Genetic determinants of cardiac traits

Rizwan Ahmed
Molecular Cardiology
Faculty of Medicine
Imperial College
London, UK

Submitted for the
Degree of Doctor of Philosophy

To Imperial College, London

July 17, 2012
Abstract

**Background:** The heritability and genomic determinants of cardiac traits such as coronary flow (CF), coronary flow reserve (CFR), myocardial relaxation (LV dP/dt_{min}) and myocardial contractility (LV dP/dt_{max}) remain unknown. Brown Norway (BN) and Spontaneously Hypertensive Rat (SHR) are genetically distant inbred rat strains with distinct cardiovascular phenotypes. Genome sequence for both the strains is available making it possible to identify genetic determinants of cardiac phenotypes.

**Aims:** To establish the heritability of cardiac traits, identify novel quantitative trait loci (QTLs) underlying these traits and to prioritize candidate genes by integrating information from physiological trait analysis with genome sequence and transcriptional profiling of cardiac tissue in an intercross derived from BN and SHR.

**Methods:** Cardiac phenotyping was carried out in the parental BN and SHR strains as well as an intercross derived from these strains using *in vivo* blood pressure (BP) profiling and *ex vivo* Langendorff preparation. Genome-wide SNP genotyping was performed using a custom genotyping assay and transcriptional profiling of non-ischaemic cardiac tissue was undertaken on a subset of the intercross (n =110).

**Results:** All cardiac traits studied were at least as heritable as BP. Novel QTLs were discovered for CF, CFR, LV dP/dt_{min} and LV dP/dt_{max} as well as confirming previously documented QTLs for BP and cardiac mass. A significant proportion of trait heritability was explained by genetic variation utilizing multiple QTL models. Expression analysis revealed *cis* and *trans* genomic control regions for cardiac gene expression. Expression QTL (eQTL) hot spots on rat chromosomes 3 and 8 were shown to be significantly enriched for regulation of mitochondrial transcripts and these regions were also linked to cardiac mass suggesting correlation between mitochondrial function and cardiac mass.
1 Introduction

1.1 Cardiovascular physiology

1.1.1 Overview of cardiac structure and function

1.1.2 Coronary flow

1.1.3 Coronary flow covariates

1.2 Animal models

1.2.1 Historical perspective

1.2.2 Inbred strains

1.2.3 Rat models and genomics

1.3 QTL mapping

1.3.1 Quantitative trait loci (QTLs)

1.3.2 Statistical aspects of QTL mapping

1.3.3 Discovering QTL genes

1.4 Microarrays

1.5 Genomic databases and data mining

2 Materials and methods

2.1 Animals
## Contents

2.2 Cardiovascular physiological techniques .................................................. 38  
2.2.1 Blood pressure ...................................................................................... 38  
2.2.2 Langendorff isolated heart preparation ..................................................... 39  
2.2.3 Myocardial infarct estimation .................................................................... 42  
2.3 High throughput genotyping ........................................................................ 46  
2.3.1 DNA extraction and quantification .............................................................. 46  
2.3.2 Genotyping workflow ............................................................................... 47  
2.3.3 SNP selection for genotyping ..................................................................... 50  
2.3.4 Metrics for SNP calling and clustering ....................................................... 51  
2.4 Microarrays ................................................................................................. 56  
2.4.1 RNA extraction and quantification .............................................................. 56  
2.4.2 RNA labelling for microarrays ................................................................... 58  
Aims of present project ...................................................................................... 61  

3 Studies of cardiovascular physiology in parental strains and F<sub>1</sub> intercross 62  
3.1 In silico data analysis .............................................................. 62  
3.1.1 Aims ....................................................................................................... 62  
3.1.2 PhysGen cardiovascular phenotyping methodology .................................. 63  
3.1.3 PhysGen phenotyping data ....................................................................... 64  
3.2 Cardiovascular physiology in inbred strains ................................................. 67  
3.2.1 Aims ....................................................................................................... 67  
3.2.2 Parental strain data analysis ..................................................................... 67  
3.3 Cardiovascular physiology in F<sub>1</sub> intercross ............................................ 70  
3.3.1 F<sub>1</sub> intercross: modifications to phenotyping protocol ............................... 70  
3.3.2 Parental strains and F<sub>1</sub> intercross: data analysis ................................... 71  
3.4 Discussion .................................................................................................... 76  

4 Cardiovascular physiology in F<sub>2</sub> intercross ............................................ 80  
4.1 Analysis of phenotypic drift ......................................................................... 80
Acknowledgements

The work presented herein would not have been possible without the generous support of my supervisor Prof. Stuart Cook. I am greatly indebted to him for the opportunity he provided, the way he nurtured my nascent scientific career and at times for the improbably high ratio of his patience to my aptitude. I consider myself fortunate to have access to an open-door, friendly and honest opinion even on Sunday nights.

I am truly fortunate to have Mr Phil Muckett as biotechnician for the group. This project would not have been feasible without his expertise in animal husbandry and his awareness of the legal requirements underlying animal work. I am also grateful to all my colleagues who supported me throughout the ups and downs typical of research.

This work owes considerable debt to some people I have not known personally but whose hard work has played a pivotal role in providing the tools to carry out this work. The developers of the statistical program “R” and document typesetting program \LaTeX fall in this category. They have made available their time, effort and expertise \textit{freely} for the wider research community. Special mention must be made here of Dr Karl Bromann (Professor of Biostatistics, University of Wisconsin), writer of the R package, \texttt{qtl}, who I have asked for advice on several occasions without failing to receive it generously.

Finally I want to express my gratitude to my wife, Shehnila, whose patience and willingness to take on my share of domestic responsibilities allowed me to follow my curiosities “down the rabbit-hole” [Lewis Carroll] of academia. She is always last in my acknowledgements but foremost in my thoughts.

This work was kindly funded by grant from the British Heart Foundation.
List of Publications

**Paper**


**Abstracts**

**American Heart Association, Orlando, November 2011**

A Novel Quantitative Trait Locus for Coronary Flow. [R Ahmed](#), Stuart A. Cook

**European Society of Cardiology, Paris, August 2011**

Novel cardiovascular quantitative trait loci in a rat intercross. [R Ahmed](#), P Muckett, Stuart A. Cook

**British Cardiovascular Society, Manchester, June 2011**

Heritability of coronary flow reserve (Heart, June 2011 Vol 97 Suppl. 1). [R Ahmed](#), P Muckett, Stuart A. Cook

**British Heart Foundation Fellows meeting, London, February 2011**

Cardiovascular phenomics approach identifies a number of quantitative trait loci for clinically important cardiac traits. [R Ahmed](#), Stuart A. Cook
Chapter 1

Introduction

1.1 Cardiovascular physiology

1.1.1 Overview of cardiac structure and function

The mammalian heart is a four chamber organ comprising two pumps, the right heart consisting of right atrium and ventricle which receives venous return and pumps blood to the pulmonary circulation and the left heart which receives oxygenated pulmonary venous return and pumps blood into the systemic circulation via aorta. Two coronary arteries, left and right, arise from the base of the aorta opposite the corresponding aortic cusps and course along the left and right atrio-ventricular grooves respectively. The left coronary artery subdivides into a left anterior descending artery (LAD) which courses along the front of the heart in the interventricular groove from the base of the heart to the apex and a left circumflex artery (LCx) which curves to the left side of the heart. Cardiac muscle cells are connected to each other through specialised “gap junctions”. This allows the cardiac muscle to connect in series and in parallel as well as allowing rapid diffusion of ions, forming a syncytium which contracts and relaxes as a whole. Cardiac musculature forming the atria and the ventricles are separated by a fibrous ring that electrically insulates the two, this fibrous ring is traversed by specialised conduction tissue providing electrical connectivity between the two chambers.

Shortening of the cardiac sarcomere is triggered by an action potential which depo-
larises the cell membrane. Cardiac cells have two types of voltage-gated channels, fast Na$^+$ and slow Ca$^{2+}$ channels. Slower opening and closure of voltage-gated Ca$^{2+}$ channels is responsible for the prolonged action potential typical of the cardiac sarcomere. Cell membrane depolarisation leads to release of Ca$^{2+}$ from the sarcoplasmic reticulum, this effect being potentiated by additional Ca$^{2+}$ influx through the T-tubules and activation of ryanodine receptor channels causing a transient “Ca$^{2+}$ spark”. The Ca$^{2+}$ ions catalyse the reactions that lead to actin and myosin filaments sliding along each other producing a muscle contraction. At the end of contraction the voltage-gated channels close, preventing further Ca$^{2+}$ influx and sarco/endoplasmic reticulum Ca$^{2+}$-ATPase (SERCA) rapidly removes Ca$^{2+}$. This active energy-dependent process results in cardiac relaxation. Oxygen and nutrient substrate required for these energy-dependent processes is provided by coronary flow (CF) which is tightly coupled to cardiac metabolism.

1.1.2 Coronary flow

Physics of coronary flow

A brief review of flow physics is required prior to discussion of CF. Flow through a blood vessel is determined by Ohm’s Law

\[ F = \frac{\Delta P}{R} \] (1.1)

where \( F \) represents flow, \( \Delta P \) represents pressure differential between two ends of the vessel and \( R \) is resistance to flow. The inverse of resistance is termed “conductance” which is a measure of flow through a vessel for a given pressure differential. Conductance is proportional to fourth power of the vessel diameter:

\[ \text{Conductance} \propto \text{Diameter}^4 \] (1.2)

This relationship is important in that a doubling in vessel diameter will result in a 16-fold increase in flow. In coronary circulation, as in systemic circulation, two-thirds of
the flow resistance resides in small vessels with muscular walls [Camici and Crea, 2007, Beltrame et al., 2009]. The muscular walls of these vessels allow the vascular diameters to change in response to metabolic requirements. The relationship between conductance and diameter means that a small change in vessel diameter results in greatly increased flow. Microvascular impairment results in inability of these small vessels to change the diameter in response to increased demand and results in reduced CF. Another important feature of the coronary circulation is that the vessels subdivide progressively to form parallel circuits. In a parallel circuit the total resistance is given as

\[
\frac{1}{R_{\text{total}}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \cdots + \frac{1}{R_N}
\]  

(1.3)

The concept of parallel circuits is important as any condition, genetic or environmental that leads to poor development of the microvasculature or loss of already developed microvasculature would result in reduction in the number of parallel circuits and reduced conductance within the system.

*In vivo* or *ex vivo* CF in isolated heart preparations can be measured by ultrasonic flow meter. Ultrasound waves are transmitted by the probe and reflected back at a different frequency. The reflected waves which have undergone Doppler shift are received by the probe and fluid velocity is calculated by the difference between frequencies of transmitted and received ultrasound waves. Flow is then calculated by taking the diameter of the vessel into account. The difference between flow and velocity is important when evaluating the literature on CF since both measures are used to describe CF. Velocity is the rate of travel of fluid per unit time whereas flow is the volume of fluid that travels per unit time. Velocity (V) and flow (F) are related by the following equation

\[
F \propto Vr^2
\]  

(1.4)

Where r is the radius of the vessel. Therefore velocity derived from Doppler measurements may represent flow if the radius of the vessel remains unchanged. The equation can be used to demonstrate that velocity and flow are not synonymous; blood flow in a narrowed
artery has a higher velocity but lower flow rates than blood flow in a normal artery. The larger lumen of the normal artery allows greater flow but has lower velocities. Higher CF velocities in patient with hypertension Houghton et al., 1990 could simply reflect narrower vessel lumen than truly reduced CF unless vessel radius is taken into account.

Another flow measure that is often reported in literature is coronary flow reserve (CFR). CFR is the ratio of maximum hyperaemic CF to baseline CF. It reflects microvascular capacitance and endothelial function. Maximum hyperaemia is induced using vasodilatory agents like adenosine or dipyridamole. It is unclear if all vasodilators are pharmacologically equivalent Holdright et al., 1993. For ex vivo heart preparations, temporary ischaemia (for one minute) is very effective in eliciting a hyperaemic response.

**Functional subdivisions of coronary arterial tree**

The coronary arterial system is functionally subdivided into three components

- The conduit epicardial vessels are 0.5 - 5 mm in diameter and provide little resistance to blood flow in the healthy heart.

- The pre-arterioles are 0.1 - 0.5 mm in diameter and play an important role in coronary resistance. Due to their extramural position, the pre-arterioles are not exposed to the local metabolic milieu.

- The microvessels are 0.01 - 0.1 mm in diameter and are responsible for 60% of the resistance to the blood flow Camici and Crea, 2007. Flow in the microvessels is affected by the local metabolic environment. Important regulators of blood flow at this level are adenosine, K_{ATP}-channels, nitric oxide and prostaglandins Tune et al., 2004.

**Regulation of coronary flow**

CF is a well studied phenotype from a physiological perspective but little is known about contribution of genetic variation to CF. From the discussion on physics of CF it is obvious that in addition to metabolic modulators any factor that affect the number, development
or size of microvessels will significantly affect CF. CF is also tightly coupled to myocardial metabolic demand although the underlying mechanisms remain largely unknown. Important regulators of CF include hypoxia, nitric oxide (NO), adenosine, endothelin, $K_{\text{ATP}}$ channels and sympathetic nervous system [Tune et al., 2004].

NO is a lipophilic gas produced from L-arginine by NO synthase (NOS) by endothelial cells. NO is a pleiotropic molecule with effects on vascular tone, angiogenesis, immune response and neurotransmission but here the discussion is limited to its effects on CF. NO release is potentiated by vascular shear stress and once released into the circulation NO has half-life of a few seconds, therefore its vascular effects are predominantly local. Being lipophilic, it diffuses into the endothelium and activates guanylyl cyclase leading to production of cGMP which in turn activates mechanisms for smooth muscle relaxation. Drugs like glyceryl tri-nitrate which release NO in the circulation are used clinically in patients with angina to promote coronary artery vasodilatation but it is unclear if NO is the factor that matches CF to myocardial metabolism since inhibition of NO with arginine analogs does not significantly reduce CF [Egashira et al., 1996, Bernstein et al., 1996]. NO seems to exert a tonic vasodilatory influence on the large epicardial coronary vessels.

Adenosine is a purine nucleoside (adenine attached to a ribose sugar) that was first implicated in metabolic control of CF in 1963 [Berne, 1963]. It mediates its actions via adenosine receptors ($A_1$, $A_2A$, $A_2B$ and $A_3$). The adenosine hypothesis for CF modulation proposes that reduced CF results in increased cardiomyocyte release of adenosine which then acts on vascular smooth muscle cells to promote vasodilatation. Numerous studies have investigated the role of adenosine as local metabolic mediator of CF during periods of increased metabolic demand such as increased heart rate and during ischaemia. There is convincing evidence that adenosine promotes coronary vasodilation during ischaemia [Laxson et al., 1993]. In this regard effects are pleiotropic as it also decreases myocardial contractility [Dobson and Schrader, 1984, Dobson et al., 1986], reduces cardiac arrhythmias [Lerman and Belardinelli, 1987] and reduces myocardial infarct size [Uematsu et al., 1998]. However evidence for adenosine as the physiological vasodilatory metabolite is lacking [Tune et al., 2000, Tune et al., 2004].
1.1. Cardiovascular physiology

$\text{K}_{\text{ATP}}$ channels are present in various tissues including vascular smooth muscle, arterial endothelium and cardiac muscle [Standen et al., 1989]. $\text{K}_{\text{ATP}}$ channel activation results in membrane hyperpolarization and relaxation of smooth muscle cells [Nichols and Lederer, 1991]. Studies have shown that $\text{K}_{\text{ATP}}$ channels contribute to basal coronary vascular tone [Samaha et al., 1992, Richmond et al., 1999] but are not required for local metabolic vasodilation during periods of increased myocardial oxygen consumption [Stepp et al., 1997]. Prostaglandins are arachidonic acid metabolites. They are released into the coronary circulation during coronary artery occlusion [Alexander et al., 1975]. Prostaglandins do not play an obvious role in physiological coronary blood flow responses [Dai and Bache, 1984]. The role of prostaglandins in modulating CF during ischaemia is at present unclear with some evidence for [Duffy et al., 1999] and some against [Pacold et al., 1986] a role in this process.

$\beta$-adrenoreceptor mediated vasodilation (mediated by intracellular activation of cAMP by receptor coupled $G_s$ protein) is a feedforward mechanism. Quantitative analysis has shown has this mechanism accounts for 25% of coronary exercise hyperaemia [Tune et al., 2004]. Interestingly $\alpha$-adrenoreceptor mediated vasoconstriction also contributes to regulation of CF during periods of increased myocardial demand such as tachycardia by vasoconstriction of large epicardial vessels which is postulated to increase tone and vascular stiffness leading to reduced retrograde coronary blood flow and increasing subendocardial blood flow.

Recently transient receptor potential vanilloid type 4 (TRPV4) channel, a $\text{Ca}^{2+}$-permeable cation channel has been shown to be involved in flow-mediated dilatation of coronary vessels [Bubolz et al., 2012]. This receptor is expressed in the endothelium of human coronary arteries [Bubolz et al., 2012], is activated by arachidonic acid metabolites [Vriens et al., 2005] and is associated with angiogenesis [Troidl et al., 2009, Schierling et al., 2011]. Bubolz et al. showed that in cultured human coronary artery endothelial cells (HCAECs) TRPV4 channel activation resulted in coronary artery vasodilatation via increased reactive oxygen species (ROS) and $\text{H}_2\text{O}_2$ production that was inhibited by TRPV4 antagonists but the intracellular pathways require further elucida-
Correlates of coronary flow impairment

From a clinical perspective CF may be impaired due to atherosclerotic narrowing as a consequence of hypertension, diabetes mellitus, smoking or hyperlipidaemia. There are a number of clinical conditions in which CF impairment occurs in the absence of atherosclerosis presumably as a result of adverse vascular remodelling [Camici and Crea, 2007, Beltrame et al., 2009], the pathophysiology of which remains undefined. Coronary vascular dysfunction has been reported in hypertrophic cardiomyopathy (HCM) [Maron et al., 1986], dilated cardiomyopathy [Neglia et al., 2002], Fabry’s disease which is due to X-linked deficiency of lysosomal α-galactosidase A as well primary coronary slow flow phenomenon [Tambe et al., 1972].

The genetic aspects of variation in CF have not been studied outside knockout models [Teng et al., 2008] and are the primary studies presented in this thesis which will also determine the covariate effects of hypertension, cardiac hypertrophy and myocardial relaxation on CF.

1.1.3 Coronary flow covariates

Hypertension

Whilst there is considerable evidence that hypertension correlates with impaired CF there is no conclusive evidence of direct cause and effect [Levy et al., 2008]. High BP may cause impaired vascular flow or vice versa and there is experimental evidence to support both views. Boegehold et al. caused experimental coarctation of the abdominal aorta above the origin of the renal arteries in rats. This caused hypertension above the site of coarctation but the arteries below the coarctation were not exposed to high BP which was measured by Femoral pressure catheter. Vascular densities were measured in cremasteric muscle and were found to be 19% lower than in sham-operated rats [Boegehold et al., 1991]. Nabha et al. showed that hypertension in spontaneously hypertensive rats (SHR) can be delayed by treatment with anti-oxidants [Nabha et al., 2005]. Antonios et al. showed that microvas-
1.1. Cardiovascular physiology

cular rarefaction occurs in normotensive offspring of individuals with essential hypertension suggesting that vascular changes predate clinical hypertension [Antonios et al., 2003].

Perhaps the most important observation underlying impaired flow and clinical covariates such as hypertension is the presence of inflammation at the level of microvasculature. Inflammation is also the unifying factor between impaired flow and hyperglycaemia [Rizzoni et al., 2001], obesity [Levy et al., 2008] and smoking and thus it would be reasonable to hypothesize that vascular impairment is a consequence of inflammation and generation of ROS whether primary as a result of genetic factors or secondary to a number of pathogenic stimuli. This inflammation may alter the balance of pro/anti angiogenic factors at the level of microcirculation [Goligorsky, 2010].

Cardiac hypertrophy

Cardiac hypertrophy may occur as a primary entity in cardiomyopathies such as HCM or may be secondary to conditions in which cardiac afterload increases such as aortic stenosis. When cardiac hypertrophy occurs in the context of hypertension, it may difficult to determine if impairment in CF is due to hypertension or cardiac hypertrophy or both. Houghton et.al studied CFR in patients with a high prevalence of hypertension (87% of the patients in the cohort were hypertensives) and did not find any correlation of CFR with cardiac hypertrophy [Houghton et al., 1990]. Schunkert et al. showed that experimental aortic stenosis in rats results in increased angiotensin converting enzyme (ACE) activity, which causes dose-dependent decrease in CFR in rats from both experimental and control groups [Schunkert et al., 1990]. Whilst these results suggest that CFR is independent of cardiac mass, the same conclusion is difficult to prove for CF and cardiac mass. During studies such as positron emission tomography (PET) [Bengel et al., 2009] scan where absolute CF is quantified, it is corrected for cardiac mass and represented in “ml/gram/min” units. Therefore direct comparisons between absolute CF and cardiac hypertrophy are often not possible.
Myocardial relaxation

During systole cardiac muscle contraction impedes CF and there is a little back flow into the aorta. CF is primarily a diastolic phenomenon and occurs when cardiac muscle is relaxing. LV diastolic impairment could therefore also play a role in impaired CF. This was first demonstrated by Galderisi et al. who showed that in newly diagnosed untreated hypertensive patients low CFR correlates with diastolic dysfunction as measured by $E_m/A_m$ ratio on tissue doppler imaging (TDI), $p < 0.0005$. $E_m$ is the rate of movement of myocardial wall during early systole and $A_m$ is the rate of movement of myocardial wall during atrial systole. Interestingly BP was not different between the two groups with impaired and normal CFR [Galderisi et al., 2002] indicating a BP-independent mechanism. Karayannis et al. studied CFR is patients with a variety of cardiovascular risk factors including hypertension, diabetes mellitus, smoking and hyperlipidaemia and found that diastolic function correlates significantly with CFR [Karayannis et al., 2011]. In this regards important evidence is provided by MacCarthy et al. [MacCarthy and Shah, 2000], who demonstrate that an increase in ROS results in impaired myocardial relaxation in guinea pigs. They also demonstrated that anti-oxidant treatment reverses this abnormality analogous to delay in hypertensive phenotype in SHRs post anti-oxidant diet [Nabha et al., 2005]. This places inflammation firmly at the centre of various cardiac pathological phenotypes (figure 1.1).

1.2 Animal models

1.2.1 Historical perspective

The genus Rattus belongs to the super-family Muridae which also includes mice and gerbils. Rats and mice are thought to have diverged from a common ancestor approximately 24 million years ago. The common rat is known as Brown Norway or Rattus Norvegicus. The rat is thought to be the first mammalian species domesticated for scientific research. The first known experimental use of rat was a study on the effects of adrenalectomy in albino rats in France in 1856. Helen Dean King, a research scientist working with Henry
1.2. Animal models

Figure 1.1 Schematic representation of the pathophysiological mechanisms underlying CF. Immune system activation increases ROS generation in response to a number of pathogenic stimuli leads to reduce NO bioavailability. This causes alterations in a number of key cardiac phenotypes. The disturbance may lead to a positive feedback loop perpetuating the cycle. The genetic determinants of these processes remain unknown.

H. Donaldson at the Wistar Institute established the first inbred rat strain in 1909 by series of 38 brother-sister matings. This strain was named after Helen Dean King as King Albino. She also captured wild Brown Norway (BN) rats and proceeded to establish the inbred Brown Norway strain in 1934 [Suckow et al., 2006]. These and other rat strains initially developed at the Wistar Institute have been the focus of biomedical research for a century. Historically, rat strains have been selectively bred for traits relevant to human disease, e.g. autoimmune diseases, hypertension, stroke and obesity. At present more than 230 inbred rat strains are available and they capture varied aspects of human pathobiology [Jacob and Kwitek, 2002].
1.2. Animal models

1.2.2 Inbred strains

Inbred rat strains comprise of individuals that are genetically identical to each other following a long history of controlled inbreeding. With each generation of brother-sister mating a certain portion of alleles become fixed and this pool of fixed alleles expands with each generation. This reflects the fact that all fixed loci will remain unchanged whereas all unfixed loci will have a certain chance of becoming fixed at each generation. After 20 generations of continued brother-sister matings 98.7% of the loci in the genome are homozygous and at this point the strain is determined to be inbred.

Because all individuals of an inbred strain are genetically identical the genotype of any individual belonging to an inbred strain is known. This allows replication, comparison and confirmation of experimental data at different times and in different places. Controlling the genetic background also allows phenotypic uniformity thereby reducing experimental variation and number of animals required to achieve statistical significance in experimental studies. Inbred rat strains have individual properties which are relevant to different human disease states e.g left ventricular hypertrophy in Spontaneously hypertensive rat (SHR) strain and stroke in stroke-prone SHR strain. Perhaps most advantageous is the fact that current advances in genome sequencing technology have provided the opportunity to annotate the rat genome with the phenotypes [Kwitek et al., 2006].

1.2.3 Rat models and genomics

The availability of the finished sequence of the human genome has enabled association of phenotypes with refined regions of the genome. Geneticists often prefer the mouse as a model because of simple housing requirements, breeding ability, shorter lifespan and the ability to manipulate the germline through transgenic and knockout technology. However, physiologists have traditionally preferred the rat as a model because of the larger size and better correlation with human physiology.

In recent times rat genetics have made significant advances including publication of the genome sequence of Rattus Norvegicus [Gibbs, 2004] and SHR [Atanur et al., 2010]. The genome resources currently available for the rat have accelerated the process of gene
discovery and understanding of mechanisms of gene action [Aitman et al., 2008]. These advances and the observation that rats provide an in-depth phenotyping system that has remarkable relevance to human disease phenotypes provide an unprecedented opportunity for understanding of human disease phenotypes and assigning function to the human genome [Jacob and Kwitek, 2002].

CF has been studied in the SHR, mostly comparing CF in SHR with the Wistar-Kyoto rat strain (WKY). WKY is the parent strain from which SHR was derived [De Jong, 1984]. SHR rats are pre-hypertensive with systolic BP around 110 - 120 for the first eight weeks of life following which there is gradual onset of hypertension over the next 12 weeks. A substrain of SHR was found to have a greater incidence of thrombotic stroke and was bred to form the SHR stroke prone strain (SHRSP). SHRSP has genetic predisposition to stroke beyond that acts additively with hypertension [Nagaoka et al., 1976].

CF differences between SHR and its parental strain, WKY, were studied in response to serotonin [Luscher et al., 1986], ischaemia [Haneda et al., 1986] and vasodilator drugs including acetylcholine and bradykinin [Pourageaud and Freslon, 1995]. These studies showed that CF in SHR is impaired in response to vasodilator stimuli and ischaemia compared with WKY. Various hypotheses have been proposed to explain these differences, chiefly being the deleterious effect of hypertension on the vascular architecture leading to vascular hypertrophy, fibrosis and impaired endothelial function. Brilla and colleagues [Brilla et al., 1991] showed that regression of vascular medial hypertrophy in SHR rats treated with an anti-hypertensive drug results in some improvement in CF, as might be expected. These studies have established SHR as a model for impaired CF but the pathobiology underlying this impairment have not been established.

In relation to cardiac phenotypes the BN rat strain is a normotensive strain that also has normal cardiac mass, insulin resistance and is the reference genome for rat. BN has been used as a model for resistance to myocardial ischaemia [Baker et al., 2000]. Additional cardiac phenotypes including CF have been studied in BN as part of the rat phenome project [Kwitek et al., 2006]. Data from this project is publicly accessible, http://pga.mcw.edu.
1.3 QTL mapping

Quantitative traits

Mendel’s experiments on peas [Mendel, 1950] and Morgan’s experiments on Drosophila eye colour [Morgan, 1910] are textbook examples of qualitative traits. However, most traits of medical importance are non-Mendelian complex traits. Complex traits are phenotypic traits that do not exhibit classic Mendelian dominant or recessive inheritance. Examples of complex traits are susceptibility to common diseases like diabetes and ischaemic heart disease [Lander and Schork, 1994]. Complex traits measured as a continuous variable are termed quantitative traits e.g. coronary flow, heart rate, systolic and diastolic pressure, in contrast to discrete traits e.g development of schizophrenia. There are two main reasons why traits are continuous: first, environment and developmental noise leads to a variation in phenotypes resulting in ambiguous relationship between genotype and phenotype [Hansen and Gardner, 1962] and second there may be many segregating loci having alleles that contribute to the phenotype. In fact, the number of alleles need not be great to result in a continuous phenotype. The study of continuous traits is termed quantitative genetics. The fundamental questions that quantitative genetics [Griffiths, 2008] concerns itself with are:

- Is the observed variation in phenotype due to genetic variation?
- How much does genetic variation contribute to total phenotypic variation - termed “heritability”?
- What are the effects of the environment on phenotypes given constant genotype - termed “norms of reaction”?
- What genomic loci contribute to the traits? What are their effect sizes?
- What is the biological explanation of the QTL influence on trait?
Heritability

The question, “Is the observed variation in phenotype due to genetic variation?” can be informed by measuring the heritability of the trait. Heritability is the proportion of phenotypic variance that is explained by genotypic variance for a given population in a specific environment. If the observed variance in phenotypes ($\sigma^2_P$) is the sum of the genotypic variance ($\sigma^2_G$) and environmental variance ($\sigma^2_E$) then broad sense heritability ($H^2$) is ratio of total genotypic and phenotypic variation [Visscher et al., 2008],

$$\sigma^2_P = \sigma^2_G + \sigma^2_E$$  \hspace{1cm} (1.5a)

$$H^2 = \frac{\sigma^2_G}{\sigma^2_P}$$  \hspace{1cm} (1.5b)

The total genetic variance can be sub-divided into genetic variance due to additive loci ($\sigma^2_A$), dominant loci ($\sigma^2_D$) and epistatic loci ($\sigma^2_I$) which is defined as the interaction between two or more alleles leading to altered gene expression. This leads us to define the narrow sense heritability ($h^2$) which is the proportion of phenotypic variance that can be ascribed to additive loci,

$$\sigma^2_G = \sigma^2_A + \sigma^2_D + \sigma^2_I$$  \hspace{1cm} (1.6a)

$$h^2 = \frac{\sigma^2_A}{\sigma^2_P}$$  \hspace{1cm} (1.6b)

It is important to emphasize that heritability is plastic and varies with the population and environment. The can be illustrated by pellagra, a disease caused by nutritional deficiency of the vitamin niacin. Pellagra was considered to be a heritable disease in 1910 and this was in part the consequence of poor socio-economic conditions in the American South which caused familial clustering of the disease. Heritability for Pellagra was reported in some studies to be 40% [Green, 1916] [Joseph, 2000]. Later on it was proven that the disease is caused by nutritional deficiency [Goldberger and Wheeler, 1990] and not “genetic”. The heritability estimate was erroneous because the effects of the environment and diet
1.3. QTL mapping

were not accounted for.

Heritability is an important predictor of the genetic variation that underlies the quantitative phenotypes as long as the limitations on interpretation of heritability are understood. Estimation of heritability is easier in animal crosses compared with human populations where siblings or half-siblings are used to estimate heritability. In a typical F$_2$ intercross experiment in animals where the intercross is derived from inbred parental strains, the genotypes in both parents and F$_1$ are constant. Therefore any excess phenotypic variation in the parents or the F$_1$ is experimental error. In contrast, in the F$_2$ intercross, because of segregated genotypes, phenotypic variation is attributable to both experimental and genetic factors. Broad sense heritability can be estimated as follows,

\[
H^2 = \frac{\sigma_{F_2}^2 - \left( \frac{\sigma_{\text{parent}1} + \sigma_{\text{parent}2} + \sigma_{F_1}}{3} \right)}{\sigma_{F_2}}
\]  

(1.7)

Genetic linkage

Genetic linkage refers to the observation that two or more genes located in proximity on a chromosome undergo less frequent recombination. In contrast genes located on different chromosomes are inherited independently of each other. Recombination describes the independent assortment of genes at meiosis. The production of new allele combinations leads to a recombinant product. The proportion of products that are recombinants is the recombination frequency (RF).

Alfred Sturtevant first developed the basic method of mapping genes using recombination frequency [Sturtevant, 1913]. The idea being that the further apart two genes are, the greater the chance that recombination will occur between them and the greater the number of recombinants produced as a consequence. Hence higher recombination frequency (RF) indicates that genes are far apart and vice versa. Sturtevant defined “one map unit” as a distance between genes for which 1% of meiosis are recombinant. A map unit is also called a centimorgan (cM) after Thomas Hunt Morgan. RF expressed in cM approximately correlates with the physical map distance expressed in megabases but this relationship is non-linear, especially at higher resolution. This is due to the fact that the
distribution of recombination sites is not random across the genome but tends to cluster in regions of a few kilobases called “recombination hotspots” [Steinmetz et al., 1987].

Constructing linkage maps by marker loci

Differences in a single nucleotide at a particular location in DNA are called single nucleotide polymorphisms (SNP), occurring at approximately 1 in a 1000 basepairs (bp) in the human genome. Other DNA polymorphisms consist of blocks of repetitive DNA called microsatellites which are short (1-4 base pairs repeated 2-10 times) or minisatellites which are longer (15-100 base pairs repeated 10 to 100 times), together termed as variable number of tandem repeats (VNTRs). VNTRs have multiple alleles that can be tracked in a pedigree. The use of DNA polymorphism to construct a genetic map was first suggested by Botstein and Lander [Lander and Botstein, 1989]. They proposed use of Restriction Fragment Length Polymorphism (RFLPs) as genetic markers which can be tested for linkage to inherited traits. Later, the discovery of highly abundant and polymorphic VNTRs [Jeffreys et al., 1985; Weber and May, 1989] provided the ideal marker loci for constructing linkage maps since they can be determined relatively rapidly by polymerase chain reaction (PCR).

1.3.1 Quantitative trait loci (QTLs)

The region of a genome that contains a gene (or genes) that contribute to a quantitative trait is called a quantitative trait locus (QTL). To discover genes underlying quantitative traits, the QTL in which the gene or genes reside is first identified using linkage analysis. Isolation of gene or genes contributing to phenotypes associated with a QTL poses specific challenges [Lander and Schork, 1994] because of genetic heterogeneity, incomplete penetrance, phenocopies and epistasis. These challenges are overcome in part by use of inbred lines where genotypes are identical, use of standard environmental conditions and precision of phenotyping.

Once a QTL is identified it is usual for the QTL interval to be in the region of 10-30 cM. A genetic interval of this size has approximately 100-300 genes in humans. Glazier
et al [Glazier et al., 2002] have proposed working criteria for discovering genes underlying complex traits which are as follows:

1. Establish statistically significant genome-wide evidence for linkage in a single study. This main aim may be subdivided in the following steps: QTL detection, confidence regions for location of QTL and estimating QTL effect size.

2. Reducing the size of the QTL interval by using genetic or bioinformatic strategies.

3. Sequence analysis of genes within the QTL interval to identify candidate nucleotide variants.

4. Functional tests of candidate genes.

Choice of genetic cross for QTL mapping

The parental strains chosen for a QTL mapping experiment in rodents must be inbred and ideally phenotypically the most divergent strains. These parental strains are then crossed to produce F\textsubscript{1} hybrids. The choice of further breeding depends on either mating F\textsubscript{1} to one of the parent strains to produce a backcross (BC) or to fellow F\textsubscript{1} to produce F\textsubscript{2} intercross.

Both of these breeding schemes have advantages and drawbacks. A backcross is simpler because each of the backcrossed animal only has two possible genotypes at a given locus. An intercross has three possible genotypes at each locus. More backcross progeny are required to detect a major QTL [Darvasi, 1998] compared with an intercross but the significance threshold for QTL detection in an intercross is higher (LOD score 4.3) compared with a backcross (LOD score 3.3). In the absence of specific knowledge about the mode of action of a QTL an intercross is the better choice since it segregates all possible genotypes and allows for detection of QTL with any mode of action (additive, dominant or recessive) [Broman and Sen, 2009].
1.3.2 Statistical aspects of QTL mapping

Linear regression model

In the linear regression model for QTL mapping, each molecular marker is considered once at a time [Broman, 2001]. The individuals are split according to their genotypes at the marker genotype. Phenotype averages for the different genotypes at that particular marker are then compared. Although this is a simple method affording the inclusion of covariates and no requirement for a genetic map, there are important disadvantages. Individuals with missing genotype information must be excluded, QTL position is difficult to determine, and because of incomplete linkage of QTL to marker (due to recombination) apparent QTL effect is reduced.

Interval mapping

The disadvantages of marker regression are largely overcome by use of interval mapping [Lander and Botstein, 1989]. Interval mapping makes use of two genetic markers jointly in detecting the QTL location and effect. The conditional probability of the QTL genotype is calculated given the genotypes of flanking molecular markers. If the marker genotype data is missing then the interval mapping method takes into account the genotype of the next molecular marker. This method requires specifically designed software [Broman and Sen, 2009].

LOD score

Logarithm of odds (LOD) score is the log-likelihood ratio of the data under the hypothesis that there is a QTL at a putative location in the genome compared to the hypothesis of there being no segregating QTL [Lander and Kruglyak, 1995] at that locus. Let us assume that a QTL is located at location z in the genome, flanked by genotype markers A and B with mean phenotypes $\mu_a$ and $\mu_b$ respectively and phenotype standard deviation, $\sigma$. In interval mapping maximum likelihood estimates (values for which the probability for the given data achieve its maximum) of the three parameters $\hat{\mu}_a$, $\hat{\mu}_b$ and $\hat{\sigma}$ are used to give
1.3. QTL mapping

the LOD score estimate \[\text{Broman, 2001}\]

\[
\text{LOD} (z) = \log_{10} \frac{\Pr (\text{Data} \mid \text{QTL at } z, \hat{\mu}_a, \hat{\mu}_b, \hat{\sigma})}{\Pr (\text{Data} \mid \text{no QTL})} \tag{1.8}
\]

Higher LOD scores correspond to greater evidence for presence of a QTL. The LOD score is calculated at each position of the genome. LOD score significance thresholds are determined by the experimental cross and can be determined by permutation tests given the genotype data. For an F\(_2\) intercross QTL mapping study, a LOD score of 4.3 was suggested to represent significant linkage by Lander et al \[\text{Lander and Kruglyak, 1995}\]. Broman \[\text{Broman, 2001}\] also suggests a customizable threshold for each mapping experiment based on the permutation test where the phenotypes are randomized to the genotypes and LOD scores are calculated across the genome for randomized phenotypes. This process is then iterated 1000 times. The observed LOD score is then compared against the LOD scores obtained from the 1000 permutation test. The proportion of the permutation LOD scores that exceed observed LOD score are reported as \(p\) value.

1.3.3 Discovering QTL genes

QTL mapping is a powerful method for localizing novel loci linked with the physiological or disease phenotypes. There have been numerous successful studies mapping QTL for a number of diseases in humans, rats and mice. The challenge in QTL approach lies in identifying the causative gene, called QTL-gene, or at least narrowing down the list to a few genes that could be rigorously tested after the QTL has been found and localized. This is a formidable task \[\text{Flint et al., 2005}\] evidenced by the fact that although the rat genome database, \[\text{http://rgd.mcw.edu/}\] lists > 1000 QTLs in the rat genome and the Mouse Genome Informatics website, \[\text{http://www.informatics.jax.org/}\] lists > 2000 QTLs in the mouse genome only a handful of these QTLs, less than 1\% by one estimate, have translated into a successful causative gene discovery. Flint et al. \[\text{Flint et al., 2005}\] show that successfully cloned QTL-genes have large phenotype effects (the proportion of phenotype that could be ascribed to the gene) and sequence variation that changes protein function.
At present the approaches to discovering QTL-genes can be broadly divided into two main categories genetic and bioinformatic. In most instances adoption of these strategies result in significant narrowing of the QTL interval and a much smaller list of candidate genes to be considered for functional characterization.

**Genetic strategies**

Genetic strategies to discovering QTL-genes involve different breeding strategies after a QTL has been discovered using a backcross or F\textsubscript{2} intercross. One such breeding strategy involves producing recombinant inbred (RI) strains. Inbred strains are crossed to produce F\textsubscript{1} and F\textsubscript{2} generations, pairs of F\textsubscript{2} generation are then intercrossed using sib-mating for 20 generations to produce RI strains [Bailey, 1971]. Each RI strain panel is a mosaic of the genomes of parental inbred strains. Once a panel of RI strains has been genotyped, the marker data can be used for all subsequent phenotyping experiments. When compared with F\textsubscript{2} intercross, RI strains can be considered to be “immortal” since they are a perpetual resource with a fixed genotype whereas each F\textsubscript{2} animal is unique and exists only once. The limitations for the RI strains are expensive maintenance and housing requirements as well as the number of genetically distinct panels of RI strains called “lines”. Most RI panels have about 30 or so lines and the ability to map QTLs is a function of the number of lines. Although it would be ideal to have more lines in a given RI panel but nevertheless they have been used successfully for mapping QTLs [Morrissey et al., 2011].

Producing a **congenic** strain which carries a chromosomal region from one parent to the genetic background of another strain is another method for narrowing QTLs that was pioneered by G. D. Snell [Snell, 1948]. It was known at the time that tissue transplanted between two strains could be compatible or incompatible but the genetics of this compatibility were not known and in absentia, the locus was named “H” locus. An outbred strain was mated to an inbred partner to produce F\textsubscript{2} which were then crossed with the inbred strain and tissue transplantation performed. The incompatible offsprings were then selected for next round of backcrossing to the inbred parent. At each mating the
1.3. QTL mapping

genomes became increasingly identical apart from the H locus which was being selected for incompatibility. Using this approach two loci responsible for incompatibility, H1 and H2, were “trapped” to narrow genomic regions. The limitation of this method is that considerable time and resources are required to produce a congenic, usually around a year for rat strains and the QTL position needs to be known before the strategy can be applied.

**Chromosome substitution strains (CSS)** [Singer et al., 2004] are derived using a similar method to the construction of congenics. Availability of complete genetic maps has facilitated construction of CSS by allowing tracing of inheritance throughout the genome. CSS construction has only become feasible with the availability of complete genetic maps that can be used to trace inheritance throughout the genome and like the RI strains also serve as a permanent genetic resource. Likewise, creation and maintenance of CSS strains demands significant resources and time.

**Bioinformatic strategies**

There are a number of bioinformatic approaches that help the discovery of QTL genes [DiPetrillo et al., 2005] by both reducing the QTL interval and prioritizing genes in the interval. One approach is to use synteny by comparing the genomic region containing the QTL with the syntenic region in another species (such as mouse) that also captures the phenotypic effect. The syntenic regions are those which have approximately identical blocks of DNA sequence in different species. This can help in narrowing down the QTL region assuming that the same casual variant is responsible for the phenotype in both species. Vitt et al. [Vitt et al., 2004] narrowed down Rf-1 (renal failure 1) QTL locus on chromosome 1 in the rat from 20 to 11.5 Mb by comparison with syntenic regions in the human genome. Rat chromosome 1 is syntenic to regions on human chromosomes 9, 10 and X and there is a locus for renal failure on human chromosome 10.

Another approach is to **combine data for crosses** in which the same phenotype has been mapped to the genome [Li et al., 2005]. The size of the QTL region is inversely related to the number of recombination events and combining data from different experi-
1.3. QTL mapping

ments narrows the QTL region as well increases power to detect smaller QTLs. Combining crosses, however, is only possible if one has access to the raw data in which the same phenotype has been studied which is a relatively uncommon opportunity. Haplotype analysis provides another tool for narrowing the QTL region. Genotypes from strains with high and low phenotypes values are examined to see the regions of similarity and dissimilarity. A causative gene cannot reside in the region of similar genotypes between phenotypically-divergent strains. This narrows down the search for linked gene to regions of dissimilar genotypes. This approach was taken by Wang et al. [Wang et al., 2004] who sequenced the Apoa2 gene in 46 genetically distinct mouse strains and found five distinct Apoa2 protein variants. They then conducted haplotype analysis of the strains in 21 crosses in which a high-density lipoprotein (HDL) QTL was found. They discovered that one of the HDL QTLs, hdlq5, was detected only in the crosses where one of the parents had a distinct Apoa2 protein variant.

The availability of genomic sequence offers another valuable tool for narrowing the QTL by direct sequence comparison. Sequence variation may alter the function or expression of a gene. The difficulty lies in the fact that our knowledge of regulatory elements is incomplete. The prediction of a sequence variation is relatively easy if such variation lies in the open reading frame (ORF) since such regions are usually well annotated. This task is more challenging if sequence variation lies upstream (5′), downstream (3′), in the intronic or intergenic regions. Sequence variations in these regions can also affect protein function by altering untranslated sequences, binding of transcription factor binding sites, RNA splicing or altering the expression levels of a gene. At present prediction of non-coding sequence variations on gene function is a bioinformatics challenge.

Analysis of gene expression aids in gene discovery by identifying genes that are expressed in the tissue of interest. Gene expression can be directly correlated to the trait by use of Quantitative trait transcript (QTT) analysis. In expression QTL (eQTL) analysis gene expression is treated as a phenotype and genomic regions that influence expression levels are identified by linkage analysis. These QTLs regulating transcript abundance are called cis eQTLs if the expression levels of a gene map to its own region,
implying that expression is controlled by the gene itself. Integration of cis eQTLs and physiological QTLs has provided another valuable tool for identifying candidate QTL-genes. It is likely that a combination of these in silico techniques will narrow QTL intervals and help identify QTL-genes [Petretto et al., 2008].

### 1.4 Microarrays

**Gene expression: Fundamentals and techniques**

Gene expression is the process by which information in the gene sequence is translated into a functional gene product which can be a protein for protein-coding genes or RNA for RNA genes. The first step in the transfer of information is to produce RNA from DNA which results in primary RNA transcript which is subsequently processed to produce mature RNA transcript. For a given gene, only one DNA strand of the two is transcribed.

Gene expression can be measured by estimating transcript abundance or protein quantitation which are highly correlated but not interchangeable [Schwanhäusser et al., 2011]. Samples may be obtained from the tissues to measure transcripts or proteins or expression may be measured in situ by in situ hybridization. Examples of these techniques are northern blot hybridization, ribonuclease production assay, quantitative PCR (qPCR) and global gene expression analysis using oligonucleotide microarrays. In northern blot hybridization, RNA is extracted from samples of interest, size fractionated by gel electrophoresis and hybridized to gene-specific probes. In qPCR RNA is first reverse-transcribed into cDNA which is subsequently amplified using PCR. The amplification products are constantly quantitated during qPCR reaction to measure transcript abundance. Oligonucleotide microarrays allow highly parallel analyses of gene expression, involving thousands of genes and allow global transcript profiling [Schena et al., 1995, Duggan et al., 1999]. This is the method that has been employed in current study and will be given more attention below. There are emerging transcript profiling techniques based on next-generation sequencing technologies, called RNA-seq which will provide a valuable and likely more powerful alternative to microarrays but at present
1.4. Microarrays

are technologically and bioinformatically challenging to implement [Wang et al., 2009b].

Expression profiling using microarrays

A microarray consists of glass or polymer slide on which genomic sequences are attached at fixed positions called “features” or “spots”. There are tens of thousands of these spots and each spot has millions of identical molecules called probes. mRNA from samples to be tested are labelled with fluorescent probes and this is then hybridized to the array. The locations of probes with specific genomic sequences are known \textit{a priori}. Labelled mRNA binds to target probes and emits fluorescence when scanned with a laser, the amount of fluorescence being proportional to the amount of mRNA abundance. Probes on the microarrays may be produced by printing cDNA or PCR products on the surface of microarrays in case of \textit{spotted} microarrays or by synthesizing the sequence matching a gene/transcript directly on the surface of an array in case of \textit{oligonucleotide} arrays.

Microarrays can be \textbf{single channel} or \textbf{two channel}. Affymetrix microarrays are examples of single channel microarrays. These microarrays have two sets of 25bp long oligonucleotide probes for each gene or exon. Expression is measured by hybridization of a single test sample to two sets of probes on the array: for the 3' array there are \textbf{perfect match} (PM) probes and paired \textbf{mis-match} (MM) probes, and for Gene and Exon arrays there are PM probes and ‘background probes’. Two channel arrays involve hybridization of two samples to each array, a test sample and a control sample. Each sample is labelled with a different coloured dye, usually red and green, and differences between test and control samples are reflected by relative fluorescence between the two channels.

For 3' arrays there are 10-15 pairs of probes for each transcript. Each PM probes is paired with a MM probe that is identical to the PM probe except for the nucleotide in the centre of the probe, for which a mismatch nucleotide is substituted in the centre of the probe. The PM measures the hybridization to intended transcript as well as non-specific hybridization whereas the MM probe measures non-specific hybridization. The amount of mRNA is calculated by measuring the hybridization differences between PM probes and paired MM probes. In case of whole-transcript arrays PM probes target entire length of
the gene (usually ∼26 PMs per gene). There are no MM probes but instead there are groups of ‘background’ probes which have the same GC content as the PM probes and each PM probe intensity is corrected by subtracting the intensity of background probes with identical GC content.

For current study Affymetrix Rat Genechip 1.0 ST arrays were used for global transcript profiling [http://www.affymetrix.com/](http://www.affymetrix.com/). A total of 27,342 rat genes are represented on the array by approximately 26 probes per gene with the probes spread across the full length of the gene. This enables more accurate estimation of gene expression than older generation 3′ based expression array designs which targeted the 3′ end of the mRNA transcripts. The array design is based on sequences and gene annotations obtained from various genome browsers including UCSC, Ensembl, RefSeq and GenBank.

**Preprocessing microarray data**

Microarray data requires pre-processing before advanced level analyses can be performed. The pre-processing steps include background correction to remove non-specific fluorescence, normalization to remove differences between arrays and summarization of the 26 probe values, in case of gene arrays, into a single value for each gene. This pre-processing requires computational software for which a variety of commercial as well as freely available tools are available. These analyses were undertaken in R [R Development Core Team, 2011](http://www.R-project.org) using packages available from the Bioconductor project [Gentleman et al., 2004](http://www.bioconductor.org). Errors in microarray data may be introduced at any step from RNA preparation to scanning of the arrays, therefore it is vital to assess the data for systematic variation.

Background correction is carried out on the assumption that overall signal intensity across each array should be even. Background correction removes signal due to non-specific binding of fluorescent molecules for a single array at a time. Any systematic signal difference between regions within each slide is corrected to equalise the average signal over an array. Robust multiarray average (RMA) method can be used for background correction of PM probes in arrays such as the rat gene chip used in our study and has
1.4. Microarrays

a greater sensitivity and specificity to detect differential expression [Irizarry et al., 2003]. The method models observed PM probe values (O) as sum of signal (S) and noise (N) and corrects each PM probe value, \( S = O - N \) to eliminate background noise.

Median probe intensity should be similar for arrays but this is often not the case because of artefacts and variability in array processing. This can assessed using MA plots. They were originally used to describe red and green dye intensities in two channel microarrays [Yang et al., 2002] but can also be used to plot the intensity of single channel arrays. Two intensity measures are plotted in MA plots, \( I_1 \) which is the intensity of the array studied and \( I_2 \) which is median intensity across all arrays. \( I_1 \) and \( I_2 \) are used to obtain the M value, 

\[
M = \log_2(I_1) - \log_2(I_2)
\]

and the A value, 

\[
A = \frac{1}{2} \left( \log_2(I_1) + \log_2(I_2) \right)
\]

The M and A values are plotted against each other to obtain the MA plot. There should be no trend in M as a function of A because mass of the distribution in an MA plot should be concentrated along the M = 0 axis. Any trends should be corrected by normalisation. An example can be used from one of the example datasets used in R [Gautier, 2011], figure 1.2.

![MA plots before normalization (left) and after normalization (right) for one of the example datasets obtained from Bioconductor.](image)

Figure 1.2 MA plots before normalization (left) and after normalization (right) for one of the example datasets obtained from Bioconductor.

Normalization can be performed by a variety of tools. Variance stabilization (VSN) tool [Huber et al., 2002] was used to normalize expression data in this study. VSN uses linear models and log-transformation to scale the data across the arrays. The final step in microarray pre-processing is summarization. Each gene is represented by multiple
probesets on each array, \( \sim 26 \) probesets/gene for GeneChip arrays. The intensities for all the probesets are summarized into a single measure using a summarization method. Commonly used statistical technique is Median Polish in which medians intensities are derived for all probesets representing a gene. These median intensities are then subtracted from the original values to estimate error. This error is then subtracted from original intensities to obtain fitted summarized values.

1.5 Genomic databases and data mining

Evidence for linkage of a trait to a region of the genome or correlation of a set of transcripts with a trait need to be evaluated further for clues to the underlying polymorphism or pathways controlling the phenotype. This is achieved by utilizing the genomic and bioinformatic approaches discussed in section 1.3. This process is facilitated by the existence of genomic databases and data mining resources which organize the genomic information accumulated thus far in approachable and searchable formats. Here we will briefly discuss the main tools in mining genomic and expression information.

Rat Genome Database (RGD)

The Rat Genome Database is a collaborative effort between leading research institutions involved in rat genetic and genomic research, http://rgd.mcw.edu/ [Twigger et al., 2007]. It collects and integrates data from rat genomic research and provides this data to researchers. RGD is also a repository of known QTLs in the rat and provides tools to compare these QTLs with similar QTLs in mouse and humans. Information on individuals rat strains and phenotypes is also provided. Particularly relevant to current study is the fact that RGD has a cardiovascular disease portal that provides access to data on genes, QTLs, strain models, biological processes and pathways related to cardiovascular diseases http://rgd.mcw.edu/rgdCuration/?module=portal&func=show&name=cardio
1.5. Genomic databases and data mining

Ensembl

Ensembl, www.ensembl.org, is a well known genome browser that arose out of a collaboration between the European Bioinformatics Institute (EBI) and the Wellcome Trust Sanger Institute both in Cambridge, UK. Ensembl was created in anticipation of the completion of human genome sequence in 1999. Fundamentally, large sequence data is annotated automatically (although manual curation is also provided for a subset of sequences) and provided to researchers freely in a searchable graphical use interface (GUI) [Flicek et al., 2011].

UCSC Genome Browser

UCSC (University of California, Santa Cruz), urlhttp://genome.ucsc.edu/, genome browser is another popular genome browser that provides customizable GUI access to sequence data for a number of species but perhaps the main advantage is that UCSC browser provides an interface to the ENCODE (Encyclopedia of DNA elements) project. The aim of the ENCODE project is to identify all functional elements in the human genome identifying not only protein-coding sequences but additionally regions coding for RNA and regulatory sequences [Birney et al., 2007]. This information can be helpful in differentiating sequence variations in non-coding regions into those which could affect biological function from those less likely to be of functional consequences.

Gene Ontology consortium

The aim of the Gene Ontology (GO) consortium is to produce a controlled vocabulary to describe gene function and products in terms of biological process, cellular components and molecular function attributed to the gene. Initially this was crucial for computationally transferring biological annotations from one organism to the other due to significant orthology [Ashburner et al., 2000] but more relevant to my study is the fact that this controlled vocabulary ensures that results from different genomic projects are easily searchable using a common vocabulary, www.geneontology.org.
1.5. Genomic databases and data mining

KEGG

Kyoto encyclopaedia of genes and genomics (KEGG) [www.genome.jp/kegg/] is a database for systemic analysis of gene functions and networks. KEGG provides tools to predict gene regulatory networks from the gene expression profiles by mapping gene expression for a given condition to biochemical pathways [Ogata et al., 1999].
Chapter 2

Materials and methods

2.1 Animals

Initial phenotyping experiments were conducted using four inbred rat strains, BN, SHR, WKY and Lewis. BN and SHR have been introduced in section 1.2. WKY and Lewis were included as normotensive control strains. 12 to 14 weeks old male rats were phenotyped at a rate of two animals/day. All rats were purchased from Charles River UK Limited (Margate, UK).

After the observation that BN and SHR are the most divergent strains with regards to CF phenotype (section 3.2), we proceeded to produce an F₂ intercross from BN and SHR. Rats were bred by a monogamous mating system. BN females were crossed with SHR males to produce BN X SHR (BXH) F₁ animals and a reciprocal cross was performed to obtain SHR x BN (HXB) animals. F₁ BXH animals were intercrossed to produce F₂ BXH animals and F₁ HXB animals were intercrossed to produce an F₂ HXB animals.

Animals were maintained at the Central Biomedical Services facility, Imperial College, London and housed at a maximum of five per cage. The animals had ad libitum access to standard rat chow and sterile water. Except for breeding purposes, animals were separated according to sex. They were maintained on a twelve hour diurnal cycles by automatic light switching. Colonies were regularly tested for specific pathogens by using sentinels kept in separate cages. All procedures were performed in accordance with the UK Animals (Scientific Procedures) Act of 1986.
2.2 Cardiovascular physiological techniques

2.2.1 Blood pressure

BP differences between strains may partially account for cardiac phenotype differences between these strains. Therefore it was important to study the BP in the F$_2$ intercross strain progeny to study the correlation between BP, CF and other cardiac phenotypes.

![Blood pressure recording during carotid cannulation](image)

**Figure 2.1** BP recording during carotid cannulation. Inhaled isoflurane is gradually reduced to 1.5%. BP gradually increases till a plateau is reached. The observations during the plateau phase are analyzed offline for systolic, diastolic and mean BP.

BP was measured using cannulation of the carotid artery in each animal before cardiac excision for *ex vivo* phenotyping. The pressure catheter was first calibrated against an external manometer. Animals were anaesthetised using inhaled 4% Isoflurane. The carotid region was exposed by removal of skin and blunt dissection in the paratracheal region. The carotid artery was freed from the surrounding tissue and glossopharyngeal nerve. Proximal end of the carotid artery was tied and distal end occluded by a 1 mm vascular clip. An arteriotomy was then performed and an ultra-miniature 2 mm pressure catheter (MPVS-Ultra Single Segment Foundation System, ADI instruments) advanced...
into the artery. The vascular clamp was released and the catheter advanced into the distal vessel and tied in place. Data was captured continuously by LabchartPro software (ADI instruments) and analyzed off line.

After carotid cannulation the concentration of inhaled Isoflurane was gradually reduced from 4% to 1.5%. This allowed the BP to reach physiological level. The BP gradually rose till a plateau was reached as shown in figure 2.1. After recording the BP data, pressure catheter was withdrawn from the vessel which was then tied off. The whole BP procedure usually took between 15 to 20 minutes.

2.2.2 Langendorff isolated heart preparation

Principles of Langendorff isolated heart technique

The Langendorff heart preparation, named after Oscar Langendorff, is a physiological technique for ex vivo investigation of function of the isolated mammalian heart. Fundamentally, the heart is perfused by a suitable perfusion fluid that is pumped into the heart through a cannula to which aorta is tied. This retrograde perfusion shuts the aortic valves and re-directs the perfusate through the coronary ostia into the coronary arteries. The perfusate is drained by cardiac veins into the right atrium and flows out thorough the pulmonary arteries. LV cavity is free of any perfusate. This is different from the isolated working heart which is a more complex preparation. In isolated working heart, the aorta and left atrium are both cannulated. Perfusate ("preload") is delivered into the left atrium (LA) and from the LA into the LV cavity and into the aorta against a hydrostatic pressure ("afterload"). Preload and afterload can be manipulated. Langendorff heart preparation allows study of cardiac physiology, including CF, LV developed pressure, heart rate as well as cardiac morphology, histology and pharmacology [Sutherland and Hearse, 2000].

Perfusion solution

The vast majority of isolated heart studies studies reported in the literature have used bicarbonate-based salt perfusion fluid initially develop by Krebs and Henseleit in 1932 [Skrzypiec-Spring et al., 2007]. There are alternatives to saline perfusion solutions in-
2.2. Cardiovascular physiological techniques

Including colloid solutions, solutions supplemented with red blood cells and blood perfusion from a donor animal. These perfusion options, however, can lead to complications including haemolysis, foaming when gassed with carbogen, requiring large amounts of blood to perfuse the heart and a doubling in the number of animals required.

The composition of Krebs-Henseleit fluid mimics the plasma ionic content. Sutherland and Hearse [Sutherland and Hearse, 2000] note that the ionic calcium concentration in this crystalline fluid is not physiological due to the absence of proteins which normally bind ionic calcium and suggest to rectify this by halving the calcium concentration initially suggested by Krebs and Henseleit. Therefore modified Krebs-Henseleit buffer solution was prepared as suggested by Sutherland and Hearse (Table 2.1). This solution was kept at 37°C in a jacketed reservoir and continuously gassed with carbogen solution (95% Oxygen, 5% CO₂) to maintain a pH of 7.35 - 7.45.

<table>
<thead>
<tr>
<th>Solute</th>
<th>micromoles</th>
<th>molar mass</th>
<th>grams/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaCl</td>
<td>118.5</td>
<td>58.45</td>
<td>6.9</td>
</tr>
<tr>
<td>NaHCO₃</td>
<td>25</td>
<td>84</td>
<td>2.09</td>
</tr>
<tr>
<td>KCl</td>
<td>4.7</td>
<td>74.5</td>
<td>0.35</td>
</tr>
<tr>
<td>MgSO₄.H₂O</td>
<td>1.2</td>
<td>138</td>
<td>0.16</td>
</tr>
<tr>
<td>KH₂PO₄</td>
<td>1.2</td>
<td>136</td>
<td>0.16</td>
</tr>
<tr>
<td>Glucose</td>
<td>11.2</td>
<td>180</td>
<td>2.01</td>
</tr>
<tr>
<td>CaCl₂.2H₂O</td>
<td>1.5</td>
<td>147</td>
<td>0.22</td>
</tr>
</tbody>
</table>

Table 2.1 Ionic composition of perfusion solution for *ex vivo* heart preparation

Anaesthesia

The animals were anaesthetized using injectable intraperitoneal Thiopental or inhaled anaesthetics such as Isoflurane. Intraperitoneal heparin was also administered to reduce the risk of intravascular thrombi. Thiopental is a barbiturate and binds at a distinct site associated with a Cl⁻ ionopore at GABA receptors and prolongs the GABA (γ amino Butyric acid) mediated CNS depressant effects. It has temporary negative inotropic action on the myocardium. Isoflurane is a volatile anaesthetic delivered with Oxygen via a calibrated vaporizer and acts vis GABA receptors. Isoflurane has *in vivo* effect of reducing
mean arterial pressure and causing reflex tachycardia. Both Isoflurane and Thiopental are cardioprotective preconditioning drugs [Raphael et al., 2005, Kim et al., 2010]. The choice between Thiopental sodium and Isoflurane is dictated by experimental need. Isoflurane is the drug of choice if the animal is to kept alive for undertaking another surgical procedure before cardiac excision is performed and this was directly applicable to current study when BP was measured prior to cardiac excision.

### Cardiac excision and aortic cannulation

The animal were placed on a warm operating board. Depth of anaesthesia was confirmed by absence of ankle reflexes following which abdomen was opened, diaphragm excised and bilateral thoracotomy performed along the bone-cartilage border. Following excision, the heart was immediately placed in ice cold Kreb’s buffer to cause cardiac standstill and transferred to the *ex vivo* perfusion apparatus. Heart was gently mounted onto the aortic cannula via the aorta. The tip of the cannula was kept below the origin of the right brachiocephalic trunk and the aortic valve. Aorta was clamped onto the cannula by a bulldog clamp and secured using 3/0 silk suture. Perfusion was established within a couple of minutes of cardiac excision. LA was then removed and a fluid-filled latex balloon placed in the LV cavity. LV contractility (LV \(\text{dP/dt}_{\text{max}}\)) and LV relaxation (LV \(\text{dP/dt}_{\text{min}}\)) were derived from LV pressure. A co-axial bipolar electrode was placed on the right atrium and pacing was commenced at 360 beats per minute. Electrodes were placed on the right atrium and left ventricle to record continuous electrocardiogram. The preparation was allowed to stabilize for 10 minutes. The whole preparation was kept inside a jacketed container which was maintained at 37°C (Fig 2.2). Langendorff set-up in our laboratory consisted of a warm jacketed reservoir for buffer, circulator pump, bath unit for warming the perfusate and a self-contained, jacketed, re-circulating Langendorff unit. The bath unit was connected to jacketed reservoir and Langendorff unit and kept the heart preparation and perfusate thermostatically warm at 37°C. This perfusate was pumped by the peristaltic pump at constant volume to the heart. Cardiac perfusion pressure was set at 90 mmHg. CF was measured by an in-line ultrasonic flowmeter which
resided in an assembly just proximal to the cardiac chamber. This assembly also housed an air bubble trap that prevented cardiac air embolism.

**Langendorff protocol**

Haemodynamic data was captured continuously by LabchartPro software (ADI instruments) and analyzed off line after the experiment. The isolated heart preparation was studied at baseline for 15 minutes (figure 2.3). Baseline phase was followed by transient one minute global ischaemia and reperfusion, to assess maximum hyperaemia. Myocardial infarction was subsequently induced by ligation of the proximal LAD for 35 minutes and this was followed by reperfusion for an hour. Cardiac eluates were collected during baseline, ischaemia, first 10 minutes and subsequent 50 minutes of reperfusion for measurement of creatinine kinase and stored at -80°C. In a subset of animals, at the end of the procedure LAD was tied again and 1% Evan’s blue dye was injected via the cannula to stain the non-ischaemic myocardium. The hearts were stored at -20°C for planimetry analysis.

As the planimetric infarct measurements did not significantly vary in the inbred strains (section 3.2), further planimetric analysis was not attempted in the F₂ intercross and dye was not infused into the heart at the end of experiment. At the end of the experiment the hearts were sectioned into left and right ventricles. The left ventricle was further divided into ischaemic and non-ischaemic areas and stored at -80°C for gene expression analysis. Animal tails were stored for DNA genotyping.

**2.2.3 Myocardial infarct estimation**

**Myocardial infarct quantification by planimetry**

Frozen hearts were sectioned from apex to base into four 1 mm slices. These were placed in 1% Triphenyl Tetrazolium Chloride (TTC) and incubated at 37°C for 15 minutes. TTC is a colourless compound which is reduced to a red formazan product in the viable tissue by the dehydrogenases present in the mitochondria. The infarcted tissue appears white [Vivaldi et al., 1985]. TTC solution was drained after 15 minutes and the slices were fixed
Figure 2.2 (a) Langendorff isolated heart preparation; the heart is attached to a perfusion cannula via the aorta. Water filled latex balloon is inserted in the left ventricle via the left atrium and the heart is paced at 360bpm by the pacing electrode; (b) A schematic description of the Langendorff system, adapted from ADI instruments.
2.2. Cardiovascular physiological techniques

**Figure 2.3** Overview of the experimental stages and phenotypes measured during the physiological experiments. LV $dP/dt_{\text{max}}$ and LV $dP/dt_{\text{min}}$ are measures of rates of LV pressure change during systole and diastole respectively.

in 10% formalin for 24 hours at room temperature. Planimetry was performed to quantify myocardial infarction.

**Fig 2.4** is an illustration of planimetry analysis for a single experiment. Briefly, the stained cardiac sections for analysis were scanned (Fig 2.4a) in .tiff format. Images were analyzed using freely available software MIPAV (Medical Image Processing, Analysis & Visualization, Center for Information Technology (CIT), National Institutes of Health (NIH), USA). Total area of the cardiac sections was measured (Fig 2.4b). This was followed by measurement of area (AAR) at risk as shown in figure 2.4c. Infarct area (IA) is obtained by the calculating the difference between AAR and normal myocardium (Fig 2.4d).

**Myocardial infarct quantification by enzymatic method**

CK catalyzes the transfer of a high energy phosphate group from creatine phosphate to ADP. The ATP produced in this reaction is used to phosphorylate glucose to produce
2.2. Cardiovascular physiological techniques

Figure 2.4 Planimetry and estimation of infarct area; (a) shows cardiac slices before image processing. The blue area is the non-ischaemic myocardium, the red area is the TTC stained alive myocardium and the white area represents infarction; (b) shows measurement of the total area; (c) represents the area at risk after removal of non-ischaemic myocardium; (d) represents the ischaemic but non-infarcted myocardium.

Glucose-6-phosphate (G6P) in the presence of hexokinase. G6P is then oxidized in a reaction which leads to formation of nicotinamide adenine dinucleotide phosphate reduced (NADPH). The rate of formation of NADPH is monitored at 340 nm and is proportional to the amount of CK in the sample.

Creatine Kinase (CK) is present in high concentration in the cardiac muscle, skeletal tissue and brain. It is used as a diagnostic and prognostic marker for acute myocardial infarction [Duma and Siegel, 1965, Shell et al., 1971]. In an ex vivo isolated heart preparation, the only source of CK is the cardiac tissue and hence CK assay provides a reliable estimate of myocardial damage. Samples for CK measurement were collected during the experimental stages as illustrated in figure 2.3. These samples were stored
2.3. High throughput genotyping

at -80°C for analysis. CK quantification in the cardiac eluate samples was performed in the biochemistry department at the Hammersmith Hospital, using Abbott creatine kinase reagent kit (Abbott Laboratories Ltd, Maidenhead, UK) according to the manufacturer’s instructions.

2.3 High throughput genotyping

2.3.1 DNA extraction and quantification

DNA extraction

DNA was extracted from the tissue or blood using robotic DNA extraction on the Maxwell® 16 system. This system is automated, allows simultaneous extraction of up to 16 samples and is largely hands-free allowing higher throughput. Pre-filled disposable plastic cartridges are supplied by the manufacturer. The cartridges contain wash buffers and paramagnetic particles. Theses paramagnetic particles can be easily separated from a suspension magnetically and become non-magnetic when removed from the magnet. Tissue or blood is placed in one of the the sample wells in the cartridge per manufacturer instructions. The cartridges are then placed in the robot, where the sample is processed. DNA is captured by the paramagnetic particles which then undergo a series of washes and elutions. The process of DNA extraction takes approximately 35 minutes on the robot following which the DNA quality is assessed.

DNA quantitation

Accurate quantitation of small amounts of DNA is important for quality assurance of all downstream applications. The quantity and quality of DNA for our study was assessed using both spectrophotometric measurements (NanoDrop) and fluorescent assay (Cubit).

NanoDrop™, which is a common instrument used in the laboratory, utilises spectrophotometric methods to quantify small amounts of DNA. The intensity of light is measured at baseline (Io) and after passing through the sample (I). The ratio of I/Io is termed the transmittance and is expressed as a percentage (%T). The absorbance is
2.3. High throughput genotyping

calculated as

\[ A = -\log(\%T) \]  \hspace{1cm} (2.1)

The absorbance is then converted into units of dsDNA, ssDNA or RNA. Additional quality indicator provided by this method is the amount of protein present in the sample. Variable amounts of protein are frequently present in the DNA sample as complete removal is not possible. The peak of light absorption for DNA is at 260 nm, while protein absorbs light at 280 nm which is attributed to tryptophan and tyrosine side chain residues. The ratio of DNA absorbance at 260 nm and protein absorbance at 280 nm, also called the A260/A280 ratio, is an indication of the purity of the sample. A260/A280 value greater than \( \sim 1.7 \) indicates that DNA is of suitable quality for downstream applications.

Spectrophotometric methods are disadvantaged by the inability to differentiate between dsDNA and RNA. Additionally DNA quantity may be over-estimated due to the presence of free nucleotides in the solution. Fluorescent dye based methods for DNA quantification such as Invitrogen’s Quant-it™ provide a solution to this problem and are used to quantitate DNA [Singer et al., 1997].

Quant-it™ assay utilizes PicoGreen, a fluorochrome that binds selectively to dsDNA in the sample. It has excitation maximum at 480 nm and emission peak at 520 nm. Bound to DNA, the fluorescent enhancement of PicoGreen is exceptionally high whilst the unbound dye has no fluorescence. Because of its highly selective binding to DNA, the accuracy of this method is superior to spectrophotometric methods. It is the method of choice for DNA quantitation.

2.3.2 Genotyping workflow

High throughput genotyping was performed using Illumina’s GolgenGate assay. GolgenGate was Illumina’s first commercial SNP genotyping assay. The system has the advantages of accurate multiplex genotyping using allele specific extension, requiring only 250 ng input of genomic DNA (gDNA) and relatively high SNP call rates. Two types of assays are available: standard panels including human SNP panel and cancer SNP
panels or **custom** panels based on sequences submitted by users. A custom SNP panel was designed for current studies as no standard panels were available for the rat strains being studied.

The GoldenGate workflow can be described in five phases: DNA quantitation, activation, extension/ligation, PCR, hybridization and imaging. DNA activation, shown in figure 2.5, required the addition of biotin to the gDNA solution resulting in random binding of biotin to gDNA. This was followed by biotin precipitation with 2-Propanol to remove excess biotin and resuspension of biotinylated DNA.

![Figure 2.5](image)

**Figure 2.5** Goldengate genotyping workflow: Illustration of ssDNA with a [T/C] SNP; the gDNA is “activated” by random binding of biotin to gDNA.

The next step in the workflow was the extension/ligation step. During this stage, streptavidin coated paramagnetic beads were added to the biotinylated DNA. These beads bind the gDNA which was then added to custom made pool of oligonucleotides called oligonucleotide pool assay (OPA). OPA contained two allele-specific oligonucleotides (ASOs) and one locus-specific oligo (LSO). The ASOs contain about 20 bp of universal primer sequence 1 or 2, followed by 20 bp of sequence that is complementary to the target DNA sequence and ending in a base complementary to one of two SNP alleles e.g. in case of [T/C] SNP, the sequence of one ASO will end in [A] and the other ASO will end in [G]. The LSO contains about 20 bp complementary to the target sequence, followed by 20 bp of “Illumicode address”. The illumicode sequence is complementary to the sequence on a
2.3. High throughput genotyping

target bead on the chip and this allows the reaction to occur at a specific bead. Illumicode sequence is followed by 20 bp of universal primer sequence 3. Once the specific ASO and LSO had bound to the target genomic region, the gap between ASO and LSO was filled by polymerase chain reaction and ASO was ligated to the LSO was by the ligase enzyme, figure 2.6. The DNA remained bound to streptavidin beads during these steps.

Figure 2.6 The extension/ligation step; ASOs bind to the target DNA depending on the terminal SNP (allele-specific binding), PCR and ligation steps then bind the ASO to the LSO.

Extension/ligation step was followed by PCR. Initially the streptavidin beads were captured magnetically allowing removal of template DNA. The beads were then separated from the DNA. Three types of universal primers were then added to the DNA, universal primers (UP) 1 and 2, both in 5′ to 3′ direction and attached to fluorophore as well as UP3 which was biotinylated. Fluorophore incorporation into the complementary strands depend on the specific SNP allele in the template strand. This was followed by 34 cycles of amplification following which double-stranded reaction product was bound to a new set of paramagnetic beads.

Hybridization step was then carried out during which dsDNA was dissociated by addition of Sodium Hydroxide. The biotinylated template was separated with the beads, leaving behind fluorophore-labelled amplicons. The products were then hybridized to the BeadChip. Each BeadChip allowed hybridization of up to 32 samples. The internal illumicode specific for each locus hybridized to its probe on the associated BeadChip.
2.3. High throughput genotyping

These probe positions were known a priori and therefore amplicons hybridizing to a specific probe along with their attached fluorophore were genotyped based on the red/green signal. The BeadChip was scanned and the attached red or green fluorophore excited providing the basis for SNP genotyping, figure 2.7.

![Image of genotyping process](image.png)

**Figure 2.7** Following allele-specific extension/ligation, universal primers with fluorophores are annealed, PCR is performed followed by hybridization and scanning steps where the signal of fluorophore indicates the specific allele at the locus.

### 2.3.3 SNP selection for genotyping

Since no off-shelf assays were available for genotyping the F₂ intercross generated in current study, it was necessary to design the ASOs, LSOs and BeadChips for custom genotyping required for QTL mapping. To design the genotyping assays for SNPs that vary between BN and SHR, two pieces of information were required; first the genome sequence data was required for BN and SHR and second the genetic variation (SNP) data between these two strains. The list of SNPs that varied between SHR and BN was obtained from the STAR consortium database [http://rgd.mcw.edu/]. Saar et al. published a total of 16,543 SNPs by genotyping a recombinant inbred (RI) panel derived from BN and SHR [Saar et al., 2008]. Assistance in obtaining the 160bp genome sequence around the SNPs obtained published by the STAR consortium was kindly provided by Santosh Atanur, who maintains the SHR genome database in Prof Aitman’s group [Atanur et al., 2010] at the Clinical Sciences Centre, Imperial College, London, [http://shr.csc.mrc.ac.uk/].

16000 SNPs and the surrounding 160 bp sequence for each SNP was submitted to
2.3. High throughput genotyping

Illumina for bioinformatic assessment. Each SNP with its surrounding sequence is scored for a number of features which are as follows:

- Is there enough flanking sequence available to design oligonucleotides?
- Does the variant have more than two alleles?
- Is the SNP located in the mitochondrial genome?
- Does the SNP lie in a repetitive region?
- Does the next marker lie less than 60 bp away?

Each SNP is then given a final score from 0.01 to 1, with 1 being the perfect score for a SNP with high probability of success in the genotyping assay. Commercial recommendation is to select SNPs with a final score of at least 0.50. 768 SNPs with final score of more than 0.95 were selected to ensure a higher probability of success in the assay (figure 2.8). These SNPs were uniformly distributed across genome as shown in figure 2.9. Additionally the assay based on these SNPs was likely to be successful as they were already genotyped for constructing linkage map in the BXH RI strains as mentioned above [Saar et al., 2008].

2.3.4 Metrics for SNP calling and clustering

Following SNP selection, bioinformatic scoring and genotyping the next step was to call the genotypes for the samples. This task was achieved by use of GenomeStudio, an Illumina propriety software that utilizes the GenCall algorithm [Fan et al., 2003]. Once the samples had been processed and the BeadChips had been scanned, there were essentially three strings of information which were integrated using GenomeStudio software. Firstly, there was the information on the individual beads on the BeadChip with each bead relating to a specific SNP, this was the code file. Secondly, there were the raw red and green intensities for each bead on the BeadChip and as already described this information relates to a specific genotype at the locus, these were saved in the intensity files. Thirdly, there was the sample file describing the samples on each BeadChip. In addition
2.3. High throughput genotyping

**Figure 2.8** Representation of the bioinformatic scores for the SNPs submitted for the first six rat chromosomes, the final SNP scores (scoring for probability of success in genotyping assay) are plotted against the physical co-ordinates of the SNPs for the individual chromosomes. The dashed lines represent the cutoff score of 0.95. Only SNPs scoring more than 0.95 were included in designing the custom assay.
Figure 2.9 Final 768 SNPs selected for genotyping plotted by their physical co-ordinates on each rat chromosome.
to samples from the F₂ intercross, the hybridized samples also included samples from the parental BN and SHR strains as well as F₁ intercross to function as controls and also to measure genotyping errors e.g. if the BN had G/G allele and the SHR had A/A allele at a locus then the F₁ must be A/G at the locus; the genotype is clearly erroneous if the F₁ is G/G at the locus. Such errors are termed as Parent-Parent-Child (P-P-C) errors. In addition to the parental and F₁ samples, sample replicates were also placed on the BeadChips to estimate reproducibility errors.

After the integration of code, intensity and sample files in the Genome Studio software, each SNP was automatically clustered into three clusters: AA, AB or BB which were then converted back to the actual BN and SHR alleles for each locus. A number of metrics were also provided for each of the 768 SNPs which required careful evaluation to ensure optimal final linkage map. The main reason for undertaking this manual assessment was that the automatic clustering algorithm was trained on human samples [Fan et al., 2003] and may not function as well in non-human custom genotyping projects. The quality metrics for individual SNP assessment were: cluster separation, mean normalised intensity (R), identification of P-P-C errors, identification of replication errors and significant deviation of heterozygotes from the Hardy-Weinberg Equilibrium. The vendor software may be unable to resolve all genotypes at a locus and may report a single cluster which results in predominant but false homozygosity which should be recognised and corrected. This can be done by screening for minor allele frequency (MAF) < 0.1.

Cluster separation measures the distance between the AA, AB and BB genotype clusters based on their red/green intensities. SNPs for which the three genotype clusters overlap to any degree were excluded from the analyses. Figure 2.10 illustrates the cluster separation for a SNP that was automatically well separated by the algorithm. Figure 2.10a illustrates an extreme example of a SNP for which it was impossible to determine any genotype clusters and was therefore excluded from analysis. Figures 2.10c and 2.10d show a SNP that was initially clustered automatically (2.10c) but since margins of the heterozygote cluster cannot be reliably distinguished from the homozygotes, borderline samples were excluded by reducing margins of the heterozygote cluster (2.10d). SNPs
with reduced Norm R (<0.20) were also excluded as the genotypes cannot be reliable called at such low signal intensity (figure 2.11).

Figure 2.10 Cluster separation analysis for homozygotes and heterozygotes in the GoldenGate assay. Norm R represents normalized signal intensity and Norm theta is the red/green colour intensity. Each dot represents a single sample and labels above each panel refer to the particular SNP for which the genotypes are being called. Panels (c) and (d) show manual adjustment of automatic SNP clusters with (c) being the result of automatic clustering and (d) being the same cluster after manual adjustment. For some samples the signal intensity is ambiguous and by excluding these samples genotyping reliability can be improved.

After excluding poorly performing SNPs on cluster separation and Norm R, data was analyzed for P-P-C and replication errors. The overall P-P-C accuracy was ~ 97%, which is higher than that reported in literature [Lepoittevin et al., 2010]. Additionally, there were only two replicate errors out of >1300 reactions, placing the error rate at
2.4. Microarrays

2.4.1 RNA extraction and quantification

RNA extraction, quantification and preparation for microarrays required care to avoid contamination and bias in expression analysis. This is in part due to omnipresence of RNAases and instability of RNA. Total of 118 RNA samples from the non-ischaemic LVs were processed in batches of twelve using Maxwell® 16 system RNA purification kits (Southampton Science Park, Southampton, UK). Work area, pipettes, racks and homogenizer were set aside exclusively for RNA processing. Non-ischaemic LV samples were snap frozen in liquid Nitrogen at the end of each experiment and stored at -80°C. Af samples were removed from -80°C freezer, they were kept on dry ice. Non-ischaemic LV samples

Figure 2.11 Example of an excluded SNP because due to low signal intensity (Norm R below 0.20) even though the three genotype clusters are well separated.

less than 0.01%. Significant deviation from Hardy-Weinberg Equilibrium was assessed using statistical package R/qtl and is described in section 5.1 and summarized in table 5.3. In brief, the quantity of excess heterozygote calls relative to expectations based on Hardy-Weinberg Equilibrium was obtained from the SNP metrics produced by the vendor software. This metric varied from -1 (complete deficiency of heterozygotes) to 1 (100% heterozygotes). If the heterozygote excess values were less than -0.3 or greater than 0.2 then the individual SNP was evaluated visually for cluster separation.
were weighed, following which 200 µL of lysis buffer (provided in the RNA purification kit) was added for each 25 µg of the tissue sample. Subsequently the sample was homogenized until no solid sample was visible. 200µL of the sample-lysate was removed and the remaining stored to meet any future requirements. 400 µL of RNA dilution buffer was then added to sample-lysate which was then vortexed to ensure complete dissolution. 50 µL of clearing agent was added to sample-lysate to bind DNA at 70 °C. This clearing agent - sample mix was then cooled and centrifuged in a spin column at 12000 rpm for two minutes. The clear fluid at the bottom of the spin column was added to Maxwell® 16 RNA cartridge. This sample in the RNA cartridge was then placed in the Maxwell robot which can automatically process twelve samples at a time resulting in a final 40 µL RNA sample in RNAase free water. This RNA was quantified using nano drop as discussed for DNA quantitation in subsection 2.3.1.

The integrity of RNA extracted from the non-ischaemic LV samples was evaluated using Agilent® 2100 Bioanalyzer platform. The Bioanalyzer is a microfluidics-based platform for sizing, quantification and quality control of DNA, RNA, proteins and cells on a single platform. The assay used a sixteen-well chip which accommodated twelve RNA samples of unknown concentration in addition to vendor supplied RNA ladder and gel-dyes. RNA samples whose integrity was to be determined were placed in the sample wells along with the dye. Once this chip was placed in the instrument, a sixteen-pin electrode was placed in the wells and a constant current passed through each well. The moving RNA fragments incorporate dyes which were then detected by laser fluorescence. RNA was quantified automatically by comparison with the ladder composed of known fragment sizes. Software provided by Agilent® automatically provided RNA integrity number, also called the RIN score that provides a measure of integrity of RNA. This score is obtained by analyzing the complete electrophoretic trace of each sample profile and each sample is then give a score from 1 to 10 with 10 for excellent RNA quality [Schroeder et al., 2006]. For my experiments the cut-off RIN score was 8.0 (figure 2.12).
2.4. Microarrays

Figure 2.12 Example of RIN score profile for one RNA sample prior to RNA labelling. The exact RIN score is obtained by analysis of this trace. Fluorescence units are plotted along the y-axis and time is displayed along the x-axis. The first peak is at 18S and the second peak at 28S. 28S/18S ratio should be greater than 1. There should be few, if any, baseline perturbations and no visible hump after the 28S peak which would indicate presence of genomic DNA. The RIN score for this particular profile was 8.8.

2.4.2 RNA labelling for microarrays

The extracted and quality-checked RNA samples needed further processing before hybridization to arrays. The fundamental principle of RNA preparation for array hybridization is as follows: sense strand cDNA is produced from RNA, this cDNA is amplified, labelled with an Affymetrix propriety reagent and hybridized to an array which is subsequently scanned to measure the intensity of bound labelled cDNA. The reagents required to process these samples were sourced from Ambion (Lingley House, 120 Birchwood Boulevard, Warrington WA3 7QH, UK) and Affymetrix (Voyager, Mercury Park, Wycombe Lane, Wooburn Green, High Wycombe HP10 0HH, UK). These reagent kits are Ambion WT Expression Kit (4411973), Affymetrix poly-A RNA Control Kit (900433), Affymetrix hybridization control kit (900454), Affymetrix WT terminal labelling kit (900671) and Affymetrix hybridization, wash, and stain kit (900720).

The first step in this process was the generation of purified sense-strand cDNA (with incorporated dUTP) ready for fragmentation and labelling. 150ng of total quality-checked RNA was incubated with random primers and T7-(N6) primer. The engineered set of random primers excluded sequences that match ribosomal RNA (rRNA) and specifically primed non-ribosomal RNA from a total RNA sample. DTT (Dithiothreitol), RNase
2.4. Microarrays

inhibitor and a reverse transcriptase were added to RNA and placed in a thermocycler. This reaction synthesized first cDNA strand, called the sense strand. RNAase H, DNA polymerase 1 and dNTP mix were then added to the sense strand. DNA polymerase synthesized second cDNA strand. The two cDNA strands were used for multiple cycles of vitro transcription using T7-RNA polymerase, resulting in production of large amounts of cRNA. This cRNA was captured using nucleic acid binding beads, washed and eluted using a buffer following which it was quantified. 10µgrams cRNA was needed for input in the second cycle cDNA synthesis. Random primers, dNTP/dU(Uracil)TP mix and reverse transcriptase were added to the input cRNA to synthesize cDNA which randomly incorporated dUTP. RNAaseH was added to hydrolize cRNA in the solution. This left behind single stranded cDNA which was removed by binding to nucleic acid binding beads, washed and eluted with a buffer. The cDNA was quantified again and assessed for integrity using Bioanalyzer (subsection 2.4.1).

![Figure 2.13 Bioanalyzer trace from a sample before (above) and after fragmentation (below). cDNA has been adequately fragmented as indicated by the shift of the peak of distribution to the left. This fragmentation can also be seen in gel images on the right obtained from the Bioanalyzer.](image)

Subsequently UDG (Uracil DNA Glycosylase) and APE (Abasic Endonuclease) were added to cDNA. UDG cleaved Uracil bases from the phosphodiester backbone of DNA and APE cleaved cDNA at the site created by UDG. TdT (Terminal Deoxynucleotidyl
2.4. Microarrays

Transferase) then added a proprietary DNA labelling reagent. The sample was then assessed for adequate fragmentation using Bioanalyzer. Adequate fragmentation of cDNA was confirmed by reduction in fragment length on the Bioanalyzer trace (figure 2.13).
Aims

The aims of studies undertaken in this project were to:

- Determine the heritability of CF, CFR, LV dP/dt\textsubscript{max} and LV dP/dt\textsubscript{min} using data from parental strains as well as F\textsubscript{1} and F\textsubscript{2} intercross.

- Determine the covariance between BP and other cardiovascular traits especially CF in the F\textsubscript{2} intercross.

- Utilize genome-wide genotyping in search for QTLs underlying cardiac traits being studied.

- Prioritize candidate genes for further studies using QTT and eQTL data from microarray gene expression.
Chapter 3

Studies of cardiovascular physiology in parental strains and F₁ intercross

3.1 *In silico* data analysis

3.1.1 Aims

Physiological studies have been undertaken in multiple inbred rat strains in the Phys-Gen Program for Genomic Applications, [http://pga.mcw.edu/](http://pga.mcw.edu/), as part of one of the five programmes funded by National Heart, Lung, and Blood Institute (NHLBI) in the United States to advance functional genomic research related to heart, lung, blood, and sleep disorders. The data was deposited online and remains a valuable resource even though the research program officially ended in July 2009. Cardiovascular phenotype data was downloaded for four inbred strains: BN, SHR, Lewis and WKY. BN is the reference rat strain, a model of normal cardiovascular physiology and was the first rat strain for which genome sequence became available [Gibbs, 2004]. SHR is a model for hypertension, cardiac hypertrophy and insulin resistance and at the time of start of this project SHR genome sequencing was nearing completion in a collaborating laboratory [Atanur et al., 2010]. WKY was included in the analysis because it is the parent strain from which SHR was derived [Okamoto and Aoki, 1963] and Lewis was included as a normotensive control strain. The availability of genome sequence for BN and SHR meant one
could quickly search for and prioritize candidate genes for functional studies if traits were found to be genetically influenced. Additionally cardiac expression data was also available for BN and SHR [Hubner et al., 2005] making them ideal candidates for a mapping strategy if additional cardiac phenotyping discovered heritable traits.

The aim of the *in silico* analysis was to review and analyze this publicly available data for inter-strain differences in cardiac phenotypes, to inform experimental design and get a baseline dataset with which to compare in-house results. Analysis was limited to male rats (to exclude sex-related effects) fed normal diet and maintained under normoxic conditions. The following *ex vivo* phenotypes were available from the database and included in the analysis: Cardiac mass, heart rate (HR), coronary flow (CF), LV developed pressure and response to myocardial ischaemia including post myocardial infarction CK release as well as histomorphometric infarct measurements.

3.1.2 PhysGen cardiovascular phenotyping methodology

PhysGen programme used a high-throughput cardiovascular phenotyping strategy. Hearts were excised following intra-peritoneal injection of Sodium Pentobarbitone and then weighed before being transferred to the *ex vivo* perfusion apparatus. The database does not provide estimates of the time it took from cardiac harvest to establishing of *ex vivo* cardiac perfusion. This is an important variable because of the time consumed by multiple steps involved in the approach adopted: confirmation of the depth of anaesthesia, excision of the heart and trimming the extra-cardiac tissue to weigh the heart accurately before mounting the heart onto the aortic cannula. These steps before re-establishing cardiac perfusion are likely to require more than three minutes and the approach therefore risked pre-conditioning the heart. Cardiac pre-conditioning is a well-described cardioprotective phenomenon [Yellon et al., 1992, Marber et al., 1994] which has consequences especially for ischaemia-related phenotypes. Hearts were allowed to beat at intrinsic denervated sinus rates and global myocardial ischaemia was employed (figure 3.1).
3.1. In silico data analysis

Figure 3.1 PhysGen phenotyping protocol, six week old animals were phenotyped, myocardial contracture is the contraction in the cardiac muscle in response to ischaemia; LDH, Lactate dehydrogenase.

3.1.3 PhysGen phenotyping data

Table 3.1 summarises the baseline cardiac phenotypes from the strains included in the analysis. At baseline CF and cardiac mass were significantly different between the strains \((p=8.9\times10^{-6}\) for CF and \(p=0.0001\) for cardiac mass using Kruskal Wallis test). BN (n=20) and SHR (n=20) were the most divergent strains with respect to cardiac mass \((p=0.0006)\) and CF \((p=5.4\times10^{-6}\) using post-hoc analysis. At baseline heart rates and LV developed pressure (the difference between LV systolic pressure and LV diastolic pressure) were not significantly different between the strains (Figure 3.2).

<table>
<thead>
<tr>
<th></th>
<th>Cardiac mass</th>
<th>Heart rate</th>
<th>CF</th>
<th>LV developed pressure</th>
</tr>
</thead>
<tbody>
<tr>
<td>BN</td>
<td>0.43 ± 0.04</td>
<td>261 ± 39</td>
<td>14.4 ± 3.6</td>
<td>137 ± 16</td>
</tr>
<tr>
<td>Lewis</td>
<td>0.51 ± 0.07</td>
<td>251 ± 27</td>
<td>9.64 ± 3</td>
<td>116 ± 12</td>
</tr>
<tr>
<td>SHR</td>
<td>0.50 ± 0.04</td>
<td>276 ± 80</td>
<td>8.16 ± 2</td>
<td>130 ± 23</td>
</tr>
<tr>
<td>WKY</td>
<td>0.53 ± 0.08</td>
<td>218 ± 42</td>
<td>10.6 ± 1.9</td>
<td>145 ± 11</td>
</tr>
</tbody>
</table>

Table 3.1 Baseline cardiac phenotypes for four inbred rat strains from the PhysGen database (BN=20, Lewis=10, SHR=20, WKY= 10). Cardiac mass (expressed as % of body weight * 100), heart rate (HR, beats/minute), CF(ml/g/min) and LV developed pressure (mmHg) expressed as mean ± SD.
3.1. *In silico* data analysis

![Graphs showing baseline phenotypes: (a) Cardiac mass indexed to body weight; (b) Heart Rate (beats/minute); (c) CF across strains (ml/gram/min); (d) LV developed pressure. Each dot represents an individual experimental animal.](image)

**Figure 3.2** Baseline phenotypes; (a) Cardiac mass indexed to body weight; (b) Heart Rate (beats/minute); (c) CF across strains (ml/gram/min) (d) LV developed pressure. Each dot represents an individual experimental animal.

Analyses were also carried out for within-group relationships between phenotypes. HR correlated significantly with LV developed pressure in BN, $p=0.03$, adjusted $R^2=0.22$, indicating a statistically significant moderate-sized effect. HR also correlated significantly with CF in the Lewis strain, $p=0.02$, adjusted $R^2=0.42$. These effects were not seen in SHR or WKY or between other phenotypes in BN and SHR. This could reflect lack of correlation between phenotypes or rather large standard deviations in the phenotypes measured which reduced statistical power to detect correlations.

Table 3.2 summarises the post-MI phenotypes. BN and SHR remained the most di-
3.1. *In silico* data analysis

<table>
<thead>
<tr>
<th></th>
<th>Heart Rate</th>
<th>CF</th>
<th>LV developed pressure</th>
</tr>
</thead>
<tbody>
<tr>
<td>BN</td>
<td>226 ± 49</td>
<td>8.39 ± 3.2</td>
<td>71 ± 17</td>
</tr>
<tr>
<td>Lewis</td>
<td>230 ± 23</td>
<td>8.50 ± 3.6</td>
<td>35 ± 19</td>
</tr>
<tr>
<td>SHR</td>
<td>221 ± 59</td>
<td>5.84 ± 2.5</td>
<td>68 ± 35</td>
</tr>
<tr>
<td>WKY</td>
<td>227 ± 65</td>
<td>10.43 ± 9.2</td>
<td>45 ± 18</td>
</tr>
</tbody>
</table>

Table 3.2 Post MI cardiac phenotypes for four inbred rat strains from the PhysGen database (BN=20, Lewis=10, SHR=20, WKY=10). Heart rate (beats/minute), CF (ml/g/min) and LV developed pressure expressed (mmHg) as mean± SD.

vergent strains as far as CF is concerned and the difference between the groups being reduced to borderline statistical significance, *p*=0.05. LV developed pressure also significantly varied within the group (*p*=0.01) with BN and Lewis being the most divergent strains on post-hoc analysis. Significant trait correlations were observed between HR and CF in SHR (*p*=0.009) and between HR and LV developed pressure in WKY (*p*=0.05).

Figure 3.3 MI related phenotypes; (a) CK release (IU/L/gram) at 40 minutes after reperfusion; (b) Area of infarction measured after TTC staining expressed as a ratio of total LV area.

Susceptibility to myocardial ischaemia was quantitatively determined by two measures: CK release for the first 40 minutes after reperfusion and histomorphometric estimation of infarct area as a proportion of total LV area (Figure 3.3). Overall CK release at 40 minutes was significantly different (*p*=0.0001) between strains mainly due to the Lewis strain being more susceptible to CK release compared with the other three strains. The hist-
tomorphometric infarct measurement also vary significantly between strains ($p=4\times10^{-5}$). BN and SHR were the most divergent strains in this regard ($p=5\times10^{-5}$). However these data needed to be tempered by the fact that a significant proportion of results was missing e.g CK at 40 minutes was available for nine out of twenty samples for BN whereas infarct area is reported as zero for four samples.

### 3.2 Cardiovascular physiology in inbred strains

#### 3.2.1 Aims

Following preliminary *in silico* analysis of PhysGen data, in-house cardiac phenotyping was carried out in the four parental inbred strains: BN, SHR, Lewis and WKY. The primary objective of this set of experiments was to establish the phenotyping protocol in our laboratory, compare the data obtained with available data from PhysGen and measure additional cardiac phenotypes including myocardial contractility (LV $dP/dt_{\text{max}}$) and myocardial relaxation (LV $dP/dt_{\text{min}}$).

Although discussed in detail in the methods section 2.2.2, the Langendorff procedure was modified slightly. External pacing was instituted at 360 bpm to keep the heart rate constant with the aim of reducing experimental variability. Regional ischaemia for 35 minutes was induced by reversible ligation of the LAD to reproduce a more realistic model of MI. Hearts were weighed at the end of the experiment instead of the beginning as the priority was to establish cardiac perfusion well within three minutes [Marber et al., 1994].

#### 3.2.2 Parental strain data analysis

Data at baseline are displayed in figure 3.4 and summarized in table 3.3. Cardiac mass varied significantly between the four strains ($p=7.8\times10^6$). Post hoc analysis utilising Tukey’s significance test showed that this was mainly due to higher cardiac mass in SHR compared with Lewis ($p=0.000004$) and BN ($p=0.003$). Difference in cardiac mass between SHR and WKY was not statistically significant ($p=0.31$).

One important observation in these set of experiments when compared with the Phys-
Figure 3.4 Baseline phenotypes; (a) Cardiac mass indexed to body weight; (b) Myocardial Relaxation (mmHg/s); (c) CF(ml/gram/min) (d) LV developed pressure.

Gen dataset was that the CF had more uniform distribution and smaller standard deviation (figure 3.4c). CF also significantly varied between strains ($p=0.0001$). Post-hoc analysis showed that and the effect was mainly due to BN having higher CF than SHR ($p=0.0003$) and Lewis ($p=0.003$). SHR also had significantly lower CF than the parental WKY strain ($p=0.03$). No significant inter-strain variation was observed for LV $dP/dt_{\text{max}}$, LV $dP/dt_{\text{min}}$ or LV developed pressure at baseline.

During the reperfusion phase, CF remained significantly different between strains ($p=0.0004$). Post hoc analysis using Tukey’s significance test showed that this difference was mainly due to difference between BN and SHR ($p=0.0001$). As seen with the
3.2. Cardiovascular physiology in inbred strains

Table 3.3 Baseline cardiac phenotypes: (BN=16, Lewis=15, SHR=19, WKY=9). Cardiac mass (heart weight/body weight)*100, CF (ml/g/min) and LV developed pressure expressed (mmHg) as mean ± SD.

<table>
<thead>
<tr>
<th></th>
<th>Cardiac mass</th>
<th>CF</th>
<th>LV developed pressure</th>
<th>LV dP/dt_{min}</th>
</tr>
</thead>
<tbody>
<tr>
<td>BN</td>
<td>0.45 ± 0.04</td>
<td>12 ± 2</td>
<td>94 ± 18</td>
<td>-2487 ± 487</td>
</tr>
<tr>
<td>Lewis</td>
<td>0.41 ± 0.04</td>
<td>9 ± 2</td>
<td>83 ± 31</td>
<td>-2174 ± 857</td>
</tr>
<tr>
<td>SHR</td>
<td>0.51 ± 0.06</td>
<td>9 ± 1.8</td>
<td>99 ± 20</td>
<td>-2480 ± 499</td>
</tr>
<tr>
<td>WKY</td>
<td>0.47 ± 0.05</td>
<td>11 ± 1.7</td>
<td>108 ± 22</td>
<td>-2667 ± 510</td>
</tr>
</tbody>
</table>

baseline LV pressure-related phenotypes, no significant inter-strain variation was observed for LV dP/dt_{max}, LV dP/dt_{min}, LV developed pressure or % drop in CF as a consequence of LAD ligation. There was also no significant inter-strain variation with regards to MI as measured by histomorphometry but CK release as a measure of myocardial injury varied significantly between strains ($p=0.009$). The significance of the CK release in this data stems from the relative resistance of SHR (64 ± 32 IU/L/gram) to cardiac injury when compared against Lewis (169 ± 120 IU/L/gram, $p=0.008$) and WKY strains (203 ± 94 IU/L/gram, $p=0.008$). SHR did not differ significantly from BN (127 ± 89, $p=0.18$).

<table>
<thead>
<tr>
<th></th>
<th>CF</th>
<th>LV developed pressure</th>
<th>LV dP/dt_{min}</th>
<th>LV dP/dt_{max}</th>
</tr>
</thead>
<tbody>
<tr>
<td>BN</td>
<td>10 ± 2</td>
<td>72 ± 16</td>
<td>-1754 ± 304</td>
<td>2981 ± 733</td>
</tr>
<tr>
<td>Lewis</td>
<td>8 ± 2</td>
<td>67 ± 21</td>
<td>-1651 ± 516</td>
<td>2684 ± 1058</td>
</tr>
<tr>
<td>SHR</td>
<td>7 ± 2</td>
<td>74 ± 17</td>
<td>-1791 ± 395</td>
<td>3044 ± 842</td>
</tr>
<tr>
<td>WKY</td>
<td>8 ± 2</td>
<td>66 ± 23</td>
<td>-1578 ± 474</td>
<td>2871 ± 958</td>
</tr>
</tbody>
</table>

Table 3.4 Post MI cardiac phenotypes: CF (ml/g/min), LV developed pressure expressed (mmHg), LV dP/dt_{min} and LV dP/dt_{max} (mmHg/s) as mean ± SD.

The most significant correlation observed in this dataset was between LV dP/dt_{min} and CF. This relationship was significant at baseline ($p=8.1\times10^{-7}$, adjusted $R^2=0.34$) as well as during reperfusion ($p=3.6\times10^{-7}$, adjusted $R^2=0.36$). CF occurs during diastolic phase of the cardiac cycle and conceptually it is easy to understand that a greater degree of myocardial relaxation correlates with higher CF. This relationship has been described before [Galderisi et al., 2002] but the effect size was smaller and the study examined *in vivo* coronary flow reserve and diastolic function using tissue doppler imaging in humans. No significant relationship was observed between % CF drop during LAD ligation and
3.3 Cardiovascular physiology in F\textsubscript{1} intercross

CK release or infarct measurement on histomphometry in parental strains.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure.png}
\caption{Average CF (ml/gram/min) for the first two minutes following establishment of \textit{ex vivo} perfusion (BN=16, SHR=19).}
\end{figure}

On examining CF data further, the difference in CF between BN and SHR was found to be most divergent during the earliest phases of establishing the \textit{ex vivo} perfusion as shown in figure 3.5. During this phase the coronary vasculature is maximally vasodilated following the inevitable but brief period of cardiac arrest post cardiac excision. This difference raised the possibility of differential vasodilatory response of the coronary circulation to ischaemia between these two strains.

### 3.3 Cardiovascular physiology in F\textsubscript{1} intercross

#### 3.3.1 F\textsubscript{1} intercross: modifications to phenotyping protocol

An F\textsubscript{1} intercross was obtained from mating BN mother and SHR father (BXH) as well BN father and SHR mother (HXB) as detailed in the methods section (2.1). The phenotyping protocol was modified to accommodate the requirement to measure BP (subsection 2.2.1) and maximal CF response to one minute global ischaemia (subsection 2.2.2). To allow accurate BP measurement, the anaesthetic protocol was changed to inhaled isofluorane mixed with Oxygen instead of intra-peritoneal Sodium Pentobarbitone. BP was then measured by direct carotid cannulation. After BP and \textit{ex vivo} HR were recorded, hearts were excised and perfused on Langendorff apparatus. In view of a published report
3.3. Cardiovascular physiology in F\textsubscript{1} intercross

[Ferrera et al., 2009] that prolonged reperfusion beyond 60 min was not useful for function assessment and did not change infarct size measurement, reperfusion was limited to 60 minutes instead of 120 minutes in earlier experiments. BN and SHR parental strains were re-phenotyped to assess the impact of change in protocol on cardiac phenotypes.

3.3.2 Parental strains and F\textsubscript{1} intercross: data analysis

Cardiac mass and BP

Cardiac mass (figure 3.6a and table 3.5) was once again significantly different between BN and SHR, 0.41 ± 0.2 and 0.53 ± 0.3 respectively, \((p=6.1\times10^{-11})\). The distribution of cardiac mass in the F\textsubscript{1} (0.43 ± 0.25) overlap with the BN to a greater degree than the SHR suggesting a possible dominant allelic effect or an effect of BP. Interestingly there is also significant difference between the two F\textsubscript{1} groups depending on direction of the cross. F\textsubscript{1}BXH cardiac mass (0.45 ± 0.01) is significantly higher than F\textsubscript{1}HXB (0.41 ± 0.2), \(p=7.6\times10^{-5}\) indicating a matrilineal effect.

<table>
<thead>
<tr>
<th></th>
<th>Cardiac mass</th>
<th>Diastolic BP</th>
<th>Systolic BP</th>
<th>Mean BP</th>
</tr>
</thead>
<tbody>
<tr>
<td>BN</td>
<td>0.42±0.02</td>
<td>93±9.65</td>
<td>128±15</td>
<td>109±11</td>
</tr>
<tr>
<td>F1BXH</td>
<td>0.45±0.01</td>
<td>103±10.06</td>
<td>150±18</td>
<td>123±12</td>
</tr>
<tr>
<td>F1HXB</td>
<td>0.41±0.02</td>
<td>100±14.31</td>
<td>150±18</td>
<td>123±16</td>
</tr>
<tr>
<td>SHR</td>
<td>0.53±0.03</td>
<td>129±10.92</td>
<td>206±21</td>
<td>159±16</td>
</tr>
</tbody>
</table>

Table 3.5 Cardiac mass ((heart weight/body weight)*100) and BP phenotypes in F\textsubscript{1} intercross (BN=18, F\textsubscript{1}BXH=7, F\textsubscript{1}HXB=10, SHR=17) expressed as mean ± SD.

BP was recorded during F\textsubscript{1} phenotyping experiments given its importance as a covariate for many cardiovascular traits. BP was partitioned into systolic BP (SBP, average of the cyclic peak), diastolic BP (DBP, average of the cyclic trough) and mean BP (MBP), calculated as sum of two-thirds DBP and one-third SBP. BN and SHR had significantly different BP phenotypes; the values for SBP were 129 ± 15 mmHg (mean ± SD) and 206 ± 22 mmHg respectively and the difference was statistically extremely significant, \(p=6\times10^{-10}\). DBP values were 93 ± 9.6 mmHg for BN and 130 ± 11 mmHg for SHR, \(p=3.8\times10^{-10}\). MBP was also significantly different between BN, 109 ± 11 mmHg and
3.3. Cardiovascular physiology in F\(_1\) intercross

Figure 3.6 Cardiac mass and BP phenotypes in F\(_1\) intercross (BN=18, F\(_1\)BXH=7, F\(_1\)HXB=10, SHR=17), (a) Cardiac mass indexed to body weight; (b) Mean BP (mmHg); (c) Systolic BP (mmHg) (d) Diastolic BP (mmHg).
3.3. Cardiovascular physiology in F\textsubscript{1} intercross

SHR, 159 ± 16 mmHg, \( p=1.0^{+8} \). The phenotype distribution of BP in the F\textsubscript{1} was closer to BN (SBP 151 ± 17 mmHg, MBP 123 ± 14 mmHg and DBP 101 ± 12 mmHg) as compared with SHR. There was no effect of maternal lineage influenced inheritance on BP (figure 3.6).

![Graphs showing CF phenotypes in F\textsubscript{1} intercross](image)

**Figure 3.7** CF phenotypes in F\textsubscript{1} intercross, (a) baseline; (b) following one minute global ischaemia to induce maximal hyperaemia; (c) reperfusion; (d) Coronary flow reserve (CFR), ratio of maximal hyperaemic CF to baseline CF.

**Coronary flow phenotypes**

CF analysis showed that BN had higher CF than SHR during all phases of the experiment (table 3.6) but particularly so during period of maximal hyperaemia after one
minute global ischaemia \( (p=1.0 \times 10^{-7}) \) and reperfusion following release of LAD ligation \( (p=4.6 \times 10^{-9}) \). No significant difference in CF was observed between the two reciprocal F\(_1\) intercrosses during any experimental stage.

<table>
<thead>
<tr>
<th></th>
<th>Baseline CF</th>
<th>Maximal CF</th>
<th>Reperfusion CF</th>
</tr>
</thead>
<tbody>
<tr>
<td>BN</td>
<td>10 ± 1.3</td>
<td>21.6 ± 3.5</td>
<td>11.3 ± 1.6</td>
</tr>
<tr>
<td>F1BXH</td>
<td>8 ± 1.1</td>
<td>19.2 ± 1.8</td>
<td>8.3 ± 0.9</td>
</tr>
<tr>
<td>F1HXB</td>
<td>8 ± 0.7</td>
<td>17.7 ± 2.3</td>
<td>9.5 ± 1.2</td>
</tr>
<tr>
<td>SHR</td>
<td>9 ± 1.6</td>
<td>13.4 ± 2.7</td>
<td>7.6 ± 1.3</td>
</tr>
</tbody>
</table>

Table 3.6 CF (ml/gram/minute) in the parentals and F\(_1\) strains at baseline, in response to one minute ischaemia and reperfusion expressed as mean ± SD.

At baseline the F\(_1\) intercross CF distribution was the same as SHR \( (p=0.6\), figure 3.7a) suggesting a dominant SHR allelic effect given the differences in BP. The same was also true for reperfusion CF when the difference between BN and F\(_1\) intercross was greater \( (p=5.13 \times 10^{-5}) \) than the difference between SHR and F\(_1\) \( (p=0.003\), figure 3.7c). However maximum hyperaemic CF in F\(_1\) intercross was more BN-like than SHR \( (p=5 \times 10^{-6})\). As CFR is a first order derivative of maximum hyperaemic CF, perhaps it is not surprising that in this regard the F\(_1\) intercross again resembled BN \( (p=0.22)\) more than SHR \( (p=5 \times 10^{-10})\). In summary, F\(_1\) intercross CF was BN-like in response to hyperaemic stimulus but SHR-like at baseline and reperfusion.

**LV pressure phenotypes**

LV developed pressure at baseline was different between BN \( (70 \pm 20 \text{ mmHg}) \) and SHR \( (89 \pm 23 \text{ mmHg}) \). This difference was of borderline statistical significance, \( p=0.05 \). With regards to LV developed pressure at baseline the F\(_1\) intercross \( (68 \pm 14 \text{ mmHg}) \) was significantly different when compared with SHR \( (p=0.005)\) but not BN \( (p=0.94)\). During reperfusion LV developed pressure did not differ significantly between BN \( (71 \pm 23 \text{ mmHg})\) and SHR \( (74 \pm 15 \text{ mmHg})\) or the F\(_1\) intercross \( (65 \pm 18 \text{ mmHg})\).

LV dP/dt\(_{\text{max}}\) being a first-order derivative of LV developed pressure was significantly different between BN \( (2270 \pm 810 \text{ mmHg/s})\) and SHR \( (2610 \pm 555 \text{ mmHg/s}\), \( p=0.0002\) at baseline. LV dP/dt\(_{\text{max}}\) did not significantly differ between the two parental strains.
3.3. Cardiovascular physiology in F\textsubscript{1} intercross

![Graphs showing LV contractility in F\textsubscript{1} intercross](a)\, (b)\, (c)\, (d)

**Figure 3.8** LV contractility in F\textsubscript{1} intercross, (a) \(dP/dt_{\text{min}}\) at baseline; (b) \(dP/dt_{\text{min}}\) during reperfusion; (c) \(dP/dt_{\text{max}}\) at baseline; (d) \(dP/dt_{\text{max}}\) during reperfusion.

during reperfusion, \(p=0.96\). Similarly although LV \(dP/dt_{\text{min}}\) was significantly different \((p=0.002)\) between BN (-1369 ± 410 mmHg/s) and SHR (-1939 ± 548 mmHg/s) at baseline, this significant did not persist during reperfusion \((p=1)\). At baseline F\textsubscript{1} intercross LV \(dP/dt_{\text{max}}\) and LV \(dP/dt_{\text{min}}\) were both significantly different from SHR, \(p=0.002\) and \(p=0.001\), respectively. These differences were not seen during reperfusion. In summary, with regards to LV pressure phenotypes, F\textsubscript{1} intercross was more similar to BN at baseline with no significant difference between parentals and F\textsubscript{1} intercross during reperfusion.
3.4 Discussion

Correlation between cardiac phenotypes

Analysis of correlation between cardiac traits in all strains revealed that LV dP/dt_{min} correlated with CF significantly at baseline (p=0.006, adjusted R^2=0.14) although within-group correlation analysis for these traits did not achieve statistical significance. This correlation was also not statistically significant across the rat strains as a whole or individually during the reperfusion phase.

3.4 Discussion

Analysis of PhysGen data provided basic preliminary view on inter-strain variation in cardiac traits. The conclusion of the PhysGen data analysis was that SHR and BN were the most divergent strains with respect to some important phenotypes including cardiac mass and CF.

The analysis of PhysGen data also revealed correlation between cardiac traits. Correlation was observed between HR and CF in the Lewis strain at baseline and between HR and CF in SHR during reperfusion. Similarly there was significant correlation between HR and LV developed pressure in BN at baseline and in WKY during reperfusion. The lack of consistent correlation could be attributed to large standard deviations in CF observed in the PhysGen data. An alternative explanation is that the correlations exist only in some strains and not in others as a results of their specific genetic make-up. For example a strain might have genetic predisposition to both tachycardia and hypertension and when such a strain is studied one might be tempted to conclude that “HR correlates with BP”. However in a cross with another normal strain which has normal HR and BP, this association will breakdown as alleles for HR and BP segregate independently.

Comparison of in-house and PhysGen data for phenotypes of the four inbred parentals strains revealed some important similarities and dissimilarities. The differences between these two strains, BN and SHR, persist despite modifications of the phenotyping protocol including cardiac pacing and weighing the heart at the end of the procedure. Since genetic mapping is likely to be successful if the parental strains show maximally divergent phe-
notypes it seemed that BN x SHR cross would be the most promising strategy for genetic studies of cardiac mass and CF. Additionally as hearts were paced during in-house experiments there was no reason to include HR in the covariate analyses thereby reducing the number of confounders. Perhaps for this reason inter-strain differences between developed pressures seen in the PhysGen during reperfusion were not seen in our laboratory.

There was strong correlation between LV dP/dt_{min} and CF at baseline. This relationship has been observed in human subjects with established hypertension [Galderisi et al., 2002] but the mechanism remains unexplained. This is an interesting observation which needs to be considered in follow-on experiments for two reasons. First, LV dP/dt_{min} would need to be taken into account when genetic mapping is performed for CF. Second, if the correlation remains significant in follow-on experiments it could potentially explain the impairment of CF in a number of pathophysiological states including cardiomyopathies [Camici and Crea, 2007].

A criticism of the original data obtained in parental strains is the lack of inclusion of another very important cardiovascular covariate, namely BP. The SHR was bred for hypertension [Okamoto and Aoki, 1963] and hypertension has been associated with cardiac hypertrophy and impaired CF. It might be argued that the phenotype differences between BN and SHR could be attributable to BP differences between these strains. However, previous quantitative trait analysis of cardiovascular traits including cardiac mass reveals that increases in left ventricular afterload account for a small fraction of the increase in left ventricular mass in rats and humans [Petretto et al., 2008]. This argument aside, having prioritised BN and SHR for further study, BP was measured and taken into account in all subsequent studies.

Analysis of F_1 data showed a difference in cardiac mass between F_1BXH and F_1HXB. There are two possible explanation for these results. The first explanation, potentially, is that these effects are likely to be matrilineal and secondary to variations in the mitochondrial genome. Mitochondrial genome has been linked to pathogenesis of type 2 diabetes and other complex traits [Alcolado et al., 2002]. The ideal tool for studying these matrilineal effects are conplastic strains in which female animals with the mitochondrial genome
3.4. Discussion

of interest are backcrossed to the male animals with background nuclear genome of interest at each generation for ten generations. This approach was taken by Pravenec et al. [Pravenec et al., 2007] who produced SHR conplastic strain that had BN mitochondrial DNA. Their results showed that replacing mitochondrial genome of the SHR with that of BN adversely influenced several metabolic phenotypes. Interestingly, the F₁ results from the data under discussion suggest that the inheritance of BN mitochondrial genome is directly related to higher cardiac mass, an adverse cardiovascular phenotype. A second potential explanation is that these effects are secondary to Y chromosome inheritance as F₁BXH males inherit Y chromosome from the SHR male parent and F₁HXB inherit it from the BN male parent. Kren et al. [Kren et al., 2001] showed that replacing Y chromosome in the parental SHR strain with Y chromosome from BN resulted in reduced BP in the consomic SHR.BN-Y. Sexual dimorphism in adverse cardiovascular phenotypes in humans is well recognised with evidence that Y chromosome is associated with risk of coronary artery disease in men of European ancestry, possible acting via inflammatory pathways [Charchar et al., 2012]. However, these two explanations for difference in cardiac mass are not mutually exclusive and different mechanisms may be at work in different models [Barrick et al., 2009].

The BN strain had a dominant allelic effect with regards to cardiac mass, BP and hyperaemic CF in the F₁ intercross. SHR has dominant allelic effect with regards to CF at baseline and reperfusion. Although observation of opposite dominant allelic effects can allow one to begin to speculate on independent genetic control of BP and CF, definitive conclusions would have to wait until the F₂ intercross had been derived and phenotyped. This also poses a question as to whether an intercross (F₂) or backcross (BC) strategy is the best breeding strategy to map QTLs. In case of dominant effects it is more efficient to use a BC strategy using the F₁ x non-dominant parent strategy. But such an approach applied in our case would reduce the power to detect loci for baseline and reperfusion CF. An intercross provides a more ‘general’ picture of the trait - the number of dominant and additive QTLs whereas a BC is preferred if the objective is to detect at least some major QTLs [Darvasi, 1998]. As the objective of current research was to get a general picture
of genetic control of cardiac traits, an intercross strategy was deemed to be potentially more efficient.

By the end of F\textsuperscript{1} phenotyping, it was clear that cardiac mass, BP and CF are consistently different between BN and SHR. This does not mean that other traits cannot be genetically determined - just that cardiac mass, BP and CF are more likely to be genetically tractable. It was also observed that CF has a consistent relationship with myocardial relaxation which would need to be accounted for if CF were to be mapped to the genome in subsequent studies. Additionally, the phenotyping protocol incorporating BP measurement and CF hyperaemia used for F\textsubscript{1} and parental re-phenotyping was finalised for much larger F\textsubscript{2} intercross study. The final experiment protocol is shown in figure 2.3.
Chapter 4

Cardiovascular physiology in F$_2$ intercross

4.1 Analysis of phenotypic drift

F$_2$ phenotyping was performed in 196 animals over the course of a year. Due to the time it took to phenotype the whole F$_2$ intercross it is important to review the long-term phenotype data to assess for two features: systematic drift and outliers. Systematic drift can arise due to a number of reasons: batch effects, seasonal variation, lack of calibration or unappreciated malfunction in the measuring apparatus, changes to animal maintenance protocols such as feed and the learning curve of the experimenter. Attempts were made to prevent such bias which could increase non-genetic variation and reduce the ability to detect QTLs. Equipment was calibrated and serviced regularly. No changes were made to animal maintenance protocols during the course of the studies. F$_2$ experiments were undertaken after substantial experience was gained in performing the procedure. It is also important to assess the data for outliers which may arise due to phenotyping errors, inaccurate data transcription or may be truly unusual individuals and so deserve careful assessment.
4.1. Analysis of phenotypic drift

**Figure 4.1** Cardiovascular phenotypes of individual animals (circles), along the y-axis, corresponding to the order in which they were measured for 196 F₂ animals over the course of one year with the phenotype values plotted along the x-axis; (a) Cardiac mass indexed to body weight; (b) Mean BP (mmHg); (c) CF across strains (ml/gram/min).
4.1. Analysis of phenotypic drift

Figure 4.2 Cardiovascular phenotypes of individual animals (circles), along the y-axis, corresponding to the order in which they were measured for 196 F2 animals over the course of one year with the phenotype values plotted along the x-axis; (a) in vivo HR during direct carotid cannulation; (b) LV dP/dt_{min} (mmHg/s).

Figure 4.1 shows the trend data for cardiac mass (4.1a), mean BP (4.1b) and CF at baseline (4.1c). There is no systematic bias but cardiac mass shows the presence of a few individuals with much higher cardiac mass (> 2 SD) when compared with the overall distribution producing a positive skew in the data. On review, these seven outliers were truly different from the average and therefore were analyzed separately.

Shapiro-Wilk normality test did not show significant deviation from normality in mean BP (p=0.8) and baseline CF data (p=0.6). Figure 4.2 shows the trend for in vivo HR and LV dP/dt_{min} again showing no systematic deviation in the long-term data trend.
4.2 F₂ intercross: data analysis

F₂ cardiac phenotype analysis

Cardiac mass did not significantly differ between F₂BXH (0.44 ± 0.05, n=85) and F₂HXB (0.45 ± 0.05, n=111), p=0.14. Figure 4.3 displays the cardiac mass data for F₂ intercross in context of the parental BN and SHR strains as well as F₁ intercross. Systolic, mean and diastolic BP for the F₂ strain were 146 ± 21 mmHg, 120 ± 16 mmHg and 100 ± 14 mmHg respectively. The statistical distribution of BP in the F₂ intercross were closer to BN (p=0.007) than SHR (p=1.0×10⁻⁸) mirroring the results obtained in F₁ generation (section 3.3). There was no significant variation in in vivo or ex vivo HR. These results are summarized in table 4.1 and displayed in figure 4.4.

![Cardiac mass data](image)

**Figure 4.3** Cardiac mass in the F₂ intercross compared with the parentals and F₁ intercross.

CF in the F₂ strain was intermediate between BN and SHR during all stages of the experiment (table 4.2). No significant difference was seen for CF between F₂BXH and F₂HXB during any experimental stage (p=0.9 at baseline and p=0.6 during reperfusion). CFR values for the F₂ intercross falls more in the SHR phenotype distribution (p=0.002) than the BN (p=2.7×10⁻¹³).
4.2. F₂ intercross: data analysis

<table>
<thead>
<tr>
<th></th>
<th>Systolic BP</th>
<th>Mean BP</th>
<th>Diastolic BP</th>
<th>ex vivo HR</th>
<th>in vivo HR</th>
</tr>
</thead>
<tbody>
<tr>
<td>BN</td>
<td>128 ± 15</td>
<td>109 ± 11</td>
<td>93 ± 10</td>
<td>288 ± 31</td>
<td>362 ± 25</td>
</tr>
<tr>
<td>F₁</td>
<td>151 ± 18</td>
<td>123 ± 14</td>
<td>101 ± 12</td>
<td>348 ± 40</td>
<td>378 ± 33</td>
</tr>
<tr>
<td>F₂</td>
<td>146 ± 21</td>
<td>120 ± 16</td>
<td>100 ± 14</td>
<td>315 ± 37</td>
<td>365 ± 31</td>
</tr>
<tr>
<td>SHR</td>
<td>206 ± 22</td>
<td>159 ± 17</td>
<td>130 ± 11</td>
<td>273 ± 11</td>
<td>373 ± 20</td>
</tr>
</tbody>
</table>

**Table 4.1** Parental, F₁ and F₂ intercross BP (mmHg) expressed as mean ± SD (BN=18, SHR=15, F₁=17, F₂=188). HR data expressed as bpm.

![Box plots of BP indices](image)

**Figure 4.4** BP (mmHg) indices in the F₂ compared with the parentals strains.

<table>
<thead>
<tr>
<th></th>
<th>Baseline CF</th>
<th>Maximal CF</th>
<th>Ischaemic CF</th>
<th>Reperfusion CF</th>
<th>CFR</th>
</tr>
</thead>
<tbody>
<tr>
<td>BN</td>
<td>10.2 ± 1.3</td>
<td>21.6 ± 3.5</td>
<td>6.2 ± 1.3</td>
<td>11.3 ± 1.7</td>
<td>1.72 ± 0.32</td>
</tr>
<tr>
<td>F₁</td>
<td>8.2 ± 0.9</td>
<td>18.3 ± 2.2</td>
<td>4.8 ± 1.0</td>
<td>8.9 ± 1.2</td>
<td>1.47 ± 0.18</td>
</tr>
<tr>
<td>F₂</td>
<td>9.9 ± 1.4</td>
<td>18.7 ± 2.9</td>
<td>5.3 ± 1.2</td>
<td>9.9 ± 1.8</td>
<td>1.23 ± 0.25</td>
</tr>
<tr>
<td>SHR</td>
<td>8.9 ± 1.6</td>
<td>13.4 ± 2.7</td>
<td>4.3 ± 0.9</td>
<td>7.6 ± 1.3</td>
<td>1.02 ± 0.12</td>
</tr>
</tbody>
</table>

**Table 4.2** CF (ml/gram/minute) in BN (n=18), SHR (n=15), F₁(n=17) and F₂(n=188) expressed as mean ± SD. CFR expressed as a ratio of maximal hyperaemic CF and baseline CF.

Myocardial relaxation in the F₂ intercross was -1430.264 ± 553 mmHg/s at baseline,
4.2. \( F_2 \) intercross: data analysis

-1770 ± 612 mmHg/s in response to global ischaemia and -1283 ± 487 during reperfusion. At baseline, LV dP/dt\(_{\text{min}}\) in the \( F_2 \) intercross was different from SHR (\( p=0.002 \)) whereas in response to global ischaemia and during reperfusion there were no significant differences between the parents and \( F_2 \) (figure 4.5). Myocardial contractility (LV dP/dt\(_{\text{max}}\)) mirrored the observations for LV dP/dt\(_{\text{min}}\) as \( F_2 \) intercross was different from SHR at baseline (2298 ± 740 mmHg/s and 2996 ± 695 mmHg/s respectively, \( p=0.001 \)) but not during the rest of the experimental stages.

**Figure 4.5** LV dP/dt\(_{\text{min}}\) (mmHg/s) in the \( F_2 \) intercross compared with the parentals strains. During baseline SHR has greater relaxation values than BN and \( F_2 \) intercross.

**Covariate analysis in the \( F_2 \) intercross**

Cardiac trait covariate analysis in the \( F_2 \) intercross revealed interesting results. There was significant correlation between LV dP/dt\(_{\text{min}}\) and CF at baseline (\( p=4.62\times10^{-11} \), adjusted \( R^2=0.21 \), figure 4.6), in response to global ischaemia (\( p=1.06\times10^{-8} \), adjusted \( R^2=0.16 \)) and during reperfusion (\( p=2.6\times10^{-14} \), adjusted \( R^2=0.26 \)). Therefore this is a
4.2. F$_2$ intercross: data analysis

**Figure 4.6** Correlation between CF and LV $dP/dt_{min}$ at baseline, $p = 4.62 \times 10^{-11}$, adjusted $R^2=0.21$.

**Figure 4.7** Correlation matrix between cardiac mass, systolic, mean and diastolic BP in the F$_2$ intercross. The numbers in the upper panel are Pearson correlation values ‘r’ and significance of results are represented by ‘*’ ($p < 0.05$=‘*’, $p < 0.01$=‘**’, $p < 0.001$=‘***’). Actual data points are plotted in the lower panel with loess lines.
Figure 4.8 Correlation matrix between systolic, mean and diastolic BP and CF in the F₂ intercross. The numbers in the upper panel are Pearson correlation values ‘r’ and significance of results are represented by ‘*’ (p < 0.05=‘*’, p < 0.01=‘**’, p < 0.001=‘***’). Actual data points are plotted in the lower panel with loess lines.

moderate-sized statistically significant effect. Similarly significant but marginally smaller effect was also seen for LV dP/dt_max and CF at baseline (p=7.5×10⁻⁹, adjusted R²=0.16) and reperfusion (p=1.02×10⁻⁷, adjusted R²=0.13). There was no statistically significant correlation between cardiac mass and BP (systolic, mean or diastolic). The cardiac mass outliers (cardiac mass > 0.59) were excluded before the analysis to ensure that requirements for parametric analyses were met (figure 4.7) and reduce the risk of a type II error. There was a statistically significant effect of cardiac mass on LV dP/dt_min, p=0.02 but the effect size was small, adjusted R²=0.12. Likewise there was no suggestion of a significant correlation between CF during any experimental stage and BP in the F₂ intercross. The only exception being a weak effect of diastolic BP on reperfusion CF (p=0.05, adjusted
4.3 Heritability of cardiac traits

R²=0.10). The relationship between LV dP/dt_{min}, LV dP/dt_{max} and BP was also examined in the F₂ intercross. There was no appreciable effect of systemic BP on LV dP/dt_{min} (adjusted R²=0.015) or LV dP/dt_{max} (adjusted R²=0.009).

4.3 Heritability of cardiac traits

As discussed in Section 1.3, heritability is the proportion of phenotypic variance that is explained by genotypic variance for a given population in a specific environment. Heritability answers the question of degree of genetic contribution to total phenotypic variance. For F₂ intercross, broad-sense heritability can be measured directly

\[
H^2 = \frac{\sigma_{F_2}^2 - \left(\frac{\sigma_{BN} + \sigma_{SHR} + \sigma_{F_1}}{3}\right)}{\sigma_{F_2}^2}
\]

Using the formula above for the trait variances in current study gives the heritability values reported in table 4.3. Since BP is known to be a heritable trait with well characterized genes, QTLs and therapeutics [Rapp, 2000], the data shows that by comparison most traits under study are at least if not more heritable than BP.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Broad sense Heritability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systolic BP</td>
<td>0.26</td>
</tr>
<tr>
<td>Mean BP</td>
<td>0.24</td>
</tr>
<tr>
<td>Diastolic BP</td>
<td>0.37</td>
</tr>
<tr>
<td>Cardiac mass</td>
<td>0.52</td>
</tr>
<tr>
<td>In vivo heart rate</td>
<td>0.31</td>
</tr>
<tr>
<td>Baseline CF</td>
<td>0.28</td>
</tr>
<tr>
<td>Reperfusion CF</td>
<td>0.39</td>
</tr>
<tr>
<td>Coronary flow reserve</td>
<td>0.64</td>
</tr>
<tr>
<td>LV dP/dt_{min}</td>
<td>0.40</td>
</tr>
<tr>
<td>LV dP/dt_{max}</td>
<td>0.51</td>
</tr>
</tbody>
</table>

Table 4.3 Broad sense heritabilities of cardiovascular traits in the F₂ intercross.
4.4 Discussion

Analysis of the F2 data shows that there had been no systematic drift in the phenotypes that could be attributed to inadvertent changes in the calibration of experimental apparatus or technique, seasonal variation, batch effects or any changes in the animal housing. There are, however, outliers in the cardiac mass phenotype. These outliers, seven in number out of a total of 196 were defined by >2 SD from the mean cardiac mass phenotype, figure 4.1a. They appear at regular intervals in the experimental timeline. The potential explanations for this observation are allelic rearrangements or de novo mutations. Rat Genome Database, [http://rgd.mcw.edu/](http://rgd.mcw.edu/) [Twigger et al., 2007] lists 73 records for cardiac mass QTLs in various rat crosses and given the number of loci, it is probable that random assortment resulting in gain of pro-hypertrophic loci or loss of anti-hypertrophic loci has resulted in an outlier phenotype in a small number of animals. Spontaneous mutation rate is unlikely to account for this observation since that would imply an unrealistic high mutation rate of $5 \times 10^{-4}$ per locus per generation in the current rat cross assuming the extreme scenario that all 73 cardiac mass QTLs contribute to cardiac mass in a cross between BN and SHR. Assuming fewer loci would lead to estimation of even higher mutation rate. This mutation rate would be greater than reported data for mice ($6.6 \times 10^{-6}$) and humans ($2 \times 10^{-8}$) [Drake et al., 1998].

There is no significant correlation in the F2 intercross between BP and cardiac mass. This observation, although interesting is not novel and has been documented in hypertensive animals [Mulvany and Korsgaard, 1983] [Pravenec et al., 1995], normotensive animals [Sebkhi et al., 1999], as well as humans [Hartford et al., 1983]. There is, however, a large amount of literature linking BP to cardiac mass and anti-hypertensive treatment to regression in cardiac hypertrophy [Ruskoaho, 1984] [Verdecchia et al., 1990]. One explanation could be that hypertension acts on a genetic background of predisposition to cardiac hypertrophy which is different in different populations e.g Afro-Carribians are known to be at higher risk of both cardiac hypertrophy and hypertension. Under such conditions correlation between BP and cardiac mass would be observed due to shared predisposition to both conditions. Alterations in the renin-angiotensin pathway are known to patho-
4.4. Discussion

Physiologically influence both cardiac mass and BP and in this regard it is interesting to note that DD genotype of the ACE allele carries a higher risk of cardiac hypertrophy for all grades of BP [Schunkert et al., 1994] indicating that predisposition to cardiac hypertrophy in hypertensive patients is accounted for to a greater degree by the genotype than BP.

There is little evidence in the published literature relating to direct correlation between BP and myocardial relaxation (LV dP/dt$\text{min}$) but indirect correlations are available if a surrogate clinical phenotype termed “heart failure with preserved (>50%) ejection fraction” (HFpEF) is used. Hypertension is considered to be a risk factor for HFpEF (as well as heart failure with reduced ejection fraction, HFrEF), along with female sex and increasing age [Borlaug and Redfield, 2011] but this relationship remains poorly defined at pathophysiological level [Volpe et al., 2010]. In the data on F$_2$ intercross no significant correlation was observed between BP and LV dP/dt$\text{min}$ and yet both traits were significantly heritable indicating that in this experimental cross the genetic variance plays a more important role in determining LV dP/dt$\text{min}$ than BP.

There was a significant relationship of minor effect size between cardiac mass and LV dP/dt$\text{min}$. This relationship has also been demonstrated previously but with divergent hypotheses posited for this correlation including direct effect of deteriorating LV dP/dt$\text{min}$ with increasing cardiac mass [Pavlopoulos et al., 2008], presence of interstitial fibrosis [Brilla et al., 1991, Brilla et al., 2000] and impaired endothelial function due to increased reactive oxidation species (ROS) [MacCarthy and Shah, 2000].

The lack of evidence for correlation between BP and LV dP/dt$\text{min}$ also holds true for the relationship between BP and CF. In the F$_2$ intercross there was no significant correlation between BP and CF, which combined with significant H$^2$ estimates point towards independently segregating genetic determinants of both traits. This is an unexpected observation as hypertension is acknowledged to lead to impaired CF in humans [Laine et al., 1998, Leschke et al., 1992, Hoenig et al., 2008]. Nevertheless the data obtained in the rat intercross is unambiguous in its lack of correlation between BP and CF. The difference in observations may be because the traits are genetically influenced
to a far greater degree in inbred rats when compared with an outbred human population. But there are additional factors to be considered: the experimental methods in humans are often imprecise and indirect measures of the phenotype; statistical correlations which link hypertension to impaired CF do not imply *causation*, which could always be due to another hidden variable (such as genetic factors) and controlled crosses of the type performed in these experiments are, for obvious reasons, impossible in humans.

One of the most significant correlations in the current data set is between CF and LV $dP/dt_{\text{min}}$ during all phases of the experiment. CF is a diastolic phenomenon and therefore greater LV $dP/dt_{\text{min}}$ allows for higher CF. The reverse can also be true in that impaired CF would impair myocardial relaxation. It would also not be implausible to speculate that CF and LV $dP/dt_{\text{min}}$ are correlated because they are determined by another unobserved variable e.g it has been demonstrated that Nitric Oxide (NO) derived from the endothelium plays an important role in enhancing LV $dP/dt_{\text{min}}$ [MacCarthy and Shah, 2000] and NO is also an important regulator of CF. Since the aim of the current study is to discover novel genetic determinants of cardiac traits, the correlation between these two traits needs to be accounted for in QTL mapping to avoid a false positive QTL discovery.
Chapter 5

QTL mapping

5.1 Data quality and genotyping errors

QTL mapping was performed in R [R Development Core Team, 2011] using the R/qtl package [Broman et al., 2003]. R is a free platform for statistical computation and graphics that is widely used by academic statisticians. R/qtl is freely available add-on package to R, accompanied by detailed documentation on how to use the functions provided with practical examples. The analysis pipeline consisted of data diagnostics to identify errors, single QTL mapping usually with interval mapping, two-dimensional two-QTL scans and finally fitting of a global multiple-QTL model.

Phenotype and genotype data was organized in the format required by R/qtl. The fidelity of long-term phenotype data has already been assessed in section 4.2 and methodology for quality assessment of genotyping data is described in section 2.3.4. Genotype data was available for 184 F\textsubscript{2} individuals. 768 markers in total were typed for each of the 184 individuals. This amounts to more than 140,000 PCR reactions and even with 0.5% error rate, there could be up to 2800 inaccurate genotypes in the current data. It is therefore important to identify and exclude the erroneous genotypes as much as possible [Broman, 1999]. The summary of the genotyping data prior to QC is summarized in table 5.1.
5.1. Data quality and genotyping errors

F₂ intercross

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of individuals</td>
<td>184</td>
</tr>
<tr>
<td>Total markers typed (including X)</td>
<td>768</td>
</tr>
<tr>
<td>Total markers typed (excluding X)</td>
<td>746</td>
</tr>
<tr>
<td>No. of autosomes</td>
<td>20</td>
</tr>
<tr>
<td>% Genotyped</td>
<td>93%</td>
</tr>
<tr>
<td>Genotypes (%)</td>
<td>AA (24.3) AB (51.5) BB (24.2)</td>
</tr>
</tbody>
</table>

**Table 5.1** Summary of the genotyping data prior to quality checking.

5.1.1 Segregation distortion and identical genotypes

Segregation distortion is the presence of significant deviation of genotype frequencies from the expected Mendelian 1:2:1 ratio expected for co-dominant SNP markers. This may indicate one of two issues: the genotype frequencies are correct and the absence of certain alleles is due to the association with a locus associated with early mortality or that there are genotyping errors leading to deviation. There is no absolute difference between the two but real segregation distortion is likely to be represented by segregation distortion of several contiguous markers. Table 5.3 shows the three markers for which segregation distortion was found. These markers were excluded from further analyses.

<table>
<thead>
<tr>
<th>Marker</th>
<th>Chromosome</th>
<th>BB</th>
<th>AB</th>
<th>AA</th>
</tr>
</thead>
<tbody>
<tr>
<td>173312790</td>
<td>4</td>
<td>92</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>245593</td>
<td>10</td>
<td>1</td>
<td>24</td>
<td>67</td>
</tr>
<tr>
<td>44100413</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>91</td>
</tr>
</tbody>
</table>

**Table 5.2** Segregation distortion results showing three markers on chromosomes 4, 10 and 16 showing extreme segregation distortion.

Given the number of individuals in current study it would be useful to compare the genotypes for all individuals to assess for unusual similarities between samples which might indicate duplication of samples. There were no 100% identical genotypes in the data and this is shown in figure 5.1.
5.1. Data quality and genotyping errors

5.1.2 Assessment of marker order

It is important to establish that markers used for genetic mapping are placed in accurate order on the right chromosome [Broman and Sen, 2009]. To have misplacement of markers would risk the entire QTL mapping results. Recombination fraction (RF) is calculated between markers to be certain that markers are in the order, since closely linked markers have smaller RF and that markers on different chromosomes are not linked. This is displayed in figure 5.2 which shows RF between markers in the upper panel (markers with abnormal low RF will be red) and LOD scores (linkage scores > 3 in red) in the lower panel. No apparent problems are detected in this data. Genetic map of the markers based on RF is shown in figure 5.3 showing normal map without any unexpected gaps and no linkage of markers on different chromosomes.

In addition to checking linkage between markers on different chromosomes, the order
5.1. Data quality and genotyping errors

**Figure 5.2** Estimated LOD scores (lower right) and RF (upper left) for the F$_2$ intercross data. Low RF (linked) markers would be shown in red.

**Figure 5.3** Map based on genetic distances between the markers in F$_2$ intercross data.
of markers on the individual chromosomes can also be checked by using the number of crossovers between markers [Broman and Sen, 2009]. The method works relatively simply: crossovers are counted between markers on a given chromosome given their current position. The marker order is the shuffled and number of crossovers are calculated again. This process is iterated for more than a thousand times. The model with the marker positions that gives the minimum number of crossovers is likely to be the most accurate model. For the $F_2$ data the marker position was adjusted for chromosomes 6 (initial number of crossovers = 327, best-marker-fit crossover number = 298).

### 5.1.3 Assessment of crossovers

Genotyping errors may also appear as tight double crossovers [Broman and Sen, 2009]. Crossovers show significant neighbouring interference and therefore for a reasonably dense map, a genotype that is that is out-of-sync with the surrounding genotypes is likely to be erroneous. Probability of a given genotype given the marker genotypes for the chromosome is compared with the probability of the genotype being accurate and reported as LOD scores in R/qtl. Majority of the erroneous genotypes were confined to six individuals. These animals were excluded from further analysis.

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>ID</th>
<th>Marker</th>
<th>LOD$_{\text{error}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>99</td>
<td>46512082</td>
<td>17</td>
</tr>
<tr>
<td>13</td>
<td>99</td>
<td>24991885</td>
<td>12</td>
</tr>
<tr>
<td>13</td>
<td>99</td>
<td>22547261</td>
<td>12</td>
</tr>
<tr>
<td>7</td>
<td>99</td>
<td>80419691</td>
<td>10.6</td>
</tr>
<tr>
<td>13</td>
<td>99</td>
<td>7744244</td>
<td>10.5</td>
</tr>
<tr>
<td>3</td>
<td>99</td>
<td>71270201</td>
<td>10.4</td>
</tr>
</tbody>
</table>

**Table 5.3** Six highest scoring erroneous genotypes likely representing genotyping errors detected by identifying tight double crossovers, interestingly all in the same individual indicating that there might have been errors during DNA extraction/processing. There were another twenty genotypes with error score $5 > $LOD$_{\text{error}}$ confined to six individuals.

The number of crossovers events in each individual can also be counted and serves as another diagnostic measure. The data for individuals with excessively high or low number of crossovers should be carefully reviewed. This may represent errors in genotyping or
sub-optimal DNA quality. There are six individuals who exhibit far greater number of crossovers than the rest of the group, figure 5.4. The very high number of crossovers (>100) is suspicious and these individuals were also excluded from any further analysis.

![Figure 5.4](image)

**Figure 5.4** Number of crossover events plotted against the $F_2$ index, there are six individuals with excessive number of crossover (> 100).

### 5.1.4 Missing genotype information

QTL mapping methods such as interval mapping work reliably only when there are no areas of low genotype information. False positive QTLs may be produced by interval mapping in areas of low genotype information. This information can be obtained by using the `plot.info` function in R/qtl where the value for a fully typed marker is 0 and for a marker with no genotype information is 1. There were only two markers with more than 20% missing information thereby eliminating the risk of false positive QTL discovery due to missing information. Table 5.4 summarizes the $F_2$ intercross data following all the steps to remove mistyped and misplaced markers as well as individuals with large numbers of errors.
5.2 Single QTL analysis

5.2.1 Establishing LOD significance thresholds

As mentioned in section 1.3.2 a larger LOD score provides greater evidence for existence of a QTL. The LOD score is the log-likelihood ratio of the probability of QTL at a locus given the phenotype values to the probability of there being no QTL. Like all other statistical tests, this begs the question of statistical significance. At what point does

<table>
<thead>
<tr>
<th>F₂ intercross</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of individuals</td>
</tr>
<tr>
<td>Total markers typed (including X)</td>
</tr>
<tr>
<td>Total markers typed (excluding X)</td>
</tr>
<tr>
<td>No. of autosomes</td>
</tr>
<tr>
<td>% Genotyped</td>
</tr>
<tr>
<td>Genotypes (%)</td>
</tr>
</tbody>
</table>

Table 5.4 Summary of the F₂ genotyping data post QC.

Figure 5.5 Missing genotype information for the F₂ intercross. Regions of missing genotype information may show spurious linkage but there are no such regions of missing information in current F₂ intercross.
5.2. Single QTL analysis

the LOD score become sufficiently large so that result is unlikely to be due to random chance? For one locus, a LOD score of more than 1.0 will be rarely seen but since LOD scores are calculated for each genotype across the genome, because of multiple tests a LOD score of $\geq 2.5$ will be seen somewhere in the genome in 10% of experiments [Broman and Sen, 2009]. The genome-wide distribution of LOD scores depends on a number of factors including the type of cross, the phenotype data, the size of the genome and the number of typed markers. This genome-wide distribution can be calculated using permutation tests [Churchill and Doerge, 1994]. In permutation test, the phenotype values are shuffled between different genotyped individuals. We determine the distribution of LOD scores assuming that there is no QTL effect and this distribution is unique to the experiment. We can then define LOD threshold as 95th percentile of this distribution and a genome-scan-adjusted $p$-value can be reported as the proportion of the distribution that exceeds the observed LOD score. The result of the permutation test will be slightly different for each phenotype but given current dataset the permutation test showed that there is less than 5% probability of encountering a LOD score of greater than 3.9 (figure 5.6) anywhere in the genome.

5.2.2 Standard interval mapping for individual phenotypes

Statistical methods for QTL mapping were discussed in section 1.3.2. For single QTL mapping interval mapping is usually the preferred choice, especially since for our data there are no regions of low marker information. Single QTL mapping was performed in R/qtl using the scanone function. Genome wide maximum LOD scores were obtained and compared with LOD threshold scores obtained from the permutation test. Genome-scale-adjusted $p$ values were also obtained for peak LOD (LOD_{peak}) results. The results for cardiovascular traits at baseline are reported in table 5.5 and results for phenotypes during reperfusion reported in table 5.6. Heart rate and maximum hyperaemic CF in response to one minute global ischaemia (figure 2.3) were not mapped to the genome and hence are not reported. Genome-wide linkage plots illustrate linkage for BP (figure 5.7), cardiac mass (figure 5.8), CF (figure 5.9), LV $dP/dt_{max}$ and LV $dP/dt_{min}$ (figure 5.10).
5.2. Single QTL analysis

Figure 5.6 Distribution of LOD scores from 1000 permutation test given the F$_2$ intercross dataset showing that a LOD score of 3.9 will only be exceeded 5% by chance.

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>LOD$_{\text{threshold}}$</th>
<th>LOD$_{\text{peak}}$</th>
<th>Chromosome</th>
<th>Position (cM)</th>
<th>P-value$_{\text{genome-wide}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiac mass</td>
<td>3.71</td>
<td>5.7</td>
<td>3</td>
<td>3</td>
<td>0.001</td>
</tr>
<tr>
<td>Systolic BP</td>
<td>3.98</td>
<td>4.7</td>
<td>2</td>
<td>50.8</td>
<td>0.009</td>
</tr>
<tr>
<td>Diastolic BP</td>
<td>3.86</td>
<td>3.94</td>
<td>2</td>
<td>50.8</td>
<td>0.04</td>
</tr>
<tr>
<td>Mean BP</td>
<td>3.94</td>
<td>4.28</td>
<td>2</td>
<td>50.8</td>
<td>0.02</td>
</tr>
<tr>
<td>Pulse pressure</td>
<td>3.92</td>
<td>5.01</td>
<td>2</td>
<td>24.6</td>
<td>0.007</td>
</tr>
<tr>
<td>CF $c_{\text{cyclic}}$</td>
<td>3.94</td>
<td>4.98</td>
<td>2</td>
<td>86.1</td>
<td>0.005</td>
</tr>
<tr>
<td>LV $dP/dt_{\text{max}}$</td>
<td>3.87</td>
<td>7.04</td>
<td>3</td>
<td>45.8</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>LV $dP/dt_{\text{min}}$</td>
<td>3.87</td>
<td>9.3</td>
<td>3</td>
<td>45.8</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>CFR</td>
<td>3.81</td>
<td>4.07</td>
<td>8</td>
<td>56.9</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Table 5.5 Single-QTL linkage analysis for cardiovascular traits at baseline. LOD$_{\text{threshold}}$ is the LOD threshold obtained for individual phenotype by use of 1000 permutation test, LOD$_{\text{peak}}$ is the LOD score obtained by standard interval mapping with corresponding chromosomes and genetic position for peak marker position. P-value$_{\text{genome-wide}}$ is the proportion of LOD$_{\text{threshold}}$ that would exceed LOD$_{\text{peak}}$ anywhere in the genome.
5.2. Single QTL analysis

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>LOD(_{\text{threshold}})</th>
<th>LOD(_{\text{peak}})</th>
<th>Chromosome</th>
<th>Position (cM)</th>
<th>P-value(_{\text{genome-wide}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>CF cyclic(_{\text{max}})</td>
<td>3.86</td>
<td>3.36</td>
<td>2</td>
<td>86.1</td>
<td>&gt;0.05</td>
</tr>
<tr>
<td>CF cyclic(_{\text{mean}})</td>
<td>3.86</td>
<td>3.75</td>
<td>2</td>
<td>78</td>
<td>&gt;0.05</td>
</tr>
<tr>
<td>CF(_{\text{recovered}})</td>
<td>3.94</td>
<td>2.84</td>
<td>3</td>
<td>114</td>
<td>&gt;0.05</td>
</tr>
<tr>
<td>LV (dP/dt)(_{\text{max}})</td>
<td>3.87</td>
<td>5.16</td>
<td>3</td>
<td>66.3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>LV (dP/dt)(_{\text{min}})</td>
<td>3.87</td>
<td>7.3</td>
<td>3</td>
<td>45.8</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

**Table 5.6** Single-QTL linkage analysis for cardiovascular traits during reperfusion. CF\(_{\text{recovered}}\) is the ratio of CF during reperfusion to CF at baseline.

**Figure 5.7** Genome-wide LOD plot for systolic, diastolic and mean pressure showing a linkage peak for all phenotypes on chromosome 2. Vertical bars on the x-axis are genotyped markers on each chromosome.

**Figure 5.8** Genome-wide LOD for cardiac mass showing linkage peak at the start of chromosome 3 and pulse pressure linkage peak at the previously linked BP locus on chromosome 2. Vertical bars on the x-axis are genotyped markers on each chromosome.
5.2. Single QTL analysis

Figure 5.9 Genome-wide LOD plot for CF cyclic$_{\text{max}}$ and CFR showing separate linkage peaks for phenotypes on chromosome 2 and 8 respectively. Vertical bars on the x-axis are genotyped markers on each chromosome.

Figure 5.10 Genome-wide LOD plot for baseline myocardial relaxation (LV dP/dt$_{\text{min}}$) and myocardial contractility (LV dP/dt$_{\text{max}}$) showing identical linkage peak on chromosome 3. Vertical bars on the x-axis are genotyped markers on each chromosome.
5.2. Single QTL analysis

Figure 5.11 Zoomed view of linkage peaks for systolic BP, diastolic BP and CF on chromosome 2 showing separate linkage regions. Vertical bars on the x-axis are genotyped markers on each chromosome.

5.2.3 QTL effect sizes and interval estimates

The single QTL analysis reveals the evidence for linkage of a number of phenotypes to genomic regions. When a trait is found to be linked to a marker, two further questions need to be answered: where is the QTL in relation to the marker and what is the effect size of the QTL [Broman and Sen, 2009; Rapp, 2000]. The position of the LOD peak gives the most likely position for the QTL but the precise region remains unknown. Usually 1-LOD or 2-LOD drop intervals are used to establish the genomic region wherein the QTL gene is present with a certain degree of confidence. 1-LOD and 2-LOD drops are the regions in which the LOD score is within 1 unit and 2 units of its maximum, respectively. Precisely which LOD drop should be used depends on the QTL effect size, for QTLs of large effect size (≥8%) 1-LOD drop gives ~95% confidence interval for QTL location [Rapp, 2000]. 1.8-LOD drop is suggested for F$_2$ by other experts [Broman and Sen, 2009] which is the measure that was have adopted for reporting LOD intervals in this study.

QTL effects are estimated from the difference in phenotype values between different genotypes. Each marker in the F$_2$ intercross has three possible genotypes: AA, AB or BB. The QTL may have an additive effect (where the combined of allelic effect is equal to sum of individual alleles) or a dominance effect (where the effect of one allele masks the effect of the non-dominant allele). Additive effects are calculated by $a = \frac{(\mu_{BB} - \mu_{AA})}{2}$
5.2. Single QTL analysis

and dominance effects are calculated by \( d = \mu_{AB} - \left( \frac{\mu_{BB} + \mu_{AA}}{2} \right) \) [Broman and Sen, 2009].

It is important to mention that there is, however, a selection bias in that only QTLs of large effect sizes are detected in mapping experiments. QTLs of small effects are almost never detected as few experiments have the power to do so.

We will now proceed to estimate single-QTL effects although the more global picture of multiple-QTLs effects will be explored later. QTL effect sizes were calculated in R/qtl using the `makeqtl` and `fitqtl` functions. The `makeqtl` function takes the data and specified chromosome numbers and positions for the QTL and pulls out the genotype probabilities or imputed genotypes for the QTL. The output is then used by `fitqtl` function which models the phenotype \( y \) by QTL \( Q \) using the formula, \( y \sim Q \) where \( Q \) is sectioned into AA, AB and BB genotypes. The effect sizes for the QTLs are reported in table 5.7. As can be seen from the data, most of the QTLs have large effect sizes. The LOD intervals for location of QTL are reported in table 5.8 using a conservative 1.8-LOD drop estimate.

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>QTL effect size(%)</th>
<th>Phenotype effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiac mass</td>
<td>14.8</td>
<td>0.17 ± 0.03 gram</td>
</tr>
<tr>
<td>Systolic BP</td>
<td>12.6</td>
<td>10.9 ± 2.9 mmHg</td>
</tr>
<tr>
<td>Diastolic BP</td>
<td>10.6</td>
<td>5.7 ± 1.6 mmHg</td>
</tr>
<tr>
<td>Mean BP</td>
<td>11.5</td>
<td>7.8 ± 1.9 mmHg</td>
</tr>
<tr>
<td>Pulse pressure</td>
<td>13.3</td>
<td>5.8 ± 1.2 mmHg</td>
</tr>
<tr>
<td>CF cyclic(_{max})</td>
<td>12.8</td>
<td>1.4 ± 0.3 ml/gram/min</td>
</tr>
<tr>
<td>LV dP/dt(_{max})</td>
<td>17.2</td>
<td>369 ± 75 mmHg/s</td>
</tr>
<tr>
<td>LV dP/dt(_{min})</td>
<td>21.8</td>
<td>310 ± 54 mmHg/s</td>
</tr>
<tr>
<td>CFR</td>
<td>10.5</td>
<td>0.10 ± 0.03</td>
</tr>
</tbody>
</table>

**Table 5.7** QTL effect size reported as % of the phenotype that is explained by the QTL with the corresponding effects in units of phenotype. Cardiac mass is indexed to bodyweight and CFR is dimensionless (ratio of peak hyperaemic flow to baseline flow).

QTL effect sizes are plotted for mean BP, cardiac mass, CF, CFR, LV dP/dt\(_{max}\) and LV dP/dt\(_{min}\) in figure 5.12. The genotypes are coded as follows: AA and BB are SHR and BN homozygotes at the locus respectively; AB is the heterozygote genotype.
### 5.2. Single QTL analysis

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Effect Plots</th>
</tr>
</thead>
</table>
| **(a)** Mean BP | ![Effect Plot for Mean BP](image1)
| **(b)** Cardiac mass | ![Effect Plot for Cardiac mass](image2)
| **(c)** CFR | ![Effect Plot for CFR](image3)
| **(d)** CF cyclicmax | ![Effect Plot for CF cyclicmax](image4)
| **(e)** LV dP/dt\text{max} | ![Effect Plot for LV dP/dt\text{max}](image5)
| **(f)** LV dP/dt\text{min} | ![Effect Plot for LV dP/dt\text{min}](image6)

**Figure 5.12** Effect plots for phenotypes given genotype at the peak of linkage analysis for single QTLs listed in table 5.8. SHR genotype is represented by AA, BN by BB and heterozygotes by AB. The peak marker ID is given for each phenotype.
5.2. Single QTL analysis

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Chromosome</th>
<th>Physical start (bp)</th>
<th>Physical end (bp)</th>
<th>Region size (cM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiac mass</td>
<td>3</td>
<td>3470823</td>
<td>24810286</td>
<td>13</td>
</tr>
<tr>
<td>Systolic BP</td>
<td>2</td>
<td>53524570</td>
<td>114897018</td>
<td>25</td>
</tr>
<tr>
<td>Diastolic BP</td>
<td>2</td>
<td>53524570</td>
<td>120856112</td>
<td>25</td>
</tr>
<tr>
<td>Mean BP</td>
<td>2</td>
<td>53524570</td>
<td>120856112</td>
<td>25</td>
</tr>
<tr>
<td>Pulse pressure</td>
<td>2</td>
<td>42514359</td>
<td>78376494</td>
<td>20</td>
</tr>
<tr>
<td>CF cyclic_{max}</td>
<td>2</td>
<td>171008806</td>
<td>217873407</td>
<td>22</td>
</tr>
<tr>
<td>LV dP/dt_{max}</td>
<td>3</td>
<td>59070301</td>
<td>68419847</td>
<td>6</td>
</tr>
<tr>
<td>LV dP/dt_{min}</td>
<td>3</td>
<td>59070301</td>
<td>68419847</td>
<td>6</td>
</tr>
<tr>
<td>CFR</td>
<td>8</td>
<td>68783831</td>
<td>113964559</td>
<td>32</td>
</tr>
</tbody>
</table>

Table 5.8 1.8 LOD-drop intervals for significant QTLs, start and end positions are the physical positions in the genome extending to the nearest genotyping markers for the interval, whereas region size is the size of 1.8-LOD drop region in cM.

![Graphs](a) (b) (c) (d)

Figure 5.13 QTL significance thresholds remain unaffected after taking into account clinically important covariates; (a) effect of BP on cardiac mass QTL; (b) effect of BP on CFR QTL; (c) effect of BP on LV dP/dt_{min}; (d) effect of LV dP/dt_{min} on CF QTL)
5.2.4 Covariate effects

Analysis of the physiological data (section 4.2) from the $F_2$ intercross revealed significant correlation between a number of cardiovascular phenotypes: CF and LV $dP/dt_{min}$ as well as cardiac mass and LV $dP/dt_{min}$. In addition, previous literature [Ruskoaho, 1984, Verdecchia et al., 1990, Hoenig et al., 2008] has shown significant correlation between BP, cardiac hypertrophy and CFR. Inclusion of a covariate with a large effect on the phenotype may result in reduction of the residual variation and enhance the ability to detect the QTL effect.

There was no evidence that BP significantly influenced the QTL effect for CF, CFR or LV $dP/dt_{min}$. Additionally, no change in QTL effect size for CFR was observed after taking effect of cardiac mass into account, figure 5.13.

5.3 Two QTL analysis

A single QTL is more tractable for candidate gene studies but phenotypes are produced by multiple loci acting additively and interactively. To achieve a greater understanding of genetic control of complex traits a more fitting model than single-QTL analysis is to consider the possibility of multiple QTLs. The next step after single-QTL scan is to perform two-QTL scans which has the following advantages [Broman and Sen, 2009]

- If a QTL of large effect is taken into account, residual variation is reduced and this may increase the power to detect a second QTL.
- Epistasis between QTL may only be assessed by considering a multiple-QTL model.

The principle of the statistical test used to perform a two-QTL scan is as follows: For each marker “i” with a potential QTL “x” in the genome, a genome scan is performed for the phenotype under the assumption that a second locus “j” also has a potential QTL “y”. Let us assume that there are hundred genetic markers, 1....100, across the genome. A two-QTL scan will first perform linkage test assuming a QTL at marker 1 and marker 2, followed by linkage test at marker 1 and marker 3 followed by marker 4 and so on till the
last test is performed assuming markers position 1 and 100. This process is then repeated assuming a QTL at marker 2 in the list of markers and then iterated for all 100 markers, one phenotype at a time. Fundamentally, a two-QTL scan asks the question, "Are there QTLs at positions A and B in the genome?" A related question is if a two-QTL model sufficiently improves the model fit over a single-QTL model. This is answered by use of two statistical measures, \( \text{LOD}_{fv1} \) and \( \text{LOD}_{av1} \) [Broman and Sen, 2009]. \( \text{LOD}_{fv1} \) is the \( \log_{10} \) likelihood ratio comparing the full model with QTL at positions i and j to the single QTL model with a QTL on either i or j. \( \text{LOD}_{av1} \) is the \( \log_{10} \) likelihood ratio comparing the additive (i.e. that there is no epistasis) model with QTL at positions i and j to the single QTL model with a QTL on either i or j. For an intercross suggested \( \text{LOD}_{fv1} \) cutoff is 7 and \( \text{LOD}_{av1} \) cutoff is 3.3 based on permutation tests. R/qtl provides a simple function \texttt{scantwo} to undertake this analysis in animal crosses.

Figure 5.14 shows two-QTL scan for cardiac mass. The upper left half of the panel shows \( \text{LOD}_{fv1} \) scores for chromosomes 3, 8, 17 and 19. Right lower half of the panel shows \( \text{LOD}_{av1} \) for the same chromosomes. The scan is suggestive for existence of two-QTL models significantly better than the single cardiac mass QTL on chr 3. These two-QTL models on chr 3: chr 8, chr 3: chr 17 and chr 3: chr 19 all show statistically significant improvement in LOD scores compared with the single-QTL model (LOD\(_{av1}=4.5\)). The rest of the phenotypes are shown in figure 5.15 using same layout for \( \text{LOD}_{fv1} \) and \( \text{LOD}_{av1} \).

For SBP chr 1:2 QTL model has a significant \( \text{LOD}_{av1} \) of 3.3, figure 5.15a. For DBP no model was found with statistically significant improvement over the single-QTL model, figure 5.15b. The additive two-QTL model pair on chr 1: chr 9 for pulse pressure, figure 5.15c was also significantly better than the single QTL model with a LOD\(_{av1}\) score of 3.3. Results for CFR revealed evidence of significant epistatic interaction between chr 1 and chr 12 with \( \text{LOD}_{fv1} = 7.14 \) but there was no additive model that showed any improvement over a single QTL scan. There were no pairs of QTLs for LV dP/dt\(_{min}\) that offered statistically significant improvement over the single-QTL model but there was suggestive evidence for additive loci on chr 3 and chr 8 with \( \text{LOD}_{av1} = 2.6 \). Two-qtl scan for CF cyclic\(_{max}\) also found a pair of qtls on chr 1 and chr 2 with significant additive...
5.4 Multiple QTL mapping

In single-QTL analysis the whole genome is scanned with a probability test of QTL presence at a specific location, one at a time. It therefore asks a binary question, “Is there
5.4. Multiple QTL mapping

Figure 5.15 Two-QTL scan results for cardiac phenotypes showing suggestive improvement in using two-qtl model over single-QTL scan. In each panel significant interactions between loci are highlighted by colour scale according to bar on the right. LOD_{lv1} is in the upper left half of each panel whereas LOD_{av1} is in the lower right half of each panel.
5.4. Multiple QTL mapping

The answers provided by single-QTL mapping are useful in that a single underlying genetic polymorphism is more traceable and amenable to functional studies. Even though a multiple-QTL is not as tractable from a gene discovery perspective as a single-QTL approach, it is essential to our understanding of “global” genomic architecture of complex traits.

For each cardiovascular trait under study, the approach taken to discovering and modelling a multiple-QTL model was as follows:

- Selecting a number of regular markers - termed *cofactors* - across the genome. The term cofactor was used in the paper by Arends et al. [Arends et al., 2010].
- All the cofactors are analyzed for their effect on the trait and the least informative cofactor is dropped, the process is then repeated for the remaining markers and represents a backward elimination process.
- The process is repeated till only a limited number of cofactors remain that have a significant effect on the trait.
- Cofactors are evaluated in groups for additive or epistatic effects.
- The most informative cofactors are selected and incorporated in a multiple-QTL model that attempts to explain the phenotypic trait variance.

Multiple-QTL modelling was carried out in R/qtl-MQM [Arends et al., 2010]. The first step was to place markers across the genome for use in multiple-QTL mapping. 150 genome-wide markers were placed using `mqmautocofactor` function in R as shown in figure 5.16.

After selection of uniformly spread cofactors, a search for the best-fitting multiple-QTL model was undertaken using the `mqmscan` function in R/qtl. Most significant cofactors were selected and incorporated in the multiple QTL model (figure 5.17). The interactions between these cofactors were explored to see if the effects are additive or epistatic. Additive interactions are due to fixed-dose effects of each allele whereas in epistatic interactions, the effects of alleles at one locus are modified by alleles at another...
5.4. Multiple QTL mapping

Figure 5.16 Map of cofactors plotted for multiple QTL mapping across the genome in the F$_2$ intercross

Figure 5.17 Genomic locations of the four most significant cofactors influencing cardiac mass.
5.4. Multiple QTL mapping

locus. Multiple-QTL analysis for cardiac mass showed that the proximal locus on chr 3, which was the most significant locus on single-QTL scan, showed additive interaction with locus on chr 8 (figure 5.18) and chr 17 (figure 5.19). In addition there was evidence for an additional locus on chr 19 which showed an epistatic interaction with the locus on chr 3. These results are displayed in figure 5.17 and summarized in table 5.9.

**Figure 5.18** Assessment of interactions between significant loci for cardiac mass (adjusted for body weight, htwt.bdwt) on chr 3 (marker 17358758) and chr 8 (marker 113964559). The graph shows significant additive interaction.

The overall multiple-QTL for cardiac mass model was cardiac mass $\sim Q_2 + Q_3 + Q_1 : Q_4$, where $Q_1 : Q_4$ indicates an epistatic interaction between the loci Q1 and Q4 had LOD = 16 and explained 36% of the variance in cardiac mass. This is clearly a much improved model that explains greater proportion of variance in cardiac mass when compared with the single-QTL model but the number of QTLs involved make the model less tractable from point of view of genetic manipulation.

Multiple-QTL model search for systolic, mean and diastolic BP revealed addition significant cofactors in addition to the previously described locus single-QTL (figure 5.7) on chr 2. For SBP six cofactors with significant effect were discovered, located on chr 1, 2, 3, 7, 17 and 19. The effects of the cofactors on chr 1, 2, 3 and 7 were additive and the locus on chr19 had an epistatic effect. There was also epistatic interaction between loci on chr 3 and 17 as found in the two-QTL scan. Mutiple-QTL model for SBP incorporating both
Figure 5.19 Assessment of interactions between significant loci for cardiac mass on chr 3 (marker 17358758) and chr 17 (marker 40120016) indicates additive action of alleles at these loci.

Figure 5.20 Genome-wide search for significant loci influencing SBP identify significant loci on chr 1, 2, 3, 7, 17 and 19.
additive and epistatic terms, \( SBP \sim Q1 + Q2 + Q3 + Q6 + Q4 + Q5 + Q4 : Q5 \), explained 45% of the variance in SBP (table 5.9) and had a LOD score of 21. For DBP, significant cofactors were present on chr 1, 2, 3, 7, 9 and 17 (figure 5.21) with the multiple-QTL model, \( DBP \sim Q2 + Q5 + Q6 + Q3 : Q4 : Q1 \), having a LOD score of 22.6, explaining 48% variation in phenotype (table 5.9). The dominant epistatic interaction between DBP cofactors on chr 2 and 7 is illustrated in figure 5.22. No additional significant cofactors were discovered using this statistical approach for LV dP/dt_{max} and LV dP/dt_{min} beyond the original single locus on chr 3 (figure 5.10). This is in contrast to the results obtained from two-QTL scans which showed an additional locus on chr 8 influencing LV dP/dt_{max} and LV dP/dt_{min}. This discrepancy may be due to more stringent statistical cut-off employed in multiple-QTL strategy or due to a difference in statistical approach.

**Figure 5.21** Genomic positions of six loci associated with DBP identified using multiple-qtl approach.

A search for additional significant CF cyclic\(_{\text{max}}\) loci revealed the presence of additively acting loci on chr 1, 2, 4 and 19 (figures 5.23 and 5.24, table 5.9). Epistatic interaction was observed between one pair of loci, on chr 1 and 19. The overall model for CF cyclic\(_{\text{max}},\)
5.4. Multiple QTL mapping

Figure 5.22 Evidence of epistatic interaction for DBP loci on chr 2 and 7. Individuals with BB phenotype at marker 72928519 on chr 7 shows no difference in phenotype between AA and BB phenotypes at marker 75860058 on chr 2. The alleles at marker 75860058 (chr 7) only manifest in the presence of AA allele at marker 72928519 (chr 2).

\[ CF_{cyclic_{\text{max}}} \sim Q2 + Q3 + Q1 + Q4 + Q1 : Q4 \] had a LOD score of 14, explaining 31% of the variation in \( CF_{cyclic_{\text{max}}} \) (table 5.9). CFR was found to have two additional significant loci on chr 1 and 12 (figure 5.25). The inclusion of these loci resulted in the multiple-QTL model explaining 29% of the variation in CFR and a LOD score =12.5 (compared with the single-QTL LOD score of 4.07). Finally, a multiple-QTL model was obtained for pulse pressure with loci on chr 1, 2, 9 and 19; LOD score for the overall model was 14, explaining 33% of the variation in pulse pressure. All these results show very significant improvements obtained by searching for and using multiple loci to explain the phenotypic variance indicating that single-qtl models tend to underestimate the overall heritability of a trait. Table 5.9 provides a summary of the multiple QTL models including the number of QTLs for each trait, the chromosomal locations of these QTLs, LOD score for the total multiple-QTL model and % variation of the phenotype explained by the multiple-QTL model.
Figure 5.23 Additively acting loci for CF cyclic\textsubscript{max} on chr 1 (marker 229399235) and 2 (marker 203879555). For both loci the BN genotype, represented by BB has higher CF compared to the SHR genotype (AA).

Figure 5.24 Genomic locations of significant loci for CF cyclic\textsubscript{max} identified using multiple QTL approach.
Figure 5.25 Significant cofactors for CFR on chr 1 and 12 in addition to the previously
described locus on chr 8.

Table 5.9 Multiple-QTL modelling for cardiovascular traits; the final results incorporates
significant loci from two-QTL genome scans as well as multiple-QTL mapping.
5.5 Discussion

A number of cardiovascular traits were mapped to the genome using linkage analysis. SBP, DBP, PP and cardiac mass are traits that have been mapped to the genome in earlier studies [Rapp, 2000, Petretto et al., 2008] and some of the QTLs for these traits have been replicated in current study. Novel QTLs for CF, CFR, LV $dP/dt_{\text{max}}$ and LV $dP/dt_{\text{min}}$ have been discovered using the approach described. In addition multiple QTL models for each phenotype were demonstrated showing that genetic variation at multiple loci contribute to the phenotype under study.

Single QTL mapping for cardiac mass shows the presence of a major QTL on short arm of chr 3 extending from 3.5 Mbp to 25 Mbp region, with an effect size of $\sim 15\%$. This is a replicated QTL for cardiac mass with a large effect size in rat species, table 5.10. Siegel et al. mapped LV mass to this locus using nutritional salt-loading in F$_2$ intercross derived from salt-sensitive (SS) Dahl and SHR [Siegel et al., 2003]. This region was also reported to have suggestive linkage for cardiac mass in a report by Cicila et. al [Cicila et al., 1999] who used a backcross between SS Dahl and salt-resistant (SR) Dahl to map cardiac hypertrophy locus. They also observed that the congenic line with the SR region on a background of SS genome had lower cardiac mass. This region on chr 3 was also the most consistent linkage region for cardiac mass reported by Inomata et. al [Inomata et al., 2005]. They undertook linkage analysis in two F$_2$ intercross populations, one derived from SHRSP and WKY fed a high salt diet and another derived from SHR and WKY fed a normal salt-diet. The data presented here confirm these earlier findings by replicating a BP-independent region for determining cardiac mass on chr 3.

<table>
<thead>
<tr>
<th>QTL name</th>
<th>LOD</th>
<th>Age</th>
<th>Diet</th>
<th>Region (on Chr 3)</th>
<th>Cross</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cm10</td>
<td>7.3</td>
<td>15 weeks</td>
<td>high salt</td>
<td>1-26510302</td>
<td>SS Dahl/SHR</td>
</tr>
<tr>
<td>Cm13</td>
<td>2.1</td>
<td>12 weeks</td>
<td>high salt</td>
<td>21201203-35237920</td>
<td>SS Dahl/SR Dahl</td>
</tr>
<tr>
<td>Cm46</td>
<td>5.4</td>
<td>7 months</td>
<td>high salt</td>
<td>6373454 - 26674018</td>
<td>SHRSP/WKY</td>
</tr>
<tr>
<td>Cm48</td>
<td>4.0</td>
<td>15 weeks</td>
<td>normal salt</td>
<td>6373454 - 26674018</td>
<td>SHR/WKY</td>
</tr>
</tbody>
</table>

Table 5.10 Previously described loci for cardiac mass coincident with the locus on chromosome 3 identified in current study using BN/SHR F$_2$ intercross.
Nearly all of rat chr 3 is syntenic to mouse chr 2 which does not allow to narrow of the QTL region. Search for syntenic regions harbouring cardiac mass QTLs in the mouse identified two loci on mouse chr 2, Hrtq1 and Hrtq2 [Rocha et al., 2004]. This adds further supportive evidence without narrowing the QTL region. Rat chr 2 is syntenic to regions on human chromosomes 2, 9, 11, 15 and 20. There have been few reported QTL mapping studies for cardiac mass in humans but ventricular mass (LVM) was mapped to human chr 12 in Dominican families [Wang et al., 2009a]. Genome-wide association studies in recent years have been performed to understand common variants underlying complex traits in humans. Vasan et al. [Vasan et al., 2009] report a meta-analysis of common genetic variants associated with echocardiographic parameters, including LVM reported by genome-wide association studies (GWAS). It is interesting to note that one of the common variants associated with LVM, rs8031633, located on human chr 15 in a region syntenic to cardiac mass QTL region on rat chr 3. However rs8031633 was excluded from further analysis in the study due to low minor allele frequency (MAF).

![Figure 5.26](image_url)

**Figure 5.26** Rat chr 3 harbouring the LVM QTL (shaded grey) coloured according to syntenic regions in the mouse (left) and human genomes (right). The coloured segments of rat chr 3 represent syntenic segments from different chromosomes of the mouse genome on the left and human genome on the right. The QTL containing region in rat extending (grey shaded area) from 3 Mbp - 25 Mbp is syntenic to mouse chr 2 (dark green) and human chr 9 (yellow), 2, 11 and 15 (coded together as green).

As far as single QTL mapping is concerned SBP, DBP and MBP all map to a single 58 Mbp interval starting from position 53 Mbp to position 121Mbp region on chr 2 with an effect size of $\sim 10\% - 20\%$. This region has recognized association with systolic, mean and
diastolic BP in previous studies in rats. Dubay et al. mapped pulse pressure to this locus in an F$_2$ intercross between Lyon hypertensive (LH) and Lyon normotensive (LH) rat strains [Dubay et al., 1993]. In an F$_2$ intercross between SHRSP and WKY this region was also linked to BP [Clark et al., 1996] and a congenic strain with the WKY region introgressed onto the SHRSP genome was associated with lower BP [McBride et al., 2003]. The region was significantly linked to BP in female rats only [Moreno et al., 2003] and independent of salt-loadings [Aneas et al., 2009].

Rat chr 2 region encoding the BP QTL identified in the current study is syntenic to regions on mouse chromosomes 3, 13 and 15 as well as human chromosomes 1, 3, 4, 5 and 8. Mouse chr 3 does not harbour any known QTLs for BP but genes involved in the renin-angiotensin-aldosterone axis are located in this region. Mouse chr 15 houses a QTL for BP, termed Bpq6 whereas no BP QTLs are known to be located on mouse chr 13. There is concordance between the rat BP QTL on chr 2, mouse BP QTL on chr 15 (32Mbp - 58Mbp) and human BP QTL on chr 8 (BP49H, 81 Mbp - 106 Mbp) [Rice et al., 2000] as these regions are syntenic and show linkage to BP. Assuming that variation in the same gene underlies BP variation, this comparison helps narrow down the rat QTL locus from 58 Mbp to $\sim$8 Mbp, figure 5.27.

![Figure 5.27](image_url)

**Figure 5.27** Rat chr 2 colour coded according to mouse syntenic regions (left) and human chr 8 (right). The coloured segments of rat chr 2 represent syntenic segments from different chromosomes of the mouse genome on the left and human genome on the right. The regions on mouse chr 15 and human chr 8 (shaded as grey) also exhibit linkage to BP. The shared syntenic region between the three species containing BP are shaded as transparent red and likeliest to contain the underlying gene.

LV dP/dt$_{\text{max}}$ has been linked to human chromosome 11 in a study reported by Arnett
et. al as part of the Hypertension Genetic Epidemiology Network [Arnett et al., 2001]. The important observation here is that the reported region in humans on chr 11 is syntenic to the region where I have shown linkage to LV dP/dt\textsubscript{max} on rat chr 2. This aids in narrowing down the region for LV dP/dt\textsubscript{max}, figure 5.28.

![Legend: Rat chr 3 colour coded according to syntenic human regions. The region containing LV dp/dt\textsubscript{max} locus is shaded in grey. The dark green region is syntenic to human chr 2 which does not have any linkage to myocardial contractility and the light green region at bottom is syntenic to human chr 11 which harbours LV dp/dt\textsubscript{max} locus in humans.]

Figure 5.28  Rat chr 3 colour coded according to syntenic human regions. The region containing LV dp/dt\textsubscript{max} locus is shaded in grey. The dark green region is syntenic to human chr 2 which does not have any linkage to myocardial contractility and the light green region at bottom is syntenic to human chr 11 which harbours LV dp/dt\textsubscript{max} locus in humans.

To date the CF, CFR and LV dP/dt\textsubscript{min} loci discovered in the study described here are the first QTLs discovered for these cardiac traits. There are a number of genes implicated in these phenotypes, primarily utilizing knockout mice models but there is no data from unbiased genomic approaches. Therefore it is not possible to utilize approaches such as syntenic or haplotype mapping to narrow the rat QTL regions for these phenotypes using available data at present time.

Furthermore this is the first systematic search for multiple-QTL models underlying cardiovascular traits using MQM approach first published by Jansen and Stam [Jansen and Stam, 1994]. The issue of heritability that could be ascribed to genetic variation is an ever-present theme in reports from current GWAS publications but search for QTLs adopting a multiple-QTL mapping approach indicates that in inbred rat strains this approach results in discovery of variants that together describe a significant proportion of heritability, as shown in table 5.9. These results confirm the expectation that multiple genetic loci contribute to trait variability by additive or epistatic mechanisms.
Chapter 6

Gene expression

6.1 Microarray quality metrics

Gene expression analysis was carried out in the non-ischaemic LV samples obtained from the F₂ intercross. For the studies here gene expression estimated were obtained by scanning microarrays to which labelled, fragmented cDNA was hybridized. Hybridization intensity is the measure of gene expression and the basis on which usual analyses such as differential expression, gene clustering and pathway analysis are based. Therefore quality assessment of hybridization intensity is crucially important. A range of procedures, together termed “pre-processing” are undertaken to reduce systematic variation in intensities. The main components of microarray pre-processing are background correction, between-array normalization and probeset-level summarization. These measures have been addressed in subsection 1.4 but will be briefly revisited.

Figure 6.1 is a heatmap of distances between arrays. The “distance” here is the difference between array intensities with greater distance for greater intensity difference. A total of 115 samples obtained from non-ischaemic left ventricles were hybridized to microarrays. The colour scale is chosen to cover the range of distances encountered in the dataset. Patterns in this plot can indicate clustering of the arrays either because of intended biological or unintended experimental factors (batch effects). The distance, “d_{ab}” between two arrays a and b is computed as the mean absolute difference (L1-distance) between the data of the arrays (using the data from all probes without filtering). In
6.1. Microarray quality metrics

Figure 6.1 Heatmap of the distances between arrays. Distance here is the difference between signal intensities obtained from the microarrays.

Formula, \( d_{ab} = \text{mean} | M_{ai} - M_{bi} | \), where \( M_{ai} \) is the value of the i-th probe on the a-th array. Outlier detection was performed by identifying arrays for which the sum of the distances relative to all other arrays was exceptionally large. One such array was detected and removed from further analysis.

Principal component analysis (PCA) is a dimension reduction and visualisation technique that is used here to project the multivariate data vector of each array into a two-dimensional plot, such that the spatial arrangement of the points on the plot reflect the overall data (dis)similarity between the arrays. This plot is used to view if the arrays cluster, and whether this is according to an intended experimental factor or unintended causes such as batch effects. PCA plot does not reveal any sample clustering in the expression dataset.

Figure 6.3 shows summaries of signal intensity distributions of the arrays. Each box
6.1. Microarray quality metrics

**Figure 6.2** Principal component analysis: A scatterplot of the arrays along the first two principal components PC1 and PC2. The scatterplot does not show any evidence for array clustering which may be intentional (two or more experimental groups) or unintentional (batch effects).

corresponds to one array. Typically, one expects the boxes to have similar positions and widths. If the distribution of an array is very different from the others, this may indicate an experimental problem. Outlier detection was performed by computing the Kolmogorov-Smirnov statistic between each array’s distribution and the distribution of the pooled data. Figure 6.4 shows smoothed histograms of the data. Typically, the distributions of the arrays should have similar shapes and ranges. Arrays whose distributions are very different from the others should be considered for possible problems. Various features of the distributions can be indicative of quality related phenomena. For instance, high levels of background will shift an array’s distribution to the right. Lack of signal diminishes its right tail. A bulge at the upper end of the intensity range often indicates signal saturation. Finally figure 6.5 shows MA plots which were previously discussed in subsection 1.4. The mass of the distribution in an MA plot to be concentrated along the M = 0 axis, and there should be no trend in M as a function of A. If there is a trend in the lower range of A, this often indicates that the arrays have different background intensities; this may be addressed by background correction.
Figure 6.3 Boxplots representing summaries of the signal intensity distributions of the arrays after background correction and normalization. Signal intensities are plotted along the x-axis for arrays plotted on the y-axis. The asterisk-marked arrays were initial outliers and improved after correction and normalization.
6.2 Data analysis

Figure 6.4 Density estimates (smoothed histograms) of the data. No sample shows significant difference from the overall distribution.

Figure 6.5 MA plots: There is no evidence of a significant trend in M as a function of A and distribution is concentrated along zero axis represented by a red line. Outlier detection was performed by computing Hoeffding’s statistic $D_a$ on the joint distribution of A and M for each array. Shown are the 4 arrays with the highest value of $D_a$ (top row), and the 4 arrays with the lowest value (bottom row). The value of $D_a$ is shown in the panel headings. There are no outliers based on this analysis.

6.2 Data analysis

eQTLs and QTTs

QTL mapping, as discussed in preceding chapters links the phenotypic trait to genomic region(s) encompassing a large number of genes. Gene expression itself is an “intermediate phenotype” between DNA sequence variation and organism level-phenotype. Two
complementary approaches can be taken to study the relationship between sequence variation, gene expression, and organismal-level phenotypes, expression QTL (eQTL) profiling and quantitative trait transcript (QTT) analysis. Using the eQTL approach, gene expression is treated as a continuous trait and linkage analysis is performed to discover genomic determinants of transcript abundance \cite{Schadt2003, Schadt2005}. If the transcript maps to within 5 Mb upstream or 5 Mb downstream of own gene region, it is termed a *cis* eQTL and if it maps to a region outside this region, it is termed a *trans* eQTL. Some genomic regions are associated with variation in levels of many transcripts and are called eQTL “hotspots” \cite{Mackay2009}. Another approach utilizing gene expression as an intermediate trait is the quantitative trait transcript (QTT) approach which involves correlation of transcript abundance with organism-level phenotype \cite{Passador-Gurgel2007}. Usually a number of QTTS correlate with a single organism-level phenotype and this may allow for construction of correlated QTTS as “co-expression networks” \cite{Mackay2009}. QTTS that correlate with a trait and whose corresponding genes lie within the physiological QTL (pQTL) for the trait are paid special attention as they could potentially aid in identifying the underlying causative variant.

**Data analysis: tools**

Each Affymetrix Rat GeneChip 1.0 ST can assess transcripts from \(\sim 27,000\) genes in the rat genome. Each gene is represented on the array by approximately 26 probes spread across the length of the gene, reflecting transcription better than 3’ biased arrays where genes are represented by probes targeting 3’ end of the transcript. For 115 samples in our dataset, there are \(> 3\) million gene-level expression measurements and nearly 80 million expression data points at probe-set level. This large amount of data necessitated gaining of bioinformatic tools and techniques. The data was analyzed in the software platform R \cite{RDevelopmentCoreTeam2011}, using add-on packages, “R/qtl” \cite{Broman2001}, “R/eqtl” \cite{KhaliliLoudet2011}, “R/eqtlM” \cite{Dawson2009}, “affy” \cite{Gautier2004}, “simpleaffy” \cite{Miller2010} and “affyQCRreport” \cite{Parman2007}. A significant part of the analysis required access
6.2. Data analysis

to a computer cluster.

**Filtering out probesets targeting variant sequences**

Affymetrix Rat GeneChip is based on DNA sequence of the reference BN rat genome. The samples used to hybridize to probe were however derived from an F$_2$ intercross between BN and SHR. The hybridization of some genes, derived from the SHR, may be affected if the probeset sequence is different from the sequence of the hybridizing sample derived from parental SHRs. Therefore sequences from the GeneChip annotation files were first examined for variation between BN and SHR and all probesets (n = 120) with variations between BN and SHR were excluded. A custom chip definition file (CDF) was created after exclusion of these probesets.

### 6.2.1 eQTL hotspots

An eQTL hotspot is a region of the genome which determine the expression of a number of genes. The underlying hypothesis is that allelic variation in an upstream “regulator” gene influences the expression of a number of “downstream” genes [Wu et al., 2008]. The number of transcripts it takes to define a hotspot has not been settled [Wu et al., 2008; Breitling et al., 2008]. Wu et al defined a hotspot as a marker to which at least 50 transcripts map but this threshold was criticised as being liberal by Breitling et al. In the absence of any clear cut-off I have opted for a more conservative threshold of 100 transcripts to define an expression hotspot. This reduces the risk of type I errors. Based on this cut-off approximately 20 hotspots listed in table 6.1 were identified. A number of markers are contiguous in this table and therefore it is likely that they represent one eQTL hotspot e.g. the hotspots on chr 3 can be grouped into two, a proximal eQTL hotspot including markers at genomic positions 11503741, 14386227, 22171227 and a distal eQTL hotspot comprising markers at positions 118284271, 120624798, 123044649, 126425831, 129385674, 132255920 and 134855115. Therefore there are approximately seven eQTL hotspots in total in the current dataset (two on chr 3, one on chr 8 and one each on chr 1, 4, 9 and 20) given the relatively conservative threshold (table 6.1). The transcripts
mapping to these hotspots were analyzed using bioinformatic tools discussed in section 1.5 for enrichment in biological pathways. The results of this analysis are discussed below by each eQTL hotspot.

<table>
<thead>
<tr>
<th>eQTL hotspots</th>
<th>Chr</th>
<th>No. of unique mapping transcripts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiac eQTL1</td>
<td>3</td>
<td>1076</td>
</tr>
<tr>
<td>Cardiac eQTL2</td>
<td>3</td>
<td>920</td>
</tr>
<tr>
<td>Cardiac eQTL3</td>
<td>8</td>
<td>600</td>
</tr>
<tr>
<td>Cardiac eQTL4</td>
<td>9</td>
<td>426</td>
</tr>
<tr>
<td>Cardiac eQTL5</td>
<td>5</td>
<td>323</td>
</tr>
<tr>
<td>Cardiac eQTL6</td>
<td>20</td>
<td>184</td>
</tr>
<tr>
<td>Cardiac eQTL 7</td>
<td>1</td>
<td>130</td>
</tr>
</tbody>
</table>

**Table 6.1** eQTL clusters obtained from expression profiling of non-ischaemic LV and the number of transcripts mapping to each cluster.

The transcripts mapping to proximal eQTL hotspot on chr 3 markers at positions 6126796, 11503741, 14386227 and 22171227, which will be referred to as **cardiac eQTL1** since this is a cardiac tissue-specific expression set, are significantly enriched for a number of biological pathways which are defined using standard GO vocabulary where CC stands for cellular component, BP is biological process and MF is molecular function (table 6.2). Approximately 1000 unique transcripts map to cardiac eQTL1 (table 6.1). Taken together the data demonstrates that cardiac eQTL1 is highly enriched for transcripts...
6.2. Data analysis

<table>
<thead>
<tr>
<th>Category</th>
<th>Term</th>
<th>p value (Benjamini)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOTERM CC</td>
<td>GO:0005739 mitochondrion</td>
<td>1.9 e^{-12}</td>
</tr>
<tr>
<td>KEGG PATHWAY</td>
<td>rno04910:Insulin signalling pathway</td>
<td>7.3 e^{-07}</td>
</tr>
<tr>
<td>GOTERM CC</td>
<td>GO:0005759 mitochondrial matrix</td>
<td>1.3 e^{-05}</td>
</tr>
<tr>
<td>GOTERM CC</td>
<td>GO:0031980 mitochondrial lumen</td>
<td>1.3 e^{-05}</td>
</tr>
<tr>
<td>GOTERM BP</td>
<td>GO:0006631 fatty acid metabolic process</td>
<td>2.6 e^{-04}</td>
</tr>
<tr>
<td>GOTERM CC</td>
<td>GO:0044429 mitochondrial part</td>
<td>1.6 e^{-05}</td>
</tr>
<tr>
<td>GOTERM CC</td>
<td>GO:0000267 cell fraction</td>
<td>1.4 e^{-04}</td>
</tr>
<tr>
<td>GOTERM MF</td>
<td>GO:0000166 nucleotide binding</td>
<td>0.004</td>
</tr>
<tr>
<td>KEGG PATHWAY</td>
<td>Valine, leucine and isoleucine degradation</td>
<td>5.7 e^{-04}</td>
</tr>
<tr>
<td>KEGG PATHWAY</td>
<td>Adipocytokine signaling pathway</td>
<td>4.0 e^{-04}</td>
</tr>
<tr>
<td>GOTERM BP</td>
<td>GO:0046320 regulation of fatty acid oxidation</td>
<td>6.6 e^{-04}</td>
</tr>
<tr>
<td>KEGG PATHWAY</td>
<td>HCM</td>
<td>3.9 e^{-03}</td>
</tr>
<tr>
<td>GOTERM MF</td>
<td>GO:0001883 purine nucleoside binding</td>
<td>0.02</td>
</tr>
<tr>
<td>GOTERM MF</td>
<td>GO:0017076 purine nucleotide binding</td>
<td>0.01</td>
</tr>
<tr>
<td>GOTERM MF</td>
<td>GO:0001882 nucleoside binding</td>
<td>0.01</td>
</tr>
<tr>
<td>GOTERM MF</td>
<td>GO:0030554 adenyl nucleotide binding</td>
<td>0.01</td>
</tr>
<tr>
<td>KEGG PATHWAY</td>
<td>ARVC</td>
<td>0.003</td>
</tr>
<tr>
<td>GOTERM MF</td>
<td>GO:0004672 protein kinase activity</td>
<td>0.02</td>
</tr>
<tr>
<td>KEGG PATHWAY</td>
<td>mTOR signaling pathway</td>
<td>0.004</td>
</tr>
</tbody>
</table>

Table 6.2 Biological pathway enrichment analysis for transcripts mapping to cardiac eQTL1, table 6.1, figure 6.6). GO term refers to the controlled vocabulary used by GO consortium to describe biological process (BP), cellular component (CC) and molecular function (MF). KEGG (Kyoto encyclopaedia of genes and genomics) describes the pathways where these transcripts are overrepresented. The final column gives the significance corrected for multiple testing.

related to mitochondrial structure, function and involved in insulin signalling and fatty acid metabolism. This eQTL hotspot co-maps with pQTL for cardiac mass (see figure 5.8) and in this respect it is significant that transcripts mapping to this region are also enriched for biological pathways involved in hypertrophic cardiomyopathy (HCM) which is a genetic disease of the cardiac sarcomere, caused by mutations in one of several genes, most of which encode components of the contractile apparatus and Arrhythmogenic Right Ventricular Dysplasia (ARVC) which is characterized by fibrofatty replacement of the myocardium and cardiac arrhythmias.

Rat chr 3 also has a second distal eQTL cluster encompassing markers at positions 118284271, 120624798, 123044649, 126425831, 129385674 and 134855115 which will be referred to as cardiac eQTL2. 920 unique transcripts map to this region. Enrichment
Figure 6.7 KEGG pathway [Ogata et al., 1999] for HCM with transcripts mapping to cardiac eQTL1 in red. Figure adapted from KEGG pathways. Abbreviations used in the figure: ITGA (Integrin alpha), ITGB (Integrin beta), SGCB (Sarcoglycan B), SGCD (Sarcoglycan D), DHPR (Voltage-gated Calcium channel), SERCA2a (Ca^{2+} transporting ATP-ase), MYBPC (Myosin-binding protein C), PRKA (cAMP activated protein kinase), ECM (extra-cellular matrix), pathway reproduced with permission.
6.2. Data analysis

Figure 6.8 KEGG pathway [Ogata et al., 1999] for insulin signalling showing transcripts from cardiac expression data that map to cardiac eQTL1 highlighted in red. Figure adapted from KEGG pathways. Abbreviations used in the figure: INSR (insulin receptor), ISR (insulin receptor substrate1), JNK (MAP kinase 10), CrkII (sarcome virus oncogene), PDK (3-phosphoinositide dependent protein kinase-1), AMPK (AMP activated protein kinase), mTOR (mechanistic target of Rapamycin), Raf (murins sarcoma virus raf), ERK1/2 (MAP kinase 1), MNK (MAP kinase-interacting serine/threonine kinase 1), eIF4E (eukaryotic translation initiation factor 4E), PDE3 (phosphodiesterase3), TSC1 (tuberous sclerosis1), PPI (protein phosphatase1), PKA (X-linked protein kinase), PHK (Calmodulin4), PYG (glycogen phosphorylase), ACC (acetyl-Coenzyme A carboxylase beta), PGC1α (peroxisome proliferator-activated receptor gamma, coactivator 1 alpha), pathway reproduced with permission.

Figure 6.8 KEGG pathway [Ogata et al., 1999] for insulin signalling showing transcripts from cardiac expression data that map to cardiac eQTL1 highlighted in red. Figure adapted from KEGG pathways. Abbreviations used in the figure: INSR (insulin receptor), ISR (insulin receptor substrate1), JNK (MAP kinase 10), CrkII (sarcoma virus oncogene), PDK (3-phosphoinositide dependent protein kinase-1), AMPK (AMP activated protein kinase), mTOR (mechanistic target of Rapamycin), Raf (murina sarcoma virus raf), ERK1/2 (MAP kinase 1), MNK (MAP kinase-interacting serine/threonine kinase 1), eIF4E (eukaryotic translation initiation factor 4E), PDE3 (phosphodiesterase3), TSC1 (tuberous sclerosis1), PPI (protein phosphatase1), PKA (X-linked protein kinase), PHK (Calmodulin4), PYG (glycogen phosphorylase), ACC (acetyl-Coenzyme A carboxylase beta), PGC1α (peroxisome proliferator-activated receptor gamma, coactivator 1 alpha), pathway reproduced with permission.
analysis reveals that transcripts mapping to this region are enriched for WD40 (W tryptophan, D Asparagine) repeat regions \((p = 0.03)\), GO CC term ‘collagen’ \((p = 0.03)\) and KEGG pathway for ECM receptor interaction \((p = 0.03)\). WD40 repeat regions are highly conserved short amino acid motifs approximately 40 residue long and are involved in a diverse range of functions including signal transduction, cell cycle control, apoptosis or platforms for protein assembly (http://www.ebi.ac.uk/interpro/IEntry?ac=IPR001680). Example of one such cardiac protein is actin-interacting protein which is involved in actin filament interactions. Collagen forms fibrous tissue and in muscular tissue such as the myocardium forms the fibrous sheaths of muscle fibres.

<table>
<thead>
<tr>
<th>Category</th>
<th>Term</th>
<th>(p) value (Benjamini)</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTERPRO</td>
<td>IPR019775:WD40 repeat, conserved site</td>
<td>0.03</td>
</tr>
<tr>
<td>GOTERM CC</td>
<td>GO:0005581:collagen</td>
<td>0.02</td>
</tr>
<tr>
<td>KEGG PATHWAY</td>
<td>rno04512:ECM-receptor interaction</td>
<td>0.03</td>
</tr>
</tbody>
</table>

**Table 6.3** Biological pathway enrichment analysis for transcripts mapping to cardiac eQTL2 on distal chr 3.

600 unique transcripts mapped to the eQTL on chr 8 referred to as **cardiac eQTL3**. With respect to enrichment for biological pathways, this cluster is similar to cardiac eQTL1 with significant enrichment for transcripts associated with mitochondrial structure and function. This information is significant for another reason: cardiac eQTL1 co-mapped with the main pQTL for cardiac mass on chr 3 identified by single-qtl mapping, figure 5.8 and cardiac eQTL3 co-maps with another locus for cardiac mass on chr 8 identified by multiple-qtl mapping, figure 5.17 whilst being enriched for a similar set of transcripts. This is a novel observation integrating multiple cardiac mass pQTLs and eQTL hotspots enriched for mitochondrial structure/function across two genomic loci in a rat model of cardiac hypertrophy. Cardiac eQTL3 is also enriched for transcripts associated with protein ubiquination. Ubiquitin is a small protein that tags other proteins and directs them to proteolysis or other destinations within the cell and it is attached to its target protein by actions of a number of enzymes together termed ubiquitin ligases [Hochstrasser, 2009]. 130 unique transcripts mapped to the eQTL hotspot on chr 1 (table
6.2. Data analysis

6.1 and figure 6.6) referred to as cardiac eQTL7 but this region was not found to be significantly enriched for any biological pathways based on information currently available. The same was also true for 323 unique transcripts mapping to the eQTL hotspot on chr 5, termed cardiac eQTL5.

<table>
<thead>
<tr>
<th>Category</th>
<th>Term</th>
<th>$p$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOTERM CC</td>
<td>GO:0005739 mitochondrion</td>
<td>$1.2 \times 10^{-7}$</td>
</tr>
<tr>
<td>GOTERM CC</td>
<td>GO:0044429 mitochondrial part</td>
<td>$1.6 \times 10^{-7}$</td>
</tr>
<tr>
<td>GOTERM CC</td>
<td>GO:0005759 mitochondrial matrix</td>
<td>$4.9 \times 10^{-5}$</td>
</tr>
<tr>
<td>GOTERM BP</td>
<td>GO:0051443 positive regulation of ubiquitin-protein ligase activity</td>
<td>0.006</td>
</tr>
<tr>
<td>GOTERM BP</td>
<td>GO:0043161 ubiquitin-dependent protein catabolic process</td>
<td>0.01</td>
</tr>
<tr>
<td>GOTERM BP</td>
<td>GO:0051340 regulation of ligase activity</td>
<td>0.02</td>
</tr>
<tr>
<td>GOTERM BP</td>
<td>GO:0031398 positive regulation of protein ubiquitination</td>
<td>0.03</td>
</tr>
<tr>
<td>GOTERM BP</td>
<td>GO:0051444 negative regulation of ubiquitin-protein ligase activity</td>
<td>0.03</td>
</tr>
<tr>
<td>GOTERM BP</td>
<td>GO:0051352 negative regulation of ligase activity</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Table 6.4 Enrichment for biological pathways in transcripts mapping to marker at position 68783831 on chr 8 (cardiac eQTL5). Ubiquitins are small proteins involved in tagging other proteins and influencing protein-macromolecule interactions.

426 unique transcripts map to the eQTL hotspot on chr 9 which will be referred to as cardiac eQTL4, however this region shows only borderline significant transcript enrichment for “electron transport chain” ($p = 0.07$, adjusted for multiple testing).

In terms of the number of transcripts per eQTL hotspot, chr 20 locus (henceforth referred to as cardiac eQTL6) at marker 16217388 is a small eQTL with only 184 transcripts mapping to this region. However gene ontology and pathway analysis enrichment analysis revealed the transcripts to be significantly enriched for transcripts associated with autoimmunity, table 6.5. Glycaemic and immune phenotypes were not the aim of the current study but a literature search reveals that this rat chromosome houses pQTLs for type 1 diabetes (Iddm37) [Blankenhorn et al., 2009] as well as anti-DNA antibody response to injected gold particles [Mas et al., 2000].

6.2.2 eQTL and QTT analysis

The next section on eQTL and QTT analysis aims to find transcripts that co-localize with trait pQTLs. In interest of clarity it would be useful to review how eQTL/pQTL
6.2. Data analysis

<table>
<thead>
<tr>
<th>Category</th>
<th>Term</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>KEGG PATHWAY</td>
<td>Graft versus host disease</td>
<td>1.6 e⁻⁰⁷</td>
</tr>
<tr>
<td>KEGG PATHWAY</td>
<td>rno04940:Type I diabetes mellitus</td>
<td>1.1 e⁻⁰⁷</td>
</tr>
<tr>
<td>KEGG PATHWAY</td>
<td>rno05320:Autoimmune thyroid disease</td>
<td>1.7 e⁻⁰⁷</td>
</tr>
<tr>
<td>KEGG PATHWAY</td>
<td>rno04612:Antigen processing and presentation</td>
<td>1.42 e⁻⁰⁷</td>
</tr>
<tr>
<td>KEGG PATHWAY</td>
<td>rno05416:Viral myocarditis</td>
<td>2.03 e⁻⁰⁶</td>
</tr>
<tr>
<td>KEGG PATHWAY</td>
<td>rno04514:Cell adhesion molecules (CAMs)</td>
<td>9.5 e⁻⁰⁶</td>
</tr>
</tbody>
</table>

Table 6.5 Biological pathway enrichment analysis for transcripts mapping to cardiac eQTL6 on chr 20.

co-mapping is different from pQTL/QTT co-localization, as the two are not synonymous. In eQTL mapping the first step is to find genomic determinants of transcript abundance. The eQTLs are then divided into cis - those which map to the same region (with 10 Mb window around the gene locus) as the gene itself and trans -those mapping to distant genomic regions. cis eQTLs co-mapping with pQTLs are then prioritized for candidate gene studies. pQTL and eQTL mapping are both based on underlying allelic variation. In QTT analysis, the trait is correlated against transcript abundance and significant QTTs are then searched for their co-localization with pQTL. A transcript that is both a cis eQTL and a QTT would meet multiple significance tests and could be an ideal candidate gene for further studies. eQTL mapping was performed using two different approaches, the traditional linkage based mapping and a Bayesian mapping algorithm both implemented in R [R Development Core Team, 2011, Dawson et al., 2009]. The significance threshold for linkage using linkage based mapping was set at 3.3.

The first pQTL to be considered is the cardiac mass pQTL located on rat chr 3 with 1.8 LOD drop interval estimate between 3 Mbp and 24 Mbp (table 5.8). As discussed above this region is also an eQTL “hotspot” and ~1000 transcripts map to this region and significantly enriched for a number of biological pathways, especially mitochondrial structure and function. However, in view of prioritizing candidate genes, the main aim to discover cis acting genes i.e. genes whose transcripts map to the same genomic locus as the gene itself. Typically trans eQTLs map at lower LOD threshold than their cis counterparts [Hubner et al., 2005] and this was replicated in our data (table 6.6).

There are 38 cis acting eQTLs co-mapping with the pQTL for LVM on chr 3 (3 Mbp -
### Phenotype/LOD

<table>
<thead>
<tr>
<th>Phenotype/LOD</th>
<th>trans eQTL</th>
<th>cis eQTL</th>
<th>cis eQTL</th>
<th>eQTLs (total)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiac mass (Chr 3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.3</td>
<td>1026</td>
<td>56</td>
<td>1082</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>369</td>
<td>38</td>
<td>407</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>19</td>
<td>12</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>BP (Chr 2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.3</td>
<td>227</td>
<td>13</td>
<td>240</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>19</td>
<td>9</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>CF cyclic&lt;sub&gt;max&lt;/sub&gt; (Chr 2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.3</td>
<td>241</td>
<td>56</td>
<td>297</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>38</td>
<td>22</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>6</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>LV dP/dt&lt;sub&gt;min&lt;/sub&gt; (Chr 3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.3</td>
<td>282</td>
<td>10</td>
<td>292</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>46</td>
<td>5</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>1</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>LV dP/dt&lt;sub&gt;max&lt;/sub&gt; (Chr 3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.3</td>
<td>282</td>
<td>10</td>
<td>292</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>46</td>
<td>5</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>1</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>CFR (Chr 8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.3</td>
<td>631</td>
<td>72</td>
<td>703</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>56</td>
<td>30</td>
<td>86</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>10</td>
<td>12</td>
<td></td>
</tr>
</tbody>
</table>

**Table 6.6** Summary of the number of eQTLs co-mapping with pQTLs. The eQTLs are divided into *trans* (mapping to a genomic region > 10 Mb away from the gene locus) and *cis* (mapping to a region within 10 Mb away from the gene locus). As the LOD threshold is raised, the proportion of transcripts mapping in *cis* increases. Since SBP, DBP and MBP map to the same locus, the eQTLs for these phenotypes are combined.
25 Mbp). Of these cis eQTLs we identified endonuclease G (Endog) as the gene underlying BP-independent cardiac hypertrophy at this locus [McDermott-Roe et al., 2011]. Endog expression is reduced in F2 animals with SHR allele (AA), figure 6.9 at the locus and as shown by our group, reduced EndoG expression is associated with increased ROS and activation of AMP-activated protein kinase (AMPK), both pro-hypertrophic stimuli [McDermott-Roe et al., 2011].

**Figure 6.9** Cardiac Endog expression is reduced in animals with SHR allele (AA) at the Endog locus. This reduction was shown to result in increased ROS production and mitochondrial dysfunction [McDermott-Roe et al., 2011]. Along y-axis are the units of gene expression which are derived from fluorescence on the expression array.

For CF cyclic\text{max} there are 56 cis acting eQTLs with a LOD score above 3.3. However this number is reduced to nine cis acting eQTLs after taking into account the QTT data and only considering transcripts which are both cis eQTLs and QTTs (table 6.7 and figure 6.12). *Dph5* is involved in tissue metabolic process and involved in Diphthamide synthesis pathway. *Sec22b* codes for a vesicle-trafficking protein and at a cellular level considered to be related to endoplasmic reticulum and Golgi apparatus. *Wars2* is an interesting candidate gene that codes for mitochondrial tryptophanyl-tRNA synthetase which plays a role in linking nucleotide triplets in tRNA to amino acids. Gene Ontology search links the expression of Wars2 to angiogenesis [Wakasugi and Schimmel, 1999] and it is interesting to note that increased Wars2 expression is higher in F2 animals having
BN allele at the locus as well as correlating positively with CF (figure 6.10). *Sypl2* (Synaptophysin-like 2) protein, on basis of homology, is speculated to be a membrane protein involved in communication between the transverse tubule and the sarcoplasmic reticulum. Likewise, *CSDE1_RAT* on basis of homology is considered to be involved in regulation of transcription. *Hist2h3c2* functions have also been inferred on basis of homology and include nucleosome assembly and DNA binding. *Slc22a15* belongs to a family of membrane transporter proteins involved in transport of metabolites. *Il6r* is interleukin 6 receptor gene which codes for a subunit of the interleukin 6 receptor and is involved in immune response, cell growth and differentiation.

<table>
<thead>
<tr>
<th>Ensembl transcript</th>
<th>chr</th>
<th>gene</th>
<th>start (bp)</th>
<th>end (bp)</th>
<th>LOD</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENSRNOT00000018367</td>
<td>2</td>
<td>Dph5</td>
<td>212033229</td>
<td>212067023</td>
<td>4.14</td>
<td>0.32</td>
</tr>
<tr>
<td>ENSRNOT00000025327</td>
<td>2</td>
<td>Sec22b</td>
<td>192743195</td>
<td>192763976</td>
<td>5.44</td>
<td>0.26</td>
</tr>
<tr>
<td>ENSRNOT00000026458</td>
<td>2</td>
<td>Wars2</td>
<td>193819876</td>
<td>193901536</td>
<td>5.93</td>
<td>0.31</td>
</tr>
<tr>
<td>ENSRNOT00000026883</td>
<td>2</td>
<td>Sypl2</td>
<td>203777722</td>
<td>203792496</td>
<td>3.36</td>
<td>-0.35</td>
</tr>
<tr>
<td>ENSRNOT00000028234</td>
<td>2</td>
<td>Il6r</td>
<td>182078051</td>
<td>182304286</td>
<td>10.19</td>
<td>-0.23</td>
</tr>
<tr>
<td>ENSRNOT00000035518</td>
<td>2</td>
<td>Igsf3</td>
<td>196492533</td>
<td>196530084</td>
<td>8.09</td>
<td>-0.23</td>
</tr>
<tr>
<td>ENSRNOT00000037824</td>
<td>2</td>
<td>CSDE1_RAT</td>
<td>198263358</td>
<td>198292478</td>
<td>6.96</td>
<td>-0.28</td>
</tr>
<tr>
<td>ENSRNOT00000040906</td>
<td>2</td>
<td>Hist2h3c2</td>
<td>191081451</td>
<td>191081861</td>
<td>6.77</td>
<td>-0.29</td>
</tr>
<tr>
<td>ENSRNOT00000046291</td>
<td>2</td>
<td>Slc22a15</td>
<td>196953291</td>
<td>197013763</td>
<td>3.38</td>
<td>-0.22</td>
</tr>
</tbody>
</table>

**Table 6.7** Prioritized candidates genes for CF based on *cis* eQTL and QTT analysis as well as co-mapping with the CF pQTL on chr 2, ‘r’ is Pearson correlation co-efficient.

The candidate genes for CFR are listed in table 6.8 derived from intersection of QTTs and *cis* eQTLs (also figure 6.12). There are eleven candidate genes on basis of overlap between QTTs and eQTLs. *Mns1* encodes a protein highly similar to the mouse meiosis-specific nuclear structural 1 protein and is postulated to play a role in meiosis. *Cmc1* codes for a Copper-binding protein involved in mitochondrial cytochrome c oxidase (COX) assembly and respiration in Saccharomyces cerevisiae. *Cmtm7* belongs to a chemokine-like factor gene superfamily and it is known to be expressed in the leucocytes but its biological function is unknown. *Acpl2* as the name suggests is a phosphatase although its relation to cardiac function is unknown. LOC100359585 is also known as calsyntenin 2 and is postulated (on basis of similarity) to modulate calcium-mediated post synaptic signals. *Ctdspl* is CTD small phosphatase-like protein, known to negatively regulate RNA
Figure 6.10 Wars2 and Il6r expression by allele, “AA” is SHR, “AB” is heterozygote and “BB” is BN.

polymerase. transcription II [Yeo et al., 2003]. D4A4J6_RAT is a predicted gene with unknown function. Mstr1 (macrophage stimulating 1 receptor) has protein-kinase activity [Wang et al., 1994]. Sema3b play a critical role in axonal guidance during embryogenesis [Puschel et al., 1995] as well as being involved in apoptosis [Castro-Rivera et al., 2004]. Bckdhb (branched chain keto acid dehydrogenase E1, beta polypeptide) is associated with the inner membrane of mitochondria, and functions in the catabolism of branched-chain amino acids. Catabolism of branched-chain amino acids also has physiologic significance in maintaining normal cardiac function [Sun et al., 2011].

LV dP/dt\textsubscript{max} and LV dP/dt\textsubscript{min} pQTLs map to the same genomic region 5.5. Search for an overlap between cis eQTLs and QTTs correlating with LV dP/dt\textsubscript{max} and LV dP/dt\textsubscript{min} did not reveal any common transcripts. Search for plausible biological candidates in the relatively small linkage region identified two candidates genes Titin and Myosin binding protein C 3 (MYBPC3), both highly expressed in the cardiac tissue. Two observations make Titin the more promising candidate of the two, presence of non-synonymous SNPs between the strains affecting a region shown to be important for myocardial relaxation and its proximity to the linkage peak within the 1.8 LOD interval. MYBPC3 does not have any coding region polymorphism and lies at the fringe of the 1.8 LOD interval. Titin is the largest protein in the body coded for by a single gene (D4A4B0_RAT in the
6.2. Data analysis

![Scatter plots showing correlation values]

Figure 6.11 QTTs correlating with CF which are additionally cis eQTLs. CF is reported along the x-axis and individual transcripts levels along the y axis (which is variable between QTTs). Label above each transcript corresponds to Ensembl transcript. Pearson correlation values are reported in the top right corner. The lines represent linear regression slopes. Although there were more transcripts correlating with CF, only the ones which are also eQTLs co-mapping with CF pQTL are reported here.
Figure 6.12 Correlation graphs and corresponding allele-wise expression for the four QTTs/cis eQTLs with most significant linkage that co-map with CFR QTL on chr 8.
6.2. Data analysis

<table>
<thead>
<tr>
<th>Ensembl transcript</th>
<th>gene</th>
<th>chr</th>
<th>start (bp)</th>
<th>end (bp)</th>
<th>r</th>
<th>p</th>
<th>LOD</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENSRNOT00000007941</td>
<td>Mns1</td>
<td>8</td>
<td>77015290</td>
<td>77035977</td>
<td>-0.27</td>
<td>0.01</td>
<td>5.38</td>
</tr>
<tr>
<td>ENSRNOT000000013467</td>
<td>Cmc1</td>
<td>8</td>
<td>122988373</td>
<td>123046369</td>
<td>-0.30</td>
<td>0.003</td>
<td>17.08</td>
</tr>
<tr>
<td>ENSRNOT000000014861</td>
<td>Cmtm7</td>
<td>8</td>
<td>119175165</td>
<td>119202081</td>
<td>0.18</td>
<td>0.03</td>
<td>5.13</td>
</tr>
<tr>
<td>ENSRNOT000000017085</td>
<td>Acpl2</td>
<td>8</td>
<td>101917843</td>
<td>101983158</td>
<td>-0.31</td>
<td>0.005</td>
<td>14.51</td>
</tr>
<tr>
<td>ENSRNOT000000036963</td>
<td>LOC100359585</td>
<td>8</td>
<td>102328707</td>
<td>102338366</td>
<td>0.23</td>
<td>0.03</td>
<td>11.43</td>
</tr>
<tr>
<td>ENSRNOT000000045368</td>
<td>Ctdspl</td>
<td>8</td>
<td>123960578</td>
<td>123993452</td>
<td>-0.27</td>
<td>0.002</td>
<td>7.42</td>
</tr>
<tr>
<td>ENSRNOT000000057704</td>
<td>D4A4J6_RAT</td>
<td>8</td>
<td>71088783</td>
<td>71106121</td>
<td>0.28</td>
<td>0.01</td>
<td>7.07</td>
</tr>
<tr>
<td>ENSRNOT000000066104</td>
<td>Mst1r</td>
<td>8</td>
<td>113175977</td>
<td>113189511</td>
<td>-0.21</td>
<td>0.03</td>
<td>6.63</td>
</tr>
<tr>
<td>ENSRNOT000000066296</td>
<td>Sema3b</td>
<td>8</td>
<td>112845711</td>
<td>112852565</td>
<td>-0.21</td>
<td>0.02</td>
<td>11.50</td>
</tr>
<tr>
<td>ENSRNOT000000067130</td>
<td>RGD1305464</td>
<td>8</td>
<td>60995695</td>
<td>61003112</td>
<td>-0.22</td>
<td>0.01</td>
<td>6.16</td>
</tr>
<tr>
<td>ENSRNOT000000067256</td>
<td>Bckdlib</td>
<td>8</td>
<td>88997979</td>
<td>89191017</td>
<td>0.22</td>
<td>0.05</td>
<td>9.74</td>
</tr>
</tbody>
</table>

Table 6.8 CFR candidates gene based on intersection of cis eQTLs and co-mapping QTTs at the CFR locus on chr 8.

rat, ENSRNOT00000032008) and consisting of 145 exons. Titin extends from the I band region of the sarcomere to the M band and for description is divided into two sections, I-band titin which is elastic and extensible and A-band titin which is inextensible. I-band titin is extensible and has three isoforms N2B, N2BA (composed of N2B and an additional N2A region) and a fetal isoform. N2B and N2BA isoforms are both expressed in adult hearts with N2BA having longer extensible region and being more compliant that the N2B isoform; the ratio of the two isoforms may determine passive stiffness in adult mammalian heart [LeWinter and Granzier, 2010, Lewinter et al., 2010]. Phosphorylation at serine residues in the N2B segment additionally reduce titin-dependent passive tension [LeWinter and Granzier, 2010] Comparison of the rat titin sequence variation between BN and SHR reveals the presence of nine non-synonymous SNPs (nsSNP) in the SHR genome compared with the reference BN genome, table 6.9. Five of these nsSNP are present in exon 44 which codes for part of N2B Titin sequence, table 6.9 and two nsSNPs result in replacement of the Serine residues, one at position 3905 resulting in replacement of Serine with Phenylalanine and the other at position 4138 replacing Serine with Leucine. These substitutions are predicted to affect protein function on basis of alteration in amino acid properties [Henikoff and Henikoff, 1992].
6.3 Discussion

<table>
<thead>
<tr>
<th>SNP position</th>
<th>Exon</th>
<th>Reference AA</th>
<th>Alternate AA</th>
<th>AA position</th>
</tr>
</thead>
<tbody>
<tr>
<td>chr3:59661015</td>
<td>3</td>
<td>A</td>
<td>V</td>
<td>121</td>
</tr>
<tr>
<td>chr3:59654243</td>
<td>6</td>
<td>T</td>
<td>A</td>
<td>394</td>
</tr>
<tr>
<td>chr3:59600175</td>
<td>44</td>
<td>L</td>
<td>S</td>
<td>3864</td>
</tr>
<tr>
<td>chr3:59600095</td>
<td>44</td>
<td>D</td>
<td>N</td>
<td>3891</td>
</tr>
<tr>
<td>chr3:59600052</td>
<td>44</td>
<td>S</td>
<td>F</td>
<td>3905</td>
</tr>
<tr>
<td>chr3:59599359</td>
<td>44</td>
<td>V</td>
<td>A</td>
<td>4136</td>
</tr>
<tr>
<td>chr3:59599353</td>
<td>44</td>
<td>S</td>
<td>L</td>
<td>4138</td>
</tr>
<tr>
<td>chr3:59468374</td>
<td>117</td>
<td>A</td>
<td>T</td>
<td>9405</td>
</tr>
<tr>
<td>chr3:59468146</td>
<td>118</td>
<td>L</td>
<td>P</td>
<td>9450</td>
</tr>
<tr>
<td>chr3:59464765</td>
<td>123</td>
<td>V</td>
<td>I</td>
<td>9939</td>
</tr>
</tbody>
</table>

Table 6.9 Consequences of nsSNPs in the Titin sequence between BN and SHR. Exon 44 codes for part of the N2B sequence of Titin and previous studies have shown that phosphorylation occurring at Serine residues in the N2B results in diastolic function improvement [Kruger and Linke, 2006]. Two of the five mutations in exon 44 result in replacement of the Serine (S) residues by F (Phenylalanine) and L (Leucine). Others include V (Valine), A (Alanine), T (Threonine), D (Aspartate), I (Isoleucine), L (Leucine), N (Asparagine) and P (Proline)

6.3 Discussion

A number of significant novel observations have been reported in the sections above. One of the most important observations is the presence of multiple cardiac eQTL hotspots pointing to genomic regulators of transcription in the heart. Two of these eQTL hotspots co-map to pQTLs for cardiac mass and are highly enriched for transcripts related to mitochondrial structure and function. There are a number of potential explanations including either the presence of key transcriptional regulatory genes at the locus, mitochondrial response to cardiac hypertrophy/stress or primary mitochondrial dysfunction leading to secondary cardiac hypertrophy.

By integrating eQTL analysis with QTT and genome sequence analysis a number of potential candidate genes can be prioritized for further studies. We have already published EndoG as a novel determinant of cardiac hypertrophy and mitochondrial function [McDermott-Roe et al., 2011]. Additionally, Titin is a plausible potential candidate gene for LV dP/dt$_{\text{min}}$ and is the subject of further studies in the laboratory. A number of genes could underlie the CF pQTL on chr 2 (table 6.7) and CFR pQTL on chr 8(table 6.8).
Chapter 7

Conclusions

The aims of the current study were to establish the heritability and covariance of cardiac traits especially with relation to BP and cardiac traits previously not studied from a genomic perspective including CF, CFR, LV dP/dt\text{min} and LV dP/dt\text{max}, to determine the heritability of these traits, to discover novel cardiac pQTLs and to discover the causative genes integrating information from genome sequence and transcriptional profiling in inbred parental rat strains (BN and SHR) and approximately 180 animals from the derived F\textsubscript{2} intercross.

The physiological tools used to answer these questions included cardiac phenotyping using \textit{in vivo} carotid cannulation with BP and HR monitoring, \textit{ex vivo} perfusion and measurement of CF, CFR, LV dP/dt\text{min} and LV dP/dt\text{max} at baseline, during regional ischaemia induced by LAD ligation and reperfusion as well as cardiac mass measured at the end of each experiment. Custom genotyping array was then designed using the genome sequences of BN and SHR to perform 768-plex genotyping assay on an automated system. Cardiac expression analysis was carried out on the non-ischaemic LV tissue from 115 samples using Affymetrix rat gene chip 1.0 ST to discover QT\texttt{Ts}, eQTLs and prioritize candidate genes for further studies. Bioinformatic tools in form of R, R/qtl, R/eqtl and R/eqtlM were utilized to undertake linkage and expression analyses.

From a physiological perspective there were a number of novel observations. The most significant perhaps was the lack of correlation between BP and other cardiac phenotypes including CF, CFR, cardiac mass, LV dP/dt\text{min} and LV dP/dt\text{max}. Other important obser-
vations include the effect of mitochondrial or Y chromosome inheritance on cardiac mass and significant heritability estimates for cardiac traits including CF, CFR, LV dP/dt\textsubscript{min} and LV dP/dt\textsubscript{max}. BP is the archetype heritable cardiovascular phenotype and the current study proved that the other cardiac traits are at least as heritable as BP.

A number of novel cardiac QTLs were discovered during this study including QTLs for CF, CFR, LV dP/dt\textsubscript{min} and LV dP/dt\textsubscript{max} as well as confirming previously observed QTLs for BP and cardiac mass. Multiple-QTL models proved that significant proportion of trait variation is explained by genetic variation especially if multiple loci and modes of gene interaction (additive or epistatic) are accounted for albeit such models being less tractable from a genetic viewpoint.

Finally, expression analysis revealed the presence of cardiac eQTL hotspots across the genome. Cardiac eQTL hotspots on chromosomes 3 and 8 are significantly enriched for transcripts underlying mitochondrial function and co-map with pQTLs for cardiac mass. This suggests a highly significant correlation between mitochondrial function and cardiac hypertrophy and we identified EndoG as a novel determinant of cardiac hypertrophy and cardiac hypertrophy. Integrating QTT and eQTL analyses there are a small number of genes which could prove to be novel determinants of CF and CFR. These studies have also prioritized Titin as a strong candidate gene underlying pQTL for LV dP/dt min with mutations affecting Serine residues in the functionally important N2B sequence. Taken together, this study has advanced our understanding of the genetic basis of cardiac traits.
List of Tables

2.1 Ionic composition of perfusion solution for *ex vivo* heart preparation . . . . 40

3.1 Baseline cardiac phenotypes for four inbred rat strains from the PhysGen database (BN=20, Lewis=10, SHR=20, WKY= 10). Cardiac mass (expressed as % of body weight * 100), heart rate (HR, beats/minute), CF(ml/g/min) and LV developed pressure (mmHg) expressed as mean ± SD. ....................................................... 64

3.2 Post MI cardiac phenotypes for four inbred rat strains from the PhysGen database (BN=20, Lewis=10, SHR=20, WKY= 10). Heart rate (beats/minute), CF (ml/g/min) and LV developed pressure expressed (mmHg) as mean ± SD. ....................................................... 66

3.3 Baseline cardiac phenotypes: (BN=16, Lewis=15, SHR=19, WKY= 9). Cardiac mass (heart weight/ body weight)*100, CF (ml/g/min) and LV developed pressure expressed (mmHg) as mean ± SD. ............................... 69

3.4 Post MI cardiac phenotypes: CF (ml/g/min), LV developed pressure expressed (mmHg), LV dP/dt_{min} and LV dP/dt_{max} (mmHg/s) as mean ± SD. ....................................................... 69

3.5 Cardiac mass ((heart weight/body weight)*100) and BP phenotypes in F_{1} intercross (BN=18, F_{1}BXH=7, F_{1}HXB=10, SHR=17) expressed as mean ± SD. ....................................................... 71

3.6 CF (ml/gram/minute) in the parentals and F_{1} strains at baseline, in response to one minute ischaemia and reperfusion expressed as mean ± SD. .. 74
### List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Parental, F₁ and F₂ intercross BP (mmHg) expressed as mean ± SD</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td>(BN=18, SHR=15, F₁=17, F₂=188). HR data expressed as bpm.</td>
<td></td>
</tr>
<tr>
<td>4.2</td>
<td>CF (ml/gram/minute) in BN (n=18), SHR (n=15), F₁ (n=17) and F₂ (n=188) expressed as mean ± SD. CFR expressed as a ratio of maximal hyperaemic CF and baseline CF.</td>
<td>84</td>
</tr>
<tr>
<td>4.3</td>
<td>Broad sense heritabilities of cardiovascular traits in the F₂ intercross.</td>
<td>88</td>
</tr>
<tr>
<td>5.1</td>
<td>Summary of the genotyping data prior to quality checking.</td>
<td>93</td>
</tr>
<tr>
<td>5.2</td>
<td>Segregation distortion results showing three markers on chromosomes 4, 10 and 16 showing extreme segregation distortion.</td>
<td>93</td>
</tr>
<tr>
<td>5.3</td>
<td>Six highest scoring erroneous genotypes likely representing genotyping errors detected by identifying tight double cross overs, interestingly all in the same individual indicating that there might have been errors during DNA extraction/processing. There were another twenty genotypes with error score 5 &gt; LOD_{error} confined to six individuals.</td>
<td>96</td>
</tr>
<tr>
<td>5.4</td>
<td>Summary of the F₂ genotyping data post QC.</td>
<td>98</td>
</tr>
<tr>
<td>5.5</td>
<td>Single-QTL linkage analysis for cardiovascular traits at baseline.</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>LOD_{threshold} is the LOD threshold obtained for individual phenotype by use of 1000 permutation test, LOD_{peak} is the LOD score obtained by standard interval mapping with corresponding chromosomes and genetic position for peak marker position. P-value_{genome-wide} is the proportion of LOD_{threshold} that would exceed LOD_{peak} anywhere in the genome.</td>
<td></td>
</tr>
<tr>
<td>5.6</td>
<td>Single-QTL linkage analysis for cardiovascular traits during reperfusion. CF_{recovered} is the ratio of CF during reperfusion to CF at baseline.</td>
<td>101</td>
</tr>
<tr>
<td>5.7</td>
<td>QTL effect size reported as % of the phenotype that is explained by the QTL with the corresponding effects in units of phenotype. Cardiac mass is indexed to bodyweight and CFR is dimensionless (ratio of peak hyperaemic flow to baseline flow).</td>
<td>104</td>
</tr>
</tbody>
</table>
5.8 1.8 LOD-drop intervals for significant QTLs, start and end positions are
the physical positions in the genome extending to the nearest genotyping
markers for the interval, whereas region size is the size of 1.8-LOD drop
region in cM. ................................................................. 106
5.9 Multiple-QTL modelling for cardiovascular traits; the final results incorpo-
rates significant loci from two-QTL genome scans as well as multiple-QTL
mapping. ................................................................. 118
5.10 Previously described loci for cardiac mass coincident with the locus on
chromosome 3 identified in current study using BN/SHR F₂ intercross. .... 119
6.1 eQTL clusters obtained from expression profiling of non-ischaemic LV and
the number of transcripts mapping to each cluster. ............................... 130
6.2 Biological pathway enrichment analysis for transcripts mapping to cardiac
eQTL1, table 6.1, figure 6.6). GO term refers to the controlled vocabu-
mary used by GO consortium to describe biological process (BP), cellular
component (CC) and molecular function (MF). KEGG (Kyoto encyclopa-
dia of genes and genomics) describes the pathways where these transcripts
are overrepresented. The final column gives the significance corrected for
multiple testing. ................................................................. 131
6.3 Biological pathway enrichment analysis for transcripts mapping to cardiac
eQTL2 on distal chr 3. ................................................................. 134
6.4 Enrichment for biological pathways in transcripts mapping to marker at
position 68783831 on chr 8 (cardiac eQTL5). Ubiquitins are small proteins
involved in tagging other proteins and influencing protein-macromolecule
interactions. ................................................................. 135
6.5 Biological pathway enrichment analysis for transcripts mapping to cardiac
eQTL6 on chr 20. ................................................................. 136
6.6 Summary of the number of eQTLs co-mapping with pQTLs. The eQTLs are divided into \textit{trans} (mapping to a genomic region > 10 Mb away from the gene locus) and \textit{cis} (mapping to a region within 10 Mb away from the gene locus). As the LOD threshold is raised, the proportion of transcripts mapping in \textit{cis} increases. Since SBP, DBP and MBP map to the same locus, the eQTLs for these phenotypes are combined.

6.7 Prioritized candidates genes for CF based on \textit{cis} eQTL and QTT analysis as well as co-mapping with the CF pQTL on chr 2, ‘r’ is Pearson correlation co-efficient.

6.8 CFR candidates gene based on intersection of \textit{cis} eQTLs and co-mapping QTTs at the CFR locus on chr 8.

6.9 Consequences of nsSNPs in the Titin sequence between BN and SHR. Exon 44 codes for part of the N2B sequence of Titin and previous studies have shown that phosphorylation occurring at Serine residues in the N2B results in diastolic function improvement [Kruger and Linke, 2006]. Two of the five mutations in exon 44 result in replacement of the Serine (S) residues by F (Phenylalanine) and L (Leucine). Others include V (Valine), A (Alanine), T (Threonine), D (Aspartate), I (Isoleucine), L (Leucine), N (Asparagine) and P (Proline).
List of Figures

1.1 Schematic representation of the pathophysiological mechanisms underlying CF. Immune system activation increases ROS generation in response to a number of pathogenic stimuli leads to reduce NO bioavailability. This causes alterations in a number of key cardiac phenotypes. The disturbance may lead to a positive feedback loop perpetuating the cycle. The genetic determinants of these processes remain unknown. .................................. 17

1.2 MA plots before normalization (left) and after normalization (right) for one of the example datasets obtained from Bioconductor. .................. 33

2.1 BP recording during carotid cannulation. Inhaled isoflurane is gradually reduced to 1.5%. BP gradually increases till a plateau is reached. The observations during the plateau phase are analyzed offline for systolic, diastolic and mean BP. ......................................................... 38

2.2 (a) Langendorff isolated heart preparation; the heart is attached to a perfusion cannula via the aorta. Water filled latex balloon is inserted in the left ventricle via the left atrium and the heart is paced at 360bpm by the pacing electrode; (b) A schematic description of the Langendorff system, adapted from ADI instruments. ......................................................... 43

2.3 Overview of the experimental stages and phenotypes measured during the physiological experiments. LV \( \frac{dP}{dt_{\text{max}}} \) and LV \( \frac{dP}{dt_{\text{min}}} \) are measures of rates of LV pressure change during systole and diastole respectively. .... 44
2.4 Planimetry and estimation of infarct area; (a) shows cardiac slices before image processing. The blue area is the non-ischaemic myocardium, the red area is the TTC stained alive myocardium and the white area represents infarction; (b) shows measurement of the total area; (c) represents the area at risk after removal of non-ischaemic myocardium; (d) represents the ischaemic but non-infarcted myocardium.

2.5 Goldengate genotyping workflow: Illustration of ssDNA with a [T/C] SNP; the gDNA is “activated” by random binding of biotin to gDNA.

2.6 The extension/ligation step; ASOs bind to the target DNA depending on the terminal SNP (allele-specific binding), PCR and ligation steps then bind the ASO to the LSO.

2.7 Following allele-specific extension/ligation, universal primers with fluorophores are annealed, PCR is performed followed by hybridization and scanning steps where the signal of fluorophore indicates the specific allele at the locus.

2.8 Representation of the bioinformatic scores for the SNPs submitted for the first six rat chromosomes, the final SNP scores (scoring for probability of success in genotyping assay) are plotted against the physical co-ordinates of the SNPs for the individual chromosomes. The dashed lines represents the cutoff score of 0.95. Only SNPs scoring more than 0.95 were included in designing the custom assay.

2.9 Final 768 SNPs selected for genotyping plotted by their physical co-ordinates on each rat chromosome.
2.10 Cluster separation analysis for homozygotes and heterozygotes in the Gold-enGate assay. Norm R represents normalized signal intensity and Norm theta is the red/green colour intensity. Each dot represents a single sample and labels above each panel refer to the particular SNP for which the genotypes are being called. Panels (c) and (d) show manual adjustment of automatic SNP clusters with (c) being the result of automatic clustering and (d) being the same cluster after manual adjustment. For some samples the signal intensity is ambiguous and by excluding these samples genotyping reliability can be improved.

2.11 Example of an excluded SNP because due to low signal intensity (Norm R below 0.20) even though the three genotype clusters are well separated.

2.12 Example of RIN score profile for one RNA sample prior to RNA labelling. The exact RIN score is obtained by analysis of this trace. Fluorescence units are plotted along the y-axis and time is displayed along the x-axis. The first peak is at 18S and the second peak at 28S. 28S/18S ratio should be greater than 1. There should be few, if any, baseline perturbations and no visible hump after the 28S peak which would indicate presence of genomic DNA. The RIN score for this particular profile was 8.8.

2.13 Bioanalyzer trace from a sample before (above) and after fragmentation (below). cDNA has been adequately fragmented as indicated by the shift of the peak of distribution to the left. This fragmentation can also be seen in gel images on the right obtained from the Bioanalyzer.

3.1 PhysGen phenotyping protocol, six week old animals were phenotyped, myocardial contracture is the contraction in the cardiac muscle in response to ischaemia; LDH, Lactate dehydrogenase.

3.2 Baseline phenotypes; (a) Cardiac mass indexed to body weight; (b) Heart Rate (beats/minute); (c) CF across strains (ml/gram/min) (d) LV developed pressure. Each dot represents an individual experimental animal.
3.3 MI related phenotypes; (a) CK release (IU/L/gram) at 40 minutes after reperfusion; (b) Area of infarction measured after TTC staining expressed as a ratio of total LV area. ......................................................... 66

3.4 Baseline phenotypes; (a) Cardiac mass indexed to body weight; (b) Myocardial Relaxation (mmHg/s); (c) CF(ml/gram/min) (d) LV developed pressure. ................................................................. 68

3.5 Average CF (ml/gram/min) for the first two minutes following establishment of ex vivo perfusion (BN=16, SHR=19). ................................................................. 70

3.6 Cardiac mass and BP phenotypes in F₁ intercross (BN=18, F₁BXH=7, F₁HXB=10, SHR=17), (a) Cardiac mass indexed to body weight; (b) Mean BP (mmHg); (c) Systolic BP (mmHg) (d) Diastolic BP (mmHg) .................................................. 72

3.7 CF phenotypes in F₁ intercross, (a) baseline; (b) following one minute global ischaemia to induce maximal hyperaemia; (c) reperfusion; (d) Coronary flow reserve (CFR), ratio of maximal hyperaemic CF to baseline CF. ...................................................... 73

3.8 LV contractility in F₁ intercross, (a) dP/dtₘₖᵢₙ at baseline; (b) dP/dtₘᵢₙ during reperfusion; (c) dP/dtₘᵃₓ at baseline; (d) dP/dtₘᵃₓ during reperfusion. .................................................. 75

4.1 Cardiovascular phenotypes of individual animals (circles), along the y-axis, corresponding to the order in which they were measured for 196 F₂ animals over the course of one year with the phenotype values plotted along the x-axis; (a) Cardiac mass indexed to body weight; (b) Mean BP (mmHg); (c) CF across strains (ml/gram/min) .................................................. 81

4.2 Cardiovascular phenotypes of individual animals (circles), along the y-axis, corresponding to the order in which they were measured for 196 F₂ animals over the course of one year with the phenotype values plotted along the x-axis; (a) in vivo HR during direct carotid cannulation; (b) LV dP/dtₘᵢₙ (mmHg/s) .................................................. 82

4.3 Cardiac mass in the F₂ intercross compared with the parentals and F₁ intercross. .................................................. 83

4.4 BP (mmHg) indices in the F₂ compared with the parentals strains. .................................................. 84
List of Figures

4.5 LV dP/dt_{min} (mmHg/s) in the F_2 intercross compared with the parentals strains. During baseline SHR has greater relaxation values than BN and F_2 intercross. ......................................................... 85

4.6 Correlation between CF and LV dP/dt_{min} at baseline, p= 4.62\times 10^{-11}, ad-
justed R^2=0.21. ................................................................. 86

4.7 Correlation matrix between cardiac mass, systolic, mean and diastolic BP in the F_2 intercross. The numbers in in the upper panel are Pearson cor-
relation values ‘r’ and significance of results are represented by ‘*’ (p <
0.05=‘*’, p < 0.01=‘**’, p < 0.001=‘***’). Actual data points are plotted
in the lower panel with loess lines. ........................................ 86

4.8 Correlation matrix between systolic, mean and diastolic BP and CF in the
F_2 intercross. The numbers in in the upper panel are Pearson correlation
values ‘r’ and significance of results are represented by ‘*’ (p < 0.05=‘*’, p
< 0.01=‘**’, p < 0.001=‘***’). Actual data points are plotted in the lower
panel with loess lines. ............................................................ 87

5.1 Comparison of genotype similarities between samples. The data shows
that there are no identical samples in the current dataset and the highest
proportion of similarity between any two samples is just above 60%. ...... 94

5.2 Estimated LOD scores (lower right) and RF (upper left) for the F_2 inter-
cross data. Low RF (linked) markers would be shown in red. .......... 95

5.3 Map based on genetic distances between the markers in F_2 intercross data. 95

5.4 Number of crossover events plotted against the F_2 index, there are six
individuals with excessive number of crossover (> 100). ................. 97

5.5 Missing genotype information for the F_2 intercross. Regions of missing
genotype information may show spurious linkage but there are no such
regions of missing information in current F_2 intercross. ............... 98

5.6 Distribution of LOD scores from 1000 permutation test given the F_2 inter-
cross dataset showing that a LOD score of 3.9 will only be exceeded 5% by
chance. .................................................................................. 100
List of Figures

5.7 Genome-wide LOD plot for systolic, diastolic and mean pressure showing a linkage peak for all phenotypes on chromosome 2. Vertical bars on the x-axis are genotyped markers on each chromosome. ........................................... 101

5.8 Genome-wide LOD for cardiac mass showing linkage peak at the start of chromosome 3 and pulse pressure linkage peak at the previously linked BP locus on chromosome 2. Vertical bars on the x-axis are genotyped markers on each chromosome. ........................................... 101

5.9 Genome-wide LOD plot for CF cyclic_{max} and CFR showing separate linkage peaks for phenotypes on chromosome 2 and 8 respectively. Vertical bars on the x-axis are genotyped markers on each chromosome. ........................................... 102

5.10 Genome-wide LOD plot for baseline myocardial relaxation (LV dP/dt_{min}) and myocardial contractility (LV dP/dt_{max}) showing identical linkage peak on chromosome 3. Vertical bars on the x-axis are genotyped markers on each chromosome. ........................................... 102

5.11 Zoomed view of linkage peaks for systolic BP, diastolic BP and CF on chromosome 2 showing separate linkage regions. Vertical bars on the x-axis are genotyped markers on each chromosome. ........................................... 103

5.12 Effect plots for phenotypes given genotype at the peak of linkage analysis for single QTLs listed in table 5.8 SHR genotype is represented by AA, BN by BB and heterozygotes by AB. The peak marker ID is given for each phenotype. ........................................... 105

5.13 QTL significance thresholds remain unaffected after taking into account the clinically important covariates; (a) effect of BP on cardiac mass QTL; (b) effect of BP on CFR QTL; (c) effect of BP on LV dP/dt_{min}; (d) effect of LV dP/dt_{min} on CF QTL) ........................................... 106
5.14 Two-QTL scan for cardiac mass; LOD\textsubscript{fv1} (\log_{10} likelihood ratio comparing the full model to a single QTL model) is displayed in the upper left half whereas LOD\textsubscript{av1} (\log_{10} likelihood ratio comparing the additive model with QTL to a single QTL model) is displayed in the lower right half. Significant interactions between loci are highlighted by colour scale according to bar on the right. This bar depicts a colour scale and numbers to the left of the bar correspond to LOD\textsubscript{fv1} (comparing full model to single QTL) and numbers to right correspond to LOD\textsubscript{av1} (comparing additive model to single QTL). 109

5.15 Two-QTL scan results for cardiac phenotypes showing suggestive improvement in using two-qtl model over single-QTL scan. In each panel significant interactions between loci are highlighted by colour scale according to bar on the right. LOD\textsubscript{fv1} is in the upper left half of each panel whereas LOD\textsubscript{av1} is in the lower right half of each panel. 110

5.16 Map of cofactors plotted for multiple QTL mapping across the genome in the F\textsubscript{2} intercross 112

5.17 Genomic locations of the four most significant cofactors influencing cardiac mass. 112

5.18 Assessment of interactions between significant loci for cardiac mass (adjusted for body weight, hwt.bdqt) on chr 3 (marker 17358758) and chr 8 (marker 113964559). The graph shows significant additive interaction. 113

5.19 Assessment of interactions between significant loci for cardiac mass on chr 3 (marker 17358758) and chr 17 (marker 40120016) indicates additive action of alleles at these loci. 114

5.20 Genome-wide search for significant loci influencing SBP identify significant loci on chr 1, 2, 3, 7, 17 and 19. 114

5.21 Genomic positions of six loci associated with DBP identified using multiple-qtl approach. 115
5.22 Evidence of epistatic interaction for DBP loci on chr 2 and 7. Individuals with BB phenotype at marker 72928519 on chr 7 shows no difference in phenotype between AA and BB phenotypes at marker 75860058 on chr 2. The alleles at marker 75860058 (chr 7) only manifest in the presence of AA allele at marker 72928519 (chr 2). ............................................ 116

5.23 Additively acting loci for CF cyclic_max on chr 1 (marker 229399235) and 2 (marker 203879555). For both loci the BN genotype, represented by BB has higher CF compared to the SHR genotype (AA). .......................... 117

5.24 Genomic locations of significant loci for CF cyclic_max identified using multiple QTL approach. .......................................................... 117

5.25 Significant cofactors for CFR on chr 1 and 12 in addition to the previously described locus on chr 8. ............................................. 118

5.26 Rat chr 3 harbouring the LVM QTL (shaded grey) coloured according to syntenic regions in the mouse (left) and human genomes (right). The coloured segments of rat chr 3 represent syntenic segments from different chromosomes of the mouse genome on the left and human genome on the right. The QTL containing region in rat extending (grey shaded area) from 3 Mbp - 25 Mbp is syntenic to mouse chr 2 (dark green) and human chr 9 (yellow), 2, 11 and 15 (coded together as green). ................. 120

5.27 Rat chr 2 colour coded according to mouse syntenic regions (left) and human chr 8 (right). The coloured segments of rat chr 2 represent syntenic segments from different chromosomes of the mouse genome on the left and human genome on the right. The regions on mouse chr 15 and human chr 8 (shaded as grey) also exhibit linkage to BP. The shared syntenic region between the three species containing BP are shaded as transparent red and likeliest to contain the underlying gene. ................................. 121
5.28 Rat chr 3 colour coded according to syntenic human regions. The region containing LV dp/dt\textsubscript{max} locus is shaded in grey. The dark green region is syntenic to human chr 2 which does not have any linkage to myocardial contractility and the light green region at bottom is syntenic to human chr 11 which harbours LV dp/dt\textsubscript{max} locus in humans. 122

6.1 Heatmap of the distances between arrays. Distance here is the difference between signal intensities obtained from the microarrays. 124

6.2 Principal component analysis: A scatterplot of the arrays along the first two principal components PC1 and PC2. The scatterplot does not show any evidence for array clustering which may be intentional (two or more experimental groups) or unintentional (batch effects). 125

6.3 Boxplots representing summaries of the signal intensity distributions of the arrays after background correction and normalization. Signal intensities are plotted along the x-axis for arrays plotted on the y-axis. The asterisk-marked arrays were initial outliers and improved after correction and normalization. 126

6.4 Density estimates (smoothed histograms) of the data. No sample shows significant difference from the overall distribution. 127

6.5 MA plots: There is no evidence of a significant trend in M as a function of A and distribution is concentrated along zero axis represented by a red line. Outlier detection was performed by computing Hoeffding’s statistic $D_a$ on the joint distribution of A and M for each array. Shown are the 4 arrays with the highest value of $D_a$ (top row), and the 4 arrays with the lowest value (bottom row). The value of $D_a$ is shown in the panel headings. There are no outliers based on this analysis. 127

6.6 Map of genome-wide cardiac eQTL hotspots. 130
6.7 KEGG pathway [Ogata et al., 1999] for HCM with transcripts mapping to cardiac eQTL1 in red. Figure adapted from KEGG pathways. Abbreviations used in the figure: ITGA (Integrin alpha), ITGB (Integrin beta), SGCB (Sarcoglycan B), SGCD (Sarcoglycan D), DHPR (Voltage-gated Calcium channel), SERCA2a (Ca\textsuperscript{2+} transporting ATP-ase), MYBPC (Myosin-binding protein C), PRKA (cAMP activated protein kinase), ECM (extra-cellular matrix), pathway reproduced with permission.

6.8 KEGG pathway [Ogata et al., 1999] for insulin signalling showing transcripts from cardiac expression data that map to cardiac eQTL1 highlighted in red. Figure adapted from KEGG pathways. Abbreviations used in the figure: INSR (insulin receptor), ISR (insulin receptor substrate1), JNK (MAP kinase 10), CrkII (sarcome virus oncogene), PDK (3-phosphoinositide dependent protein kinase-1), AMPK (AMP activated protein kinase), mTOR (mechanistic target of Rapamycin), Raf (murins sarcoma virus raf), ERK1/2 (MAP kinase 1), MNK (MAP kinase-interacting serine/threonine kinase 1), eIF4E (eukaryotic translation initiation factor 4E), PDE3 (phosphodiesterase3), TSC1 (tuberous sclerosis1), PP1 (protein phosphatase1), PKA (X-linked protein kinase), PHK (Calmodulin4), PYG (glycogen phosphorylase), ACC (acetyl-Coenzyme A carboxylase beta), PGC1\textalpha (peroxisome proliferator-activated receptor gamma, coactivator 1 alpha), pathway reproduced with permission.

6.9 Cardiac Endog expression is reduced in animals with SHR allele (AA) at the Endog locus. This reduction was shown to result in increased ROS production and mitochondrial dysfunction [McDermott-Roe et al., 2011]. Along y-axis are the units of gene expression which are derived from fluorescence on the expression array.

6.10 Wars2 and Il6r expression by allele, “AA” is SHR, “AB” is heterozygote and “BB” is BN.
List of Figures

6.11 QTTs correlating with CF which are additionally \textit{cis} eQTLs. CF is reported along the x-axis and individual transcripts levels along the y axis (which is variable between QTTs). Label above each transcript corresponds to Ensembl transcript. Pearson correlation values are reported in the top right corner. The lines represent linear regression slopes. Although there were more transcripts correlating with CF, only the ones which are also eQTLs co-mapping with CF pQTL are reported here.

6.12 Correlation graphs and corresponding allele-wise expression for the four QTTs/\textit{cis} eQTLs with most significant linkage that co-map with CFR QTL on chr 8.


Bibliography


[Saar et al., 2008] Saar, K., Beck, A., Bihoreau, M. T., Birney, E., Brocklebank, D., Chen, Y., Cuppen, E., Demonchy, S., Dopazo, J., Flicek, P., Foglio, M., Fujiyama,


tetrazolium staining of irreversible ischemic injury following coronary artery occlusion

underlying factor in heart failure with preserved ejection fraction. *J Clin Hypertens

[Vriens et al., 2005] Vriens, J., Owsianik, G., Fisslthaler, B., Suzuki, M., Janssens, A.,
Voets, T., Morisseau, C., Hammock, B. D., Fleming, I., Busse, R., and Nilius, B.
(2005). Modulation of the ca2 permeable cation channel trpv4 by cytochrome p450

51.

Rundek, T., and Sacco, R. L. (2009a). Novel quantitative trait locus is mapped to
chromosome 12p11 for left ventricular mass in Dominican families: the Family Study


DNA polymorphisms which can be typed using the polymerase chain reaction. *Am J

