Integrated Genome Sequencing and Gene Expression Analysis in The Stroke-Prone Spontaneously Hypertensive Rat

Mohammed Dashti MSc. BSc (Hons)

Thesis submitted for the degree of Doctor of Philosophy to the University of Glasgow

Institute of Cardiovascular and Medical Sciences
College of Medical, Veterinary and Life Sciences
University of Glasgow

September 2014

Thesis Summary

Human hypertension is known to have a large heritable component. However, the identification of the genetic factors is complicated, because the disease is complex and multi-factorial. Human essential hypertension (HTN) affects 25-35% of the adult population in developed countries and it is the leading risk factor for many cardiovascular diseases that are associated with high morbidity and mortality rates worldwide. Linkage analysis and genome-wide association study (GWAS) have been used as strategies to investigate the genetic basis of HTN. Linkage analysis has failed to identify convincing candidate genes for HTN, however, GWAS studies have reported several candidate genes for HTN.

The identification of causative genes for HTN has been facilitated by the use of rats that have been selectively bred for high blood pressure. The stroke-prone spontaneously hypertensive rat ($SHRSP_{Gla}$) is an excellent model for human cardiovascular disease. Previously, a genome-wide scan between $SHRSP_{Gla}$ and normotensive Wistar Kyoto (WKY_{Gla}) rats identified quantitative trait loci (QTL) on chromosomes 2, 3 and 14 for blood pressure, pulse pressure and cardiac mass, respectively. These QTLs were validated by the construction of chromosome 2, 3 and 14 congenic strains.

This research project attempts to identify the genetic determinants of $SHRSP_{Gla}$ phenotypes by using mRNA and micro(mi)RNA expression profiling data, in combination with the genome sequence of the $SHRSP_{Gla}$ and WKY_{Gla}, to facilitate human translational studies for hypertension and vice versa.

In chapter 3, we implemented Partek® batch effect removal tool to improve previously analysed heart transcriptomic data analysis between parental strains and chromosome 14 congenic strains at 3 different ages. The new improved analysis was able to accurately capture (a) the significant expression of osteopontin, a previously identified positional candidate gene for cardiac mass and (b) the phenotype of $WKY.SP_{Gla}14a$ strain using disease analysis of ingenuity pathway analysis (IPA). Furthermore, we integrated salt-challenged renal gene expression profiling data of parental strains from two platforms with chromosome 3 congenic ($SP.WKY_{Gla}3d$) and sub-congenic ($SP.WKY_{Gla}3f$) strains.

This new workflow using Partek® genomics suite, identified 3 novel positional candidate genes for pulse pressure and salt-induced pulse pressure. In addition, cluster analysis was carried out using Biolayout*Express*^{3D} on renal gene expression profiling data between parental strains with chromosome 2 congenic strain (SP.WKY$_{Gla}$2a) at different time points. The WKY$_{Gla}$ expression pattern identified (109) cluster sharing similarities with the SP.WKY$_{Gla}$2a expression pattern when compared to the SHRSP$_{Gla.}$ Cluster 109 includes 4 positional candidate genes for blood pressure; ATPase, class VI, type 11B (*Atp11b*), phosphoinositide-3-kinase, regulatory subunit 1 (*Pik3r1*), glutathione S-transferase mu 7 and glutathione S-transferase mu 1 (*Gstm1*). *Gstm1*, was previously identified as a functional and positional candidate gene for hypertension in the SHRSP$_{Gla}$, providing proof of concept for our analysis. IPA analysis on cluster 109 identified genes and pathways associated with traits observed in the SHRSP$_{Gla}$. Quantitative real-time polymerase chain reaction (qRT-PCR) confirmed the differential expression of *Atp11b* and *Pik3r1* between parental strains and SP.WKY$_{Gla}$2a at 21 weeks of age salt-challenged. In addition, we ran Biolayout*Express*^{3D} cluster analysis across multiple tissues between young (5 weeks of age), adult (16 weeks of age) and mature (21 weeks of age) parental strains. IPA analysis on the identified clusters implicated genes with cardiovascular system function, molecular transport (young, adult and mature animals), nervous system function, and inflammatory response (only at adult and mature animals). *Atp11b*, which is responsible for ion transport, was implicated in all clusters at different hypertensive stages. Guanylate cyclase 1, soluble, alpha 3 (*Gucy1a3*) implicated in cluster 104 at severely hypertensive stage was the only gene that is already known to cause hypertension. qRT-PCR confirmed the differential expression of *Atp11b* in heart and kidney tissues. In this chapter, we were able to identify positional genes for the SHRSP$_{Gla}$ phenotype that were not implicated by our previous workflows and gene expression analysis tools.

In chapter 4, Next Generation Sequencing (NGS) of SHRSP$_{Gla}$ and WKY$_{Gla}$ was carried out, generating 1,163,332 single nucleotide polymorphisms (SNP), 213,130 insertions and deletions of DNA segments (INDEL). A notable feature of the data included uneven distribution and density between SHRSP$_{Gla}$ at 27x and WKY$_{Gla}$ at 26x coverage assembled to Brown Norway (BN) RGSC 3.4. Structural analysis in the SHRSP$_{Gla}$ and WKY$_{Gla}$ identified thousands of large deletions as

well as copy number variation (CNV) of duplication and deletion. Genomic variants were annotated with Ensembl version 69 to prioritise sequence variants based on their predicted effect and to identify genes affected by large structural variants that are unique to each rat strain. Sequence variants between our strains (a) were higher in intergenic/ intronic regions compared to exonic regions including in our QTL regions on chromosomes 2,3 and 14, (b) IPA functional analysis on sequence variants implicate networks, diseases and pathways that regulate cardiovascular disease, and (c) has high sensitivity and specificity in detecting three variants data by different sequencing technologies. The new BN reference improved NGS coverage by at least 0.50x in both strains, detecting additional sequence variants with more defined predicted effect annotation with Ensembl version 71 when compared to BN reference RGSC 3.4. Also we attempted to isolate stroke candidate genes by isolated sequence variants unique to $SHRSP_{Gla}$ compared to non-stroke prone 27 rat genomes against BN RGSC 3.4. $SHRSP_{Gla}$ unique sequence variants were further prioritised by (a) deleterious protein coding sequence, (b) QTLs for blood pressure and stroke from our and other hypertensive rat strains, (c) in-common with published sequence variants of $SHRSP_{Izm}$ strain. Examples of the prioritised sequence variants included (a) a frameshift variant within *Atp11b* caused by a deletion of 28 bp, and (b) a common SNP between SHRSP strains identified in our analysis causing a stop gain in stromal interaction molecule 1 that acts as calcium channel regulator which identified as candidate gene for stroke in $SHRSP_{Izm}$. The latter emphasised the importance of integrating sequence variants from all SHRSP strains, as they share high proportion of SNPs, to prioritise stroke causative variants.

In chapter 5, we investigated the types of genomic variants that are likely to have an influence on the cardiac and renal transcriptomic output between $SHRSP_{Gla}$ and WKY_{Gla} using IPA. Overall, all genomic variants types are likely to correlate with transcriptomic data. Furthermore, we customised commercial microarray gene expression array to our rat strains by removing probes where their target sequence harbour sequence variants to avoid false positive results. The customised Affymetrix analysis of renal transcriptomic data has an impact on overall results when compared to the default analysis. Nevertheless, the customisation of Illumina array showed that true positive genes would be

removed as well. We presented a reductive and integrative strategy called "Swiss cheese" model that integrates proteins coding variants, QTL regions, transcriptomic data between $SHRSP_{Gla}$ and WKY_{Gla} was able to identified positional candidate genes for the $SHRSP_{Gla}$ phenotype. Moreover, we integrated the $SHRSP_{Gla}$ unique sequence with published brain gene profiling data and prioritised 39 stroke candidate genes in the $SHRSP_{Gla}$. The $SHRSP_{Gla}$ was assessed as model for HTN by integrating human GWAS candidate genes for blood pressure with our transcriptomic data, QTLs for blood pressure and sequence variants between from $SHRSP_{Gla}$ and WKY_{Gla}. This demonstrated human GWAS candidate genes for blood pressure are (a) conserved and mapped to rat blood pressure QTLs (b) half of these genes harbour sequence variants and some are differentially expressed in our rat strains tissues. *Gucy1a3-Gucy1b3* were prioritised for translation study as they map to chromosome 2 congenic interval with functional sequence variants and differentially expressed in heart and/or kidney between $SHRSP_{Gla}$ and WKY_{Gla}. We also investigated differences in renal miRNAs in human hypertension and in the $SHRSP_{Gla}$ using NGS technology. Human miRNA studies were able to implicate miRNAs associated with kidney damage. Whereas, in our rat studies, specifically between $SP.WKY_{Gla}2c^*$ versus $SHRSP_{Gla,}$ we implicated miRNAs targeting positional candidate genes such as *Gstm1* and *Gucy1a3*. Finally, we integrated cardiac gene expression profiling from $SHRSP_{Gla}$ and WKY_{Gla} and publically available human heart failure data that led to the translation of SPARC related modular calcium binding 2, cardiac response to stress and wound healing, from our rat strain to human.

In summary, we implemented, improved and integrated high-throughput genomic variants with transcriptomic data that identified and prioritised miRNA, genes, clusters and pathways that may underline pathophysiological processes in the $SHRSP_{Gla}$ that can enhance the current knowledge of human patho-physiological mechanisms of cardiovascular disease and eventually lead to novel therapeutic targets.

Table of Contents

List of Tables

List of Figures

Publications

Original articles

Mohammed Dashti, Klio Maritou, Santosh Atanur, Tim J. Aitman, Anna F. Dominiczak, Rainer Breitling, John D. McClure, Delyth Graham, Martin W. McBride. Whole Genome Next Generation Sequencing and Gene Expression Analysis in the Stroke-Prone Spontaneously Hypertensive Rat. Abstracts from the Cardiovascular Genetics Satellite - the 16th International SHR Symposium, Rome June 18, 2014. High blood pressure & cardiovascular prevention. *journal of the Italian Society of Hypertension*. 2014 June 21, 171-88.

Sampson AK, Mohammed D, Beattie W, Graham D, Kenyon CJ, Al-Dujaili EA, Guryev V, Mcbride MW, Dominiczak AF. Introgressed chromosome 2 quantitative trait loci restores aldosterone regulation and reduces response to salt in the stroke-prone spontaneously hypertensive rat. *Journal of Hypertension*. 2014 October ;32(10):2013-21.

Trude Skogstrand, Sabine Leha, John McClure, Mohammed Dashti, Bjarne M. Iversen (deceased), Delyth Graham, Martin W. McBride, Michael Hultström. Identification of a common molecular pathway in hypertensive renal damage: comparison of rat and human gene expression profiles. *Journal of Hypertension*. 2014 (in press).

Abstracts

Mohammed Dashti, Klio Maritou, Santosh Atanur, Tim J. Aitman, Anna F. Dominiczak, Delyth Graham, John D. McClure, Martin W. McBride. Next-generation Sequencing and Integrative Genomics Analysis in the Stroke-prone Spontaneously Hypertensive Rat. Glasgow Polyomics Launch Symposium, Glasgow, UK. September 2012. (Poster Communication)

Mohammed Dashti, Klio Maritou, Santosh Atanur, Tim J. Aitman, Anna F. Dominiczak, Delyth Graham, Rainer Breitling, John D. McClure, Martin W. McBride. Next-generation Sequencing and Integrative Genomics Analysis in the Stroke-prone Spontaneously Hypertensive and Wistar Kyoto Rats. Rat Genomics and Models, Oxford, UK. December 2012. (Poster Communication)

Mohammed Dashti, Klio Maritou, Santosh Atanur, Tim J. Aitman, Anna F. Dominiczak, Delyth Graham, Rainer Breitling, John D. McClure, Martin W. McBride. A Combined Approach on Integrated Functional Genomics Data To Prioritise Candidate Genes in the Stroke-prone Spontaneously Hypertensive and Wistar Kyoto Rats. Scottish Cardiovascular Forum, University of Strathclyde, Glasgow, UK. February 2013. (Poster Communication)

Mohammed Dashti, Klio Maritou, Santosh Atanur, Tim J. Aitman, Anna F. Dominiczak, Delyth Graham, Rainer Breitling, John D. McClure, Martin W. McBride. Genome Sequencing and Gene Expression Analysis in the Stroke-Prone Spontaneously Hypertensive Rat. Annual Scientific Meeting of the British Hypertension Society. Greenwich University, London, UK. September 2013. (Oral Communication)

Martin W. McBride, Mohammed Dashti, Klio Maritou, Santosh Atanur, Tim J. Aitman, Wendy Beattie, Rainer Breitling, Anna F. Dominiczak, John D. McClure, Delyth Graham. Integrated Genome Sequencing and Gene Expression Analysis In The Stroke-Prone Spontaneously Hypertensive Rat. Joint meeting ESH-ISH Hypertension, Athens, Greece. June 2014 (Oral Communication)

Awards

EURATRANS Short-term travel fellowships - Training on Next generation analysis at Prof Tim Aitman's laboratory, Imperial College, London, UK. (September 2011)

Selected for Wellcome Trust Advanced Course of Functional Genomics & Systems Biology, at the Wellcome Trust Genome Campus, Hinxton, UK. (June 2012)

Selected for Rat Annotation Workshop, at the Wellcome Trust Genome Campus, Hinxton, UK. (March 2013)

Acknowledgements

First of all, I would like to thank my supervisors Dr. Martin W. McBride, Dr. John D. McClure, and Prof. Rainer Breitling for giving me the opportunity to work in their research group and for their guidance and support throughout my PhD project.

I would like to extend my gratitude to my unofficial supervisor Dr. Delyth Graham for her time, vital input, and support during my PhD period.

I would like to express my appreciation to Prof. Tim Aitman for allowing me to visit his laboratory, as well as his group for their hospitality and assistance in the bioinformatic analyses. Also I would like to thank Dr. Paul Barton for allowing us to validate our cardiac candidate genes in their laboratory.

I would like to thank Mrs. Wendy Beattie and Dr. Erin Olson for passing on their laboratory skills. An extended thanks to Dr. Caline Koh-Tan for her assistance and for sharing with me her expertise on congenic animals.

Many thanks to all my office mates and Glasgow family, Louise, Sophia, Scott, Omar, and Zaki. It would not have been the same without you guys. Thanks also to students Samantha, Riya, and Ting for allowing me to assist you in your projects enabling me to develop my teaching abilities.

Finally, I would like to thank my family and friends for their patience and encouragements.

Author's Declaration

I declare that the work presented in this thesis is, to the best of my knowledge and belief, original and my own work, unless otherwise stated in the text. All the rat phenotyping data and laboratory experiments of microarray gene expression arrays among our rat strains, were carried out by the Glasgow Cardiovascular Research Centre staff. Whole genome sequencing of the $SHRSP_{Gla}$ was conducted at Prof. Norbert Huebner's laboratory (Max Delbrück Center for Molecular Medicine, Germany), whereas, the WKY_{Gla} was sequenced at Prof. Tim Aitman's laboratory (Imperial College, London, UK). Small sequence variants calling between our rat strains against the Brown Norway Rat Genome Sequencing Consortium version 3.4, including annotation with Ensembl version 61 as well raw large deletions analysis, was carried out by Mr Santosh Atanur (Prof. Tim Aitman's group). Dr. Prashant Srivastava, (Prof. Tim Aitman's group), identified affected Affymetrix probes by sequence variants between our rat strains. In addition, laboratory experiment of small RNA Next-generation Sequencing in human and our rat strains was carried out at the Glasgow Polyomics facility of the University of Glasgow. This work has not been submitted previously for a higher degree. It was carried out under the supervision of Dr. Martin W. McBride and Dr. John D. McClure in the Institute of Cardiovascular and Medical Sciences at the British Heart Foundation Glasgow Cardiovascular Research Centre.

Mohammed Dashti
September 2014

Abbreviations

ANOVA	Analysis of variance
ACE	Angiotensin converting enzyme
BN	Brown Norway
CVD	Cardiovascular disease
cDNA	Complementary DNA
CNV	Copy number variation
ENCODE	Encyclopedia of DNA elements
FDR	False discovery rate
GWAS	Genome-wide association study
GRA	Glucocorticoid-remediable aldosteronism
HTN	Human essential hypertension
IPA	Ingenuity pathway analysis
IPKB	Ingenuity pathways knowledge base
INDEL	Insertions and deletions of DNA segments
IGV	Integrative genomic viewer
LVH	Left ventricular hypertrophy
LVMI	Left ventricular mass index
LOD	Logarithm of the odds
mmHg	Measured in millimetres of mercury
NGS	Next generation sequencing
PCA	Principal component analysis
qRT-PCR	Quantitative reverse transcription-PCR
QTL	Quantitative trait loci
RP	Rank Products
RGD	Rat genome database
RI	Recombinant inbred
RAAS	Renin angiotensin aldosterone system
RMA	Robust multi-array average
SNP	Single nucleotide polymorphisms
SIFT	Sorting intolerant from tolerant
SHR	Spontaneously hypertensive rat
SHRSP	Spontaneously hypertensive rat stroke-prone

1 Introduction

1.1 Cardiovascular Disease

According to the World Health Organization, cardiovascular disease (CVD) is the major leading cause of death in world accounting for 30% (17.3 million people) of worldwide deaths in 2008 (WHO, 2012). CVD encompasses a range of pathological conditions that affects cardiovascular system components; heart and blood vessels (Nichols et al., 2013). Coronary heart disease and cerebrovascular disease (stroke) are the main forms of CVD, representing the world's first (7.3 million people) and second (6.2 million people) most common causes of death, respectively (WHO, 2012). A number of controllable risk factors, including environmental, life style choices, and medical conditions, have been identified for developing CVD, and vary from poor nutrition, physical inactivity, smoking, high alcohol consumption, air pollution, obesity, to high blood pressure/cholesterol, and diabetes (Yusuf et al., 2004; Yusuf et al., 2014). Ageing, sex, family history, and ethnicity have been identified as non-controllable risk factors for developing CVD (Roger et al., 2012). During the last 30 years, intervention and prevention of CVD risk factors have led to rapid reduction of the premature mortality rates in developed countries including the UK (Nichols et al., 2013).

1.1.1 Blood Pressure

Blood is a bodily fluid that delivers important substances such as oxygen, nutrients and hormones to all cells, tissues, organs and transports toxic metabolic waste products such as carbon dioxide away from those cells, tissues and organs. The force exerted by the blood flow within blood vessels is called blood pressure. Blood pressure is usually measured in millimetres of mercury (mmHg) and recorded as a fraction with the maximum pressure following systole of the left ventricle of the heart over the minimum pressure that accompanies diastole- hence, it varies depending on the cardiac output and the total vascular resistance. Blood pressure is also classified as systolic, which is the highest measurement of the pressure in arteries during heart contraction, whereas, diastolic blood pressure is the lowest measurement of pressure in arties when the heart relaxes (Strandberg & Pitkala, 2003). Pulse pressure is the difference between systolic and diastolic blood pressure measurements (Sesso et al., 2000).

Blood pressure readings vary between normal healthy individuals, however, an ideal pressure level in adults is suggested to be 120 systolic over 80 diastolic mmHg (Chobanian et al., 2003; Frese et al., 2011).

1.1.2 Regulation of Blood Pressure

The transportation of various molecules, such as gases, nutrients and metabolism by-products across cells and organ tissues within an organism, is performed via circulation. This makes blood circulation a function indispensible for life maintenance, and therefore, its regulation has to be tightly controlled. Several physiological factors partake in the regulation of blood pressure; renal, neural, vascular and endocrine systems, in combination with homeostatic mechanisms, interplay in order to respond to changes induced by environmental factors, including diet, stress and physical activity. Cardiovascular homeostasis can be interrupted by common factors, such as change in posture, or increased physical activity. If such changes are not compensated, they can result in reduced blood pressure, and in severe cases, reduced oxygen delivery to vital organs such as the heart and the brain.

Normal blood pressure is regulated and maintained by a number of interacting physiological mechanisms involving neural, vascular, endocrine, and renal systems (Figure 1-1). Sudden changes in blood pressure are detected by the sympathetic and the parasympathetic nervous systems, offering a rapid response to fluctuations in blood pressure. The baroreceptor reflex system plays a vital role in the compensation of rapid blood pressure changes (Stauss, 2002). Baroreceptors can be located at the carotid sinus, the aortic arch, and the right atrium, and they can have both circulatory and renal effects. For instance, when the blood pressure is high, baroreceptors inhibit cardiac, renal and vasomotor sympathetic afferents to restore normal blood pressure. Baroreceptors have the opposite effect when a sudden drop in blood pressure is detected (Reviewed by Guyenet, 2006). Furthermore, stimulation of the sympathetic nervous system induces release of adrenaline and noradrenaline from the adrenal, causing increased flow resistance in the vessels (and therefore, increased blood pressure) via α – receptors signalling (Reid, 1986; Guyenet, 2006).

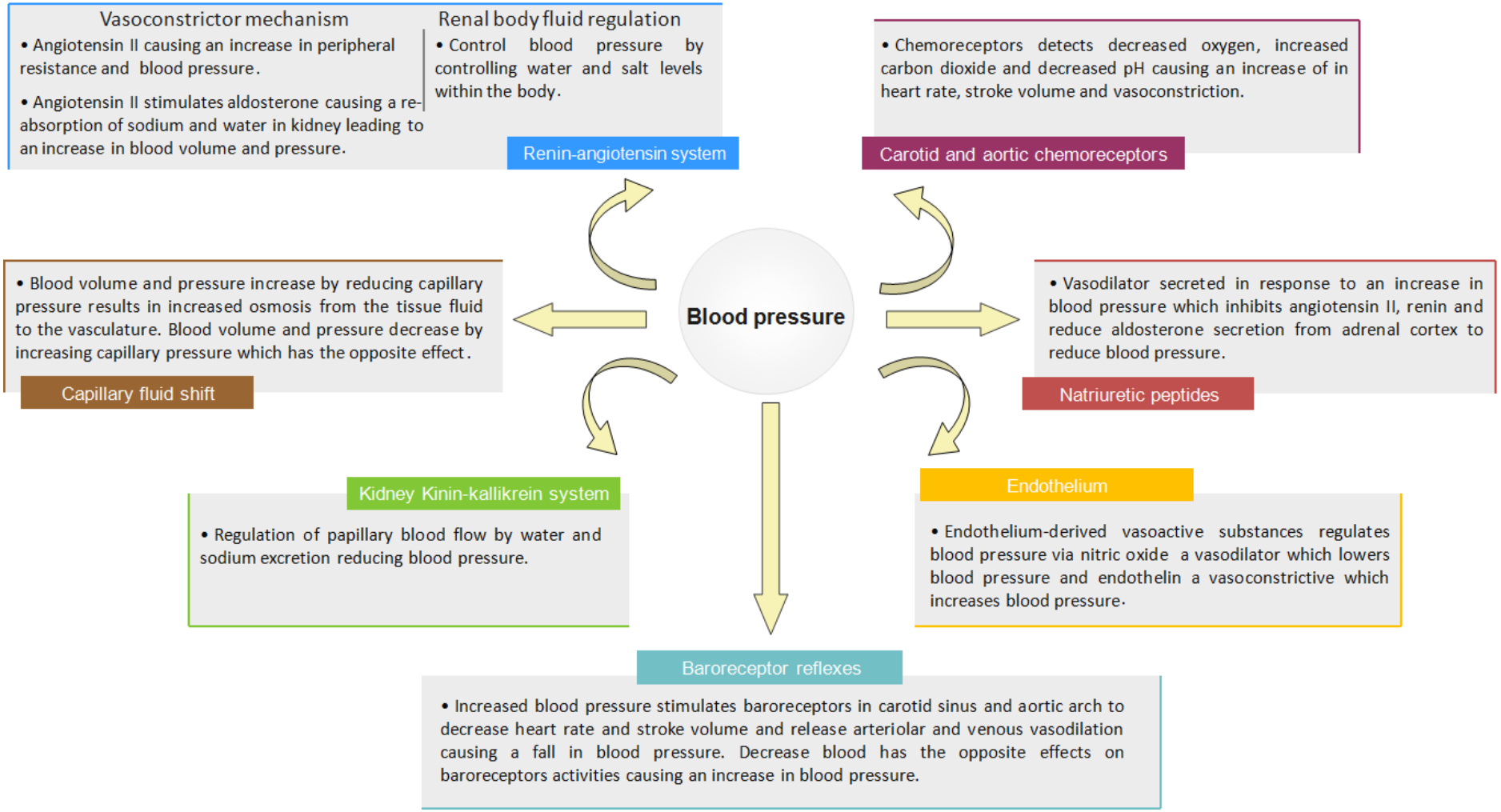


Figure 1-1 Physiological regulation of blood pressure.
Schematic summarising the principle physiological mechanisms of RAAS, capillary fluid shift, Kinin-kalikrein system, baroreceptor reflexes, carotid and aortic chemoreceptors, endothelium, and natriuretic peptides that all regulate and maintain blood pressure.

Chemoreceptors that monitor the levels of oxygen and carbon dioxide (via pH sensing) also play a role in blood pressure regulation and are located close to the carotid sinus and aortic arch (Fletcher et al., 1992).

Endothelin (ET), a 21 amino acid peptide that is produced by the vascular endothelium from a 39 amino acid precursor, can also regulate blood pressure and vascular tone by maintaining the balance between vasoconstriction (e.g endothelin) and vasodilation (e.g endothelium-derived relaxing factor such as Nitric Oxide) (Kohan et al., 2011).

Another endogenous blood pressure regulation mechanism is regulated by the release of aldosterone. In response to angiotensin II (see paragraph 1.1.2.1 for more details), or elevated potassium levels within the serum, aldosterone is released by the adrenal cortex, inducing (a) the retention of sodium, and (b) the excretion of potassium, via the renal system. This osmotic change leads to increased fluid retention, and therefore, increased blood pressure (Freel & Connell, 2004).

Finally, the capillary fluid shift mechanism can also regulate blood pressure by exchanging blood and fluids across capillary membranes.

1.1.2.1 Kidney blood pressure regulation; the renin angiotensin aldosterone system (RAAS)

Kidneys are considered the most important organ in regulating long term blood pressure. They regulate body fluid volume and osmolality by removing excess salt (such as sodium, potassium and chloride) and water out of the blood mass, and turn them into urine (Guyton, 1991). The most important mechanism by which kidneys exert the long-term control of blood pressure and volume homeostasis is the renin angiotensin aldosterone system (RAAS) (Hall, 1991).

RAAS is one of the major regulators of blood pressure (Figure 1-2). RAAS hormonal cascade begins with renin secreted by juxtaglomerular kidney in response to low blood volume and/or a decrease in blood pressure (Guyton, 1991). Renin in the blood stream cleavesangiotensinogen (secreted by the liver)

to form angiotensin I (Reid, 1998). Angiotensin I is then converted to angiotensin II, a vasoconstrictor and major active product of the RAAS, by angiotensin converting enzyme (ACE) that is found mainly within the lungs (Cambien et al., 1992). Angiotensin II will increase blood pressure and stimulate aldosterone secretion from the adrenal glands causing renal tubular re-absorption of sodium and chloride, water retention and vascular contractions. This action is mediated by angiotensin II binding to AT1 receptors, whereas the interaction between angiotensin II and AT2 receptor have the opposite effect stimulating vasodilation. Angiotensin 1-7 peptide is another vasodilator with cardioprotective effects generated by ACE homologue (ACE 2) and probably ACE as well either directly via angiotensin II or indirectly via Angiotensin 1-9 peptide (Reviewed by Re, 2004; Schindler et al., 2007).

A number of research groups are investigating the utilisation of the function of angiotensin II in RAAS system to stimulate hypertension and hypertrophy in animal models to identify and validate therapeutic targets. More specifically, the Nicklin group focuses on *in vitro* function of angiotensin 1-7 and 1-9, as well as the in vivo effect of the peptides on myocardial infarction and in hypertension mouse models via gene editing or direct peptide infusion in order to explore novel therapeutic options (Flores-Munoz et al., 2012; Clarke et al., 2013).

Finally, the adrenal glands also partake in blood pressure regulation via the secretion of epinephrine and norepinephrine. These hormones cause vasoconstriction of arteries and veins, increasing heart rate and as a consequence, blood pressure.

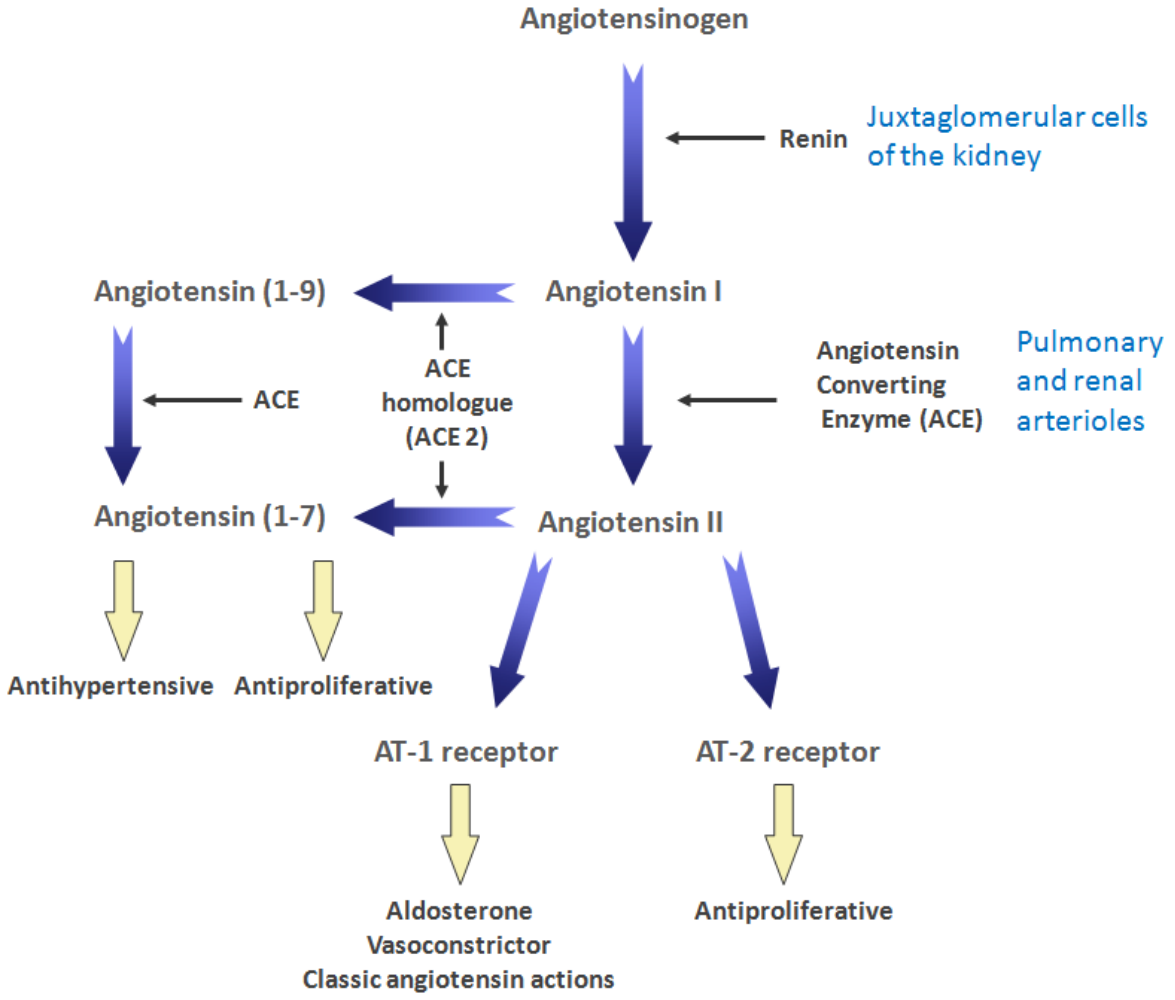


Figure 1-2 The renin angiotensin aldosterone system.
RAAS hormonal cascades showing the relationship between renin, angiotensin peptides and converting enzymes and their role in regulation of blood pressure (Adapted from Re, 2004).

1.2 Essential Hypertension

Human essential hypertension (HTN) is the leading cause for CVD; it is characterised by chronically elevated arterial blood pressure of an unknown cause and is associated with high morbidity and mortality rates (Caulfield et al., 2003). A study by Lawes et al., 2008 found several attributes to high blood pressure worldwide, including approximately 13.5% of premature deaths, 47% of ischaemic heart disease, and 54% of stroke. During recent decades, HTN prevalence has increased dramatically; according to the American Heart Association, about 77.9 million (1 out of every 3) adults have high blood pressure in the USA (Roger et al., 2012), while in Europe it varies among countries from 25-35% of the adult population (Ruppert & Maisch, 2003; Joffres et al., 2013). HTN is diagnosed when average blood pressure readings for systolic and diastolic are 140/90 mmHg, respectively (Table 1).

HTN is a multi-factorial disease in which multiple genes and environmental factors contribute; resulting in a complex, multi-factorial disease in which the genetic parameters have not been clarified and the genetic susceptibility variants have not been determined (Ruppert & Maisch, 2003). HTN accounts for approximately 90-95% of all cases of hypertension, whereas secondary hypertension, which is caused by identifiable secondary pathological conditions, accounts for approximately 5-10% of all cases of hypertension (Faselis et al., 2011).

Table 1-1 Blood pressure classification for adults.

Classification	Systolic pressure mmHg	Diastolic pressure mmHg
Normal	90-119	<80
Pre-hypertensive	120-139	80-90
Hypertension Stage 1	140-159	90-99
Hypertension Stage 2	≥160	≥100

(Table adapted from Chobanian et al., 2003)

1.2.1 Environmental Risk Factors

A number of inter-related environmental factors have been found to be responsible for 80% of hypertension (Beilin, 2004). These include obesity (calorie excess), insulin resistance, high alcohol consumption, physical inactivity, mental stress/depression, cigarette smoking, caffeine intake, vitamin D deficiency, low potassium intake, low calcium intake, and high salt intake (in salt-sensitive patients).

The prevalence of obesity-associated hypertension in United States of America among adults is 60%-70% (Must et al., 1999). A strong association has been observed between increased body mass index and the frequency of hypertension within the middle-aged population (Timpson et al., 2008). This might be due to the fact that body weight and blood volume are correlated; an increase in body weight will increase the cardiac output (for more blood) thus, increasing the pressure on blood vessels. Nevertheless, the mechanism in which obesity can lead to hypertension is by the activation of the sympathetic nervous system and RAAS (Rahmouni et al., 2005). Furthermore, obesity can cause insulin resistance, a condition where cells fail to response to normal insulin (a hormone with vasodilatory properties), which can cause hypertension (Salvetti et al., 1993). In addition, insulin resistance can be caused by vitamin D deficiency that also predispose to hypertension by inhibiting renin, which subsequently activates RAAS (Lee, 2008).

1.2.1.1 Salt Sensitivity

Salt sensitivity is defined as ≥10 mmHg rise in blood pressure during salt-loading, which is then expected to reduce, compared to a salt-depleted state. High salt intake is a major risk factor for HTN, and the rest of CVDs. An International Study of Salt and Blood Pressure (INTERSALT) examined urinary sodium excretion level with blood pressure. The study found a significant positive association between increased sodium excretion and blood pressure (Stamler et al., 1989). Studies on hypertension prevalence among populations have observed that nations with natural food diets that are low in sodium have a lower rate of hypertension, compared to industrialised nations that mainly consume processed

food that is high in sodium (Adrogue & Madias, 2007). These findings are further supported by observations from dietary intervention approaches in hypertensive patients, in which reduction in sodium intake has been found to lower blood pressure (Sacks et al., 2001). The possible ways high sodium intake can induce hypertension includes sodium retention by impairing the ability to excrete sodium and to impact the RAAS which suppresses angiotensin II function leading to an increase in cardiac output and pressure on arteries walls (Drenjancevic-Peric et al., 2011).

A study from Ji et al., 2008, following screening for variations in three genes: *SLC12A3* (*NCCT*), *SLC12A1* (*NKCC2*) and *KCNJ1* (*ROMK*) from members of the Framingham Heart Study, demonstrated that many rare alleles that affect kidney sodium handling are involved in blood pressure variation. They also identified new alleles that benefit health, however, they are under purifying selection.

Animal studies have identified candidate genes for salt-sensitive hypertension in Dahl salt-sensitive rats using zinc-finger nuclease technology. For example Jin et al., 2014 provided evidence that voltage-gated proton channel, HV1, acts as a sodium sensor and promotes superoxide production in medullary thick ascending limb of Dahl salt-sensitive rats (Jin et al., 2014).

1.2.2 Genetic risk factors

The genetic determinants of HTN have been estimated to be 30% - 50% of the blood pressure variation among individuals (Lifton, 1996; Marteau et al., 2005). Family and twin studies indicate that the heritability of systolic blood pressure is approximately between 15 to 40%, while that of diastolic blood pressure is approximately 15 to 30% (Feinleib et al., 1977; Mongeau et al., 1986; Staessen et al., 2003). In addition, hypertension sibling recurrent risk (a measure of familial aggregation of hypertension) for a person who has an older sibling with hypertension are 1.2-1.5 times more likely to be diagnosed with hypertension than the general population (Caulfield et al., 2003).

HTN is rarely due to monogenic mutations inherited by Mendelian forms; it is widely believed that it results from the inheritance of a number of susceptibility

genes, in addition to the involvement of multiple environmental determinants (Marteau et al., 2005).

1.2.2.1 Mendelian forms of inheritance

A number of monogenic or Mendelian forms of hypertension have been identified, in which most share a common pathway of increased distal tubular re-absorption of sodium and chloride and volume expansion that suppresses plasma renin activity (Garovic et al., 2006).

1.2.2.1.1 Glucocorticoid-Remediable Aldosteronism

Glucocorticoid-remediable aldosteronism is an autosomal dominant disorder characterised by an early onset of hypertension with normal or elevated aldosterone levels. This is the result of chimeric gene duplication from unequal cross over between 11b-hydroxylase (*CYP11B1*) and aldosterone synthase gene (*CYP11B2*), causing aldosterone synthase to be expressed in the adrenal gland under the regulation adrenocorticotropin. This eventually increases both sodium re-absorption and potassium secretion (Lifton et al., 1992; Simonetti et al., 2012).

1.2.2.1.2 Apparent Mineralocorticoid Excess

Apparent mineralocorticoid excess is an autosomal recessive disorder characterised by an early onset of hypertension with hypokalaemia, metabolic alkalosis, low renin and aldosterone. This is a result of a mutated 11 ß - hydroxysteroid dehydrogenase type 2 enzyme causing high concentrations of cortisol in the kidney that activates the mineralocorticoid receptor resulting in aldosterone like effect and leads to hypertension (Wilson et al., 1995; Simonetti et al., 2012).

1.2.2.1.3 Liddle Syndrome

Liddle syndrome is an autosomal dominant disorder, characterised by an early onset of hypertension, with hypokalaemia, metabolic alkalosis, and low levels of renin and aldosterone. It is believed that this phenotype is the a result of mutated sodium channel, non-voltage-gated 1, beta and/ or gamma subunit

causing an increased epithelial sodium channel mediated sodium re-absorption leading to hypertension (Shimkets et al., 1994; Hansson et al., 1995; Snyder et al., 1995; Simonetti et al., 2012).

1.2.2.1.4 Hypertension with Brachydactyly

Hypertension with brachydactyly is an autosomal dominant disorder, characterised by shortened finger bones with hypertension that develops and progress into severe hypertension that closely resembles HTN. It has been suggested that a structural variation within chromosome 12 between p12.2-p11.2 region results in impaired baroreceptor reflex function of blood pressure buffering that may leads to hypertension (Schuster et al., 1996; Jordan et al., 2000; Pereda et al., 2013).

1.2.2.1.5 Pseudohypoaldosteronism type II

Pseudohypoaldosteronism type II, also known as Gordon's syndrome, is an autosomal dominant disorder characterised by hypertension, hyperkalaemia, hyperchloraemia, and metabolic acidosis. This is the result of mutation in either one of the two serine-threonine kinase genes, *WNK1* and/or *WNK4*, causing an increase in sodium re-absorption resulting in hypertension (Wilson et al., 2001). Boyden et al., 2012 showed that kelch-like 3 and cullin 3 are also causative genes for Gordon's syndrome.

1.2.2.2 Identifying candidate genes associated with HTN

Identifying candidate genes in multi-factorial complex diseases, such as HTN, is much more difficult when compared to monogenic disorders. Linkage analysis and genome-wide association study (GWAS) are the main approaches used to identify candidate genes in HTN.

Linkage analysis is used to locate quantitative trait loci (QTLs) when the disease causing genes are unknown. This is done by using known polymorphic markers and linking them to the disease in family studies. Any genetic marker can be used including single nucleotide polymorphism (SNP) and/or microsatellites (nucleotide repeats within the genome).

One of the notable linkage studies to identify QTLs for HTN in large families with affected members, has been carried out by the Medical Research Council BRItish Genetics of HyperTension (BRIGHT). Genome wide scans of 3599 individuals identified a significant QTL for HTN on chromosome 6q with a logarithm of the odds (LOD) score of 3.21 (>3.0 is usually used as cut-off for significant linkage) and genome-wide significance (p=0.042), and other suggestive QTLs on chromosomes 2q, 5q and 9q with LOD scores higher than 1.57 (but lower than 3.0) with genome-wide significance (p=0.017) when assessed under a locus-counting strategy (Caulfield et al., 2003). When the number of affected participants were increased in follow-up studies, a significant QTL on chromosome 5q13 for HTN and novel suggestive QTLs were identified (Munroe et al., 2006). Binder (2007), reviewed a number of linkage studies and showed that QTLs associated with HTN were implicated in all autosomal chromosomes. Nevertheless, QTLs associated with hypertension on chromosomes 1, 2, 3, 17 and 18 were reported by at least two linkages studies (Binder, 2007). Recently, research groups are moving away from linkage analysis due to (a) usually the identified QTL regions are large and contain many genes which are more laborious to investigate (poor genetic resolution in the centimorgan range), (b) the detecting power of linkage analysis in common alleles that have low penetrance is poor (Padmanabhan et al., 2008), (c) inappropriate marker density in linkage analysis studies can lead to loss of information (McBride et al., 2006), and d) genetic recombination may result in misleading statistical probability.

GWAS is a more preferred approach to identify polygenic disease susceptibility genes by comparing large number of SNP or microsatellite markers across the whole genome between patients and controls in a population (no need for family studies) (Ruppert & Maisch, 2003). GWAS using a large number of SNPs by the Wellcome Trust Case Control Consortium (WTCCC) failed to identify convincing candidate genes for HTN (Burton et al., 2007). This might be due to (a) GWAS principle is based on the idea that a single gene is the causative reason of the phenotype; whereas, HTN involves multiple complex interactions of genes, gene-environment (Ruppert & Maisch, 2003), and ethnic diversity in populations (Padmanabhan et al., 2008), (b) it has been suggested that the standards that define the controls could be more strict , in order to increase the phenotypic distance between the "hypercontrols" and the affected individuals, increasing

the chance of finding genetics association (Delles et al., 2008a), (c) GWAS demonstrate a small portion of the total heritability, and in order to be used for a successful study the samples should be large, carefully dissected phenotype and from homogeneous populations (Amos, 2007).

Successful GWAS studies were reviewed in Simino et al., 2012 in which more than 50 genes were identified as candidate genes for blood pressure and HTN. Some of the published GWAS data were combined to create a more powerful meta-analysis to prioritise candidate genes for blood pressure and HTN. Meta-analysis was performed using the the Global Blood Pressure Genetics (Global BPgen) consortium study, which took into account data from 34,433 subjects (Newton-Cheh et al., 2009), with Cohorts for Heart and Aging Research in Genome Epidemiology (CHARGE) consortium in 29,136 subjects (Levy et al., 2009) revealed genome-wide significance ($P < 5 \times 10^{-8}$) for systolic, diastolic blood pressure, and hypertension (Table 1-2). Moreover, GWAS studies have good genetic resolution (usually in the kilobase range) in which the SNPs associated with HTN are next to a few genes or sometime within a gene giving an insight into possible pathogenesis pathways based on known function. For instance, Padmanabhan et al., 2012 predicted possible pathways of the GWAS candidate genes for blood pressure and HTN based on their known function, which is helpful in uncovering novel mechanisms (Table 1-3). It is essential to note that although human GWAS studies for hypertension demonstrated blood pressure susceptibility genes, nevertheless, the associated SNPs have modest effect on blood pressure (< 1 mmHg), explaining a small fraction (approximately 3%) of blood pressure heritability within populations, and fall mainly outside the gene coding regions (Hindorff et al., 2009).

Table 1-2 Meta-analysis of Global BPgen and CHARGE of significant Loci for Systolic and Diastolic Blood Pressure and Hypertension.

SNP	Gene	Blood pressure	CHARGE *P*-value	Global BPgen *P*-value	CHARGE + Global BPgen *P*-value
rs1004467	CYP17A1	Systolic	1.99E-06	1.08E-05	1.28E-10
rs381815	PLEKHA7	Systolic	5.76E-07	2.72E-04	1.89E-09
rs2681492	ATP2B1	Systolic	3.01E-11	4.07E-03	3.76E-11
rs3184504	SH2B3	Systolic	5.73E-07	6.36E-04	4.52E-09
rs9815354	ULK4	Diastolic	7.88E-07	3.79E-04	2.54E-09
rs11014166	CACNB2	Diastolic	8.82E-07	1.46E-03	1.24E-08
rs2681472	ATP2B1	Diastolic	3.74E-08	2.43E-03	1.47E-09
rs3184504	SH2B3	Diastolic	1.68E-08	2.83E-07	2.58E-14
rs2384550	TBX3/TBX5	Diastolic	1.32E-07	1.06E-02	3.75E-08
rs6495122	CSK/ULK3	Diastolic	8.10E-07	3.98E-05	1.84E-10
rs2681472	ATP2B1	Hypertension	1.65E-08	2.15E-04	1.75E-11

(Table adapted from Levy et al., 2009)

Table 1-3 Possible pathways of GWAS candidate genes for blood pressure and HTN.

Chromosome	Gene/ nearest gene	GWAS study	Possible pathway	Known function
10p12.3	CACNB2	Levy D 2009; Ehret et al., 2011	Vascular/cardiac function?	Subunit of voltage-gated calcium channel expressed in heart.
11p15.5	LSP1/TNNT3	Johnson et al., 2011	Endothelial function?	Expressed in leukocytes and endothelial cells. Involved in signaling, regulating the cytoskeletal architecture and neutrophil migration.
12q21.3	ATP2B1	Levy D 2009; Newton-Cheh et al., 2009; Ehret et al., 2011, Johnson et al., 2011, Kato et al., 2011	Vascular function?	Encodes plasma membrane calcium- or calmodulin-dependent ATPase expressed in endothelium.
12q24.1	SH2B3	Levy D 2009; Newton-Cheh et al., 2009; Ehret et al., 2011	Endothelial function?	Also known as lymphocyte-specific adaptor protein (LNK), may regulate hematopoietic progenitors and inflammatory signaling pathways in endothelium.

(Table adapted from Padmanabhan et al., 2012)

1.2.3 Endothelial Dysfunction in HTN

Endothelial dysfunction can be the cause, as well as consequence of HTN. As a result, elucidating the underlying mechanisms in endothelial dysfunction is necessary. NO production and bioavailability by vascular endothelium is critical for blood flow regulation and the vasodilation (Mensah, 2007). NO is formed by enzymatic action of endothelial NO synthase (eNOS), although, uncoupled eNOS, as well as nicotinamide adenine dinucleotide phosphate-oxidase (NADPH), and xanthine oxidases, are important sources for reactive oxygen species (ROS). These are chemically reactive molecules containing oxygen produced as by-products of normal physiological oxidation/reduction (redox) reactions, which are involved in cell signalling and homeostasis (Hancock et al., 2001). Under normal conditions, the rates of ROS production of free radicals and normal cellular defences are balanced. However, imbalances in ROS production, called oxidative stress, can lead to changes in vascular structure and function. In HTN, a number of ROS-dependent mechanisms for hypertension have been identified (Endemann & Schiffrin, 2004). These include (a) decreased NO bioavailability and/or vascular tone; NO reacts with superoxide producing peroxynitrite that has lower vasodilator effect than NO and will also increase ROS production, (b) vascular inflammation; Increase in inflammatory cytokines (e.g. interleukin-6 (IL-6), tumor necrosis factor alpha (TNF-α) influences the adhesion molecules (e.g. intercellular adhesion molecule (ICAM)-1 and vascular cell adhesion molecule-1 (VCAM-1) in a ROS-sensitive (pathological) manner leading to increased endothelial layer permeability which will inhibit normal endothelial function. This leads to the synthesis of pro-inflammatory and pro-thrombotic factors and increased endothelin-1 secretion that will also increase ROS generation (c) vascular remodelling; NO bioavailability and vascular inflammation over time, which will eventually lead to vascular stiffness (remodelling) (Reviewed by Delles et al., 2008b; Montezano & Touyz, 2011).

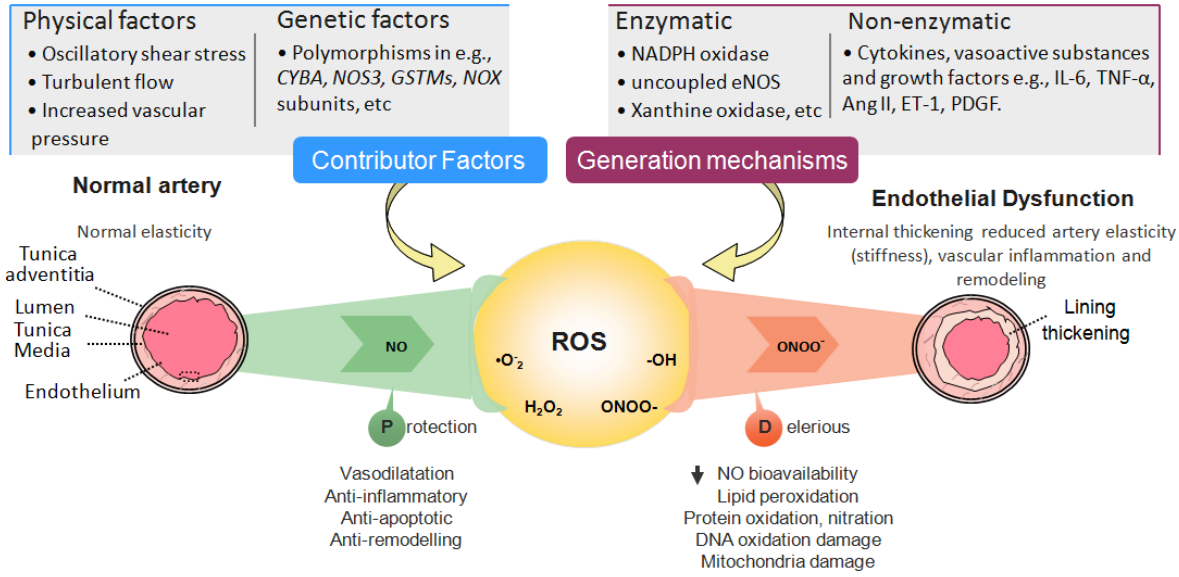


Figure 1-3 ROS sources and oxidative stress effects on the vasculature.
ROS regulates a number of important cellular functions and generated by enzymatic (such as NADPH) and non-enzymatic (such as angiotensin II (AngII), endothelin-1 (ET-1) pathways due to physical factors and/or genetic factors. The former includes changes in the blood flaw and the latter due to polymorphisms of oxidative stress related genes such as NADPH oxidases (NOX), nitric oxide synthase 3 (NOS3), and etc. When the ROS generation exceeds endogenous antioxidant mechanism (oxidative stress) will result in a number of deleterious changes, mainly loss of NO bioavailability and vascular inflammation (Original diagram constructed from information within Delles et al., 2008b; Montezano & Touyz, 2011).

1.2.4 Treatments for hypertension

Findings from epidemiological, prevalence and intervention studies of environmental risks factors associated with hypertension (Stamler et al., 1989; Salvetti et al., 1993; Must et al., 1999; Sacks et al., 2001; Beilin, 2004; Adrogue & Madias, 2007; Lee, 2008; Timpson et al., 2009) suggest that better life style choices, such as good diet, exercise, and reasonable consumption of alcohol, tobacco and caffeine, can lead to reduction in blood pressure reducing CVD risk. Nevertheless, pharmacologic therapy is required for the management of blood pressure in hypertensive patients.

Recently, panel members appointed to the Eighth Joint National Committee (JNC 8) with expertise in hypertension, cardiology, pharmacology, clinical trials, evidence-based medicine, informatics, and development, and implemented clinical guidelines in systems of care released 2014 evidence-based guidelines for the management of high blood pressure in adults (James et al., 2014). Although, other management guidelines for the pharmacological management of hypertension have been already into practice, such as the collaboration between the British Hypertension Society and The National Institute for Health and Care Excellence (NICE) in 2013 (NICE, 2013). Drugs listed in Table 1-4 are for hypertensive patients with varying ages and ethnicity groups and include complications from other diseases from two management guidelines. Although, different guidelines have different strategy in managing hypertension, the first in-line drug usually used to treat high blood pressure is a thiazide diuretic, which increase sodium and water excretion by the kidneys. Diuretic drugs show a consistent drop in blood pressure in patients with an African origin, which may be due to increased susceptibility of salt sensitivity putting hypertensive patients with African origin at risk to develop hypokalaemia (Gibbs et al., 1999). Therefore, angiotensin receptor blocker, or ACE inhibitor, is prescribed instead, Both reduce the production of angiotensin II allowing for blood vessels to widen thus reduce blood pressure, which is also beneficial for hypertensive patients with diabetes and chronic kidney disease (Molnar et al., 2014). Severely hypertensive patients (with stage 2 hypertension) might require drug combinations therapy to achieve normal blood pressure control (James et al.,

2014). However, in all cases, current anti-hypertensive drugs therapy do not provide permanent (long-term) effect on lowering blood pressure.

Table 1-4 Hypertensive guidelines for drug therapy in adults.

	Population	Blood Pressure mmHg	Initial drug treatment options
JNC 8	General ≥ 60 years	<150/90	Non-black: thiazide-type diuretic, angiotensin converting enzyme inhibitor, angiotensin receptor blocker or calcium channel blocker Black: Angiotensin converting enzyme inhibitor or calcium channel blocker
	General < 60 years	<140/90	
	Diabetes	<140/90	
	Chronic kidney disease	<140/90	Angiotensin converting enzyme inhibitor or calcium channel blocker
NICE	≥ 55 years or black of any age	<150/90	Step 1: Calcium channel blocker Step 2: Combine treatments of angiotensin converting enzyme inhibitor or angiotensin receptor blocker with calcium channel blocker Step 3: Combine step 2 with Thiazide like diuretic Step 4: Resistant hypertension, follow step 3 and consider the addition of other diuretics
	General ≤ 55 years	<140/90	Step 1: Angiotensin converting enzyme inhibitor, or angiotensin receptor blocker Step 2: Combine treatments of angiotensin converting enzyme inhibitor or angiotensin receptor blocker with calcium channel blocker Step 3: Combine step 2 with Thiazide like diuretic Step 4: Resistant hypertension, follow step 3 and consider the addition of other diuretics

(Table adapted from NICE, 2013; James et al., 2014)

1.2.5 HTN end-organ damage

When HTN is not treated it can cause end-organ damage, including left ventricular hypertrophy and stroke.

1.2.5.1 Left ventricular hypertrophy

Chronically elevated cardiac workload can lead to thickening of the left ventricle myocardium leading to left ventricular hypertrophy (LVH). This cardiac remodelling can be caused by (a) physiological factors; such as reaction of exercise/physical training (athletic hypertrophy), which is reversible (b) pathological factors; high blood pressure (hypertension), aortic valve stenosis (does not open fully) and insufficiency (does not close tightly), and/or by underlying mechanisms in patients with dilated cardiomyopathy (by increased filling of the heart left ventricle) and coronary artery disease (as the myocardium compensates for the ischemic or infarcted tissue) (reviewed by Bauml, 2010). Furthermore, genetics, age, sex, race, body mass index and stimulation of sympathetic nervous systems and RAAS are factors havecrucial roles in the pathogenesis of LVH (Gradman & Alfayoumi, 2006). Left ventricular mass index (LVMI) measurement, is calculated by measuring left ventricular mass corrected for body size, with specific cut-off levels to diagnose LVH (Reviewed by Bauml, 2010). LVH is not a disease *per se*; however, it is an important predictor for hypertrophic cardiomyopathies, heart failure, stroke that can lead to sudden death (Gradman & Alfayoumi, 2006). Therefore, understanding the genetic influence and developmental mechanisms of LVH is essential.

One of the important components in the development of LVH is the initiation of myocardial fibrosis (abnormal excess of fibroblasts cells in the cardiac muscle) in response to increased stress on the left ventricular wall by stimulating cardiomyocytes, collagen formation and fibroblasts. The fibrosis increases myocardial mass, lowering the blood volume in the left ventricular during diastole (diastolic dysfunction), and eventually, impairing the heart to relax between contractions (Kahan & Bergfeldt, 2005). The development of LVH has been associated with the sympathetic nervous system, and RAAS via norepinephrine and angiotensin II, respectively, where both (a) to be increased in LVF patients effect increased in LVH patients (Gonzalez et al., 2002; Schlaich

et al., 2003) (b) enhances fibrosis mediated by transforming growth factor ß1 in cardiac fibroblasts (Akiyama-Uchida et al., 2002; Gonzalez et al., 2002). A number of hypertrophic cardiomyopathy causing genes that encodes sarcomeric proteins (basic unit of muscle) were identified; nevertheless, the underlying mechanism in which these genes cause LVH is not known yet (Maron, 2002).

1.2.5.2 Stroke

Stroke is defined as the interruption of blood supply to the brain, causing permanent brain damage that can lead to mental and physical disability, and even sudden death. Stroke can be the result of a rupture of a blood vessel (haemorrhagic stroke) and/or a blockage in an artery by clots (ischaemic stroke) (Alqadri et al., 2013). Hypertension is one the main risk factors for ischaemic stroke, which is the second leading, cause of death (WHO, 2012). This suggests the importance of understanding the genetic controls and mechanisms of stroke development. Human GWAS were successful in identifying a few candidate genes that may cause stroke; however, the implicated genes have not yet been functionally characterised (Fornage et al., 2011; Traylor et al., 2012). Animal studies show that angiotensin II receptors (AT1 receptor and AT2 receptor) are involved in the pathogenesis (increasing oxidative stress as a result of AT1 receptor pathway) (Wakisaka et al., 2010), and neuroprotection (suppressing the oxidative stress as a result of AT2 receptor pathway) in stroke (Iwai et al., 2004). Currently, angiotensin II receptor blockers and ACE inhibitors are used to prevent stroke in patients with and without history of stroke, as well as to lower blood pressure (Dahlof, 2009).

1.2.6 Animal models

Animal models have been selectively inbred over several generations to study human CVD, including HTN, in order to identify the genetics determinants and patho-physiological pathways is vital. This is due to (a) increased genetic homogeneity, (b) reduced environmental factors using standardised housing criteria, (c) production of specific inter-crosses between diseased and control animals for linkage analysis (Graham et al., 2005), (d) access to disease relevant tissues even at early age with diet induce disease (e) Most of human genes have

corresponding orthologues in mouse (Waterston et al., 2002) and rat genomes enhancing translation studies (Gibbs et al., 2004).

1.2.6.1 Mouse

The early availability and development of both genome and genetic manipulation (knock-out or knock-in gene of interests) in the mouse, made it the preferred animal model for molecular research. Although, genetic and non-genetic mouse strains for human CVD are already in existence, due to the small animal body size and increase heart rate, mouse models are not as suitable as rats for the study human CVD (Rapp, 2000).

1.2.6.2 Rat

Using rats as an animal model for the study of human CVD has more advantages over mouse models due to the fact that (a) rats' bigger body size allows for easier handling, surgery, and physiological measurements and (b) rats have longer life span, allowing for a better patho-physiological similarity to human CVD (Rapp, 2000). A number of inbred rat strains, brother and sister mating over 20 generations to achieve homozygous (genetically identical) strain, have been selectively bred to display human CVD phenotypes. In 2008, the STAR consortium surveyed over 20,000 SNPs across 167 inbred laboratory rat strains (including our hypertensive and normotensive rat strains) facilitating the construction of high-density genetic maps, characterising SNPs for disease gene mapping, and QTL identification (Saar et al., 2008). Most commonly used inbred hypertensive rat strains, including selectively inbred and/or induced by surgery or diet, were developed from either Wistar-related or Sprague-Dawley stocks and these have a common origin (Reviewed by McBride et al., 2004). Recently, a phylogenetic tree analysis was carried out on 28 laboratory rat strains, genome sequence, with a number of complex traits (including hypertensive and normotensive rat strains) to gain an insight into the pressures underlying selection for these phenotypes (Figure 1-4) (Atanur et al., 2013). This study includes divergent rat strains, for example Dahl salt-sensitive (SS/Jr) and resistant (SR/Jr), as well as Sabra hypertension-prone (SBH/Ygl) and resistant (SBN/Ygl), in which sensitive and prone rat strains develop high blood pressure in response to high salt intake while resistant rat strains remain normotensive. Nevertheless, spontaneously

hypertensive rats (SHR) are a selectively inbred rat strain that develops high blood pressure without any provocative dietary and environmental factors.

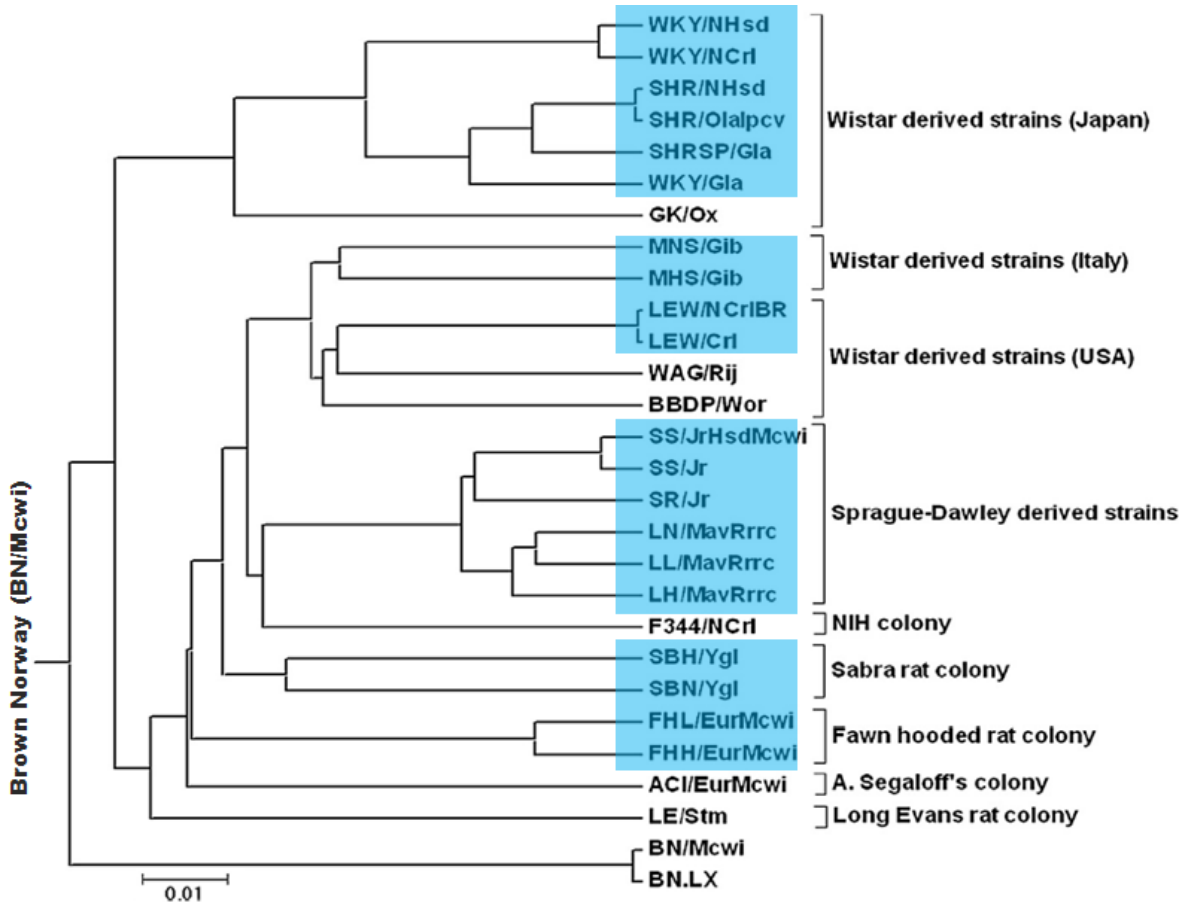


Figure 1-4 Phylogenetics tree of 28 rat strains for complex human disease.
Phylogenetics tree of 28 laboratory rat strains for human complex disease including models for hypertension with their control strains (highlighted in blue), diabetes, and insulin resistance. The scale represents genetic distance based on rats whole genome sequence compared to Brown Norway (BN/Mcwi) as the reference strain (Modified figure from Atanur et al., 2013).

1.2.7 The Stroke-Prone Spontaneously Hypertensive Rat (SHRSP)

The SHR rat was selectively developed from the Wistar-Kyoto (WKY) normotensive rat. Further inbreeding in SHR with high blood pressure (180-220 mmHg) led to the development of the stroke-prone spontaneously hypertensive rats (SHRSP), a sub-strain of SHR with increased blood pressure (220-240 mmHg) and a greater propensity to cerebral stroke (approximately 80%) than SHR (approximately 10%) (Yamori & Okamoto, 1974; Okamoto et al., 1975). Moreover, the incidence of cerebral stroke in the SHRSP reaches 100% with high salt intake (Yamori et al., 1974). Further phenotype and end-organ damage assessment studies showed that the SHRSP displays other human CVD features as well, such as left ventricular hypertrophy (Clark et al., 1996), endothelial dysfunction (Kerr et al., 1999) and insulin resistance (James et al., 2001). As these patho-physiolologal traits shared similarity with the human disease traits (Rapp, 2000), the SHRSP was considered an excellent model to study the development, prognosis and treatment of HTN, in addition, to gene-environment interaction. Nevertheless, human CVD, such as spontaneous heart failure, and atherosclerosis, have not been observed in the SHRSP.

The University of Glasgow animal facilities have maintained SHRSP and WKY (normotensive reference strain) colonies since 1991. Both rat strains are known officially as $SHRSP_{Gla}$ and WKY_{Gla} rat strains. The spontaneous onset hypertension state in the $SHRSP_{Gla}$ is maintained by selective mating of a male with systolic pressure between 170-190 mmHg and a female with systolic pressure between 130-150 mmHg at 12 weeks of age, compared to 130 mmHg in WKY_{Gla} rats (male and female) of the same age (Davidson et al., 1995; Graham et al., 2007). The blood pressure difference between $SHRSP_{Gla}$ genders has been confirmed to be due to the genetics influence on chromosome Y (Negrin et al., 2001).

A previous genome wide scan was carried out by our group, in order to identify QTLs underlying the $SHRSP_{Gla}$ phenotype (Clark et al., 1996). The linkage analysis study start with cross between parental strains ($SHRSP_{Gla}$ and WKY_{Gla}) to generated heterozygotic first generation (F1). The F1 offspring undergo brother-sister mating to generate F2 progeny, which have random genetic compliment of the parental strains. The offspring's of F2 are genotyped with microsatellite

markers spanning the genome and phenotyped for $SHRSP_{Gla}$ traits. Both genotype and phenotype data are imported into a statistical software to calculate the probability of LOD of each trait set against the genomic positions. This approach identified a number of QTLs with LOD score higher than 3.0 for $SHRSP_{Gla}$; two QTLs on chromosome 2 for blood pressure (with and without salt diet), QTL on chromosome 3 for pulse pressure (with and without salt diet), QTL on chromosome 14 for LVMI (Figure 1-5a), and QTL for stroke on chromosome 5 (Jeffs et al., 1997). The QTL regions identified were too large to carry out positional cloning, in order to identify trait controlling genes. Therefore, to validate and confine an implicated QTL, a substitution of small segments of a chromosome from one strain to another is used to fine-map the QTL region to identify genes that influence hypertension, cardiac hypertrophy and stroke (Rapp, 2000). Congenic strains were constructed using marker-assisted approach to confirm the existence of these QTLs so as to allow genetic dissection of the implicated regions (Graham et al., 2005). This is carried out by introgression of donar sequences from the WKY_{Gla} strain to the $SHRSP_{Gla}$ background by backcrossing WKY_{Gla} strain to $SHRSP_{Gla}$ with selective breeding of the genetically best animals from each generation to be maintained for backcrossing using microsatellite markers (Figure 1-6). The QTLs were validated when the phenotype of the congenic strains of chromosome 2, 3 and 14 were compared to the $SHRSP_{Gla}$, for blood pressure (McBride et al., 2003), pulse pressure and LVMI (Foote et al., 2011), respectively (Figure 1-5b).

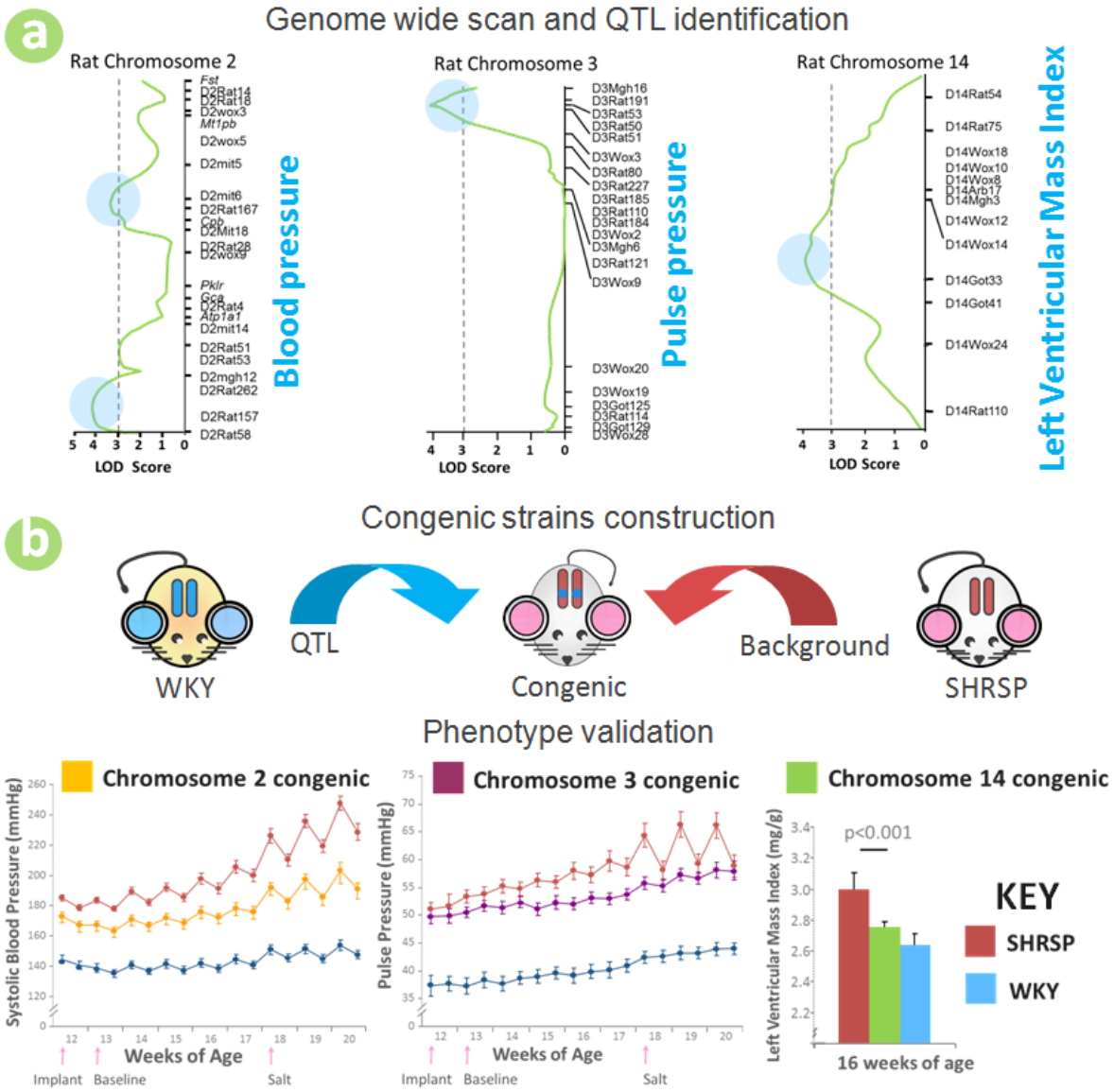


Figure 1-5 The identification and validation of the SHRSP$_{Gla}$ phenotype QTLs.
a) Previous genome wide scan study between SHRSP$_{Gla}$ and WKY$_{Gla}$ identified SHRSP$_{Gla}$ phenotype QTLs for blood pressure on chromosome 2 (two QTLs), pulse pressure on chromosome 3, and LVMI on chromosome 14 rats with LOD score cut-off >3.0 (Clark et al., 1996). **b**) Phenotype validation on the constructed congenic strains: chromosome 2 congenic strain showed significant reduction systolic blood pressure (with and without salt diet) compared to the SHRSP$_{Gla}$, chromosome 3 congenic strain showed significant reduction in pulse pressure (with and without salt diet) compared to the SHRSP$_{Gla}$, and chromosome 14 congenic strain showed significant reduction in LVMI compared to the SHRSP$_{Gla}$ at 16 weeks of age (Data courtesy of Glasgow University laboratory).

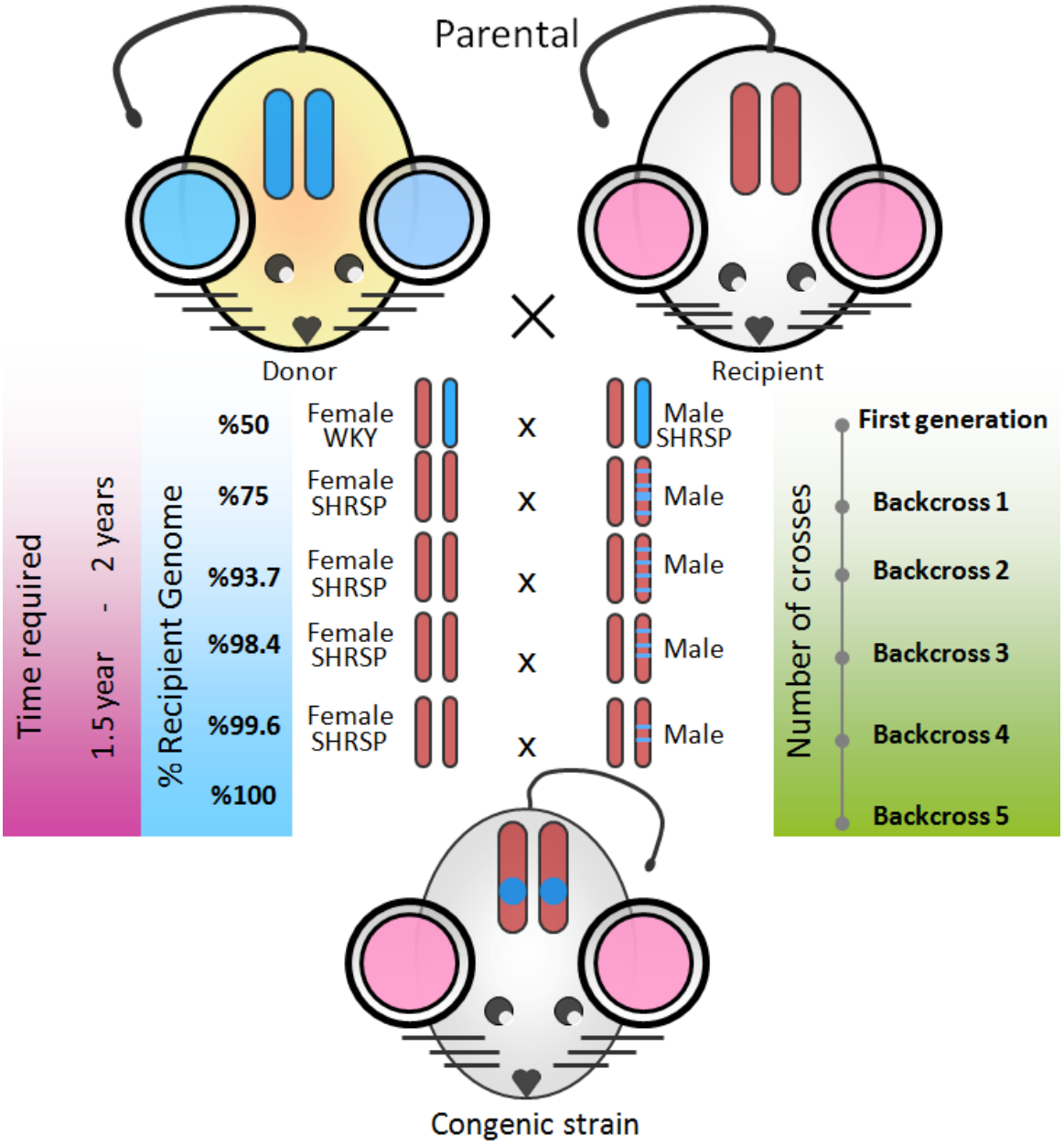


Figure 1-6 Congenic strain construction.
Congenic strain constructed using marker assisted approach by introgressing the QTL of interested from the donor (WKY_{Gla}) rat strain into the genetic background of the recipient ($SHRSP_{Gla}$) strain by mating the two to generate the first generation. Then first backcross offsprings are genotyped with microsatellite markers spanning the entire genome including chromosome of interests in which additional markers are used. Selective breeding is used for the second backcross where only the rat males with the least heterozygous marker alleles on the chromosomal region of interest and the most homozygous for recipient alleles on remaining genome are used to mate with the recipient female rat strain. The backcrosses is repeated with the same criteria for selective breeding to produce a congenic strain with homozygousity in all background markers. Finally, QTL of interest is fixed to obtain homozygous by crossing appropriate males and females (Original figure, constructed from information within Graham et al., 2005).

1.2.7.1 Microarray Expression Analysis and Congenic Strains

Microarray gene expression profiling is defined as the simultaneous measurement of thousands of genes expression levels within a given messenger RNA (mRNA) sample (from two or more experimental conditions) using DNA microarrays. Since the introduction of microarray technology, the advancements in clinical diagnostics (confirmation of genes or biomarkers) and medical research (study the mechanisms involved in complex disorders and the response of therapeutic drug or diet), have flourished.

Microarray analyses are a multiplex technique that requires the use of specialist machines to analyse and detect the signals from the hybridization of oligonucleotides probes bound to the target and control DNA, detecting the variation in expression by measuring the difference in the intensity of the signal. Typically, microarray expression data include a number of variation sources, such as random error,introduced during the performance of the experiment and/or measurement error associated with acquisition of raw intensity data from microarray hybridized images (Tarca et al., 2006). As a result, normalisation method are used to reduce these systemic effects and maintain biological variability.

The implementation of integrating gene expression profiling and congenic rat strains has been successful in isolating candidate genes for hypertension in QTL regions. *Cd36* gene (thrombospondin receptor), which regulates fatty acid metabolism and ectopic accumulation of fat, was identified as a causative gene (mutated) for insulin resistance in the SHR using rat chromosome 4 congenic strains (with QTL from the normotensive BN rat) (Aitman et al., 1999). *Srebf1* (sterol regulatory element binding factor 1c isoform) was identified as causative gene (mutated) for hepatic steatosis (common metabolic disorder in hypertensive patients) in the SHR using rat chromosome 10 congenic strain (with QTL from the normotensive BN rat) (Pravenec et al., 2008b).

Alternatively, other research groups integrated recombinant inbred (RI) rat strains and gene expression profiling as a strategy to identify candidate genes for hypertension. RI strains have been generated by mating two inbred strains (hypertensive and a control), followed by at least 20 generations of brother-

sister interbreeding, so as to produce inbred strains in which each individual RI strain has unique parental strains genetic mixture. This approach led to the identification of *Ogn* (osteoglycin) as a candidate gene for cardiac hypertrophy of the left ventricle using BXH/HXB panel of rat RI strains, derived from SHR and BN strains (Petretto et al., 2008).

1.2.7.2 Candidate Genes for Blood Pressure Regulation and Salt Sensitivity in SHRSP$_{Gla}$

Our group carried out renal gene expression profiling between SHRSP$_{Gla}$, WKY$_{Gla}$ and chromosome 2 congenic strains to identify differentially expressed genes mapping within the blood pressure QTL (positional candidate genes). *Gstm1* (glutathione S-transferase μ-type 1) which is involved in defence against oxidative stress found to be expressed in kidney at low level in the SHRSP$_{Gla}$ when compared to chromosome 2 congenic and WKY$_{Gla}$ strains at early age (McBride et al., 2003). A follow-up study identified sequence variants within *Gstm1* promoter and coding regions (McBride et al., 2005). The reduction of renal and vascular expression level of *Gstm1* in the SHRSP$_{Gla}$ were not affected with anti-hypertensive drugs suggesting its involvement in hypertension pathogenesis is via oxidative stress pathway (Koh-Tan et al., 2009). Furthermore, *Gstm1* transgenic SHRSP$_{Gla}$ strain model, *Gstm1* from WKY$_{Gla}$ integrated into the SHRSP$_{Gla}$ background, shows significant reduction in systolic pressure compared to SHRSP$_{Gla}$ (Olson, 2014).

Sphingosine-1-Phosphate Receptor 1 (*S1pr1*) and *Vcam1* have been identified as candidate genes for salt-sensitive hypertension in the SHRSP$_{Gla}$. Sequence variants were identified in *S1pr1* and *Vcam1* and their expression levels were significantly differentially expressed between chromosome 2 congenic strains and the SHRSP$_{Gla}$ in response to salt intake (Graham et al., 2007). *S1pr1*, also known as endothelial differentiation gene 1 (*Edg1*), is essential for the regulation and protection of vascular growth and development during embryogenesis via vascular endothelial cell growth factor (Chae et al., 2004; Gaengel et al., 2012). *S1pr1* knock-out mice died due to vascular de-stabilisation (Garcia et al., 2001). *S1pr1* is important inducer of pro-inflammatory signaling

pathway that may contribute to vascular inflammation in $SHRSP_{Gla}$ (Yogi et al., 2011). *Vcam1* is also considered as pro-inflammatory marker, which is mainly expressed in endothelial cells. Meta-analysis studies on renal gene expression profiling from a number of hypertensive rat strains and their control (publically available), found variation in *Vcam1* expression levels between all hypertension rats compared to all normotensive rats at established hypertension stage only (Marques et al., 2010). This may suggest that *Vcam1* is involved in kidney inflammation.

1.2.7.3 Systems Genetics Analysis

Associating sequence variation within a positional gene identified by the integration of microarray expression analysis and congenic strains was used as strategy to identify causative genes for $SHRSP_{Gla}$ phenotype. The recent availability of the $SHRSP_{Gla}$ and WKY_{Gla} whole genome sequence (Atanur et al., 2013) opened the door for other integrative approaches. For instance, integrating deleterious sequence variants from whole genome sequence with differentially expressed genes from gene expression profiling data can identify a genetic loci that control mRNA expression levels on functional tissues between $SHRSP_{Gla}$ and WKY_{Gla}. This integration can be enhanced with addition of expression profiling of microRNAs (miRNA), short non-coding RNA sequence that regulate genes expression at post translation level, especially if a positional differentially expressed miRNA is affected by genetic variants and/or a positional differentially expressed gene with miRNA (that is also differentially expressed) binding region affected by a sequence variant. This integration can be further enriched with the addition of other 'omics' data, such as proteomics and metabolomics (intermediates and products of metabolism) to fully understand the genetic determinants of disease in the $SHRSP_{Gla}$ (Figure 1-7).

Nevertheless, the integrative 'omics' strategy varies among researchers; however, they are divided into two general approaches a) independent integration: to determine the results separately in each 'omics' then combine the results from each 'omic' into one interpretative model (simple and popular among researchers), b) dependent integration: combine all 'omics' data into one mathematical model (challenging requiring advanced bioinformatics analysis).

Integrative studies in human using functional sequence variants from whole genome population and GWAS studies are more likely to regulate gene expression (Nicolae et al., 2010; Lappalainen et al., 2013) in a tissue specific manner (Dimas et al., 2009). This makes the application of systems genetics analysis directly to humans challenging, due to obtaining samples of CVD functionally relevant tissues. Alternatively, animal models serve as invaluable source for systems genetics analysis, in which the prioritised genes and/or pathway can be translated to human. For example, *Ogn* that was identified from integrative approach in rat studies, then was validated in mouse by knock-out experiment, and afterwards, translated to human when its expression level in RI rat strains correlated with heart tissue from patients with aortic stenosis (Petretto et al., 2008).

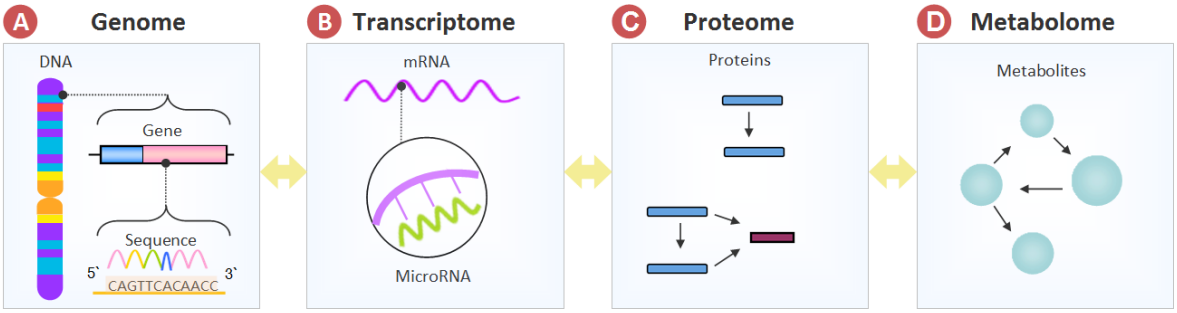


Figure 1-7 Multiple 'omics' analysis in the SHRSP$_{Gla}$.
Systems genetics analysis via integrative 'omics' data aims to give answers to **A**) what happen using genomics analysis (associate causative genetic variants to the SHRSP$_{Gla}$ traits), **B**) what appears to be happening using transcriptomic analysis, **C**) what makes it happen using proteomics analysis and **D**) what has happened and is happening using metabolomics analysis. B,C and D is carried out by monitoring mRNA, microRNA, proteins and metabolites levels in a functionally relevant tissues to the SHRSP$_{Gla}$ traits (when compared to a control). Molecular data of **A**, **B**, **C** and **D** that target different cellular information along with SHRSP$_{Gla}$ traits can help to model biological networks which can be later translated to human study (original diagram compiled from information from MacLellan et al., 2012).

1.2.7.4 Gene Expression analysis tools

A number of bio-statistical methods were used to analyse the differentially expressed genes in the dataset, by Glasgow Cardiovascular Research Centre (GCRC) staff. Previously, our group used two classic approaches in analysing gene expression profiling from microarray experiments. In-house Rank Products

analysis (RP), was used by our group, which is a simple and rapid technique based on calculating rank products from replicate experiments. This method includes many advantages, and it is mainly reliable in small and highly noisy microarray data, uses biological reasoning; hence is very powerful in determining biological relevant expression changes, and needed few numbers of replica experiments to obtain reproducible results (Breitling et al., 2004). Rank product is a script that runs on R programming language (http://www.r-project.org/), and is widely used for statistical software development and data analysis that needs prior knowledge of R language and there is no graphic user interface i.e not user friendly. Alternatively, since diseases are often caused by the disruption of a system, or network of genes, identifying only the individual differentially expressed genes from a tissue may not be sufficient to discover the underlying mechanisms of cardiovascular disease. To overcome this obstacle, our group used co-expression analysis, called meta-covariate (Hopcroft et al., 2010), in which a set/cluster of genes that expressed in coordinated fashion i.e "response- relevant" under controlled or uncontrolled perturbation in an experiment.

Recently our group started to use Partek® Genomic Suit, a commercially statistical and data mining software that offers interactive data visualization of the results. Partek® supports a complete workflow for microarray analysis which measure changes in gene, exon and miRNA expression. In addition, to next generation sequencing that facilitate molecule quantitation analysis at the transcript level, alternative splicing, allele specific expression, and SNP discovery are also possible (http://www.partek.com/). Partek uses Analysis of variance (ANOVA), which is an ideal method for gene expression analysis; it can identify both sources of variation, characterize the signal to noise ratio in a system, and can include more than two factors (multi-way ANOVA) that may contribute to total experimental variation such as scanner, scan date, operator, etc.

In simple differential expression analysis of large-scale gene expression survey, the analysis must be corrected for multi-testing. False Discovery Rate (FDR) is used by our group for the control of false positive in multi-tests that result in significant result (Benjamini & Hochberg, 1995; Pawitan et al., 2005).

1.2.7.5 Ingenuity Pathway Analysis (IPA)

We also used IPA functional analysis for our gene expression profiling data, genes affected by genomics variants, as well as the integration between the two. IPA is a web-based software widely used for the analysis of microarray results with respect to biological pathway context and prediction of novel interactions between genes that appear to be differentially expressed. IPA reflects a multifaceted knowledge database, known as Ingenuity Pathway Knowledge Base (IPKB), and contains information on genes, proteins, chemicals, drugs, diseases, biological functions and pathways which have been extracted manually from different published journals. One of the advantage of IPA over other data mining tools is they have their own unique annotation for genes, in which all uploaded data including probe identifiers from all commercial gene expression arrays, Ensembl gene identifiers, and gene names map to their database giving them unique identifiers. This unique identifier can facilitate integration/comparison between genes within QTL regions, genes affected by genomics variants and with differentially expressed genes from different array platforms (http://www.ingenuity.com/).

1.2.7.6 Gene Expression Validation

A number of steps were used to increase the detection sensitivity of true significantly differentially expressed gene, such as normalisation of the data, and correcting for multi-testing. Nevertheless, large-scale gene expression survey has the disadvantages of identifying false-positive results (Larkin et al., 2005). As a result, independent confirmation by quantitative Reverse Transcription - Polymerase Chain Reaction (qRT-PCR) is necessary. qRT-PCR permits accurate and sensitive measurement of gene expression level, which will be used to validate candidate gene expression.

1.2.7.7 Translation studied from rat to human and vice versa

It is important to note that direct translation from rat to human is not always possible due to the difference of genetics between the two species. For example, when we tried to associate human orthologues of *Gstm1* (candidate gene for HTN in $SHRSP_{Gla}$) (McBride et al., 2003; McBride et al., 2005) with hypertension in human via sequencing, genotyping and studies in three different independent cohorts, we failed to show significant results (Delles et al., 2008c). To our best knowledge, no successful translation of candidate genes for hypertension was made from rat to human. Nevertheless, human orthologues of *Cd36*, candidate gene for insulin resistance and blood pressure in the SHR) (Pravenec et al., 2008a) were successfully associated with insulin resistance in human populations via genetic *CD36* renal deficiency (Miyaoka et al., 2001; Lepretre et al., 2004). In addition, epoxide hydrolase 2, cytoplasmic (*Ephx2)* candidate gene for heart failure in the spontaneously hypertensive heart failure rats (Monti et al., 2008) and *Ogn* candidate gene for left ventricular hypertrophy in the SHR (Petretto et al., 2008) both translated to human studies.

Alternatively, other groups questioned the usefulness of identifying individual candidate genes for hypertension in hypertensive rats and should focus only on the pathogenesis roles of clusters, networks and pathways instead as human GWAS studies can directly identify candidate genes for hypertension (Nabika et al., 2012). This led to another translation direction of candidate gene for hypertension from human to rat. This translation approach was carried out by (a) integrative comparative genomics between human GWAS candidate genes for hypertension and hypertensive rat to explore the transcriptional mechanisms that mediates the outcome of the genes identified (Langley et al., 2013) (b) Applying Zinc-finger nuclease mediated mutagenesis to test the role of human GWAS candidate genes for hypertension in hypertensive rats (Jin et al., 2014) and to differentiate between multiple causative genes at a single GWAS locus (Flister et al., 2013). For example, Endres et al., 2014 showed evidence concluding that pleckstrin homology domain containing, family A member 7 (*Plekha7*), using Zinc-finger nuclease technology, may affect the variation in blood pressure and cardiovascular functions by effecting vasculature via changes in intracellular calcium handling and of nitric oxide bioavailability in Dhal salt-sensitive rat.

1.3 Aims

The overall aim of this study is to investigate the genetic determinants of the $SHRSP_{Gla}$ traits that will help in human translational studies for hypertension and vice versa.

In chapter 3: Improving and implementing new gene expression analysis workflows that allows for (a) properly adjusting of batch bias transcriptomic data to improve the accuracy of identification positional candidate genes, (b) maximising the efficiency of identifying positional candidate genes in our integrative approach of gene expression profiling data with congenic strains, and (c) identifying candidate genes for the $SHRSP_{Gla}$ phenotype via gene expression cluster analysis between $SHRSP_{Gla}$ and WKY_{Gla} across multiple tissues at different hypertensive stages.

In chapter 4: Identifying, annotating and validating the genomic variations between $SHRSP_{Gla}$ and WKY_{Gla} whole genome sequence compared to BN reference assemblies and extracting and prioritising $SHRSP_{Gla}$ unique sequence variants from 27 (non stroke-prone) rat genomes.

In chapter 5: applying a multifaceted approach that includes and integrates systems genetics analysis from rat and human studies that allows for (a) investigation of different types of genomics variants that are likely to influence transcriptome data outcome between our rat strains, (b) customisation of commercial gene expression arrays to our rat strains to improving the accuracy of our analysis, (c) prioritisation of candidate genes for blood pressure, cardiac mass, and stroke in the $SHRSP_{Gla}$, and (d) prioritisation of CVD candidate genes and miRNAs for translation study from our rat strains to human and vice versa.

2 Methods

2.1 Materials

2.1.1 Rat strains

The standardised housing criteria for both $SHRSP_{Gla}$ and WKY_{Gla}, which have been developed at University of Glasgow since 1991 and the construction of congenic rat strains using microsatellite marker-assisted, were previously described elsewhere (Graham et al., 2005). The nomenclature of these congenic rat strains consists first of the abbreviation of the background strain (SP used as a short for SHRSP), second by the donor strain, third the origin of the colony (Gla used as short for Glasgow) and a number which refers to the rat chromosome number that was introgressed from the donor strain.

In this study a number of congenic rat strains were used for chromosome 2, 3 and 14. Chromosome 2 congenic consists of $SP.WKY_{Gla}2a$ using microsatellite markers D2Rat13 (located at 37,753,748 bp) and D2Rat157 (located at 216,711,836 bp) with 179 Mbp transferred congenic interval and $SP.WKY_{Gla}2c^*$ using microsatellite markers D2Wox9 (located at 165,636,169 bp) and D2Mgh12 (located at 210,636,169 bp) with 45 Mbp transferred congenic interval. Chromosome 3 congenic strains consists of $SP.WKY_{Gla}3d$ using microsatellite markers D3Mgh16 (located at 6,373,335 bp) and D3Wox3 (located at 27,506,822 bp) with 21 Mbp transferred congenic interval and $SP.WKY_{Gla}3d$ using microsatellite markers D3Mgh16 (located at 6,373,335 bp) and rs197649383 (located at 13,643,063 bp) with 7 Mbp transferred congenic interval. Chromosome 14 congenic strains consists of $SP.WKY_{Gla}14a$ using microsatellite markers D14Rat54 (located at 5,116,492 bp) and D14Rat110 (located at 107,119,581 bp) with 102 Mbp transferred congenic and $WKY.SP_{Gla}14a$ using microsatellite markers D14Rat54 (located at 5,116,492 bp) and D14Got41 (located at 34,375,297 bp) with 29 Mbp transferred congenic. Researchers at the British Heart Foundation - Glasgow Cardiovascular Research Centre (BHF/GCRC) had collected tissues from the above rat strains previous to this study.

2.1.2 Microarray gene expression profiling data

All gene expression profiling data used in this study were carried out previously by BHF/GCRC staff, some published, however, their analysis in this study were improved, combined, integrated and extended to other genes expression analysis tools (Table 2-1). All gene expression profiling data in these rat strains were from male specimens.

Table 2-1 Gene expression profiling data from different tissues, ages and platforms used in this study.

		Neonatal	5 weeks		16 weeks		21 weeks	21 weeks salt challenged	
Kidney	Strain	×	$SHRSP_{Gla}$, WKY_{Gla} (n=3)		$SHRSP_{Gla}$, $SP.WKY_{Gla}$2a, WKY_{Gla} (n=4)		$SHRSP_{Gla}$, $SP.WKY_{Gla}$2a, WKY_{Gla} (n=3)	$SHRSP_{Gla}$, $SP.WKY_{Gla}$2a, WKY_{Gla} (n=3)	$SHRSP_{Gla}$, $SP.WKY_{Gla}$3d, $SP.WKY_{Gla}$3f, WKY_{Gla} (n=4)
	Chip		Affymetrix RGU34 A,B,C		Affymetrix Rat230_2		Affymetrix Rat230_2	Affymetrix Rat230_2	Illumina
	Published		McBride et al., 2003		×		Graham et al., 2007; Hopcroft et al., 2010		×
Heart	Strain	$SHRSP_{Gla}$, $SP.WKY_{Gla}$14a, $WKY.SP_{Gla}$14a, WKY_{Gla} (n=4)	$SHRSP_{Gla}$, $SP.WKY_{Gla}$14a, $WKY.SP_{Gla}$14a, WKY_{Gla} (n=4)	$SHRSP_{Gla}$, WKY_{Gla} (n=4)	$SHRSP_{Gla}$, $SP.WKY_{Gla}$14a, $WKY.SP_{Gla}$14a, WKY_{Gla} (n=4)	$SHRSP_{Gla}$, WKY_{Gla} (n=4)	×	$SHRSP_{Gla}$, WKY_{Gla} (n=4)	
	Chip	Illumina	Illumina	Affymetrix Exon 1.0 ST	Illumina	Affymetrix Exon 1.0 ST		Illumina	
	Published	Crawford, 2011						×	
Brain	Strain	×	$SHRSP_{Gla}$, WKY_{Gla} (n=4)		$SHRSP_{Gla}$, WKY_{Gla} (n=4)		×	$SHRSP_{Gla}$, WKY_{Gla} (n=4)	
	Chip		Illumina		Illumina			Illumina	
	Published		Bailey et al., 2014					×	
Aorta	Strain	×	×		$SHRSP_{Gla}$, WKY_{Gla} (n=3)		×	×	
	Chip				Affymetrix RAE230 A and B				
	Published				×				

(×) indicate the data was never published before or no gene expression data was carried out at this time point.

2.1.3 Genes annotation and functional analysis

A number of resources used for gene expression probes where their database houses information for annotation, mapping genes to functional related physiological rat regions and provides biological interpretations for genes of interests.

2.1.3.1 Affymetrix and Illumina

Affymetrix Rat Gene Chip Array (http://www.affymetrix.com) probesets (group of oligomeric probes, each 25 bases in length, per gene or small group of highly similar genes) were annotated with annotation files version 33 released in February, 2013.. Whereas, Illumina RatRef-12v1 Expression BeadChip Array (http://www.illumina.com) probes (usually one specific 50 bases oligmer probe per gene attached to silica bead type that are randomly replicated on each array) were annotated with annotation file ratref-12_v1_0_r5_11222119_a released in January, 2010. The Affymetrix and Illumina arrays version that were used in this study between our rat strains are listed in Table 2-1.

2.1.3.2 Ensembl

Ensembl (http://www.Ensembl.org/) is a joint project between EMBL - EBI and the Wellcome Trust Sanger Institute which aims to maintain and update annotation of several species genomes. Ensembl offer automated genes prediction (these are donated with LOC symbol) which are later processed manually and released in the next Ensembl annotation version. Ensembl also host a data mining tool BioMart that allows the querying of large gene expression probes from Affymetrix and Illumina for genes symbols, position in the genome, and orthologous. As Affymetrix and Illumina gene expression annotation files update relies on Ensembl gene build annotation upgrade this study also utilised the Basic Local Alignment Search Tool (Blast) available on Ensembl to query prioritised probes (sequence) with predicted and/or no gene symbols to get the recent annotation.

2.1.3.3 Rat Genome Database

Rat Genome Database (RGD; http://rgd.mcw.edu/) offers tools for gene prediction which are incorporated into Ensembl genome annotation (these are donated with RGD symbol). RGD also hosts a traits and phenotypes database for several rat strains including QTL for several disease models. This study accessed blood pressure and stroke QTLs from RGD to prioritise candidate genes in the $SHRSP_{Gla}$.

2.1.3.4 Ingenuity pathway analysis

IPA software (http://www.ingenuity.com) was used to provide biological and functional interpretations for genes found to be differentially expressed and/or harbour genomic variants. The functional analysis starts with uploading the dataset which includes an identifier (this can be gene name and/or gene expression probe ID) with corresponding fold change or expression intensity. Each identifier in the dataset map to its corresponding gene in the IPKB, or else it is excluded from the analysis. IPA then generates networks based on the molecular relationships between genes, whether it is direct (gene to gene) or indirect (genes products), according to the published literature in IPKB (as of 27/6/2014). In each network the genes associated with biological function, disease and/or canonical pathway were identified and a p value derived from Fisher's exact test to determine the probability that each biological function, disease and/or canonical pathway assigned to a given network is due to chance alone. Then networks were then scored for top biological functions, diseases and canonical pathways. In IPA, genes and/or gene products are shown as nodes (shape for each functional class) and the biological connectivity between two nodes are shown as edges (line supported by a literature reference) (Figure 2-1).

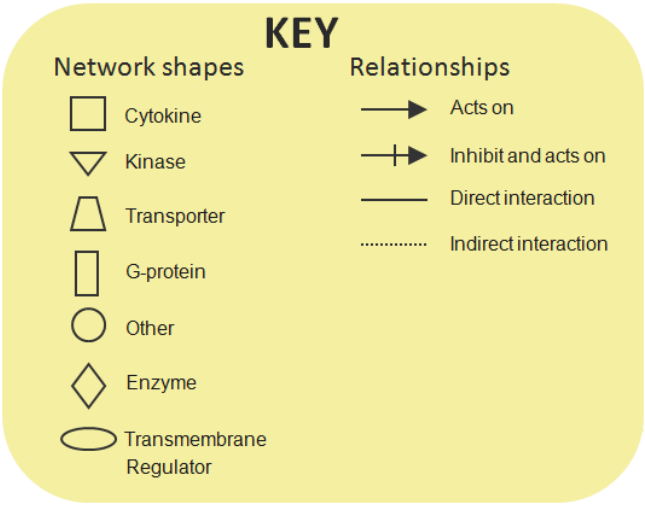


Figure 2-1 IPA network shapes and relationships.
A diagram showing IPA different functional classes for molecules and relationships types (IPA, http://www.ingenuity.com).

2.1.4 Gene expression analysis tools

2.1.4.1 Partek

Partek® Genomic suite software (http://www.partek.com/) was used in this study to analyse gene expression in pairwise comparison. Normalisation for all Affymetrix gene expression data were carried out in Partek® Genomic suite software using Robust Multi-array Average (RMA; Irizarry et al., 2003) and all Illumina gene expression data were quantile normalised in Illumina Genome Studio was imported to Partek®.

2.1.4.1.1 Analysis of variance

One-way ANOVA was used to analyse all gene expression data using the strains in rat and condition in human as factor in the ANOVA model. FDR of <0.05 was used a cut off to identify significant differentially expression genes. Furthermore, in batch biased gene expression data two-way ANOVA model was used for the analysis in which the batch and the rat strain used as factors.

2.1.4.1.2 Batch effect removal tool

Previously batch biased gene expression profiling of heart between data of $SHRSP_{Gla}$, WKY_{Gla} and chromosome 14 congenic rat strains at neonatal, 5 and 16 weeks of age was analysed by BHF/GCRC staff using two-way ANOVA (batch and

strains as factors) in Partek® with a cut-off of FDR <0.05 (Crawford, 2011). The analysis of the data was extended to implement Partek® Batch effect removal tool after quantile normalisation in Genome Studio and analysed with the same two-way ANOVA model. Partek® *3D* principle components analysis (PCA) was used to identify and compare batch effect from both analyses. In addition, the comparison of two analyses included differentially expressed genes focusing and prioritising on the least differentially expressed comparison WKY_{Gla} versus $WKY.SP14a_{Gla}$ across different time points.

2.1.4.2 BioLayout*Express*^{3D}

BioLayout*Express*^{3D} is a tool used for visualisation and clustering of large network graphs of microarray gene expression profiling data (Freeman et al., 2007). Initially, pairwise transcript-to-transcript correlation matrix is calculated based on each gene/transcript expression intensity between our rat strains across different tissues and/or timepoints. Later, undirected networks with weighted edges are constructed based on a Pearson's correlation threshold cut-off of r= 0.75, where nodes represent transcripts and edges represent Pearson correlations above the threshold. Finally, network graphs are clustered using Markov Clustering algorithm (MCL).However, different cut-off values were used in the analysis presented in this thesis to generate network graphs due to the number of samples in the imported data and when no informative cluster with genes expression pattern of interests is found after applying Markov clustering algorithm on the network graphs.

2.1.4.2.1 Cluster analysis between $SHRSP_{Gla}$ and WKY_{Gla} and 2c congenic rat strains on renal transcriptomic data

BioLayout*Express*^{3D} cluster analysis carried out on previously generated data of $SHRSP_{Gla}$, WKY_{Gla} and chromosome 2 congenic $SP.WKY_{Gla}2a$ rat strains at 16 and 21 weeks of age with and without salt intake using a cut-off value of 0.85. Gene expression data were normalised with RMA in Partek® before being imported to BioLayout*Express*^{3D}.

2.1.4.2.2 Cluster analysis between $SHRSP_{Gla}$ and WKY_{Gla} across tissues at different ages

Benefiting from the gene expression platform free feature in IPA, normalised Affymetrix and Illumina gene expression intensity data were uploaded to IPA separately. Common genes, which were used as identifiers, between all dataset across tissues at 5 and 16 weeks of age were chosen for BioLayout*Express*^{3D} cluster analysis. At 21 weeks of age and salt challenged each Illumina gene expression data of a tissue was normalised separately and BioLayout*Express*^{3D} cluster analysis was run across all tissues using Illumina probes as an identifier. Figure 2-2 shows flow diagram of the steps taken in the BioLayout*Express*^{3D} analyses.

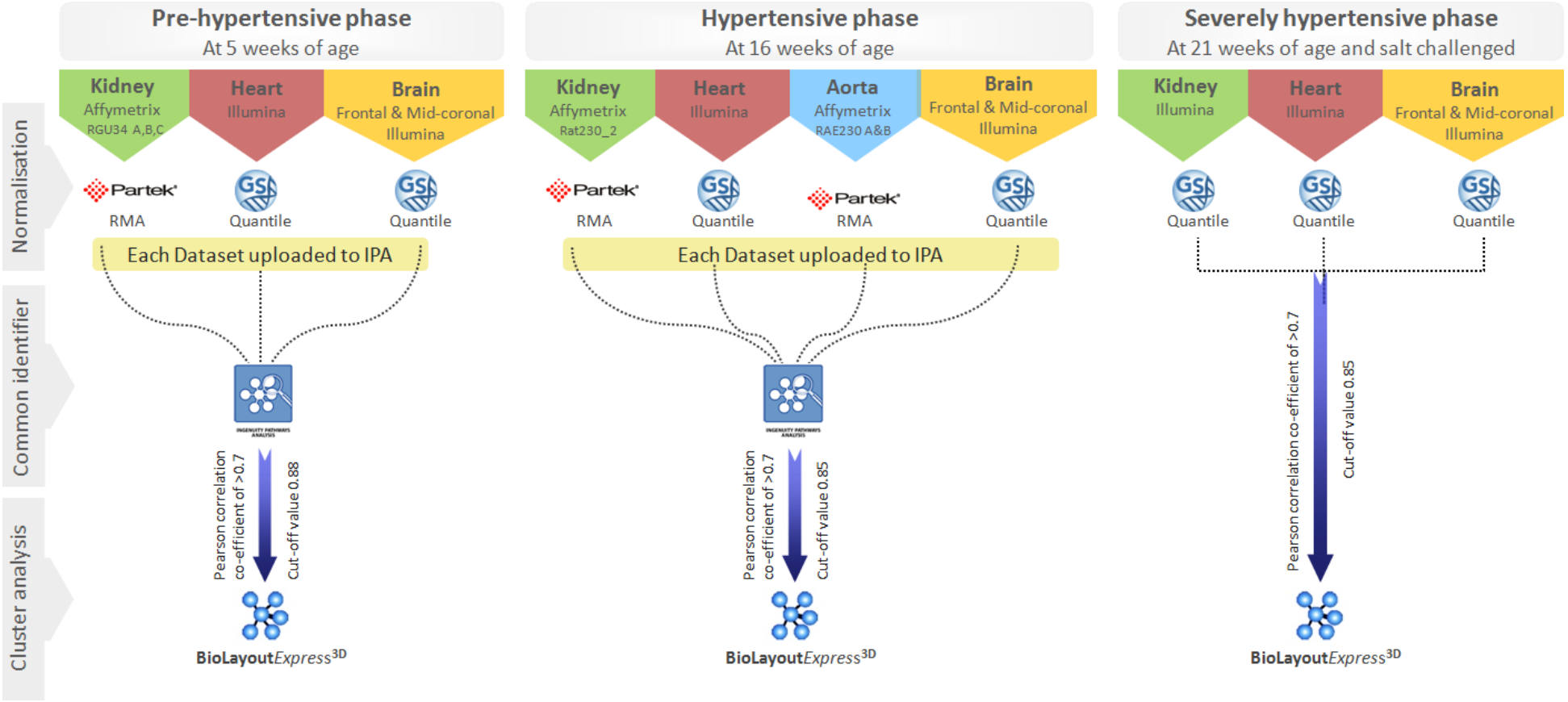


Figure 2-2 Flow diagram of the steps taken in the BioLayout*Express*^{3D} analyses.
A flow diagram detailing how gene expression data from different tissues, platform age were normalised, processed and imported in BioLayout*Express*^{3D} for cluster analysis.

2.1.5 Whole genome Next generation sequencing of $SHRSP_{Gla}$ and WKY_{Gla} rat strains

2.1.5.1 Samples

Both rat strains sequencing library preparations were processed on the Illumina platform (HiSeq 2000) by collaboration. The $SHRSP_{Gla}$ was sequenced at Prof. Norbert Huebner's laboratory, Max Delbrück Center for Molecular Medicine, Germany with read length 75 bp. Whereas, the WKY_{Gla} was sequenced at Prof. Tim Aitman's laboratory, Imperial College, London, UK with read length 100 bp.

2.1.5.2 Rat reference genome RGSC 3.4

Training on handling $SHRSP_{Gla}$ and WKY_{Gla} whole genome sequence data including mapping to Rat Genome Sequencing Consortium (RGSC) 3.4 rat reference genome, calling and annotating sequence variants and large deletion on $SHRSP_{Gla}$ and WKY_{Gla} was offered at Tim Aitman's laboratory. The calling and annotation of NGS small sequence variants, as well as the calling of raw large deletion analysis against BN RGSC 3.4 was executed by Mr Santosh Atanur from Tim Aitman's laboratory.

2.1.5.2.1 Mapping

Raw paired-end reads, trimmed at the end of read of all the bases with quality less than 20, of $SHRSP_{Gla}$ and WKY_{Gla} mapped to the draft reference genome of Brown Norway RGSC 3.4 (Gibbs et al., 2004) with Burrows wheeler aligner version 0.5.8c (Li & Durbin, 2009). Sequence Alignment Map data were then compiled and converted to a single compressed binary file, for each strain, using SAMtools (Li et al., 2009). Picard software (http://picard.sourceforge.net) was used to remove duplicate reads.

2.1.5.2.2 Sequencing variants calling

Genome Analysis Toolkit (GATK; McKenna et al., 2010; Depristo et al., 2011) version 1.0.6001 was used to reduce the rate of false positive results by (a) local realignment of a read that overlapped with an INDEL since there are mis-

alignments around the location (b) recalibration of base quality since Illumina quality scores are biased toward high quality. GATK with multi sample feature were used to call filtered SNPs and INDELs (<50 bp) which are unique between our strains discarding common variants generated by each strain against rat reference genome RGSC 3.4.

2.1.5.2.3 Variants effect predictor annotation Ensembl 69

Variants effect predictor tool version 2.4 (McLaren et al., 2010) was used to annotate and functionally categorize the effect of small variants based on their location in a gene based on Ensembl 69 gene set.

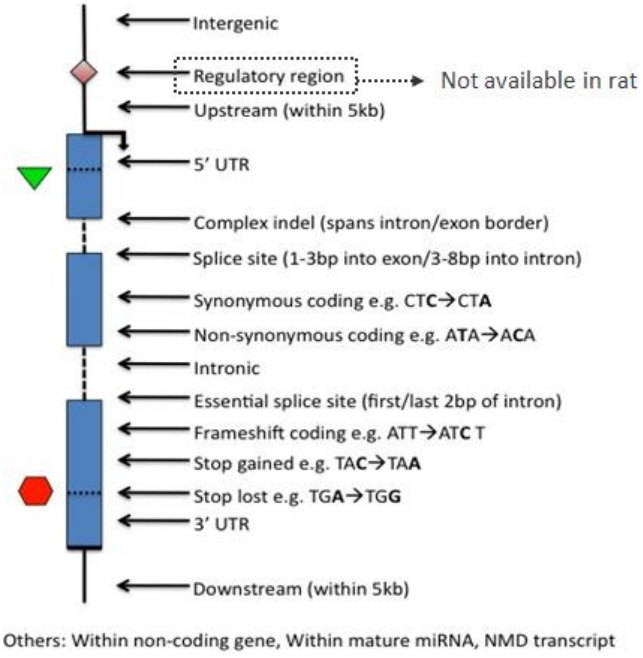


Figure 2-3 Ensembl 69 predicted sequence variants consequence types.
A diagram showing the location of the types of consequences relative to the transcript structure (Modified figure from McLaren et al., 2010).

2.1.5.2.4 Large deletion

A custom tool described previously (Atanur et al., 2010) with algorithm to detect large deletions (>50 bp) was used by Imperial College staff to generate raw data (deleted regions and the number of reads implicated) for each strain independently. Deleted regions in $SHRSP_{Gla}$ and WKY_{Gla} with > 4 reads was chosen as cut off and annotated with Ensembl BioMart version 69.

2.1.5.2.5 Copy number variation

Copy number variations (CNV) were called with CNVnator version 0.2.7 (Abyzov et al., 2011) for each chromosome with bin size of 100 bp guided by rat reference genome RGSC 3.4 for each strain independently. Regions with CNV duplication and deletion were annotated with Ensembl BioMart version 69.

2.1.5.3 Rat reference genome Rnor 5.0

2.1.5.3.1 Mapping

Raw paired-end reads, trimmed at the end of read of all the bases with quality less than 20, of $SHRSP_{Gla}$ and WKY_{Gla} mapped to the new reference genome of Brown Norway Rnor 5.0 with improved Burrows wheeler aligner version 0.7.4 (Li & Durbin, 2009). Sequence alignment Map data were then compiled and converted to a single compressed binary file, for each strain, using SAMtools (Li et al., 2009). Picard software (http://picard.sourceforge.net/) was used to remove duplicate reads.

2.1.5.3.2 Sequencing variants calling

Genome Analysis Toolkit version 2.5-2 (GATK; McKenna et al., 2010; Despristo et al., 2011) was used to reduce the rate of false positive results by (a) local realignment of a read that overlapped with an INDEL since there are mis-alignments around the location (b) recalibration of base quality since Illumina quality scores are biased toward high quality. GATK with multi sample feature were used to call filtered SNPs and INDELs (<50 bp) which are unique between our strains discarding common variants generated by each strain against rat reference genome Rnor 5.0.

2.1.5.3.3 Variants effect predictor annotation Ensembl 71

Variants effect predictor tool version 7.1 (McLaren et al., 2010), which offers more predicted consequence types than version 2.5 was used to annotate and

functionally categorize the effect of small variants based on their location in a gene based on Ensembl gene set 71.

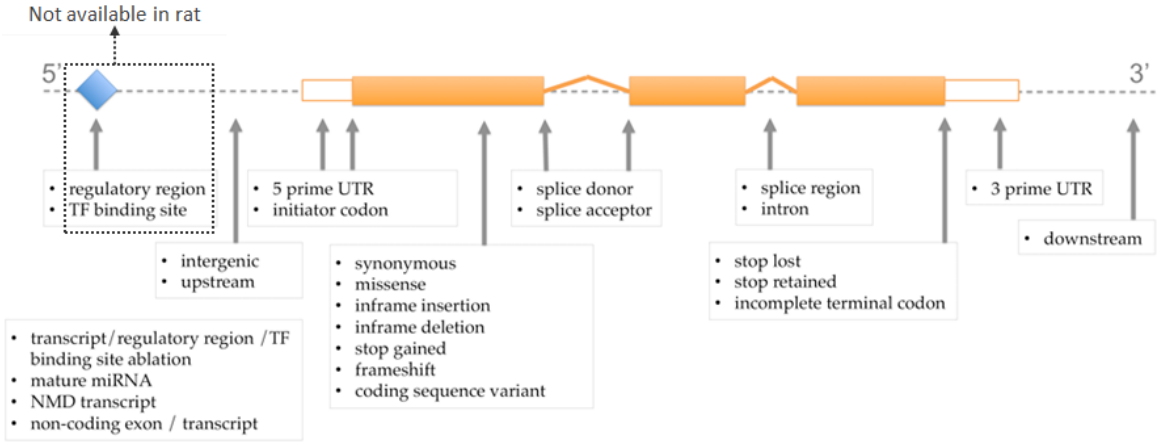


Figure 2-4 Ensembl 71 predicted sequence variants consequence types.
A diagram showing the location of the types of consequences relative to the transcript structure (Modified figure from McLaren et al., 2010).

2.1.5.4 Quality control

Three different sequencing data were used as quality control for whole NGS of $SHRSP_{Gla}$ and WKY_{Gla}.

2.1.5.4.1 SureSelect

SureSelect (Agilent Technologies, UK) targeted sequence capture data generated by Mr Santosh Atanur from Imperial College staff for $SHRSP_{Gla}$ and WKY_{Gla} on chromosome 2 region (between 211,822,164-213,846,183 bp) was compared to the rat reference genome RGSC 3.4 (using the same pipeline in paragraph 1.1.2.1). DNA used to generates targeted paired-end short reads (100 bp in length) constructed from libraries with 300-600 bp insert size that were sequenced on an Illumina HiSeq2000. The annotated sequence variants between $SHRSP_{Gla}$ and WKY_{Gla} by Ensembl variants predictor (version 64) was provided in Excel sheet format in 2010 (Martin McBride, personal communication). The sequence capture array has high coverage (10,000x) and was used as quality control for NGS whole genome sequence data.

2.1.5.4.2 Sanger sequencing

In addition, previously published data (McBride et al., 2005; Graham et al., 2007) of 34 SNPs/INDEL of traditional capillary sequencing for genes on congenic region on chromosome 2 were compared to NGS data and sequence capture against rat reference genome RGSC 3.4 as quality control.

2.1.5.4.3 STAR SNP database

2,338 unique SNPs between $SHRSP_{Gla}$ and WKY_{Gla} against rat reference genome RGSC 3.4 on all chromosomes were extracted from STAR Consortium data (Saar et al., 2008) and compared to NGS data against rat reference genome RGSC 3.4 as quality control.

2.1.5.5 Visualisation of variants

2.1.5.5.1 Genome level

NGS genomic variants were visualised on genome level by Circos (Krzywinski et al., 2009).

2.1.5.5.2 Single nucleotide level

NGS genomic variants were visualised on gene and single nucleotide level by Integrative Genomic Viewer (IGV; Robinson et al., 2011).

2.1.5.6 $SHRSP_{Gla}$ unique sequence variants

2.1.5.6.1 $SHRSP_{Gla}$ sequence variants compared to 27 whole genome sequenced rat strains

$SHRSP_{Gla}$ unique SNPs and INDELs were extracted from raw NGS data of 28 whole genome rat strains against rat reference genome RGSC 3.4 (Atanur et al., 2013). $SHRSP_{Gla}$ unique sequence variants were then annotated with variants effect predictor annotation on Ensembl version 69 gene set (McLaren et al., 2010).

Moreover, $SHRSP_{Gla}$ unique non-synonymous SNPs (Ensembl gene identifiers) were uploaded to IPA to carry out functional analysis. In addition, $SHRSP_{Gla}$ and WKY_{Gla} QTLs for blood and pulse pressure, left ventricular mass index and stroke as well as rat strains blood pressure and stroke QTLs from RGD website were used to prioritise protein coding $SHRSP_{Gla}$ unique sequence variants. Furthermore, predicted protein coding $SHRSP_{Gla}$ unique sequence variants compared to the rat reference genome RGSC 3.4 were mapped to rat reference genome Rnor 5.0 and annotated with variants effect predictor annotation on Ensembl gene set 71 to prioritise variants with consistent and deleterious predicted consequences.

2.1.5.6.2 SHRSP rat strains STAR SNP database

$SHRSP_{Gla}$ and another 4 SHRSP inbred rat strains SNPs against rat reference genome RGSC 3.4 were extracted from STAR Consortium data (Saar et al., 2008) to carry out a comparison analysis.

2.1.6 Integrating transcriptomic and genomic data

2.1.6.1 In rats

2.1.6.1.1 Integration of transcriptomic data and genomic variants between $SHRSP_{Gla}$ and WKY_{Gla}

Small sequence variants and structural variants including large deletions and CNV were integrated with gene expression profiling data from heart and kidney. Illumina gene expression profiling of the heart at neonatal, 5 and 16 weeks of age were quantile normalised in Genome Studio and Affymetrix gene expression profiling of the kidney at 16 and 21 weeks of age were RMA normalised in Partek®. Both gene expression data were analysed by ANOVA in Partek® Genomics Suite using cut-off FDR <0.05. Ensembl gene identifier of genes affected by different types of genomic variants data and differentially expressed probes of the heart and kidney transcriptomic data were uploaded to IPA to carry out the integration step.

2.1.6.1.2 Improvement of microarray analysis by utilising integration of NGS sequence variants and gene expression data

Affymetrix Chip Description File (CDF) was downloaded from http://brainarray.mbni.med.umich.edu/ and called default CDF in this study (Dai et al., 2005). The default CDF were filtered by Imperial College staff of probes that harbour a SNP within their target region and called custom CDF in this study. Both CDFs were used to analyse Affymetrix gene expression data of kidney at 21 weeks of age between $SHRSP_{Gla}$ and chromosome 2 congenic $SP.WKY_{Gla}2a$ where both were RMA normalised with the same ANOVA model in Partek® Genomics Suite using cut-off FDR <0.05. Both Affymetrix default and custom CDF analyses were compared to investigate the impact of SNPs on differentially expressed Affymetrix probes prioritising positional candidate genes.

To further assess the impact of SNPs on differentially expressed probes, SNPs between $SHRSP_{Gla}$ and WKY_{Gla} were examined in 24,000 Illumina probes (50 bp) target regions. The impacted Illumina probes were uploaded to IPA along with genes within the congenic strains on chromosome 2, 3 and 14 to check for positional candidate genes that may falsely identify as differentially expressed gene in Illumina analysis.

2.1.6.1.3 "Swiss cheese" integrative and reductive model

Significantly differentially expressed positional candidate genes identified by Partek® Genomics Suite with a cut-off of FDR <0.05 and 2 fold change in kidney and heart tissues that harbour protein coding sequence variants were prioritised. Affymetrix kidney gene expression data at 16 and 21 weeks of age were normalised with RMA and Illumina heart gene expression data at neonatal and 5 and 16 weeks of age was quantile normalised in Genome Studio. IPA facilities the integration by uploading variants (Ensemble gene identifier), differentially expressed genes (probe identifier), genes within the congenic regions (Ensemble gene identifier) on chromosome 2, 3 and 14, as well as functional analysis on the result.

2.1.6.1.4 Integrating $SHRSP_{Gla}$ unique small sequence variants compared to 27 genome sequence rat strains with brain gene expression of $SHRSP_{Gla}$ and WKY_{Gla}

The isolated $SHRSP_{Gla}$ unique sequence variants were integrated with those previously generated and analysed by BHF/GCRC staff for brain transcriptomic data between $SHRSP_{Gla}$ and WKY_{Gla} at different time points (Bailey et al., 2014) and salt loaded using Rank Products (Breitling et al., 2004). The identified differentially expressed genes with FDR cut-off of <0.05 that harbour $SHRSP_{Gla}$ unique sequence variants were checked whether they mapped to $SHRSP_{Gla}$ and WKY_{Gla} QTLs for blood and pulse pressure, left ventricular mass index and stroke as well as rat strains stroke QTLs from RGD website were used to prioritise candidate genes for stroke in the $SHRSP_{Gla}$.

2.1.6.2 In human to rat and vice versa

Three different approaches were used to translate candidate genes and miRNA associated with hypertension from rat to human and vice versa.

2.1.6.2.1 Human GWAS genes for blood pressure and hypertension integrated with $SHRSP_{Gla}$ and WKY_{Gla} transcriptomic and genomic data

Human candidate genes for blood pressure and hypertension were obtained from several GWAS studies (Levy et al., 2009; Newton-Cheh et al., 2009; Padmanabhan et al., 2010; Ehret et al., 2011; Johnson et al., 2011; Kato et al., 2011; Wain et al., 2011; Salvi et al., 2012; Franceschini et al., 2013) and integrated with differentially expressed genes in heart and kidney and sequence variants between $SHRSP_{Gla}$ and WKY_{Gla}. Human candidate genes for blood pressure and hypertension were checked for orthology in rat using EnsemblCompara (precomputed data: phylogeny based orthology and paralogy predictions) available at Ensembl BioMart website, and whether they map to QTLs for blood and pulse pressure, left ventricular mass index between $SHRSP_{Gla}$ and WKY_{Gla} as well as other blood pressure QTL from other rat strains obtained from RGD website. Illumina gene expression profiling of the heart at neonatal, 5 and 16 weeks of age were quantile normalised in Genome Studio and Affymetrix gene expression profiling of the kidney at 16 and 21 weeks of age and were RMA

normalised in Partek®. Both gene expression data were analysed by ANOVA in Partek® Genomics Suite using cut-off FDR <0.05.

Affymetrix and Illumina probes targeted regions of prioritised candidate genes were checked for sequence variants between $SHRSP_{Gla}$ and WKY_{Gla}. Furthermore, prioritised candidate genes were checked for differential spliced exons events from previously published Affymetrix Exon array data on heart at 5 and 16 weeks of age between $SHRSP_{Gla}$ and WKY_{Gla} (Crawford, 2011).

2.1.6.2.2 Integrating renal miRNA sequencing data from human and rat

Small RNA molecule NGS in human between hypertensive and control kidney (cortex and medulla) and in rat between $SHRSP_{Gla}$ and WKY_{Gla} and chromosome 2 congenic $SP.WKY_{Gla}2c$* kidney at 16 weeks of age both with no replica was executed by Glasgow Polyomics Facility (Glasgow, UK) using Illumina Solexa sequencing technology. Short Solex read were processed in statistical computing software called R Console (http://www.r-project.org) using the Kraken suite components (Davis et al., 2013). Reaper tool was used to trim the smallRNA kit '3 adapter, PolyA, PolyN, low scoring and complexity tracts and obtain clean. Tally another tool from Kraken suite components were used to collapse redundant into unique reads with their depth information. Unique clean reads of human and rat Solexa reads were later mapped against miRBase (version 20) using Krakenbot Mapping Tool available on (http://wwwdev.ebi.ac.uk/enright-srv/krakenbot/). DEseq statistical package (Anders & Huber, 2010) was used to identify differential expression in simple pairwise approach for human study between hypertensive and control and for rat study between $SHRSP_{Gla}$, WKY_{Gla} and $SP.WKY_{Gla}2c$* using adjusted p value of <0.05 as cut-off.

Positional targeted genes of differentially expressed miRNA in rat kidney study between $SP.WKY_{Gla}2c$ and $SHRSP_{Gla}$ were integrated with sequence variants within 3' UTR and kidney gene expression profiling at 16 and 21 weeks of age between $SHRSP_{Gla}$ and WKY_{Gla} to prioritise candidate miRNA for the $SHRSP_{Gla}$ phenotype. Affymetrix kidney gene expression profiling at 16 and 21 weeks of age were RMA normalised in Partek®. Both gene expression data were analysed by ANOVA in Partek® Genomics Suite using cut-off FDR <0.05.

2.1.6.2.3 Integrating heart gene expression data from human and rat

In the rat study, whole heart gene expression profiling on Illumina platform of $SHRSP_{Gla}$ and WKY_{Gla} at neonatal, 5 and 16 weeks were used for the analysis. Moreover, in the human study, left ventricular gene expression data of heart failure were used for the analysis. The first, Heart data 1, was between normal and post-ischemic heart failure (Greco et al., 2012). The second, Heart failure data 2, was between normal and heart failure that arises from familial, idiopathic dilated, ischemic, viral, hypertrophic cardiomyopathies (Schinke et al., 2004). Both rat and human gene expression data were RMA normalised and analysed by Partek® Genomic suite where a FDR of <0.05 and a fold change of 1.5 were used as cut-off.

2.1.7 Gene expression validation

Reagents and laboratory equipments used in this study were of the highest available grade. Hazardous chemicals were handled in accordance with Control of Substances Hazardous to Health regulations with laboratory coat and latex gloves were worn while conducting laboratory experiments. In addition, nuclease-free reagents and consumables were used in gene expression validation.

2.1.7.1 Validation in rat

2.1.7.1.1 Quantification of nucleic acids

Total RNA (DNase free) that were generated previously by BHF/GCRC staff was accurately quantified by measuring 1.5 µL of RNA of each sample on a Nanodrop (ND-1000) spectrophotometer (Thermo Scientific, U.K.). Absorbance at 260 nm was used for quantification of nucleic acids, an optical density of 1 corresponding to 40ng/µL RNA. Absorbance ratios (260 nm/280 nm) was used as an indicator for sample purity in which a ratio of 2 for RNA was considered pure and free from protein contamination.

2.1.7.1.2 Complementary deoxyribonucleic acid (cDNA) synthesis

cDNA was synthesised using Taqman reverse transcription reagents (Applied Biosystems, USA) on PTC-225 Peltier Thermal Cycler (MJ Research) in 96 well

plates using the following conditions: 25°C for 10min, followed by 48°C for 30min, then 95°C for 5min, finally 4°C forever. Reverse transcription reaction consists of 10X Reaction Buffer 2.0µL, MgCl2 (25mM) 4.4µL, dNTPs (2.5mM ea.) 4.0µL, Primers (50µM) 1.0µL, RNase Inhibitor (20U/µl) 0.4µL, Multiscribe reverse transcriptase (50U/µl) 0.5µL and 1µg of RNA template and adjusted with water (RNase-free) to final reaction volume of 20 µL per well.

2.1.7.1.3 Quantitative real-time polymerase chain reaction QRTPCR

All Real-time PCR data performed in this study was carried out on an ABI PRISM 7900HT Sequence Detection system with Taqman Gene Expression Assays from (Applied Biosystems, USA) in 384 well plates. The Taqman Gene Expression Assays used are *Atp11b* (Rn01493364_m1, labelled FAM), *Pik3r1* (Rn00564547_m1, labelled FAM), *Fubp3* (Rn01436208_m1, labelled FAM), *Ugt2b7* (Rn00821928_m1, labelled FAM), *Afm* (Rn00597217_m1, labelled FAM), *Gucy1a3* (Rn00367252, labelled FAM), *Gucy1b3* (Rn00562775, labelled FAM), *Gpr98* (Rn01757838_m1, labelled FAM), *Smoc2* (Rn01455681_m1, labelled FAM), *Lamc2* (Rn01759129_m1, labelled FAM), *Β-actin* (4352340E, labelled VIC) and *B2m* (Rn00560865_m1, labelled FAM).

In Kidney, validation reactions consisted of 2.5 µL Taqman Expression PCR Master Mix, 1X (gene of interest) FAM labelled, 1x *Β-actin* (housekeeping gene) VIC labelled and 2 µL of cDNA in a final volume of 5µL in duplex. Whereas, singleplex reactions was used for validation in heart which consists of 2.5 µL Taqman Expression PCR Master Mix, 1X (gene of interest or *B2m* a housekeeping gene) FAM labelled, 1x water (RNase-free) and 2 µL of cDNA in a final volume of 5µL. Both were run on the same temperature cycling conditions; 50°C for 2min, then 95°C for 10min, then 40 cycles of 95°C for 0.15 seconds then 60°C for 1min.

qRT-PCR results were analysed using the relative quantification expression method of comparative C_t system ($2^{-\Delta\Delta Ct}$) (Livak & Schmittgen, 2001) with *Β-actin* (for kidney tissue) and *B2m* (for heart tissue) utilised as endogenous controls. This allows for the standardisation of the cDNA amounts added to the reaction, thus acting as a normaliser to correct for sampling discrepancies.

During the extension phase of the PCR reaction the fluorescence released from the qRT-PCR probe is directly proportional to the amount of product generated in each cycle that is detected and measured by ABI PRISM 7900HT thermo cycler. During exponential phase of amplification in PCR reaction, a threshold of fluorescence is set automatically by Applied Biosystems Sequence Detection Software. The fluorescence signal detected at a cycle number that is above this threshold is called the cycle threshold (C_t). Genes of interest were normalised to *β-actin* (for kidney tissue) and *B2m* (for heart tissue) gene expression that calculated by the ΔC_t using the following equation:

$$\Delta C_t = C_t \text{ (gene of interest)} - C_t \text{ (housekeeping gene)}.$$

The relative quantity value (RQ) was then calculated:

$$RQ = 2^{-\Delta\Delta Ct}$$

Data was analysed in at least triplicate for each rat. The comparison between the congenic and parental strains for qRT-PCR was carried out by 2 sample Student's t-test.

qRT-PCR results of Osteopontin (*Spp1*) on heart at neonatal and 5 weeks of age and *Dnm1*, *Rabgap1* and *Tor1b* on kidney at 21 weeks of age and salt challenged in our rat strains was carried out by BHF/GCRC staff.

2.1.7.2 Validation in human

To validate the potential candidate genes, heart failure hearts from Paul Barton's lab at Imperial College, London were used. These heart failure types includes dilated cardiomyopathy at different recovery and unstable stages as well as controls. The validation of *Smoc2* and *lamc2* was carried out in London by a colleague.

3 Gene expression analysis in the $SHRSP_{Gla}$ and WKY_{Gla} rat strains

3.1 Introduction

A number of experiments have been conducted across different strains, tissues, time points and conditions to identify candidate genes for the $SHRSP_{Gla}$ phenotype. However, it should be noted that there are many factors that make the integration of these experiments challenging. These include the use of different microarray platforms and statistical analysis tools, as well as changes in the state of annotation over time.

Affymetrix released a number of rat gene chip expression arrays over the years in which each upgraded version provides more probes that mapped to newly discovered genes in fewer or single array. Affymetrix continues releasing updates for genes annotation for all their rat arrays with every rat genome upgrade.

Illumina released one rat gene expression array before the release of a rat genome and has no plans to update the array genes annotation. However, a computational biology group from Cambridge University released a re-annotation of Illumina probe sequences (Barbosa-Morais et al., 2010). Illumina 50 bp oligo-nucleotide probe increases hybridisation efficiency and specificity to its target compared to Affymetrix 25 bp oligo-nucleotide length probe (Chou et al., 2004). Illumina 12 samples format in single array at an economical price compared to Affymetrix allowed for large gene expression studies.

Large gene expression studies usually suffer from systemic error or batch effect when analysing samples from multiple batches, having potentially an impact on the results. Principal component analysis (PCA), a mathematical algorithm used to scale down a high dimensional data while maintaining most variability in the data (Ringner, 2008), is used to identify non-biological biases such as batch effects.

Partek® genomic suite, commercial software, offers two strategies to adjust for batch effect. First method to correct for batch effect is by accounting for batch bias as factor in ANOVA model and the second is by removing the batch bias by Partek® effect removal tool. However, no comparison study between the two batch bias adjusting methods has been carried out in our transcriptomic bias data.

There is no gold standard method to analyse gene expression data; therefore, our group implemented different statistical tests and clustering methods on our transcriptomic data. Renal expression datasets from $SHRSP_{Gla}$, WKY_{Gla} and chromosomal 2 congenic ($SP.WKY_{Gla}2a$) rat strains, which were with and without salt challenge, were analysed a Bayesian meta-covariate cluster method and conventional pairwise approach by Rank Products. Previously in our laboratory, Bayesian meta-covariate tool identified a cluster of 13 genes (Hopcroft et al., 2010) of which most were not identified in Rank Products analysis (Graham et al., 2007). Functional analysis by IPA showed that genes involved in transcriptional activation and circadian rhythm may contribute in the protection against hypertension in response to high salt intake (Hopcroft et al., 2010). Nevertheless, the meta-covariate clustering method used is very intensive computationally and requires the use of specialist expertise that has not since been available.

BioLayout*Express*^{3D} application is a visualization and clustering tool for gene expression data that produces clusters (groupings) of genes based on pattern of expression (Freeman et al., 2007). BioLayout*Express*^{3D} software can be run on most personal computers and only requires a common identifier of genes and normalised expression intensity for the analysis. This allows for a network analysis on gene expression from different platforms and across multiple tissues (Kapetanovic et al., 2013).

Over the years our group has built a library of gene expression profiling data between $SHRSP_{Gla}$ and WKY_{Gla}, documenting early and advanced phases of hypertension. Additionally, this library also includes chromosome 2, 3 and 14 congenic and sub-congenic rat strains, allowing for the prioritisation of positional candidate genes for $SHRSP_{Gla}$ phenotype (Figure 3-1; Table 3-1).

3.2 Aims

This chapters specific aims are:

- Improve and implement new gene expression analysis workflows to identify candidate genes in the $SHRSP_{Gla}$
    - Compare two strategies to adjust for batch biased heart gene expression data between $SHRSP_{Gla}$, WKY_{Gla} and chromosome 14 congenic strains using Partek® genomic suit
    - Combined expression chip platforms approach to identify renal positional candidate genes between $SHRSP_{Gla}$, WKY_{Gla} and chromosome 3 congenic strains
    - Run BioLayout*Express*^{3D} cluster analysis to identify renal positional candidate between $SHRSP_{Gla}$, WKY_{Gla} and chromosome 2 congenic strains
- Run cluster analysis by BioLayout*Express*^{3D} between $SHRSP_{Gla}$ and WKY_{Gla} across tissues at pre-hypertensive, hypertensive and severely hypertensive phases

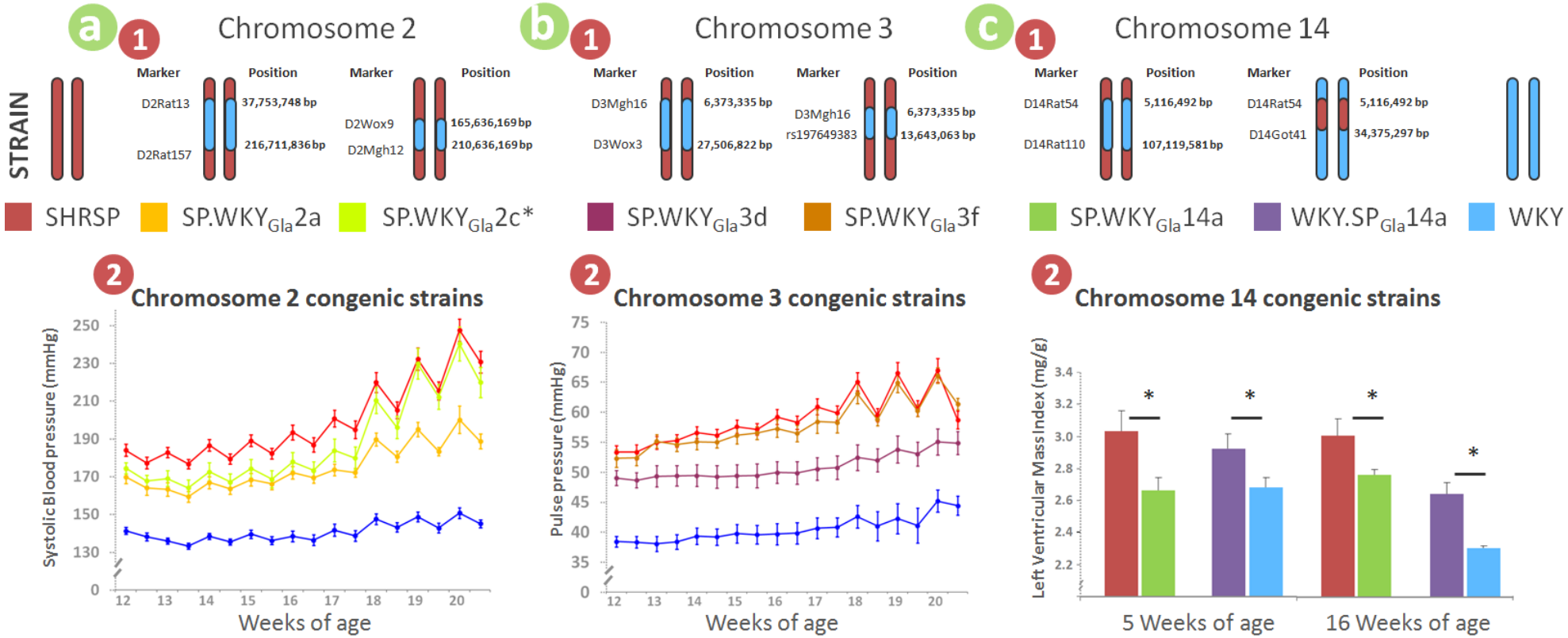


Figure 3-1 $SHRSP_{Gla}$, WKY_{Gla} and chromosome 2, 3 and 14 sub-congenic strains phenotype.
Schematic showing the position of microsatellite markers used in constructing chromosome **a1**) 2, **b1**) 3 and **c1**) 14 sub-congenic strains. **a2**) Systolic blood pressure among $SHRSP_{Gla}$, WKY_{Gla} and chromosome 2 sub-congenic strains. **b2**) Pulse pressure among $SHRSP_{Gla}$, WKY_{Gla} and chromosome 3 sub-congenic strains. **c2**) Left ventricular mass index among $SHRSP_{Gla}$, WKY_{Gla} and chromosome 14 sub-congenic strains. *P<0.05

Table 3-1 Summary of all gene expression profiling from different platforms of $SHRSP_{Gla}$, WKY_{Gla} and congenic strains.

		Neonatal	5 weeks		16 weeks		21 weeks	21 weeks salt challenged	
Kidney	Strain	×	$SHRSP_{Gla}$, WKY_{Gla}		$SHRSP_{Gla}$, $SP.WKY_{Gla}2a$, WKY_{Gla}		$SHRSP_{Gla}$, $SP.WKY_{Gla}2a$, WKY_{Gla}	$SHRSP_{Gla}$, $SP.WKY_{Gla}2a$, WKY_{Gla}	$SHRSP_{Gla}$, $SP.WKY_{Gla}3d$, $SP.WKY_{Gla}3f$, WKY_{Gla}
	Chip		Affymetrix RGU34 A,B,C		Affymetrix Rat230_2		Affymetrix Rat230_2	Affymetrix Rat230_2	Illumina
Heart	Strain	$SHRSP_{Gla}$, $SP.WKY_{Gla}14a$, $WKY.SP_{Gla}14a$, WKY_{Gla}	$SHRSP_{Gla}$, $SP.WKY_{Gla}14a$, $WKY.SP_{Gla}14a$, WKY_{Gla}	$SHRSP_{Gla}$, WKY_{Gla}	$SHRSP_{Gla}$, $SP.WKY_{Gla}14a$, $WKY.SP_{Gla}14a$, WKY_{Gla}	$SHRSP_{Gla}$, WKY_{Gla}	$SHRSP_{Gla}$, WKY_{Gla}	$SHRSP_{Gla}$, WKY_{Gla}	$SHRSP_{Gla}$, WKY_{Gla}
	Chip	Illumina	Illumina	Affymetrix Exon 1.0 ST	Illumina	Affymetrix Exon 1.0 ST	Affymetrix Rat230_2	Affymetrix Rat230_2	Illumina
Brain	Strain	×	$SHRSP_{Gla}$, WKY_{Gla}		$SHRSP_{Gla}$, WKY_{Gla}		$SHRSP_{Gla}$, WKY_{Gla}	$SHRSP_{Gla}$, WKY_{Gla}	
	Chip		Illumina		Illumina		Illumina	Illumina	
Aorta	Strain	×	×		$SHRSP_{Gla}$, WKY_{Gla}		×	×	
	Chip				Affymetrix RAE230 A and B				

3.3 Results

3.3.1 Improving and implementing new gene expression analysis workflows to identify candidate genes in the $SHRSP_{Gla}$

3.3.1.1 Partek® effect removal analysis on batch biased heart gene expression data between $SHRSP_{Gla}$, WKY_{Gla} and chromosome 14 congenic strains

PCA analysis was carried out on heart gene expression data between $SHRSP_{Gla}$, WKY_{Gla}, chromosome 14 congenics (SP.WKY_{Gla}14a called F for short and WKY.SP_{Gla}14a called D for short) across neonatal, 5 and 16 weeks of age. Overlaying the PCA plot with sentrix barcode (batches) and sample group used (Figure 3-2 A1&A2) shows that the age of rat strains divided the data into three sections, which is an excepted effect. However, batch bias is another effect observed in which each clustered at specific position regardless of sample group. After the batch bias was removed with Partek® effect removal tool, PCA plot overlaid with batches and sample group used (Figure 3-2 B1&B2) revealed a better sample group clustering and no batch biases compared to the PCA plot without batch effect removed (Figure 3-2 A1&A2).

To assess the improvement of Partek® effect removal tool, we compared WKY_{Gla} versus WKY.SP_{Gla}14a comparison at neonatal, 5 and 16 weeks with batch effect accounted for as factor in an ANOVA model against batch effect is removed by Partek® effect removal tool same with a cut-off of FDR <0.05 using Partek® Genomic suite software. The overall results show that Partek® batch removed analysis identified more significant genes including positional candidate genes than batch as factor analysis (Figure 3-2 A3&B3). Osteopontin (*Spp1*), a significantly differentially expressed positional gene, was identified by batch removed analysis at two time points and at 1 time point by batch as factor analysis. Previous qRT-PCR validation work by a colleague in our group on *Spp1* showed that the change in expression between WKY_{Gla} and WKY.SP_{Gla}14a at the same two time points identified by batch removal analysis (Figure 3-3).

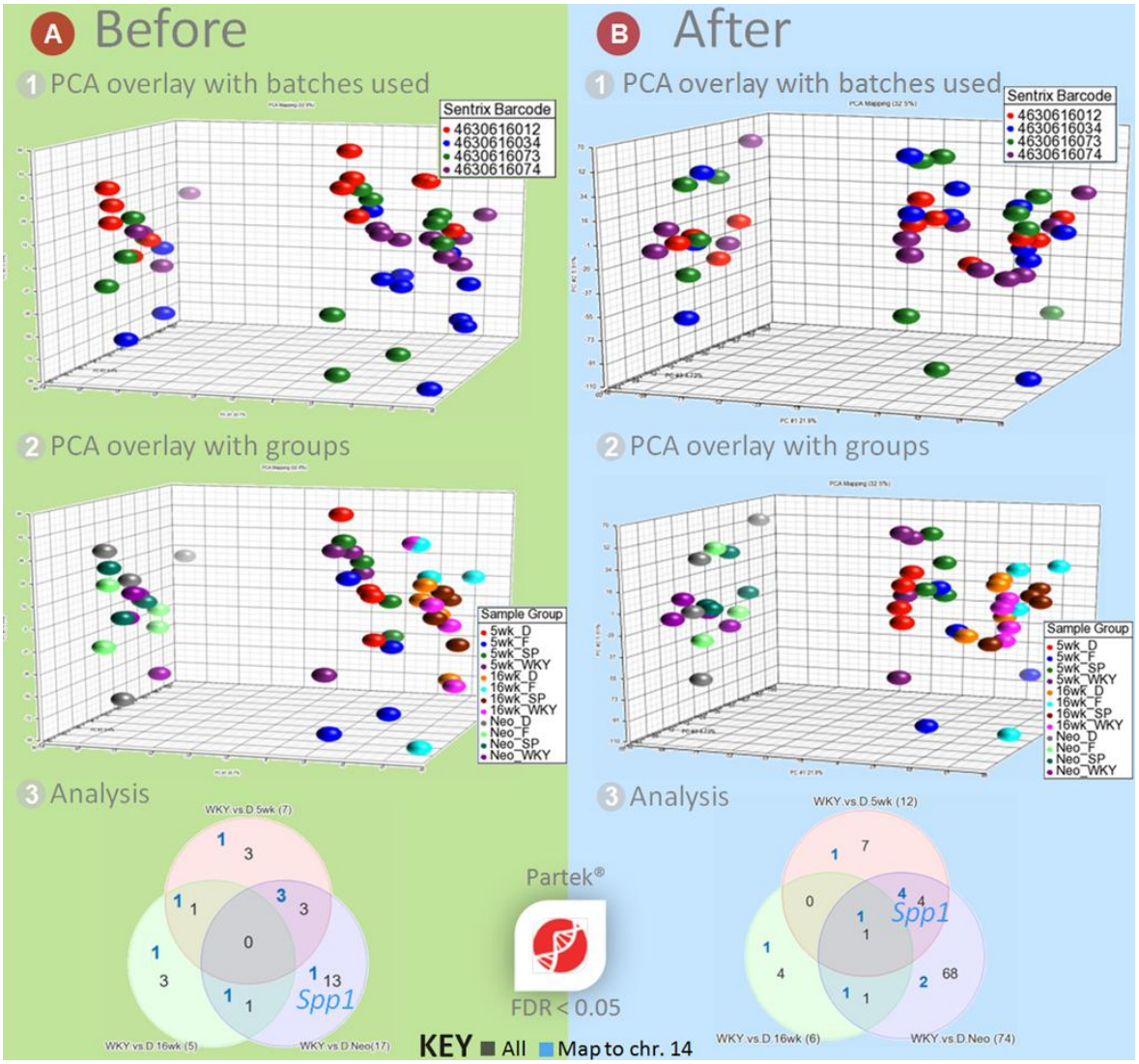


Figure 3-2 Partek® effect removal analysis on batch biased heart gene expression data between SHRSP$_{Gla}$, WKY$_{Gla}$ and chromosome 14 congenic strains.
PCA analysis on heart data from SHRSP$_{Gla}$, WKY$_{Gla}$ and chromosome 14 congenic strains at three different time points which are randomised across 4 array chips. **A1:** PCA plot with sentrix barcode (batch) shows that there is a minor chip bias in which each chip are clustered at specific location such as 4630616012 chip represented by red colour are at the top, 4630616034 chip represented in blue is at the bottom and the other chips are in the middle. This means a sample may show a bias, depending on the chip it is hybridized to. **A2:** PCA mapping of the sample groups (each strain at each time point) confirms the bias caused by chip location of each sample. **A3:** Venn diagram of WKY$_{Gla}$ versus WKY.SP$_{Gla}$14a (D) shows few significant differentially expressed genes in (black) especially positional candidate (blue) at different time points for example *Spp1* only significant at neonatal. **B1:** After removing the batch effect, overlaying the PCA plot with the sentrix barcode does not indicates chip bias where each chip is not clustered at specific position. **B2:** This is confirmed when the PCA plot overlaid with sample groups that shows a better groups clustering compared to the PCA plot with sample groups before removing the batch effect (A2). **B3:** Venn diagram of WKY$_{Gla}$ versus WKY.SP$_{Gla}$14a (D) shows more significant genes in all time points for instance *spp1* is significant at neonatal and 5 weeks of age.

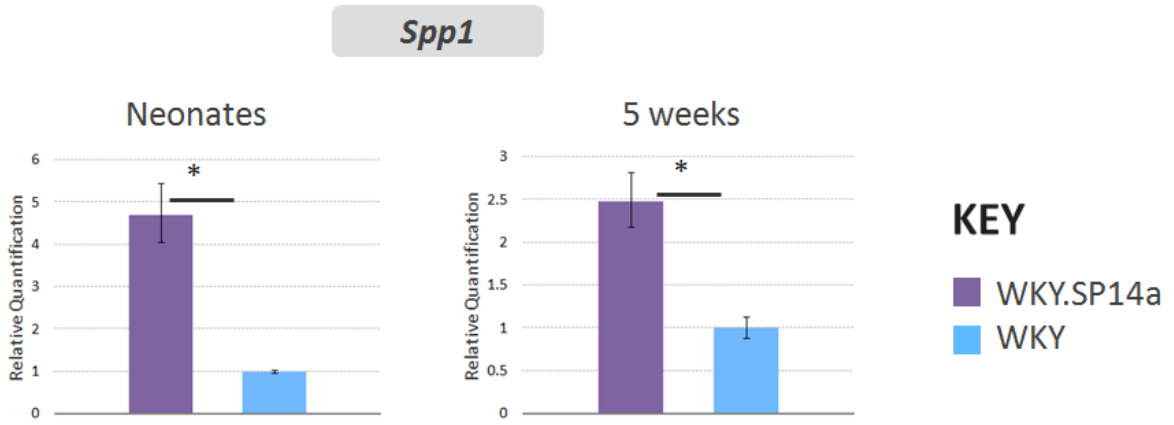


Figure 3-3 qRT-PCR of *Spp1* in heart between WKY_{Gla} and $WKY.SP_{Gla}14a$.
Spp1 significantly differentially expressed between WKY_{Gla} and $WKY.SP_{Gla}14a$ (D) in heart at neonatal and 5 weeks of age. *P<0.05. Data courtesy of Ms. Wendy Beattie.

In order to further investigate which type of batch bias corrections is more appropriate for our data, we compared the Partek® effect removal or Batch as factor in an ANOVA model. We looked at top diseases that could be related to $WKY.SP_{Gla}14a$ phenotype from genes identified in each method using IPA. At neonatal and 5 weeks of age, Partek® batch removal analysis identified more cardiovascular and inflammatory disease genes than accounting for batch bias as factor in an ANOVA model analysis. However, at the age of 16 weeks both analyses identified same disease names and gene numbers (Figure 3-4).

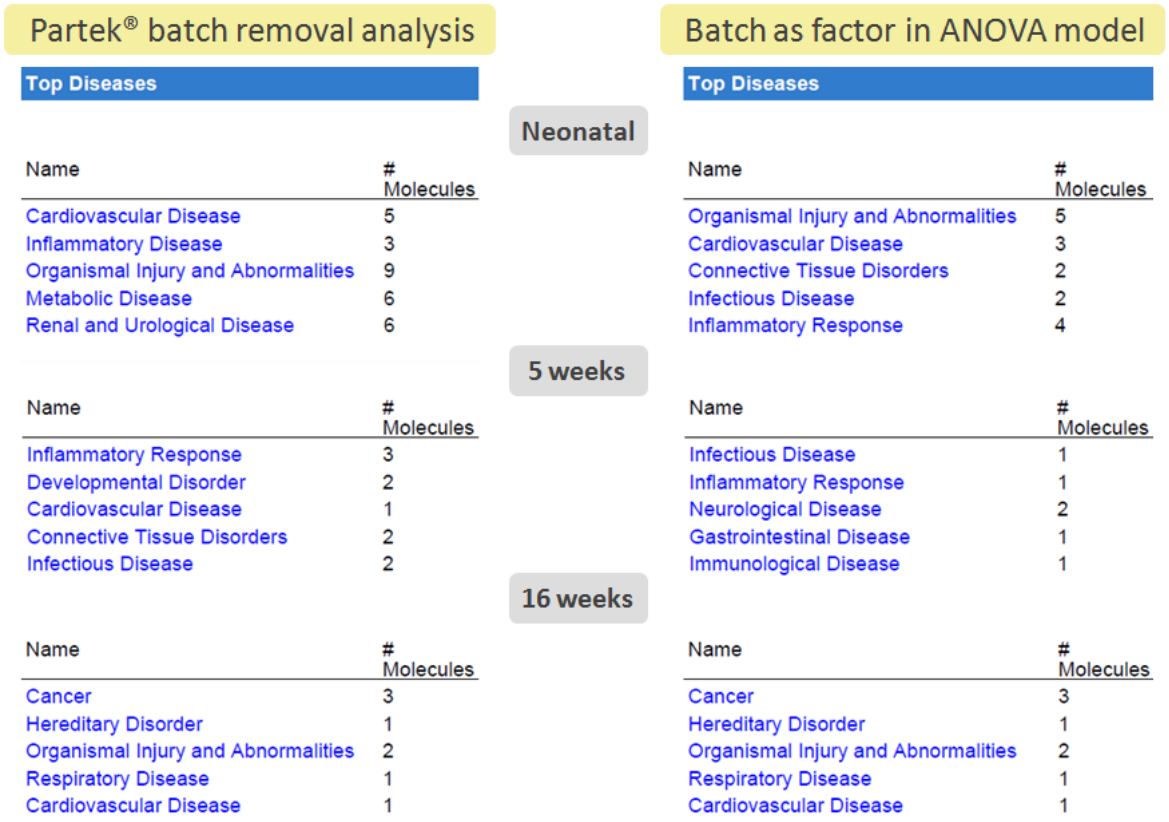


Figure 3-4 IPA top diseases analysis from two batch bias correction methods between WKY_{Gla} and $WKY.SP_{Gla}14a$ in heart.

Partek® batch removal analysis between WKY_{Gla} and $WKY.SP_{Gla}14a$ in heart at neonatal and 5 weeks of age implicated more cardiovascular and inflammatory disease genes than accounting for batch bias as factor in an ANOVA model analysis. At 16 weeks of age, both analyses have the same top disease name and genes number.

3.3.1.2 Combined expression chip platforms approach to identify renal positional candidate genes between $SHRSP_{Gla}$, WKY_{Gla} and chromosome 3 congenic strains

Partek® ANOVA analysis in the Illumina BeadChip expression data between $SHRSP_{Gla}$, WKY_{Gla}, and chromosome 3 congenic strains identified positional candidate differentially expressed genes mapped to the congenic interval common between SP.WKY_{Gla}3d and SP.WKY_{Gla}3f (Figure 3-5). These genes are dynamin1 (*Dnm1*) and torsin family 1 member B (*Tor1b*) (Table 3-2). Partek® ANOVA analysis in the Affymetrix GeneChip expression data between $SHRSP_{Gla}$ and WKY_{Gla} identified between GTPase activating protein 1 (*Rabgap1*) as the only gene significantly differentially expressed mapped to the unique SP.WKY_{Gla}3d congenic interval when compared to SP.WKY_{Gla}3f (Table 3-3). The Illumina RatRef Beadchip RatRef-12v1 does not contain a probe that targets the *Rabgap1* gene.

qRT-PCR validation was carried out by our group for the prioritised genes; *Dnm1, Tor1b* and *Rabgap1* from the microarray expression profiling data (Figure 3-6). Expression of *Dnm1* mRNA level was significantly higher in WKY_{Gla}, SP.WKY_{Gla}3d and SP.WKY_{Gla}3f kidneys compared to $SHRSP_{Gla}$ at 21-weeks salt-challenge. Expression of *Tor1b* mRNA level was significantly lower in WKY_{Gla}, SP.WKY_{Gla}3d and SP.WKY_{Gla}3f compared to $SHRSP_{Gla}$ at 21-weeks salt-challenge. Expression of *Rabgap1* mRNA level was significantly higher in WKY_{Gla} and SP.WKY_{Gla}3d compared to SP.WKY_{Gla}3f and $SHRSP_{Gla}$ at 21-weeks salt-challenge.

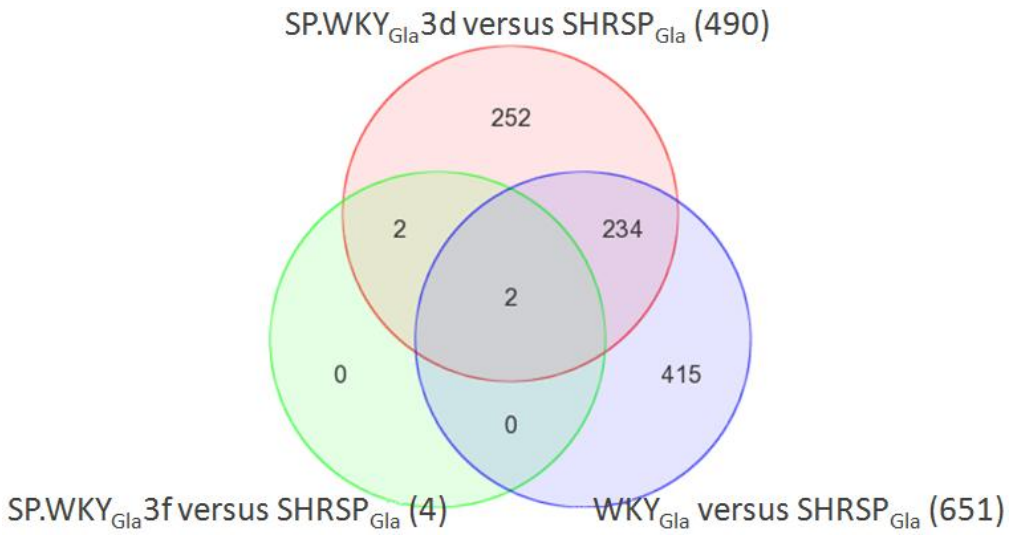


Figure 3-5 Significant Illumina probes between parental and chromosome 3 congenic strains in kidney at 21 weeks old of age salt.
Partek® ANOVA analysis between SHRSP$_{Gla}$, WKY$_{Gla}$, SP.WKY$_{Gla}$3d and SP.WKY$_{Gla}$3f in kidney at 21 weeks of age salt challenge. Two genes were in-common between WKY$_{Gla}$ versus SHRSP$_{Gla}$, SP.WKY$_{Gla}$3d versus SHRSP$_{Gla}$ and SP.WKY$_{Gla}$3f versus SHRSP$_{Gla}$ comparisons.

Table 3-2 Illumina results for positional candidate genes which are in common between SP.WKY$_{Gla}$3d and SP.WKY$_{Gla}$3f congenic intervals.

Rat			WKY$_{Gla}$ v.s SHRSP$_{Gla}$		SP.WKY$_{Gla}$3d v.s SHRSP$_{Gla}$		SP.WKY$_{Gla}$3f v.s SHRSP$_{Gla}$	
Gene	Position	Gene Description	FDR* (q-value)	Fold Change	FDR* (q-value)	Fold Change	FDR* (q-value)	Fold Change
Dnm1	chr3:11,434,778-11,478,452	Dynamin 1	0.0004	3.2	0.0003	3.9	0.0014	4.1
Tor1b	chr3:10,018,706-10,024,731	Torsin family 1, member B	0.0009	-1.2	0.0031	-1.2	0.0372	-1.2

Table 3-3 Affymetrix results for a positional candidate gene which is unique to the SP.WKY$_{Gla}$3d congenic strain interval.

Rat			WKY$_{Gla}$ v.s SHRSP$_{Gla}$	
Gene	Position	Gene Description	FDR* (q-value)	Fold Change
Rabgap1	chr3:17,204,810 - 17,322,894	RAB GTPase activating protein 1	0.001	-1.4

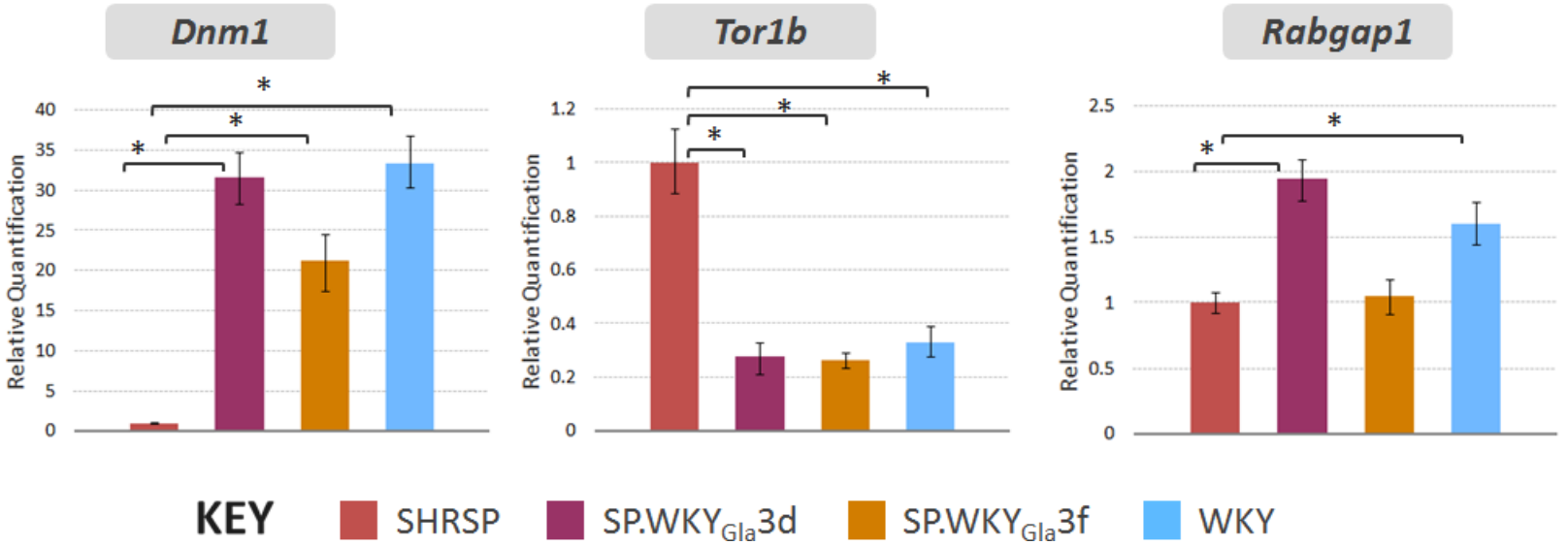


Figure 3-6 qRT-PCR of *Dnm1*, *Tor1b* and *Rabgap1* in kidney between SHRSP$_{Gla}$, WKY$_{Gla}$ and chromosome 3 congenic strains at 21 weeks of age salt-challenge.

qRT-PCR validated microarray gene profiling expressing results for the prioritised positional candidate genes *Dnm1*, *Tor1b* and *Rabgap1* between SHRSP$_{Gla}$, WKY$_{Gla}$, SP.WKY$_{Gla}$3d and SP.WKY$_{Gla}$3f in kidney at 21 weeks of age salt challenge. *P<0.05. Data courtesy of Mr Ting Wang and Ms. Wendy Beattie.

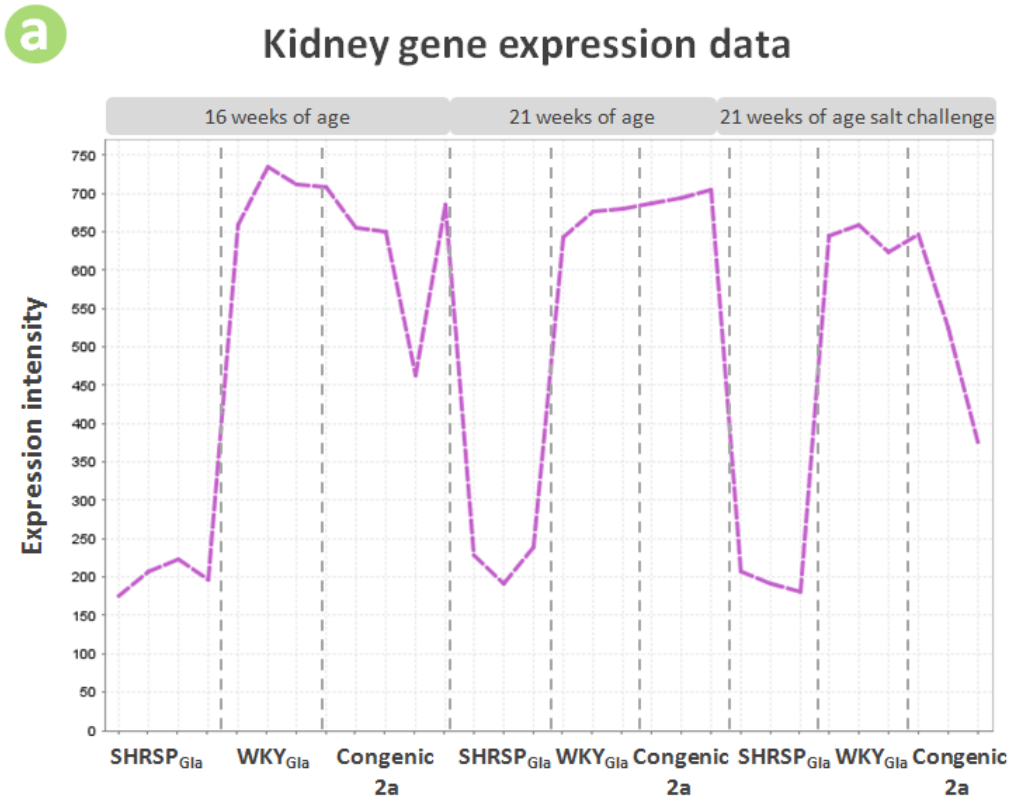
3.3.1.3 BioLayout*Express*^{3D} cluster analysis to identify renal positional candidate between SHRSP$_{Gla}$, WKY$_{Gla}$ and chromosome 2 congenic strains

BioLayout*Express*^{3D} cluster analysis on SHRSP$_{Gla}$ and WKY$_{Gla}$ and chromosome 2 congenic SP.WKY$_{Gla}$2a kidney resulted in 1 cluster (109) of 6 genes in which their pattern of expression in WKY$_{Gla}$ is the same in SP.WKY$_{Gla}$2a strain at 16 and 21 weeks of age with and without salt treatment (Figure 3-7a). Four of 6 genes in the cluster 109 mapped to the SP.WKY$_{Gla}$2a congenic interval (Figure 3-7b).

IPA functional and pathway analysis implicated *Gstm1*, *Gstm7* and *Pik3r1* with pathological functions and a canonical pathway that are directly related to the SHRSP$_{Gla}$ phenotype (Figure 3-8).

Positional candidate genes, *Atp11b* and *Pik3r1*, were prioritised for qRT-PCR validation between SHRSP$_{Gla}$, WKY$_{Gla}$ and SP.WKY$_{Gla}$2a at 21 weeks of age with and without salt treatment. As there is no longer SP.WKY$_{Gla}$2a kidney available at 16 weeks of age, SP.WKY$_{Gla}$2c* congenic strain was chosen instead at this time point with parental strains.

qRT-PCR results at 16 weeks of age of *Atp11b* and *Pik3r1* (Figure 3-9) revealed that their mRNA level is significantly differentially expressed between SHRSP$_{Gla}$ and WKY$_{Gla}$ and is not between SHRSP$_{Gla}$ and SP.WKY$_{Gla}$2c* strain which is expected as *Atp11b* and *Pik3r1* does not mapped to their congenic interval. At 21 weeks of age, *Pik3r1* mRNA level is significantly differentially expressed between SHRSP$_{Gla}$ and WKY$_{Gla}$, and SHRSP$_{Gla}$ and SP.WKY$_{Gla}$2a and *Atp11b* mRNA level is only significantly differentially expressed between SHRSP$_{Gla}$ and WKY$_{Gla}$. However, at 21 weeks of age salt challenge both mRNA level is significantly differentially expressed between SHRSP$_{Gla}$ and WKY$_{Gla}$, and SHRSP$_{Gla}$ and SP.WKY$_{Gla}$2a.



b

Cluster 109

Gene	Gene Description	Congenic 2a interval
Atp11b	ATPase, class VI, type 11B	✓
Cela1	Chymotrypsin-like elastase family, member 1	×
Gstm1	Glutathione S-transferase mu 1	✓
Gstm7	Glutathione S-transferase, mu 7	✓
Pik3r1	Phosphoinositide-3-kinase, regulatory subunit 1 (alpha)	✓
Vwf	Von Willebrand factor	×

Figure 3-7 BioLayout*Express*^{3D} Cluster 109.
a&b: A cluster (109) of 6 genes including 4 are mapped to the congenic interval SP.WKY$_{Gla}$2a in which their expression pattern in WKY$_{Gla}$ is the same in SP.WKY$_{Gla}$2a in relation to SHRSP$_{Gla}$ expression pattern.

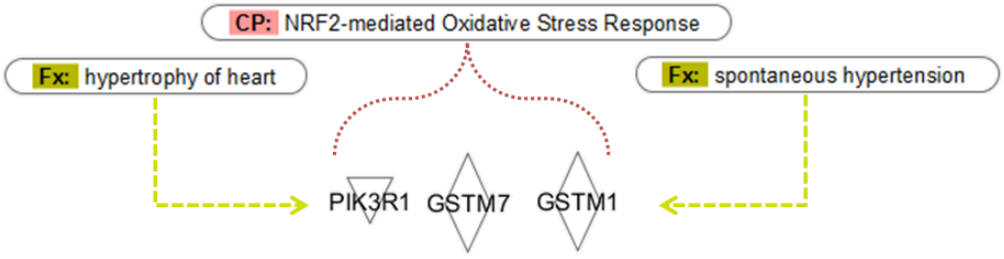


Figure 3-8 IPA biological and pathway analysis on cluster 109.
Pathological function of hypertrophy of the heart and spontaneously hypertensive both observed in SHRSP$_{Gla}$ phenotype. All of the three genes implicated in oxidative stress response pathway.

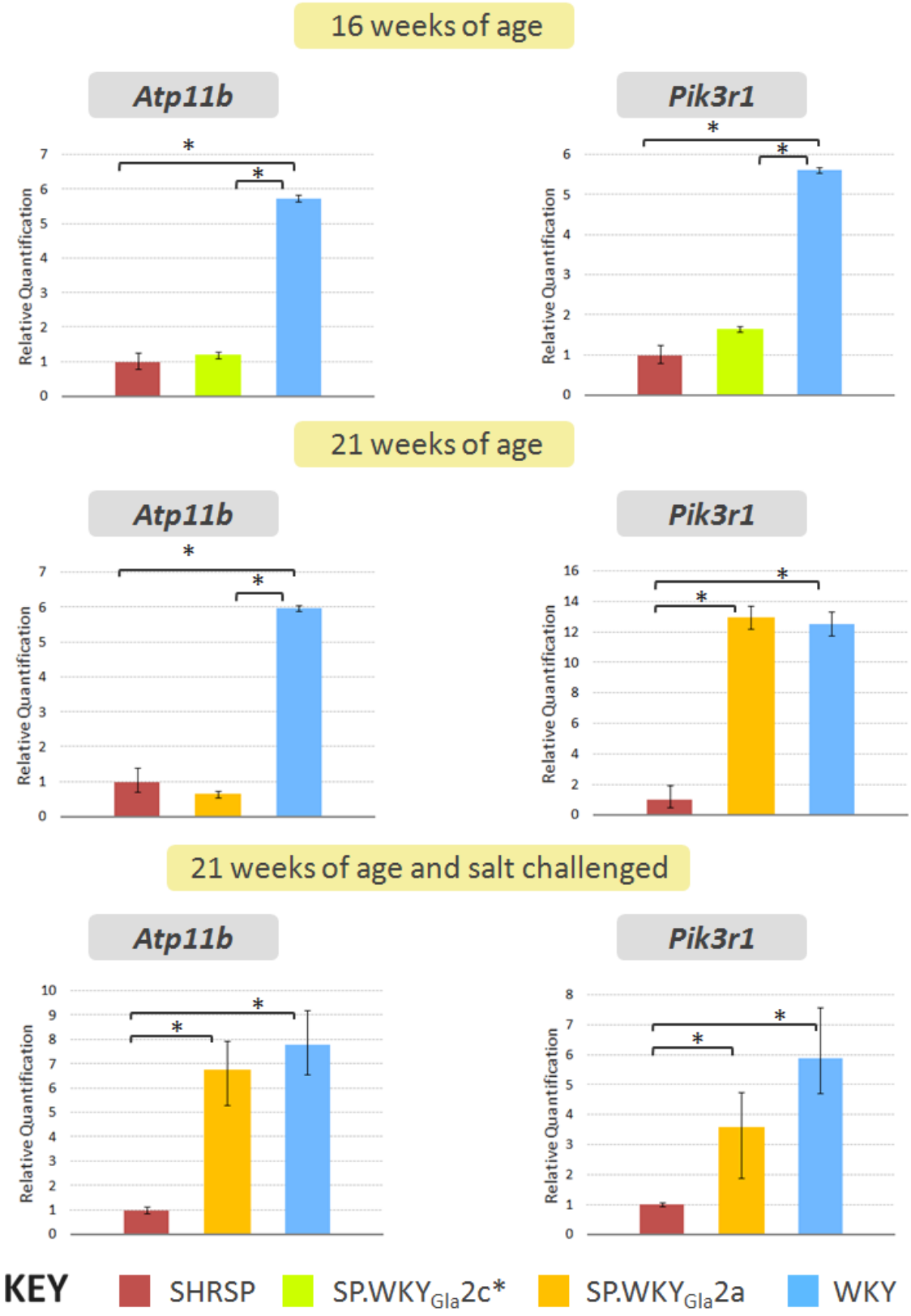


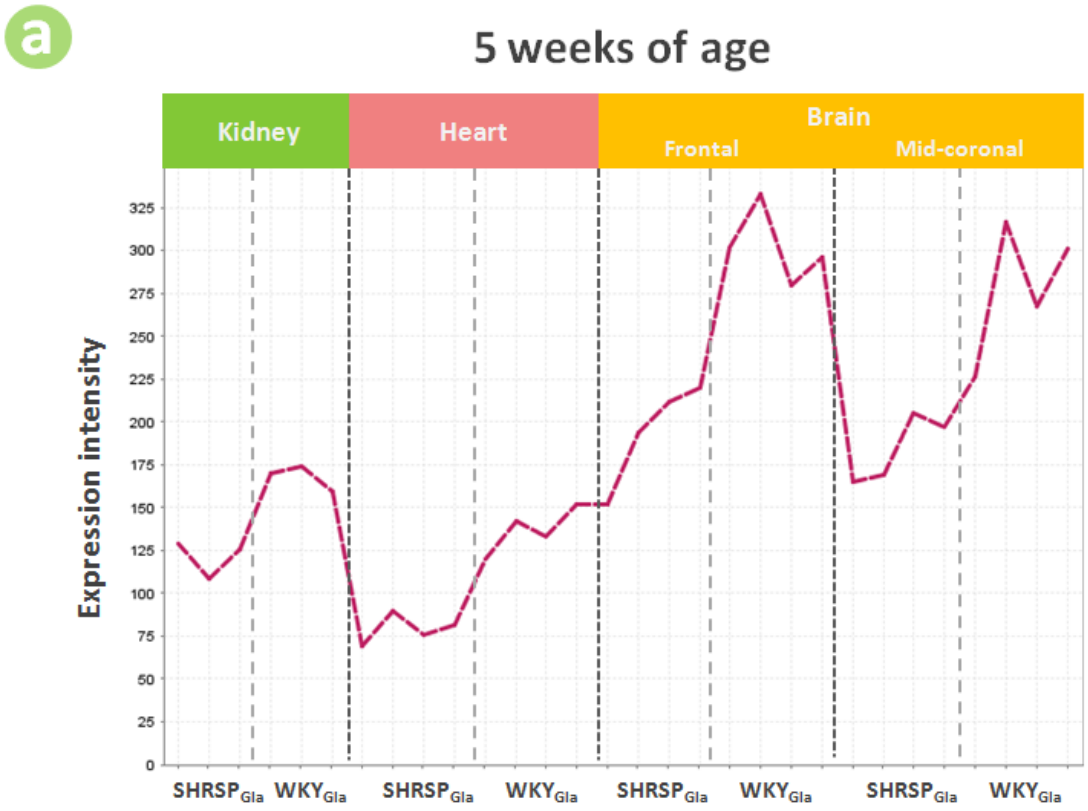
Figure 3-9 qRT-PCR validation of BioLayout*Express*^{3D} Cluster 109.
At 16 weeks, *Atp11b* and *Pik3r1* are significantly differently expressed between $SHRSP_{Gla}$ and WKY_{Gla}. At 21 weeks, *Atp11b* and *Pik3r1* are significantly differently expressed between $SHRSP_{Gla}$ and WKY_{Gla}; however, only *Pik3r1* is significantly differently expressed between $SHRSP_{Gla}$ and $SP.WKY_{Gla}2a$. At 21 weeks of salt challenge, *Atp11b* and *Pik3r1* are significantly differently expressed between $SHRSP_{Gla}$ and WKY_{Gla}, and $SHRSP_{Gla}$ and $SP.WKY_{Gla}2a$. *P<0.05

3.3.2 BioLayout*Express*^{3D} cluster analysis on gene expression data across multiple tissues and ages between SHRSP$_{Gla}$ and WKY$_{Gla}$

Utilising IPA probe annotation of different experimental gene expression platforms, we were able to run cluster analysis on common genes from multiple tissues gene expression data between SHRSP$_{Gla}$ and WKY$_{Gla}$ from different gene expression platforms.

3.3.2.1 BioLayout*Express*^{3D} cluster analysis on kidney, heart and brain gene expression data between SHRSP$_{Gla}$ and WKY$_{Gla}$ at 5 weeks of age

BioLayout*Express*^{3D} cluster analysis on kidney, heart and brain gene expression of SHRSP$_{Gla}$ and WKY$_{Gla}$ at 5 weeks of age identified one cluster (108) of genes where their expression intensity pattern is different between SHRSP$_{Gla}$ and WKY$_{Gla}$ (Figure 3-10).



b

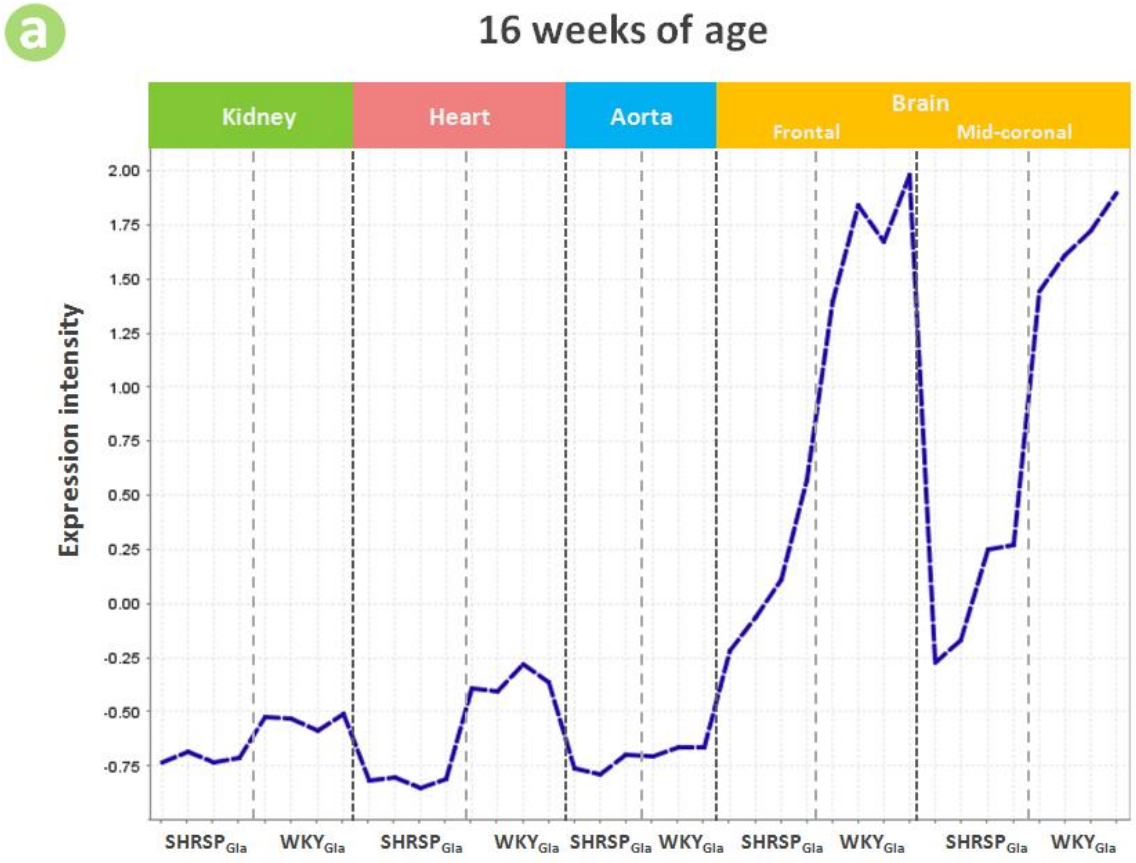
Cluster 108

Gene	Gene Description	Position
Atp11b	ATPase, class VI, type 11B	chr2:122,139,635-122,306,306
Iqub	IQ motif and ubiquitin domain containing	chr4:50,903,801-50,948,991

Figure 3-10 BioLayout*Express*^{3D} Cluster 108.
a&b: A cluster (108) of 2 genes in which their expression pattern in $SHRSP_{Gla}$ is different than WKY_{Gla} and consistent across kidney, heart, brain (frontal and mid-coronal) gene expression data at 5 weeks of age.

3.3.2.2 BioLayout*Express*^{3D} cluster analysis on kidney, heart, aorta and brain gene expression data between $SHRSP_{Gla}$ and WKY_{Gla} at 16 weeks of age

BioLayout*Express*^{3D} cluster analysis on kidney, heart, aorta and brain gene expression of $SHRSP_{Gla}$ and WKY_{Gla} at 16 weeks of age identified one cluster (53) of genes were their expression intensity pattern is different between $SHRSP_{Gla}$ and WKY_{Gla} (Figure 3-11).

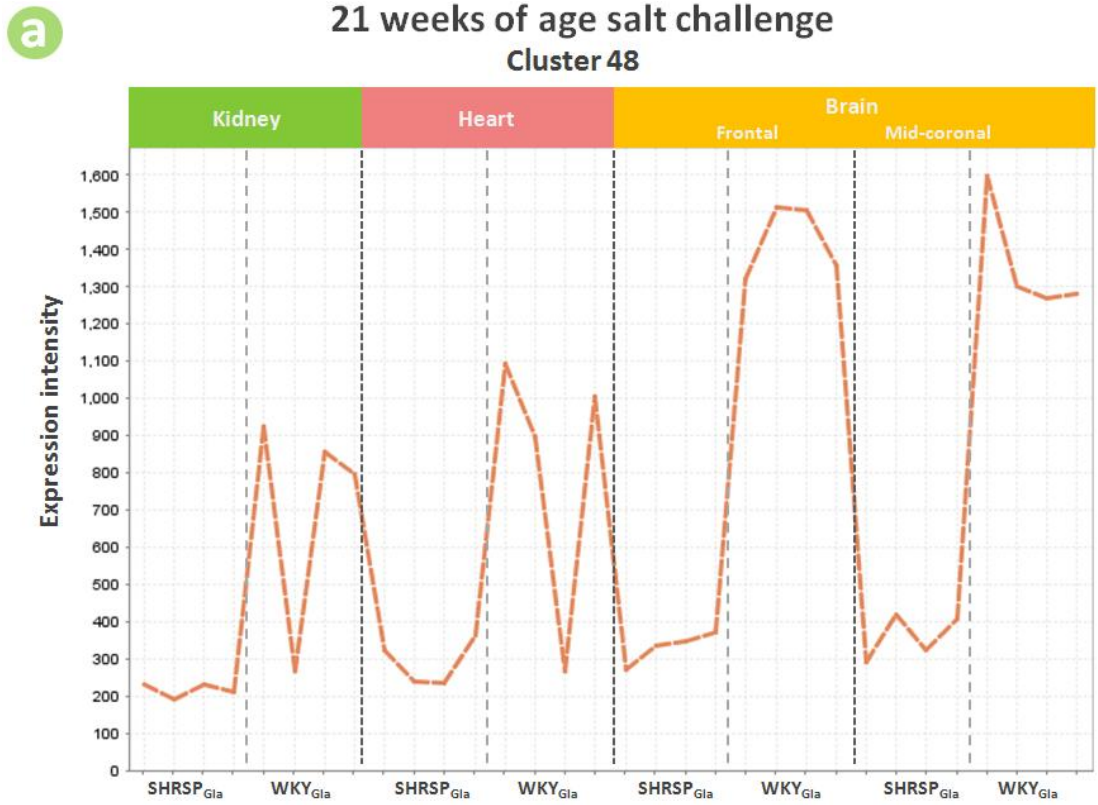


Gene	Gene Description	Position
Atp11b	ATPase, class VI, type 11B	chr2:122,139,635-122,306,306
Polr2i	Polymerase (RNA) II (DNA directed) polypeptide I	chr1:85,489,089-85,490,526
Mobp	Myelin-associated oligodendrocyte basic protein	chr8:125,118,618-125,144,099
Ndufaf5	NADH dehydrogenase (ubiquinone) complex I, assembly factor 5	chr3:128,353,866-128,384,213
Tcf4	Transcription factor 4	chr18:66,135,863-66,359,984

Figure 3-11 BioLayout*Express*^{3D} Cluster 53.
a&b: A cluster (53) of 5 genes in which their expression pattern in $SHRSP_{Gla}$ is different than WKY_{Gla} and consistent across kidney, heart, aorta and brain (frontal and mid-coronal) gene expression data at 16 weeks of age.

3.3.2.3 BioLayout*Express*^{3D} cluster analysis on kidney, heart and brain gene expression between $SHRSP_{Gla}$ and WKY_{Gla} at 21 weeks of age and salt challenge

BioLayout*Express*^{3D} cluster analysis on kidney, heart, and brain gene expression of $SHRSP_{Gla}$ and WKY_{Gla} at 21 weeks of age salt challenge identified 4 clusters (48, 104, 300, 410) of genes were their expression intensity pattern is different between $SHRSP_{Gla}$ and WKY_{Gla} (Figure 3-12, 3-13, 3-14, 3-15).

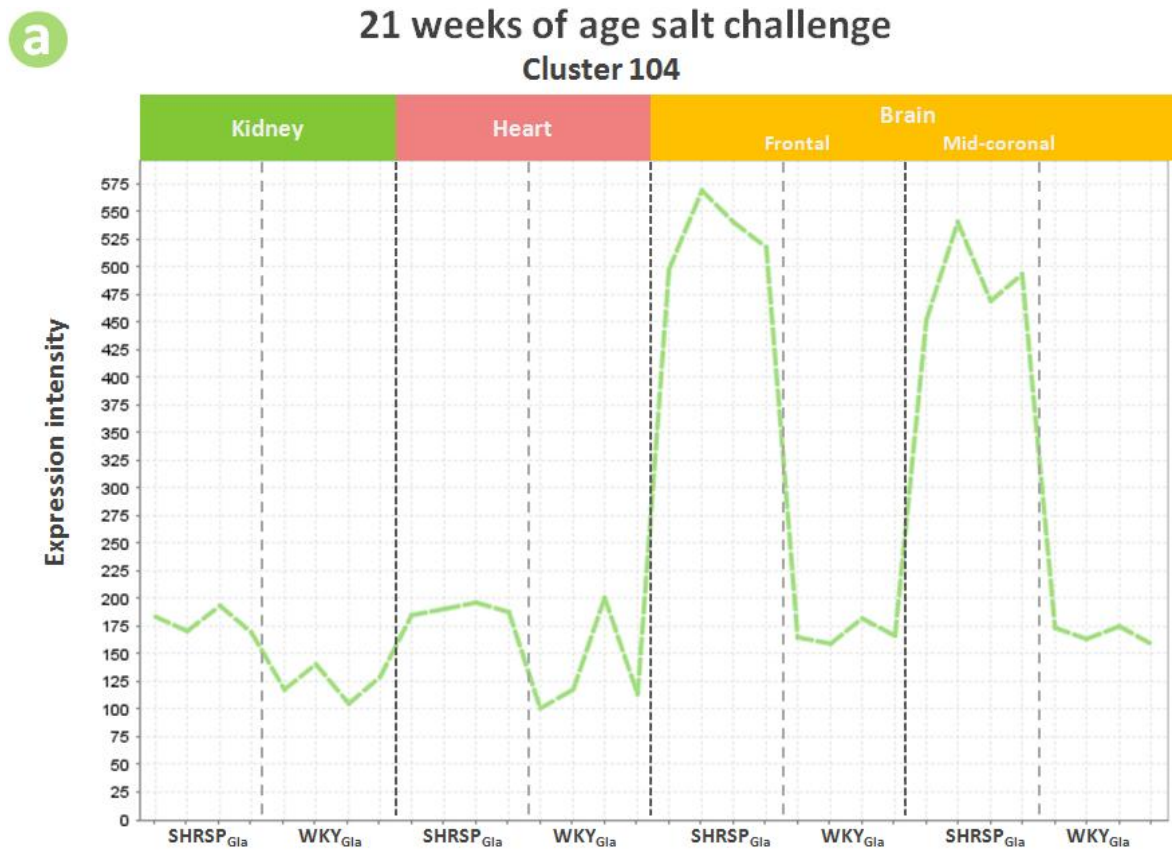


b

Cluster 48

Gene	Gene Description	Position
Atp11b	ATPase, class VI, type 11B	chr2:122,139,635-122,306,306
Rpl17l1	Ribosomal protein L17-like 1	chr1:64,205,419-64,205,973
Ndufaf5	NADH dehydrogenase (ubiquinone) complex I, assembly factor 5	chr3:128,353,866-128,384,213
Mrpl18	Mitochondrial ribosomal protein L18	chr1:42,111,647-42,116,465
Rbm8a	RNA binding motif protein 8A	chr2:191,429,805-191,432,261
Ttf1	Transcription termination factor, RNA polymerase I	chr3:8,067,017-8,086,185
Gpr98	G protein-coupled receptor 98	chr2:8,851,539-8,915,885
Haghl	Hydroxyacylglutathione hydrolase-like	chr10:15,050,382-15,053,046
Tmem106c	Transmembrane protein 106C	chr7:136,668,971-136,674,567
RGD1563903	Hypothetical gene supported by X60212	chr2:429,538-430,107
Vps13c	Vacuolar protein sorting 13 homolog C (S. cerevisiae)	chr8:72,196,264-72,365,531

Figure 3-12 BioLayout*Express*^{3D} Cluster 48.
a&b: A cluster (48) of 11 genes in which their expression pattern in $SHRSP_{Gla}$ is different than WKY_{Gla} and consistent across kidney, heart, brain (frontal and mid-coronal) gene expression data at 21 weeks of age salt challenge.

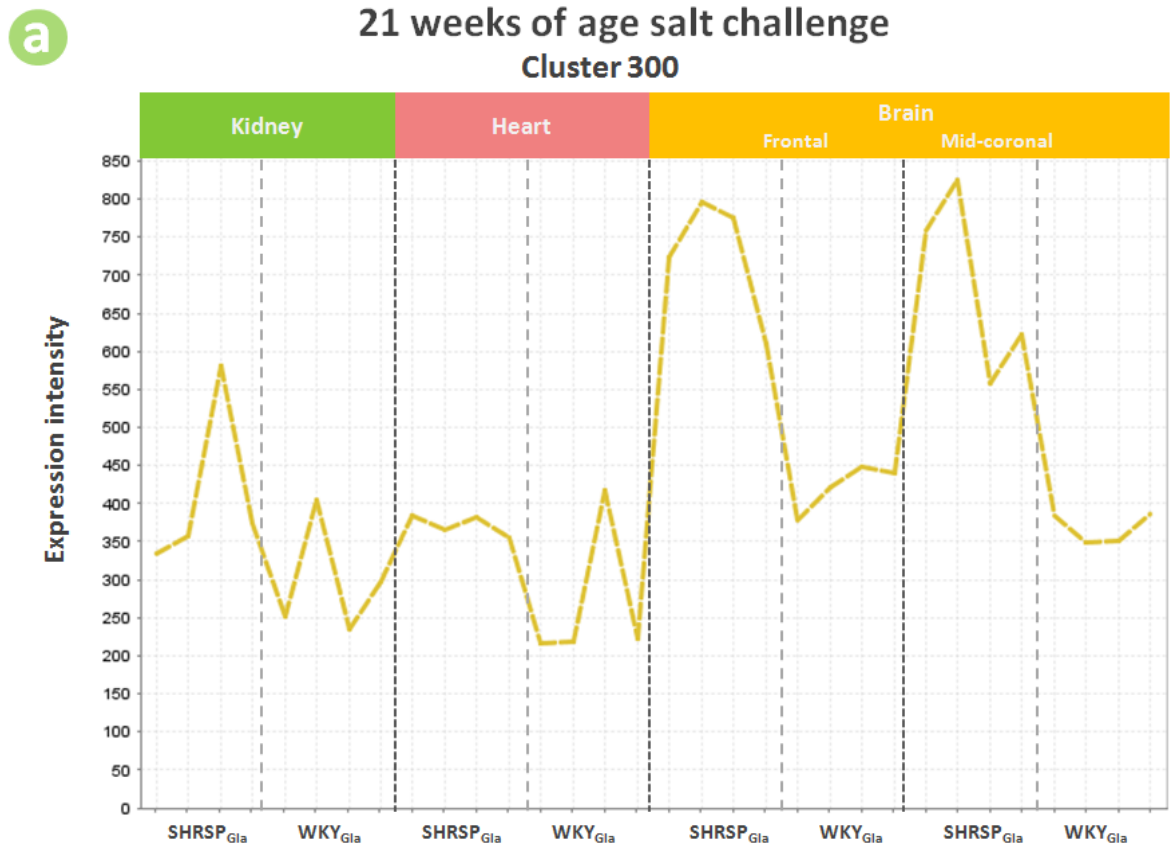


b

Cluster 104

Gene	Gene Description	Position
Ilf3	Interleukin enhancer binding factor 3	chr8:20,481,746-20,508,850
Zfp61	Zinc finger protein 61	chr1:79,504,886-79,512,625
Fam151b	Family with sequence similarity 151, member B	chr2:22,686,502-22,722,586
Zfp597	Zinc finger protein 597	chr10:11,920,710-11,926,384
Gucy1a3	Guanylate cyclase 1, soluble, alpha 3	chr2:173,756,824-173,818,316
Zfp566	Zinc finger protein 566	chr1:85,114,488-85,127,961
Adpgk	ADP-dependent glucokinase	chr8:63,120,989-63,148,896

Figure 3-13 BioLayout*Express*^{3D} Cluster 104.
a&b: A cluster (104) of 7 genes in which their expression pattern in $SHRSP_{Gla}$ is different than WKY_{Gla} and consistent across kidney, heart, brain (frontal and mid-coronal) gene expression data at 21 weeks of age salt challenge.

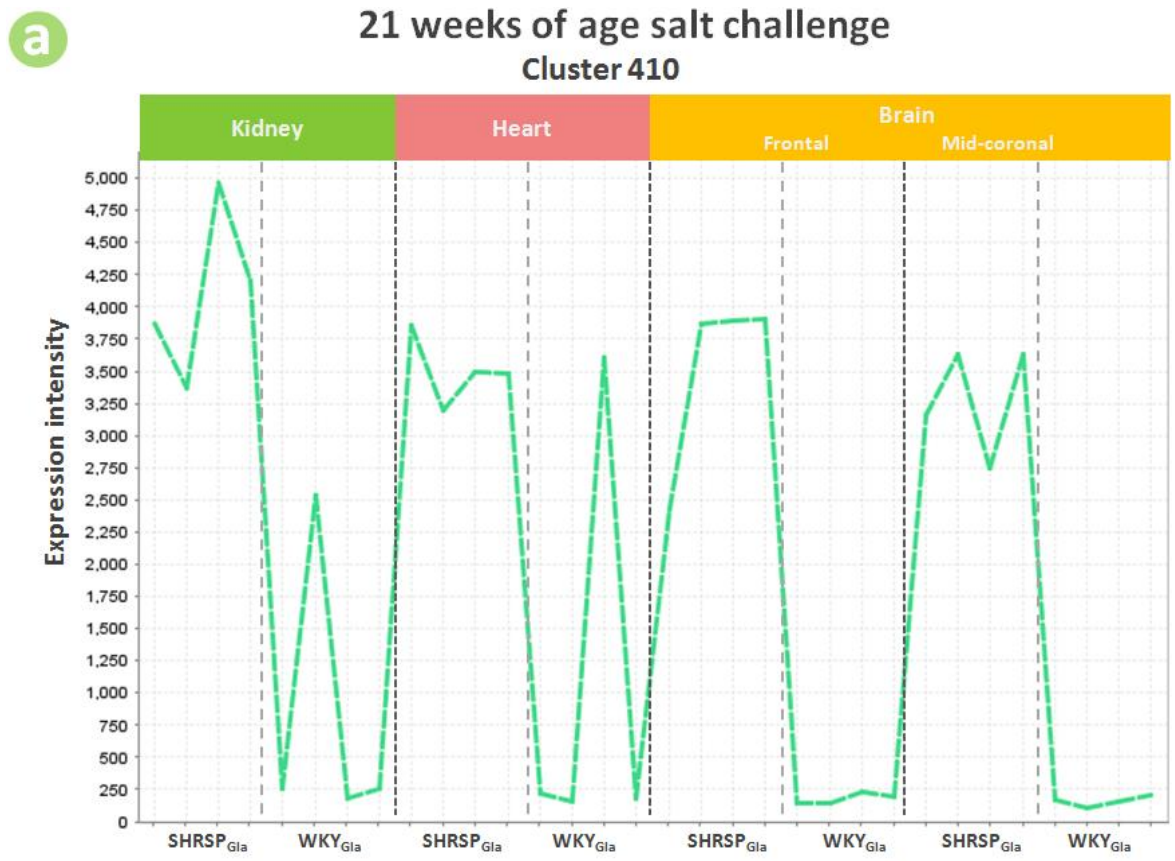


b

Cluster 300

Gene	Gene Description	Position
Fam32a	Family with sequence similarity 32, member A	chr16:18,025,468-18,029,788
Fkbp8	FK506 binding protein 8	chr16:19,401,883-19,407,192
Sort1	Sortilin 1	chr2:203,853,009-203,931,559
Poll	Polymerase (DNA directed), lambda	chr1:250,718,414-250,726,874
Rps16	Ribosomal protein S16	chr1:83,444,204-83,463,592

Figure 3-14 BioLayout*Express*^{3D} Cluster 300.
a&b: A cluster (300) of 5 genes in which their expression pattern in $SHRSP_{Gla}$ is different than WKY_{Gla} and consistent across kidney, heart, brain (frontal and mid-coronal) gene expression data at 21 weeks of age salt-challenged.



b

Cluster 410

Gene	Gene Description	Position
RGD1566136	Similar to 40S ribosomal protein S9	chrX:36,637,326-36,637,910
RGD1564649	Similar to 40S ribosomal protein S9	chr8:69,442,237-69,442,814
Rps9	Ribosomal protein S9	chr1:63,824,657-63,828,046

Figure 3-15 BioLayout*Express*^{3D} Cluster 410.
a&b: A cluster (410) of 3 genes in which their expression pattern in $SHRSP_{Gla}$ is different than WKY_{Gla} and consistent across kidney, heart, brain (frontal and mid-coronal) gene expression data at 21 weeks of age salt challenge.

3.3.2.4 Functional and validation analysis on BioLayout*Express*^{3D} clusters across multiple tissues and ages between $SHRSP_{Gla}$ and WKY_{Gla}

Functional analysis on BioLayout*Express*^{3D} clusters with genes differentially expressed between $SHRSP_{Gla}$ and WKY_{Gla} across tissues at 5, 16 and 21 weeks of age salt-challenged (Figure 3-16) all implicated genes associated with molecular transport and cardiovascular system function. Nevertheless, clusters with genes at the hypertensive and severely hypertensive phases implicated genes associated nervous system function and inflammatory response.

Atp11b was prioritised for qRT-PCR validation in heart as it has been identified in BioLayout*Express*^{3D} clusters across tissues at pre-hypertensive, hypertensive and severely hypertensive phases. qRT-PCR results showed that *Atp11b* mRNA level in heart is significantly differentially expressed between $SHRSP_{Gla}$ and WKY_{Gla} at neonatal, 5 and 16 weeks of age (Figure 3-17).

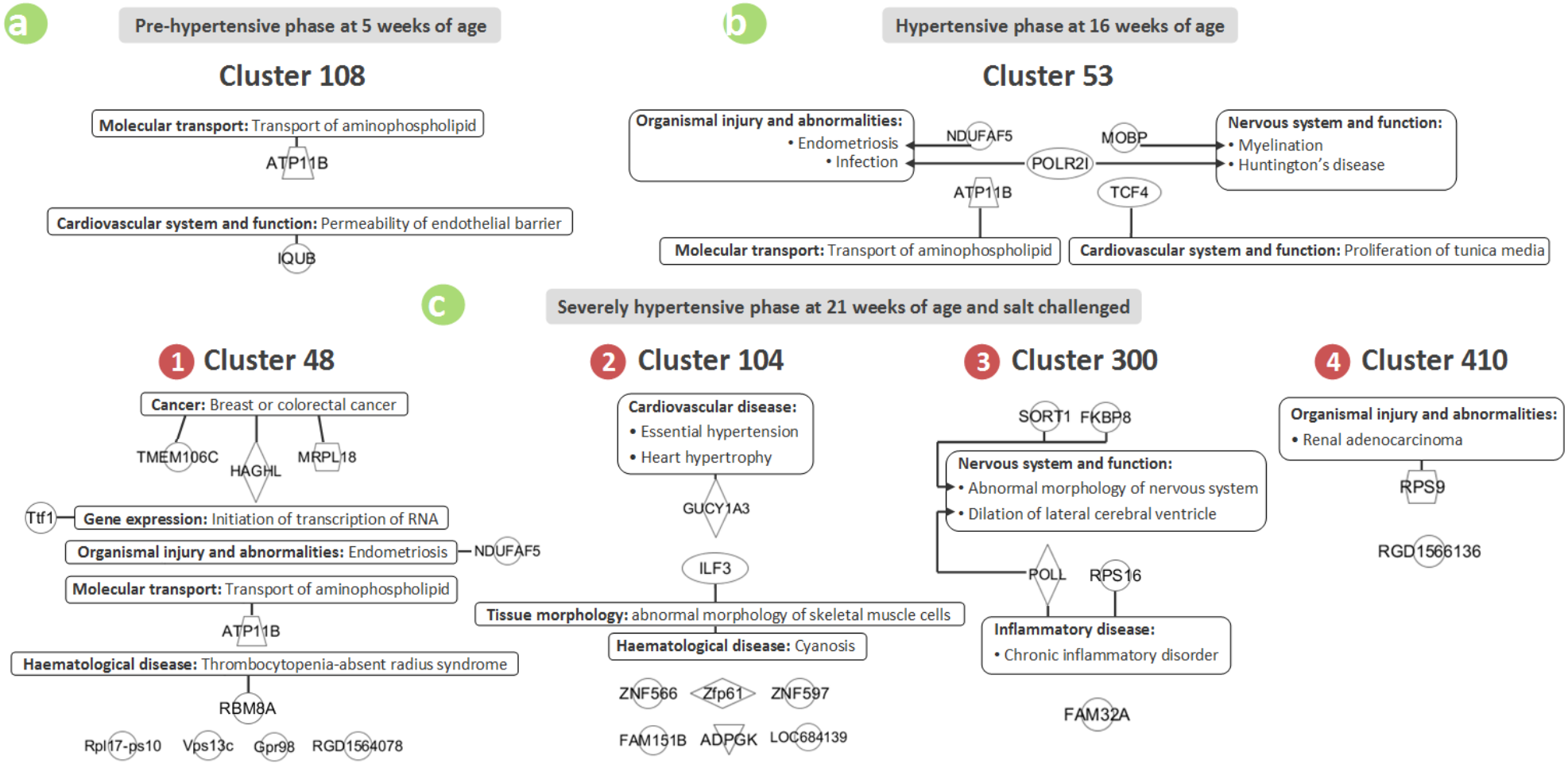


Figure 3-16 Functional analysis on BioLayout*Express*^{3D} clusters with genes differentially expressed between $SHRSP_{Gla}$ and WKY_{Gla} across tissues and ages.

IPA functional analysis on BioLayout*Express*^{3D} clusters across tissues and ages implicated genes with cardiovascular system function at **a**) 5, **b**) 16 and **c**) 21 weeks of age and salt challenge. IPA also implicated genes with nervous system function and inflammatory disease at 16 and 21 weeks of age and salt challenge.

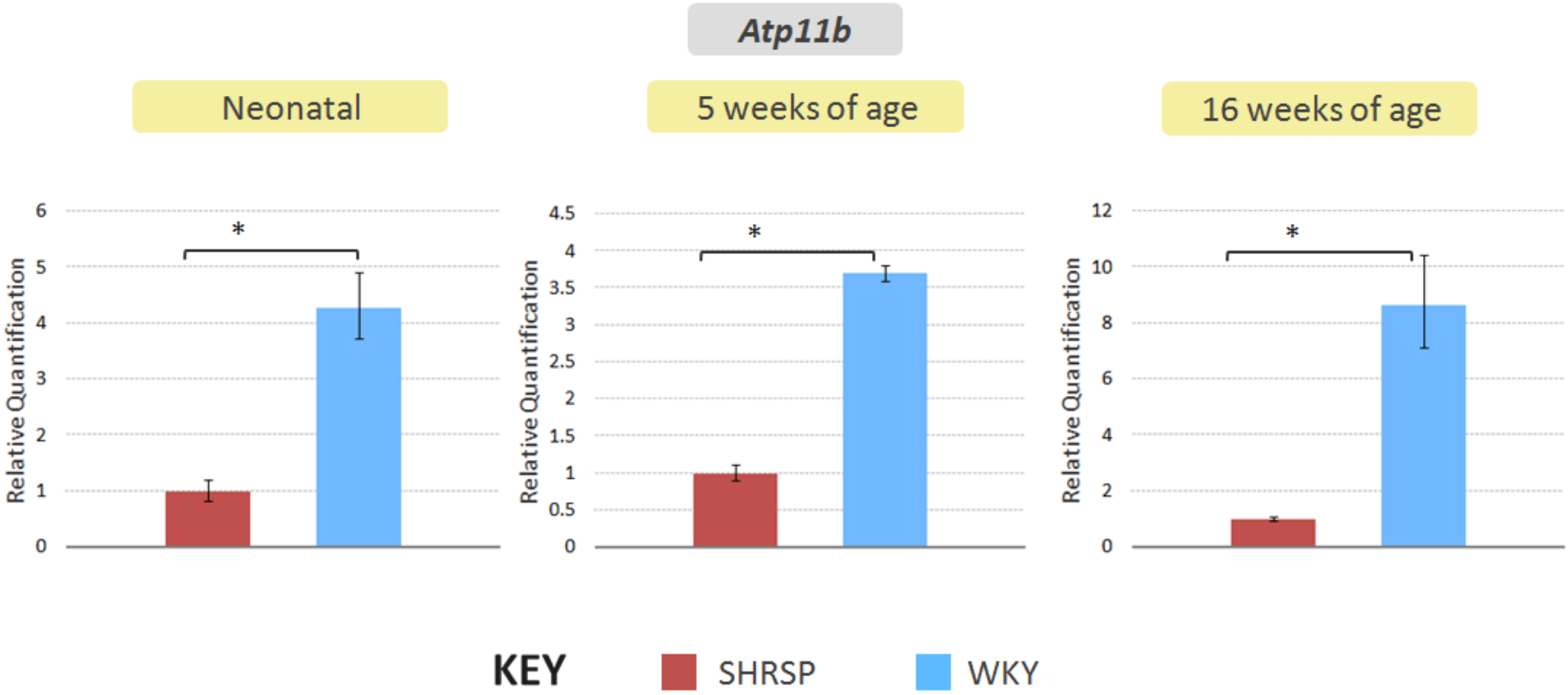


Figure 3-17 qRT-PCR of *Atp11b* in heart at neonatal, 5 and 16 weeks of age between $SHRSP_{Gla}$ and WKY_{Gla}.
Atp11b mRNA level is significantly differentially expressed between $SHRSP_{Gla}$ and WKY_{Gla} in heart at neonatal, 5 and 16 weeks of age. *P<0.05

3.4 Discussion

This chapter has investigated a number of gene expression analysis workflows to identify candidate genes for the $SHRSP_{Gla}$ phenotype.

The implementation of Partek® effect removal tool for batch bias in transcriptomic data analysis, identified more significant differentially expressed genes including positional candidate genes for LVMI between WKY_{Gla} and $WKY.SP_{Gla}14a$ when compared accounting for batch bias as factor in ANOVA model. Partek® batch effect removal analysis also enhanced the IPA functional analysis by identifying more cardiac function genes which are related to the $WKY.SP_{Gla}14a$ phenotype. qRT-PCR validation of positional candidate gene *Spp1* for LVMI (paper in preparation) between WKY_{Gla} and $WKY.SP_{Gla}14a$ at two time points suggests that using Partek® batch removal analysis gives better prediction of candidate genes compared to our regular analysis in which we could missed positional candidate genes for LVMI if we used one time point.

Improving a biased transcriptomic data is not the only strategy to identify potentially missed candidate genes for $SHRSP_{Gla}$ phenotype. The combination of Affymetrix and Illumina gene expression platform increased the efficiency of identification of renal positional candidate genes for pulse pressure where some could be missed if we relied only on one experimental platform.

The identified positional candidate *Dnm1* and *Tor1b* mapping to the congenic interval common between $SP.WKY_{Gla}3d$ and $SP.WKY_{Gla}3f$ and *Rabgap1* mapping to the congenic interval unique to $SP.WKY_{Gla}3d$ all not associated with essential hypertension before. *Dnm1* encodes Dynamin 1 protein that is essential for the formation of vesicles via clathrin-mediated endocytosis that regulates kidney filtration barrier maintenance (Soda et al., 2012), synaptic transmission (Ferguson et al., 2007) and mitochondrial fission (Lefevre et al., 2012). Hyndman et al., 2011 showed that Dynamin isoforms (*Dnm1*, *Dnm2* and *Dnm3*) interact with nitric oxide synthase (NOS) isoform in rat kidney and the inhibition of *Dnm2* led to increased NO production which is essential for the regulation of blood flow and pressure. *Tor1b* encodes Torsin B protein, which is found in the nuclear membrane and cytoplasmic endoplasmic reticulum, plays a role in synaptic

transmission (Kakazu et al., 2012). A mutation in *Tor1b* has been associated with torsion dystonia, a neurological condition in which muscles contracts involuntarily (Kamm et al., 2006). *Rabgap1* encodes for a protein that belongs to the Rab GTPase family, which they regulates a number of intracellular vesicle trafficking steps (Nagai et al., 2013).

BioLayout*Express*^{3D} cluster analysis between SHRSP$_{Gla}$, WKY$_{Gla}$ and chromosome 2 congenic SP.WKY$_{Gla}$2a in kidney at 16 and 21 weeks of age with and without salt challenge, identified a cluster (109) with new positional candidate genes that were not identified previous with our Rank Products (Graham et al., 2007) and meta-covariate cluster analysis (Hopcroft et al., 2010) on the same rat strains at 21 weeks of age and salt challenge. In addition, cluster (109) identified *Gstm1*, a previously validated functional and positional candidate gene for hypertension in the SHRSP$_{Gla}$ (McBride et al., 2005) providing proof of concept for our BioLayout*Express*^{3D} cluster analysis. Functional analysis of IPA on cluster 109 implicated some genes with pathological functions such as heart hypertrophy and spontaneous hypertension and oxidative stress canonical pathway were both phenotypes are observed in the SHRSP$_{Gla}$ were it is also a model for oxidative stress. The prioritised positional candidate genes *Pik3r1* and *Atp11b* for validation have biological functions that can be a causative for the SHRSP$_{Gla}$ phenotype. *Pik3r1* plays an important role in the metabolic actions of insulin, and a mutation in this gene has been associated with insulin resistance (Malodobra et al., 2011). Moreover, *Atp11b* transports ions across membranes and is part of a subfamily of P-type ATPases that transport ions such as H(+), Na(+), K(+), or Ca(+) (Apell, 2004) and are important in regulation of blood pressure. qRT-PCR results confirm that *Pik3r1* and *Atp11b* mRNA levels are significantly differentially expressed between parental strains and SP.WKY$_{Gla}$2a and SHRSP$_{Gla}$ at 21 weeks old and salt challenge. Furthermore, qRT-PCR results shows no change of mRNA levels of *Pik3r1* and *Atp11b* at 16 weeks of age between SHRSP$_{Gla}$ and an improved blood pressure congenic strain SP.WKY$_{Gla}$2c* (both genes do not map to SP.WKY$_{Gla}$2c* congenic interval) which may suggest these genes are not related to inflammatory and stress response. Unlike *Pik3r1*, *Atp11b* mRNA level does not show significant change between SP.WKY$_{Gla}$2a and SHRSP$_{Gla}$ at 21 weeks of age which may imply that *Atp11b* is salt response candidate gene in the SHRSP$_{Gla}$. However, the latter cannot be ruled out without

further qRT-PCR validation carried out between SP.WKY_{Gla}2a and $SHRSP_{Gla}$ in kidney at 16 weeks of age as well as earlier time points.

The success of the identification and validation of differentially expressed genes with BioLayout*Express*^{3D} encouraged us to carry out more complex cluster analysis on multiple and available functionally related tissues to $SHRSP_{Gla}$ at different hypertensive phases. BioLayout*Express*^{3D} cluster analysis identified clusters of genes with differentially expressed pattern between $SHRSP_{Gla}$ and WKY_{Gla} that is consistent across tissues at all hypertensive phases. IPA functional analysis on these clusters implicated genes related to cardiovascular system function and molecular transport from all hypertensive phases and nervous system function, organismal injury and abnormalities and inflammatory response at hypertensive and severely hypertensive phases were all captured $SHRSP_{Gla}$ phenotype.

BioLayout*Express*^{3D} cluster analysis across tissues identified a cluster of genes in which their pattern of expression is high in WKY_{Gla} compared to $SHRSP_{Gla}$ at all hypertensive phases in which *Atp11b* was common in all these clusters. This may suggest that *Atp11b* is driving disease as it is differentially expressed from an early age on a number of tissues that can regulate and influence blood pressure. The latter is confirmed by qRT-PCR results of *Atp11b*, which shows that it is significantly differentially expressed between $SHRSP_{Gla}$ and WKY_{Gla} in kidney and heart at different time of age. At severely hypertensive phase 3 clusters identified by BioLayout*Express*^{3D} in which their genes expression pattern is high in $SHRSP_{Gla}$ compared WKY_{Gla}. One of these genes in these clusters is *Gucy1a3,* that has been shown to cause heart hypertrophy by preventing the pulmonary vasodilator response to NO in mice knockout experiment (Vermeersch et al., 2007) and it has been associated with blood pressure in Genome wide association study in human (Ehret et al., 2011).

BioLayout*Express*^{3D} cluster analysis is fast to carry out in and all the presented data takes less than an hour for each, whereas in our meta-covariate cluster analysis it took months to get results for only small set of data (Hopcroft et al., 2010). However, BioLayout*Express*^{3D} is not efficient in finding clusters with gene expression pattern related to the user criteria. As a result, the user has to

browse hundreds of possible clusters expression patterns in order to find one of interests. Moreover, our BioLayout*Express*^{3D} cluster analysis across tissues at 5 and 16 weeks of age was limited to genes that were common between gene expression platforms from different versions of Affymetrix chips and Illumina platform. However, the BioLayout*Express*^{3D} cluster analysis across tissues at 21 weeks of age and salt challenge used one gene expression platform (Illumina), which may reduce the chances of missing genes that contribute to the $SHRSP_{Gla}$ phenotype. Like all clustering methods, follow-up experiments are using qRT-PCR is necessary to differentiate between the highlighted clusters (i.e genes) if they are truly useful or just spurious.

4 Whole Genome Next Generation Sequencing of $SHRSP_{Gla}$ and WKY_{Gla} Rat Strains

4.1 Introduction

Currently, the term Next generation sequencing (NGS) used to described sequencing technologies that reads DNA templates in a massively parallel manner which generates short reads (35-500 bp) randomly along the whole genome compared to the very low throughput of Sanger sequencing (Zhang et al., 2011). NGS of the whole genome or a targeted region is more efficient in terms of cost, time and effort when compared to traditional Sanger sequencing. NGS can detect structural variations, such as large deletions, as well as CNV, in addition to the detection of high-throughput SNPs and INDELs (Metzker, 2010). The application of NGS for the analysis of human (Ahn et al., 2009), mouse (Keane et al., 2011; Yalcin et al., 2011) and rat (Atanur et al., 2010; Atanur et al., 2013; Guo et al., 2013) in research has been successful and recently been used as a genetic diagnostic tool (Shanks et al., 2013). A number of platforms offer NGS, most of which are based on common principles; enrichment of the genome, template library preparation, sequencing and imaging (Figure 4-1), however, they differ in sequencing technology (Metzker, 2010). Illumina NGS templates can be short paired-end (PE) reads (two ends of the same DNA fragment) or long mate pair reads, with each variation having its advantages. We used the PE template since it is standard for most platforms (Van Heesch et al., 2013) and therefore will be more informative when compared to other sequenced rat strains so as to isolate genomic variants for a phenotype and/or phylogenetic analysis (Atanur et al., 2013). In order to do this, NGS of rat strains reads should to be aligned to an accurate reference genome sequences instead of *de novo* assembly.

Brown Norway ($BN_{SsNHsd/Mcwi}$) rat strain is the reference genome; it has been used widely as control strain for a number of fields and studies as well as it is the founder of RI rat strains. Baylor College of Medicine and Celera genomics constructed the initial draft genome of BN Rat Genome Sequencing Consortium (RGSC) 3.1 with a combination of Whole Genome Shotgun (WGS) sequencing with a bacterial artificial chromosome (BAC) refined by DNA fingerprints, genetics markers and syntenic regions (Worthey et al., 2010). Further improvements of Atlas assembler version, targeted problematic regions and draft sequence of Y chromosome led to the release of draft sequence RGSC 3.4, however, the Y chromosome was not included (Gibbs et al., 2004). Both RGSC 3.1 and 3.4

suffered from significant gaps (Worthey et al., 2010). Recently, RGSC released an upgraded version of the assembly Rnor 5.0 with

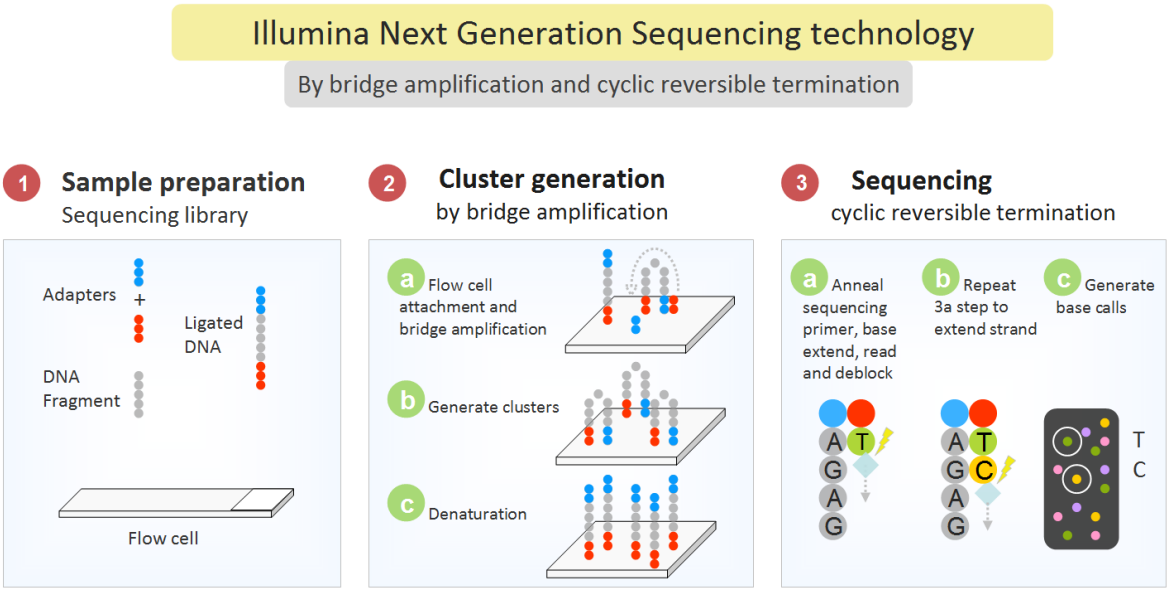


Figure 4-1 Overview of NGS Illumina Technology
1) Sample preparation begins with sequencing library by adapters ligate at opposite ends of DNA fragment and the use of Illumina flow cell. **2a**) Primers on flow cell capture the adapters of DNA fragment and immobilised the template on flow cell surface then bridge amplification is performed. **2b**) the amplification form dense DNA clusters on flow cell channel. **2c**) Denaturation of the bridges leaves one strand anchored to the flow cell surface ready for sequencing. **3a**) Universal sequencing primer anneal to the DNA fragment adapter and first base extension is performed with reverse termination of four deoxynucleotides carrying different flurophores and a blocking group to terminate the extension of DNA polymerase after incorporating 1 identified nucleotide by capturing the image of the laser activated fluorescence. **3b**) The latter step is repeated after washing down the unincorporated nucleotides and cleavage step removes both the blocking group and the fluorescent dyes. **3c**) The resulting images are used to generate base calls of each DNA cluster sequence (Original diagram compiled from information on Metzker, 2010).

improvement in the total length and gaps before scientific publication. To present day, Y chromosome sequence is not included in any BN assembly versions. This might be due to the fact that BN Y chromosome, includes polymorphism, is significantly larger than other rat strains (Kren et al., 2001).

No and/or low coverage (average number of times a nucleotide base is read) of quantity and quality of reads are likely to lead to missed sequence variants or sequencing error. Hence, assessing average coverage and NGS sequence variants with variants from other sequencing methods as a means of quality control, is very important. Moreover, mapping the NGS reads to a newer assembly will improve sequencing coverage and thus, increase accuracy of detection genomic variants.

The rate of BN assembly updates compared to the mouse and human reference assemblies is slow and does not meet standards of a finished level genome (Worthey et al., 2010). As a result, Guo et al., 2013 constructed genomes draft of their own a rat strains via *de novo* assembly aided by BN reference genome and GapCloser tool achieving larger genomes than the BN, and eventually leading to the detection of novel genomic variants. Nevertheless, these novel variants will not have functional significance since functional annotation of sequence variants is assessed by BN reference genome.

Annotation tools such as Ensembl variant effect predictor (McLaren et al., 2010) of sequence variants, are essential for identifying disease causative genes. Consequence effect of sequence variants can be categorised into protein coding and non-coding sequences, a change in these classes can induce a phenotype. Predicted evolutionary conserved protein coding consequences, such as non-synonymous, stop gain and/or loss, and frameshift variants, are usually prioritised according to which change is more likely to influence protein function (Cooper & Shendure, 2011). Predicted non coding consequences are also important since they can alter gene expression. For instance, upstream and 5'-UTR variants can disturb the transcription factor binding site or epigenetic regulation and 3'-UTR consequence can disturb miRNA binding regions causing mRNA destabilisation followed by degradation, and/or by mRNA translational repression (Boffa et al., 2008; Guo et al., 2013). On the other hand, variants in intronic and intergenic are usually given a lower priority. Recently, Encyclopedia

of DNA Elements (ENCODE) project suggested that such regions are functional (Ian et al., 2012); however, some evolutionary biologists disagree (Graur et al., 2013). The release of the BN reference genome Rnor 5.0 led to a major upgrade in Ensemble variant effect predictor version 71 classification of consequences and offered new prediction tools such as Sorting Intolerant From Tolerant (SIFT) algorithm to assess likely damaging non-synonymous SNPs by using conserved homology sequence (Kumar et al., 2009).

The first whole genome inbred rat strain sequenced by NGS is the spontaneously hypertensive rat ($SHR_{OlaIpcv}$) which was compared to BN reference RGSC 3.4 (Atanur et al., 2010). Atanur et al., 2010 identified in the $SHR_{OlaIpcv}$ more than 3 million SNPs with high level of sensitivity and specificity validation at low (10x) coverage. The $SHR_{OlaIpcv}$ whole genome had another NGS update which improved the coverage to 23x and led to the identification of new sequence variants that might be causative for the $SHR_{OlaIpcv}$ phenotype (Simonis et al., 2012). Atanur et al., 2010 success in carrying out NGS, encouraged other groups, including ours, to collaborate and sequence a number of rat strains in order to identify genetic determinants which underlie human complex diseases and to carry out phylogenetic analysis (Atanur et al., 2013). Nevertheless, sequencing $SHRSP_{Gla}$ on its own against BN reference will generate a large number of genomic variants. Moreover, preliminary analysis between $SHRSP_{Gla}$ and WKY_{Gla} sequence variants from targeted sequencing (sequence capture) and STAR project consortium sequencing data (Saar et al., 2008) estimated that over 40% of $SHRSP_{Gla}$ sequence capture SNPs would not be informative as they are common with WKY_{Gla}. As a result, sequencing WKY_{Gla} is necessary to prioritise genomic variants. In addition, $SHRSP_{Gla}$ and WKY_{Gla} with the publically available sequence of $SHR_{OlaIpcv}$ will further prioritise candidate genes for hypertension; common variants between $SHR_{OlaIpcv}$ and $SHRSP_{Gla}$, and stroke; unique $SHRSP_{Gla}$ variants. The latter can be further prioritised with other strains of SHR along other hypertensive rats that were sequenced in the phylogenetic study of 28 rat genomes (Atanur et al., 2013) and non-synonymous SNPs of $SHRSP_{Izm}$ compared to SHR_{Izm} against BN reference RGSC 3.4 at two QTLs on chromosome 1 and 18 for stroke latency (Gandolgor et al., 2013). Nevertheless, $SHRSP_{Gla}$ was the only rat strain in that study with unique phenotypes of severe hypertension and stroke prone (Atanur et al., 2013).

4.2 Aims

This chapter specific aims are:.

- Identification and annotation of genomics variants between $SHRSP_{Gla}$ and WKY_{Gla} genome sequences compared the BN reference assemblies and isolating variants at congenic intervals for blood pressure on chromosome 2 QTL, pulse pressure on chromosome 3 QTL, LVH on chromosome 14 QTL.

- Estimate genomic variants' distribution and frequencies between $SHRSP_{Gla}$ and WKY_{Gla} compared to the reference BN RGSC 3.4.

- Provide independent validation of genome sequencing of $SHRSP_{Gla}$ and WKY_{Gla} using sequence variants from STAR project consortium SNP genotyping data, sequence capture and Sanger sequencing.

- Identify a strategy that allows for comparison, at the nucleotide level, between the closely related $SHR_{OlaIpcv}$ and $SHRSP_{Gla}$ strains.

- Identify $SHRSP_{Gla}$ unique and deleterious protein coding sequence variants against to 27 rat strains genome compared to BN reference genome RGSC 3.4 and prioritising common $SHRSP_{Izm}$ variants.

4.3 Results

4.3.1 Sequence variants between $SHRSP_{Gla}$ and WKY_{Gla} compared to BN reference RGSC 3.4

4.3.1.1 Variants calling between $SHRSP_{Gla}$ and WKY_{Gla}

When both strains were compared to the BN reference RGSC 3.4 genome, a high proportion of the SNPs and INDELs common between the $SHRSP_{Gla}$ and WKY_{Gla}. Furthermore, the total number of unique genomic sequence variants in the $SHRSP_{Gla}$ were less than WKY_{Gla} when compared to the BN reference RGSC 3.4 genome (Figure 4-2a). Heterozygous sequence variants were detected in our inbred rat strains (Figure 4-2b).

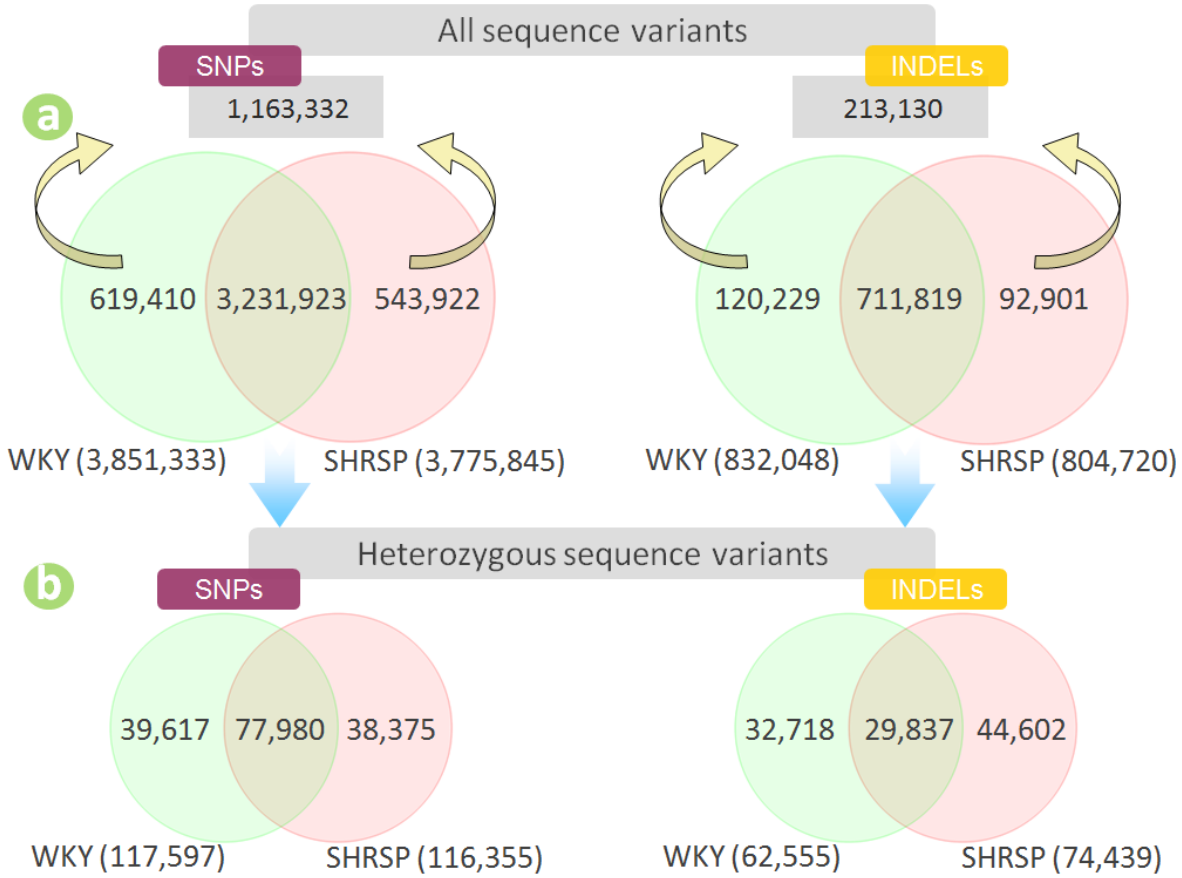


Figure 4-2 $SHRSP_{Gla}$ and WKY_{Gla} genomic sequence variants compared to BN reference genome RGSC 3.4.
a) Venn diagrams showings all SNPs and INDELs, for each rat strain compared to BN assembly RGSC 3.4. Common variants (intersect) were filtered out from the analysis to focus on unique variants between $SHRSP_{Gla}$ and WKY_{Gla} resulted in 1,163,332 SNPs and 213,130 INDELs. **b**) Venn diagrams showing only heterozygous SNPs and INDELs for $SHRSP_{Gla}$ and WKY_{Gla}.

4.3.1.2 Distribution and density of sequence variants

Circos plots with the genomics sequence variants (first two tracks) distribution and density over 100 kb window across the BN genome RGSC 3.4. It was observed that genomic variants were not uniform; for instance, in chromosome 1 there were certain areas with high cluster of SNPs and others without. The same was observed in the INDEL track, which additionally seemed to be correlated with the SNPs track (Figure 4-3).

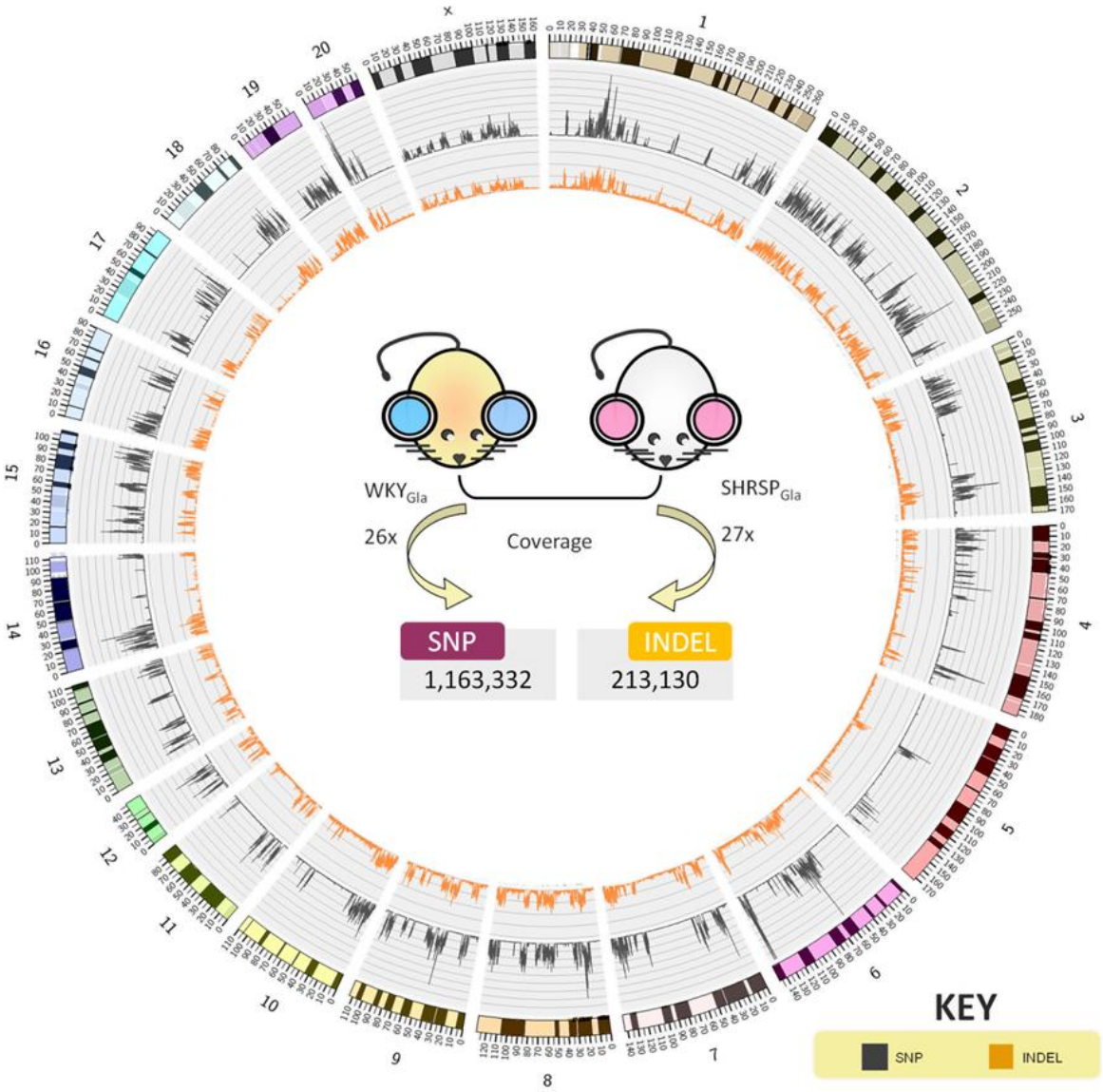


Figure 4-3 SHRSP$_{Gla}$ and WKY$_{Gla}$ genomic variants and their density distribution across BN reference genome RGSC 3.4.
A circos plot showing the unique sequence variants density distribution between SHRSP$_{Gla}$ and WKY$_{Gla}$ over 100 kb window across rat genome RGSC 3.4. The distribution of SNPs and INDELs was not uniform and correlated at 27x SHRSP$_{Gla}$ and 26x WKY$_{Gla}$ median coverage.

4.3.1.3 Small INDEL size distributions

As demonstrated in Figure 4-4A, size distribution of small insertions that were, unique to $SHRSP_{Gla}$ versus WKY_{Gla} when compared to BN reference RGSC 3.4, revealed that 1 bp was the highest in genome wide and in variants that affect protein function. However, 3 bp was also high; nevertheless, this could be biased for the reading frame. Moreover, as depicted in Figure 4-4B, in small deletions, unique to WKY_{Gla} versus $SHRSP_{Gla}$ when compared to BN reference RGSC 3.4, 1 bp size distribution was the highest, and in variants that have an effect on protein function, 3 bp.

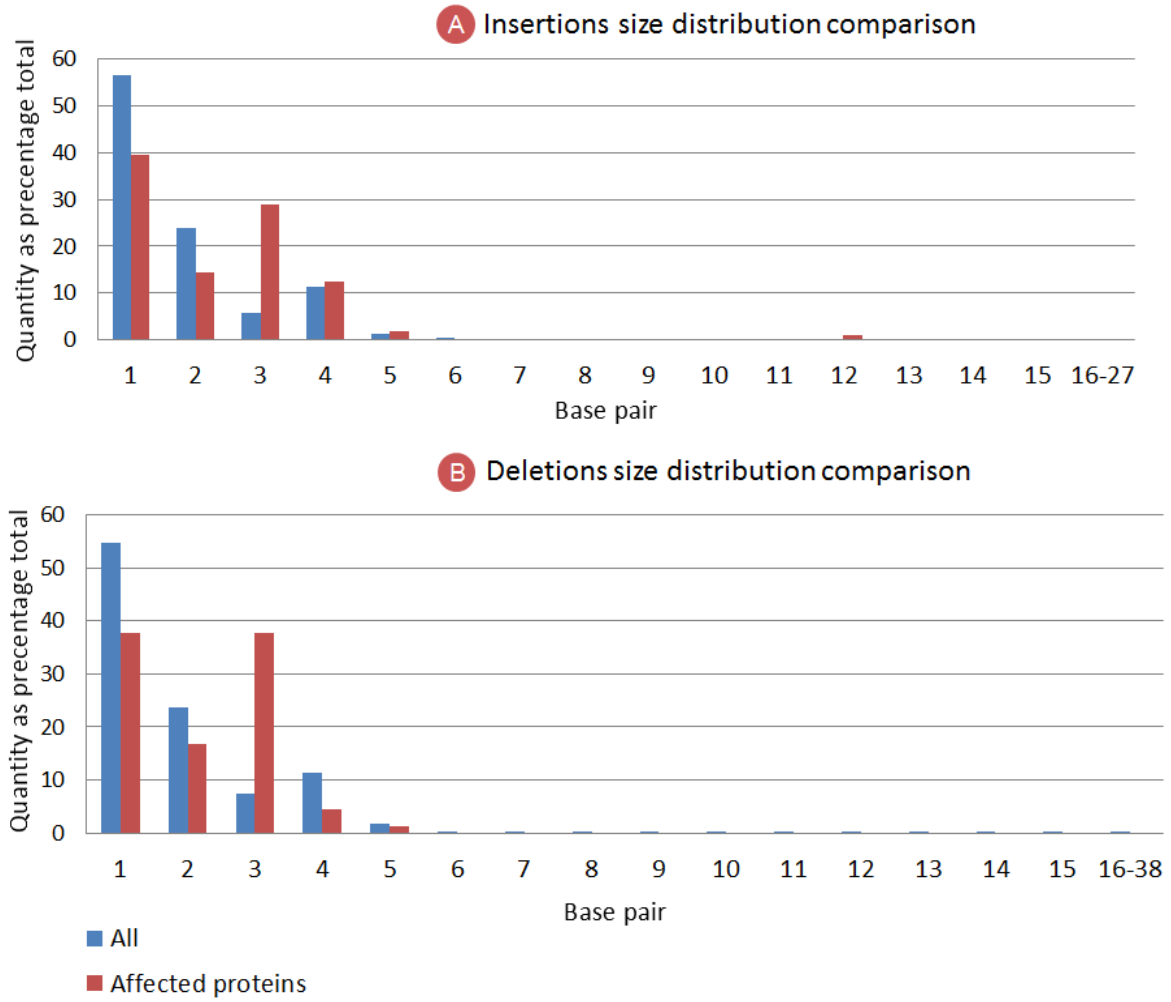


Figure 4-4 Small insertions and deletions size distribution comparison.
A) Distribution of small insertions sizes shown as a percent of total for all (blue) insertions and for predicted insertions that have an effect on proteins function (red). **B**: Distribution of small deletions sizes shown as a percentage of total for all (blue) deletions and for predicted deletions that have an effect on proteins function (red).

4.3.1.4 Variants effect predictor annotation Ensembl 69

Table 4-1 shows the predicted effect of sequence variants unique to $SHRSP_{Gla}$ versus WKY_{Gla} when compared to BN reference genome RGSC 3.4 (Figure 4-2a) by Ensembl variants effect predictor tool (version 69). Predicted consequences from SNPs and INDELs suggested that there were more intergenic and intronic variants compared to exonic variants, in both SNPs and INDELs. Protein coding sequence variants were then prioritised according to which ones were more likely to have an effect on protein function.

Table 4-1 Predicted consequence of all $SHRSP_{Gla}$ versus WKY_{Gla} sequence variants compared to BN reference genome RGSC 3.4.

SNP Consequences		INDEL Consequences	
INTERGENIC	737,915	INTERGENIC	131,551
UPSTREAM	64,410	UPSTREAM	13,311
DOWNSTREAM	63,810	DOWNSTREAM	12,663
3PRIME UTR	3,544	3PRIME UTR	710
5PRIME UTR	707	5PRIME UTR	83
INTRONIC	305,439	INTRONIC	59,623
SYNONYMOUS CODING	9,207	SYNONYMOUS CODING	70
NONSYNONYMOUS CODING	3,788	NONSYNONYMOUS CODING	70
STOP GAINED	55	FRAMESHIFT CODING	139
STOP LOST	3	COMPLEX INDEL	2
ESSENTIAL SPLICE SITE	57	ESSENTIAL SPLICE SITE	10
SPLICE SITE	1,018	SPLICE SITE	187
WITHIN MATURE miRNA	8	WITHIN MATURE miRNA	3
WITHIN NON CODING GENE	913	WITHIN NON CODING GENE	60

The numbers generated in this table are based on the predicted sequence variants location within a gene relative to the transcript structure by Ensembl annotation.

4.3.1.5 IPA results for all unique $SHRSP_{Gla}$ versus WKY_{Gla} NGS variants

Biological interpretation of unique SNPs and INDELs between $SHRSP_{Gla}$ and WKY_{Gla} when compared to BN reference genome RGSC 3.4, implicated networks, diseases and canonical pathways that regulate cardiovascular disease (Table 4-2).

Table 4-2 IPA results for all unique $SHRSP_{Gla}$ versus WKY_{Gla} sequence variants compared to BN reference genome RGSC 3.4.

Variants	Top Networks	Top Diseases	Top Canonical Pathways
SNP	Organ Morphology Organismal Injury and Abnormalitie	Cardiovascular Disease Endocrine system disorder	Endothelin -1 signalling pathway Apoptosis Signalling
INDEL	Cardiovascular Disease	Cardiovascular Disease Haematological Disease	Nuclear Factor Of Activated T-Cells in Cardiac Hypertrophy

4.3.1.6 Variants effect predictor annotation on $SHRSP_{Gla}$ and WKY_{Gla} QTLs

Table 4-3 shows extracted protein coding sequence variants in addition to other sequence variations in genes within $SHRSP_{Gla}$ and WKY_{Gla} validated QTLs regions on chromosome 2 (1,000,000bp-225,000,000bp), 3 (1,000,000bp-36,000,000bp), and 14 (5,000,000bp-112,194,335bp) from the the whole annotated sequence variants between $SHRSP_{Gla}$ versus WKY_{Gla} in Table 4-1 as they are more likely to be the causative of the $SHRSP_{Gla}$ phenotype.

Table 4-3 Predicted consequences of $SHRSP_{Gla}$ versus WKY_{Gla} sequence variants compared to BN reference genome RGSC 3.4 on QTL regions at chromosomes 2, 3 and 14.

SNP Consequences				INDEL Consequences			
QTLs	2	3	14	QTLs	2	3	14
INTERGENIC	120,019	14,541	20,923	INTERGENIC	20,277	2,482	3,696
UPSTREAM	9,021	2,617	2,396	UPSTREAM	1,865	520	421
DOWNSTREAM	8,683	2,524	2,589	DOWNSTREAM	1,835	428	498
3PRIME UTR	572	201	141	3PRIME UTR	125	26	26
5PRIME UTR	83	25	14	5PRIME UTR	12	0	6
INTRONIC	62,434	16,263	16,974	INTRONIC	12,347	3,033	3,104
SYNONYMOUS CODING	1,310	409	375	SYNONYMOUS CODING	11	2	2
NONSYNONYMOUS CODING	489	121	144	NONSYNONYMOUS CODING	11	2	2
STOP GAINED	2	0	3	FRAMESHIFT CODING	15	8	1
STOP LOST	0	0	0	COMPLEX INDEL	0	0	0
ESSENTIAL SPLICE SITE	4	4	1	ESSENTIAL SPLICE SITE	3	0	0
SPLICE SITE	160	68	46	SPLICE SITE	36	6	4
WITHIN MATURE miRNA	1	0	0	WITHIN MATURE miRNA	0	0	0
WITHIN NON CODING GENE	103	36	37	WITHIN NON CODING GENE	10	3	5

The numbers generated in this table are based on the predicted sequence variants location within a gene relative to the transcript structure by Ensembl annotation.

4.3.2 Mapping $SHRSP_{Gla}$ and WKY_{Gla} NGS reads to BN new assembly Rnor 5.0

Table 4-4 demonstrate an improved of mapping and average coverage of NGS reads of $SHRSP_{Gla}$ and WKY_{Gla} with the BN new reference Rnor 5.0 when compared to the BN reference genome RGSC 3.4. In addition, there were many improvements in the BN new reference genome Rnor 5.0 in base pairs (66,296,177) and total length (145,534,490) compared to the BN reference RGSC 3.4.

Table 4-4 Mapping and coverage results of $SHRSP_{Gla}$ and WKY_{Gla} reads when compared to BN difference assemblies.

Brown Norway RGSC 3.4 assembly			Brown Norway Rnor 5.0 new assembly		
Base Pairs	Golden Path Length	Date of release	Base Pairs	Golden Path Length	Date of release
2,507,066,667	2,718,897,334	Dec 2004	2,573,362,844	2,909,698,938	Mar 2013
Rat Strains	Gb of bases mapped*	Average coverage	Rat Strains	Gb of bases mapped*	Average coverage
WKY	68.17	26.51	WKY	69.88	27.16
SHRSP	70.03	27.23	SHRSP	71.37	27.74

*Gb of bases mapped and average coverage was calculated after removing clonal reads.

4.3.3 Sequence variants between $SHRSP_{Gla}$ and WKY_{Gla} compared to BN new reference genome Rnor 5.0

4.3.3.1 Variants calling between $SHRSP_{Gla}$ and WKY_{Gla}

When $SHRSP_{Gla}$ and WKY_{Gla} strains were compared to the BN new reference Rnor 5.0, a high proportion of the SNPs and INDELs were common between our rat strains. Furthermore, the total number of unique genomic variants with BN new reference genome Rnor 5.0 was more than the BN reference genome RGSC 3.4. However, in both, the unique variants in $SHRSP_{Gla}$ were less than WKY_{Gla} (Figure 4-5).

Figure 4-5 Variants calling between $SHRSP_{Gla}$ and WKY_{Gla} against BN new assembly Rnor 5.0.
Venn diagrams showing intersections of SNPs and INDELs that common and unique to each strain when compared to the BN new reference genome Rnor 5.0 as well as the total unique variants between $SHRSP_{Gla}$ and WKY_{Gla}.

4.3.3.2 Ensembl Variants effect predictor annotation 71

Table 4-5 shows the predicted effect of sequence variants which are unique to WKY_{Gla} versus $SHRSP_{Gla}$ when compared to BN new reference genome Rnor 5.0 (Figure 4-5) by Ensembl variants effect predictor tool (version 71). Furthermore, predicted consequences suggested that there were more intergenic and intronic variants compared to exonic variants. In additionally, the new Ensembl variants annotation 71 added new and more refined predicted consequences compared to Ensembl variants predictor annotator 69.

Table 4-5 Variants Ensembl effect predictor of $SHRSP_{Gla}$ and WKY_{Gla} total unique sequence variants.

SNP Consequences		INDEL Consequences	
INTERGENIC	831,606	INTERGENIC	173,717
UPSTREAM	57,459	UPSTREAM	12,924
DOWNSTREAM	56,806	DOWNSTREAM	11,922
3PRIME UTR	3,846	3PRIME UTR	854
5PRIME UTR	622	5PRIME UTR	80
INTRONIC	316,558	INTRONIC	68,801
SYNONYMOUS CODING	5,493	TRUNCTION	34,267
STOP RETAINED	4	FRAMESHIFT	102
MISSENSE CODING	3,875	MISSENSE CODING	4
INITIATOR CODON	7	ELONGATION	34,368
STOP GAINED	56	INFRAME INSERTION	31
STOP LOST	4	INFRAME DELETION	56
SPLICE REGION	897	SPLICE REGION	154
SPLICE ACCEPTOR	16	SPLICE ACCEPTOR	10
SPLICE DONOR	25	SPLICE DONOR	8
WITHIN CODING SEQUENCE	13	WITHIN CODING SEQUENCE	7
WITHIN MATURE miRNA	12	WITHIN MATURE miRNA	4
WITHIN NON CODING TRANSCRIPT	831	WITHIN NON CODING TRANSCRIPT	50
WITHIN NON CODING EXON	773	WITHIN NON CODING EXON	38

The numbers generated in this table are based on the predicted sequence variants location within a gene relative to the transcript structure by Ensembl annotation.

4.3.3.3 Improvements of BN new assembly on selected genes of interest

Figure 4-6 Shows examples of improvements of BN new reference Rnor 5.0 compared to BN reference genome RGSC 3.4 were observed in the *Spp1* and hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 5 (*Hsd3b5*) genes. *Spp1* annotation and *Hsd3b5* missing sequences are improved in the BN new reference genome Rnor 5.0 when compared to BN reference genome RGSC 3.4.

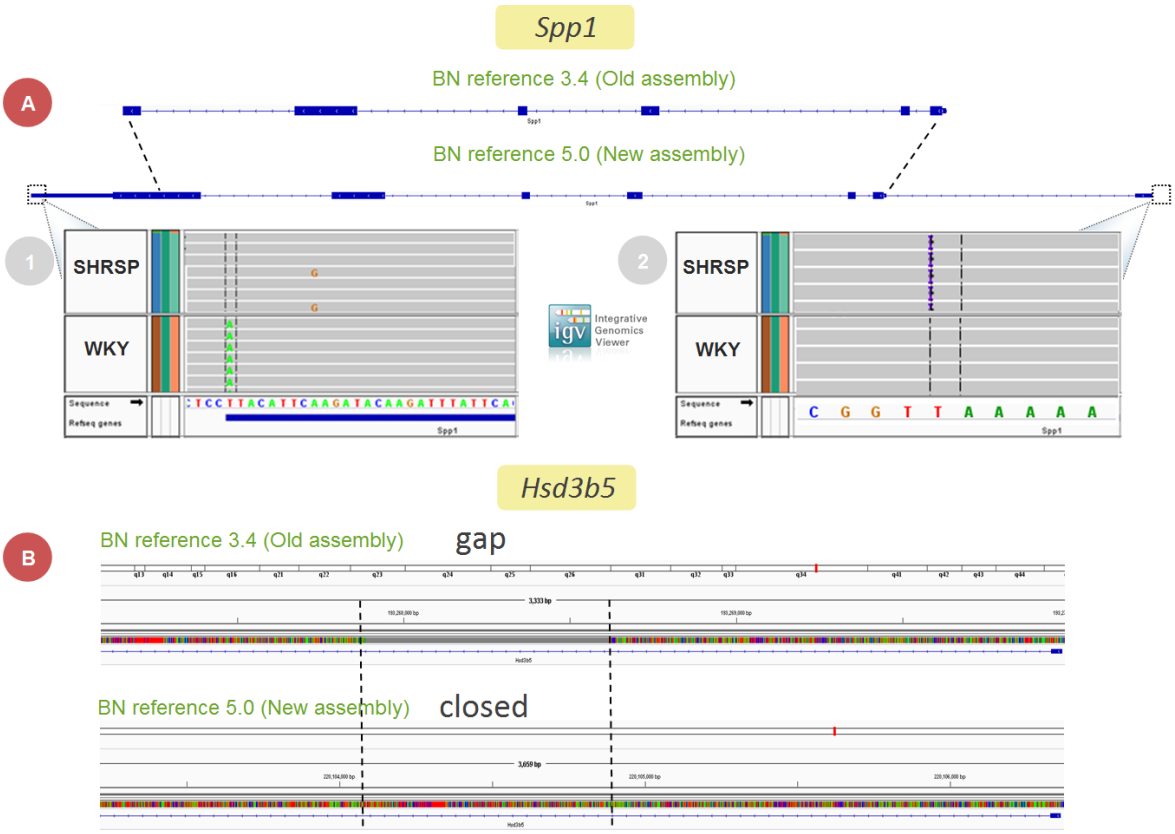


Figure 4-6 Improvements of BN new reference sequence compared to the old reference visualised by IGV.
A) *Spp1* gene sequence in BN reference genome RGSC 3.4 was incomplete when compared to the new BN reference genome Rnor 5.0. The new *Spp1* regions harbour potential functional sequence variants such as **A1** SNP at 3' UTR region and **A2** upstream insertion when viewed with IGV. **B**) The interogenic gap of hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 5 (*Hsd3b5*) on the BN reference genome RGSC 3.4 was closed in the new reference Rnor 5.0 thus new regions to look for sequence variants were revealed.

4.3.4 Quality control of NGS data

4.3.4.1 Quality control comparing NGS data with sequence capture data

As a quality control, a region on chromosome 2 (211,822,164-213,846,183 bp) was sequenced by both Illumina NGS whole genome and Agilent SureSelect sequence capture between $SHRSP_{Gla}$ and WKY_{Gla} compared to BN reference genome RGSC 3.4 was used. Figure 4-7, presents 75% of SNPs that were common between both Illumina NGS whole genome and Agilent SureSelect sequence capture, as well as the ones that were unique to each method. NGS data detected more unique SNPs than sequence capture, which might be due to the presence of gaps (Figure 4-7a). Sequence capture detected SNPs which were not captured by NGS data; this could be due to a low number of reads for variants calling parameters (Figure 4-7b).

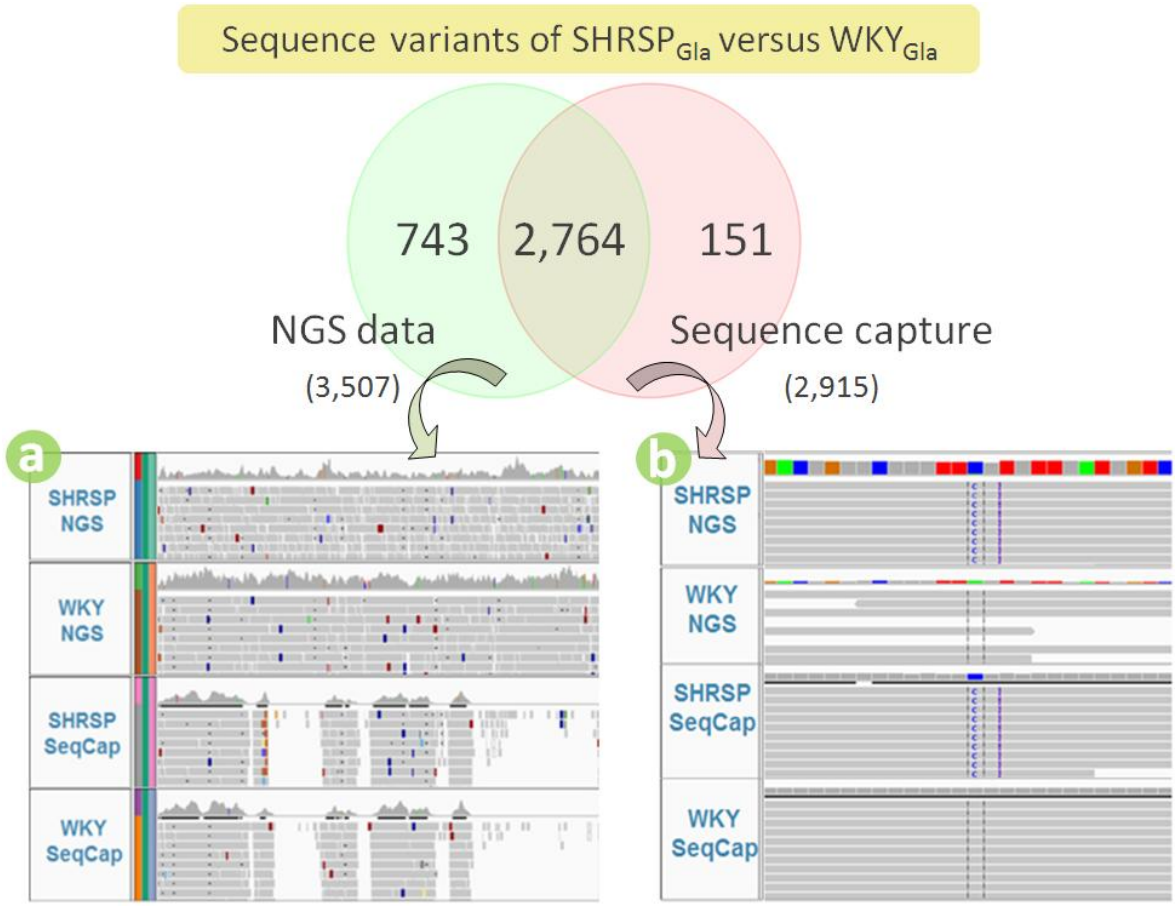


Figure 4-7 Quality control: NGS data versus Sequence Capture.
Venn diagram showing SNPs that were common and unique to each of the sequencing method. **a**) IGV shows an example of a SNP that is detected in the NGS analysis but not in the sequence capture analysis due to SureSelect design, which covers exonic regions more than intergenic regions. **b**) IGV shows an example of a SNP that was detected by sequence capture but not in the NGS analysis due to the low number of reads in the WKY_{Gla} genome at the $SHRSP_{Gla}$ SNP position in which it was excluded from the analysis.

4.3.4.2 Quality control comparing NGS data with STAR sequence data

Unique SNPs from NGS data and STAR project consortium sequencing data of $SHRSP_{Gla}$ and WKY_{Gla} compared to BN reference RGSC 3.4 was also used as quality control. NGS data captured 98% of STAR data and 1.7% SNPs that were unique to STAR data (Figure 4-8a). Nevertheless, only 42.5% (17 SNPs) (Figure 4-8b2) of the 1.7% unique SNPs to STAR data were true positive SNPs that were missed by NGS analysis. This is due to the fact that since half of the unique STAR data SNPs are either not properly annotated (Figure 4-8b1) or does not map to the BN reference genome RGSC 3.4 (i.e not comparable) (Figure 4-8b3).

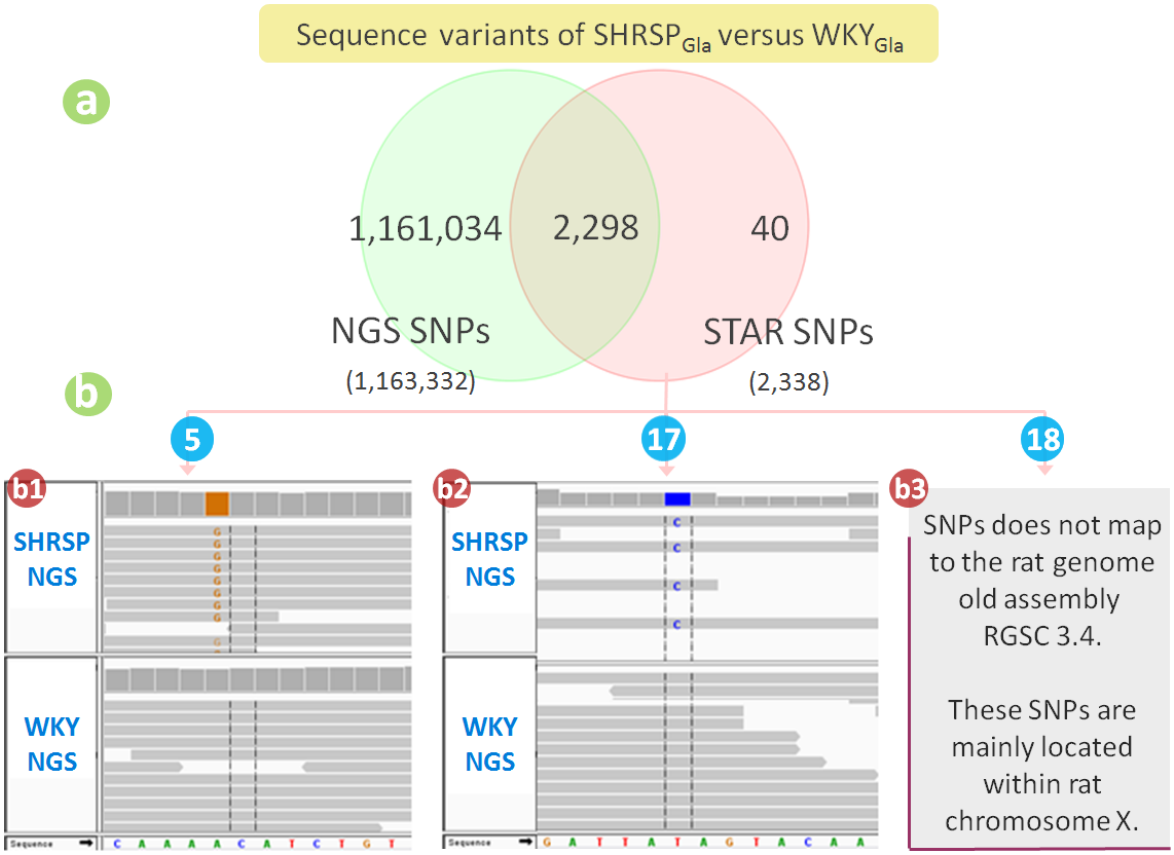


Figure 4-8 Quality control: NGS data versus STAR data.
a) Venn diagram showing SNPs that were common and unique to each of the sequencing method. The NGS data captured most of the STAR SNPs. **b**) not all the SNPs that are unique to STAR data are true and/or comparable SNPs. **b1**) IGV shows 1of out 5 STAR unique SNPs that were incorrectly called by 1 bp. **b2**) IGV shows one of out 17 STAR unique SNPs which was missed by NGS data due to insufficient reads for variant calling parameters. **b3**) 18 STAR unique SNPs were not mapped to the BN reference genome RGSC 3.4 and thus, were incomparable SNPs to NGS data.

4.3.4.3 Quality control comparing NGS data, sequence capture with Sanger sequencing

When comparing 34 SNPs/INDELs from Sanger sequencing on a region chromosome 2 to both NGS data and sequence capture unique sequence from $SHRSP_{Gla}$ and WKY_{Gla} compared to BN reference RGSC 3.4. Both were able to detect most of the variants, however, NGS data missed 4 SNPs and 1 INDEL, while sequence capture data missed only 3 SNPs (Figure 4-9).

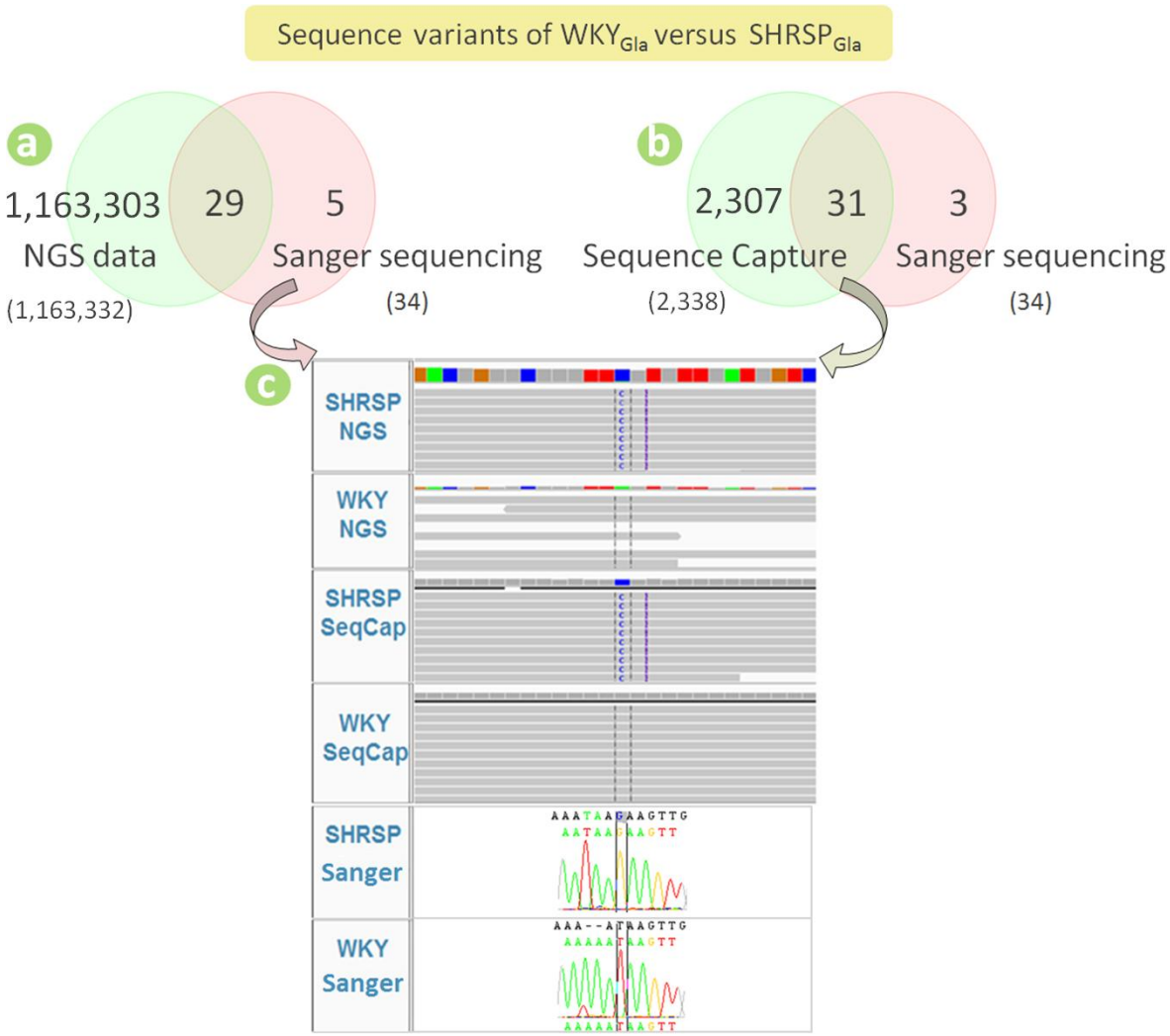


Figure 4-9 Quality control: NGS data and sequence capture compared to Sanger sequencing data.
a,b) Venn diagram showing SNPs and INDELs of Sanger sequencing that were common and unique to NGS data and sequence capture. **c**) Sanger sequence SNP (complementary) visualised by IGV, were captured by sequence capture analysis but missed by NGS analysis due to the low number of reads in the WKY_{Gla} genome at the $SHRSP_{Gla}$ SNP position in which it was excluded from the analysis. (Capillary image courtesy of Ms. Wendy Beattie)

4.3.5 Structural variations between $SHRSP_{Gla}$ and WKY_{Gla}

Structural variations analysis on NGS data of $SHRSP_{Gla}$ and WKY_{Gla} detected large deletion (more than 45 bp) and CNV when compared to the BN reference genome RGSC 3.4.

4.3.5.1 Large deletion

Figure 4-10a demonstrates that the number of large deletions implicated present in the WKY_{Gla} was higher compared to $SHRSP_{Gla}$. Additionally, results of mapping deleted regions in $SHRSP_{Gla}$ and WKY_{Gla} to genes via Ensembl BioMart (version 69) are shown in Figure 4-10b.

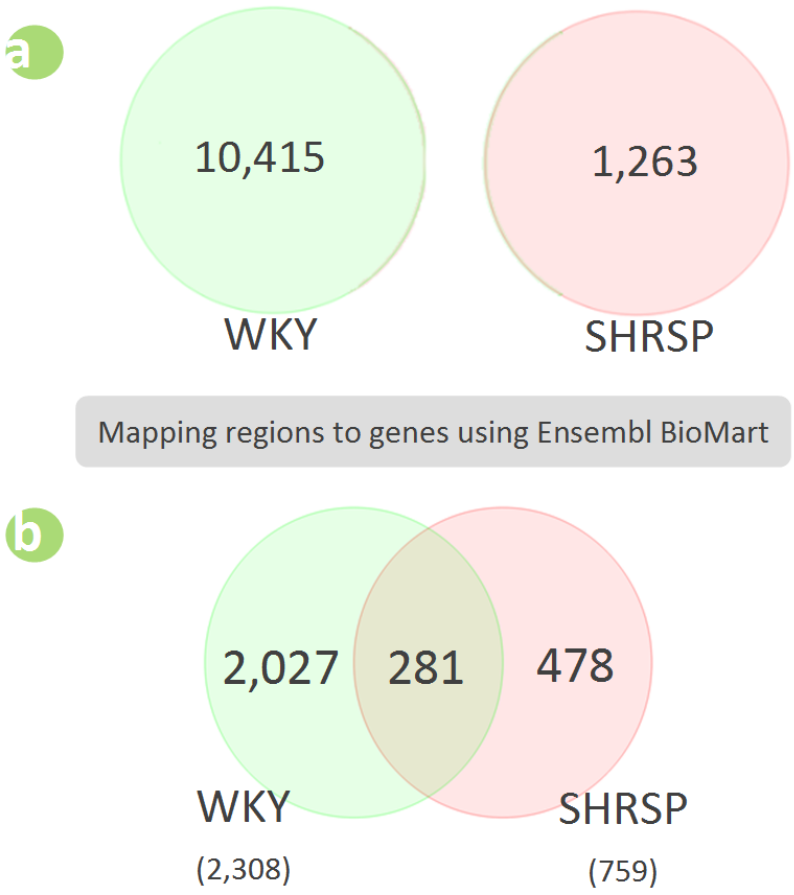


Figure 4-10 Large deletions analysis in $SHRSP_{Gla}$ and WKY_{Gla}.
a) Large deletion analysis between the $SHRSP_{Gla}$ and WKY_{Gla} revealed that there were more deletions in the WKY_{Gla}. **b**) Venn diagrams showings deleted and/partial deleted genes that are in-common and unique to each rat strain compared to BN reference genome RGSC 3.4.

4.3.5.1.1 Large deletion size distributions

Figure 4-11 demonstrate large deletions at chromosomes and size distribution in the $SHRSP_{Gla}$. Analysis revealed that $SHRSP_{Gla}$ had a high proportion of large deletion across rat chromosomes, when corrected to their length, were located on chromosomes 1 and 2 (Figure 4-11a). Moreover, high proportion of large deletion size distribution in the $SHRSP_{Gla}$ was between 100-250 bp and larger than 10 kb (Figure 4-11b).

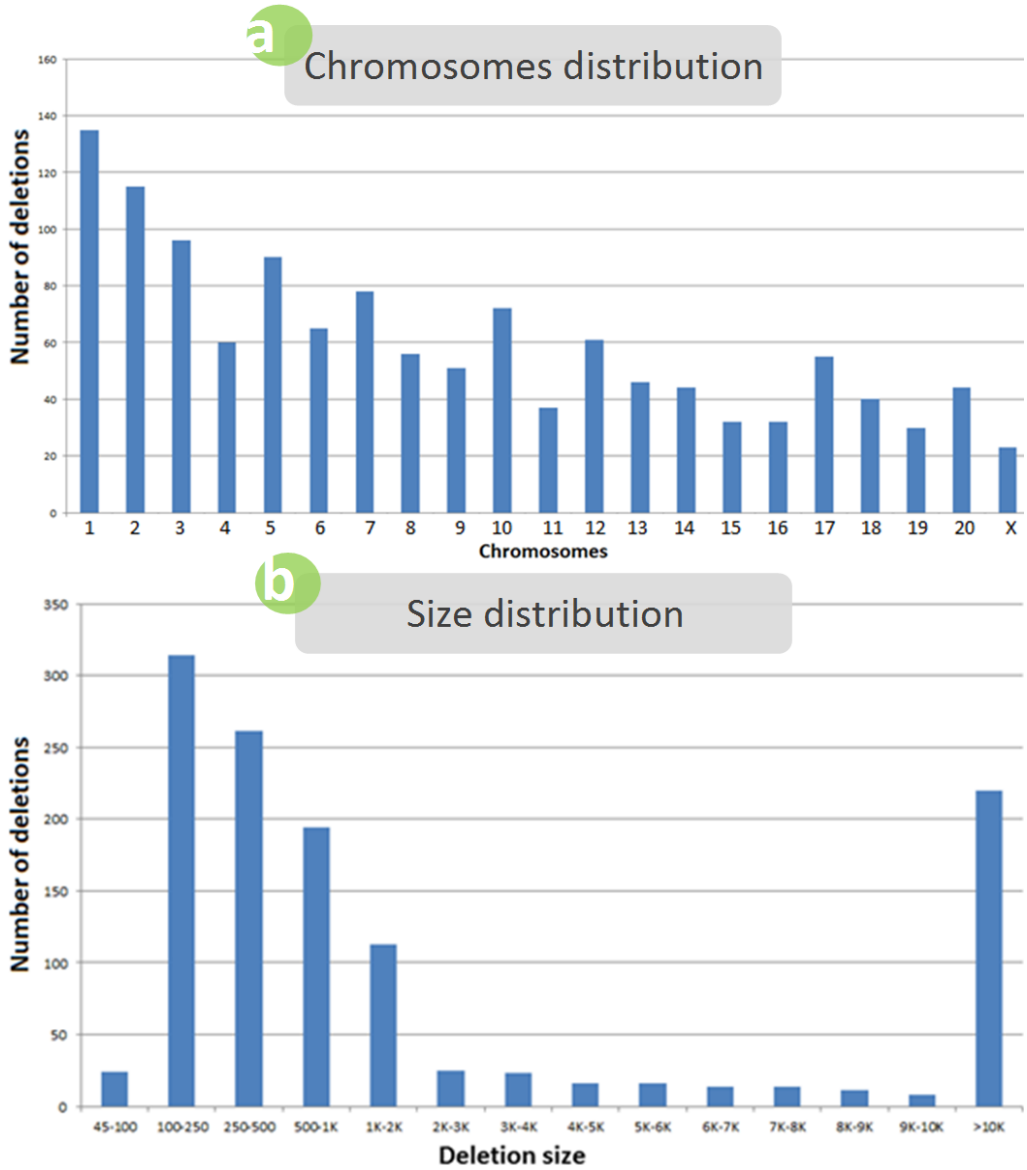


Figure 4-11 Large deletion distributions in size and across rat genome.
a) Distribution of large deletion in the $SHRSP_{Gla}$ across rat genome in which high proportion of deletion were in chromosomes 1 and 2. **b**) Distribution of sizes of large deletions in the $SHRSP_{Gla}$ were between 100-250 bp and larger than 10 kb.

4.3.5.2 Copy number variation

CNV analysis in $SHRSP_{Gla}$ and WKY_{Gla} when compared to the BN reference genome RGSC 3.4 showed that there were more copy number duplications in the $SHRSP_{Gla}$ than WKY_{Gla} (Figure 4-12a). In addition, the copy number deletions in the WKY_{Gla} were more compared to the $SHRSP_{Gla}$ (Figure 4-12b). When the CNV regions of duplication and deletion annotated were by using Ensembl BioMart (version 69), we refined only what was unique to each rat strain for duplication (Figure 4-12c) and deletion (Figure 4-12d), and therefore, of interest.

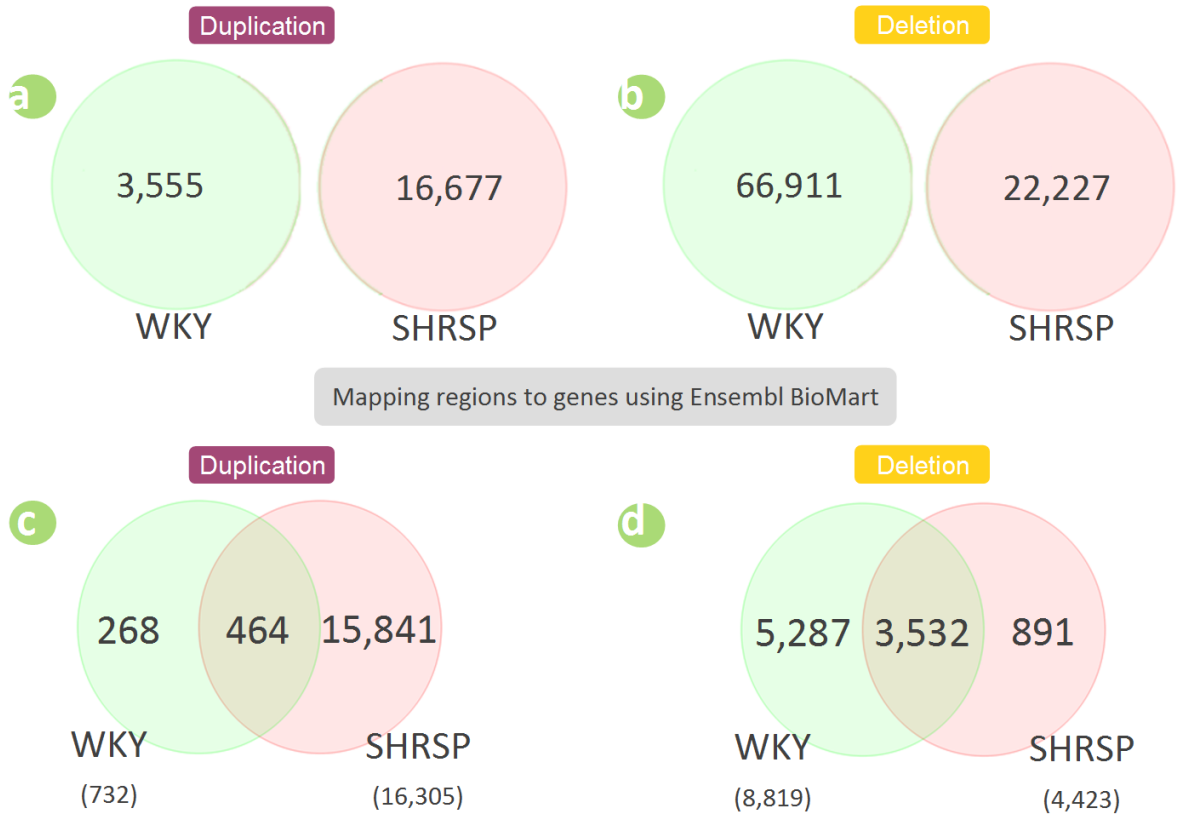


Figure 4-12 CNV analysis in $SHRSP_{Gla}$ and WKY_{Gla}.
a) Copy number variants of duplication in $SHRSP_{Gla}$ were more when compared to WKY_{Gla}. **b**) Copy number variants of deletion in WKY_{Gla} were more when compared $SHRSP_{Gla}$. **c**) Venn diagram showed what was common and what was unique of CNV of duplication to each rat strain. **d**) Venn diagram showed what was common and what was unique of CNV of deletion to each rat strain.

4.3.6 NGS whole genome sequence of WKY_{Gla}, $SHRSP_{Gla}$, and $SHR_{Olalpcv}$

When comparing the NGS whole genome SNPs across all three rat strains compared to BN reference genome RGSC 3.4, the intersection that was common among them had the highest number of SNPs, around 2.5 million (Figure 4-13). Such an observation, was expected, given the fact strains' were relative similar compared to BN reference genome RGSC 3.4. Prioritising intersections of what was common between $SHRSP_{Gla}$ and $SHR_{Olalpcv}$ and what was unique to $SHRSP_{Gla}$ was of interest since they could harbour causative variants for spontaneous hypertension and stroke proneness, respectively.

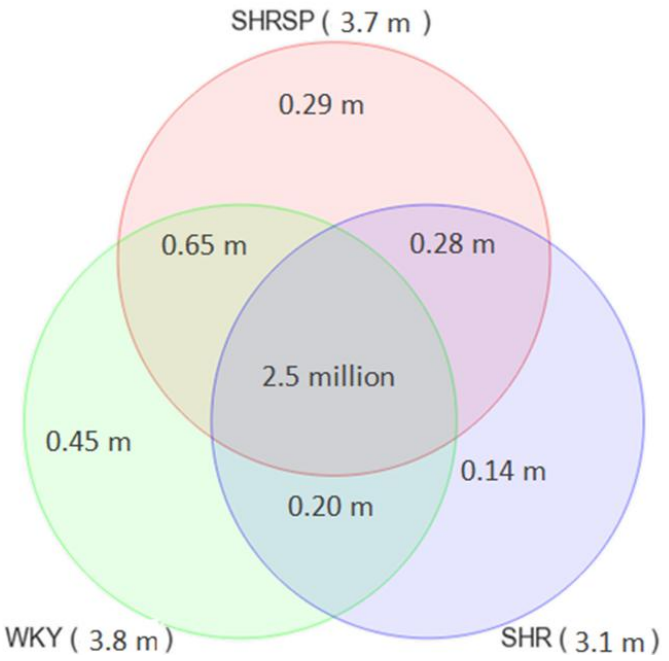


Figure 4-13 SNPs calling between $SHRSP_{Gla}$, WKY_{Gla} and $SHR_{Olalpcv}$ against BN assembly RGSC 3.4.
Venn diagram of SNPs across the three rat strains showing different intersection of what was common and what was unique among the three rat strains.

4.3.7 $SHRSP_{Gla}$ unique sequence variants compared to 27 NGS rat genomes

$SHRSP_{Gla}$ NGS sequence variants compared to NGS whole genome of 27 rat strains against BN reference genome RGSC 3.4 resulted in SNPs 23,621 INDELs 3,532 unique to the $SHRSP_{Gla}$.

4.3.7.1 $SHRSP_{Gla}$ unique variants effect predictor annotation Ensembl 69

Table 4-6 shows annotated sequence variants unique to the $SHRSP_{Gla}$ when compared to NGS 27 rat against BN reference RGSC 3.4. Predicted consequences from SNPs and INDELs showed that there were more intergenic and intronic variants compared to exonic variants in both SNPs and INDELs. Protein coding sequence variants from SNPs and INDELs were prioritised as these were more likely to have an effect on protein function.

Table 4-6 Variants Ensembl effect predictor of $SHRSP_{Gla}$ unique sequence variants compared to 27 NGS rat genomes.

SNP Consequences		INDEL Consequences	
INTERGENIC	13,025	INTERGENIC	1,967
UPSTREAM	1,382	UPSTREAM	247
DOWNSTREAM	1,526	DOWNSTREAM	235
3PRIME UTR	59	3PRIME UTR	14
5PRIME UTR	15	5PRIME UTR	2
INTRONIC	8,017	INTRONIC	1,113
SYNONYMOUS CODING	140	SYNONYMOUS CODING	0
NONSYNONYMOUS CODING	103	NONSYNONYMOUS CODING	4
STOP GAINED	3	FRAMESHIFT CODING	3
STOP LOST	0	COMPLEX INDEL	0
ESSENTIAL SPLICE SITE	3	ESSENTIAL SPLICE SITE	1
SPLICE SITE	17	SPLICE SITE	2
WITHIN MATURE miRNA	0	WITHIN MATURE miRNA	0
WITHIN NON CODING GENE	21	WITHIN NON CODING GENE	0

The numbers generated in this table are based on the predicted sequence variants location within a gene relative to the transcript structure by Ensembl annotation.

4.3.7.2 $SHRSP_{Gla}$ unique SNPs compared to 27 NGS rat genomes

Subsequently, we prioritised protein coding $SHRSP_{Gla}$ unique SNPs compared to 27 NGS rat genomes specifically the 103 non-synonymous and 3 stop gain SNPs as they were likely to have a deleterious effect on protein function. IPA analysis on the 88 genes with non-synonymous mutations Figure 4-14 in which (only 67 genes were mapped to IPA) implicated functional networks (Figure 4-14a), disease (Figure 4-14b), and pathways (Figure 4-14c), that regulate cardiovascular disease. $SHRSP_{Gla}$ and WKY_{Gla} QTL regions for blood pressure, left ventricular index, and stroke (chromosome 5: 62,555,131 bp-148,718,705 bp) as well as other QTLs for blood pressure and stroke from other rat strains on RGD were used to prioritise candidate genes. Furthermore, the consistent consequence prediction in the BN new reference genome Rnor 5.0 and SIFT predication score (for non-synonymous) were also used to prioritised $SHRSP_{Gla}$ unique non-synonymous (Table 4-7) and stop gained (Table 4-8) SNPs.

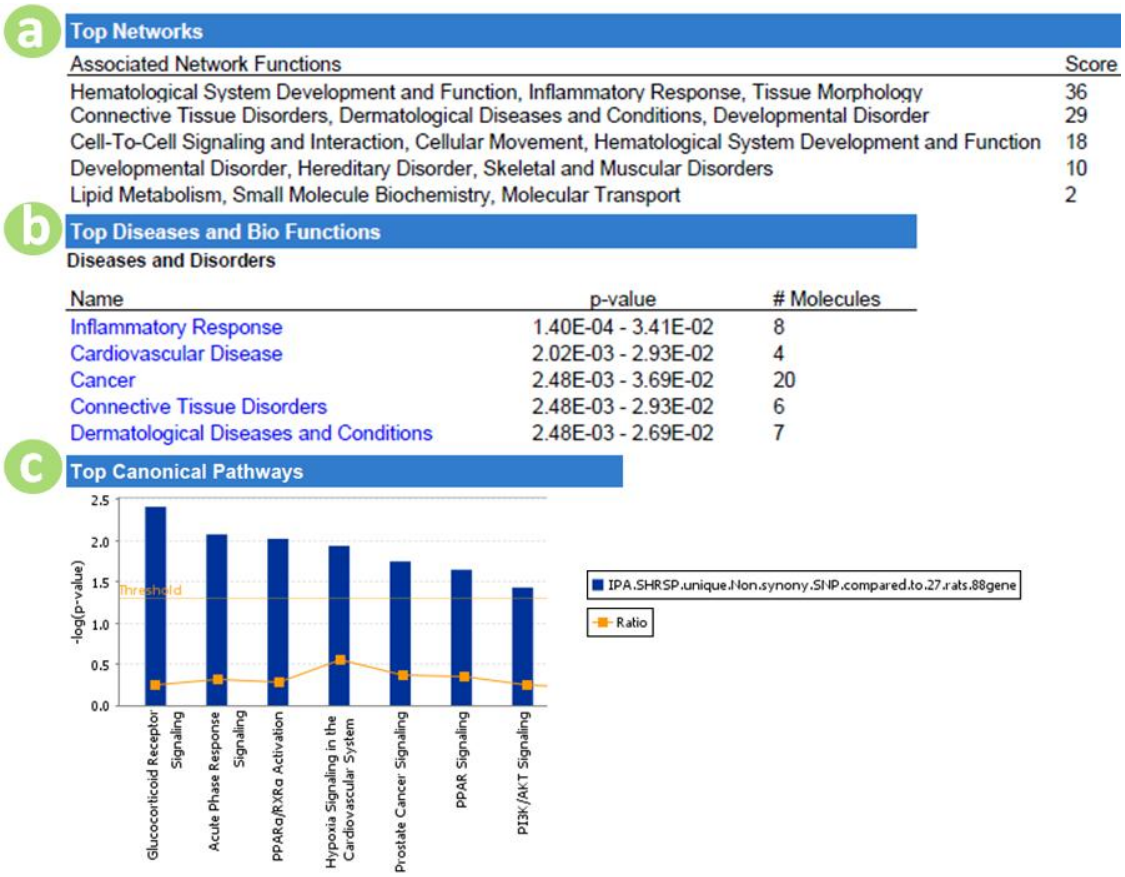


Figure 4-14 IPA analysis of $SHRSP_{Gla}$ unique non-synonymous SNPs.
The 67 genes with non-synonymous SNPs implicated **a**) top networks such as haematological functions and tissue morphology, **b**) top disease such as cardiovascular and **c**) top canonical pathways hypoxia signalling in cardiovascular system all of which can regulate of the cardiovascular disease.

Table 4-7 $SHRSP_{Gla}$ unique non-synonymous SNPs compared to 27 NGS rat genomes.

Gene	Position	WKY_{Gla} vs $SHRSP_{Gla}$ QTL	RGD Blood pressure and stroke QTL	Rnor 5.0 Consequences	SIFT (prediction and score)
Mad1l1	12_15012314_C/A	-	BP QTL 218	non synonymous	deleterious(0)
Slc10a6	14_7291990_A/G	Left ventricular mass index	no relevant QTL	non synonymous	deleterious(0)
Gpr132	6_137856331_G/A	-	no relevant QTL	non synonymous	deleterious(0)
Olr847	5_70182159_C/A	Stroke	BP QTL 7,49,254	non synonymous	deleterious(0)
Ahnak2(a)	6_137763287_G/T	-	no relevant QTL	non synonymous	deleterious(0)
Dmp1	14_6638436_C/T	Left ventricular mass index	no relevant QTL	non synonymous	deleterious(0.01)
LOC685722	9_11714748_T/C	-	BP QTL 332	non synonymous	deleterious(0.01)
Vom2r7	1_49201316_T/G	-	BP QTL 58,96,145	non synonymous	deleterious(0.01)
IGG-2A	6_138347843_C/T	-	no relevant QTL	non synonymous	deleterious(0.02)
Olr852	5_70367184_G/A	Stroke	BP QTL 7,49,254	non synonymous	deleterious(0.03)
Hap1	10_89293462_G/A	-	BP QTL 1,9,72,82,91,134,168,249,311	non synonymous	deleterious(0.04)
Mill1	1_78075767_G/A	-	BP QTL 58,96,267,361	non synonymous	deleterious(0.04)
Erf	1_80520430_C/T	-	BP QTL 58,96,267,361 Stroke QTL Strs4	non synonymous	deleterious(0.04)
Rinl	1_83904959_A/G	-	BP QTL 58,96,267,361 Stroke QTL Strs4	non synonymous	deleterious(0.04)
Krt42	10_89194497_T/A	-	BP QTL 1,9,72,82,83,91,134,168,249	non synonymous	deleterious(0.05)
Abca1	5_70519063_C/T	Stroke	BP QTL 7,49,254	non synonymous	tolerated(0.07)
Pou2f2	1_80430867_C/G	-	BP QTL 58,96,267,361 Stroke QTL Strs4	non synonymous	tolerated(0.07)
Cic	1_80564423_G/T	-	BP QTL 58,96,267,361 Stroke QTL Strs4	non synonymous	tolerated(0.07)
Agap2	7_67032021_G/A	-	BP QTL 38,181,182	non synonymous	tolerated(0.07)
F1M2F7_RAT	2_78745124_G/A	Blood Pressure	BP QTL 13,14,36,99,132,163,206,239	non synonymous	tolerated(0.07)
D4A1G8_RAT	6_137724015_G/A	-	no relevant QTL	non synonymous	tolerated(0.07)
Hjurp	9_87145600_T/C	-	BP QTL 108,185,334,356	non synonymous	tolerated(0.08)
Fcgbpl1	1_83258111_C/T	-	BP QTL 58,96,267,361 Stroke QTL Strs4	non synonymous	tolerated(0.08)
Mllt4	1_48906849_T/G	-	BP QTL 58,96,145	non synonymous	tolerated(0.09)
Soga1	3_147603642_C/T	-	BP QTL 20,81,167,208,285	non synonymous	tolerated(0.1)
Zfp341	3_145086849_T/G	-	BP QTL 20,81,208,285	non synonymous	tolerated(0.11)
Tgm2	3_148854709_C/T	-	BP QTL 20,81,167,176,208,285,320	non synonymous	tolerated(0.13)
Sfxn4	1_267458176_A/T	-	no relevant QTL	non synonymous	tolerated(0.13)
Emilin3	3_151705881_C/T	-	BP QTL 20,81,167,176,208,285,320	non synonymous	tolerated(0.14)

Continued

Tril	4_82181137_G/T	-	BP QTL 5,86,124,135 Stroke QTL Strs3	non synonymous	tolerated(0.15)
Gpr31	1_47582567_G/A	-	BP QTL 58,96,145	non synonymous	tolerated(0.15)
Fbxo17	1_83843123_G/A	-	BP QTL 58,96,267,361 Stroke QTL Strs4	non synonymous	tolerated(0.17)
Cpa5	4_57535858_C/T	-	BP QTL 21,79,86,330 Stroke QTL Strs3	non synonymous	tolerated(0.18)
Taf7l	X_121950694_A/G	-	no relevant QTL	non synonymous	tolerated(0.18)
Cyp2b3	1_81438320_G/A	-	BP QTL 58,96,267,361 Stroke QTL Strs4	non synonymous	tolerated(0.18)
Ahnak2(b)	6_137757644_C/T	-	no relevant QTL	non synonymous	tolerated(0.19)
Ighe	6_138262637_C/T	-	no relevant QTL	non synonymous	tolerated(0.2)
Ighg3	6_138392889_T/A	-	no relevant QTL	non synonymous	tolerated(0.2)
RGD1304622	13_70193563_G/A	-	BP QTL 11,25,31,55,80,220,241,244	non synonymous	tolerated(0.22)
Mroh8	3_147814250_T/G	-	BP QTL 20,81,167,208,285,320	non synonymous	tolerated(0.24)
Shcbp1l	13_68437681_G/A	-	BP QTL 5,11,25,31,55,80,159,220,241	non synonymous	tolerated(0.25)
Rsph3	1_41332744_T/C	-	BP QTL 58,96,145,360	non synonymous	tolerated(0.25)
Leprel4	10_89352901_T/C	-	BP QTL 1,9,72,82,91,134,168,249,311	non synonymous	tolerated(0.29)
Alpi	9_85894533_G/A	-	BP QTL 108,185,334,356	non synonymous	tolerated(0.3)
Sla2	3_147290682_A/C	-	BP QTL 20,81,167,208,285	non synonymous	tolerated(0.31)
LOC691485	6_137337602_G/A	-	no relevant QTL	non synonymous	tolerated(0.32)
Ntrk2	17_11770480_G/A	-	BP QTL 194,247	non synonymous	tolerated(0.32)
Slc28a3	17_12030232_G/A	-	BP QTL 194,247	non synonymous	tolerated(0.34)
Fktn	5_71185451_A/C	Stroke	BP QTL 7,49,254	non synonymous	tolerated(0.35)
Vom1r15	1_58357132_G/C	-	BP QTL 58,96,145	non synonymous	tolerated(0.35)
F1M4H2_RAT	1_78175741_G/A	-	BP QTL 58,96,267,361 Stroke QTL Strs4	non synonymous	tolerated(0.39)
IGG2B_RAT	6_138287273_T/C	-	no relevant QTL	non synonymous	tolerated(0.4)
Hsp90ab1	9_11036204_G/A	-	BP QTL 332	non synonymous	tolerated(0.41)
Nfkbib	1_83867940_G/C	-	BP QTL 58,96,267,361 Stroke QTL Strs4	non synonymous	tolerated(0.42)
SAAL1	1_97232817_C/T	-	BP QTL 58,96,154,171,199,200,267	non synonymous	tolerated(0.46)
Myo5b	18_71561010_G/A	-	BP QTL 15,41,48,54,229,230,232,233	non synonymous	tolerated(0.47)
Epn3	10_83133194_C/T	-	BP QTL 1,9,12,57,71,72,82,91,148,168	non synonymous	tolerated(0.51)
Ugt1a7c	9_87030209_A/G	-	BP QTL 108,185,334,356	non synonymous	tolerated(0.52)

Continued

Try10	4_69056646_G/A	-	BP QTL 5,21,79,86,124 Stroke QTL Strs3	non synonymous	tolerated(0.58)
Supt3h	9_11552350_C/T	-	BP QTL 332	non synonymous	tolerated(0.62)
Egflam	2_56537970_C/T	Blood Pressure	BP QTL 2,14,36,174,206,240,243,336	non synonymous	tolerated(0.64)
Mrpl14	9_10903315_A/G	-	BP QTL 332	non synonymous	tolerated(0.66)
Osmr	2_56090512_T/C	Blood Pressure	BP QTL 2,14,36,132,174,240,243,336	non synonymous	tolerated(0.73)
F2r	2_25988115_G/A	Blood Pressure	BP QTL 18,36,115,240,243	non synonymous	tolerated(0.74)
Dym	18_71952016_T/C	-	BP QTL 41,48,54,229,230,232,233,234	non synonymous	tolerated(0.82)
Ppp1r13l	1_78888003_G/A	-	BP QTL 58,96,267,361 Stroke QTL Strs4	non synonymous	tolerated(0.99)
Lifr	2_56452598_G/T	Blood Pressure	BP QTL 2,14,36,174,206,240,243,336	non synonymous	tolerated(1)
Hps5	1_97289086_C/T	-	BP QTL 58,96,154,171,199,200,267,314	non synonymous	tolerated(1)
Zdhhc13	1_98589015_C/T	-	BP QTL 58,96,154,171,199,200,267	non synonymous	tolerated(1)
Fcgbp	1_83185291_A/T	-	BP QTL 58,96,267,361	non synonymous	tolerated(1)
Tcp10b	1_47607977_C/T	-	BP QTL 58,96,145	not in database	No result
Plg	1_42820003_C/A	-	BP QTL 58,96,145,360	not in database	No result
Sipa1l3	1_84540258_C/T	-	BP QTL 58,96,267,361 Stroke QTL Strs4	not in database	No result
F1LYG7_RAT	2_77590358_C/T	Blood Pressure	BP QTL 2,14,36,99,132,163,206,239,240	not in database	No result
F1M057_RAT	1_49160720_T/C	-	BP QTL 58,96,145	not in database	No result
F1LWY7_RAT	1_49033451_C/T	-	BP QTL 58,96,145	not in database	No result
Ndrg3	3_147316888_A/T	-	20,81,167,208,285	not in database	No result
D3ZH33_RAT	8_89073850_G/A	-	BP QTL 35,62,125,184,253,262,316,331	not in database	No result
D3ZPF8_RAT	11_13361353_T/C	-	BP QTL 187	not in database	No result
F1LPR6_RAT	6_138250689_G/C	-	no relevant QTL	not in database	No result
F1MA07_RAT	1_48959926_C/A	-	BP QTL 58,96,145	not in database	No result
D3ZN72_RAT	6_139814054_G/C	-	no relevant QTL	not in database	No result
Aff1	14_7016123_C/T	Left ventricular mass index	no relevant QTL	not in database	No result
Tldc2	3_147677628_G/A	-	BP QTL 20,208	stop gained	-
F1M0S7_RAT	3_148664776_G/T	-	BP QTL 20,81,167,176,208,285,320	synonymous	-
Lrfn1	1_83580117_G/A	-	BP QTL 58,96,267,361 Stroke QTL Strs4	synonymous	-
RGD1563145	3_149436622_A/G	-	BP QTL 20,81,167,176,208,285,320	synonymous	-

SIFT (Sorting Intolerant from Tolerant) algorithm predicts and score the effect of coding variants on protein function result of a non-synonymous SNP based on sequence homology and the physical properties of amino acids.

Table 4-8 $SHRSP_{Gla}$ unique stop gain SNPs compared to 27 NGA rat genomes.

Gene	Position	WKY_{Gla} vs $SHRSP_{Gla}$ QTL	RGD Blood pressure and stroke QTL	Rnor 5.0 Consequences
Catsperg1	1_84329456_C/T	-	BP QTL 58,96,267,361 Stroke QTL Strs4	Intronic
Hps5	1_97289152_G/A	-	BP QTL 58,96,154,171,199,200,267,314	Stop gain
Stim1	1_159924291_C/T	-	BP QTL 26,28,30,44,58 Stroke QTL Strs1	Stop gain

4.3.7.3 $SHRSP_{Gla}$ unique INDELs compared to 27 NGS rat genomes

We prioritised protein coding $SHRSP_{Gla}$ unique INDELs compared to 27 NGS rat genomes specifically the 4 non-synonymous and 3 frame shift INDELs as they were likely to have a deleterious effect on protein function. $SHRSP_{Gla}$ and WKY_{Gla} QTL regions for blood pressure, left ventricular index, and stroke as well as other QTLs for blood pressure and stroke from other rat strains were used to prioritise candidate genes. In addition, the consistent consequence prediction in the BN new reference genome Rnor 5.0 were also used to prioritised $SHRSP_{Gla}$ unique non-synonymous (Table 4-9) and frame shift (Table 4-10) INDELs.

Table 4-9 $SHRSP_{Gla}$ unique non-synonymous INDELs compared to 27 NGS rat genomes.

Gene	Position	WKY_{Gla} vs $SHRSP_{Gla}$ QTL	RGD Blood pressure and stroke QTL	Rnor 5.0 Consequences
LOC308401	1_78176494_CTC/-	-	BP QTL 58,96,267,361 Stroke QTL Strs4	Inframe deletion
Ccdc97	1_80924956_CAT/-	-	BP QTL 58,96,267,361 Stroke QTL Strs4	Inframe deletion
Abracl	1_13066145_GAG/-	-	BP QTL 58,360	Intronic
E9PT93_RAT	13_69973467_-/GAG	-	BP QTL 25,31,55,80,159,220,241,244	Intergenic

Table 4-10 $SHRSP_{Gla}$ unique frame shift INDELs compared to 27 NGS rat genomes.

Gene	Position	WKY_{Gla} vs $SHRSP_{Gla}$ QTL	RGD Blood pressure and stroke QTL	Rnor 5.0 Consequences
Atp11b	2_122263768_TGGCTGGGATTAAAGTATGGGTACTTAC/-	Blood Pressure	BP QTL 6,14,90,105,132,163,206,239	Frameshift
IGH-6	6_138250892_-/C	-	-	Frameshift
Ecel1	9_85939782_-/GACA	-	BP QTL 108,185,334,356	Intronic

4.3.8 SHRSP rat strains STAR SNPs analysis

Five SHRSP rat stains SNPs from STAR data were compared to each other against BN reference genome RGSC 3.4. Figure 4-15 shows that a high proportion of SNPs were common among SHRSP rat strains

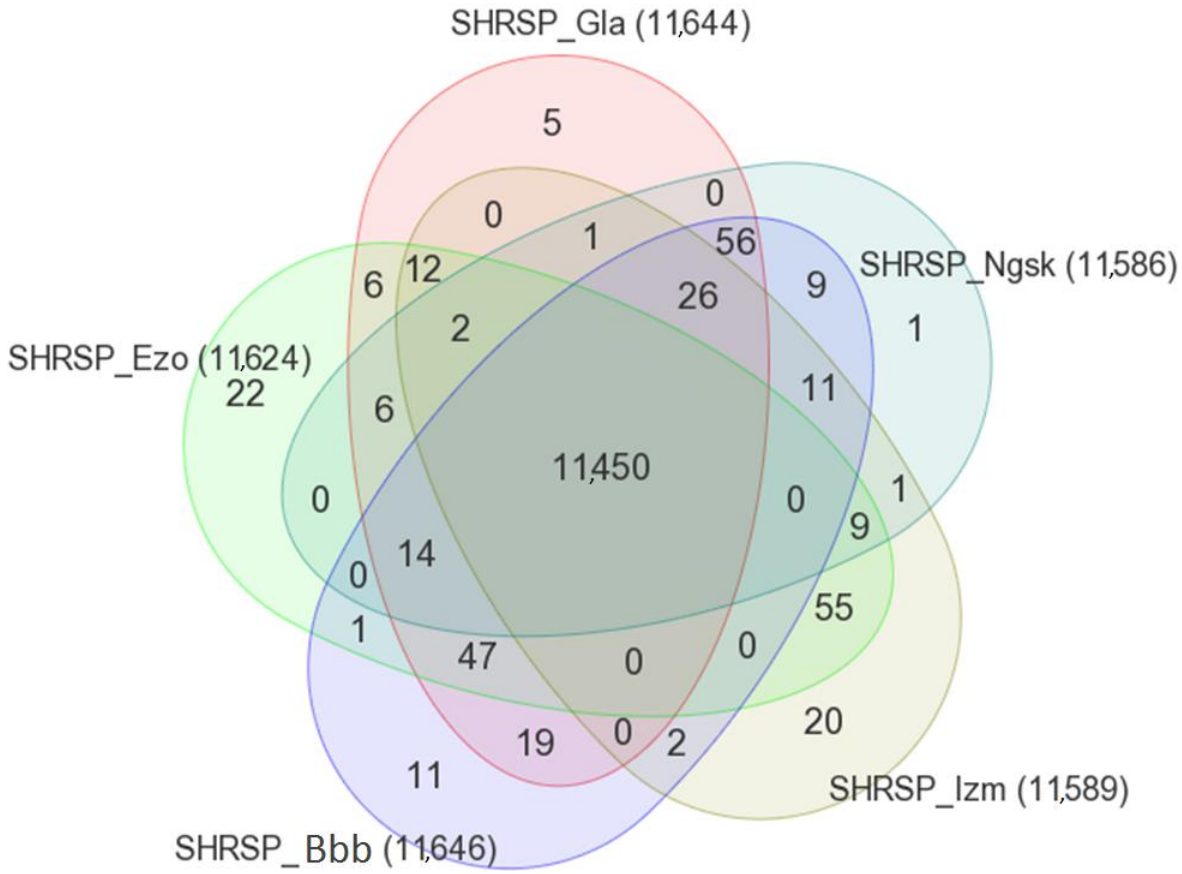


Figure 4-15 STAR data analysis among SHRSP rat strains.
Venn diagram of 5 SHRSP rat strains STAR data SNPs in which it shows high proportion of SNPs were common between all SHRSP rat strains.

4.4 Discussion

Even though the BN and WKY_{Gla} are normotensive rats and the SHRSP derived from the WKY strain, the WKY_{Gla} strain has more genomic variants compared to reference BN than to $SHRSP_{Gla}$. Nevertheless, another WKY strain (WKY_{NHsd}), also has more genomic variants than $SHRSP_{Gla}$ and WKY_{Gla} when compared to BN reference genome (Atanur et al., 2013). This suggests that the two WKY strains are genetically different from BN reference genome than the $SHRSP_{Gla}$; nonetheless, these genomics differences do not have an effect on blood pressure regulation. Nevertheless, a large number of sequence variants is common between $SHRSP_{Gla}$ and WKY_{Gla}. This observation is further validated by phylogenetic analyses performed by Atanur et al., 2013 in which both strains belong to the same subtree that derived from Japanese Wistar strains, models for hypertension. In addition, density distribution analysis of genomic variants revealed regions on all rat chromosomes with either low clusters or no SNPs and INDELs, which also suggesting that these regions are shared by descent. Moreover, the non uniform peak pattern of SNPs and INDEL tracks across the genome suggests that they are likely to be correlated, a phenomenon observed also in another Japanese Wister derived strain $SHR_{Olalpcv}$ (Atanur et al., 2010).

$SHRSP_{Gla}$ and WKY_{Gla} are inbred rat strains over several generations and they are expected to be homozygous at every locus in the genome. However, a percentage of 3.04 SNPs and 3.09 INDELs in WKY_{Gla} and 3.09 SNPs and 9.25 INDELs in $SHRSP_{Gla}$ were detected in heterozygous state. Possible explanations of heterozygous sequence variants are combinations of 1) biological true positive due to *de novo* mutations and incomplete inbreeding 2) technical false positive due to sequencing errors, mapping errors in problematic regions that suffer from CNV and highly repetitive sequence (Atanur et al., 2010; Simonis et al., 2012; Atanur et al., 2013). Guo et al., 2013 Sanger sequenced 45 NGS heterozygous SNPs and found that 6 were true positive while 39 were false positive, suggesting that caution should be taken into account when analysing NGS heterozygous sequence variants.

Functional annotation by Ensembl variants effect predictor showed that sequence variants of $SHRSP_{Gla}$ and WKY_{Gla} are higher in intergenic and intronic regions compared to the exonic region the same is observed in $SHR_{Olalpcv}$ (Atanur

et al., 2010), Dark Agouti (DA) and Fischer (F344) rat strains (Guo et al., 2013). This is most likely to the evolutionary pressure on protein coding genes, something that could explain the selective sizes of INDEL that are biased (triplet) towards the reading frame.

The functional implication of biological and canonical pathways' results from IPA of all $SHRSP_{Gla}$ and WKY_{Gla} NGS sequence variants captured some of the $SHRSP_{Gla}$ phenotypes. Isolating validated QTLs regions implicated protein coding variants and these are also likely to be causative for $SHRSP_{Gla}$ phenotypes.

The new BN assembly Rnor 5.0 improved in both base pairs sequence and total length when compared to the BN assembly RGSC 3.4. When our NGS short reads mapped to both BN old and new assemblies, there were additional bases mapped with the new BN reference 1.71 Gb for WKY_{Gla} and 1.34 Gb for $SHRSP_{Gla}$. There was also an improvement of at least 0.50x of average coverage in both $SHRSP_{Gla}$ and WKY_{Gla}. Moreover, from variant sequences that are unique between $SHRSP_{Gla}$ and WKY_{Gla} total 1,295,337 SNPs and 382,972 INDELs in the new BN assembly Rnor 5.0; this is more than the 1,163,332 SNPs and 213,130 INDELs with BN assembly RGSC 3.4. Variant effect predictor annotation 71 was also improved to give more defined consequences of sequence variants when compared to the previous version 69. The improvement of the reference genome was reflected in our genes of interest. For example in *Spp1*, new regions show sequence variants of interest and in *Hsd3b5* sequence gap was closed with BN new assembly Rnor 5.0.

The quality control results between the NGS and sequence capture when compared to RGSC 3.4 showed that 75% of the SNPs are common. This might be due to the design of the sequence capture bait in a way that maximizes the coverage area by focusing more on the exonic regions over intragenic regions resulting in more unique SNPs to the NGS analysis. In addition, sequence capture has unique SNPs because the NGS reads in one of the strain is not sufficient to be identified as a SNP, this is likely to be due to sequence capture much deeper coverage. The quality assessment of STAR Consortium data of true and comparable SNPs against NGS variants of $SHRSP_{Gla}$ and WKY_{Gla} compared to RGSC 3.4 confirmed that 99.3% of these SNPs are true positive (Saar et al., 2008). Both NGS data and sequence capture performed well in detecting Sanger sequencing

variants between the $SHRSP_{Gla}$ and WKY_{Gla}; however, sequence capture detected more variants than NGS data which might be due to its higher coverage resulting in higher sensitivity and specificity. Some unique NGS SNPs and INDELs between $SHRSP_{Gla}$ and WKY_{Gla} might not be called due to not fulfilling the sequence variants calling cut-off. Using IGV to manually browse reads across genes of interest can identify such variants which should be further validated by Sanger sequencing.

The frequencies of large deletions are much more common in the WKY_{Gla} when compared to $SHRSP_{Gla}$ against BN reference genome RGSC 3.4 this is the same with WKY_{NCrl} and WKY_{NHsd} (Atanur et al., 2013). For the purpose of this study, unique large deletions in the $SHRSP_{Gla}$ were prioritised. The large deletions in $SHRSP_{Gla}$, which are more common in 100-200kb, would either result in partial and/or complete deleted 478 genes. The read length of paired-end whole-genome which is 75 bp in $SHRSP_{Gla}$ and 100 bp in WKY_{Gla} with libraries of different insert size can have an impact on the accuracy of the identified structural variations. Guo et al., 2013 estimated that the accuracy of detected structural variant of one insert size library had 45-50% sensitivity. As a result, having multiple insert sizes of both small and large can increase the sensitivity of the detected structural variants (Atanur et al., 2010). Several groups have reported CNV can correlate with gene expression levels in mouse (Yalcin et al., 2011), rats (Guryev et al., 2008; Charchar et al., 2010; Simonis et al., 2012) and humans (Stranger et al., 2007) such integration can used as a reductive approach to prioritise causative candidate genes for $SHRSP_{Gla}$ phenotype.

When comparing SNPs from $SHRSP_{Gla}$, WKY_{Gla} and $SHR_{OlaIpcv}$ against the BN reference assembly RGSC 3.4, the intersection that is shared by all, has the highest proportion of SNPs which fits the WKY clade and supports that these strains are similar evolutionary and genetically. Additionally, the SNPs in the common shared intersection between the $SHRSP_{Gla}$ and $SHR_{OlaIpcv}$ can be prioritised as potentially hypertensive causal SNPs. Furthermore, SNPs that are unique to $SHRSP_{Gla}$ might be the causative of stroke-prone phenotype, nevertheless, another reductive strategy is needed to work with sensible variants number.

To further enrich for causative sequence variants of stroke-prone phenotype sequence variants, $SHRSP_{Gla}$ were compared to 27 rat genomes against BN reference genomes RGSC 3.4. To assess whether the protein coding sequence variants that are unique to the $SHRSP_{Gla}$ contribute to the $SHRSP_{Gla}$ cardiovascular phenotypes, we used our validated QTL regions on chromosome 2, 3 and 14 as well as QTL 5 for stroke to identify positional and functional candidate gene or genes. Consistency of sequence variants consequence prediction in the BN new reference genome Rnor 5.0 and SIFT deleterious predication non-synonymous SNPs were also prioritised. Moreover, we investigated whether the implicated gene encompassed a rat cardiovascular QTL and identified the most relevant citation that describes the implicated function.

$SHRSP_{Gla}$ unique non-synonymous SNPs with predicted deleterious effect on protein function implicated dentin matrix acidic phosphoprotein 1 (*Dmp1*) and solute carrier family 10, member 6 (*Slc10a6*) within our congenic regions on chromosome 14, and olfactory receptors 847 (*Olr847*) 852 (*Olr852*) within our stroke QTL. Dmp1 is a protein important in the mineralization of tissues; mutations within this gene in humans or complete genetic ablation in knock-out mouse resulted in hypophosphatemic rickets (low levels of phosphate in the blood) and elevated levels of fibroblast growth factor 23 (Rangiani et al., 2012) which induced left ventricular hypertrophy (Faul et al., 2011). *Slc10a6* has no known function that can cause LVH; nonetheless, it has been implicated with insulin resistance in a genome wide association study (Chen et al., 2012). The olfactory receptor genes which are responsible for detection of odour, form the largest gene family. However, we did not find a study implicating these genes in stroke. ATP-binding cassette, subfamily A (ABC1), member 1 (*Abca1*) was candidate gene to have a damaging effect on non-synonymous SNP at the border line score that mapped to our stroke QTL. Abca1 protein functions as cholesterol efflux pump, and the mouse knock-out model had higher instance of developing left and right ventricular hypertrophy (Christiansen-Weber et al., 2000), while mutations within *Abca1* increased the risk of stroke (Pasdar et al., 2007)

We extended the stroke candidate genes analysis in the $SHRSP_{Gla}$ to include deleterious non-synonymous SNPs within stroke QTLs of other SHRSP rat strains in which we identified two genes. One of these genes was the Ets2 repressor factor (*Erf*), a member of the EST family that regulates cell proliferation and

differentiation. Mutations in *Erf* have been associated with premature closure of the lambdoid cranial sutures (craniosynostosis) (Twigg et al., 2013), and knock-out mouse model developed hypoplasia of forebrain and anaemia (Papadaki et al., 2007) which increases significantly the risk for stroke (Maguire et al., 2007). Another implicated gene is Ras and Rab interactor-like (*Rinl*); however, this gene it does not have a relevant function to stroke. Nevertheless, it was recently published that NGS non-synonymous SNPs of $SHRSP_{izm}$ within its stroke QTL (Gandolgor et al., 2013) and two of these SNPs are common with deleterious SNPs that are unique to the $SHRSP_{Gla}$. These SNPs are mapped to *Rinl* and MHC I like leukocyte 1 (*Mill1*) which have been predicted to be involved in nutrient metabolism (Rabinovich et al., 2008). Another interesting non-synonymous SNP in a gene that does lie within stroke QTLs but function for stroke is plasminogen (*Plg*). Plg protein has an important role in dissolving blood clots and inhibiting angiogenesis; defects in this gene are likely to cause hypertension (Levi et al., 2001) and/or stroke (Su et al., 2008)

$SHRSP_{Gla}$ unique stop gain SNP implicated stromal interaction molecule 1 (*Stim1*) within other SHRSP stroke QTL and one for blood pressure. STIM1 protein mediates store operated calcium channels which plays an important role in arterial thrombosis and ischemic brain infarction (Varga-Szabo et al., 2008). Recently, *Stim1* was identified as candidate gene for stroke in the $SHRSP_{Izm}$ strain which harbour the same stop gain SNP identified by our analysis (Ferdaus et al., 2014).

$SHRSP_{Gla}$ unique non-synonymous INDELs implicated coiled-coil domain containing 97 (*Ccdc97*) and predicted gene Loc30401 within QTL stroke of other SHRSP strains; however, no relevant function was found. Finally, frame shift INDELs that are unique to $SHRSP_{Gla}$ implicated ATPase, class VI, type 11B (*ATP11b*) within our QTL for blood pressure. *ATP11b* is involved in the transport of aminophospholipid and ions across membranes, a function that can be important in the regulation of blood pressure.

STAR Consortium SNPs of the five SHRSP rat strains against BN RGSC 3.4 analysis shows that high proportion of SNPs are common; suggesting close phylogenetic proximity between these SHRSP rat strains. As we prioritised $SHRSP_{Gla}$ unique non synonymous SNPs that are common with $SHRSP_{izm}$, we can extend this to other

SHRSP rat strains to further prioritised sequence variants on whole genome level. Furthermore, over 600 inbred strains on Rat genome Database and in the near future these strains will be sequenced and comparing these with $SHRSP_{Gla}$ unique sequence variants will further prioritise candidate genes for $SHRSP_{Gla}$ phenotype.

5 Integration of transcriptomic and genomic variants to prioritise candidate genes in Rat and translation to Human

5.1 Introduction

Single dimension -omic data have proven beneficial in the identification of candidate genes, however, they may lead to complex, less reliable, and/or misleading biological interpretation when compared to more integrated approaches. It is a common belief that the overall surveillance of multiple biological systems, through the integration of genomic, transcriptomic and proteomic data, can facilitate a better biological interpretation which will eventually lead to the identification of prioritised candidate genes for the $SHRSP_{Gla}$ phenotype from an improved analysis to translate in human and vice versa.

The integration of different types of genomic variants and transcriptomic across tissues and time points at QTLs and genome level is helpful to understand, study and elucidate the genetic architecture of complex disease including hypertension in $SHRSP_{Gla}$. This posed a challenging task before the availability of NGS of $SHRSP_{Gla}$ and WKY_{Gla} genomes.

The identification of NGS sequence variants in our strains is beneficial for designing PCR assays primers and private microarray probes. The latter can be extended to public microarray probes by the identification of sequence variants within probes' target region which may alter probe hybridisation affinities leading to false results (Benovoy et al., 2008). Therefore, research groups such as Tim Aitman's, customised their public microarray gene expression chip to be unique to their rat strains, and the affected probes were masked/filtered to improve the analysis (Maratou et al., 2011). Although this issue can be solved by mRNA profiling using NGS on our rat strains, over the years we have compiled a transcriptomic library cataloguing thousands of differentially expressed genes between $SHRSP_{Gla}$ and WKY_{Gla} across tissues, times points and their response to salt using Affymetrix and Illumina chips.

Reductive methods have been proposed to identify and prioritise candidate genes from high-throughput transcriptomic data (Lombard et al., 2007) and genomics data (Jaffe et al., 2011) separately but not for integrating the two omics fields. Hence, we are proposing the Swiss cheese model as an integrative and reductive method to identify candidate genes for $SHRSP_{Gla}$ phenotype

(Figure 5-1). The model consists of an integrational approach for prioritising predicted protein coding sequence variants with positional and functional renal and cardiac differentially expressed genes that could potentially contribute to blood pressure regulation and cardiac mass in the $SHRSP_{Gla}$.

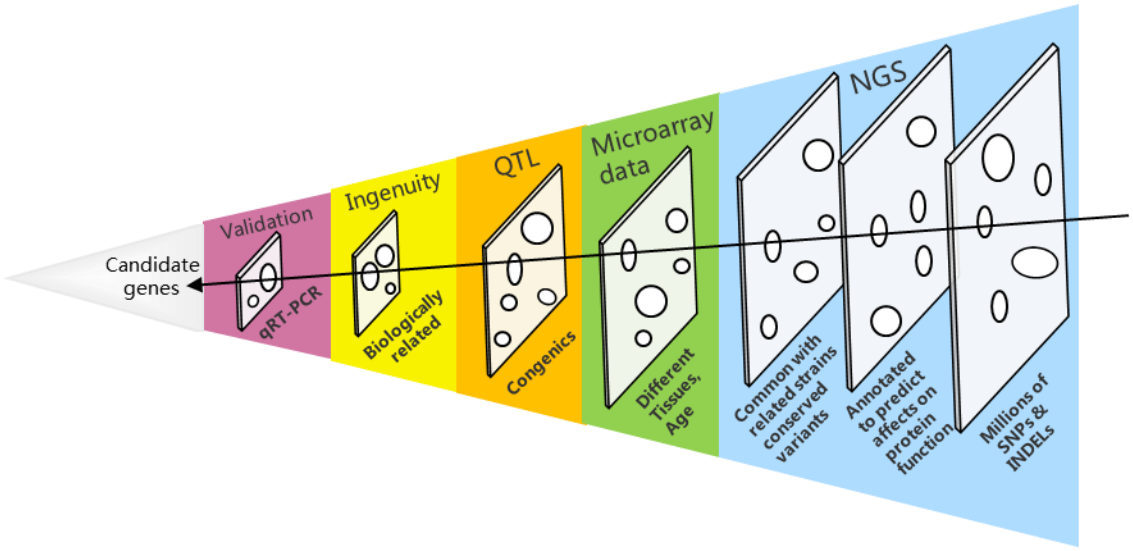


Figure 5-1 Swiss Cheese model: reductive and integrative method.
The pipeline starts with high number of unique genomic variants that are more likely to cause loss/gain of protein function. Then the resulted candidate genes should be significant differently expressed, using cut-offs of p-value < 0.05 and fold change > 2-fold, across time points and tissues within our QTLs. Functional and biologically relevance of the resulted candidate genes will be assessed and validate. If the genes passed all these criteria then the genes will be qualified to be candidate for $SHRSP_{Gla}$ phenotype.

There is no one perfect integrative method for transcriptomic and genomics data. Thus, utilising publicly availably related -omics data can be used as reductive and integrative method. Knowing that the $SHRSP_{Gla}$ have exceptionally higher blood pressure and cerebral stroke compared to the 27 rats that were sequenced in Atanur et al., 2013 including SHR strains. The integration of isolated $SHRSP_{Gla}$ unique sequence variants with transcriptomic brain data across frontal and mid-coronal sections and age between $SHRSP_{Gla}$ and WKY_{Gla} (Bailey et al., 2014) can serve as another integrative approach to isolate candidate genes causing hypertension and stroke in the $SHRSP_{Gla}$.

Clinician and geneticist may argue whether the SHRSP still useful model to identify candidate genes for hypertension when GWAS in human genome can directly accomplish this (Nabika et al., 2012). However, implicating human GWAS candidate genes in the $SHRSP_{Gla}$ at functional related regions, with sequence variations which are differentially expressed across tissues and age can further prioritise a more promising candidate genes for blood pressure and hypertension which are conserved between rat and human.

Recently, the role of miRNA targeting the renin-angiotensin-aldosterone system, vascular smooth muscle cells as well as endothelial and renal cells in human and rat for hypertension was studied (Batkai & Thum, 2012).Expression profiling of renal miRNA and mRNA between hypertensive human patients and control were integrated to identify and demonstrated that miRNAs can regulate renin expression (Marques et al., 2011). The identification of differentially expressed mir-181a miRNA and its targeted renin gene as candidate for hypertension in human (Marques et al., 2011) were translated with the same level of expressions in BPH/2J genetically hypertensive mice confirming their contribution in hypertension (Jackson et al., 2013).

The availability of human hypertensive expression profiling of renal miRNA and mRNA data (Marques et al., 2011) is a rare valuable translational opportunity to integrate with $SHRSP_{Gla}$ and WKY_{Gla} renal miRNA and mRNA to prioritise and translate conserved candidate miRNA and mRNA for hypertension. Even though human renal (medulla and cortex section) expression profiling for mRNA and miRNA expression profiling analysis in Marques et al., 2011 were corrected for ethnicity, age, weight, however, the p-value cut-off used in all comparisons was not consistent, and all above 0.1 as oppose to the standard < 0.05. It is unclear why the authors did so; however, previously Marques et al., 2010 published a meta analysis on all public transcriptomic data of hypertensive rats where some of the identified genes were later validate candidate genes for hypertension in human (Marques et al., 2011).

Human tissue availability for hypertension studies is severely restricted, making it hard to obtain samples for our renal and cardiac translation. Thus, cardiac transcriptomic data for heart failure is the next best data available to integrate with our rat strains data in order to prioritise and translate cardiac candidate

genes. However, our rat strains do not have a heart failure phenotype at the ages measured; nevertheless, given the fact the genes differently expressed consistently from early time till the hypertrophy phenotype established, this might lead eventually to heart failure at older age.

5.2 Aims

This chapter specific aims are:

- To integrate cardiac and renal transcriptomic data with genomics variants between $SHRSP_{Gla}$ and WKY_{Gla} and examine which types of genomic variants are likely to alter transcriptomic output at genome wide and congenic regions.

- To assess the improvements of gene expression analysis utilising $SHRSP_{Gla}$ and WKY_{Gla} sequence variants to customising public gene expression chips

- To apply our reductive and integrative Swiss cheese model to prioritise candidate genes that control blood pressure in the $SHRSP_{Gla}$.

- To isolate blood pressure and stroke candidate genes in SHRSP by the integration of $SHRSP_{Gla}$ unique sequence variants compared to 27 rat genomes and brain gene expression data.

- To assess how informative the $SHRSP_{Gla}$ compared to human GWAS candidate genes for blood pressure and hypertension.

- To translate renal and cardiac candidate genes and miRNA from our rat strains and human and vice versa.

5.3 Results

5.3.1 Prioritising rat candidate genes

A number of integrative approaches were used to prioritise candidate genes for the $SHRSP_{Gla}$.

5.3.1.1 Integration of rat transcriptomic data and genomic variants between $SHRSP_{Gla}$ and WKY_{Gla}

The integration of transcriptomic data and different types of genomic variants data between $SHRSP_{Gla}$ and WKY_{Gla} were combined using IPA to investigate a potential direct relation. Significantly differentially expressed genes in the heart of neonates (486 genes), at 5 weeks (182 genes) and at 16 weeks of age (330 genes), as well as in the kidney at 16 weeks of age (482 genes) and at 21 weeks of age (1,172 genes) were combined with different types of genomic variants.

5.3.1.1.1 Integration of transcriptomic data and small sequence variants between $SHRSP_{Gla}$ and WKY_{Gla}

Affected genes with small sequence variants that mapped to IPA for SNP and INDEL, in coding and non-coding regions, were compared to renal and cardiac transcriptomic data at different time points. Overlapping genes were assessed for implication in our congenic regions (Figure 5-2). The highest proportion of overlapping genes in both tissues was affected by non-coding variants, especially upstream, downstream, in intronic and splice sites – this included genes within our congenic regions and the proportion affected by coding variants especially stop lost/gained, frameshift and essential site splice.

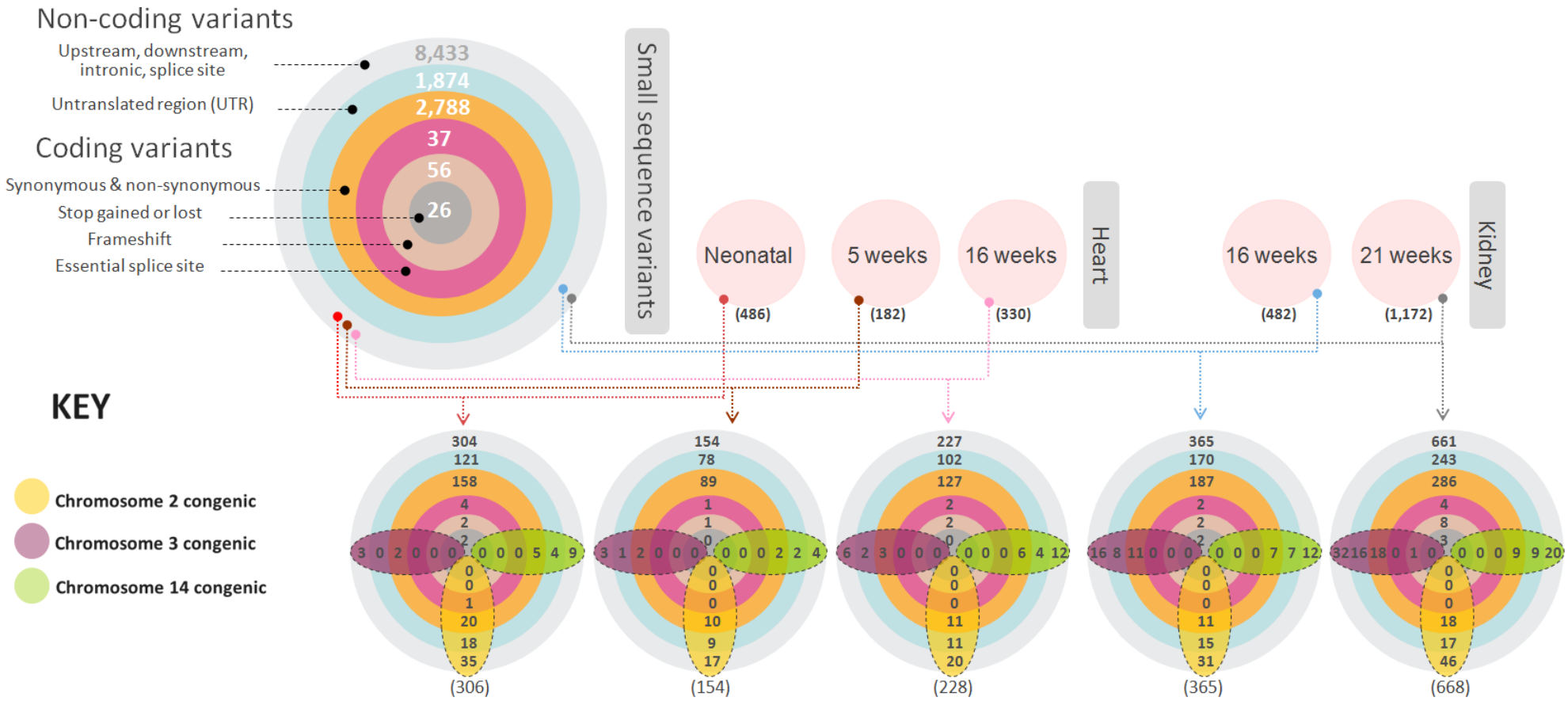


Figure 5-2 Integration of transcriptomic data and small sequence variants.
Top left; affected genes with small sequence variants (SNPs and INDELs) in coding and non coding regions with different predictive consequences between $SHRSP_{Gla}$ and WKY_{Gla}. Top right; gene expression data of $SHRSP_{Gla}$ and WKY_{Gla} from kidney and heart at different time points. Bottom circles; genes common between transcriptomic data and small single genomic variants as well as genes mapped to our congenic regions.

5.3.1.1.2 Integration of transcriptomic data and structural variants

To investigate other types of genomic variants that were likely to alter gene expression levels, we combined transcriptomic data with structural variants between $SHRSP_{Gla}$ and WKY_{Gla} using IPA.

5.3.1.1.2.1 Integration of transcriptomic data and large deletions

Subsequently, we combined partially and/or fully deleted genes that were unique to $SHRSP_{Gla}$ (273 genes) and WKY_{Gla} (1,583 genes), with transcriptomic data from $SHRSP_{Gla}$ and WKY_{Gla}, using IPA (Figure 5-3). The highest proportion of overlapping genes was observed between large deletions unique to WKY_{Gla} (including genes within our congenic regions) and large deletions unique to $SHRSP_{Gla}$.

5.3.1.1.2.2 Integration of transcriptomic data and copy number variations

We decided to combine CNV with transcriptomic data from $SHRSP_{Gla}$ and WKY_{Gla} using IPA (Figure 5-4, Figure 5-5). Genes affected by CNV duplications in $SHRSP_{Gla}$ (Figure 5-5) and CNV deletions in WKY_{Gla} (Figure 5-4) had the highest proportion of significantly differentially expressed genes in both tissues at different time points, including genes mapped to our congenic regions. Furthermore, structural variations of copy number overlapped more with genes that present with changes in expression levels, including genes mapping to our congenic regions rather than large deletions.

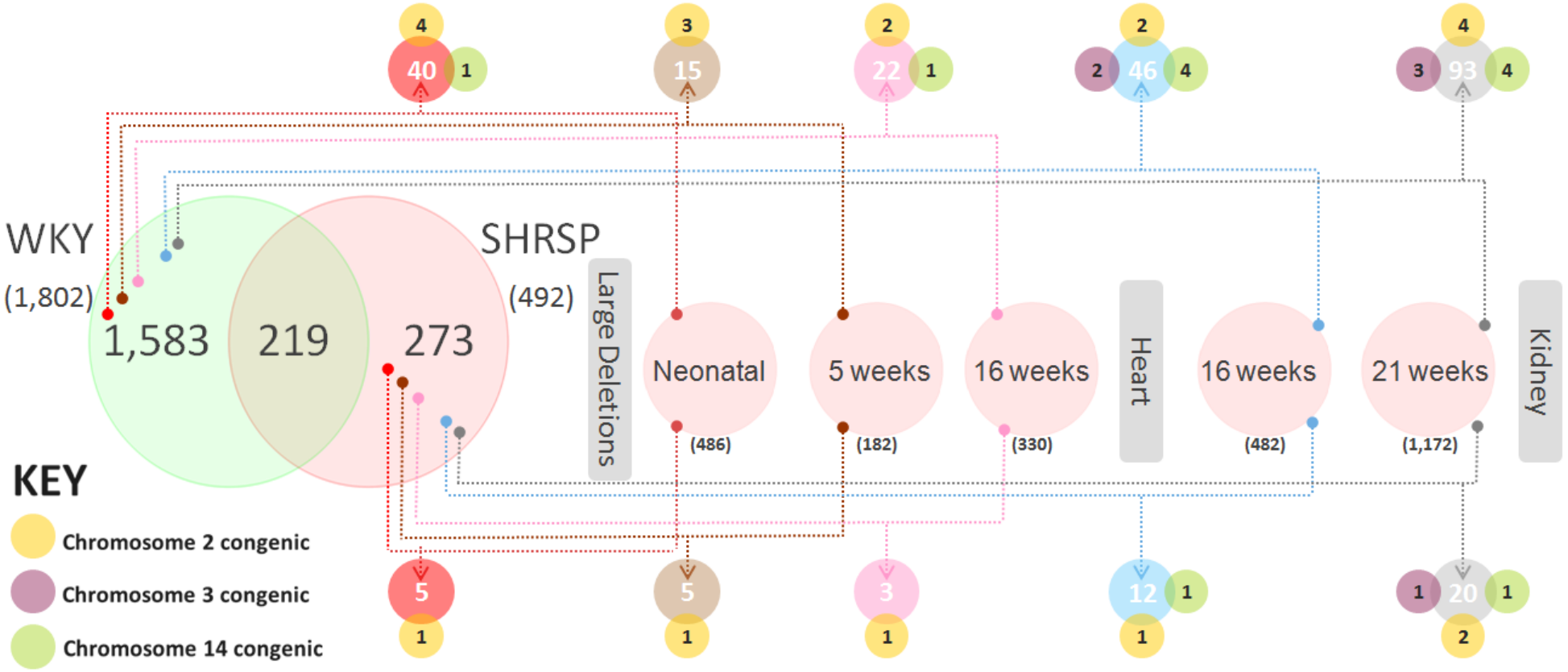


Figure 5-3 Integration of transcriptomic data and large deletions.
Left; Venn diagram of large deletions, showing genes unique and common between the $SHRSP_{Gla}$ and WKY_{Gla}. Right; significantly differentially expressed genes from kidney and heart at different time points between $SHRSP_{Gla}$ and WKY_{Gla}. Top and bottom circles; overlapping genes as well as genes implicated in our congenic regions.

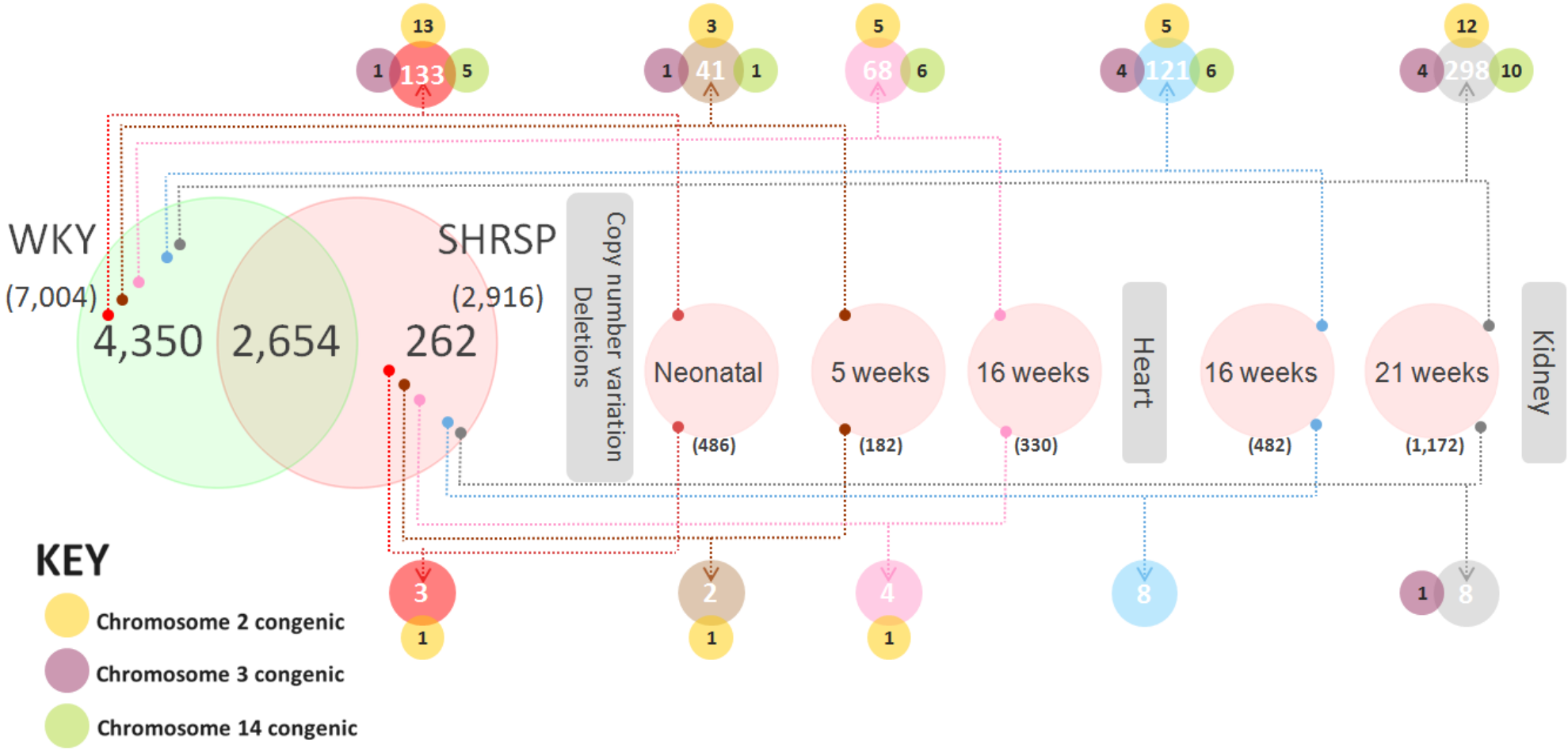


Figure 5-4 Integration of transcriptomic data and CNV of deletion.
Combining CNV of deletions with transcriptomic data between $SHRSP_{Gla}$ and WKY_{Gla} revealed that more significantly differentially expressed genes overlapped with CNV of deletions unique to WKY_{Gla}, including genes within our congenic regions than $SHRSP_{Gla}$.

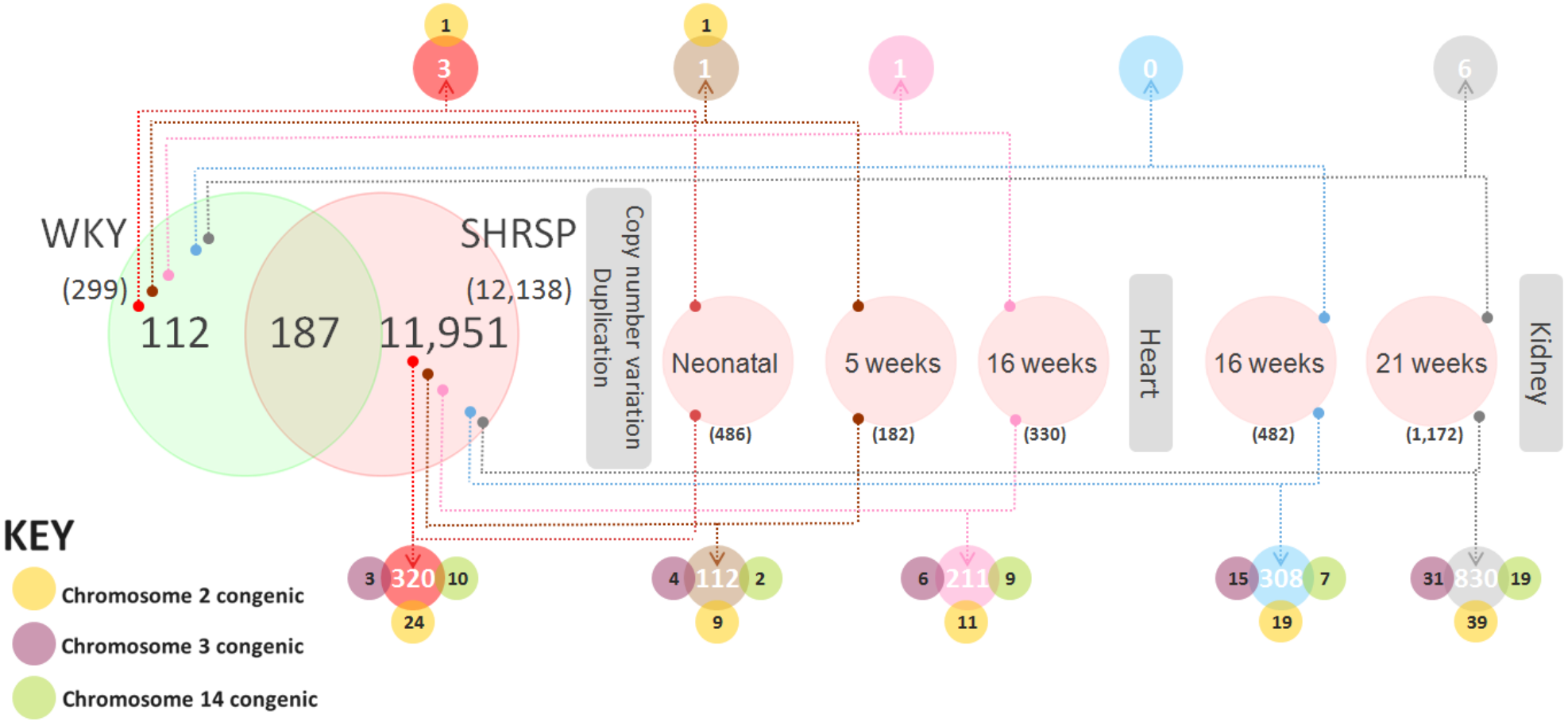


Figure 5-5 Integration of transcriptomic data and CNV of duplication.
Combining CNV of duplications with transcriptomic data between $SHRSP_{Gla}$ and WKY_{Gla} showed more significantly differentially expressed genes overlapped with CNV deletions unique to $SHRSP_{Gla}$ including genes within our congenic regions than WKY_{Gla}.

5.3.1.2 Improvement of microarray analysis by utilising NGS sequence variants.

5.3.1.2.1 Utilising NGS data to improve rat Affymetrix microarray analysis

We then assessed the potential of optimisation of the Affymetrix microarray by removing probes whose targeted regions harboured SNPs, and therefore, could lead to false results. Significantly differentially expressed genes of chromosomes 2 congenic strain (SP.WKY$_{Gla}$2a) versus SHRSP$_{Gla}$ at 21 weeks of age in kidney identified using custom Affymetrix chip description file (CDF) unique to our strains were compared to default (BrainArray) Affymetrix CDF. The overall result revealed a 23% difference between the two Affymetrix CDF. Moreover, additional positional candidates were identified from the custom CDF when the results were filtered with regards to the SP.WKY$_{Gla}$2a congenic interval (Figure 5-6).

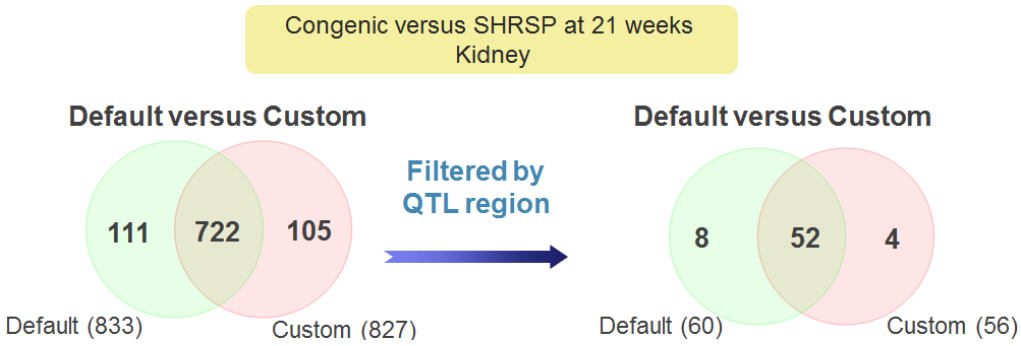


Figure 5-6 Integration of transcriptomic data with sequence variants to improve microarray analysis.
Affymetrix CDF unique to our strains analysis had an impact on overall results when compared to the default (BrainArray) Affymetrix CDF analysis between chromosome 2 congenic strain and SHRSP$_{Gla}$ at 21 weeks of age in kidney including positional candidate genes. The two different CDF analysis were conducted on Partek® Genomic Suite using a cut-off of FDR <0.05.

5.3.1.2.2 Integrating Illumina rat expression BeadChip probes with NGS SNPs

We assessed the optimisation of microarray analysis of our rat strains on Illumina platform. 633 probes out of 24,000 total probes/array, whose targeted regions harbour SNPs, were identified and integrated with our QTL/congenic regions on chromosome 2, 3 and 14. Genes mapped to affected Illumina rat probes' within our congenic regions were identified by using IPA, which should be removed to customised Illumina rat BeadChip to our rat strains (Figure 5-7).

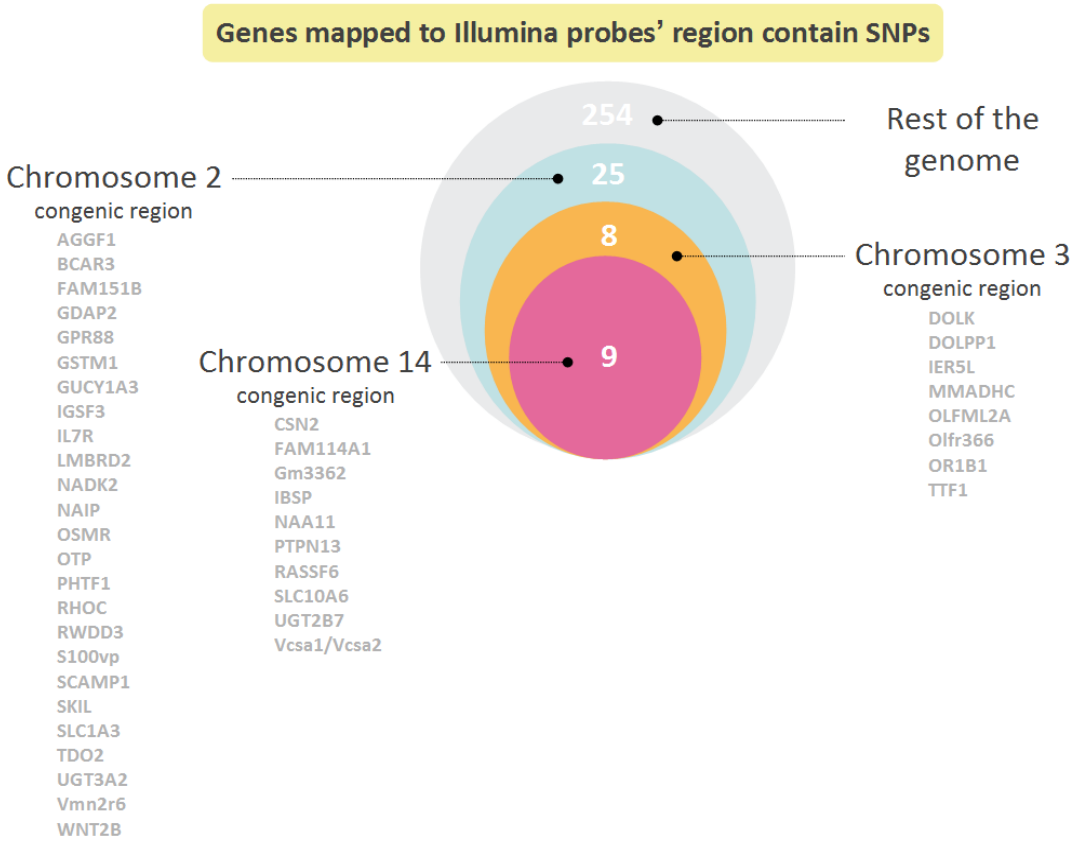


Figure 5-7 Utilising NGS data to identify affected probes on rat Illumina BeadChip.
Affected Illumina rat probes (663 probes), that mapped to 302 genes in IPA, were integrated with our congenic regions of chromosome 2 (25 genes), 3 (8 genes) and 14 (9 genes).

5.3.1.2.2.1 qRT-PCR validation of genes whose probes' region harbour SNPs

In order to investigate the microarray analysis of our rat strains on the Illumina platform, we validated our results by performing qRT-PCR analysis. qRT-PCR analysis was performed on genes in which their Illumina rat probes' region contained SNPs within our congenic regions that should be removed to avoid false positive results. Our data demonstrated that the expression of *Gstm1*, previously identified as functional and positional candidate gene in the $SHRSP_{Gla}$, mRNA levels was significantly elevated in the 21 weeks of age kidney of WKY_{Gla} when compared to $SHRSP_{Gla}$ (Figure 5-8), although this could have been done at other age of animals (McBride et al., 2003; McBride et al.,2005). Similarly, we observed a significant increase at the protein tyrosine phosphatase, non-receptor type 13 (*Ptpn13*), potential positional candidate gene for LVF which maps to LVMI QTL on chromosome 14, mRNA expression in the 16 week heart at 16 weeks of WKY_{Gla} compared to $SHRSP_{Gla}$. Statistical analysis was performed by using the Student's t-test.

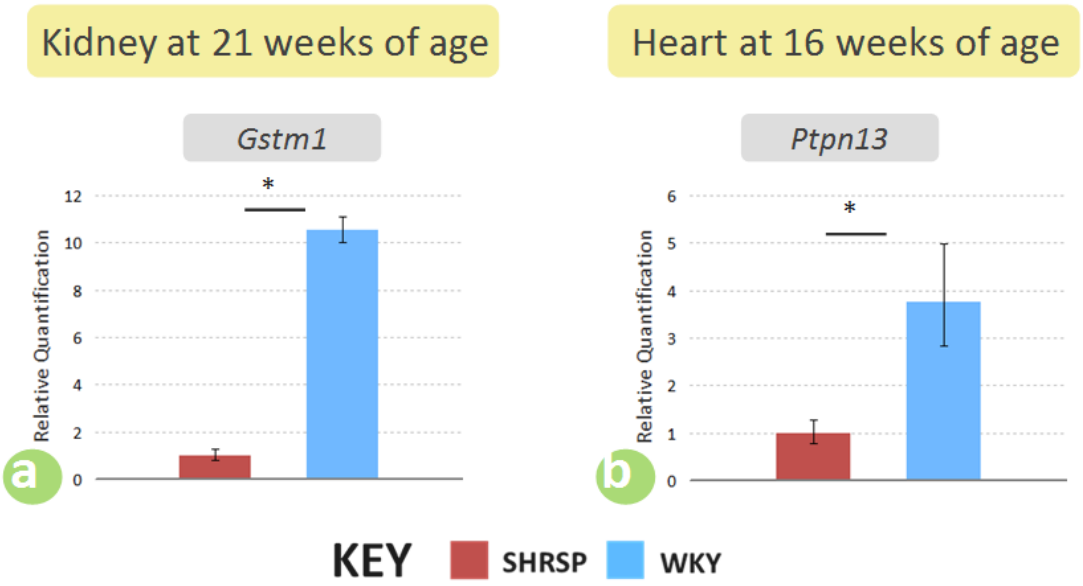


Figure 5-8 qRT-PCR validation of genes whose probes' region harbour SNPs
a) *Gstm1* is significantly elevated in WKY_{Gla} when compared to $SHRSP_{Gla}$ at 21 weeks of age in kidney. **b**) *Ptpn13* is significantly elevated in WKY_{Gla} when compared to $SHRSP_{Gla}$ at 16 weeks of age in heart. *P<0.05.

5.3.1.3 Application of "Swiss cheese" integrative and reductive model in our rat strains

5.3.1.3.1 "Swiss cheese" candidate genes for the $SHRSP_{Gla}$ phenotype

Protein coding variants; 769 non-synonymous coding, 5 stop gained, 24 frame shift coding variants within chromosome 2, 3, and 14 congenic regions were associated with significantly differently expressed positional candidate genes; *Gstm1*, *Fubp3*, *Ugt2b7*, and *Afm* in kidney and *Atp6ap1l* and *Gpr98* in heart (Table 5-1).

Table 5-1 Integrative and reductive "Swiss cheese" model candidate genes.

Partek (ANOVA) fold change 2 and FDR <0.05					
Tissue	Heart			Kidney	
Age	Neonatal	5 weeks	16 weeks	16 weeks	21 weeks
Genes	**Chromosome 2 QTL** *Atp6ap1*|: *ATPase, H+ transporting, lysosomal accessory protein 1-like* *Gpr98:* G protein-coupled receptor 98			**Chromosome 2 QTL** *Gstm1: Glutathione S-Transferase Mu type 1* *Ect2: Epithelial cell transforming sequence 2* **Chromosome 3 QTL** *Fubp3: Far upstream element-binding protein 3* *Lrsam1: Leucine rich repeat and sterile alpha motif* **Chromosome 14 QTL** *Klhl8: Kelch-like 8 (Drosophila)* *Ugt2b7: Uridine diphosphate glycosyltransferase 2* *Afm: Afamin* *Abcg3: ATP-binding cassette, sub-family G*	

5.3.1.3.1.1 NGS variants in prioritised "Swiss cheese" candidate genes

The "Swiss cheese" model candidate genes - *Fubp3*, *Ugt2b7*, *Afm*, and *Gpr98* - were prioritised for further analysis. These genes also harboured sequence variants in important non coding protein regions which could also contribute to the differential gene expression levels between $SHRSP_{Gla}$ and WKY_{Gla} (Table 5-2, Table 5-3, Table 5-4, and Table 5-5).

Table 5-2 *Fubp3* variants between $SHRSP_{Gla}$ and WKY_{Gla} compared to BN RGSC 3.4 assembly.

SNPs		INDELs	
Downstream	24	Downstream	1
Intronic	228	Intronic	5
Upstream	5	Upstream	4
Splice site	2	Splice site	1
Non synonymous coding	3		
Synonymous coding	5		
3Prime UTR	1		

Table 5-3 *Ugt2b7* variants between $SHRSP_{Gla}$ and WKY_{Gla} compared to BN RGSC 3.4 assembly.

SNPs		INDELs	
Downstream	35	Downstream	3
Intronic	162	Intronic	17
Upstream	37	Upstream	6
Splice site	1		
Non synonymous coding	2		
Synonymous coding	4		
3Prime UTR	3		

Table 5-4 *Afm* variants between $SHRSP_{Gla}$ and WKY_{Gla} compared to BN RGSC 3.4 assembly.

SNPs		INDELs	
Downstream	2	Downstream	2
Intronic	26	Intronic	7
Upstream	23		
Non synonymous coding	1		
Synonymous coding	1		
5Prime UTR	2		

Table 5-5 *Gpr98* variants between $SHRSP_{Gla}$ and WKY_{Gla} compared to BN RGSC 3.4 assembly.

SNPs		INDELs	
Downstream	10	Downstream	4
Intronic	11	Intronic	11
Upstream	11	Upstream	7
Splice site	1	Splice site	1
Non synonymous coding	9		
Synonymous coding	10		
3Prime UTR	11		

5.3.1.3.1.2 Validation of prioritised "Swiss cheese" candidate genes by using qRT-PCR

The "Swiss cheese" model prioritised candidate genes *Ugt2b7* and *Afm*, were further validated by qRT-PCR. *Ugt2b7* and *Afm* mRNA level expression in kidney at 21 weeks of age and *Ugt2b7* at 16 weeks of age were significantly elevated in WKY_{Gla} when compared to $SHRSP_{Gla}$. However, the expression of *Fubp3* mRNA levels in kidney at 16 and 21 weeks of age showed no significant changes between $SHRSP_{Gla}$ and WKY_{Gla} (Figure 5-9), nevertheless, the trend of fold change supports the microarray results (data not shown). Furthermore, *Gpr98* mRNA levels were not detectable in the hearts of neonates, or at 5 and 16 weeks of age using qRT-PCR in both $SHRSP_{Gla}$ and WKY_{Gla}. Statistical analysis was performed by using the Student's t-test.

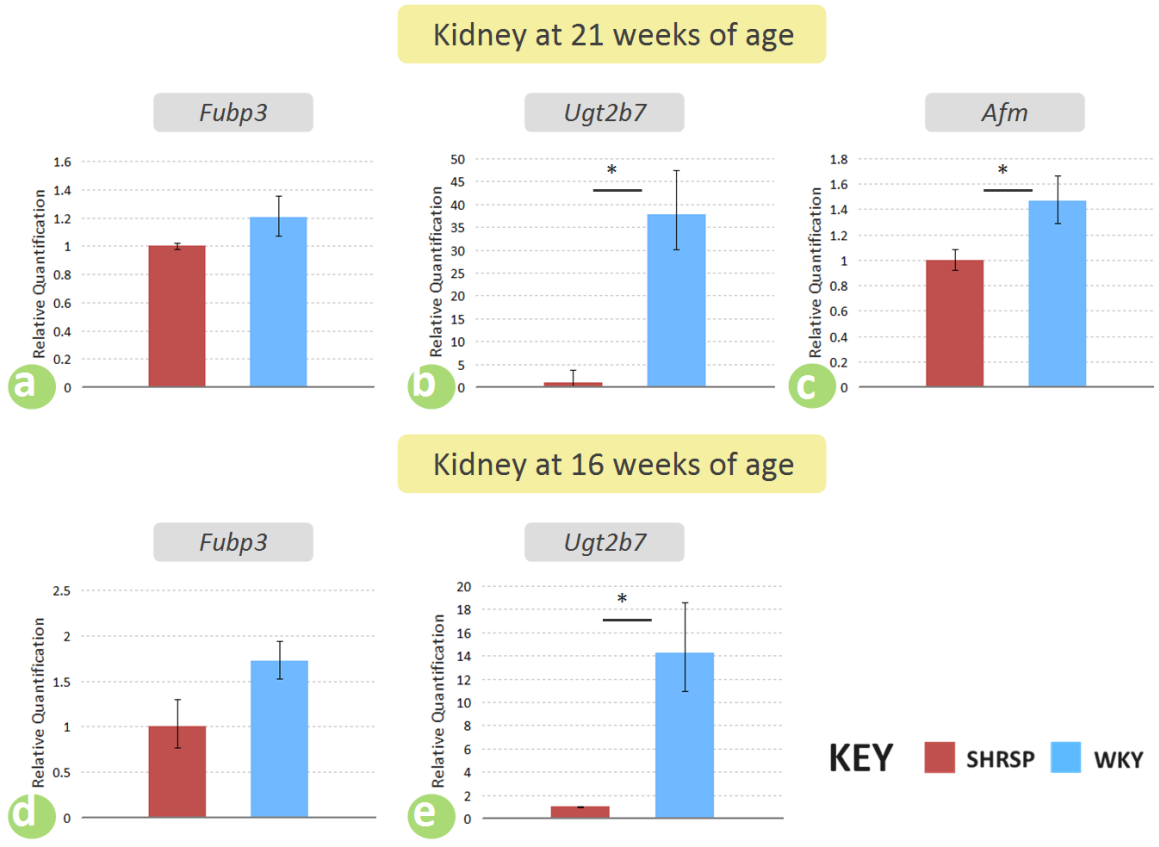


Figure 5-9 qRT-PCR of Swiss cheese candidate genes in kidney at 16 and 21 weeks of age of $SHRSP_{Gla}$ and WKY_{Gla} strains.
At 21 weeks of age **a**) the expression pattern of *Fubp3* shows no significant change between $SHRSP_{Gla}$ and WKY_{Gla}. However, mRNA levels in **b**) *Ugt2b7* and **c**) *Afm* is significantly elevated in WKY_{Gla} when compared to $SHRSP_{Gla}$. At 21 weeks of age **d**) *Fubp3* shows no significant change between our rat strains, however, **e**) *Ugt2b7* is significantly elevated in WKY_{Gla} when compared to $SHRSP_{Gla}$. *P<0.05.

5.3.1.3.1.3 Troubleshooting Gpr98 experimental data result

Further investigation of the *Gpr98* qRT-PCR results was carried out by using the Ensembl gene annotation versions 69 and 71 and visualising the mapped NGS reads in *Gpr98*. Ensembl annotation version 69 of BN RGSC 3.4 assembly assigned the Ensembl gene id for *Gpr98* as ENSRNOG00000016306; this has been changed in Ensembl annotation version 71 for new BN Rnor 5.0 to novel gene and *Gpr98* has been assigned a different Ensembl id (ENSRNOG00000046496). The discrepancy between the *Gpr98* Ensembl ids may have affected the probe design, and this could explain why the qRT-PCR result did not validate. Additionally, an explanation regarding the qRT-PCR result in $SHRSP_{Gla}$, could be the large deletion of 2,199 bp in the $SHRSP_{Gla}$ (Figure 5-10).

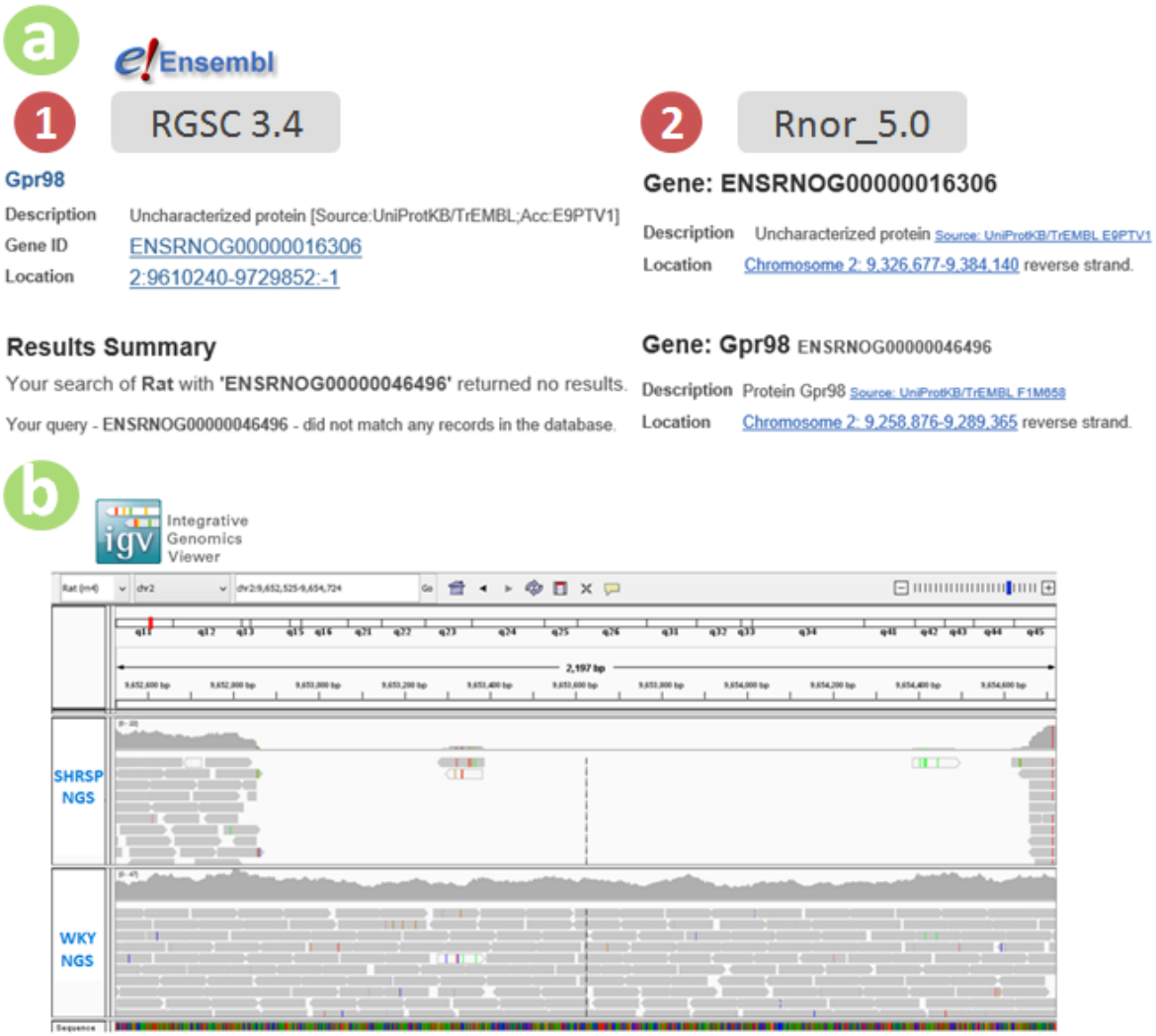


Figure 5-10 Trouble shooting of *Gpr98* experimental data.
a) Ensembl gene id for *Gpr98* in **a1**) BN RGSC 3.4 assembly differ than Ensemble gene id in **a2**) BN Rnor 5.0. **b**) NGS unmapped (white gap) reads (represented in grey) of *Gpr98* indicating large deletion in $SHRSP_{Gla}$ when visualised by IGV, but, that is not the case in the WKY_{Gla}.

5.3.1.4 Integrating $SHRSP_{Gla}$ unique small sequence variants with brain gene expression between $SHRSP_{Gla}$ and WKY_{Gla}.

In order to prioritise stroke candidate genes in $SHRSP_{Gla}$, we integrated unique small sequence variants to the $SHRSP_{Gla}$ when compared to 27 rat strains genome (Atanur et al., 2013) with previously published transcriptomic data of $SHRSP_{Gla}$ and WKY_{Gla} frontal and mid-coronal brain sections at different time points (Bailey et al., 2014) and salt loaded analysed with Rank Products (Breitling et al., 2004) and FDR cut-off of <0.05. $SHRSP_{Gla}$ and WKY_{Gla} QTL regions for blood pressure, left ventricular index, and stroke as well as stroke QTLs from other rat strains on RGD were used to prioritise stroke candidate genes in the $SHRSP_{Gla}$ (Table 5-6).

Table 5-6 Integrating $SHRSP_{Gla}$ unique variants compared 27 genome sequence rat strains with brain gene expression of $SHRSP_{Gla}$ and WKY_{Gla}.

Rat		WKY_{Gla} versus $SHRSP_{Gla}$											
Gene	RGD Stroke QTL	QTL	Small sequence variants unique to $SHRSP_{Gla}$								Brain gene expression			
			upstream	downstream	intronic	3' UTR	splice site	synonymous	non-synonymous	frame shift	Section	Time point	FDR p-value	Fold change
Apoc1	Strs4	-	1	-	-	-	-	-	-	-	Frontal	16 wk	1.50E-02	-1.871
											Mid-coronal	16 wk	3.93E-02	-1.696
Gmfg	Strs4	-	5	-	-	-	-	-	-	-	Mid-coronal	16 wk	2.37E-02	-1.683
Map4k1	Strs4	-		5	6	-	-	1	-	-	Frontal	5 wk	3.95E-02	1.453
Mas1	-	-	2	-	6	-	-	-	-	-	Mid-coronal	5 wk	4.11E-02	1.552
Med29	Strs4	-	-	2	-	-	-	-	-	-	Frontal	5 wk	4.72E-02	1.461
Mrpl18	-	-	1	-	-	-	-	-	-	-	Frontal	5 wk	0.00E+00	27.746
											Mid-coronal	5 wk	0.00E+00	29.126
											Frontal	16 wk	0.00E+00	25.422
											Mid-coronal	16 wk	0.00E+00	31.425
											Frontal	21 wk	0.00E+00	30.669
											Mid-coronal	21 wk	0.00E+00	25.892
											Frontal	21 wk and salt	0.00E+00	39.417
											Mid-coronal	21 wk and salt	0.00E+00	41.708
Rps16	Strs4	-	-	1	1	-	-	-	-	-	Frontal	5 wk	1.17E-04	-3.007
											Mid-coronal	5 wk	0.00E+00	-3.033
											Frontal	16 wk	0.00E+00	-3.059
											Mid-coronal	16 wk	0.00E+00	-2.924
											Frontal	21 wk	0.00E+00	-3.188
											Mid-coronal	21 wk	0.00E+00	-3.303
											Frontal	21 wk and salt	3.98E-04	-2.48
											Mid-coronal	21 wk and salt	3.16E-04	-2.749
Znf235	Strs4	-	-	-	-	-	-	1	-	-	Frontal	21 wk	1.16E-02	2.049

Continued

Rat		WKY$_{Gla}$ versus SHRSP$_{Gla}$											
Gene	RGD Stroke QTL	QTL	Small sequence variants unique to SHRSP$_{Gla}$								Brain gene expression			
			upstream	downstream	intronic	3' UTR	splice site	synonymous	non-synonymous	frame shift	Section	Time point	FDR p-value	Fold change
Znf566	Strs4	-	-	-	5	-	-	-	-	-	Frontal	5 wk	1.60E-04	-2.663
											Mid-coronal	5 wk	1.47E-04	-2.34
											Frontal	16 wk	1.19E-04	-2.516
											Mid-coronal	16 wk	4.39E-04	-2.645
											Frontal	21 wk	3.02E-03	-1.937
											Mid-coronal	21 wk	2.60E-04	-2.767
											Frontal	21 wk and salt	3.54E-02	-1.801
											Mid-coronal	21 wk and salt	1.07E-03	-2.27
Atp11b	-	Blood pressure (chromosome 2)	-	-	-	-	-	-	-	1	Frontal	5 wk	1.57E-02	1.581
											Mid-coronal	5 wk	4.78E-03	1.636
											Mid-coronal	16 wk	3.38E-02	1.672
											Frontal	21 wk and salt	1.34E-02	2.056
Egflam	-	Blood pressure (chromosome 2)	1	7	173	-	-	3	1		Mid-coronal	5 wk	3.44E-02	1.474
Mctp1	-	Blood pressure (chromosome 2)	-	-	1	-	-	-	-	-	Frontal	5 wk	3.19E-02	-1.827
Negr1	-	-	-	-	1	-	-	-	-	-	Frontal	5 wk	8.28E-03	-1.898
Ntng1	-	Blood pressure (chromosome 2)	-	-	1	-	-	-	-	-	Mid-coronal	21 wk and salt	1.33E-02	-1.86
Pxmp4	-	-	10	3	30	1	-	-	-	-	Frontal	5 wk	0.00E+00	2.831
											Mid-coronal	5 wk	4.33E-05	3.054
											Frontal	16 wk	0.00E+00	3.904
											Mid-coronal	16 wk	4.33E-05	3.136
											Frontal	21 wk	1.45E-04	3.638
											Mid-coronal	21 wk	1.75E-04	2.998
											Frontal	21 wk and salt	0.00E+00	4.659
											Mid-coronal	21 wk and salt	0.00E+00	3.604
Ttf1	-	Blood pressure (chromosome 3)	-	-	2	-	-	-	-	-	Frontal	5 wk	1.56E-02	1.523
											Mid-coronal	5 wk	4.37E-02	1.439
Cntn4	-	-	-	-	1	-	-	-	-	-	Frontal	5 wk	2.69E-02	-1.535
Iqub	Strs3	-	1	1	10	-	-	-	-	-	Mid-coronal	5 wk	1.82E-02	1.558
											Frontal	21 wk	3.12E-02	1.674
											Mid-coronal	21 wk and salt	4.60E-02	1.595
Rybp	-	-	-	-	1	-	-	-	-	-	Frontal	5 wk	1.80E-02	-1.556
											Mid-coronal	5 wk	2.35E-02	-1.568
Snx10	Strs3	-	-	5	29	1	1	-	-	-	Frontal	5 wk	8.82E-03	-1.849
											Frontal	16 wk	9.46E-03	-1.621
											Mid-coronal	16 wk	4.30E-02	-1.506
											Frontal	21 wk and salt	1.24E-04	-2.744
											Mid-coronal	21 wk and salt	6.86E-03	-1.959
Nudt14	-	-	2	2	-	-	-	-	-	-	Frontal	5 wk	7.36E-03	1.585
											Mid-coronal	5 wk	1.18E-02	1.589
Tmem121	-	-	1	22	-	-	-	-	-	-	Frontal	5 wk	3.78E-02	1.466
Tmem179	-	-	-	2	1	-	-	-	-	-	Frontal	21 wk	7.10E-03	2.29

Continued

Rat		WKY$_{Gla}$ versus SHRSP$_{Gla}$											
Gene	RGD Stroke QTL	QTL	Small sequence variants unique to SHRSP$_{Gla}$								Brain gene expression			
			upstream	downstream	intronic	3' UTR	splice site	synonymous	non-synonymous	frame shift	Section	Time point	FDR p-value	Fold change
Zcchc8	-	-	-	-	1	-	-	-	-	-	Frontal	16 wk	4.13E-02	-1.616
Lamc2	-	-	1	-	70	-	-	3	-	-	Frontal	21 wk and salt	2.97E-02	1.732
Abcg3l1	-	Left ventricular mass index	3	-	23	-	-	1	-	-	Frontal	5 wk	4.78E-02	1.43
Hsd17b13	-	Left ventricular mass index	2	1	7	1	-	-	-	-	Mid-coronal	5 wk	1.18E-02	1.537
Loc360919	-	Left ventricular mass index	2	-	-	-	-	-	-	-	Frontal	5 wk	2.25E-03	1.736
											Mid-coronal	5 wk	1.02E-02	1.717
											Frontal	16 wk	4.13E-02	1.57
											Mid-coronal	16 wk	2.77E-02	1.628
											Mid-coronal	21 wk	3.17E-02	1.803
											Frontal	21 wk and salt	1.22E-02	1.742
											Mid-coronal	21 wk and salt	1.42E-02	1.93
Znf644	-	-	1	14	10	-	-	-	-	-	Frontal	16 wk	2.23E-02	1.688
Nrg1	-	-	-	-	1	-	-	-	-	-	Mid-coronal	16 wk	2.51E-02	1.605
											Mid-coronal	21 wk	2.04E-02	1.699
Phf7	-	-	-	1	-	-	-	-	-	-	Frontal	21 wk	2.20E-02	-1.557
Rmi1	-	-	2	6	-	-	-	-	1	-	Frontal	21 wk	3.36E-02	-1.598
Tcf4	Strs5	-	4	-	1	-	-	-	-	-	Frontal	5 wk	1.13E-02	1.643
											Mid-coronal	5 wk	8.71E-04	1.975
											Frontal	16 wk	1.75E-04	2.598
											Mid-coronal	16 wk	8.24E-04	2.405
											Frontal	21 wk	3.83E-03	2.228
											Mid-coronal	21 wk	4.82E-03	2.183
											Frontal	21 wk and salt	2.82E-04	2.858
											Mid-coronal	21 wk and salt	2.12E-02	1.846
Fto	-	-	-	-	1	-	-	-	-	-	Frontal	5 wk	3.24E-02	1.353
Inpp5d	-	-	2	-	23	-	-	-	-	-	Frontal	5 wk	2.19E-03	1.757
											Mid-coronal	5 wk	4.43E-03	1.767
Myo1b	-	-	-	-	1	-	-	-	-	-	Frontal	5 wk	8.34E-03	-1.78
Neu2	-	-	2	2	1	-	-	1	-	-	Mid-coronal	21 wk and salt	2.93E-02	1.941
Tmem63b	-	-	13	1	50	-	-	-	-	-	Frontal	16 wk	7.98E-03	1.734
											Frontal	21 wk	4.15E-02	1.868
Sgca	-	-	1	-	13	-	-	-	-	-	Mid-coronal	5 wk	3.09E-02	1.537

(wk) short for weeks of age and (salt) short for salt loaded.

5.3.2 Prioritising human candidate genes

5.3.2.1 Translation of candidate genes/miRNA from human to rat and vice versa

5.3.2.1.1 Human GWAS genes for blood pressure and hypertension integrated with $SHRSP_{Gla}$ and WKY_{Gla} transcriptomic and genomic data

Subsequently, we attempted to translate human GWAS candidate genes for blood pressure and hypertension to our rat strains. We assessed whether a total of 72 genes map to rat genome, blood pressure QTLs, harbour functional small sequence variants and if they are significantly differentially expressed between $SHRSP_{Gla}$ and WKY_{Gla} in kidney and heart tissues (Table 5-7).

Table 5-7 Human GWAS of blood pressure and hypertension candidate genes integrated with $SHRSP_{Gla}$ and WKY_{Gla} transcriptomic and genomic data.

Human	Rat		WKY_{Gla} versus $SHRSP_{Gla}$											
GWAS Gene	ortholog	RGD Blood pressure QTL	QTL	Small sequence variants									Gene expression	
				upstream	downstream	intronic	3' UTR	5' UTR	splice site	essential splice site	synonymous	non-synonymous	Kidney	Heart
MTHFR[2, 4, 5]	✓	BP 7,103,131,139,147,155,210,292	-	-	-	-	-	-	-	-	-	-	-	-
NPPA[2, 4, 6]	✓	BP 7,103,131,139,147,155,210,292	-	-	-	-	-	-	-	-	-	-	-	-
NPPB[2, 4, 7]	✓	BP 7,103,131,139,147,155,210,292	-	-	5	-	-	-	-	-	-	-	-	-
TMEM133[4]	×	-	-	-	-	-	-	-	-	-	-	-	×	×
FLJ32810[4]	✓	-	-	-	-	-	-	-	-	-	-	-	-	-
ST7L[6]	✓	BP 10,14,16,50,63,101,105,132,158	Blood pressure (Chromosome 2)	-	-	-	-	-	-	-	-	-	-	-
CAPZA1[6]	✓	BP 10,14,16,50,63,101,105,132,158	Blood pressure (Chromosome 2)	-	-	-	-	-	-	-	-	-	-	-
FIGN[6, 7]	✓	BP 37,118,122,177,251,264	-	1	6	224	1	-	-	-	-	-	No probe	-
GRB14[6]	✓	BP 37,118,122,177,251,264	-	1	6	157	1	-	1	-	1	-	-	-
GUCY1B3[4]	✓	BP 14,73,101,105,132,163,164,165	Blood pressure (Chromosome 2)	4	8	74	-	-	-	-	-	-	-	5wk, 16wk
GUCY1A3[4]	✓	BP 14,73,101,105,132,163,164,165	Blood pressure (Chromosome 2)	1	5	44	3	1	-	-	-	-	21wk	Neo, 5wk, 16wk
NPR3[4, 5, 6]	✓	BP 14,36,99,132,174,206,240,243	Blood pressure (Chromosome 2)	-	-	-	-	-	-	-	-	-	-	-
BAT5 (ABHD16A)[4]	✓	BP 195	-	8	2	9	-	-	-	-	1	-	-	Neo, 5wk
BAT2 (PRRC2A)[4]	✓	BP 195	-	2	-	1	-	-	1	-	1	-	-	-
CYP17A1[1, 2, 4, 6]	✓	BP 255	-	-	-	-	-	-	-	-	-	-	-	-
NT5C2[4, 6]	✓	BP 255	-	-	-	-	-	-	-	-	-	-	-	No probe
TBX3[4, 6]	✓	BP 83,188,218,268,269	-	-	-	-	-	-	-	-	-	-	-	-
TBX5[1, 4]	✓	BP 83,188,218,268,269	-	-	-	1	-	-	-	-	-	-	-	-
ULK3[1, 4]	✓	BP 35,39,62,217,253,262,273,316	-	-	1	1	-	-	-	-	-	-	No probe	-
CYP1A1[4]	✓	BP 35,39,62,217,253,262,273,316	-	2	3	-	-	-	-	-	-	-	-	-
GNAS[4]	✓	BP 81,167,285,320	-	-	-	-	-	-	-	-	-	-	-	-
EDN3[4]	✓	BP 81,167,285,320	-	-	-	-	-	-	-	-	-	-	No probe	-
FURIN[4]	✓	BP 58,84,89,94,97,107,117,138,154	-	-	-	-	-	-	-	-	-	-	-	-
FES[4]	✓	BP 58,84,89,94,97,107,117,138,154	-	-	-	-	-	-	-	-	-	-	21wk	-
C10orf107[2, 4]	✓	-	-	-	2	1	-	-	-	-	-	-	-	-

Continued

Human	Rat		WKY_{Gla} versus $SHRSP_{Gla}$											
GWAS Gene	ortholog	RGD Blood pressure QTL	QTL	Small sequence variants									Gene expression	
				upstream	downstream	intronic	3' UTR	5' UTR	splice site	essential splice site	synonymous	non-synonymous	Kidney	Heart
ALDH2[6]	✓	BP 83,188,218,268,269	-	-	-	-	-	-	-	-	-	-	-	-
CHIC2[7]	✓	BP 189	Left ventricular mass index	-	-	-	-	-	-	-	-	-	-	-
PLCD3[2]	✓	BP 1,9,72,82,91,134,168,249,311	-	-	-	-	-	-	-	-	-	-	-	-
GOSR2[4]	✓	BP 1,9,72,82,91,134,168,249,311	-	-	-	-	-	-	-	-	-	-	-	-
ATP2B1[1, 4, 5, 6]	✓	BP 102,182	-	-	1	-	-	-	-	-	-	-	-	-
ADAMTS8[7]	✓	BP 35,262,271,315	-	10	8	17	-	-	1	-	1	2	-	-
SLC4A7[4]	✓	BP 191	-	-	5	68	-	-	1	-	5	-	-	-
JAG1[4]	✓	BP 81,123,207,285	-	-	1	4	-	-	-	-	2	-	-	-
NOV[7]	✓	BP 38,181,214,265,266	-	-	-	-	-	-	-	-	-	-	-	-
NOS3[5, 8]	✓	BP 79,178,330	-	-	-	1	-	-	-	-	-	-	-	-
PIK3CG[7]	✓	BP 141,325	-	-	7	54	2	-	-	-	1	-	No probe	-
RSPO3[9]	✓	BP 58,96,145,360	-	-	-	1	-	-	-	-	-	-	-	-
SLC39A8[4]	✓	BP 50,175,239,337	-	-	-	2	-	-	-	-	-	-	21wk	-
MECOM[2, 4]	✓	BP 6,14,90,132,163,206,239,243	Blood pressure (Chromosome 2)	15	2	74	1	-	-	-	1	-	-	-
MOV10[4]	✓	BP 10,14,16,50,63,101,105,132,158	Blood pressure (Chromosome 2)	-	-	-	-	-	-	-	-	-	-	-
CASZ1[6]	✓	BP 7,103,131,139,147,155,210,292	-	-	-	-	-	-	-	-	-	-	-	-
PLCE1[4]	✓	BP 42,255,259,306,362	-	-	-	-	-	-	-	-	-	-	-	-
UMOD[3]	✓	BP 26,28,30,44,78,84,88,89,107,154	-	-	-	-	-	-	-	-	-	-	-	-
PLEKHG1[9]	✓	BP 58,96,145,360	-	6	5	21	-	-	-	-	-	1	No probe	No probe
CYP1A2[2]	✓	BP 35,39,62,217,253,262,273,316	-	-	-	6	-	-	2	-	2	1	-	-
HFE[4, 5]	✓	BP 8,98,192,242,245,247,278,343	-	-	-	-	-	-	-	-	-	-	-	-
ZNF652[2, 4]	✓	BP 1,9,12,71,72,82,91,134,148,168	-	-	-	-	-	-	-	-	-	-	-	-
ADRB1[7]	✓	-	-	2	2	-	-	-	-	-	-	-	-	-
CACNB2[1, 4]	✓	BP 369	-	-	-	-	-	-	-	-	-	-	-	Neonates
AGT[5]	✓	BP 32	-	-	-	-	-	-	-	-	-	-	-	-

Continued

Human	Rat		WKY_{Gla} versus $SHRSP_{Gla}$											
GWAS Gene	ortholog	RGD Blood pressure QTL	QTL	Small sequence variants									Gene expression	
				upstream	downstream	intronic	3' UTR	5' UTR	splice site	essential splice site	synonymous	non-synonymous	Kidney	Heart
ULK4[1, 4, 9]	✓	BP 252,263	-	-	3	8	-	-	-	-	-	-	-	-
CSK[1, 2]	✓	BP 35,39,62,217,253,262,273,316	-	-	1	1	-	-	-	-	-	-	-	-
CNNM2[6]	✓	BP 255	-	12	5	78	-	-	-	-	-	-	-	-
LSP1[5]	✓	BP 26,30,89,143,154,196,197,246	-	-	-	-	-	-	-	-	-	-	21wk	No probe
TNNT3[5]	✓	BP 26,30,89,143,154,196,197,246	-	-	-	-	-	-	-	-	-	-	-	-
SOX6[5, 9]	✓	BP 26,28,30,44,78,84,88,89,107,154	-	-	-	1	-	-	-	-	-	-	-	-
MAP4[7]	✓	BP 125,184,252,263,331	-	-	-	-	-	-	-	-	-	-	21 wk	-
FGF5[2, 4, 6]	✓	-	Left ventricular mass index	1	6	12	-	-	-	-	1	-	-	-
PLEKHA7[1, 4]	✓	BP 26,28,30,44,78,84,88,89,107,154	-	-	-	-	-	-	-	-	-	-	No probe	No probe
RPL6[6]	✓	BP 83,188,218,268,269	-	-	-	-	-	-	-	-	-	-	-	-
ADM[4]	✓	BP 26,28,30,44,77,84,88,89,107,138	-	-	-	-	-	-	-	-	-	-	-	-
SH2B3[1, 2, 4]	✓	BP 83,188,218,268,269	-	-	-	-	-	-	-	-	-	-	-	-
EBF1[4]	✓	BP 857,82,133,168	-	-	-	51	-	-	-	-	-	-	-	-
PTPN11[6]	✓	BP 83,188,218,268,269	-	-	-	-	-	-	-	-	-	-	-	-
EVX1[9]	✓	BP 86,124,135,146,330,338	-	-	-	-	-	-	-	-	-	-	No probe	-
ENPEP[6]	✓	BP 50,175,202,203,204,239,275,337	-	-	-	-	-	-	-	-	-	-	21wk	
HOXA1[9]	✓	BP 86,124,135,146,330,338	-	-	-	-	-	-	-	-	-	-	-	-
HOXA2[9]	✓	BP 86,124,135,146,330,338	-	-	-	-	-	-	-	-	-	-	-	-
HOXA3[9]	✓	BP 86,124,135,146,330,338	-	-	-	-	-	-	-	-	-	-	No probe	No probe
HOXA4[9]	✓	BP 86,124,135,146,330,338	-	4	12	-	-	-	-	1	1	-	-	No probe
HOXA5[9]	✓	BP 86,124,135,146,330,338	-	1	1	1	-	-	-	-	-	-	-	No probe
HOXA13[9]	✓	BP 86,124,135,146,330,338	-	-	3	-	-	-	-	-	-	-	No probe	-

Symbol (✓) indicates genes/probes map to the genome, symbol (×) indicates gene and/or probe does not map to the genome, (wk) short for weeks of age and (neo) short for Neonates. Human GWAS genes for blood pressure and hypertension obtained from a number of studies (Levy et al., 2009[1]; Newton-Cheh et al., 2009[2]; Padmanabhan et al., 2010[3]; Ehret et al., 2011[4]; Johnson et al., 2011[5]; Kato et al., 2011[6]; Wain et al., 2011[7]; Salvi et al., 2012[8]; Franceschini et al., 2013[9]).

5.3.2.1.1.1 qRT-PCR analysis of genes common between integrating human GWAS genes blood pressure and hypertension $SHRSP_{Gla}$ and WKY_{Gla} transcriptomic and genomic data

Human GWAS genes for blood pressure and hypertension *Gucy1a3* and *Gucy1b3* were prioritised for qRT-PCR validation. *Gucy1a3* and *Gucy1b3* mapped to the rat genome and to the QTL for blood pressure in $SHRSP_{Gla}$ and WKY_{Gla} strains while harbouring small genomic variants and differentially expressing in relevant tissues to hypertension. qRT-PCR of *Gucy1a3* and *Gucy1b3* showed no significant changes in mRNA level between $SHRSP_{Gla}$ and WKY_{Gla} in heart at 5 and 16 weeks of age (Figure 5-11).

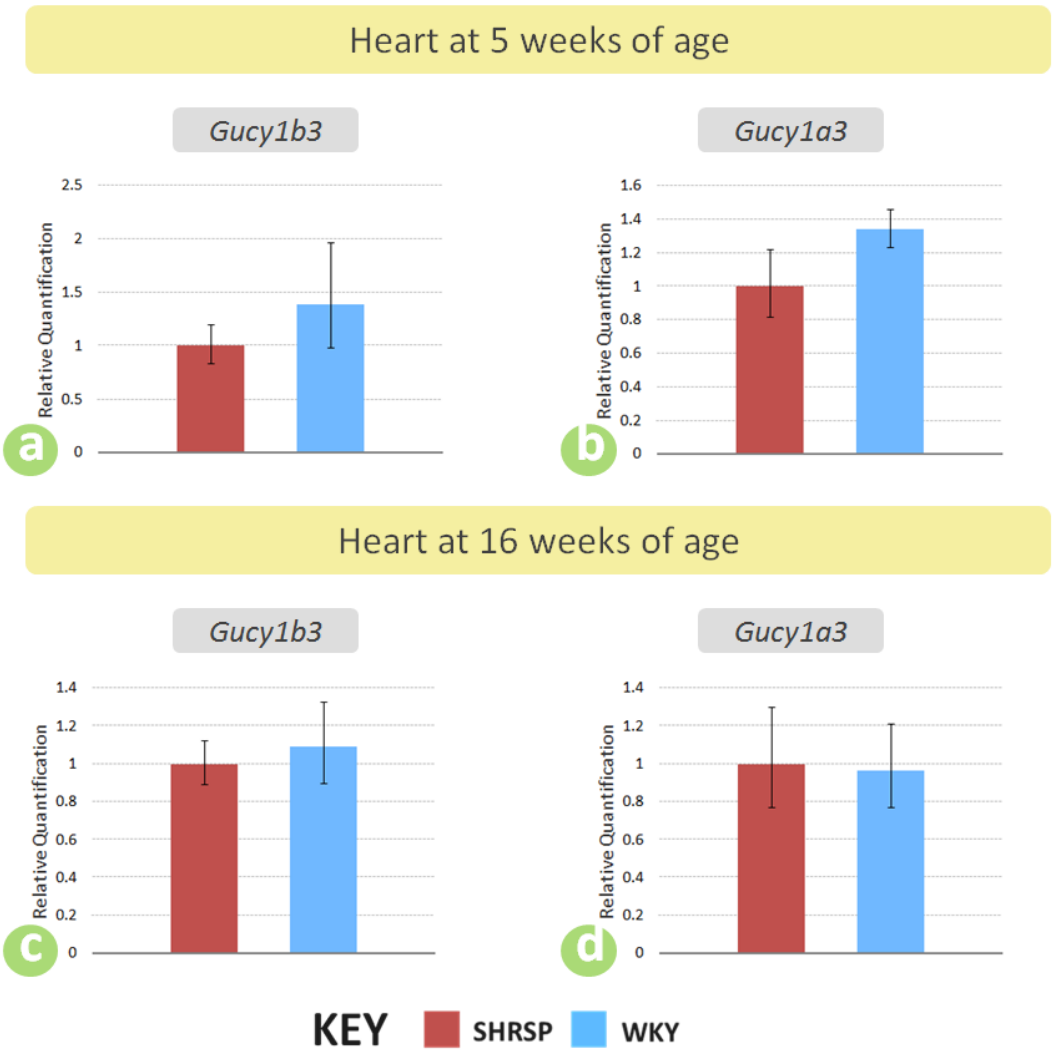


Figure 5-11 qRT-PCR of translated candidate genes for blood pressure from human GWAS to our rat strains.
At 5 weeks of age the mRNA expression level of **a**) *Gucy1b3* **b**) *Gucy1a3* in heart shows no significant change between $SHRSP_{Gla}$ and WKY_{Gla}. At 16 weeks of age the mRNA expression level of **c**) *Gucy1b3* **d**) *Gucy1a3* in heart shows no significant change between $SHRSP_{Gla}$ and WKY_{Gla}. *P<0.05.

5.3.2.1.1.2 Troubleshooting Gucy1a3 and Gucy1b3 experimental data

To investigate why the experimental data of qRT-PCR of *Gucy1a3* and *Gucy1b3* did not match with the microarray results between $SHRSP_{Gla}$ and WKY_{Gla} in heart at 5 and 16 weeks of age, we followed the subsequent strategy. We identified variants of $SHRSP_{Gla}$ and WKY_{Gla} within significant Affymetrix and Illumina probes' targeting regions as they could lead to false results (Table 5-8). *Gucy1a3* significant probes; in kidney only at 21 weeks of age with Affymetrix and in heart at all time points with Illumina. Both of these significant probes' target region, from two platforms, are affected by harboured sequence variants; however, each was affected by a different SNP. *Gucy1b3* significant probe in heart at 5 and 16 weeks of age with Illumina platform was not affected by variants within its targeted region.

Table 5-8 Troubleshooting *Gucy1a3* and *Gucy1b3* microarray probes' target region

WKY_{Gla} versus $SHRSP_{Gla}$							
Gene	Position	Significant Probe ID	Tissue	Time point	FDR p-value	Fold change	Variants within probes' target region
Gucy1a3	chr2:173,756,824-173,818,316	1387079_at	Kidney	21 weeks of age	2.50E-02	1.43751	SNP - chr2:173,757,246
		ILMN_1371357	Heart	Neonates	1.92E-21	-5.40889	SNP - chr2:173,755,198
				5 weeks of age	7.44E-21	-4.63913	
				16 weeks of age	1.90E-20	-4.43004	
Gucy1b3	chr2:173,683,153-173,735,298	ILMN_1371639	Heart	5 weeks of age	1.90E-07	1.60285	-
				16 weeks of age	0.0002	1.35309	

In addition, to further investigate why the experimental data of qRT-PCR of *Gucy1a3* and *Gucy1b3* did not match the microarray results between $SHRSP_{Gla}$ and WKY_{Gla} in heart at 5 and 16 weeks of age, we looked for evidence of alternate splicing of *Gucy1a3* and *Gucy1b3* using rat Affymetrix Exon ST array in heart at 5 and 16 weeks of age (previously generated data by our group) (Figure 5-12). *Gucy1a3* microarray transcript cluster probe ID (7210088) signal intensity, which is a relative measure of expression of the transcript is significantly higher level in $SHRSP_{Gla}$ than WKY_{Gla} at the last exon (exon 9)/3' UTR in heart at 5 and 16 weeks of age (Figure 5-12A1-A2). Taqman gene expression probe (Rn00367252) for *Gucy1a3* gene span over exons 6 and 7 (Figure 5-12A3) which may not detect the alternative splicing at exon 9/3' UTR between $SHRSP_{Gla}$ and WKY_{Gla} in heart at 5 and 16 weeks of age. There is no evidence of alternative splicing for *Gucy1b3* signal intensity between $SHRSP_{Gla}$ than WKY_{Gla} in heart at 5 and 16 weeks of age (Figure 5-12B1-B2). However, overall probes signal intensity of *Gucy1b3* gene is higher in the WKY_{Gla} than $SHRSP_{Gla}$ in heart at 5 and 16 weeks of age. *Gucy1b3* Taqman gene expression probe (Rn00562775) map to span over

exons 3 and 4 which has the little change of signal intensity between $SHRSP_{Gla}$ than WKY_{Gla} in heart at both time point (Figure 5-12B3)

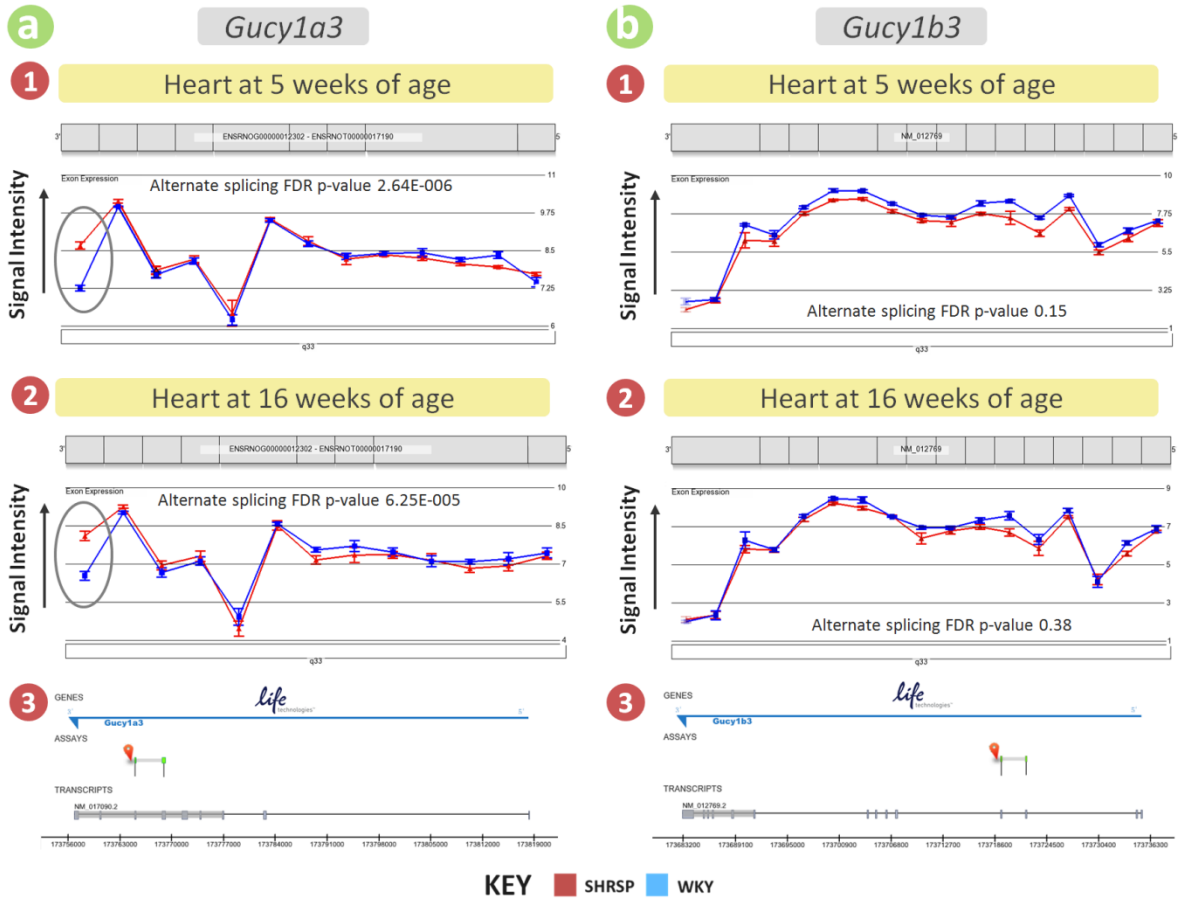


Figure 5-12 Troubleshooting *Gucy1a3* and *Gucy1b3* using rat Affymetrix Exon ST array.
a) *Gucy1a3* signal intensity from the exon array between $SHRSP_{Gla}$ and WKY_{Gla} in heart shows evidence of alternative splicing at **a1**) 5 weeks of age and **a2**) 16 weeks of age in which **a3**) *Gucy1a3* Taqman gene expression probe does not span over the affected exon. **b**) *Gucy1b3* signal intensity from the exon array between $SHRSP_{Gla}$ and WKY_{Gla} in heart shows no evidence of alternative splicing but an overall change at **b1**) 5 weeks of age and **b2**) 16 weeks of age in which **b3**) *Gucy1b3* Taqman gene expression probe spans over the least signal intensity difference between $SHRSP_{Gla}$ and WKY_{Gla} at exon 3 and 4.

5.3.2.1.2 Integrating renal miRNA sequencing data from human and rat to prioritise hypertension candidate miRNA

In order to translate differentially expressed renal miRNA between human to our rat, we integrated miRNA NGS renal data from normotensive and hypertensive subjects. The samples were derived from medulla and cortex sections and a whole kidney from WKY_{Gla}, chromosome 2 congenic (SP.WKY_{Gla}2c*, we call it 2c* for short) and $SHRSP_{Gla}$ rat strains. Our data suggested that there was no common differential expression of miRNA between normotensive and hypertensive from human and rat analysis (Table 5-9-Table 5-13).

In human study, normotensive and hypertensive subjects have no common differentially expressed miRNA between renal sections; medulla and cortex (Table 5-9, Table 5-10). In our rat study, there was common differentially expressed miRNA in whole kidney between WKY_{Gla} versus $SHRSP_{Gla}$, 2c* versus $SHRSP_{Gla}$, and WKY_{Gla} versus 2c* strains (Table 5-11- Table 5-13).

We searched for sequence variants within miRNAs, and then investigated miRNAs mapping to our congenic region to prioritise candidate miRNA for hypertension in our rat strains. In all rat comparisons, all miRNAs did not harbour sequence variants and none of them mapped to chromosome 2 (Table 5-11, Table 5-12, Table 5-13).

We then investigated whether the identified differentially expressed miRNA targets positional genes mapped to chromosome 2. Significantly differentially expressed miRNAs between 2c* vers $SHRSP_{Gla}$ were prioritised by integrating miRanda (Betel et al., 2008) predicted target genes mapped to chromosome 2 congenic region (422 genes); these included genes with 3' UTR harbour sequence variants unique to $SHRSP_{Gla}$ (928 genes), and renal gene expression at 16 and 21 weeks of age of $SHRSP_{Gla}$ and WKY_{Gla} (Table 5-14). 14 targeted positional candidate genes by differentially expressed miRNA between 2c* and $SHRSP_{Gla}$ in which their 3 'UTR region harbour variants unique to $SHRSP_{Gla}$. Three of the 14 positional genes have variants on or close to the predicted miRNA binding site and 3 genes implicated in renal gene expression data. *Gucy1a3* and *Gstm1* 3' UTR along with 2c* versus $SHRSP_{Gla}$ significant renal miRNA binding sites and sequence variants from $SHRSP_{Gla}$ and WKY_{Gla} is showing in Figure 5-13 and Figure 5-14.

Table 5-9 Differentially expressed miRNA in human renal medulla between normotensive and hypertensive subjects.

Human		Normotensive versus hypertensive			
miRNA	Map to chromosome	Normotensive Mean	Hypertensive Mean	Fold Change	Adjusted p-value
hsa-miR-129-2-3p	11	19	249	3.72	0.0147

The fold change value is in $\log_2$ fold change.

Table 5-10 Differentially expressed in human renal cortex between normotensive and hypertensive subjects.

Human		Normotensive versus hypertensive			
miRNA	Map to chromosome	Normotensive Mean	Hypertensive Mean	Fold Change	Adjusted p-value
hsa-miR-217	2	1243.97	150.06	-3.06	0.027
hsa-miR-216a	2	1157.08	126.47	-3.20	0.027

The fold change value is in $\log_2$ fold change.

Table 5-11 Differentially expressed miRNA in rat kidney between $SHRSP_{Gla}$ and WKY_{Gla}.

Rat		WKY_{Gla} versus $SHRSP_{Gla}$				
miRNA	Map to chromosome	WKY Mean	SHRSP Mean	Fold Change	Adjusted p-value	Variants
rno-miR-3596b	17	481.52	31.57	-3.94	4.00E-09	-
rno-miR-423-5p	10	4096.66	760.78	-2.43	1.81E-05	-
rno-miR-219-2-3p	3	1.27	104.18	6.362	1.81E-05	-
rno-miR-7a-5p	1, 17	926.28	178.36	-2.38	0.00013	-

The fold change value is in $\log_2$ fold change.

Table 5-12 Differentially expressed miRNA in rat kidney between 2c* and $SHRSP_{Gla}$.

Rat		2c* versus $SHRSP_{Gla}$				WKY_{Gla} versus $SHRSP_{Gla}$
miRNA	Map to chromosome	2c* Mean	SHRSP Mean	Fold Change	Adjusted p-value	Variants
rno-miR-7a-5p	1, 17	1495.79	217.27	-2.9	3.34E-17	-
rno-miR-423-5p	10	5184.27	926.77	-2.5	1.39E-16	-
rno-miR-221-5p	X	707.33	172.09	-2.1	2.71E-08	-
rno-miR-148b-3p	7	74026.63	25663.79	-1.6	7.64E-07	-
rno-miR-322-3p	X	383.83	104.79	-1.9	1.14E-05	-
rno-miR-672-5p	X	275.65	107.68	-1.36	0.018739	-
rno-miR-30c-2-3p	5, 9	3622.96	1846.80	-0.98	0.021465	-

The fold change value is in $\log_2$ fold change.

Table 5-13 Differentially expressed miRNA in rat kidney between WKY_{Gla} and 2c*

Rat		WKY_{Gla} versus 2c*				WKY_{Gla} versus $SHRSP_{Gla}$
miRNA	Map to chromosome	WKY Mean	2c* Mean	Fold Change	Adjusted p-value	Variants
rno-miR-3596b	17	465.41	22.87	-4.35	1.35E-15	-
rno-miR-219-2-3p	3	1.23	113.50	6.54	5.95E-07	-

The fold change value is in $\log_2$ fold change.

Table 5-14 Positional targeted genes of differentially expressed miRNA in rat kidney between 2c* and $SHRSP_{Gla}$.

Rat		2c* versus $SHRSP_{Gla}$							WKY_{Gla} versus $SHRSP_{Gla}$		
Targeted genes	Size of 3' UTR	miRNA							Variants on/close to miRNA binding site	Kidney Gene expression	
		rno-mir-7a-5p	rno-mir-423-5p	rno-mir-221-5p	rno-mir-148b-3p	rno-mir-322-3p	rno-mir-672-5p	rno-mir-30c-2-3p		16 weeks	21 weeks
Gstm1	362	-	-	-	✓	-	✓	-	-	✓	✓
Snx27	3498	✓	-	-	✓	✓	✓	✓	-	✓	-
Gucy1a3	2236	✓	-	-	-	✓	✓	✓	rno-mir-322-3p rno-mir-30c-2-3p rno-mir-7a-5p	-	✓
Txnip	1365	✓	-	-	✓	✓	-	✓	-	-	-
Slc16a1	1608	-	✓	✓	✓	✓	-	-	-	-	-
Casq2	1062	-	-	-	✓	-	✓	✓	rno-mir-672-5p	-	-
Mllt11	1054	✓	-	✓	-	✓	✓	✓	-	-	-
Polr3c	403	-	-	-	-	✓	-	-	-	-	-
Aph1a	1101	-	-	-	-	-	-	✓	-	-	-
Lysmd1	883	✓	-	-	✓	✓	-	✓	-	-	-
Chrnb2	506	-	-	-	-	-	-	✓	rno-mir-30c-2-3p	-	-
Tdo2	427	-	-	✓	-	-	-	-	-	-	-
Tspan2	3143	✓	-	✓	✓	✓	✓	✓	-	-	-
Pias3	960	-	-	✓	✓	✓	-	✓	-	-	-

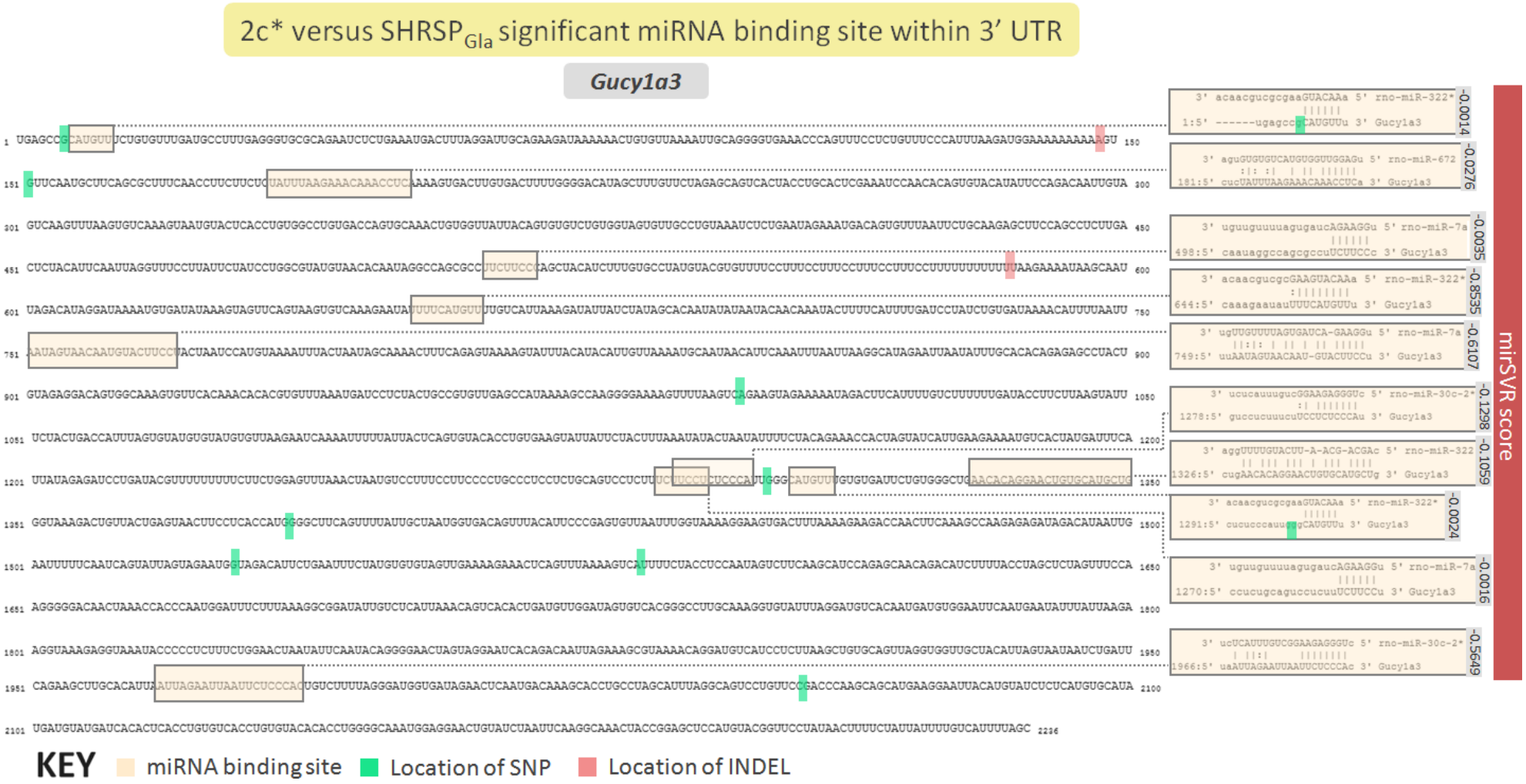


Figure 5-13 Predicted target sequence for miRNA binding site and small sequence variants between SHRSP$_{Gla}$ and WKY$_{Gla}$ within the 3' UTR of *Gucy1a3* mRNA. miRanda predicted binding sites are scored for likelihood of mRNA downregulation using mirSVR for differentially expressed miRNA in kidney between 2c* and SHRSP$_{Gla}$ within *Gucy1a3* 3'UTR along with sequence variants between SHRSP$_{Gla}$ and WKY$_{Gla}$

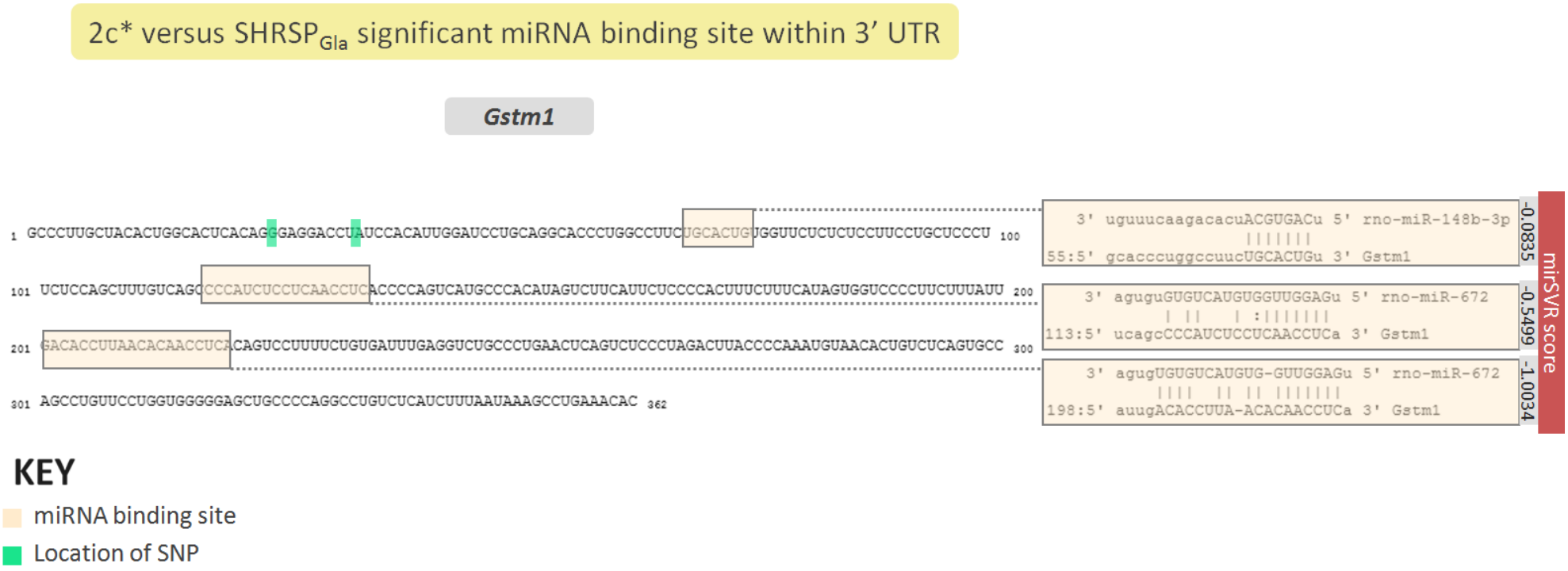


Figure 5-14 Predicted target sequence for miRNA binding site and small sequence variants between SHRSP$_{Gla}$ and WKY$_{Gla}$ within the 3' UTR of *Gstm1* mRNA.
miRanda predicted binding sites are scored for likelihood of mRNA downregulation using mirSVR for differentially expressed miRNA in kidney between 2c* and SHRSP$_{Gla}$ within *Gstm1* 3'UTR along with sequence variants between SHRSP$_{Gla}$ and WKY$_{Gla}$.

5.3.2.1.3 Integrating heart gene expression data from human and rat to prioritise cardiac candidate genes

In order to translate cardiac candidate genes between human heart failure to our rat strains, we integrated differentially expressed genes from $SHRSP_{Gla}$ and WKY_{Gla} whole heart across 3 time points with differently expressed genes of two human heart failure data (Schinke et al., 2004; Greco et al., 2012).from left ventricular of the heart (Figure 5-15).

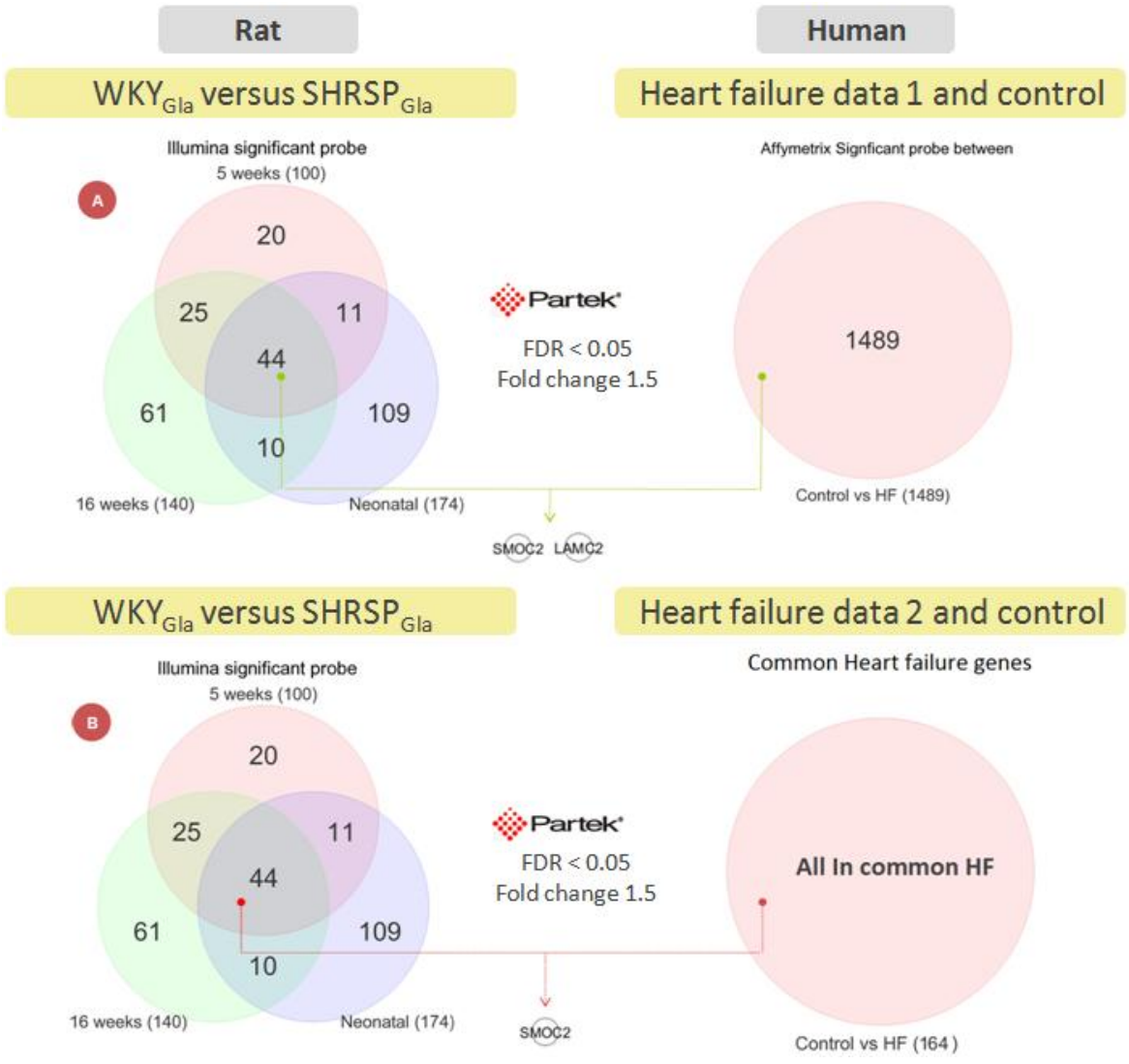


Figure 5-15 Translation of cardiac candidate genes from rat to human.
A) Cardiac $SHRSP_{Gla}$ and WKY_{Gla} common differentially expressed genes in all time points overlapped with human heart failure 1 (Greco et al., 2012) resulted in common genes; *Smoc2* and *Lamc2* in which the fold change direction is the same in rat and human. **B**) Cardiac $SHRSP_{Gla}$ and WKY_{Gla} common differentially expressed genes in all time points overlapped with human heart failure 2 (Schinke et al., 2004) in which only *Smoc2* was in-common and the fold change direction is the same in rat and human. Human and rat data analysis were conducted on Partek® Genomic Suite using a cut-off of FDR <0.05 and fold change of 1.5.

5.3.2.1.3.1 Genomic variants of the translated cardiac candidate genes in human and rat.

Functional sequence variants of common genes between human and rat heart data were identified and isolated for *Smoc2* and *Lamc2* from $SHRSP_{Gla}$ and WKY_{Gla} NGS variants compared to BN RGCS 3.4 (Table 5-15 and Table 5-16).

Table 5-15 SPARC related modular calcium binding 2 (*Smoc2*) $SHRSP_{Gla}$ and WKY_{Gla} variants compared to BN assembly RGSC 3.4.

SNPs		INDELs	
Downstream	17	Downstream	1
Intronic	177	Intronic	33
Upstream	1	Upstream	4
3Prime UTR	7		
Synonymous coding	2		

Table 5-16 Laminin, gamma 2 (*Lamc2*) $SHRSP_{Gla}$ and WKY_{Gla} variants compared to BN assembly RGSC 3.4.

SNPs		INDELs	
Downstream	20	Downstream	3
Intronic	161	Intronic	24
Upstream	13	Upstream	2
Synonymous coding	6		

5.3.2.1.4.1 qRT-PCR of the translated cardiac candidate genes in human and rat.

Subsequently, we decided to further validate by qRT-PCR the potential cardiac candidate genes from human heart failure and our rats, *Smoc2* and *Lamc2*. A collaborator at Paul Barton's laboratory in Imperial College carried out these experiments.

Samples were obtained from different dilated cardiomyopathy (DCM) heart failure stages; first 'stable' end-stage heart failure DCM (HSC), second 'unstable' end-stage HF – Left Ventricular Assist Device (IMP) that either on a recovery which had the device explanted (IMP-recovery) or non-recovery (IMP-non-recovery) and donor left ventricular controls (DVL). qRT-PCR of *Smoc2* and *Lamc2* were conducted on $SHRSP_{Gla}$ and WKY_{Gla} whole heart at neonates, 5and 16 weeks of age. Figure 5-16 demonstrates that the qRT-PCR validation analysis of rat to human translated cardiac genes revealed that *Smoc2* is differentially expressed in human and rat hearts, whereas *Lamc2* only significant in rats.

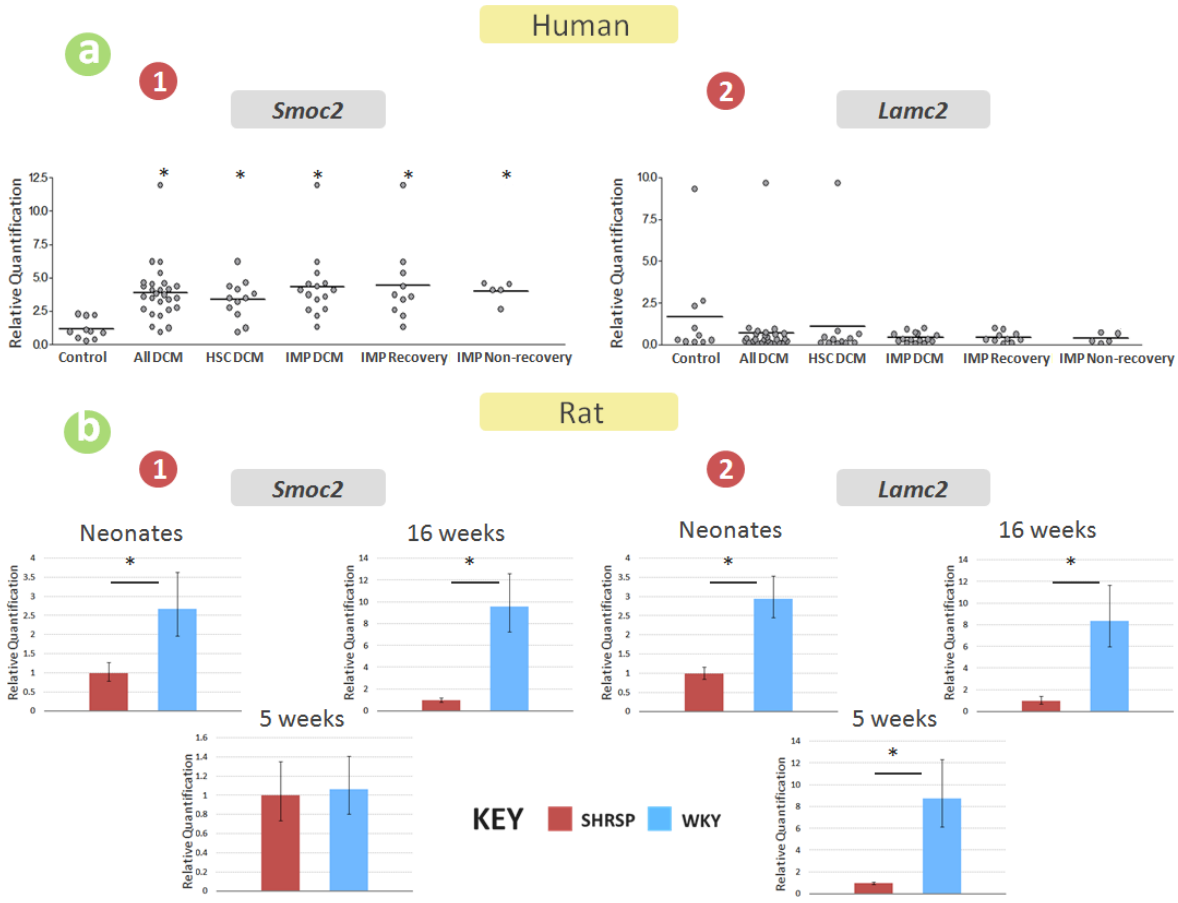


Figure 5-16 qRT-PCR analysis of translated cardiac candidate genes in rat and human. **a1**) *Smoc2* was significantly differentially expressed between all the stages of dilated cardiomyopathy heart failure and control. **a2**) *Lamc2* was not significant between all stages of dilated cardiomyopathy heart failure and control. **b1**) *Smoc2* was significantly differentially expressed in heart between $SHRSP_{Gla}$ and WKY_{Gla} at neonates and 16 weeks of age but not 5 weeks of age. **b2**) *Lamc2* was significantly differentially expressed in heart between $SHRSP_{Gla}$ and WKY_{Gla} at all time points. *P<0.05.

5.3.2.1.4.2 Troubleshooting Smoc2 experimental data in rat

Smoc2 qRT-PCR results of rat and human were significantly differentially expressed, however, rat fold change directions in microarray (Figure 5-17) is different than qRT-pCR (Figure 5-16). Figure 5-17 also shows the difference in signal intensity between $SHRSP_{Gla}$ and WKY_{Gla} significant but also very subtle.

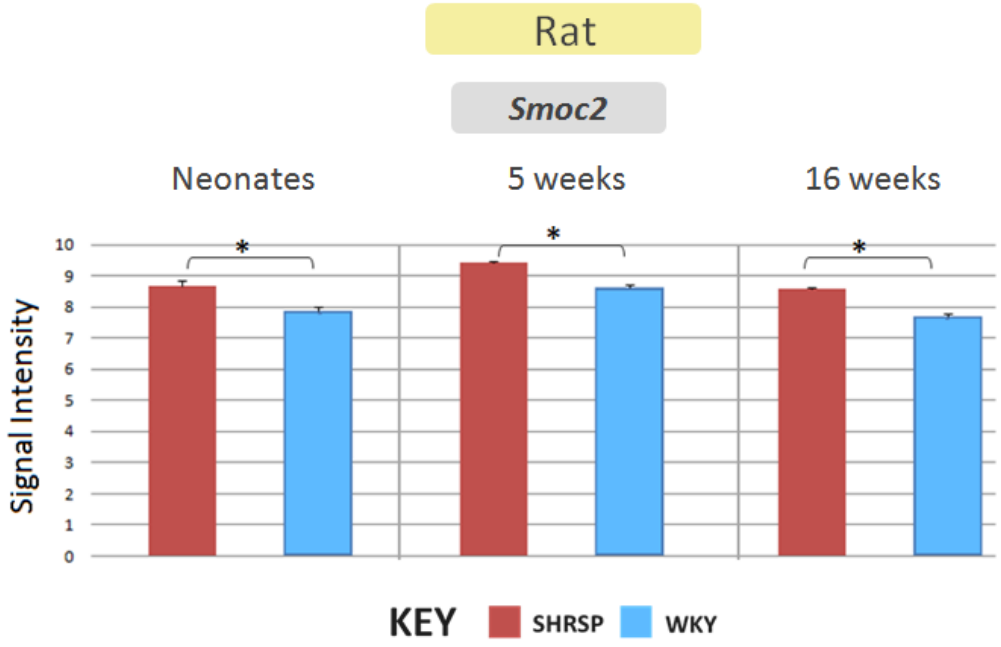


Figure 5-17 Microarray signal intensities of *Smoc2* on $SHRSP_{Gla}$ and WKY_{Gla} rat strains at neonates, 5 and 16 weeks of age.
Graph showing low microarray signal intensities of *Smoc2* between $SHRSP_{Gla}$ and WKY_{Gla} whole heart at neonates, 5 and 16 weeks of age from Illumina BeadChip array. *Smoc2* was significantly differentially expressed in heart between $SHRSP_{Gla}$ and WKY_{Gla} at all time points.*P<0.05.

5.4 Discussion

Herein we have investigated the types of genomic variants that are likely to have an influence on the transcriptomic output between $SHRSP_{Gla}$ and WKY_{Gla}. The overall results show that all types of genomics variations can correlate with transcriptomic data; however, CNV of duplication unique to the $SHRSP_{Gla}$ have the largest portion of differentially expressed renal and cardiac genes compared to small sequence variation (SNPs and INDELs) and large deletion. These data are further supported by observations of Simonis et al., 2012 when different types of NGS genomic variants integrated with NGS RNA sequencing (RNA-Seq) on liver between $SHR_{OlaIpcv}$ and a BN derived strain with polydactyly-luxate $BN_{Lx/Cub}$ that showed gene duplications unique to $SHR_{OlaIpcv}$ which were highly correlating with transcriptomic output compared to other types of genomic variants.

Small sequence variations in non-coding regions are highly correlating with the overall transcriptomic output at congenic strains regions on chromosome 2, 3 and 14 when compared to other types of genomic variants. Furthermore, at the whole genome level, small sequence variations in non coding regions, which is an important part that can regulate gene expression, is more likely to influence cardiac and renal gene expression compared to small sequence variations in coding region. These results may change and/or improve once the recent findings of Encyclopedia of DNA Elements (ENCODE) project which shows 80% of the human genome have biochemical functions and annotate regulatory elements for gene expression (Dunham et al., 2012) identified in the rat.

In our analysis we observed an overall change in results between Affymetrix custom microarray unique to our strains and default Affymetrix, when comparing chromosome 2 congenic strain ($SP.WKY_{Gla}2a$) to $SHRSP_{Gla}$ gene expression in kidney at 21 weeks of age. Additional unique positional candidate genes were identified using with the custom Affymetrix analysis. Nevertheless, removing the affected probes may have left some genes' expression levels undetected, explaining, thus, the unique positional candidate genes identified by default Affymetrix analysis only. Additionally, Illumina rat chip affected probes mapped to *Gstm1* and *Ptpn13* are shown by qRT-PCR to be differentially expressed between $SHRSP_{Gla}$ and WKY_{Gla}. Our results coincide with what is reported in

Schurmann et al., 2012, that the majority of sequence variants within probes' target region had no significant effect on hybridization efficiency. As a result, relying only on the custom microarray chip unique to our rat strains put the analysis at risk of missing true positive differentially expressed genes.

Gstm1 has been previous identified as functional and positional gene for hypertension in the $SHRSP_{Gla}$ kidney (McBride et al., 2005). Nevertheless, it has been identified by our reductive and integrative a Swiss cheese model. Furthermore, UDP glucuronosyltransferase 2 family, polypeptide B7 (*Ugt2b7*) and afamin (*Afm*) both map to congenic region of chromosome 14 and they are differentially expressed between $SHRSP_{Gla}$ and WKY_{Gla} in kidney. There is no direct link for *Ugt2b7* and *Afm* to hypertension. However, *Ugt2b7* is involved in metabolic pathways for glucuronidation and conjugation that detoxify and eliminate variety of endogenous compounds such as steroid hormones (including estrogen) and xenobiotics drugs such as Carvedilo, a therapeutic agent used to treat hypertension (Cheng et al., 1998; Holthe et al., 2003; Ohno et al., 2004). *Afm* is vitamin E binding protein and also regulates the transport of vitamin E, playing an important role in lowering blood pressure as it is promoting the production of nitric oxide resulting in vasodilatation (Boshtam et al., 2002; Jerkovic et al., 2005). As a result, functional sequence variants in *Ugt2b7* and *Afm* may affect their biological functions and result in hypertension.

The 39 stroke candidate genes identified by the integration of $SHRSP_{Gla}$ unique sequence variants compared to 27 rat strains genome with frontal and mid-coronal brain gene profiling at different time points (Bailey et al., 2014) with salt loaded were not prioritised previously in Bailey et al., 2014 as stroke candidate genes. Given the number of time points of brain sections, including salt loaded, genes that were differentially expressed at more than one time point and/or brain section used to as filtering criteria in which 16 candidate genes were prioritised for stroke. Six of these genes that were mapped to RGD stroke QTLs; ribosomal protein S16 (*Rps16*), zinc finger protein 566 (*Znf566*), transcription factor 4 (*Tcf4*), sorting nexin 10 (*Snx10*), IQ motif and ubiquitin domain containing (*Iqub*), apolipoprotein C-I (*Apoc1*). Three out of the 16 candidate genes for stroke mapped to QTLs between $SHRSP_{Gla}$ and WKY_{Gla} for blood pressure; ATPase, class VI, type 11B (*Atp11b*), transcription termination

factor, RNA polymerase I (*Ttf1*) and for left ventricular mass index; predicted gene Loc360919. *Snx10* has a role in regulation of osteoclast (Zhu et al., 2012) and when mutated lead to ostepetrosis (Pangrazio et al., 2013). *Apoc1* is involved in the regulation of lipid transport and associated with Alzheimer's disease (Zhou et al., 2014) and a mutated *Ttf1* has been associated with brain-lung-thyroid syndrome (Shetty et al., 2014). *Atp11b,* which is involved in ion transport, is the only gene with 28 bp deletion causing a frameshift. Apart from these genes with biological function, the rest of the identified genes mapped to QTLs of our interests are with uncertain function. This is in line with the few identified GWAS candidate genes for stroke in human in which genes have no biological functions related to stroke (Traylor et al., 2012) and none were implicated in our analysis.

The integration of human GWAS studies for blood pressure and hypertension with our cardiac and renal transcriptomic data and sequence variants between from $SHRSP_{Gla}$ and WKY_{Gla} was used as a strategic approach to get insights into their genetics mechanisms and translate candidate genes from human to rat. We have shown most human GWAS genes for hypertension were conserved and mapped to BP QTLs in rat. Moreover, half of these genes harbour sequence variants but few are differentially expressed in heart and kidney between $SHRSP_{Gla}$ and WKY_{Gla}. *Gucy1a3-Gucy1b3* were identified when we considered our congenic regions, sequence variants, transcriptomic data between $SHRSP_{Gla}$ and WKY_{Gla} are associated with human GWAS candidate genes for hypertension. Both *Gucy1a3-Gucy1b3* serve as nitric oxide receptors (Saino et al., 2004), and both can facilitate communication between heart and kidney to regulate blood volume and Na+ balance (Garbers et al., 2006). qRT-PCR validation of the prioritised genes in heart did not match our microarray gene expression result. The identification of two significantly differentially expressed probes affected for *Gucy1a3* by two different SNPs from two microarray gene expression in heart and kidney might explain the qRT-PCR results (although we tested *Gucy1a3* in heart only), however, we have shown that having a SNP within probes' target region does not necessary lead to false positive result. Moreover, the alternative splicing of *Gucy1a3* in heart from Exon array results showed that the Taqman probe used for testing might provide a better explanation for the qRT-PCR results as it does not cover the alternative spliced exon.

The integration of renal next generation data to translate differentially expressed miRNA between hypertensive and normotensive subjects from human to rat was not successful. This might be due to functional differences in kidney sections between cortex and medulla used in human compared to the whole kidney used in rat. Other reasons that might contribute to this are; the miRBase (version 20) total number of miRNA identified in human (1872) is about the triple compared to in rat (449) (Griffiths-Jones et al., 2008) and only one replica was used for each subject. Therefore, human and rat data were analysed separately.

In human miRNA study, normotensive and hypertensive subjects have did not have common differentially expressed miRNA between renal medulla and cortex which, is consistent with their differences in anatomical boundaries and biological functions (Marques et al., 2011). None of the identified renal miRNA differentially expressed between hypertensive and control in our human study were in common with human hypertensive renal miRNA identified in Marques et al., 2011. Nevertheless, both hsa-mir-216a and -217 identified in human cortex together been associated with hepatocellular carcinoma recurrence and metastases (Xia et al., 2013) and are potential biomarkers for chronic kidney disease (Szeto et al., 2012). Furthermore, miR-217 act as an endogenous inhibitor of silent information regulator 1 promoting endothelial cell senescence which can cause endothelial dysfunction (Menghini et al., 2009). Additionally, hsa-mir-129-2-3p identified in human medulla has been suggested to regulate glycerol levels (Raitoharju et al., 2014), which can function as vasodilator to improve blood flow (Maruhashi et al., 2013).

Low number of replica did not seem to have an impact on identifying common differentially expressed miRNA WKY_{Gla} versus $SHRSP_{Gla}$, 2c* versus $SHRSP_{Gla}$ and WKY_{Gla} versus 2c* comparisons. However, none of the identified differentially expressed miRNA in all rat strains comparisons harboured sequence variants and none were mapped to chromosome 2 including 2c* versus $SHRSP_{Gla}$ comparison. Therefore, we considered miRNA targeting genes and prioritised 2c* versus $SHRSP_{Gla}$ comparison to identify positional genes with sequence variants in their 3 'UTR that may affect cardiac and/or renal genes expression between $SHRSP_{Gla}$ and WKY_{Gla}. Sorting nexin family member 27 (*Snx27*) has role in glucose and ion transport (Steinberg et al., 2013) and it is targeted by most of the miRNA identified which might be due its large 3' UTR size. The difference in mRNA level

of *Gstm1* between $SHRSP_{Gla}$ and WKY_{Gla} was reflected at the protein level (McBride et al., 2005), however, this change was not associated with miRNAs (rno-mir148b-3p and rno-mir-672) that targets small size 3' UTR. The identification of sequence variants next to the binding site of rno-mir-7a-5p, rno-mir-322-3p and rno-mir-30c-2-3p that target *Gucy1a3* 3'UTR might increase the chance of affect their binding affinity (Preskill & Weidhaas, 2013) and have an impact on mRNA degradation and translation (Valencia-Sanchez et al., 2006). The current miRNA study needs further work on laboratory validation of implicated miRNAs, mRNAs, sequence variants and proteins.

When considering translation of cardiac differentially expressed genes common between our rat strains and two human heart failure data. We found *Smoc2* and *Lamc2* were common between our rat strains and human ischemic heart failure data 1 (Greco et al., 2012) where fold change direction is the same in rat and human. *Smoc2* was the only common gene between our rat strains and human heart failure data 2 (Schinke et al., 2004), which arose from different aetiologies of cardiomyopathies. Sequence variants from $SHRSP_{Gla}$ and WKY_{Gla} were implicated in both genes that may affect their function. When validating using qRT-PCR with human heart failure samples from Imperial College, *Smoc2* was significantly differentially expressed between controls and all different stages of dilated cardiomyopathy heart failure. Low signal intensities between $SHRSP_{Gla}$ and WKY_{Gla} and sequence variants in the *Smoc2* 3'UTR might contribute to the discrepancies in fold change results obtained from microarray and qRT-PCR in rat. Nevertheless, *Lamc2* did not validate in dilated cardiomyopathy heart failure, which might suggest that it may be specific for ischemic heart failure. Also the samples used for validation are small subset of a larger study and a bigger sample can be used in the future for *Lamc2*.

When considering the biology of these genes and how they can be functionally related to heart failure, there is evidence that *Smoc2* plays role in Embryogenesis (angiogenesis) and wound healing (Rocnik et al., 2006). A mutated *Lamc2* can cause abnormal morphology of tissue (Meng et al., 2003) that is localised to its epithelial cells. This suggests that these genes may play a role in cardiac stress and wound healing.

6 General Discussion

HTN is a major risk factor for premature CVD such as renal disease, coronary heart disease and cerebrovascular disease, and it has been estimated that it will have affected 29% of the population worldwide by 2025 (Kearney et al., 2005). Blood pressure regulation is influenced by the interaction of multiple genes as well as environmental factors, such as high salt diet. These factors complicate the elucidation of the genetic basis of HTN, as well as prevention and treatment of high blood pressure. Human GWAS has identified a number of candidate genes for blood pressure and hypertension (Levy et al., 2009; Newton-Cheh et al., 2009; Padmanabhan et al., 2010; Ehret et al., 2011; Johnson et al., 2011; Kato et al., 2011; Wain et al., 2011; Salvi et al., 2012; Franceschini et al., 2013). Nevertheless, the effect of the associated SNPs on blood pressure is small and represents a modest fraction of the heritability estimated from family and twin studies (Feinleib et al., 1977; Mongeau et al., 1986; Staessen et al., 2003). Furthermore, the associated SNPs GWAS studied are usually located next to several genes that need further investigation to identify the causative genes and alleles. The use of rats that have been selectively bred to provide animal models for HTN offer more control over genetics background and environmental factors; this includes the production of specific inter-crosses between diseased and control rat for linkage analysis (Graham et al., 2005). The $SHRSP_{Gla}$ is an excellent model for human cardiovascular diseases as it mimics a number of human cardiovascular disease features, including high blood pressure, endothelial dysfunction and left ventricular hypertrophy. Previous work in our laboratory identified a number of QTLs between $SHRSP_{Gla}$ and WKY_{Gla} (control) (Clark et al., 1996) and consequently congenic strains were constructed on chromosomes 2, 3 and 14 for blood pressure, pulse pressure and left ventricular hypertrophy, respectively (Graham et al., 2005).

The experiments presented in this project applied a wide range of systems genetic analysis to investigate the genetic determinants of the $SHRSP_{Gla}$ traits that will help in human translational studies and vice versa. This was achieved by (1) implementing new gene expression analysis workflows that allow for (a) integrative analysis of gene expression profiling data and congenic strains, and (b) to run gene expression cluster analysis between $SHRSP_{Gla}$ and WKY_{Gla} across tissues at different hypertensive stages, (2) identifying, annotating, and validating the genomic variations between $SHRSP_{Gla}$ and WKY_{Gla} whole genomes

and extracting $SHRSP_{Gla}$ unique sequence variants from 28 (non stroke-prone) rat genomes, (3) applying a multifaceted approach that includes and integrates systems genetics analysis from rat and human studies.

The availability of large transcriptomic data from functional tissues relevant to CVD and genome sequences of $SHRSP_{Gla}$ and WKY_{Gla} as well as other rat strains (Atanur et al., 2013) have accelerated our systems genetics analysis and translational studies between rat and human.

Previously, a number of research groups attempted to translate human GWAS candidate genes for blood pressure and hypertension to their hypertensive rats so as to (a) to prioritise and confirm human candidate genes for hypertension (Endres et al., 2014; Jin et al., 2014) even at a single GWAS locus using zinc-finger nuclease technology (Flister et al., 2013), and (b) to explore the regulatory mechanisms that mediate the outcome of genes identified using transcriptomic data from their hypertensive rat strain (Langley et al., 2013).

In this study, we followed a different strategy but with the same objectives (Langley et al., 2013; Endres et al., 2014; Jin et al., 2014) for a translation study from human GWAS candidate genes for blood pressure and hypertension to the $SHRSP_{Gla}$ (from "human to rat") using sequence variants, mRNA (renal and cardiac) and miRNA (renal) expression profiling data, and blood pressure QTLs between $SHRSP_{Gla}$ and WKY_{Gla}. Findings from our human to rat translation study showed that human GWAS candidate genes for blood pressure are mainly conserved in rat and mapped to QTL regions for blood pressure, half of these genes harbour sequence variants, and few are differentially expressed in tissues between $SHRSP_{Gla}$ and WKY_{Gla}. Due to time limitations we could not prove the involvement of these genes in our rat strains; nevertheless, our strategy could be used to prioritise human candidate genes for hypertension even at a single GWAS locus, in a cluster with correlated genes across functional tissues relevant to CVD and with detailed troubleshooting, that are likely be missed by other translation studies. For example, *Gucy1a3-Gucy1b3* (human GWAS candidate genes for blood pressure) encode α and ß sub-units of a heterodimeric protein, soluble guanylate cyclase, that function as nitric oxide receptor and a mouse knock-out of these subunits resulted in hypertension (Friebe et al., 2007).

Gucy1a3-Gucy1b3 were prioritised as translational candidate genes for hypertension since both map to a congenic interval (SP.WKY$_{Gla}$2c*) for blood pressure, are differentially expressed in heart and/or kidney and harbour functional sequence variants between SHRSP$_{Gla}$ and WKY$_{Gla}$. Furthermore, BioLayout*Express*^{3D} analysis across multiple tissues at severely hypertensive stage identified *Gucy1a3* in a cluster of genes (total 8) that share common differentially expression pattern between SHRSP$_{Gla}$ and WKY$_{Gla}$. This cluster of genes might contribute to a network or pathway that underlies the SHRSP$_{Gla}$ phenotype which is considered an ideal translation approach (Delles et al., 2010; Pravenec et al., 2014) giving an insights into the pathogenesis and possible therapeutic targets for human hypertension. Validation by qRT-PCR results for *Gucy1a3* and *Gucy1b3* in heart did not match the microarray expression results. For *Gucy1a3*, this can be due to (a) sequence variants identified in probe–target sequences that may alter probe hybridization affinities causing false positive result, and (b) evidence from Exon array heart data (Crawford, 2011) showing that the Taqman probe used is inappropriate. Moreover, differentially expressed miRNA in kidney between SP.WKY$_{Gla}$2c* and SHRSP$_{Gla}$ which target *Gucy1a3*'s 3'UTR that may suggest a change at post-transcription and/or post-translation level.

Alternatively, in the opposite direction of translation, from "rat to human", only a few groups have been able to successful translate their prioritised findings in hypertensive rats to human. Examples include one research group using specific spontaneously hypertensive heart failure rats to identify and translate *Ephx2* as a heart failure susceptibility gene (Monti et al., 2008), while another group used recombinant inbred rat strains panel produced over years from SHR and BN rat strains to identify and validate *Ogn* as a left ventricular hypertrophy susceptibility gene (Petretto et al., 2008).

In this study, we carried out a different strategy for translation from rat to human, using cardiac gene expression data at neonatal, 5 and 16 weeks of age, sequence variants between SHRSP$_{Gla}$ and WKY$_{Gla}$ integrated with two human gene expression data of heart failure from public repository data (Schinke et al., 2004; Greco et al., 2012). This approach identified two commonly differentially expressed genes with the same fold-change direction in both rat and human

data. These are the *Smoc2* and *Lamc2* genes, and they are both involved in wound healing (Rocnik et al., 2006; Adair-Kirk et al., 2012). qRT-PCR results in rats confirmed the differential expression significance of *Smoc2* and *Lamc2*. Nevertheless, *Smoc2* qRT-PCR fold change direction in our rat study did not match the microarray expression data; the low expression level and/or influence of sequence variants identified within its 3' UTR may contribute to this. On the other hand, qRT-PCR in human dilated cardiomyopathy heart failure tissues validated *Smoc2* differential expression from two microarray gene expression data of heart failure. This may suggest that *Smoc2* plays a cross-species role in cardiac stress and wound healing.

Previously, our group's integrative approach of renal microarray gene expression profiling and chromosome 2 congenic strains led to the identification of a number of positional candidate genes. *Gstm1* was identified as a positional and functional candidate gene for blood pressure using SP.WKY$_{Gla}$2c* (congenic interval does not contain salt-sensitivity QTL) (McBride et al., 2003). Similarly, *S1pr1* and *Vcam1* were identified as positional candidate genes for salt-sensitivity hypertension using SP.WKY$_{Gla}$2a (congenic interval contain salt-sensitivity QTL) (Graham et al., 2007).

The success of our integrative approach of microarray gene expression profiling data with congenic strains usually focuses on single gene/genetic variants. This workflow puts the analysis at risk of missing potential positional candidate genes due to the limitation of the chip used. In addition, we wanted to move away from simple differential expression analysis to cluster analysis and integrate the results with sequence variants to allow findings to be interpreted in context of causative networks and pathways.

In the current study, we improved our workflow for the integrative approach of gene expression profiling data with congenic strains. This was carried out by combining two renal gene expression profiling datasets during salt challenged between SHRSP$_{Gla}$ and WKY$_{Gla}$ from two microarray platforms as well as chromosome 3 sub-congenic strains from Illumina maximised the identification efficiency of positional candidate genes. This approach identified *Dnm1* and *Tor1b* as novel candidate genes for blood pressure, localised within the common

congenic interval shared by SP.WKY$_{Gla}$3d and SP.WKY$_{Gla}$3f strains, and *Rabgap1* as a novel candidate gene for salt-induced pulse pressure variability and renal pathology, located within the region unique to the SP.WKY$_{Gla}$3d strain. This workflow shed the light on failed integrative studies of microarray gene expression profiling and congenic strain to identify positional candidate genes, suggesting that it might be due to the limited efficiency of the gene array used (Moujahidine et al., 2004).

Moving beyond simple differential expression analysis, cluster analysis was carried out by using the BioLayout*Express*^{3D} tool on multiple functional cardiovascular relevant tissues to SHRSP$_{Gla}$ traits at different hypertensive stages. Functional analysis on the identified clusters at pre-hypertensive, hypertensive and severely hypertensive stages between SHRSP$_{Gla}$ and WKY$_{Gla}$ implicated genes with a cardiovascular system and nervous system function, and genes involved in molecular transport, and inflammatory response. The latter was unique to the hypertensive and severely hypertensive stages. *Atp11b*, a P-type ATPase membrane protein critical for ion transport (Halleck et al., 2002) and mediator of platinum chemotherapy resistance in ovarian cancer (Moreno-Smith et al., 2013), was a common gene in all the identified clusters at different hypertensive stages. This may suggest that *Atp11b* is the disease-driving gene since it is differentially expressed from an early age in all relevant tissues and has a predicted frameshift mutation caused by a deletion of 28bp. qRT-PCR on heart and kidney between SHRSP$_{Gla}$ and WKY$_{Gla}$ at different time points validated the cluster analysis results. In addition, *Atp11b* was identified as a novel positional candidate gene for salt-induced blood pressure variability in another renal gene expression cluster analysis between WKY$_{Gla}$, SHRSP$_{Gla}$ and SP.WKY$_{Gla}$2a congenic strains that was not identified previously using other gene expression analysis tools (Graham et al., 2007; Hopcroft et al., 2010). qRT-PCR showed that *Atp11b*, which is located within the congenic interval of SP.WKY$_{Gla}$2a strain, is significantly differentially expressed between SHRSP$_{Gla}$ and SP.WKY$_{Gla}$2a congenic strains at the severely hypertensive stage and under salt challenged. Although a similar strategy was used on hypertensive mouse multiple tissues to identify potential pharmacological targets of hypertension (Puig et al., 2010). Nevertheless, their cluster analysis was at only one hypertensive stage, when it

is difficult to differentiate between cause and effect genes, especially in the absence of sequence variants.

High-throughput NGS identified several genomic variants, including 1,163,332 SNPs and 213,130 INDEL between $SHRSP_{Gla}$ and WKY_{Gla} with uneven distribution across all BN reference genome RGSC 3.4 chromosomes including our QTLs regions on chromosome 2, 3 and 14. The release of new BN reference genome Rnor 5.0 improved our NGS coverage and genomic variants' analysis. However, the overwhelming number of genomics variants makes it extremely difficult to find the $SHRSP_{Gla}$ causal trait variants. Results from integrating genomic and transcriptomic data between $SHRSP_{Gla}$ and WKY_{Gla} revealed that they are likely to be correlated, this coincides with findings from other NGS analysis in rat (Simonis et al., 2012) and human studies (Dimas et al., 2009; Nicolae et al., 2010; Lappalainen et al., 2013). In addition, the small sequence variants that are likely to influence the gene expression between our rats were higher in non-coding regions compared to coding regions which is notable in the light of the recent ENCODE findings that non-coding regions are functional (Ian et al., 2012). Nevertheless, integrating genomics and transcriptomics data can be used as a reductive approach especially when focusing on protein coding variants within our QTLs as they are most likely to contribute (protein function) to the $SHRSP_{Gla}$ phenotype. This reductive and integrative strategy, can be thought of as a "Swiss cheese" model. Using this model we were able to identify a number of positional candidate genes. *Gstm1*, is one of the identified genes, and it was previously shown to influence blood pressure in the $SHRSP_{Gla}$ (McBride et al., 2003; McBride et al., 2005) providing proof of concept for our model. In addition, we were able to validate some of the identified positional genes such as *Ugt2b7* and *Afm* in kidney tissue between parental strains. *Ugt2b7* plays important role in glucuronidation of lipids such as estrogens (Ohno et al., 2004) and *Afm* involves in the regulation and transport of vitamins (Boshtam et al., 2002; Jerkovic et al., 2005); both of these functions, if impaired due to the identified sequence variants, can affect their protein function levels that may result in hypertension.

We also used another integrative and reductive approach to identify disease-causing genes in the $SHRSP_{Gla}$. This was carried out by isolating and annotating $SHRSP_{Gla}$ unique sequence variants when compared to 27 rat genomes with no

stroke phenotype (Atanur et al., 2013). Deleterious protein coding variants were prioritised, especially if they mapped to blood pressure and stroke QTLs or/and were in common with published sequence variants of the $SHRSP_{Izm}$ (Gandolgor et al., 2013). One of the common variants between SHRSP strains identified in our analysis maps to *Stim1*, which was recently identified as a candidate gene for stroke in the $SHRSP_{Izm}$ (Ferdaus et al., 2014). Furthermore, we integrated all $SHRSP_{Gla}$ unique sequence with published frontal and mid-coronal brain gene profiling at different time points (Bailey et al., 2014). This integrative strategy prioritised 39 stroke candidate genes that had not be chosen for validation previously (Bailey et al., 2014) and were further prioritised when considering deleterious sequence variants and our QTL regions.

Our direct approach to translate renal candidate miRNA for hypertension between our rat and human studies using NGS, did not result in common differentially expressed miRNA. This might be due to (a) the differences of genetics between the two species, (b) different sections of kidney being used (in human study we analysed renal medulla and cortex separately, while in rats we used whole kidney), and (c) the human miRNA database being more fully established compared to rat miRNA database. Nevertheless, human renal miRNAs with functions related to hypertension were differentially expressed between hypertensive patients and controls. For example, the hsa-mir-216a and -217 are considered to be biomarkers for chronic kidney disease, and have been involved in the endothelial dysfunction (Menghini et al., 2009; Szeto et al., 2012). In addition, the identified miRNA in our study were not implicated previously in other miRNA studies between hypertensive patients and controls using Agilent Human miRNA microarray (Marques et al., 2011). In our rat miRNA study, we identified differentially expressed miRNA, rno-miR-672-5p and -148b-3p, between $SP.WKY_{Gla}2c^{*}$ and $SHRSP_{Gla}$ where both target *Gstm1* 3'UTR that harbour sequence variants between $SHRSP_{Gla}$ and WKY_{Gla} providing other possible mechanisms for its low mRNA and protein levels in the $SHRSP_{Gla}$ kidney (McBride et al., 2005).

False positive results are likely to be included in large-scale genome and genes' (Larkin et al., 2005) survey analysis. However, due to time limitations we could not carry out all the necessary laboratory validation. The present work identifies

a number of high priority follow up studies. In human studies, *Lamc2* was validated in a small subset of dilated cardiomyopathy heart failure samples. Increasing samples size and/or use of another type of cardiomyopathy that leads to heart failure could validate our study findings. Also, due to the limitations of time, we were not able to address the issue of *Smoc2* fold change direction in our rat between qRT-PCR and microarray results and hence further laboratory validation should be carried out. Moreover, the identified candidate miRNA from human and rat studies should be validated. In rat studies, qRT-PCR validation for *Gucy1a3* with appropriate Taqman probe that covers the alternate spliced exon as well as protein level validation. Luciferase reporter gene experiments (for *Gucy1a3* and *Gstm1*) can also be used on rat kidney to confirm the predicted binding of the targeting miRNA identified from our rat study. Furthermore, qRT-PCR validation is required for *Atp11b* on other implicated tissues between parental strains, at different time points, along with chromosome 2 congenic strains. The positional candidate genes identified by Swiss cheese model were validated between parental strains only this could be extended to include congenic strains. Although the quality control results of NGS sequence variants of $SHRSP_{Gla}$ and WKY_{Gla} demonstrate high sensitivity and specificity when compared to three independent sequencing technologies. Nevertheless, Sanger sequencing is required to validate functional sequence variants implicated within our prioritised candidate genes for the $SHRSP_{Gla}$ phenotype. Finally, transgenic strains can be constructed by introgressing of one of our strong candidate genes from the WKY_{Gla} into the $SHRSP_{Gla}$ background. Transgenic strains are a powerful model to examine and demonstrate the phenotype improvements of the transgene when compared to the $SHRSP_{Gla}$.

Future bioinformatics work in rat studies would also be sensible. Cluster analysis can be carried out using BioLayout*Express*^{3D} between $SHRSP_{Gla}$ and WKY_{Gla} from pre-hypertensive stage till severely hypertensive stage on each tissue separately, to identify tissue specific networks and pathways that may underline $SHRSP_{Gla}$ traits. Moreover, we can integrate targeting genes of miRNAs affected by sequence variants with renal and cardiac transcriptomic data between $SHRSP_{Gla}$ and WKY_{Gla} to prioritise candidate miRNAs for the $SHRSP_{Gla}$ trait. Furthermore, the preliminary results of the customised Affymetrix array to our rat strains compared to the default led to the identification of additional positional

candidates. This can be extended to include other gene expression profiling data. Also, our rat gene expression profiling data can benefit from the new feature offered by IPA, called "Mechanistic networks", where probable signaling cascades are produced from connecting upstream regulators to predict how they might work together to achieve the changes in gene expression data (http://www.ingenuity.com/). In addition, a recent proteomics analysis on mesenteric resistance arteries from $SHRSP_{Gla}$, WKY_{Gla} and chromosome 2 congenic strains, carried out in our laboratory, can be integrated with parental sequence variants to prioritise significant differentially expressed proteins for $SHRSP_{Gla}$ traits. Moreover, a new BN reference genome was released (RGSC 6.0 on July 2014) as well as new variants effect predictor annotation version 76. Extending our NGS analysis to include the latter can improve both of our coverage and genomics variants.

The current study has some limitations. Although, we were able to identify significant bioinformatics results from our rat and human miRNA studies, however, additional biological replicates can improve the results. Also, in our miRNA study we used separated renal cortex and medulla sections in human study while we used a whole kidney in the rat study; this, might contribute to not finding a common significant miRNA between the two. In addition, we used human heart failure tissue for our translation study, which might not be ideal since $SHRSP_{Gla}$ does not progress to the heart failure phenotype; instead, it has left ventricular hypertrophy, which is however a risk factor for heart failure. This is typical of issues with human translation using human tissues. In addition, quality control results show that NGS analysis did not capture a few Sanger sequence variants because these variants did not meet our calling parameters. As a result, lowering the NGS variants calling cut-off and re-running the quality control to adjust our parameters according to the best results could be done. Also, the BN reference genome used (RGSC 3.4 and Rnor 5.0) is a draft version, so we could have missed genomic variants for this reason. Finally, our gene expression analysis on chromosome 3 congenic strains shows that we can miss positional candidate genes due to no probes from Affymetrix or Illumina gene expression arrays representing these genes.

From the current work's findings we can suggest a number of future projects. For instance, RNA-Seq NGS can be carried out on our rat strains tissues to identify positional candidate genes that may not be covered by commercial gene expression arrays, and to avoid false positive results caused by sequence variants. In addition, BioLayout*Express*^{3D} cluster analysis could be conducted on gene expression profiling data from $SHRSP_{Gla}$ with other hypertensive rat strains on public repository databases such as Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) and ArrayExpress (http://www.ebi.ac.uk/arrayexpress/). The clusters identified from this approach may uncover patho-physiological pathways that underline hypertension in rats, which is useful for human studies. This study can be further enhanced by integrating common genomics variants from hypertensive rats (Atanur et al., 2013), which can also prioritise disease causal candidate genes. Furthermore, re-analysis of NGS data from the 28 rats genomes (Atanur et al., 2013) could be performed against the BN reference genome RGSC 6.0, followed by isolation and annotation of $SHRSP_{Gla}$. $SHRSP_{Gla}$ unique genomics variants can then be integrated with other SHRSP strain sequenced, such as $SHRSP_{Izm}$ (Gandolgor et al., 2013), in order to identify common stroke causal variants. This can be further enriched by integrating the findings with brain gene expression data from $SHRSP_{Gla}$ and WKY_{Gla} (Bailey et al., 2014) and $SHRSP_{Izm}$ and WKY_{Izm} strains (Yoshida et al., 2014) to identify common stroke candidate genes. Finally, integrating large scale data from multi-level -omics; transcriptomics, genomics variants, proteomics, and metabolomics into a unified mathematical model to obtain a better understating and wider view of the underlying drivers of pathways for $SHRSP_{Gla}$ traits is challenging. Nevertheless, a number of integrative models has been developed (Bertero et al., 2014), in which most relay on Orthogonal projections to latent structures (Trygg & Wold, 2002), or use commercial software such as IPA.

In conclusion, the current work established and improved methods that identified, prioritised and translated candidate genes for cardiovascular disease specifically hypertension, cardiomyopathies, and stroke using multi-level -omics integration strategies from rat and human studies. The implicated genes, clusters and pathways that may underline the pathogenesis process in the $SHRSP_{Gla}$ can add to the current knowledge of human patho-physiological

mechanisms of cardiovascular disease and eventually lead to therapeutic targets.

7 References

ABYZOV, A., URBAN, A. E., SNYDER, M. & GERSTEIN, M. 2011. CNVnator: An approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Research,* 21, 974-984.

ADAIR-KIRK, T. L., GRIFFIN, G. L., MEYER, M. J., KELLEY, D. G., MINER, J. H., KEENE, D. R., MARINKOVICH, M. P., RUPPERT, J. M., UITTO, J. & SENIOR, R. M. 2012. Keratinocyte-Targeted Expression of Human Laminin gamma 2 Rescues Skin Blistering and Early Lethality of Laminin gamma 2 Deficient Mice. *Plos One,* 7.

ADROGUE, H. J. & MADIAS, N. E. 2007. Mechanisms of disease: Sodium and potassium in the pathogenesis of hypertension. *New England Journal of Medicine,* 356, 1966-1978.

AHN, S.-M., KIM, T.-H., LEE, S., KIM, D., GHANG, H., KIM, D.-S., KIM, B.-C., KIM, S.-Y., KIM, W.-Y., KIM, C., PARK, D., LEE, Y. S., KIM, S., REJA, R., JHO, S., KIM, C. G., CHA, J.-Y., KIM, K.-H., LEE, B., BHAK, J. & KIM, S.-J. 2009. The first Korean genome sequence and analysis: Full genome sequencing for a socio-ethnic group. *Genome Research,* 19, 1622-1629.

AITMAN, T. J., GLAZIER, A. M., WALLACE, C. A., COOPER, L. D., NORSWORTHY, P. J., WAHID, F. N., AL-MAJALI, K. M., TREMBLING, P. M., MANN, C. J., SHOULDERS, C. C., GRAF, D., ST LEZIN, E., KURTZ, T. W., KREN, V., PRAVENEC, M., IBRAHIMI, A., ABUMRAD, N. A., STANTON, L. W. & SCOTT, J. 1999. Identification of Cd36 (Fat) as an insulin-resistance gene causing defective fatty acid and glucose metabolism in hypertensive rats. *Nature Genetics,* 21, 76-83.

AKIYAMA-UCHIDA, Y., ASHIZAWA, N., OHTSURU, A., SETO, S., TSUKAZAKI, T., KIKUCHI, H., YAMASHITA, S. & YANO, K. 2002. Norepinephrine enhances fibrosis mediated by TGF-beta in cardiac fibroblasts. *Hypertension,* 40, 148-154.

ALQADRI, S. L., SREENIVASAN, V. & QURESHI, A. I. 2013. Acute hypertensive response management in patients with acute stroke. *Current cardiology reports,* 15, 426-426.

AMOS, C. I. 2007. Successful design and conduct of genome-wide association studies. *Human Molecular Genetics,* 16, R220-R225.

ANDERS, S. & HUBER, W. 2010. Differential expression analysis for sequence count data. *Genome Biology,* 11.

APELL, H. J. 2004. How do P-type ATPases transport ions? *Bioelectrochemistry,* 63, 149-156.

ATANUR, S. S., BIROL, I., GURYEV, V., HIRST, M., HUMMEL, O., MORRISSEY, C., BEHMOARAS, J., FERNANDEZ-SUAREZ, X. M., JOHNSON, M. D., MCLAREN, W. M., PATONE, G., PETRETTO, E., PLESSY, C., ROCKLAND, K. S., ROCKLAND, C., SAAR, K., ZHAO, Y., CARNINCI, P., FLICEK, P., KURTZ, T., CUPPEN, E., PRAVENEC, M., HUBNER, N., JONES, S. J. M., BIRNEY, E. & AITMAN, T. J. 2010. The genome sequence of the spontaneously hypertensive rat: Analysis and functional significance. *Genome Research,* 20, 791-803.

ATANUR, S. S., DIAZ, A. G., MARATOU, K., SARKIS, A., ROTIVAL, M., GAME, L., TSCHANNEN, M. R., KAISAKI, P. J., OTTO, G. W., MA, M. C. J., KEANE, T. M., HUMMEL, O., SAAR, K., CHEN, W., GURYEV, V., GOPALAKRISHNAN, K., GARRETT, M. R., JOE, B., CITTERIO, L., BIANCHI, G., MCBRIDE, M., DOMINICZAK, A., ADAMS, D. J., SERIKAWA, T., FLICEK, P., CUPPEN, E., HUBNER, N., PETRETTO, E., GAUGUIER, D., KWITEK, A., JACOB, H. & AITMAN, T. J. 2013. Genome Sequencing Reveals Loci under Artificial Selection that Underlie Disease Phenotypes in the Laboratory Rat. *Cell,*

154, 691-703.
BAILEY, E. L., MCBRIDE, M. W., CRAWFORD, W., MCCLURE, J. D., GRAHAM, D., DOMINICZAK, A. F., SUDLOW, C. L. M., SMITH, C. & WARDLAW, J. M. 2014. Differential gene expression in multiple neurological, inflammatory and connective tissue pathways in a spontaneous model of human small vessel stroke. *Neuropathology and Applied Neurobiology*, n/a-n/a.
BARBOSA-MORAIS, N. L., DUNNING, M. J., SAMARAJIWA, S. A., DAROT, J. F. J., RITCHIE, M. E., LYNCH, A. G. & TAVARE, S. 2010. A re-annotation pipeline for Illumina BeadArrays: improving the interpretation of gene expression data. *Nucleic Acids Research,* 38.
BATKAI, S. & THUM, T. 2012. MicroRNAs in Hypertension: Mechanisms and Therapeutic Targets. *Current Hypertension Reports,* 14, 79-87.
BAUML, M. A. & UNDERWOOD, D. A. 2010. Left ventricular hypertrophy: An overlooked cardiovascular risk factor. *Cleveland Clinic Journal of Medicine,* 77, 381-387.
BEILIN, L. J. 2004. Hypertension research in the 21st century: where is the gold? *Journal of Hypertension,* 22, 2243-2251.
BENJAMINI, Y. & HOCHBERG, Y. 1995. Controlling the false discovery rate - a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological,* 57, 289-300.
BENOVOY, D., KWAN, T. & MAJEWSKI, J. 2008. Effect of polymorphisms within probe-target sequences on olignonucleotide microarray experiments. *Nucleic Acids Research,* 36, 4417-4423.
BERTERO, T., LU, Y., ANNIS, S., HALE, A., BHAT, B., SAGGAR, R., SAGGAR, R., WALLACE, W. D., ROSS, D. J., VARGAS, S. O., GRAHAM, B. B., KUMAR, R., BLACK, S. M., FRATZ, S., FINEMAN, J. R., WEST, J. D., HALEY, K. J., WAXMAN, A. B., CHAU, B. N., COTTRILL, K. A. & CHAN, S. Y. 2014. Systems-level regulation of microRNA networks by miR-130/301 promotes pulmonary hypertension. *Journal of Clinical Investigation,* 124, 3514-3528.
BETEL, D., WILSON, M., GABOW, A., MARKS, D. S. & SANDER, C. 2008. The microRNA.org resource: targets and expression. *Nucleic Acids Research,* 36, D149-D153.
BINDER, A. 2007. A review of the genetics of essential hypertension. *Current Opinion in Cardiology,* 22, 176-184.
BOFFA, M. B., MARET, D., HAMILL, J. D., BASTAJIAN, N., CRAINICH, P., JENNY, N. S., TANG, Z., MACY, E. M., TRACY, R. P., FRANCO, R. F., NESHEIM, M. E. & KOSCHINSKY, M. L. 2008. Effect of single nucleotide polymorphisms on expression of the gene encoding thrombin-activatable fibrinolysis inhibitor: a functional analysis. *Blood,* 111, 183-189.
BOSHTAM, M., RAFIEI, M., SADEGHI, K. & SARRAF-ZADEGAN, N. 2002. Vitamin E can reduce blood pressure in mild hypertensives. *International Journal for Vitamin and Nutrition Research,* 72, 309-314.
BOYDEN, L. M., CHOI, M., CHOATE, K. A., NELSON-WILLIAMS, C. J., FARHI, A., TOKA, H. R., TIKHONOVA, I. R., BJORNSON, R., MANE, S. M., COLUSSI, G., LEBEL, M., GORDON, R. D., SEMMEKROT, B. A., POUJOL, A., VALIMAKI, M. J., DE FERRARI, M. E., SANJAD, S. A., GUTKIN, M., KARET, F. E., TUCCI, J. R., STOCKIGT, J. R., KEPPLER-NOREUIL, K. M., PORTER, C. C., ANAND, S. K., WHITEFORD, M. L., DAVIS, I. D., DEWAR, S. B., BETTINELLI, A., FADROWSKI, J. J., BELSHA, C. W., HUNLEY, T. E., NELSON, R. D., TRACHTMAN, H., COLE, T. R. P., PINSK, M., BOCKENHAUER, D., SHENOY, M., VAIDYANATHAN, P., FOREMAN, J. W., RASOULPOUR, M., THAMEEM, F., AL-SHAHROURI, H. Z., RADHAKRISHNAN, J., GHARAVI, A. G., GOILAV, B. &

LIFTON, R. P. 2012. Mutations in kelch-like 3 and cullin 3 cause hypertension and electrolyte abnormalities. *Nature,* 482**,** 98-U126.

BREITLING, R., ARMENGAUD, P., AMTMANN, A. & HERZYK, P. 2004. Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *Febs Letters,* 573**,** 83-92.

BURTON, P. R., CLAYTON, D. G., CARDON, L. R., CRADDOCK, N., DELOUKAS, P., DUNCANSON, A., KWIATKOWSKI, D. P., MCCARTHY, M. I., OUWEHAND, W. H., SAMANI, N. J., TODD, J. A., DONNELLY, P., BARRETT, J. C., DAVISON, D., EASTON, D., EVANS, D., LEUNG, H.-T., MARCHINI, J. L., MORRIS, A. P., SPENCER, C. C. A., TOBIN, M. D., ATTWOOD, A. P., BOORMAN, J. P., CANT, B., EVERSON, U., HUSSEY, J. M., JOLLEY, J. D., KNIGHT, A. S., KOCH, K., MEECH, E., NUTLAND, S., PROWSE, C. V., STEVENS, H. E., TAYLOR, N. C., WALTERS, G. R., WALKER, N. M., WATKINS, N. A., WINZER, T., JONES, R. W., MCARDLE, W. L., RING, S. M., STRACHAN, D. P., PEMBREY, M., BREEN, G., ST CLAIR, D., CAESAR, S., GORDON-SMITH, K., JONES, L., FRASER, C., GREEN, E. K., GROZEVA, D., HAMSHERE, M. L., HOLMANS, P. A., JONES, I. R., KIROV, G., MOSKVINA, V., NIKOLOV, I., O'DONOVAN, M. C., OWEN, M. J., COLLIER, D. A., ELKIN, A., FARMER, A., WILLIAMSON, R., MCGUFFIN, P., YOUNG, A. H., FERRIER, I. N., BALL, S. G., BALMFORTH, A. J., BARRETT, J. H., BISHOP, D. T., ILES, M. M., MAQBOOL, A., YUDASHEVE, N., HALL, A. S., BRAUND, P. S., DIXON, R. J., MANGINO, M., STEVENS, S., THOMPSON, J. R., BREDIN, F., TREMELLING, M., PARKES, M., DRUMMOND, H., LEES, C. W., NIMMO, E. R., SATSANGI, J., FISHER, S. A., FORBES, A., LEWIS, C. M., ONNIE, C. M., PRESCOTT, N. J., SANDERSON, J., MATHEW, C. G., BARBOUR, J., MOHIUDDIN, M. K., TODHUNTER, C. E., MANSFIELD, J. C., AHMAD, T., CUMMINGS, F. R., JEWELL, D. P., et al. 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature,* 447**,** 661-678.

CAMBIEN, F., POIRIER, O., LECERF, L., EVANS, A., CAMBOU, J. P., ARVEILER, D., LUC, G., BARD, J. M., BARA, L., RICARD, S., TIRET, L., AMOUYEL, P., ALHENCGELAS, F. & SOUBRIER, F. 1992. Deletion polymorphism in the gene for angiotensin-converting enzyme is a potent risk factor for myocardial-infarction. *Nature,* 359**,** 641-644.

CAULFIELD, M., MUNROE, P., PEMBROKE, J., SAMANI, N., DOMINICZAK, A., BROWN, M., BENJAMIN, N., WEBSTER, J., RATCLIFFE, P., O'SHEA, S., PAPP, J., TAYLOR, E., DOBSON, R., KNIGHT, J., NEWHOUSE, S., HOOPER, J., LEE, W., BRAIN, N., CLAYTON, D., LATHROP, G. M., FARRALL, M., CONNELL, J. & STU, M. R. C. B. G. H. 2003. Genome-wide mapping of human loci for essential hypertension. *Lancet,* 361**,** 2118-2123.

CHAE, S. S., PAIK, J. H., ALLENDE, M. L., PROIA, R. L. & HLA, T. 2004. Regulation of limb development by the sphingosine 1-phosphate receptor Slp(1)/EDG-1 occurs via the hypoxia/NEGF axis. *Developmental Biology,* 268**,** 441-447.

CHARCHAR, F. J., KAISER, M., BINGHAM, A. J., FOTINATOS, N., AHMADY, F., TOMASZEWSKI, M. & SAMANI, N. J. 2010. Whole Genome Survey of Copy Number Variation in the Spontaneously Hypertensive Rat Relationship to Quantitative Trait Loci, Gene Expression, and Blood Pressure. *Hypertension,* 55**,** 1231-U251.

CHEN, G., BENTLEY, A., ADEYEMO, A., SHRINER, D., ZHOU, J., DOUMATEY, A., HUANG, H., RAMOS, E., ERDOS, M., GERRY, N., HERBERT, A., CHRISTMAN, M. & ROTIMI, C. 2012. Genome-wide association study identifies novel loci association with fasting insulin and insulin resistance in African Americans. *Human Molecular Genetics,* 21**,** 4530-4536.

CHENG, Z. Q., RIOS, G. R., KING, C. D., COFFMAN, B. L., GREEN, M. D.,

MOJARRABI, B., MACKENZIE, P. I. & TEPHLY, T. R. 1998. Glucuronidation of catechol estrogens by expressed human UDP-glucuronosyltransferases (UGTs) 1A1, 1A3, and 2B7. *Toxicological Sciences,* 45**,** 52-57.

CHOBANIAN, A. V., BAKRIS, G. L., BLACK, H. R., CUSHMAN, W. C., GREEN, L. A., IZZO, J. L., JONES, D. W., MATERSON, B. J., OPARIL, S., WRIGHT, J. T., ROCCELLA, E. J. & NATL HIGH BLOOD PRESSURE EDUC, P. 2003. Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. *Hypertension,* 42**,** 1206-1252.

CHOU, C. C., CHEN, C. H., LEE, T. T. & PECK, K. 2004. Optimization of probe length and the number of probes per gene for optimal microarray analysis of gene expression. *Nucleic Acids Research,* 32.

CHRISTIANSEN-WEBER, T. A., VOLAND, J. R., WU, Y., NGO, K., ROLAND, B. L., NGUYEN, S., PETERSON, P. A. & FUNG-LEUNG, W. P. 2000. Functional loss of ABCA1 in mice causes severe placental malformation, aberrant lipid distribution, and kidney glomerulonephritis as well as high-density lipoprotein cholesterol deficiency. *American Journal of Pathology,* 157**,** 1017-1029.

CLARK, J. S., JEFFS, B., DAVIDSON, A. O., LEE, W. K., ANDERSON, N. H., BIHOREAU, M. T., BROSNAN, M. J., DEVLIN, A. M., KELMAN, A. W., LINDPAINTNER, K. & DOMINICZAK, A. F. 1996. Quantitative trait loci in genetically hypertensive rats - Possible sex specificity. *Hypertension,* 28**,** 898-906.

CLARKE, C., FLORES-MUNOZ, M., MCKINNEY, C. A., MILLIGAN, G. & NICKLIN, S. A. 2013. Regulation of cardiovascular remodeling by the counter-regulatory axis of the renin-angiotensin system. *Future cardiology,* 9**,** 23-38.

COOPER, G. M. & SHENDURE, J. 2011. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nature Reviews Genetics,* 12**,** 628-640.

CRAWFORD, W. 2011. Microarray analysis of chromosome 14 congenic strains in the SHRSP. *University of Glasgow* MSc(R).

DAHLOF, B. 2009. Prevention of stroke: new evidence. *European Heart Journal Supplements,* 11**,** F33-F38.

DAI, M. H., WANG, P. L., BOYD, A. D., KOSTOV, G., ATHEY, B., JONES, E. G., BUNNEY, W. E., MYERS, R. M., SPEED, T. P., AKIL, H., WATSON, S. J. & MENG, F. 2005. Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Research,* 33.

DAVIDSON, A. O., SCHORK, N., JAQUES, B. C., KELMAN, A. W., SUTCLIFFE, R. G., REID, J. L. & DOMINICZAK, A. F. 1995. Blood-pressure in genetically hypertensive rats - influence of the Y-chromosome. *Hypertension,* 26**,** 452-459.

DAVIS, M. P. A., VAN DONGEN, S., ABREU-GOODGER, C., BARTONICEK, N. & ENRIGHT, A. J. 2013. Kraken: A set of tools for quality control and analysis Of high-throughput sequence data. *Methods,* 63**,** 41-49.

DELLES, C., MCBRIDE, M. W., GRAHAM, D., PADMANABHAN, S. & DOMINICZAK, A. F. 2010. Genetics of hypertension: From experimental animals to humans. *Biochimica Et Biophysica Acta-Molecular Basis of Disease,* 1802**,** 1299-1308.

DELLES, C., MCBRIDE, M. W., PADMANABHAN, S. & DOMINICZAK, A. F. 2008a. The genetics of cardiovascular disease. *Trends in Endocrinology and Metabolism,* 19**,** 309-316.

DELLES, C., MILLER, W. H. & DOMINICZAK, A. F. 2008b. Targeting reactive oxygen species in hypertension. *Antioxidants & Redox Signaling,* 10**,** 1061-1077.

DELLES, C., PADMANABHAN, S., LEE, W. K., MILLER, W. H., MCBRIDE, M. W., MCCLURE, J. D., BRAIN, N. J., WALLACE, C., MARCANO, A. C. B., SCHMIEDER, R. E., BROWN, M. J., CAULFIELD, M. J., MUNROE, P. B., FARRALL, M., WEBSTER, J., CONNELL, J. M. & DOMINICZAK, A. F. 2008c. Glutathione S-transferase variants and hypertension. *Journal of Hypertension,* 26**,** 1343-1352.

DEPRISTO, M. A., BANKS, E., POPLIN, R., GARIMELLA, K. V., MAGUIRE, J. R., HARTL, C., PHILIPPAKIS, A. A., DEL ANGEL, G., RIVAS, M. A., HANNA, M., MCKENNA, A., FENNELL, T. J., KERNYTSKY, A. M., SIVACHENKO, A. Y., CIBULSKIS, K., GABRIEL, S. B., ALTSHULER, D. & DALY, M. J. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics,* 43**,** 491-+.

DIMAS, A. S., DEUTSCH, S., STRANGER, B. E., MONTGOMERY, S. B., BOREL, C., ATTAR-COHEN, H., INGLE, C., BEAZLEY, C., ARCELUS, M. G., SEKOWSKA, M., GAGNEBIN, M., NISBETT, J., DELOUKAS, P., DERMITZAKIS, E. T. & ANTONARAKIS, S. E. 2009. Common Regulatory Variation Impacts Gene Expression in a Cell Type-Dependent Manner. *Science,* 325**,** 1246-1250.

DRENJANCEVIC-PERIC, I., JELAKOVIC, B., LOMBARD, J. H., KUNERT, M. P., KIBEL, A. & GROS, M. 2011. High-Salt Diet and Hypertension: Focus on the Renin-Angiotensin System. *Kidney & Blood Pressure Research,* 34**,** 1-11.

DUNHAM, I., KUNDAJE, A., ALDRED, S. F., COLLINS, P. J., DAVIS, C., DOYLE, F., EPSTEIN, C. B., FRIETZE, S., HARROW, J., KAUL, R., KHATUN, J., LAJOIE, B. R., LANDT, S. G., LEE, B.-K., PAULI, F., ROSENBLOOM, K. R., SABO, P., SAFI, A., SANYAL, A., SHORESH, N., SIMON, J. M., SONG, L., TRINKLEIN, N. D., ALTSHULER, R. C., BIRNEY, E., BROWN, J. B., CHENG, C., DJEBALI, S., DONG, X., ERNST, J., FUREY, T. S., GERSTEIN, M., GIARDINE, B., GREVEN, M., HARDISON, R. C., HARRIS, R. S., HERRERO, J., HOFFMAN, M. M., IYER, S., KELLIS, M., KHERADPOUR, P., LASSMANN, T., LI, Q., LIN, X., MARINOV, G. K., MERKEL, A., MORTAZAVI, A., PARKER, S. C. J., REDDY, T. E., ROZOWSKY, J., SCHLESINGER, F., THURMAN, R. E., WANG, J., WARD, L. D., WHITFIELD, T. W., WILDER, S. P., WU, W., XI, H. S., YIP, K. Y., ZHUANG, J., BERNSTEIN, B. E., GREEN, E. D., GUNTER, C., SNYDER, M., PAZIN, M. J., LOWDON, R. F., DILLON, L. A. L., ADAMS, L. B., KELLY, C. J., ZHANG, J., WEXLER, J. R., GOOD, P. J., FEINGOLD, E. A., CRAWFORD, G. E., DEKKER, J., ELNITSKI, L., FARNHAM, P. J., GIDDINGS, M. C., GINGERAS, T. R., GUIGO, R., HUBBARD, T. J., KENT, W. J., LIEB, J. D., MARGULIES, E. H., MYERS, R. M., STAMATOYANNOPOULOS, J. A., TENENBAUM, S. A., WENG, Z., WHITE, K. P., WOLD, B., YU, Y., WROBEL, J., RISK, B. A., GUNAWARDENA, H. P., KUIPER, H. C., MAIER, C. W., XIE, L., CHEN, X., MIKKELSEN, T. S., GILLESPIE, S., et al. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature,* 489**,** 57-74.

EHRET, G. B., MUNROE, P. B., RICE, K. M., BOCHUD, M., JOHNSON, A. D., CHASMAN, D. I., SMITH, A. V., TOBIN, M. D., VERWOERT, G. C., HWANG, S. J., PIHUR, V., VOLLENWEIDER, P., O'REILLY, P. F., AMIN, N., BRAGG-GRESHAM, J. L., TEUMER, A., GLAZER, N. L., LAUNER, L., ZHAO, J. H., AULCHENKO, Y., HEATH, S., SOBER, S., PARSA, A., LUAN, J. A., ARORA, P., DEHGHAN, A., ZHANG, F., LUCAS, G., HICKS, A. A., JACKSON, A. U., PEDEN, J. F., TANAKA, T., WILD, S. H., RUDAN, I., IGL, W., MILANESCHI, Y., PARKER, A. N., FAVA, C., CHAMBERS, J. C., FOX, E. R., KUMARI, M., GO, M. J., VAN DER HARST, P., KAO, W. H. L., SJOGREN, M., VINAY, D. G., ALEXANDER, M., TABARA, Y., SHAW-HAWKINS, S., WHINCUP, P. H., LIU, Y. M., SHI, G., KUUSISTO, J., TAYO, B., SEIELSTAD, M., SIM, X., NGUYEN, K. D. H., LEHTIMAKI, T., MATULLO, G., WU, Y., GAUNT, T. R., ONLAND-MORET,

N. C., COOPER, M. N., PLATOU, C. G. P., ORG, E., HARDY, R., DAHGAM, S., PALMEN, J., VITART, V., BRAUND, P. S., KUZNETSOVA, T., UITERWAAL, C., ADEYEMO, A., PALMAS, W., CAMPBELL, H., LUDWIG, B., TOMASZEWSKI, M., TZOULAKI, I., PALMER, N. D., ASPELUND, T., GARCIA, M., CHANG, Y. P. C., O'CONNELL, J. R., STEINLE, N. I., GROBBEE, D. E., ARKING, D. E., KARDIA, S. L., MORRISON, A. C., HERNANDEZ, D., NAJJAR, S., MCARDLE, W. L., HADLEY, D., BROWN, M. J., CONNELL, J. M., HINGORANI, A. D., DAY, I. N. M., LAWLOR, D. A., BEILBY, J. P., LAWRENCE, R. W., CLARKE, R., et al. 2011. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature,* 478**,** 103-109.

ENDEMANN, D. H. & SCHIFFRIN, E. L. 2004. Endothelial dysfunction. *Journal of the American Society of Nephrology,* 15**,** 1983-1992.

ENDRES, B. T., PRIESTLEY, J. R. C., PALYGIN, O., FLISTER, M. J., HOFFMAN, M. J., WEINBERG, B. D., GRZYBOWSKI, M., LOMBARD, J. H., STARUSCHENKO, A., MORENO, C., JACOB, H. J. & GEURTS, A. M. 2014. Mutation of Plekha7 attenuates salt-sensitive hypertension in the rat. *Proceedings of the National Academy of Sciences of the United States of America,* 111**,** 12817-12822.

FASELIS, C., DOUMAS, M. & PAPADEMETRIOU, V. 2011. Common secondary causes of resistant hypertension and rational for treatment. *International journal of hypertension,* 2011**,** 236239-236239.

FAUL, C., AMARAL, A. P., OSKOUEI, B., HU, M.-C., SLOAN, A., ISAKOVA, T., GUTIERREZ, O. M., AGUILLON-PRADA, R., LINCOLN, J., HARE, J. M., MUNDEL, P., MORALES, A., SCIALLA, J., FISCHER, M., SOLIMAN, E. Z., CHEN, J., GO, A. S., ROSAS, S. E., NESSEL, L., TOWNSEND, R. R., FELDMAN, H. I., SUTTON, M. S. J., OJO, A., GADEGBEKU, C., DI MARCO, G. S., REUTER, S., KENTRUP, D., TIEMANN, K., BRAND, M., HILL, J. A., MOE, O. W., KURO-O, M., KUSEK, J. W., KEANE, M. G. & WOLF, M. 2011. FGF23 induces left ventricular hypertrophy. *Journal of Clinical Investigation,* 121**,** 4393-4408.

FEINLEIB, M., GARRISON, R. J., FABSITZ, R., CHRISTIAN, J. C., HRUBEC, Z., BORHANI, N. O., KANNEL, W. B., ROSENMAN, R., SCHWARTZ, J. T. & WAGNER, J. O. 1977. NHLBI TWIN STUDY OF CARDIOVASCULAR-DISEASE RISK-FACTORS - METHODOLOGY AND SUMMARY OF RESULTS. *American Journal of Epidemiology,* 106**,** 284-295.

FERDAUS, M. Z., XIAO, B., OHARA, H., NEMOTO, K., HARADA, Y., SAAR, K., HUEBNER, N., ISOMURA, M. & NABIKA, T. 2014. Identification of Stim1 as a Candidate Gene for Exaggerated Sympathetic Response to Stress in the Stroke-Prone Spontaneously Hypertensive Rat. *Plos One,* 9.

FERGUSON, S. M., BRASNJO, G., HAYASHI, M., WOELFEL, M., COLLESI, C., GIOVEDI, S., RAIMONDI, A., GONG, L.-W., ARIEL, P., PARADISE, S., O'TOOLE, E., FLAVELL, R., CREMONA, O., MIESENBOECK, G., RYAN, T. A. & DE CAMILLI, P. 2007. A selective activity-dependent requirement for dynamin 1 in synaptic vesicle endocytosis. *Science,* 316**,** 570-574.

FLETCHER, E. C., LESSKE, J., BEHM, R., MILLER, C. C., STAUSS, H. & UNGER, T. 1992. CAROTID CHEMORECEPTORS, SYSTEMIC BLOOD-PRESSURE, AND CHRONIC EPISODIC HYPOXIA MIMICKING SLEEP-APNEA. *Journal of Applied Physiology,* 72**,** 1978-1984.

FLISTER, M. J., TSAIH, S.-W., O'MEARA, C. C., ENDRES, B., HOFFMAN, M. J., GEURTS, A. M., DWINELL, M. R., LAZAR, J., JACOB, H. J. & MORENO, C. 2013. Identifying multiple causative genes at a single GWAS locus. *Genome Research,* 23**,** 1996-2002.

FLORES-MUNOZ, M., WORK, L. M., DOUGLAS, K., DENBY, L., DOMINICZAK, A. F.,

GRAHAM, D. & NICKLIN, S. A. 2012. Angiotensin-(1-9) Attenuates Cardiac Fibrosis in the Stroke-Prone Spontaneously Hypertensive Rat via the Angiotensin Type 2 Receptor. *Hypertension,* 59**,** 300-+.

FOOTE, K., MCBRIDE, M. W., GRAHAM, D., DOUGLAS, K., KETTLEWELL, S., SMITH, G. L., DOMINICZAK, A. F. & LOUGHREY, C. M. 2011. Assessment of Cardiac Function in Chromosome 14 Congenic Strains using Pressure-Volume Measurements. *Biophysical Journal,* 100**,** 299-299.

FORNAGE, M., DEBETTE, S., BIS, J. C., SCHMIDT, H., IKRAM, M. A., DUFOUIL, C., SIGURDSSON, S., LUMLEY, T., DESTEFANO, A. L., FAZEKAS, F., VROOMAN, H. A., SHIBATA, D. K., MAILLARD, P., ZIJDENBOS, A., SMITH, A. V., GUDNASON, H., DE BOER, R., CUSHMAN, M., MAZOYER, B., HEISS, G., VERNOOIJ, M. W., ENZINGER, C., GLAZER, N. L., BEISER, A., KNOPMAN, D. S., CAVALIERI, M., NIESSEN, W. J., HARRIS, T. B., PETROVIC, K., LOPEZ, O. L., AU, R., LAMBERT, J.-C., HOFMAN, A., GOTTESMAN, R. F., GARCIA, M., HECKBERT, S. R., ATWOOD, L. D., CATELLIER, D. J., UITTERLINDEN, A. G., YANG, Q., SMITH, N. L., ASPELUND, T., ROMERO, J. R., RICE, K., TAYLOR, K. D., NALLS, M. A., ROTTER, J. I., SHARRETT, R., VAN DUIJN, C. M., AMOUYEL, P., WOLF, P. A., GUDNASON, V., VAN DER LUGT, A., BOERWINKLE, E., PSATY, B. M., SESHADRI, S., TZOURIO, C., BRETELER, M. M. B., MOSLEY, T. H., SCHMIDT, R., LONGSTRETH, W. T., DECARLI, C. & LAUNER, L. J. 2011. Genome-Wide Association Studies of Cerebral White Matter Lesion Burden: The CHARGE Consortium. *Annals of Neurology,* 69**,** 928-939.

FRANCESCHINI, N., FOX, E., ZHANG, Z., EDWARDS, T. L., NALLS, M. A., SUNG, Y. J., TAYO, B. O., SUN, Y. V., GOTTESMAN, O., ADEYEMO, A., JOHNSON, A. D., YOUNG, J. H., RICE, K., DUAN, Q., CHEN, F., LI, Y., TANG, H., FORNAGE, M., KEENE, K. L., ANDREWS, J. S., SMITH, J. A., FAU, J. D., GUANGFA, Z., GUO, W., LIU, Y., MURRAY, S. S., MUSANI, S. K., SRINIVASAN, S., EDWARDS, D. R. V., WANG, H., BECKER, L. C., BOVET, P., BOCHUD, M., BROECKE, U., BURNIER, M., CARTY, C., CHASMAN, D. I., EHRET, G., CHEN, W.-M., CHEN, G., CHEN, W., DING, J., DREISBACH, A. W., EVANS, M. K., GUO, X., GARCIA, M. E., JENSEN, R., KELLER, M. E., LETTRE, G., LOTAY, V., MARTIN, L. W., MOORE, J. H., MORRISON, A. C., MOSLEY, T. H., OGUNNIYI, A., PALMAS, W., PAPANICOLAOU, G., PENMAN, A., POLAK, J. F., RIDKER, P. M., SALAKO, B., SINGLETON, A. B., SHRINER, D., TAYLOR, K. D., VASAN, R., WIGGINS, K., WILLIAMS, S. M., YANEK, L. R., ZHAO, W., ZONDERMAN, A. B., BECKER, D. M., BERENSON, G., BOERWINKLE, E., BOTTINGER, E., CUSHMAN, M., EATON, C., NYBERG, F., HEISS, G., HIRSCHHRON, J. N., HOWARD, V. J., KARCZEWSK, K. J., LANKTREE, M. B., LIU, K., LIU, Y., LOOS, R., MARGOLIS, K., SNYDER, M., PSATY, B. M., SCHORK, N. J., WEIR, D. R., ROTIMI, C. N., SALE, M. M., HARRIS, T., KARDIA, S. L. R., HUNT, S. C., ARNETT, D., REDLINE, S., COOPER, R. S., RISCH, N. J., RAO, D. C., et al. 2013. Genome-wide Association Analysis of Blood-Pressure Traits in African-Ancestry Individuals Reveals Common Associated Genes in African and Non-African Populations. *American Journal of Human Genetics,* 93**,** 545-554.

FREEL, E. M. & CONNELL, J. M. C. 2004. Mechanisms of hypertension: The expanding role of aldosterone. *Journal of the American Society of Nephrology,* 15**,** 1993-2001.

FREEMAN, T. C., GOLDOVSKY, L., BROSCH, M., VAN DONGEN, S., MAZIERE, P., GROCOCK, R. J., FREILICH, S., THORNTON, J. & ENRIGHT, A. J. 2007. Construction, visualisation, and clustering of transcription networks from Microarray expression data. *Plos Computational Biology,* 3**,** 2032-2042.

FRESE, E. M., FICK, A. & SADOWSKY, H. S. 2011. Blood pressure measurement

guidelines for physical therapists. *Cardiopulmonary physical therapy journal,* 22**,** 5-12.

FRIEBE, A., MERGIA, E., DANGEL, O., LANGE, A. & KOESLING, D. 2007. Fatal gastrointestinal obstruction and hypertension in mice lacking nitric oxide-sensitive guanylyl cyclase. *Proceedings of the National Academy of Sciences of the United States of America,* 104**,** 7699-7704.

GAENGEL, K., NIAUDET, C., HAGIKURA, K., SIEMSEN, B. L., MUHL, L., HOFMANN, J. J., EBARASI, L., NYSTROM, S., RYMO, S., CHEN, L. L., PANG, M.-F., JIN, Y., RASCHPERGER, E., ROSWALL, P., SCHULTE, D., BENEDITO, R., LARSSON, J., HELLSTROM, M., FUXE, J., UHLEN, P., ADAMS, R., JAKOBSSON, L., MAJUMDAR, A., VESTWEBER, D., UV, A. & BETSHOLTZ, C. 2012. The Sphingosine-1-Phosphate Receptor S1PR1 Restricts Sprouting Angiogenesis by Regulating the Interplay between VE-Cadherin and VEGFR2. *Developmental Cell,* 23**,** 587-599.

GANDOLGOR, T.-A., OHARA, H., CUI, Z.-H., HIRASHIMA, T., OGAWA, T., SAAR, K., HUEBNER, N., WATANABE, T., ISOMURA, M. & NABIKA, T. 2013. Two Genomic Regions of Chromosomes 1 and 18 Explain Most of the Stroke Susceptibility Under Salt Loading in Stroke-Prone Spontaneously Hypertensive Rat/Izm. *Hypertension,* 62**,** 55-61.

GARBERS, D. L., CHRISMAN, T. D., WIEGN, P., KATAFUCHI, T., ALBANESI, J. P., BIELINSKI, V., BARYLKO, B., REDFIELD, M. M. & BURNETT, J. C., JR. 2006. Membrane guanylyl cyclase receptors: an update. *Trends in Endocrinology and Metabolism,* 17**,** 251-258.

GARCIA, J. G. N., LIU, F., VERIN, A. D., BIRUKOVA, A., DECHERT, M. A., GERTHOFFER, W. T., BAMBURG, J. R. & ENGLISH, D. 2001. Sphingosine 1-phosphate promotes endothelial cell barrier integrity by Edg-dependent cytoskeletal rearrangement. *Journal of Clinical Investigation,* 108**,** 689-701.

GAROVIC, V. D., HILLIARD, A. A. & TURNER, S. T. 2006. Monogenic forms of low-renin hypertension. *Nature Clinical Practice Nephrology,* 2**,** 624-630.

GIBBS, C. R., BEEVERS, D. G. & LIP, G. Y. H. 1999. The management of hypertensive disease in Black patients. *Qjm-Monthly Journal of the Association of Physicians,* 92**,** 187-192.

GIBBS, R. A., WEINSTOCK, G. M., METZKER, M. L., MUZNY, D. M., SODERGREN, E. J., SCHERER, S., SCOTT, G., STEFFEN, D., WORLEY, K. C., BURCH, P. E., OKWUONU, G., HINES, S., LEWIS, L., DERAMO, C., DELGADO, O., DUGAN-ROCHA, S., MINER, G., MORGAN, M., HAWES, A., GILL, R., HOLT, R. A., ADAMS, M. D., AMANATIDES, P. G., BADEN-TILLSON, H., BARNSTEAD, M., CHIN, S., EVANS, C. A., FERRIERA, S., FOSLER, C., GLODEK, A., GU, Z. P., JENNINGS, D., KRAFT, C. L., NGUYEN, T., PFANNKOCH, C. M., SITTER, C., SUTTON, G. G., VENTER, J. C., WOODAGE, T., SMITH, D., LEE, H. M., GUSTAFSON, E., CAHILL, P., KANA, A., DOUCETTE-STAMM, L., WEINSTOCK, K., FECHTEL, K., WEISS, R. B., DUNN, D. M., GREEN, E. D., BLAKESLEY, R. W., BOUFFARD, G. G., DE JONG, J., OSOEGAWA, K., ZHU, B. L., MARRA, M., SCHEIN, J., BOSDET, I., FJELL, C., JONES, S., KRZYWINSKI, M., MATHEWSON, C., SIDDIQUI, A., WYE, N., MCPHERSON, J., ZHAO, S. Y., FRASER, C. M., SHETTY, J., SHATSMAN, S., GEER, K., CHEN, Y. X., ABRAMZON, S., NIERMAN, W. C., HAVLAK, P. H., CHEN, R., DURBIN, K. J., EGAN, A., REN, Y. R., SONG, X. Z., LI, B. S., LIU, Y., QIN, X., CAWLEY, S., COONEY, A. J., D'SOUZA, L. M., MARTIN, K., WU, J. Q., GONZALEZ-GARAY, M. L., JACKSON, A. R., KALAFUS, K. J., MCLEOD, M. P., MILOSAVLJEVIC, A., VIRK, D., VOLKOV, A., WHEELER, D. A., ZHANG, Z. D., BAILEY, J. A., EICHLER, E. E., TUZUN, E., BIRNEY, E., et al. 2004. Genome sequence of

the Brown Norway rat yields insights into mammalian evolution. *Nature,* 428**,** 493-521.

GONZALEZ, A., LOPEZ, B., QUEREJETA, R. & DIEZ, J. 2002. Regulation of myocardial fibrillar collagen by angiotensin II. A role in hypertensive heart disease? *Journal of Molecular and Cellular Cardiology,* 34**,** 1585-1593.

GRADMAN, A. H. & ALFAYOUMI, F. 2006. From left ventricular hypertrophy to congestive heart failure: Management of hypertensive heart disease. *Progress in Cardiovascular Diseases,* 48**,** 326-341.

GRAHAM, D., MCBRIDE, M. W., BRAIN, N. J. R. & DOMINICZAK, A. F. 2005. Congenic/consomic models of hypertension. *Hypertension: Methods and Protocols,* 108**,** 3-15.

GRAHAM, D., MCBRIDE, M. W., GAASENBEEK, M., GILDAY, K., BEATTIE, E., MILLER, W. H., MCCLURE, J. D., POLKE, J. M., MONTEZANO, A., TOUYZ, R. M. & DOMINICZAK, A. F. 2007. Candidate genes that determine response to salt in the stroke-prone spontaneously hypertensive rat - Congenic analysis. *Hypertension,* 50**,** 1134-1141.

GRAUR, D., ZHENG, Y., PRICE, N., AZEVEDO, R. B. R., ZUFALL, R. A. & ELHAIK, E. 2013. On the Immortality of Television Sets: "Function" in the Human Genome According to the Evolution-Free Gospel of ENCODE. *Genome Biology and Evolution,* 5**,** 578-590.

GRECO, S., FASANARO, P., CASTELVECCHIO, S., D'ALESSANDRA, Y., ARCELLI, D., DI DONATO, M., MALAVAZOS, A., CAPOGROSSI, M. C., MENICANTI, L. & MARTELLI, F. 2012. MicroRNA Dysregulation in Diabetic Ischemic Heart Failure Patients. *Diabetes,* 61**,** 1633-1641.

GRIFFITHS-JONES, S., SAINI, H. K., VAN DONGEN, S. & ENRIGHT, A. J. 2008. miRBase: tools for microRNA genomics. *Nucleic Acids Research,* 36**,** D154-D158.

GUO, X., BRENNER, M., ZHANG, X., LARAGIONE, T., TAI, S., LI, Y., BU, J., YIN, Y., SHAH, A. A., KWAN, K., LI, Y., JUN, W. & GULKO, P. S. 2013. Whole-Genome Sequences of DA and F344 Rats with Different Susceptibilities to Arthritis, Autoimmunity, Inflammation and Cancer. *Genetics,* 194**,** 1017-+.

GURYEV, V., SAAR, K., ADAMOVIC, T., VERHEUL, M., VAN HEESCH, S. A. A. C., COOK, S., PRAVENEC, M., AITMAN, T., JACOB, H., SHULL, J. D., HUBNER, N. & CUPPEN, E. 2008. Distribution and functional impact of DNA copy number variation in the rat. *Nature Genetics,* 40**,** 538-545.

GUYENET, P. G. 2006. The sympathetic control of blood pressure. *Nature Reviews Neuroscience,* 7**,** 335-346.

GUYTON, A. C. 1991. Blood-pressure control - special role of the kidneys and body-fluids. *Science,* 252**,** 1813-1816.

HALL, J. E. 1991. The renin angiotensin system renal actions and blood pressure regulation. *Comprehensive Therapy,* 17**,** 8-17.

HALLECK, M. S., SCHLEGEL, R. A. & WILLIAMSON, P. L. 2002. Reanalysis of ATP11B, a type IV P-type ATPase. *Journal of Biological Chemistry,* 277**,** 9736-9740.

HANCOCK, J. T., DESIKAN, R. & NEILL, S. J. 2001. Role of reactive oxygen species in cell signalling pathways. *Biochemical Society Transactions,* 29**,** 345-350.

HANSSON, J. H., NELSONWILLIAMS, C., SUZUKI, H., SCHILD, L., SHIMKETS, R., LU, Y., CANESSA, C., IWASAKI, T., ROSSIER, B. & LIFTON, R. P. 1995. Hypertension caused by a truncated epithelial sodium-channel gamma-subunit - genetic-heterogeneity of Liddle syndrome. *Nature Genetics,* 11**,** 76-82.

HINDORFF, L. A., SETHUPATHY, P., JUNKINS, H. A., RAMOS, E. M., MEHTA, J. P., COLLINS, F. S. & MANOLIO, T. A. 2009. Potential etiologic and functional

implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences of the United States of America,* 106**,** 9362-9367.

HOLTHE, M., RAKVAG, T. N., KLEPSTAD, P., IDLE, J. R., KAASA, S., KROKAN, H. E. & SKORPEN, F. 2003. Sequence variations in the UDP-glucuronosyltransferase 2B7 (UGT2B7) gene: identification of 10 novel single nucleotide polymorphisms (SNPs) and analysis of their relevance to morphine glucuronidation in cancer patients. *Pharmacogenomics Journal,* 3**,** 17-26.

HOPCROFT, L. E. M., MCBRIDE, M. W., HARRIS, K. J., SAMPSON, A. K., MCCLURE, J. D., GRAHAM, D., YOUNG, G., HOLYOAKE, T. L., GIROLAMI, M. A. & DOMINICZAK, A. F. 2010. Predictive response-relevant clustering of expression data provides insights into disease processes. *Nucleic Acids Research,* 38**,** 6831-6840.

HYNDMAN, K. A., MUSALL, J. B., XUE, J. & POLLOCK, J. S. 2011. Dynamin activates NO production in rat renal inner medullary collecting ducts via protein-protein interaction with NOS1. *American Journal of Physiology-Renal Physiology,* 301**,** F118-F124.

IRIZARRY, R. A., HOBBS, B., COLLIN, F., BEAZER-BARCLAY, Y. D., ANTONELLIS, K. J., SCHERF, U. & SPEED, T. P. 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics,* 4**,** 249-264.

IWAI, M., LIU, H. W., CHEN, R., IDE, A., OKAMOTO, S., HATA, R., SAKANAKA, M., SHIUCHI, T. & HORIUCHI, M. 2004. Possible inhibition of focal cerebral ischemia by angiotensin II type 2 receptor stimulation. *Circulation,* 110**,** 843-848.

JACKSON, K. L., MARQUES, F. Z., WATSON, A. M. D., PALMA-RIGO, K., THU-PHUC, N.-H., MORRIS, B. J., CHARCHAR, F. J., DAVERN, P. J. & HEAD, G. A. 2013. A Novel Interaction Between Sympathetic Overactivity and Aberrant Regulation of Renin by miR-181a in BPH/2J Genetically Hypertensive Mice. *Hypertension,* 62**,** 775-+.

JAFFE, A., WOJCIK, G., CHU, A., GOLOZAR, A., MAROO, A., DUGGAL, P. & KLEIN, A. P. 2011. Identification of functional genetic variation in exome sequence analysis. *BMC proceedings,* 5 Suppl 9**,** S13-S13.

JAMES, D. J., CAIRNS, F., SALT, I. P., MURPHY, G. J., DOMINICZAK, A. F., CONNELL, J. M. C. & GOULD, G. W. 2001. Skeletal muscle of stroke-prone spontaneously hypertensive rats exhibits reduced insulin-stimulated glucose transport and elevated levels of caveolin and flotillin. *Diabetes,* 50**,** 2148-2156.

JAMES, P. A., OPARIL, S., CARTER, B. L., CUSHMAN, W. C., DENNISON-HIMMELFARB, C., HANDLER, J., LACKLAND, D. T., LEFEVRE, M. L., MACKENZIE, T. D., OGEDEGBE, O., SMITH, S. C., JR., SVETKEY, L. P., TALER, S. J., TOWNSEND, R. R., WRIGHT, J. T., JR., NARVA, A. S. & ORTIZ, E. 2014. 2014 Evidence-Based Guideline for the Management of High Blood Pressure in Adults Report From the Panel Members Appointed to the Eighth Joint National Committee (JNC 8). *Jama-Journal of the American Medical Association,* 311**,** 507-520.

JEFFS, B., CLARK, J. S., ANDERSON, N. H., GRATTON, J., BROSNAN, M. J., GAUGUIER, D., REID, J. L., MACRAE, I. M. & DOMINICZAK, A. F. 1997. Sensitivity to cerebral ischaemic insult in a rat model of stroke is determined by a single genetic locus. *Nature Genetics,* 16**,** 364-367.

JERKOVIC, L., VOEGELE, A. F., CHWATAL, S., KRONENBERG, F., RADCLIFFE, C. M., WORMALD, M. R., LOBENTANZ, E. M., EZEH, B., ELLER, P., DEJORI, N.,

DIEPLINGER, B., LOTTSPEICH, F., SATTLER, W., UHR, M., MECHTLER, K., DWEK, R. A., RUDD, P. M., BAIER, G. & DIEPLINGER, H. 2005. Afamin is a novel human vitamin E-binding glycoprotein characterization and in vitro expression. *Journal of Proteome Research,* 4**,** 889-899.

JI, W., FOO, J. N., O'ROAK, B. J., ZHAO, H., LARSON, M. G., SIMON, D. B., NEWTON-CHEH, C., STATE, M. W., LEVY, D. & LIFTON, R. P. 2008. Rare independent mutations in renal salt handling genes contribute to blood pressure variation. *Nature Genetics,* 40**,** 592-599.

JIN, C., SUN, J., STILPHEN, C. A., SMITH, S. M. E., OCASIO, H., BERMINGHAM, B., DARJI, S., GUHA, A., PATEL, R., GEURTS, A. M., JACOB, H. J., LAMBERT, N. A. & O'CONNOR, P. M. 2014. HV1 Acts as a Sodium Sensor and Promotes Superoxide Production in Medullary Thick Ascending Limb of Dahl Salt-Sensitive Rats. *Hypertension,* 64**,** 541-50.

JOFFRES, M., FALASCHETTI, E., GILLESPIE, C., ROBITAILLE, C., LOUSTALOT, F., POULTER, N., MCALISTER, F. A., JOHANSEN, H., BACLIC, O. & CAMPBELL, N. 2013. Hypertension prevalence, awareness, treatment and control in national surveys from England, the USA and Canada, and correlation with stroke and ischaemic heart disease mortality: a cross-sectional study. *Bmj Open,* 3.

JOHNSON, T., GAUNT, T. R., NEWHOUSE, S. J., PADMANABHAN, S., TOMASZEWSKI, M., KUMARI, M., MORRIS, R. W., TZOULAKI, I., O'BRIEN, E. T., POULTER, N. R., SEVER, P., SHIELDS, D. C., THOM, S., WANNAMETHEE, S. G., WHINCUP, P. H., BROWN, M. J., CONNELL, J. M., DOBSON, R. J., HOWARD, P. J., MEIN, C. A., ONIPINLA, A., SHAW-HAWKINS, S., ZHANG, Y., SMITH, G. D., DAY, I. N. M., LAWLOR, D. A., GOODALL, A. H., FOWKES, F. G., ABECASIS, G. R., ELLIOTT, P., GATEVA, V., BRAUND, P. S., BURTON, P. R., NELSON, C. P., TOBIN, M. D., VAN DER HARST, P., GLORIOSO, N., NEUVRITH, H., SALVI, E., STAESSEN, J. A., STUCCHI, A., DEVOS, N., JEUNEMAITRE, X., PLOUIN, P. F., TICHET, J., JUHANSON, P., ORG, E., PUTKU, M., SOBER, S., VELDRE, G., VIIGIMAA, M., LEVINSSON, A., ROSENGREN, A., THELLE, D. S., HASTIE, C. E., HEDNER, T., LEE, W. K., MELANDER, O., WAHLSTRAND, B., HARDY, R., WONG, A., COOPER, J. A., PALMEN, J., CHEN, L., STEWART, A. F. R., WELLS, G. A., WESTRA, H. J., WOLFS, M. G. M., CLARKE, R., FRANZOSI, M. G., GOEL, A., HAMSTEN, A., LATHROP, M., PEDEN, J. F., SEEDORF, U., WATKINS, H., OUWEHAND, W. H., SAMBROOK, J., STEPHENS, J., CASAS, J. P., DRENOS, F., HOLMES, M. V., KIVIMAKI, M., SHAH, S., SHAH, T., TALMUD, P. J., WHITTAKER, J., WALLACE, C., DELLES, C., LAAN, M., KUH, D., HUMPHRIES, S. E., NYBERG, F., CUSI, D., ROBERTS, R., NEWTON-CHEH, C., FRANKE, L., STANTON, A. V., DOMINICZAK, A. F., FARRALL, M., et al. 2011. Blood Pressure Loci Identified with a Gene-Centric Array. *American Journal of Human Genetics,* 89**,** 688-700.

JORDAN, J., TOKA, H. R., HEUSSER, K., TOKA, O., SHANNON, J. R., TANK, J., DIEDRICH, A., STABROTH, C., STOFFELS, M., NARAGHI, R., OELKERS, W., SCHUSTER, H., SCHOBEL, H. P., HALLER, H. & LUFT, F. C. 2000. Severely impaired baroreflex-buffering in patients with monogenic hypertension and neurovascular contact. *Circulation,* 102**,** 2611-2618.

KAHAN, T. & BERGFELDT, L. 2005. Left ventricular hypertrophy in hypertension: Its arrhythmogenic potential. *Heart,* 91**,** 250-256.

KAKAZU, Y., KOH, J.-Y., HO, K. W. D., GONZALEZ-ALEGRE, P. & HARATA, N. C. 2012. Synaptic vesicle recycling is enhanced by torsinA that harbors the DYT1 dystonia mutation. *Synapse,* 66**,** 453-464.

KAMM, C., ASMUS, F., MUELLER, J., MAYER, P., SHARMA, M., MULLER, U. J., BECKERT, S., EHLING, R., ILLIG, T., WICHMANN, H. E., POEWE, W.,

MUELLER, J. C. & GASSER, T. 2006. Strong genetic evidence for association of TOR1A/TOR1B with idiopathic dystonia. *Neurology,* 67**,** 1857-1859.

KAPETANOVIC, R., FAIRBAIRN, L., DOWNING, A., BERALDI, D., SESTER, D. P., FREEMAN, T. C., TUGGLE, C. K., ARCHIBALD, A. L. & HUME, D. A. 2013. The impact of breed and tissue compartment on the response of pig macrophages to lipopolysaccharide. *Bmc Genomics,* 14.

KATO, N., TAKEUCHI, F., TABARA, Y., KELLY, T. N., GO, M. J., SIM, X., TAY, W. T., CHEN, C.-H., ZHANG, Y., YAMAMOTO, K., KATSUYA, T., YOKOTA, M., KIM, Y. J., ONG, R. T. H., NABIKA, T., GU, D., CHANG, L.-C., KOKUBO, Y., HUANG, W., OHNAKA, K., YAMORI, Y., NAKASHIMA, E., JAQUISH, C. E., LEE, J.-Y., SEIELSTAD, M., ISONO, M., HIXSON, J. E., CHEN, Y.-T., MIKI, T., ZHOU, X., SUGIYAMA, T., JEON, J.-P., LIU, J. J., TAKAYANAGI, R., KIM, S. S., AUNG, T., SUNG, Y. J., ZHANG, X., WONG, T. Y., HAN, B.-G., KOBAYASHI, S., OGIHARA, T., ZHU, D., IWAI, N., WU, J.-Y., TEO, Y. Y., TAI, E. S., CHO, Y. S. & HE, J. 2011. Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians. *Nature Genetics,* 43**,** 530-U57.

KEANE, T. M., GOODSTADT, L., DANECEK, P., WHITE, M. A., WONG, K., YALCIN, B., HEGER, A., AGAM, A., SLATER, G., GOODSON, M., FURLOTTE, N. A., ESKIN, E., NELLAKER, C., WHITLEY, H., CLEAK, J., JANOWITZ, D., HERNANDEZ-PLIEGO, P., EDWARDS, A., BELGARD, T. G., OLIVER, P. L., MCINTYRE, R. E., BHOMRA, A., NICOD, J., GAN, X., YUAN, W., VAN DER WEYDEN, L., STEWARD, C. A., BALA, S., STALKER, J., MOTT, R., DURBIN, R., JACKSON, I. J., CZECHANSKI, A., GUERRA-ASSUNCAO, J. A., DONAHUE, L. R., REINHOLDT, L. G., PAYSEUR, B. A., PONTING, C. P., BIRNEY, E., FLINT, J. & ADAMS, D. J. 2011. Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature,* 477**,** 289-294.

KEARNEY, P. M., WHELTON, M., REYNOLDS, K., MUNTNER, P., WHELTON, P. K. & HE, J. 2005. Global burden of hypertension: analysis of worldwide data. *Lancet,* 365**,** 217-223.

KERR, S., BROSNAN, M. J., MCINTYRE, M., REID, J. L., DOMINICZAK, A. F. & HAMILTON, C. A. 1999. Superoxide anion production is increased in a model of genetic hypertension - Role of the endothelium. *Hypertension,* 33**,** 1353-1358.

KOH-TAN, H. H. C., GRAHAM, D., HAMILTON, C. A., NICOLL, G., FIELDS, L., MCBRIDE, M. W., YOUNG, B. & DOMINICZAK, A. F. 2009. Renal and vascular glutathione S-transferase mu is not affected by pharmacological intervention to reduce systolic blood pressure. *Journal of Hypertension,* 27**,** 1575-1584.

KOHAN, D. E., ROSSI, N. F., INSCHO, E. W. & POLLOCK, D. M. 2011. Regulation of Blood Pressure and Salt Homeostasis by Endothelin. *Physiological Reviews,* 91**,** 1-77.

KREN, V., QI, N. N., KRENOVA, D., ZIDEK, V., SLADKA, M., JACHYMOVA, M., MIKOVA, B., HORKY, K., BONNE, A., VAN LITH, H. A., VAN ZUTPHEN, B. F. M., LAU, Y. F. C., PRAVENEC, M. & ST LEZIN, E. 2001. Y-chromosome transfer induces changes in blood pressure and blood lipids in SHR. *Hypertension,* 37**,** 1147-1152.

KRZYWINSKI, M., SCHEIN, J., BIROL, I., CONNORS, J., GASCOYNE, R., HORSMAN, D., JONES, S. J. & MARRA, M. A. 2009. Circos: An information aesthetic for comparative genomics. *Genome Research,* 19**,** 1639-1645.

KUMAR, P., HENIKOFF, S. & NG, P. C. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols,* 4**,** 1073-1082.

LANGLEY, S. R., BOTTOLO, L., KUNES, J., ZICHA, J., ZIDEK, V., HUBNER, N., COOK, S. A., PRAVENEC, M., AITMAN, T. J. & PETRETTO, E. 2013. Systems-level approaches reveal conservation of trans-regulated genes in the rat and genetic determinants of blood pressure in humans. *Cardiovascular Research,* 97**,** 653-665.

LAPPALAINEN, T., SAMMETH, M., FRIEDLAENDER, M. R., T HOEN, P. A. C., MONLONG, J., RIVAS, M. A., GONZALEZ-PORTA, M., KURBATOVA, N., GRIEBEL, T., FERREIRA, P. G., BARANN, M., WIELAND, T., GREGER, L., VAN ITERSON, M., ALMLOEF, J., RIBECA, P., PULYAKHINA, I., ESSER, D., GIGER, T., TIKHONOV, A., SULTAN, M., BERTIER, G., MACARTHUR, D. G., LEK, M., LIZANO, E., BUERMANS, H. P. J., PADIOLEAU, I., SCHWARZMAYR, T., KARLBERG, O., ONGEN, H., KILPINEN, H., BELTRAN, S., GUT, M., KAHLEM, K., AMSTISLAVSKIY, V., STEGLE, O., PIRINEN, M., MONTGOMERY, S. B., DONNELLY, P., MCCARTHY, M. I., FLICEK, P., STROM, T. M., LEHRACH, H., SCHREIBER, S., SUDBRAK, R., CARRACEDO, A., ANTONARAKIS, S. E., HAESLER, R., SYVAENEN, A.-C., VAN OMMEN, G.-J., BRAZMA, A., MEITINGER, T., ROSENSTIEL, P., GUIGO, R., GUT, I. G., ESTIVILL, X., DERMITZAKIS, E. T. & GEUVADIS, C. 2013. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature,* 501**,** 506-511.

LARKIN, J. E., FRANK, B. C., GAVRAS, H., SULTANA, R. & QUACKENBUSH, J. 2005. Independence and reproducibility across microarray platforms. *Nature Methods,* 2**,** 337-343.

LAWES, C. M. M., VANDER HOORN, S., RODGERS, A. & INT SOC, H. 2008. Global burden of blood-pressure-related disease, 2001. *Lancet,* 371**,** 1513-1518.

LEE, J. H., O'KEEFE, J. H., BELL, D., HENSRUD, D. D. & HOLICK, M. F. 2008. Vitamin D Deficiency An Important, Common, and Easily Treatable Cardiovascular Risk Factor? *Journal of the American College of Cardiology,* 52**,** 1949-1956.

LEFEVRE, S., SLIWA, D., RUSTIN, P., CAMADRO, J.-M. & SANTOS, R. 2012. Oxidative stress induces mitochondrial fragmentation in frataxin-deficient cells. *Biochemical and Biophysical Research Communications,* 418**,** 336-341.

LEPRETRE, F., LINTON, K. J., LACQUEMANT, C., VATIN, V., SAMSON, C., DINA, C., CHIKRI, M., ALI, S., SCHERER, P., SERON, K., VASSEUR, F., AITMAN, T. & FROGUEL, P. 2004. Genetic study of the CD36 gene In a French diabetic population. *Diabetes & Metabolism,* 30**,** 459-463.

LEVI, M., MOONS, L., BOUCHE, A., SHAPIRO, S. D., COLLEN, D. & CARMELIET, P. 2001. Deficiency of urokinase-type plasminogen activator-mediated plasmin generation impairs vascular remodeling during hypoxia-induced pulmonary hypertension in mice. *Circulation,* 103**,** 2014-2020.

LEVY, D., EHRET, G. B., RICE, K., VERWOERT, G. C., LAUNER, L. J., DEHGHAN, A., GLAZER, N. L., MORRISON, A. C., JOHNSON, A. D., ASPELUND, T., AULCHENKO, Y., LUMLEY, T., KOTTGEN, A., VASAN, R. S., RIVADENEIRA, F., EIRIKSDOTTIR, G., GUO, X. Q., ARKING, D. E., MITCHELL, G. F., MATTACE-RASO, F. U. S., SMITH, A. V., TAYLOR, K., SCHARPF, R. B., HWANG, S. J., SIJBRANDS, E. J. G., BIS, J., HARRIS, T. B., GANESH, S. K., O'DONNELL, C. J., HOFMAN, A., ROTTER, J. I., CORESH, J., BENJAMIN, E. J., UITTERLINDEN, A. G., HEISS, G., FOX, C. S., WITTEMAN, J. C. M., BOERWINKLE, E., WANG, T. J., GUDNASON, V., LARSON, M. G., CHAKRAVARTI, A., PSATY, B. M. & VAN DUIJN, C. M. 2009. Genome-wide association study of blood pressure and hypertension. *Nature Genetics,* 41**,** 677-687.

LI, H. & DURBIN, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics,* 25**,** 1754-1760.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & GENOME PROJECT DATA, P. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25**,** 2078-2079.

LIFTON, R. P. 1996. Molecular genetics of human blood pressure variation. *Science,* 272**,** 676-680.

LIFTON, R. P., DLUHY, R. G., POWERS, M., RICH, G. M., COOK, S., ULICK, S. & LALOUEL, J. M. 1992. A chimeric 11-beta-hydroxylase aldosterone synthase gene causes Glucocorticoid-remediable aldosteronism and human hypertension. *Nature,* 355**,** 262-265.

LIVAK, K. J. & SCHMITTGEN, T. D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. *Methods,* 25**,** 402-408.

LOMBARD, Z., TIFFIN, N., HOFMANN, O., BAJIC, V. B., HIDE, W. & RAMSAY, M. 2007. Computational selection and prioritization of candidate genes for Fetal Alcohol Syndrome. *Bmc Genomics,* 8.

MACLELLAN, W. R., WANG, Y. & LUSIS, A. J. 2012. Systems-based approaches to cardiovascular disease. *Nature Reviews Cardiology,* 9**,** 172-184.

MAGUIRE, J. L., DEVEBER, G. & PARKIN, P. C. 2007. Association between iron-deficiency anemia and stroke in young children. *Pediatrics,* 120**,** 1053-1057.

MARATOU, K., BEHMOARAS, J., FEWINGS, C., SRIVASTAVA, P., D'SOUZA, Z., SMITH, J., GAME, L., COOK, T. & AITMAN, T. 2011. Characterization of the macrophage transcriptome in glomerulonephritis-susceptible and -resistant rat strains. *Genes and Immunity,* 12**,** 78-89.

MARON, B. J. 2002. Hypertrophic cardiomyopathy - A systematic review. *Jama-Journal of the American Medical Association,* 287**,** 1308-1320.

MARQUES, F. Z., CAMPAIN, A. E., TOMASZEWSKI, M., ZUKOWSKA-SZCZECHOWSKA, E., YANG, Y. H. J., CHARCHAR, F. J. & MORRIS, B. J. 2011. Gene Expression Profiling Reveals Renin mRNA Overexpression in Human Hypertensive Kidneys and a Role for MicroRNAs. *Hypertension,* 58**,** 1093-U337.

MARQUES, F. Z., CAMPAIN, A. E., YANG, Y. H. J. & MORRIS, B. J. 2010. Meta-Analysis of Genome-Wide Gene Expression Differences in Onset and Maintenance Phases of Genetic Hypertension. *Hypertension,* 56**,** 319-U385.

MARTEAU, J. B., ZAIOU, M., SIEST, G. & VISVIKIS-SIEST, S. 2005. Genetic determinants of blood pressure regulation. *Journal of Hypertension,* 23**,** 2127-2143.

MARUHASHI, T., SOGA, J., FUJIMURA, N., IDEI, N., MIKAMI, S., IWAMOTO, Y., KAJIKAWA, M., MATSUMOTO, T., HIDAKA, T., KIHARA, Y., CHAYAMA, K., NOMA, K., NAKASHIMA, A., GOTO, C. & HIGASHI, Y. 2013. Nitroglycerine-Induced Vasodilation for Assessment of Vascular Function A Comparison With Flow-Mediated Vasodilation. *Arteriosclerosis Thrombosis and Vascular Biology,* 33**,** 1401-+.

MCBRIDE, M. W., BROSNAN, M. J., MATHERS, J., MCLELLAN, L. I., MILLER, W. H., GRAHAM, D., HANLON, N., HAMILTON, C. A., POLKE, J. M., LEE, W. K. & DOMINICZAK, A. F. 2005. Reduction of Gstm1 expression in the stroke-prone spontaneously hypertension rat contributes to increased oxidative stress. *Hypertension,* 45**,** 786-792.

MCBRIDE, M. W., CARR, F. J., GRAHAM, D., ANDERSON, N. H., CLARK, J. S., LEE, W. K., CHARCHAR, F. J., BROSNAN, M. J. & DOMINICZAK, A. F. 2003. Microarray analysis of rat chromosome 2 congenic strains. *Hypertension,*

41, 847-853.
MCBRIDE, M. W., CHARCHAR, F. J., GRAHAM, D., MILLER, W. H., STRAHORN, P., CARR, F. J. & DOMINICZAK, A. F. 2004. Functional genomics in rodent models of hypertension. *Journal of Physiology-London,* 554, 56-63.
MCBRIDE, M. W., GRAHAM, D., DELLES, C. & DOMINICZAK, A. F. 2006. Functional genomics in hypertension. *Current Opinion in Nephrology and Hypertension,* 15, 145-151.
MCKENNA, A., HANNA, M., BANKS, E., SIVACHENKO, A., CIBULSKIS, K., KERNYTSKY, A., GARIMELLA, K., ALTSHULER, D., GABRIEL, S., DALY, M. & DEPRISTO, M. A. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research,* 20, 1297-1303.
MCLAREN, W., PRITCHARD, B., RIOS, D., CHEN, Y., FLICEK, P. & CUNNINGHAM, F. 2010. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics,* 26, 2069-2070.
MENG, X. M., KLEMENT, J. F., LEPERI, D. A., BIRK, D. E., SASAKI, T., TIMPL, R., UITTO, J. & PULKKINEN, L. 2003. Targeted inactivation of murine laminin gamma 2-chain gene recapitulates human junctional epidermolysis bullosa. *Journal of Investigative Dermatology,* 121, 720-731.
MENGHINI, R., CASAGRANDE, V., CARDELLINI, M., MARTELLI, E., TERRINONI, A., AMATI, F., VASA-NICOTERA, M., IPPOLITI, A., NOVELLI, G., MELINO, G., LAURO, R. & FEDERICI, M. 2009. MicroRNA 217 Modulates Endothelial Cell Senescence via Silent Information Regulator 1. *Circulation,* 120, 1524-U102.
MENSAH, G. A. 2007. Healthy endothelium: The scientific basis for cardiovascular health promotion and chronic disease prevention. *Vascular Pharmacology,* 46, 310-314.
METZKER, M. L. 2010. APPLICATIONS OF NEXT-GENERATION SEQUENCING Sequencing technologies - the next generation. *Nature Reviews Genetics,* 11, 31-46.
MIYAOKA, K., KUWASAKO, T., HIRANO, K., NOZAKI, S., YAMASHITA, S. & MATSUZAWA, Y. 2001. CD36 deficiency associated with insulin resistance. *Lancet,* 357, 686-687.
MOLNAR, M. Z., KALANTAR-ZADEH, K., LOTT, E. H., LU, J. L., MALAKAUSKAS, S. M., MA, J. Z., QUARLES, D. L. & KOVESDY, C. P. 2014. Angiotensin-Converting Enzyme Inhibitor, Angiotensin Receptor Blocker Use, and Mortality in Patients With Chronic Kidney Disease. *Journal of the American College of Cardiology,* 63, 650-658.
MONGEAU, J. G., BIRON, P. & SING, C. F. 1986. THE INFLUENCE OF GENETICS AND HOUSEHOLD ENVIRONMENT UPON THE VARIABILITY OF NORMAL BLOOD-PRESSURE - THE MONTREAL ADOPTION SURVEY. *Clinical and Experimental Hypertension Part a-Theory and Practice,* 8, 653-660.
MONTEZANO, A. C. & TOUYZ, R. M. 2012. Reactive Oxygen Species and Endothelial Function - Role of Nitric Oxide Synthase Uncoupling and Nox Family Nicotinamide Adenine Dinucleotide Phosphate Oxidases. *Basic & Clinical Pharmacology & Toxicology,* 110, 87-94.
MONTI, J., FISCHER, J., PASKAS, S., HEINIG, M., SCHULZ, H., GOESELE, C., HEUSER, A., FISCHER, R., SCHMIDT, C., SCHIRDEWAN, A., GROSS, V., HUMMEL, O., MAATZ, H., PATONE, G., SAAR, K., VINGRON, M., WELDON, S. M., LINDPAINTNER, K., HAMMOCK, B. D., ROHDE, K., DIETZ, R., COOK, S. A., SCHUNCK, W.-H., LUFT, F. C. & HUBNER, N. 2008. Soluble epoxide hydrolase is a susceptibility factor for heart failure in a rat model of human disease. *Nature Genetics,* 40, 529-537.

MORENO-SMITH, M., HALDER, J. B., MELTZER, P. S., GONDA, T. A., MANGALA, L. S., RUPAIMOOLE, R., LU, C., NAGARAJA, A. S., GHARPURE, K. M., KANG, Y., RODRIGUEZ-AGUAYO, C., VIVAS-MEJIA, P. E., ZAND, B., SCHMANDT, R., WANG, H., LANGLEY, R. R., JENNINGS, N. B., IVAN, C., COFFIN, J. E., ARMAIZ, G. N., BOTTSFORD-MILLER, J., KIM, S. B., HALLECK, M. S., HENDRIX, M. J. C., BORNMAN, W., BAR-ELI, M., LEE, J.-S., SIDDIK, Z. H., LOPEZ-BERESTEIN, G. & SOOD, A. K. 2013. ATP11B mediates platinum resistance in ovarian cancer. *Journal of Clinical Investigation,* 123**,** 2119-2130.

MOUJAHIDINE, M., LAMBERT, R., DUTIL, J., PALIJAN, A., SIVO, Z., ARIYARAJAH, A. & DENG, A. Y. 2004. Combining congenic coverage with gene profiling in search of candidates for blood pressure quantitative trait loci in Dahl rats. *Hypertension Research,* 27**,** 203-212.

MUNROE, P. B., WALLACE, C., XUE, M.-Z., MARCANO, A. C. B., DOBSON, R. J., ONIPINLA, A. K., BURKE, B., GUNGADOO, J., NEWHOUSE, S. J., PEMBROKE, J., BROWN, M., DOMINICZAK, A. F., SAMANI, N. J., LATHROP, M., CONNELL, J., WEBSTER, J., CLAYTON, D., FARRALL, M., MEIN, C. A., CAULFIELD, M. & MED RES COUNCIL BRITISH, G. 2006. Increased support for linkage of a novel locus on chromosome 5q13 for essential hypertension in the British genetics of hypertension study. *Hypertension,* 48**,** 105-111.

MUST, A., SPADANO, J., COAKLEY, E. H., FIELD, A. E., COLDITZ, G. & DIETZ, W. H. 1999. The disease burden associated with overweight and obesity. *Jama-Journal of the American Medical Association,* 282**,** 1523-1529.

NABIKA, T., OHARA, H., KATO, N. & ISOMURA, M. 2012. The stroke-prone spontaneously hypertensive rat: still a useful model for post-GWAS genetic studies? *Hypertension Research,* 35**,** 477-484.

NAGAI, H., YASUDA, S., OHBA, Y., FUKUDA, M. & NAKAMURA, T. 2013. All members of the EPI64 subfamily of TBC/RabGAPs also have GAP activities towards Ras. *Journal of Biochemistry,* 153**,** 283-288.

NEGRIN, C. D., MCBRIDE, M. W., CARSWELL, H. V. O., GRAHAM, D., CARR, F. J., CLARK, J. S., JEFFS, B., ANDERSON, N. H., MACRAE, I. M. & DOMINICZAK, A. F. 2001. Reciprocal consomic strains to evaluate Y chromosome effects. *Hypertension,* 37**,** 391-397.

NEWTON-CHEH, C., JOHNSON, T., GATEVA, V., TOBIN, M. D., BOCHUD, M., COIN, L., NAJJAR, S. S., ZHAO, J. H., HEATH, S. C., EYHERAMENDY, S., PAPADAKIS, K., VOIGHT, B. F., SCOTT, L. J., ZHANG, F., FARRALL, M., TANAKA, T., WALLACE, C., CHAMBERS, J. C., KHAW, K. T., NILSSON, P., VAN DER HARST, P., POLIDORO, S., GROBBEE, D. E., ONLAND-MORET, N. C., BOTS, M. L., WAIN, L. V., ELLIOTT, K. S., TEUMER, A., LUAN, J., LUCAS, G., KUUSISTO, J., BURTON, P. R., HADLEY, D., MCARDLE, W. L., BROWN, M., DOMINICZAK, A., NEWHOUSE, S. J., SAMANI, N. J., WEBSTER, J., ZEGGINI, E., BECKMANN, J. S., BERGMANN, S., LIM, N., SONG, K., VOLLENWEIDER, P., WAEBER, G., WATERWORTH, D. M., YUAN, X., GROOP, L., ORHO-MELANDER, M., ALLIONE, A., DI GREGORIO, A., GUARRERA, S., PANICO, S., RICCERI, F., ROMANAZZI, V., SACERDOTE, C., VINEIS, P., BARROSO, I., SANDHU, M. S., LUBEN, R. N., CRAWFORD, G. J., JOUSILAHTI, P., PEROLA, M., BOEHNKE, M., BONNYCASTLE, L. L., COLLINS, F. S., JACKSON, A. U., MOHLKE, K. L., STRINGHAM, H. M., VALLE, T. T., WILLER, C. J., BERGMAN, R. N., MORKEN, M. A., DORING, A., GIEGER, C., ILLIG, T., MEITINGER, T., ORG, E., PFEUFER, A., WICHMANN, H. E., KATHIRESAN, S., MARRUGAT, J., O'DONNELL, C. J., SCHWARTZ, S. M., SISCOVICK, D. S., SUBIRANA, I., FREIMER, N. B., HARTIKAINEN, A. L., MCCARTHY, M. I., O'REILLY, P. F., PELTONEN, L., POUTA, A., DE JONG, P. E., SNIEDER, H., VAN GILST, W. H.,

CLARKE, R., GOEL, A., HAMSTEN, A., PEDEN, J. F., et al. 2009. Genome-wide association study identifies eight loci associated with blood pressure. *Nature Genetics,* 41**,** 666-676.

NICE. 2013. Hypertensive guideline. Available at http://www.nice.org.uk/guidance/QS28

NICHOLS, M., TOWNSEND, N., SCARBOROUGH, P. & RAYNER, M. 2013. European Cardiovascular Disease Statistics 4th edition 2012: EuroHeart II. *European heart journal,* 34**,** 3007-3007.

NICOLAE, D. L., GAMAZON, E., ZHANG, W., DUAN, S., DOLAN, M. E. & COX, N. J. 2010. Trait-Associated SNPs Are More Likely to Be eQTLs: Annotation to Enhance Discovery from GWAS. *Plos Genetics,* 6.

OHNO, A., SAITO, Y., HANIOKA, N., JINNO, H., SAEKI, M., ANDO, M., OZAWA, S. & SAWADA, J. 2004. Involvement of human hepatic UGT1A1, UGT2B4, and UGT2B7 in the glucuronidation of carvedilol. *Drug Metabolism and Disposition,* 32**,** 235-239.

OKAMOTO, K., HAZAMA, F., YAMORI, Y., HAEBARA, H. & NAGAOKA, A. 1975. Pathogenesis and prevention of stroke in spontaneously hypertensive rats. *Clinical science and molecular medicine. Supplement,* 2**,** 161s-163s.

OLSON, E. 2014. Characterization of the GSTM1 transgenic SHRSP rat. *University of Glasgow* PhD thesis.

PADMANABHAN, S., MELANDER, O., HASTIE, C., MENNI, C., DELLES, C., CONNELL, J. M. & DOMINICZAK, A. F. 2008. Hypertension and genome-wide association studies: combining high fidelity phenotyping and hypercontrols. *Journal of Hypertension,* 26**,** 1275-1281.

PADMANABHAN, S., MELANDER, O., JOHNSON, T., DI BLASIO, A. M., LEE, W. K., GENTILINI, D., HASTIE, C. E., MENNI, C., MONTI, M. C., DELLES, C., LAING, S., CORSO, B., NAVIS, G., KWAKERNAAK, A. J., VAN DER HARST, P., BOCHUD, M., MAILLARD, M., BURNIER, M., HEDNER, T., KJELDSEN, S., WAHLSTRAND, B., SJOGREN, M., FAVA, C., MONTAGNANA, M., DANESE, E., TORFFVIT, O., HEDBLAD, B., SNIEDER, H., CONNELL, J. M. C., BROWN, M., SAMANI, N. J., FARRALL, M., CESANA, G., MANCIA, G., SIGNORINI, S., GRASSI, G., EYHERAMENDY, S., WICHMANN, H. E., LAAN, M., STRACHAN, D. P., SEVER, P., SHIELDS, D. C., STANTON, A., VOLLENWEIDER, P., TEUMER, A., VOLZKE, H., RETTIG, R., NEWTON-CHEH, C., ARORA, P., ZHANG, F., SORANZO, N., SPECTOR, T. D., LUCAS, G., KATHIRESAN, S., SISCOVICK, D. S., LUAN, J. A., LOOS, R. J. F., WAREHAM, N. J., PENNINX, B. W., NOLTE, I. M., MCBRIDE, M., MILLER, W. H., NICKLIN, S. A., BAKER, A. H., GRAHAM, D., MCDONALD, R. A., PELL, J. P., SATTAR, N., WELSH, P., MUNROE, P., CAULFIELD, M. J., ZANCHETTI, A., DOMINICZAK, A. F. & GLOBAL, B. C. 2010. Genome-Wide Association Study of Blood Pressure Extremes Identifies Variant near UMOD Associated with Hypertension. *Plos Genetics,* 6**,** 11.

PADMANABHAN, S., NEWTON-CHEH, C. & DOMINICZAK, A. F. 2012. Genetic basis of blood pressure and hypertension. *Trends in Genetics,* 28**,** 397-408.

PANGRAZIO, A., FASTH, A., SBARDELLATI, A., ORCHARD, P. J., KASOW, K. A., RAZA, J., ALBAYRAK, C., ALBAYRAK, D., VANAKKER, O. M., DE MOERLOOSE, B., VELLODI, A., NOTARANGELO, L. D., SCHLACK, C., STRAUSS, G., KUEHL, J.-S., CALDANA, E., LO IACONO, N., SUSANI, L., KORNAK, U., SCHULZ, A., VEZZONI, P., VILLA, A. & SOBACCHI, C. 2013. SNX10 Mutations Define a Subgroup of Human Autosomal Recessive Osteopetrosis With Variable Clinical Severity. *Journal of Bone and Mineral Research,* 28**,** 1041-1049.

PAPADAKI, C., ALEXIOU, M., CECENA, G., VERYKOKAKIS, M., BILITOU, A., CROSS, J. C., OSHIMA, R. G. & MAVROTHALASSITIS, G. 2007. Transcriptional

repressor Erf determines extraembryonic ectoderm differentiation. *Molecular and Cellular Biology,* 27**,** 5201-5213.

PASDAR, A., YADEGARFAR, G., CUMMING, A., WHALLEY, L., ST CLAIR, D. & MACLEOD, M.-J. 2007. The effect of ABCAI gene polymorphisms on ischaemic stroke risk and relationship with lipid profile. *Bmc Medical Genetics,* 8.

PAWITAN, Y., MICHIELS, S., KOSCIELNY, S., GUSNANTO, A. & PLONER, A. 2005. False discovery rate, sensitivity and sample size for microarray studies. *Bioinformatics,* 21**,** 3017-3024.

PEREDA, A., GARIN, I., GARCIA-BARCINA, M., GENER, B., BERISTAIN, E., MIREN IBANEZ, A. & PEREZ DE NANCLARES, G. 2013. Brachydactyly E: isolated or as a feature of a syndrome. *Orphanet Journal of Rare Diseases,* 8.

PETRETTO, E., SARWAR, R., GRIEVE, I., LU, H., KUMARAN, M. K., MUCKETT, P. J., MANGION, J., SCHROEN, B., BENSON, M., PUNJABI, P. P., PRASAD, S. K., PENNELL, D. J., KIESEWETTER, C., TASHEVA, E. S., CORPUZ, L. M., WEBB, M. D., CONRAD, G. W., KURTZ, T. W., KREN, V., FISCHER, J., HUBNER, N., PINTO, Y. M., PRAVENEC, M., AITMAN, T. J. & COOK, S. A. 2008. Integrated genomic approaches implicate osteoglycin (Ogn) in the regulation of left ventricular mass. *Nature Genetics,* 40**,** 546-552.

PRAVENEC, M., CHURCHILL, P. C., CHURCHILL, M. C., VIKLICKY, O., KAZDOVA, L., AITMAN, T. J., PETRETTO, E., HUBNER, N., WALLACE, C. A., ZIMDAHL, H., ZIDEK, V., LANDA, V., DUNBAR, J., BIDANI, A., GRIFFIN, K., QI, N., MAXOVA, M., KREN, V., MLEJNEK, P., WANG, J. & KURTZ, T. W. 2008a. Identification of renal Cd36 as a determinant of blood pressure and risk for hypertension. *Nature Genetics,* 40**,** 952-954.

PRAVENEC, M., KAZDOVA, L., LANDA, V., ZIDEK, V., MLEJNEK, P., SIMAKOVA, M., JANSA, P., FOREJT, J., KREN, V., KRENOVA, D., QI, N., WANG, J.-M., CHAN, D., AITMAN, T. J. & KURTZ, T. W. 2008b. Identification of mutated Srebf1 as a QTL influencing risk for hepatic steatosis in the spontaneously hypertensive rat. *Hypertension,* 51**,** 148-153.

PRAVENEC, M., KREN, V., LANDA, V., MLEJNEK, P., MUSILOVA, A., SILHAVY, J., SIMAKOVA, M. & ZIDEK, V. 2014. Recent Progress in the Genetics of Spontaneously Hypertensive Rats. *Physiological Research,* 63**,** S1-S8.

PRESKILL, C. & WEIDHAAS, J. B. 2013. SNPs in MicroRNA Binding Sites as Prognostic and Predictive Cancer Biomarkers. *Critical Reviews in Oncogenesis,* 18**,** 327-340.

RABINOVICH, B. A., KETCHEM, R. R., WOLFSON, M., GOLDSTEIN, L., SKELLY, M. & COSMAN, D. 2008. A role for the MHC class I-like Mill molecules in nutrient metabolism and wound healing. *Immunology and Cell Biology,* 86**,** 489-496.

RAHMOUNI, K., CORREIA, M. L. G., HAYNES, W. G. & MARK, A. L. 2005. Obesity-associated hypertension - New insights into mechanisms. *Hypertension,* 45**,** 9-14.

RAITOHARJU, E., SEPPALA, I., OKSALA, N., LYYTIKAINEN, L.-P., RAITAKARI, O., VIIKARI, J., ALA-KORPELA, M., SOININEN, P., KANGAS, A. J., WALDENBERGER, M., KLOPP, N., ILLIG, T., LEIVISKA, J., LOO, B.-M., HUTRI-KAHONEN, N., KAHONEN, M., LAAKSONEN, R. & LEHTIMAKI, T. 2014. Blood microRNA profile associates with the levels of serum lipids and metabolites associated with glucose metabolism and insulin resistance and pinpoints pathways underlying metabolic syndrome The cardiovascular risk in Young Finns Study. *Molecular and Cellular Endocrinology,* 391**,** 41-49.

RANGIANI, A., CAO, Z., SUN, Y., LU, Y., GAO, T., YUAN, B., RODGERS, A., QIN, C., KURO-O, M. & FENG, J. Q. 2012. Protective Roles of DMP1 in High

Phosphate Homeostasis. *Plos One*, 7.
RAPP, J. P. 2000. Genetic analysis of inherited hypertension in the rat. *Physiological Reviews*, 80, 135-172.
RE, R. N. 2004. Mechanisms of disease: local renin-angiotensin-aldosterone systems and the pathogenesis and treatment of cardiovascular disease. *Nature Clinical Practice Cardiovascular Medicine*, 1, 42-47.
REID, I. A. 1998. The renin-angiotensin system: Physiology, pathophysiology, and pharmacology. *Advances in Physiology Education*, 20, S236-S245.
REID, J. L. 1986. Alpha-adrenergic receptors and blood-pressure control. *American Journal of Cardiology*, 57, E6-E12.
RINGNER, M. 2008. What is principal component analysis? *Nature Biotechnology*, 26, 303-304.
ROBINSON, J. T., THORVALDSDOTTIR, H., WINCKLER, W., GUTTMAN, M., LANDER, E. S., GETZ, G. & MESIROV, J. P. 2011. Integrative genomics viewer. *Nature Biotechnology*, 29, 24-26.
ROCNIK, E. F., LIU, P., SATO, K., WALSH, K. & VAZIRI, C. 2006. The novel SPARC family member SMOC-2 potentiates angiogenic growth factor activity. *Journal of Biological Chemistry*, 281, 22855-22864.
ROGER, V. L., GO, A. S., LLOYD-JONES, D. M., BENJAMIN, E. J., BERRY, J. D., BORDEN, W. B., BRAVATA, D. M., DAI, S., FORD, E. S., FOX, C. S., FULLERTON, H. J., GILLESPIE, C., HAILPERN, S. M., HEIT, J. A., HOWARD, V. J., KISSELA, B. M., KITTNER, S. J., LACKLAND, D. T., LICHTMAN, J. H., LISABETH, L. D., MAKUC, D. M., MARCUS, G. M., MARELLI, A., MATCHAR, D. B., MOY, C. S., MOZAFFARIAN, D., MUSSOLINO, M. E., NICHOL, G., PAYNTER, N. P., SOLIMAN, E. Z., SORLIE, P. D., SOTOODEHNIA, N., TURAN, T. N., VIRANI, S. S., WONG, N. D., WOO, D., TURNER, M. B., AMER HEART ASSOC STAT, C. & STROKE STAT, S. 2012. Heart Disease and Stroke Statistics-2012 Update A Report From the American Heart Association. *Circulation*, 125, E2-E220.
RUPPERT, V. & MAISCH, B. 2003. Genetics of human hypertension. *Herz*, 28, 655-662.
SAAR, K., BECK, A., BIHOREAU, M.-T., BIRNEY, E., BROCKLEBANK, D., CHEN, Y., CUPPEN, E., DEMONCHY, S., DOPAZO, J., FLICEK, P., FOGLIO, M., FUJIYAMA, A., GUT, I. G., GAUGUIER, D., GUIGO, R., GURYEV, V., HEINIG, M., HUMMEL, O., JAHN, N., KLAGES, S., KREN, V., KUBE, M., KUHL, H., KURAMOTO, T., KUROKI, Y., LECHNER, D., LEE, Y.-A., LOPEZ-BIGAS, N., LATHROP, G. M., MASHIMO, T., MEDINA, I., MOTT, R., PATONE, G., PERRIER-CORNET, J.-A., PLATZER, M., PRAVENEC, M., REINHARDT, R., SAKAKI, Y., SCHILHABEL, M., SCHULZ, H., SERIKAWA, T., SHIKHAGAIE, M., TATSUMOTO, S., TAUDIEN, S., TOYODA, A., VOIGT, B., ZELENIKA, D., ZIMDAHL, H., HUBNER, N. & CONSORTIUM, S. 2008. SNP and haplotype mapping for genetic analysis in the rat. *Nature Genetics*, 40, 560-566.
SACKS, F. M., SVETKEY, L. P., VOLLMER, W. M., APPEL, L. J., BRAY, G. A., HARSHA, D., OBARZANEK, E., CONLIN, P. R., MILLER, E. R., SIMONS-MORTON, D. G., KARANJA, N., LIN, P. H. & GRP, D. A.-S. C. R. 2001. Effects on blood pressure of reduced dietary sodium and the dietary approaches to stop hypertension (DASH) diet. *New England Journal of Medicine*, 344, 3-10.
SAINO, M., MARUYAMA, T., SEKIYA, T., KAYAMA, T. & MURAKAMI, Y. 2004. Inhibition of angiogenesis in human glioma cell lines by antisense RNA from the soluble guanylate cyclase genes, GUCYIA3 and GUCYIB3. *Oncology Reports*, 12, 47-52.
SALVETTI, A., BROGI, G., DI LEGGE, V. & BERNINI, G. P. 1993. The inter-relationship between insulin resistance and hypertension. *Drugs*, 46 Suppl

2, 149-59.
SALVI, E., KUTALIK, Z., GLORIOSO, N., BENAGLIO, P., FRAU, F., KUZNETSOVA, T., ARIMA, H., HOGGART, C., TICHET, J., NIKITIN, Y. P., CONTI, C., SEIDLEROVA, J., TIKHONOFF, V., STOLARZ-SKRZYPEK, K., JOHNSON, T., DEVOS, N., ZAGATO, L., GUARRERA, S., ZANINELLO, R., CALABRIA, A., STANCANELLI, B., TROFFA, C., THIJS, L., RIZZI, F., SIMONOVA, G., LUPOLI, S., ARGIOLAS, G., BRAGA, D., D'ALESSIO, M. C., ORTU, M. F., RICCERI, F., MERCURIO, M., DESCOMBES, P., MARCONI, M., CHALMERS, J., HARRAP, S., FILIPOVSKY, J., BOCHUD, M., IACOVIELLO, L., ELLIS, J., STANTON, A. V., LAAN, M., PADMANABHAN, S., DOMINICZAK, A. F., SAMANI, N. J., MELANDER, O., JEUNEMAITRE, X., MANUNTA, P., SHABO, A., VINEIS, P., CAPPUCCIO, F. P., CAULFIELD, M. J., MATULLO, G., RIVOLTA, C., MUNROE, P. B., BARLASSINA, C., STAESSEN, J. A., BECKMANN, J. S. & CUSI, D. 2012. Genomewide Association Study Using a High-Density Single Nucleotide Polymorphism Array and Case-Control Design Identifies a Novel Essential Hypertension Susceptibility Locus in the Promoter Region of Endothelial NO Synthase. *Hypertension,* 59**,** 248-+.
SCHINDLER, C., BRAMLAGE, P., KIRCH, W. & FERRARIO, C. M. 2007. Role of the vasodilator peptide angiotensin-(1-7) in cardiovascular drug therapy. *Vascular health and risk management,* 3**,** 125-37.
SCHINKE, M., PAUL, A., BROWN, J., RIGGI, L. E., IZUMO, S., BARTUNEK, J., ALLEN, P. & TSUBAKIHARA, M. 2004. GSE1145: changes in cardiac transcription profiles brought about by heart failure. *Gene Expression Omnibus*.
SCHLAICH, M. P., KAYE, D. M., LAMBERT, E., SOMMERVILLE, M., SOCRATOUS, F. & ESLER, M. D. 2003. Relation between cardiac sympathetic activity and hypertensive left ventricular hypertrophy. *Circulation,* 108**,** 560-565.
SCHURMANN, C., HEIM, K., SCHILLERT, A., BLANKENBERG, S., CARSTENSEN, M., DOERR, M., ENDLICH, K., FELIX, S. B., GIEGER, C., GRALLERT, H., HERDER, C., HOFFMANN, W., HOMUTH, G., ILLIG, T., KRUPPA, J., MEITINGER, T., MUELLER, C., NAUCK, M., PETERS, A., RETTIG, R., RODEN, M., STRAUCH, K., VOELKER, U., VOELZKE, H., WAHL, S., WALLASCHOFSKI, H., WILD, P. S., ZELLER, T., TEUMER, A., PROKISCH, H. & ZIEGLER, A. 2012. Analyzing Illumina Gene Expression Microarray Data from Different Tissues: Methodological Aspects of Data Analysis in the MetaXpress Consortium. *Plos One,* 7.
SCHUSTER, H., WIENKER, T. F., BAHRING, S., BILGINTURAN, N., TOKA, H. R., NEITZEL, H., JESCHKE, E., TOKA, O., GILBERT, D., LOWE, A., OTT, J., HALLER, H. & LUFT, F. C. 1996. Severe autosomal dominant hypertension and brachydactyly in a unique Turkish kindred maps to human chromosome 12. *Nature Genetics,* 13**,** 98-100.
SESSO, H. D., STAMPFER, M. J., ROSNER, B., HENNEKENS, C. H., GAZIANO, J. M., MANSON, J. E. & GLYNN, R. J. 2000. Systolic and diastolic blood pressure, pulse pressure, and mean arterial pressure as predictors of cardiovascular disease risk in men. *Hypertension,* 36**,** 801-807.
SHANKS, M. E., DOWNES, S. M., COPLEY, R. R., LISE, S., BROXHOLME, J., HUDSPITH, K. A. Z., KWASNIEWSKA, A., DAVIES, W. I. L., HANKINS, M. W., PACKHAM, E. R., CLOUSTON, P., SELLER, A., WILKIE, A. O. M., TAYLOR, J. C., RAGOUSSIS, J. & NEMETH, A. H. 2013. Next-generation sequencing (NGS) as a diagnostic tool for retinal degeneration reveals a much higher detection rate in early-onset disease. *European Journal of Human Genetics,* 21**,** 274-280.
SHETTY, V. B., KIRALY-BORRI, C., LAMONT, P., BIKKER, H. & CHOONG, C. S. Y.

2014. NKX2-1 mutations in brain-lung-thyroid syndrome: a case series of four patients. *Journal of Pediatric Endocrinology & Metabolism,* 27**,** 373-378.

SHIMKETS, R. A., WARNOCK, D. G., BOSITIS, C. M., NELSONWILLIAMS, C., HANSSON, J. H., SCHAMBELAN, M., GILL, J. R., ULICK, S., MILORA, R. V., FINDLING, J. W., CANESSA, C. M., ROSSIER, B. C. & LIFTON, R. P. 1994. Liddles syndrome - heritable human hypertension caused by mutations in the beta-subunit of the epithelial sodium-channel. *Cell,* 79**,** 407-414.

SIMINO, J., RAO, D. C. & FREEDMAN, B. I. 2012. Novel findings and future directions on the genetics of hypertension. *Current Opinion in Nephrology and Hypertension,* 21**,** 500-507.

SIMONETTI, G. D., MOHAUPT, M. G. & BIANCHETTI, M. G. 2012. Monogenic forms of hypertension. *European Journal of Pediatrics,* 171**,** 1433-1439.

SIMONIS, M., ATANUR, S. S., LINSEN, S., GURYEV, V., RUZIUS, F.-P., GAME, L., LANSU, N., DE BRUIJN, E., VAN HEESCH, S., JONES, S. J. M., PRAVENEC, M., AITMAN, T. J. & CUPPEN, E. 2012. Genetic basis of transcriptome differences between the founder strains of the rat HXB/BXH recombinant inbred panel. *Genome Biology,* 13.

SNYDER, P. M., PRICE, M. P., MCDONALD, F. J., ADAMS, C. M., VOLK, K. A., ZEIHER, B. G., STOKES, J. B. & WELSH, M. J. 1995. Mechanism by which Liddles syndrome mutations increase activity of a human epithelial NA+ channel. *Cell,* 83**,** 969-978.

SODA, K., BALKIN, D. M., FERGUSON, S. M., PARADISE, S., MILOSEVIC, I., GIOVEDI, S., VOLPICELLI-DALEY, L., TIAN, X., WU, Y., MA, H., SON, S. H., ZHENG, R., MOECKEL, G., CREMONA, O., HOLZMAN, L. B., DE CAMILLI, P. & ISHIBE, S. 2012. Role of dynamin, synaptojanin, and endophilin in podocyte foot processes. *Journal of Clinical Investigation,* 122**,** 4401-4411.

STAESSEN, J. A., WANG, J. G., BIANCHI, G. & BIRKENHAGER, W. H. 2003. Essential hypertension. *Lancet,* 361**,** 1629-1641.

STAMLER, J., ROSE, G., STAMLER, R., ELLIOTT, P., DYER, A. & MARMOT, M. 1989. INTERSALT study findings - public-health and medical-care implications. *Hypertension,* 14**,** 570-577.

STAUSS, H. M. 2002. Baroreceptor reflex function. *American Journal of Physiology-Regulatory Integrative and Comparative Physiology,* 283**,** R284-R286.

STEINBERG, F., GALLON, M., WINFIELD, M., THOMAS, E. C., BELL, A. J., HEESOM, K. J., TAVARE, J. M. & CULLEN, P. J. 2013. A global analysis of SNX27-retromer assembly and cargo specificity reveals a function in glucose and metal ion transport. *Nature Cell Biology,* 15**,** 461-+.

STRANDBERG, T. E. & PITKALA, K. 2003. What is the most important component of blood pressure: systolic, diastolic or pulse pressure? *Current Opinion in Nephrology and Hypertension,* 12**,** 293-297.

STRANGER, B. E., FORREST, M. S., DUNNING, M., INGLE, C. E., BEAZLEY, C., THORNE, N., REDON, R., BIRD, C. P., DE GRASSI, A., LEE, C., TYLER-SMITH, C., CARTER, N., SCHERER, S. W., TAVARE, S., DELOUKAS, P., HURLES, M. E. & DERMITZAKIS, E. T. 2007. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science,* 315**,** 848-853.

SU, E. J., FREDRIKSSON, L., GEYER, M., FOLESTAD, E., CALE, J., ANDRAE, J., GAO, Y., PIETRAS, K., MANN, K., YEPES, M., STRICKLAND, D. K., BETSHOLTZ, C., ERIKSSON, U. & LAWRENCE, D. A. 2008. Activation of PDGF-CC by tissue plasminogen activator impairs blood-brain barrier integrity during ischemic stroke. *Nature Medicine,* 14**,** 731-737.

SZETO, C.-C., CHING-HA, K. B., KA-BIK, L., MAC-MOUNE, L. F., CHEUNG-LUNG, C. P., GANG, W., KAI-MING, C. & KAM-TAO, L. P. 2012. Micro-RNA expression in the urinary sediment of patients with chronic kidney diseases. *Disease Markers,* 33**,** 137-144.

TARCA, A. L., ROMERO, R. & DRAGHICI, S. 2006. Analysis of microarray experiments of gene expression profiling. *American Journal of Obstetrics and Gynecology,* 195**,** 373-388.

TIMPSON, N. J., HARBORD, R., SMITH, G. D., ZACHO, J., TYBJAERG-HANSEN, A. & NORDESTGAARD, B. G. 2009. Does Greater Adiposity Increase Blood Pressure and Hypertension Risk? Mendelian Randomization Using the FTO/MC4R Genotype. *Hypertension,* 54**,** 84-U131.

TRAYLOR, M., FARRALL, M., HOLLIDAY, E. G., SUDLOW, C., HOPEWELL, J. C., CHENG, Y.-C., FOMAGE, M., IKRAM, M. A., MALIK, R., BEVAN, S., THORSTEINSDOTTIR, U., NALLS, M. A., LONGSTRETH, W. T., WIGGINS, K. L., YADAV, S., PARATI, E. A., DESTEFANO, A. L., WORRALL, B. B., KITTNER, S., KHAN, M. S., REINER, A. P., HELGADOTTIR, A., ACHTERBERG, S., FERNANDEZ-CADENAS, I., ABBOUD, S., SCHMIDT, R., WALTERS, M., CHEN, W.-M., RINGELSTEIN, E. B., O'DONNELL, M., HO, W. K., PERA, J., LEMMENS, R., NORRVING, B., HIGGINS, P., BENN, M., SALE, M., KUHLENBAEUMER, G., DONEY, A. S. F., VICENTE, A. M., DELAVARAN, H., ALGRA, A., DAVIES, G., OLIVEIRA, S. A., PALMER, C. N. A., DEARY, I., SCHMIDT, H., PANDOLFO, M., MONTANER, J., CARTY, C., DE BAKKER, P. I. W., KOSTULAS, K., FERRO, J. M., VAN ZUYDAM, N. R., VALDIMARSSON, E., NORDESTGAARD, B. G., LINDGREN, A., THIJS, V., SLOWIK, A., SALEHEEN, D., PARE, G., BERGER, K., THORLEIFSSON, G., HOFMAN, A., MOSLEY, T. H., MITCHELL, B. D., FURIE, K., CLARKE, R., LEVI, C., SESHADRI, S., GSCHWENDTNER, A., BONCORAGLIO, G. B., SHARMA, P., BIS, J. C., GRETARSDOTTIR, S., PSATY, B. M., ROTHWELL, P. M., ROSAND, J., MESCHIA, J. F., STEFANSSON, K., DICHGANS, M., MARKUS, H. S., AUSTRALIAN STROKE GENETICS, C., WTCCC & INT STROKE GENETICS, C. 2012. Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE Collaboration): a meta-analysis of genome-wide association studies. *Lancet Neurology,* 11**,** 951-962.

TRYGG, J. & WOLD, S. 2002. Orthogonal projections to latent structures (O-PLS). *Journal of Chemometrics,* 16**,** 119-128.

TWIGG, S. R. F., VORGIA, E., MCGOWAN, S. J., PERAKI, I., FENWICK, A. L., SHARMA, V. P., ALLEGRA, M., ZARAGKOULIAS, A., AKHA, E. S., KNIGHT, S. J. L., LORD, H., LESTER, T., IZATT, L., LAMPE, A. K., MOHAMMED, S. N., STEWART, F. J., VERLOES, A., WILSON, L. C., HEALY, C., SHARPE, P. T., HAMMOND, P., HUGHES, J., TAYLOR, S., JOHNSON, D., WALL, S. A., MAVROTHALASSITIS, G. & WILKIE, A. O. M. 2013. Reduced dosage of ERF causes complex craniosynostosis in humans and mice and links ERK1/2 signaling to regulation of osteogenesis. *Nature Genetics,* 45**,** 308-313.

VALENCIA-SANCHEZ, M. A., LIU, J. D., HANNON, G. J. & PARKER, R. 2006. Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes & Development,* 20**,** 515-524.

VAN HEESCH, S., KLOOSTERMAN, W. P., LANSU, N., RUZIUS, F.-P., LEVANDOWSKY, E., LEE, C. C., ZHOU, S., GOLDSTEIN, S., SCHWARTZ, D. C., HARKINS, T. T., GURYEV, V. & CUPPEN, E. 2013. Improving mammalian genome scaffolding using large insert mate-pair next-generation sequencing. *Bmc Genomics,* 14.

VARGA-SZABO, D., BRAUN, A., KLEINSCHNITZ, C., BENDER, M., PLEINES, I., PHAM, M., RENNE, T., STOLL, G. & NIESWANDT, B. 2008. The calcium sensor STIM1

is an essential mediator of arterial thrombosis and ischemic brain infarction. *Journal of Experimental Medicine,* 205**,** 1583-1591.

VERMEERSCH, P., BUYS, E., POKREISZ, P., MARSBOOM, G., ICHINOSE, F., SIPS, P., PELLENS, M., GILLIJNS, H., SWINNEN, M., GRAVELINE, A., COLLEN, D., DEWERCHIN, M., BROUCKAERT, P., BLOCH, K. D. & JANSSENS, S. 2007. Soluble guanylate cyclase-alpha 1 deficiency selectively inhibits the pulmonary vasodilator response to nitric oxide and increases the pulmonary vascular remodeling response to chronic hypoxia. *Circulation,* 116**,** 936-943.

WAIN, L. V., VERWOERT, G. C., O'REILLY, P. F., SHI, G., JOHNSON, T., JOHNSON, A. D., BOCHUD, M., RICE, K. M., HENNEMAN, P., SMITH, A. V., EHRET, G. B., AMIN, N., LARSON, M. G., MOOSER, V., HADLEY, D., DOERR, M., BIS, J. C., ASPELUND, T., ESKO, T., JANSSENS, A. C. J. W., ZHAO, J. H., HEATH, S., LAAN, M., FU, J., PISTIS, G., LUAN, J. A., ARORA, P., LUCAS, G., PIRASTU, N., PICHLER, I., JACKSON, A. U., WEBSTER, R. J., ZHANG, F., PEDEN, J. F., SCHMIDT, H., TANAKA, T., CAMPBELL, H., IGL, W., MILANESCHI, Y., HOTTENGA, J.-J., VITART, V., CHASMAN, D. I., TROMPET, S., BRAGG-GRESHAM, J. L., ALIZADEH, B. Z., CHAMBERS, J. C., GUO, X., LEHTIMAKI, T., KUEHNEL, B., LOPEZ, L. M., POLASEK, O., BOBAN, M., NELSON, C. P., MORRISON, A. C., PIHUR, V., GANESH, S. K., HOFMAN, A., KUNDU, S., MATTACE-RASO, F. U. S., RIVADENEIRA, F., SIJBRANDS, E. J. G., UITTERLINDEN, A. G., HWANG, S.-J., VASAN, R. S., WANG, T. J., BERGMANN, S., VOLLENWEIDER, P., WAEBER, G., LAITINEN, J., POUTA, A., ZITTING, P., MCARDLE, W. L., KROEMER, H. K., VOELKER, U., VOELZKE, H., GLAZER, N. L., TAYLOR, K. D., HARRIS, T. B., ALAVERE, H., HALLER, T., KEIS, A., TAMMESOO, M.-L., AULCHENKO, Y., BARROSO, I., KHAW, K.-T., GALAN, P., HERCBERG, S., LATHROP, M., EYHERAMENDY, S., ORG, E., SOBER, S., LU, X., NOLTE, I. M., PENNINX, B. W., CORRE, T., MASCIULLO, C., SALA, C., GROOP, L., VOIGHT, B. F., MELANDER, O., et al. 2011. Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nature Genetics,* 43**,** 1005-U122.

WAKISAKA, Y., CHU, Y., MILLER, J. D., ROSENBERG, G. A. & HEISTAD, D. D. 2010. Spontaneous intracerebral hemorrhage during acute and chronic hypertension in mice. *Journal of Cerebral Blood Flow and Metabolism,* 30**,** 56-69.

WATERSTON, R. H., LINDBLAD-TOH, K., BIRNEY, E., ROGERS, J., ABRIL, J. F., AGARWAL, P., AGARWALA, R., AINSCOUGH, R., ALEXANDERSSON, M., AN, P., ANTONARAKIS, S. E., ATTWOOD, J., BAERTSCH, R., BAILEY, J., BARLOW, K., BECK, S., BERRY, E., BIRREN, B., BLOOM, T., BORK, P., BOTCHERBY, M., BRAY, N., BRENT, M. R., BROWN, D. G., BROWN, S. D., BULT, C., BURTON, J., BUTLER, J., CAMPBELL, R. D., CARNINCI, P., CAWLEY, S., CHIAROMONTE, F., CHINWALLA, A. T., CHURCH, D. M., CLAMP, M., CLEE, C., COLLINS, F. S., COOK, L. L., COPLEY, R. R., COULSON, A., COURONNE, O., CUFF, J., CURWEN, V., CUTTS, T., DALY, M., DAVID, R., DAVIES, J., DELEHAUNTY, K. D., DERI, J., DERMITZAKIS, E. T., DEWEY, C., DICKENS, N. J., DIEKHANS, M., DODGE, S., DUBCHAK, I., DUNN, D. M., EDDY, S. R., ELNITSKI, L., EMES, R. D., ESWARA, P., EYRAS, E., FELSENFELD, A., FEWELL, G. A., FLICEK, P., FOLEY, K., FRANKEL, W. N., FULTON, L. A., FULTON, R. S., FUREY, T. S., GAGE, D., GIBBS, R. A., GLUSMAN, G., GNERRE, S., GOLDMAN, N., GOODSTADT, L., GRAFHAM, D., GRAVES, T. A., GREEN, E. D., GREGORY, S., GUIGO, R., GUYER, M., HARDISON, R. C., HAUSSLER, D., HAYASHIZAKI, Y., HILLIER, L. W., HINRICHS, A., HLAVINA, W., HOLZER, T., HSU, F., HUA, A., HUBBARD, T., HUNT, A., JACKSON, I.,

JAFFE, D. B., JOHNSON, L. S., JONES, M., JONES, T. A., JOY, A., KAMAL, M., KARLSSON, E. K., et al. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature,* 420**,** 520-562.

WHO. 2012. World Health Statistics Report. Geneva, Switzerland. Available at http://www.who.int/gho/publications/world_health_statistics/2012/en/

WILSON, F. H., DISSE-NICODEME, S., CHOATE, K. A., ISHIKAWA, K., NELSON-WILLAMS, C., DESITTER, I., GUNEL, M., MILFORD, D. V., LIPKIN, G. W., ACHARD, J. M., FEELY, M. P., DUSSOL, B., BERLAND, Y., UNWIN, R. J., MAYAN, H., SIMON, D. B., FARFEL, Z., JEUNEMAITRE, X. & LIFTON, R. P. 2001. Human hypertension caused by mutations in WNK kinases. *Science,* 293**,** 1107-1112.

WILSON, R. C., KROZOWSKI, Z. S., LI, K., OBEYESEKERE, V. R., RAZZAGHYAZAR, M., HARBISON, M. D., WEI, J. Q., SHACKLETON, C. H. L., FUNDER, J. W. & NEW, M. I. 1995. A mutation in the HSD11B2 gene in a family with apparent mineralocorticoid excess. *Journal of Clinical Endocrinology & Metabolism,* 80**,** 2263-2266.

WORTHEY, E. A., STODDARD, A. J. & JACOB, H. J. 2010. Sequencing of the Rat Genome and Databases. *Rat Genomics: Methods and Protocols,* 597**,** 33-53.

XIA, H., OOI, L. L. P. J. & HUI, K. M. 2013. MicroRNA-216a/217-induced epithelial-mesenchymal transition targets PTEN and SMAD7 to promote drug resistance and recurrence of liver cancer. *Hepatology,* 58**,** 629-641.

YALCIN, B., WONG, K., AGAM, A., GOODSON, M., KEANE, T. M., GAN, X., NELLAKER, C., GOODSTADT, L., NICOD, J., BHOMRA, A., HERNANDEZ-PLIEGO, P., WHITLEY, H., CLEAK, J., DUTTON, R., JANOWITZ, D., MOTT, R., ADAMS, D. J. & FLINT, J. 2011. Sequence-based characterization of structural variation in the mouse genome. *Nature,* 477**,** 326-329.

YAMORI, Y. & OKAMOTO, K. 1974. Spontaneous hypertension in the rat. A model for human "essential" hypertension. *Verhandlungen der Deutschen Gesellschaft fur Innere Medizin,* 80**,** 168-70.

YAMORI, Y., TOMIMOTO, K., OOSHIMA, A., HAZAMA, F. & OKAMOTO, K. 1974. Developmental course of hypertension in SHR-substrains susceptible to hypertensive cerebrovascular lesions. *Japanese Heart Journal,* 15**,** 209-210.

YOGI, A., CALLERA, G. E., ARANHA, A. B., ANTUNES, T. T., GRAHAM, D., MCBRIDE, M., DOMINICZAK, A. & TOUYZ, R. M. 2011. Sphingosine-1-Phosphate-Induced Inflammation Involves Receptor Tyrosine Kinase Transactivation in Vascular Cells Upregulation in Hypertension. *Hypertension,* 57**,** 809-+.

YOSHIDA, M., WATANABE, Y., YAMANISHI, K., YAMASHITA, A., YAMAMOTO, H., OKUZAKI, D., SHIMADA, K., NOJIMA, H., YASUNAGA, T., OKAMURA, H., MATSUNAGA, H. & YAMANISHI, H. 2014. Analysis of genes causing hypertension and stroke in spontaneously hypertensive rats: Gene expression profiles in the brain. *International Journal of Molecular Medicine,* 33**,** 887-896.

YUSUF, S., HAWKEN, S., OUNPUU, S., DANS, T., AVEZUM, A., LANAS, F., MCQUEEN, M., BUDAJ, A., PAIS, P., VARIGOS, J., LIU, L. S. & INVESTIGATORS, I. S. 2004. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *Lancet,* 364**,** 937-952.

YUSUF, S., RANGARAJAN, S., TEO, K., ISLAM, S., LI, W., LIU, L., BO, J., LOU, Q., LU, F., LIU, T., YU, L., ZHANG, S., MONY, P., SWAMINATHAN, S., MOHAN, V., GUPTA, R., KUMAR, R., VIJAYAKUMAR, K., LEAR, S., ANAND, S., WIELGOSZ, A., DIAZ, R., AVEZUM, A., LOPEZ-JARAMILLO, P., LANAS, F., YUSOFF, K.,

ISMAIL, N., IQBAL, R., RAHMAN, O., ROSENGREN, A., YUSUFALI, A., KELISHADI, R., KRUGER, A., PUOANE, T., SZUBA, A., CHIFAMBA, J., OGUZ, A., MCQUEEN, M., MCKEE, M., DAGENAIS, G. & INVESTIGATORS, P. 2014. Cardiovascular Risk and Events in 17 Low-, Middle-, and High-Income Countries. *New England Journal of Medicine,* 371**,** 818-827.

ZHANG, J., CHIODINI, R., BADR, A. & ZHANG, G. 2011. The impact of next-generation sequencing on genomics. *Journal of genetics and genomics,* 38**,** 95-109.

ZHOU, Q., ZHAO, F., LV, Z.-P., ZHENG, C.-G., ZHENG, W.-D., SUN, L., WANG, N.-N., PANG, S., DE ANDRADE, F. M., FU, M., HE, X.-H., HUI, J., JIANG, W.-Y., YANG, C.-Y., SHI, X.-H., ZHU, X.-Q., PANG, G.-F., YANG, Y.-G., XIE, H.-Q., ZHANG, W.-D., HU, C.-Y. & YANG, Z. 2014. Association between APOC1 Polymorphism and Alzheimer's Disease: A Case-Control Study and Meta-Analysis. *Plos One,* 9.

ZHU, C. H., MORSE, L. R. & BATTAGLINO, R. A. 2012. SNX10 is required for osteoclast formation and resorption activity. *Journal of Cellular Biochemistry,* 113**,** 1608-1615.