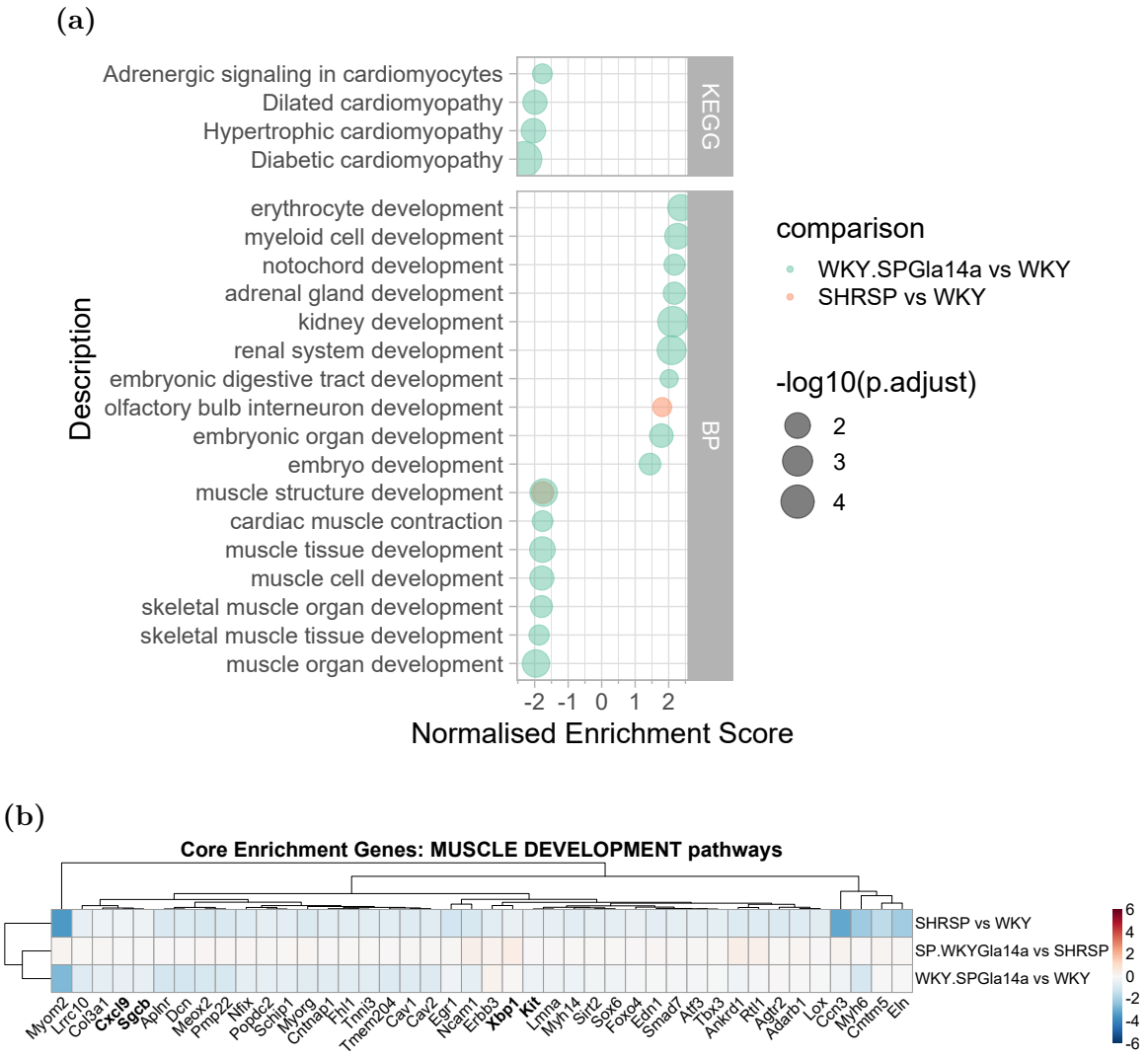



**Figure 7.16: Pathways Enriched for Developmental and Cardiovascular Biology** GSEA of GO and KEGG terms produced no enrichment in SP.WKYGla14a vs SHRSP comparison. (a) Bubble plot of Normalised Enrichment Score (NES) in terms containing keywords 'development', 'cardiac', or 'cardio'. (b) Heatmap of LFC in all 3 comparison groups for core enrichment genes involved in pathways associated with muscle tissue development. Genes in bold-face are contained within the WKY.SPGla14a chromosome 14 congenic region.

Genes annotated within the 'muscle tissue development' pathway tended to be downregulated in the SHRSP and WKY.SPGla14a compared to the WKY whilst there was no change (LFC = 0) in SP.WKYGla14a vs SHRSP comparison (Figure 7.16b).

### 7.3.7 STRINGdb Analysis of DGE Analysis

Ensembl IDs with adjusted $p$-value $\leq 0.05$ were uploaded to STRINGdb and ORA for all categories were downloaded. There were 2,201, 1,944 and 256 unique terms with FDR $\leq$0.05 in SHRSP vs WKY, WKY.SPGla14a vs WKY, and SP.WKYGla14a vs SHRSP respectively (Figure 7.17). There were 198 terms shared by all 3 comparisons, 183 of which contained at least one gene from the approximated congenic region of the WKY.SPGla14a (chromosome 14:5000000-29000000).

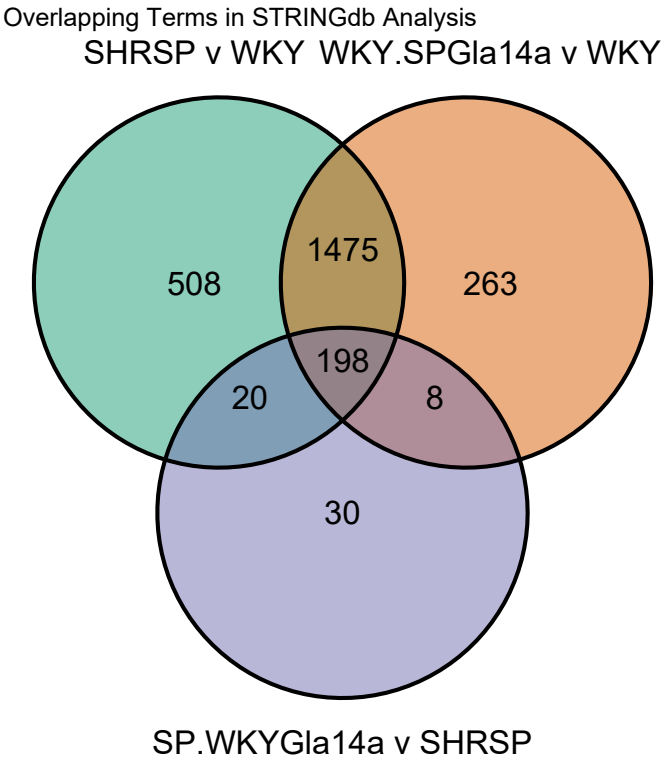



**Figure 7.17: Comparison of Shared and Unique STRINGdb Terms** Venn Diagram of shared and unique terms which had a $p$-adjusted significance $\leq$0.05 in the three comparison groups

Similarly to GSEA analysis, enriched terms that contained at least one gene from the defined chromosome 14 region included a number of terms associated with RNA and DNA binding, cellular metabolism and metabolic processes, response to external stimuli, and ion binding (Figure 7.18). As expected due to the reduced number of significantly differentially expressed genes, these tend to be missing or of nominal significance in the SP.WKYGla14a vs SHRSP comparison (Figure 7.18).

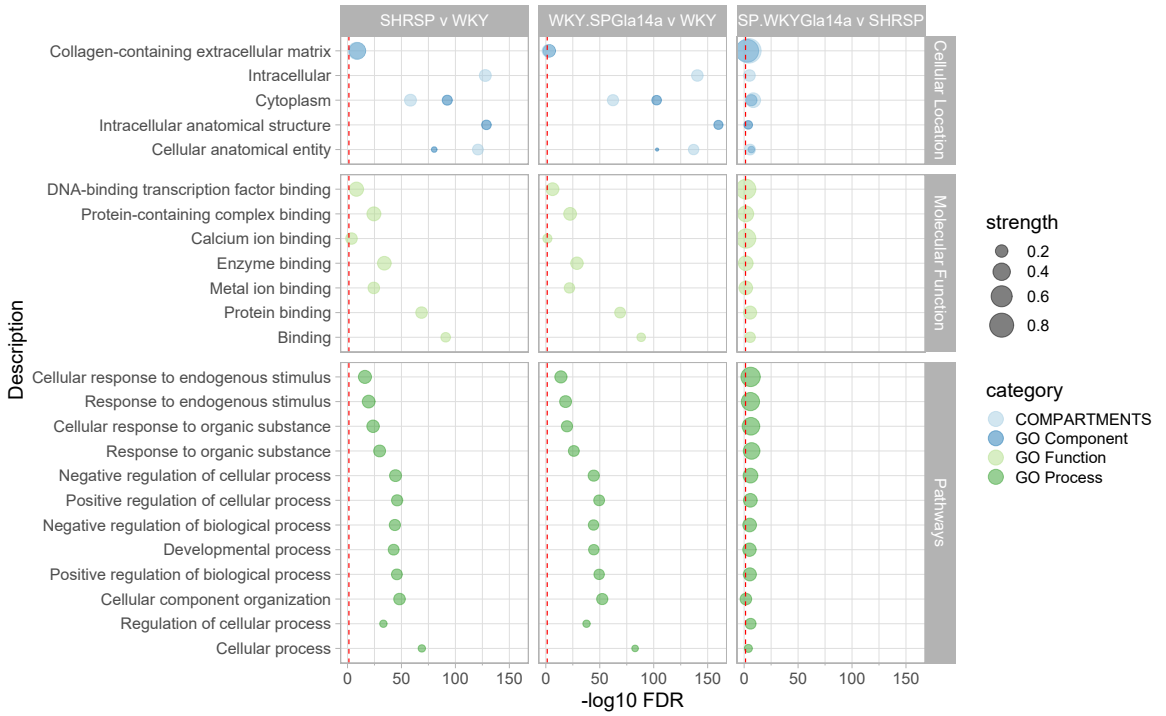



**Figure 7.18: Shared Terms in ORA of Differentially Expressed Genes** Bubble plot constructed using the top 10 shared terms with the greatest strength score in each comparison group, separated by enrichment category. Size of points indicates 'strength' (i.e. the ratio between the number of proteins in each network that are annotated with a term), and the number of proteins that we expect to be annotated with this term in a random network of the same size. The value is log10(observed/expected) and describes how large an enrichment effect is. Red line depicts line of significance, and size of point indicates strength.

Terms not shared in ORA by all three comparisons were mostly shared between the WKY.SPGla14a and SHRSP vs WKY comparisons (Figure 7.17). There was significant enrichment of metabolic processes and transcriptional regulation, as well as terms associated with the extracellular space and ATP-dependant activity. STRINGdb performs ORA based on SMART database protein domains. There were no significant enrichment of specific protein domains in the SP.WKYGla14a vs SHRSP background comparison. The most significantly enriched SMART domains were associated with RNA binding and recognition, growth factor or epidermal growth factor (EGF)-like domains, and zinc-finger binding (Figure 7.19).

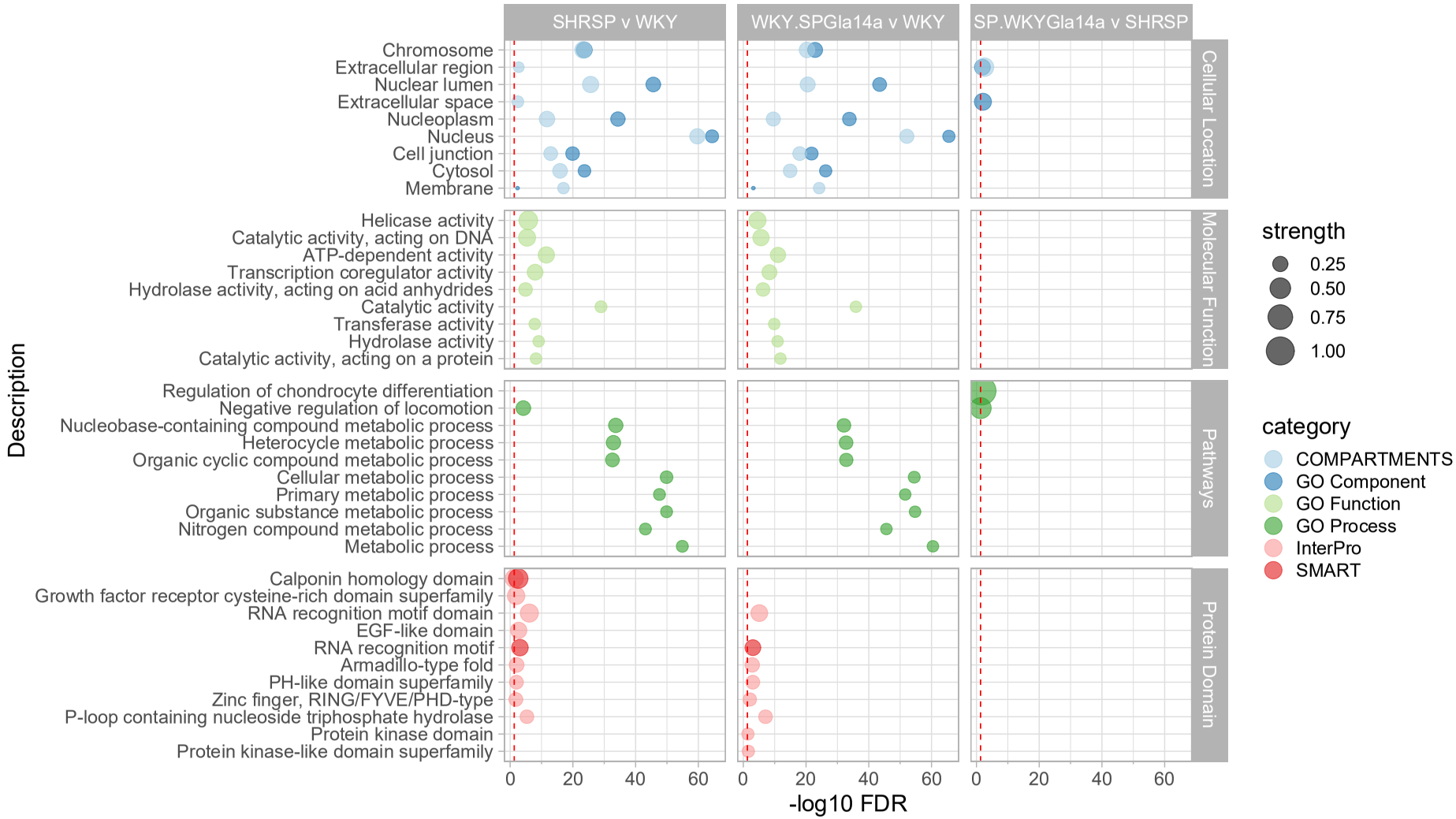



**Figure 7.19: Comparison-Specific Terms in ORA of Differentially Expressed Genes** Bubble plot of the most over-represented terms in each comparison. Size of point indicates 'strength', red dashed line is equal to $p$-value $\leq 0.05$.

### 7.3.8 Ingenuity Pathways Analysis of DGE Analysis

Datasets were uploaded to proprietary software Ingenuity Pathways Analysis to perform upstream regulator analysis. Z-scores suggested activation transcriptional regulators *Myc, CEBPB, E2F1*, and *TFEB* (Figure 7.20) were predictive of changes in the gene expression profiles of the WKY.SPGla14a and SHRSP comprised to the WKY. These transcription factors have a 0 or negative z-score in the opposing SP.WKYGla14a vs SHRSP comparison.

The chromosome 14 gene *EFEMP1*, an extracellular matrix protein, was predicted to be suppressed in both congenic strains compared to the respective background but activated in parental comparison.

The miRNA let-7, phosphatase *Pten*, growth factor *TGF-*$\beta$, and transcriptional regulator *Tp53* had a strong negative z-score, predictive of their suppression in WKY.SPGla14a and SHRSP vs WKY gene expression patterns. $\beta$-estradiol was predicted to be activated in all three comparisons.

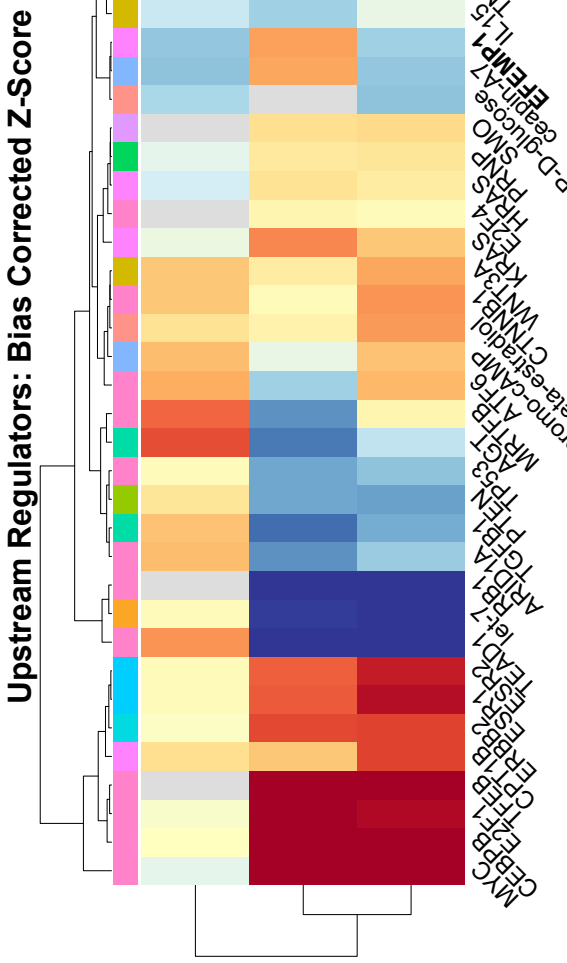



**Figure 7.20: Ingenuity Pathway Analysis of Upstream Regulators** Heatmap of z-scores in IPA analysis of upstream regulators. Z-score infers the activation state of putative regulator where a higher z-score reflects activation, a lower z-score reflects suppression and grey indicate no Z-score was produced in corresponding analysis. Regulators are not limited to transcription factors and is any gene or small molecule experientially observed to affect gene expression. Molecule type is indicated by color along the top of the heatmap.

### 7.3.9 Differential Transcript Usage Analysis

Across all 3 comparisons, there were 552 genes which showed evidence of DTU in DTUrtle, 315 of these also showed differential gene expression in at least of one the 3 comparisons Figure 7.21a. A further 237 genes were implicated in DTUrtle which were not significantly differentially expressed, and were implicated only by DTU.

Between comparisons, only 5 genes with displayed DTU across all three comparisons. An additional 50 genes with evidence of DTU were shared by the SHRSP and WKY.SPGla14a vs WKY comparisons (Figure 7.21b). Construction of a volcano style-plot shows genes with the max difference in transcript proportion included *Zbtb7c, Gpat2, p2rx6, Acp5, Atp4a*, and *Rps24* (Figure 7.21c).

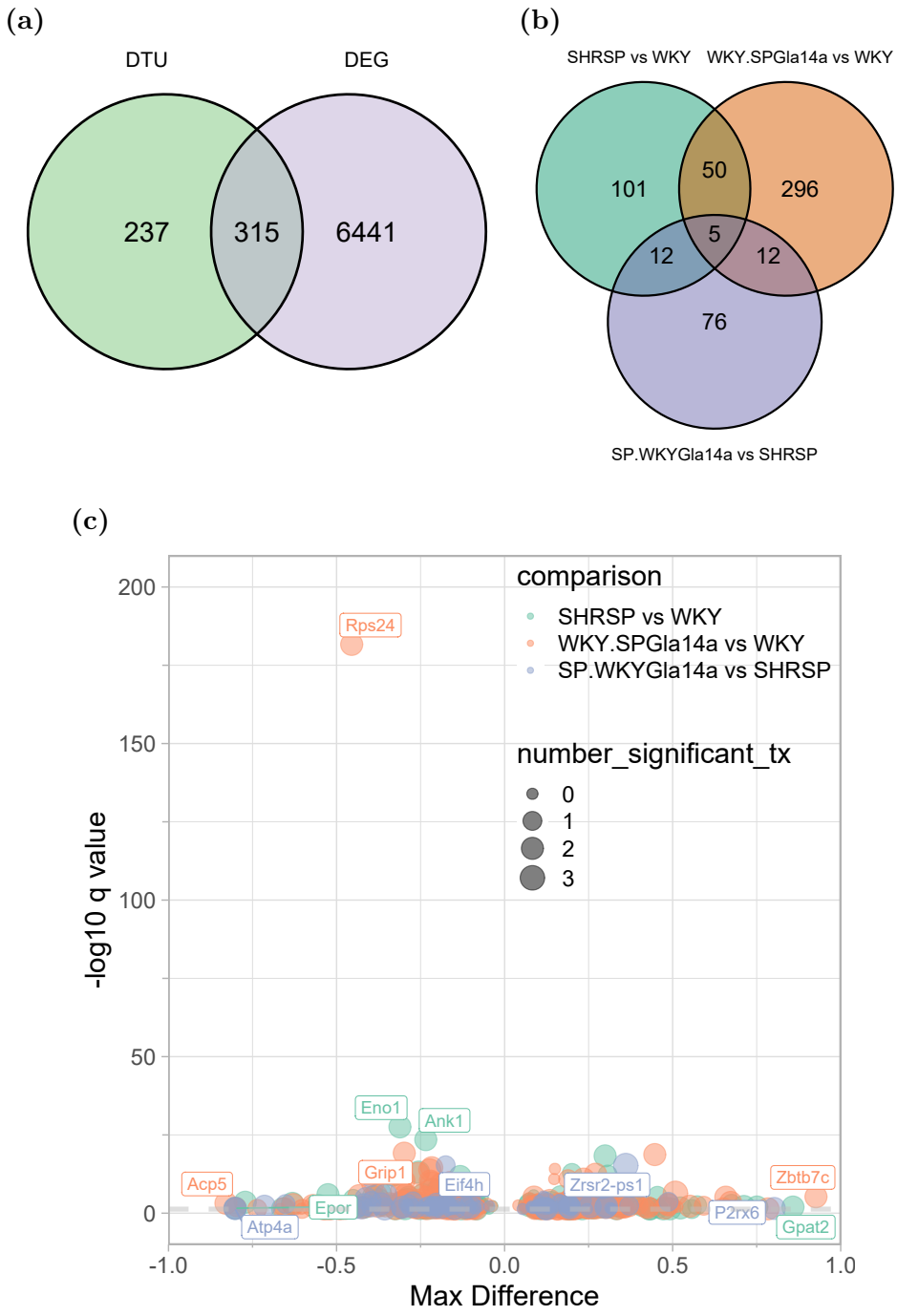



**Figure 7.21: Overview of Differential Transcript Usage in GD18.5 Hearts** Venn diagram of overlapping (a) genes unique or shared between DTU analyses and DGE analyses, and (b) genes unique or overlapping between comparisons groups in DTU analyses. (c) Volcano-style plot of genes with DTU in all 3 comparison groups, each dot represents a gene with $p$-adjusted value $\leq$0.05. Max difference in transcript proportion is plotted on the x-axis. Size of point represents number of transcripts which reached significance in post-hoc testing. Top genes with greatest -log10(q-value) and max-difference in each comparison are labelled on the plot area.

### 7.3.10 Pathway Analysis in Genes with Differential Transcript Usage

Genes with evidence of significant differential transcript usage (q-value (FDR)$\leq$0.05) were uploaded to STRINGdb and all results tables were downloaded for analysis. There

were 155 terms enriched in the SHRSP vs WKY. In the WKY.SPGla14a vs WKY comparison, genes with significantly altered transcript usage were over-represented across 398 terms. In the SP.WKYGla14a vs SHRSP comparison, 69 terms were significantly over-represented.

Terms were filtered to contain $\geq$1 gene unique to DTU analysis (n=237 genes *not* differentially expressed) to focus on additional pathways not highlighted in DGE (Figure 7.22). The top pathways included 'Ras protein transduction' and 'DENN'-domain proteins in the SP.WKYGla14a vs SHRSP comparison. Enriched pathways tended to be shared between the WKY.SPGla14a and SHRSP vs WKY comparisons, including 'nucleotide, protein and ion binding', and 'heart development'.

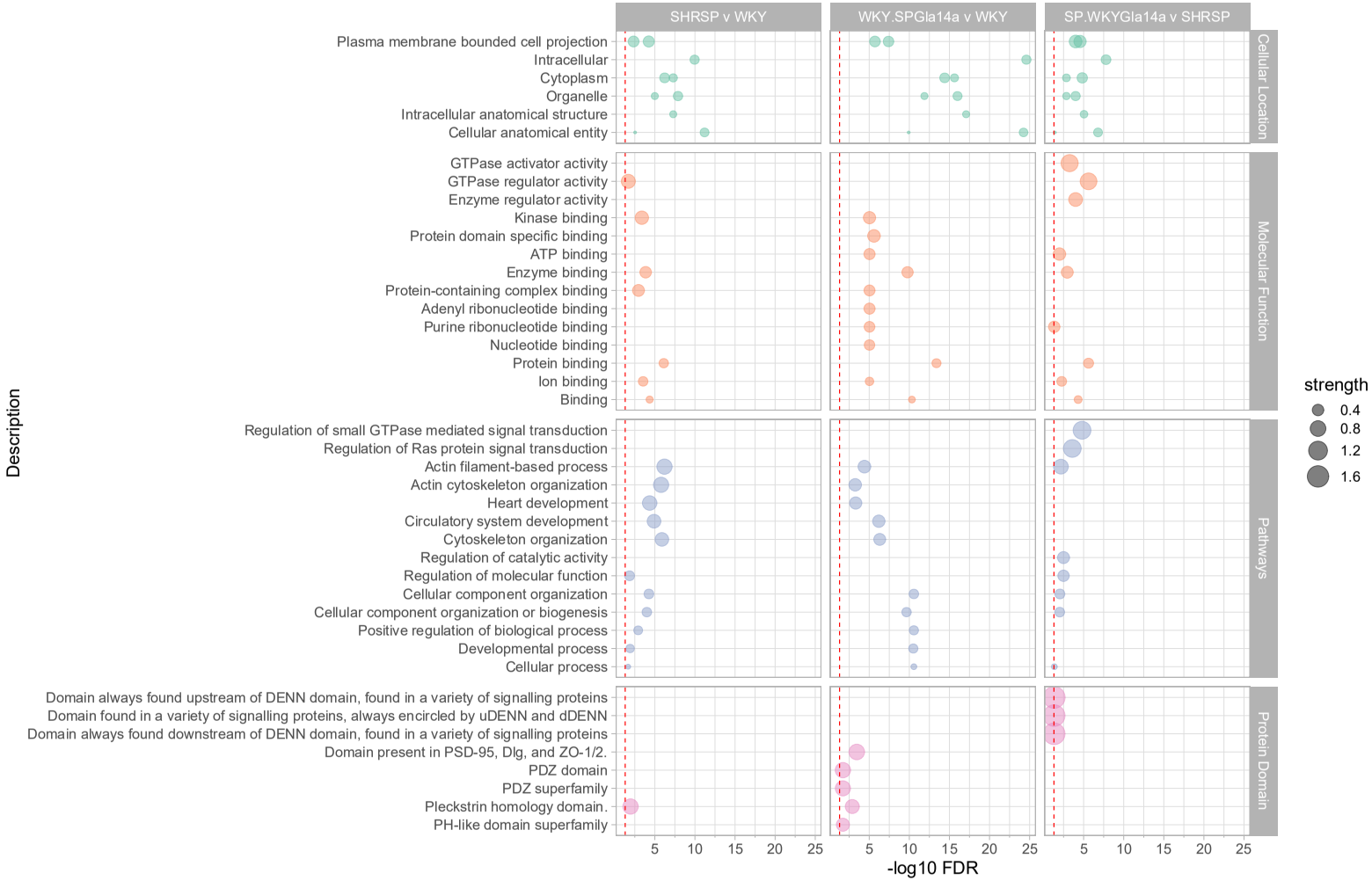
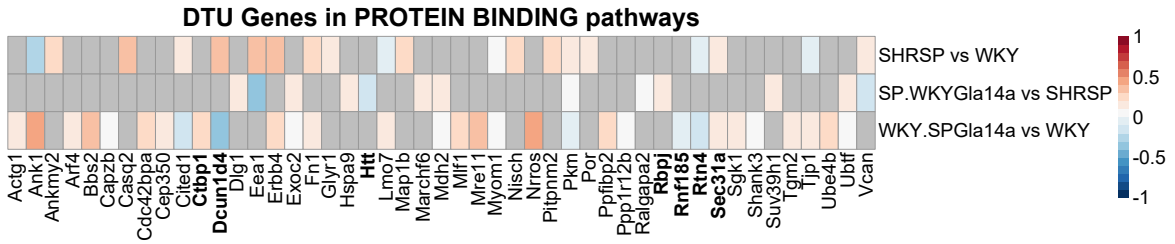



**Figure 7.22: Over-representation Analysis of Genes with Differential Transcript Usage** Bubble plot of enriched terms from STRINGdb in each comparison group, the top terms from each of; cellular location, pathways, molecular function, and protein domain are displayed. Line of significance is indicated by dashed red line and 'strength' of enrichment is illustrated by the size of points.

Inspection of maximum differences in transcript proportion in enriched pathways show that genes in 'protein binding'-pathway included chromosome 14 congenic region genes *Ctbp1, Dcun1d4, Htt, Rbpj, Rnf185, Rtn4*, and *Sec31a*. In the SHRSP and WKY.SPGla14a vs WKY comparisons, *Rtn4*, *Sec31a*, *Erbb4*, and *Fn1* have proportional differences in the same direction (Figure 7.23a). None of these genes reached significance in the opposing SP.WKYGla14a vs SHRSP comparison. *Erbb4* and *Ctbp1* are also annotated within the nucleotide-associated pathways.

Nucleotide-binding associated pathways shows a more strain specific response, with differentially regulated genes which reached significance within the pathway varying between comparison groups (Figure 7.23b). A number of genes were shared with a positive maximum difference in transcript usage between SHRSP and WKY.SPGla14a vs WKY comparisons including *Xylb*, and *Erbb4*. Most of the genes in this pathway were significant in the WKY.SPGla14a vs WKY comparison.

**(a)**




**(b)**

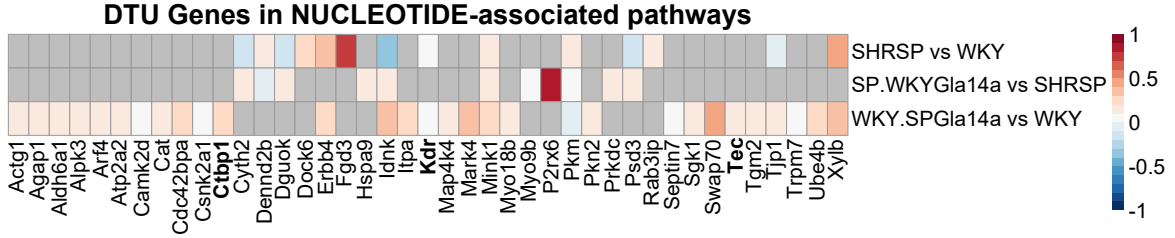



**Figure 7.23: Pathway Specific Maximum Difference in Transcript Usage** Heatmap of maximum-difference in transcript usage from core genes in (a) protein-binding and (b) nucleotide-associated pathways. Genes in the WKY.SPGla14a chromosome 14 region of highlighted in bold type-face.

'Heart Development' was significantly enriched in the WKY.SPGla14a and SHRSP vs WKY comparisons. Genes associated with this pathway tended to show negative maximum differences in transcript usage across comparison groups (Figure 7.24a). The heart development pathway contained a number of genes annotated within the protein and nucleotide binding pathways, including chromosome 14 genes *Rtn4* and *Kdr* and *Erbb4*.

Other cardiac-specific terms were significantly enriched in both WKY.SPGla14a

and SHRSP vs WKY comparison groups, including muscle development terms, cardiomyopathy, and hypertrophy. Core enrichment genes are displayed in Figure 7.24b, and indicate differential transcript usage in calcium -signalling associated genes *Ryr2*, *Camk2d*, and *Cacnb2*.

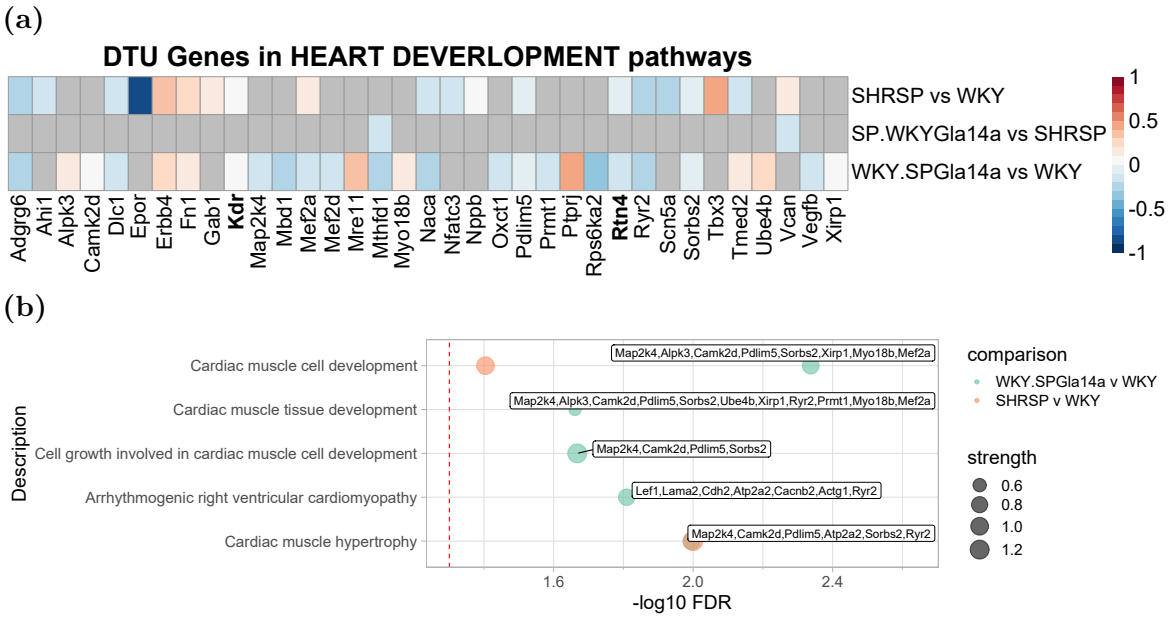



**Figure 7.24: Cardiac-Specific Enrichment of Genes with DTU** (a) Bubble plot of cardiac-associated terms in STRINGdb ORA of DTU genes in all 3 comparisons. SP.WKYGla14a showed no enrichment of cardiac-associated terms. Genes with evidence of DTU within pathways are labelled in the plot area. (b) heatmap of maximum difference in transcript use in genes annotated within the 'heart-development' pathway. Genes in the WKY.SPGla14a chromosome 14 congenic region are highlighted in bold typeface.

## 7.4 Discussion

Foetal parameters at GD18.5 suggested a larger body weight but smaller head:body weight and body length:body weight ratios in WKY pups versus pups from the SHRSP and SP.WKYGla14a strains. Similar differences in foetal body weight have been reported between SHR and WKY rats, (Bassan et al., 2005). However, previous study of Glasgow WKY and SHRSP litters suggest no evidence of foetal growth restriction in pups from SHRSP dams (H. Morgan, 2018). Head weight was also comparatively increased in the SHRSP and SP.WKYGla14a compared to the WKY. Such evidence of 'head-sparing' was not reported in SHR pups where smaller body weights were accompanied by a decreased head circumference and brain weight (Bassan et al., 2005). Differences in body length:weight ratio also support a normal foetal growth rate in the SHRSP and SP.WKYGla14a. Gestation in the rat is relatively short (21 days), and thus growth over a 24-hour period is potentially considerable. Gestational time-points are started from appearance of copulation plug, with twice daily checks in mating cages, where appearance of plug is deemed day 0.5. In this case, litters assessed were from n=1 dams per genotype and thus, there is the potential for 'true' gestational age

to be up to ±12-hours apart. Additionally, time of collection within the 24-hour period determined to be 'GD18.5' was varied relative to conception. Studies of pregnancy disorders mitigate this internal variability by assessing litters from multiple dams, which was beyond the scope of this study. The comparison of foetal growth parameters from the n=1 dam/genotype was likely influenced by the relatively larger variance in 'true' gestational age. Whilst the foetus will grow continually during gestation, the heart is fully formed from GD16 (Marcela et al., 2012), and thus gene expression in all foetuses will represent a fully developed cardiac tissue *in-utero* and comparison of gene expression at GD18.5 in this study should be robust to minor differences in true gestational age.

Following formation of the cardiac muscle *in-utero*, there is evidence that almost 25% of all expressed genes were differentially expressed in SHRSP and WKY.SPGla14a congenic strains compared to the WKY. DTU analysis showed there was an additional, smaller set of expressed genes that were differentially regulated at the transcript level, resulting in a difference in gene-level expression approximately half of the time. Combining DGE and DTU implicated the most diverse and inclusive gene-set to be considered in pathway analyses. The majority of genes differentially expressed at the $\leq$0.05 significance level were shared between the comparisons of the WKY background chromosome 14 congenic and its donor strain, SHRSP, to the WKY (SHRSP vs WKY and WKY.SPGla14a vs WKY). However, this was not the case in DTU analyses, where genes with significant evidence of proportional differences in transcript usage were more likely to be unique to each comparison than shared. PCA of gene expression grouped SHRSP-genome containing strains at opposite ends of the first principal component to the WKY, and in both DGE and DTU analyses, the introduction of the WKY chromosome 14 region into the SHRSP background had a reduced effect compared to the introduction of SHRSP chromosome 14 region into the WKY.

It has been hypothesised that a reversion to the foetal gene program is involved in pathological cardiac remodelling. IPA analysis of upstream regulators implicated gene expression patterns in the SHRSP and WKY.SPGla14a were associated with increased MYC (also referred to as c-Myc), a transcription factor that is increased in response to aortic banding, and acts to downregulate fatty acid oxidation (Ahuja et al., 2010). MYC upregulation was also associated with increased DNA synthesis using an inducible transgenic mouse model (Xiao et al., 2001), and is considered a master regulatory transcription factor activating pluripotency in embryonic stem cells (Fazilaty and Basler, 2023). Conversely, MYC-deficient mice have attenuated hypertrophic response to aortic banding surgery (Zhong et al., 2006). Blocking ET-1 stimulated MYC upregulation in isolated rat neonatal cardiomyocytes reduced ET-1 stimulated hypertrophy (Zhong et al., 2006). A number of pathways associated with DNA and RNA binding and transcription were enriched in GD18.5 hearts, potentially suggesting key differences in

cell proliferative activity and progression through the cell cycle. The upregulation of MYC was also consistent with negative enrichment of oxidative phosphorylation and mitochondrial proteins in the SHRSP and WKY.SPGla14a vs the WKY. It has been hypothesised that MYC can induce cardiac hypertrophy through decreasing mitochondrial function and stimulating glycolysis (H.G. Lee et al., 2009). Despite no difference in cardiac phenotypes in early life, the SHRSP chromosome 14 region is associated with a gene expression profile consistent with decreased mitochondrial content and enrichment of genes involved in oxidative phosphorylation. Genetic predisposition to LVH might thus be associated with a reversion to a disruptive, SHRSP-like foetal gene programme, which emphasises reliance on glycolytic pathways.

In the SHRSP and WKY.SPGla14a, there was evidence of alternative splicing in the type II ryanodine receptor (*Ryr2*), *Camk2d*, *Mef2a*, *Scn5a*, *Pdlim5*, and chromosome 14 congenic region genes *Rtn4* and *Kdr* (previously known as *Flk1*), all of which were over-represnted in 'heart development' pathways. Changes in splicing of *Ryr2, Rtn4, Mef2a, Pdlim5*, and *Camk2d* have been associated with DCM (C.H. George et al., 2007; W. Guo et al., 2012; Beraldi et al., 2014; Maatz et al., 2014; Baralle and Giudice, 2017; S. Liu and J.F. Martin, 2022). Alternative splicing of the *Scn5a* gene has also been associated with myotonic dystrophy and arrhythmia development in DCM patients (Freyermuth et al., 2016). The foetal isoform of human SCN5A displays altered electrophysiological properties, which are analogous to loss-of-function mutations in the SCN5A gene (Freyermuth et al., 2016). In DCM, variants in the RNA-binding motif protein, *Rbm20*, were associated with perturbations during cardiogenesis that resulted in a severe cardiac phenotype and changes to cardiomyocyte morphogenesis when modelled *in-vitro* (Beraldi et al., 2014). Introduction of a 95kb deletion in the rat *Rbm20* gene was associated with reduced survival due to cardiac arrhythmia, increased fibrosis and increased concentric remodelling (W. Guo et al., 2012). RNA sequencing data from samples obtained from both human cardiac biopsies with RBM20 missense mutations, and rat *Rbm20* mutation showed shared dysregulation of genes including *Pdlim5* and *Camk2d*, associated with cardiac remodelling and arrhythmia development (W. Guo et al., 2012). Cardiac transcriptomes of WKY, SHRSP, and chromosome 14 congenic strains showed no differences in *Rbm20* gene expression, although pathway analysis demonstrated RNA and DNA binding was strongly enriched within differentially expressed genes. Differences in the ability of SHRSP and WKY.SPGla14a hearts to mature *in utero* should be assessed in relation to the isoform expression of important developmental cardiac genes such as *Ryr2, Rtn4, Mef2a, Pdlim5*, *Scn5a* and *Camk2d*.

There were 76 genes which were differentially expressed in all three comparison groups. With a strict FDR (padj) $\leq$0.01, differential regulation of >2,000 genes were shared between the SHRSP and WKY.SPGla14a compared to the WKY. GSEA and ORA found significant enrichment of extracellular matrix-associated (ECM) in exper-

imental gene-sets, which was negative in GSEA. A consistent negative fold change in chemokines from the CXC-family was found in the SHRSP and WKY.SPGla14a compared to the WKY. These were contained within the defined chromosome 14 region approximated from the WKY.SPGla14a. The ECM supports the spatiotemporal regulation of cellular processes during cardiac development, with determinant functions for the morphogenesis of specific heart substructures (A.C. Silva et al., 2020). The ECM secretes a number of growth factors involved in the control of cardiac development, including fibronectin, fibrillin, collagen, and heparin-binding EGF-like growth factors which increase cardiomyocyte proliferation *in-vitro* (Ieda et al., 2009). Knock-out of ECM-secreted factors *Fbn1* or *Fn1* can cause severe cardiac defects and embryonic/perinatal lethality (Lockhart et al., 2011). Genes with significantly altered expression were enriched for genes encoding proteins that contain an 'EGF-like domain', including growth factors secreted by the ECM such as fibrillin (*Fbn1*) and fibronectin (*Fn1*). The WKY chromosome 14 region is associated with a higher expression of the fibulin family gene *Efemp1* (EGF-containing fibulin-like ECM protein 1) compared to strains containing SHRSP-chromosome 14. The *Efemp1* gene is contained within the defined congenic region and supports an association of decreased ECM-paracrine function during *in-utero* cardiac development in the SHRSP and WKY.SPGla14a compared to the WKY.

Reduction in ECM paracrine function reduced proliferative capacity of embryonic cardiomyocytes (Ieda et al., 2009). The hypertrophic heart rat (HHR) exhibit a significant, postnatal, 'catch-up' growth as a result a reduced cardiomyocyte number at birth, that were shorter and thinner than controls (NHR) at post-natal day 2 (Porrello et al., 2009). However, in the HHR, the reduced number of cardiomyocytes was attributed to an increase in apoptosis rather than a decrease in proliferative capacity during cardiogenesis (Porrello et al., 2009). These studies focused on post-natal day 2, rather than embryonic cardiomyocytes. Potential differences in proliferative capacity of embryonic cardiomyocytes associated with the SHRSP chromosome 14 should be explored as a mechanism of pre-disposition to cardiac hypertrophy. Presumably, insufficient expansion during cardiogenesis and in early post-natal development would result in cardiomyocytes have to rely on catch-up hypertrophic growth to meet haemodynamic demands.

Fibroblast growth factor (FGF), Insulin-like growth factor (IGF) and epidermal growth factor (EGF) are considered important growth factors involved in cardiac development (G.S. Brown, Jang, and D. Li, 2022). These growth factors exert function through activation of receptor tyrosine kinases including the epidermal growth factor family of receptor tyrosine kinases (ErbB1-4). Activation of ErbBs cause ERK-translocation to the nucleus via activation of RAS-GTP signalling cascade. FGF, IGR and EGF receptors can also activate the PI3K-Akt signalling pathway (G.S. Brown,

Jang, and D. Li, 2022). ErbBs have been extensively studied in cancer, however they play essential roles in regulation of cell proliferation, differentiation, and migration during development (Wieduwilt and Moasser, 2008). In SHRSP and WKY.SPGla14a, the epidermal growth factor gene (*Egfr*), encoding the ErbB1 receptor was reduced compared to the WKY. Calcium- and diacylglycerol (DAG)-regulated nucleotide exchange factor, *Rasgrp1*, which activates Ras through the exchange of bound GDP for GTP, was also reduced in the SHRSP and WKY.SPGla14a compared to the WKY. ERK-signalling cascades were also predicted to be negatively enriched in comparisons to the WKY.

Depending on genetic background, global knock-out of *Egfr* in mice causes gestational or post-natal lethality affecting multiple organs, including the heart (Sibilia and E.F. Wagner, 1995; Sibilia, B. Wagner, et al., 2003; Schreier, Rabe, et al., 2013). A cardiomyocyte and vascular smooth muscle cell-specific knock-down of *Egfr* in mice, resulted in hypotension, cardiac hypertrophy, and premature death beginning from 100 days, but did not confer embryonic lethality (Schreier, Rabe, et al., 2013). When a more specific vascular smooth muscle cell promoter was used, and no reduction of *Egfr* was achieved in the heart, the cardiac phenotype was rescued in knock-out mice (Schreier, Hunerberg, et al., 2016). In the adult SHR rat, *Egfr* receptor expression is decreased and treatment with anti-sense *Egfr* from 5-12 weeks reduced LVM and blood pressure (Kagiyama et al., 2003). In the human and rat failing heart, expression of ErbB2 and ErbB4 receptors was reduced (Rohrbach, X. Yan, et al., 1999; Rohrbach, Niemann, et al., 2005) and a cardiomyocyte-specific *Egfr* knockdown in mice did not develop cardiac hypertrophy until 9-weeks of age, after development of cardiac dysfunction due to loss of cardiac contractility (S. Guo et al., 2022). Another mouse model where cardiomyocyte-specific knock-out of *Egfr* was inducible, knock-out animals developed DCM and abnormal cardiac function (Rajagopalan et al., 2008). Evidence suggests cardiac specific reduction of *Egfr* results in less severe cardiac phenotypes than global knock-outs or genetic manipulation of other ErbBs (Rajagopalan et al., 2008; Schreier, Rabe, et al., 2013; S. Guo et al., 2022). As well as differential expression of the *Egfr*, differential transcript usage of the *Erbb4* gene was demonstrated in SHRSP and WKY.SPGla14a compared to the WKY, and was annotated with protein and nucleotide binding pathways, as well as 'heart-development' pathways.

Although not reaching significance after adjustment in ORA, *Pik3r3* and *Isr1* were differentially expressed in all three comparison groups and are associated with the development of DCM and the PI3K-Akt signaling pathway. *Spp1* was also annotated within the PI3K-Akt signaling pathway, and its overexpression in the cardiac muscle during gestation was validated by qRT-PCR. ORA in both genes with DGE, and genes showing DTU indicated enrichment of genes within the Pleckstrin homology (PH) and PH-like domains. Pleckstrin homology-like domain family A, member 3 (*PHLDA3*)

expression was downregulated in hypertrophic murine hearts and angiotensin II-treated cardiomyocytes, mediated by Akt-signaling both *in-vivo* and *in-vitro* (J. Liu et al., 2019). PH domain-containing proteins are one of the largest superfamilies in humans, and seem to occur in a number of small GTPases, GTPase binding proteins and kinases (Lenoir et al., 2015). Contained in approximately 11% of human proteins, PH-domain containing proteins have been implicated in a range of cardiovascular, metabolic and degenerative disease (Powis et al., 2023).

Three common pipelines of short-read alignment and quantification were compared. The application of either HiSAT + FeatureCounts or StringTie, or Kallisto pseudo-alignment, affected the number of genes used to build statistical models. Despite aligning only ∼50% of reads, Kallisto quantified the expression of the most genes compared to FeatureCounts and StringTie. As well as an improved sensitivity to the transcriptome, scaled gene-counts from transcript-level quantification also offers reduced FDR in DESeq2 models compared to gene-level counts, as well as potentially recovering more multi-mapping reads which are discarded in gene-based counts from HTSeq/FeatureCounts (Soneson, Love, and M.D. Robinson, 2015).

Kallisto successfully aligned only 50% of paired-end reads. Analysis of HISAT2 alignment metrics support this was likely a consequence of the high number of retained introns and non-coding transcripts from the ribosomal depletion library preparation, where only 50% of reads were aligned to mRNA/UTR regions of the genome. All samples had ≥85% alignment rate to the whole genome using HiSAT2. Rather than enriching for Poly(A)-tailed mRNA, the ribosomal depletion allowed pre-mRNA, miRNA, and long non-coding RNA (lncRNA/ncRNA) to be sequenced. The Kallisto index for pseudoalignment was built using the Ensembl reference transcriptome, to preserve comparative power against HISAT2 alignment to the Ensembl reference by Glasgow Polyomics. In the human and mouse genomes, there is an average of 5.5 and 3.8 transcripts per gene, respectively (GRCh38.p14 and GRCm39, Ensembl version 110). Comparatively, the most up-to-date rat gene-build (mRatBN7.2) has only 1.9 transcripts per gene, which increases to 2.4 transcripts per gene when only coding genes are considered (mRatBN7.2, Ensembl version 110). *De-novo* transcript assembly with StringTie identified some 15,000 potential unannotated transcripts expressed in the GD18.5 heart. Whilst a proportion of these will represent errors caused by StringTie extending transcripts to include as a many read as possible, some will be novel transcripts that warrant further investigation. Comparison of the StringTie data to the newest release of NCBI's RefSeq annotation, which contains annotation for approximately equal number of protein-coding genes but includes annotation for more than 3x more lncRNA and pseudogenes than the Ensembl version (de Jong et al., 2023), might improve quantification herein. The alignment rate of Kallisto might therefore also be improved from increased annotation of the non-coding transcriptome in the rat using

more recently published RefSeq annotation.

Combined with the parallels which occur during pathology of LVH and foetal development, the parental and chromosome 14 congenic genetic models are complex genetic models of cardiovascular remodelling, making this dataset a rich resource for future analysis. RNA-sequencing has been described as a hypothesis-generating experiment. The chromosome 14 SHRSP-genome alters transcriptional regulation of embryonic hearts, and increased risk of cardiac hypertrophy could be caused by the suppression of ECM- and growth factor signalling during development, and upregulation of the MYC transcription factor. Although not possible during the experimental work herein, differences in mitochondrial function, energy metabolism and DNA-RNA machinery in the GD18.5 heart are good candidates for validation of RNA-sequencing analysis presented.

# Chapter 8

# Targeted Variant Discovery from Alignment to Improved Rat Reference Genome Assemblies

## 8.1 Introduction

A reference genome for the *Rattus Norvegicus* was first published in 2003, using the Brown Norway (BN) inbred strain. Since then there have been a number of iterations of improvements and updates, each providing improved genome continuity (Figure 8.1) (de Jong et al., 2023). Compared to the human and the mouse, initial efforts to sequence the rat genome advanced at a slower rate. In 2021, the 7th generation of the rat genome (mRatBN7.2) was adopted by the Genome Reference Consortium (GRC) as the official rat reference genome and is a considerable improvement to Rnor_6.0, first published in 2014 (Figure 8.1). Alongside technological advancements in genome sequencing, a large variety of sub-strains have been bred over the last 150 years which display specific and complex phenotypes, including heterogenous stock and recombinant inbred panels which allow for high quality QTL mapping, transcript and protein expression QTL, and GWAS studies. Since the first iteration of the rat genome, a number of large 'catalogue' studies quantifying inter-strain variance between commonly used rat strains have been published, the most recent in 2023 (Figure 8.1). It should be noted that until Hermsen et al. (2015), large studies into multiple inbred and heterogenous stock strains had relied on the v3.4 assembly (Figure 8.1).

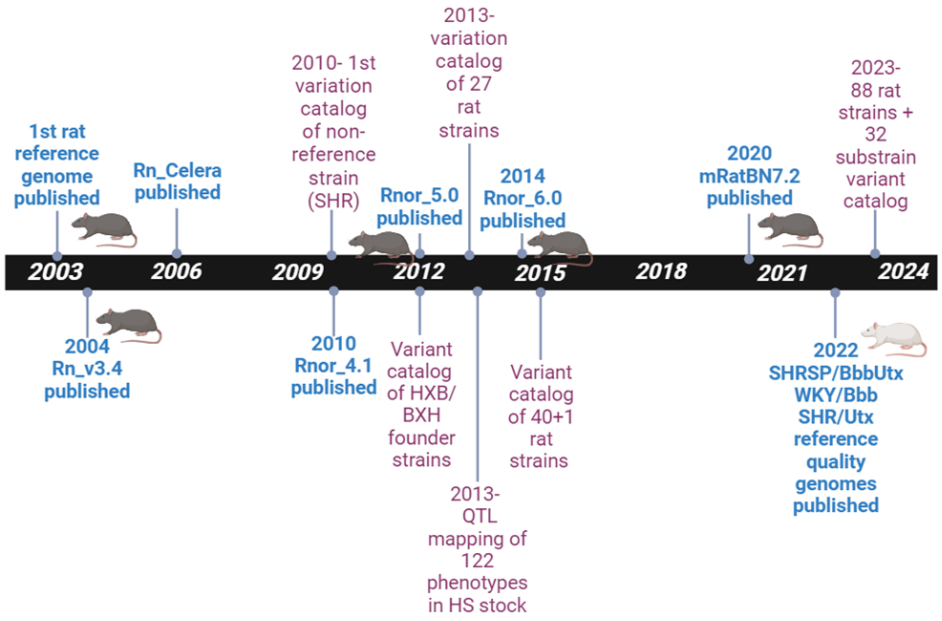



**Figure 8.1: Timeline of *Rattus Norvegicus* Reference Genome Publication** Since 2003, 6 iterations of the Brown Norway genome has been published as a reference genome for studies using the laboratory rat. Publication of new reference genomes are labelled in blue. Whole genome sequencing and variant analyses to generate 'catalogues' of rat variants, and other studies of significance are labelled in pink.

Bred for their genetic predisposition to high blood pressure and stroke, the SHRSP and its genetic control strain WKY, are divided into sub-strains depending on their geographical location. For example, the SHR/NCrl and SHR/NHsd are housed in the US and Europe respectively (Hermsen et al., 2015). The BN genome is genetically diverse from a number of inbred strains (Nijman et al., 2008) such that the reference itself comprises a phenotype and genotype unique to that of both the experimental and control strains used in CVD research (Kalbfleisch et al., 2023). Genetic analyses of experimental and control strains can become limited by a reduced focus on the variants between two comparative strains. In addition, there is a potential loss of variant information because variant containing reads may not effectively align to the reference. Atanur et al. (2013) constructed a phylogenetic tree of 28 rat strains, demonstrating the Wistar derived strains from Japan were the most distant from the BN strains. The largest joint analysis of rat whole genome sequencing (WGS) showed the WKY/N contributed the largest number of strain-specific variant sites across all included strains (de Jong et al., 2023), supporting the relative difference between the BN and wistar-derived strains.

Attributable to the distribution of WKY and SHRSP strains before complete inbreeding, the WKY/Gla genome is distinct from other WKY sub-strains, with a mosaic distribution of genomic regions associated with both WKY and SHR clusters (Hermsen et al., 2015). Whilst no reference genome is going to perfectly represent all laboratory rat strains, a particular challenge is posed with Wistar-derived strains, where much

of the variation will arise from the phylogenetic distance between the WKY, SHR, and SHRSP in relation to the BN reference (Star Consortium et al., 2008). A greater concern is the potential challenges such differences present during short read alignment, where regions of high variation might reduce alignment accuracy and efficiency. With the reduction in cost and accuracy of WGS, and the advancement in sequencing capabilities, it has become feasible to generate accurate, reference quality, *de-novo* genome assemblies of individual strains. As long-read and whole exome sequencing become more affordable and accessible, publication of reference quality genomes of distinct rat sub-strains has the potential to improve alignment efficacy to the chosen 'reference' genomes. To this end, the newest update to RefSeq and Ensembl databases include the publication of reference quality genomes for WKY/Bbb, SHR/Utx and SHRSP/BbbUtx sub-strains. Compared to the mRatBN7.2, the SHRSP/BbbUtx has a genome wide divergence of ∼0.18% containing nearly 5 million single nucleotide polymorphisms (SNPs) and 1.66 million insertions and deletions (InDels) (Kalbfleisch et al., 2023).

The SHRSP/BbbUtx is an SHRSP sub-strain maintained at the University of Texas, originally obtained from the University of Heidelberg, Germany. This represents a reference genome ancestrally closer to the WKY/Gla and SHRSP/Gla strains housed at the University of Glasgow. The benefit of calling variants following alignment of reads to a sub-strain based reference is in part due to the reduced number of variants and consequent reduction in time and memory requirements of data processing. Of greater importance to functional genomics, the biological impact of sub-strain variants are higher than comparison to a common reference, as sub-strain variants more often affected protein sequence and nucleotides with a higher phastCons scores (Hermsen et al., 2015).

Following alignment of 28 strains (including SHR and SHRSP) to the Rn_V3.4, genes harboring non-synonymous coding SNPs were over-represented in human GWAS studies of hypertension, metabolic phenotypes, and their complications (Atanur et al., 2013). With improved reference genome for alignment, de Jong et al. (2023) found over 2,000 rat-human orthologues contained 'high impact' variants associated with 1,393 traits. This included >800 variants associated with heart-related phenotypes (de Jong et al., 2023). Few studies of genetic variance in the rat have focused on genetic predictors of LVMI, although a number of QTLs have identified regions of the rat genome associated with cardiac mass and are catalogued on the Rat Genome Database (Vedi et al., 2023). Comparison of the hypertrophic heart rat (HHR) to the 5th generation of the BN reference genome (Rnor_5.0) produced a database of more than 8 million variants in the HHR, exhibiting a higher change rate from the BN versus its control normal heart rat (NHR) (Prestes, Marques, Lopez-Campos, Lewandowski, et al., 2018). Variants unique to the HHR were overrepresented in pathways such as

extracellular matrix-receptor interaction and tryptophan metabolism. Using a previous QTL, an exonic variant of *Trim55* was associated with gene expression in the HHR and an F2 population (Prestes, Marques, Lopez-Campos, Booth, et al., 2016).

Maintained as inbred strains at the University of Glasgow since 1999, the WKY/Gla and SHRSP/Gla formed the background of chromosome 14 congenic strains described in previous chapters. Although a recent catalog of variants in inbred strains was published following the adoption of mRatBN7.2 (de Jong et al., 2023), only the SHRSP/Gla was included in 88+32 strains analysed. In addition to the improvements in the mRatBN7.2, the SHRSP/BbbUtx reference quality genome offers the potential to improve alignment of short read data and higher quality variant identification than previously reported. Alignment to a genetically similar reference has the potential to both improve alignment, and generate a refined variant panel with greater biological relevance to the defined phenotype. The work presented with this chapter was supported by funding from the Graham Wilson Travel Scholarship and MVLS training fund. Analyses were performed in collaboration with the Rat Genome Database (RGD) housed at the Medical College of Wisconsin, Milwaukee.

## 8.2 Specific Hypothesis and Aims

It is hypothesised that an improved alignment and high confidence variant calls will generate a panel of genetic variants underlying physiological differences in blood pressure and cardiac structure. Alignment of previously generated WKY/Gla and SHRSP/Gla short read DNA sequencing data was undertaken with the specific aims:

- Identify regions of shared and unique variation between SHRSP/Gla and WKY/Gla with the Brown Norway reference genome (mRatBN7.2)
- Use the Wistar-derived SHRSP/BbbUtx sub-strain as a reference genome to identify specific variants that contribute to heart and vascular phenotypes, with a focus on previously identified QTLs associated with LVMI in the rat
- Perform alignment of short read RNA sequencing data from GD18.5 hearts to the SHRSP/BbbUtx reference genome for use in short variant discovery.

## 8.3 Specific Methods

DNA sequencing was performed on the Illumina platform (Illumina HiSeq2X00) using DNA isolated from WKY/Gla and SHRSP/Gla. Detailed methods were described in Atanur et al. (2013). Description of RNA-sequencing analysis can be found in Chapter 7. Whole genome alignment and variant calling was performed using the pipeline proposed by the Genome Analysis Toolkit (GATK), the standard for identifying SNPs

and InDels in germline DNA and RNAseq data (Van der Auwera et al., 2013). An overview of the pipeline is shown in Figure 8.2

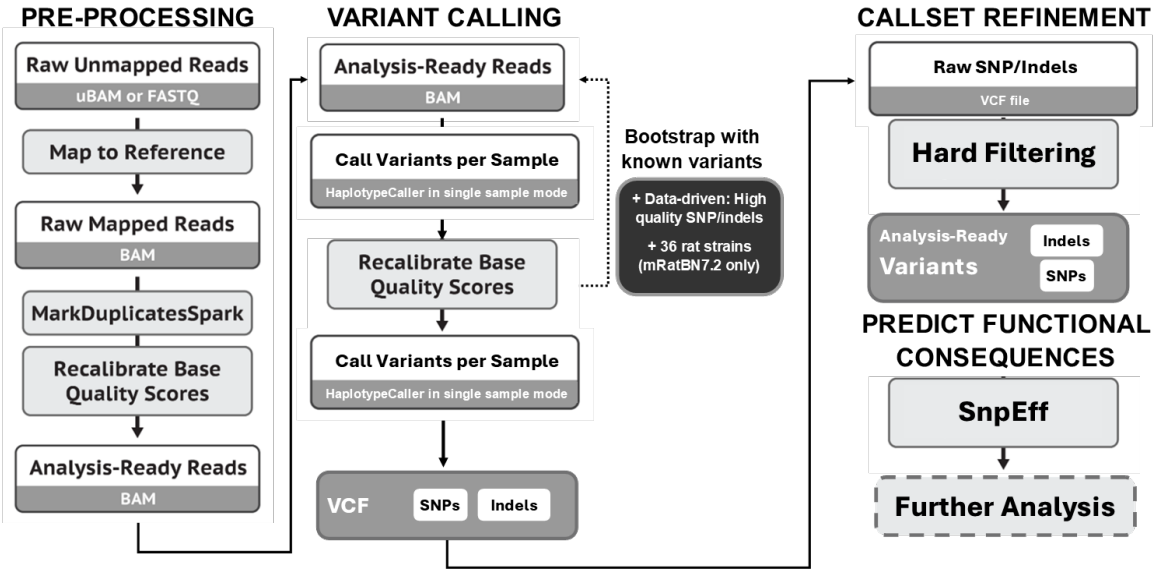



**Figure 8.2: GATK Variant Calling Pipeline** Adapted from gatk.broadinstitute.org.

### 8.3.1 Obtaining Reference and Raw Genome Sequence Data

The mRatBN7.2 rat reference assembly was downloaded from the NCBI Genome Reference Consortium. The reference quality genome assembly for the SHRSP is available from the NCBI under the submitted GenBank assembly GCA_021556685.1. Both reference genomes were indexed for alignment using Picard and Samtools. WKY/Gla and SHRSP/Gla genome sequence data was downloaded from the EBI sequence read archive (accession ERP002160). Fastq files were either downloaded directly or parsed from BAM files depending on availability in the EBI archive.

At the time of study, the SHRSP/BbbUtx reference genome was downloaded from directly from the BioProject archive detailed in (Kalbfleisch et al., 2023)[1], before the genome and annotation were included in the GenBank and Ensembl databases. Annotation files were received with thanks from the RGD courtesy of Dr. Peter Doris at the University of Texas.

### 8.3.2 Mapping to Reference Genome

Fastq files were mapped to both the mRatBN7.2 and SHRSP/BbbUtx reference genomes using Burrows-Wheeler Aligner (BWA-0.5.8c) maximal exact match (mem) algorithm (DePristo et al., 2011). All default parameters were used to align reads using BWA. The aligned reads were then processed for variant analysis following Genome Analysis Toolkit best practices (Van der Auwera et al., 2013). Mapped reads were merged

[1]BioProject link `https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_021556685.1/`.

with unaligned BAM (uBAM) files to apply read group information from the uBAM whilst retaining information from the alignment process. This ensures congruency in meta-data and generates additional information to result in a 'clean' alignment file.

### 8.3.3 Genomic Variant Calling

Single nucleotide polymorphisms (SNPs), insertion, and deletion (InDels) mutations were called from aligned BAM files using GATK version 4. Prior to calling variants, alignment files underwent a series of base recalibration steps to improve the accuracy of base quality scores used for variant filtering. There is a strong reliance of variant calling algorithms on the quality score assigned to base calls in sequence data reads. As such, known systematic technical errors from sequencing reads can lead to over- or under-estimation of base quality scores and can strongly affect downstream processes. GATK therefore recommends a base quality score recalibration (BQSR) step, which is a machine learning process that models errors in base call quality scores and adjusts them accordingly. Users can input .vcf files of 'known' variants that can mask bases at sites of expected variation to avoid BQSR counting real variants as erroneous.

High quality reference variants are not available for the SHRSP/BbbUtx genome. To ensure proper BQSR, SNPs and InDels were called from merged BAM files via local re-assembly of haplotypes (GATK HaplotypeCaller). The resulting variants were filtered to contain those with the highest confidence which could be used as input files of 'known' variants for feeding into BQSR. In the case of the SHRSP/BbbUtx, where there were no reference variants to input into the Base Recalibrator, three rounds of data-driven BQSR was performed using the highest confidence variants from each bootstrap step. The final adjusted model was used for variant calling in single sample mode.

In the mRatBN7.2-based variant calling, a second pass of base recalibration was performed with both input data-driven high quality variants (as described above), and two additional VCF files provided by RGD containing (1) high quality variants from analysis of 36 rat strains, and (2) the most recent European Variation Archive (EVA release 4). Following the re-adjustment of base scores to generate a final adjusted model for variant calling, SNPs and InDels in the SHRSP/Gla and WKY/Gla were detected in single sample mode (HaplotypeCaller).

VCF files were then processed for callset refinement which estimates the probability each putative call is a true genetic variant. Variant calls which passed 6 hard-filters were 'analysis-ready'-variants utilised for downstream functional consequence predictions and analyses. The hard filtering was based on the parameters; 'QualByDepth' (QD) <2.0, 'FisherStrand' (FS) >60, 'StrandOddsRatio' (SOR) >3.0, 'RMSMappingQuality' (MQ) <40.0, 'MappingQualityRankSumTest' (MQRankSum) <-12.5, and 'Read-

PosRankSumTest' (ReadPosRankSum) <-8.0, and are based on GATK best practice recommendations.

### 8.3.4 Functional Consequence of Predicted Variants

The functional consequence of predicted SNP and InDel variants were analysed using SnpEff Version 4 (published 2020) (Cingolani, Platts, et al., 2012b). SnpEff is a computer program used to annotate the functional consequences of variants based on their genomic locations. The software has a number of reference genomes 'built-in' including the Rnor_v6.0. Using .GTF/.GFF and reference genome files, a custom database can be built. Custom databases for mRatBN7.2 and SHRSP/BbbUtx were built for the purpose of this analysis by Dr. Monika Tutaj at the RGD. Final VCF files were used as input files for annotation by SnpEff based on variant genomic location. Coding effects such as synonymous or non-synonymous amino acid replacement, start codon gain or loss, stop codon gain or loss, or frame-shift variants were predicted, as well as annotation of variants to intronic, untranslated region (UTR), upstream, downstream, splice site, or intergenic regions (Table 8.1).

SnpSift (Cingolani, Patel, et al., 2012a) was used to parse VCF files from SnpEff analysis into tab delimited files, and subsequent analyses were conducted using custom scripts in R.

**Table 8.1:** Description of SnpEff Predictions

| SnpEff Prediction | Description |
|---|---|
| Intergenic | The variant is in an intergenic region |
| Upstream | Upstream of a gene (default length: 5K bases) |
| UTR 5 prime | Variant hits 5'UTR region |
| UTR 5 deleted | The variant deletes an exon which is in the 5'UTR of the transcript |
| Start gained | A variant in 5'UTR region produces a three base sequence that can be a START codon. |
| Splice site acceptor | The variant hits a splice acceptor site (defined as two bases before exon start, except for the first exon). |
| Splice site donor | The variant hits a Splice donor site (defined as two bases after coding exon end, except for the last exon). |
| Start lost | Variant causes start codon to be mutated into a non-start codon. |
| Synonymous start | Variant causes start codon to be mutated into another start codon. |

| | |
|---|---|
| CDS | The variant hits a CDS. |
| Gene | The variant hits a gene. |
| Transcript | The variant hits a transcript. |
| Exon | The variant hist an exon. |
| Exon deleted | A deletion removes the whole exon. |
| Non-synonymous coding | Variant causes a codon that produces a different amino acid |
| Synonymous coding | Variant causes a codon that produces the same amino acid |
| Frame shift | Insertion or deletion causes a frame shift |
| Codon change | One or many codons are changed |
| Codon insertion | One or many codons are inserted |
| Codon change plus codon insertion | One codon is changed and one or many codons are inserted |
| Codon deletion | One or many codons are deleted |
| Codon change plus codon deletion | One codon is changed and one or more codons are deleted |
| Stop gained | Variant causes a STOP codon |
| Synonymous stop | Variant causes stop codon to be mutated into another stop codon. |
| Stop lost | Variant causes stop codon to be mutated into a non-stop codon |
| Intron | Variant hist and intron. Technically, hits no exon in the transcript. |
| UTR 3 prime | Variant hits 3'UTR region |
| UTR 3 deleted | The variant deletes an exon which is in the 3'UTR of the transcript |
| Downstream | Downstream of a gene (default length: 5K bases) |
| Intron conserved | The variant is in a highly conserved intronic region |
| Intergenic conserved | The variant is in a highly conserved intergenic region |

UTR; untranslated region, CDS; coding sequence.

### 8.3.5 Variant Calling from Short Read RNA-sequencing in GD18.5 Hearts

Dr. Monika Tutaj performed alignment of trimmed short read RNA-sequencing data generated from GD18.5 heart tissue (Chapter 7). Fastq files were aligned to the SHRSP/BbbUtx genome and colleagues at the RGD uploaded aligned BAM files into

the JBrowse viewer. RNA sequencing short variant discovery (SNPs & InDels) was conducted using the GATK pipeline (Van der Auwera et al., 2013). Similar to DNA pre-processing, following HISAT alignment to the reference, BAM files were 'cleaned' by merging with uBAM files and marking duplicates using picard. RNA aligners have different conventions to DNA aligners and reads that span introns require reformatting for the GATK HaplotypeCaller. 'SplitNCigarReads' splits reads with N in the cigar into multiple supplementary alignments and hard clips the mismatching overhang. This step also reassigns mapping qualities to match DNA conventions. The analysis ready BAM files then enter the same variant calling pipeline detailed above (Figure 8.2). Final VCF files were analysed using custom scripts in R.

### 8.3.6 Processing Variant Callsets

VCF files were read into R and homozygous variants were retained for downstream analyses. VCF files contain measures of genotype calling quality at the sample level. Quality is measured on the Phred scale, termed the 'Phred-scaled likelihood of the genotype'. The phred-scale is -log10(1-p), such that a value of 10 indicates a 1 in 10 chance of error, while a 100 indicates a 1 in $10^{10}$ chance. The 'QUAL' score assigned to each variant is the Phred-scaled probability that a reference or alterative (REF/ALT) allele is present, given sequencing data. The QUAL score is therefore an indication of how confident we are that there is some kind of variation at a given site. The QUAL-score is less important in single sample variant calling, where GQ values are a better observation of genotype assignment.

The 'PL' is the Phred scaled likelihoods of all the possible genotypes (0/0, 0/1, and 1/1). The PL values are 'normalised' such that the PL of the most likely genotype assigned in the genotype field is 0 in the Phred scale. Other values are scaled relative to the most likely genotype, and determines how likely any variation is present at that site. The GQ is the difference between the PL of the second most likely genotype, and the PL of the most likely genotype and indicates the difference between the likelihoods of the two most likely genotypes. A low GQ indicates reduced confidence in genotype assignment. A GQ call of <20, was used as a filter for variants called, corresponding to <1% chance that a call is incorrect.

Once high confidence, homozygous variants are obtained from VCF files, the ratio of transition (Ti) to transversion (Tv) SNPs can be assessed to determine the potential rate of false positives. Transitions are an interchange of two purines (A/G, G/A) or of two pyrimidines (C/T, T/C) whilst transversion mutations are interchanges of purine for pyrimidine (A/C, A/T, G/C, G/T, C/A, C/G, T/A, T/G). If the distribution of transition and transversion mutations were random, a Ti:Tv ratio of 0.5 would be expected. This is due the presence of twice as many possible transversion mutations than there are transitions. In the biological context, it is common to see a methylated

cytosine undergo deamination to become thymine (C<T), a transition mutation. This tendency increases the expected ratio to increase from 0.5 to ∼2.01. CpG islands, which are found in primer regions, have higher concentrations of 5-methylcytosines at risk of transition. Whole exome sequencing shows a stronger lean towards transition mutations, increasing the expected ratio to ∼3.0 in these cases. A significant deviation from the expected values could indicate artifactual variants causing bias. A Ti:Tv ratio that is low could be used as an indicator of higher rates of false positives.

### 8.3.7 Fitting Variants in Quantitative Trait Loci for Cardiac Mass

QTLs associated with cardiac mass and/or ventricle weight were downloaded from the Rat Genome Database. As a chain file between the mRatBN7.2 and SHRSP/BbbUtx assemblies was not publicly available, liftOver between genomic coordinates could not be performed. Instead, lists of the genes contained within each QTL were retrieved from the Ensembl database using biomaRt. These gene lists were used as a filter for determining variants contained within QTL regions. SnpEff annotates effects as 'upstream' and 'downstream' of genes within 5k bp and thus variants 'within QTLs' are potentially up to 5k bp outside of true QTL region. Intergenic variants within QTLs but not within 5kb of a gene will also not have been captured. This approximate localisation of variants to QTL regions should therefore be interpreted with caution, and needs to be validated by the generation of a chain file between the two genomes. The time and computing power required to do this were outside the constraints of this project.

### 8.3.8 Data Availability

DNA sequences from the WKY/Gla and SHRSP/Gla, and all reference genomes and transcriptomes utilised herein are publicly available resources. The GATK pipeline was followed according to published best practices and any deviations from the recommended best practice was detailed herein. Functions generated in R for data processing and visualisation are available at: `https://github.com/c-triv97/Source-Codes`.

## 8.4 Results

The RGD site contains variant information for 48 strains compared to the mRatBN7.2. The Glasgow WKY and SHRSP substrains are not currently included in this catalogue. Data from RGD (Figure 8.3) demonstrate the WKY/N, WKY/NCrl, SHRSP/A3NCrl, and SHR/OlalpcvMcwi genomes contain the highest number of SNPs and InDels, with almost 1.1 million InDels and ∼5 million SNPs.

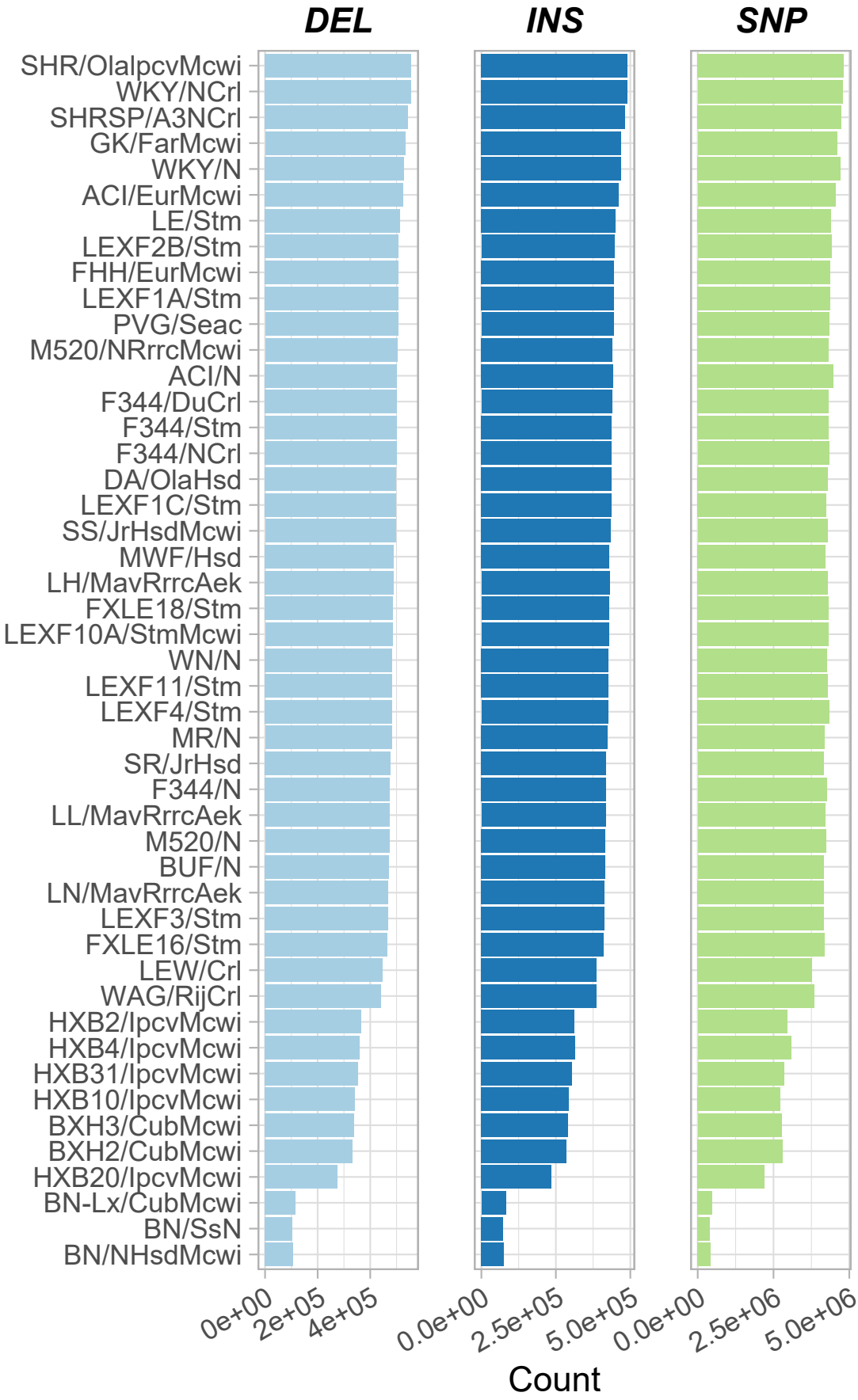



**Figure 8.3: Catalogue of Variants Across 48 Strains** Data downloaded from the Rat Genome Database (November 2023), strains are organised along the y-axis in order of total number of SNP, insertions (INS) and deletions (DEL).

### 8.4.1 Read Quality and Alignment Statistics

Results from MultiQC report were reported in R and plotted (Figure 8.4).

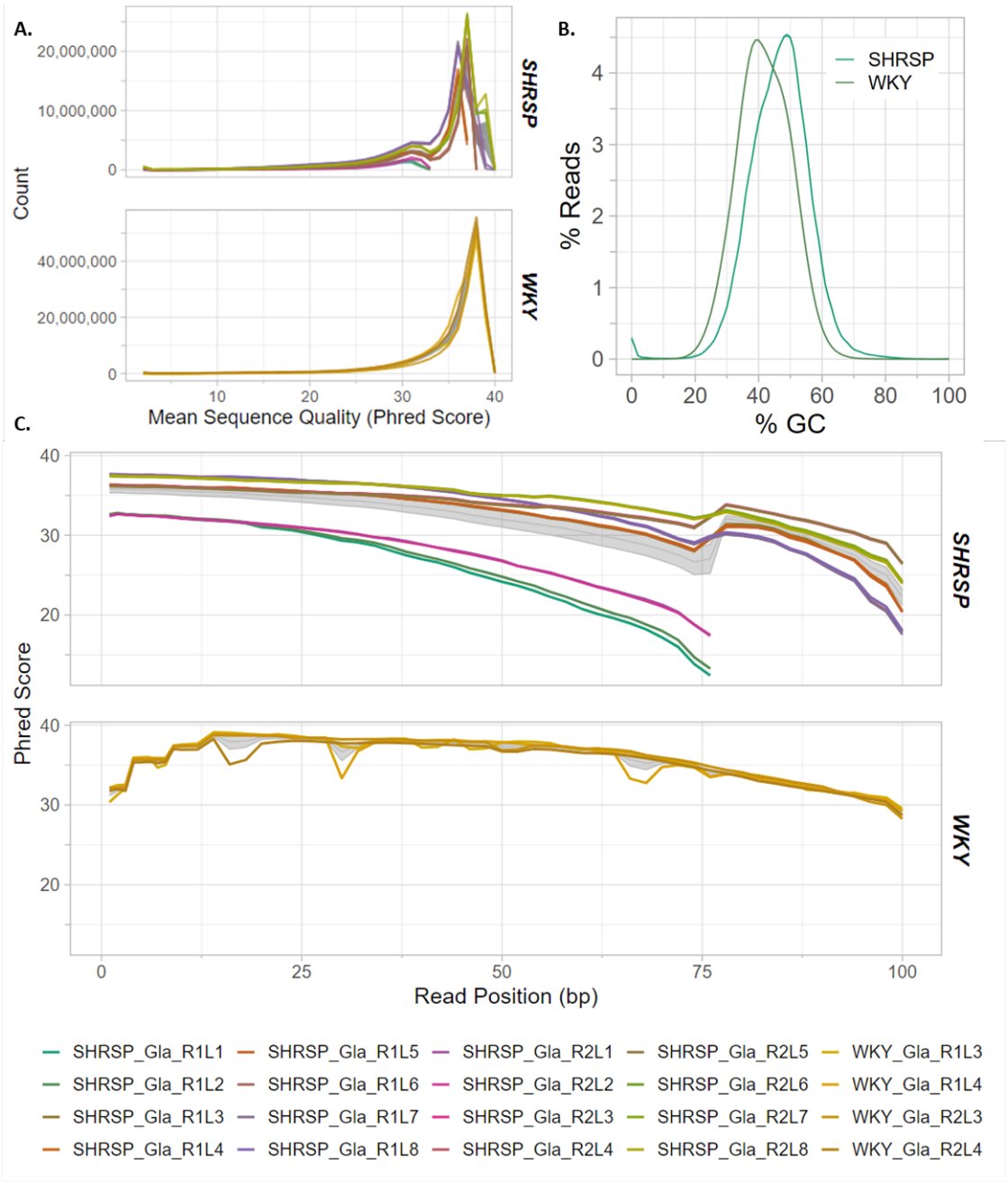



**Figure 8.4: DNA Short-Read Sequencing Quality** (A) Counts of per-base Phred-score (B) average GC content, and (C) quality per base score across all WKY/Gla and SHRSP/Gla DNA sequencing reads used for alignment of to mRatBN7.2 and SHRSP/BbbUtx, There were approximately 200 million reads in each sample, with an average read length of 100bp.

The alignment rate of short read DNA sequence to the mRatBN7.2 and SHRSP/BbbUtx is shown in Figure 8.5

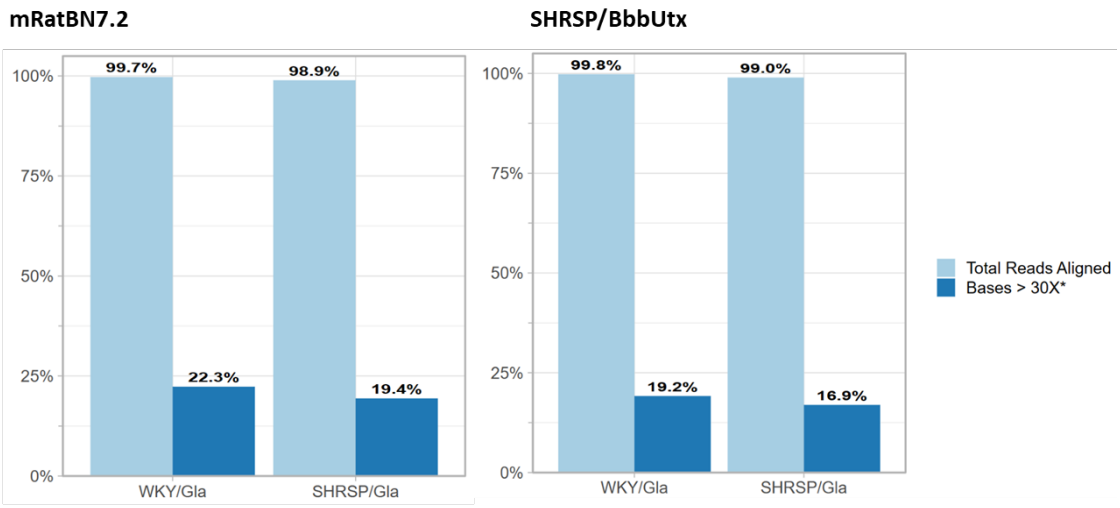



**Figure 8.5: Whole Genome Alignment Statistics** Percentage of total reads aligned to (left) mRatBN7.2 and (right) SHRSP/BbbUtx reference genomes

### 8.4.2 Whole Genome Alignment to Brown Norway (mRatBN7.2)

Variant analysis called 1,622,051 InDels and 4,765,814 SNPs in the WKY/Gla. The SHRSP/Gla contained 1,536,913 and 4,799,608 InDels and SNPs respectively. In both alignments there was a small % of variants that were heterozygous and were filtered out during initial processing. Using a phred-scaled probability >20 as a cut-off for high confidence, 1,361,663 homozygous variants unique to the WKY/Gla and 1,034,766 unique to the SHRSP/Gla were identified. A greater density of variants identified in the SHRSP/Gla had phred score <20 and were filtered out (Figure 8.6a).

Over 60% of variants which passed filtering were shared in between the WKY/Gla and SHRSP/Gla compared to the BN (3,982,193 in common, Figure 8.6b). The WKY/Gla genome contained ∼300,000 more variants than the SHRSP compared to the mRatBN7.2 (Figure 8.6b), and in both strains, most variants were SNPs (Figure 8.6d). The remaining variants were equally distributed between insertions and deletions (Figure 8.6d). The number of SNPs and InDels in the WKY/Gla and SHRSP/Gla is similar to that of the WKY/N, WKY/NCrl, SHRSP/A3NCrl, and SHR/OlalpcvMcwi published on the RGD.

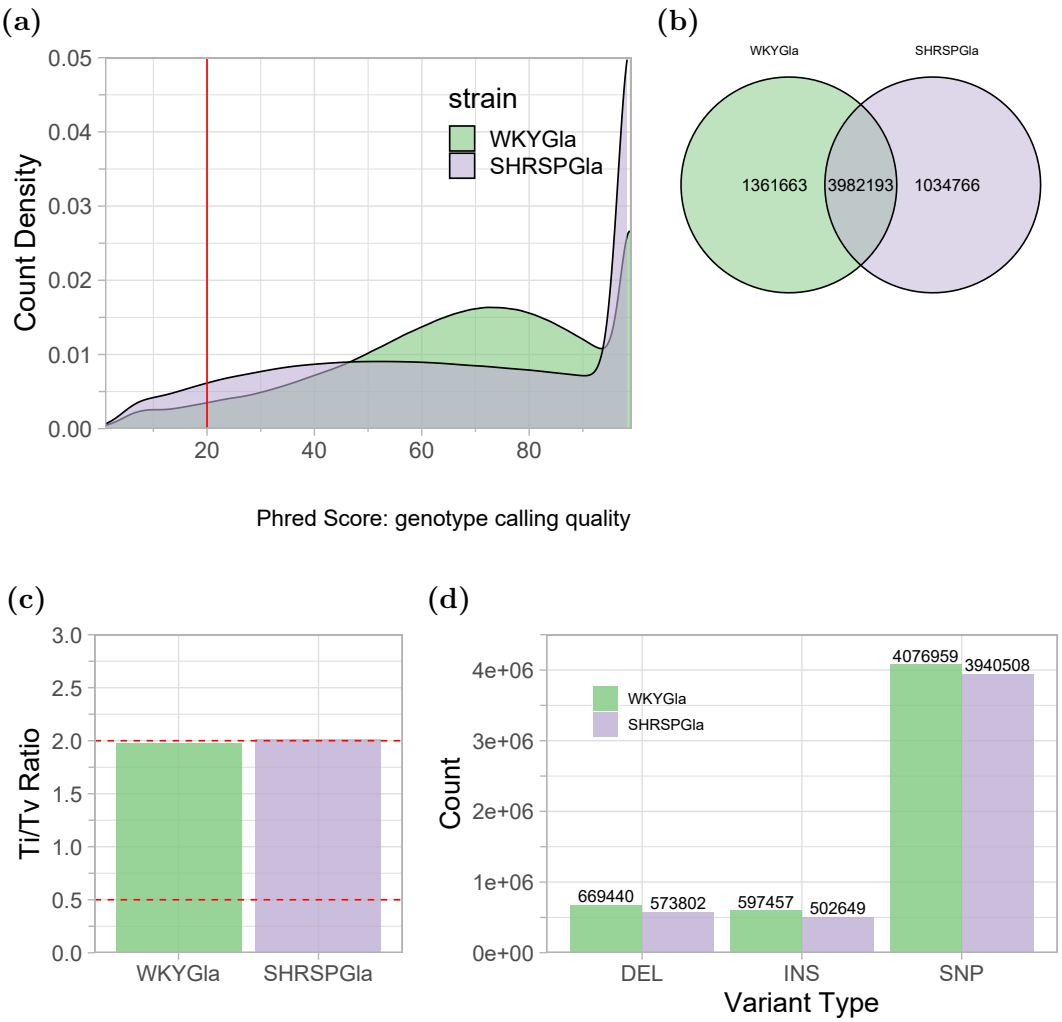



**Figure 8.6: Alignment to mRatBN7.2 Reference Genome** (a) Density plot of Phred-scaled likelihood of genotype quality represented by the 'GQ' score. GQ call of <20 was used as a filter for variants called, corresponding to <1% chance that a call is incorrect, and is indicated by horizontal red line, (b) Venn-diagram of total variants called across both strains, (c) Transition to transversion ratio in WKY/Gla and SHRSP/Gla SNP variants, and (d) counts per variant type of variants identified in WKY/Gla and SHRSP/Gla

The ratio of transition (Ti) to transversion (Tv) was ∼2.0 in both WKY/Gla and SHRSP/Gla. The number of variants per length of chromosome (change rate) ranged from less than 1 variant per 1000 bases on X & Y chromosomes to >2 variants per 1000 bases (Figure 8.7). Change rate tended to be higher in WKY/Gla excepting chromosome 12, where SHRSP/Gla had a higher variant rate. Chromosomes 2, 5, 12, and 14 had the highest variant rate in both strains.

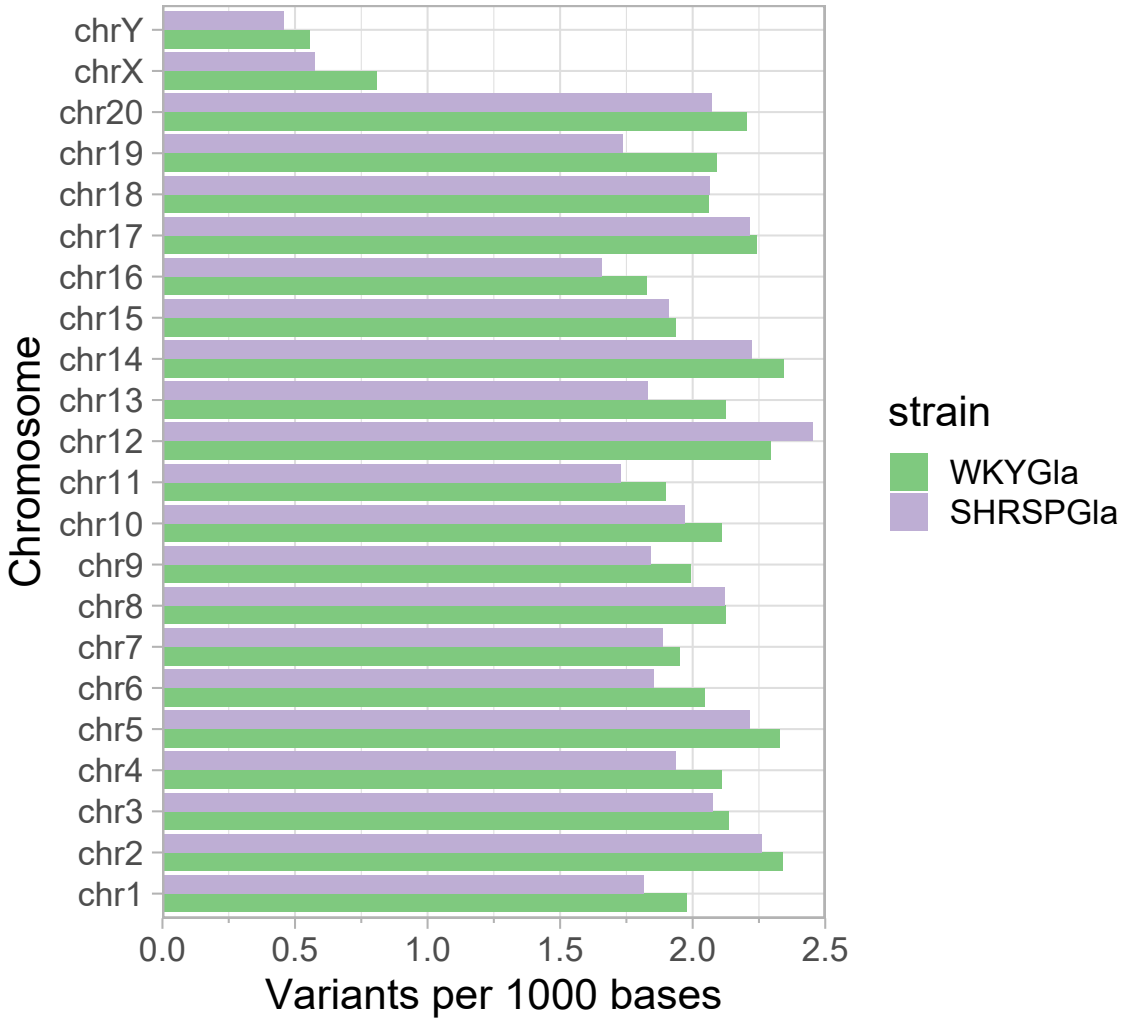



**Figure 8.7: Variant Rate Per Chromosome from mRatBN7.2** Variance change rate was normalised to the length of individual chromosomes and was calculated per 1000 base pairs for the WKY/Gla and SHRSP/Gla. Change rate per chromosome in WKY/Gla and SHRSP/Gla are plotted on the same scale to allow comparison of change rate along the genome.

### 8.4.3 Whole Genome Alignment to SHRSP/BbbUtx Reference Quality Sequence

Variant analysis called 522,123 InDels and 1,388,270 SNPs in the WKY/Gla genome when the SHRSP/BbbUtx genome was used as reference for alignment. The majority of variants were homozygous and after removing non-homozygous variants, 425,359 InDels and 1,290,914 SNPs were identified.

Prior to removal of heterozygous variants, alignment of the SHRSP/Gla genome to the SHRSP/BbbUtx reference resulted in the identification of 59,775 SNPs and 112,658 InDels. After removing non-homozygous mutations, compared to the SHRSP/BbbUtx, there were 23,053 InDels and 34,527 SNPs in the SHRSP/Gla genome.

A higher proportion of total variants with a GQ score <20 were found in the SHRSP/Gla than WKY/Gla (Figure 8.8a) amd the TiTv ratio of SNPs was lower in the SHRSP/Gla (∼1.8) compared to the WKY/Gla (∼2.0, Figure 8.8b). In both strains, the majority of variants were SNPs, and there was an even distribution of insertions and deletions (Figure 8.8d), which mirrored the distribution of variance when genomes were compared to the mRatBN7.2 reference.

Only 0.4% of variants (5,070) were shared between the Glasgow strains (Figure 8.8c). When Glasgow WKY and SHRSP genomes were aligned to the reference quality SHRSP/BbbUtx strain, 1,620,958 unique variants were detected in the WKY/Gla with a phred score >20. Compared to the SHRSP/BbbUtx reference, the SHRSP/Gla genome contained an 40,384 unique variants.

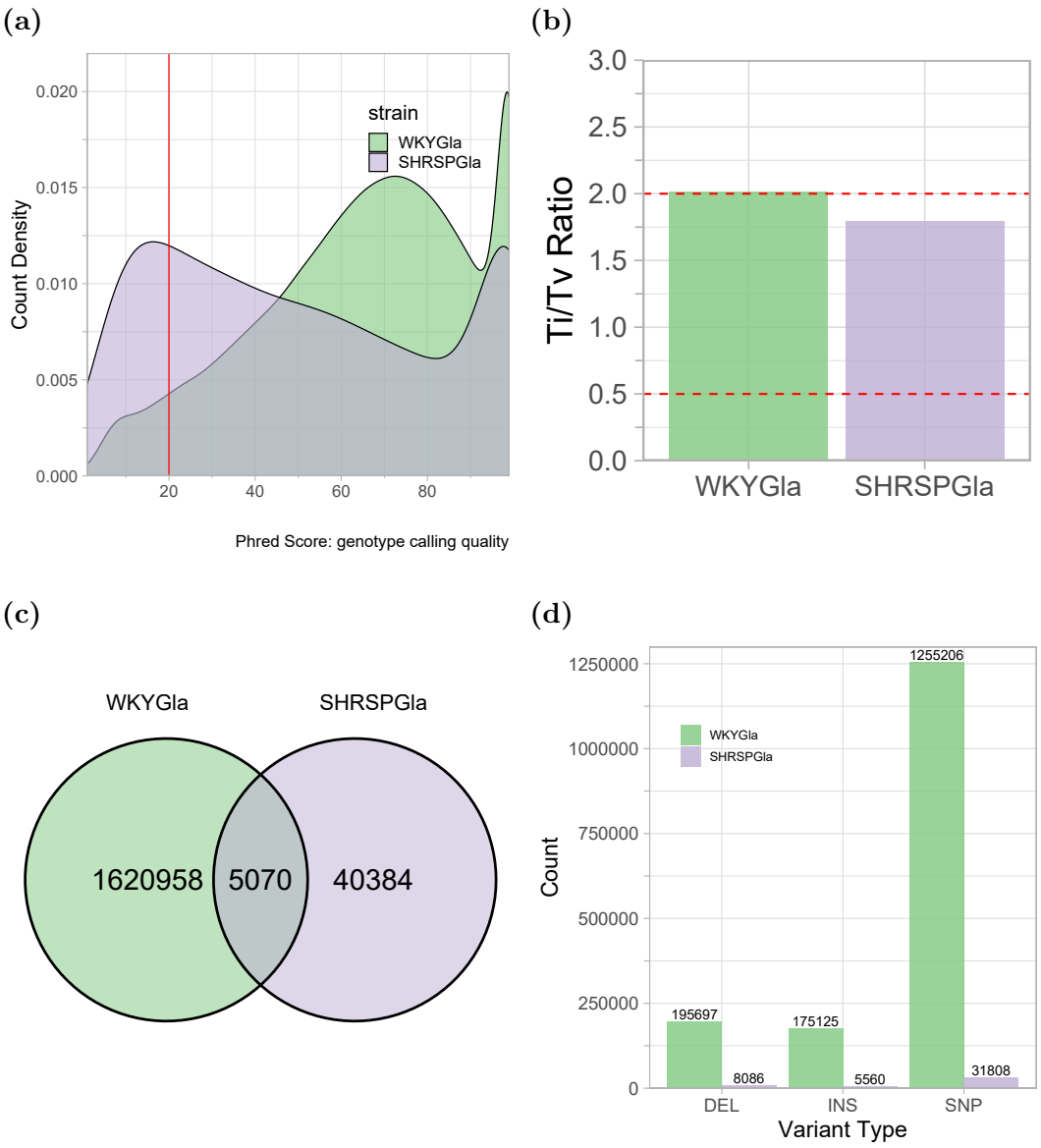



**Figure 8.8: Alignment to SHRSP/BbbUtx Reference Genome** (a) Density plot of Phred-scaled likelihood of genotype quality represented by the 'GQ' score. GQ call of <20 was used as a filter for variants called, corresponding to <1% chance that a call is incorrect, and is indicated by horizontal red line, (b) Transition to transversion ratio in WKY/Gla and SHRSP/Gla SNP variants, (c) Venn-diagram of total variants called across both strains, (d) Counts per variant type of variants identified in WKY/Gla and SHRSP/Gla

As a result of the vast reduction in variance between the SHRSP/Gla compared to the SHRSP/BbbUtx, the change rate of variance in the SHRSP/Gla was <1 in 100,000 in almost all chromosomes apart from chromosome 7, 11, and 15, which had a change rate ∼1 in 10,000 (Figure 8.9). The rate of variance is also greatly reduced in the WKY/Gla when using the Wistar-derived SHRSP/BbbUtx as the reference genome, and was below 1 in 1000 base pairs across all chromosomes apart from chromosome 2. Chromosome 11 had the highest rate of variance in SHRSP/Gla but a relatively low rate of variance in the WKY/Gla. It is worth noting that the absolute change rate of all chromosomes was greater in the WKY/Gla than the SHRSP/Gla.

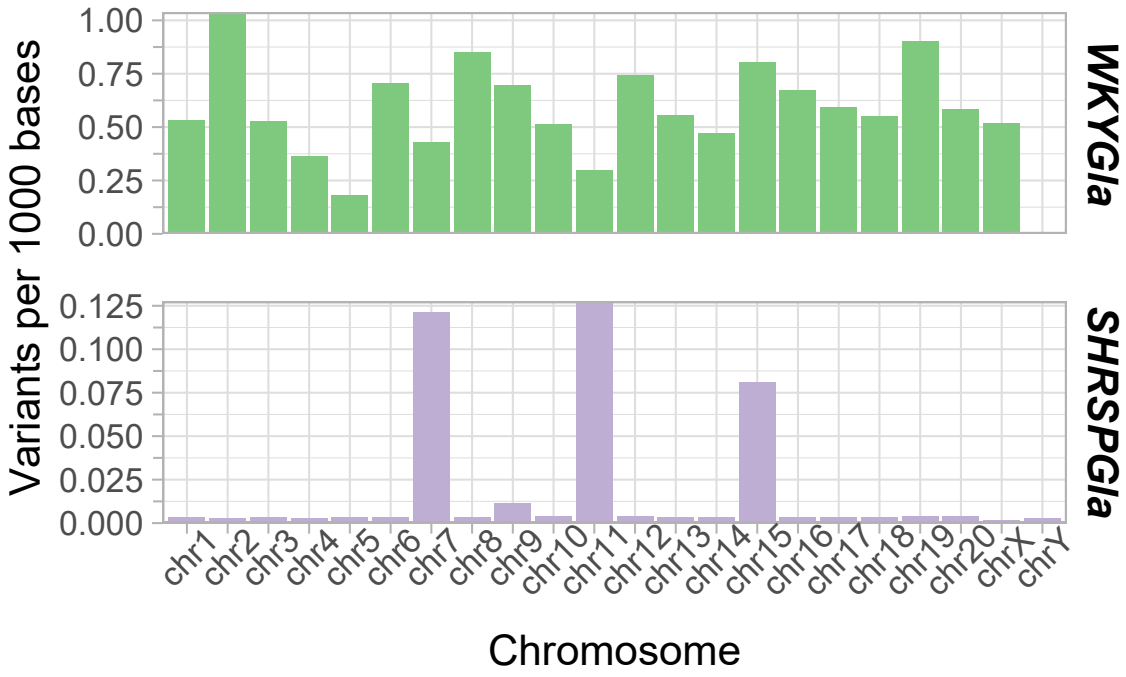



**Figure 8.9: Variant Rate Per Chromosome in SHRSP/BbbUtx Alignment** Variance change rate was normalised to the length of individual chromosomes and was calculated per 1000 base pairs for the WKY/Gla and SHRSP/Gla. Rate of variance on chromosomes 1-20 plus X and Y are plotted for WKY/Gla (top) and SHRSP/Gla (bottom). Variance rates were plotted on different y-scales for ease of visualisation.

### 8.4.4 Predicted Consequences of Observed Genetic Variation

SnpEff was used to predict functional consequences of observed variance, the program reports putative variant impact as 'High', 'Moderate', 'Low', and 'Modifier' respectively depending on severity of variant effect. A 'high'-impact variant is predicted to have disruptive impact in the protein, causing protein truncation, loss of function or triggering nonsense mediated decay. In contrast, 'modifier'-impact variants are non-coding/intergenic variants, or variants affecting non-coding genes. SnpEff analysis was performed in variant files produced by alignment to both mRatBN7.2 and SHRSP/BbbUtx. For clarity, only SnpEff analysis of variants compared to the SHRSP/BbbUtx genome are presented. Most called variants were intergenic (Table 8.2, Table 8.3) and thus classified as 'modifier' variants.

**Table 8.2:** SnpEff Predicted Functional Outcomes of Called Variants in WKY/Gla from SHRSP/BbbUtx Reference Genome

| Predicted Impact | Predicted Effect | n |
|---|---|---|
| **High** | 3 prime UTR variant | 3 |
| | disruptive inframe deletion | 1 |
| | frameshift variant | 51 |

| | | |
|---|---|---|
| | inframe deletion | 1 |
| | intron variant | 24 |
| | splice acceptor variant | 12 |
| | splice donor variant | 12 |
| | splice region variant | 11 |
| | start lost | 1 |
| | stop gained | 12 |
| | stop lost | 9 |
| **Moderate** | disruptive inframe deletion | 37 |
| | disruptive inframe insertion | 47 |
| | inframe deletion | 19 |
| | inframe insertion | 17 |
| | missense variant | 2365 |
| | splice region variant | 42 |
| **Low** | 5 prime UTR premature start codon gain variant | 126 |
| | downstream gene variant | 1 |
| | intron variant | 754 |
| | non coding exon variant | 4 |
| | splice region variant | 889 |
| | stop retained variant | 3 |
| | synonymous variant | 4699 |
| **Modifier** | 3 prime UTR variant | 7630 |
| | 5 prime UTR variant | 1051 |
| | downstream gene variant | 63645 |
| | intergenic region | 1270211 |
| | intragenic variant | 1 |
| | intron variant | 520112 |
| | non coding exon variant | 256 |
| | non coding transcript variant | 12 |
| | upstream gene variant | 64113 |

**Table 8.3:** SnpEff Predicted Functional Outcomes of Called Variants in SHRSP/Gla from SHRSP/BbbUtx Reference Genome

| **Predicted Impact** | **Predicted Effect** | **n** |
|---|---|---|
| **High** | frameshift variant | 2 |
| **Moderate** | disruptive inframe insertion | 1 |
| | missense variant | 33 |
| **Low** | 5 prime UTR premature start codon gain variant | 1 |
| | initiator codon variant | 1 |
| | intron variant | 11 |
| | splice region variant | 12 |
| | synonymous variant | 53 |
| **Modifier** | 3 prime UTR variant | 121 |
| | 5 prime UTR variant | 19 |
| | downstream gene variant | 1245 |
| | intergenic region | 50016 |
| | intragenic variant | 1 |
| | intron variant | 11696 |
| | non coding exon variant | 5 |
| | upstream gene variant | 1408 |

There were 92 'high' impact genetic variants identified in the WKY/Gla (Table 8.2) and 2 in the SHRSP/Gla (Table 8.3). One of the 'high' impact variants identified in the SHRSP/Gla genome was shared with the WKY/Gla; a frameshift deletion (chr4:168032186_AT/A) in the *Tas2r120* gene. The c.562delT was predicted to cause loss of function in the *Tas2r120* bitter taste receptor. The other 'high' impact variant in the SHRSP/Gla was an insertion mutation (chr7:124043460_A/AC) predicted to cause a frameshift in the *Mlc1* gene on chromosome 7. This variant was in a region of high variance clustered on chromosome 7 (Figure 8.10). Comparing all available strains on the RGD to the mRatBN7.2, no frameshift variants in the *Mlc1* gene were found between available rat strains (Figure 8.3) and the mRatBN7.2.

In the WKY/Gla, the 92 'high' impact variants were within 68 protein coding genes (Figure 8.10). There is a 'high' impact insertion & splice donor (chr14:6837347_G/GCTGTCGC
variant within the *Dspp* gene on chromosome 14, predicted to result in Loss of Function (LOF). The *Dspp* gene is part of the SIBLING family containing *Spp1*, within the

chromosome 14 cardiac mas QTL and congenic region.

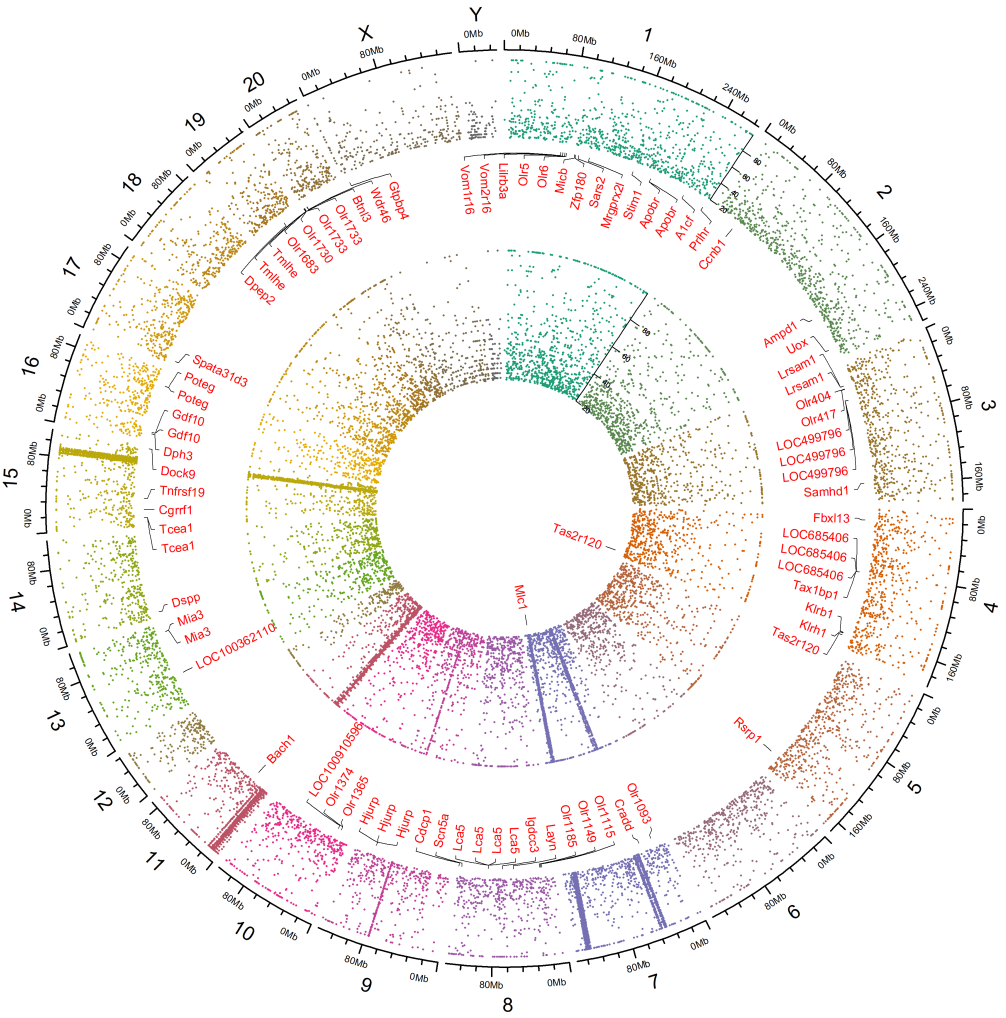



**Figure 8.10: Circos Plot of Variants in the Glasgow WKY and SHRSP strains to the SHRSP/BbbUtx** All variants found in the SHRSP/Gla (inner ring) and WKY/Gla (outer ring). Variants are plotted along their chromosomal location (x-axis track) and GQ score (y-axis). 'High' impact variants are labelled with gene name annotated from SnpEff.

Chromosomes 1, 3, 4, and 16 have a relatively high rate of 'high' impact variants (Figure 8.10), localised to specific regions. There were 69 variants in the WKY/Gla predicted to result in loss of function (LOF) or nonsense mediated decay (NMD), including the *Tas2r120* and *Dspp*. There is a 'splice acceptor' SNP (chr8:123069679_A/G) in the *Scn5a* gene which affects all transcripts and predicts LOF in the protein. The reverse SNP (G>A) is present in the SHRSP/Gla, SHRSP/A3NCrl, and SHR/OlalpcvMcwi compared to the mRatBN7.2 (EVAdb; rs197446846) but not any WKY substrains. There were 5 genes containing SNPs which were predicted to result in NMD due to the generation of a 'stop gained' mutation in the WKY/Gla: *Vom2r16, Lilrb3a, Klrb1, Tnfrsf19*, and *Btnl3* (Figure 8.10). A number of genes contained >1 'high' impact

mutation (Figure 8.10).

### 8.4.5 Variants within Regions Associated with Cardiac Mass

QTLs associated with cardiac mass were downloaded from the Rat Genome Database (`https://rgd.mcw.edu/rgdweb/search/qtls.html?100`). From the dataset, QTLs discovered using one of either the WKY or SHRSP for F2 generation were filtered and regions were overlaid with generated WKY/Gla and SHRSP/Gla variant catalogues. We identified 21 QTLs over 10 chromosomes for cardiac mass (Table 8.4). The relatively small Cm57 QTL region on chromosome 17 contained no annotated genes (using the region as a filter in Ensembl BiomaRt) and could not be investigated further. There were 233,271 variants unique to the WKY/Gla, 854 variants unique to the SHRSP/Gla, and 1,233 shared variants in genes contained within QTLs for cardiac mass. All variants in the SHRSP/Gla contained within QTLs were 'modifier' variants.

**Table 8.4:** Cardiac Mass QTLs

| Symbol | Chromosome | Size | Strains Crossed |
|---|---|---|---|
| Cm80 | 1 | 71815718 | F344/NCrl\|SHRSP/Bbb |
| Cm43 | 3 | 19577950 | SHRSP/Tkyo\|WKY/Tkyo |
| Cm48 | 3 | 19577950 | WKY/Tkyo\|SHR/Tkyo |
| Cm46 | 3 | 19577950 | SHRSP/Tkyo\|WKY/Tkyo |
| Cm58 | 3 | 77228957 | WKY/Cfd\|WKHA/Cfd |
| Cm25 | 3 | 27980720 | WKY/NCruk\|F344/NHsd |
| Cm24 | 5 | 14628930 | WKY.WKHA-(D5Rat45-D5Rat245/Cfd) |
| Cm54 | 5 | 22867670 | WKY/Cfd\|WKHA/Cfd |
| Cm74 | 6 | 46355327 | SHRSP/Tkyo\|WKY/Tkyo |
| Cm77 | 6 | 47823340 | WKY/Tkyo\|SHR/Tkyo |
| Cm47 | 6 | 46355327 | SHRSP/Tkyo\|WKY/Tkyo |
| Cm75 | 10 | 38358252 | SHRSP/Tkyo\|WKY/Tkyo |
| Cm44 | 10 | 38358252 | SHRSP/Tkyo\|WKY/Tkyo |
| Cm78 | 10 | 36937156 | WKY/Tkyo\|SHR/Tkyo |
| Cm56 | 12 | 29333306 | WKY/Cfd\|WKHA/Cfd |
| Cm45 | 13 | 45000000 | SHRSP/Tkyo\|WKY/Tkyo |
| Cm76 | 13 | 45000000 | SHRSP/Tkyo\|WKY/Tkyo |
| Cm79 | 13 | 18302022 | WKY/Tkyo\|SHR/Tkyo |
| Cm59 | 15 | 6987058 | WKY/Cfd\|WKHA/Cfd |
| Cm55 | 17 | 29808941 | WKY/Cfd\|WKHA/Cfd |
| Cm57 | 17 | 54231 | WKY/Cfd\|WKHA/Cfd |

Focusing on 'high', 'moderate', and 'low' effect variants, genes in cardiac mass QTLs contained a number of synonymous and missense variants in the WKY/Gla which were not found in the SHRSP/Gla genome (Figure 8.11). QTL regions were also associated with intron variants affecting splice regions (Figure 8.11). There were a few cases of stop gained effects, including *Mtus2* on chromosome 12 and *Lrsam1* on chromosome 3. The *Lrsam1* gene on chromosome 3 also contained a frameshift insertion. Chromosome 3 contained a high density of SNPs including variants in Xin repeat protein 2 *Xirp2*, and a number of sodium-voltage gated ion channel alpha subunits; *Scn7a* and *Scn3a*. Three of the QTLs on chromosome 3 overlapped and most variants were clustered together (Figure 8.11).

*LOC100362110* contained a frameshift deletion within the chromosome 13 QTL, as did the *Apobr* gene on chromosome 1, with loss of function predicted in both cases. On chromosome 1, the *Stim1* gene contained a stop lost SNP that formed part of the cluster of 'high' impact variants within the QTL on chromosome 1. The Cm80 QTL was identified in an F344/NCrl|SHRSP/Bbb F2 population and the region represents a site of high variance in the SHRSP substrains, potentially associated with hypertension and associated end organ damage.


**Figure 8.11: Variants in QTLs Associated with Cardiac Mass** Genes contained within QTL associated with cardiac mass containing variants are plotted, where x- position determines position along chromosome and y- position indicates relative gtQuality score. 'High' and 'Moderate' genes are labelled, 'low' effect genes are not labelled and 'modifier' genes were not plotted for clarity. Colour denotes predicted effect, which may overlap when variants have more than one predicted effect.

On chromosome 5 there are two distinct QTLs where the associated genetic regions did not overlap one another (Figure 8.11). Both QTLs were identified using a WKY but not SHRSP as a parental strain to F2 progeny (Table 8.4). Despite a number of variants in the WKY/Gla, there are only two variants in the SHRSP/Gla genome within the chromosome 5 QTL region. In the WKY/Gla, the *Runx3* transcription factor contained a number of synonymous SNPs and a SNP affecting a splice region (chr5:151888834_G/C, c.698-8G>C) with a 'low' predicted effect. *Runx3* also contained a missense SNP (chr5:151889229_C/G, p.Thr362Ser) predicted to have a 'moderate' effect. Using the mRatBN7.2 as reference, WKY/N and WKY/NCrl substrains share a number of variants in the *Runx3* gene, including the missense variant in the final exon (EVAdb; rs199045902 C>G), that is not shared by the SHRSP/A3NCrl or the SHRSP/Gla.

### 8.4.6 Overlaying Variants with Genes Identified in Differential Expression and Transcript Usage Analysis at GD18.5

Next, genes which displayed evidence of differential transcript usage and gene expression in the RNA sequencing data (Chapter 7) were overlaid with genes containing variants discovered in SHRSP/Gla and WKY/Gla alignment to the SHRSP/BbbUtx. There were 48 genes that contained a 'high' or 'moderate' impact variant and significant evidence of DTU in the comparison of GD18.5 transcriptome. Associated with genetically determined cardiac hypertrophy in the HHR, there was evidence of DTU in the *Trim55* gene between the SHRSP vs WKY heart (Figure 8.12a). The WKY/Gla but not the SHRSP/Gla has a missense SNP (chr2:106793391_T/C) in the *Trim55* gene which causes a p.Ser513Pro mutation in the *Trim55* gene. The *Trim55* gene is the QTL region Cm22, and the reciprocal variant can be mapped to rs198060252 in the EVAdb on the mRatBN7.2 genome. The HHR rat, SHR/OlalpcvMcwi, SHRSP/A3NCrl and SHRSP/Gla have the C T mutation which is not present in the WKY/Gla or any WKY sub-strain.

The *Scn3a* and *Xirp2* genes from the chromosome 3 QTL region were significantly downregulated in the SHRSP vs WKY GD18.5 cardiac transcriptome (Figure 8.12b). Localising to a different chromosome to *Scn3a*, the sodium-voltage gated channel family member, *Scn5a*, showed evidence of significant DTU in the SHRSP vs WKY GD18.5 RNA sequencing data and contained a splice acceptor variant in the WKY/Gla (Figure 8.12a). There are more than 200 variants in the *Scn5a* gene found in the WKY/Gla genome, including 4 SNPs in the 3'prime UTR, a missense SNP (chr8:123004027_C/T), and an additional splice region SNP (chr8:123018789_A/G). Downstream of *Scn5a*, an inframe deletion variant was found in the functionally associated *Scn10a* gene (chr8:123184978_CAGG/C) in the WKY/Gla genome.

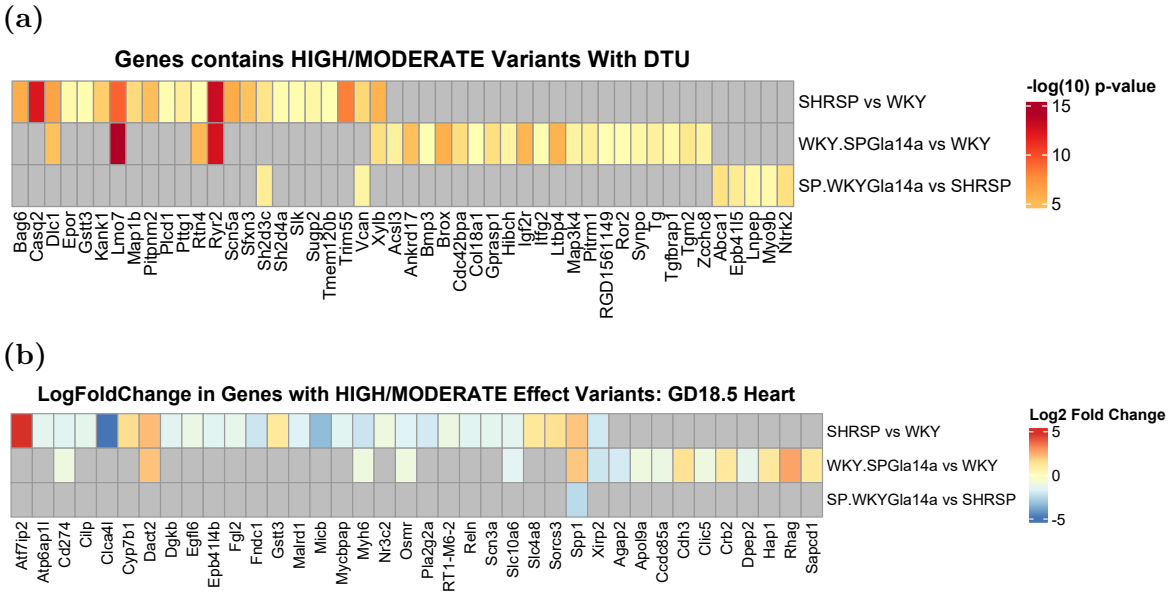



**Figure 8.12: Expression of High and Moderate Impact Variants in GD18.5 Hearts** Heatmap of (a) log(10)p-adjusted value and (b) Log2FoldChange of genes with both evidence of significant DTU or DGE at the 95% significance level, and contained 'high' or 'moderate' effect variants in the WKY/Gla and SHRSP/Gla whole genome alignment to the SHRSP/BbbUtx.

One differentially regulated gene which contained a 'high' impact variant was *Micb* (MHC class I polypeptide-related sequence B) (Figure 8.12b), which was downregulated in the SHRSP vs WKY GD18.5 cardiac transcriptome. The 'high' impact SNP (chr1:92457350_A/G, p.Ter174Trpext*/c.A521G) was predicted to create a 'stop lost' effect. There were no variants found SHRSP/Gla genome compared to the SHRSP/BbbUtx. When mapped to the mRatBN7.2, the reciprocal SNP was present in the WKY/Gla alignment to the mRatBN7.2 but not the SHRSP/Gla. The SNP (rs197565950., A>G) is also found in all WKY substrains, and the SHR/OlalpcvMcwi, but not the SHRSP/A3NCrl or SHRSP/Gla. The *Micb* variant localises within the chromosome 1 QTL, along with the *Stim1* gene, which also contains a high impact 'stop-lost' variant. Both genes (*Stim1* and *Micb*) were downregulated in the SHRSP heart at GD18.5 compared to the WKY, although downregulation of the *Stim1* gene was not significant.

The WKY/Gla genome contained a number of variants in the *Spp1* gene compared to the SHRSP/BbbUtx reference, which were not found in the SHRSP/Gla comparison. *Spp1* was the only gene containing 'high' or 'moderate' impact variants which showed significant evidence of differential expression in all three of the comparison groups (Figure 8.12b). The *Spp1* gene has 11 variants contained between the 5'- to 3'-UTR, all of which are unique to the WKY/Gla. There are also 50 variants within 5K bp downstream and 22 variants within 5K bp upstream of *Spp1* gene which are unique to the WKY/Gla. The JBrowse view of *Spp1* with the SHRSP/BbbUtx as a reference is

displayed in Figure 8.13.

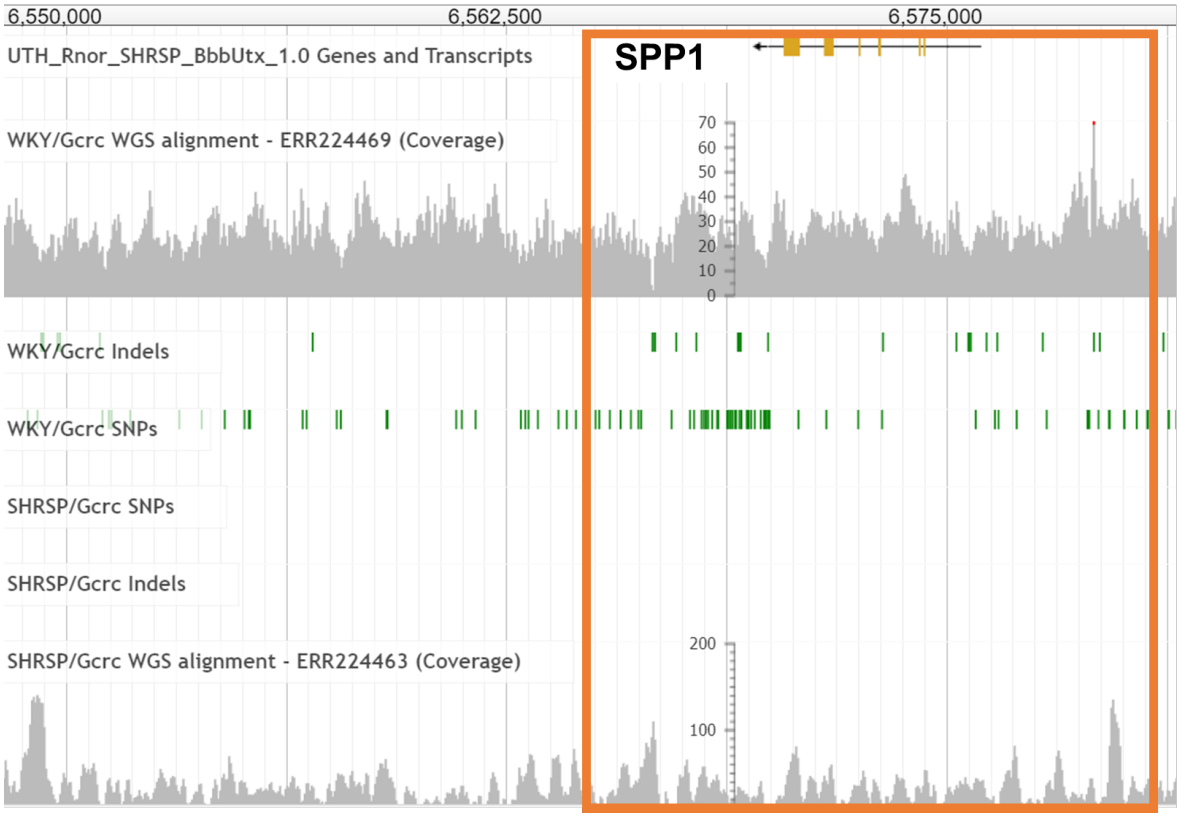



**Figure 8.13: Variants in *Spp1* between the SHRSP/BbbUtx and Glasgow Strains** Rat Genome Database JBrowse with WKY/Gla and SHRSP/Gla WGS coverage and high quality SNP/InDels displayed as tracks. Orange box highlights region 5KB up and downstream of the *Spp1* transcript in the SHRSP/BbbUtx genome. Green lines represent location of SNPs, insertions, and deletions present in the WKY/Gla (WKY/Gcrc) and SHRSP/Gla (SHRSP/Gcrc) discovered following whole genome alignment to the SHRSP/BbbUtx

Apart from 3, all variants in the *Spp1* CDS are predicted to have a 'modifier' effect and are primarily contained within the introns (4 InDels, 3 SNPs). The 3'-UTR/splice region contained a SNP (chr14:6569945_T>A, c.*427 A>T, Figure 8.14a) in the WKY/Gla genome compared to the SHRSP/BbbUtx. A synonymous SNP (chr14:6571559_G/A, p.Tyr150Tyr) was also discovered in exon 7, predicted to have a 'low' impact.

The 3-UTR/spice region variant T>A, is also present in the WKY/Gla, WKY/NCrl and WKY/N comparison to the mRatBN7.2 reference genome. The SNP, catalogued as rs3322317194 in the EVAdb, is not present in the SHRSP/Gla comparison to the mRatBN7.2, and is not found in either the SHRSP/A3NCrl, or SHRSP/Gla (Figure 8.14a).


**Figure 8.14: Mapping *Spp1* 3-UTR and Missense Variants in the mRatBN7.2** JBrowse view of variants discovered in strain comparison to the mRatBN7.2. Variants in the WKY substrains, WKY/NCrl and WKY/N, and the substrains SHR/OlalpcvMcwi, SHRSP/A3NCrl are also displayed on the tracks. Orange box denotes (a) chr14:6569945_T>A, c.*427 A>T and (b) chr14:6570767_G/T, p.Leu187Ile from the alignment to the SHRSP/BbbUtx.

A missense SNP in the WKY/Gla genome compared to the SHRSP/BbbUtx was found in exon 7. The SNP (chr14:6570767_G/T, p.Leu187Ile) was predicted to have a 'moderate' impact. When compared to the mRatBN7.2, the SHRSP/Gla but not the WKY/Gla, has a p.Ile187Leu, 'moderate' missense mutation which is annotated as rs8174959 in the EVAdb. It is not shared with the SHR/OlalpcvMcwi but is common

SNP to the SHRSP/BbbUtx, the SHRSP/Gla, and the SHRSP/A3NCrl (Figure 8.14b).

### 8.4.7 Short Variant Discovery in RNA-sequencing Data Aligned to SHRSP/BbbUtx Reference

Alignment of RNA-sequencing data from GD18.5 hearts was performed using the SHRSP/BbbUtx as a reference. The substrains WKY/Gla and SHRSP/Gla were previously referred to as WKY and SHRSP in Chapter 7. For continuity, unless otherwise stated, WKY and SHRSP will refer to the Glasgow sub-strain in variant analysis from RNA sequencing data presented.

From these alignments, variant discovery using GATK pipeline generated a number of variants in WKY/Gla, SHRSP/Gla and chromosome 14 congenic strains. A high proportion of variant calls had a phred score <20 (Figure 8.15a). In the SHRSP/Gla TiTV ratio was <2 and indicates a potential increase in false positives. The TiTV ratio was grater than random (>0.5) and variants were with the phred scaled probability >20 were taken forward for analyses. TiTV ratio was >2 in WKY/Gla, WKY.SPGla14a and SP.WKYGla14a (Figure 8.15b).

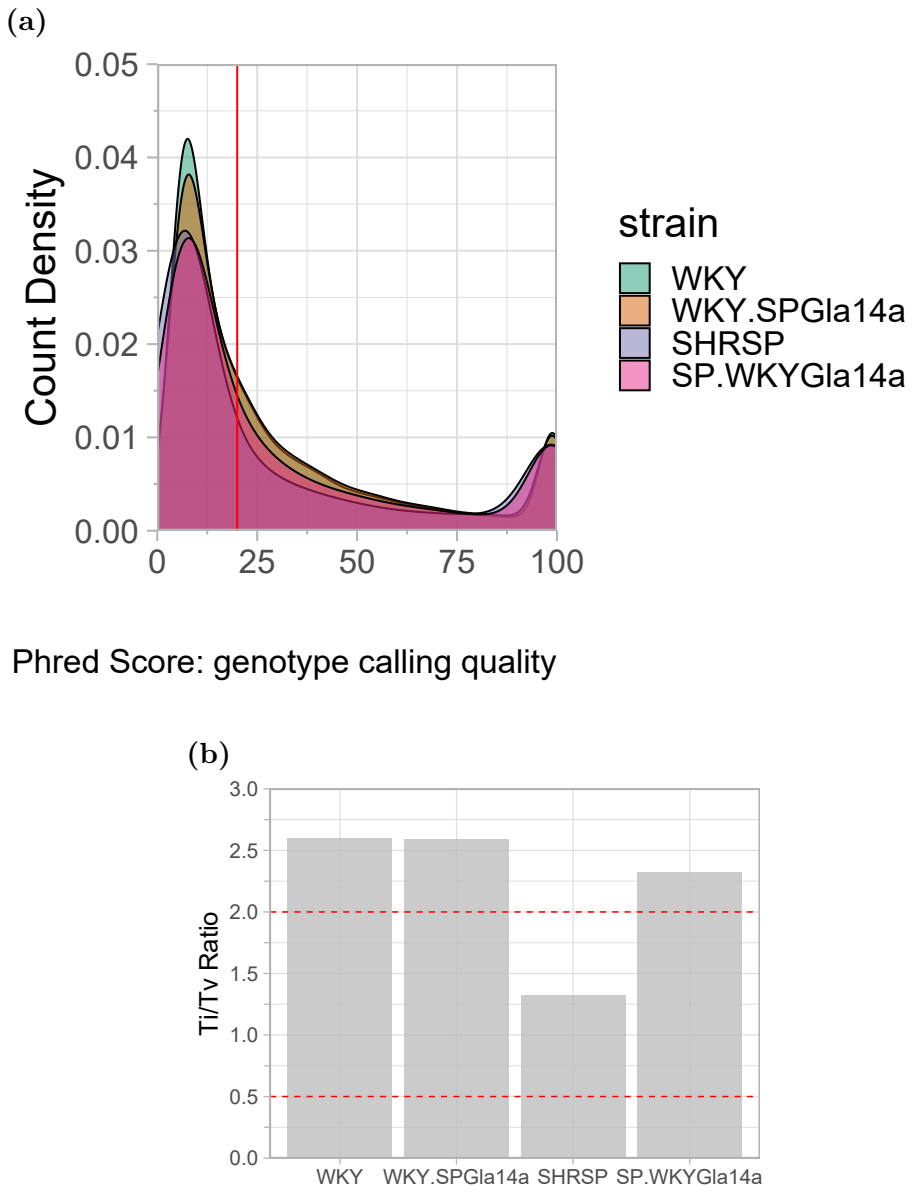



**Figure 8.15: GD18.5 Alignment of Short Read RNA Sequencing to SHRSP/BbbUtx Genome** (a) Density plot of Phred-scaled likelihood of genotype quality represented by the 'GQ' score. GQ call of <20 was used as a filter for variants called, corresponding to <1% chance that a call is incorrect, and is indicated by horizontal red line, (b) Transition to transversion ratio in variants called from alignment of RNA-sequence data to the SHRSP/BbbUtx.

Variant analysis called over 120,000 variants from RNA-sequencing reads in the WKY/Gla and WKY.SPGla14a alignments to the SHRSP/BbbUtx. Within the transcribed regions of the genome, 33,627 and 5,305 variants were discovered unique to the WKY/Gla and SHRSP/Gla respectively. The majority of variance in the WKY/Gla was shared with the WKY.SPGla14a, such that 74,165 variants were shared between these two strains (Figure 8.16). The WKY.SPGla14a had the greatest number of variants, and the SHRSP/Gla the least. A full overview of shared and unique variants is demonstrated in Figure 8.16.

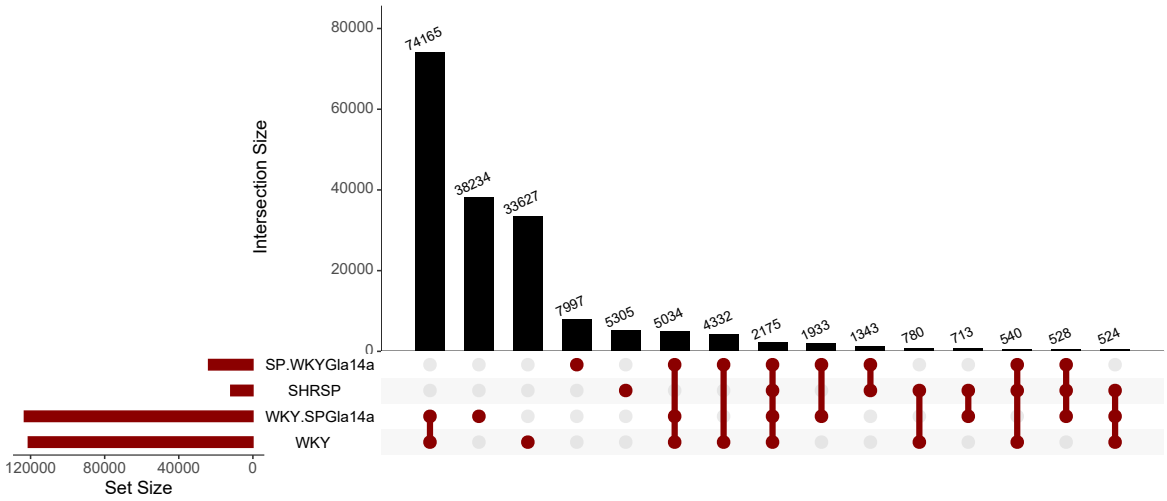



**Figure 8.16: Upset Plot of Shared and Unique Variants Called from RNA-seq Alignments** UpSet plot shows intersections in a matrix, the rows corresponding to the sets of variants called from WKY/Gla, SHRSP/Gla, WKY.SPGla14a and SP.WKYGla14a RNA-sequences. The columns correspond to the intersections between these sets.

One a whole genome level, the WKY/Gla and WKY.SPGla14a had the highest number of variants (Figure 8.17a). When only chromosome 14 variants were considered, the WKY/Gla and SP.WKYGla14a strains had the highest number of variants (Figure 8.17b). There were 177,230 variants called across the 4 strains from alignment of short read RNA sequencing data to the SHRSP/BbbUtx reference. Of these, 34,969 were not included in the variants called from WGS analysis ($\sim$ 19.7%). The mean gtQuality and altCounts did not appear to differ significantly across variants which were found in WGS analysis and those which were not.

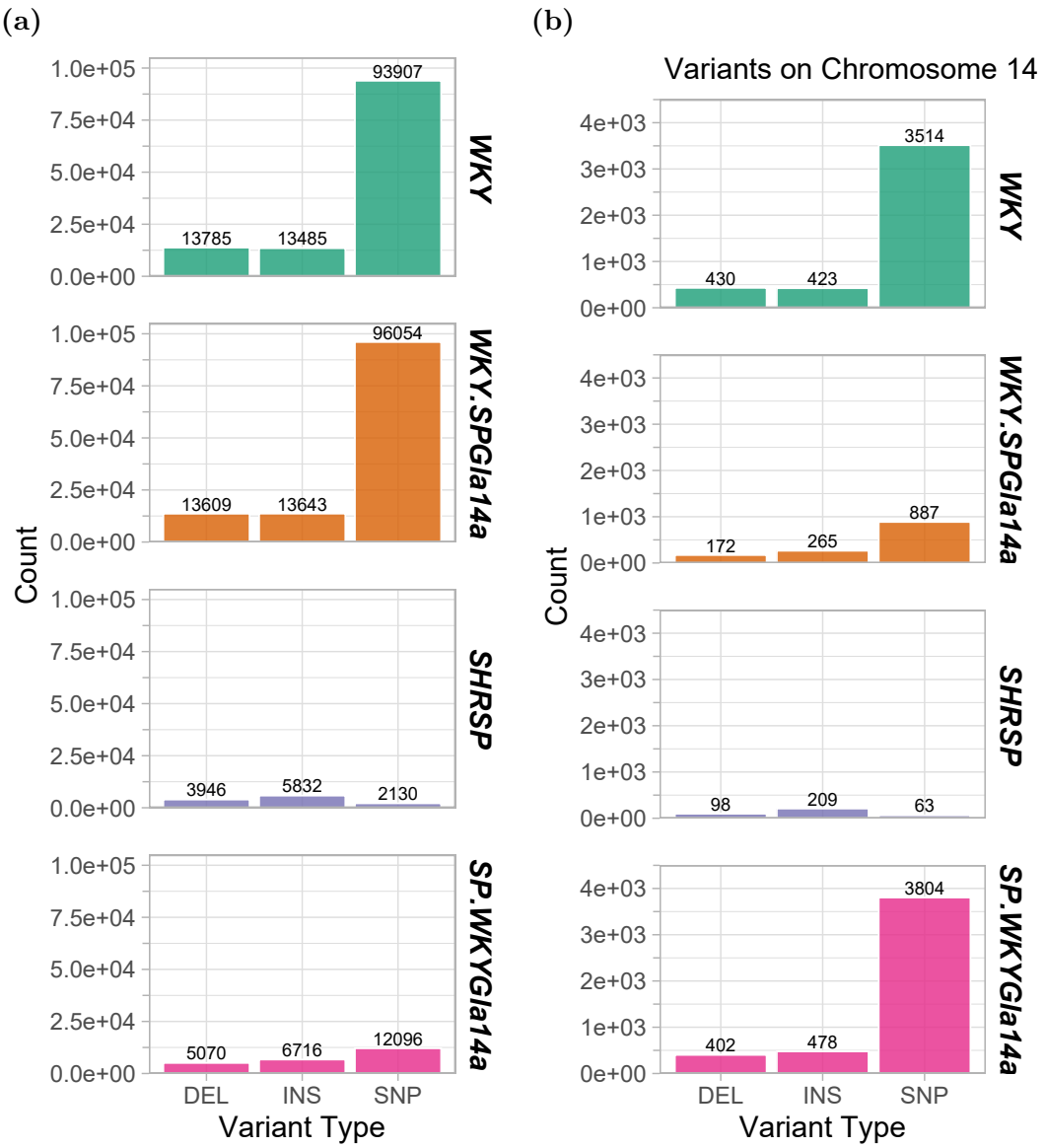



**Figure 8.17: Variant Discovery in GD18.5 RNA Sequencing Data** Variants called from alignment of short read RNA sequencing data to the SHRSP/BbbUtx (a) across the whole genome and (b) restricted to chromosome 14. Variant calling was performed in WKY/Gla, SHRSP/Gla and chromosome 14 congenic strains WKY.SPGla14a, and SP.WKYGla14a. Total number is displayed on the top of each bar.

### 8.4.8 SnpEff Predictions Replicated by Variant Discovery in RNA Sequencing Data

SnpEff predictions were merged with variants found in short read RNA sequencing data such that only variants identified in the original WGS were annotated with predictions. There were no variants with 'high' predicted effect in SHRSP-derived RNA sequencing data (Figure 8.18). The stop lost on chromosome 1 gene *Stim1* was present in the WKY/Gla and WKY.SPGla14a but not SP.WKYGla14a, which follows parental background as expected. The variant in the *Micb* gene and *Scn5a* on chromosomes 1 & 8 were also present in the RNA sequencing variant calls from WKY/Gla and WKY.SPGla14a strains. There were no 'high' impact variants on chromosome 14.

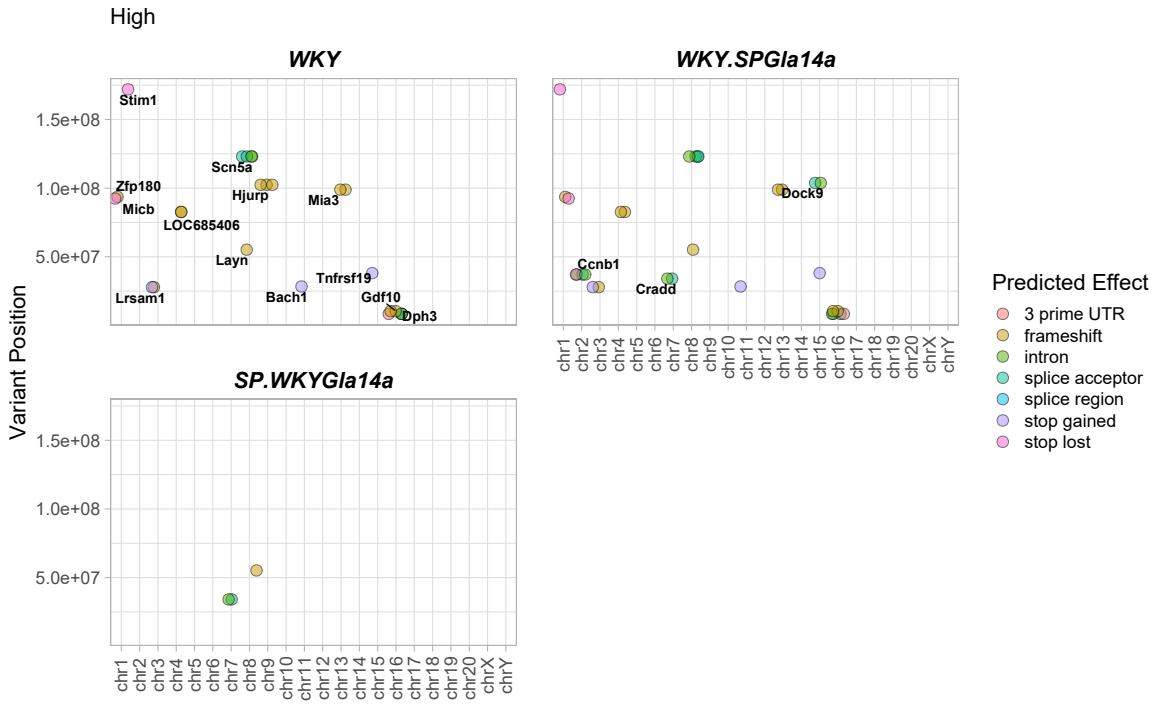



**Figure 8.18: 'High' Impact Variants Called in GD18.5 RNA Sequencing Data** Bubble plot of variants found along each chromosome in RNA sequence variant calls from alignment of reads to the SHRSP/BbbUtx. 'High'-impact variants are plotted and labelled with gene names, the maximum number of labels are overlaid on plot avoiding overlaps. Position along the y-axis indicates start point along chromosome.

Most 'moderate' variants in RNAseq variants were missense variants (Figure 8.19). Chromosome 14 contained 39 variants across 23 genes with a 'moderate' predicted effect, all of which were missense mutations apart from 1 splice region variant in the *Ptpn13* gene (chr14:12949628_G/A). This variant was identified in the SP.WKYGla14a only. The p.Leu187Ile missense SNP in the *Spp1* gene was found on both the WKY/Gla and SP.WKYGla14a transcriptome but not the SHRSP/Gla and WKY.SPGla14a. The *Rtn4* variant within chromosome 14 was also found in both chromosome 14 congenic strains and the WKY/Gla strain, but not SHRSP/Gla.

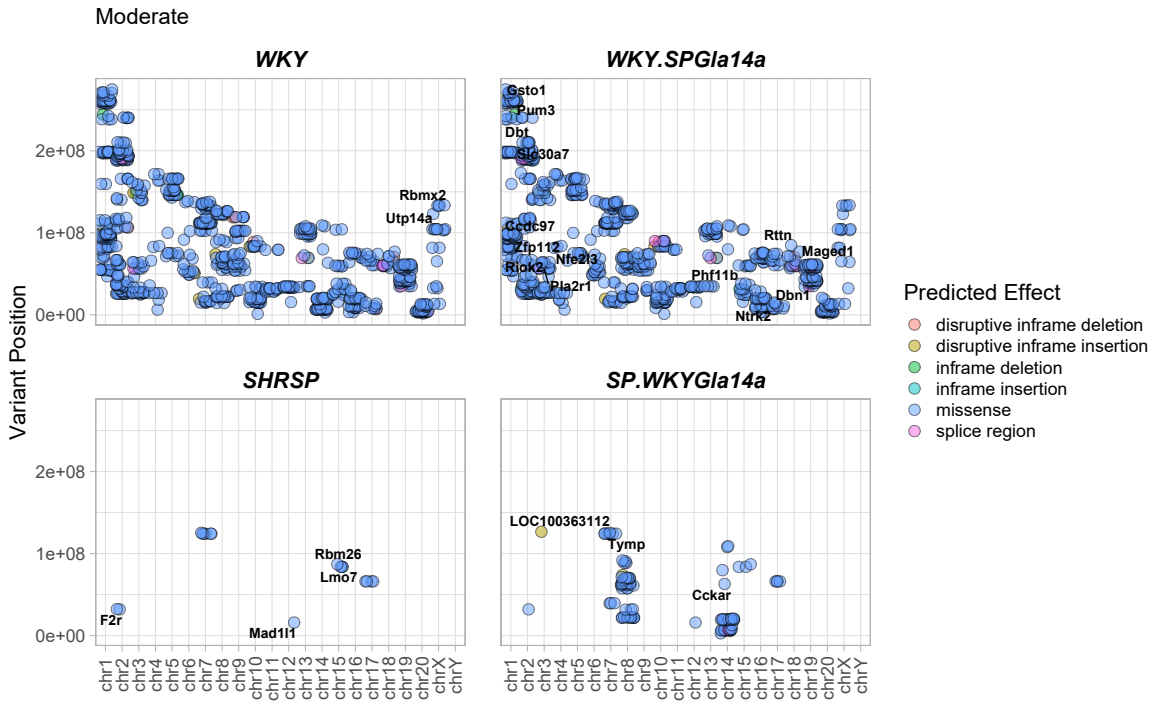



**Figure 8.19: 'Moderate' Impact Variants Called in GD1.8 RNA Sequencing Data** Bubble plot of variants found along each chromosome in RNA sequence variant calls from alignment of reads to the SHRSP/BbbUtx. Position along the y-axis indicates start point along chromosome. 'Moderate'-impact variants are plotted and labelled with gene names, the maximum number of labels are overlaid on plot avoiding overlaps

The 'low'-impact predictions were mainly synonymous variants and again followed a pattern that indicated agreement between background congenic and parental strains (Figure 8.20). The second most common effect was a splice region variant, or premature start codon gain at the 5'-UTR. On chromosome 14, there were 72 'low' impact variants, which affected 36 genes. These were mostly synonymous mutations (67 variants) affecting 34 genes. A 5'-UTR premature start codon gain & intron variant was identified in 2 genes; *LOC689986*, and *Antxr2*. Finally, variants predicted to affect the splice regions were replicated in 5 genes; *Antxr2, Cnot61, Rufy3, Abcg3* and, *LOC679894*.

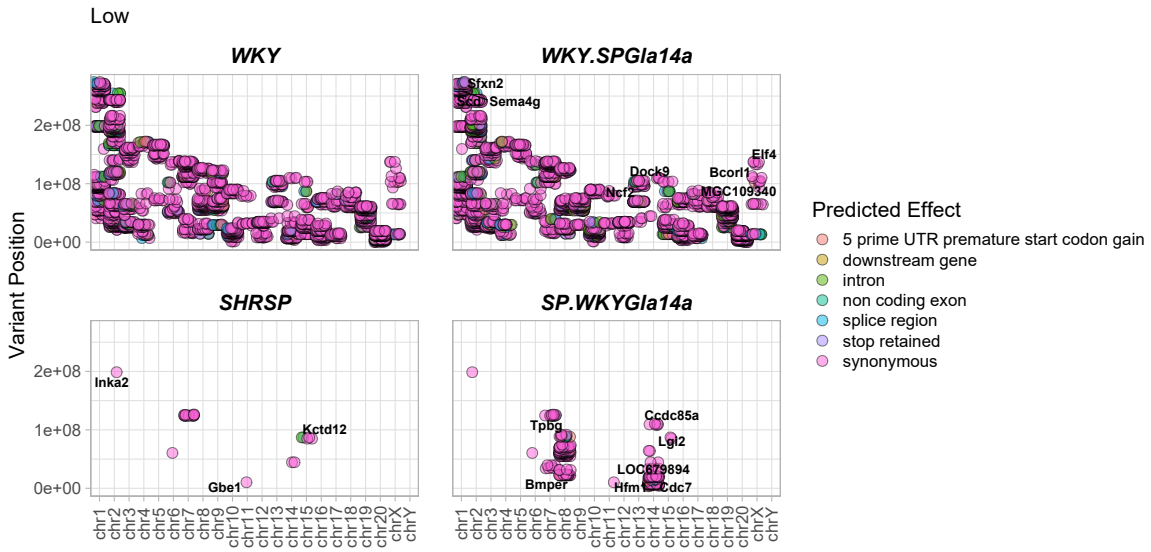



**Figure 8.20: 'Low' Impact Variants Called in GD1.8 Sequencing Data** 'Low'-impact variants are plotted and labelled with gene names, the maximum number of labels are overlaid on plot avoiding overlaps. Bubble plot of variants found along each chromosome in RNA sequence variant calls from alignment of reads to the SHRSP/BbbUtx. Position along the y-axis indicates start point along chromosome.

## 8.5 Discussion

As the rat reference assembly has improved, publications cataloguing variance between inbred, outbred, and heterogenous stock strains from the BN reference have allowed continuous annotation of 'known' variants in the rat genome. Using the newest and most complete reference genome, mRatBN7.2, alignment and variant calling was performed in WKY/Gla and SHRSP/Gla sub-strains. Almost 4 million variants were shared between the strains and reiterates the relative phylogenetic distance between the Japanese Wistar-derived strains and the reference Brown Norway (BN). Data retrieved from the Rat Genome Database supported the relative genetic distance between the BN and Wistar-derived strains where SHR/OlalpcvMcwi, WKY/NCrl, SHRSP/A3NCrl, GK/FarMcwi and WKY/N have the highest InDel and SNP count across 30 inbred strains. A phylogenetic tree of 120 strains and substrains recently constructed also indicates the Wistar strains have relatively higher genetic divergence from the mRatBN7.2 reference (de Jong et al., 2023).

The primary aim of experimental work in this Chapter was to determine genetic factors contributing to cardiac structure (whole heart and left ventricle mass), which is increased in the SHRSP/Gla from 5-weeks of age, and persists to adulthood after the onset of increased blood pressure. The SHRSP genome represents a 'high-risk' genome, predisposing to development of cardiac hypertrophy during adulthood. Conversely, the WKY genome represents a relatively 'protected' genome, where LVH is inducible by diet or surgical intervention but not predicted by genetic factors alone. In

addition to almost 4 million shared genetic variants, the WKY/Gla and SHRSP/Gla genome contained an additional 1.36 and 1.03 million variants that were strain specific. Comparison to the mRatBN7.2 implicated over 2 million variants which could be collectively or individually contributing to relative protection or risk of cardiac hypertrophy. A number of variants were also potentially 'missed', where reads which could not be aligned were removed from analysis.

To overcome these potential limitations in variant analysis and improve annotation of functional variants, the WKY/Gla and SHRSP/Gla WGS reads were aligned to the reference quality SHRSP/BbbUtx genome (also known as SHR-A3/Utx; SHR-A3), an SHRSP sub-strain housed at the University of Texas (Houston, TX). Using the genetically more similar strain as a reference, the number of shared variants was reduced to 5,070, implicating a further 1.6 million sites of variance between the WKY/Gla and the SHRSP/BbbUtx. Whilst phenotypes between Glasgow SHRSP and WKY strains have not been directly compared to the SHRSP/BbbUtx, the SHRSP/BbbUtx develops hypertension, cardiac hypertrophy, renal injury, and stochastic stroke similarly to the SHRSP/Gla. The 1.6 million variants found against the SHRSP/BbbUtx reference in the WKY/Gla likely represent genetic sources of relative protection against hypertension and developed end-organ injuries. To increase confidence in variants called, variants discovered from alignment of RNA sequencing data (GD18.5 hearts) to SHRSP/BbbUtx was performed and compared to WGS. The ribosomal depletion library prep maintained a greater amount of genetic sequence however any variants not within transcribed regions of the genome will not be found in these data, and 'validation' of variants in this respect is caveated by the origin of the sequence reads.

Consistently, *Spp1* is expressed at a higher level in strains with SHRSP-chromosome 14 genome. There were a number of SNPs, insertions, and deletions in the WKY/Gla compared to the SHRSP/BbbUtx in the *Spp1* gene which were not present in the SHRSP/Gla and other SHR/SHRSP substrains. The congenic region containing *Spp1* is located in the first 4-32Mb of chromosome 14, a region of comparatively high variance in the WKY/Gla versus the SHRSP/BbbUtx, with an average rate of >7 variants per 10,000 bases. In contrast, the variant rate of the SHRSP/Gla from the SHRSP/BbbUtx in this region was <2 per million bases. Compared to the SHRSP/BbbUtx, most of the variance in the WKY/Gla *Spp1* gene was located within introns or the intergenic regions up and downstream of coding sequences. Exonic variants were mostly synonymous, however a 3' prime UTR and missense SNP was discovered in the WKY/Gla compared to the SHRSP/BbbUtx, resulting in a Leucine to Isoleucine amino acid substitution (p.Leu187Ile) in the WKY/Gla compared to the SHRSP strains (SHRSP/BbbUtx and SHRSP/Gla). The relative similarity of these amino acids suggest the variants within introns, UTR-, and intergenic regions are likely to be driving observed differences in *Spp1* expression in SHRSP and chromosome 14 congenic strains compared to

the WKY.

The risk-enhancing role of *Spp1* is conferred by overexpression of the gene. In the H9c2 cell model, overexpression of both WKY and SHRSP-derived *Spp1* results in cell growth irrespective of genetic background (data not published). During cardiac development, the neonatal period, and in adulthood, the WKY genome is associated with reduced *Spp1* expression compared to the SHRSP. The 3' UTR, located downstream of coding sequence, is involved in RNA stability, translation, and localisation (Steri et al., 2018). miRNAs can act as regulatory units following transcription, and the miRNA-433 has been shown to negatively regulate *Spp1* expression by directly binding the 3' UTR of *Spp1* mRNA in osteoblast cells (Han et al., 2021). It is also plausible intron variants and variants upstream of *Spp1* act to promote or reduce transcription. Binding sites for the transcription factors RUNX2, SP1 and MYT1-zinc finger are altered by variants in the human SPP1 gene (Giacopelli et al., 2004). Preliminary data from within this lab has suggested variants in the SHRSP and WKY *Spp1* promoter region can affect promoter activity in H9c2 cells, but not HeLa cells (Asirvatham, 2022). Like miRNAs, RNA binding proteins and transcription factors can have an inhibitory effect, the SRY-related transcription factor, Sox9, is associated with reduced *Spp1* expression in the mouse heart, and is dependent on RUNX2 binding (Peacock et al., 2011; Pritchett et al., 2012). Analysis of transcription factor binding in the SHRSP/Gla and WKY/Gla sequences could determine transcript promoting or inhibiting variants with improved accuracy compared to initial efforts (Asirvatham, 2022), where high variance in both WKY/Gla and SHRSP/Gla from the Rnor_6 reduced ability to accurately generate sequences.

Variants not on chromosome 14 but known to be associated with CVD were found in *Trim55*, *Stim1* and *Micb* genes, which were also subject to differential transcript usage and expression in SHRSP vs WKY cardiac transcriptome comparisons. The 'stop lost' variant on the *Stim1* gene maps to rs3319087534 in EVAdb, and is a 'stop-gain' SNP in SHRSP/Gla and SHRSP/BbbUtx strains compared to the mRatBN7.2 (and WKY strains). Present in the SHRSP/A3NCrl and SHRSP/Gla but not the WKY substrains or SHR/OlalpcvMcwi, the C>T polymorphism results in truncation of the carboxy terminus of the STIM1 protein, and has been associated with renal injury in the SHRSP (Dhande et al., 2020). Compared to the SHRSP/BbbUtx reference, the loss of the premature stop codon the WKY/Gla was replicated in GD18.5 RNA-sequencing. In GD18.5 hearts there is a relative decrease in *Stim1* expression in SHRSP foetus' compared to the WKY, which was not significant. A congenic model of *Stim1* rescue in the SHRSP (by introduction of SHR region) rescued renal phenotype but did not reduce *Stim1* expression in the kidneys (Dhande et al., 2020). Along with a cluster of variants in *Micb*, *Bag3*, *Orai3*, and *Apobr*, the *Stim1* mutation localises in a chromosomal region associated with cardiac mass by QTL analysis (Cm80, F344/NCrl|SHRSP/Bbb

F2 progeny). Variants in both *Stim1* and *Orai* gene family of have been associated with dysregulation of calcium handling in cardiac response to pressure overload, and cardiac hypertrophy (Rosado et al., 2015; Benard et al., 2016; Ohba et al., 2017). In the rat, *Stim1* and *Orai3* are co-localised on chromosome 1, which is not the case in humans and does not represent a highly conserved gene cluster. The rat *Orai1* and *Orai2* genes are localised to chromosome 12 within one of the localised regions of high variance in the WKY/Gla compared to the SHRSP/BbbUtx. Chromosome 12 also contains a QTL for cardiac mass and might be contributing to increased genetic risk of cardiac physiology in SHRSP substrains through the contribution of variants affecting function of genes within the store-operated $Ca^{2+}$ entry (SOCE) pathway (Rosenberg et al., 2021).

A SNP in the WKY/Gla *Scn5a* gene, located at an intron & splice acceptor region (chr8:123069679_A/G), corresponded with the high impact SNP, rs197446846, discovered in alignment of 163 strains to the mRatBN7.2 (de Jong et al., 2023). SHRSP genome contained a G>A transition SNP compared to mRatBN7.2 and WKY substrains. Mapping high impact rat variants to human GWAS hits shows variants in the SCN5A gene are associated with a number of electrocardiographic traits including P wave duration, PR interval, atrial fibrillation, QRS duration, QT interval, T wave morphology measurement, and heart function measurement (de Jong et al., 2023). The positionally adjacent SCN10A has also been associated electrocardiographic traits (Sollis et al., 2023). Common variants in SCN5A and neighbouring SCN10A have also been associated with cardiovascular age, and the rare polygenetic Brugada syndrome, a disease with a high risk of sudden cardiac death (Bezzina et al., 2013; Libiseller-Egger et al., 2022). Similarly to *Spp1*, *Scn5a* is located in a region of high variance between the WKY/Gla and SHRSP/BbbUtx (1/1000 bases), coupled with lower variance in the SHRSP/Gla (<0.025/1000 bases).

In addition to *Scn5a*, other sodium voltage-gated channel alpha subunits were highlighted in variant analyses, in particular *Scn7a* and *Scn3a* in the QTL region on chromosome 3 associated with cardiac mass. *Scn3a* was also significantly down-regulated in the SHRSP compared to WKY at GD18.5. Using F2 progeny from SHRSP and WKY substrains, a genetically determined alteration of baseline HR regulation was linked to a chromosome 3 region containing a number of SCN-$\alpha$ subunits including 3a, 2a, 1a, 9a and 7a (Kreutz et al., 1997). In the WKY/Gla genome, there were over 25,000 variants on the chromosome 3 region containing QTLs associated with cardiac mass, with a number of these affecting splice regions. This includes the Xin repeat protein 2 (*Xirp2*, also known as *Cmya3*), which was significantly downregulated in GD18.5 hearts (SHRSP and WKY.SPGla14a vs WKY). The localisation of SCN- genes and XIRP2 is highly conserved across human, mouse, and chicken species (Q. Wang et al., 2014), with XIRP2 co-localising within 2q24.3 on the human chromosome 2 (Sollis

et al., 2023). *Xirp2* is an important gene in cardiac development, ensuring the correct localisation of mature intercalated disks (Farrell et al., 2018). Similarly to the SIBLING family of genes on chromosome 14, the localisation of variants in sodium voltage channels and *Xirp2* on chromosome 3 represent highly conserved syntenic genes which are likely to be functionally related and under the control of the same transcription units (Moreno-Hagelsieb et al., 2001). These represent good candidates for translation between humans and rat model organism studies.

Comparison of a number sub-strains to the mRatBN7.2 suggested variance from the BN reference strain was concentrated in olfactory and haemoglobin genes (Hermsen et al., 2015). These rapidly evolve and are highly variable in many species. The SHRSP/Gla contained two 'high' impact variants compared to the SHRSP/BbbUtx, including the *Tas2r120*, the bitter taste receptor. The single exon gene contains no SNPs or InDels in the SHRSP/Gla and WKY/Gla alignments to the mRatBN7.2. Of the strains available on JBrowse and in the RGD Variant visualiser, a single SNP exists within the *Tas2r120* (C>T), which is not present in any WKY or SHRSP sub-strain (SHRSP/BbbUtx not available). The frameshift deletion, called in both the SHRSP and WKY/Gla strains compared to the SHRSP/BbbUtx, potentially represents an insertion unique to the SHRSP/BbbUtx sub-strain or an error in sequencing and/or alignment. Given that olfactory (and presumably taste) genes evolve with greater rapidity, the separation between the SHRSP/BbbUtx from SHRSP colonies the Glasgow strains were originally obtained from potentially underlies the variance in the *Tas2r120* gene.

Alignment to a closely related sub-strain provides an opportunity to determine genetic effects of specific phenotypes. The distribution of variance was mostly even across the SHRSP/Gla genome, with dense regions appearing on chromosomes 7, 9 and 15. Of the >45,000 variants found in the SHRSP/Gla genome compared to the SHRSP/BbbUtx, ∼7,000 genes containing variants were expressed in GD18.5 hearts, and <20,000 variants were called from the alignment of SHRSP/Gla GD18.5 short reads to the SHRSP/BbbUTx. That being said, only 120,000 of the 1.6 millions variants called in WGS, were found in RNA-sequencing data from the WKY/Gla. Thus, a greater proportion of variants discovered between SHRSP/Gla and the SHRSP/BbbUtx are contained within the transcribed regions of the genome. A high impact variant in the *Mlc1* gene on chromosome 7 was discovered in the SHRSP/Gla compared to the SHRSP/BbbUtx, and was localised within a genetic region containing QTLs associated with stroke and BP (Vedi et al., 2023). The *Mlc1* gene was not expressed in GD18.5 cardiac tissue. The SHRSP/Gla displays a particular sensitivity to stroke, particularly in response to salt or pressure loading. Although this is not experimentally verified, SHRSP/Gla strains cannot tolerate the salt loading published in other strains and SHRSP sub-strains. Whether the *Mlc1* or other SHRSP/Gla specific variants impact

stroke and salt sensitivity is as yet unknown.

Alignment of WKY/Gla and SHRSP/Gla produced a refined set of called variants which are likely to affect differences in phenotype and disease risk/protection. By overlaying cardiac mass QTLs and GD18.5 RNA sequencing data, genetic regions underlying genetic predisposition to left ventricular hypertrophy have been identified. These appear to be centered on gene clusters that are highly conversed between species and are good candidates for future translational studies. Chromosome 14 congenic strains have localised genetic variance that supports their congenic background and donor strains. Variant calling from RNA derived short read sequencing enabled some validation of variants called from DNA sequencing performed over 10 years ago. Detailed analysis of *Spp1* variants within introns, upstream, and downstream of transcriptional start sites will determine variants underlying the pathological overexpression associated with the SHRSP-genome.

# Chapter 9

# General Discussion

In hypertensive and non-hypertensive populations, development of excessively increased LVMI is associated with increased cardiovascular morbidity and mortality (Bella and Goring, 2012; Bombelli et al., 2023). Although progression to clinically defined LVH is influenced by a number of overlapping risk factors for hypertension and CVD, LVMI is under relatively high genetic control (Schmieder, 2010; Drazner, 2011). The importance of understanding the interplay between LVH and hypertension is of clinical relevance, especially given the potentially higher risk associated with more aggressive antihypertensive therapy in patients with LVH (Heimark et al., 2023). In the landscape of a shift toward more aggressive blood pressure lowering, being able to effectively identify and treat patients at high risk of LVH will become important.

Mirroring the genetic influence observed in human populations, a number of genetic rat models exist which develop phenotypes and diseases associated with human LVH (Jasinska-Stroschein, 2022). This includes the SHRSP, a sub-strain of which has been maintained at the University of Glasgow for over 30 years. What is interesting about the SHRSP versus other animal models of disease, is the onset of discernible differences in LVMI precedes the onset of measurably increased blood pressure. Thus, these animals represent a hypertensive patient population at a high genetic risk of LVH development. This contrasts the HHR, developed as a local inbred strain by Harrap et al. (2002), which is normotensive but experiences an increase in LVMI potentially more reflective of genetic cardiomyopathies (HCM or DCM phenotypes). Following the discovery of a QTL for LVMI on the rat chromosome 14 (J.S. Clark et al., 1996), two congenic strains were generated, whereby regions associated with high or low genetic risk of LVH were introgressed into the WKY and SHRSP backgrounds. In these congenic strains, the high risk SHRSP chromosome 14 functions on a normotensive WKY background and the low risk WKY chromosome 14 functions on a hypertensive SHRSP background. The four strains mimic more closely a human population of additive relative risk, which can be utilised to determine genetic influences on LVMI and implicate pathways which

can be targeted therapeutically.

Prior to this work, since the generation and establishment of the chromosome 14 congenic strains, only the male WKY.SPGla14a and SP.WKYGla14a had been characterised. Phenotype and genotype had been explored using gold-standard pressure-volume loop technology, as well as more standard ultrasound echocardiographic assessment. RNA microarray of the cardiac transcriptome at 3 time-points was also performed. These analyses implicated the chromosome 14 gene, *Spp1*, as a positional candidate gene in developed cardiac phenotypes. The upregulation of *Spp1* was associated with SHRSP chromosome 14 region. *Spp1* is part of the SIBLING genes family, which is highly conserved between human, mouse, and rat models (Bellahcene et al., 2008). Its localisation to the congenic region, and positional conservation suggested *Spp1* was a good functional candidate contributing to susceptibility to LVH. Herein, variant analysis demonstrated the congenic region on chromosome 14 was a region of high variance between the WKY/Gla and SHRSP strains. Functional studies using the H9c2 model (described in Chapter 5), demonstrated over-expression of *Spp1* was sufficient to cause an increase in cell size. In initial studies, the effect of *Spp1* on H9c2 cells was consistent irrespective of whether *Spp1* was isolated from the WKY/Gla or SHRSP/Gla transcriptome (data not published). This work is supported by the WGS analysis (Chapter 8), which showed the CDS of *Spp1* did not contain any variants predicted to alter *Spp1* function, despite a non-synonymous SNP.

Interestingly, a number of SNPs and InDels were found in splice regions, introns, and within ±5kb of *Spp1*, that could contribute to the increased *Spp1* expression associated with the SHRSP genome. Excess *Spp1* in H9c2 cells appeared to stimulate the production of pro-growth extracellular vesicles, which resulted in H9c2 cell growth when taken up by naïve cells. Control of *Spp1* expression in the heart, throughout the life-course, is dependent on *trans* and *cis*-acting factors including recruitment of transcription factors. In the rat, previous attempts to determine key functional variants affecting transcription were hindered in part by the quality of the reference genome. In 2020, the mRatBN7.2 was released and represented an increase in reference quality beyond previous iterations (de Jong et al., 2023). This was coupled with the publication of the SHRSP/BbbUtx genome in 2022, which was of equal quality to the mRatBN7.2 (Kalbfleisch et al., 2023), making it suitable for use as a reference genome in alignment and variant calling. Transcriptomics in GD18.5 hearts demonstrated foetal hearts have distinct transcriptional phenotypes, which separate strains containing SHRSP-genome from the WKY/Gla in cluster-based analysis. Novel to this work, foetal gene expression was found to be disrupted by the SHRSP chromosome 14 region following complete cardiac development *in-utero*. Alignment of whole genome and RNA sequencing to the genetically closer SHRSP/BbbUtx, implicated a number of genetic variants within genes that demonstrated differential regulation during cardiac

development. Variants in the *Spp1* promoter region, introns surrounding transcription start sites, or untranslated regions (UTR) were fine mapped between WKY/Gla and SHRSP sub-strains using two alignments to the mRatBN7.2 and SHRSP/BbbUtx reference quality genomes. In relation to cancer (Liang et al., 2019; Deepti et al., 2022), Duchenne Muscular Dystrophy (Pegoraro et al., 2011; Vianello et al., 2017), DCM (Barp et al., 2015), and hypertension (Nakayama et al., 2011; Y. Yang, Y. Wang, and P.J. Gao, 2020), variants in the human *Spp1* gene have been associated with severity of disease. In addition, a variant found further upstream (rs10011284) has been associated with serum levels of OPN protein (Y. Cheng et al., 2022b).

Research has suggested women are at greater risk from an increase in LVMI compared to men (Liao et al., 1995; Gerdts, Okin, et al., 2008; Geske et al., 2017; Gerdts, Izzo, et al., 2018), and experience greater CVD risk at lower blood pressures than men (Elfassy et al., 2023). Female and male rats demonstrate some gender differences in cardiac size and function (Boluyt et al., 2004; K.M. Rice et al., 2014), with SHRSP females demonstrating a lower rate of stroke and improved inflammatory profile compared to males (Olivera and Graham, 2023). Sex differences between Wistar-derived strains have been attributed, in part, to genetic effects of the Y-chromosome (A.O. Davidson et al., 1995). Work herein has fine-mapped functionally significant variants in comparison of the WKY/Gla to the SHRSP/BbbUtx reference genome, where the WKY/Gla contained a number of Y-chromosome variants that were distinct from the SHRSP/Gla alignment, representing candidates for determining sex-based differences in CVD development in a genetically influenced hypertensive model. That being said, the QTL on chromosome 14 was not strictly sexually dimorphic, despite greater significance of the association in males. Given that in humans, risk of developing LVH is greater in hypertensive women (Gerdts, Okin, et al., 2008; Gerdts, Izzo, et al., 2018; X. Wang, Hao, et al., 2022), investigation of the phenotype in the female congenic strains was important to determine. Chapter 3 showed female WKY, SHRSP, and chromosome 14 congenic strains had indices of LVM that were greater than males, despite a comparably reduced blood pressure in female animals. This was coupled with a greater relative increase in *Spp1* expression in female SHRSP and WKY.SPGla14a compared to the WKY than in the males. Work within this thesis supports a developed hypothesis of enhanced *Spp1* signalling in females underlying an increase in the risk of cardiac remodelling at lower blood pressures. Determining whether increases in *Spp1* expression occurs prior to remodelling as a stimulus, or following cardiac remodelling as a biomarker is still to be determined. Data from GD18.5 and neonate hearts suggest the increase in *Spp1* preceded changes in LVMI, however this data was not sex disaggregated and cannot determine whether sexual dimorphism in *Spp1* expression occurs prior to or following differences in cardiac mass.

Neither male nor female WKY.SPGla14a and SP.WKYGla14a congenic strains had

developed divergent cardiac phenotypes from their background strain at 16-weeks of age. The candidate *Spp1* gene was reliably increased in SHRSP and WKY.SPGla14a strains at all assessed time-points. Echocardiographic assessment undertaken did not allow diastolic function to be accurately assessed. Given that variants in *Spp1* have been associated with diastolic dysfunction, independent of LVH in human hypertensives (Nakayama et al., 2011), future investigations with the improved available technology should determine potential diastolic differences in chromosome 14 congenic and background strains. Assessment of diastolic function in female congenic strain would complement PV loop data generated during initial investigations in male chromosome 14 congenic strains (Douglas et al., 2010).

No phenotype differences were discernible between strains with *Spp1* upregulation and without at GD18.5 or neonatal time-points, which was expected from previous parental comparisons at this early time point. At 5-weeks, WKY.SPGla14a animals represented an intermediary phenotype between WKY and SHRSP strains. The transgenic mouse model of Renault et al. (2010) demonstrated increasing *Spp1* in cardiomyocytes did not cause distinguishable differences in cardiac phenotype within the first 2 weeks of life. However, cardiomyocyte specific *Spp1* over-expression led to premature death and symptoms consistent with DCM (Renault et al., 2010). The DCM-like cardiac phenotype was rescued in these animals if increased *Spp1* was reduced at 11-weeks of age (Renault et al., 2010). Given the number of variants associated with heritable cardiomyopathies, which are also associated with LVMI in the general population (Aung, Vargas, et al., 2019; Aung, Lopes, et al., 2023), the role of *Spp1* in genetic models is important to discern. The severe model of cardiac specific over-expression in mice is reduced in described chromosome 14 congenic models, where genetically programmed *Spp1* over-expression may not be limited to cardiomyocytes and results in effects that are not lethal. The H9c2 model, described in Chapter 5 and Chapter 6, proposes a direct function of *Spp1* in increasing cardiac cell size, potentially involving the secretion of *Spp1* in extracellular vesicles. Hypertensive patients produce pro-hypertrophic EVs (Fandl et al., 2023) and differentiating rat osteoclasts release EVs that contain *Spp1* protein. These support the developed hypothesis that one of the functional consequences of increased *Spp1* expression in cardiomyocytes is its increased trafficking in vesicles as part of cell-cell communication. Although preliminary, IHC in Chapter 4 indicated *Spp1* protein was localised both outwith and within proximity to markers of cardiomyocyte origin in SHRSP and WKY.SPGla14a hearts. Once snRNA-sequencing data becomes available, localisation of cells with increased *Spp1* will indicate whether cardiomyoctes are the site or destination of *Spp1* mRNA for translation into active protein.

In neonate and adult cardiac tissue, western blotting and IHC imaging could not determine a difference in OPN protein expression which mirrored differences in *Spp1*

mRNA. Assays of OPN protein following transfection of H9c2 cells with alternative *Spp1* transcripts did not support a requirement for *Spp1* protein to be increased, at least in intracellular compartments, for adverse effects on cell size to be determined. The mutant strain on the SHRSP background (SHRSP-*Spp1*$^{Em1Mcwi}$) did not confer an immediately obvious phenotype. However, in some cases, *Spp1* knock-out mice cannot not be distinguished from littermates until external pressures are applied (Liaw et al., 1998; Sam et al., 2004; Leen et al., 2008; Kramerova et al., 2019). Spp1$^{-/-}$ mice respond more favourably to aldosterone infusion (Sam et al., 2004), ANG-II infusion (A.R. Collins et al., 2004), and genetic perturbation of other disease causing genes (Mohamed, Gadeau, et al., 2015; Yousefi et al., 2019). In the described transgenic model, the SHRSP background was used as a genetic driver of LVH hypothesised to be reduced when *Spp1* expression was reduced by CRISPR/Cas9-mediated deletion mutation in exon 4 of the *Spp1* gene. Alternative transcription of a functionally active form of *Spp1*, analogous to the human OPN-c form, is potentially key to determine the protective effect of *Spp1* knock-out on the SHRSP background. Current estimation of differential transcript usage in GD18.5 hearts depended on the available annotation of the transcriptome. Construction of the transcription by *de-novo* transcript alignment suggested a high number of currently unannotated transcripts are expressed in the GD18.5 heart. DTU analyses were also more strain-specific than DGE analyses supporting highly state-specific control of transcript usage. Exploration of alternative *Spp1* transcripts in this data, should be combined with analysis of the transcripts produced by the SHRSP-*Spp1*$^{Em1Mcwi}$ strain. Different forms of OPN likely localise to different subcellular locations (Shinohara, H.J. Kim, et al., 2008a) and whilst H9c2 cell size is similarly affected by overexpression of both *Spp1* isoforms from the SHRSP-*Spp1*$^{Em1Mcwi}$, other known functions of *Spp1* and the ability of the protein to interact with its targets are as yet undetermined.

Reactivation of the foetal gene program is often associated with cardiac injury, and although its activation in acute settings is protective, the foetal gene program in the adult heart promotes excessive proliferation and cell migration which can be damaging (Fazilaty and Basler, 2023). RNA-sequencing in Chapter 7 indicated the SHRSP and WKY.SPGla14a have a foetal gene expression programme which is distinct from the WKY. IPA analysis determined the transcription factor MYC was associated with gene expression differences between the WKY and the SHRSP and WKY.SPGla14a strains. MYC is frequently upregulated in cancers as regulator of 'stem-ness' and immune suppression (Fazilaty and Basler, 2023). Genetic perturbation of MYC alters hypertrophic responses in the mouse heart (T. Jackson et al., 1990; Zhong et al., 2006) and the transcription factor is also upregulated in pressure-overload LVH, altering cardiac energy metabolism (Ahuja et al., 2010). Given that mitochondrial dysfunction is a key hallmark of pathological cardiac remodelling, determination of cause and effect might be an important consideration in the context herein. The GD18.5 transcrip-

tome highlighted cardiac gene expression profiles in the SHRSP and WKY.SPGla14a were consistent with mitochondrial dysregulation preceding changes in cardiomyocyte morphology. In future work, characterisation of mitochondrial function and energy metabolism following *Spp1* upregulation in the developed H9c2 model will determine the precise mechanisms by which *Spp1* causes increased risk in LVH development. Targeting of pathways in hypertensive patients at risk of developing LVH, will be the translational output of an improved understanding of how genes activated within a foetal gene program can be leveraged to reduce the occurrence of adverse cardiac remodelling.

LVMI and genetically driven cardiomyopathies share a significant number of variants in GWAS studies (Aung, Vargas, et al., 2019; Tadros et al., 2021). The genetic variance explored in Chapter 8 demonstrated high variance in the sodium voltage-gated ion channels expressed in the heart. These channels are part of an overlapping functional genomic cluster in the human, which have been associated with electrocardiographic traits and myotonic dystrophy (Freyermuth et al., 2016). Similarly to the ion channel genes, *Spp1* is contained within a cluster of genes that is conserved between species and represents a good translational target of preserved genetically influenced function. Analysis of long non-coding RNA support the importance of positional conservation and function, even if sequence identity is comparatively reduced (Amaral et al., 2018). Polygenetic phenotypes like hypertension, CVD, and LVH are likely to arise as a result of pleiotropic effects of key causal variants. Indeed, variants associated with LVMI are often also associated with blood pressure, liver and/or metabolic traits (Kanai et al., 2018). Although individual variants will likely not directly translate between humans and model organisms, the network functionality of genes affecting common phenotypes are likely the key to translating the findings in genetically controlled phenotypes (S.N. Wright et al., 2023).

## 9.1 Conclusions

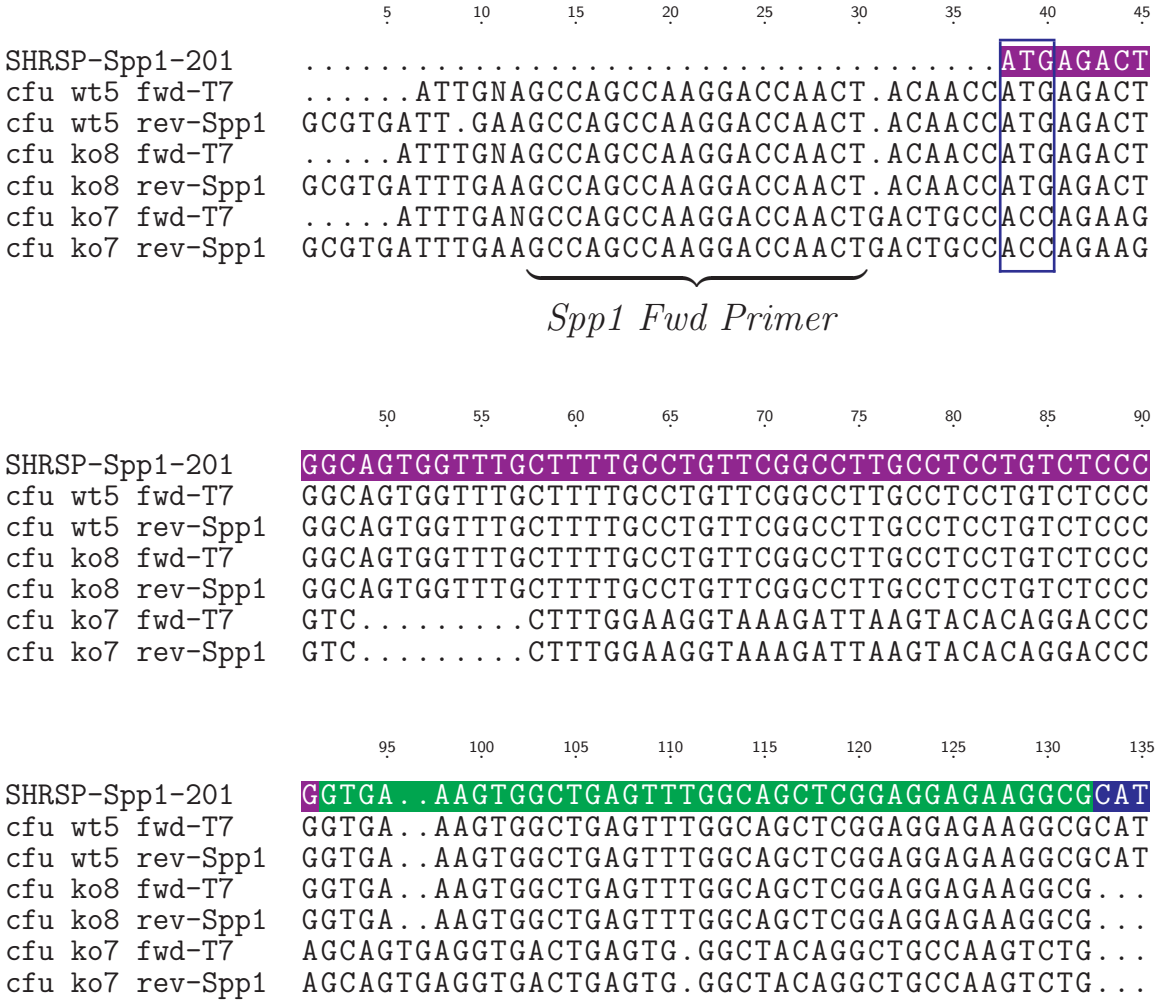
The characterisation of 5 rat strains demonstrated the complex association of genetic factors associated with LVMI and hypertension. The genetic risk conferred by the congenic region of chromosome 14 is supported by the dramatic effect on the transcriptome of the developing foetal heart. Supporting previous work, upregulation of *Spp1* is consistent in strains containing SHRSP genome at the chromosome 14 locus, which is sufficient in a cardiac cell model to cause cell elongation. Genetic variant analysis from the genome and transcriptome generated herein should look to prioritise variants and genes through convergence of the functionality of targets between rat and human data, making use of large consortia data including the UK BioBank and GETx. Integrated analysis of relevant pathways and functions represent a promising application of resources developed within this thesis. The experiments herein are a strong foundation

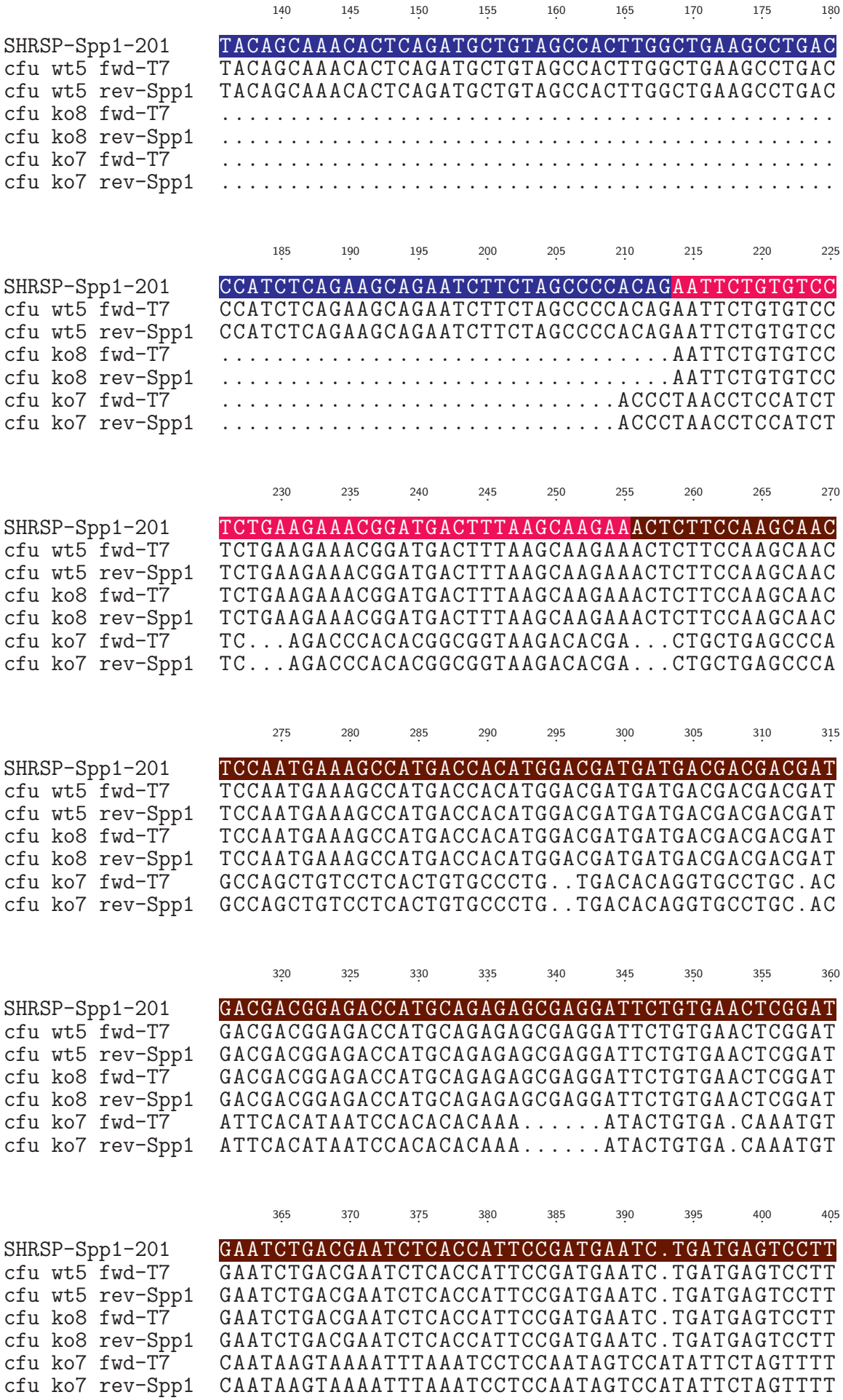
for future studies to further investigate; molecular pathway/s by which extracellular vesicles are produced following *Spp1* overexpression function in H9c2 and other relevant cell models, specific genetic variants resulting in *Spp1* overexpression in the described genetic models, and localise the origin of cardiac *Spp1* in healthy and pathologically remodelled myocardia using developed IHC and snRNA-seq technologies.

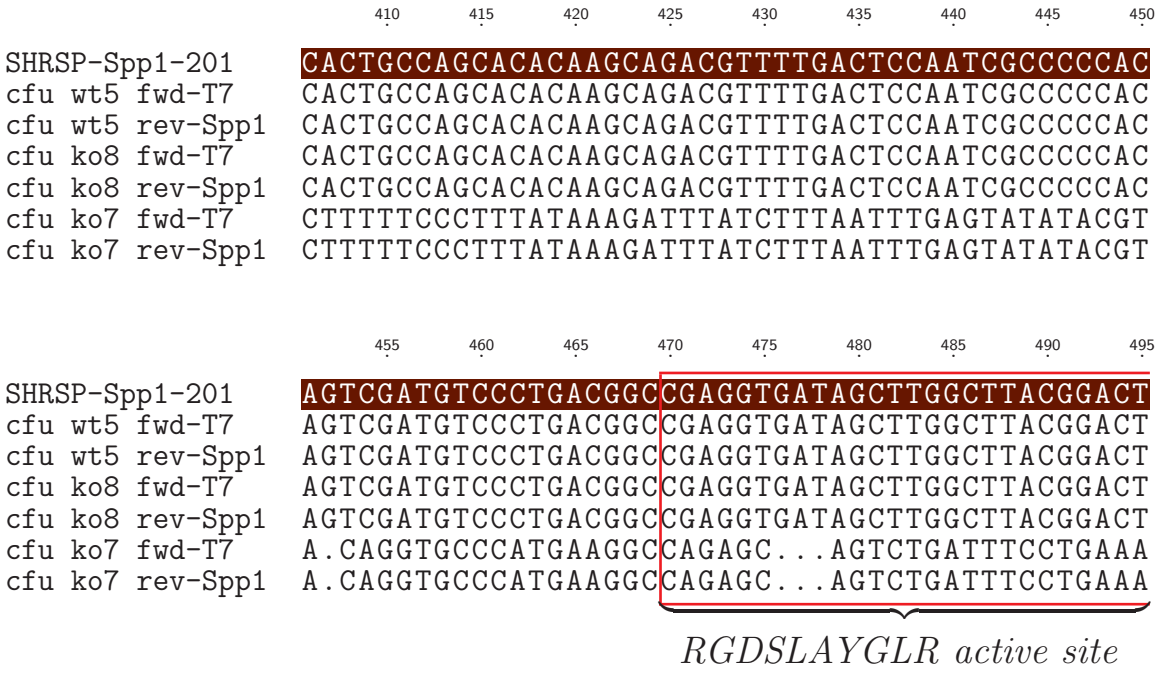
# Appendices

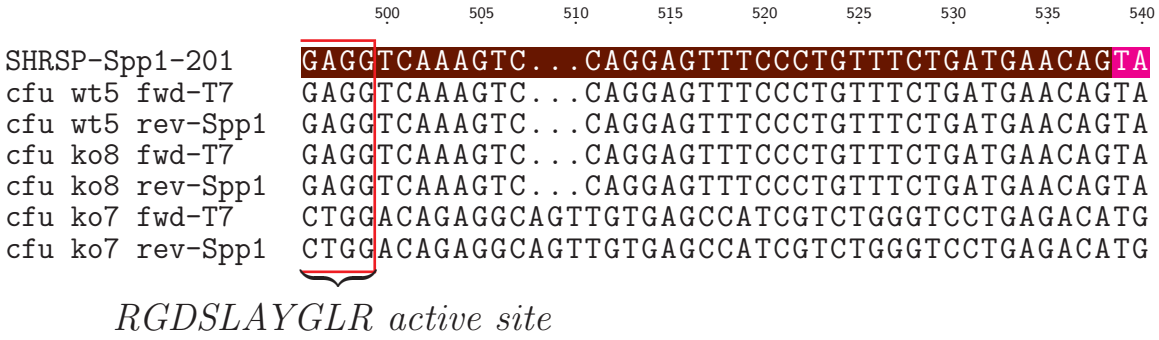
# Appendix A

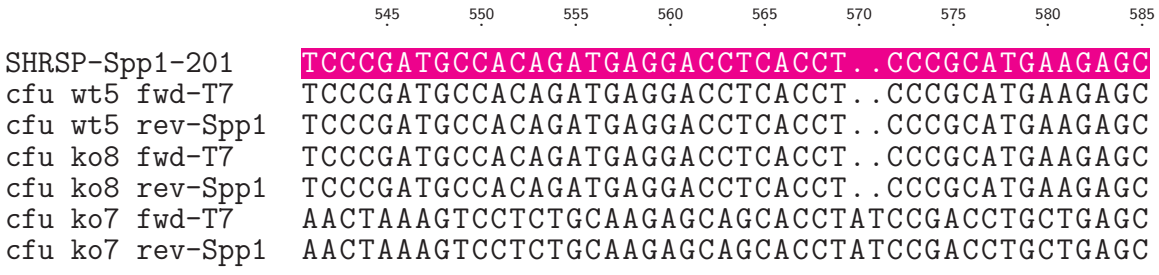
# Multiple Sequence Alignment

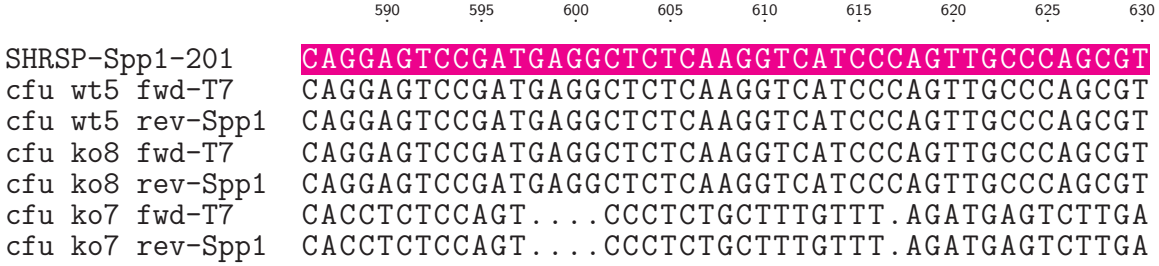


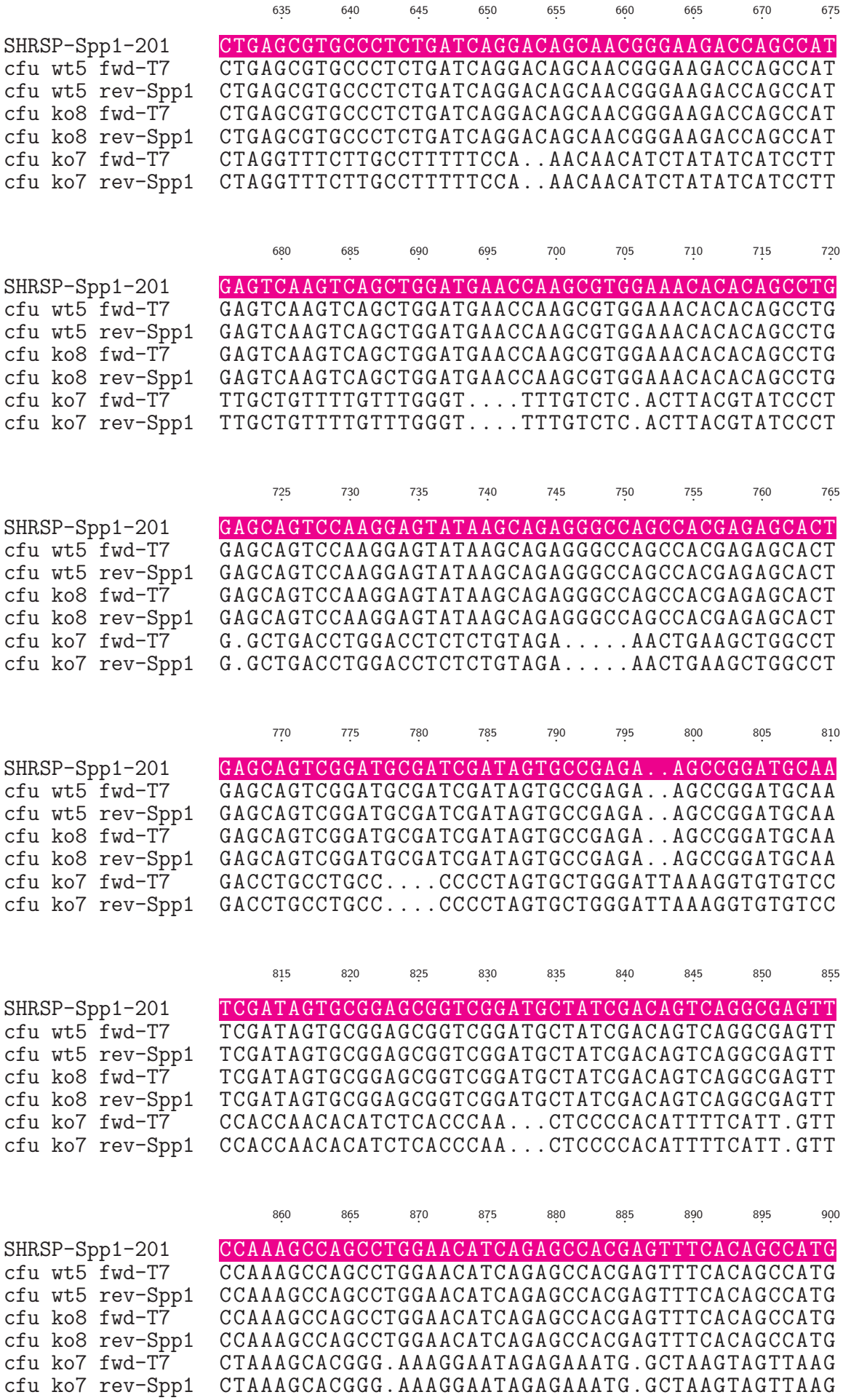

140 145 150 155 160 165 170 175 180
SHRSP-Spp1-201 TACAGCAAACACTCAGATGCTGTAGCCACTTGGCTGAAGCCTGAC
cfu wt5 fwd-T7 TACAGCAAACACTCAGATGCTGTAGCCACTTGGCTGAAGCCTGAC
cfu wt5 rev-Spp1 TACAGCAAACACTCAGATGCTGTAGCCACTTGGCTGAAGCCTGAC
cfu ko8 fwd-T7 .....
cfu ko8 rev-Spp1 .....
cfu ko7 fwd-T7 .....
cfu ko7 rev-Spp1 .....
185 190 195 200 205 210 215 220 225
SHRSP-Spp1-201 CCATCTCAGAAGCAGAATCTTCTAGCCCCACAGAATTCTGTGTCC
cfu wt5 fwd-T7 CCATCTCAGAAGCAGAATCTTCTAGCCCCACAGAATTCTGTGTCC
cfu wt5 rev-Spp1 CCATCTCAGAAGCAGAATCTTCTAGCCCCACAGAATTCTGTGTCC
cfu ko8 fwd-T7 .....AATTCTGTGTCC
cfu ko8 rev-Spp1 .....AATTCTGTGTCC
cfu ko7 fwd-T7 .....ACCCTAACCTCCATCT
cfu ko7 rev-Spp1 .....ACCCTAACCTCCATCT
230 235 240 245 250 255 260 265 270
SHRSP-Spp1-201 TCTGAAGAAACGGATGACTTTAAGCAAGAAACTCTTCCAAGCAAC
cfu wt5 fwd-T7 TCTGAAGAAACGGATGACTTTAAGCAAGAAACTCTTCCAAGCAAC
cfu wt5 rev-Spp1 TCTGAAGAAACGGATGACTTTAAGCAAGAAACTCTTCCAAGCAAC
cfu ko8 fwd-T7 TCTGAAGAAACGGATGACTTTAAGCAAGAAACTCTTCCAAGCAAC
cfu ko8 rev-Spp1 TCTGAAGAAACGGATGACTTTAAGCAAGAAACTCTTCCAAGCAAC
cfu ko7 fwd-T7 TC...AGACCCACACGGCGGTAAGACACGA...CTGCTGAGCCCA
cfu ko7 rev-Spp1 TC...AGACCCACACGGCGGTAAGACACGA...CTGCTGAGCCCA
275 280 285 290 295 300 305 310 315
SHRSP-Spp1-201 TCCAATGAAAGCCATGACCACATGGACGATGATGACGACGACGAT
cfu wt5 fwd-T7 TCCAATGAAAGCCATGACCACATGGACGATGATGACGACGACGAT
cfu wt5 rev-Spp1 TCCAATGAAAGCCATGACCACATGGACGATGATGACGACGACGAT
cfu ko8 fwd-T7 TCCAATGAAAGCCATGACCACATGGACGATGATGACGACGACGAT
cfu ko8 rev-Spp1 TCCAATGAAAGCCATGACCACATGGACGATGATGACGACGACGAT
cfu ko7 fwd-T7 GCCAGCTGTCCTCACTGTGCCCTG..TGACACAGGTGCCTGC.AC
cfu ko7 rev-Spp1 GCCAGCTGTCCTCACTGTGCCCTG..TGACACAGGTGCCTGC.AC
320 325 330 335 340 345 350 355 360
SHRSP-Spp1-201 GACGACGGAGACCATGCAGAGAGCGAGGATTCTGTGAACTCGGAT
cfu wt5 fwd-T7 GACGACGGAGACCATGCAGAGAGCGAGGATTCTGTGAACTCGGAT
cfu wt5 rev-Spp1 GACGACGGAGACCATGCAGAGAGCGAGGATTCTGTGAACTCGGAT
cfu ko8 fwd-T7 GACGACGGAGACCATGCAGAGAGCGAGGATTCTGTGAACTCGGAT
cfu ko8 rev-Spp1 GACGACGGAGACCATGCAGAGAGCGAGGATTCTGTGAACTCGGAT
cfu ko7 fwd-T7 ATTCACATAATCCACACACAAA......ATACTGTGA.CAAATGT
cfu ko7 rev-Spp1 ATTCACATAATCCACACACAAA......ATACTGTGA.CAAATGT
365 370 375 380 385 390 395 400 405
SHRSP-Spp1-201 GAATCTGACGAATCTCACCATTCCGATGAATC.TGATGAGTCCTT
cfu wt5 fwd-T7 GAATCTGACGAATCTCACCATTCCGATGAATC.TGATGAGTCCTT
cfu wt5 rev-Spp1 GAATCTGACGAATCTCACCATTCCGATGAATC.TGATGAGTCCTT
cfu ko8 fwd-T7 GAATCTGACGAATCTCACCATTCCGATGAATC.TGATGAGTCCTT
cfu ko8 rev-Spp1 GAATCTGACGAATCTCACCATTCCGATGAATC.TGATGAGTCCTT
cfu ko7 fwd-T7 CAATAAGTAAAATTTAAATCCTCCAATAGTCCATATTCTAGTTTT
cfu ko7 rev-Spp1 CAATAAGTAAAATTTAAATCCTCCAATAGTCCATATTCTAGTTTT


410 415 420 425 430 435 440 445 450
SHRSP-Spp1-201 CACTGCCAGCACACAAGCAGACGTTTTGACTCCAATCGCCCCCAC
cfu wt5 fwd-T7 CACTGCCAGCACACAAGCAGACGTTTTGACTCCAATCGCCCCCAC
cfu wt5 rev-Spp1 CACTGCCAGCACACAAGCAGACGTTTTGACTCCAATCGCCCCCAC
cfu ko8 fwd-T7 CACTGCCAGCACACAAGCAGACGTTTTGACTCCAATCGCCCCCAC
cfu ko8 rev-Spp1 CACTGCCAGCACACAAGCAGACGTTTTGACTCCAATCGCCCCCAC
cfu ko7 fwd-T7 CTTTTTCCCTTTATAAAGATTTATCTTTAATTTGAGTATATACGT
cfu ko7 rev-Spp1 CTTTTTCCCTTTATAAAGATTTATCTTTAATTTGAGTATATACGT
455 460 465 470 475 480 485 490 495
SHRSP-Spp1-201 AGTCGATGTCCCTGACGGCCGAGGTGATAGCTTGGCTTACGGACT
cfu wt5 fwd-T7 AGTCGATGTCCCTGACGGCCGAGGTGATAGCTTGGCTTACGGACT
cfu wt5 rev-Spp1 AGTCGATGTCCCTGACGGCCGAGGTGATAGCTTGGCTTACGGACT
cfu ko8 fwd-T7 AGTCGATGTCCCTGACGGCCGAGGTGATAGCTTGGCTTACGGACT
cfu ko8 rev-Spp1 AGTCGATGTCCCTGACGGCCGAGGTGATAGCTTGGCTTACGGACT
cfu ko7 fwd-T7 A.CAGGTGCCCATGAAGGCCAGAGC...AGTCTGATTTCCTGAAA
cfu ko7 rev-Spp1 A.CAGGTGCCCATGAAGGCCAGAGC...AGTCTGATTTCCTGAAA
RGDSLAYGLR active site


500 505 510 515 520 525 530 535 540
SHRSP-Spp1-201 GAGGTCAAAGTC...CAGGAGTTTCCCTGTTTCTGATGAACAGTA
cfu wt5 fwd-T7 GAGGTCAAAGTC...CAGGAGTTTCCCTGTTTCTGATGAACAGTA
cfu wt5 rev-Spp1 GAGGTCAAAGTC...CAGGAGTTTCCCTGTTTCTGATGAACAGTA
cfu ko8 fwd-T7 GAGGTCAAAGTC...CAGGAGTTTCCCTGTTTCTGATGAACAGTA
cfu ko8 rev-Spp1 GAGGTCAAAGTC...CAGGAGTTTCCCTGTTTCTGATGAACAGTA
cfu ko7 fwd-T7 CTGGACAGAGGCAGTTGTGAGCCATCGTCTGGGTCCTGAGACATG
cfu ko7 rev-Spp1 CTGGACAGAGGCAGTTGTGAGCCATCGTCTGGGTCCTGAGACATG
RGDSLAYGLR active site


545 550 555 560 565 570 575 580 585
SHRSP-Spp1-201 TCCCGATGCCACAGATGAGGACCTCACCT..CCCGCATGAAGAGC
cfu wt5 fwd-T7 TCCCGATGCCACAGATGAGGACCTCACCT..CCCGCATGAAGAGC
cfu wt5 rev-Spp1 TCCCGATGCCACAGATGAGGACCTCACCT..CCCGCATGAAGAGC
cfu ko8 fwd-T7 TCCCGATGCCACAGATGAGGACCTCACCT..CCCGCATGAAGAGC
cfu ko8 rev-Spp1 TCCCGATGCCACAGATGAGGACCTCACCT..CCCGCATGAAGAGC
cfu ko7 fwd-T7 AACTAAAGTCCTCTGCAAGAGCAGCACCTATCCGACCTGCTGAGC
cfu ko7 rev-Spp1 AACTAAAGTCCTCTGCAAGAGCAGCACCTATCCGACCTGCTGAGC


590 595 600 605 610 615 620 625 630
SHRSP-Spp1-201 CAGGAGTCCGATGAGGCTCTCAAGGTCATCCCAGTTGCCCAGCGT
cfu wt5 fwd-T7 CAGGAGTCCGATGAGGCTCTCAAGGTCATCCCAGTTGCCCAGCGT
cfu wt5 rev-Spp1 CAGGAGTCCGATGAGGCTCTCAAGGTCATCCCAGTTGCCCAGCGT
cfu ko8 fwd-T7 CAGGAGTCCGATGAGGCTCTCAAGGTCATCCCAGTTGCCCAGCGT
cfu ko8 rev-Spp1 CAGGAGTCCGATGAGGCTCTCAAGGTCATCCCAGTTGCCCAGCGT
cfu ko7 fwd-T7 CACCTCTCCAGT....CCCTCTGCTTTGTTT.AGATGAGTCTTGA
cfu ko7 rev-Spp1 CACCTCTCCAGT....CCCTCTGCTTTGTTT.AGATGAGTCTTGA


635 640 645 650 655 660 665 670 675
SHRSP-Spp1-201 CTGAGCGTGCCCTCTGATCAGGACAGCAACGGGAAGACCAGCCAT
cfu wt5 fwd-T7 CTGAGCGTGCCCTCTGATCAGGACAGCAACGGGAAGACCAGCCAT
cfu wt5 rev-Spp1 CTGAGCGTGCCCTCTGATCAGGACAGCAACGGGAAGACCAGCCAT
cfu ko8 fwd-T7 CTGAGCGTGCCCTCTGATCAGGACAGCAACGGGAAGACCAGCCAT
cfu ko8 rev-Spp1 CTGAGCGTGCCCTCTGATCAGGACAGCAACGGGAAGACCAGCCAT
cfu ko7 fwd-T7 CTAGGTTTCTTGCCTTTTTCCA..AACAACATCTATATCATCCTT
cfu ko7 rev-Spp1 CTAGGTTTCTTGCCTTTTTCCA..AACAACATCTATATCATCCTT
680 685 690 695 700 705 710 715 720
SHRSP-Spp1-201 GAGTCAAGTCAGCTGGATGAACCAAGCGTGGAAACACACAGCCTG
cfu wt5 fwd-T7 GAGTCAAGTCAGCTGGATGAACCAAGCGTGGAAACACACAGCCTG
cfu wt5 rev-Spp1 GAGTCAAGTCAGCTGGATGAACCAAGCGTGGAAACACACAGCCTG
cfu ko8 fwd-T7 GAGTCAAGTCAGCTGGATGAACCAAGCGTGGAAACACACAGCCTG
cfu ko8 rev-Spp1 GAGTCAAGTCAGCTGGATGAACCAAGCGTGGAAACACACAGCCTG
cfu ko7 fwd-T7 TTGCTGTTTTGTTTGGGT....TTTGTCTC.ACTTACGTATCCCT
cfu ko7 rev-Spp1 TTGCTGTTTTGTTTGGGT....TTTGTCTC.ACTTACGTATCCCT
725 730 735 740 745 750 755 760 765
SHRSP-Spp1-201 GAGCAGTCCAAGGAGTATAAGCAGAGGGCCAGCCACGAGAGCACT
cfu wt5 fwd-T7 GAGCAGTCCAAGGAGTATAAGCAGAGGGCCAGCCACGAGAGCACT
cfu wt5 rev-Spp1 GAGCAGTCCAAGGAGTATAAGCAGAGGGCCAGCCACGAGAGCACT
cfu ko8 fwd-T7 GAGCAGTCCAAGGAGTATAAGCAGAGGGCCAGCCACGAGAGCACT
cfu ko8 rev-Spp1 GAGCAGTCCAAGGAGTATAAGCAGAGGGCCAGCCACGAGAGCACT
cfu ko7 fwd-T7 G.GCTGACCTGGACCTCTCTGTAGA.....AACTGAAGCTGGCCT
cfu ko7 rev-Spp1 G.GCTGACCTGGACCTCTCTGTAGA.....AACTGAAGCTGGCCT
770 775 780 785 790 795 800 805 810
SHRSP-Spp1-201 GAGCAGTCGGATGCGATCGATAGTGCCGAGA..AGCCGGATGCAA
cfu wt5 fwd-T7 GAGCAGTCGGATGCGATCGATAGTGCCGAGA..AGCCGGATGCAA
cfu wt5 rev-Spp1 GAGCAGTCGGATGCGATCGATAGTGCCGAGA..AGCCGGATGCAA
cfu ko8 fwd-T7 GAGCAGTCGGATGCGATCGATAGTGCCGAGA..AGCCGGATGCAA
cfu ko8 rev-Spp1 GAGCAGTCGGATGCGATCGATAGTGCCGAGA..AGCCGGATGCAA
cfu ko7 fwd-T7 GACCTGCCTGCC....CCCCTAGTGCTGGGATTAAAGGTGTGTCC
cfu ko7 rev-Spp1 GACCTGCCTGCC....CCCCTAGTGCTGGGATTAAAGGTGTGTCC
815 820 825 830 835 840 845 850 855
SHRSP-Spp1-201 TCGATAGTGCGGAGCGGTCGGATGCTATCGACAGTCAGGCGAGTT
cfu wt5 fwd-T7 TCGATAGTGCGGAGCGGTCGGATGCTATCGACAGTCAGGCGAGTT
cfu wt5 rev-Spp1 TCGATAGTGCGGAGCGGTCGGATGCTATCGACAGTCAGGCGAGTT
cfu ko8 fwd-T7 TCGATAGTGCGGAGCGGTCGGATGCTATCGACAGTCAGGCGAGTT
cfu ko8 rev-Spp1 TCGATAGTGCGGAGCGGTCGGATGCTATCGACAGTCAGGCGAGTT
cfu ko7 fwd-T7 CCACCAACACATCTCACCCAA...CTCCCCACATTTTCATT.GTT
cfu ko7 rev-Spp1 CCACCAACACATCTCACCCAA...CTCCCCACATTTTCATT.GTT
860 865 870 875 880 885 890 895 900
SHRSP-Spp1-201 CCAAAGCCAGCCTGGAACATCAGAGCCACGAGTTTCACAGCCATG
cfu wt5 fwd-T7 CCAAAGCCAGCCTGGAACATCAGAGCCACGAGTTTCACAGCCATG
cfu wt5 rev-Spp1 CCAAAGCCAGCCTGGAACATCAGAGCCACGAGTTTCACAGCCATG
cfu ko8 fwd-T7 CCAAAGCCAGCCTGGAACATCAGAGCCACGAGTTTCACAGCCATG
cfu ko8 rev-Spp1 CCAAAGCCAGCCTGGAACATCAGAGCCACGAGTTTCACAGCCATG
cfu ko7 fwd-T7 CTAAAGCACGGG.AAAGGAATAGAGAAATG.GCTAAGTAGTTAAG
cfu ko7 rev-Spp1 CTAAAGCACGGG.AAAGGAATAGAGAAATG.GCTAAGTAGTTAAG

```
                          905       910       915       920       925       930       935       940       945
SHRSP-Spp1-201            AGGACAAGCTAGTCCTAG.ACCCTA.AGAGTAAGGAAGATGATAG
cfu wt5 fwd-T7            AGGACAAGCTAGTCCTAG.ACCCTA.AGAGTAAGGAAGATGATAG
cfu wt5 rev-Spp1          AGGACAAGCTAGTCCTAG.ACCCTA.AGAGTAAGGAAGATGATAG
cfu ko8 fwd-T7            AGGACAAGCTAGTCCTAG.ACCCTA.AGAGTAAGGAAGATGATAG
cfu ko8 rev-Spp1          AGGACAAGCTAGTCCTAG.ACCCTA.AGAGTAAGGAAGATGATAG
cfu ko7 fwd-T7            AACGCT.GTCGATCCTAGCACTCCACAAAGTGATTAACAAG.TCT
cfu ko7 rev-Spp1          AACGCT.GTCGATCCTAGCACTCCACAAAGTGATTAACAAG.TCT

                          950       955       960       965       970       975       980       985       990
SHRSP-Spp1-201            GTATCTGAAATTCCGCATTTCTCATGAATTAGAGAGTTCATCTTC
cfu wt5 fwd-T7            GTATCTGAAATTCCGCATTTCTCATGAATTNNNNNGTTCATCTTC
cfu wt5 rev-Spp1          GTATCTGAANT.CCGCATTTCT.....................
cfu ko8 fwd-T7            GTATCTGAAATTCCGCATTTCTCATGAATTAGAGAGTTCATCTTC
cfu ko8 rev-Spp1          GTATCTGANNN.CCGCATTTCT.....................
cfu ko7 fwd-T7            GTAACTCCAGTTTCACAGTACCCACCCTCTTCTGGTCTCTGTGGC
cfu ko7 rev-Spp1          GTAACTCCAGTTTCACAGNANCCACCCTCTTCTGG.........

                          995       1000      1005      1010      1015      1020
SHRSP-Spp1-201            TGAGGTCAATTAA..................
cfu wt5 fwd-T7            TGAGGTCA.TTAANNANNNNNNN........
cfu wt5 rev-Spp1          ..........................................
cfu ko8 fwd-T7            TGAGGTCAATTAANGNNNNNNNNNN......
cfu ko8 rev-Spp1          ..........................................
cfu ko7 fwd-T7            ACAAACATGGGGCACAGATACATATGGNNNNN
cfu ko7 rev-Spp1          ..........................................
```

# Appendix B

# QuPath Scripting

Tissue Detector was built using GUI and .json report is included below.


```
            {
"pixel_classifier_type": "OpenCVPixelClassifier",
"metadata": {
    "inputPadding": 0,
    "inputResolution": {
    "pixelWidth": {
        "value": 3.535755325731459,
        "unit": "µm"
    },
    "pixelHeight": {
        "value": 3.535755325731459,
        "unit": "µm"
    },
    "zSpacing": {
        "value": 1.0,
        "unit": "z-slice"
    },
    "timeUnit": "SECONDS",
    "timepoints": []
    },
    "inputWidth": 512,
    "inputHeight": 512,
    "inputNumChannels": 3,
    "outputType": "CLASSIFICATION",
    "outputChannels": [],
    "classificationLabels": {
```

```
    "0": "Ignore*",
    "1": "Region*"
    }
},
"op": {
    "type": "data.op.channels",
    "colorTransforms": [
    {
        "combineType": "MEAN"
    }
    ],
    "op": {
    "type": "op.core.sequential",
    "ops": [
        {
        "type": "op.filters.gaussian",
        "sigmaX": 3.0,
        "sigmaY": 3.0
        },
        {
        "type": "op.threshold.constant",
        "thresholds": [
            12.0
        ]
        }
    ]
    }
}
}
```


Multiplex classifier 'multiplex_test' was built using GUI and .json report is included below.


```
            {
"object_classifier_type": "CompositeClassifier",
"classifiers": [
    {
    "object_classifier_type": "SimpleClassifier",
    "function": {
        "classifier_fun": "ClassifyByMeasurementFunction",
```

```
        "measurement": "ROI: 0.50 µm per pixel: CY7-VIM: Mean",
        "pathClassEquals": "CY7-VIM",
        "pathClassAbove": "CY7-VIM",
        "threshold": 5.0
    },
    "pathClasses": [
        "CY7-VIM"
    ],
    "filter": "DETECTIONS_ALL",
    "timestamp": 1702051724328
    },
    {
    "object_classifier_type": "SimpleClassifier",
    "function": {
        "classifier_fun": "ClassifyByMeasurementFunction",
        "measurement": "ROI: 0.50 µm per pixel: CFP-aSMA: Mean",
        "pathClassEquals": "CFP-aSMA",
        "pathClassAbove": "CFP-aSMA",
        "threshold": 30.0
    },
    "pathClasses": [
        "CFP-aSMA"
    ],
    "filter": "DETECTIONS_ALL",
    "timestamp": 1702051750971
    },
    {
    "object_classifier_type": "SimpleClassifier",
    "function": {
        "classifier_fun": "ClassifyByMeasurementFunction",
        "measurement": "ROI: 0.50 µm per pixel: TRITC-My: Mean",
        "pathClassEquals": "TRITC-My",
        "pathClassAbove": "TRITC-My",
        "threshold": 40.0
    },
    "pathClasses": [
        "TRITC-My"
    ],
    "filter": "DETECTIONS_ALL",
    "timestamp": 1702051573016
    },
```

```
    {
    "object_classifier_type": "SimpleClassifier",
    "function": {
        "classifier_fun": "ClassifyByMeasurementFunction",
        "measurement": "ROI: 0.50 µm per pixel: CY5-OPN: Mean",
        "pathClassEquals": "CY5-OPN",
        "pathClassAbove": "CY5-OPN",
        "threshold": 50.0
    },
    "pathClasses": [
        "CY5-OPN"
    ],
    "filter": "DETECTIONS_ALL",
    "timestamp": 1702051588715
    }
]
}
```


User generated '.groovy' script ran across all images used in analysis. Functionality is described above and in QuPath documentation (Bankhead et al., 2017). Singleplex Analysis:

```
setImageType('FLUORESCENCE');
setChannelNames(
    'DAPI',
    'OPN-FITC'
)

createAnnotationsFromPixelClassifier("TissueDectector",
                                     0.0, 0.0,
                                     "INCLUDE_IGNORED")

selectAnnotations();
addPixelClassifierMeasurements("TissueDectector",
                                  "TissueDetector")

def originalAnnotations = getAnnotationObjects()

selectAnnotations();
runPlugin('qupath.lib.plugins.objects.DilateAnnotationPlugin',
```

```
        '{"radiusMicrons":-20.0,
        "lineCap":"ROUND",
        "removeInterior":false,
        "constrainToParent":true}')

removeObjects(originalAnnotations, true)

selectAnnotations();
runPlugin('qupath.lib.plugins.objects.DilateAnnotationPlugin',
        '{"radiusMicrons":-20.0,
            "lineCap":"ROUND",
            "removeInterior":false,
            "constrainToParent":true}')

removeObjects(originalAnnotations, true)

selectObjectsByClassification("Region*");
runPlugin('qupath.lib.algorithms.TilerPlugin',
        '{"tileSizeMicrons":10.0,
            "trimToROI":true,"makeAnnotations":false,
            "removeParentAnnotation":false}')

selectDetections();
runPlugin('qupath.lib.algorithms.IntensityFeaturesPlugin',
          '{"pixelSizeMicrons":0.5,
            "region":"ROI",
            "tileSizeMicrons":25.0,
            "channel1":true,
            "channel2":true,
            "doMean":true,
            "doStdDev":true,
            "doMinMax":true,
            "doMedian":true}')

selectDetections();
addShapeMeasurements("AREA")
```

Multiplex analysis:

```
setImageType('FLUORESCENCE');
```

```
setChannelNames('DAPI',
                'TRITC-My',
                'CY5-OPN',
                'CY7-VIM',
                'CFP-aSMA'
            )
def tiles = getTileObjects()
removeObjects(tiles, true)
def annotations = getAnnotationObjects()
removeObjects(annotations, true)
def detections = getDetectionObjects()
removeObjects(detections, true)

createAnnotationsFromPixelClassifier("TissueDetector",
                                     0.0,
                                     0.0,
                                     "INCLUDE_IGNORED")

selectAnnotations();
addPixelClassifierMeasurements("TissueDetector",
                               "TissueDetector")

def originalAnnotations = getAnnotationObjects()

selectAnnotations();
runPlugin('qupath.lib.plugins.objects.DilateAnnotationPlugin',
          '{"radiusMicrons":-20.0,
            "lineCap":"ROUND",
            "removeInterior":false,
            "constrainToParent":true}')

removeObjects(originalAnnotations, true)

selectObjectsByClassification("Region*");
runPlugin('qupath.lib.algorithms.TilerPlugin',
          '{"tileSizeMicrons":10.0,"trimToROI":true,
            "makeAnnotations":false,
            "removeParentAnnotation":false}')

selectDetections();
runPlugin('qupath.lib.algorithms.IntensityFeaturesPlugin',
```

```
        '{"pixelSizeMicrons":0.5,
          "region":"ROI",
          "tileSizeMicrons":25.0,
          "channel1":true,
          "channel2":true,
          "channel3":true,
          "channel4":true,
          "channel5":true,
          "doMean":true,
          "doStdDev":false,
          "doMinMax":false,
          "doMedian":false,
          "doHaralick":false,
          "haralickMin":NaN,
          "haralickMax":NaN,
          "haralickDistance":1,
          "haralickBins":32}')

selectDetections();
runObjectClassifier("multiplex_test")
```

# List of References

Abdulrahman, N., Jaspard-Vinassa, B., Fliegel, L., Jabeen, A., Riaz, S., Gadeau, A.P., and Mraiche, F., 2018. Na(+)/h(+) exchanger isoform 1-induced osteopontin expression facilitates cardiac hypertrophy through p90 ribosomal s6 kinase. *Physiol genomics* [Online], 50(5), pp.332–342. Available from: https://doi.org/10.1152/physiolgenomics.00133.2017.

Aghagolzadeh, P., Plaisance, I., Bernasconi, R., Treibel, T.A., Pulido Quetglas, C., Wyss, T., Wigger, L., Nemir, M., Sarre, A., Chouvardas, P., Johnson, R., Gonzalez, A., and Pedrazzini, T., 2023. Assessment of the cardiac noncoding transcriptome by single-cell rna sequencing identifies fixer, a conserved profibrogenic long noncoding rna. *Circulation* [Online], 148(9), pp.778–797. Available from: https://doi.org/10.1161/CIRCULATIONAHA.122.062601.

Ahuja, P., Zhao, P., Angelis, E., Ruan, H., Korge, P., Olson, A., Wang, Y., Jin, E.S., Jeffrey, F.M., Portman, M., and Maclellan, W.R., 2010. Myc controls transcriptional regulation of cardiac metabolism and mitochondrial biogenesis in response to pathological stress in mice. *J clin invest* [Online], 120(5), pp.1494–505. Available from: https://doi.org/10.1172/JCI38331.

Amaral, P.P., Leonardi, T., Han, N., Vire, E., Gascoigne, D.K., Arias-Carrasco, R., Buscher, M., Pandolfini, L., Zhang, A., Pluchino, S., Maracaja-Coutinho, V., Nakaya, H.I., Hemberg, M., Shiekhattar, R., Enright, A.J., and Kouzarides, T., 2018. Genomic positional conservation identifies topological anchor point rnas linked to developmental loci. *Genome biol* [Online], 19(1), p.32. Available from: https://doi.org/10.1186/s13059-018-1405-5.

Amarasinghe, S.L., Su, S., Dong, X., Zappia, L., Ritchie, M.E., and Gouil, Q., 2020. Opportunities and challenges in long-read sequencing data analysis. *Genome biol* [Online], 21(1), p.30. Available from: https://doi.org/10.1186/s13059-020-1935-5.

Ames, E.G., Lawson, M.J., Mackey, A.J., and Holmes, J.W., 2013. Sequencing of mrna identifies re-expression of fetal splice variants in cardiac hypertrophy. *J mol cell cardiol* [Online], 62, pp.99–107. Available from: https://doi.org/10.1016/j.yjmcc.2013.05.004.

Anders, S., Reyes, A., and Huber, W., 2012. Detecting differential usage of exons from rna-seq data. *Genome res* [Online], 22(10), pp.2008–17. Available from: https://doi.org/10.1101/gr.133744.111.

Anders, S., Pyl, P.T., and Huber, W., 2015. Htseq—a python framework to work with high-throughput sequencing data. *Bioinformatics* [Online], 31 (2), pp.166–169. Available from: https://doi.org/10.1093/bioinformatics/btu638.

Anto Michel, N., Ljubojevic-Holzer, S., Bugger, H., and Zirlik, A., 2022. Cellular heterogeneity of the heart. *Front cardiovasc med* [Online], 9, p.868466. Available from: https://doi.org/10.3389/fcvm.2022.868466.

Arnett, D.K., Fuentes, L. de las, and Broeckel, U., 2004. Genes for left ventricular hypertrophy. *Curr hypertens rep* [Online], 6(1), pp.36–41. Available from: https://doi.org/10.1007/s11906-004-0009-5.

Arnett, D.K., Li, N., Tang, W., Rao, D.C., Devereux, R.B., Claas, S.A., Kraemer, R., and Broeckel, U., 2009. Genome-wide association study identifies single-nucleotide polymorphism in kcnb1 associated with left ventricular mass in humans: the hypergen study. *Bmc med genet* [Online], 10, p.43. Available from: https://doi.org/10.1186/1471-2350-10-43.

Ashizawa, N., Graf, K., Do, Y.S., Nunohiro, T., Giachelli, C.M., Meehan, W.P., Tuan, T.L., and Hsueh, W.A., 1996. Osteopontin is produced by rat cardiac fibroblasts and mediates a(ii)-induced dna synthesis and collagen gel contraction. *J clin invest* [Online], 98(10), pp.2218–27. Available from: https://doi.org/10.1172/JCI119031.

Asirvatham, A., 2022. *Investigating the role of osteopontin in cardiac hypertrophy* [Online]. Thesis. University of Glasgow. Available from: https://doi.org/https://theses.gla.ac.uk/83377/.

Atanur, S.S., Diaz, A.G., Maratou, K., Sarkis, A., Rotival, M., Game, L., Tschannen, M.R., Kaisaki, P.J., Otto, G.W., Ma, M.C., Keane, T.M., Hummel, O., Saar, K., Chen, W., Guryev, V., Gopalakrishnan, K., Garrett, M.R., Joe, B., Citterio, L., Bianchi, G., McBride, M., Dominiczak, A., Adams, D.J., Serikawa, T., Flicek, P., Cuppen, E., Hubner, N., Petretto, E., Gauguier, D., Kwitek, A., Jacob, H., and Aitman, T.J., 2013. Genome sequencing reveals loci under artificial selection that underlie disease phenotypes in the laboratory rat. *Cell* [Online], 154(3), pp.691–703. Available from: https://doi.org/10.1016/j.cell.2013.06.040.

Aung, N., Lopes, L.R., Duijvenboden, S. van, Harper, A.R., Goel, A., Grace, C., Ho, C.Y., Weintraub, W.S., Kramer, C.M., Neubauer, S., Watkins, H.C., Petersen, S.E., and Munroe, P.B., 2023. Genome-wide analysis of left ventricular maximum wall thickness in the uk biobank cohort reveals a shared genetic background with hypertrophic cardiomyopathy. *Circ genom precis med* [Online], 16(1), e003716. Available from: https://doi.org/10.1161/CIRCGEN.122.003716.

Aung, N., Vargas, J.D., Yang, C., Cabrera, C.P., Warren, H.R., Fung, K., Tzanis, E., Barnes, M.R., Rotter, J.I., Taylor, K.D., Manichaikul, A.W., Lima, J.A.C., Bluemke, D.A., Piechnik, S.K., Neubauer, S., Munroe, P.B., and Petersen, S.E., 2019. Genome-wide analysis of left ventricular image-derived phenotypes identifies fourteen loci associated with cardiac morphogenesis and heart failure development. *Circulation* [Online], 140(16), pp.1318–1330. Available from: https://doi.org/10.1161/CIRCULATIONAHA.119.041161.

Backman, J.D., Li, A.H., Marcketta, A., Sun, D., Mbatchou, J., Kessler, M.D., Benner, C., Liu, D., Locke, A.E., Balasubramanian, S., Yadav, A., Banerjee, N., Gillies, C.E., Damask, A., Liu, S., Bai, X., Hawes, A., Maxwell, E., Gurski, L., Watanabe, K., Kosmicki, J.A., Rajagopal, V., Mighty, J., Regeneron Genetics, C., DiscovEhr, Jones, M., Mitnaul, L., Stahl, E., Coppola, G., Jorgenson, E., Habegger, L., Salerno, W.J., Shuldiner, A.R., Lotta, L.A., Overton, J.D., Cantor, M.N., Reid, J.G., Yancopoulos, G., Kang, H.M., Marchini, J., Baras, A., Abecasis, G.R., and Ferreira, M.A.R., 2021. Exome sequencing and analysis of 454,787 uk biobank participants. *Nature* [Online], 599(7886), pp.628–634. Available from: https://doi.org/10.1038/s41586-021-04103-z.

Bairey Merz, C.N., Nelson, M.D., Cheng, S., and Wei, J., 2020. Sex differences and the left ventricle: morphology matters. *Eur heart j cardiovasc imaging* [Online], 21(9), pp.991–993. Available from: https://doi.org/10.1093/ehjci/jeaa195.

Bairey Merz, C.N., Pepine, C.J., Walsh, M.N., and Fleg, J.L., 2017. Ischemia and no obstructive coronary artery disease (inoca): developing evidence-based therapies and research agenda for the next decade. *Circulation* [Online], 135(11), pp.1075–1092. Available from: https://doi.org/10.1161/CIRCULATIONAHA.116.024534.

Banerjee, I., Fuseler, J.W., Price, R.L., Borg, T.K., and Baudino, T.A., 2007. Determination of cell types and numbers during cardiac development in the neonatal and adult rat and mouse. *Am j physiol heart circ physiol* [Online], 293(3), H1883–91. Available from: https://doi.org/10.1152/ajpheart.00514.2007.

Bankhead, P., Loughrey, M.B., Fernandez, J.A., Dombrowski, Y., McArt, D.G., Dunne, P.D., McQuaid, S., Gray, R.T., Murray, L.J., Coleman, H.G., James, J.A., Salto-Tellez, M., and Hamilton, P.W., 2017. Qupath: open source software for digital pathology image analysis. *Sci rep* [Online], 7(1), p.16878. Available from: https://doi.org/10.1038/s41598-017-17204-5.

Baralle, F.E. and Giudice, J., 2017. Alternative splicing as a regulator of development and tissue identity. *Nat rev mol cell biol* [Online], 18(7), pp.437–451. eprint: 2017May10. Available from: https://doi.org/10.1038/nrm.2017.27.

Barp, A., Bello, L., Politano, L., Melacini, P., Calore, C., Polo, A., Vianello, S., Soraru, G., Semplicini, C., Pantic, B., Taglia, A., Picillo, E., Magri, F., Gorni, K., Messina, S., Vita, G.L., Vita, G., Comi, G.P., Ermani, M., Calvo, V., Angelini, C., Hoffman, E.P., and Pegoraro, E., 2015. Genetic modifiers of duchenne muscular dystrophy and dilated cardiomyopathy. *Plos one* [Online], 10(10), e0141240. Available from: https://doi.org/10.1371/journal.pone.0141240.

Barry, S.P., Davidson, S.M., and Townsend, P.A., 2008. Molecular regulation of cardiac hypertrophy. *Int j biochem cell biol* [Online], 40(10), pp.2023–39. Available from: https://doi.org/10.1016/j.biocel.2008.02.020.

Barve, R.A., Gu, C.C., Yang, W., Chu, J., Davila-Roman, V.G., and Fuentes, L. de las, 2016. Genetic association of left ventricular mass assessed by m-mode and two-dimensional echocardiography. *J hypertens* [Online], 34(1), pp.88–96. Available from: https://doi.org/10.1097/HJH.0000000000000765.

Baschong, W., Suetterlin, R., and Laeng, R.H., 2001. Control of autofluorescence of archival formaldehyde-fixed, paraffin-embedded tissue in confocal laser scanning microscopy (clsm). *J histochem cytochem* [Online], 49(12), pp.1565–72. Available from: https://doi.org/10.1177/002215540104901210.

Bassan, H., Bassan, M., Pinhasov, A., Kariv, N., Giladi, E., Gozes, I., and Harel, S., 2005. The pregnant spontaneously hypertensive rat as a model of asymmetric intrauterine growth retardation and neurodevelopmental delay. *Hypertension in pregnancy* [Online], 24(3), pp.201–211. Available from: https://doi.org/10.1080/10641950500281142.

Basting, T. and Lazartigues, E., 2017. Doca-salt hypertension: an update. *Curr hypertens rep* [Online], 19(4), p.32. Available from: https://doi.org/10.1007/s11906-017-0731-4.

Bastos, A., Gomes, A.V.P., Silva, G.R., Emerenciano, M., Ferreira, L.B., and Gimba, E.R.P., 2023. The intracellular and secreted sides of osteopontin and their putative physiopathological roles. *Int j mol sci* [Online], 24(3). Available from: https://doi.org/10.3390/ijms24032942.

Baud, A., Guryev, V., Hummel, O., Johannesson, M., Rat Genome Sequencing and Mapping Consortium, and Flint, J., 2014. Genomes and phenomes of a population of outbred rats and its progenitors. *Sci data* [Online], 1, p.140011. Available from: https://doi.org/10.1038/sdata.2014.11.

Bell, J.R., Curl, C.L., Harding, T.W., Vila Petroff, M., Harrap, S.B., and Delbridge, L.M.D., 2016. Male and female hypertrophic rat cardiac myocyte functional responses to ischemic stress and beta-adrenergic challenge are different. *Biol sex differ* [Online], 7, p.32. Available from: https://doi.org/10.1186/s13293-016-0084-8.

Bella, J.N. and Goring, H.H., 2012. Genetic epidemiology of left ventricular hypertrophy. *Am j cardiovasc dis* [Online], 2(4), pp.267–78. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23173100.

Bellahcene, A., Castronovo, V., Ogbureke, K.U., Fisher, L.W., and Fedarko, N.S., 2008. Small integrin-binding ligand n-linked glycoproteins (siblings): multifunctional proteins in cancer. *Nat rev cancer* [Online], 8(3), pp.212–26. Available from: https://doi.org/10.1038/nrc2345.

Benard, L., Oh, J.G., Cacheux, M., Lee, A., Nonnenmacher, M., Matasic, D.S., Kohlbrenner, E., Kho, C., Pavoine, C., Hajjar, R.J., and Hulot, J.S., 2016. Cardiac stim1 silencing impairs adaptive hypertrophy and promotes heart failure through inactivation of mtorc2/akt signaling. *Circulation* [Online], 133(15), 1458–71, discussion 1471. Available from: https://doi.org/10.1161/CIRCULATIONAHA.115.020678.

Benavides, F., Rulicke, T., Prins, J.B., Bussell, J., Scavizzi, F., Cinelli, P., Herault, Y., and Wedekind, D., 2020. Genetic quality assurance and genetic monitoring of laboratory mice and rats: felasa working group report. *Lab anim* [Online], 54(2), pp.135–148. Available from: https://doi.org/10.1177/0023677219867719.

Beraldi, R., Li, X., Martinez Fernandez, A., Reyes, S., Secreto, F., Terzic, A., Olson, T.M., and Nelson, T.J., 2014. Rbm20-deficient cardiogenesis reveals early disruption of rna processing and sarcomere remodeling establishing a developmental etiology for dilated cardiomyopathy. *Hum mol genet* [Online], 23(14), pp.3779–91. Available from: https://doi.org/10.1093/hmg/ddu091.

Berg, C.W. van den, Okawa, S., Chuva de Sousa Lopes, S.M., Iperen, L. van, Passier, R., Braam, S.R., Tertoolen, L.G., Sol, A. del, Davis, R.P., and Mummery, C.L., 2015. Transcriptome of human foetal heart compared with cardiomyocytes from pluripotent stem cells. *Development* [Online], 142(18), pp.3231–8. Available from: https://doi.org/10.1242/dev.123810.

Bezzina, C.R., Barc, J., Mizusawa, Y., Remme, C.A., Gourraud, J.B., Simonet, F., Verkerk, A.O., Schwartz, P.J., Crotti, L., Dagradi, F., Guicheney, P., Fressart, V., Leenhardt, A., Antzelevitch, C., Bartkowiak, S., Borggrefe, M., Schimpf, R., Schulze-Bahr, E., Zumhagen, S., Behr, E.R., Bastiaenen, R., Tfelt-Hansen, J., Olesen, M.S., Kaab, S., Beckmann, B.M., Weeke, P., Watanabe, H., Endo, N., Minamino, T., Horie, M., Ohno, S., Hasegawa, K., Makita, N., Nogami, A., Shimizu, W., Aiba, T., Froguel, P., Balkau, B., Lantieri, O., Torchio, M., Wiese, C., Weber, D., Wolswinkel, R., Coronel, R., Boukens, B.J., Bezieau, S., Charpentier, E., Chatel, S., Despres, A., Gros, F., Kyndt, F., Lecointe, S., Lindenbaum, P., Portero, V., Violleau, J., Gessler, M., Tan, H.L., Roden, D.M., Christoffels, V.M., Le Marec, H., Wilde, A.A., Probst, V., Schott, J.J., Dina, C., and Redon, R., 2013. Common variants at scn5a-scn10a and hey2 are associated with brugada syndrome, a rare disease with high risk of sudden cardiac death. *Nat genet* [Online], 45(9), pp.1044–9. Available from: https://doi.org/10.1038/ng.2712.

BHF, 2023. Available from: https://www.bhf.org.uk/-/media/files/for-professionals/research/heart-statistics/bhf-cvd-statistics-global-factsheet.pdf.

Bioquest, A., 2023. *Fluorescence spectrum viewer* [Online]. Available from: https://www.aatbio.com/fluorescence-excitation-emission-spectrum-graph-viewer?compare=5_fitc_fluorescein_5_isothiocyanate;cy5_cyanine_5;cy7_cyanine_7;dapi_4_6_diamidino_2_phenylindole;tritc_tetramethylrhodamine_isothiocyanate;cfp [Accessed 2023].

Blood Pressure Lowering Treatment Trialists Collaboration, 2021. Pharmacological blood pressure lowering for primary and secondary prevention of cardiovascular disease across different levels of blood pressure: an individual participant-level data meta-analysis. *Lancet* [Online], 397(10285), pp.1625–1636. Available from: https://doi.org/10.1016/S0140-6736(21)00590-0.

Bodenhofer, U., Bonatesta, E., Horejs-Kainrath, C., and Ho, S., 2015. Msa: an r package for multiple sequence alignment. *Bioinformatics* [Online], 31(24), pp.3997–3999. Available from: https://doi.org/10.1093/bioinformatics/btv494.

Boileau, E., Doroudgar, S., Riechert, E., Jurgensen, L., Ho, T.C., Katus, H.A., Volkers, M., and Dieterich, C., 2020. A multi-network comparative analysis of transcriptome and translatome identifies novel hub genes in cardiac remodeling. *Front genet* [Online], 11, p.583124. Available from: https://doi.org/10.3389/fgene.2020.583124.

Boluyt, M.O., Converso, K., Hwang, H.S., Mikkor, A., and Russell, M.W., 2004. Echocardiographic assessment of age-associated changes in systolic and diastolic function of the female f344 rat heart. *J appl physiol (1985)* [Online], 96(2), pp.822–8. Available from: https://doi.org/10.1152/japplphysiol.01026.2003.

Bombelli, M., Vanoli, J., Facchetti, R., Maloberti, A., Cuspidi, C., Grassi, G., and Mancia, G., 2023. Impact of the increase in left ventricular mass on the risk of long-term cardiovascular mortality: a prospective cohort study. *Hypertension* [Online], 80(6), pp.1321–1330. Available from: https://doi.org/10.1161/HYPERTENSIONAHA.122.19988.

Bornstein, A.B., Rao, S.S., and Marwaha, K., 2023. Left ventricular hypertrophy. In: *Statpearls* [Online]. Treasure Island (FL). Available from: https://www.ncbi.nlm.nih.gov/pubmed/32491466.

Borrego, I., Frobert, A., Ajalbert, G., Valentin, J., Kaltenrieder, C., Fellay, B., Stumpe, M., Cook, S., Dengjel, J., and Giraud, M.N., 2022. Fibrin, bone marrow cells and macrophages interactively modulate cardiomyoblast fate. *Biomedicines* [Online], 10(3). Available from: https://doi.org/10.3390/biomedicines10030527.

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L., 2016. Near-optimal probabilistic rna-seq quantification. *Nature biotechnology* [Online], 34, pp.525–527. Available from: https://doi.org/10.1038/nbt.3519.

Brooks, W.W., Shen, S.S., Conrad, C.H., Goldstein, R.H., and Bing, O.H., 2010. Transition from compensated hypertrophy to systolic heart failure in the spontaneously hypertensive rat: structure, function, and transcript analysis. *Genomics* [Online], 95(2), pp.84–92. Available from: https://doi.org/10.1016/j.ygeno.2009.12.002.

Brown, G.S., Jang, J., and Li, D., 2022. Growth factors and their roles in cardiac development and regeneration: a narrative review. *Pediatric medicine* [Online], 0, pp.1–12. Available from: https://doi.org/10.21037/pm-22-17.

Buccitelli, C. and Selbach, M., 2020. Mrnas, proteins and the emerging principles of gene expression control. *Nat rev genet* [Online], 21(10), pp.630–644. Available from: https://doi.org/10.1038/s41576-020-0258-4.

Burke, G.L., Arcilla, R.A., Culpepper, W.S., Webber, L.S., Chiang, Y.K., and Berenson, G.S., 1987. Blood pressure and echocardiographic measures in children: the bogalusa heart study. *Circulation* [Online], 75(1), pp.106–14. Available from: https://doi.org/10.1161/01.cir.75.1.106.

Chen, C., Wang, J., Pan, D., Wang, X., Xu, Y., Yan, J., Wang, L., Yang, X., Yang, M., and Liu, G.P., 2023. Applications of multi-omics analysis in human diseases. *Medcomm (2020)* [Online], 4(4), e315. Available from: https://doi.org/10.1002/mco2.315.

Chen, M., Wang, L., Li, Y., Chen, Y., Zhang, H., Zhu, Y., He, R., Li, H., Lin, J., Zhang, Y., and Zhang, C., 2020. Genetic modifiers of duchenne muscular dystrophy in chinese patients. *Front neurol* [Online], 11, p.721. Available from: https://doi.org/10.3389/fneur.2020.00721.

Chen, S., Zhou, Y., Chen, Y., and Gu, J., 2018. Fastp: an ultra-fast all-in-one fastq preprocessor. *Bioinformatics* [Online], 34 (17), pp.i884–i890. Available from: https://doi.org/10.1093/bioinformatics/bty560.

Cheng, Q., Chen, M., Wang, H., Chen, X., Wu, H., Du, Y., and Xue, J., 2022a. Microrna-27a-3p inhibits lung and skin fibrosis of systemic sclerosis by negatively regulating spp1. *Genomics* [Online], 114(4), p.110391. Available from: https://doi.org/10.1016/j.ygeno.2022.110391.

Cheng, Y., Li, Y., Scherer, N., Grundner-Culemann, F., Lehtimaki, T., Mishra, B.H., Raitakari, O.T., Nauck, M., Eckardt, K.U., Sekula, P., Schultheiss, U.T., and investigators, G., 2022b. Genetics of osteopontin in patients with chronic kidney disease: the german chronic kidney disease study. *Plos genet* [Online], 18(4), e1010139. Available from: https://doi.org/10.1371/journal.pgen.1010139.

Chenouard, V., Remy, S., Tesson, L., Menoret, S., Ouisse, L.H., Cherifi, Y., and Anegon, I., 2021. Advances in genome editing and application to the generation of genetically modified rat models. *Front genet* [Online], 12, p.615491. Available from: https://doi.org/10.3389/fgene.2021.615491.

Ching, S., Chia, Y., and Wan Azman, W., 2012. Prevalence and determinants of left ventricular hypertrophy in hypertensive patients at a primary care clinic. *Malays fam physician* [Online], 7(2-3), pp.2–9. Available from: https://www.ncbi.nlm.nih.gov/pubmed/25606250.

Chitre, A.S., Polesskaya, O., Holl, K., Gao, J., Cheng, R., Bimschleger, H., Garcia Martinez, A., George, T., Gileta, A.F., Han, W., Horvath, A., Hughson, A., Ishiwari, K., King, C.P., Lamparelli, A., Versaggi, C.L., Martin, C., St Pierre, C.L., Tripi, J.A., Wang, T., Chen, H., Flagel, S.B., Meyer, P., Richards, J., Robinson, T.E., Palmer, A.A., and Solberg Woods, L.C., 2020. Genome-wide association study in 3,173 outbred

rats identifies multiple loci for body weight, adiposity, and fasting glucose. *Obesity (silver spring)* [Online], 28(10), pp.1964–1973. Available from: https://doi.org/10.1002/oby.22927.

Cicha, I., 2021. The grand challenges in cardiovascular drug delivery. *Frontiers in drug delivery* [Online], 1. Available from: https://doi.org/10.3389/fddev.2021.784731.

Cingolani, P., Patel, V., Coon, M., Nguyen, T., Land, S., Ruden, D., and Lu, X., 2012a. Using drosophila melanogaster as a model for genotoxic chemical mutational studies with a new program, snpsift. *Frontiers in genetics*, 3.

Cingolani, P., Platts, A., Coon, M., Nguyen, T., Wang, L., Land, S., Lu, X., and Ruden, D., 2012b. A program for annotating and predicting the effects of single nucleotide polymorphisms, snpeff: snps in the genome of drosophila melanogaster strain w1118; iso-2; iso-3. *Fly*, 6(2), pp.80–92.

Claridge, B., Rai, A., Fang, H., Matsumoto, A., Luo, J., McMullen, J.R., and Greening, D.W., 2021. Proteome characterisation of extracellular vesicles isolated from heart. *Proteomics* [Online], 21(13-14), e2100026. Available from: https://doi.org/10.1002/pmic.202100026.

Clark, J.S., Jeffs, B., Davidson, A.O., Lee, W.K., Anderson, N.H., Bihoreau, M.T., Brosnan, M.J., Devlin, A.M., Kelman, A.W., Lindpaintner, K., and Dominiczak, A.F., 1996. Quantitative trait loci in genetically hypertensive rats. possible sex specificity. *Hypertension* [Online], 28(5), pp.898–906. Available from: https://doi.org/10.1161/01.hyp.28.5.898.

Claycomb, W.C., Lanson, N.A., Stallworth, B.S., Egeland, D.B., Delcarpio, J.B., Bahinski, A., and Izzo, N.J., 1998. Hl-1 cells: a cardiac muscle cell line that contracts and retains phenotypic characteristics of the adult cardiomyocyte. *Proceedings of the national academy of sciences of the united states of america* [Online], 95(6), pp.2979–2984. Available from: https://doi.org/DOI10.1073/pnas.95.6.2979.

Collins, A.R., Schnee, J., Wang, W., Kim, S., Fishbein, M.C., Bruemmer, D., Law, R.E., Nicholas, S., Ross, R.S., and Hsueh, W.A., 2004. Osteopontin modulates angiotensin ii-induced fibrosis in the intact murine heart. *J am coll cardiol* [Online], 43(9), pp.1698–705. Available from: https://doi.org/10.1016/j.jacc.2003.11.058.

Compeau, P., Pevzner, P., and Tesler, G., 2011. How to apply de bruijn graphs to genome assembly. *Nature biotechnology* [Online], 29, pp.987–991. Available from: https://doi.org/10.1038/nbt.2023.

Conesa, A. and Beck, S., 2019. Making multi-omics data accessible to researchers. *Sci data* [Online], 6(1), p.251. Available from: https://doi.org/10.1038/s41597-019-0258-4.

Corchete, L.A., Rojas, E.A., Alonso-Lopez, D., De Las Rivas, J., Gutierrez, N.C., and Burguillo, F.J., 2020. Systematic comparison and assessment of rna-seq procedures for gene expression quantitative analysis. *Sci rep* [Online], 10(1), p.19737. Available from: https://doi.org/10.1038/s41598-020-76881-x.

Crawford, W., 2010. *Microarray analysis of chromosome 14 congenic strains in the shrsp*. Master's thesis. Available at https://theses.gla.ac.uk/2761/1/2010CrawfordwMSc.pdf. Glasgow, UK: Glasgow, UK.

Cui, Y., Zheng, Y., Liu, X., Yan, L., Fan, X., Yong, J., Hu, Y., Dong, J., Li, Q., Wu, X., Gao, S., Li, J., Wen, L., Qiao, J., and Tang, F., 2019. Single-cell transcriptome analysis maps the developmental track of the human heart. *Cell rep* [Online], 26(7), 1934–1950 e5. Available from: https://doi.org/10.1016/j.celrep.2019.01.079.

Cuspidi, C., Faggiano, A., and Tadic, M., 2023. Hypertensive organ damage: the vulnerable heart of women. *J hum hypertens* [Online], 37(11), pp.1047–1048. Available from: https://doi.org/10.1038/s41371-023-00808-y.

Cuspidi, C., Sala, C., Negri, F., Mancia, G., Morganti, A., and Italian Society of, H., 2012. Prevalence of left-ventricular hypertrophy in hypertension: an updated review of echocardiographic studies. *J hum hypertens* [Online], 26(6), pp.343–9. Available from: https://doi.org/10.1038/jhh.2011.104.

Dai, J., Peng, L., Fan, K., Wang, H., Wei, R., Ji, G., Cai, J., Lu, B., Li, B., Zhang, D., Kang, Y., Tan, M., Qian, W., and Guo, Y., 2009. Osteopontin induces angiogenesis through activation of pi3k/akt and erk1/2 in endothelial cells. *Oncogene* [Online], 28(38), pp.3412–22. Available from: https://doi.org/10.1038/onc.2009.189.

Dalal, S., Zha, Q., Daniels, C.R., Steagall, R.J., Joyner, W.L., Gadeau, A.P., Singh, M., and Singh, K., 2014. Osteopontin stimulates apoptosis in adult cardiac myocytes via the involvement of cd44 receptors, mitochondrial death pathway, and endoplasmic reticulum stress. *Am j physiol heart circ physiol* [Online], 306(8), H1182–91. Available from: https://doi.org/10.1152/ajpheart.00954.2013.

Dam, S. van, Vosa, U., Graaf, A. van der, Franke, L., and Magalhaes, J.P. de, 2018. Gene co-expression analysis for functional classification and gene-disease predictions. *Brief bioinform* [Online], 19(4), pp.575–592. Available from: https://doi.org/10.1093/bib/bbw139.

Dannenberg, A.L., Levy, D., and Garrison, R.J., 1989. Impact of age on echocardiographic left ventricular mass in a healthy population (the framingham study). *Am j cardiol* [Online], 64(16), pp.1066–8. Available from: https://doi.org/10.1016/0002-9149(89)90816-3.

Datta, R., Bansal, T., Rana, S., Datta, K., Datta Chaudhuri, R., Chawla-Sarkar, M., and Sarkar, S., 2017. Myocyte-derived hsp90 modulates collagen upregulation via biphasic activation of stat-3 in fibroblasts during cardiac hypertrophy. *Mol cell biol* [Online], 37(6). Available from: https://doi.org/10.1128/MCB.00611-16.

Davidson, A.O., Schork, N., Jaques, B.C., Kelman, A.W., Sutcliffe, R.G., Reid, J.L., and Dominiczak, A.F., 1995. Blood pressure in genetically hypertensive rats. influence of the y chromosome. *Hypertension* [Online], 26(3), pp.452–9. Available from: https://doi.org/10.1161/01.hyp.26.3.452.

Davis, M.B., Arany, Z., McNamara, D.M., Goland, S., and Elkayam, U., 2020. Peripartum cardiomyopathy: jacc state-of-the-art review. *J am coll cardiol* [Online], 75(2), pp.207–221. Available from: https://doi.org/10.1016/j.jacc.2019.11.014.

de Jong, T.V., Pan, Y., Rastas, P., Munro, D., Tutaj, M., Akil, H., Benner, C., Chen, D., Chitre, A.S., Chow, W., Colonna, V., Dalgard, C.L., Demos, W.M., Doris, P.A., Garrison, E., Geurts, A.M., Gunturkun, H.M., Guryev, V., Hourlier, T., Howe, K., Huang, J., Kalbfleisch, T., Kim, P., Li, L., Mahaffey, S., Martin, F.J., Mohammadi, P., Ozel, A.B., Polesskaya, O., Pravenec, M., Prins, P., Sebat, J., Smith, J.R., Solberg Woods, L.C., Tabakoff, B., Tracey, A., Uliano-Silva, M., Villani, F., Wang, H., Sharp, B.M., Telese, F., Jiang, Z., Saba, L., Wang, X., Murphy, T.D., Palmer, A.A., Kwitek, A.E., Dwinell, M.R., Williams, R.W., Li, J.Z., and Chen, H., 2023. A revamped rat reference genome improves the discovery of genetic diversity in laboratory rats. *Biorxiv* [Online]. Available from: https://doi.org/10.1101/2023.04.13.536694.

Deepti, P., Pasha, A., Kumbhakar, D.V., Doneti, R., Heena, S.K., Bhanoth, S., Poleboyina, P.K., Yadala, R., S, D.A., and Pawar, S.C., 2022. Overexpression of secreted phosphoprotein 1 (spp1) predicts poor survival in hpv positive cervical cancer. *Gene* [Online], 824, p.146381. Available from: https://doi.org/10.1016/j.gene.2022.146381.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., Angel, G. del, Rivas, M.A., Hanna, M., McKenna, A., Fennell, T.J., Kernytsky, A.M., Sivachenko, A.Y., Cibulskis, K., Gabriel, S.B., Altshuler, D., and Daly, M.J., 2011. A framework for variation discovery and genotyping using next-generation dna sequencing data. *Nat genet* [Online], 43(5), pp.491–8. Available from: https://doi.org/10.1038/ng.806.

Devereux, R.B. and Reichek, N., 1977. Echocardiographic determination of left ventricular mass in man. anatomic validation of the method. *Circulation* [Online], 55(4), pp.613–8. Available from: https://doi.org/10.1161/01.cir.55.4.613.

Dewey, F.E., Perez, M.V., Wheeler, M.T., Watt, C., Spin, J., Langfelder, P., Horvath, S., Hannenhalli, S., Cappola, T.P., and Ashley, E.A., 2011. Gene coexpression network topology of cardiac development, hypertrophy, and failure. *Circulation-cardiovascular genetics* [Online], 4(1), 26–U129. Available from: https://doi.org/10.1161/Circgenetics.110.941757.

Dewing, J.M., Saunders, V., O'Kelly, I., and Wilson, D.I., 2022. Defining cardiac cell populations and relative cellular composition of the early fetal human heart. *Plos one* [Online], 17(11), e0259477. Available from: https://doi.org/10.1371/journal.pone.0259477.

Dhande, I.S., Zhu, Y., Kneedler, S.C., Joshi, A.S., Hicks, M.J., Wenderfer, S.E., Braun, M.C., and Doris, P.A., 2020. Stim1 polymorphism disrupts immune signaling and creates renal injury in hypertension. *J am heart assoc* [Online], 9(5), e014142. Available from: https://doi.org/10.1161/JAHA.119.014142.

Di Nicolantonio, R., Kostka, V., Kwitek, A., Jacob, H., Thomas, W.G., and Harrap, S.B., 2006. Fine mapping of lvm1: a quantitative trait locus controlling heart size independently of blood pressure. *Pulm pharmacol ther* [Online], 19(1), pp.70–3. Available from: https://doi.org/10.1016/j.pupt.2005.02.010.

Diaz, L., Zambrano, E., Flores, M.E., Contreras, M., Crispin, J.C., Aleman, G., Bravo, C., Armenta, A., Valdes, V.J., Tovar, A., Gamba, G., Barrios-Payan, J., and Bobadilla, N.A., 2020. Ethical considerations in animal research: the principle of 3r's. *Rev invest clin* [Online], 73(5). Available from: https://doi.org/s113961211409.

Diez, J. and Butler, J., 2023. Growing heart failure burden of hypertensive heart disease: a call to action. *Hypertension* [Online], 80(1), pp.13–21. Available from: https://doi.org/10.1161/HYPERTENSIONAHA.122.19373.

Diez, J. and Frohlich, E.D., 2010. A translational approach to hypertensive heart disease. *Hypertension* [Online], 55(1), pp.1–8. Available from: https://doi.org/10.1161/HYPERTENSIONAHA.109.141887.

Dirkx, E., Costa Martins, P.A. da, and De Windt, L.J., 2013. Regulation of fetal gene expression in heart failure. *Biochim biophys acta* [Online], 1832(12), pp.2414–24. Available from: https://doi.org/10.1016/j.bbadis.2013.07.023.

Doetschman, T. and Azhar, M., 2012. Cardiac-specific inducible and conditional gene targeting in mice. *Circ res* [Online], 110(11), pp.1498–512. Available from: https://doi.org/10.1161/CIRCRESAHA.112.265066.

Doggrell, S.A. and Brown, L., 1998. Rat models of hypertension, cardiac hypertrophy and failure. *Cardiovasc res* [Online], 39(1), pp.89–105. Available from: https://doi.org/10.1016/s0008-6363(98)00076-5.

Dornas, W.C. and Silva, M.E., 2011. Animal models for the study of arterial hypertension. *J biosci* [Online], 36(4), pp.731–7. Available from: https://doi.org/10.1007/s12038-011-9097-y.

Douglas, K., Graham, D., Crawford, W., McBride, M., and Dominiczak, A., 2010. Left ventricular function in a model of left ventricular hypertrophy: 4c.02. *Journal of hypertension* [Online]. Vol. 28, e213. Available from: https://doi.org/10.1097/01.hjh.0000378854.46256.4f.

Drazner, M.H., 2011. The progression of hypertensive heart disease. *Circulation* [Online], 123(3), pp.327–34. Available from: https://doi.org/10.1161/CIRCULATIONAHA.108.845792.

Drozdov, I., Didangelos, A., Yin, X., Zampetaki, A., Abonnenc, M., Murdoch, C., Zhang, M., Ouzounis, C.A., Mayr, M., Tsoka, S., and Shah, A.M., 2013. Gene network and proteomic analyses of cardiac responses to pathological and physiological stress. *Circ cardiovasc genet* [Online], 6(6), pp.588–97. Available from: https://doi.org/10.1161/CIRCGENETICS.113.000063.

Du, Y., Mao, L., Wang, Z., Yan, K., Zhang, L., and Zou, J., 2022. Osteopontin - the stirring multifunctional regulatory factor in multisystem aging. *Front endocrinol (lausanne)* [Online], 13, p.1014853. Available from: https://doi.org/10.3389/fendo.2022.1014853.

Duong, P., Chung, A., Bouchareychas, L., and Raffai, R.L., 2019. Cushioned-density gradient ultracentrifugation (c-dguc) improves the isolation efficiency of extracellular vesicles. *Plos one* [Online], 14(4). Available from: https://doi.org/10.1371/journal.pone.0215324.

Elfassy, T., German, C., Muntner, P., Choi, E., Contreras, G., Shimbo, D., and Yang, E., 2023. Blood pressure and cardiovascular disease mortality among us adults: a sex-stratified analysis, 1999-2019. *Hypertension* [Online]. Available from: https://doi.org/10.1161/HYPERTENSIONAHA.123.21228.

Estill, S.J. and Garcia, J.A., 2000. A marker assisted selection protocol (masp) to generate c57bl/6j or 129s6/svevtac speed congenic or consomic strains. *Genesis* [Online], 28(3-4), pp.164–6. Available from: https://doi.org/10.1002/1526-968x(200011/12)28:3/4<164::aid-gene110>3.0.co;2-r.

Ettehad, D., Emdin, C.A., Kiran, A., Anderson, S.G., Callender, T., Emberson, J., Chalmers, J., Rodgers, A., and Rahimi, K., 2016. Blood pressure lowering for prevention of cardiovascular disease and death: a systematic review and meta-analysis. *Lancet* [Online], 387(10022), pp.957–967. Available from: https://doi.org/10.1016/S0140-6736(15)01225-8.

Evangelou, E., Warren, H.R., Mosen-Ansorena, D., Mifsud, B., Pazoki, R., Gao, H., Ntritsos, G., Dimou, N., Cabrera, C.P., Karaman, I., Ng, F.L., Evangelou, M., Witkowska, K., Tzanis, E., Hellwege, J.N., Giri, A., Velez Edwards, D.R., Sun, Y.V., Cho, K., Gaziano, J.M., Wilson, P.W.F., Tsao, P.S., Kovesdy, C.P., Esko, T., Magi, R., Milani, L., Almgren, P., Boutin, T., Debette, S., Ding, J., Giulianini, F., Holliday, E.G., Jackson, A.U., Li-Gao, R., Lin, W.Y., Luan, J., Mangino, M., Oldmeadow, C., Prins, B.P., Qian, Y., Sargurupremraj, M., Shah, N., Surendran, P., Theriault, S., Verweij, N., Willems, S.M., Zhao, J.H., Amouyel, P., Connell, J., Mutsert, R. de, Doney, A.S.F., Farrall, M., Menni, C., Morris, A.D., Noordam, R., Pare, G., Poulter, N.R., Shields, D.C., Stanton, A., Thom, S., Abecasis, G., Amin, N., Arking, D.E., Ayers, K.L., Barbieri, C.M., Batini, C., Bis, J.C., Blake, T., Bochud, M., Boehnke, M., Boerwinkle, E., Boomsma, D.I., Bottinger, E.P., Braund, P.S., Brumat, M., Campbell, A., Campbell, H., Chakravarti, A., Chambers, J.C., Chauhan, G., Ciullo, M., Cocca, M., Collins, F., Cordell, H.J., Davies, G., Borst, M.H. de, Geus, E.J. de, Deary, I.J., Deelen, J., Del Greco, M.F., Demirkale, C.Y., Dorr, M., Ehret, G.B., Elosua, R., Enroth, S., Erzurumluoglu, A.M., Ferreira, T., Franberg, M., Franco, O.H., Gandin, I., et al., 2018. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat genet* [Online], 50(10), pp.1412–1425. Available from: `https://doi.org/10.1038/s41588-018-0205-x`.

Evans, A.L., Brown, W., Kenyon, C.J., Maxted, K.J., and Smith, D.C., 1994. Improved system for measuring systolic blood pressure in the conscious rat. *Med biol eng comput* [Online], 32(1), pp.101–2. Available from: `https://doi.org/10.1007/BF02512487`.

Ewels, P., Magnusson, M., Lundin, S., and Käller, M., 2016. Multiqc: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* [Online], 32(19), pp.3047–3048. Available from: `https://doi.org/10.1093/bioinformatics/btw354`.

Fagard, R.H., Celis, H., Thijs, L., and Wouters, S., 2009. Regression of left ventricular mass by antihypertensive treatment: a meta-analysis of randomized comparative studies. *Hypertension* [Online], 54(5), pp.1084–91. Available from: `https://doi.org/10.1161/HYPERTENSIONAHA.109.136655`.

Fan, H.Y., Lin, W.Y., Lu, T.P., Chen, Y.Y., Hsu, J.B., Yu, S.L., Su, T.C., Lin, H.J., Chen, Y.C., and Chien, K.L., 2022. Targeted next-generation sequencing for genetic variants of left ventricular mass status among community-based adults in taiwan. *Front genet* [Online], 13, p.1064980. Available from: `https://doi.org/10.3389/fgene.2022.1064980`.

Fandl, H.K., Garcia, V.P., Treuth, J.W., Brewster, L.M., Greiner, J.J., Davy, K.P., Stauffer, B.L., and DeSouza, C.A., 2023. Endothelial derived extracellular vesicles from obese/hypertensive adults increase factors associated with hypertrophy and fibrosis

in cardiomyocytes. *Am j physiol heart circ physiol* [Online]. Available from: https://doi.org/10.1152/ajpheart.00035.2023.

Faqeer, A., Wang, M., Alam, G., Padhiar, A.A., Zheng, D., Luo, Z., Zhao, I.S., Zhou, G., Beucken, J.J.J.P. van den, Wang, H., and Zhang, Y., 2023. Cleaved spp1-rich extracellular vesicles from osteoclasts promote bone regeneration via tgfβ1/smad3 signaling. *Biomaterials* [Online], 303. Available from: https://doi.org/10.1016/j.biomaterials.2023.122367.

Farrell, E., Armstrong, A.E., Grimes, A.C., Naya, F.J., Lange, W.J. de, and Ralphe, J.C., 2018. Transcriptome analysis of cardiac hypertrophic growth in mybpc3-null mice suggests early responders in hypertrophic remodeling. *Front physiol* [Online], 9, p.1442. Available from: https://doi.org/10.3389/fphys.2018.01442.

Fazilaty, H. and Basler, K., 2023. Reactivation of embryonic genetic programs in tissue regeneration and disease. *Nat genet* [Online], 55(11), pp.1792–1806. Available from: https://doi.org/10.1038/s41588-023-01526-4.

Ferreira Filho, C., Abreu, L.C., Valenti, V.E., Ferreira, M., Meneghini, A., Silveira, J.A., Riera, A.R., Colombari, E., Murad, N., Santos-Silva, P.R., Silva, L.J., Vanderlei, L.C., Carvalho, T.D., and Ferreira, C., 2010. Anti-hypertensive drugs have different effects on ventricular hypertrophy regression. *Clinics (sao paulo)* [Online], 65(7), pp.723–8. Available from: https://doi.org/10.1590/S1807-59322010000700012.

Forman, D.E., Cittadini, A., Azhar, G., Douglas, P.S., and Wei, J.Y., 1997. Cardiac morphology and function in senescent rats: gender-related differences. *J am coll cardiol* [Online], 30(7), pp.1872–7. Available from: https://doi.org/10.1016/s0735-1097(97)00411-7.

Foryst-Ludwig, A., Kreissl, M.C., Sprang, C., Thalke, B., Bohm, C., Benz, V., Gurgen, D., Dragun, D., Schubert, C., Mai, K., Stawowy, P., Spranger, J., Regitz-Zagrosek, V., Unger, T., and Kintscher, U., 2011. Sex differences in physiological cardiac hypertrophy are associated with exercise-mediated changes in energy substrate availability. *Am j physiol heart circ physiol* [Online], 301(1), H115–22. Available from: https://doi.org/10.1152/ajpheart.01222.2010.

Freyermuth, F., Rau, F., Kokunai, Y., Linke, T., Sellier, C., Nakamori, M., Kino, Y., Arandel, L., Jollet, A., Thibault, C., Philipps, M., Vicaire, S., Jost, B., Udd, B., Day, J.W., Duboc, D., Wahbi, K., Matsumura, T., Fujimura, H., Mochizuki, H., Deryckere, F., Kimura, T., Nukina, N., Ishiura, S., Lacroix, V., Campan-Fournier, A., Navratil, V., Chautard, E., Auboeuf, D., Horie, M., Imoto, K., Lee, K.Y., Swanson, M.S., Munain, A.L. de, Inada, S., Itoh, H., Nakazawa, K., Ashihara, T., Wang, E., Zimmer, T., Furling, D., Takahashi, M.P., and Charlet-Berguerand, N., 2016. Splicing misregulation of scn5a contributes to cardiac-conduction delay and heart arrhythmia in myotonic dystrophy.

*Nat commun* [Online], 7, p.11067. Available from: https://doi.org/10.1038/ncomms11067.

Fry, A., Littlejohns, T.J., Sudlow, C., Doherty, N., Adamska, L., Sprosen, T., Collins, R., and Allen, N.E., 2017. Comparison of sociodemographic and health-related characteristics of uk biobank participants with those of the general population. *Am j epidemiol* [Online], 186(9), pp.1026–1034. Available from: https://doi.org/10.1093/aje/kwx246.

Fu, S.H., Zhang, Y.J., Li, Y.L., Luo, L.M., Zhao, Y.L., and Yao, Y., 2020. Extracellular vesicles in cardiovascular diseases. *Cell death discovery* [Online], 6(1). Available from: https://doi.org/6810.1038/s41420-020-00305-y.

Fujiu, K. and Nagai, R., 2014. Fibroblast-mediated pathways in cardiac hypertrophy. *J mol cell cardiol* [Online], 70, pp.64–73. Available from: https://doi.org/10.1016/j.yjmcc.2014.01.013.

Garimella, P.S., Toit, C. du, Le, N.N., and Padmanabhan, S., 2023. A genomic deep field view of hypertension. *Kidney int* [Online], 103(1), pp.42–52. Available from: https://doi.org/10.1016/j.kint.2022.09.029.

Garmany, R., Bos, J.M., Tester, D.J., Giudicessi, J.R., Dos Remedios, C.G., Dasari, S., Nagaraj, N.K., Nair, A.A., Johnson, K.L., Ryan, Z.C., Maleszewski, J.J., Ommen, S.R., Dearani, J.A., and Ackerman, M.J., 2023. Multi-omic architecture of obstructive hypertrophic cardiomyopathy. *Circ genom precis med* [Online], 16(2), e003756. Available from: https://doi.org/10.1161/CIRCGEN.122.003756.

Garvin, A.M., De Both, M.D., Talboom, J.S., Lindsey, M.L., Huentelman, M.J., and Hale, T.M., 2021. Transient ace (angiotensin-converting enzyme) inhibition suppresses future fibrogenic capacity and heterogeneity of cardiac fibroblast subpopulations. *Hypertension* [Online], 77(3), pp.904–918. Available from: https://doi.org/10.1161/HYPERTENSIONAHA.120.16352.

Georgakoudi, I., Jacobson, B.C., Muller, M.G., Sheets, E.E., Badizadegan, K., Carr-Locke, D.L., Crum, C.P., Boone, C.W., Dasari, R.R., Van Dam, J., and Feld, M.S., 2002. Nad(p)h and collagen as in vivo quantitative fluorescent biomarkers of epithelial precancerous changes. *Cancer res* [Online], 62(3), pp.682–7. Available from: https://www.ncbi.nlm.nih.gov/pubmed/11830520.

George, C.H., Rogers, S.A., Bertrand, B.M., Tunwell, R.E., Thomas, N.L., Steele, D.S., Cox, E.V., Pepper, C., Hazeel, C.J., Claycomb, W.C., and Lai, F.A., 2007. Alternative splicing of ryanodine receptors modulates cardiomyocyte ca2+ signaling and susceptibility to apoptosis. *Circ res* [Online], 100(6), pp.874–83. Available from: https://doi.org/10.1161/01.RES.0000260804.77807.cf.

Gerdts, E., Izzo, R., Mancusi, C., Losi, M.A., Manzi, M.V., Canciello, G., De Luca, N., Trimarco, B., and Simone, G. de, 2018. Left ventricular hypertrophy offsets the sex difference in cardiovascular risk (the campania salute network). *Int j cardiol* [Online], 258, pp.257–261. Available from: https://doi.org/10.1016/j.ijcard.2017.12.086.

Gerdts, E., Okin, P.M., Simone, G. de, Cramariuc, D., Wachtell, K., Boman, K., and Devereux, R.B., 2008. Gender differences in left ventricular structure and function during antihypertensive treatment: the losartan intervention for endpoint reduction in hypertension study. *Hypertension* [Online], 51(4), pp.1109–14. Available from: https://doi.org/10.1161/HYPERTENSIONAHA.107.107474.

Geske, J.B., Ong, K.C., Siontis, K.C., Hebl, V.B., Ackerman, M.J., Hodge, D.O., Miller, V.M., Nishimura, R.A., Oh, J.K., Schaff, H.V., Gersh, B.J., and Ommen, S.R., 2017. Women with hypertrophic cardiomyopathy have worse survival. *Eur heart j* [Online], 38(46), pp.3434–3440. Available from: https://doi.org/10.1093/eurheartj/ehx527.

Ghatpande, S., Goswami, S., Mascareno, E., and Siddiqui, M.A., 1999. Signal transduction and transcriptional adaptation in embryonic heart development and during myocardial hypertrophy. *Mol cell biochem* [Online], 196(1-2), pp.93–7. Available from: https://www.ncbi.nlm.nih.gov/pubmed/10448907.

Giachelli, C.M., Liaw, L., Murry, C.E., Schwartz, S.M., and Almeida, M., 1995. Osteopontin expression in cardiovascular diseases. *Ann n y acad sci* [Online], 760, pp.109–26. Available from: https://doi.org/10.1111/j.1749-6632.1995.tb44624.x.

Giacopelli, F., Marciano, R., Pistorio, A., Catarsi, P., Canini, S., Karsenty, G., and Ravazzolo, R., 2004. Polymorphisms in the osteopontin promoter affect its transcriptional activity. *Physiol genomics* [Online], 20(1), pp.87–96. Available from: https://doi.org/10.1152/physiolgenomics.00138.2004.

Gibbs, R.A., Weinstock, G.M., Metzker, M.L., Muzny, D.M., Sodergren, E.J., Scherer, S., Scott, G., Steffen, D., Worley, K.C., Burch, P.E., Okwuonu, G., Hines, S., Lewis, L., DeRamo, C., Delgado, O., Dugan-Rocha, S., Miner, G., Morgan, M., Hawes, A., Gill, R., Celera, Holt, R.A., Adams, M.D., Amanatides, P.G., Baden-Tillson, H., Barnstead, M., Chin, S., Evans, C.A., Ferriera, S., Fosler, C., Glodek, A., Gu, Z., Jennings, D., Kraft, C.L., Nguyen, T., Pfannkoch, C.M., Sitter, C., Sutton, G.G., Venter, J.C., Woodage, T., Smith, D., Lee, H.M., Gustafson, E., Cahill, P., Kana, A., Doucette-Stamm, L., Weinstock, K., Fechtel, K., Weiss, R.B., Dunn, D.M., Green, E.D., Blakesley, R.W., Bouffard, G.G., De Jong, P.J., Osoegawa, K., Zhu, B., Marra, M., Schein, J., Bosdet, I., Fjell, C., Jones, S., Krzywinski, M., Mathewson, C., Siddiqui, A., Wye, N., McPherson, J., Zhao, S., Fraser, C.M., Shetty, J., Shatsman, S., Geer, K., Chen, Y., Abramzon, S., Nierman, W.C., Havlak, P.H., Chen, R., Durbin, K.J., Egan,

A., Ren, Y., Song, X.Z., Li, B., Liu, Y., Qin, X., Cawley, S., Worley, K.C., Cooney, A.J., D'Souza, L.M., Martin, K., Wu, J.Q., Gonzalez-Garay, M.L., Jackson, A.R., Kalafus, K.J., McLeod, M.P., Milosavljevic, A., Virk, D., Volkov, A., Wheeler, D.A., Zhang, Z., Bailey, J.A., Eichler, E.E., et al., 2004. Genome sequence of the brown norway rat yields insights into mammalian evolution. *Nature* [Online], 428(6982), pp.493–521. Available from: https://doi.org/10.1038/nature02426.

Gladka, M.M., 2021. Single-cell rna sequencing of the adult mammalian heart-state-of-the-art and future perspectives. *Curr heart fail rep* [Online], 18(2), pp.64–70. Available from: https://doi.org/10.1007/s11897-021-00504-3.

Gladka, M.M., Molenaar, B., Ruiter, H. de, Elst, S. van der, Tsui, H., Versteeg, D., Lacraz, G.P.A., Huibers, M.M.H., Oudenaarden, A. van, and Rooij, E. van, 2018. Single-cell sequencing of the healthy and diseased heart reveals cytoskeleton-associated protein 4 as a new modulator of fibroblasts activation. *Circulation* [Online], 138(2), pp.166–180. Available from: https://doi.org/10.1161/CIRCULATIONAHA.117.030742.

Goble, M.M., Mosteller, M., Moskowitz, W.B., and Schieken, R.M., 1992. Sex differences in the determinants of left ventricular mass in childhood. the medical college of virginia twin study. *Circulation* [Online], 85(5), pp.1661–5. Available from: https://doi.org/10.1161/01.cir.85.5.1661.

Grabowski, K., Koplin, G., Aliu, B., Schulte, L., Schulz, A., and Kreutz, R., 2013. Mapping and confirmation of a major left ventricular mass qtl on rat chromosome 1 by contrasting shrsp and f344 rats. *Physiol genomics* [Online], 45(18), pp.827–33. Available from: https://doi.org/10.1152/physiolgenomics.00067.2013.

Grabowski, K., Riemenschneider, M., Schulte, L., Witten, A., Schulz, A., Stoll, M., and Kreutz, R., 2015. Fetal-adult cardiac transcriptome analysis in rats with contrasting left ventricular mass reveals new candidates for cardiac hypertrophy. *Plos one* [Online], 10(2), e0116807. Available from: https://doi.org/10.1371/journal.pone.0116807.

Graf, K., Do, Y.S., Ashizawa, N., Meehan, W.P., Giachelli, C.M., Marboe, C.C., Fleck, E., and Hsueh, W.A., 1997. Myocardial osteopontin expression is associated with left ventricular hypertrophy. *Circulation* [Online], 96(9), pp.3063–71. Available from: https://doi.org/10.1161/01.cir.96.9.3063.

Grisel, J.E. and Crabbe, J.C., 1995. Quantitative trait loci mapping. *Alcohol health res world* [Online], 19(3), pp.220–227. Available from: https://www.ncbi.nlm.nih.gov/pubmed/31798043.

Gs, A.K., Raj, B., Santhosh, K.S., Sanjay, G., and Kartha, C.C., 2014. Ascending aortic constriction in rats for creation of pressure overload cardiac hypertrophy model. *J vis exp* [Online], (88), e50983. Available from: https://doi.org/10.3791/50983.

GTEx Consortium, 2020. The gtex consortium atlas of genetic regulatory effects across human tissues. *Science* [Online], 369(6509), pp.1318–1330. Available from: https://doi.org/10.1126/science.aaz1776.

Guo, S., Okyere, A.D., McEachern, E., Strong, J.L., Carter, R.L., Patwa, V.C., Thomas, T.P., Landy, M., Song, J., Lucchese, A.M., Martin, T.G., Gao, E., Rajan, S., Kirk, J.A., Koch, W.J., Cheung, J.Y., and Tilley, D.G., 2022. Epidermal growth factor receptor-dependent maintenance of cardiac contractility. *Cardiovasc res* [Online], 118(5), pp.1276–1288. Available from: https://doi.org/10.1093/cvr/cvab149.

Guo, W., Schafer, S., Greaser, M.L., Radke, M.H., Liss, M., Govindarajan, T., Maatz, H., Schulz, H., Li, S., Parrish, A.M., Dauksaite, V., Vakeel, P., Klaassen, S., Gerull, B., Thierfelder, L., Regitz-Zagrosek, V., Hacker, T.A., Saupe, K.W., Dec, G.W., Ellinor, P.T., MacRae, C.A., Spallek, B., Fischer, R., Perrot, A., Ozcelik, C., Saar, K., Hubner, N., and Gotthardt, M., 2012. Rbm20, a gene for hereditary cardiomyopathy, regulates titin splicing. *Nat med* [Online], 18(5), pp.766–73. Available from: https://doi.org/10.1038/nm.2693.

Gupta, S., Rawat, S., Arora, V., Kottarath, S.K., Dinda, A.K., Vaishnav, P.K., Nayak, B., and Mohanty, S., 2018. An improvised one-step sucrose cushion ultracentrifugation method for exosome isolation from culture supernatants of mesenchymal stem cells. *Stem cell res ther* [Online], 9(1), p.180. Available from: https://doi.org/10.1186/s13287-018-0923-0.

Han, Z., Shi, F., Chen, Y., Dong, X., Zhang, B., and Li, M., 2021. Relationship between mirna-433 and spp1 in the presence of fracture and traumatic brain injury. *Exp ther med* [Online], 22(3), p.928. Available from: https://doi.org/10.3892/etm.2021.10360.

Hansen, C. and Spuhler, K., 1984. Development of the national institutes of health genetically heterogeneous rat stock. *Alcohol clin exp res* [Online], 8(5), pp.477–9. Available from: https://doi.org/10.1111/j.1530-0277.1984.tb05706.x.

Hardison, R.C., 2016. A guide to translation of research results from model organisms to human. *Genome biol* [Online], 17(1), p.161. Available from: https://doi.org/10.1186/s13059-016-1026-9.

Harper, A.R., Goel, A., Grace, C., Thomson, K.L., Petersen, S.E., Xu, X., Waring, A., Ormondroyd, E., Kramer, C.M., Ho, C.Y., Neubauer, S., Investigators, H., Tadros, R., Ware, J.S., Bezzina, C.R., Farrall, M., and Watkins, H., 2021. Common genetic variants and modifiable risk factors underpin hypertrophic cardiomyopathy susceptibility and expressivity. *Nat genet* [Online], 53(2), pp.135–142. Available from: https://doi.org/10.1038/s41588-020-00764-0.

Harrap, S.B., Danes, V.R., Ellis, J.A., Griffiths, C.D., Jones, E.F., and Delbridge, L.M., 2002. The hypertrophic heart rat: a new normotensive model of genetic cardiac and cardiomyocyte hypertrophy. *Physiol genomics* [Online], 9(1), pp.43–8. Available from: https://doi.org/10.1152/physiolgenomics.00006.2002.

Hegenbarth, J.-C., Lezzoche, G., De Windt, L.J., and Stoll, M., 2022. Perspectives on bulk-tissue rna sequencing and single-cell rna sequencing for cardiac transcriptomics. *Frontiers in molecular medicine* [Online], 2. Available from: https://doi.org/10.3389/fmmed.2022.839338.

Heimark, S., Mehlum, M.H., Mancia, G., Soraas, C.L., Liestol, K., Wachtell, K., Larstorp, A.C., Rostrup, M., Mariampillai, J.E., Kjeldsen, S.E., Julius, S., and Weber, M.A., 2023. Middle-aged and older patients with left ventricular hypertrophy: higher mortality with drug treated systolic blood pressure below 130 mm hg. *Hypertension* [Online]. Available from: https://doi.org/10.1161/HYPERTENSIONAHA.123.21454.

Heineke, J. and Molkentin, J.D., 2006. Regulation of cardiac hypertrophy by intracellular signalling pathways. *Nat rev mol cell biol* [Online], 7(8), pp.589–600. Available from: https://doi.org/10.1038/nrm1983.

Hermsen, R., Ligt, J. de, Spee, W., Blokzijl, F., Schafer, S., Adami, E., Boymans, S., Flink, S., Boxtel, R. van, Weide, R.H. van der, Aitman, T., Hubner, N., Simonis, M., Tabakoff, B., Guryev, V., and Cuppen, E., 2015. Genomic landscape of rat strain and substrain variation. *Bmc genomics* [Online], 16(1), p.357. Available from: https://doi.org/10.1186/s12864-015-1594-1.

Hoeijmakers, J.H., 2009. Dna damage, aging, and cancer. *N engl j med* [Online], 361(15), pp.1475–85. Available from: https://doi.org/10.1056/NEJMra0804615.

Hoffman, E.P., Gordish-Dressman, H., McLane, V.D., Devaney, J.M., Thompson, P.D., Visich, P., Gordon, P.M., Pescatello, L.S., Zoeller, R.F., Moyna, N.M., Angelopoulos, T.J., Pegoraro, E., Cox, G.A., and Clarkson, P.M., 2013. Alterations in osteopontin modify muscle size in females in both humans and mice. *Med sci sports exerc* [Online], 45(6), pp.1060–8. Available from: https://doi.org/10.1249/MSS.0b013e31828093c1.

Homberg, J.R., Wohr, M., and Alenina, N., 2017. Comeback of the rat in biomedical research. *Acs chem neurosci* [Online], 8(5), pp.900–903. Available from: https://doi.org/10.1021/acschemneuro.6b00415.

Hosur, V., Low, B.E., Li, D., Stafford, G.A., Kohar, V., Shultz, L.D., and Wiles, M.V., 2020. Genes adapt to outsmart gene-targeting strategies in mutant mouse strains by skipping exons to reinitiate transcription and translation. *Genome biol* [Online], 21(1), p.168. Available from: https://doi.org/10.1186/s13059-020-02086-0.

Huang da, W., Sherman, B.T., and Lempicki, R.A., 2009. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids res* [Online], 37(1), pp.1–13. Available from: https://doi.org/10.1093/nar/gkn923.

Hubbard, N.E., Chen, Q.J., Sickafoose, L.K., Wood, M.B., Gregg, J.P., Abrahamsson, N.M., Engelberg, J.A., Walls, J.E., and Borowsky, A.D., 2013. Transgenic mammary epithelial osteopontin (spp1) expression induces proliferation and alveologenesis. *Genes cancer* [Online], 4(5-6), pp.201–12. Available from: https://doi.org/10.1177/1947601913496813.

Hulsmans, M., Schloss, M.J., Lee, I.H., Bapat, A., Iwamoto, Y., Vinegoni, C., Paccalet, A., Yamazoe, M., Grune, J., Pabel, S., Momin, N., Seung, H., Kumowski, N., Pulous, F.E., Keller, D., Bening, C., Green, U., Lennerz, J.K., Mitchell, R.N., Lewis, A., Casadei, B., Iborra-Egea, O., Bayes-Genis, A., Sossalla, S., Ong, C.S., Pierson, R.N., Aster, J.C., Rohde, D., Wojtkiewicz, G.R., Weissleder, R., Swirski, F.K., Tellides, G., Tolis G., J., Melnitchouk, S., Milan, D.J., Ellinor, P.T., Naxerova, K., and Nahrendorf, M., 2023. Recruited macrophages elicit atrial fibrillation. *Science* [Online], 381(6654), pp.231–239. Available from: https://doi.org/10.1126/science.abq3061.

Humphries, M.P., Maxwell, P., and Salto-Tellez, M., 2021. Qupath: the global impact of an open source digital pathology system. *Comput struct biotechnol j* [Online], 19, pp.852–859. Available from: https://doi.org/10.1016/j.csbj.2021.01.022.

Icer, M.A. and Gezmen-Karadag, M., 2018. The multiple functions and mechanisms of osteopontin. *Clin biochem* [Online], 59, pp.17–24. Available from: https://doi.org/10.1016/j.clinbiochem.2018.07.003.

Iczkiewicz, J., Jackson, M.J., Smith, L.A., Rose, S., and Jenner, P., 2006. Osteopontin expression in substantia nigra in mptp-treated primates and in parkinson's disease. *Brain res* [Online], 1118(1), pp.239–50. Available from: https://doi.org/10.1016/j.brainres.2006.08.036.

Ieda, M., Tsuchihashi, T., Ivey, K.N., Ross, R.S., Hong, T.T., Shaw, R.M., and Srivastava, D., 2009. Cardiac fibroblasts regulate myocardial proliferation through beta1 integrin signaling. *Dev cell* [Online], 16(2), pp.233–44. Available from: https://doi.org/10.1016/j.devcel.2008.12.007.

Innes, B.A., McLaughlin, M.G., Kapuscinski, M.K., Jacob, H.J., and Harrap, S.B., 1998. Independent genetic susceptibility to cardiac hypertrophy in inherited hypertension. *Hypertension* [Online], 31(3), pp.741–6. Available from: https://doi.org/10.1161/01.hyp.31.3.741.

Inomata, H., Watanabe, T., Iizuka, Y., Liang, Y.Q., Mashimo, T., Nabika, T., Ikeda, K., Yanai, K., Gotoda, T., Yamori, Y., Isobe, M., and Kato, N., 2005. Identification of quantitative trait loci for cardiac hypertrophy in two different strains of the spontaneously hypertensive rat. *Hypertens res* [Online], 28(3), pp.273–81. Available from: https://doi.org/10.1291/hypres.28.273.

Irion, C.I., Dunkley, J.C., John-Williams, K., Condor Capcha, J.M., Shehadeh, S.A., Pinto, A., Loebe, M., Webster, K.A., Brozzi, N.A., and Shehadeh, L.A., 2020. Nuclear osteopontin is a marker of advanced heart failure and cardiac allograft vasculopathy: evidence from transplant and retransplant hearts. *Front physiol* [Online], 11, p.928. Available from: https://doi.org/10.3389/fphys.2020.00928.

Israeli-Rosenberg, S., Manso, A.M., Okada, H., and Ross, R.S., 2014. Integrins and integrin-associated proteins in the cardiac myocyte. *Circ res* [Online], 114(3), pp.572–586. Available from: https://doi.org/10.1161/CIRCRESAHA.114.301275.

Jackson, T., Allard, M.F., Sreenan, C.M., Doss, L.K., Bishop, S.P., and Swain, J.L., 1990. The c-myc proto-oncogene regulates cardiac development in transgenic mice. *Mol cell biol* [Online], 10(7), pp.3709–16. Available from: https://doi.org/10.1128/mcb.10.7.3709-3716.1990.

Jafary, F.H. and Jafar, T.H., 2008. Disproportionately high risk of left ventricular hypertrophy in indo-asian women: a call for more studies. *Echocardiography* [Online], 25(8), pp.812–9. Available from: https://doi.org/10.1111/j.1540-8175.2008.00713.x.

Jasinska-Stroschein, M., 2022. Searching for an experimental rodent model of heart failure with preserved ejection fraction: re-visited. *Biomed pharmacother* [Online], 152, p.113251. Available from: https://doi.org/10.1016/j.biopha.2022.113251.

Jeffs, B., Negrin, C.D., Graham, D., Clark, J.S., Anderson, N.H., Gauguier, D., and Dominiczak, A.F., 2000. Applicability of a 'speed' congenic strategy to dissect blood pressure quantitative trait loci on rat chromosome 2. *Hypertension* [Online], 35(1 Pt 2), pp.179–87. Available from: https://doi.org/10.1161/01.hyp.35.1.179.

Jiang, P., Zhang, D., Qiu, H., Yi, X., Zhang, Y., Cao, Y., Zhao, B., Xia, Z., and Wang, C., 2017. Tiron ameliorates high glucose-induced cardiac myocyte apoptosis by pkcdelta-dependent inhibition of osteopontin. *Clin exp pharmacol physiol* [Online], 44(7), pp.760–770. Available from: https://doi.org/10.1111/1440-1681.12762.

Jimenez-Tellez, N. and Greenway, S.C., 2019. Cellular models for human cardiomyopathy: what is the best option? *World j cardiol* [Online], 11(10), pp.221–235. Available from: https://doi.org/10.4330/wjc.v11.i10.221.

Joshi, A., Rienks, M., Theofilatos, K., and Mayr, M., 2021. Systems biology in cardiovascular disease: a multiomics approach. *Nat rev cardiol* [Online], 18(5), pp.313–330. Available from: https://doi.org/10.1038/s41569-020-00477-1.

Junaid, A., Moon, M.C., Harding, G.E., and Zahradka, P., 2007. Osteopontin localizes to the nucleus of 293 cells and associates with polo-like kinase-1. *Am j physiol cell physiol* [Online], 292(2), pp.C919–26. Available from: https://doi.org/10.1152/ajpcell.00477.2006.

Kagiyama, S., Qian, K., Kagiyama, T., and Phillips, M.I., 2003. Antisense to epidermal growth factor receptor prevents the development of left ventricular hypertrophy. *Hypertension* [Online], 41(3 Pt 2), pp.824–9. Available from: https://doi.org/10.1161/01.HYP.0000047104.42047.9B.

Kalbfleisch, T.S., Hussien AbouEl Ela, N.A., Li, K., Brashear, W.A., Kochan, K.J., Hillhouse, A.E., Zhu, Y., Dhande, I.S., Kline, E.J., Hudson, E.A., Murphy, T.D., Thibaud-Nissen, F., Smith, M.L., and Doris, P.A., 2023. The assembled genome of the stroke-prone spontaneously hypertensive rat. *Hypertension* [Online], 80(1), pp.138–146. Available from: https://doi.org/10.1161/HYPERTENSIONAHA.122.20140.

Kanai, M., Akiyama, M., Takahashi, A., Matoba, N., Momozawa, Y., Ikeda, M., Iwata, N., Ikegawa, S., Hirata, M., Matsuda, K., Kubo, M., Okada, Y., and Kamatani, Y., 2018. Genetic analysis of quantitative traits in the japanese population links cell types to complex human diseases. *Nat genet* [Online], 50(3), pp.390–400. Available from: https://doi.org/10.1038/s41588-018-0047-6.

Kanitz, A., Gypas, F., Gruber, A.J., Gruber, A.R., Martin, G., and Zavolan, M., 2015. Comparative assessment of methods for the computational inference of transcript isoform abundance from rna-seq data. *Genome biol* [Online], 16(1), p.150. Available from: https://doi.org/10.1186/s13059-015-0702-5.

Kassambara, A., 2023. *Rstatix: pipe-friendly framework for basic statistical tests* [Online]. R package version 0.7.2. Available from: https://CRAN.R-project.org/package=rstatix.

Kawamura, K., Iyonaga, K., Ichiyasu, H., Nagano, J., Suga, M., and Sasaki, Y., 2005. Differentiation, maturation, and survival of dendritic cells by osteopontin regulation. *Clin diagn lab immunol* [Online], 12(1), pp.206–12. Available from: https://doi.org/10.1128/CDLI.12.1.206-212.2005.

Kerckhoffs, R.C., 2012. Computational modeling of cardiac growth in the post-natal rat with a strain-based growth law. *J biomech* [Online], 45(5), pp.865–71. Available from: https://doi.org/10.1016/j.jbiomech.2011.11.028.

Khurshid, S., Lazarte, J., Pirruccello, J.P., Weng, L.C., Choi, S.H., Hall, A.W., Wang, X., Friedman, S.F., Nauffal, V., Biddinger, K.J., Aragam, K.G., Batra, P., Ho, J.E., Philippakis, A.A., Ellinor, P.T., and Lubitz, S.A., 2023. Clinical and genetic associations of deep learning-derived cardiac magnetic resonance-based left ventricular mass. *Nat commun* [Online], 14(1), p.1558. Available from: https://doi.org/10.1038/s41467-023-37173-w.

Kim, D., Paggi, J.M., Park, C., and al., et, 2019. Graph-based genome alignment and genotyping with hisat2 and hisat-genotype. *Nat biotechnol* [Online], 37, pp.907–915. Available from: https://doi.org/10.1038/s41587-019-0201-4.

Kim, H.M., Hwang, I.C., Choi, H.M., Yoon, Y.E., and Cho, G.Y., 2022. Prognostic implication of left ventricular hypertrophy regression after antihypertensive therapy in patients with hypertension. *Front cardiovasc med* [Online], 9, p.1082008. Available from: https://doi.org/10.3389/fcvm.2022.1082008.

Kimes, B.W. and Brandt, B.L., 1976. Properties of a clonal muscle-cell line from rat-heart. *Experimental cell research* [Online], 98(2), pp.367–381. Available from: https://doi.org/https://doi.org/10.1016/0014-4827(76)90447-X.

Klingbeil, A.U., Schneider, M., Martus, P., Messerli, F.H., and Schmieder, R.E., 2003. A meta-analysis of the effects of treatment on left ventricular mass in essential hypertension. *Am j med* [Online], 115(1), pp.41–6. Available from: https://doi.org/10.1016/s0002-9343(03)00158-x.

Komatsubara, I., Murakami, T., Kusachi, S., Nakamura, K., Hirohata, S., Hayashi, J., Takemoto, S., Suezawa, C., Ninomiya, Y., and Shiratori, Y., 2003. Spatially and temporally different expression of osteonectin and osteopontin in the infarct zone of experimentally induced myocardial infarction in rats. *Cardiovasc pathol* [Online], 12(4), pp.186–94. Available from: https://doi.org/10.1016/s1054-8807(03)00042-5.

Konoshenko, M.Y., Lekchnov, E.A., Vlassov, A.V., and Laktionov, P.P., 2018. Isolation of extracellular vesicles: general methodologies and latest trends. *Biomed res int* [Online], 2018, p.8545347. Available from: https://doi.org/10.1155/2018/8545347.

Kramer, A., Green, J., Pollard J., J., and Tugendreich, S., 2014. Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics* [Online], 30(4), pp.523–30. Available from: https://doi.org/10.1093/bioinformatics/btt703.

Kramerova, I., Kumagai-Cresse, C., Ermolova, N., Mokhonova, E., Marinov, M., Capote, J., Becerra, D., Quattrocelli, M., Crosbie, R.H., Welch, E., McNally, E.M., and Spencer, M.J., 2019. Spp1 (osteopontin) promotes tgfbeta processing in fibroblasts of dystrophin-deficient muscles through matrix metalloproteinases. *Hum mol genet* [Online], 28(20), pp.3431–3442. Available from: https://doi.org/10.1093/hmg/ddz181.

Krause, S.W., Rehli, M., Kreutz, M., Schwarzfischer, L., Paulauskis, J.D., and Andreesen, R., 1996. Differential screening identifies genetic markers of monocyte to macrophage maturation. *J leukoc biol* [Online], 60(4), pp.540–5. Available from: https://doi.org/10.1002/jlb.60.4.540.

Kreutz, R., Struk, B., Stock, P., Hubner, N., Ganten, D., and Lindpaintner, K., 1997. Evidence for primary genetic determination of heart rate regulation: chromosomal mapping of a genetic locus in the rat. *Circulation* [Online], 96(4), pp.1078–81. Available from: https://doi.org/10.1161/01.cir.96.4.1078.

Krishnan, A., Samtani, R., Dhanantwari, P., Lee, E., Yamada, S., Shiota, K., Donofrio, M.T., Leatherbury, L., and Lo, C.W., 2014. A detailed comparison of mouse and human cardiac development. *Pediatr res* [Online], 76(6), pp.500–7. Available from: https://doi.org/10.1038/pr.2014.128.

Krumholz, H.M., Larson, M., and Levy, D., 1995. Prognosis of left ventricular geometric patterns in the framingham heart study. *J am coll cardiol* [Online], 25(4), pp.879–84. Available from: https://doi.org/10.1016/0735-1097(94)00473-4.

Lau, E., Cao, Q., Lam, M.P.Y., Wang, J., Ng, D.C.M., Bleakley, B.J., Lee, J.M., Liem, D.A., Wang, D., Hermjakob, H., and Ping, P., 2018. Integrated omics dissection of proteome dynamics during cardiac remodeling. *Nat commun* [Online], 9(1), p.120. Available from: https://doi.org/10.1038/s41467-017-02467-3.

Lee, H.G., Chen, Q., Wolfram, J.A., Richardson, S.L., Liner, A., Siedlak, S.L., Zhu, X., Ziats, N.P., Fujioka, H., Felsher, D.W., Castellani, R.J., Valencik, M.L., McDonald, J.A., Hoit, B.D., Lesnefsky, E.J., and Smith, M.A., 2009. Cell cycle re-entry and mitochondrial defects in myc-mediated hypertrophic cardiomyopathy and heart failure. *Plos one* [Online], 4(9), e7172. Available from: https://doi.org/10.1371/journal.pone.0007172.

Leen, L.L., Filipe, C., Billon, A., Garmy-Susini, B., Jalvy, S., Robbesyn, F., Daret, D., Allieres, C., Rittling, S.R., Werner, N., Nickenig, G., Deutsch, U., Duplaa, C., Dufourcq, P., Lenfant, F., Desgranges, C., Arnal, J.F., and Gadeau, A.P., 2008. Estrogen-stimulated endothelial repair requires osteopontin. *Arterioscler thromb vasc biol* [Online], 28(12), pp.2131–6. Available from: https://doi.org/10.1161/ATVBAHA.108.167965.

Leibowitz, D., Bursztyn, M., Jacobs, J.M., Ein-Mor, E., and Stessman, J., 2010. High prevalence of left ventricular hypertrophy in octogenarian women: the jerusalem longitudinal cohort study. *Blood press* [Online], 19(2), pp.86–91. Available from: https://doi.org/10.3109/08037050903516292.

Lenga, Y., Koh, A., Perera, A.S., McCulloch, C.A., Sodek, J., and Zohar, R., 2008. Osteopontin expression is required for myofibroblast differentiation. *Circ res* [Online], 102(3), pp.319–27. Available from: https://doi.org/10.1161/CIRCRESAHA.107.160408.

Lenoir, M., Kufareva, I., Abagyan, R., and Overduin, M., 2015. Membrane and protein interactions of the pleckstrin homology domain superfamily. *Membranes (basel)* [Online], 5(4), pp.646–63. Available from: https://doi.org/10.3390/membranes5040646.

Leong, X.F., Ng, C.Y., and Jaarin, K., 2015. Animal models in cardiovascular research: hypertension and atherosclerosis. *Biomed res int* [Online], 2015, p.528757. Available from: https://doi.org/10.1155/2015/528757.

Lewis, A.A., Ayers, C.R., Selvin, E., Neeland, I., Ballantyne, C.M., Nambi, V., Pandey, A., Powell-Wiley, T.M., Drazner, M.H., Carnethon, M.R., Berry, J.D., Seliger, S.L., DeFilippi, C.R., and Lemos, J.A. de, 2020. Racial differences in malignant left ventricular hypertrophy and incidence of heart failure: a multicohort study. *Circulation* [Online], 141(12), pp.957–967. Available from: https://doi.org/10.1161/CIRCULATIONAHA.119.043628.

Li, F., Wang, X., Capasso, J.M., and Gerdes, A.M., 1996. Rapid transition of cardiac myocytes from hyperplasia to hypertrophy during postnatal development. *J mol cell cardiol* [Online], 28(8), pp.1737–46. Available from: https://doi.org/10.1006/jmcc.1996.0163.

Li, S. and Jakobs, T.C., 2022. Secreted phosphoprotein 1 slows neurodegeneration and rescues visual function in mouse models of aging and glaucoma. *Cell rep* [Online], 41(13), p.111880. Available from: https://doi.org/10.1016/j.celrep.2022.111880.

Liang, L., Lu, G., Pan, G., Deng, Y., Liang, J., Liang, L., Liu, J., Tang, Y., and Wei, G., 2019. A case-control study of the association between the spp1 gene snps and the susceptibility to breast cancer in guangxi, china. *Front oncol* [Online], 9, p.1415. Available from: https://doi.org/10.3389/fonc.2019.01415.

Liao, Y., Cooper, R.S., Mensah, G.A., and McGee, D.L., 1995. Left ventricular hypertrophy has a greater impact on survival in women than in men. *Circulation* [Online], 92(4), pp.805–810. Available from: https://doi.org/10.1161/01.CIR.92.4.805.

Liaw, L., Birk, D.E., Ballas, C.B., Whitsitt, J.S., Davidson, J.M., and Hogan, B.L., 1998. Altered wound healing in mice lacking a functional osteopontin gene (spp1). *J clin invest* [Online], 101(7), pp.1468–78. Available from: https://doi.org/10.1172/JCI2131.

Libiseller-Egger, J., Phelan, J.E., Attia, Z.I., Benavente, E.D., Campino, S., Friedman, P.A., Lopez-Jimenez, F., Leon, D.A., and Clark, T.G., 2022. Deep learning-derived cardiovascular age shares a genetic basis with other cardiac phenotypes. *Sci rep* [Online], 12(1), p.22625. Available from: https://doi.org/10.1038/s41598-022-27254-z.

Lin, H., Castro-Diehl, C., Short, M.I., Xanthakis, V., Yola, I.M., Kwan, A.C., Mitchell, G.F., Larson, M.G., Vasan, R.S., and Cheng, S., 2021. Shared genetic and environmental architecture of cardiac phenotypes assessed via echocardiography: the framingham heart study. *Circ genom precis med* [Online], 14(2), e003244. Available from: https://doi.org/10.1161/CIRCGEN.120.003244.

Litvinukova, 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., Seidman, C.E., Noseda, M., Hubner, N., and Teichmann, S.A., 2020. Cells of the adult human heart. *Nature* [Online], 588(7838), pp.466–472. Available from: https://doi.org/10.1038/s41586-020-2797-4.

Liu, J., Liu, X., Hui, X., Cai, L., Li, X., Yang, Y., Shu, S., Wang, F., Xia, H., and Li, S., 2019. Novel role for pleckstrin homology-like domain family a, member 3 in the regulation of pathological cardiac hypertrophy. *J am heart assoc* [Online], 8(16), e011830. Available from: https://doi.org/10.1161/JAHA.118.011830.

Liu, S. and Martin, J.F., 2022. Rna splicing to cytoskeleton: a new path to cardiomyocyte ploidy and division? *Dev cell* [Online], 57(8), pp.945–946. Available from: https://doi.org/10.1016/j.devcel.2022.04.002.

Liu, X., Sun, Y., Li, H., Li, Y., Li, M., Yuan, Y., Cui, S., and Yao, D., 2017. Effect of spp1 on nerve degeneration and regeneration after rat sciatic nerve injury. *Bmc neurosci* [Online], 18(1), p.30. Available from: https://doi.org/10.1186/s12868-017-0348-1.

Liu, X., Zhao, J., Xue, L., Zhao, T., Ding, W., Han, Y., and Ye, H., 2022. A comparison of transcriptome analysis methods with reference genome. *Bmc genomics* [Online], 23(1), p.232. Available from: https://doi.org/10.1186/s12864-022-08465-0.

Lockhart, M., Wirrig, E., Phelps, A., and Wessels, A., 2011. Extracellular matrix and heart development. *Birth defects res a clin mol teratol* [Online], 91(6), pp.535–50. Available from: https://doi.org/10.1002/bdra.20810.

Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., Foster, B., Moser, M., Karasik, E., Gillard, B., Ramsey, K., Sullivan, S., Bridge, J., Magazine, H., Syron, J., Fleming, J., Siminoff, L., Traino, H., Mosavel, M., Barker, L., Jewell, S., Rohrer, D., Maxim, D., Filkins, D., Harbach, P., Cortadillo, E., Berghuis, B., Turner, L., Hudson, E., Feenstra, K., Sobin, L., Robb, J., Branton, P., Korzeniewski, G., Shive, C., Tabor, D., Qi, L., Groch, K., Nampally, S., Buia, S., Zimmerman, A., Smith, A., Burges, R., Robinson, K., Valentino, K., Bradbury, D., Cosentino, M., Diaz-Mayoral, N., Kennedy, M., Engel, T., Williams, P., Erickson, K., Ardlie, K., Winckler, W., Getz, G., DeLuca, D., MacArthur, D., Kellis, M., Thomson, A., Young, T., Gelfand, E., Donovan, M., Meng, Y., Grant, G., Mash, D., Marcus, Y., Basile, M., Liu, J., Zhu, J., Tu, Z., Cox, N.J., Nicolae, D.L., Gamazon, E.R., Im, H.K., Konkashbaev, A., Pritchard, J., Stevens, M., Flutre, T., Wen, X., Dermitzakis, E.T., Lappalainen, T., Guigo, R., Monlong, J., Sammeth, M., Koller, D., Battle, A., Mostafavi, S., McCarthy, M., Rivas, M., Maller, J., Rusyn, I., Nobel, A., Wright, F., Shabalin, A., Feolo, M., Sharopova, N., et al., 2013. The genotype-tissue expression (gtex) project. *Nature genetics* [Online], 45(6), pp.580–585. Available from: `https://doi.org/10.1038/ng.2653`.

Lopes, L.R., Aung, N., Duijvenboden, S. van, Munroe, P.B., Elliott, P.M., and Petersen, S.E., 2021. Prevalence of hypertrophic cardiomyopathy in the uk biobank population. *Jama cardiol* [Online], 6(7), pp.852–854. Available from: `https://doi.org/10.1001/jamacardio.2021.0689`.

Lopez, B., Gonzalez, A., Lindner, D., Westermann, D., Ravassa, S., Beaumont, J., Gallego, I., Zudaire, A., Brugnolaro, C., Querejeta, R., Larman, M., Tschope, C., and Diez, J., 2013. Osteopontin-mediated myocardial fibrosis in heart failure: a role for lysyl oxidase? *Cardiovasc res* [Online], 99(1), pp.111–20. Available from: `https://doi.org/10.1093/cvr/cvt100`.

Lorenzen, J.M., Schauerte, C., Hübner, A., Kölling, M., Martino, F., Scherf, K., Batkai, S., Zimmer, K., Foinquinos, A., Kaucsar, T., Fiedler, J., Kumarswamy, R., Bang, C., Hartmann, D., Gupta, S.K., Kielstein, J., Jungmann, A., Katus, H.A., Weidemann, F., Müller, O.J., Haller, H., and Thum, T., 2015. Osteopontin is indispensible for ap1-mediated angiotensin ii-related mir-21 transcription during cardiac fibrosis. *European heart journal* [Online], 36(32), pp.2184–2196. Available from: `https://doi.org/10.1093/eurheartj/ehv109`.

Love, M.I., Huber, W., and Anders, S., 2014. Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome biology* [Online], 15. Available from: `https://doi.org/10.1186/s13059-014-0550-8`.

Lovisa, F., Garbin, A., Crotti, S., Di Battista, P., Gallingani, I., Damanti, C.C., Tosato, A., Carraro, E., Pillon, M., Mafakheri, E., Romanato, F., Gaffo, E., Biffi, A., Bortoluzzi,

S., Agostini, M., and Mussolin, L., 2021. Increased tenascin c, osteopontin and hsp90 levels in plasmatic small extracellular vesicles of pediatric alk-positive anaplastic large cell lymphoma: new prognostic biomarkers? *Diagnostics (basel)* [Online], 11(2). Available from: https://doi.org/10.3390/diagnostics11020253.

Lu, J., Liu, J., Guo, Y., Zhang, Y., Xu, Y., and Wang, X., 2021. Crispr-cas9: a method for establishing rat models of drug metabolism and pharmacokinetics. *Acta pharm sin b* [Online], 11(10), pp.2973–2982. Available from: https://doi.org/10.1016/j.apsb.2021.01.007.

Lyu, L., Wang, H., Li, B., Qin, Q., Qi, L., Nagarkatti, M., Nagarkatti, P., Janicki, J.S., Wang, X.L., and Cui, T., 2015. A critical role of cardiac fibroblast-derived exosomes in activating renin angiotensin system in cardiomyocytes. *J mol cell cardiol* [Online], 89(Pt B), pp.268–79. Available from: https://doi.org/10.1016/j.yjmcc.2015.10.022.

Maatz, H., Jens, M., Liss, M., Schafer, S., Heinig, M., Kirchner, M., Adami, E., Rintisch, C., Dauksaite, V., Radke, M.H., Selbach, M., Barton, P.J., Cook, S.A., Rajewsky, N., Gotthardt, M., Landthaler, M., and Hubner, N., 2014. Rna-binding protein rbm20 represses splicing to orchestrate cardiac pre-mrna processing. *J clin invest* [Online], 124(8), pp.3419–30. Available from: https://doi.org/10.1172/JCI74523.

Magno, R. and Maia, A.-T., 2019. Gwasrapidd: an r package to query, download and wrangle gwas catalog data. *Bioinformatics* [Online], pp.1–2. Available from: https://doi.org/10.1093/bioinformatics/btz605.

Maillet, M., Berlo, J.H. van, and Molkentin, J.D., 2013. Molecular basis of physiological heart growth: fundamental concepts and new players. *Nat rev mol cell biol* [Online], 14(1), pp.38–48. Available from: https://doi.org/10.1038/nrm3495.

Mamazhakypov, A., Sartmyrzaeva, M., Sarybaev, A.S., Schermuly, R., and Sydykov, A., 2022. Clinical and molecular implications of osteopontin in heart failure. *Curr issues mol biol* [Online], 44(8), pp.3573–3597. Available from: https://doi.org/10.3390/cimb44080245.

Manabe, I., Shindo, T., and Nagai, R., 2002. Gene expression in fibroblasts and fibrosis: involvement in cardiac hypertrophy. *Circ res* [Online], 91(12), pp.1103–13. Available from: https://doi.org/10.1161/01.res.0000046452.67724.b8.

Mancia, G., Facchetti, R., Vanoli, J., Dell'Oro, R., Seravalle, G., and Grassi, G., 2022. White-coat hypertension without organ damage: impact on long-term mortality, new hypertension, and new organ damage. *Hypertension* [Online], 79(5), pp.1057–1066. Available from: https://doi.org/10.1161/HYPERTENSIONAHA.121.18792.

Mancia, G., Kreutz, R., Brunstrom, M., Burnier, M., Grassi, G., Januszewicz, A., Muiesan, M.L., Tsioufis, K., Agabiti-Rosei, E., Algharably, E.A.E., Azizi, M., Benetos, A., Borghi, C., Hitij, J.B., Cifkova, R., Coca, A., Cornelissen, V., Cruickshank, J.K., Cunha, P.G., Danser, A.H.J., Pinho, R.M., Delles, C., Dominiczak, A.F., Dorobantu, M., Doumas, M., Fernandez-Alfonso, M.S., Halimi, J.M., Jarai, Z., Jelakovic, B., Jordan, J., Kuznetsova, T., Laurent, S., Lovic, D., Lurbe, E., Mahfoud, F., Manolis, A., Miglinas, M., Narkiewicz, K., Niiranen, T., Palatini, P., Parati, G., Pathak, A., Persu, A., Polonia, J., Redon, J., Sarafidis, P., Schmieder, R., Spronck, B., Stabouli, S., Stergiou, G., Taddei, S., Thomopoulos, C., Tomaszewski, M., Van de Borne, P., Wanner, C., Weber, T., Williams, B., Zhang, Z.Y., and Kjeldsen, S.E., 2023. 2023 esh guidelines for the management of arterial hypertension the task force for the management of arterial hypertension of the european society of hypertension: endorsed by the international society of hypertension (ish) and the european renal association (era). *J hypertens* [Online], 41(12), pp.1874–2071. Available from: https://doi.org/10.1097/HJH.0000000000003480.

Many, G.M., Yokosaki, Y., Uaesoontrachoon, K., Nghiem, P.P., Bello, L., Dadgar, S., Yin, Y., Damsker, J.M., Cohen, H.B., Kornegay, J.N., Bamman, M.M., Mosser, D.M., Nagaraju, K., and Hoffman, E.P., 2016. Opn-a induces muscle inflammation by increasing recruitment and activation of pro-inflammatory macrophages. *Exp physiol* [Online], 101(10), pp.1285–1300. Available from: https://doi.org/10.1113/EP085768.

Mao, Y. and Zhang, G., 2022. A complete, telomere-to-telomere human genome sequence presents new opportunities for evolutionary genomics. *Nat methods* [Online], 19(6), pp.635–638. Available from: https://doi.org/10.1038/s41592-022-01512-4.

Marcela, S.G., Cristina, R.M., Angel, P.G., Manuel, A.M., Sofia, D.C., Patricia de, L.R., Bladimir, R.R., and Concepcion, S.G., 2012. Chronological and morphological study of heart development in the rat. *Anat rec (hoboken)* [Online], 295(8), pp.1267–90. Available from: https://doi.org/10.1002/ar.22508.

Mardis, E.R., 2017. Dna sequencing technologies: 2006-2016. *Nat protoc* [Online], 12(2), pp.213–218. Available from: https://doi.org/10.1038/nprot.2016.182.

Maron, B.J. and Maron, M.S., 2013. Hypertrophic cardiomyopathy. *Lancet* [Online], 381(9862), pp.242–55. Available from: https://doi.org/10.1016/S0140-6736(12)60397-3.

Marques, F.Z., Booth, S.A., Prestes, P.R., Curl, C.L., Delbridge, L.M., Lewandowski, P., Harrap, S.B., and Charchar, F.J., 2016. Telomere dynamics during aging in polygenic left ventricular hypertrophy. *Physiol genomics* [Online], 48(1), pp.42–9. Available from: https://doi.org/10.1152/physiolgenomics.00083.2015.

Marques-Coelho, D., Iohan, L., Melo de Farias, A.R., Flaig, A., Brainbank Neuro, C.E.B.N.N., Lambert, J.C., and Costa, M.R., 2021. Differential transcript usage unravels gene expression alterations in alzheimer's disease human brains. *Npj aging mech dis* [Online], 7(1), p.2. Available from: https://doi.org/10.1038/s41514-020-00052-5.

Marsh, S.A., Powell, P.C., Agarwal, A., Dell'Italia, L.J., and Chatham, J.C., 2007. Cardiovascular dysfunction in zucker obese and zucker diabetic fatty rats: role of hydronephrosis. *Am j physiol heart circ physiol* [Online], 293(1), H292–8. Available from: https://doi.org/10.1152/ajpheart.01362.2006.

Masenga, S.K. and Kirabo, A., 2023. Hypertensive heart disease: risk factors, complications and mechanisms. *Front cardiovasc med* [Online], 10, p.1205475. Available from: https://doi.org/10.3389/fcvm.2023.1205475.

Mateescu, B., Kowal, E.J., Balkom, B.W. van, Bartel, S., Bhattacharyya, S.N., Buzas, E.I., Buck, A.H., Candia, P. de, Chow, F.W., Das, S., Driedonks, T.A., Fernandez-Messina, L., Haderk, F., Hill, A.F., Jones, J.C., Van Keuren-Jensen, K.R., Lai, C.P., Lasser, C., Liegro, I.D., Lunavat, T.R., Lorenowicz, M.J., Maas, S.L., Mager, I., Mittelbrunn, M., Momma, S., Mukherjee, K., Nawaz, M., Pegtel, D.M., Pfaffl, M.W., Schiffelers, R.M., Tahara, H., Thery, C., Tosar, J.P., Wauben, M.H., Witwer, K.W., and Nolte-'t Hoen, E.N., 2017. Obstacles and opportunities in the functional analysis of extracellular vesicle rna - an isev position paper. *J extracell vesicles* [Online], 6(1), p.1286095. Available from: https://doi.org/10.1080/20013078.2017.1286095.

Mazzarotto, F., Olivotto, I., Boschi, B., Girolami, F., Poggesi, C., Barton, P.J.R., and Walsh, R., 2020. Contemporary insights into the genetics of hypertrophic cardiomyopathy: toward a new era in clinical testing? *J am heart assoc* [Online], 9(8), e015473. Available from: https://doi.org/10.1161/JAHA.119.015473.

McBride, M.W., Brosnan, M.J., Mathers, J., McLellan, L.I., Miller, W.H., Graham, D., Hanlon, N., Hamilton, C.A., Polke, J.M., Lee, W.K., and Dominiczak, A.F., 2005. Reduction of gstm1 expression in the stroke-prone spontaneously hypertension rat contributes to increased oxidative stress. *Hypertension* [Online], 45(4), pp.786–92. Available from: https://doi.org/10.1161/01.HYP.0000154879.49245.39.

McBride, M.W., Carr, F.J., Graham, D., Anderson, N.H., Clark, J.S., Lee, W.K., Charchar, F.J., Brosnan, M.J., and Dominiczak, A.F., 2003. Microarray analysis of rat chromosome 2 congenic strains. *Hypertension* [Online], 41(3 Pt 2), pp.847–53. Available from: https://doi.org/10.1161/01.HYP.0000047103.07205.03.

McBride, M.W., Charchar, F.J., Graham, D., Miller, W.H., Strahorn, P., Carr, F.J., and Dominiczak, A.F., 2004. Functional genomics in rodent models of hypertension. *J*

*physiol* [Online], 554(Pt 1), pp.56–63. Available from: https://doi.org/10.1113/jphysiol.2003.049361.

McKenna, W.J. and Judge, D.P., 2021. Epidemiology of the inherited cardiomyopathies. *Nat rev cardiol* [Online], 18(1), pp.22–36. Available from: https://doi.org/10.1038/s41569-020-0428-2.

McLellan, M.A., Skelly, D.A., Dona, M.S.I., Squiers, G.T., Farrugia, G.E., Gaynor, T.L., Cohen, C.D., Pandey, R., Diep, H., Vinh, A., Rosenthal, N.A., and Pinto, A.R., 2020. High-resolution transcriptomic profiling of the heart during chronic stress reveals cellular drivers of cardiac fibrosis and hypertrophy. *Circulation* [Online], 142(15), pp.1448–1463. Available from: https://doi.org/10.1161/CIRCULATIONAHA.119.045115.

Meder, B., Haas, J., Sedaghat-Hamedani, F., Kayvanpour, E., Frese, K., Lai, A., Nietsch, R., Scheiner, C., Mester, S., Bordalo, D.M., Amr, A., Dietrich, C., Pils, D., Siede, D., Hund, H., Bauer, A., Holzer, D.B., Ruhparwar, A., Mueller-Hennessen, M., Weichenhan, D., Plass, C., Weis, T., Backs, J., Wuerstle, M., Keller, A., Katus, H.A., and Posch, A.E., 2017. Epigenome-wide association study identifies cardiac gene patterning and a novel class of biomarkers for heart failure. *Circulation* [Online], 136(16), pp.1528–1544. Available from: https://doi.org/10.1161/CIRCULATIONAHA.117.027355.

Michelhaugh, S.A., Camacho, A., Ibrahim, N.E., Gaggin, H., D'Alessandro, D., Coglianese, E., Lewis, G.D., and Januzzi J. L., J., 2020. Proteomic signatures during treatment in different stages of heart failure. *Circ heart fail* [Online], 13(8), e006794. Available from: https://doi.org/10.1161/CIRCHEARTFAILURE.119.006794.

Miga, K.H., Koren, S., Rhie, A., Vollger, M.R., Gershman, A., Bzikadze, A., Brooks, S., Howe, E., Porubsky, D., Logsdon, G.A., Schneider, V.A., Potapova, T., Wood, J., Chow, W., Armstrong, J., Fredrickson, J., Pak, E., Tigyi, K., Kremitzki, M., Markovic, C., Maduro, V., Dutra, A., Bouffard, G.G., Chang, A.M., Hansen, N.F., Wilfert, A.B., Thibaud-Nissen, F., Schmitt, A.D., Belton, J.M., Selvaraj, S., Dennis, M.Y., Soto, D.C., Sahasrabudhe, R., Kaya, G., Quick, J., Loman, N.J., Holmes, N., Loose, M., Surti, U., Risques, R.A., Graves Lindsay, T.A., Fulton, R., Hall, I., Paten, B., Howe, K., Timp, W., Young, A., Mullikin, J.C., Pevzner, P.A., Gerton, J.L., Sullivan, B.A., Eichler, E.E., and Phillippy, A.M., 2020. Telomere-to-telomere assembly of a complete human x chromosome. *Nature* [Online], 585(7823), pp.79–84. Available from: https://doi.org/10.1038/s41586-020-2547-7.

Mihl, C., Dassen, W.R., and Kuipers, H., 2008. Cardiac remodelling: concentric versus eccentric hypertrophy in strength and endurance athletes. *Neth heart j* [Online], 16(4), pp.129–33. Available from: https://doi.org/10.1007/bf03086131.

Mirza, M., Shaughnessy, E., Hurley, J.K., Vanpatten, K.A., Pestano, G.A., He, B., and Weber, G.F., 2008. Osteopontin-c is a selective marker of breast cancer. *Int j cancer* [Online], 122(4), pp.889–97. Available from: https://doi.org/10.1002/ijc.23204.

Mohamed, I.A., Gadeau, A.P., Fliegel, L., Lopaschuk, G., Mlih, M., Abdulrahman, N., Fillmore, N., and Mraiche, F., 2015. Na+/h+ exchanger isoform 1-induced osteopontin expression facilitates cardiomyocyte hypertrophy. *Plos one* [Online], 10(4), e0123318. Available from: https://doi.org/10.1371/journal.pone.0123318.

Mohamed, I.A. and Mraiche, F., 2015. Targeting osteopontin, the silent partner of na+/h+ exchanger isoform 1 in cardiac remodeling. *J cell physiol* [Online], 230(9), pp.2006–18. Available from: https://doi.org/10.1002/jcp.24958.

Molkentin, J.D., Lu, J.R., Antos, C.L., Markham, B., Richardson, J., Robbins, J., Grant, S.R., and Olson, E.N., 1998. A calcineurin-dependent transcriptional pathway for cardiac hypertrophy. *Cell* [Online], 93(2), pp.215–28. Available from: https://doi.org/10.1016/s0092-8674(00)81573-1.

Monserrat, L., Lopez, B., Gonzalez, A., Hermida, M., Fernandez, X., Ortiz, M., Barriales-Villa, R., Castro-Beiras, A., and Diez, J., 2011. Cardiotrophin-1 plasma levels are associated with the severity of hypertrophy in hypertrophic cardiomyopathy. *Eur heart j* [Online], 32(2), pp.177–83. Available from: https://doi.org/10.1093/eurheartj/ehq400.

Morel, L., Yu, Y., Blenman, K.R., Caldwell, R.A., and Wakeland, E.K., 1996. Production of congenic mouse strains carrying genomic intervals containing sle-susceptibility genes derived from the sle-prone nzm2410 strain. *Mamm genome* [Online], 7(5), pp.335–9. Available from: https://doi.org/10.1007/s003359900098.

Moreno-Hagelsieb, G., Trevino, V., Perez-Rueda, E., Smith, T.F., and Collado-Vides, J., 2001. Transcription unit conservation in the three domains of life: a perspective from escherichia coli. *Trends genet* [Online], 17(4), pp.175–7. Available from: https://doi.org/10.1016/s0168-9525(01)02241-7.

Morgan, H., 2018. *Pregnancy in the stroke-prone spontaneously hypertensive rat: investigating impaired vascular remodelling and establishing a novel model of superimposed pre-eclampsia*. University of Glasgow.

Mosley, J.D., Levinson, R.T., Farber-Eger, E., Edwards, T.L., Hellwege, J.N., Hung, A.M., Giri, A., Shuey, M.M., Shaffer, C.M., Shi, M., Brittain, E.L., Chung, W.K., Kullo, I.J., Arruda-Olson, A.M., Jarvik, G.P., Larson, E.B., Crosslin, D.R., Williams, M.S., Borthwick, K.M., Hakonarson, H., Denny, J.C., Wang, T.J., Stein, C.M., Roden, D.M., and Wells, Q.S., 2020. The polygenic architecture of left ventricular mass mirrors the clinical epidemiology. *Sci rep* [Online], 10(1), p.7561. Available from: https://doi.org/10.1038/s41598-020-64525-z.

Mosterd, A., D'Agostino, R.B., Silbershatz, H., Sytkowski, P.A., Kannel, W.B., Grobbee, D.E., and Levy, D., 1999. Trends in the prevalence of hypertension, antihypertensive therapy, and left ventricular hypertrophy from 1950 to 1989. *N engl j med* [Online], 340(16), pp.1221–7. Available from: https://doi.org/10.1056/NEJM199904223401601.

Mott, R., Talbot, C.J., Turri, M.G., Collins, A.C., and Flint, J., 2000. A method for fine mapping quantitative trait loci in outbred animal stocks. *Proc natl acad sci u s a* [Online], 97(23), pp.12649–54. Available from: https://doi.org/10.1073/pnas.230304397.

Mou, H., Smith, J.L., Peng, L., Yin, H., Moore, J., Zhang, X.O., Song, C.Q., Sheel, A., Wu, Q., Ozata, D.M., Li, Y., Anderson, D.G., Emerson, C.P., Sontheimer, E.J., Moore, M.J., Weng, Z., and Xue, W., 2017. Crispr/cas9-mediated genome editing induces exon skipping by alternative splicing or exon deletion. *Genome biol* [Online], 18(1), p.108. Available from: https://doi.org/10.1186/s13059-017-1237-8.

Mutlak, M. and Kehat, I., 2015. Extracellular signal-regulated kinases 1/2 as regulators of cardiac hypertrophy. *Front pharmacol* [Online], 6, p.149. Available from: https://doi.org/10.3389/fphar.2015.00149.

Nabika, T., Ohara, H., Kato, N., and Isomura, M., 2012. The stroke-prone spontaneously hypertensive rat: still a useful model for post-gwas genetic studies? *Hypertens res* [Online], 35(5), pp.477–84. Available from: https://doi.org/10.1038/hr.2012.30.

Nakada, Y., Nhi Nguyen, N.U., Xiao, F., Savla, J.J., Lam, N.T., Abdisalaam, S., Bhattacharya, S., Mukherjee, S., Asaithamby, A., Gillette, T.G., Hill, J.A., and Sadek, H.A., 2019. Dna damage response mediates pressure overload-induced cardiomyocyte hypertrophy. *Circulation* [Online], 139(9), pp.1237–1239. Available from: https://doi.org/10.1161/CIRCULATIONAHA.118.034822.

Nakamura, M. and Sadoshima, J., 2018. Mechanisms of physiological and pathological cardiac hypertrophy. *Nat rev cardiol* [Online], 15(7), pp.387–407. Available from: https://doi.org/10.1038/s41569-018-0007-y.

Nakayama, H., Nagai, H., Matsumoto, K., Oguro, R., Sugimoto, K., Kamide, K., Ohishi, M., Katsuya, T., Okamoto, H., Maeda, M., Komamura, K., Azuma, J., Rakugi, H., and Fujio, Y., 2011. Association between osteopontin promoter variants and diastolic dysfunction in hypertensive heart in the japanese population. *Hypertens res* [Online], 34(10), pp.1141–6. Available from: https://doi.org/10.1038/hr.2011.102.

Naqvi, S., Godfrey, A.K., Hughes, J.F., Goodheart, M.L., Mitchell, R.N., and Page, D.C., 2019. Conservation, acquisition, and functional impact of sex-biased gene expression in mammals. *Science* [Online], 365(6450). Available from: https://doi.org/10.1126/science.aaw7317.

Nethononda, R.M., McGurk, K.A., Whitworth, P., Francis, J., Mamasoula, C., Cordell, H.J., Neubauer, S., Keavney, B.D., Mayosi, B.M., Farrall, M., and Watkins, H., 2019. Marked variation in heritability estimates of left ventricular mass depending on modality of measurement. *Sci rep* [Online], 9(1), p.13556. Available from: https://doi.org/10.1038/s41598-019-49961-w.

Neubauer, S., 2007. The failing heart–an engine out of fuel. *N engl j med* [Online], 356(11), pp.1140–51. Available from: https://doi.org/10.1056/NEJMra063052.

Nicin, L., Wagner, J.U.G., Luxan, G., and Dimmeler, S., 2022. Fibroblast-mediated intercellular crosstalk in the healthy and diseased heart. *Febs lett* [Online], 596(5), pp.638–654. Available from: https://doi.org/10.1002/1873-3468.14234.

Niel, G. van, Carter, D.R.F., Clayton, A., Lambert, D.W., Raposo, G., and Vader, P., 2022. Challenges and directions in studying cell-cell communication by extracellular vesicles. *Nat rev mol cell biol* [Online]. Available from: https://doi.org/10.1038/s41580-022-00460-3.

Nielsen, J.B., Fritsche, L.G., Zhou, W., Teslovich, T.M., Holmen, O.L., Gustafsson, S., Gabrielsen, M.E., Schmidt, E.M., Beaumont, R., Wolford, B.N., Lin, M., Brummett, C.M., Preuss, M.H., Refsgaard, L., Bottinger, E.P., Graham, S.E., Surakka, I., Chu, Y., Skogholt, A.H., Dalen, H., Boyle, A.P., Oral, H., Herron, T.J., Kitzman, J., Jalife, J., Svendsen, J.H., Olesen, M.S., Njølstad, I., Løchen, M.L., Baras, A., Gottesman, O., Marcketta, A., O'Dushlaine, C., Ritchie, M.D., Wilsgaard, T., Loos, R.J.F., Frayling, T.M., Boehnke, M., Ingelsson, E., Carey, D.J., Dewey, F.E., Kang, H.M., Abecasis, G.R., Hveem, K., and Willer, C.J., 2018. Genome-wide study of atrial fibrillation identifies seven risk loci and highlights biological pathways and regulatory elements involved in cardiac development. *Am j hum genet* [Online], 102(1), pp.103–115. Available from: https://doi.org/10.1016/j.ajhg.2017.12.003.

Nigam, R., Munzenmaier, D.H., Worthey, E.A., Dwinell, M.R., Shimoyama, M., and Jacob, H.J., 2013. Rat strain ontology: structured controlled vocabulary designed to facilitate access to strain data at rgd. *J biomed semantics* [Online], 4(1), p.36. Available from: https://doi.org/10.1186/2041-1480-4-36.

Nijman, I.J., Kuipers, S., Verheul, M., Guryev, V., and Cuppen, E., 2008. A genome-wide snp panel for mapping and association studies in the rat. *Bmc genomics* [Online], 9, p.95. Available from: https://doi.org/10.1186/1471-2164-9-95.

Nikfarjam, S. and Singh, K.K., 2023. Dna damage response signaling: a common link between cancer and cardiovascular diseases. *Cancer med* [Online], 12(4), pp.4380–4404. Available from: https://doi.org/10.1002/cam4.5274.

Nomura, S., Satoh, M., Fujita, T., Higo, T., Sumida, T., Ko, T., Yamaguchi, T., Tobita, T., Naito, A.T., Ito, M., Fujita, K., Harada, M., Toko, H., Kobayashi, Y., Ito, K., Takimoto, E., Akazawa, H., Morita, H., Aburatani, H., and Komuro, I., 2018. Cardiomyocyte gene programs encoding morphological and functional signatures in cardiac hypertrophy and failure. *Nat commun* [Online], 9(1), p.4435. Available from: https://doi.org/10.1038/s41467-018-06639-7.

Nowicka, M. and Robinson, M.D., 2016. Drimseq: a dirichlet-multinomial framework for multivariate count outcomes in genomics. *F1000research*, 5. [version 2; referees: 2 approved], p.1356.

Oh, G.C. and Cho, H.J., 2020. Blood pressure and heart failure. *Clin hypertens* [Online], 26, p.1. Available from: https://doi.org/10.1186/s40885-019-0132-x.

Ohba, T., Watanabe, H., Murakami, M., Iino, K., Adachi, T., Baba, Y., Kurosaki, T., Ono, K., and Ito, H., 2017. Stromal interaction molecule 1 haploinsufficiency causes maladaptive response to pressure overload. *Plos one* [Online], 12(11), e0187950. Available from: https://doi.org/10.1371/journal.pone.0187950.

Okamoto, H. and Imanaka-Yoshida, K., 2012. Matricellular proteins: new molecular targets to prevent heart failure. *Cardiovasc ther* [Online], 30(4), e198–209. Available from: https://doi.org/10.1111/j.1755-5922.2011.00276.x.

Olah, A., Matyas, C., Kellermayer, D., Ruppert, M., Barta, B.A., Sayour, A.A., Torok, M., Koncsos, G., Giricz, Z., Ferdinandy, P., Merkely, B., and Radovits, T., 2019. Sex differences in morphological and functional aspects of exercise-induced cardiac hypertrophy in a rat model. *Front physiol* [Online], 10, p.889. Available from: https://doi.org/10.3389/fphys.2019.00889.

Oldfield, C.J., Duhamel, T.A., and Dhalla, N.S., 2020. Mechanisms for the transition from physiological to pathological cardiac hypertrophy. *Can j physiol pharmacol* [Online], 98(2), pp.74–84. Available from: https://doi.org/10.1139/cjpp-2019-0566.

Olivera, S. and Graham, D., 2023. Sex differences in preclinical models of hypertension. *J hum hypertens* [Online], 37(8), pp.619–625. Available from: https://doi.org/10.1038/s41371-022-00770-1.

Pagel, C.N., Wasgewatte Wijesinghe, D.K., Taghavi Esfandouni, N., and Mackie, E.J., 2014. Osteopontin, inflammation and myogenesis: influencing regeneration, fibrosis and size of skeletal muscle. *J cell commun signal* [Online], 8(2), pp.95–103. Available from: https://doi.org/10.1007/s12079-013-0217-3.

Paik, D.T., Cho, S., Tian, L., Chang, H.Y., and Wu, J.C., 2020. Single-cell rna sequencing in cardiovascular development, disease and medicine. *Nat rev cardiol* [Online], 17(8), pp.457–473. Available from: https://doi.org/10.1038/s41569-020-0359-y.

Parikh, V.N. and Ashley, E.A., 2017. Next-generation sequencing in cardiovascular disease: present clinical applications and the horizon of precision medicine. *Circulation* [Online], 135(5), pp.406–409. Available from: https://doi.org/10.1161/CIRCULATIONAHA.116.024258.

Pathan, M., Fonseka, P., Chitti, S.V., Kang, T., Sanwlani, R., Van Deun, J., Hendrix, A., and Mathivanan, S., 2019. Vesiclepedia 2019: a compendium of rna, proteins, lipids and metabolites in extracellular vesicles. *Nucleic acids res* [Online], 47(D1), pp.D516–D519. Available from: https://doi.org/10.1093/nar/gky1029.

Peacock, J.D., Huk, D.J., Ediriweera, H.N., and Lincoln, J., 2011. Sox9 transcriptionally represses spp1 to prevent matrix mineralization in maturing heart valves and chondrocytes. *Plos one* [Online], 6(10), e26769. Available from: https://doi.org/10.1371/journal.pone.0026769.

Pegoraro, E., Hoffman, E.P., Piva, L., Gavassini, B.F., Cagnin, S., Ermani, M., Bello, L., Soraru, G., Pacchioni, B., Bonifati, M.D., Lanfranchi, G., Angelini, C., Kesari, A., Lee, I., Gordish-Dressman, H., Devaney, J.M., McDonald, C.M., and Cooperative International Neuromuscular Research, G., 2011. Spp1 genotype is a determinant of disease severity in duchenne muscular dystrophy. *Neurology* [Online], 76(3), pp.219–26. Available from: https://doi.org/10.1212/WNL.0b013e318207afeb.

Pertea, G. and Pertea, M., 2020. Gff utilities: gffread and gffcompare. *F1000research* [Online], 9, p.304. Available from: https://doi.org/10.12688/f1000research.23297.1.

Pertea, M., Kim, D., Pertea, G.M., Leek, J.T., and Salzberg, S.L., 2016. Transcript-level expression analysis of rna-seq experiments with hisat, stringtie, and ballgown. *Nature protocols* [Online], 11, pp.1650–1667. Available from: https://doi.org/10.1038/nprot.2016.095.

Pervolaraki, E., Dachtler, J., Anderson, R.A., and Holden, A.V., 2018. The developmental transcriptome of the human heart. *Sci rep* [Online], 8(1), p.15362. Available from: https://doi.org/10.1038/s41598-018-33837-6.

Pimpalwar, N., Czuba, T., Smith, M.L., Nilsson, J., Gidlof, O., and Smith, J.G., 2020. Methods for isolation and transcriptional profiling of individual cells from the human heart. *Heliyon* [Online], 6(12), e05810. Available from: https://doi.org/10.1016/j.heliyon.2020.e05810.

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. *Circ res* [Online], 118(3), pp.400–9. Available from: https://doi.org/10.1161/CIRCRESAHA.115.307778.

Podzimkova, J., Palecek, T., Kuchynka, P., Marek, J., Danek, B.A., Jachymova, M., Kalousova, M., Zima, T., and Linhart, A., 2019. Plasma osteopontin levels in patients with dilated and hypertrophic cardiomyopathy. *Herz* [Online], 44(4), pp.347–353. Available from: https://doi.org/10.1007/s00059-017-4645-3.

Pollard, C.M., Desimine, V.L., Wertz, S.L., Perez, A., Parker, B.M., Maning, J., McCrink, K.A., Shehadeh, L.A., and Lymperopoulos, A., 2019. Deletion of osteopontin enhances beta(2)-adrenergic receptor-dependent anti-fibrotic signaling in cardiomyocytes. *Int j mol sci* [Online], 20(6). Available from: https://doi.org/10.3390/ijms20061396.

Pollard, M.O., Gurdasani, D., Mentzer, A.J., Porter, T., and Sandhu, M.S., 2018. Long reads: their purpose and place. *Hum mol genet* [Online], 27(R2), R234–R241. Available from: https://doi.org/10.1093/hmg/ddy177.

Porrello, E.R., Bell, J.R., Schertzer, J.D., Curl, C.L., McMullen, J.R., Mellor, K.M., Ritchie, R.H., Lynch, G.S., Harrap, S.B., Thomas, W.G., and Delbridge, L.M.D., 2009. Heritable pathologic cardiac hypertrophy in adulthood is preceded by neonatal cardiac growth restriction. *American journal of physiology-regulatory integrative and comparative physiology* [Online], 296(3), R672–R680. Available from: https://doi.org/10.1152/ajpregu.90919.2008.

Powis, G., Meuillet, E.J., Indarte, M., Booher, G., and Kirkpatrick, L., 2023. Pleckstrin homology [ph] domain, structure, mechanism, and contribution to human disease. *Biomed pharmacother* [Online], 165, p.115024. Available from: https://doi.org/10.1016/j.biopha.2023.115024.

Prestes, P.R., Marques, F.Z., Lopez-Campos, G., Booth, S.A., McGlynn, M., Lewandowski, P., Delbridge, L.M., Harrap, S.B., and Charchar, F.J., 2016. Tripartite motif-containing 55 identified as functional candidate for spontaneous cardiac hypertrophy in the rat locus cardiac mass 22. *J hypertens* [Online], 34(5), pp.950–8. Available from: https://doi.org/10.1097/HJH.0000000000000875.

Prestes, P.R., Marques, F.Z., Lopez-Campos, G., Lewandowski, P., Delbridge, L.M.D., Charchar, F.J., and Harrap, S.B., 2018. Involvement of human monogenic cardiomyopathy genes in experimental polygenic cardiac hypertrophy. *Physiol genomics* [Online], 50(9), pp.680–687. Available from: https://doi.org/10.1152/physiolgenomics.00143.2017.

Pritchett, J., Harvey, E., Athwal, V., Berry, A., Rowe, C., Oakley, F., Moles, A., Mann, D.A., Bobola, N., Sharrocks, A.D., Thomson, B.J., Zaitoun, A.M., Irving, W.L., Guha, I.N., Hanley, N.A., and Hanley, K.P., 2012. Osteopontin is a novel downstream target of sox9 with diagnostic implications for progression of liver fibrosis in humans. *Hepatology* [Online], 56(3), pp.1108–16. Available from: https://doi.org/10.1002/hep.25758.

Puente, B.N., Kimura, W., Muralidhar, S.A., Moon, J., Amatruda, J.F., Phelps, K.L., Grinsfelder, D., Rothermel, B.A., Chen, R., Garcia, J.A., Santos, C.X., Thet, S., Mori, E., Kinter, M.T., Rindler, P.M., Zacchigna, S., Mukherjee, S., Chen, D.J., Mahmoud, A.I., Giacca, M., Rabinovitch, P.S., Aroumougame, A., Shah, A.M., Szweda, L.I., and Sadek, H.A., 2014. The oxygen-rich postnatal environment induces cardiomyocyte cell-cycle arrest through dna damage response. *Cell* [Online], 157(3), pp.565–79. Available from: https://doi.org/10.1016/j.cell.2014.03.032.

Raff, U., Ott, C., Ruilope, L.M., Menne, J., Haller, H., and Schmieder, R.E., 2014. Prevention of electrocardiographic left ventricular remodeling by the angiotensin receptor blocker olmesartan in patients with type 2 diabetes. *J hypertens* [Online], 32(11), 2267–76, discussion 2276. Available from: https://doi.org/10.1097/HJH.0000000000000313.

Rajabi, M., Kassiotis, C., Razeghi, P., and Taegtmeyer, H., 2007. Return to the fetal gene program protects the stressed heart: a strong hypothesis. *Heart fail rev* [Online], 12(3-4). Available from: https://doi.org/10.1007/s10741-007-9034-1.

Rajagopalan, V., Zucker, I.H., Jones, J.A., Carlson, M., and Ma, Y.J., 2008. Cardiac erbb-1/erbb-2 mutant expression in young adult mice leads to cardiac dysfunction. *Am j physiol heart circ physiol* [Online], 295(2), H543–54. Available from: https://doi.org/10.1152/ajpheart.91436.2007.

Rat Genome Sequencing and Mapping Consortium, Baud, A., Hermsen, R., Guryev, V., Stridh, P., Graham, D., McBride, M.W., Foroud, T., Calderari, S., Diez, M., Ockinger, J., Beyeen, A.D., Gillett, A., Abdelmagid, N., Guerreiro-Cacais, A.O., Jagodic, M., Tuncel, J., Norin, U., Beattie, E., Huynh, N., Miller, W.H., Koller, D.L., Alam, I., Falak, S., Osborne-Pellegrin, M., Martinez-Membrives, E., Canete, T., Blazquez, G., Vicens-Costa, E., Mont-Cardona, C., Diaz-Moran, S., Tobena, A., Hummel, O., Zelenika, D., Saar, K., Patone, G., Bauerfeind, A., Bihoreau, M.T., Heinig, M., Lee, Y.A., Rintisch, C., Schulz, H., Wheeler, D.A., Worley, K.C., Muzny, D.M., Gibbs, R.A., Lathrop, M., Lansu, N., Toonen, P., Ruzius, F.P., Bruijn, E. de, Hauser, H., Adams, D.J., Keane, T., Atanur, S.S., Aitman, T.J., Flicek, P., Malinauskas, T., Jones, E.Y., Ekman, D., Lopez-Aumatell, R., Dominiczak, A.F., Johannesson, M., Holmdahl, R., Olsson, T., Gauguier, D., Hubner, N., Fernandez-Teruel, A., Cuppen, E., Mott, R., and Flint, J., 2013. Combined sequence-based and genetic mapping analysis of complex traits in outbred rats. *Nat genet* [Online], 45(7), pp.767–75. Available from: https://doi.org/10.1038/ng.2644.

Razo, C., Welgan, C.A., Johnson, C.O., McLaughlin, S.A., Iannucci, V., Rodgers, A., Wang, N., LeGrand, K.E., Sorensen, R.J.D., He, J., Zheng, P., Aravkin, A.Y., Hay, S.I., Murray, C.J.L., and Roth, G.A., 2022. Effects of elevated systolic blood pressure

on ischemic heart disease: a burden of proof study. *Nat med* [Online], 28(10), pp.2056–2065. Available from: https://doi.org/10.1038/s41591-022-01974-1.

Redman, M., King, A., Watson, C., and King, D., 2016. What is crispr/cas9? *Arch dis child educ pract ed* [Online], 101(4), pp.213–5. Available from: https://doi.org/10.1136/archdischild-2016-310459.

Renault, M.A., Robbesyn, F., Reant, P., Douin, V., Daret, D., Allieres, C., Belloc, I., Couffinhal, T., Arnal, J.F., Klingel, K., Desgranges, C., Dos Santos, P., Charpentier, F., and Gadeau, A.P., 2010. Osteopontin expression in cardiomyocytes induces dilated cardiomyopathy. *Circ heart fail* [Online], 3(3), pp.431–9. Available from: https://doi.org/10.1161/CIRCHEARTFAILURE.109.898114.

Rhie, A., McCarthy, S.A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., Uliano-Silva, M., Chow, W., Fungtammasan, A., Kim, J., Lee, C., Ko, B.J., Chaisson, M., Gedman, G.L., Cantin, L.J., Thibaud-Nissen, F., Haggerty, L., Bista, I., Smith, M., Haase, B., Mountcastle, J., Winkler, S., Paez, S., Howard, J., Vernes, S.C., Lama, T.M., Grutzner, F., Warren, W.C., Balakrishnan, C.N., Burt, D., George, J.M., Biegler, M.T., Iorns, D., Digby, A., Eason, D., Robertson, B., Edwards, T., Wilkinson, M., Turner, G., Meyer, A., Kautt, A.F., Franchini, P., Detrich H. W., 3., Svardal, H., Wagner, M., Naylor, G.J.P., Pippel, M., Malinsky, M., Mooney, M., Simbirsky, M., Hannigan, B.T., Pesout, T., Houck, M., Misuraca, A., Kingan, S.B., Hall, R., Kronenberg, Z., Sovic, I., Dunn, C., Ning, Z., Hastie, A., Lee, J., Selvaraj, S., Green, R.E., Putnam, N.H., Gut, I., Ghurye, J., Garrison, E., Sims, Y., Collins, J., Pelan, S., Torrance, J., Tracey, A., Wood, J., Dagnew, R.E., Guan, D., London, S.E., Clayton, D.F., Mello, C.V., Friedrich, S.R., Lovell, P.V., Osipova, E., Al-Ajli, F.O., Secomandi, S., Kim, H., Theofanopoulou, C., Hiller, M., Zhou, Y., Harris, R.S., Makova, K.D., Medvedev, P., Hoffman, J., Masterson, P., Clark, K., Martin, F., Howe, K., Flicek, P., Walenz, B.P., Kwak, W., Clawson, H., et al., 2021. Towards complete and error-free genome assemblies of all vertebrate species. *Nature* [Online], 592(7856), pp.737–746. Available from: https://doi.org/10.1038/s41586-021-03451-0.

Ribeiro-Rodrigues, T.M., Laundos, T.L., Pereira-Carvalho, R., Batista-Almeida, D., Pereira, R., Coelho-Santos, V., Silva, A.P., Fernandes, R., Zuzarte, M., Enguita, F.J., Costa, M.C., Pinto-do, O.P., Pinto, M.T., Gouveia, P., Ferreira, L., Mason, J.C., Pereira, P., Kwak, B.R., Nascimento, D.S., and Girao, H., 2017. Exosomes secreted by cardiomyocytes subjected to ischaemia promote cardiac angiogenesis. *Cardiovasc res* [Online], 113(11), pp.1338–1350. Available from: https://doi.org/10.1093/cvr/cvx118.

Rice, J., Courter, D.L., Giachelli, C.M., and Scatena, M., 2006. Molecular mediators of alphavbeta3-induced endothelial cell survival. *J vasc res* [Online], 43(5), pp.422–36. Available from: https://doi.org/10.1159/000094884.

Rice, K.M., Fannin, J.C., Gillette, C., and Blough, E.R., 2014. Efficacy of female rat models in translational cardiovascular aging research. *J aging res* [Online], 2014, p.153127. Available from: https://doi.org/10.1155/2014/153127.

Rohrbach, S., Niemann, B., Silber, R.E., and Holtz, J., 2005. Neuregulin receptors erbb2 and erbb4 in failing human myocardium – depressed expression and attenuated activation. *Basic res cardiol* [Online], 100(3), pp.240–9. Available from: https://doi.org/10.1007/s00395-005-0514-4.

Rohrbach, S., Yan, X., Weinberg, E.O., Hasan, F., Bartunek, J., Marchionni, M.A., and Lorell, B.H., 1999. Neuregulin in cardiac hypertrophy in rats with aortic stenosis. differential expression of erbb2 and erbb4 receptors. *Circulation* [Online], 100(4), pp.407–12. Available from: https://doi.org/10.1161/01.cir.100.4.407.

Rosado, J.A., Diez, R., Smani, T., and Jardin, I., 2015. Stim and orai1 variants in store-operated calcium entry. *Front pharmacol* [Online], 6, p.325. Available from: https://doi.org/10.3389/fphar.2015.00325.

Rosenberg, P., Zhang, H., Bryson, V.G., and Wang, C., 2021. Soce in the cardiomyocyte: the secret is in the chambers. *Pflugers arch* [Online], 473(3), pp.417–434. Available from: https://doi.org/10.1007/s00424-021-02540-3.

Roser, M., Ritchie, H., and Spooner, F., 2021. Burden of disease. *Our world in data*. https://ourworldindata.org/burden-of-disease.

Rysa, J., Leskinen, H., Ilves, M., and Ruskoaho, H., 2005. Distinct upregulation of extracellular matrix genes in transition from hypertrophy to hypertensive heart failure. *Hypertension* [Online], 45(5), pp.927–33. Available from: https://doi.org/10.1161/01.HYP.0000161873.27088.4c.

Saitoh, Y., Kuratsu, J., Takeshima, H., Yamamoto, S., and Ushio, Y., 1995. Expression of osteopontin in human glioma. its correlation with the malignancy. *Lab invest* [Online], 72(1), pp.55–63. Available from: https://www.ncbi.nlm.nih.gov/pubmed/7837791.

Sam, F., Xie, Z., Ooi, H., Kerstetter, D.L., Colucci, W.S., Singh, M., and Singh, K., 2004. Mice lacking osteopontin exhibit increased left ventricular dilation and reduced fibrosis after aldosterone infusion. *Am j hypertens* [Online], 17(2), pp.188–93. Available from: https://doi.org/10.1016/j.amjhyper.2003.10.007.

Santos, M.D., Gioftsidi, S., Backer, S., Machado, L., Relaix, F., Maire, P., and Mourikis, P., 2021. Extraction and sequencing of single nuclei from murine skeletal muscles. *Star protoc* [Online], 2(3), p.100694. Available from: https://doi.org/10.1016/j.xpro.2021.100694.

Satoh, M., Nakamura, M., Akatsu, T., Shimoda, Y., Segawa, I., and Hiramori, K., 2005. Myocardial osteopontin expression is associated with collagen fibrillogenesis in human dilated cardiomyopathy. *Eur j heart fail* [Online], 7(5), pp.755–62. Available from: https://doi.org/10.1016/j.ejheart.2004.10.019.

Savarese, G., Becher, P.M., Lund, L.H., Seferovic, P., Rosano, G.M.C., and Coats, A.J.S., 2023. Global burden of heart failure: a comprehensive and updated review of epidemiology. *Cardiovasc res* [Online], 118(17), pp.3272–3287. Available from: https://doi.org/10.1093/cvr/cvac013.

Sawaki, D., Czibik, G., Pini, M., Ternacle, J., Suffee, N., Mercedes, R., Marcelin, G., Surenaud, M., Marcos, E., Gual, P., Clement, K., Hue, S., Adnot, S., Hatem, S.N., Tsuchimochi, I., Yoshimitsu, T., Henegar, C., and Derumeaux, G., 2018. Visceral adipose tissue drives cardiac aging through modulation of fibroblast senescence by osteopontin production. *Circulation* [Online], 138(8), pp.809–822. Available from: https://doi.org/10.1161/CIRCULATIONAHA.117.031358.

Schauer, A., Adams, V., Poitz, D.M., Barthel, P., Joachim, D., Friedrich, J., Linke, A., and Augstein, A., 2019. Loss of sox9 in cardiomyocytes delays the onset of cardiac hypertrophy and fibrosis. *Int j cardiol* [Online], 282, pp.68–75. Available from: https://doi.org/10.1016/j.ijcard.2019.01.078.

Schirmer, H., Lunde, P., and Rasmussen, K., 1999. Prevalence of left ventricular hypertrophy in a general population; the tromso study. *Eur heart j* [Online], 20(6), pp.429–38. Available from: https://doi.org/10.1053/euhj.1998.1314.

Schmieder, R.E., 2010. End organ damage in hypertension. *Dtsch arztebl int* [Online], 107(49), pp.866–73. Available from: https://doi.org/10.3238/arztebl.2010.0866.

Schreier, B., Hunerberg, M., Rabe, S., Mildenberger, S., Bethmann, D., Heise, C., Sibilia, M., Offermanns, S., and Gekle, M., 2016. Consequences of postnatal vascular smooth muscle egfr deletion on acute angiotensin ii action. *Clin sci (lond)* [Online], 130(1), pp.19–33. Available from: https://doi.org/10.1042/CS20150503.

Schreier, B., Rabe, S., Schneider, B., Bretschneider, M., Rupp, S., Ruhs, S., Neumann, J., Rueckschloss, U., Sibilia, M., Gotthardt, M., Grossmann, C., and Gekle, M., 2013. Loss of epidermal growth factor receptor in vascular smooth muscle cells and cardiomyocytes causes arterial hypotension and cardiac hypertrophy. *Hypertension* [Online], 61(2), pp.333–40. Available from: https://doi.org/10.1161/HYPERTENSIONAHA.112.196543.

Schytte, G.N., Christensen, B., Bregenov, I., Kjoge, K., Scavenius, C., Petersen, S.V., Enghild, J.J., and Sorensen, E.S., 2020. Fam20c phosphorylation of the rgdsvvyglr motif in osteopontin inhibits interaction with the alphavbeta3 integrin. *J cell biochem* [Online]. Available from: https://doi.org/10.1002/jcb.29708.

Shen, J.F., Fan, Z.B., Wu, C.W., Qi, G.X., Cao, Q.Y., and Xu, F., 2022. Sacubitril valsartan enhances cardiac function and alleviates myocardial infarction in rats through a suv39h1/spp1 axis. *Oxid med cell longev* [Online], 2022, p.5009289. Available from: https://doi.org/10.1155/2022/5009289.

Shimizu, I. and Minamino, T., 2016. Physiological and pathological cardiac hypertrophy. *J mol cell cardiol* [Online], 97, pp.245–62. Available from: https://doi.org/10.1016/j.yjmcc.2016.06.001.

Shinohara, M.L., Kim, H.J., Kim, J.H., Garcia, V.A., and Cantor, H., 2008a. Alternative translation of osteopontin generates intracellular and secreted isoforms that mediate distinct biological activities in dendritic cells. *Proc natl acad sci usa* [Online], 105(20), pp.7235–9. Available from: https://doi.org/10.1073/pnas.0802301105.

Shinohara, M.L., Kim, J.H., Garcia, V.A., and Cantor, H., 2008b. Engagement of the type i interferon receptor on dendritic cells inhibits t helper 17 cell development: role of intracellular osteopontin. *Immunity* [Online], 29(1), pp.68–78. Available from: https://doi.org/10.1016/j.immuni.2008.05.008.

Shirakawa, K., Endo, J., Kataoka, M., Katsumata, Y., Yoshida, N., Yamamoto, T., Isobe, S., Moriyama, H., Goto, S., Kitakata, H., Hiraide, T., Fukuda, K., and Sano, M., 2018. Il (interleukin)-10-stat3-galectin-3 axis is essential for osteopontin-producing reparative macrophage polarization after myocardial infarction. *Circulation* [Online], 138(18), pp.2021–2035. Available from: https://doi.org/10.1161/CIRCULATIONAHA.118.035047.

Sibilia, M., Wagner, B., Hoebertz, A., Elliott, C., Marino, S., Jochum, W., and Wagner, E.F., 2003. Mice humanised for the egf receptor display hypomorphic phenotypes in skin, bone and heart. *Development* [Online], 130(19), pp.4515–25. Available from: https://doi.org/10.1242/dev.00664.

Sibilia, M. and Wagner, E.F., 1995. Strain-dependent epithelial defects in mice lacking the egf receptor. *Science* [Online], 269(5221), pp.234–8. Available from: https://doi.org/10.1126/science.7618085.

Siegel, A.K., Planert, M., Rademacher, S., Mehr, A.P., Kossmehl, P., Wehland, M., Stoll, M., and Kreutz, R., 2003. Genetic loci contribute to the progression of vascular and cardiac hypertrophy in salt-sensitive spontaneous hypertension. *Arterioscler thromb vasc biol* [Online], 23(7), pp.1211–7. Available from: https://doi.org/10.1161/01.ATV.0000079509.20542.C9.

Silva, A.C., Pereira, C., Fonseca, A., Pinto-do, O.P., and Nascimento, D.S., 2020. Bearing my heart: the role of extracellular matrix on cardiac development, homeostasis, and injury response. *Front cell dev biol* [Online], 8, p.621644. Available from: https://doi.org/10.3389/fcell.2020.621644.

Silverman, E.S. and Collins, T., 1999. Pathways of egr-1-mediated gene transcription in vascular biology. *Am j pathol* [Online], 154(3), pp.665–70. Available from: https://doi.org/10.1016/S0002-9440(10)65312-6.

Simone, G. de, Daniels, S.R., Devereux, R.B., Meyer, R.A., Roman, M.J., Divitiis, O. de, and Alderman, M.H., 1992. Left ventricular mass and body size in normotensive children and adults: assessment of allometric relations and impact of overweight. *J am coll cardiol* [Online], 20(5), pp.1251–60. Available from: https://doi.org/10.1016/0735-1097(92)90385-z.

Simone, G. de, Devereux, R.B., Daniels, S.R., and Meyer, R.A., 1995. Gender differences in left ventricular growth. *Hypertension* [Online], 26(6 Pt 1), pp.979–83. Available from: https://doi.org/10.1161/01.hyp.26.6.979.

Simone, G. de, Devereux, R.B., Izzo, R., Girfoglio, D., Lee, E.T., Howard, B.V., and Roman, M.J., 2013. Lack of reduction of left ventricular mass in treated hypertension: the strong heart study. *J am heart assoc* [Online], 2(3), e000144. Available from: https://doi.org/10.1161/JAHA.113.000144.

Singh, K., Balligand, J.L., Fischer, T.A., Smith, T.W., and Kelly, R.A., 1995. Glucocorticoids increase osteopontin expression in cardiac myocytes and microvascular endothelial cells. role in regulation of inducible nitric oxide synthase. *J biol chem* [Online], 270(47), pp.28471–8. Available from: https://doi.org/10.1074/jbc.270.47.28471.

Singh, K., Sirokman, G., Communal, C., Robinson, K.G., Conrad, C.H., Brooks, W.W., Bing, O.H., and Colucci, W.S., 1999. Myocardial osteopontin expression coincides with the development of heart failure. *Hypertension* [Online], 33(2), pp.663–70. Available from: https://doi.org/10.1161/01.hyp.33.2.663.

Singh, M., Dalal, S., and Singh, K., 2014. Osteopontin: at the cross-roads of myocyte survival and myocardial function. *Life sciences* [Online], 118(1), pp.1–6. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4254317/pdf/nihms631803.pdf.

Skavdahl, M., Steenbergen, C., Clark, J., Myers, P., Demianenko, T., Mao, L., Rockman, H.A., Korach, K.S., and Murphy, E., 2005. Estrogen receptor-beta mediates male-female differences in the development of pressure overload hypertrophy. *Am j physiol heart circ physiol* [Online], 288(2), H469–76. Available from: https://doi.org/10.1152/ajpheart.00723.2004.

Skelly, D.A., Squiers, G.T., McLellan, M.A., Bolisetty, M.T., Robson, P., Rosenthal, N.A., and Pinto, A.R., 2018. Single-cell transcriptional profiling reveals cellular diversity and intercommunication in the mouse heart. *Cell rep* [Online], 22(3), pp.600–610. Available from: https://doi.org/10.1016/j.celrep.2017.12.072.

Slyper, M., Porter, C.B.M., Ashenberg, O., Waldman, J., Drokhlyansky, E., Wakiro, I., Smillie, C., Smith-Rosario, G., Wu, J., Dionne, D., Vigneau, S., Jane-Valbuena, J., Tickle, T.L., Napolitano, S., Su, M.J., Patel, A.G., Karlstrom, A., Gritsch, S., Nomura, M., Waghray, A., Gohil, S.H., Tsankov, A.M., Jerby-Arnon, L., Cohen, O., Klughammer, J., Rosen, Y., Gould, J., Nguyen, L., Hofree, M., Tramontozzi, P.J., Li, B., Wu, C.J., Izar, B., Haq, R., Hodi, F.S., Yoon, C.H., Hata, A.N., Baker, S.J., Suva, M.L., Bueno, R., Stover, E.H., Clay, M.R., Dyer, M.A., Collins, N.B., Matulonis, U.A., Wagle, N., Johnson, B.E., Rotem, A., Rozenblatt-Rosen, O., and Regev, A., 2020. A single-cell and single-nucleus rna-seq toolbox for fresh and frozen human tumors. *Nat med* [Online], 26(5), pp.792–802. Available from: https://doi.org/10.1038/s41591-020-0844-1.

Small, H.Y., Morgan, H., Beattie, E., Griffin, S., Indahl, M., Delles, C., and Graham, D., 2016. Abnormal uterine artery remodelling in the stroke prone spontaneously hypertensive rat. *Placenta* [Online], 37, pp.34–44. Available from: https://doi.org/10.1016/j.placenta.2015.10.022.

Smith, J.R., Hayman, G.T., Wang, S.J., Laulederkind, S.J.F., Hoffman, M.J., Kaldunski, M.L., Tutaj, M., Thota, J., Nalabolu, H.S., Ellanki, S.L.R., Tutaj, M.A., De Pons, J.L., Kwitek, A.E., Dwinell, M.R., and Shimoyama, M.E., 2020. The year of the rat: the rat genome database at 20: a multi-species knowledgebase and analysis platform. *Nucleic acids res* [Online], 48(D1), pp.D731–D742. Available from: https://doi.org/10.1093/nar/gkz1041.

Smits, A.H., Ziebell, F., Joberty, G., Zinn, N., Mueller, W.F., Clauder-Munster, S., Eberhard, D., Falth Savitski, M., Grandi, P., Jakob, P., Michon, A.M., Sun, H., Tessmer, K., Burckstummer, T., Bantscheff, M., Steinmetz, L.M., Drewes, G., and Huber, W., 2019. Biological plasticity rescues target activity in crispr knock outs. *Nat methods* [Online], 16(11), pp.1087–1093. Available from: https://doi.org/10.1038/s41592-019-0614-5.

Snell, G.D., 1948. Methods for the study of histocompatibility genes. *J genet* [Online], 49(2), pp.87–108. Available from: https://doi.org/10.1007/BF02986826.

Solberg Woods, L.C., Stelloh, C., Regner, K.R., Schwabe, T., Eisenhauer, J., and Garrett, M.R., 2010. Heterogeneous stock rats: a new model to study the genetics of renal phenotypes. *Am j physiol renal physiol* [Online], 298(6), F1484–91. Available from: https://doi.org/10.1152/ajprenal.00002.2010.

Sollis, E., Mosaku, A., Abid, A., Buniello, A., Cerezo, M., Gil, L., Groza, T., Gunes, O., Hall, P., Hayhurst, J., Ibrahim, A., Ji, Y., John, S., Lewis, E., MacArthur, J.A.L., McMahon, A., Osumi-Sutherland, D., Panoutsopoulou, K., Pendlington, Z., Ramachandran, S., Stefancsik, R., Stewart, J., Whetzel, P., Wilson, R., Hindorff, L., Cunningham, F., Lambert, S.A., Inouye, M., Parkinson, H., and Harris, L.W., 2023. The nhgri-

ebi gwas catalog: knowledgebase and deposition resource. *Nucleic acids res* [Online], 51(D1), pp.D977–D985. Available from: https://doi.org/10.1093/nar/gkac1010.

Sollner, J.F., Leparc, G., Hildebrandt, T., Klein, H., Thomas, L., Stupka, E., and Simon, E., 2017. An rna-seq atlas of gene expression in mouse and rat normal tissues. *Sci data* [Online], 4, p.170185. Available from: https://doi.org/10.1038/sdata.2017.185.

Soneson, C., Love, M.I., and Robinson, M.D., 2015. Differential analyses for rna-seq: transcript-level estimates improve gene-level inferences. *F1000res* [Online], 4, p.1521. Available from: https://doi.org/10.12688/f1000research.7563.2.

Song, H.K., Hong, S.E., Kim, T., and Kim, D.H., 2012. Deep rna sequencing reveals novel cardiac transcriptomic signatures for physiological and pathological hypertrophy. *Plos one* [Online], 7(4), e35552. Available from: https://doi.org/10.1371/journal.pone.0035552.

Soto, P.F., Herrero, P., Schechtman, K.B., Waggoner, A.D., Baumstark, J.M., Ehsani, A.A., and Gropler, R.J., 2008. Exercise training impacts the myocardial metabolism of older individuals in a gender-specific manner. *Am j physiol heart circ physiol* [Online], 295(2), H842–50. Available from: https://doi.org/10.1152/ajpheart.91426.2007.

Stack, E.C., Wang, C., Roman, K.A., and Hoyt, C.C., 2014. Multiplexed immunohistochemistry, imaging, and quantitation: a review, with an assessment of tyramide signal amplification, multispectral imaging and multiplex analysis. *Methods* [Online], 70(1), pp.46–58. Available from: https://doi.org/10.1016/j.ymeth.2014.08.016.

Star Consortium, Saar, K., Beck, A., Bihoreau, M.T., Birney, E., Brocklebank, D., Chen, Y., Cuppen, E., Demonchy, S., Dopazo, J., Flicek, P., Foglio, M., Fujiyama, A., Gut, I.G., Gauguier, D., Guigo, R., Guryev, V., Heinig, M., Hummel, O., Jahn, N., Klages, S., Kren, V., Kube, M., Kuhl, H., Kuramoto, T., Kuroki, Y., Lechner, D., Lee, Y.A., Lopez-Bigas, N., Lathrop, G.M., Mashimo, T., Medina, I., Mott, R., Patone, G., Perrier-Cornet, J.A., Platzer, M., Pravenec, M., Reinhardt, R., Sakaki, Y., Schilhabel, M., Schulz, H., Serikawa, T., Shikhagaie, M., Tatsumoto, S., Taudien, S., Toyoda, A., Voigt, B., Zelenika, D., Zimdahl, H., and Hubner, N., 2008. Snp and haplotype mapping for genetic analysis in the rat. *Nat genet* [Online], 40(5), pp.560–6. Available from: https://doi.org/10.1038/ng.124.

Steri, M., Idda, M.L., Whalen, M.B., and Orru, V., 2018. Genetic variants in mrna untranslated regions. *Wiley interdiscip rev rna* [Online], 9(4), e1474. Available from: https://doi.org/10.1002/wrna.1474.

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., and Mesirov, J.P., 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc natl acad sci u s a* [Online], 102(43), pp.15545–50. Available from: https://doi.org/10.1073/pnas.0506580102.

Summers, K.M., Bush, S.J., Wu, C., and Hume, D.A., 2022. Generation and network analysis of an rna-seq transcriptional atlas for the rat. *Nar genom bioinform* [Online], 4(1), lqac017. Available from: https://doi.org/10.1093/nargab/lqac017.

Suryawanshi, H., Clancy, R., Morozov, P., Halushka, M.K., Buyon, J.P., and Tuschl, T., 2020. Cell atlas of the foetal human heart and implications for autoimmune-mediated congenital heart block. *Cardiovasc res* [Online], 116(8), pp.1446–1457. Available from: https://doi.org/10.1093/cvr/cvz257.

Suzuki, T., Kamimura, D., Wakui, H., and Tamura, K., 2021. May need more comprehensive approach to residual risks in well controlled hypertensive patients. *Hypertens res* [Online], 44(2), pp.253–255. Available from: https://doi.org/10.1038/s41440-020-00567-0.

Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A.L., Fang, T., Doncheva, N.T., Pyysalo, S., Bork, P., Jensen, L.J., and Mering, C. von, 2023. The string database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic acids res* [Online], 51(D1), pp.D638–D646. Available from: https://doi.org/10.1093/nar/gkac1000.

Szpirer, C., 2020. Rat models of human diseases and related phenotypes: a systematic inventory of the causative genes. *J biomed sci* [Online], 27(1), p.84. Available from: https://doi.org/10.1186/s12929-020-00673-8.

Tabakoff, B., Smith, H., Vanderlinden, L.A., Hoffman, P.L., and Saba, L.M., 2019. Networking in biology: the hybrid rat diversity panel. *Methods mol biol* [Online], 2018, pp.213–231. Available from: https://doi.org/10.1007/978-1-4939-9581-3_10.

Tadros, R., Francis, C., Xu, X., Vermeer, A.M.C., Harper, A.R., Huurman, R., Kelu Bisabu, K., Walsh, R., Hoorntje, E.T., Te Rijdt, W.P., Buchan, R.J., Velzen, H.G. van, Slegtenhorst, M.A. van, Vermeulen, J.M., Offerhaus, J.A., Bai, W., Marvao, A. de, Lahrouchi, N., Beekman, L., Karper, J.C., Veldink, J.H., Kayvanpour, E., Pantazis, A., Baksi, A.J., Whiffin, N., Mazzarotto, F., Sloane, G., Suzuki, H., Schneider-Luftman, D., Elliott, P., Richard, P., Ader, F., Villard, E., Lichtner, P., Meitinger, T., Tanck, M.W.T., Tintelen, J.P. van, Thain, A., McCarty, D., Hegele, R.A., Roberts, J.D., Amyot, J., Dube, M.P., Cadrin-Tourigny, J., Giraldeau, G., L'Allier, P.L., Garceau, P., Tardif, J.C., Boekholdt, S.M., Lumbers, R.T., Asselbergs, F.W., Barton, P.J.R., Cook, S.A., Prasad, S.K., O'Regan, D.P., Velden, J. van der, Verweij, K.J.H., Talajic,

M., Lettre, G., Pinto, Y.M., Meder, B., Charron, P., Boer, R.A. de, Christiaans, I., Michels, M., Wilde, A.A.M., Watkins, H., Matthews, P.M., Ware, J.S., and Bezzina, C.R., 2021. Shared genetic pathways contribute to risk of hypertrophic and dilated cardiomyopathies with opposite directions of effect. *Nat genet* [Online], 53(2), pp.128–134. Available from: https://doi.org/10.1038/s41588-020-00762-2.

Taki, H., Tuomilehto, J., Zimmet, P., Tamosiunas, A., Kowlessur, S., Magliano, D.J., Shaw, J.E., Soderberg, S., and Nilsson, U., 2023. Left ventricular hypertrophy: an ecg-based study of prevalence and risk factors in a multiethnic population. *Open heart* [Online], 10(2). Available from: https://doi.org/10.1136/openhrt-2023-002495.

Tallquist, M.D. and Molkentin, J.D., 2017. Redefining the identity of cardiac fibroblasts. *Nat rev cardiol* [Online], 14(8), pp.484–491. Available from: https://doi.org/10.1038/nrcardio.2017.57.

Tang, X., Chen, X.F., Chen, H.Z., and Liu, D.P., 2017. Mitochondrial sirtuins in cardiometabolic diseases. *Clin sci (lond)* [Online], 131(16), pp.2063–2078. Available from: https://doi.org/10.1042/CS20160685.

Taylor, H.C.M., Chaturvedi, N., Davey Smith, G., Ferreira, D.L.S., Fraser, A., Howe, L.D., Hughes, A.D., Lawlor, D.A., Timpson, N.J., and Park, C.M., 2023. Is height(2.7) appropriate for indexation of left ventricular mass in healthy adolescents? the importance of sex differences. *Hypertension* [Online]. Available from: https://doi.org/10.1161/HYPERTENSIONAHA.121.17109.

Tekath, T. and Dugas, M., 2021a. Differential transcript usage analysis of bulk and single-cell rna-seq data with dturtle. *Bioinformatics* [Online], 37(21), pp.3781–3787. Available from: https://doi.org/10.1093/bioinformatics/btab629.

Tekath, T. and Dugas, M., 2021b. Differential transcript usage analysis of bulk and single-cell rna-seq data with dturtle. *Bioinformatics* [Online], 37 (21), pp.3781–3787. Available from: https://doi.org/10.1093/bioinformatics/btab629.

Tingleff, J., Munch, M., Jakobsen, T.J., Torp-Pedersen, C., Olsen, M.E., Jensen, K.H., Jorgensen, T., and Kirchoff, M., 1996. Prevalence of left ventricular hypertrophy in a hypertensive population. *Eur heart j* [Online], 17(1), pp.143–9. Available from: https://doi.org/10.1093/oxfordjournals.eurheartj.a014672.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., Baren, M.J. van, Salzberg, S.L., Wold, B.J., and Pachter, L., 2010. Transcript assembly and quantification by rna-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat biotechnol* [Online], 28(5), pp.511–5. Available from: https://doi.org/10.1038/nbt.1621.

Tucker, N.R., Chaffin, M., Fleming, S.J., Hall, A.W., Parsons, V.A., Bedi K. C., J., 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* [Online], 142(5), pp.466–482. Available from: https://doi.org/10.1161/CIRCULATIONAHA.119.045401.

Unger, T., Borghi, C., Charchar, F., Khan, N.A., Poulter, N.R., Prabhakaran, D., Ramirez, A., Schlaich, M., Stergiou, G.S., Tomaszewski, M., Wainford, R.D., Williams, B., and Schutte, A.E., 2020. 2020 international society of hypertension global hypertension practice guidelines. *Hypertension* [Online], 75(6), pp.1334–1357. Available from: https://doi.org/10.1161/HYPERTENSIONAHA.120.15026.

Vad, O.B., Angeli, E., Liss, M., Ahlberg, G., Andreasen, L., Christophersen, I.E., Hansen, C.C., Møller, S., Hellsten, Y., Haunsoe, S., Tveit, A., Svendsen, J.H., Gotthardt, M., Lundegaard, P.R., and Olesen, M.S., 2023. Loss of cardiac splicing regulator rbm20 is associated with early-onset atrial fibrillation. *Jacc: basic to translational science* [Online]. Available from: https://doi.org/10.1016/j.jacbts.2023.08.008.

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. *J am coll cardiol* [Online], 80(25), pp.2361–2371. Available from: https://doi.org/10.1016/j.jacc.2022.11.005.

Valen, G., Yan, Z.Q., and Hansson, G.K., 2001. Nuclear factor kappa-b and the heart. *J am coll cardiol* [Online], 38(2), pp.307–14. Available from: https://doi.org/10.1016/s0735-1097(01)01377-8.

Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K.V., Altshuler, D., Gabriel, S., and DePristo, M.A., 2013. From fastq data to high confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr protoc bioinformatics* [Online], 43(1110), pp.11 10 1–11 10 33. Available from: https://doi.org/10.1002/0471250953.bi1110s43.

Vasan, R.S., Larson, M.G., Aragam, J., Wang, T.J., Mitchell, G.F., Kathiresan, S., Newton-Cheh, C., Vita, J.A., Keyes, M.J., O'Donnell, C.J., Levy, D., and Benjamin, E.J., 2007. Genome-wide association of echocardiographic dimensions, brachial artery endothelial function and treadmill exercise responses in the framingham heart study. *Bmc med genet* [Online], 8 Suppl 1(Suppl 1), S2. Available from: https://doi.org/10.1186/1471-2350-8-S1-S2.

Vasan, R.S., Larson, M.G., Leip, E.P., Evans, J.C., O'Donnell, C.J., Kannel, W.B., and Levy, D., 2001. Impact of high-normal blood pressure on the risk of cardiovascular disease. *N engl j med* [Online], 345(18), pp.1291–7. Available from: https://doi.org/10.1056/NEJMoa003417.

Vedi, M., Smith, J.R., Thomas Hayman, G., Tutaj, M., Brodie, K.C., De Pons, J.L., Demos, W.M., Gibson, A.C., Kaldunski, M.L., Lamers, L., Laulederkind, S.J.F., Thota, J., Thorat, K., Tutaj, M.A., Wang, S.J., Zacher, S., Dwinell, M.R., and Kwitek, A.E., 2023. 2022 updates to the rat genome database: a findable, accessible, interoperable, and reusable (fair) resource. *Genetics* [Online], 224(1). Available from: https://doi.org/10.1093/genetics/iyad042.

Vianello, S., Pantic, B., Fusto, A., Bello, L., Galletta, E., Borgia, D., Gavassini, B.F., Semplicini, C., Soraru, G., Vitiello, L., and Pegoraro, E., 2017. Spp1 genotype and glucocorticoid treatment modify osteopontin expression in duchenne muscular dystrophy cells. *Hum mol genet* [Online], 26(17), pp.3342–3351. Available from: https://doi.org/10.1093/hmg/ddx218.

Viwatrangkul, P., Lawanwisut, S., Leekhaphan, P., Prasart-Intara, T., Phiensuparp, P., Prakiatpongsa, S., Amnaj, P., Phoominart, V., Chanyou, K., Jiratrakan, P., Klumnaimueang, P., Pipitdaecha, N., Panchamawat, R., Tangkongpanich, P., Mungthin, M., Rangsin, R., and Sakboonyarat, B., 2021. Prevalence and associated factors of electrocardiographic left ventricular hypertrophy in a rural community, central thailand. *Sci rep* [Online], 11(1), p.7083. Available from: https://doi.org/10.1038/s41598-021-86693-2.

Vogel, B., Acevedo, M., Appelman, Y., Bairey Merz, C.N., Chieffo, A., Figtree, G.A., Guerrero, M., Kunadian, V., Lam, C.S.P., Maas, A., Mihailidou, A.S., Olszanecka, A., Poole, J.E., Saldarriaga, C., Saw, J., Zuhlke, L., and Mehran, R., 2021. The lancet women and cardiovascular disease commission: reducing the global burden by 2030. *Lancet* [Online], 397(10292), pp.2385–2438. Available from: https://doi.org/10.1016/S0140-6736(21)00684-X.

Vuckovic, S., Dinani, R., Nollet, E.E., Kuster, D.W.D., Buikema, J.W., Houtkooper, R.H., Nabben, M., Velden, J. van der, and Goversen, B., 2022. Characterization of cardiac metabolism in ipsc-derived cardiomyocytes: lessons from maturation and disease modeling. *Stem cell res ther* [Online], 13(1), p.332. Available from: https://doi.org/10.1186/s13287-022-03021-9.

Wadley, G.D., McConell, G.K., Goodman, C.A., Siebel, A.L., Westcott, K.T., and Wlodek, M.E., 2013. Growth restriction in the rat alters expression of metabolic genes during postnatal cardiac development in a sex-specific manner. *Physiol genomics* [Online], 45(3), pp.99–105. Available from: https://doi.org/10.1152/physiolgenomics.00095.2012.

Wagner, G.P., Kin, K., and Lynch, V.J., 2012. Measurement of mrna abundance using rna-seq data: rpkm measure is inconsistent among samples. *Theory biosci* [Online], 131(4), pp.281–5. Available from: https://doi.org/10.1007/s12064-012-0162-3.

Wan, E.Y.F., Fung, W.T., Schooling, C.M., Au Yeung, S.L., Kwok, M.K., Yu, E.Y.T., Wang, Y., Chan, E.W.Y., Wong, I.C.K., and Lam, C.L.K., 2021. Blood pressure and risk of cardiovascular disease in uk biobank: a mendelian randomization study. *Hypertension* [Online], 77(2), pp.367–375. Available from: https://doi.org/10.1161/HYPERTENSIONAHA.120.16138.

Wang, Q., Lin, J.L., Erives, A.J., Lin, C.I., and Lin, J.J., 2014. New insights into the roles of xin repeat-containing proteins in cardiac development, function, and disease. *Int rev cell mol biol* [Online], 310, pp.89–128. Available from: https://doi.org/10.1016/B978-0-12-800180-6.00003-7.

Wang, R.S., Maron, B.A., and Loscalzo, J., 2023. Multiomics network medicine approaches to precision medicine and therapeutics in cardiovascular diseases. *Arterioscler thromb vasc biol* [Online], 43(4), pp.493–503. Available from: https://doi.org/10.1161/ATVBAHA.122.318731.

Wang, X., Hao, G., Chen, L., Yang, Y., Zhou, H., Kang, Y., Shaver, L., Chen, Z., Zheng, C., Zhang, L., Li, S., Wang, Z., and Gao, R., 2022. Hypertension-mediated organ damage and established cardiovascular disease in patients with hypertension: the china hypertension survey, 2012-2015. *J hum hypertens* [Online], 36(12), pp.1092–1098. Available from: https://doi.org/10.1038/s41371-021-00635-z.

Wang, X., You, X., Langer, J.D., Hou, J., Rupprecht, F., Vlatkovic, I., Quedenau, C., Tushev, G., Epstein, I., Schaefke, B., Sun, W., Fang, L., Li, G., Hu, Y., Schuman, E.M., and Chen, W., 2019. Full-length transcriptome reconstruction reveals a large diversity of rna and protein isoforms in rat hippocampus. *Nat commun* [Online], 10(1), p.5009. Available from: https://doi.org/10.1038/s41467-019-13037-0.

Wang, Z., Wang, H., Zhang, Y., Yu, F., Yu, L., and Zhang, C., 2021. Single-cell rna sequencing analysis to characterize cells and gene expression landscapes in atrial septal defect. *J cell mol med* [Online], 25(20), pp.9660–9673. Available from: https://doi.org/10.1111/jcmm.16914.

Wehrens, M., Leeuw, A.E. de, Wright-Clark, M., Eding, J.E.C., Boogerd, C.J., Molenaar, B., Kraak, P.H. van der, Kuster, D.W.D., Velden, J. van der, Michels, M., Vink, A., and Rooij, E. van, 2022. Single-cell transcriptomics provides insights into hypertrophic cardiomyopathy. *Cell rep* [Online], 39(6), p.110809. Available from: https://doi.org/10.1016/j.celrep.2022.110809.

Whittington, N.C. and Wray, S., 2017. Suppression of red blood cell autofluorescence for immunocytochemistry on fixed embryonic mouse tissue. *Curr protoc neurosci* [Online], 81, pp.2 28 1–2 28 12. Available from: https://doi.org/10.1002/cpns.35.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., and Yutani, H., 2019. Welcome to the tidyverse. *Journal of open source software* [Online], 4(43), p.1686. Available from: https://doi.org/10.21105/joss.01686.

Wieduwilt, M.J. and Moasser, M.M., 2008. The epidermal growth factor receptor family: biology driving targeted therapeutics. *Cell mol life sci* [Online], 65(10), pp.1566–84. Available from: https://doi.org/10.1007/s00018-008-7440-8.

Wild, P.S., Felix, J.F., Schillert, A., Teumer, A., Chen, M.H., Leening, M.J.G., Volker, U., Grossmann, V., Brody, J.A., Irvin, M.R., Shah, S.J., Pramana, S., Lieb, W., Schmidt, R., Stanton, A.V., Malzahn, D., Smith, A.V., Sundstrom, J., Minelli, C., Ruggiero, D., Lyytikainen, L.P., Tiller, D., Smith, J.G., Monnereau, C., Di Tullio, M.R., Musani, S.K., Morrison, A.C., Pers, T.H., Morley, M., Kleber, M.E., Aragam, J., Benjamin, E.J., Bis, J.C., Bisping, E., Broeckel, U., Cheng, S., Deckers, J.W., Del Greco, M.F., Edelmann, F., Fornage, M., Franke, L., Friedrich, N., Harris, T.B., Hofer, E., Hofman, A., Huang, J., Hughes, A.D., Kahonen, M., Investigators, K., Kruppa, J., Lackner, K.J., Lannfelt, L., Laskowski, R., Launer, L.J., Leosdottir, M., Lin, H., Lindgren, C.M., Loley, C., MacRae, C.A., Mascalzoni, D., Mayet, J., Medenwald, D., Morris, A.P., Muller, C., Muller-Nurasyid, M., Nappo, S., Nilsson, P.M., Nuding, S., Nutile, T., Peters, A., Pfeufer, A., Pietzner, D., Pramstaller, P.P., Raitakari, O.T., Rice, K.M., Rivadeneira, F., Rotter, J.I., Ruohonen, S.T., Sacco, R.L., Samdarshi, T.E., Schmidt, H., Sharp, A.S.P., Shields, D.C., Sorice, R., Sotoodehnia, N., Stricker, B.H., Surendran, P., Thom, S., Toglhofer, A.M., Uitterlinden, A.G., Wachter, R., Volzke, H., Ziegler, A., Munzel, T., Marz, W., Cappola, T.P., Hirschhorn, J.N., Mitchell, G.F., Smith, N.L., Fox, E.R., et al., 2017. Large-scale genome-wide analysis identifies genetic variants associated with cardiac structure and function. *J clin invest* [Online], 127(5), pp.1798–1812. Available from: https://doi.org/10.1172/JCI84840.

Williams, B., Mancia, G., Spiering, W., Agabiti Rosei, E., Azizi, M., Burnier, M., Clement, D., Coca, A., De Simone, G., Dominiczak, A., Kahan, T., Mahfoud, F., Redon, J., Ruilope, L., Zanchetti, A., Kerins, M., Kjeldsen, S., Kreutz, R., Laurent, S., Lip, G.Y.H., McManus, R., Narkiewicz, K., Ruschitzka, F., Schmieder, R., Shlyakhto, E., Tsioufis, K., Aboyans, V., and Desormais, I., 2018. 2018 practice guidelines for the management of arterial hypertension of the european society of cardiology and the european society of hypertension. *Blood press* [Online], 27(6), pp.314–340. Available from: https://doi.org/10.1080/08037051.2018.1527177.

Wingett, S.W. and Andrews, S., 2018. Fastq screen: a tool for multi-genome mapping and quality control. *F1000research* [Online], 7. [version 2; referees: 4 approved], p.1338. Available from: https://doi.org/10.12688/f1000research.15931.2.

Witek, P., Korga, A., Burdan, F., Ostrowska, M., Nosowska, B., Iwan, M., and Dudka, J., 2016. The effect of a number of h9c2 rat cardiomyocytes passage on repeatability of cytotoxicity study results. *Cytotechnology* [Online], 68(6), pp.2407–2415. Available from: https://doi.org/10.1007/s10616-016-9957-2.

Wright, S.N., Leger, B.S., Rosenthal, S.B., Liu, S.N., Jia, T., Chitre, A.S., Polesskaya, O., Holl, K., Gao, J., Cheng, R., Garcia Martinez, A., George, A., Gileta, A.F., Han, W., Netzley, A.H., King, C.P., Lamparelli, A., Martin, C., St Pierre, C.L., Wang, T., Bimschleger, H., Richards, J., Ishiwari, K., Chen, H., Flagel, S.B., Meyer, P., Robinson, T.E., Solberg Woods, L.C., Kreisberg, J.F., Ideker, T., and Palmer, A.A., 2023. Genome-wide association studies of human and rat bmi converge on synapse, epigenome, and hormone signaling networks. *Cell rep* [Online], 42(8), p.112873. Available from: https://doi.org/10.1016/j.celrep.2023.112873.

Xiao, G., Mao, S., Baumgarten, G., Serrano, J., Jordan, M.C., Roos, K.P., Fishbein, M.C., and MacLellan, W.R., 2001. Inducible activation of c-myc in adult myocardium in vivo provokes cardiac myocyte hypertrophy and reactivation of dna synthesis. *Circ res* [Online], 89(12), pp.1122–9. Available from: https://doi.org/10.1161/hh2401.100742.

Xie, Z., Pimental, D.R., Lohan, S., Vasertriger, A., Pligavko, C., Colucci, W.S., and Singh, K., 2001. Regulation of angiotensin ii-stimulated osteopontin expression in cardiac microvascular endothelial cells: role of p42/44 mitogen-activated protein kinase and reactive oxygen species. *J cell physiol* [Online], 188(1), pp.132–8. Available from: https://doi.org/10.1002/jcp.1104.

Xie, Z., Singh, M., and Singh, K., 2004a. Erk1/2 and jnks, but not p38 kinase, are involved in reactive oxygen species-mediated induction of osteopontin gene expression by angiotensin ii and interleukin-1beta in adult rat cardiac fibroblasts. *J cell physiol* [Online], 198(3), pp.399–407. Available from: https://doi.org/10.1002/jcp.10419.

Xie, Z., Singh, M., and Singh, K., 2004b. Osteopontin modulates myocardial hypertrophy in response to chronic pressure overload in mice. *Hypertension* [Online], 44(6), pp.826–31. Available from: https://doi.org/10.1161/01.HYP.0000148458.03202.48.

Xu, Q., Modrek, B., and Lee, C., 2002. Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. *Nucleic acids res* [Online], 30(17), pp.3754–66. Available from: https://doi.org/10.1093/nar/gkf492.

Yang, J., Yu, X., Xue, F., Li, Y., Liu, W., and Zhang, S., 2018. Exosomes derived from cardiomyocytes promote cardiac fibrosis via myocyte-fibroblast cross-talk. *Am j transl res* [Online], 10(12), pp.4350–4366. Available from: https://www.ncbi.nlm.nih.gov/pubmed/30662677.

Yang, L.Y., Ge, X., Wang, Y.L., Ma, K.L., Liu, H., Zhang, X.L., and Liu, B.C., 2013. Angiotensin receptor blockers reduce left ventricular hypertrophy in dialysis patients: a meta-analysis. *Am j med sci* [Online], 345(1), pp.1–9. Available from: https://doi.org/10.1097/MAJ.0b013e318249d387.

Yang, Y., Wang, Y., and Gao, P.J., 2020. Osteopontin associated with left ventricular hypertrophy and diastolic dysfunction in essential hypertension. *J hum hypertens* [Online], 34(5), pp.388–396. Available from: https://doi.org/10.1038/s41371-019-0246-3.

Yim, J., Cho, H., and Rabkin, S.W., 2018. Gene expression and gene associations during the development of heart failure with preserved ejection fraction in the dahl salt sensitive model of hypertension. *Clin exp hypertens* [Online], 40(2), pp.155–166. Available from: https://doi.org/10.1080/10641963.2017.1346113.

Yin, F.C., Spurgeon, H.A., Rakusan, K., Weisfeldt, M.L., and Lakatta, E.G., 1982. Use of tibial length to quantify cardiac hypertrophy: application in the aging rat. *Am j physiol* [Online], 243(6), H941–7. Available from: https://doi.org/10.1152/ajpheart.1982.243.6.H941.

Yokosaki, Y., Matsuura, N., Sasaki, T., Murakami, I., Schneider, H., Higashiyama, S., Saitoh, Y., Yamakido, M., Taooka, Y., and Sheppard, D., 1999. The integrin alpha(9)beta(1) binds to a novel recognition sequence (svvyglr) in the thrombin-cleaved amino-terminal fragment of osteopontin. *J biol chem* [Online], 274(51), pp.36328–34. Available from: https://doi.org/10.1074/jbc.274.51.36328.

Yousefi, K., Irion, C.I., Takeuchi, L.M., Ding, W., Lambert, G., Eisenberg, T., Sukkar, S., Granzier, H.L., Methawasin, M., Lee, D.I., Hahn, V.S., Kass, D.A., Hatzistergos, K.E., Hare, J.M., Webster, K.A., and Shehadeh, L.A., 2019. Osteopontin promotes left ventricular diastolic dysfunction through a mitochondrial pathway. *J am coll cardiol* [Online], 73(21), pp.2705–2718. Available from: https://doi.org/10.1016/j.jacc.2019.02.074.

Zhang, C., Zhang, B., Lin, L.L., and Zhao, S., 2017. Evaluation and comparison of computational tools for rna-seq isoform quantification. *Bmc genomics* [Online], 18(1), p.583. Available from: https://doi.org/10.1186/s12864-017-4002-1.

Zhang, H., Hu, L., and Wei, X., 2020. Prognostic value of left ventricular hypertrophy in hypertensive patients: a meta-analysis of electrocardiographic studies. *J clin hypertens (greenwich)* [Online], 22(2), pp.254–260. Available from: https://doi.org/10.1111/jch.13795.

Zhang, Z., Fan, H., Richardson, W., Gao, B.Z., and Ye, T., 2023. Management of autofluorescence in formaldehyde-fixed myocardium: choosing the right treatment. *Eur j histochem* [Online], 67(4). Available from: https://doi.org/10.4081/ejh.2023.3812.

Zhang, Z. and Wang, C., 2022. Exploring key genes and pathways of cardiac hypertrophy based on bioinformatics. *Dis markers* [Online], 2022, p.2081590. Available from: https://doi.org/10.1155/2022/2081590.

Zhao, K., Zhang, M., Zhang, L., Wang, P., Song, G., Liu, B., Wu, H., Yin, Z., and Gao, C., 2016. Intracellular osteopontin stabilizes traf3 to positively regulate innate antiviral response. *Sci rep* [Online], 6, p.23771. Available from: https://doi.org/10.1038/srep23771.

Zhao, M., Huang, W., Zou, S., Shen, Q., and Zhu, X., 2020. A five-genes-based prognostic signature for cervical cancer overall survival prediction. *Int j genomics* [Online], 2020, p.8347639. Available from: https://doi.org/10.1155/2020/8347639.

Zhao, X., Johnson, J.N., Singh, K., and Singh, M., 2007. Impairment of myocardial angiogenic response in the absence of osteopontin. *Microcirculation* [Online], 14(3), pp.233–40. Available from: https://doi.org/10.1080/10739680601139369.

Zhong, W., Mao, S., Tobis, S., Angelis, E., Jordan, M.C., Roos, K.P., Fishbein, M.C., Alboran, I.M. de, and MacLellan, W.R., 2006. Hypertrophic growth in cardiac myocytes is mediated by myc through a cyclin d2-dependent pathway. *Embo j* [Online], 25(16), pp.3869–79. Available from: https://doi.org/10.1038/sj.emboj.7601252.

Zohar, R., Lee, W., Arora, P., Cheifetz, S., McCulloch, C., and Sodek, J., 1997. Single cell analysis of intracellular osteopontin in osteogenic cultures of fetal rat calvarial cells. *J cell physiol* [Online], 170(1), pp.88–100. Available from: https://doi.org/10.1002/(SICI)1097-4652(199701)170:1<88::AID-JCP10>3.0.CO;2-K.