Title
A Systems Genetics Approach to the Identification of Causal Genes in Heart Failure Using a Large Mouse Panel

Author
Rau, Christoph Daniel

Publication Date
2013

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

A Systems Genetics Approach to the Identification of Causal Genes in Heart Failure Using a Large Mouse Panel

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Microbiology, Immunology and Molecular Genetics

by

Christoph Daniel Rau

2013

ABSTRACT OF THE DISSERTATION

A Systems Genetics Approach

to the Identification of Causal

Genes in Heart Failure

Using a Large Mouse Panel

By

Christoph Daniel Rau

Doctor of Philosophy in Microbiology, Immunology and Molecular Genetics

University of California, Los Angeles, 2013

Professor Aldons Jake Lusis, Chair

Heart failure (HF) accounts for 1 in 9 deaths in the United States and is the leading cause of hospitalization for people over the age of 65 and the incidence of HF is predicted to rise over the coming years. The complexity which underlies common forms

of HF has hindered the study of the disease in humans, and approaches, such as genome-wide association studies (GWAS), have had only modest success in identifying genes which are related to this disease. Here we describe the use of a panel of mice to facilitate the study of this complex disorder, reducing heterogeneity and facilitating systems-level approaches.

We used the β-adrenergic agonist isoproterenol to induce HF in 105 unique strains drawn from the Hybrid Mouse Diversity Panel, a novel mouse resource population for the analysis of complex traits. Our first study reports the results of a GWAS on heart weights, cardiac fibrosis and other surrogate traits relevant to HF. Among the 32 significant loci, we identified several strong candidates which had previously been shown to contribute to mendelian forms of cardiomyopathy. We were also able to validate two novel candidate genes, the orphan transporter *Abcc6*, and the long noncoding RNA *Miat*, using gene targeting, transgenic and *in vitro* approaches.

As part of our systems genetics approach, we developed a novel gene network construction algorithm, which improves on prior methods by allowing non-linear interactions and the ability for genes to operate in multiple modules at once. We were able to demonstrate using previously published data that our results either matched or exceeded another well-known network construction algorithm. In a subsequent study, we applied this method to transcriptomes taken from the HF study. We identified a module of 41 genes which significantly regulates the response of the heart to isoproterenol and

HF and which contains several genes of interest such as *Lgals3*, a diagnostic marker for human HF.

Our results provide a valuable resource toward a better understanding of the pathways and gene-by-environment interactions influencing heart failure.

The dissertation of Christoph Daniel Rau is approved.

Yibin Wang

Ken Bradley

Genhong Cheng

Aldons J. Lusis, Committee Chair

University of California, Los Angeles

2013

Dedication

To everyone who has inspired me

but especially to Hilary,

who does so every day.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Acknowledgements

During my doctoral studies I received help and support from a number of people who I would like to thank at this time.

I would first like to thank Jake Lusis for being my mentor. It has been an amazing experience working in a lab driven by a brilliant PI who has boundless and infectious enthusiasm for research. Jake has been integral to my development as a teacher and a scientist.

I would also like to thank Yibin Wang for being my unofficial 'co-mentor' during the HMDP heart failure project. His expertise and encouragement were of great help in the completion of this dissertation and in the analysis of the results.

I would also like to thank the other members of my committee, Ken Bradley and Genhong Cheng for their guidance and support of my dissertation project. Their advice was greatly appreciated in shaping the direction and impact of the overall project.

I would like to thank the members of the Lusis and Wang labs for their help over the course of my graduate studies. I would especially like to thank Jessica Wang, who spearheaded the HMDP heart failure project with me. It has been invaluable (and often necessary) for this project to always have two sets of eyes and hands available to collect, organize and analyze data as it was generated, and this project would not have been possible without her. I would also like to thank Anatole, the Brians, Jonathan,

Maragarete, Mete, Millie, Nick, Vincent and Zhihua in a special way for their help and encouragement during the past five years.

I have been lucky in my graduate studies to work with amazing undergraduates each of whom made many contributions to my research and helped me develop my skills. Thank you, Elizabeth, Brandi, Rozeta, Mary and Tram for your eagerness and hard work. I would also like to thank Jake Lusis, Karen Reue and John Merriam for letting me TA human genetics for an extra year.

I would like to thank my AP Biology teacher in 10th grade, Tim McFaul, for igniting a love of biology in me. I'd like to think that the reason I ended up where I am in biology is because of how bogged down AP Bio became when we were studying the molecular biology and genetics segments. I would also like to thank all of the other professors and teachers I have had who kept that flame burning.

I would like to thank my friends Tuck and Lauren for their friendship and support during the course of my graduate work. I would also like to thank my parents, who raised me to value science and the pursuit of knowledge, and whose advice I have always treasured. I would also like to thank my sister, whose toy-stealing ways prompted my first forays into science, and who has always provided an excellent source of competition.

This work was funded by the Cell and Molecular Training Grant and the Vascular Biology Training Grant (both NIH T32HL69766). I would also like to thank Springer for allowing me to reprint "Hybrid Mouse Diversity Panel: a Panel of Inbred Mouse Strains Suitable for Analysis of Complex Genetic Traits" (License #3206681205552) as part of

Chapter 2. Chapter 4 is used under the Creative Commons license, and is copyrighted by myself and my co-authors.

But most importantly, I want to thank my wife Hilary, who has been my best friend, my biggest supporter, and my closest advisor throughout this entire process. This wouldn't have been possible without you. Thank you.

VITA

2004-2006	Tutoring for Principles of Computer Science Harvey Mudd College Claremont, California
2007	B.S., Mathematical Biology Harvey Mudd College Claremont, California
2007-Present	Graduate Student Researcher Department of Microbiology, Immunology and Molecular Genetics University of California, Los Angeles
2008-2011	NIH Ruth L. Kirchstein NRSA GM07185 Cell and Molecular Biology Training Grant
2008	Teaching Assistant in Microbiology Immunology and Molecular Genetics University of California, Los Angeles
2010-2011	Teaching Assistant in Molecular, Cell and Developmental Biology University of California, Los Angeles
2011-2013	NIH Ruth L. Kirchstein NRSA GM07185 Vascular Biology Training Grant
2012	Kern Lipid Conference Early Career Investigator Award of Excellence

PUBLICATIONS

Davis, R.C., van Nas, A., Bennett, B., Orozco, L., Pan, C., **Rau, C.D.**, Eskin, E., and Lusis, A.J. (2013). Genome-wide association mapping of blood cell traits in mice. Mammalian Genome : Official Journal of the International Mammalian Genome Society *24*, 105–118.

Lu, G., Ota, A., Run, S., Franklin, S., **Rau, C.D.**, Ping, P., Lane, T.F., Zhou, H., Reue,K., Lusis, A.J., Vondriska, T., Wang, Y. (2013). PPM1l encodes an inositol requiring protein 1 (IRE1) specific phosphatase that regulates the functional outcome of the ER stress response. Molecular Metabolism. *In Press*

Ghazalpour, A., **Rau, C.D.**, Farber, C.R., Bennett, B.J., Orozco, L.D., van Nas, A., Pan, C., Allayee, H., Beaven, S.W., Civelek, M., et al. (2012). Hybrid mouse diversity panel: a panel of inbred mouse strains suitable for analysis of complex genetic traits. Mammalian Genome : Official Journal of the International Mammalian Genome Society *23*, 680–692.

Ho, M.C.W., Johnsen, H., Goetz, S.E., Schiller, B.J., Bae, E., Tran, D. a, Shur, A.S., Allen, J.M., **Rau, C.**, Bender, W., et al. (2009). Functional evolution of cis-regulatory modules at a homeotic gene in Drosophila. PLoS Genetics *5*, e1000709.

Lagarrigue, S., Hormozdiari, F., Martin, L.J., Lecerf, F., Hasin, Y., **Rau, C.**, Hagopian, R., Xiao, Y., Yan, J., Drake, T. a, et al. (2013). Limited RNA Editing in Exons of Mouse Liver and Adipose Tissue. Genetics *193*, 1107–1115.

Mungrue, I.N., Zhao, P., Yao, Y., Meng, H., **Rau, C.**, Havel, J. V, Gorgels, T.G.M.F., Bergen, A. a B., MacLellan, W.R., Drake, T. a, et al. (2011). Abcc6 deficiency causes increased infarct size and apoptosis in a mouse cardiac ischemia-reperfusion model. Arteriosclerosis, Thrombosis, and Vascular Biology *31*, 2806–2812.

Parks, B.W., Nam, E., Org, E., Kostem, E., Norheim, F., Hui, S.T., Pan, C., Civelek, M., **Rau, C.D.**, Bennett, B.J., et al. (2013). Genetic control of obesity and gut microbiota composition in response to high-fat, high-sucrose diet in mice. Cell Metabolism *17*, 141–152.

Rau, C.D., Wisniewski, N., Orozco, L.D., Bennett, B., Weiss, J., and Lusis, A.J. (2013). Maximal information component analysis: a novel non-linear network analysis method. Frontiers in Genetics *4*, 28.

Weiss, J.N., Karma, A., MacLellan, W.R., Deng, M., **Rau, C.D.**, Rees, C.M., Wang, J., Wisniewski, N., Eskin, E., Horvath, S., et al. (2012). "Good enough solutions" and the genetics of complex diseases. Circulation Research *111*, 493–504.

Chapter 1

Introduction

Heart Failure

Heart disease is the primary cause of death in developed countries and will soon overtake infection as the primary cause of death in the developing world[1]. In 2009, heart disease accounted for 32.3% of all deaths in the United States, equivalent to a rate of roughly one death every 40 seconds[1]. Furthermore, the cost of treating heart disease in the United States is $312.6 billion per year, 37% higher than the cost of treating the second most common cause of death, cancer [1]. Over the past thirty years, however, the age-adjusted death rate due to heart disease has steadily decreased, from a peak (in men) of around 900 deaths per hundred thousand people a year to around 350 deaths per hundred thousand[1,2]. This dramatic decrease has been attributed roughly equally to changes in public health attitudes, such as the steady decline in cigarette use, and medical interventions, for example development of statins, beta-blockers and advanced medical practices for the treatment of heart disease [1,2].

While most forms of heart disease have shown a significant reduction in mortality and morbidity over the past thirty years, heart failure (HF), which accounts for roughly 1 in 9 deaths in the United States, has resisted this trend[1]. Between 2000 and 2010, the number of hospitalizations in the United States for HF remained constant[3], and is currently the leading cause of hospitalization in people over the age of 65[4]. The incidence of HF in the population is predicted to rise 25% from its current levels by 2030[1].

Heart failure is characterized by the inability of the heart to supply enough blood to the rest of the body. A significant period of morbidity, in which the ability to perform normal life tasks is progressively curtailed, is followed by eventual death due to HF. The symptoms of HF differ depending on which chamber of the heart is failing. Left sided HF is characterized by shortness of breath, cardiomegaly, confusion and hypoxia, while right-sided failure has symptoms of severe edema, loss of appetite and impaired liver function. Common to both left and right-sided HF is a dramatically increased heart rate and fatigue when performing common tasks.

Factors Underlying Heart Failure

The heart is an incredibly resilient organ which is usually able to adapt to the stress that it is subjected to on a daily basis. HF, which occurs when the heart is unable to further compensate for stress and damage, represents the end stage of a slowly developing disorder that often shows no symptoms until it has reached a severe state [5,6]. Because the heart is highly resistant to stress and HF is typically diagnosed in old age, it can be difficult to determine the proximal causes of the disease. The most common risk factor for HF is hypertension, which forces the heart muscle to pump harder to compensate for the increased pressure[1]. This increase in pressure overtaxes the heart, overwhelms its repair mechanisms and eventually causes it to fail. Closely following hypertension as a risk factor for HF is an antecedent myocardial infarction[1]. In a myocardial infarction, some portion of the tissue of the heart dies. This death of cardiac tissue puts additional strain on the remaining cells of the heart. Other less common

causes include heart valve disorders, thyroid disorders, alcohol and other drug abuse, viral myocarditis and treatments from other diseases[7,8]. Each of these risk factors leads to an insufficiency in the heart's ability to pump sufficient quantities of blood at a rate which meets the body's needs. The body compensates for this insufficiency in part through the activation of the β-adrenergic signaling pathway through the release of catecholamines from the sympathetic nervous system and the adrenal gland[9–11]. Catecholamines interact with adrenergic receptors on the surface of the cardiomyocytes and activate a signaling pathway within the cell to increase the rate and strength with which the cardiomyocyte beats as well as promoting the cardiomyocyte to grow in size[11]. In this manner, when an initial stressor to the heart leads to decreased flow of blood, the adrenergic signaling pathway is able to restore the heart to 'normal' function until the stressor is removed. If the initial challenge is not removed, the continual catecholamine stimulation leads to calcium overload, oxidative stress and cardiac cell dysfunction and eventually results in cardiomyocyte death[12]. These cardiomyocytes are replaced by fibroblasts, which form a collagen scar where the cardiomyocyte once was. These scars exacerbate the initial challenge both because they are unable to beat on their own and because collagenous tissue is stiffer than the natural myocardium. This remodeling response necessitates upregulation of catecholamine stimulation, which, in turn, leads to additional myocyte toxicity in a powerful positive feedback loop[12].

This preliminary stage of HF, in which the heart is able to adapt to increasingly severe stressors, is called cardiac hypertrophy and is not usually diagnosed or treated unless specifically screened for. While the person with developing HF continues living a

normal life, the remodeling of the heart continues as the body attempts to compensate for an increasingly damaged myocardium. Eventually, the damage to the heart becomes too severe for the heart to compensate, symptoms begin to present themselves, a diagnosis can be given and treatment can begin. By this point there is significant disregulation of the heart's normal function and the damage to the heart is generally not reversible[10,13,14].

Current strategies for the prevention of HF are focused on the reduction of environmental risk factors in order to mitigate the chances of an individual developing the disease. For instance, quitting smoking reduces the risk of developing heart disease by between 50 to 75% [15]. However, not all people are equally at risk of developing HF even when these environmental factors are controlled for. The existence of a number of rare mendelian forms of HF [7,8,16–19], termed cardiomyopathies, points to an important role for genetics in the incidence and progression of HF. A number of important pathways and genes have been identified using standard molecular biology practices, such as calcineurin, the sarcoglycan family, the beta-adrenergic signaling cascade and deficiencies in calcium signaling and oxidative stress responses[13,20–22]. Despite these advances, it is likely that the majority of genes which predispose individuals to developing the common form of HF remain unknown.

Genome Wide Association Studies of Heart Failure

The sequencing of the human genome in the early 2000s began a new chapter in the attempt to identify genes which predispose individuals to disease. Researchers at the time were convinced that the entirety of the genetic architecture underlying the

susceptibility to disease and other complex phenotypes would be uncovered and explained quickly. David Baltimore stated in the issue of *Nature* that announced the sequencing of the genome that " the analysis of [our genome] will provide us with the power to uncover the genetic basis of our individual capabilities such as mathematical ability, memory, physical coordination, and even, perhaps, creativity."[23] Yet, as further analysis of the genome was performed and the complexity of the genome was more fully appreciated, researchers began to develop increasingly sophisticated methods for the identification of disease-influencing genes.

Genome-wide association studies (GWAS) utilize the natural variation which exists in a population to identify regions and, through follow-up experiments, individual genes which influence a phenotype based on whether the variation in the genome correlates with a corresponding change in phenotype. For example, if there is a mutation shared by a number of people in a population which results in the premature truncation of a gene, GWAS asks if people with such a mutation are more (or less) likely to have an altered phenotype when compared to those with a full-length copy of the gene. GWAS are performed by the genotyping of individuals' DNA for hundreds of thousands of polymorphisms, as well as the careful phenotyping of those same individuals for the trait of interest. As of December 2012, GWAS studies in humans had identified 4,902 significant($P<1e-8$) associations between polymorphisms and phenotypes from traits as diverse as height and weight to blood pressure or anxiety or cannabis dependence[24].

Despite a number of successful GWA studies performed on other diseases, including several in other cardiovascular diseases such as atherosclerosis, GWAS performed on HF have had only modest success in elucidating the genetics underlying this complex disease. Only two heart-failure related loci [25] have reached accepted levels of genome-wide significance despite studies involving tens of thousands of patients and large meta-analyses [26–28]. The challenges of performing GWAS on HF in human populations is complicated by the late onset of the disease, which allows for a number of environmental confounders to mask polymorphism to phenotype correlations, particularly if the effect of the environmental confounder is strong and the effect of the individual gene is weak. For example, smoking increases the risk of developing heart disease by 100-300%[15]. If a particular polymorphism in a gene increases your risk of developing heart disease by 3%, then the effect of the polymorphism will be completely overshadowed by whether or not you smoke, how much you smoke, and/or how long ago you quit smoking[15]. Additionally, HF is strongly influenced by a number of underlying etiologies, such as myocardial infarction, hypertension or metabolic disorders, each of which are, in and of themselves, complex traits with significant environmental confounders[29].

Animal Models and Heart Failure

In humans, HF occurs late in life and can be caused or influenced by a number of underlying etiologies and environmental factors. This has made it difficult to analyze HF in human studies. Animal models provide a means of carefully controlling the specific

cause of a disease as well as eliminating a significant fraction of the environmental variances through the control and standardization of factors such as housing, diet and age. Species as diverse as pig[30], rat[31], mouse[32], zebrafish[33] and even fly[34] have been used as a means to understand the ways in which the heart responds to stress and how individual genes can have a protective or deleterious role. Animal models have been fruitful for the identification of many HF-related genes[6,21,22,32,35–37], however many of these genes were discovered by analyzing established pathways known to be involved in HF or through serendipitous analysis of a spontaneous mutation in an organism. The former is unlikely to reveal new insights into novel disease pathways, while the latter relies on random chance to identify new genes and pathways.

Linkage analysis, a method which, like GWAS, links polymorphisms in the genome to phenotypes has been used in the past in humans and model organisms to attempt to identify regions of the genome that influence HF[32,38–40]. Unlike mendelian traits, where such analyses have led to the identification of hundreds of underlying genes, complex traits such as HF have proven resistant to linkage analysis. GWAS analysis has a number of benefits that make it better suited to addressing complex disorders such as HF. First, the mice used in a linkage study are unique, and therefore each must be genotyped individually for polymorphic markers. More importantly, each mouse can only be used once in an experiment, which eliminates the ability to do case/control studies using identical mice. Furthermore, linkage studies suffer from poor resolution: a single locus for a trait could potentially span tens to hundreds of Mb, encompassing hundreds of genes. Despite reducing the complexity of the identification of the causal

gene by an order of magnitude, the work necessary to identify the causal gene within a single locus remains a major undertaking[32,35,41].

The Hybrid Mouse Diversity Panel

To overcome the limitations of linkage analysis in mice and incorporate the advances of GWAS, we have developed a panel of over a hundred mouse inbred lines which we term the Hybrid Mouse Diversity Panel (HMDP) which allows us to perform association analyses in mice. This resource is described in detail in chapter 2. Some of the major advantages of the HMDP include an approximately 10-fold improvement in mapping resolution relative to linkage studies and an unlimited pool of genetically identical mice to study, allowing us to further control environmental noise and perform case-control studies. Previously published work using the HMDP has shown that the panel is well-suited to the identification of causal genes in phenotypes such as plasma lipid levels[42], bone mineral density[43], obesity[44] and blood cell traits[45], among others.

Heart Failure in the HMDP

The ability to perform case/control studies and carefully control environmental variables makes the HMDP an excellent resource in which to study HF. The synthetic catecholamine isoproterenol (ISO) has been used extensively[46–48] to cause mice to develop HF. Although adrenergic overstimulation is rarely a principal cause of HF in humans, overstimulation of the adrenergic signaling pathway is a common driver of HF progression after the initial stressor forces the heart to begin to remodel. ISO offers advantages over other methods of inducing HF in rodents such as trans-aortic constriction

(TAC) because it is a purely chemical means of inducing HF. TAC and other physical models require researchers to physically interact with the heart and could lead to the introduction of experimenter error among the samples. Chapter 3 describes the results of a GWAS performed on the HMDP using an ISO challenge to develop HF phenotypes in the panel.

Cardiac Fibrosis in the HMDP

Cardiac fibrosis and scarring is caused by and exacerbates many cardiovascular disorders, including tissue replacement after myocardial infarction[49] and the structural remodeling which occurs during HF[50,51]. Although post-mortem analyses of cardiac fibrosis in humans have correlated fibrosis to the severity of heart-related conditions, few systems-wide approaches have addressed cardiac fibrosis[52] and no GWAS of this important factor in HF have been reported to date, likely due to the difficulty in analyzing this phenotype non-invasively in humans.

Through the use of an animal model such as the HMDP, cardiac fibrosis can be studied in a controlled environment that reduces environmental variation and provides easy access to heart tissue. ISO has previously been demonstrated[9] to lead to extensive changes in cardiac fibrosis. Chapter 3 also describes the results of GWAS performed on the amount of fibrosis present in the hearts of control strains and strains treated with ISO.

Gene Network Analysis

Although GWAS have had success in the identification of genes involved in the modulation of phenotypes, some[53,54] have claimed that GWAS have proven disappointing, at least compared to the initial promise of what they could offer. Much of this criticism is based on the fact that while GWAS analysis may have revealed a number of genes that appear to influence phenotypes, each gene's contribution to the overall variation of the phenotype is small, typically conferring only a modest increase in risk[54,55] and in total explaining only a small fraction of the overall genetic heritability of the phenotype. For instance, although it is thought that human height is roughly 80-90% heritable, GWAS studies have only been able to explain roughly 10% of that heritability[56].

There are several possible explanations for why GWAS do not return strong candidates with large effect sizes, including the hypothesis that most of the mutations which lead to a disease state are highly selected against and fall below cutoff thresholds used in GWAS. Another possible explanation is that the variations in a phenotype that we observe in a population are based on the interactions of many genes at once, acting together and in opposition to one another to result in a final phenotype in a manner too complex to be observed by a method that queries a single polymorphism (and therefore a single gene) at a time.

A powerful method to analyze genetic data is gene coexpression network analysis, where gene expression from a tissue across many individuals are compared to one another. Genes which are expressed in similar ways across either genetic perturbations

(eg. different strains) and/or an environmental perturbations (eg. ISO stimulation) are clustered together to form groups of genes called modules, which can then be related back to the initial phenotype of interest to identify subsets of genes which together may influence the phenotype. This method has been used in the past successfully to identify genes related to HF, autism, bone density, the inflammatory response and more[31,55,57–60].

Despite these successes, most co-expression methods rely on assumptions which do not accurately reflect the biological systems they are meant to model. These assumptions introduce artificiality into the analysis and potentially result in skewed or altered modules, diminishing the efficacy and usefulness of the method itself. One major assumption that is frequently made in network algorithms is that genes interact with one another in a linear fashion. Basic chemistry and biology suggest that this is rarely the case, and typically should only be assumed in certain situations and certain conditions for certain molecules. Another assumption made by network analysis methods is that genes can have only one role, and therefore belong to only a single module, an assumption which has many known counterexamples in biology[61,62]. Chapter 4 describes a new network analysis method, Maximal Information Component Analysis (MICA) that specifically avoids both of these assumptions and allows for non-linear interactions between genes and for genes to exist in multiple modules. Chapter 5 then describes the application of MICA to the ISO-treated HMDP panel and the novel pathways that are revealed by that analysis.

Dissertation Goals and Strategy

The focus of this dissertation project is the elucidation of genes which underlie heart failure using a large mouse panel and systems-level genetic techniques. Highlights include: a GWAS study on mouse hypertrophy which revealed roles in common forms of HF for many previously identified mendelian cardiomyopathy genes; the first report of a cardiac fibrosis GWAS; the development of a novel gene network analysis tool; and the analysis of transcriptomes of the HMDP HF project using this tool and eQTL hotspot analysis.

This chapter provides an introduction to heart failure and the difficulties in studying this disease in human populations. It also summarizes the tools developed to study common diseases in mice. Chapter 2 is a published review article on the HMDP, the panel of mice which was used to perform the heart failure study. Chapter 3: "Genome Wide Association Studies of Cardiac Hypertrophy and Cardiac Fibrosis Using a Large Mouse Panel" contains a manuscript describing the results of two GWA studies from the data we generated as a part of this panel. Chapter 4: "Maximal Information Component Analysis: a Novel Network Analysis Algorithm" contains a paper which describes the development of and details of MICA, the network analysis method developed as a part of the dissertation. Chapter 5: "eQTL Hotspot Analysis and Gene Networks Reveal Key Drivers of Catecholamine-Induced Heart Failure" contains a manuscript which describes the application of systems biology techniques to the transcriptomes of the HMDP HF study. Chapter 6 provides a synthesis of the findings of this dissertation, and describes the results in the context of future studies that will expand our knowledge about the way in which the genes identified during the course of this dissertation lead to heart failure.

References

1. Go, A. S. *et al.* Heart disease and stroke statistics--2013 update: a report from the American Heart Association. *Circulation* **127**, e6–e245 (2013).

2. Ford, E. S. & Capewell, S. Proportion of the decline in cardiovascular mortality disease due to prevention versus treatment: public health versus clinical care. *Annual review of public health* **32**, 5–22 (2011).

3. States, U., Hall, M. J., Ph, D., Levant, S. & Defrances, C. J. Hospitalization for Congestive Heart Failure : (2012).

4. Krumholz, H. M. *et al.* Predictors of readmission among elderly survivors of admission with heart failure. *American heart journal* **139**, 72–7 (2000).

5. Creemers, E. E., Wilde, A. a & Pinto, Y. M. Heart failure: advances through genomics. *Nature reviews. Genetics* **12**, 357–62 (2011).

6. Breckenridge, R. Heart failure and mouse models. *Disease models & mechanisms* **3**, 138–43 (2010).

7. Schönberger, J. & Seidman, C. E. Many roads lead to a broken heart: the genetics of dilated cardiomyopathy. *American journal of human genetics* **69**, 249–60 (2001).

8. Davies, M. J. The cardiomyopathies: an overview. *Heart (British Cardiac Society)* **83**, 469–74 (2000).

9. Zhang, X. *et al.* Cardiotoxic and Cardioprotective Features of Chronic β-Adrenergic Signaling. *Circulation research* **112**, 498–509 (2013).

10. Keys, J. R. & Koch, W. J. The adrenergic pathway and heart failure. *Recent progress in hormone research* **59**, 13–30 (2004).

11. Satwani, S., Dec, G. W. & Narula, J. Beta-adrenergic blockers in heart failure: review of mechanisms of action and clinical outcomes. *Journal of cardiovascular pharmacology and therapeutics* **9**, 243–55 (2004).

12. Adameova, A., Abdellatif, Y. & Dhalla, N. S. ` SE REVIEW / SYNTHE Role of the excessive amounts of circulating catecholamines and glucocorticoids in stress-induced heart disease. **514**, 493–514 (2009).

13. Velagaleti, R. S. & O'Donnell, C. J. Genomics of Heart Failure. *Heart Failure Clinics* **6**, 115–124 (2010).

14. Gjesdal, O., Bluemke, D. a & Lima, J. a Cardiac remodeling at the population level--risk factors, screening, and outcomes. *Nature reviews. Cardiology* **8**, 673–85 (2011).

15. Ockene, I. S. & Miller, N. H. Cigarette Smoking, Cardiovascular Disease, and Stroke : A Statement for Healthcare Professionals From the American Heart Association. *Circulation* **96**, 3243–3247 (1997).

16. Burkett, E. L. & Hershberger, R. E. Clinical and genetic issues in familial dilated cardiomyopathy. *Journal of the American College of Cardiology* **45**, 969–81 (2005).

17. Lefeber, D. J. *et al.* Autosomal recessive dilated cardiomyopathy due to DOLK mutations results from abnormal dystroglycan O-mannosylation. *PLoS genetics* **7**, e1002427 (2011).

18. Arad, M. *et al.* Constitutively active AMP kinase mutations cause glycogen storage disease mimicking hypertrophic cardiomyopathy. *Journal of Clinical Investigation* **109**, 357–362 (2002).

19. Friedrich, F. W. *et al.* A new polymorphism in human calmodulin III gene promoter is a potential modifier gene for familial hypertrophic cardiomyopathy. *European heart journal* **30**, 1648–55 (2009).

20. Lu, M. *et al.* The function of calcineurin and ERK1/2 signal in the antihypertrophic effects of kappa-opioid receptor stimulation on myocardial hypertrophy induced by isoprenaline. *Die Pharmazie* **67**, 182–6 (2012).

21. Goehringer, C. *et al.* Prevention of cardiomyopathy in delta-sarcoglycan knockout mice after systemic transfer of targeted adeno-associated viral vectors. *Cardiovascular research* **82**, 404–10 (2009).

22. Chu, G. & Kranias, E. G. Phospholamban as a therapeutic modality in heart failure. *Novartis Foundation symposium* **274**, 156–71; discussion 172–5, 272–6 (2006).

23. Baltimore, D. Our genome unveiled. *Nature* **409**, 814–6 (2001).

24. Hindorff, L. *et al.* A Catalog of Published Genome-Wide Association Studies. at < www.genome.gov/gwastudies>

25. Villard, E. *et al.* A genome-wide association study identifies two loci associated with heart failure due to dilated cardiomyopathy. *European heart journal* **32**, 1065–76 (2011).

26. Morrison, A. C. *et al.* Genomic variation associated with mortality among adults of European and African ancestry with heart failure: the cohorts for heart and aging research in genomic epidemiology consortium. *Circulation. Cardiovascular genetics* **3**, 248–55 (2010).

27. Parsa, A. *et al.* Hypertrophy-associated polymorphisms ascertained in a founder cohort applied to heart failure risk and mortality. *Clinical and translational science* **4**, 17–23 (2011).

28. Smith, N. L. *et al.* Association of genome-wide variation with the risk of incident heart failure in adults of European and African ancestry: a prospective meta-analysis from the cohorts for heart and aging research in genomic epidemiology (CHARGE) consortium. *Circulation. Cardiovascular genetics* **3**, 256–66 (2010).

29. Mudd, J. O. & Kass, D. A. Tackling heart failure in the twenty-first century. *Nature* **451**, 919–28 (2008).

30. Santos, L. A. S. *et al.* Cavo-pulmonary anastomosis associated with left ventricular in comparison with biventricular circulatory support in acute heart failure. *Revista Brasileira de Cirurgia Cardiovascular* **27**, 552–561 (2012).

31. McDermott-Roe, C. *et al.* Endonuclease G is a novel determinant of cardiac hypertrophy and mitochondrial function. *Nature* **478**, 114–118 (2011).

32. Wheeler, F. C. *et al.* QTL mapping in a mouse model of cardiomyopathy reveals an ancestral modifier allele affecting heart function and survival. *Mammalian genome : official journal of the International Mammalian Genome Society* **16**, 414–23 (2005).

33. Kloos, W., Katus, H. A. & Meder, B. Genetic cardiomyopathies. Lessons learned from humans, mice, and zebrafish. *Herz* **37**, 612–7 (2012).

34. Yu, L., Daniels, J., Glaser, A. E. & Wolf, M. J. Raf-mediated cardiac hypertrophy in adult Drosophila. *Disease models & mechanisms* **6**, 964–76

35. Wheeler, F. C. *et al.* Tnni3k modifies disease progression in murine models of cardiomyopathy. *PLoS genetics* **5**, e1000647 (2009).

36. Sarwar, R. & Cook, S. a Genomic analysis of left ventricular remodeling. *Circulation* **120**, 437–44 (2009).

37. Patten, R. D. & Hall-Porter, M. R. Small animal models of heart failure: development of novel therapies, past and present. *Circulation. Heart failure* **2**, 138–44 (2009).

38. Cervino, A. C. *et al.* Integrating QTL and high-density SNP analyses in mice to identify Insig2 as a susceptibility gene for plasma cholesterol levels. *Genomics* **86**, 505–17 (2005).

39. Solberg, L. C. *et al.* Genetic analysis of the stress-responsive adrenocortical axis. *Physiological genomics* **27**, 362–9 (2006).

40. Xiong, Q. *et al.* Genetic and molecular basis of quantitative trait loci of arthritis in rat: genes and polymorphisms. *Journal of immunology (Baltimore, Md. : 1950)* **181**, 859–64 (2008).

41. Le Corvoisier, P., Park, H.-Y., Carlson, K. M., Marchuk, D. a & Rockman, H. a Multiple quantitative trait loci modify the heart failure phenotype in murine cardiomyopathy. *Human molecular genetics* **12**, 3097–107 (2003).

42. Bennett, B. J. *et al.* A high-resolution association mapping panel for the dissection of complex traits in mice. *Genome research* **20**, 281–90 (2010).

43. Farber, C. R. *et al.* Mouse genome-wide association and systems genetics identify Asxl2 as a regulator of bone mineral density and osteoclastogenesis. *PLoS genetics* **7**, e1002038 (2011).

44. Parks, B. W. *et al.* Genetic control of obesity and gut microbiota composition in response to high-fat, high-sucrose diet in mice. *Cell metabolism* **17**, 141–52 (2013).

45. Davis, R. C. *et al.* Genome-wide association mapping of blood cell traits in mice. *Mammalian genome : official journal of the International Mammalian Genome Society* **24**, 105–18 (2013).

46. Berthonneche, C. *et al.* Cardiovascular response to beta-adrenergic blockade or activation in 23 inbred mouse strains. *PloS one* **4**, e6610 (2009).

47. Brittsan, a G. *et al.* The effect of isoproterenol on phospholamban-deficient mouse hearts with altered thyroid conditions. *Journal of molecular and cellular cardiology* **31**, 1725–37 (1999).

48. Kumaran, K. S. & Prince, P. S. M. Caffeic acid protects rat heart mitochondria against isoproterenol-induced oxidative damage. *Cell stress & chaperones* **15**, 791–806 (2010).

49. Frangogiannis, N. G. Regulation of the inflammatory response in cardiac repair. *Circulation research* **110**, 159–73 (2012).

50. Beltrami, C. a. *et al.* Structural basis of end-stage failure in ischemic cardiomyopathy in humans. *Circulation* **89**, 151–163 (1994).

51. Kong, P., Christia, P. & Frangogiannis, N. G. The pathogenesis of cardiac fibrosis. *Cellular and molecular life sciences : CMLS* (2013).doi:10.1007/s00018-013-1349-6

52. Ellinor, P. T. *et al.* A novel locus for dilated cardiomyopathy, diffuse myocardial fibrosis, and sudden death on chromosome 10q25-26. *Journal of the American College of Cardiology* **48**, 106–11 (2006).

53. Goldstein, D. B. Common genetic variation and human traits. *The New England journal of medicine* **360**, 1696–8 (2009).

54. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–53 (2009).

55. Weiss, J. N. *et al.* "Good enough solutions" and the genetics of complex diseases. *Circulation research* **111**, 493–504 (2012).

56. Yang, J. *et al.* Genome partitioning of genetic variation for complex traits using common SNPs. *Nature genetics* **43**, 519–25 (2011).

57. Park, C. C. *et al.* Gene networks associated with conditional fear in mice identified using a systems genetics approach. *BMC Systems Biology* **5**, 43 (2011).

58. Dewey, F. E., Wheeler, M. T. & Ashley, E. a Systems biology of heart failure, challenges and hopes. *Current opinion in cardiology* **26**, 314–21 (2011).

59. Barabási, A.-L., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based approach to human disease. *Nature reviews. Genetics* **12**, 56–68 (2011).

60. Keller, M. P. *et al.* A gene expression network model of type 2 diabetes links cell cycle regulation in islets with diabetes susceptibility. *Genome research* **18**, 706–16 (2008).

61. Beaulieu, J.-M. *et al.* A beta-arrestin 2 signaling complex mediates lithium action on behavior. *Cell* **132**, 125–36 (2008).

62. Winbanks, C. E. *et al.* Follistatin-mediated skeletal muscle hypertrophy is regulated by Smad3 and mTOR independently of myostatin. *The Journal of cell biology* **197**, 997–1008 (2012).

Chapter 2

The Hybrid Mouse Diversity Panel

Introduction

The hybrid mouse diversity panel (HMDP) represents a significant improvement in the ability for researchers to use mouse genetics for the study of all manner of complex traits. Its development was essential for the research described in subsequent chapters. This chapter consists of a reprint of a review article about the HMDP, which describes the panel in detail. Specifically, it addresses the improved ability to perform GWAS using the HMDP compared to prior attempts of using mice in association studies. This improvement is due to the incorporation of recombinant inbred strains and the use of an algorithm, EMMA, to correct for the extensive population structure in mice. It then describes some of the studies that have been performed with the HMDP, as well as other work, including this dissertation, that was ongoing at time of publication.

Mamm Genome
DOI 10.1007/s00335-012-9411-5

Hybrid mouse diversity panel: a panel of inbred mouse strains suitable for analysis of complex genetic traits

Anatole Ghazalpour · Christoph D. Rau · Charles R. Farber · Brian J. Bennett · Luz D. Orozco · Atila van Nas · Calvin Pan · Hooman Allayee · Simon W. Beaven · Mete Civelek · Richard C. Davis · Thomas A. Drake · Rick A. Friedman · Nick Furlotte · Simon T. Hui · J. David Jentsch · Emrah Kostem · Hyun Min Kang · Eun Yong Kang · Jong Wha Joo · Vyacheslav A. Korshunov · Rick E. Laughlin · Lisa J. Martin · Jeffrey D. Ohmen · Brian W. Parks · Matteo Pellegrini · Karen Reue · Desmond J. Smith · Sotirios Tetradis · Jessica Wang · Yibin Wang · James N. Weiss · Todd Kirchgessner · Peter S. Gargalovic · Eleazar Eskin · Aldons J. Lusis · Renée C. LeBoeuf

Received: 29 March 2012 / Accepted: 4 July 2012

Abstract We have developed an association-based approach using classical inbred strains of mice in which we correct for population structure, which is very extensive in mice, using an efficient mixed-model algorithm. Our approach includes inbred parental strains as well as recombinant inbred strains in order to capture loci with effect sizes typical of complex traits in mice (in the range of 5 % of total trait variance). Over the last few years, we have typed the hybrid mouse diversity panel (HMDP) strains for a variety of clinical traits as well as intermediate phenotypes and have shown that the HMDP has sufficient power to map genes for highly complex traits with resolution that is in most cases less than a megabase. In this essay, we review our experience with the HMDP, describe various ongoing projects, and discuss how the HMDP may fit into the larger picture of common diseases and different approaches.

Anatole Ghazalpour, Christoph D. Rau, Charles R. Farber, Brian J. Bennett, Luz D. Orozco, and Atila van Nas contributed equally to this work.

A. Ghazalpour · M. Civelek · R. C. Davis · S. T. Hui · B. W. Parks · A. J. Lusis
Department of Medicine, David Geffen School of Medicine, University of California, Los Angeles, CA, USA

C. D. Rau · J. N. Weiss · A. J. Lusis
Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, CA, USA

C. R. Farber
Departments of Medicine and Biochemistry and Molecular Genetics, and Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA

B. J. Bennett
Department of Genetics, and Nutrition Research Institute, University of North Carolina, Chapel Hill, NC, USA

L. D. Orozco · A. van Nas · C. Pan · J. W. Joo · L. J. Martin · K. Reue · J. Wang · E. Eskin · A. J. Lusis
Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA, USA

H. Allayee
Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

S. W. Beaven
Division of Digestive Diseases, Department of Medicine, David Geffen School of Medicine, University of California, Los Angeles, CA, USA

T. A. Drake
Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California, Los Angeles, CA, USA

R. A. Friedman
Department of Otology/Skull Base Surgery, House Research Institute, Los Angeles, CA, USA

N. Furlotte · E. Kostem · E. Y. Kang · E. Eskin
Department of Computer Sciences, University of California, Los Angeles, CA, USA

J. D. Jentsch
Department of Psychology & Behavioral Neuroscience and Psychiatry & Biobehavioral Sciences, David Geffen School of Medicine, University of California, Los Angeles, CA, USA

H. M. Kang
Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI, USA

Published online: 15 August 2012

Introduction

The human genome-wide association studies (GWAS) of the last several years have provided the first unbiased views of the genetics of common complex diseases, such as coronary artery disease, diabetes, and cancer. Many of the loci contain novel genes not previously connected to their respective disease, indicating that there is great potential to discover new pathways and new targets for therapeutic intervention. These GWAS do, however, have some important limitations. First, human GWAS are not well powered to study genetic interactions, such as gene-by-gene or gene-by-environment interactions (Zuk et al. 2012). Second, it will be difficult to move from locus to a disease pathway directly in humans (Altshuler et al. 2008). And third, for most diseases, GWAS have identified only a small fraction of the total genetic contributions and, thus, there is a great deal more to be discovered (Altshuler et al. 2008; Manolio et al. 2009).

To simplify genetic analysis, natural variations relevant to disease have been studied in mice and rats (Ahlqvist et al. 2011; Flint and Mackay 2009; Keane et al. 2011). This has generally involved traditional linkage mapping methods with crosses between different strains to identify quantitative trait loci (QTLs). An important problem with such analysis has been poor mapping resolution because the QTLs generally contain hundreds of genes, making the identification of the causal genes difficult.

To address these limitations, we have developed an association-based approach using classical inbred strains of mice (Bennett et al. 2010). We follow previous attempts to apply association in mice (Cervino et al. 2007; Grupe et al. 2001; Guo et al. 2007; Liao et al. 2004; Liu et al. 2007; Pletcher et al. 2004) with two differences. First, we correct for population structure, which is very extensive in mice, using an efficient mixed-model algorithm (EMMA) (Kang et al. 2008). Second, to capture loci with effect sizes typical of complex traits in mice (in the range of 5 % of total trait variance), we supplemented the population with recombinant inbred (RI) strains.

Over the last few years, we have typed the hybrid mouse diversity panel (HMDP) strains for a variety of clinical traits as well as intermediate phenotypes, and have shown that the HMDP has sufficient power to map genes for highly complex traits with resolution that is in most cases less than a megabase. In this essay, we review our experience with the HMDP, describe various ongoing projects, and discuss how the HMDP may fit into the larger picture of common diseases and different approaches.

Overview of the HMDP

The hybrid mouse diversity panel (HMDP) consists of a population of over 100 inbred mouse strains selected for usage in systematic genetic analyses of complex traits

J. W. Joo
Bioinformatics Program, David Geffen School of Medicine, University of California, Los Angeles, CA, USA

V. A. Korshunov
Department of Medicine, Aab Cardiovascular Research Institute, University of Rochester School of Medicine and Dentistry, Rochester, NY, USA

R. E. Laughlin
Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, CA, USA

J. D. Ohmen
Department of Cell Biology and Genetics, House Research Institute, Los Angeles, CA, USA

M. Pellegrini
Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, CA, USA

D. J. Smith
Department of Molecular and Medical Pharmacology, David Geffen School of Medicine, University of California, Los Angeles, CA, USA

S. Tetradis
Molecular Biology Institute, David Geffen School of Medicine, University of California, Los Angeles, CA, USA

S. Tetradis · A. J. Lusis
Division of Diagnostic and Surgical Science, School of Dentistry, University of California, Los Angeles, CA, USA

J. Wang
Department of Medicine/Cardiology, David Geffen School of Medicine, University of California, Los Angeles, CA, USA

Y. Wang
Division of Molecular Medicine, Department of Anesthesiology, Physiology and Medicine, David Geffen School of Medicine, University of California, Los Angeles, CA, USA

T. Kirchgessner · P. S. Gargalovic
Department of Cardiovascular Drug Discovery, Bristol-Myers Squibb Co, Pennington, NJ, USA

R. C. LeBoeuf (✉)
Division of Metabolism, Endocrinology and Nutrition, Department of Medicine, University of Washington, 850 Republican Street, Seattle, WA 98109-4725, USA
e-mail: leboeuf@u.washington.edu

(Table 1). Our goals in selecting the strains were to (1) increase resolution of genetic mapping, (2) have a renewable resource that is available to all investigators worldwide, and (3) provide a shared data repository that would allow the integration of data across multiple scales, including genomic, transcriptomic, metabolomic, proteomic, and clinical phenotypes. The core of our panel for association mapping (Bennett et al. 2010; Cervino et al. 2007; Grupe et al. 2001) consists of 29 classic parental inbred strains which are a subset of a group of mice commonly called the mouse diversity panel. We settled on our strains by eliminating closely related strains and removing wild-derived strains. The decision to remove wild-derived stains is based on the tradeoff between statistical power and genetic diversity. While we were sacrificing the genetic diversity by leaving out wild-derived strains, our panel increased the statistical power (assuming the same number of animals) to identify genetic variants polymorphic among the classical inbred strains which affect traits, and these variants account for a tremendous amount of phenotypic diversity among the classical inbred strains.

In order to increase power, we included panels of RI mice, including the BXD, CXB, BXA/AXB, and BXH panels. Power calculations with the inclusion of these additional strains indicated that we have 70 % power to detect SNPs that contribute ~10 % of the overall variance of a complex trait (Bennett et al. 2010). We have recently shown that power can be further increased by performing meta-analysis in which data from the HMDP are combined with data from traditional crosses (Furlotte et al. 2012). Power can also be increased by typing additional commercially available RI panels, as discussed below.

A key feature of the panel is that genotyping is not necessary due to the wealth of single nucleotide polymorphism (SNP) genotypes known across the mouse strains (Keane et al. 2011; Kirby et al. 2010). The inbred strains used for the HMDP were previously genotyped by the Broad Institute and then combined with genotypes from the Wellcome Trust Center for Human Genetics (WTCHG) (Table 2). Genotypes of RI strains at the Broad Institute were inferred from WTCHG genotypes by interpolating alleles at polymorphic SNPs among parental strains, calling ambiguous genotypes missing. Of the 140,000 SNPs available, 107,145 were informative with an allele frequency >5 % and were used for GWAS in our publications (Bennett et al. 2010; Farber et al. 2011; Park et al. 2011). Additional genotyping classifications have been performed recently (Keane et al. 2011; Kirby et al. 2010), and the resulting 4 million SNP genotypes are freely available (Table 2).

In the current HMDP panel consisting of over 100 inbred and RI strains, we used ~860,000 SNPs to examine linkage disequilibrium (LD) blocks. These SNPs have greater than 5 % minor allele frequency (mean and median minor allele frequency of 28 and 31 %) and are polymorphic between the strains. Using 0.8 as the r^2 cutoff to define LD, there are a total of 13,706 LD blocks with the median size of 42.8 kb per block (mean of 143.3 kb per block) scattered throughout the genome (Fig. 1). Since LD blocks are more likely to define the window in which a candidate gene for a locus resides, the presence of small LD in the HMDP suggests that the number of causal candidate genes will be on average fewer than five genes per locus. This is a considerable improvement over mapping resolution in traditional linkage studies and/or in outbred stock mice where, on average, the number of candidate genes for each locus ranges from 10 to 50. Large blocks (defined as LD blocks >1 Mb) are also present in the HMDP panel but not frequently. Only 1.5 % of the LD blocks (211 of 13,706 total) have a large size encompassing 12.5 % of the genome. The X chromosome is noted to contain multiple large LD blocks, suggesting that mapping resolution for this chromosome is reduced as compared to the autosomes. These large blocks reflect the regions in the genome that were inherited by all strains from the shared ancestors as described in (Frazer et al. 2007). The high-resolution mapping property of the HMDP becomes particularly important in systems genetics as one common goal is to identify causal genes that coordinately regulate network function. Such drivers have been reported in numerous studies as candidate genes for "QTL hot spots," but the true identity of such drivers has been elusive mainly due to lack of resolution in mapping.

In addition to the excellent resolution, the HMDP has important advantages for systems genetics and for analysis of genetic interactions. The progeny from a genetic cross are unique and as such can be characterized for a limited number of phenotypes, whereas the inbred strains of the HMDP can be examined for an unlimited number of phenotypes since the data are cumulative. The same concept applies to interactions. Thus, mice of the same genotype can be examined under a variety of conditions to identify gene-by-environment interactions, and epistatic interactions can be tested using targeted perturbations on specific genetic backgrounds. This is also an important feature of other replicate mouse genetic systems such as consomic strains, collaborative cross strains, and wild-derived inbred strains.

Discoveries using the HMDP

The successful use of the HMDP to identify complex trait genes was recently highlighted in a study of bone mineral density (BMD) (Farber et al. 2011). BMD is a polygenic phenotype that is commonly investigated in human and

Table 1 The 114 strains typed within the HMDP for metabolic phenotypes

	JAX ID	JAX strain name	EMMA name	JAX URL	Type
1	691	129X1/SvJ	129X1/SvJ	http://jaxmice.jax.org/strain/000691.html	Classic
2	646	A/J	A/J	http://jaxmice.jax.org/strain/000646.html	
3	648	AKR/J	AKR/J	http://jaxmice.jax.org/strain/000648.html	
4	651	BALB/cJ	BALB/cJ	http://jaxmice.jax.org/strain/000651.html	
5	2282	BTBR T<+> tf/J	BTBRT<+>tf/J	http://jaxmice.jax.org/strain/002282.html	
6	653	BUB/BnJ	BUB/BnJ	http://jaxmice.jax.org/strain/000653.html	
7	659	C3H/HeJ	C3H/HeJ	http://jaxmice.jax.org/strain/000659.html	
8	664	C57BL/6J	C57BL/6J	http://jaxmice.jax.org/strain/000664.html	
9	662	C57BLKS/J	C57BLKS/J	http://jaxmice.jax.org/strain/000662.html	
10	668	C57L/J	C57L/J	http://jaxmice.jax.org/strain/000668.html	
11	669	C58/J	C58/J	http://jaxmice.jax.org/strain/000669.html	
12	656	CBA/J	CBA/J	http://jaxmice.jax.org/strain/000656.html	
13	657	CE/J	CE/J	http://jaxmice.jax.org/strain/000657.html	
14	671	DBA/2J	DBA/2J	http://jaxmice.jax.org/strain/000671.html	
15	1800	FVB/NJ	FVB/NJ	http://jaxmice.jax.org/strain/001800.html	
16	674	I/LnJ	I/LnJ	http://jaxmice.jax.org/strain/000674.html	
17	2106	KK/HlJ	KK/HlJ	http://jaxmice.jax.org/strain/002106.html	
18	675	LG/J	LG/J	http://jaxmice.jax.org/strain/000675.html	
19	676	LP/J	LP/J	http://jaxmice.jax.org/strain/000676.html	
20	677	MA/MyJ	MA/MyJ	http://jaxmice.jax.org/strain/000677.html	
21	1976	NOD/ShiLtJ	NOD/LtJ	http://jaxmice.jax.org/strain/001976.html	
22	2423	NON/ShiLtJ	NON/LtJ	http://jaxmice.jax.org/strain/002423.html	
23	684	NZB/BINJ	NZB/BINJ	http://jaxmice.jax.org/strain/000684.html	
24	1058	NZW/LacJ	NZW/LacJ	http://jaxmice.jax.org/strain/001058.html	
25	680	PL/J	PL/J	http://jaxmice.jax.org/strain/000680.html	
26	683	RIIIS/J	RIIIS/J	http://jaxmice.jax.org/strain/000683.html	
27	644	SEA/GnJ	SEA/GnJ	http://jaxmice.jax.org/strain/000644.html	
28	686	SJL/J	SJL/J	http://jaxmice.jax.org/strain/000686.html	
29	687	SM/J	SM/J	http://jaxmice.jax.org/strain/000687.html	
30	689	SWR/J	SWR/J	http://jaxmice.jax.org/strain/000689.html	
31	1673	AXB1/PgnJ	AXB-1/PgnJ	http://jaxmice.jax.org/strain/001673.html	AXB/BXA
32	1681	AXB10/PgnJ	AXB-10/PgnJ	http://jaxmice.jax.org/strain/001681.html	
33	1683	AXB12/PgnJ	AXB-12/PgnJ	http://jaxmice.jax.org/strain/001683.html	
34	1826	AXB13/PgnJ	AXB-13/PgnJ	http://jaxmice.jax.org/strain/001826.html	
35	1685	AXB15/PgnJ	AXB-15/PgnJ	http://jaxmice.jax.org/strain/001685.html	
36	1687	AXB19/PgnJ	AXB-19/PgnJ	http://jaxmice.jax.org/strain/001687.html	
37	1686	AXB19a/PgnJ	AXB-18/PgnJ	http://jaxmice.jax.org/strain/001686.html	
38	1688	AXB19b/PgnJ	AXB-20/PgnJ	http://jaxmice.jax.org/strain/001688.html	
39	1674	AXB2/PgnJ	AXB-2/PgnJ	http://jaxmice.jax.org/strain/001674.html	
40	1690	AXB23/PgnJ	AXB-23/PgnJ	http://jaxmice.jax.org/strain/001690.html	
41	1691	AXB24/PgnJ	AXB-24/PgnJ	http://jaxmice.jax.org/strain/001691.html	
42	1676	AXB4/PgnJ	AXB-4/PgnJ	http://jaxmice.jax.org/strain/001676.html	
43	1677	AXB5/PgnJ	AXB-5/PgnJ	http://jaxmice.jax.org/strain/001677.html	
44	1678	AXB6/PgnJ	AXB-6/PgnJ	http://jaxmice.jax.org/strain/001678.html	
45	1679	AXB8/PgnJ	AXB-8/PgnJ	http://jaxmice.jax.org/strain/001679.html	
46	1692	BXA1/PgnJ	BXA-1/PgnJ	http://jaxmice.jax.org/strain/001692.html	
47	1699	BXA11/PgnJ	BXA-11/PgnJ	http://jaxmice.jax.org/strain/001699.html	

Table 1 continued

	JAX ID	JAX strain name	EMMA name	JAX URL	Type
48	1700	BXA12/PgnJ	BXA-12/PgnJ	http://jaxmice.jax.org/strain/001700.html	
49	1701	BXA13/PgnJ	BXA-13/PgnJ	http://jaxmice.jax.org/strain/001701.html	
50	1702	BXA14/PgnJ	BXA-14/PgnJ	http://jaxmice.jax.org/strain/001702.html	
51	1703	BXA16/PgnJ	BXA-16/PgnJ	http://jaxmice.jax.org/strain/001703.html	
52	1693	BXA2/PgnJ	BXA-2/PgnJ	http://jaxmice.jax.org/strain/001693.html	
53	1710	BXA24/PgnJ	BXA-24/PgnJ	http://jaxmice.jax.org/strain/001710.html	
54	1711	BXA25/PgnJ	BXA-25/PgnJ	http://jaxmice.jax.org/strain/001711.html	
55	1999	BXA26/PgnJ	BXA-26/PgnJ	http://jaxmice.jax.org/strain/001999.html	
56	1694	BXA4/PgnJ	BXA-4/PgnJ	http://jaxmice.jax.org/strain/001694.html	
57	1696	BXA7/PgnJ	BXA-7/PgnJ	http://jaxmice.jax.org/strain/001696.html	
58	1697	BXA8/PgnJ	BXA-17/PgnJ	http://jaxmice.jax.org/strain/001697.html	
59	1697	BXA8/PgnJ	BXA-8/PgnJ	http://jaxmice.jax.org/strain/001697.html	
60	36	BXD1/TyJ	BXD-1/TyJ	http://jaxmice.jax.org/strain/000036.html	BXD
61	12	BXD11/TyJ	BXD-11/TyJ	http://jaxmice.jax.org/strain/000012.html	
62	45	BXD12/TyJ	BXD-12/TyJ	http://jaxmice.jax.org/strain/000045.html	
63	40	BXD13/TyJ	BXD-13/TyJ	http://jaxmice.jax.org/strain/000040.html	
64	329	BXD14/TyJ	BXD-14/TyJ	http://jaxmice.jax.org/strain/000329.html	
65	95	BXD15/TyJ	BXD-15/TyJ	http://jaxmice.jax.org/strain/000095.html	
66	13	BXD16/TyJ	BXD-16/TyJ	http://jaxmice.jax.org/strain/000013.html	
67	15	BXD18/TyJ	BXD-18/TyJ	http://jaxmice.jax.org/strain/000015.html	
68	10	BXD19/TyJ	BXD-19/TyJ	http://jaxmice.jax.org/strain/000010.html	
69	75	BXD2/TyJ	BXD-2/TyJ	http://jaxmice.jax.org/strain/000075.html	
70	330	BXD20/TyJ	BXD-20/TyJ	http://jaxmice.jax.org/strain/000330.html	
71	77	BXD21/TyJ	BXD-21/TyJ	http://jaxmice.jax.org/strain/000077.html	
72	43	BXD22/TyJ	BXD-22/TyJ	http://jaxmice.jax.org/strain/000043.html	
73	31	BXD24/TyJ-Cep290<rd16>/J	BXD-24/TyJ	http://jaxmice.jax.org/strain/000031.html	
74	41	BXD27/TyJ	BXD-27/TyJ	http://jaxmice.jax.org/strain/000041.html	
75	47	BXD28/TyJ	BXD-28/TyJ	http://jaxmice.jax.org/strain/000047.html	
76	29	BXD29-Tlr4<lps-2J>/J	BXD-29/TyJ	http://jaxmice.jax.org/strain/000029.html	
77	83	BXD31/TyJ	BXD-31/TyJ	http://jaxmice.jax.org/strain/000083.html	
78	78	BXD32/TyJ	BXD-32/TyJ	http://jaxmice.jax.org/strain/000078.html	
79	3222	BXD33/TyJ	BXD-33/TyJ	http://jaxmice.jax.org/strain/003222.html	
80	3223	BXD34/TyJ	BXD-34/TyJ	http://jaxmice.jax.org/strain/003223.html	
81	3225	BXD36/TyJ	BXD-36/TyJ	http://jaxmice.jax.org/strain/003225.html	
82	3227	BXD38/TyJ	BXD-38/TyJ	http://jaxmice.jax.org/strain/003227.html	
83	3228	BXD39/TyJ	BXD-39/TyJ	http://jaxmice.jax.org/strain/003228.html	
84	3229	BXD40/TyJ	BXD-40/TyJ	http://jaxmice.jax.org/strain/003229.html	
85	3230	BXD42/TyJ	BXD-42/TyJ	http://jaxmice.jax.org/strain/003230.html	
86	37	BXD5/TyJ	BXD-5/TyJ	http://jaxmice.jax.org/strain/000037.html	
87	7	BXD6/TyJ	BXD-6/TyJ	http://jaxmice.jax.org/strain/000007.html	
88	84	BXD8/TyJ	BXD-8/TyJ	http://jaxmice.jax.org/strain/000084.html	
89	105	BXD9/TyJ	BXD-9/TyJ	http://jaxmice.jax.org/strain/000105.html	
90	3787	B6cC3-1/KccJ	BXHE1	http://jaxmice.jax.org/strain/003787.html	BXH
91	32	BXH10/TyJ	BXH-10/TyJ	http://jaxmice.jax.org/strain/000032.html	
92	9	BXH14/TyJ	BXH-14/TyJ	http://jaxmice.jax.org/strain/000009.html	
93	33	BXH19/TyJ	BXH-19/TyJ	http://jaxmice.jax.org/strain/000033.html	
94	34	BXH2/TyJ	BXH-2/TyJ	http://jaxmice.jax.org/strain/000034.html	

Table 1 continued

	JAX ID	JAX strain name	EMMA name	JAX URL	Type
95	3784	BXH20/KccJ	BXHA1	http://jaxmice.jax.org/strain/003784.html	
96	3786	BXH22/KccJ	BXHB2	http://jaxmice.jax.org/strain/003786.html	
97	11	BXH4/TyJ	BXH-4/TyJ	http://jaxmice.jax.org/strain/000011.html	
98	38	BXH6/TyJ	BXH-6/TyJ	http://jaxmice.jax.org/strain/000038.html	
99	14	BXH7/TyJ	BXH-7/TyJ	http://jaxmice.jax.org/strain/000014.html	
100	76	BXH8/TyJ	BXH-8/TyJ	http://jaxmice.jax.org/strain/000076.html	
101	8	BXH9/TyJ	BXH-9/TyJ	http://jaxmice.jax.org/strain/000008.html	
102	351	CXB1/ByJ	CXB-1/ByJ	http://jaxmice.jax.org/strain/000351.html	CxB
103	1631	CXB10/HiAJ	CXB-10/HiAJ	http://jaxmice.jax.org/strain/001631.html	
104	1632	CXB11/HiAJ	CXB-11/HiAJ	http://jaxmice.jax.org/strain/001632.html	
105	1633	CXB12/HiAJ	CXB-12/HiAJ	http://jaxmice.jax.org/strain/001633.html	
106	1634	CXB13/HiAJ	CXB-13/HiAJ	http://jaxmice.jax.org/strain/001634.html	
107	352	CXB2/ByJ	CXBE	http://jaxmice.jax.org/strain/000352.html	
108	353	CXB3/ByJ	CXB-3/ByJ	http://jaxmice.jax.org/strain/000353.html	
109	354	CXB4/ByJ	CXBH	http://jaxmice.jax.org/strain/000354.html	
110	355	CXB5/ByJ	CXB-5/ByJ	http://jaxmice.jax.org/strain/000355.html	
111	356	CXB6/ByJ	CXB-6/ByJ	http://jaxmice.jax.org/strain/000356.html	
112	357	CXB7/ByJ	CXB-7/ByJ	http://jaxmice.jax.org/strain/000357.html	
113	1629	CXB8/HiAJ	CXB-8/HiAJ	http://jaxmice.jax.org/strain/001629.html	
114	1630	CXB9/HiAJ	CXB-9/HiAJ	http://jaxmice.jax.org/strain/001630.html	

These strains are commercially available from The Jackson Laboratory (JAX) (Bar Harbor, ME). Additional strains will soon be entered into the HMDP data base (see Table 2). For entry of phenotypic measures into the HMDP database and for calculating associations using EMMA programming (see Table 2), the JAX strain names need to be converted to the EMMA strain names that we have provided here

Table 2 URL sites useful for the HMDP as well as sites used to develop the HMDP, including sites at The Jackson Laboratory (JAX) (Bar Harbor, ME)

Name	Description	URL
EMMA	Obtain EMMA in an R package	http://mouse.cs.ucla.edu/emma/
EMMA server	Input data and obtain analysis	http://mouse.cs.ucla.edu/emmaserver/
EMMA power simulator	R package allowing statistical power experiment via in silico mapping	http://mouse.cs.ucla.edu/power/
EMMA study design Webserver	Provides power simulations with various background genetic effects and threshold estimations	http://mouse.cs.ucla.edu/emmaserver/powerSimulation/
Mouse HapMap	Latest genetic maps that are regularly updated	http://mouse.cs.ucla.edu/mousehapmap/
Mouse genome informatics (JAX)	Search tools for gene names and chromosomal positions, phenotypes, expression, orthology, and other features	http://www.informatics.jax.org/
SGR HMDP database	Database for HMDP and several F2 studies with cQTL, eQTL, heatmap and correlation analyses	http://systems.genetics.ucla.edu
Genenetwork	Study relationships among phenotypes and genotypes	http://www.genenetwork.org/webqtl/main.py
Mouse phenome database (JAX)	Phenotypes across hundreds of strains contributed by over 200 investigators	http://phenome.jax.org/
Wellcome trust center for human genetics	Center to understand the genetic foundations of human variation and disease	http://www.well.ox.ac.uk/home
Perlegen	Perlegen mouse SNP browser covering 12 classical inbred strains and 4 wild-derived strains	http://mouse.cs.ucla.edu/perlegen/
Mouse genomes project	Nucleotide sequence database for many key mouse strains	http://sanger.ac.uk/resources/mouse/genomes/

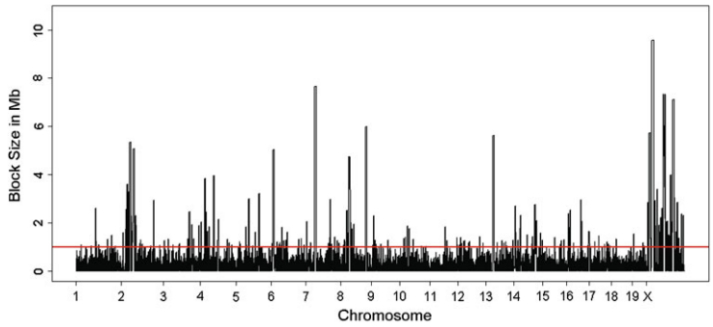

Fig. 1 Map of linkage disequilibrium (LD) blocks along the genome in the hybrid mouse diversity panel (HMDP) population. The LD blocks were determined using ~860,000 SNPs that have >5 % minor allele frequency (mean and median minor allele frequency of 28 and 31 %) and are polymorphic between the strains. Using 0.8 as the r^2 cutoff to define LD, there are a total of 13,706 LD blocks with a median size of 42.8 kb per block (mean of 143.3 kb per block) scattered throughout the genome. Only 1.5 % of the LD blocks (211 of 13,706 total) have a large size encompassing 12.5 % of the genome. The location of the large blocks in the genome can be identified by lines that cross the *red bar*. See text for more details

rodent genetic studies and is the single strongest predictor of osteoporotic fracture (Cummings et al. 2002; Farber and Rosen 2010). For this study, total body, spine, and femur areal BMD data were generated on 16 week-old male mice from 96 HMDP strains. The whole bone transcriptome was also profiled using Illumina gene expression microarrays. The authors used EMMA to perform genome-wide association for the three BMD measures. A total of four genome-wide significant associations were identified on Chromosomes (Chrs) 7, 11, 12, and 17, each affecting BMD at one or more sites. The Chr 12 association for total-body BMD was chosen for further analysis since the 3 Mb window surrounding the association contained only 14 candidate genes. Interestingly, the most significant association was with a nonsynonymous SNP (rs29131970) in the additional sex-combs like 2 (*Asxl2*) gene (Fig. 2). This polymorphism was predicted to have deleterious effects on ASXL2 protein function. To gain further support for *Asxl2* being the causal gene, existing human genome-wide association data (generated in ~6,000 Icelandic subjects) was used to evaluate SNPs within the human syntenic region for association with BMD (Styrkarsdottir et al. 2008). One SNP (rs7563012) was significant after Bonferroni correction and was located in intron 3 of the human *ASXL2* gene (Fig. 2). Together these data suggested that *Asxl2* influenced BMD in both humans and mice. Consistent with this hypothesis, BMD was found to be lower in $Asxl2^{-/-}$ knockout mice.

Network analysis in the HMDP is another powerful approach for investigating complex traits from a systems-level perspective. Coexpression networks can be used to annotate genes of unknown function based on the known functions of the genes to which they are most closely connected. This "guilt by association" approach has been shown to be a robust gene annotation tool (Wolfe et al. 2005) and was used to determine the mechanism through which *Asxl2* influenced BMD. Weighted Gene Co-expression Network Analysis (WGNCA) was first used to generate a coexpression network using the bone microarray data, which identified *Asxl2* as being connected to genes involved in myeloid cell differentiation. In bone, osteoclasts are bone-resorbing cells of myeloid origin (Teitelbaum and Ross 2003). Additionally, in a human protein-protein interaction network, ASXL2 interacts with TRAF6, a key component of the major signaling pathway regulating osteoclastogenesis (Teitelbaum and Ross 2003). Thus, based on network inferences, *Asxl2* was predicted to be involved in the differentiation of osteoclasts. To test this prediction, expression of *Asxl2* was knocked down in osteoclast precursors. An ~50 % reduction in *Asxl2* transcript levels inhibited the formation of TRAP+ (a marker of mature osteoclasts) multinuclear cells. These data suggested that *Asxl2* influences BMD, at least in part, through its regulation of osteoclastogenesis. This work highlights the ability of using the HMDP and systems genetics to move from association to gene to mechanism in a single step.

In another study, (Park et al. 2011) utilized the HMDP resource to report on gene networks associated with conditional fear. In this study, the authors combined behavioral phenotypes with gene expression data in two regions of the brain (striatum and hippocampus) to identify groups of genes that coordinately regulate the behavior of the animals. Overall, they observed significant overlap between

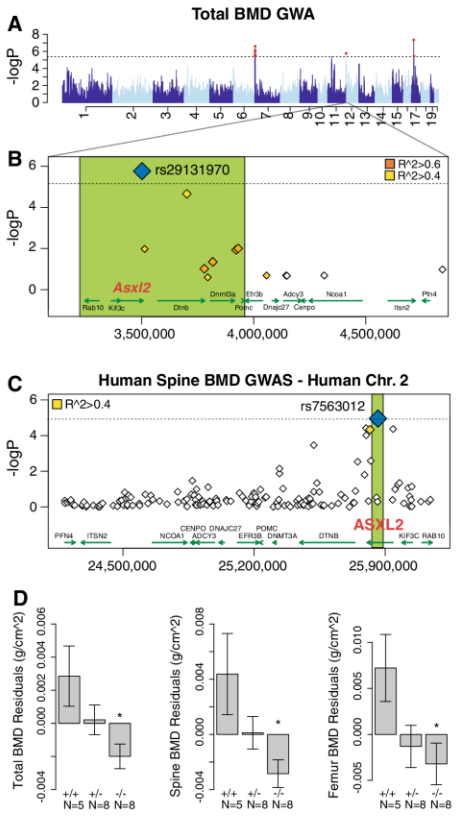

Fig. 2 Variation in *Asxl2* in mice and humans is associated with bone mineral density (BMD). **a** Genome-wide association in the HMDP for total BMD identifies an association on chromosome (Chr) 12. **b** A nonsynonymous SNP (rs29131970) in *Asxl2* that was predicted to alter protein function was the most significantly associated Chr 12 SNP in the HMDP. **c** Human SNPs within ASXL2 were also associated with BMD in ~6,000 Icelandic individuals. **d** Male mice deficient in *Asxl2* (−/−) display significant decreases relative to wild-type controls (+/+) in total BMD, spine BMD, and femur BMD residuals after adjustments for age and body weight. Data shown in **d** are residual mean ± SEM, $^*P < 0.05$

the local QTLs and QTL hotspots in the two tissues, as well as module conservation and preservation of highly connected genes (also known as "hubs") in the striatum and hippocampus networks. The authors were also able to identify tissue-specific network modules between the striatum and the hippocampus, and after performing functional enrichment analysis of the modules, they arrived at pathways likely to contribute to the differences in hippocampus and striatum function. Finally, using modules as functional units, the authors were able to demonstrate correlations with behavioral traits, thus helping to prioritize candidate genes and pathways for behavioral traits.

In addition to identifying cellular mechanisms and genes underlying physiological traits, the HMDP has been used to investigate the relationships across various biological scales at the global level (Ghazalpour et al. 2011). For example, among the intermediate phenotypes that have been examined in liver are transcript levels (in triplicate, using the Affymetrix platform) and a set of peptides corresponding to about 1,000 proteins (using quantitative mass spectrometry analysis) (Ghazalpour et al. 2011). The correlation between protein and transcript levels was quite weak, with a correlation coefficient of less than 0.4 in most cases, similar to what has been observed in previous studies with yeast and worms. More surprising was the finding that transcript levels were much more strongly correlated with clinical traits (primarily metabolic) than were protein levels. One possible explanation is that transcript levels may be reactive rather than causal with respect to physiologic traits (Ghazalpour et al. 2011).

Overall, the HMDP is being used to develop a multiscale understanding of a number of complex traits, including a recent report on elevated heart rate (Smolock et al. 2012). There are already over 70 traditional clinical traits reported and there are ongoing studies related to diet-induced obesity, hearing loss, heart failure, atherosclerosis, lipoprotein metabolism, bone metabolism, vascular injury, hematopoietic stem cells, air pollution, gut flora, addictive behavior, hepatotoxicity, and diabetic complications. In addition, gene expression microarrays have been used to quantify mRNA levels in liver, bone, adipose, brain, peritoneal macrophages, aorta, and heart, and proteomic and metabolomic profiling has been performed in liver.

Integration of the HMDP with other resources

The HMDP is just one of several recently proposed approaches to improve the resolution of mouse genetic studies. Other approaches include the Collaborative Cross (CC) (Churchill et al. 2004), outbred designs (Valdar et al. 2006; Yalcin et al. 2010), and the use of consomic strains (Gregorova et al. 2008; Singer et al. 2004; Takada et al. 2008). Each approach has advantages and disadvantages relative to the HMDP. The CC is a recently developed panel of RI strains that are descendants from eight founder strains. A key difference between the HMDP and the CC is that three of the CC founders are wild-derived strains. Wild-derived strains introduce a significantly larger amount of genetic variation and corresponding phenotypic variation compared to the HMDP. For this reason, it is

likely that more genes in the CC will have effects on traits than in the HMDP. However, the additional variation may make it relatively more difficult to map quantitative loci polymorphic in both panels since the increased variation of the CC and outbred panels will reduce the relative effect size of the same variant compared to the HMDP (Kang et al. 2008). The CC will ultimately contain approximately 300 strains compared to the current 100 strains of the HMDP. However, the HMDP may be enlarged to 260 strains (see Future directions and conclusions section) (Collaborative Cross Consortium 2012). Both the CC and the HMDP use inbred strains so they share the advantage of accumulation of data on each strain over time as more and more studies are performed. Utilizing inbred strains also facilitates performing studies with perturbations because identical animals can be phenotyped both with and without a perturbation.

An advantage of outbred designs is that they have higher resolution than the HMDP (up to 100 kb). On the other hand, a disadvantage in outbred designs, either utilizing a specially designed Heterogeneous Stock (Valdar et al. 2006) or commercially available outbred mice (Yalcin et al. 2010), is that each animal is unique. Overall, several mapping strategies are, or will be, available to tackle the genetics of complex diseases.

Consomic strains are also proving valuable for reducing genetic intervals containing candidate genes (Hoover-Plow et al. 2006; Prows et al. 2008) and for identifying causal genes (Burrage et al. 2010). However, consomic strains have limited uses for initial genomic studies, as only two alleles are sampled and consomic strains by definition consist of large areas of LD which were captured from the donor strain during breeding. Nonetheless, consomic strains have enriched the ability to test for epistasis and to reduce genetic intervals by the generation of congenic strains.

Resources for design and analysis of HMDP

UCLA maintains several resources for the design and analysis of HMDP studies (Table 2). The two most relevant to investigators are the EMMA association webserver and EMMA design webserver. Investigators utilizing the HMDP can upload their collected phenotypes to the association webserver which will perform association mapping using EMMA and return population structure-corrected *P* values for each SNP. The analysis is performed on a high-performance computing infrastructure at UCLA, eliminating the need for investigators applying the HMDP to invest in computational resources to perform the analysis. The EMMA design webserver allows an investigator to estimate the power of a proposed study design through simulations (Kirby et al. 2010) that can help guide an investigator in the design of HMDP studies. In addition, curated genotypes of the HMDP strains are also available.

A systems genetics database

Because the HMDP mice are inbred, with fixed genotypes, the data generated from their study are cumulative. To facilitate the integration and analysis of such data, we also developed a database called the Systems Genetics Resource (SGR). The database comprises mouse genomic, transcriptomic, metabolomic, proteomic, and clinical trait data from the HMDP as well as selected traditional mouse crosses and several human studies. The data are accompanied by detailed descriptions of how the data were acquired, with protocols and links to related published papers. A summary of current data sets contained in the database is presented in Table 3.

We developed a web-based interface where data can be queried for information on specific gene and trait correlations, gene expression in various tissues, or quantitative trait loci, as well as be downloaded for other types of analyses. Such information can be used, for example, to prioritize candidate genes in genetic studies and *trans*-acting loci can be used to generate hypotheses about regulatory pathways (Fig. 3). The intermediate phenotypes can also be used to model gene networks and causal interactions. The power of the SGR is expected to expand as more data are added. Some of the data is also available in the Genenetwork and the Mouse Genome Informatics databases (Table 2).

The SGR is a resource that can be used to answer specific questions and to understand the relationships among different genes. For example, data generated from primary macrophages of the HMDP helped us to determine that a gene of interest, the interferon inducible helicase 1 (*Ifih1*), shows gene-by-environment interactions and that it is under the control of the inflammatory stimulus bacterial lipopolysaccharide (LPS) (Fig. 3a). Using eQTL, we also found that the expression of *Ifih1* is controlled by three loci on Chrs 5, 8, and 13 (Fig. 3b). We can also compare the expression patterns of *Ifih1* among the different tissues available on the database, which include adipose, aorta, heart, liver, and macrophages treated in three different conditions (Fig. 3c). Similarly, we found that the three *trans*-eQTL on Chrs 5, 8, and 13 are specific to macrophages treated with LPS, but there is also a strong *cis*-eQTL in the liver and other regulatory loci in Chr 2. Such information can allow us to identify candidate regulators, to examine gene-by-environment interactions and tissue specificities, and to prioritize candidate genes for clinical QTL.

Table 3 Data sets currently on line on the systems genetics resource (SGR) database

(A) Hybrid mouse diversity panel: original panel analysis: chow fed, males, 16 weeks
(B) Hybrid mouse diversity panel: chow fed, males, 16 weeks: second set
(C) Hybrid mouse diversity panel: isoproterenol-induced hypertrophy and heart failure
(D) Hybrid mouse diversity panel: Akita Diabetes Study, pilot
(E) Hybrid mouse diversity panel: diet-induced obesity
(F) Hybrid mouse diversity panel: human ApoB100 transgenics
(G) Hybrid mouse diversity panel: peritoneal macrophages, inflammatory responses
(H) Hybrid mouse diversity panel: atherosclerosis using ApoE Leiden, CETP transgenics
(I) Hybrid mouse diversity panel: hearing phenotypes
(J) Hybrid mouse diversity panel: mandible morphology
(K) Hybrid mouse diversity panel: attention-related behavioral traits
(L) Hybrid mouse diversity panel: parental strain survey
(M) Hybrid mouse diversity panel: vascular injury response
(N) Hybrid mouse diversity panel: F1 hybrids
(O) Metabolic syndrome in men (Metsim) study
(P) Human aortic endothelial cells (EC) culture
(Q) Human aortic smooth muscle cell (SMC) culture
(R) (BALB/cBy.Ldlr$^{-/-}$ × C57BL/6JLdlr$^{-/-}$) F2
(S) (C57BL/6J × DBA/2J)F2 on db/db background
(T) (CeH/HeJ.Apoe$^{-/-}$ × C57BL/6JApoe$^{-/-}$) F2
(U) (C3H/HeJ × C57BL/6J) F2
(V) (C57BL/6J × CAST/Ei) F2
(W) Genome tagged mice, global congenic strains: DBA/2J on C57BL/6J background and CAST/Ei on C57BL/6J background
(X) Recombinant inbred congenic strains, Demant, C3H × C57BL/6
(Y) (C57BL/6J × DBA/2J) F2

This resource is updated regularly with submissions from the HMDP user groups

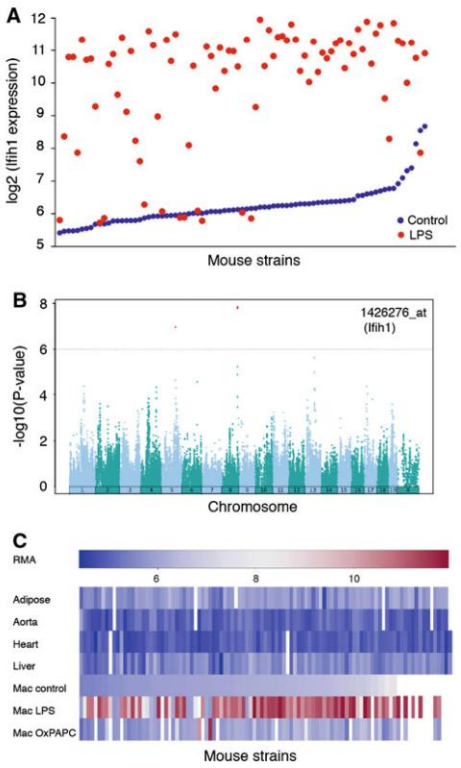

Fig. 3 Database plots for interferon-inducible helicase 1 (*Ifih1*). Sample plots for a given gene of interest that can be obtained from our online database. **a** Lipopolysaccharide (LPS) response of *Ifih1* in macrophages of the HMDP. **b** Genome-wide association for the expression of *Ifih1* in LPS-treated macrophages. **c** Relative expression levels among mouse strains of the HMDP in adipose, aorta, heart, and liver, and for macrophages treated with control, LPS or oxidation products of 1-palmitoyl-2-arachidonyl-*sn*-glycero-3-phosphocholine (OxPAPC) media. Robust microarray average (RMA) refers to an algorithm for gene expression microarray background corrections. We used the Affymetrix GCOS RMA algorithm

Future directions and conclusions

While the resolution of the HMDP is excellent at most loci, the power is marginal. As judged by QTL studies, few loci contributing to complex clinical traits have effect sizes as large as 10 % and most are below 5 % (Flint and Mott 2008). Thus, using the panel of 100 strains (Bennett et al. 2010), only a subset of the loci contributing to complex traits are likely to be identified. As mentioned above, the power can be enhanced by integrating the results from traditional crosses or by expanding the number of inbred and recombinant inbred strains. Recently, two panels of advanced intercross RI lines have become available from The Jackson Laboratory: The LXS RI panel includes 62 strains and 50 more BXD strains have recently been developed. Also available from The Jackson Laboratory are several sets of cryopreserved strains: 19 strains from the AKXD RI panel, 13 from the AKXL panel, and 15 from the NXSM panel. Thus, it is possible to increase the size of the HMDP to over 260 strains. Recombinant Inbred Congenic Lines, Chromosome Substitution Strains, and Genome Tagged Mice (Peters et al. 2007) could also be employed to increase both power and resolution. A particularly useful complement to the HMDP will be the CC strain set now being generated (Casci 2012; Threadgill and

Churchill 2012). Preliminary analyses of the partially inbred CC lines have been promising and substantial resources for their characterization, including complete genomic sequencing, have been planned (Collaborative Cross Consortium 2012).

As discussed above, human genetic studies of complex traits are limited in several respects, and the HMDP provides a partial solution to some of these limitations. First, human studies are poorly powered to identify genetic interactions, and these can be addressed more effectively in mice. As mentioned above, the HMDP is very convenient for genetic analysis of environmental interactions because mice of the same genotype can be examined under different conditions. Also, epistasis is likely to complicate studies of diseases such as diabetic complications and atherosclerosis, where one set of genes contributes to a predisposing factor (diabetes and elevated cholesterol, respectively), and another set of genes affects the response to these factors. In the HMDP, sensitizing genes can be introduced by breeding dominant mutations onto each of the HMDP strains and examining the F1 progeny. For example, we have bred a dominant hyperlipidemia-inducing gene, *APOE-Leiden*, onto a number of the HMDP strains and find atherosclerosis to be concordant with previous studies in which recessive mutations were transferred onto different backgrounds (B. Bennett and A. J. Lusis unpublished). Second, it will be difficult to identify, directly in humans, the pathways perturbed by novel GWAS genes. For example, the striking relationship between an allele of APOE and Alzheimer's has been known for nearly 20 years and yet the mechanism remains uncertain. Clearly, studies in mice, where access to tissues and environmental conditions can be standardized, will simplify such analyses. Moreover, studying the genes in the context of natural variation, as opposed to transgenic or gene-targeted mice, may well offer important advantages. Third, for most common disease traits, human GWAS have been able to identify only a small fraction of the total heritability. There are undoubtedly many explanations but, clearly, much remains to be discovered. Studies in mice will most likely identify different, although overlapping, gene sets and perhaps different pathways. Traits that are substantially influenced by environmental factors will be addressed with greater power in mice because such factors can be controlled.

The HMDP panel should be useful as a tool for investigation of basic biological processes as well as complex clinical traits. An example is the study by (Ghazalpour et al. 2011) that examines the relationship between transcript levels, protein levels, and metabolic traits. By providing many thousands of genetic perturbations in various combinations, the HMDP enables global dissection of the relationships between biological scales such as DNA methylation, transcription factor binding, histone modification, and transcription.

Acknowledgments Funding for this work was supported in part by the U.S. National Institutes of Health (NIH) grants NIAMS RO1 AR057759 (CRF) and NHLBI K99 HL102223 (BJB); the Ruth L. Kirschstein National Research Service Awards T32 HL69766 (CDR), T32-HL007895 (JW), NINDS R01 NS050148 and NIGMS R01 GM098273 (DJS), NHLBI RO1 HL098227 (RCL), NIDDK DP3 DK094311 (AJL, RCD, RCL), NIH RO1 GM095656 (MP), NIDDK RO1 DK072206 (TAD), RO1 DK080339 (STH), 5RO1 DC010856-02 (JDO, RAF), PO1 HL028481 (KR), and PO1 HL090553 (KR). EE was supported by National Science Foundation grants 0513612, 0731455, 0729049, 0916676, and 1065276, and NIH grants K25-HL080079, U01 DA024417, P01-HL30568, and PO1-HL28481. LDO was supported by USPHS National Research Service Award GM07104 and USPHS National Research Service Award T32-HG002536. This research was supported in part by the University of California, Los Angeles, CA, subcontract of contract N01-ES-45530 from the National Toxicology Program and National Institute of Environmental Health Sciences to Perlegen Sciences (EE) and from the Russian Program #2011-1.5-501-006-024 (VAK) and NIH RO1 HL105623 (VAK).

References

Ahlqvist E, Ekman D, Lindvall T, Popovic M, Forster M, Hultqvist M, Klaczkowska D, Teneva I, Johannesson M, Flint J, Valdar W, Nandakumar KS, Holmdahl R (2011) High-resolution mapping of a complex disease, a model for rheumatoid arthritis, using heterogeneous stock mice. Hum Mol Genet 20:3031–3041

Altshuler D, Daly MJ, Lander ES (2008) Genetic mapping in human disease. Science 322:881–888

Bennett BJ, Farber CR, Orozco L, Kang HM, Ghazalpour A, Siemers N, Neubauer M, Neuhaus I, Yordanova R, Guan B, Truong A, Yang WP, He A, Kayne P, Gargalovic P, Kirchgessner T, Pan C, Castellani LW, Kostem E, Furlotte N, Drake TA, Eskin E, Lusis AJ (2010) A high-resolution association mapping panel for the dissection of complex traits in mice. Genome Res 20:281–290

Burrage LC, Baskin-Hill AE, Sinasac DS, Singer JB, Croniger CM, Kirby A, Kulbokas EJ, Daly MJ, Lander ES, Broman KW, Nadeau JH (2010) Genetic resistance to diet-induced obesity in chromosome substitution strains of mice. Mamm Genome 21:115–129

Casci T (2012) Mouse genetics: fruits of the Collaborative Cross. Nat Rev Genet 13:223

Cervino AC, Darvasi A, Fallahi M, Mader CC, Tsinoremas NF (2007) An integrated in silico gene mapping strategy in inbred mice. Genetics 175:321–333

Churchill GA, Airey DC, Allayee H, Angel JM, Attie AD, Beatty J, Beavis WD, Belknap JK, Bennett B, Berrettini W, Bleich A, Bogue M, Broman KW, Buck KJ, Buckler E, Burmeister M, Chesler EJ, Cheverud JM, Clapcote S, Cook MN, Cox RD, Crabbe JC, Crusio WE, Darvasi A, Deschepper CF, Doerge RW, Farber CR, Forejt J, Gaile D, Garlow SJ, Geiger H, Gershenfeld H, Gordon T, Gu J, Gu W, de Haan G, Hayes NL, Heller C, Himmelbauer H, Hitzemann R, Hunter K, Hsu HC, Iraqi FA, Ivandic B, Jacob HJ, Jansen RC, Jepsen KJ, Johnson DK, Johnson TE, Kempermann G, Kendziorski C, Kotb M, Kooy RF, Llamas B, Lammert F, Lassalle JM, Lowenstein PR, Lu L, Lusis A, Manly KF, Marcucio R, Matthews D, Medrano JF, Miller DR, Mittleman G, Mock BA, Mogil JS, Montagutelli X, Morahan G, Morris DG, Mott R, Nadeau JH, Nagase H, Nowakowski RS, O'Hara BF, Osadchuk AV, Page GP, Paigen B, Paigen K, Palmer AA, Pan HJ, Peltonen-Palotie L, Peirce J, Pomp D, Pravenec M, Prows DR, Qi Z, Reeves RH, Roder J, Rosen GD, Schadt EE, Schalkwyk LC, Seltzer Z, Shimomura K, Shou S, Sillanpaa MJ, Siracusa LD, Snoeck HW, Spearow JL, Svenson

K, Tarantino LM, Threadgill D, Toth LA, Valdar W, de Villena FP, Warden C, Whatley S, Williams RW, Wiltshire T, Yi N, Zhang D, Zhang M, Zou F (2004) The Collaborative Cross, a community resource for the genetic analysis of complex traits. Nat Genet 36:1133–1137

Collaborative Cross Consortium (2012) The genome architecture of the collaborative cross mouse genetic reference population. Genetics 190:389–401

Cummings SR, Bates D, Black DM (2002) Clinical use of bone densitometry: scientific review. JAMA 288:1889–1897

Farber CR, Rosen C (2010) Genetics of osteoporosis. In: Robertson RP (ed) Translational endocrinology & metabolism. The Endocrine Society, Chevy Chase, pp 87–116

Farber CR, Bennett BJ, Orozco L, Zou W, Lira A, Kostem E, Kang HM, Furlotte N, Berberyan A, Ghazalpour A, Suwanwela J, Drake TA, Eskin E, Wang QT, Teitelbaum SL, Lusis AJ (2011) Mouse genome-wide association and systems genetics identify Asxl2 as a regulator of bone mineral density and osteoclastogenesis. PLoS Genet 7:e1002038

Flint J, Mackay TF (2009) Genetic architecture of quantitative traits in mice, flies, and humans. Genome Res 19:723–733

Flint J, Mott R (2008) Applying mouse complex-trait resources to behavioural genetics. Nature 456:724–727

Frazer KA, Eskin E, Kang HM, Bogue MA, Hinds DA, Beilharz EJ, Gupta RV, Montgomery J, Morenzoni MM, Nilsen GB, Pethiyagoda CL, Stuve LL, Johnson FM, Daly MJ, Wade CM, Cox DR (2007) A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. Nature 448:1050–1053

Furlotte NA, Kang EY, Van Nas A, Farber CR, Lusis AJ, Eskin E (2012) Increasing association mapping power and resolution in mouse genetic studies through the use of meta-analysis for structured populations. Genetics 191(3):959–967

Ghazalpour A, Bennett B, Petyuk VA, Orozco L, Hagopian R, Mungrue IN, Farber CR, Sinsheimer J, Kang HM, Furlotte N, Park CC, Wen PZ, Brewer H, Weitz K, Camp DG 2nd, Pan C, Yordanova R, Neuhaus I, Tilford C, Siemers N, Gargalovic P, Eskin E, Kirchgessner T, Smith DJ, Smith RD, Lusis AJ (2011) Comparative analysis of proteome and transcriptome variation in mouse. PLoS Genet 7:e1001393

Gregorova S, Divina P, Storchova R, Trachtulec Z, Fotopulosova V, Svenson KL, Donahue LR, Paigen B, Forejt J (2008) Mouse consomic strains: exploiting genetic divergence between *Mus m. musculus* and *Mus m. domesticus* subspecies. Genome Res 18:509–515

Grupe A, Germer S, Usuka J, Aud D, Belknap JK, Klein RF, Ahluwalia MK, Higuchi R, Peltz G (2001) In silico mapping of complex disease-related traits in mice. Science 292:1915–1918

Guo Y, Lu P, Farrell E, Zhang X, Weller P, Monshouwer M, Wang J, Liao G, Zhang Z, Hu S, Allard J, Shafer S, Usuka J, Peltz G (2007) In silico and in vitro pharmacogenetic analysis in mice. Proc Natl Acad Sci USA 104:17735–17740

Hoover-Plow J, Shchurin A, Hart E, Sha J, Hill AE, Singer JB, Nadeau JH (2006) Genetic background determines response to hemostasis and thrombosis. BMC Blood Disord 6:6

Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E (2008) Efficient control of population structure in model organism association mapping. Genetics 178:1709–1723

Keane TM, Goodstadt L, Danecek P, White MA, Wong K, Yalcin B, Heger A, Agam A, Slater G, Goodson M, Furlotte NA, Eskin E, Nellaker C, Whitley H, Cleak J, Janowitz D, Hernandez-Pliego P, Edwards A, Belgard TG, Oliver PL, McIntyre RE, Bhomra A, Nicod J, Gan X, Yuan W, van der Weyden L, Steward CA, Bala S, Stalker J, Mott R, Durbin R, Jackson IJ, Czechanski A, Guerra-Assuncao JA, Donahue LR, Reinholdt LG, Payseur BA, Ponting CP, Birney E, Flint J, Adams DJ (2011) Mouse genomic variation and its effect on phenotypes and gene regulation. Nature 477:289–294

Kirby A, Kang HM, Wade CM, Cotsapas C, Kostem E, Han B, Furlotte N, Kang EY, Rivas M, Bogue MA, Frazer KA, Johnson FM, Beilharz EJ, Cox DR, Eskin E, Daly MJ (2010) Fine mapping in 94 inbred mouse strains using a high-density haplotype resource. Genetics 185:1081–1095

Liao G, Wang J, Guo J, Allard J, Cheng J, Ng A, Shafer S, Puech A, McPherson JD, Foernzler D, Peltz G, Usuka J (2004) In silico genetics: identification of a functional element regulating H2-Ealpha gene expression. Science 306:690–695

Liu P, Vikis H, Lu Y, Wang D, You M (2007) Large-scale in silico mapping of complex quantitative traits in inbred mice. PLoS One 2:e651

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM (2009) Finding the missing heritability of complex diseases. Nature 461:747–753

Park CC, Gale GD, de Jong S, Ghazalpour A, Bennett BJ, Farber CR, Langfelder P, Lin A, Khan AH, Eskin E, Horvath S, Lusis AJ, Ophoff RA, Smith DJ (2011) Gene networks associated with conditional fear in mice identified using a systems genetics approach. BMC Syst Biol 5:43

Peters LL, Robledo RF, Bult CJ, Churchill GA, Paigen BJ, Svenson KL (2007) The mouse as a model for human biology: a resource guide for complex trait analysis. Nat Rev Genet 8:58–69

Pletcher MT, McClurg P, Batalov S, Su AI, Barnes SW, Lagler E, Korstanje R, Wang X, Nusskern D, Bogue MA, Mural RJ, Paigen B, Wiltshire T (2004) Use of a dense single nucleotide polymorphism map for in silico mapping in the mouse. PLoS Biol 2:e393

Prows DR, Hafertepen AP, Winterberg AV, Gibbons WJ Jr, Wesselkamper SC, Singer JB, Hill AE, Nadeau JH, Leikauf GD (2008) Reciprocal congenic lines of mice capture the aliq1 effect on acute lung injury survival time. Am J Respir Cell Mol Biol 38:68–77

Singer JB, Hill AE, Burrage LC, Olszens KR, Song J, Justice M, O'Brien WE, Conti DV, Witte JS, Lander ES, Nadeau JH (2004) Genetic dissection of complex traits with chromosome substitution strains of mice. Science 304:445–448

Smolock EM, Ilyushkina IA, Ghazalpour A, Gerloff J, Murashev AN, Lusis AJ, Korshunov VA (2012) A genetic locus on mouse chromosome 7 controls elevated heart rate. Physiol Genomics 44(13):689–698

Styrkarsdottir U, Halldorsson BV, Gretarsdottir S, Gudbjartsson DF, Walters GB, Ingvarsson T, Jonsdottir T, Saemundsdottir J, Center JR, Nguyen TV, Bagger Y, Gulcher JR, Eisman JA, Christiansen C, Sigurdsson G, Kong A, Thorsteinsdottir U, Stefansson K (2008) Multiple genetic loci for bone mineral density and fractures. N Engl J Med 358:2355–2365

Takada T, Mita A, Maeno A, Sakai T, Shitara H, Kikkawa Y, Moriwaki K, Yonekawa H, Shiroishi T (2008) Mouse inter-subspecific consomic strains for genetic dissection of quantitative complex traits. Genome Res 18:500–508

Teitelbaum SL, Ross FP (2003) Genetic regulation of osteoclast development and function. Nat Rev Genet 4:638–649

Threadgill DW, Churchill GA (2012) Ten years of the collaborative cross. G3 (Bethesda) 2:153–156

Valdar W, Solberg LC, Gauguier D, Burnett S, Klenerman P, Cookson WO, Taylor MS, Rawlins JN, Mott R, Flint J (2006) Genome-wide genetic association of complex traits in heterogeneous stock mice. Nat Genet 38:879–887

Wolfe CJ, Kohane IS, Butte AJ (2005) Systematic survey reveals general applicability of "guilt-by-association" within gene coexpression networks. BMC Bioinformatics 6:227

Yalcin B, Nicod J, Bhomra A, Davidson S, Cleak J, Farinelli L, Osteras M, Whitley A, Yuan W, Gan X, Goodson M, Klenerman P, Satpathy A, Mathis D, Benoist C, Adams DJ, Mott R, Flint J (2010) Commercially available outbred mice for genome-wide association studies. PLoS Genet 6(9):e1001085

Zuk O, Hechter E, Sunyaev SR, Lander ES (2012) The mystery of missing heritability: genetic interactions create phantom heritability. Proc Natl Acad Sci USA 109:1193–1198

Chapter 3

Genome Wide Association Studies of Cardiac Hypertrophy and Cardiac Fibrosis Using a Large Mouse Panel

Introduction

This chapter contains a copy of a manuscript currently in submission regarding the major findings of the hybrid mouse diversity panel (HMDP) heart failure (HF) project with regards to organ weights and cardiac fibrosis. It describes the observation of significant diversity within the HMDP both before and after treatment with isoproterenol with regards to organ weights and fibrosis, as well as differences in the degree of effect that isoproterenol had on these phenotypes. It further describes the identification of 30 genome-wide significant loci using GWAS, as well as describing likely candidate genes within these loci. Several of our loci overlap with genes that have previously been well-characterized as being involved in HF, including calcineurin A and B and phospholamban. The manuscript also demonstrates using both knockout and transgenic lines of mice that *Abcc6* has a significant effect on cardiac fibrosis in the presence of ISO. Additionally, it describes preliminary analysis of a novel noncoding RNA, *Miat*, which may regulate cardiac hypertrophy.

A genome-wide association study of isoproterenol-induced cardiac hypertrophy and fibrosis in mice

Authors: Christoph D. Rau[1*], Jessica Wang[2*], Rozeta Avetisyan[1], Zhihua Wang[3], Milagros Romay[1], Lisa Martin[2], Shuxun Ren[3], Yibin Wang[3], Aldons J. Lusis[1,2X]

Affiliations:

[1]Department of Microbiology, Immunology and Molecular Genetics, and Department of Human Genetics, University of California, Los Angeles, CA 90095, USA

[2]Department of Medicine/Division of Cardiology, David Geffen School of Medicine, University of California, Los Angeles, CA 90095, USA

[3]Departments of Anesthesiology, Physiology and Medicine, Cardiovascular Research Laboratories, David Geffen School of Medicine, University of California, Los Angeles, CA 90095

* These authors contributed equally to this work

X Corresponding author

Summary

The complexity underlying common forms of heart failure (HF) has hindered genetic approaches such as genome-wide association studies (GWAS) in human populations.

We addressed this by performing a GWAS in a novel mouse resource population, the Hybrid Mouse Diversity Panel, following exposure to the β-adrenergic agonist isoproterenol. Our analyses revealed 32 significant loci, each containing between 1 and 30 genes (average 14) affecting cardiac hypertrophy, fibrosis and other surrogate traits relevant to HF. We also performed global transcriptome analysis of heart tissues from the mouse strains to prioritize candidate genes. Several loci contained very strong candidates which had previously been shown to contribute to Mendelian forms of HF in humans or implicated in HF in transgenic mouse models. We also validated two novel genes, one for fibrosis and one for cardiac hypertrophy, using gene targeting, transgenic and *in vitro* approaches. The former, *Abcc6*, is an orphan transporter previously known to contribute to calcification but not fibrosis. The second, *Miat*, is a long noncoding RNA that is induced by ISO treatment. Our data provide a rich resource for identifying genes and interactions contributing to polygenic forms of HF.

Introduction

Heart failure (HF) is a common cause of death in developed countries [1]. In contrast to many other common disorders, genome-wide association studies (GWAS) of HF have had modest success in elucidating the genetics underlying this complex disease. Only two heart-failure related loci [2] have reached accepted levels of genome-wide significance, despite studies involving tens of thousands of patients and large meta-analyses [3–5]. The challenges of performing GWA in HF is likely due to the very complex nature of the disease, which can arise as a result of multiple underlying etiologies, such as myocardial infarction, hypertension or diabetes, each of which are, in and of themselves, complex traits with significant environmental confounders [1].

Fibrous tissue accumulation and scarring in the myocardium is a fundamental component of many cardiac disorders, including remodeling in both myocardial infarction[6] and heart failure [7,8]. This important phenotype has proven difficult to study in humans at the level of a GWAS or other population-based study. Although recent reports have suggested that cardiac fibrosis may be roughly monitored through the use of plasma biomarkers[9] or MRI imaging[10], these methods are not considered as accurate as traditional analysis, which requires physical examination and sectioning of heart tissue. The invasive and lengthy procedures necessary to quantify cardiac fibrosis in humans has limited studies.

Animal models such as the mouse, in which the disease onset and environmental effects can be controlled, have provided a useful means for the genetic dissection of heart failure and cardiac fibrosis[11–14]. However, although a number of loci for heart failure

traits have been mapped in mice and rats, only in a few instances have the underlying genes been identified. The primary problem has been the poor mapping resolution in traditional linkage analyses. We have recently reported the development of a resource, the Hybrid Mouse Diversity Panel (HMDP), for genome-wide association mapping in mice, providing at least an order of magnitude of improved resolution when compared to traditional linkage analysis [15].

We now report a genetic analysis of isoproterenol-induced heart failure and cardiac fibrosis in the HMDP. Although rarely an initial impetus for HF, β-adrenergic stimulation characterizes the HF state especially in later stages and is considered a driving force behind ongoing hypertrophy and progressive cardiac failure [16]. Isoproterenol (ISO), a synthetic non-selective β-adrenergic agonist, provides a means to activate the β-adrenergic receptors without the need to cause an initial insult to the heart [17–20]. Additionally, it has been reported that the pressure overload response of the heart results in extensive changes to cardiac fibrosis[21]. We now report the identification of about thirty genome-wide significant loci for hypertrophy and fibrosis. To complement the GWAS we have carried out transcriptional profiling of ventricles from the entire panel, before and after treatment with ISO. This has enabled improved prioritization of candidates at GWAS loci as well as pathway analyses. Several of the loci contained well-studied HF-associated genes such as *Ppp3ca* and *Sgcd*, as well as others involved in other forms of cardiac dysfunction. We have also identified a number of novel genetic loci significantly associated with cardiac hypertrophy and fibrosis, several containing only a few genes. We selected two promising novel candidates for validation studies. One of

these is *Miat*, a non-coding RNA previously implicated in myocardial infarction risk[22]. *Miat* was upregulated in diseased hearts and had significant effects on the progression of cardiomyocyte hypertrophy in response to catecholamine-induced stress in cultured myocytes. The second is *Abcc6*, an orphan transporter previously shown to contribute to tissue calcification in the disorder pseudoxanthoma elasticum[23]. We now demonstrate using gene targeted and transgenic mice that variations in *Abcc6* expression contribute to ISO-induced fibrosis. Our data constitute a freely available resource linking genomic variations, gene expression, and HF traits in an ISO-induced HF mouse model (http://systems.genetics.ucla.edu/data).

Results

Phenotypic characterization of HMDP mice following ISO treatment

Our preliminary studies have revealed that inbred strains of mice exhibit dramatic and highly reproducible differences in cardiac hypertrophy in response to ISO treatment. In order to identify the genetic contributions to common forms of HF, 748 mice from 105 different strains of the HMDP were divided into control (average 2.2 per strain) and treated (average 4.1 per strain) cohorts, details of which are available in our database. Treated mice were implanted with an Alzet micropump and given 20 mg/kg/day of ISO for three weeks, at which point all mice were sacrificed. Cardiac hypertrophy was quantified by left- and right- ventricular weights, while left and right-atria weights were measured as an indication of cardiac function. Liver and lung weights, as a reflection of fluid retention, were also used as indicators of congestive heart failure.

Cardiac fibrosis was measured by quantifying fibrotic tissue as a percentage of all tissue in left ventricular sections stained with Masson Trichrome. Quantification was performed using the NIS Elements AR program. To confirm our results, we examined a subset of the strains using Sirius Red, another fibrosis-marking stain, and observed very high concordance between our samples (R=0.75, P=0.00014, data not shown). As expected[8,24], we observe strong correlations between cardiac fibrosis and total heart weight (P=1.1E-07). Our results also compare favorably to prior quantifications of fibrosis in a limited number of strains[25].

We observed that of the 470 mice assigned to the treatment cohort, 139 (29.6%) died prior to the end of the protocol, most (127) within the first 48 hours of treatment. By contrast, all of the control cohort survived to the end of the protocol. Survival to the end of the protocol clearly exhibited a genetic component with many strains (42) showing no premature death susceptibility to ISO challenge, while a small number of strains (9) displayed 100% lethality (Fig 3.1a).

Association analysis was performed using roughly 132,000 SNPs across the genome with the EMMA algorithm [26] to correct for population structure. In addition to the absolute weight measurements, analyses were performed on the ratios of each treated weight to its corresponding control weight as a measure of responsiveness to ISO treatment. Previous work performed using the HMDP and EMMA has reported that a genome-wide significance threshold for the HMDP is 4.1E-06 [15]. Utilizing this cutoff and a minimum minor allele frequency (MAF) of 7.5%, 32 significant loci were discovered from our analysis (Table 3.1). While linkage analysis in mice typically

exhibits a resolution of tens of Mb, the loci identified in this study averaged 2 Mb in size, based on linkage disequilibrium with the majority being less than 1 Mb.

To help identify candidate genes at the loci, global transcriptome analysis was performed on left ventricular tissue of mice from both control and ISO-treated groups. Local (presumably *cis*-acting) and *trans*-acting loci controlling transcript levels in control and ISO-treated mouse hearts were mapped using EMMA, similar to the phenotypic traits. Additionally, the Wellcome Trust Mouse Genomes Project sequencing database [27], which has the full genomic sequence of 10 strains in our panel, was utilized to examine genomic variations, such as missense, nonsense or splicing variations, in each locus. Together, these two approaches provided a powerful and systematic method for the identification of causal genes within each locus.

The right and left ventricular weight (RVW, LVW) variations mirrored each other closely, with associations being somewhat stronger for RVW (Fig 3.1c). In total we observed eight loci corresponding to treated RVW (Fig 3.3a), two loci for the ratio of treated to untreated RVW (Fig 3.3b) and a single locus for the ratio of right atrial weight.

Similar to the heart weights, we observed marked variation of liver and lung weights following ISO treatment across the HMDP (Fig 3.1d, 3.1e). Lung weight in particular showed a robust increase in weight with ISO treatment, with over half of treated strains reporting a 20% or higher increase in weight. We observed five loci corresponding to ISO-treated lung weights (Fig 3.3c), and two loci corresponding to ISO-treated liver weights (Fig 3.3d), including one which replicated an RV peak.

Cardiac fibrosis also varied significantly both in baseline and treated mice, with the extent of fibrosis being much greater in treated mice (Fig 3.1f). We observed a total of seven loci for cardiac fibrosis in untreated animals (Fig 3.3e), and six in treated animals (Fig 3.3f). Of these loci, two occurred in both treated and untreated analyses.

Overall, we observed nineteen loci (Table 3.1) associated with ISO-induced cardiac hypertrophy and heart failure and eleven loci associated with cardiac fibrosis. Loci typically ranged between 500kb to 2 Mb in size and contained from one to twenty genes within the linkage disequilibrium (LD) block containing the peak SNP. Our results represent an improvement of more than an order of magnitude compared to traditional linkage analysis, which has a reported resolution of 10 to 20 Mb [12,28]. All significant and suggestive loci as well as gene expression data are available in our resource for public query and analysis.

Overlap with previously reported GWAS and linkage loci

In human studies, only two loci have been reported which exceed the accepted threshold of 1E-8 [2], while in mice broad linkage peaks have complicated efforts to identify single causal genes [20]. We explored whether the loci we identified overlap with human GWAS results by examining the top twelve previously identified significant and suggestive human loci [3–5,29]. The human loci were mapped onto the mouse genome using the NCBI Homologene resource and compared to the suggestive (P<1E-05, MAF>5%) loci identified in the weight portion of our study. We observe six of twelve human loci, including one of the genome-wide significant loci near USP3, replicating in our study (Table 3.4) at a statistical significance of P=3.5E-4.

Sixteen loci for HF-related traits have previously been identified via linkage in mice for CHF-related traits. We observed that seven of these sixteen QTLs overlap with weight loci identified in this study (P=0.0012) (Table 3.5). It should be noted that some of the linkage studies utilized different hypertrophy-inducing stressors, which would be expected to implicate different genetic loci as compared to ISO stimulation. For example, several previous QTLs were observed after sensitization using a calsequestrin transgene [12]. We observe a modest downregulation of calsequestrin after ISO treatment instead of the upregulation observed in the calsequestrin transgenic model. (Figure 3.8).

Gene Expression Analysis

To help identify candidate genes at our loci, we carried out global expression analysis of left ventricular heart tissue from both control and ISO treated mice from 92 strains. The loci controlling gene expression levels were mapped using EMMA, and are referred to as expression quantitative trait loci (eQTL). eQTLs are termed '*cis*' if the locus maps within 1Mb of the gene and otherwise are termed *'trans'*. Overall, we observed 2985 *cis* and 5912 *trans* eQTL in control animals and 3093 *cis* and 4842 *trans* eQTL in treated animals (*cis* threshold P<3.6E-3, *trans* threshold P<4.22 E-6).

Total heart weight and cardiac fibrosis with and without ISO treatment was correlated with global gene expression (top 50 genes for each phenotype shown in Table 3.2). We observed many genes which are known to play important roles in HF are highly correlated with heart weight. For instance, *Fstl1* (R=0.57, P=6.65E-17) has been reported to attenuate cardiac hypertrophy in mice [30]. We also observe known HF biomarkers such as *Timp1* (R=0.61,P=2.18E-20) and *Lgals3* (R=0.53,P=1.18E-14) correlated strongly to

cardiac hypertrophy[9]. In addition, several known biomarkers for cardiac fibrosis[9,31] such as *Lgals3* (R=0.46,P=6.2E-11) as well as several collagen species(*Col1a1*, R=0.43,P=1.45E-9; *Col3a1*, R=0.39, P=2.2E-8) were correlated with measured levels of cardiac fibrosis. DAVID [32,33] analysis of all significantly correlated genes (P<1.23E-07) revealed strong enrichment for eighteen GO categories for heart weight (Table 3.3), including: signaling molecules (P=5.23E-16), the extracellular matrix (P=4.96E-13), cell adhesion (P=2.83E-4) and calcium binding (P=0.0058). We also observed (Table 3.3) enrichment of ten GO categories for cardiac fibrosis, including genes involved in signaling(P=3.9E-7) and the proteinaceous extracellular matrix (P=4.88E-5) All gene expression, as well as eQTLs, can be found on our online database.

Identification of hypertrophy, cardiomyopathy and cardiac fibrosis genes using the HMDP HF resource.

Our resource provides researchers with access to tabular or graphical results of GWAS peaks (Fig 3.3), eQTLs (Fig 3.2b), gene expression (Fig 3.2a) and collected phenotypes (Fig 3.1). The combination of these four types of data provide a powerful tool for the identification of important and novel candidate genes in HF. For example, the data can be queried to identify genes, which show a consistent up- or down-regulation in response to ISO stimulus (Fig 3.2a), or to determine which genes at a locus show *cis* regulation (Fig 3.2b), as well as which genes of interest show strong correlations to HF-related phenotypes (Fig 3.2c). Using our database, we analyzed several of the GWA peaks contribution to HF traits and demonstrated strong evidence that the candidates

within them were genes previously demonstrated to contribute to HF in humans or transgenic mouse models. Some of these results are discussed below.

We observed a genome-wide significant SNP (p = 1.9E-6) on chromosome 3 for the trait treated-to-untreated right RVW ratio that lies between the second and third exons of *Ppp3ca*, encoding the alpha isozyme of calcineurin A. Calcineurin A is a known target of β-adrenergic signaling, with a well-described role in ISO-induced hypertrophy[20,34–36]. Calcineurin A is the only gene in LD with the peak SNP and has suggestive *cis*-eQTLs in both control (P=0.029) and treated (P=0.046) animals as well as a significant (P=1.3E-3) *cis*-eQTL for the ratio of treated to control calcineurin A expression (Figure 3.4a,b). We also observed a modest correlation between the ratio of *Ppp3ca* expression and the ratio of heart weights in control and ISO-treated animals (P=0.01). We further observed *Ppp3cb*, the beta isozyme of calcineurin A, in a locus on chromosome 14 (p=3.3E-6) for the trait treated lung weight. *Ppp3cb* has a highly suggestive *cis*-eQTL (4.7E-3) as well as an insertion in a splice site in several strains of the HMDP.

On chromosome 10, we observed a treated RVW locus (P-value=2.8E-07) containing the gene for phospholamban (*Pln*) in an LD block with 17 genes. *Pln* is another well-studied gene involved in heart failure[34–38]. *Pln* expression is regulated in *cis* in ISO-treated (p = 1.45E-05) mice and is modestly correlated with RVW (R = -0.145, P = 0.03). The same locus also had a significant impact on a post-treatment liver weight locus (P=3.48E-06) in the same LD block, and suggested *Pln* as the candidate gene for that locus as well.

Another RVW locus was observed on chromosome 11, with the peak SNP (P = 2.15E-06) lying between the 2nd and 3rd exons of sarcoglycan delta (*Sgcd)*. *Sgcd* has previously been implicated in familial dilated cardiomyopathy in humans. The role of *Sgcd* in cardiomyopathy and muscular dystrophy is well-studied in a knockout mouse model[39–42].

Several other loci contained candidates highly suggestive of functional roles in cardiac hypertrophy. Our most significant locus for RVW (P=3.49E-10) lies on chromosome 5 and contains 11 genes within LD. Two attractive candidates lie within the region based on results from our study. Knockout of one of these candidates, *Mospd3,* resulted in significant right-ventricle-specific dilation and thinning in neonatal mice, leading to 50% neonatal lethality, although heterozygotes showed no phenotypic changes [43]. *Mospd3* has a non-synonymous SNP among the HMDP strains and was correlated to overall heart weight (R=0.29, P=6.2E-5). Another candidate at this locus, *Pcolce,* was upregulated in the failing heart and plays an important role in the regulation of cardiac fibrosis in mouse models of HF [44]. It is also correlated with RVW (R=0.35, P=5.85E-7).

We observed several interesting candidates for cardiac fibrosis. For instance, on chromosome 7 we identified a locus (P=1.4E-6) containing *Snrpn* and *Ube3a* which overlaps with a previously identified suggestive heart failure locus in humans[5]. On Chromosome 17, a shared locus in both treated (P=2.5E-6) and untreated (P=1.4E-6) hearts contains 4 genes within LD(Table 3.1). *Fert2,* a gene within this locus, is a known regulator of fibroblast migration[45] and plays a vital role in cadherin/beta-catenin association[46], a pathway which has been shown to be important for fibrogenesis[47,48]. We

observed a strong *cis*-eQTL for this gene (P=1E-5), which has a non-synonymous mutation within several strains and note that it falls within a heart weight QTL in rats[49]. We also identified a collagen species *Col22a1* as a candidate on a locus on chromosome 16(P=9.6E-7) and *Tll1*, a metalloprotease which interacts with collagen[50] as a candidate within a locus on chromosome 8 (P=1.43E-6). Finally, a locus of interest for RVW on chromosome 5 (p = 1.23E-06) contains *Prkag2*, which is under strong *cis* regulation in both untreated (p =7.44E-26) and treated (p=2.8E-19) mice and contains two non-synonymous SNPs among the strains of the HMDP (Figure 3.5a,b). Human families with mutations in *Prkag2* have a dominant form of glycogen storage disease, primarily involving the heart, giving rise to heart wall thickening and eventual heart failure [51,52]. Mice containing a transgene of *Prkag2* with a T400N point mutation also display glycogen storage-mediated cardiomyopathy cardiac wall thickening and dysfunction [53]. Mutations in *Prkag2* have also been linked to premature activation of the NF-κB and Akt signaling pathway, a known mediator of cardiac hypertrophy [53].

Our resource facilitates the identification of novel genes involved in cardiac hypertrophy through the integration of GWA loci, eQTL peaks, gene expression and phenotype correlations. For example, we have identified the DNA repair gene *Mgmt* as a strong candidate for the change in right atrial size (P=1.41E-6) on chromosome 7 in a locus containing 15 genes. *Mgmt* removes the O^6-alkyl-guanine DNA mutation that is the major carcinogenic lesion introduced by alkylating mutagens [54]. It has previously been studied in the contexts of oncogenesis [55–57] and chemotherapy efficacy [54], but has not been implicated in heart disease. We observe a strong *cis*-eQTL (P=4.8E-5) linking

the expression of the gene to the same SNPs which regulate the phenotype. Additionally, *Mgmt* shows a strong correlation with many phenotypes, most notably total heart weight (R=-0.46, P=3.4E-11).

In vitro validation of *Miat* as a novel regulator of cardiac hypertrophy

We identified *Miat*, a multi-exonic long non-coding RNA previously associated with an increased risk of myocardial infarction[22], as a novel HF candidate in a chromosome 5 locus (P=1.3E-06), containing 22 genes, for treated lung weight. In addition, we observed that the *Miat* locus overlapped with a suggestive locus for both RVW (P=2E-4) and LVW(P=7.8E-3). We tested whether *Miat* expression directly modulates cardiomyocyte function, using a neonatal rat ventricular cardiomyocyte (NRVM) model. These cardiomyocytes respond *in vitro* to β-adrenergic agonist phenylephrine (PE) by undergoing hypertrophy (Figure 3.9b), as characterized by increased cell size and expression of molecular markers including natriuretic peptide A and B (*Anp* and *Bnp*) (Fig. 3.9c). The induction of *Miat* expression by PE treatment was significantly abrogated by siRNA knockdown (Fig. 3.9b). *Miat* siRNA knockdown reduced the PE-induced cardiomyocyte hypertrophy from 29.0% to 17.0% (Fig 3.9b), and this was accompanied by decreased expressions of the *Anp* and *Bnp* as well as *Myh7*, a marker of hypertrophic cardiomyopathy (Fig. 3.9c).

To understand how *Miat* contributes to hypertrophy, we examined several key regulators of cardiomyocyte differentiation and proliferation. The data showed that *Miat* siRNA selectively downregulated mRNA levels of NK2 homeobox 5 (*Nkx2.5*) and

GATA binding protein (*Gata4*) and *NFkB* in NRVMs compared to control siRNA after PE treatment (Fig. 3.9d), suggesting that *Miat* modulates cardiac hypertrophy via targeted regulation of key transcription factors in cardiomyocytes.

In vivo validation of *Abcc6* as a novel regulator of cardiac fibrosis

One of the ISO dependent fibrosis loci, on chromosome 7 (Table 3.1, P=7.1E-7) at 53 Mb, contains 28 genes within LD (Figure 3.6A). One of these genes, *Abcc6,* has a splice site variation resulting in a premature stop codon that is found in a number of strains in the HMDP[58,59]. We selected *Abcc6* as a candidate based on the fact that mutations of the gene in the disorder *Pseudoxanthoma elasticum* (PXE) are known to contribute to systemic vascular calcification, including the skin, eyes and heart. *Abcc6* is thought to mediate calcification by influencing the secretion into the circulation of a substance that remains to be identified[60]. Since *Abcc6* is expressed at highest levels in the liver and kidney it has been assumed to function in these tissues. However, we observed nominally significant *cis*-eQTLs for *Abcc6* in the heart (P=0.01 in treated animals, P=0.03 in control) suggesting that Abcc6 is also expressed in heart and may function there as well. Recent evidence has shown that *Abcc6* is localized in the mitochondrial associated membrane, and that deficiency of *Abcc6* influences mitochondrial morphology[61]. Moreover, deficiency of *Abcc6* leads to increased infarct size in a cardiac ischemia-reperfusion model[23].

To examine the role of *Abcc6* in cardiac fibrosis we studied C57BL/6J mice carrying a targeted mutation of *Abcc6* (KO mice) as well as control mice (strain C57BL/6J do not have the splicing mutation discussed above)[23]. We observed that

neither the wild type nor the KO mice exhibited significant calcification in the absence of ISO treatment as judged by Alizarin Red staining. This is consistent with previous studies indicating that the calcification is age dependent and occurs only in older mice[58]. After ISO treatment there was a very modest increase in staining (P=0.06) (Fig. 3.6B). Neither wild type nor KO mice exhibited significant fibrosis at baseline as judged by Masson Trichrome staining (Fig. 3.6D, E) or Alizarin Red staining (data not shown). Notably, C57BL/6J mice are among the most resistant strains to fibrosis in the HMDP (Fig. 3.1). However, after ISO treatment the KO mice but not the wildtype mice exhibited significantly increased amounts of fibrosis, but did not display significant increases in calcification (Fig 3.6D). Thus, these results are consistent with the fact that the fibrosis locus is ISO-dependent.

To complement these studies we examined an *Abcc6* bacterial artificial chromosome (BAC) transgenic mouse on a fibrosis-susceptible C3H/HeJ background. Strain C3H/HeJ mice carry the splicing variation that results in a deficiency of *Abcc6*[58] In the absence of ISO neither C3H/HeJ mice nor C3H/HeJ mice carrying the transgene exhibited significant calcification or fibrosis, whereas after ISO treatment the C3H/Hej mice but not the transgenic mice exhibited substantial fibrosis and calcification (Fig 3.6E). This suggests that while *Abcc6* is sufficient to cause significant changes to cardiac fibrosis after ISO stimulation, a potential modifier locus affects calcification.

To examine pathways by which *Abcc6* may contribute to fibrosis, we identified 569 genes that were significantly correlated with *Abcc6* transcript levels in hearts of

HMDP mice in treated animals (Table 3.6). Using DAVID[32,33], we observed that these genes were highly enriched for mitochondrial functions (P=4.65 E-15), transit peptides(P=2.7E-7) and acetylation(P=4.5E-4) (Table 3.7). These findings are consistent with the localization of *Abcc6* in the mitochondrial associated membrane.

Discussion

Heart failure is the leading cause of hospitalization in developed countries and new therapeutic and diagnostic approaches are urgently required. However, efforts to identify genes contributing to common forms of the disease have been only modestly successful, presumably because of the extremely complex etiology of the disease[3–5]. There has been considerable success in identifying genes contributing to Mendelian forms of cardiac hypertrophy, but it is not yet clear if common variants in these genes contribute to non-Mendelian forms of cardiac dysfunction in the general population. Linkage studies in mice have mapped a number of loci contributing to heart failure in various models, but very few genes have been positionally cloned due to the poor mapping resolution of QTL analysis in mice. We have used a novel strategy involving GWA studies in mice to perform fine mapping of loci contributing to heart failure traits in a ISO-treated mouse population. We have combined this strategy with global gene expression analysis in the heart to help identify candidate genes and pathways. A number of the loci identified in our study contain genes known to be involved in cardiomyopathy based on previous biochemical or genetic studies, validating our findings. We have also partially characterized two novel candidates, one at a hypertrophic locus and the other at a fibrosis locus. Our results indicate that the *Abcc6* transporter can markedly contribute

to fibrosis in an ISO-dependent manner. They also provide strong suggestive evidence for a role of *Miat*, a non coding RNA, in hypertrophy. These points are discussed below.

Previous work on HF using rodent models have successfully identified several candidate genes for the disease. Rats have provided key insights into the genetics of HF, identifying candidates such as *Endog* [14] and others [62,63]. Research done on mouse models has typically focused on the use of sensitizers such as calsequestrin transgenics [11–13] to identify additional modifier mutations which either exacerbate or protect against the original sensitizing mutation. Using these approaches, researchers have been able to identify a number of strong candidates such as *Tnni3k* [13] and others [12].

By using association rather than linkage, we were able to significantly improve the resolution of our mapping data over prior rodent studies, in some cases approaching the level of human GWAS with 1-5 genes in a given locus. However, unlike the tens of thousands of individuals studied in human GWAS, we were able to identify many more significant loci while using approximately 750 animals. We attribute the success of our study to an increase in power due to a carefully controlled and shared environment in our vivarium, which is not possible in human studies. Our results have suggested a number of new potential loci for heart failure, as well as implicated several genes previously associated with Mendelian forms of cardiomyopathy as being involved in polygenic forms of HF as well. Additionally, our loci provide additional evidence in support of some suggestive GWAS hits in human studies, as we observed a 50% overlap between the published suggestive human GWAS hits and our peaks (Table 3.2).

HF in humans has a heritability of 28% [64] with clear evidence for a strong environmental influence on the disease. Our experimental design included a specific HF inducer (ISO) administered at a specific dose and for a specific time period (3 weeks). Our ability to carefully control the type, severity and length of treatment allows us to explore specific gene-by-environment interactions, which may be expanded to other disease models (e.g. angiotensin II, trans-aortic-constriction, myocardial infarction and others) or treatments (e.g. beta-blockers) to explore the ways in which genetics interacts with different disease models and treatments. Our results, obtained through careful analysis of each locus for genes which correlated with heart phenotypes, had *cis*-eQTL regulation and/or nonsynonymous mutations(Table 3.1), suggest that a number of genetic factors lead to catecholamine-mediated HF susceptibility. Genes involved in the beta-adrenergic stimulation pathway (*Ppp3ca, Ppp3cb* and *Pln),* in heart development (*Mospd3),* in cellular integrity (*Sgcd)* and energy and calcium homeostasis (*Prkag2* and *Calm3*) all appear to play roles in the development of HF in our mouse model. Similarly, a number of different pathways interact to lead to cardiac fibrosis. We identified several genes (*Fert2*[45], *Tll1*[50], *Tjp1*[65]) which have previously been shown to interact with known fibrogenesis pathways or collagen itself and several others (*Cntln*[66], *Fert2*[67],*Srpx*[68]) which play roles in the regulation of known drivers of heart stress or damage.

We implicated a novel long non-coding RNA, *Miat*, in ISO-dependent hypertrophy. Previous work has shown that *Miat* RNA is localized to the nucleus[69,70] and that it interacts with splicing factor 1 through tandem UACUAAC repeats[71] influencing global cell splicing[70]. *Miat* has previously been described as a susceptibility

gene for myocardial infarction[22] and there is also suggestive evidence for its involvement in schizophrenia[70] and p53-negative cancer[69]. Our *in vitro* studies with cultured cardiomyocytes show that *Miat* is upregulated following treatment with catecholamines, and that knockdown of *Miat* affects the hypertrophic response. Further, we provided evidence that *Miat* expression influences a number of transcription factors known to influence cardiac hypertrophy. Our results suggest that *Miat* acts as a splicing regulator in the heart as well, and that its upregulation after catecholamine treatment effects a change in mRNA splicing leading to an upregulation of the hypertrophic response.

Our results have revealed a novel role for *Abcc6* in a stress-specific cardiac fibrosis. *Abcc6* was originally identified through positional cloning as the cause of the Mendelian disorder PXE, characterized by calcified skin lesions, angioid streaks in the eye and leg pain, and also angina and myocardial infarction in older individuals[72]. Studies of *Abcc6* KO mice and of a naturally occurring splice variant of *Abcc6* in certain strains of mice have revealed a phenotype similar to human PXE, namely, the occurrence of tissue calcification in older mice[23,58]. Cardiac fibrosis has not been observed in *Abcc6* KO mice, even at 24 weeks of age[73]. Our studies were performed on relatively young mice, which exhibited negligible evidence of calcification. We observed that on a C57BL/6J genetic background, which is highly resistant to fibrosis, the *Abcc6* KO allele results in marked fibrosis in ISO treated mice. We also expressed, as a transgene, the normal *Abcc6* allele on the background of strain C3H/HeJ, which carries the naturally occurring splice mutant. In that study, the transgene markedly protected against ISO-dependent fibrosis. It is important to note that the effect of *Abcc6* on fibrosis was

dependent on ISO treatment, and thus *Abcc6* can be considered a modifier of the ISO stress.

Precisely how *Abcc6* acts to protect against fibrosis is unknown. The substance transported by the protein has not been identified despite intensive efforts by several groups. Based on parabiosis and other studies[60] it appears that a deficiency of *Abcc6* has a systemic effect due to the absence of a substance normally present in the circulation. Because *Abcc6* is expressed at highest levels in liver and kidney, it has been speculated that its primary function is in these tissues, but our results clearly show genetically regulated expression in the heart as well. We recently showed that a deficiency of *Abcc6* results in the activation of bone morphogenic protein signaling in heart, and that in the absence of *Abcc6* infarct size is increased in an ischemia-reperfusion model[23]. A potential mechanism by which *Abcc6* deficiency could lead to increased fibrosis is increased necrosis or apoptosis. It is noteworthy that in the ischemia-reperfusion studies, Abcc6 led to an increase in the number of apoptotic cells[23]. Our results also suggest that *Abcc6* deficiency affects mitochondrial function, as expression of the gene in the heart was strikingly associated with changes in the expression of mitochondrial genes. This is consistent with a recent report indicating that the protein is located in the mitochondrial associated membrane[61]. Proper mitochondrial function is very important for the proper functioning of the heart[74], and disruptions to the normal functioning of mitochondria are associated with cardiomyocyte death and heart failure[14,17,75]. The human heart requires a daily synthesis of approximately 30kg of ATP, which is mostly produced through the mitochondria[74], and chronic (but not acute[76]) catecholamine stimulation leads to

mitochondrial damage, energy deficits and HF[17,77]. Additionally, the mitochondria are important regulators of cell death and apoptosis[74,75]. The importance of the mitochondria to HF and the localization of *Abcc6* to the mitochondria associated membrane and its correlation to mitochondrial gene expression further suggest a role for *Abcc6* in HF.

An important benefit of our approach using the HMDP is that the data gathered in this model system are cumulative (since the strains of mice are inbred and permanent), allowing the integration of our data collected in this study with prior and future studies. The power and usefulness of our resource will only increase as additional experiments are performed. Future experiments on the same panel, examining the effects of an alpha-adrenergic receptor agonist such as angiotensin or a physical intervention such as trans-aortic constriction, will help elucidate the pathways and genes that are common between models and those which are unique to each model system of HF. In this study, we focused exclusively on the effects of chronic catecholamine stimulation on the weights of the total heart, individual heart chambers, organ weights that are markers of heart failure severity and the global transcription profile in control and treated left ventricles. It will be of interest to examine other metrics of HF such as echocardiographic traits, metabolite levels and to generate transcriptomic analyses of other potentially relevant tissues such as the adrenal gland and hypothalamus in the future. In particular, the application of systems genetics approaches, such as network modeling, to common variations contributing to HF in mice has the potential to complement traditional molecular biology approaches.

Experimental Procedures

Online database

All results and data can be accessed at http://systems.genetics.ucla.edu/data

Mice and isoproterenol treatment

The following mouse strains were obtained from The Jackson Laboratory and then bred in our colony: 30 common inbred strains (129X1/SvJ, A/J, AKR/J, BALB/cJ, BALB/cByJ, BTBR T+ tf/J, BuB/BnJ, C3H/HeJ, C57BL/6J, C57BLKS/J, C57L/J, C58/J, CBA/J, CE/J, DBA/2J, FVB/NJ, KK/HlJ, LG/J, LP/J, MA/MyJ, NOD/ShiLtJ, NON/ShiLtJ, NZB/BINJ, NZW/LacJ, PL/J, RIIIS/J, SEA/GnJ, SJL/J, SM/J, SWR/J) and 76 RI lines [RI (number of strains) - BXD (44), AXB(9), BXA(10), BXH(5), CxB(7)]. All animal experiments were conducted following guidelines established and approved by the University of California, Los Angeles Institutional Animal Care and Use Committee. All mice have been previously genotyped at over 130,000 locations.

Isoproterenol (20 mg per kg body weight per day) was administered for 21 d in 8- to 10-week-old female mice using ALZET osmotic minipumps, which were surgically implanted intra-peritoneally.

Abcc6 KO and transgenic mice[23,61] underwent the same protocol as described above, although both male and female mice were used in the analysis. No significant difference between genders was observed as a result of ISO treatment in these KO and transgenic animals.

Heart Weights

At day 21, mice were sacrificed after body weight was recorded. The heart was removed and weighed, then separated into its four component chambers, each of which was individually weighed as well. Each chamber of the heart was immediately frozen in liquid nitrogen for any future analysis. Lung and liver were removed and weighed. Additionally, the adrenal glands were removed, weighed and frozen in liquid nitrogen. All frozen tissues were immediately transferred to a –80 freezer.

Fibrosis and Calcification

A portion of the left ventricle was placed in formalin for at least 48hrs for preservation of ultrastructure. These samples were then washed with distilled water and sent to UCLA Department of Pathology and Laboratory Medicine for paraffin embedding and staining using Masson's Trichrome. Sections were analyzed using a Nikon Eclipse, TE2000-U microscope at 2-100x magnifications and images captured of the entire cross-section of the heart. Fibrosis was quantified using the Nikon Imagine System Elements AR program by comparing the amount of tissue stained blue (for collagen) to the total tissue area to obtain a quantification of percent fibrosis. Staining for Sirius and Alizarin Red used a similar protocol, instead quantifying red-stained collagen to the total area.

Efficient Mixed Model Association.

We performed association analysis using EMMA (Efficient Mixed Model Association), which uses a linear mixed model to estimate and correct for the effects of population structure between the mice of the HMDP prior to determining association p-

values. We closely followed the method described in Kang et al [26]. In brief, a kinship matrix between the strains were calculated to weight the pair-wise relationships between the strains. Values will increase when unrelated strains share a phenotypic effect at a SNP, while values will decrease if the opposite is true. Simulation of a causal locus was used to determine an adjusted genome-wide significance threshold for the EMMA algorithm on the HMDP. Due to linkage disequilibrium, the simulation shows that a cutoff of 1E-5 represents an accurate correction for multiple comparisons in the HMDP [26].

Microarray and eQTL analysis

Following homogenization of left ventricular tissue samples in QIAzol, RNA was extracted using the Qiagen miRNAeasy extraction kit, and verified as having a RIN>7 by Agilent Bioanalyzer. Two RNA samples were pooled for each strain/experimental condition, whenever possible, and arrayed on Illumina Mouse Reference 8 version 2.0 chips. Analysis was conducted using the Neqc algorithm included in the limma R package [78] and batch effects addressed through the use of COMbat [79]. eQTLs were then calculated using EMMA, as described above. Significance thresholds were calculated as in Parks et al., 2013[80]. Briefly, *cis*-eQTLs were calculated using a FDR of 5% for all SNPs that lay within 1 Mb of any probe, while *trans*-eQTLs were calculated using the overall HMDP cutoff as determined in Kang et al[26].

***Miat* Studies**

NRVM isolation and culture were performed as reported previously [81,82] Briefly, 2-

to 4-day-old rats of a mixed population were killed by decapitation and rinsed in ethanol. Hearts were excised and atria and excess fat were trimmed from the ventricles, then minced in a buffer solution and transferred into a tube. The buffer solution was evacuated and replaced with a collagenase-pancreatin solution in which the ventricles were incubated at 37°C for 30 min with moderate agitation, followed by replacement of the collagenase-pancreatin solution and incubation for an additional 20 min. Newborn calf serum was then added to the dissociated cells in the protease solution to constitute a suspension with 20% serum, which was centrifuged and resuspended in buffer solution. Myocytes and fibroblasts were separated with a Percoll density gradient, of which the former was transferred to a plating medium consisting of Dulbecco's modification of Eagle's/Ham's F12 medium (DMEM) with 10% newborn calf serum and 1% antibiotic/antimycotic. All culture media and additives were purchased from Sigma-Aldrich (MO). NRVMs were plated at a density of $5.0 – 7.5 * 10^4$ cells/cm^2, with an average plating viability of 85%, as determined by trypan blue exclusion. Cultures were maintained in a humidified incubator at 37°C and 5% carbon dioxide, with media changed daily.

Lipofectamine 2000 (Invitrogen, Life sciences) was used to transfect *Miat* siRNA (Qiagen) into NRVMs following the manufacturer's instruction.

The following primers were used for quantitative PCR reactions: Miat sense: TAACAGGTCTCACACCCCTCT, Miat antisense: CAGACCGGTCACGTGTCTACT; Anp sense: ATACAGTGCGGTGTCCAACA, Anp antisense:

AGCCCTCAGTTTGCTTTTCA. Bnp sense: CAGCTCTCAAAGGACCAAGG, Bnp antisense: GCAGCTTGAACTATGTGCCA. Gata4 sense: CTCCTACTCCAGCCCCTACC, Gata4 antisense: GTGATAGAGGCCACAGGCAT

Acknowledgments

This work was supported by NIH grant HL110667 and HL28481. CDR was supported by NIH training grant T32HL69766 and JW was supported by NIH training grant HL007895.

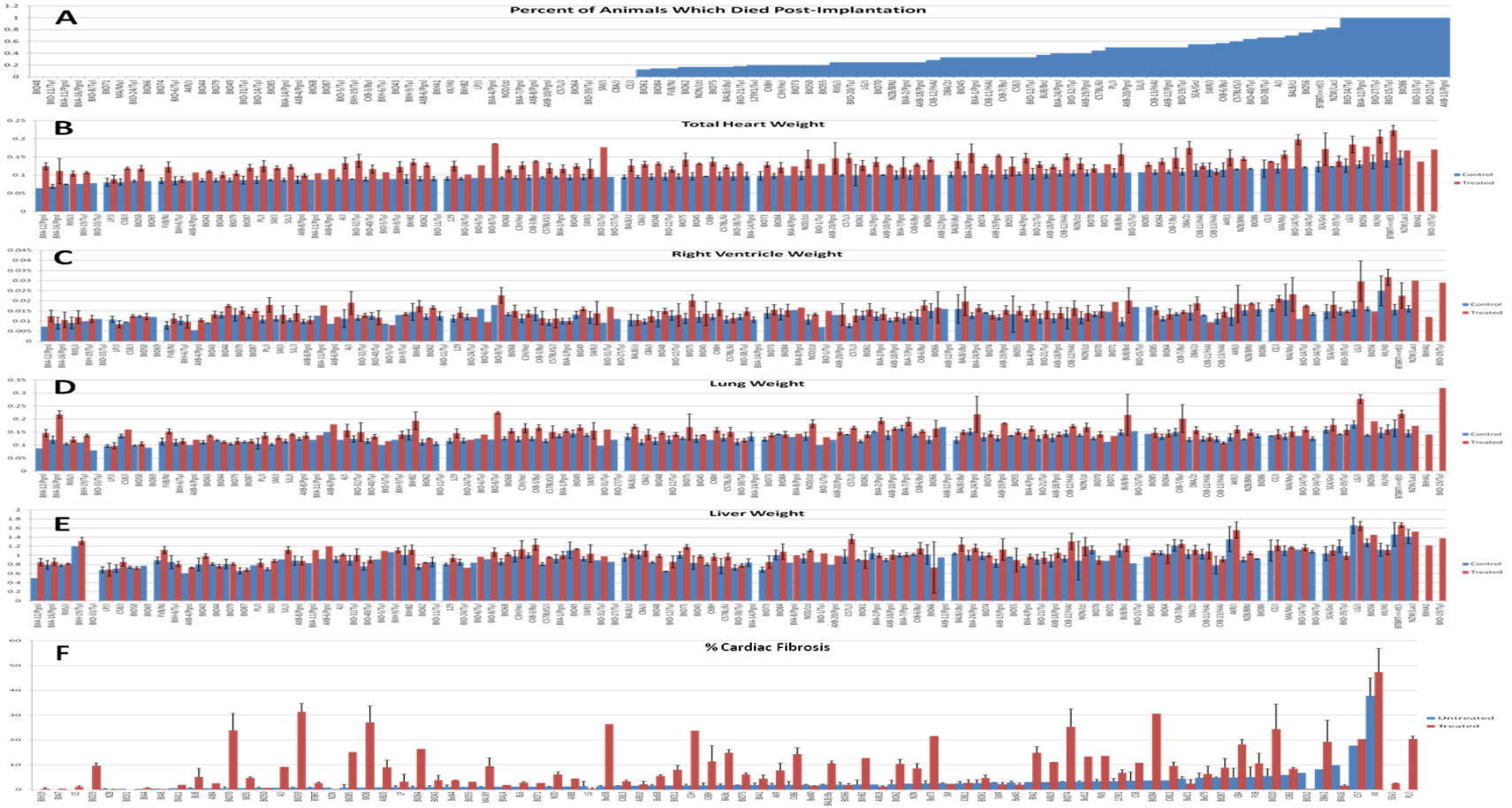

Figure 3.1. Wide variation in HF traits between the strains of the HMDP. A) Isoproterenol-induced lethality B) Total Heart Weight C) Right Ventricular Weight D) Lung Weight E) Liver Weight. F) Cardiac Fibrosis B-E are organized by the untreated heart weight of the strain and display mean +/- standard deviation.

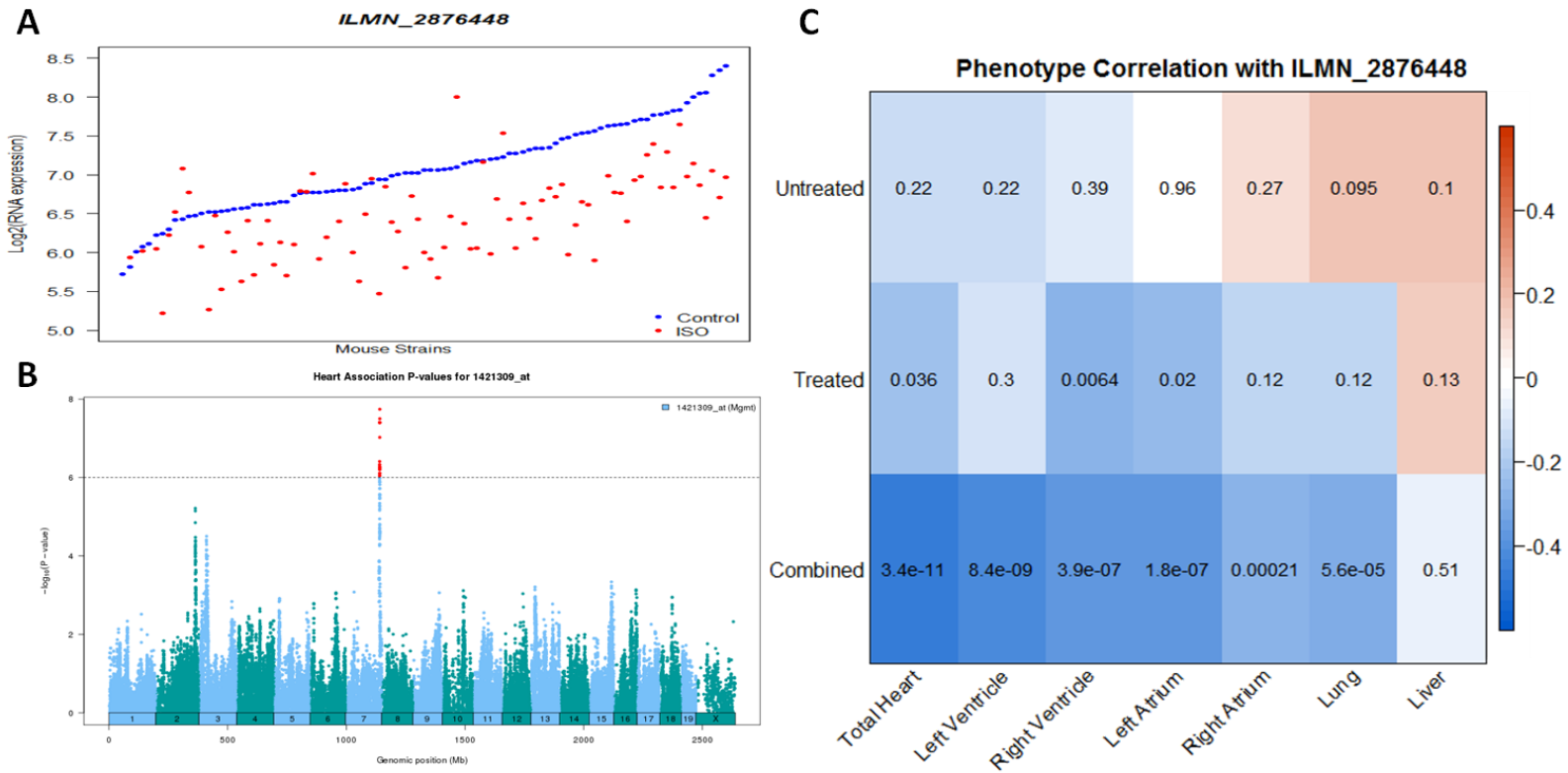

Figure 3.2. Sample plots for *Mgmt*. A variety of data on genes of interest can be obtained from our online database. A) ISO response of *Mgmt.* B) Genome wide association for the expression of *Mgmt.* C) Correlations (indicated by color) and significances (P-value indicated by number inside of box) of untreated, treated and combined gene expressions with HF-related phenotypes for *Mgmt.*

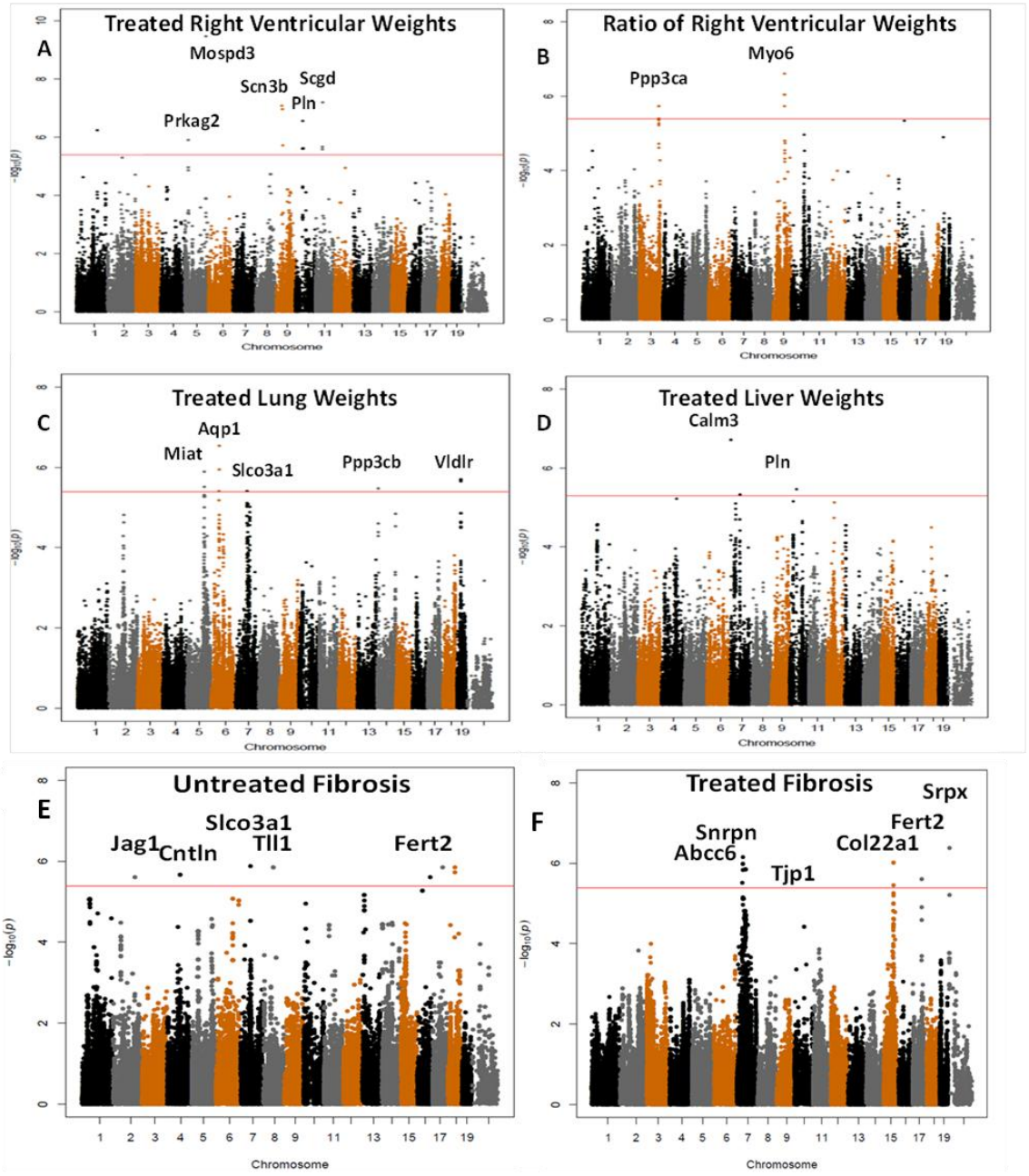

Figure 3.3. Manhattan plots of HF traits A) treated right ventricular weights B) Ratio of treated to untreated right ventricular weights C) treated lung weights D) treated liver weights. E) baseline cardiac fibrosis and F) treated cardiac fibrosis The red line indicates the threshold for significance. Likely candidate genes are indicated by gene symbols above peaks. Proposed candidate genes are indicated next to their respective peaks

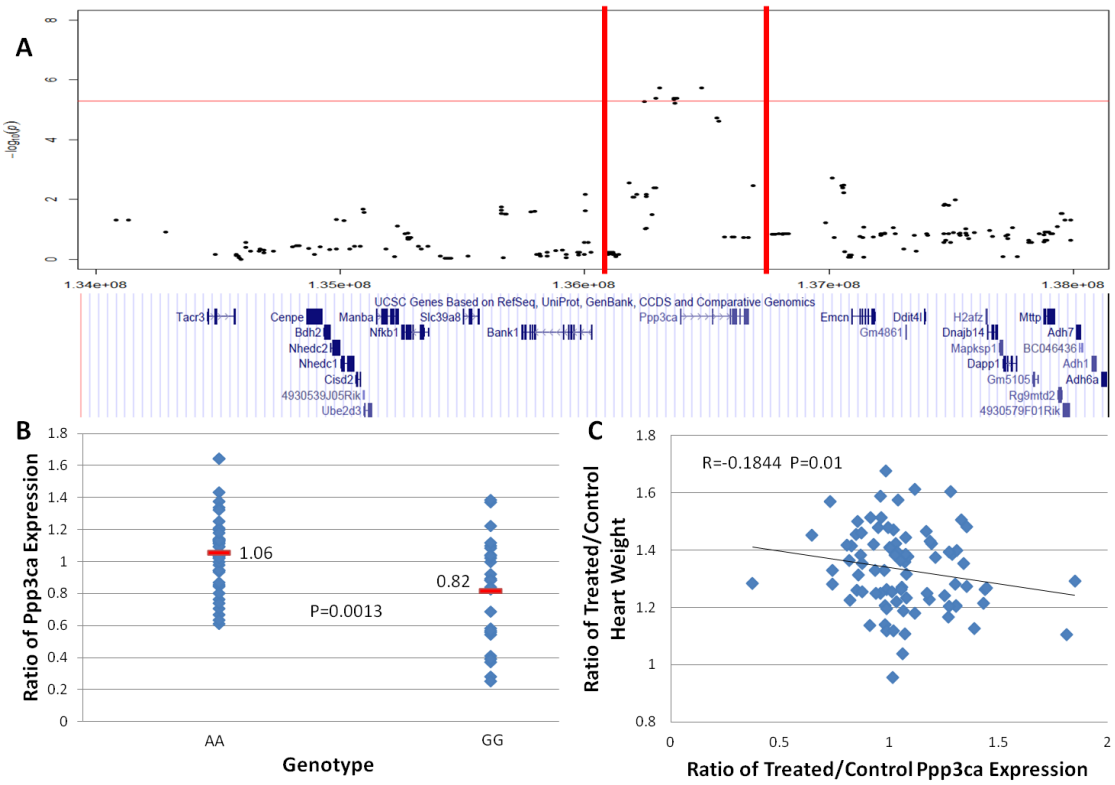

Figure 3.4. *Ppp3ca* is a candidate gene at the Chromosome 3 right ventricular weight ratio locus. A) *Ppp3ca* is the only gene within the locus. The red horizontal line represents the significance threshold, while the red vertical lines indicate the limits of the Linkage disequilibrium block. B) *Ppp3ca* has a significant *cis*-eQTL (P=0.0013) for the ratio of gene expression after and before treatment. Red line and number indicate average gene expression C) The ratio of *Ppp3ca* expression has a weak but significant (P=0.01) negative correlation with the ratio of treated to untreated heart weights.

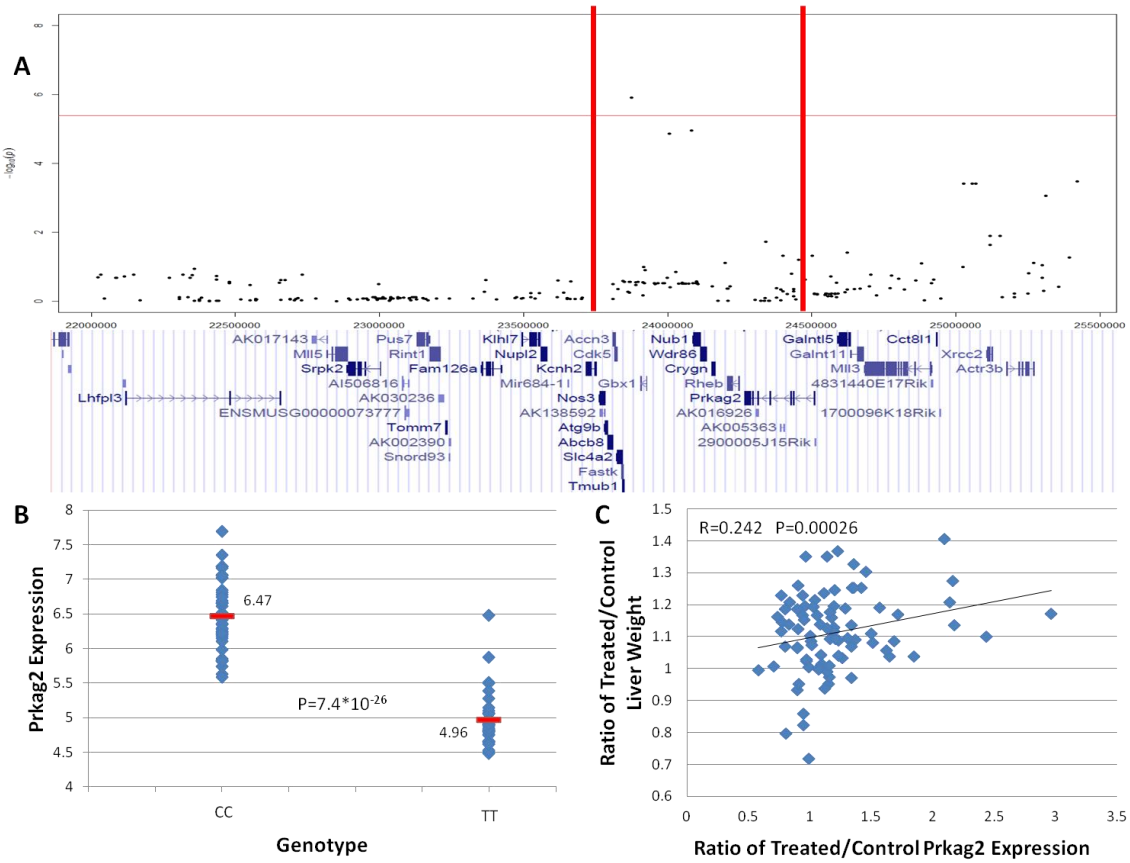

Figure 3.5. *Prkag2* is a candidate gene at the Chromosome 5 right ventricular weight locus. A) *Prkag2* is one of twenty genes within the locus. The red horizontal line represents the significance threshold, while the red vertical lines indicate the limits of the Linkage disequilibrium block. B) *Prkag2* has a very significant *cis*-eQTL (P=7.4E-26) Red line and number indicate average gene expression C) The ratio of *Prkag2* expression has a significant (P=0.00026) positive correlation with the ratio of treated to untreated liver weights, a measure of HF progression.

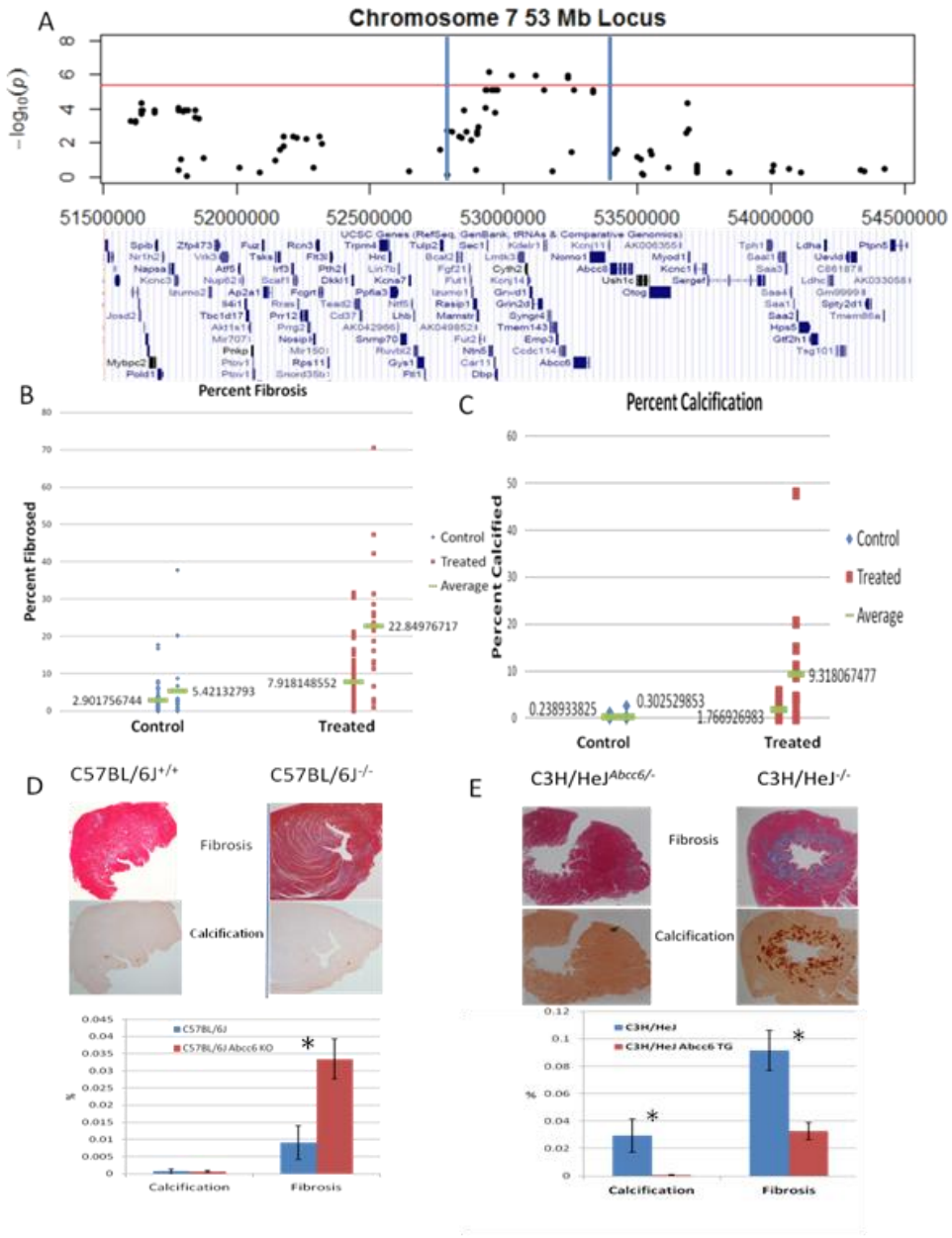

Figure 3.6. *Abcc6* plays a role in the regulation of cardiac fibrosis after ISO stimulation. A) The locus on chromosome 7 which contains *Abcc6* spans ~800kb and contains 28 genes within LD. B) Calcification in post-ISO treated hearts is increased in mice lacking *Abcc6*. C) Expression of *Abcc6* in a mouse that lacks the gene is sufficient to rescue the mouse from the Iso-induced fibrosis and calcification. D) Knockout of *Abcc6* in a mouse is sufficient to cause ISO-induced fibrosis, but does not cause a significant increase in calcification. E) The addition of *Abcc6* as a transgene is sufficient to significantly reduce both calcification and fibrosis. (*=P< 0.05)

Tables

Table 3.1. Significant HF trait loci identified in HMDP GWA. For each locus, the estimated LD block surrounding the peak is reported as well as any candidate genes in the locus. **Bold** entries represent genes which contain nonsynonymous mutations within the HMDP as reported by the Wellcome Trust Mouse Genome Project[27], while underlined entries contain significant *cis* eQTLs. See Table 3.8 for additional details.

Phenotype	Peak SNP	Chromo-some	LD Block (Mb)	Peak P-value	Number of Genes in Locus	Candidate Genes
Hypertrophic Loci						
Treated right ventricle weight	mm37-1-134467906	1	133.78-134.53	5.75E-07	14	
Treated right ventricle weight	mm37-5-137934905	5	137.93-138.15	3.49E-10	11	***Mospd3***

Treated right ventricle weight	mm37-5-23873494	5	23.82-24.47	1.23E-06	20	***Prkag2***
Treated right ventricle weight	mm37-9-40202022	9	39.77-40.52	8.41E-08	15	***Scn3b***
Treated right ventricle weight	mm37-10-49818583	10	48.19-54.24	2.80E-07	22	*Pln*
Treated right ventricle weight	mm37-11-47181489	11	46.18-49.3	2.15E-06	41	*Sgcd*
Ratio of right ventricle weights	mm37-3-136305887	3	136.04-136.79	7.83E-07	1	*Ppp3ca*
Ratio of right ventricle weights	mm37-9-80542295	9	80.00-80.99	2.94E-07	2	*Myo6*
Ratio of right atrial weights	mm37-7-142011844	7	141.50-144.81	1.41E-06	15	***Mgmt***
Fluid Retention and Heart Failure Loci						
Treated liver weight	mm37-7-15251391	7	15.13-18.75	1.93E-07	57	*Calm3*

Treated liver weight	mm37-10-49468021	10	48.19-54.24	3.48E-06	22	*Pln*
Treated Lung weight	mm37-5-111867706	5	110.87-112.87	1.28E-06	22	*Miat*
Treated Lung weight	mm37-6-53975816	6	53.88-55.57	2.90E-07	17	*Aqp1*
Treated Lung weight	mm37-7-81841621	7	79.8-82.2	3.88E-06	4	*Furin* ***Iqgap1***
Treated Lung weight	mm37-14-14941056	14	8.5-21.5	3.34E-06	50	***Ppp3cb***
Treated Lung weight	mm37-19-27061190	19	26.68-27.43	2.01E-06	1	***Vldlr***
Baseline Fibrosis Loci						
Baseline Fibrosis		2	13.7-14.0	2.51E-06	6	*Jag1*
Baseline Fibrosis		4	84-85	2.20E-06	2	***Cntln***
Baseline Fibrosis		7	72.3-74.3	1.31E-06	7	***Tjp1***

Baseline Fibrosis		8	63-66	1.43E-06	12	***Tll1***
Baseline Fibrosis		16	82.5-84.25	2.52E-06	0	
Baseline Fibrosis		17	64.2-65.8	1.43E-06	4	***Fert2***
Baseline Fibrosis		18	47.2-48.2	1.87E-06	3	
Treated Fibrosis Loci						
Treated Fibrosis		7	52.85-53.42	7.11E-07	28	***Abcc6***
Treated Fibrosis		7	60.5-69.5	1.40E-06	18	***Snrpn***
Treated Fibrosis		7	72.3-74.3	1.40E-06	7	***Tjp1***
Treated Fibrosis		15	68.4-71.4	9.60E-07	3	***Col22a1***
Treated Fibrosis		17	64.2-65.8	2.47E-06	4	***Fert2***
Treated Fibrosis		X	10-12.5	4.17E-07	12	***Srpx***

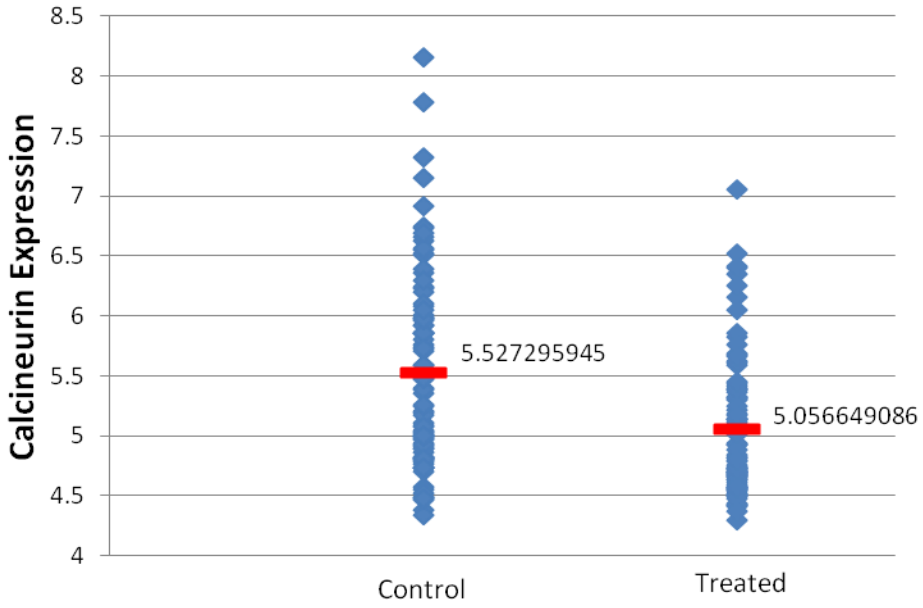

Figure 3.7. Gene expression of Calsequestrin between control and treated groups reveals no upregulation of Calcineurin after ISO treatment. Red line and numbers indicate average expression

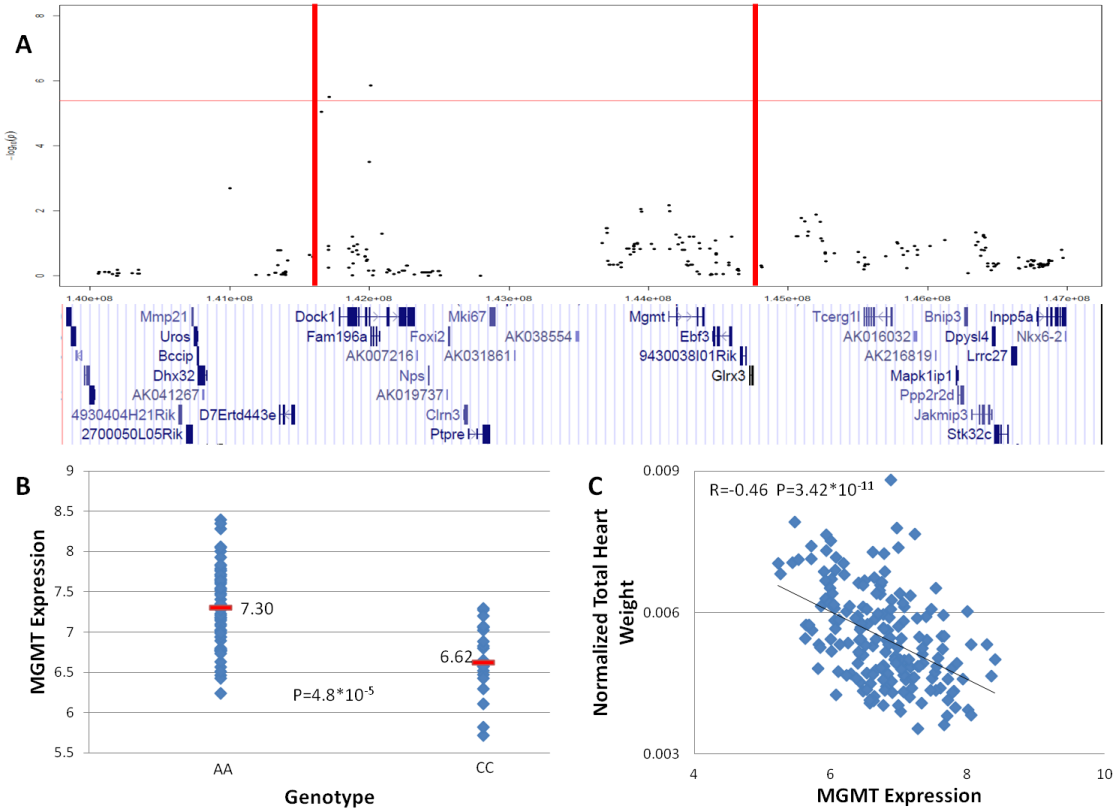

Figure 3.8. *Mgmt* is a candidate gene at the Chromosome 7 right atrial weight ratio locus. A) *Mgmt* is one of fifteen genes within the locus. The red horizontal line represents the significance threshold, while the red vertical lines indicate the limits of the Linkage disequilibrium block. B) *Mgmt* has a significant *cis*-eQTL (P=4.8E-5). Red line and number indicate average gene expression. C) *Mgmt* expression is significantly (P=3.4E-11) negatively correlated with total heart weight.

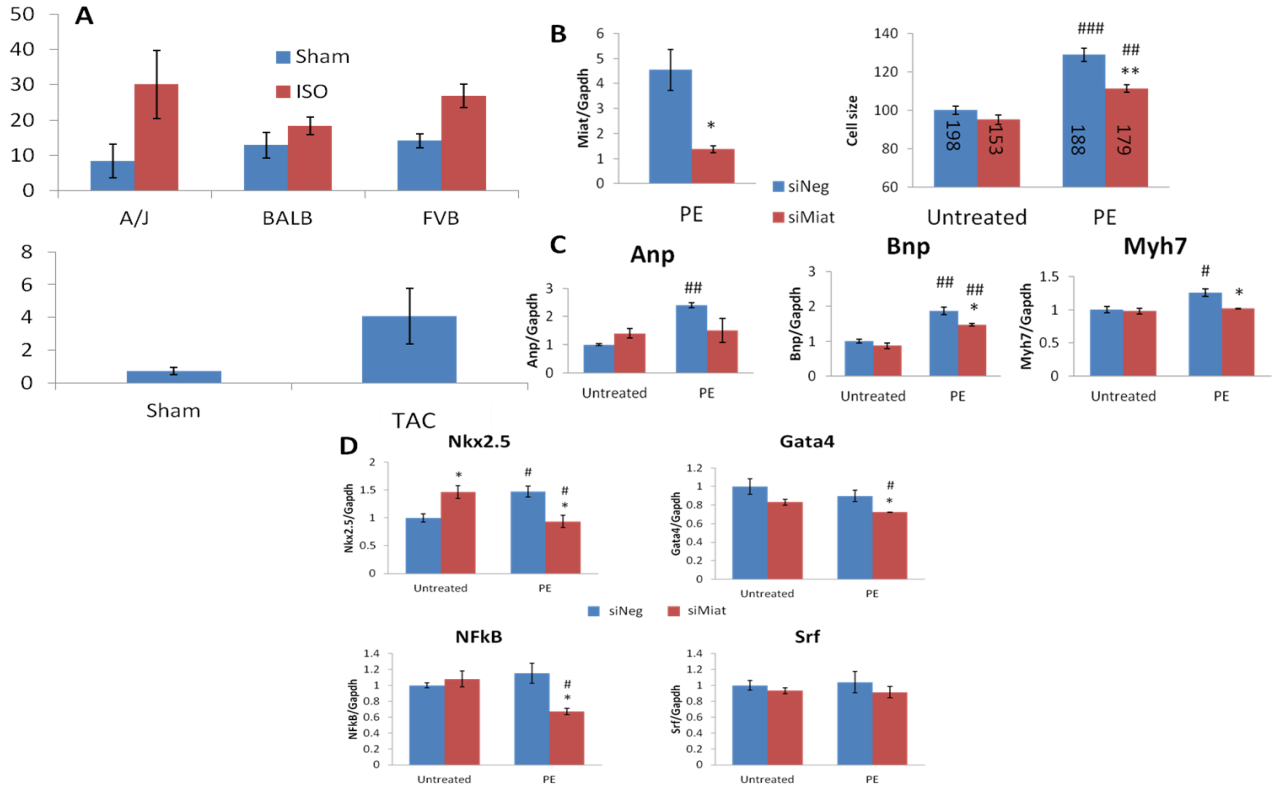

Figure 3.9. *Miat* plays a role in the regulation of cardiac hypertrophy in a NRVM model. A) *Miat* shows upregulation in several strains after Isoproterenol treatment as well as after TAC. B) siRNA mediated knockdown of *Miat* results in reduced levels of MIAT after PE stimulation and results in a significant decrease in the hypertrophy in NRVMs treated with PE. C) Expression of *Anp*, *Bnp* and *Myh7* are reduced after *Miat* knockdown. D) *Miat* knockdown results in the downregulation of *Nkx2.5*, *Gata4* and *NfkB* after PE stimulation in NRVMs, but *Srf* expression is not affected. # P<0.05 compared to untreated. ##P<0.01 compared to untreated. ###P<0.001 compared to untreated. * P<0.05 compared to PE-treated siNeg. ** P<0.01 compared to PE-treated siNeg.

Table 3.2. Top 50 genes associated with change in heart weight and fibrosis after ISO stimulation

Gene	Correlation	P-value
Heart Weight		
Polydom	0.617596	9.85E-21
Timp1	0.613233	2.18E-20
Col14a1	0.611989	2.73E-20
Pqlc3	0.608349	5.22E-20
Cysltr1	0.60517	9.14E-20
Ms4a7	0.601412	1.76E-19
Col1a1	0.598886	2.72E-19
Fbln1	0.597058	3.72E-19
Adamts2	0.596923	3.81E-19
Srpx2	0.593703	6.57E-19
Itgbl1	0.577932	8.72E-18
Arl11	0.575703	1.24E-17
Arhgdig	0.574618	1.48E-17
Tnfrsf22	0.574097	1.60E-17
Akr1b8	0.573019	1.90E-17
Clec4n	0.572797	1.97E-17
Mfap5	0.569964	3.06E-17
Serpina3n	0.569745	3.17E-17
Fstl1	0.564921	6.65E-17
Enpp1	0.563422	8.36E-17
Tnc	0.560378	1.32E-16
2610001E17Rik	0.559099	1.60E-16
Tspo	0.556363	2.41E-16
Lox	0.548158	8.02E-16
Tlr13	0.547795	8.45E-16
BC020188	0.545962	1.10E-15
Clecsf8	0.545292	1.21E-15
Mrc1	0.54384	1.49E-15
Lcp1	0.535766	4.63E-15
Lhfpl2	0.535412	4.86E-15
Sfrp2	0.533209	6.58E-15
Lyzs	0.532612	7.15E-15
Ceecam1	0.530104	1.01E-14
Sod3	0.529277	1.13E-14

Lgals3	0.528909	1.18E-14
AI593442	0.526939	1.54E-14
Ctsk	0.525117	1.97E-14
Rab15	0.523403	2.48E-14
Sparc	0.521834	3.05E-14
Nkd2	0.519423	4.20E-14
Pdgfrl	0.517852	5.15E-14
Col16a1	0.516872	5.86E-14
Fbn1	0.515796	6.74E-14
Tyms	0.513343	9.25E-14
Panx1	0.50964	1.49E-13
Pak3	0.509256	1.56E-13
Cd84	0.507879	1.86E-13
Ptn	0.504475	2.85E-13
Thbs2	0.503324	3.29E-13
Igfbp6	0.5029	3.47E-13
Fibrosis		
Cd84	0.534179617	5.76E-15
4833426J09Rik	-0.504187457	2.95E-13
Trem2	0.501584387	4.08E-13
Lhfpl2	0.492466394	1.24E-12
Pira3	0.477522167	7.19E-12
Fbln1	0.475128456	9.44E-12
Tlr13	0.475071603	9.50E-12
Dpep2	0.472414904	1.28E-11
Lpxn	0.469994676	1.68E-11
Spp1	0.468330459	2.03E-11
Nkd2	0.466066577	2.61E-11
Slc15a3	0.465737638	2.70E-11
Hk3	0.465622455	2.74E-11
Cd68	0.463499051	3.46E-11
Ms4a7	0.463092647	3.61E-11
Timp1	0.460680936	4.70E-11
Lgals3	0.45815764	6.17E-11
Serpina3n	0.457110571	6.91E-11
Lyzs	0.45648218	7.39E-11
Ctsk	0.455264318	8.42E-11
Gpr137b	0.452550428	1.12E-10
D130016K21Rik	-0.447442414	1.92E-10

Gp38	0.447207702	1.96E-10
0610039P13Rik	0.443718264	2.82E-10
Pira11	0.44189737	3.39E-10
Tyrobp	0.439826344	4.19E-10
Arl11	0.434154212	7.41E-10
Rragb	-0.431144226	9.98E-10
Lox	0.430011606	1.12E-09
Col1a1	0.427348939	1.45E-09
Polydom	0.426614942	1.56E-09
Gusb	0.424636321	1.88E-09
Sh3bp2	0.424455578	1.92E-09
Olfml1	0.419376011	3.12E-09
Igf1	0.419125275	3.19E-09
Mfap4	0.416523477	4.08E-09
Lcp1	0.413469105	5.43E-09
Enpp1	0.412453067	5.96E-09
Cotl1	0.412414232	5.99E-09
Pdgfrl	0.410965452	6.84E-09
Rab3il1	0.406262198	1.05E-08
Clecsf8	0.405770469	1.10E-08
Dusp22	-0.405705789	1.11E-08
Wbscr5	0.405578072	1.12E-08
Rhobtb2	-0.405307804	1.15E-08
Sfrp2	0.405274921	1.15E-08
Tmod4	-0.40358668	1.34E-08
Hemp1	0.403176105	1.39E-08
Rbp1	0.402200027	1.52E-08
Anpep	0.401731327	1.58E-08

Table 3.3. DAVID enrichment of all genes significantly correlated to a change in heart weight or fibrosis post-ISO.

Term	Count	%	Total Genes in Category	Fold Enrichment	Benjamini-corrected P-value
Heart Weight					
Signal	86	40.9	2970	2.584923	5.23E-16
Secreted	55	26.1	1420	3.457641	6.02E-14
glycoprotein	91	43.3	3600	2.256547	9.03E-14
Proteinaceous extracellular matrix	27	12.8	297	6.973787	4.95E-13
disulfide bond	71	33.8	2469	2.5671	1.35E-12
glycosylation site:N-linked (GlcNAc...)	84	40	3444	1.963598	2.49E-08
short sequence motif: Cell attachment site	12	5.7	76	12.71172	4.35E-07
ECM-receptor interaction	9	4.2	83	9.017287	2.59E-04
Focal adhesion	13	6.1	198	5.459962	2.83E-04
hydroxylation	8	3.8	64	11.15875	3.15E-04
collagen	8	3.8	84	8.501905	0.001631507
calcium binding	6	2.8	47	11.39617	0.005795532
inflammatory response	7	3.3	78	8.01141	0.007037039
cell adhesion	14	6.6	561	3.288895	0.00946431
domain:TSP N-terminal	5	2.3	1602	21.18619	0.011852841
egf-like domain	10	4.7	91	4.021171	0.02154191
microfibril	3	1.4	4	57.53374	0.026390381
Laminin G, thrombospondin-type, N-terminal	5	2.3	21	21.14643	0.037640676

Fibrosis					
glycoprotein	34	5.1	3600	2.76428051	2.05E-07
signal	31	4.7	2970	3.054998068	3.86E-07
extracellular region	23	3.5	213	3.423714286	5.81E-06
short sequence motif:Cell attachment site	7	1.1	76	24.59364035	3.57E-05
proteinaceous extracellular matrix	10	1.5	78	8.42020202	4.88E-05
disulfide bond	21	3.5	2379	2.726543566	2.52E-04
protein complex binding	5	0.8	78	18.51727982	0.017972598
sequence variant	8	1.2	362	5.90092081	0.021688753
collagen	3	0.5	21	39.48631579	0.03118142
lytic vacuole	5	0.8	179	6.98547486	0.049030232

Table 3.4. Significant overlap of HMDP and human GWAS loci for HF traits. Human loci were considered to overlap with mouse loci if they fell within 5 MB of a mouse locus peak. Overall, six of the twelve currently reported loci are matched in the HMDP study.

Loci shared between Human Studies and the HMDP Study						
Human Study	Human Phenotype	Human Signific-ance	Syntenic Location	HMDP Phenotype	HMDP P-value	HMDP Location
Ellinor et al. & Parsa et al.	Dilated Cardiomyopathy Cardiac Hypertrophy	LOD: 8.2 1E-07	Chr 1: ~187-193 MB Chr1:~189.7MB	RV Weight	1.56E-10	Chr 1: 186.73MB
				Heart Weight	1.08E-06	Chr 1: 186.73MB
Parsa et al.	Cardiac Hypertrophy	1E-07	Chr 4: ~120 MB	Change in Right Ventricle Weight	8.99E-07	Chr 4: 122.0 MB
Parsa et al.	Cardiac Hypertrophy	2E-07	Chr 7: ~66.5 MB	Change in Right Atrial Weight	9.57E-07	Chr 7: 64.06 MB
Smith et al.	Heart failure	1E-08	Chr 9: ~66.25 MB	Change in Right Atrial Weight	4.93E-06	Chr 9: 66.33 MB
Morrison et al.	Mortality among HF patients	3E-07	Chr9:~114.5 MB	Change in Right Atrial Weight	9.97E-06	Chr9: 116.8 MB

Table 3.5. Significant overlap of HF GWAS loci in HMDP with QTL from previous mouse linkage analyses. Overall, 7 of the 16 reported QTLs relating to heart failure in mice are duplicated in the study.

Loci shared between other Mouse Studies and the HMDP Study					
QTL Designation	QTL Phenotype	Location	HMDP Phenotype	HMDP P-value	HMDP Location
Hrq3	Heart Rate	Chr 1: 90.5-170 MB	Initial Left Ventricle Weight	8.47E-06	Chr 1: 154.47 MB
Hrtq1	Heart Weight	Chr 2:64.5-141.8 MB	Final Heart Weight	6.45E-06	Chr2:79.63 MB
Hrq1	Heart Rate	Chr 2: 64.6-159.4 MB	Final Heart Weight	6.45E-06	Chr2:79.63 MB
Cmn1	Dilated Cardiomyopathy	Chr9: 110.4-118.2 MB	Change in Right Atrial Weight	9.97E-06	Chr9: 116.8 MB
Hrtq3	Heart Weight	Chr 10: 26-88.5 MB	Initial Ventricle Weight	1.85E-06	Chr 10: 41.11 MB
Hrtfm6	Heart Failure	Chr 13: 108.1-134.2 MB	Initial Right Atrial Weight	6.75E-07	Chr 13: 117.73 MB
Hrtfm4	Heart Failure	Chr 18: 29.6-65.3 MB	Change in Atrial Weight	2.46E-07	Chr 18: 47.78 MB

Table 3.6. Top 50 genes correlated with *Abcc6* expression in treated hearts

Gene	Correlation
Pdk3	-0.39923
Hnf4a	0.389375
Il17rd	-0.35226
Ppp2r5c	-0.34849
Fthfd	0.340877
Prcc	0.334707
9130427A09Rik	-0.32961
Raf1	0.327894
Pdlim3	-0.32716
Spata6	-0.3261
Khk	0.325041
Sat1	-0.32466
Skd3	0.323688
Yif1a	0.322354
Lsm7	0.321032
Ndrg1	-0.31614
Rad52b	0.311824
Sc5d	0.3113
Pcbd2	0.310854
Tarbp2	0.309935
Tp53i11	0.309146
Usp3	-0.30538
Msh3	-0.30449
Tm6sf2	0.304208
Aox3	0.303498
Shc1	-0.3033
C030002N13Rik	-0.30094
Pcx	0.300781
Acate3	-0.30068
0610012D14Rik	0.30068
Dcn	-0.29921
Dos	0.298249
BC012016	-0.29746
Cd34	-0.29663

Tes	-0.29622
Rhoj	-0.29548
Eif1a	-0.29524
Scube2	0.295225
Mrpl34	0.29516
Slc25a15	0.294806
BC017612	0.294683
1110007C05Rik	0.294225
Ndn	-0.29361
Actr3b	0.293373
Fadd	-0.29336
Spag7	0.29336
Mthfd1	0.292604
Gnmt	0.292274
Cugbp1	-0.29128
Armcx2	-0.29006

Table 3.7. DAVID enrichment of all genes significantly correlated *Abcc6* expression in treated animals.

Term	Count	%	Total Genes in Category	Fold Enrichment	Benjamini-corrected P-value
Abcc6					
mitochondrion	97	18.30189	1322	2.466302	4.65E-15
transit peptide	40	7.54717	457	3.112975	1.36E-07
acetylation	104	19.62264	2325	1.590897	1.51E-04
transferase	65	12.26415	1385	1.66915	0.004939
mitochondrial envelope	27	5.09434	391	2.321096	0.005577
ribosome	16	3.018868	192	2.801075	0.01712
coenzyme metabolic process	16	3.018868	143	3.969545	0.022271
cytoplasm	114	21.50943	3029	1.338559	0.041307

Table 3.8. Expanded details on each locus

Phenotype	Peak SNP	Chromosome	LD Block (Mb)	Peak P-value	Minor Allele Frequency	Genes in Locus
Treated right ventricle weight	mm37-1-134467906	1	133.78-134.53	5.75E-07	0.077	*Adipor1* *Adora1* *Btg2* *Chi3l1* *Chit1* *Cyb5r1* *Fmod* *Klhl12* *Mybph* *Myog* *Optc* *Ppfia4* *Prelp* *Rabif*
Treated right ventricle weight	mm37-5-137934905	5	137.93-138.15	3.49E-10	0.078	*Actl6b* *Agfg2* *Fbxo24* *Gigyf1* *Gnb2* *Irs3* *Lrch4* ***Mospd3*** ***Pcolce*** *Pop7* *Trfr2*
Treated right ventricle weight	mm37-5-23873494	5	23.82-24.47	1.23E-06	0.088	*Abcb8* *Abcf2* *Accn3* *Agap3*

						Asb10 *Atg9b* *Cdk5* *Chpf2* *Crygn* *Fastk* *Gbx1* *Kcnh2* *Nos3* *Nub1* ***Prkag2*** *Rheb* *Slc4a2* *Smarcd3* *Tmub1* *Wdr86*
Treated right ventricle weight	mm37-9-40202022	9	39.77-40.52	8.41E-08	0.08	*1700001J11Rik* *AK007162* *Gramd1b* *Olfr978* *Olfr979* *Olfr980* *Olfr981* *Olfr982* *Olfr983* *Olfr984* *Olfr985* *Olfr986* ***Scn3b*** *Tmem225* *Zfp202*
Treated right ventricle	mm37-10-4981858	10	48.19-54.24	2.80E-07	0.077	*AK005653* *Ascc3* *Asf1a*

weight	3					*Dcbld1* *Fam162b* *Fam184a* *Gopc* *Gp49a* *Gprc6a* *Grik2* *Lilrb4* *Man1a* *Mcm9* *Nepn* *Nus1* ***Pln*** *Rfx6* *Ros1* *Sim1* *Slc35f1* *Vgll2* *Zfa*
Treated right ventricle weight	mm37-11-47181489	11	46.18-49.3	2.15E-06	0.11	*9930111J21Rik2* *AK029860* *BC053393* *Btnl9* *Dppa1* *Fam71b* *Gm12169* *Gm12171* *Gm12185* *Gm4926* *Gm5431* *Gnb2l1* *Havcr1* *Havcr2* *Ifi47*

						Irgm1 *Itk* *Med7* *Mgat1* *Olfr10* *Olfr1386* *Olfr1387* *Olfr1388* *Olfr1389* *Olfr1390* *Olfr1391* *Olfr1392* *Olfr1393* *Olfr1394* *Olfr1395* *Olfr1396* *Olfr56* ***Sgcd*** *Snord95* *Snord96a* *Tgtp2* *Timd2* *Timd4* *Trim41* *Trim7* *Zfp62*
Ratio of right ventricle weights	mm37-3-136305887	3	136.04-136.79	7.83E-07	0.271	***Ppp3ca***
Ratio of right ventricle weights	mm37-9-80542295	9	80.00-80.99	2.94E-07	0.099	*Impg1* ***Myo6***
Ratio of right atrial	mm37-7-	7	142.00-144.81	1.41E-06	0.078	*Clrn3* *Dock1*

weights	142011844					*Ebf3* *Fam196a* *Foxi2* *Glrx3* ***Mgmt*** *Mki67* *Nps* *Ptpre*
Treated liver weight	mm37-7-15251391	7	15.13-18.75	1.93E-07	0.209	*Ap2s1* *Bbc3* *C5ar1* ***Calm3*** *Ccdc8* *Ccdc9* *Ceacam11* *Ceacam12* *Ceacam13* *Ceacam14* *Ceacam15* *Ceacam5* *Ceacam9* *Ceacam-ps1* *Crx* *Crxos1* *Dact3* *Dhx34* *Egam1* *Egam-1c* *Ehd2* *Fkrp* *Gltscr1* *Gltscr2* *Gm4745* *Gng8*

						Gpr77 *Grlf1* *Hif3a* *Kptn* *Meis3* *Napa* *Npas1* *Obox1* *Obox2* *Obox3* *Obox5* *Obox6* *Pnmal1* *Pnmal2* *Ppp5c* *Prkd2* *Prr24* *Psg16* *Psg29* *Psg-ps1* *Ptgir* *Sae1* *Sepw1* *Slc1a5* *Slc8a2* *Snord23* *Strn4* *Tmem160* *Vmn1r90* *Zc3h4* *Zfp541*
Treated liver weight	mm37-10-4946802	10	48.19-54.24	3.48E-06	0.135	*AK005653* *Ascc3* *Asf1a*

	1					*Dcbld1* *Fam162b* *Fam184a* *Gopc* *Gp49a* *Gprc6a* *Grik2* *Lilrb4* *Man1a* *Mcm9* *Nepn* *Nus1* ***Pln*** *Rfx6* *Ros1* *Sim1* *Slc35f1* *Vgll2* *Zfa*
Treated Lung weight	mm37-5-111867706	5	110.87-112.87	1.28E-06	0.258	*Asphd2* *Chek2* *Cryba4* *Crybb1* *Ddx51* *Ep400* *Fbrsl1* *Galnt9* *Hps4* *Hscb* ***Miat*** *Mir701* *Mn1* *Noc4l* *Pitpnb*

						Pus1 *Sez6l* *Srrd* *Tfip11* *Tpst2* *Ttc28* *Ulk1*
Treated Lung weight	mm37-6-53975816	6	53.88-55.57	2.90E-07	0.233	*Adcyap1r1* ***Aqp1*** *Chn2* *Cpvl* *Crhr2* *Fam188b* *Fkbp14* *Gars* *Ggct* *Ghrhr* *Inmt* *Nod1* *Plekha8* *Prr15* *Scrn1* *Wipf3* *Znrf2*
Treated Lung weight	mm37-7-81841621	7	79.8-82.2	3.88E-06	0.244	2610034B18Rik 2900076A07Rik 3110040N11Rik AB349811 AK006397 AK017243 Alpk3 Anpep Ap3b2 Ap3s2 BC048679

						Blm Bnc1 Btbd1 Cib1 Cpeb1 Crtc3 D330012F22Rik Fam103a1 Fes Fsd2 Furin Gm15880 Hddc3 Hdgfrp3 Homer2 Idh2 Iqgap1 LOC100859962 LOC434205 Man2a2 Mesp2 Mir1839 Mir1965 Ngrn Nmb Pde8a Prc1 Rccd1 Rps17 Sec11a Sema4b Slc28a1 Tm6sf1 Ttll13 Unc45a Vps33b

						Wdr73 Whamm Zfp592 Zfp710 Zscan2
Treated Lung weight	mm37-14-14941056	14	8.5-21.5	3.34E-06	0.173	*Abhd6* *Acox2* *Anxa7* *Atxn7* *Cadps* *Dnajc9* *Dnase1l3* *Ecd* *Fezf2* *Fhit* *Flnb* *Gng2* *Il3ra* *Kcnk16* *Kcnk5* *Kctd6* *Lrrc3b* *Mrps16* *Myoz1* *Nek10* *Ngly1* *Nid2* *Nkiras1* *Nr1d2* *Nudt13* *Oit1* *Olfr31* *Olfr720* *Oxsm* *Pdhb*

						Ppp3cb *Psmd6* *Ptprg* *Pxk* *Rarb* *Rpl15* *Rpp14* *Sec24c* *Slc4a7* *Sntn* *Synpo2l* *Synpr* *Thoc7* *Thrb* *Top2b* *Ttc18* *Ube2e1* *Ube2e2* *Usp54* *Zmynd17*
Treated Lung weight	mm37-19-27061190	19	26.68-27.43	2.01E-06	0.205	***Vldlr***
Baseline Fibrosis	mm37-2-139163425	2	13.7-14.0	2.51E-06	0.0625	*AK014986* *Jag1* *Ptpla* *St8sia6* *Stam*
Baseline Fibrosis	mm37-4-84420058	4	84-85	2.20E-06	0.10417	*Bnc2* ***Cntln*** *Sh3gl2*

Baseline Fibrosis	mm37-7-73365047	7	72.3-74.3	1.31E-06	0.10638	*AK007376* *BC046251* *Chsy1* *H47* *Lrrk1* *Pcsk6* *Snrpa1* *Tarsl2* ***Tjp1*** *Tm2d3*
Baseline Fibrosis	mm37-8-64382692	8	63-66.5	1.43E-06	0.0625	*2700029M09Rik* *Aadat* *AK005580* *AK015772* *AK084432* *AK205379* *Anxa10* *Cbr4* *Clcn3* *Ddx60* *Mfap3l* *Nek1* *Palld* *Sh3rf1* *Spock3* ***Tll1***
Baseline Fibrosis	mm37-16-83682440	16	82.5-84.25	2.52E-06	0.06522	*NO GENES*
Baseline Fibrosis	mm37-17-6437673	17	64.2-65.8	1.43E-06	0.0625	*AU016765* *AK019598*

	5					*AK050349* ***Fert2*** *Man2a1* *Pja2* *Tmem232*
Baseline Fibrosis	mm37-18-47786513	18	47.2-48.2	1.87E-06	0.0625	*9130209A04Rik* *AK090337* *AK148393* *Commd10* *Gm5095* *Sema6a*
Treated Fibrosis	mm37-7-52946331	7	52.85-53.42	7.11E-07	0.17021	***Abcc6*** *Abcc8* *AK049852* *AK085337* *AK133391* *Bcat2* *Car11* *Ccdc114* *Cyth2* *Dbp* *Emp3* *Fam83e* *Fgf21* *Fut1* *Fut2* *Grin2d* *Grwd1* *Hsd17b14* *Izumo1* *Kcnj11* *Kcnj14* *Kdelr1*

						Lmtk3 *Mamstr* *Nomo1* *Ntn5* *Plekha4* *Rasip1* *Rpl18* *Sec1* *Spaca4* *Sphk2* *Sult2b1* *Syngr4* *Tmem143*
Treated Fibrosis	mm37-7-68593223	7	60.5-69.5	1.40E-06	0.10638	*A230056P14Rik* *A230057D06Rik* *A230073K19Rik* *A330076H08Rik* *AK009067* *AK045535* *AK081140* *AK086712* *Atp10a* *C230091D08Rik* *Cyfip1* *D7Ertd715e* *DOKist4* *Gabra5* *Gabrb3* *Gabrg3* *Gm9962* *Herc2* *lpw* *Luzp2* *Ndn*

						Nipa1 *Nipa2* *Oca2* *Shyc* *Siglech* *Snord116* *Snord64* ***Snrpn*** *Snurf* *Tubgcp5* *U80893* *Ube3a*
Treated Fibrosis	mm37-7-7225556 4	7	72.3-74.3	1.40E-06	0.11364	*AK007376* *BC046251* *Chsy1* *H47* *Lrrk1* *Pcsk6* *Snrpa1* *Tarsl2* ***Tjp1*** *Tm2d3*
Treated Fibrosis	mm37-15-6990705 6	15	68.4-71.4	9.60E-07	0.2766	***Col22a1*** *Fam135b* *Khdrbs3* *Mir30b* *Mir30d* *Zfat*
Treated Fibrosis	mm37-17-6437673 5	17	64.2-65.8	2.47E-06	0.0625	*AU016765* *AK019598* *AK050349* ***Fert2***

						Man2a1 *Pja2* *Tmem232*
Treated Fibrosis	mm37-X-10277028	X	10-12.5	4.17E-07	0.17021	*Akap4* *AU022751* *1700012L04Rik* *1700054O13Rik* *1810030O07Rik* *2900002K06Rik* *4930403L05Rik* *4930524L23Rik* *4930557A04Rik* *4930578C19Rik* *AK008545* *AK010638* *AK017500* *AK036020* *AK075658* *AK136006* *AK139591* *AK156771* *AK170409* *Atp6ap2* *B630019K06Rik* *BC058988* *Bcor* *Bmp15* *Cacna1f* *Cask* *Ccdc120* *Ccdc22* *Ccnb3* *cip7* *Clcn5*

						Cybb *Cypt7* *Cypt8* *Ddx3x* *Dgkk* *Dusp21* *Dynlt3* *Ebp* *Efhc2* *Eras* *Foxp3* *Fthl17* *Ftsj1* *Fundc1* *Gata1* *Glod5* *Gm14374* *Gm14458* *Gm14459* *Gm14483* *Gm14484* *Gm14501* *Gm14511* *Gm4906* *Gm4984* *Gm5132* *Gm5382* *Gm5634* *Gm5635* *Gm6592* *Gm6787* *Gpkow* *Gpr34* *Gpr82*

						Gripap1 *Hdac6* *Kcnd1* *Kdm6a* *Lancl3* *LOC100270707* *Magix* *Maoa* *Maob* *Med14* *Mid1ip1* *Mir188* *Mir362* *Mir500* *Mir501* *Mir532* *Mir684-1* *Mporc-b* *Mycs* *Ndp* *Nudt10* *Nudt11* *Nyx* *Otc* *Otud5* *Pcsk1n* *Pdzx* *Pim2* *Plp2* *Porcn* *Ppp1r3f* *Pqbp1* *Praf2* *Prickle3*

						Rbm3 *Rpgr* *Shroom4* *Slc35a2* *Slc38a5* ***Srpx*** *Ssx9* *Ssxb1* *Ssxb10* *Ssxb2* *Ssxb3* *Ssxb5* *Ssxb8* *Ssxb9* *Suv39h1* *Syp* *Sytl5* *Tbc1d25* *Tcfe3* *Tfe3* *Timm17b* *Tspan7* *Usp27x* *Usp9x* *Was* *Wdr13* *Wdr45* *Xk*

Table 3.9. List of all strains used in the study

Strain	Untreated Mice	Treated Mice	Early Death
129X1/SvJ	3	8	1
A/J	3	8	3
AKR/J	2	6	0
AXB-10/PgnJ	2	3	0
AXB-12/PgnJ	1	2	2
AXB-13/PgnJ	0	1	1
AXB-18/PgnJ	4	4	1
AXB-19/PgnJ	5	5	2
AXB-20/PgnJ	1	4	2
AXB-4/PgnJ	1	2	0
AXB-6/PgnJ	1	2	1
AXB-8/PgnJ	3	5	0
BALB/cByJ	2	4	1
BALB/cJ	6	12	7
BTBRT<+>tf/J	5	14	12
BUB/BnJ	3	8	3
BXA-1/PgnJ	2	3	1
BXA-11/PgnJ	1	1	0
BXA-12/PgnJ	1	3	3
BXA-14/PgnJ	4	5	0
BXA-16/PgnJ	2	4	0
BXA-2/PgnJ	3	4	1
BXA-24/PgnJ	5	8	4
BXA-4/PgnJ	4	4	0
BXA-7/PgnJ	5	5	0
BXA-8/PgnJ	1	2	1
BXD-1/TyJ	1	0	0
BXD-11/TyJ	2	2	1
BXD-12/TyJ	2	2	0
BXD-14/TyJ	1	3	0
BXD-15/TyJ	1	2	2
BXD-19/TyJ	0	2	1
BXD-20/TyJ	0	1	1
BXD-21/TyJ	6	9	2
BXD-22/TyJ	0	1	1
BXD-24/TyJ	2	1	0
BXD-27/TyJ	1	3	3

BXD-31/TyJ	1	1	0
BXD-32/TyJ	4	5	2
BXD-33/TyJ	1	1	1
BXD-34/TyJ	2	5	5
BXD-38/TyJ	4	6	4
BXD-39/TyJ	2	3	0
BXD-40/TyJ	7	14	9
BXD43	2	5	0
BXD44	2	2	0
BXD45	3	4	2
BXD48	2	4	0
BXD49	3	3	0
BXD-5/TyJ	1	1	0
BXD50	4	5	1
BXD55	2	5	1
BXD56	3	4	3
BXD-6/TyJ	1	1	0
BXD61	4	7	2
BXD62	3	4	1
BXD64	3	3	0
BXD66	3	3	0
BXD68	3	5	0
BXD69	1	0	0
BXD70	3	4	1
BXD71	1	1	0
BXD73	3	5	1
BXD74	2	2	0
BXD75	3	6	1
BXD79	3	4	0
BXD-8/TyJ	1	2	0
BXD84	3	6	1
BXD85	1	2	0
BXD86	2	2	2
BXD87	3	3	0
BXH-19/TyJ	1	3	0
BXH-6/TyJ	4	5	0
BXH-9/TyJ	1	2	0
BXHA1	0	1	0
BXHB2	4	5	0
C3H/HeJ	4	8	1

C57BL/6J	6	11	4
C57BLKS/J	3	5	3
C57L/J	2	3	0
C58/J	2	4	1
CBA/J	3	7	0
CE/J	2	3	0
CXB-11/HiAJ	2	3	1
CXB-12/HiAJ	4	6	2
CXB-13/HiAJ	2	5	1
CXB-3/ByJ	4	3	0
CXB-6/ByJ	3	7	4
CXB-7/ByJ	2	3	1
CXBH	2	5	1
DBA/2J	6	11	3
FVB/NJ	6	10	1
KK/HlJ	2	4	0
LG/J	4	4	1
LP/J	3	3	0
MA/MyJ	3	4	0
NOD/LtJ	3	7	0
NON/LtJ	4	6	1
NZB/BlNJ	2	4	1
NZW/LacJ	3	6	5
PL/J	2	7	2
RIIIS/J	4	9	3
SEA/GnJ	3	9	5
SJL/J	2	7	2
SM/J	3	3	0
SWR/J	4	9	5

References

1. Mudd, J. O. & Kass, D. A. Tackling heart failure in the twenty-first century. *Nature* **451,** 919–28 (2008).

2. Villard, E. *et al.* A genome-wide association study identifies two loci associated with heart failure due to dilated cardiomyopathy. *European heart journal* **32,** 1065–76 (2011).

3. Smith, N. L. *et al.* Association of genome-wide variation with the risk of incident heart failure in adults of European and African ancestry: a prospective meta-analysis from the cohorts for heart and aging research in genomic epidemiology (CHARGE) consortium. *Circulation. Cardiovascular genetics* **3,** 256–66 (2010).

4. Morrison, A. C. *et al.* Genomic variation associated with mortality among adults of European and African ancestry with heart failure: the cohorts for heart and aging research in genomic epidemiology consortium. *Circulation. Cardiovascular genetics* **3,** 248–55 (2010).

5. Parsa, A. *et al.* Hypertrophy-associated polymorphisms ascertained in a founder cohort applied to heart failure risk and mortality. *Clinical and translational science* **4,** 17–23 (2011).

6. Frangogiannis, N. G. Regulation of the inflammatory response in cardiac repair. *Circulation research* **110,** 159–73 (2012).

7. Beltrami, C. a. *et al.* Structural basis of end-stage failure in ischemic cardiomyopathy in humans. *Circulation* **89,** 151–163 (1994).

8. Kong, P., Christia, P. & Frangogiannis, N. G. The pathogenesis of cardiac fibrosis. *Cellular and molecular life sciences : CMLS* (2013). doi:10.1007/s00018-013-1349-6

9. Milting, H. *et al.* Plasma biomarkers of myocardial fibrosis and remodeling in terminal heart failure patients supported by mechanical circulatory support devices. *The Journal of heart and lung transplantation : the official publication of the International Society for Heart Transplantation* **27,** 589–96 (2008).

10. Waterhouse, D. F., Ismail, T. F., Prasad, S. K., Wilson, M. G. & O'Hanlon, R. Imaging focal and interstitial fibrosis with cardiovascular magnetic resonance in athletes with left ventricular hypertrophy: implications for sporting participation. *British journal of sports medicine* **46 Suppl 1,** i69–77 (2012).

11. Wheeler, F. C. *et al.* QTL mapping in a mouse model of cardiomyopathy reveals an ancestral modifier allele affecting heart function and survival. *Mammalian genome : official journal of the International Mammalian Genome Society* **16,** 414–23 (2005).

12. Le Corvoisier, P., Park, H.-Y., Carlson, K. M., Marchuk, D. a & Rockman, H. a. Multiple quantitative trait loci modify the heart failure phenotype in murine cardiomyopathy. *Human molecular genetics* **12,** 3097–107 (2003).

13. Wheeler, F. C. *et al.* Tnni3k modifies disease progression in murine models of cardiomyopathy. *PLoS genetics* **5,** e1000647 (2009).

14. McDermott-Roe, C. *et al.* Endonuclease G is a novel determinant of cardiac hypertrophy and mitochondrial function. *Nature* **478,** 114–118 (2011).

15. Bennett, B. J. *et al.* A high-resolution association mapping panel for the dissection of complex traits in mice. *Genome research* **20,** 281–90 (2010).

16. Dorn, G. W. & Liggett, S. B. Pharmacogenomics of beta-adrenergic receptors and their accessory signaling proteins in heart failure. *Clinical and translational science* **1,** 255–62 (2008).

17. Kumaran, K. S. & Prince, P. S. M. Caffeic acid protects rat heart mitochondria against isoproterenol-induced oxidative damage. *Cell stress & chaperones* **15,** 791–806 (2010).

18. Brittsan, a G. *et al.* The effect of isoproterenol on phospholamban-deficient mouse hearts with altered thyroid conditions. *Journal of molecular and cellular cardiology* **31,** 1725–37 (1999).

19. Zheng, J., Shen, H., Xiong, Y., Yang, X. & He, J. The beta1-adrenergic receptor mediates extracellular signal-regulated kinase activation via Galphas. *Amino acids* **38,** 75–84 (2010).

20. Breckenridge, R. Heart failure and mouse models. *Disease models & mechanisms* **3,** 138–43 (2010).

21. Zhang, X. *et al.* Cardiotoxic and Cardioprotective Features of Chronic β-Adrenergic Signaling. *Circulation research* **112,** 498–509 (2013).

22. Ishii, N. *et al.* Identification of a novel non-coding RNA, MIAT, that confers risk of myocardial infarction. *Journal of human genetics* **51,** 1087–99 (2006).

23. Mungrue, I. N. *et al.* Abcc6 deficiency causes increased infarct size and apoptosis in a mouse cardiac ischemia-reperfusion model. *Arteriosclerosis, thrombosis, and vascular biology* **31,** 2806–12 (2011).

24. Berk, B. C., Fujiwara, K. & Lehoux, S. Review series ECM remodeling in hypertensive heart disease. **117,** (2007).

25. Bronson, R. T. in *Genetic Effects of Aging II* (Harrison, D.) 279–358 (Telford Press, 1990).

26. Kang, H. M. *et al.* Efficient control of population structure in model organism association mapping. *Genetics* **178,** 1709–23 (2008).

27. Yalcin, B. *et al.* Sequence-based characterization of structural variation in the mouse genome. *Nature* **477,** 326–9 (2011).

28. Flint, J. & Eskin, E. Genome-wide association studies in mice. *Nature reviews. Genetics* **13,** 807–17 (2012).

29. Ellinor, P. T. *et al.* A novel locus for dilated cardiomyopathy, diffuse myocardial fibrosis, and sudden death on chromosome 10q25-26. *Journal of the American College of Cardiology* **48,** 106–11 (2006).

30. Shimano, M. *et al.* Cardiac myocyte follistatin-like 1 functions to attenuate hypertrophy following pressure overload. *Proceedings of the National Academy of Sciences of the United States of America* **108,** E899–906 (2011).

31. De Boer, R. a, Yu, L. & van Veldhuisen, D. J. Galectin-3 in cardiac remodeling and heart failure. *Current heart failure reports* **7,** 1–8 (2010).

32. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4,** 44–57 (2009).

33. Dennis, G. *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome biology* **4,** P3 (2003).

34. Eijgelsheim, M. *et al.* Genome-wide association analysis identifies multiple loci related to resting heart rate. *Human molecular genetics* **19,** 3885–94 (2010).

35. Miyazaki, Y., Ikeda, Y., Shiraishi, K., Fujimoto, S. N. & Aoyama, H. Heart Failure-Inducible Gene Therapy Targeting Protein Phosphatase 1 Prevents Progressive Left Ventricular Remodeling. *Gene Therapy* **7,** 1–13 (2012).

36. Vasan, R. S. *et al.* Genetic variants associated with cardiac structure and function: a meta-analysis and replication of genome-wide association data. *JAMA : the journal of the American Medical Association* **302,** 168–78 (2009).

37. Chu, G. & Kranias, E. G. Phospholamban as a therapeutic modality in heart failure. *Novartis Foundation symposium* **274,** 156–71; discussion 172–5, 272–6 (2006).

38. Schönberger, J. & Seidman, C. E. Many roads lead to a broken heart: the genetics of dilated cardiomyopathy. *American journal of human genetics* **69,** 249–60 (2001).

39. Ivandic, B. T. *et al.* Whole-genome analysis of gene expression associates the ubiquitin-proteasome system with the cardiomyopathy phenotype in disease-sensitized congenic mouse strains. *Cardiovascular research* **94,** 87–95 (2012).

40. Burkett, E. L. & Hershberger, R. E. Clinical and genetic issues in familial dilated cardiomyopathy. *Journal of the American College of Cardiology* **45,** 969–81 (2005).

41. Liew, C.-C. & Dzau, V. J. Molecular genetics and genomics of heart failure. *Nature reviews. Genetics* **5,** 811–25 (2004).

42. Bauer, R., Macgowan, G. A., Blain, A., Bushby, K. & Straub, V. Steroid treatment causes deterioration of myocardial function in the {delta}-sarcoglycan-deficient mouse model for dilated cardiomyopathy. *Cardiovascular research* **79,** 652–61 (2008).

43. Pall, G. S. *et al.* A novel transmembrane MSP-containing protein that plays a role in right ventricle development. *Genomics* **84,** 1051–9 (2004).

44. Baicu, C. F. *et al.* Effects of the absence of procollagen C-endopeptidase enhancer-2 on myocardial collagen accumulation in chronic pressure overload. *American journal of physiology. Heart and circulatory physiology* **303,** H234–40 (2012).

45. Itoh, T., Hasegawa, J., Tsujita, K., Kanaho, Y. & Takenawa, T. The tyrosine kinase Fer is a downstream target of the PLD-PA pathway that regulates cell migration. *Science signaling* **2,** ra52 (2009).

46. Xu, G. *et al.* Continuous association of cadherin with beta-catenin requires the non-receptor tyrosine-kinase Fer. *Journal of cell science* **117,** 3207–19 (2004).

47. El Sayegh, T. Y., Kapus, a & McCulloch, C. a. Beyond the epithelium: cadherin function in fibrous connective tissues. *FEBS letters* **581,** 167–74 (2007).

48. Cho, I. J. *et al.* E-cadherin antagonizes transforming growth factor β1 gene induction in hepatic stellate cells by inhibiting RhoA-dependent Smad3 phosphorylation. *Hepatology (Baltimore, Md.)* **52,** 2053–64 (2010).

49. Garrett, M. R., Dene, H. & Rapp, J. P. Time-course genetic analysis of albuminuria in Dahl salt-sensitive rats on low-salt diet. *Journal of the American Society of Nephrology : JASN* **14,** 1175–87 (2003).

50. Clark, T. G. *et al.* The mammalian Tolloid-like 1 gene, Tll1, is necessary for normal septation and positioning of the heart. *Development (Cambridge, England)* **126,** 2631–42 (1999).

51. Arad, M. *et al.* Constitutively active AMP kinase mutations cause glycogen storage disease mimicking hypertrophic cardiomyopathy. *Journal of Clinical Investigation* **109,** 357–362 (2002).

52. Scheffold, T., Waldmüller, S. & Borisov, K. A case of familial hypertrophic cardiomyopathy emphasizes the importance of parallel screening of multiple disease genes. *Clinical research in cardiology : official journal of the German Cardiac Society* **100,** 627–8 (2011).

53. Banerjee, S. K., McGaffin, K. R., Huang, X. N. & Ahmad, F. Activation of cardiac hypertrophic signaling pathways in a transgenic mouse with the human PRKAG2 Thr400Asn mutation. *Biochimica et biophysica acta* **1802,** 284–91 (2010).

54. Brown, C. C. *et al.* A genome-wide association analysis of temozolomide response using lymphoblastoid cell lines shows a clinically relevant association with MGMT. *Pharmacogenetics and genomics* **22,** 796–802 (2012).

55. Pegg, A. E., Fang, Q. & Loktionova, N. a. Human variants of O6-alkylguanine-DNA alkyltransferase. *DNA repair* **6,** 1071–8 (2007).

56. Bugni, J. M., Han, J., Tsai, M., Hunter, D. J. & Samson, L. D. Genetic association and functional studies of major polymorphic variants of MGMT. *DNA repair* **6,** 1116–26 (2007).

57. Chahal, M. *et al.* O6-Methylguanine-DNA methyltransferase is a novel negative effector of invasion in glioblastoma multiforme. *Molecular cancer therapeutics* **11,** 2440–50 (2012).

58. Meng, H. *et al.* Identification of Abcc6 as the major causal gene for dystrophic cardiac calcification in mice through integrative genomics. *Proceedings of the National Academy of Sciences of the United States of America* **104,** 4530–5 (2007).

59. Aherrahrou, Z. *et al.* An alternative splice variant in Abcc6, the gene causing dystrophic calcification, leads to protein deficiency in C3H/He mice. *The Journal of biological chemistry* **283,** 7608–15 (2008).

60. Jiang, Q. *et al.* Parabiotic heterogenetic pairing of Abcc6-/-/Rag1-/- mice and their wild-type counterparts halts ectopic mineralization in a murine model of pseudoxanthoma elasticum. *The American journal of pathology* **176,** 1855–62 (2010).

61. Martin, L. J. *et al.* ABCC6 localizes to the mitochondria-associated membrane. *Circulation research* **111,** 516–20 (2012).

62. Sarwar, R. & Cook, S. a. Genomic analysis of left ventricular remodeling. *Circulation* **120,** 437–44 (2009).

63. Marques, F. Z., Campain, A. E., Yang, Y. H. J. & Morris, B. J. Meta-analysis of genome-wide gene expression differences in onset and maintenance phases of genetic hypertension. *Hypertension* **56,** 319–24 (2010).

64. Velagaleti, R. S. & O'Donnell, C. J. Genomics of Heart Failure. *Heart Failure Clinics* **6,** 115–124 (2010).

65. Fernandez, I. E. & Eickelberg, O. New cellular and molecular mechanisms of lung injury and fibrosis in idiopathic pulmonary fibrosis. *Lancet* **380,** 680–8 (2012).

66. Menni, C. *et al.* Variant on chromosome 9p is associated with left ventricular mass: results from two cohorts of essential hypertensive individuals. *Journal of hypertension* **30,** 2144–50 (2012).

67. Sangrar, W., Gao, Y., Scott, M., Truesdell, P. & Greer, P. a. Fer-mediated cortactin phosphorylation is associated with efficient fibroblast migration and is dependent on reactive oxygen species generation during integrin-mediated cell adhesion. *Molecular and cellular biology* **27,** 6140–52 (2007).

68. Pawłowski, K. *et al.* A widespread peroxiredoxin-like domain present in tumor suppression- and progression-implicated proteins. *BMC genomics* **11,** 590 (2010).

69. Chaudhry, M. A. Expression Pattern of Small Nucleolar RNA Host Genes and Long Non-Coding RNA in X-rays-Treated Lymphoblastoid Cells. *International journal of molecular sciences* **14,** 9099–110 (2013).

70. Barry, G. *et al.* The long non-coding RNA Gomafu is acutely regulated in response to neuronal activation and involved in schizophrenia-associated alternative splicing. *Molecular psychiatry* 1–9 (2013). doi:10.1038/mp.2013.45

71. Tsuiji, H. *et al.* Competition between a noncoding exon and introns: Gomafu contains tandem UACUAAC repeats and associates with splicing factor-1. *Genes to cells devoted to molecular cellular mechanisms* **16,** 479–490 (2011).

72. Finger, R. P. *et al.* Pseudoxanthoma elasticum: genetics, clinical manifestations and therapeutic approaches. *Survey of ophthalmology* **54,** 272–85

73. Prunier, F. *et al.* Pseudoxanthoma elasticum: cardiac findings in patients and abcc6-deficient mouse model. *PloS one* **8,** e68700 (2013).

74. Rosca, M. G. & Hoppel, C. L. Mitochondria in heart failure. *Cardiovascular research* **88,** 40–50 (2010).

75. Chen, Y. *et al.* Dual autonomous mitochondrial cell death pathways are activated by Nix/BNip3L and induce cardiomyopathy. *Proceedings of the National Academy of Sciences of the United States of America* **107,** 9035–42 (2010).

76. Robinson, M. M. *et al.* Acute {beta}-adrenergic stimulation does not alter mitochondrial protein synthesis or markers of mitochondrial biogenesis in adult men. *American journal of physiology. Regulatory, integrative and comparative physiology* **298,** R25–33 (2010).

77. Oka, T. *et al.* Mitochondrial DNA that escapes from autophagy causes inflammation and heart failure. *Nature* **485,** 251–255 (2012).

78. Smyth, G. K. in *Bioinformatics and Computational Biology Solutions using R and Bioconductor* (Gentleman, R., Irizarry, W. & Huber, W.) 397–420 (Springer, 2005).

79. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics (Oxford, England)* **8,** 118–27 (2007).

80. Parks, B. W. *et al.* Genetic control of obesity and gut microbiota composition in response to high-fat, high-sucrose diet in mice. *Cell metabolism* **17,** 141–52 (2013).

81. Ross, R. S. *et al.* 1 Integrins Participate in the Hypertrophic Response of Rat Ventricular Myocytes. *Circulation Research* **82,** 1160–1172 (1998).

82. Brown, D. a *et al.* Modulation of gene expression in neonatal rat cardiomyocytes by surface modification of polylactide-co-glycolide substrates. *Journal of biomedical materials research. Part A* **74,** 419–29 (2005).

Chapter 4

Maximal Information Component Analysis:

a Novel Network Analysis Algorithm

Introduction

Gene network analysis represents an additional means of analyzing data which is returned from a study on the hybrid mouse diversity panel (HMDP). A number of methods to predict, partition and analyze gene function have been proposed, however many of them make assumptions that do not have biological justifications. This chapter contains a reprint of a paper describing a new network construction algorithm, Maximal Information Component Analysis, or MICA. MICA avoids a number of dangerous assumptions made by other methods. We compare MICA to a commonly used network analysis algorithm, WGCNA, and demonstrate that we observe either comparable, or, in the case of case/control studies, improved results with regards to a number of network construction parameters.

frontiers in
GENETICS

METHODS ARTICLE
published: 12 March 2013
doi: 10.3389/fgene.2013.00028

Maximal information component analysis: a novel non-linear network analysis method

Christoph D. Rau[1,2†], Nicholas Wisniewski[1†], Luz D. Orozco[3], Brian Bennett[1,3], James Weiss[1,4] and Aldons J. Lusis[1,2,3]*

[1] Division of Cardiology, Department of Medicine, David Geffen School of Medicine, University of California, Los Angeles, CA, USA
[2] Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, CA, USA
[3] Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA, USA
[4] Department of Physiology, David Geffen School of Medicine, University of California, Los Angeles, CA, USA

Edited by:
Kenneth S. Kompass, University of California, San Francisco, USA

Reviewed by:
Michelle Lacey, Tulane University, USA
Yinglei Lai, The George Washington University, USA

***Correspondence:**
Aldons J. Lusis, Division of Cardiology, Department of Medicine, David Geffen School of Medicine, A2-237 Center for Health Sciences, University of California, 10833 Le Conte Avenue, Los Angeles, CA 90095-1679, USA.
e-mail: jlusis@mednet.ucla.edu
† These authors have contributed equally to this work.

Background: Network construction and analysis algorithms provide scientists with the ability to sift through high-throughput biological outputs, such as transcription microarrays, for small groups of genes (modules) that are relevant for further research. Most of these algorithms ignore the important role of non-linear interactions in the data, and the ability for genes to operate in multiple functional groups at once, despite clear evidence for both of these phenomena in observed biological systems.

Results: We have created a novel co-expression network analysis algorithm that incorporates both of these principles by combining the information-theoretic association measure of the maximal information coefficient (MIC) with an Interaction Component Model. We evaluate the performance of this approach on two datasets collected from a large panel of mice, one from macrophages and the other from liver by comparing the two measures based on a measure of module entropy, Gene Ontology (GO) enrichment, and scale-free topology (SFT) fit. Our algorithm outperforms a widely used co-expression analysis method, weighted gene co-expression network analysis (WGCNA), in the macrophage data, while returning comparable results in the liver dataset when using these criteria. We demonstrate that the macrophage data has more non-linear interactions than the liver dataset, which may explain the increased performance of our method, termed Maximal Information Component Analysis (MICA) in that case.

Conclusions: In making our network algorithm more accurately reflect known biological principles, we are able to generate modules with improved relevance, particularly in networks with confounding factors such as gene by environment interactions.

Keywords: gene expression, ICMg, scale-free topology, MINE, GxE interactions

INTRODUCTION

High throughput biological technologies, such as transcriptome microarrays, have enabled researchers to query biological networks that underlie cellular processes and pathways involved in diseases. Examination of these pathways has led to the discovery of novel biological targets (Gargalovic et al., 2006; Horvath et al., 2006; Dewey et al., 2011; Park et al., 2011). A common form of biological network is the co-expression network, constructed by analyzing the pairwise relationships between RNA transcripts across a set of perturbations (Stuart et al., 2003; Zhang and Horvath, 2005; Keller et al., 2008; Langfelder and Horvath, 2008; Barabási et al., 2011; Park et al., 2011). In these networks, genes whose expression patterns are related to one another form the links or edges of the graph, while the genes themselves form the nodes or vertices. A common means of analyzing co-expression networks relies on algorithms that partition the network into clusters or modules, consisting of genes having strong associations with each other. These modules assist researchers in the identification of key genes and interactions in a biological process by dramatically reducing the overall complexity of the data from thousands of individual genes to a small number of functional components.

Many computational methods (Steffen et al., 2002; Schäfer and Strimmer, 2005; Berger et al., 2007; Langfelder and Horvath, 2008; Parkkinen and Kaski, 2010; Weng et al., 2011) for the analysis of transcriptomes have been developed. A basic assumption made by many of these co-expression methods is that relationships in a biological network can be accurately described using linear dependence measures such as Pearson correlation or a monotonic dependence measure such as Spearman's correlation. However, linear or monotonic relationships approximate only a fraction of the true relationship types observed in

a biological system (**Figure 1**). By limiting subsequent analysis to the linear fraction of the relationships in the biological network, researchers limit their ability to accurately recreate the network and identify the proper gene modules. One means of circumventing this problem has been through the use of Mutual Information (MI), which is capable of identifying non-linear connections in the data, and has been used in several previously described algorithms (Butte et al., 2000; Daub et al., 2004; Margolin et al., 2006; Meyer et al., 2007). A drawback of MI, which has proven difficult to address in some cases, has been its sensitivity to bin size and number as well as an unsatisfying [0-Infinity] range (Reshef et al., 2011). Recently, a modification to MI termed Maximal Information-based Non-parametric Exploration (MINE) has been described that eliminates these two limitations of MI by identifying the ideal bin size and renormalizing the MI measure into a [0,1] state space (Reshef et al., 2011). We utilize MINE in Maximal Information Component Analysis (MICA) to construct networks that are based on a more accurate set of relationships.

Another common assumption made by many module construction algorithms involves the method by which genes are clustered into modules after the underlying network structure has been identified. Many methods adopt a strict clustering approach, where genes are partitioned uniquely into a single module per gene. In some cases, this is done out of necessity (hierarchical tree-based methods), but in many cases it is done purely for computational efficiency. Although convenient and fast, clustering methods that force genes to uniquely exist in a single module result in incomplete modules, missing key genes that link the modules to one another (Parkkinen and Kaski, 2010). An alternate approach assigns "fuzzy" module membership (MM), in which genes can exist in multiple modules simultaneously (Yang, 1993; Daub et al., 2004; Yang et al., 2004; Parkkinen and Kaski, 2010). In MICA, we apply interaction component modeling for genes (ICMg), an iterative module identification method that assigns "fuzzy" MM based on the empirical results of the Latent Dirichlet Allocation algorithm (Parkkinen and Kaski, 2010). By not relying on traditional one to one gene-module approaches, we allow for a more accurate reconstruction of module dynamics and relationships to clinical traits of interest to the researcher.

In this paper, we describe a novel module identification method, MICA, which avoids some of the above unlikely assumptions made by other network algorithms. We then demonstrate its functionality over prior methods by analyzing two large gene expression datasets collected from macrophages and liver from about 100 inbred strains of mice.

RESULTS

We developed MICA as a means to address what we viewed as problematic assumptions made by many other network analysis algorithms. The method relies on the combination of two previously described methods, each of which addresses a one of our

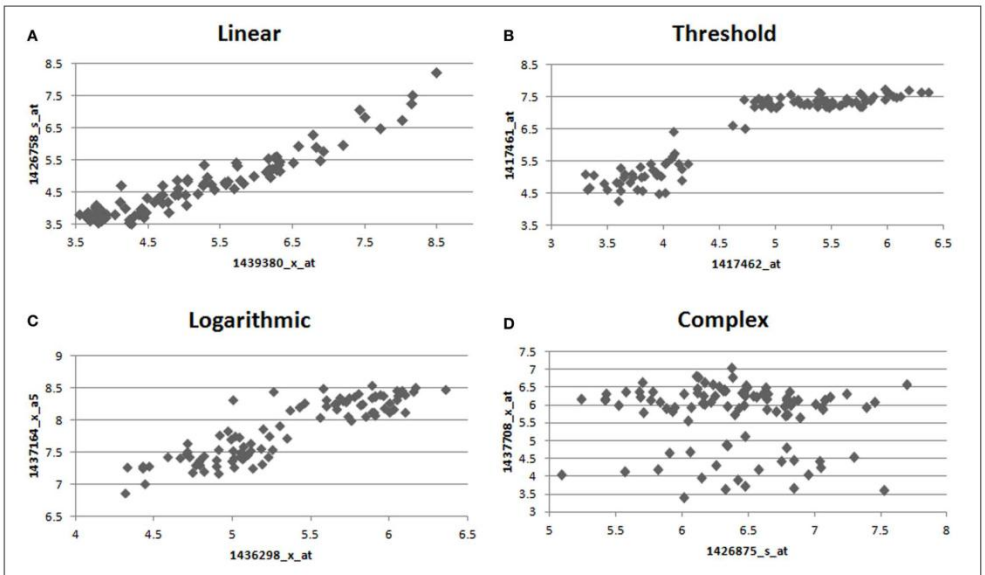

FIGURE 1 | Example relationships observed between pairs of genes in the macrophage dataset. We observe many different forms of relationships, such as **(A)** linear **(B)** high threshold **(C)** logarithmic **(D)** complex, where different relationships are observed between treated and control macrophages. Only interactions of type **(A)** are preferred by Pearson correlation, while all interactions are treated equally by MIC.

primary concerns with current methods. In order to account for the many non-linear interactions which we have observed in our data, we utilized a recently described algorithm, MINE, which can identify and measure both linear and non-linear interactions. We paired this with the ICMg algorithm, which utilizes the data generated by MINE to place genes into multiple modules, which accounts for the multiple interactions in different pathways that genes may have.

For this paper, we utilize two datasets, one on control and treated macrophages and the other from livers. Both datasets were taken from the Hybrid Mouse Diversity Panel (HMDP), a large mouse panel of over 100 strains of mice (Ghazalpour et al., 2012). Millions of SNPs and other genetic perturbations exist between the strains of mice in the HMDP while confounding factors such as environmental variation are minimized, making these datasets ideal for network biology and module identification. Both datasets were verified to have large enough sample sizes to reliably address issues of non-linearity. We first describe the components which make up the MICA algorithm, then compare the results of MICA to the well-regarded weighted gene co-expression network analysis (WGCNA) method (Langfelder and Horvath, 2008) on each dataset.

MICA ALLOWS GENES TO EXIST WITHIN MULTIPLE MODULES

Gene modules attempt to represent groups of genes that act together in a concerted manner. The degree to which a gene belongs to a particular module, a measure known as MM, is a powerful tool for determining the relative importance of individual genes in a given module. In the context of WGCNA, the MM is defined as the correlation of a gene with the module representative (eigengene), and is sometimes also referred to as the module eigengene-based connectivity (kME) (Horvath and Dong, 2008).

While many genes perform only a single role, and would be expected to reside in a single module (have high MM for one module, very low MM for all others), there are other genes that may play roles in multiple pathways. For instance, a transcription factor can activate multiple different pathways; Cytochrome C, which usually is responsible for energy metabolism, also plays an important role in the activation of apoptosis. These genes would have high MM in several modules corresponding to their important roles in each. Therefore, a critical step of network analysis is the calculation of the MM measure. However, by definition the MM measure in WGCNA and other methods are defined on already determined modules. As a result, this approach often produces confusing results. Genes that are placed in other modules will sometimes have higher MMs than many genes within a module. The genes with low MM within a module are counted fully, while those outside are ignored when summarizing the module in question (**Figure 2A**). MICA calculates MMs prior to actual module assignment, which allows all genes with high MM for a module to affect the module, while limiting the effects of genes with low MM. **Figure 2B** shows a sample set of 20 genes taken from our macrophage dataset after analysis with MICA. Several distinct patterns of gene expression can be observed, with most genes showing strong membership in a single module, while others appear to act across two or more modules, including several which do not appear to belong predominantly in any module. By using a weighted PCA algorithm, it is possible to fully incorporate the contributions of each gene to each module, regardless of the magnitude of that contribution.

MICA REPRODUCES SCALE-FREE TOPOLOGY

Work by Barabási and others has suggested that the underlying topology of biological networks is approximately "scale-free" (Langfelder and Horvath, 2008; Barabási et al., 2011; Dewey et al., 2011). In other words, the distribution of node connectivities approximates a power law distribution. Approximate scale-free topology (SFT) has been empirically observed in studies performed on metabolite networks and protein–protein interaction networks (Barabási et al., 2011). Several popular module construction algorithms, including WGCNA, evaluate the fit of their preliminary co-expression network against a SFT. These approaches then systematically modify their co-expression networks in order to maximize the goodness-of-fit to the *a priori* scale-free assumption prior to module partitioning. In the case of WGCNA, Zhang and Horvath (2005) observed that the scale-free fit of a correlation network is highly dependent on the significance threshold used for thresholding the correlation coefficient. They proposed the SFT criterion, which functions by raising each element of the correlation table to a series of powers and comparing the resulting correlation distributions to an idealized SFT distribution. Users are recommended to choose the smallest exponent that allows the scale-free goodness-of-fit criterion to surpass a given threshold (usually an R^2 of 0.9). Raising the correlation matrix to a user-defined power in this way is a significant and severe modification to the original network relationships, with higher powers increasingly distorting the data to favor only the strongest possible connections while devaluing weaker connections.

When Pearson correlation is used for constructing a correlation network, the SFT criterion typically requires one to choose a relatively high power (6 or greater). Using Pearson correlation, we observed that for the macrophage dataset, SFT was only achieved after raising the correlation matrix to the power of 7 (**Figure 3A**). Likewise, the liver dataset requires a power of 16 (**Figure 3C**). At a power of 1 (the original correlation table without modification), the SFT score is negative for the macrophage data and very close to zero for the liver data, indicating a profound disagreement between the raw output of Pearson correlation and an acceptable SFT fit.

In comparison, Maximal Information is a modified version of MI that accurately identifies both linear and non-linear relationships. Strikingly, we find that very low powers are needed to achieve SFT when MINE is used both for the macrophage (**Figure 3B**) and liver (**Figure 3D**) data. We observe that MINE gives a nearly ideal fit to a scale-free network at a power of 2 for macrophage and already passes the recommended threshold at a power of 1 (unmodified) for the liver data. Without modification, the macrophage dataset nearly passes the threshold as well. This suggests that the MINE algorithm naturally captures the hypothesized approximate SFT of biological networks, and eliminates the need to explicitly soft threshold the data with a power function.

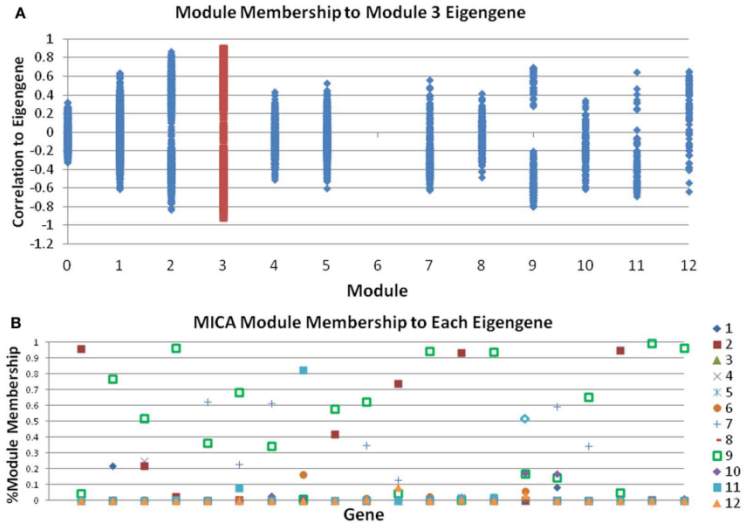

FIGURE 2 | Module Memberships are more meaningful in MICA. (A) Module membership comparisons between all modules and the eigengene of module 3 for the macrophage dataset as calculated by WGCNA. Many genes in other modules, notably modules 2 and 9, have stronger correlations with the eigengene of module 3 than genes in module 3 itself. **(B)** Sample output from 20 genes from MICA, which calculates Module Membership before module analysis and the generation of the eigengene, meaning that all genes contribute to the calculation of each eigengene.

PRINCIPLE COMPONENT ANALYSIS IS CONSERVED ACROSS A WIDE RANGE OF POTENTIAL MM CUTOFFS

Two common goals of module construction algorithms are the identification of enriched pathways, domains, and molecular functions within modules, and the discovery of modules which are strongly correlated with disease severity or other phenotypes of interest (Sharma et al., 2005; Weng et al., 2011). Gene-set enrichment algorithms calculate the overabundance of a particular category of genes within a group when compared to that category's presence in the entire dataset (Huang et al., 2009a,b). To calculate overrepresentation, these methods require strict binary categorization of genes as either being present or absent from a given module. Using MICA, any MM cutoff could theoretically be selected to perform this categorization; however, proper MM cutoffs should preserve the overall action of the MICA-identified module. In order to determine the stability of the network at various MM cutoffs, we calculated the first principle component [called an eigengene (Park et al., 2011)] of each module in our MICA-derived macrophage network at seventeen MM cutoffs (5% intervals from 10% to 90% MM). We then calculated the average correlation of eigengenes to one another and to the weighted PCA which represents the true activity of the module as a whole without partitioning. This stability measurement was high across the panel of MM cutoffs, with a significant loss occurring only when MM cutoffs were greater than 70% or less than 20%. Between 35% and 55% MM cutoffs, eigengene correlations to one another and to the weighted PCA were greater than 0.99 (**Figure 4**). This near perfect correlation implies that within this range of cutoffs, any binary partitioning of the modules is equally capable of describing the action of the network as a whole.

STABILITY OF EIGENGENES ALLOWS FOR SELECTION OF OPTIMAL MODULES IN TERMS OF SIZE AND GENE-SET ENRICHMENT

Ideally, network analysis and module construction should prioritize specific pathways and genes for further analysis by targeted approaches. To achieve that goal, ideal modules should be both highly enriched for specific gene categories, and also small enough to reasonably examine all the genes in the module for interesting candidates and drivers without eliminating large numbers of genes from consideration. In MICA, average module size is inversely correlated with MM cutoff, but the relationship between MM cutoff and gene-set enrichment is significantly more complex. We observe near perfect correlation (greater than 0.99) in the MICA modules for cutoffs that lie between 35% and 55%. This implies that we may select any cutoff within this range for gene enrichment analysis and remain confident that the modules selected accurately represent the entire network as a whole. To determine this ideal cutoff and identify the optimal modules for further analysis, we calculated DAVID enrichment scores for each set of modules (Dennis et al., 2003; Huang et al., 2009b). We then applied a metric that incorporates both module sizes and enrichment significances while penalizing the network for

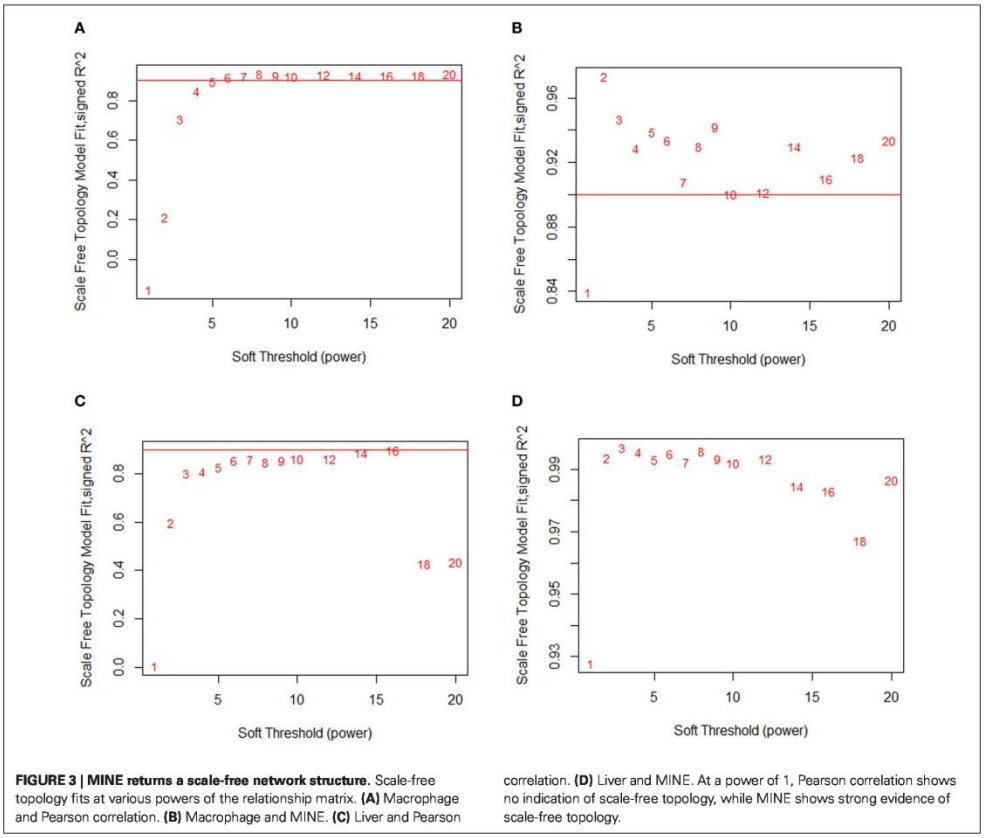

FIGURE 3 | MINE returns a scale-free network structure. Scale-free topology fits at various powers of the relationship matrix. **(A)** Macrophage and Pearson correlation. **(B)** Macrophage and MINE. **(C)** Liver and Pearson correlation. **(D)** Liver and MINE. At a power of 1, Pearson correlation shows no indication of scale-free topology, while MINE shows strong evidence of scale-free topology.

the number of genes not included in the overall network in order to determine the optimal MM cutoff to use for further analysis (**Figure 5A**). For example, in the macrophage dataset, the optimal MM cutoff is 35% a point where the binary partitioned model represents the network as a whole and possesses several small but highly enriched modules.

COMPARISON OF MICA TO WGCNA

WGCNA is an extensively used module identification and network analysis method (Zhang and Horvath, 2005; Langfelder and Horvath, 2008; Dewey et al., 2011; Park et al., 2011). We compared the MICA method to WGCNA using two recently described gene expression microarray datasets from a large mouse panel, one from control and OxPAPC-treated macrophages (Orozco et al., 2012) and another from liver (Bennett et al., 2010). We constructed modules in WGCNA using the standard methodology described in Langfelder et al. (2008). WGCNA infers the number of modules in a co-expression network automatically based on dynamic branch cutting of a hierarchical cluster tree (Langfelder et al., 2008). Additional modules can create bias due to additional degrees of freedom. In order to prevent bias, we fixed the number of modules for MICA to the same number that was inferred through WGCNA.

In comparing WGCNA to MICA, we rely on several measures of network fitness. The first measure of network fitness is the SFT criterion defined by comparing the observed distribution of edge connections across the inferred network to the power-law distribution of an ideal scale-free system. WGNCA suggests raising the correlation matrix to a power in order to reach an appropriate approximation to SFT. We use this method when comparing MICA to WGCNA, observing at which power each method reaches an appropriate approximation to a scale-free system. A method that better captures the SFT of the underlying network is the one that reaches this scale-free criterion threshold at the power closest to unity. The next comparison metric is perplexity, a measure of the entropy of a system,

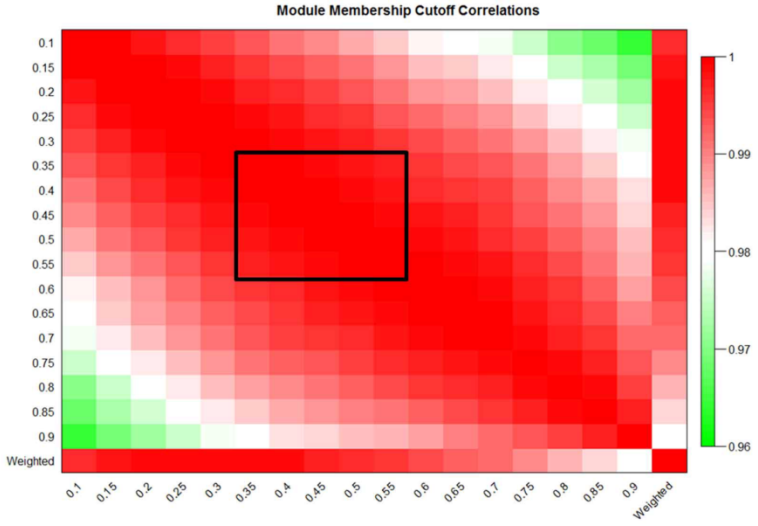

FIGURE 4 | Eigengene values are preserved over a large range of MM cutoffs in MICA. Eigengene correlations at 5% cutoffs from 10% to 90% MM and also the weighted eigengene for the macrophage dataset. The black box indicates the region of preserved module eigengenes where correlation to one another and to the weighted PCA was greater than 0.99.

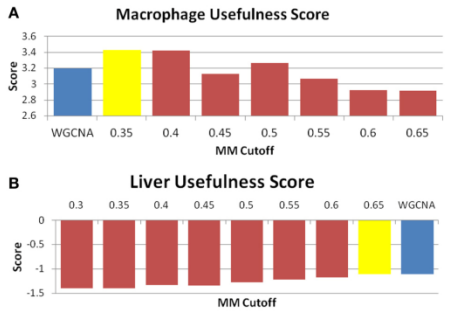

FIGURE 5 | MICA displays equal or higher "usefulness" than WGCNA. "Usefulness" score of data for the **(A)** macrophage and **(B)** liver datasets. The blue bar is the score received by WGCNA, while the yellow bar is the score returned for MICA at the optimal MM cutoff. A higher score indicates a more desirable result. We observe an improvement in the macrophage and conservation in the liver datasets.

and equivalent to a misclassification rate (Brown et al., 1992; Parkkinen and Kaski, 2010). We constructed standard gene classes as described in Shiga et al. (2007) and calculated the ability of either WGCNA or MICA to recapture these classes in their modules. Network analysis methods are often used, particularly in datasets which only vary due to biological variability to determine Gene Ontology (GO) categories for further study (Gargalovic et al., 2006; Horvath et al., 2006; Dewey et al., 2011; Yee et al., 2011; Xiao et al., 2012). We utilized differences in GO enrichments as one measure of network fitness, but felt that a strict comparison of GO enrichment values only captured part of the overall "usefulness" of the constructed modules. To address this issue, the final measure of network fitness compares modules identified through MICA and WGCNA by their "usefulness" as determined by a combination of DAVID gene-set enrichment, module size and number of genes unplaced in modules (Huang et al., 2009b). Ideally, as many modules as possible in a network should be highly enriched and reasonably small to assist in further study.

MACROPHAGE DATASET

We examined a dataset consisting of the 5070 most variably expressed genes in a panel of macrophages isolated from inbred mouse strains before and after treatment with OxPAPC, an oxidized phospholipid. MICA strongly captures the SFT in the system, crossing the recommended threshold at a power of 2 and attaining a nearly perfect fit to an ideal scale-free system with an R^2 of 0.97 (**Figure 2B**). At a power of 1 (the raw relationship values), the scale-free fit is very high at 0.84. By comparison, the Pearson correlation used by WGCNA does not reach the scale-free threshold until a power of 6 (**Figure 2A**). At the power of 1, it is clear that the Pearson correlation is not an accurate means by

which one may capture the SFT of this system, as the signed R^2 value of the degree distributions is less than 0.

The perplexity of the MICA-derived modules varied significantly based on the MM cutoff selected. In comparing WGCNA and MICA, we chose to compare WGCNA to the ideal MM cutoff selected by our "usefulness" measure, which combines gene enrichment and module sizes. WGCNA returned a perplexity score of 193.44 based on 256 standard GO categories included in the analysis (**Figure 6A**). The ideal MM cutoff for the macrophage dataset is 35%. At that cutoff, the MICA modules have a perplexity score of 171.77, a 11.2% improvement over WGCNA. We also calculated perplexities at 5% intervals across the stable range of MICA (35–60%) (**Figure 6A**). In terms of our module "usefulness" measure, we saw improvement (3.43 vs. 3.20) between the 35% MM cutoff MICA modules and the WGCNA modules. Comparable improvement was observed at 40%, while equivalent enrichment was observed at 50%. (**Figure 5A**). Also observed were increases in average GO enrichment at 35 and 40% cutoffs compared to WGCNA (5.457 and 5.384 vs. 5.264, **Figure 7A**).

Orozco et al. (2012) describe a set of genes and GO terms which are involved in the OxPAPC response. In order to determine the ability of MICA and WGCNA to return relevant modules, we examined each set of modules and compared them to the results of Orozco et al. Terms of interest included regulation of kinase activity, cytokine production, genes containing a SH2 domain, glutathione biosynthesis, and oxidative stress response. We compiled lists of all enriched GO terms in both WGCNA and MICA modules. We observed that the MICA-analyzed network contained more modules that were significantly enriched for these GO terms, with six modules being enriched for one or more term of interest as opposed to four in WGCNA. Both methods were able to identify modules involved in oxidative response, regulation of kinase activity and cytokine production, while WGCNA identified an additional module involved in glutathione metabolism and MICA identified two modules associated with SH2 domain and an additional module for regulation of kinase activity. We also observed that MICA segregated all identified OxPAPC-related genes (Hmox1, Ifi205, and Il1a) into a single module, while WGCNA split these genes into multiple modules. The identification of a core "OxPAPC response" module, as defined as the module which contains all the OxPAPC-related genes, represents a significant improvement for MICA over WGCNA, which was unable to find such a module.

LIVER DATASET

We also examined a dataset consisting of 7000 highly expressed genes from livers taken from a large panel of mouse strains. In these data we observed MICA strongly capturing the SFT of the network, reaching an R^2 fit of 0.93 without any modification and an R^2 greater than 0.99 at a power of 2 (**Figure 3D**). By comparison, Pearson correlation did not reach the recommended cutoff of $R^2 = 0.9$ until a power of 16, representing a substantial modification of the co-expression data in order to fit the underlying hypothesis (**Figure 3C**). As in the macrophage data, the unmodified Pearson correlation data showed no relationship to a scale-free network, with an R^2 close to 0.

Unlike the macrophage dataset, we do not observe the same level of conservation of eigengenes across MM cutoffs in the liver dataset. We selected for further analysis a set of MM cutoffs (35–65%) in which eigengene correlations were over 0.9 (**Figure A1**). We also observe that the GO enrichment terms generally improve rather than decrease over the range of conserved MM cutoffs, and our ideal MM cutoff occurs at 65%. At 65%, MICA returns a perplexity score of 105.38, while WGCNA returns a perplexity score of 121.02 (**Figure 6B**). MICA shows a 12.9%

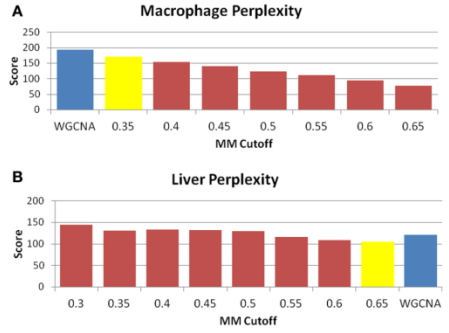

FIGURE 6 | MICA returns a small improvement in module entropy over WGCNA. Perplexity measures for **(A)** macrophage and **(B)** liver. The blue bar is the score received by WGCNA, while the yellow bar is the score returned for MICA at the optimal MM cutoff. As perplexity is a measure of entropy, a lower score is more desirable. In both cases, a small improvement in perplexity is observed in the optimal MICA modules vs. the WGCNA modules.

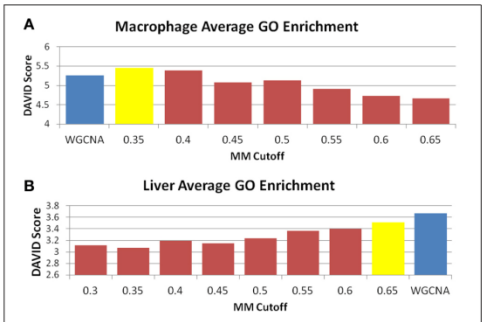

FIGURE 7 | MICA has higher average GO enrichment in the macrophage dataset. Average GO enrichments for modules derived from **(A)** macrophage and **(B)** liver datasets. The blue bar is the score received by WGCNA, while the yellow bar is the score returned for MICA at the optimal MM cutoff. A higher score is more desirable, and we observe improved average GO enrichment for MICA in the macrophage data, and improved average GO enrichment for WGCNA in the liver data.

improvement over WGCNA in terms of perplexity for the liver dataset.

WGNCA was unable to place 66.5% of genes into modules, which affected its "usefulness" score compared to MICA, which was unable to place 10.9%. However, WGCNA returned higher average GO enrichments (3.67 vs. 3.51, **Figure 7B**) when compared to MICA. Modules were indistinguishable from one another in terms of overall module usefulness (−1.10201 vs. −1.10213, **Figure 5B**).

MODULE STABILITY

In order to determine the overall stability of the modules observed in both WGCNA and MICA, we randomly partitioned the macrophage dataset into two equal parts and ran both MICA and WGCNA on each half. No universally accepted means of comparing two sets of modules to one another exists, particularly in the case of modules with non-binary gene-module occurrence. We adopt a method previously used to compare modules created by WGCNA (Langfelder and Horvath, 2008) to compare the MICA and WGCNA modules to one another. We note that this method was designed for network methods which place genes into single modules, and that forcing our MICA results to conform to this requirement will inevitably weaken the network stability observed via MICA.

We observed that when run through the soft thresholding function, both MICA runs return a power of 3, while the two runs of WGCNA differ, with one returning 4 and the other 6. This suggests that MICA is capturing similar levels of SFT for each portion of the data while WGCNA is unable to do so. The hard thresholding criteria for both MICA runs is also identical at a cutoff of 0.45. We further observe that WGCNA returns differing numbers of modules (13 vs. 14) for the two halves of the macrophage dataset.

We observe broadly similar levels of stability in both WGCNA- and MICA-derived modules (**Figure 8**), with the majority of modules in both methods showing strong preservation between the halves of the macrophage dataset. A notable exception is the salmon module from part 1 of the WGCNA data (**Figure 8A**), which is not preserved at all in the part 2 WGCNA network. The salmon module of part 2 of WGCNA also shows relatively weak preservation as well.

EFFECTS OF DATASET ON MICA

We find that MICA appears to show an overall improvement in module construction when compared to WGCNA in the macrophage dataset, but is comparable to WGCNA in the liver dataset. To evaluate whether underlying differences in the network architecture between the two datasets led to the differences in improvement, we returned to the original data to look for differences in the number of non-linear interactions captured by MINE vs. Pearson correlation. If there are more non-linear interactions in a dataset, then MICA should perform better than WGCNA, which does not take into account the non-linear interactions in the data. On the other hand, if a network has very few non-linear interactions, then both MINE and Pearson correlation should return comparable results to one another.

In order to determine whether we were seeing more non-linear interactions in the macrophage dataset, we selected all relationships from both the macrophage and liver datasets that had a high (greater than 0.9) maximal information coefficient (MIC) score. Our first observation was that the macrophage dataset had significantly more strong MIC scores than the liver dataset (1274 vs. 360 interactions). We then examined the distribution of the Pearson correlation values measured for these strong MIC interactions, after sampling the macrophage dataset such that it had an equal number of observations as the liver data (**Figure 9A**). Compared to the liver data, the macrophage data showed enrichment for both very high (greater than 0.9) and low (less than 0.6) Pearson correlations. This suggests that the macrophage dataset both contains more non-linear interactions, and also a greater fraction of interactions that are very close to perfectly linear. While the linear interactions will be picked up by Pearson correlation, the increased number of non-linear interactions can only be detected appropriately through MINE.

There are two possible explanations for the differences between the two datasets. The first is that the macrophage dataset is an *in vitro* system containing a single cell type, while the liver samples contain multiple cell types. The second possible explanation is that the improvement comes because we analyzed both treated and untreated data together, rather than separately. Accordingly, we separated the control and OxPAPC-treated macrophages and compared each separately to the liver dataset (**Figures 9B,C**). We observed slightly increased numbers of strong MIC interactions for the control (448) and treated (549) data compared to the liver data (360). However, although we continued to observe enrichment of very high correlations in both the control and treated OxPAPC datasets compared to the liver data, and we no longer observed enrichment of low correlations in either data (with the exception of a single interaction in the OxPAPC-treated dataset). This is an example of gene by environment interactions where a treatment or environmental perturbation interacts with underlying genetic variation to result in different relationships between genes in different environmental conditions. These observations suggest that the improvement observed when using MICA on the macrophage dataset is a result of MICA's ability to capture gene by environment interactions between the treated and control samples. It further suggests that Pearson correlation and WGCNA are less successful in the macrophage dataset because they are incapable of using these interactions.

DISCUSSION

We report a novel network analysis method, MICA, which combines two previously published methods: MINE, a modification of MI which accounts for non-linear interactions in datasets without many of the shortcomings of the canonical methods, and ICMg, which relies on an iterative process to assign distributed MMs as opposed to a rigid in-or-out dichotomy. Together, this combination is less restrictive than module construction algorithms that include linear but exclude non-linear co-expression relationships and allow only single-MM. Thus, MICA has the advantage that it embraces concepts that are better rooted in actual biological observations.

To validate the MICA approach, we analyzed datasets from macrophages treated with OxPAPC and from livers, which in

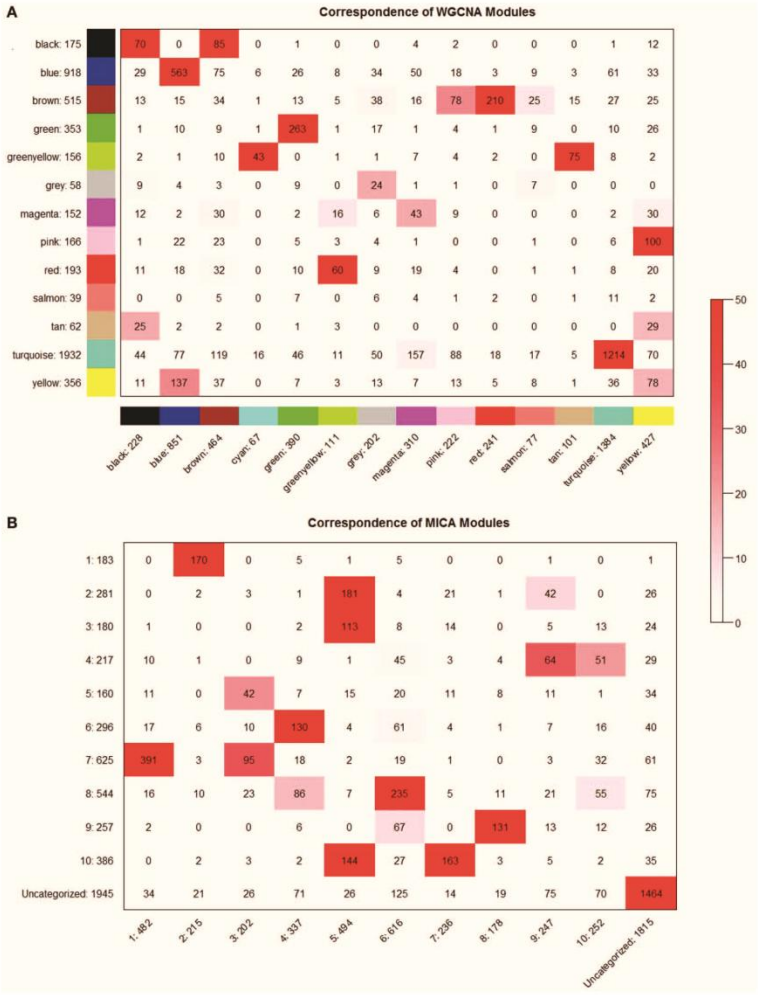

FIGURE 8 | MICA and WGCNA have comparable module preservation statistics. Module preservation between even partitionings of the macrophage dataset for **(A)** WGCNA and **(B)** MICA. Preservation significances are calculated using a Fisher's exact test and are represented by the depth of the red color as indicated.

one case revealed distinct advantages of MICA over WGNCA, a benchmark correlation network approach, and yielded comparable results in the other case. Specifically, MICA may be particularly well suited for the analysis of networks in which gene by environment interactions are expected to occur, which traditional module construction methods are ill-equipped to detect. In the case of macrophages treated with OxPAPC, analysis with MICA resulted in modules that are more highly enriched in pathways of interest, and better able to place genes with similar functions into the same modules compared to other methods. In both macrophage and liver datasets, there is a dramatic improvement in the ability for MICA to detect an overall network structure

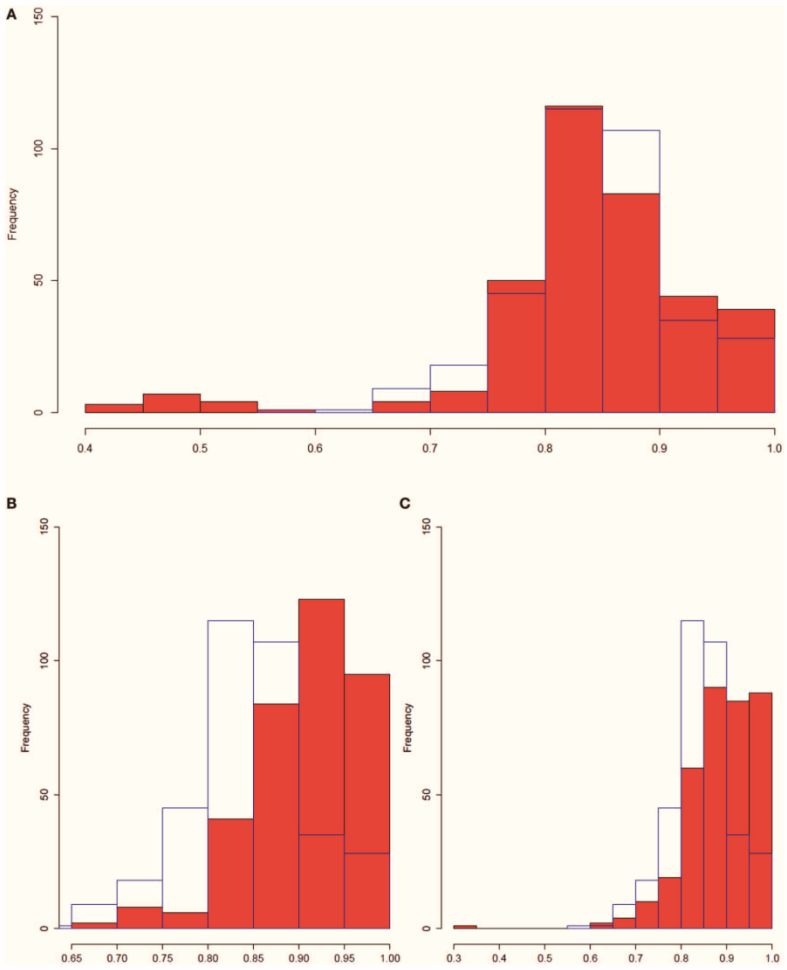

FIGURE 9 | The entire macrophage dataset has a greater number of non-linear interactions. Histograms comparing distributions of Pearson correlations for interactions with high (>0.9) MIC scores. All distributions were sampled such that there were 360 entries for each entry. The blue border in each part of the figure is the liver dataset. **(A)** Full macrophage dataset vs. liver dataset. **(B)** Control macrophage vs. liver. **(C)** Treated macrophage vs. liver. We observe an enrichment for low Pearson correlation data with high MIC scores (implying significant non-linearity) in the full macrophage set which is not observed in either the control, treated, or liver datasets.

that better approximates the hypothesized topology underlying the biological network.

We have further observed that in contrast to MICA, which utilizes MINE and ICMg, no significant improvements were achieved when WGCNA was modified by using ICMg on Pearson correlations, or topological overlap and hierarchical clustering on Maximal Information scores (**Figure 10**). Thus, both MINE and ICMg each provide partial solutions that are synergistic when combined.

As an attempt to incorporate known biological principles such as feedback loops and multi-functional proteins into a transcriptome co-expression network analysis method, MICA shows initial

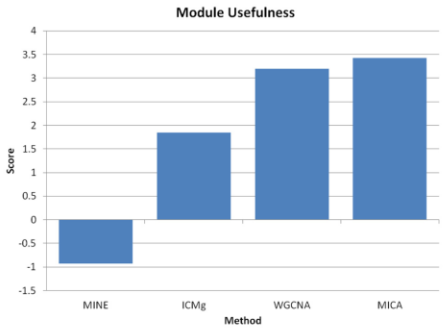

FIGURE 10 | MINE or ICMg used separately do not improve network analysis. Module "usefulness" measure for macrophage data when using either MINE plus hierarchical clustering, Pearson correlation plus ICMg, WGCNA, or MICA. Higher usefulness is more desirable, and we observe significantly reduced usefulness scores when either MINE or ICMg are used separately from one another.

promise, but will undoubtedly benefit from further refinement. A major drawback is that MICA lacks the ability to independently determine the number of modules in a network; i.e., an arbitrary number of modules must be specified. Additionally, it is not straightforward to calculate *p*-values (significance) for the MICA measure, in contrast to the many approaches that have been developed (e.g., regression models) to calculate *p*-values for correlation measures. Finally, MICA is significantly more computationally intensive than correlation-based methods such as WGCNA. Such improvements may allow MICA to identify smaller, more enriched and more relevant modules for further analysis and discovery of novel genes with roles in important phenotypes.

In conclusion, MICA is an attractive network analysis method because (1) it does not discard non-linear interactions; (2) it removes the need for soft thresholding; (3) employs a fuzzy clustering algorithm for module detection; and (4) shows improvements over correlation algorithms in certain cases, particularly those involving gene by environment interactions.

METHODS

MAXIMAL INFORMATION NON-PARAMETRIC EXPLORATION (MINE)

We utilize the recently described MINE algorithm to determine a normalized relationship matrix which incorporates non-linear interactions (Reshef et al., 2011). MINE relies on a modified version of MI called the MIC. MIC calculates normalized MI values for all partitions of a finite set of ordered pairs with the *x*-values going into x bins and the *y*-values partitioned into y bins, such that $x \times y < n^{0.6}$, as recommended by the authors and where n is the number of arrays. The algorithm then normalizes across partitions with the same number of bins, but different bin sizes, by dividing the data by $\log(\min\{x, y\})$, which is the maximum possible score for any MI query with x horizontal and y vertical bins. The MIC of an interaction is then defined as the maximum normalized value across the set of partitions. MINE is implemented in a Java environment. MIC scores are calculated for all pairs of expression data and compiled into a matrix format.

SCALE-FREE TOPOLOGY

The soft thresholding SFT function of the WGCNA R package (Langfelder and Horvath, 2008) was used to determine the fit of all datasets and relationship generating methods (either Pearson correlation or MIC) to an idealized SFT. Briefly, the function acts by calculating the sum of the link strengths for each gene in the data, and finds an R^2 between the distribution of total node link strengths and a power-law distribution. It then repeats this process, raising the original relationships to a power of $n = 1 - 20$. The ideal soft thresholding criterion is defined as the first power which passes the recommended R^2 threshold of 0.9.

ICMg

ICMg (Parkkinen and Kaski, 2010) relies on an iterative component model to calculate MMs. As ICMg does not allow for weighted edges, network edges were trimmed using the hard thresholding function of the WGCNA R package, which calculates an R^2 fit between the degree of node connectivity in a dataset based on a thresholding function at increasing intervals of 0.05 and a power-law distribution. The ideal hard thresholding criterion is the lowest cutoff which passes a recommended R^2 threshold of 0.9. ICMg allows users to select an arbitrary number of modules. As WGCNA automatically selects the number of modules it will return, we selected a number of modules equal to that observed with WGNCA for ICMg in order to avoid biasing the results toward the method with more modules and therefore more degrees of freedom. Module assignments were then initially assigned to the network using a Dirichlet distribution.

ICMg is an iterative process. In each iteration, each edge is independently interrogated utilizing Gibbs sampling with the following equation:

$$p(z_0|\{z\}', \{L\}', \alpha, \beta) \propto \frac{n'_{z_0} + \alpha}{N' + C_\alpha} \times \frac{(q'_{z_0 i_0} + \beta)(q'_{z_0 j_0} + \beta)}{(2n'_{z_0} + 1 + M\beta)(2n'_{z_0} + M\beta)}$$

where $\{L\}'$ is the set of all links excluding the one being interrogated, $\{z\}'$ is the set of module assignments for the links excluding the link being interrogated, n_z is the count of links assigned to component z, i, and j represent the genes linked by edge z_0 and q_{z_i} counts the module-node co-occurrences between module z and node i. C is the total number of modules, and M is the total number of nodes. α and β are control parameters which modify the overall distribution of module sizes and the average MM per gene per module, however, these were not modified and the default values ($\alpha = 10$, $\beta = 0.1$) found in (Parkkinen and Kaski, 2010) were used. 40,000 burn-in rounds were performed to eliminate any dependence on initial conditions and to allow values of q/M to reach steady-state values at which point MM of each node was sampled every 10 iterations for another 10,000 iterations of the network to determine proportional MM in each module for each gene.

DETERMINING THE ABILITY OF MICA TO ADDRESS NON-LINEARITY IN DATASETS

In order to determine whether the sample sizes of the macrophage and liver datasets were large enough to reliably address non-linearity, we utilized pre-computed bootstrapped tables of *p*-values for arrays of varying sample sizes from MINE available at http://www.exploredata.net/Downloads/P-Value-Tables. We observe that at our hard thresholding cutoff of 0.45, this means that for all edges in the liver MICA network, *p*-values are less than 2.72E-6, while for the macrophage dataset, all edges have *p*-values less than 2.74E-7, far over the nominal significance value of 0.05 or the Bonferroni corrected values of 7.1E-6 and 9.8E-6.

CALCULATING "EIGENGENES" VIA WEIGHTED PRINCIPLE COMPONENT ANALYSIS

We borrow the concept of the eigengene from WGCNA (Langfelder and Horvath, 2008) to describe the overall behavior of a set of genes. As in WGCNA, we define an eigengene of a module to be the first principal component of the transcript levels of the genes contained within the module, however, we utilize the dudi.pca function of the ade4 R package to implement a weighted PCA which utilizes the MMs from ICMg to weight the contribution of each gene to the eigengene (Chessel et al., 2004). We also calculate the unweighted eigengene for each module at 5% intervals across the genome in which genes whose MM for that module passes the current threshold are included in the eigengene calculation.

OPTIMAL MM CUTOFF SELECTION

While there are methods to compare networks to one another (Langfelder et al., 2011), these typically are concerned with determining preservation of modules and comparing individual genes to one another and not asking which module is objectively "better." In order to compare MICA networks to one another and to WGCNA-derived networks, we define a parameter "usefulness" (U), which incorporates both GO enrichment scores, the number of genes present in a given module and the number of genes not placed in any module. We define "usefulness" as follows:

$$U = \sum_{i}^{1,\ldots,n} \left(\frac{\mathrm{DAVID}_i}{\log_2 N_i} \right) - \log_{10} M,$$

where DAVID_i is the maximum DAVID (Dennis et al., 2003; Huang et al., 2009b) enrichment score for module i (equivalently, the negative log of a GO enrichment score could be used), N_i is the number of genes in module i, and M is the number of genes not included in any module for the current method.

WGCNA

We followed the network analysis methods described in Langfelder et al. (Langfelder and Horvath, 2008) and the parameters found in the online WGCNA tutorials at http://labs.genetics.ucla.edu/horvath/htdocs/CoexpressionNetwork/Rpackages/WGCNA/Tutorials/. Pearson correlations were determined for each pair of genes, and after performing a soft thresholding SFT fit, the correlations were raised to the recommended power. Adjusted correlations are then converted into Topological Overlap measures by the following equation:

$$\mathrm{TOM}_{ij} = \frac{\sum_u \{A_{iu} A_{uj}\} + A_{ij}}{\min(k_i, k_j) + 1 - A_{ij}}$$

where i and j are the pair of genes to be analyzed, u is the set of all other genes, A is the adjusted correlation matrix, and k is the degree of the node. TOM scores are then converted to DistTOM scores by subtracting TOM from 1. The DistTOM array undergoes hierarchical clustering, and modules are determined using the dynamic tree cut algorithm and eigengenes are determined from the first principle component of the genes in each module. Modules whose eigengenes have a Pearson correlation of greater than 0.8 are merged.

The WGCNA method is implemented in the freely available WGCNA R package (Langfelder and Horvath, 2008). Here we used many of the R functions from this package (e.g., for evaluating SFT and the creation of **Figures 4**, **A1**).

STANDARD GENE CLASSES

The GO database is organized into three distinct directed acyclic graphs. We derived standard gene classes for our data in a method similar to Shiga et al. (2007). Starting at the root of the Biological Process GO graph, we proceeded from parent node to child nodes, checking the number of genes in that GO category that also appear in any module in our gene networks. As we progress away from the root, the number of genes in each category decreases and the number of categories increases. We used the parameters utilized in Shiga et al. for our analysis. When a GO category contains less than 30 genes present in our network, we stop progressing down that branch and add its parent GO category to the standard gene-set, unless there are more than 300 included genes in that category, in which case it was omitted as being too broad for logical compartmentalization into a single module. In this way, we generate a set of reasonably sized functionally-related gene-sets with which to explore the accuracy of the module construction method using perplexity.

PERPLEXITY

Perplexity represents a measure of the entropy in a system, and has been used extensively in fields as diverse as natural language processing (Brown et al., 1992) to previous clustering algorithms (Parkkinen and Kaski, 2010). In this case perplexity represents the ability of a module creation algorithm to accurately recover underlying functional gene categories as determined by our standard gene classes. We applied perplexity to the confusion matrix formed of the frequency of co-occurrence between standard classes on the columns (c) and the modules as the rows (r). From this confusion matrix, perplexity is defined as

$$\mathrm{perplexity} = 2^{-\frac{\sum_l \log \hat{p}(c_l|r_l)}{N}}$$

where N is the total number of non-zero samples, l is an indexing variable for all such entries in the confusion matrix, and the probabilities $\hat{p}$ are empirically determined by normalizing the rows of the confusion matrix. Perplexity is proportional to the size of the

APPENDIX

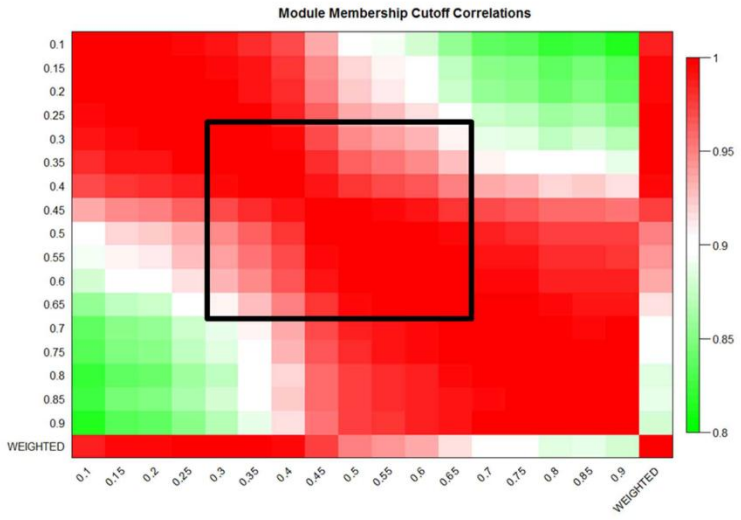

FIGURE A1 | MICA shows strong preservation of eigengenes in the liver dataset. Eigengene correlations at 5% cutoffs from 10% to 90% MM and also the weighted eigengene for the liver dataset. The black box indicates the region of preserved module eigengenes.

overall dataset. To compare perplexities between network methods, we normalize the data by multiplying each perplexity value by the proportion of genes initially included in the dataset and the genes actually placed by each method. A lower perplexity score represents a more accurate capture of the functional categories.

MODULE STABILITY

Module stability was calculated using the method described by (Langfelder and Horvath, 2008) and documented at http://labs.genetics.ucla.edu/horvath/htdocs/CoexpressionNetwork/Rpackages/WGCNA/Tutorials/Consensus-RelateToFemMods.pdf. Briefly, the macrophage dataset was randomly divided into two halves. Each half was independently processed using MICA and WGCNA. MICA genes were forced into the module in which it had the highest MM to allow for the use of the method. In order to determine module preservation, each half was compared to one another by creating a table of gene-overlaps between genes in modules of the first half and genes in modules of the first half. A Fisher's exact test was applied to each overlap to calculate a significance of preservation for each module–module pairing. Overall module preservation was then visually determined based on the significance of preservation for each module in the other half of the dataset.

MICA

Code for MICA is available from systems.genetics.ucla.edu

DATASETS

The macrophage dataset was obtained from Orozco et al, which isolated primary macrophages from a large panel of inbred mouse strains (Orozco et al., 2012). The macrophage dataset includes 80 strains of control macrophages and macrophages treated with 50 ug/ml oxidized 1-palmitoyl-2-arachidonoyl-sn-glycero-3-phosphatidylcholine (OxPAPC) for 4 h.

The liver dataset was taken from Bennett et al. and includes livers taken from 97 strains of mice (Bennett et al., 2010). Transcriptome data was obtained using the Affymetrix HT MOE-430A microarray platform, and normalized using the robust multichip average (RMA) method.

A major limitation of MICA is the time involved in the generation of the MIC scores using MINE, which has a large $O(n^2)$ computation time. In order to run MICA in a reasonable amount of time, it is important to limit the genes selected to the smallest informative set. As such, we selected for genes which were expressed in the dataset and which showed variation across the dataset (as genes which do not vary are generally uninformative for network analysis). We calculated average signal intensity and coefficient of variation (CV) for each probeset. We then reduced our dataset to relevant genes by first selecting probes with above average intensity, and then selecting probes with greater than 5% CV, resulting in 5070 genes for the Macrophage dataset. For the liver dataset, we selected the 7000 most highly expressed genes for analysis.

Both datasets are available at http://systems.genetics.ucla.edu/data/

ACKNOWLEDGMENTS

We would like to thank Steve Horvath for comments and suggestions on this manuscript. We would also like to thank Mario Deng, Alan Garfinkel, Jonathan Hoffman, Steve Horvath, Hrayr Karagueuzian, Zhilin Qu, Yibin Wang, and Jessica Wang for comments and feedback during the development of MICA. This work was supported by the Ruth L. Kirschstein National Research Service Award T32HL69766 (CR), 1 R21 HL110667 (CR and AL), HL28481 and HL30568 (AL), the Jim Easton CDF Investigator Award (NW), K99/R00 HL102223 (BJB), the Laubisch and Kawata endowments (JW) and the USPHS National Research Service Award GM07104, and USPHS National Research Service Award T32-HG002536 (LO).

REFERENCES

Barabási, A.-L., Gulbahce, N., and Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.* 12, 56–68.

Bennett, B. J., Farber, C. R., Orozco, L., Kang, H. M., Ghazalpour, A., Siemers, N., et al. (2010). A high-resolution association mapping panel for the dissection of complex traits in mice. *Genome Res.* 20, 281–290.

Berger, S. I., Posner, J. M., and Ma'ayan, A. (2007). Genes2Networks: connecting lists of gene symbols using mammalian protein interactions databases. *BMC Bioinformatics* 8:372. doi: 10.1186/1471-2105-8-372

Brown, P. F., Pietra, V. J. D., Mercer, R. L., Pietra, S. A. D., and Lai, J. C. (1992). An estimate of an upper bound for the entropy of english. *Comput. Linguist.* 18, 31–40.

Butte, A. J., Tamayo, P., Slonim, D., Golub, T. R., and Kohane, I. S. (2000). Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. *Proc. Natl. Acad. Sci. U.S.A.* 97, 12182–12186.

Chessel, D., Dufour, A. B., and Thioulouse, J. (2004). The ade4 package - I: one-table methods. *R News* 4, 5–10.

Daub, C. O., Steuer, R., Selbig, J., and Kloska, S. (2004). Estimating mutual information using B-spline functions–an improved similarity measure for analysing gene expression data. *BMC Bioinformatics* 5:118. doi: 10.1186/1471-2105-5-118

Dennis, G., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., et al. (2003). DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* 4, P3.

Dewey, F. E., Perez, M. V., Wheeler, M. T., Watt, C., Spin, J., Langfelder, P., et al. (2011). Gene coexpression network topology of cardiac development, hypertrophy, and failure. *Circ. Cardiovasc. Genet.* 4, 26–35.

Gargalovic, P. S., Imura, M., Zhang, B., Gharavi, N. M., Clark, M. J., Pagnon, J., et al. (2006). Identification of inflammatory gene modules based on variations of human endothelial cell responses to oxidized lipids. *Proc. Natl. Acad. Sci. U.S.A.* 103, 12741–12746.

Ghazalpour, A., Rau, C. D., Farber, C. R., Bennett, B. J., Orozco, L. D., van Nas, A., et al. (2012). Hybrid mouse diversity panel: a panel of inbred mouse strains suitable for analysis of complex genetic traits. *Mamm. Genome* 23, 680–692.

Horvath, S., and Dong, J. (2008). Geometric interpretation of gene coexpression network analysis. *PLoS Comput. Biol.* 4:e1000117. doi: 10.1371/journal.pcbi.1000117

Horvath, S., Zhang, B., Carlson, M., Lu, K. V., Zhu, S., Felciano, R. M., et al. (2006). Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. *Proc. Natl. Acad. Sci. U.S.A.* 103, 17402–17407.

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1–13.

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57.

Keller, M. P., Choi, Y., Wang, P., Davis, D. B., Rabaglia, M. E., Oler, A. T., et al. (2008). A gene expression network model of type 2 diabetes links cell cycle regulation in islets with

diabetes susceptibility. *Genome Res.* 18, 706–716.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559. doi: 10.1186/1471-2105-9-559

Langfelder, P., Luo, R., Oldham, M. C., and Horvath, S. (2011). Is my network module preserved and reproducible? *PLoS Comput. Biol.* 7:e1001057. doi: 10.1371/journal.pcbi.1001057

Langfelder, P., Zhang, B., and Horvath, S. (2008). Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* 24, 719–720.

Margolin, A. A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R., et al. (2006). ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* 7(Suppl. 1):S7. doi: 10.1186/1471-2105-7-S1-S7

Meyer, P. E., Kontos, K., Lafitte, F., and Bontempi, G. (2007). Information-theoretic inference of large transcriptional regulatory networks. *EURASIP J. Bioinform. Syst. Biol.* 2007:79879. doi: 10.1155/2007/79879

Orozco, L. D., Bennett, B. J., Farber, C. R., Ghazalpour, A., Pan, C., Che, N., et al. (2012). Unraveling inflammatory responses using systems genetics and gene-environment interactions in macrophages. *Cell* 151, 658–670.

Park, C. C., Gale, G. D., de Jong, S., Ghazalpour, A., Bennett, B. J., Farber, C. R., et al. (2011). Gene networks associated with conditional fear in mice identified using a systems genetics approach. *BMC Syst. Biol.* 5:43. doi: 10.1186/1752-0509-5-43

Parkkinen, J. A., and Kaski, S. (2010). Searching for functional gene modules with interaction component models. *BMC Syst. Biol.* 4:4. doi: 10.1186/1752-0509-4-4

Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., et al. (2011). Detecting novel associations in large data sets. *Science* 334, 1518–1524.

Schäfer, J., and Strimmer, K. (2005). An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics* 21, 754–764.

Sharma, U. C., Pokharel, S., Evelo, C. T. A., and Maessen, J. G. (2005). A systematic review of large scale and heterogeneous gene array data in heart failure. *J. Mol. Cell. Cardiol.* 38, 425–432.

Shiga, M., Takigawa, I., and Mamitsuka, H. (2007). Annotating gene function by combining expression data with a modular gene network. *Bioinformatics* 23, i468– i478.

Steffen, M., Petti, A., Aach, J., D'haeseleer, P., and Church, G. (2002). Automated modelling of signal transduction networks. *BMC Bioinformatics* 3:34. doi: 10.1186/1471-2105-3-34

Stuart, J. M., Segal, E., Koller, D., and Kim, S. K. (2003). A gene-coexpression network for global discovery of conserved genetic modules. *Science* 302, 249–255.

Weng, L., Macciardi, F., Subramanian, A., Guffanti, G., Potkin, S. G., Yu, Z., et al. (2011). SNP-based pathway enrichment analysis for genome-wide association studies. *BMC Bioinformatics* 12:99. doi: 10.1186/1471-2105-12-99

Xiao, Y., Xu, C., Guan, J., Ping, Y., Fan, H., Li, Y., et al. (2012). Discovering dysfunction of multiple microRNAs cooperation in disease by a conserved microRNA co-expression network. *PLoS ONE* 7:e32201. doi: 10.1371/journal.pone.0032201

Yang, M.-S. (1993). A survey of fuzzy clustering. *Math. Comput. Model.* 18, 1–16.

Yang, M.-S., Hwang, P.-Y., and Chen, D.-H. (2004). Fuzzy clustering algorithms for mixed feature variables. *Fuzzy Sets Syst.* 141, 301–317.

Yee, K., Dombek, K. M., Lo, K., Mittler, J. E., Zhu, J., Schadt, E. E., et al. (2011). Construction of regulatory networks using expression time-series data of a genotyped population. *Proc. Natl. Acad. Sci. U.S.A.* 118, 19436–19441.

Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* 4:Article17. doi: 10.2202/1544–6115.1128

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 01 December 2012; accepted: 21 February 2013; published online: 12 March 2013.

Citation: Rau CD, Wisniewski N, Orozco LD, Bennett B, Weiss J and Lusis AJ (2013) Maximal information component analysis: a novel non-linear network analysis method. Front. Genet. 4:28. doi: 10.3389/fgene.2013.00028

This article was submitted to Frontiers in Statistical Genetics and Methodology, a specialty of Frontiers in Genetics.

Chapter 5

eQTL Hotspot Analysis and Gene Networks Reveal Key Drivers of Catecholamine-Induced Heart Failure

Introduction

This chapter contains a manuscript describing the application of MICA from chapter 4 to the data generated from the HMDP HF study. In it, we identify a module of ~40 genes which is highly correlated with a number of cardiac phenotypes. Using additional network-based algorithms, we identify potential regulators of this module, including a gene *Magi2*, that is not expressed in the heart, but appears to play a role in the control of the heart's response to ISO. We also perform a transcriptome-wide analysis of genes expressed in the heart, and identify several regions of the genome that strongly regulate overall gene expression in the panel. We use SNPs and expression data to identify possible candidate genes for the master controllers of gene expression in the heart under background conditions and during ISO-induced heart failure.

A systems genetics approach to identify genetic pathways and key drivers of isoproterenol-induced cardiac hypertrophy and cardiomyopathy in mice.

Authors: Christoph D. Rau[1], Jessica Wang[2], Shuxun Ren[3], Yibin Wang[3], Aldons J. Lusis[1,2]

[1]Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, CA 90095, USA

[2]Department of Human Genetics, David Geffen School of Medicine, David Geffen School of Medicine, University of California, Los Angeles, CA 90095, USA

[3]Departments of Anesthesiology, Physiology and Medicine, Cardiovascular Research Laboratories, David Geffen School of Medicine, University of California, Los Angeles, CA 90095

Abstract

We recently reported a Genome Wide Association Study (GWAS) on heart failure (HF) and cardiac fibrosis. We used the β-adrenergic agonist isoproterenol (ISO) to induce HF-associated phenotypes in the Hybrid Mouse Diversity Panel (HMDP), a novel mouse resource population. Here we report the results of analyses performed on transcriptomes taken from 92 matched strains of the HMDP both with and without ISO treatment. We performed eQTL hotspot analysis to identify master regulators of gene expression in treated and untreated hearts, and identified three loci which significantly regulate over 10% of all expressed genes in the heart. One of these loci contains the miRNA-processing gene *Drosha*, and we identify a nonsynonymous SNP in *Drosha* which is significantly correlated with heart weight. Additionally, we utilized a recently described gene network construction algorithm, Maximal Information Component Analysis (MICA), to identify genes and pathways involved in the response to chronic ISO exposure. We observed six modules which show significant correlation to HF-related phenotypes, including a module of 41 genes which contained several genes of interest, including *Lgals3*, a diagnostic marker for HF. Utilizing eQTL hotspot analysis, a locus containing *Magi2*, a gene which helps to control the turnover of the β-adrenergic receptor in the brain, but is not expressed in the heart, was shown to be involved in the regulation of this module.

Introduction

Heart failure (HF) is the leading cause of death in developed countries[1,2]. In contrast to many other disorders, genome-wide association studies (GWAS) to identify novel genes and pathways which regulate HF have had only modest success in human studies[3–5]. The complex nature of the disease suggests that a more successful approach would be to examine the failing heart using systems-level techniques in a controllable model that lacks the differing etiologies and environmental confounders that frustrate human studies.

Animal models allow researchers to monitor and control the environmental factors that contribute to the disease, and provide a possible means for genomic analysis of HF. We have recently published an article describing a successful GWAS of HF and cardiac fibrosis phenotypes in a novel mouse resource, the Hybrid Mouse Diversity Panel (HMDP) exposed to chronic β-adrenergic stimulation. We demonstrated that due to the improved resolution compared to traditional linkage studies[6,7], the ability to perform case/control studies and the reduction of environmental variation , that it was possible to identify a number of loci which contribute to HF.

Previous research using the HMDP has suggested that our resource is well suited to other systems genetics approaches, such as eQTL hotspot analysis and transcriptome network analysis[7–11] for the identification of genes which contribute to phenotypes of interest. We now report an analysis of transcriptomes taken from treated and untreated mice of the HMDP HF study using both *trans*-eQTL hotspot analysis and co-expression network analysis to identify genes and pathways which contribute to HF.

Although rarely an initial cause of HF, β-adrenergic stimulation is commonly seen in HF patients and is considered a driver of ongoing hypertrophy and eventual cardiac failure[12]. Isoproterenol, a synthetic β-adrenergic agonist and a well-studied model in mice[13–16], provides a means to induce HF-related phenotypes without relying on a more error-prone initial insult to the heart. *Trans-eQTL* hotspot analysis reveals seven significant hotspots where a single locus significantly regulates over 5% of the genes expressed in the heart, three of which control the expression of 10% of the expressed genes within the heart. Within these loci we have identified as candidates genes with known roles in cardiac growth and the response to adrenergic stimulation, including the *Drosha* miRNA-processing gene and *Akap5*. We also identified the *Serpina3n* gene, whose gene expression is very highly correlated to changes in heart weight and fibrosis in our panel, as being a strong candidate within one of our *trans*-eQTL loci.

Additionally, we utilized a recently described co-expression network analysis method, Maximal Information Component Analysis (MICA) [17], to form modules of functionally-related genes from our transcriptome data. MICA has several advantages over other co-expression analysis methods, including the ability to analyze both linear and non-linear relationships between genes and to allow genes to proportionally exist within multiple modules at a time. We have demonstrated[17] that MICA is particularly well suited to studying case/control studies such as the HF HMDP. MICA identifies six modules which show significant correlations to HF-related phenotypes. One module of 41 genes was examined in detail, and contained a number of genes previously implicated in HF such as *Lgals3*, a plasma biomarker of HF[18] and *Timp1*, an extracellular matrix

regulator[19,20]. We also observe a number of genes previously unimplicated in HF in the module, including the eQTL hotspot gene *Serpina3n*. An eQTL hotspot analysis of this module specifically identifies a single locus on chromosome 5 which significantly regulates a quarter of the genes in the module. This locus contains a single gene, *Magi2,* which regulates the turnover of the β-adrenergic receptors in the brain[21].

Results

eQTL hotspot analysis identifies key regulators of gene expression

Left ventricular tissue from untreated and treated (3 weeks of ISO at 20 mg/kg body weight/day) mice were dissected from 92 matched strains of the HMDP and processed using Illumina Mouse Ref 8 2.0 gene expression arrays. Genes were first filtered for those significantly expressed in the heart in at least 25% of either control or treated strains. In addition to the control and treated data sets (~13,000 genes), an additional set was calculated for the difference in gene expression in all genes whose expression varied between ISO and control animals by a coefficient of variation of at least 5% (~8000 genes). Using the EMMA algorithm [22] and ~130,000 SNPs, we identified loci, which we term expression quantitative trait loci (eQTL), which regulate gene expression at an FDR of 1% (P=6.1E-5).

eQTL hotspot analysis tests for loci which significantly regulate a significant number of genes in the tissue or condition of interest. Prior research[23] has shown that these hotspots are capable of identifying important regulators of tissue development,

cellular processes and/or the response to environmental stressors. To facilitate the identification of hotspots, we partitioned the genome into 4,691 500kb windows. For each gene, we selected the SNP with the lowest P-value to represent the entire window for further analysis. Significant eQTL hotspots were determined by summing the number of genes per window which exceeded the FDR cutoff and comparing that number to a bootstrapped distribution of enrichments in order to determine a final significance threshold (3.85E-7 for untreated and treated mice, 6.23E-7 for delta, Figure 5.6). We observe 445 significant windows for untreated strains, 177 for treated strains and 172 for the difference in gene expression between ISO and control mice (Figure 5.1, Table 5.3). A full list of these windows is presented in- Table S1. We will focus only on 'master regulator' loci which regulate over 5% of the genes expressed in the heart (Table 1). We observe three loci which regulate at least 5% of baseline gene expression, and four loci which regulate at least 5% of the change in gene expression between control and treated mice. We do not, however, observe any loci for treated animals which affect at least 5% of expressed genes (maximum: 2.2%). We hypothesize that the absence of master regulators in treated animals (Figure 5.1B, Table 5.1) results from the systems-wide effects of ISO stimulation, which results in few loci which regulate large numbers of genes in the heart specifically.

To help identify potential master regulators at these loci, we correlated gene expression with the heart weights of the HMDP HF study which are available along with a number of other phenotypes on our public database (http://systems.genetics.ucla.edu/data). Additionally, we used the Wellcome Trust

Mouse Genomes Project sequencing database[24], which has the full genomic sequence of ten strains contained in our panel, to examine genetic variations present within each locus. Together, these approaches provide a powerful and systematic method for the identification of causal genes within these eQTL hotspot loci.

Our most significant locus for untreated gene expression spans 6 Mb (12-18Mb) on chromosome 15, and regulates 11.3% of all expressed transcripts (Figure 5.2A). *Drosha*, the initiator of miRNA processing, lies near the center of the locus, and is the only gene within the locus to be expressed at significant levels in the heart, although its expression does not correlate with heart weight. We identified a nonsynonymous mutation within exon 5 of *Drosha* in the C57BL/6J strain which is significantly correlated to untreated heart weights (P=0.034, Figure 5.2B). miRNA signaling plays a vital role in cardiac development and disease[25–27]. We propose that this nonsynonymous variation results in subtle changes to the way in which miRNAs are processed in the heart, leading to observable differences in heart development and gene expression.

Our most significant locus is associated with the change in gene expression between treated and untreated strains and spans 1 Mb (77.5-78.5) on chromosome 12. It regulates 11.4% of all expressed and varying transcripts (Figure 5.2C). *Akap5*, an anchoring protein which localizes protein kinase A to particular regions of the cell as well as binding to calcineurin, a known regulator of the β-adrenergic signaling pathway[28–30], is a strong candidate within this locus. We observe a strong correlation between heart

weight and *Akap5* gene expression(R=0.35, P=6E-7, Figure 5.2D), which further supports a role for *Akap5* in the regulation of the cellular response to ISO stimulation.

Another locus of interest regulating the change in gene expression of 9.7% of all transcripts expressed in the heart was also located on chromosome 12(Figure 5.2E). Although this locus contains *Dicer1,* which, along with *Drosha,* regulates miRNA processing[25,27], we see no evidence that it is the causal gene at this locus, as its gene expression does not vary significantly across the panel, nor does it have any nonsynonymous SNPs in the founder strains of the HMDP. Instead, we propose the serpin peptidase inhibitor *Serpina3n* as a likely candidate gene at this locus. *Serpina3n* has one of the highest correlations to heart weights of any gene in the genome (R=0.57, P=3E-17, Figure 5.2F). The function of *Serpina3n* beyond the most general level of 'protein inhibition' is currently unknown, although its family of *Serpina3*s has been linked in several studies to aneurisms[31,32] and has been shown to be disregulated in dilated cardiomyopathy[33].

Gene Network Analysis using Maximal Information Component Analysis

Previous research using the HMDP has shown the benefit of using systems-level transcriptomics to generate co-expression networks to better understand the genes and pathways underlying phenotypes of interest[6,8,10,34]. Here we apply MICA[17], a unbiased gene network construction method that captures both linear and nonlinear interactions within the data, and allows genes to be spread proportionally across multiple modules, to generate a network analysis of the HMDP HF transcriptomes.

Prior research[35] has suggested that inclusion of non-expressed transcripts or invariant genes across strain/treatment differences in network construction algorithms can potentially lead to inaccurate results due to the addition of significant noise to the analysis, as well as dramatically increasing the time it takes to run the algorithm. We chose to use the set of ~8000 genes previously identified as being expressed in the heart and showing a high coefficient of variation as the basis of our network.

We observe that out of 8,126 genes, 1,392 are not placed into modules because they did not display sufficiently strong relationships to other genes to be included in the module detection step. We used the Database for Annotation, Visualization and Integrated Discovery (DAVID)[36,37] to query whether these unincluded genes were enriched for particular Gene Ontology (GO) categories when compared to the dataset as a whole, however no GO term was enriched for these genes (most significant enrichment p-value: alternative splicing, P=0.12). This result for our excluded genes matches those observed in other module construction algorithms [38] which leave out a subset of poorly connected genes, and suggests that these genes were properly excluded.

MICA allows genes to reside in multiple modules, with each gene theoretically influencing every module, although most genes will exist primarily in one or two modules[17]. The first principle component of a module, sometimes called an eigengene, has been used in the past to combine groups of genes into a single 'pseudogene' which can be correlated to phenotypes of interest.[17,38,39] Weighted principle components, taking into account the proportional contribution of each gene to each module, were calculated

and correlated to HF-related phenotypes gathered during the HMDP HF project (Figure 5.3). Six modules (3,5,6,8,11,19) were observed to contain at least one significant (P <1.7E-4) correlation to a HF-related trait. In particular, module 5 showed very strong correlation to a number of traits, including total heart weight (R=.66, P=4e-24), left ventricular internal dimension at diastole (LVID;d, R=0.55, P=1e-16) and cardiac fibrosis (R=.45, P=1e-10).

A goal of network analysis is the identification of a core set of genes (hubs) within one or multiple pathways which drives the module as a whole. DAVID was used to search for enriched pathways within each module. As DAVID is unable to process the proportional gene memberships returned by MICA, modules were thresholded based on the weight of each genes' membership in that module (%MM).

Choosing the best %MM cutoff involved balancing the desire for small modules (which occur at high %MM cutoffs) and preservation of the eigengenes returned by MICA (which are highest at low %MM cutoffs). We selected several %MM thresholds and selected the highest %MM cutoff for which the correlation between the eigengenes at that cutoff and the weighted eigengenes were over 90% (Table 5.4). This cutoff was 50%.

Module 5 is highly correlated to HF traits and contains numerous genes relevant to HF

At a 50%MM cutoff, module 5 (red in Figure 5.4) contains 41 genes (Table 5.2). Module 5 is enriched for glycoproteins(P=2.1E-9), signaling molecules(P=2.6E-10), and

ECM (P=2.9E-6) GO terms (Table 5.5). Strikingly, out of the 41 genes in the module, fourteen (34.1%) have previously been described as involved in heart hypertrophy or cardiac remodeling based on transgenic studies in animal models or mendelian forms of HF (Table 5.2). These genes include *Lgals3*, a recently-described diagnostic marker of heart failure[18], *Timp1*, an extracellular matrix regulator[19,20], *Pde4e*, knockout of which induces a progressive cardiomyopathy[40,41] as well as several collagens. Serendipitously, we observe that *Serpina3n*, the peptidase inhibitor previously identified as a candidate gene in a *trans*-eQTL hotspot is also a member of module 5.

Using genetic markers as anchors to determine reactive versus causal relationships within module 5

Most co-expression network analysis methods, including MICA, return a non-directed network, where links between genes are reported, but cause-effect relationships between genes within a network are unknown. Additionally, it is unknown whether a module which is highly correlated with a clinical trait of interest is driving the phenotype, or acting in response to it. To determine these relationships, the NEO algorithm[42], which uses SNPs as anchors to infer directionality between two measured traits was applied to module 5. 28 genes (68.3%) had significant directionality as determined by NEO (Figure 5.5). *Adamts2*, a metalloproteinase, has the largest number of directed links, with 11 causal and one reactive relationship with *Pdgfrl*, a growth-factor receptor which does not react to any other gene in module 5. The most reactive gene is *Lgals3*, which is affected by 10 other genes but affects no other genes in turn. NEO was also used to determine

relationship between genes in the module and phenotypes (Table S4). This analysis showed that some genes within the module appeared to drive the phenotypes, such as *Dct* and *2610028H24Rik,* however others, notably *Col14A1*, a collagen family member was downstream from the HF traits, suggesting that its expression was driven by changes in cardiac mass, and not the other way around.

The *trans* eQTL hotspot of module 5 reveals a potential regulator of HF

To further elucidate the means by which module 5 genes are influencing HF phenotypes, we used our transcriptome data, tied with the genotypes of the HMDP, to perform eQTL hotspot analysis for the 41 genes within module 5 (Figure 5.7). We observe that 13 of the 42 genes in module 5(30.9%) are regulated by a single locus on chromosome 5. This region contains only a single gene: membrane-associated guanylate kinase inverted 2 (*Magi2).* *Magi2* is primarily expressed in the hormone excreting glands of the brain[43], where it has been demonstrated to interact with beta-1 adrenergic receptors and regulate their endocytosis after stimulation by adrenergic agonists such as ISO [21]. Expression in the heart is low or absent. Knockout of *Magi2* results in neonatal lethality within 24 hours of birth with abnormal spine morphology and improper brain development[44]. No reports of variations in this gene have previously been described as influencing heart-related phenotypes.

Discussion

Heart failure is the primary cause of hospitalization in people over the age of 65 in developed countries[45]. Complicating the understanding of HF from a genetic level are decades of unobserved cardiac stressors such as hypertension, atherosclerosis and diabetes that can contribute to the disease [1] and which differ between each patient. As a result, efforts to identify genes contributing to non-mendelian forms of HF using systems-level approaches have only been modestly successful[4,46,47]. We attempted to identify genes and pathways which are implicated in HF using the HDMP, a large inbred mouse panel[6], using both *trans*-eQTL hotspot analysis and gene module detection using MICA[17] through the chronic stimulation of the β-adrenergic pathway, which is known to be upregulated in most heart failure patients[48–50]. We observe several loci through eQTL hotspot analysis which appear to regulate large fractions of the expressed genes in the heart, and have identified six modules of genes which are highly correlated with HF phenotypes. We further analyzed one of these modules and identified the underlying causal relationship network in the module as well as *Magi2*, a master regulator for the module as a whole that is primarily expressed in the brain. These points are discussed in turn below.

eQTL hotspots reveal loci which act as 'master regulators' of gene expression in a particular tissue[23]. In applying this method of analysis to our data, we identified seven loci, each of which significantly regulated more than 5% of the expressed genes in the heart. One of these loci contains *Drosha*, a major regulator of miRNA processing. *Drosha* contains a nonsynonymous SNP which is significantly associated (P=0.035) with heart weight. miRNAs play an important role in the heart[26,51,52] and it follows that a

mutation to one of the miRNA processing enzymes would lead to dramatic variations in gene expression in the heart. We also identified a peptidase inhibitor, *Serpina3n*, which is very highly correlated to HF-related traits and has previously been implicated in cardiac remodeling in CVB3 myocarditis[53].

We also examined the transcriptomes of both treated and untreated mice of the HMDP using a recently-developed network analysis tool MICA[17]. We observed strong correlations between the eigengenes of our identified modules and relevant clinical traits observed in our mice. The module most strongly associated with changes in hypertrophy and heart function is Module 5, which contains 41 genes. We observed that many of the genes within the module (~33%) had previously been described as being implicated in hypertrophy, cardiac remodeling or fibrosis. Interestingly, one such gene is *Serpina3n*, one of the candidate genes from the *trans*-eQTL hotspot analysis. Several additional genes in module 5 have been implicated in other aspects of cardiac functioning. For instance, *JAG2* has been reported as upregulated in the resident cardiac progenitor cell niche in adult mouse hearts[54] and *ACOT1* has been shown as specifically induced in heart and skeletal muscle by a high-fat diet[55]. The remaining genes in module 5 should be carefully examined with the goal of finding additional modifiers of cardiac remodeling.

Further analysis of module 5 revealed a possible mechanism for its regulation. NEO[42], a tool for providing directionality to undirected gene networks, was used to uncover the ways in which the genes in module 5 interacted with one another. This analysis showed that *Adamts2* plays a central role in the regulation of module 5, and is

regulated itself by *Pdgfrl*, a poorly described receptor molecule[56]. *Trans* eQTL hotspot analysis of module 5 revealed *Magi2* as a likely driver of module 5. *Magi2* is involved in the regulation of the response to adrenergic stimulation in the brain, specifically in hormone-excreting regions[21]. We hypothesize that these regions, which are exposed to ISO during our protocol, secrete a factor into the plasma which then interacts with *Pdgfrl*, activates *Adamts2* and drives the phenotypic effects of module 5. NEO analysis is not definitive, therefore it is possible that gene feedback loops and complex genetic relationships may complicate the analysis and reduce the validity of these conclusions[34,42]. Further research should be performed on *Magi2, Pdgfrl* and *Adamts2* to validate the results of NEO and determine potential regulatory functions of each of these genes.

One of the benefits of using an inbred mouse panel such as the HMDP is that the data gathered are cumulative, which allows for the integration of data collected in this study with studies performed in the future on additional mice or alternate treatment conditions. The power and usefulness of these data will only increase as future experiments are performed to ascertain whether the results observed from this study are conserved across different conditions. Future experiments, such as examining the networks created under the effects of an α-adrenergic receptor agonist like angiotensin or a physical challenge such as trans-aortic constriction will reveal whether all of these etiologies that lead to HF are perturbing the same genes/modules and identify which genes/pathways are unique to each model system. By combining these data with one

another, common genes may be identified that will lead to improved levels of care for patients with HF in the future.

Materials and Methods

Online database

All results and data can be accessed at http://systems.genetics.ucla.edu/data

Mice and isoproterenol treatment

The following mouse strains were obtained from The Jackson Laboratory and then bred in our colony: 29 common inbred strains (129X1/SvJ, A/J, AKR/J, BALB/cJ, BALB/cByJ, BTBR T+ tf/J, BuB/BnJ, C3H/HeJ, C57BL/6J, C57BLKS/J, C58/J, CBA/J, CE/J, DBA/2J, FVB/NJ, KK/HlJ, LG/J, LP/J, MA/MyJ, NOD/ShiLtJ, NON/ShiLtJ, NZB/BINJ, NZW/LacJ, PL/J, RIIIS/J, SEA/GnJ, SJL/J, SM/J, SWR/J) and 69 RI lines [RI (number of strains) - BXD (40), AXB(8), BXA(10), BXH(5), CxB(6)]. All animal experiments were conducted following guidelines established and approved by the University of California, Los Angeles Institutional Animal Care and Use Committee. All mice have been previously genotyped at over 130,000 locations. Isoproterenol (20 mg per kg body weight per day) was administered for 21 d in 8- to 10-week-old female mice using ALZET osmotic minipumps, which were surgically implanted intra-peritoneally.

Microarray and eQTL analysis

Following homogenization of left ventricular tissue samples in QIAzol, RNA was extracted using the Qiagen miRNAeasy extraction kit, and verified as having a RIN>7

by Agilent Bioanalyzer. Two RNA samples were pooled for each strain/experimental condition, whenever possible, and arrayed on Illumina Mouse Reference 8 version 2.0 chips. Analysis was conducted using the Neqc algorithm included in the limma R package [57] and batch effects addressed through the use of COMbat [58]. eQTLs were calculated EMMA[22], closely following the methods described in that paper. In brief, a kinship matrix between the strains were calculated to weight the pair-wise relationships between the strains. Values will increase when unrelated strains share a phenotypic effect at a SNP, while values will decrease if the opposite is true. Significance thresholds were calculated as in Parks et al., 2013 using an FDR of 1%.

eQTL hotspot analysis

The genome was binned into windows of 500kb. For each gene, the lowest P-value within a window was used to represent the total number of SNPs within that window as a whole. eQTL hotspots were determined by summing up the number of genes within each window whose lowest p-value in that window was less than the 6.1E-5 significance threshold. Window significance was determined by randomly sampling 5,000,000 windows, and determining the enrichment of each window, while the significance threshold was determined by correcting for the total number of SNPs.

A similar process was used to determine the hotspot of module 5; however, in this case only the 41 genes contained in module 5 were used.

MICA

Maximal Information Component Analysis was performed on the combined set of untreated and treated arrays on 8126 genes which were both expressed in at least 25% of either the untreated or treated data, and whose coefficient of variation was greater than 5%. Analysis was performed as described in Rau et al.[17]. Briefly, transcriptomes were processed using the MINE algorithm[60] to determine relationships between each gene in the dataset. The resulting relationship matrix was trimmed using the hard thresholding function (threshold=0.45) of the WGCNA R package[38] and inputted into the ICMg algorithm[61]. The ICMg algorithm determined the proportional module membership for each gene. Weighted principle components ('Eigengenes') of each module were determined through the use of the dudi.pca function from the ade4 R package[62] and compared to HF-related phenotypes using the heatmap function of WGCNA[38]. For all control variables, the standard values were used.

NEO

Gene expression data for the genes in module 5 along with any eQTLs for those genes whose p-value exceeded a threshold of 1E-6 were selected as inputs for the Network Edge Orienting (NEO) software package[42]. Standard protocol for NEO was used, and we selected marker, gene and phenotype combinations that yielded a LEO, NB.AtoB or BtoA score >0.75 and mlogp.M.AtoB or BtoA score <.05 for further analysis.

Acknowledgments

This work was supported by NIH grant HL110667 and HL28481. CDR was supported by NIH training grant T32HL69766 and JW was supported by NIH training grant HL007895.

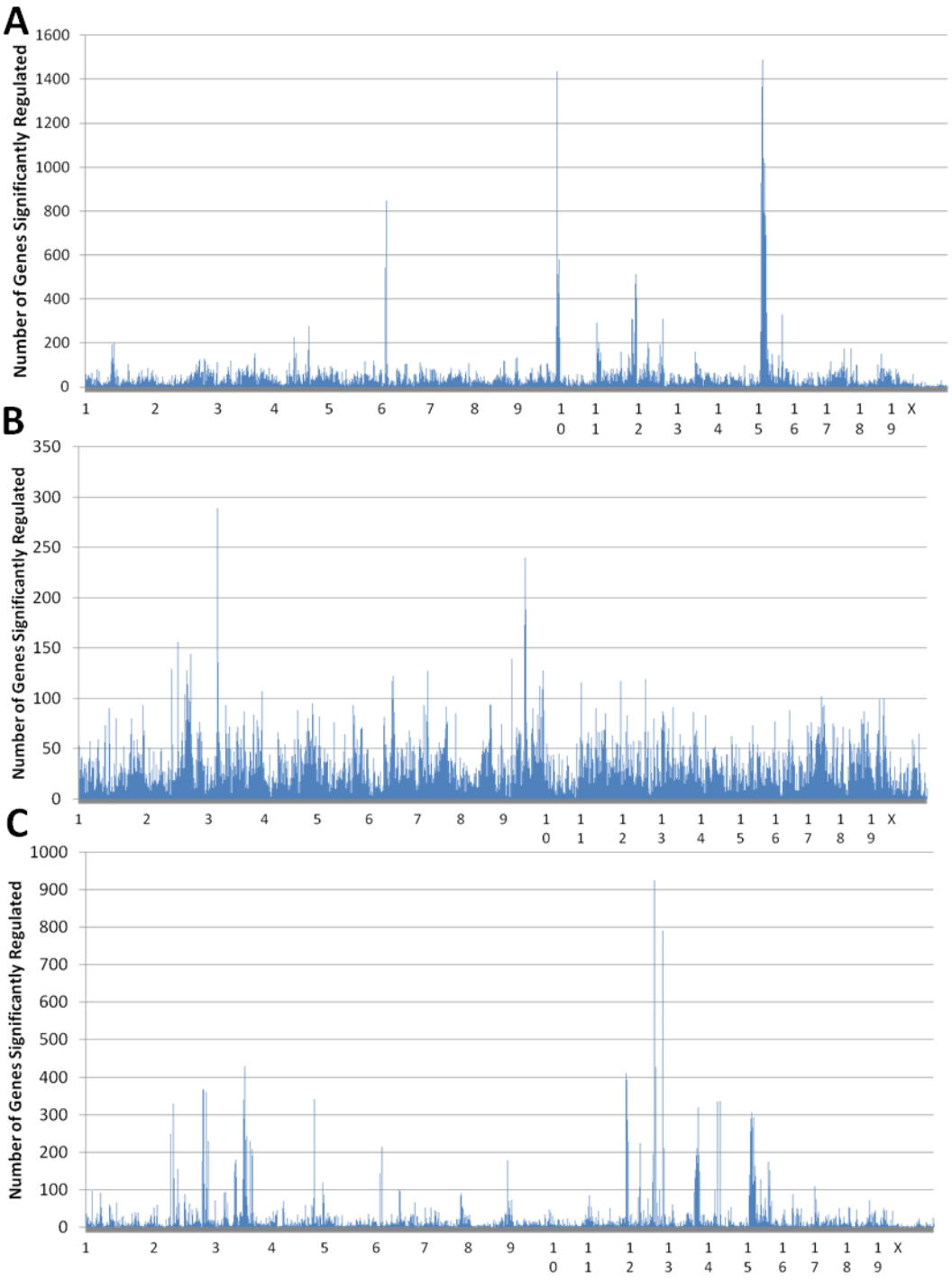

Figure 5.1. eQTL hotspots show regions of strong regulation. A) Control B) Treated C) Ratio of control to treated.

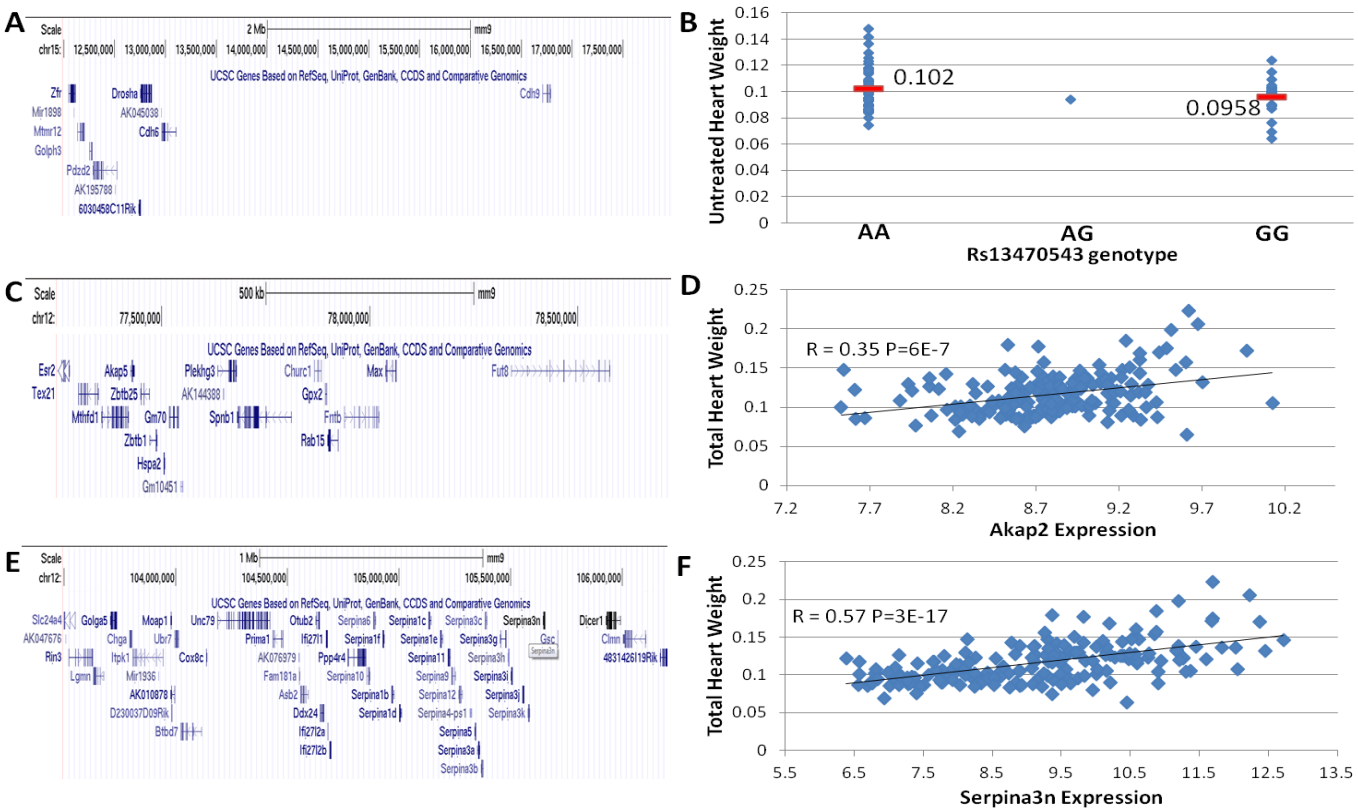

Figure 5.2 **<u>eQTL hotspots are correlated to cardiac` phenotypes of interest</u>** A) A locus on chromosome 15 regulates 11.3% of all transcripts and contains the gene *Drosha.* B) A nonsynonymous SNP within *Drosha* is significantly associated with control heart weight. C) A locus on chromosome 12 regulates 11.4% of all transcripts and contains the gene *Akap5* D) *Akap5* expression correlates strongly with heart weight. E) Another locus on chromosome 12 regulates 9.7% of all transcripts and contains the gene *Serpina3n*. F) *Serpina3n* expression correlates very strongly with heart weight

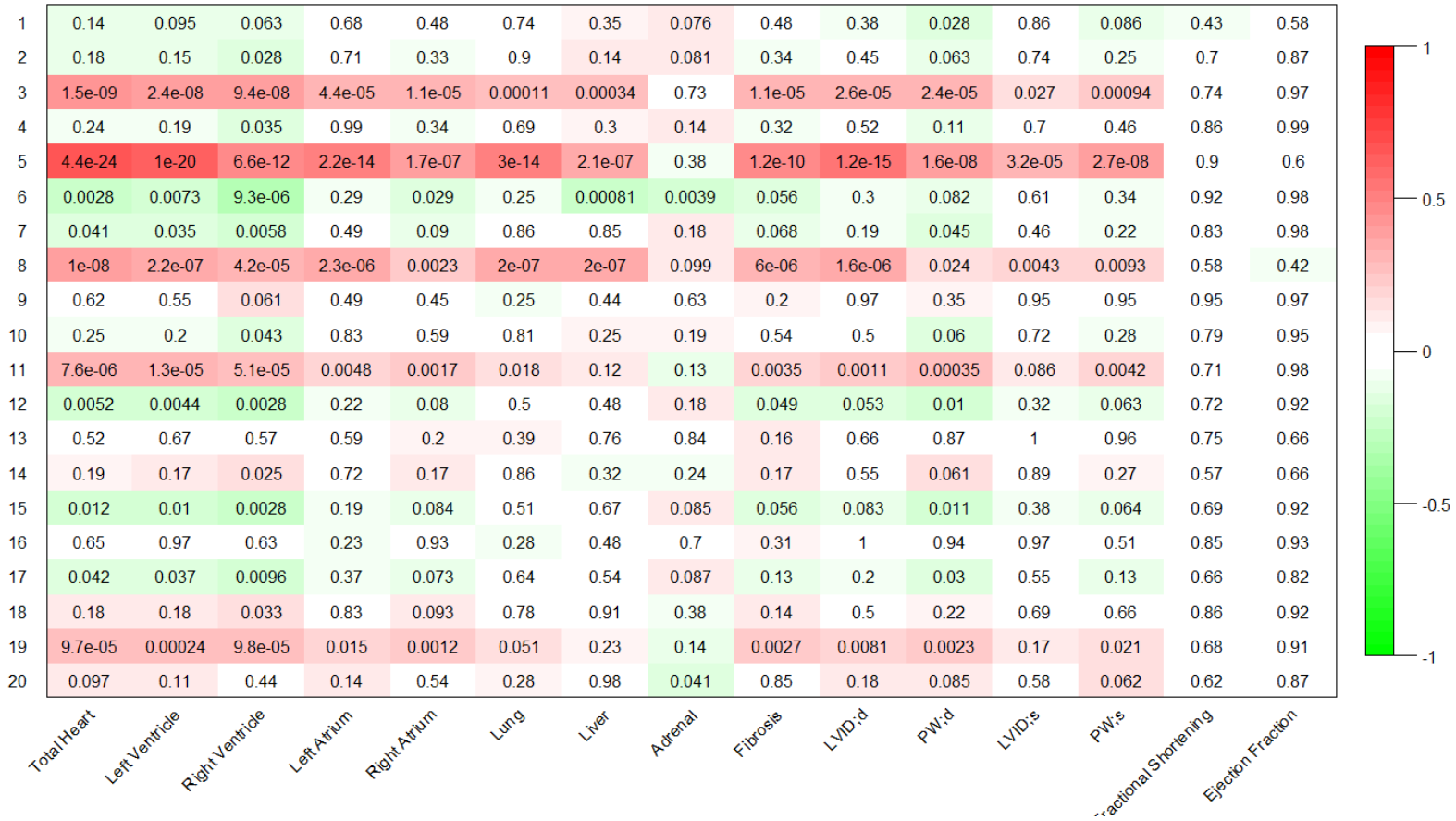

Figure 5.3 **Six Modules correlate significantly with at least one HF-related phenotype.** Module-trait correlations are displayed where color indicates the strength of the correlation (red = positive, green = negative) and the number within the box is the P-value for statistical significance of the relationship. Left Ventricular Internal Dimension (LVID), Peritoneal Wall (PW), at Diastole (;d), at Systole (;s)

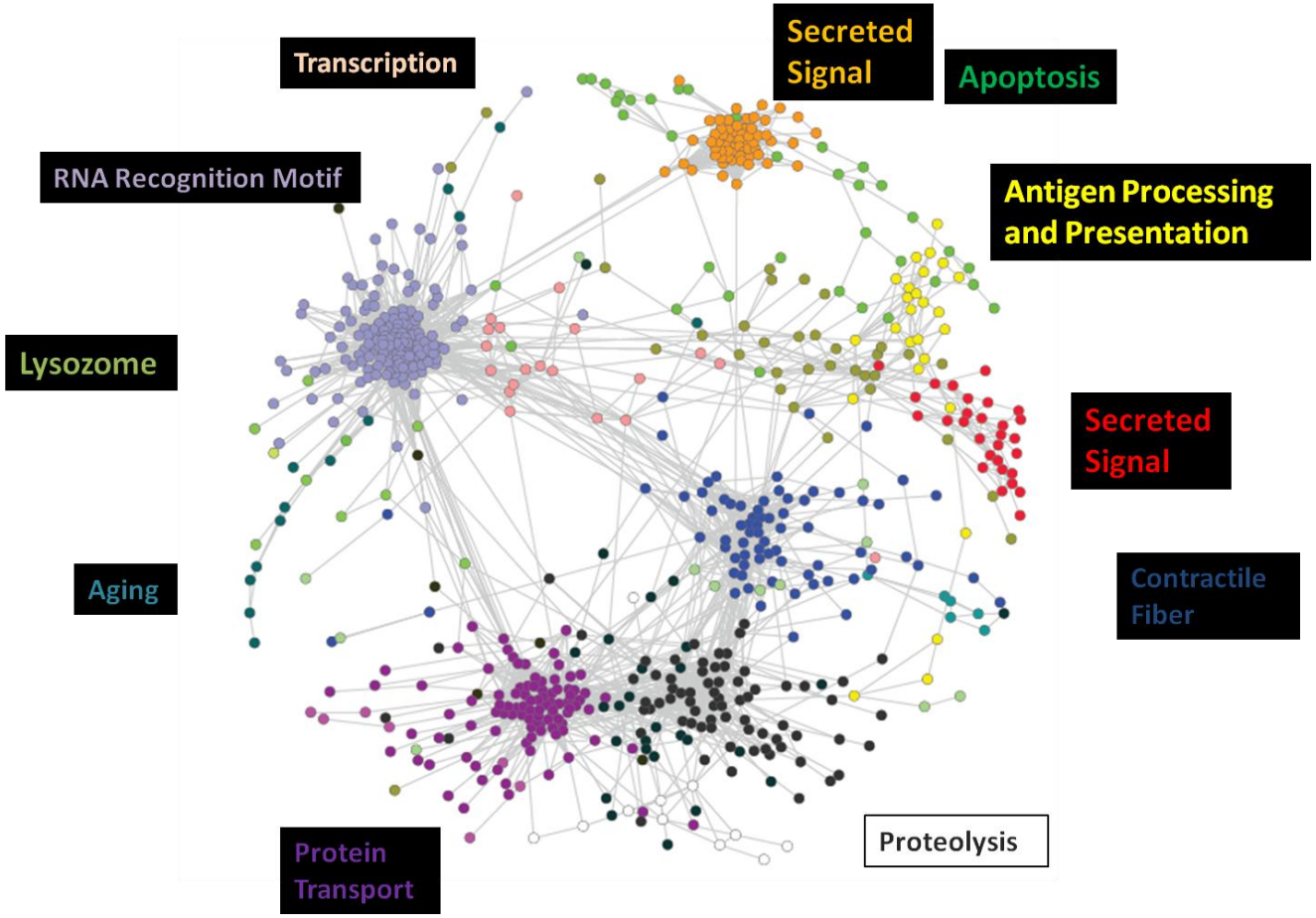

Figure 5.4 **Gene modules are enriched for a number of GO terms**. A network plot showing genes as nodes, relationships between those genes as edges, and using colors to represent modules. GO terms for each module are displayed around the image in the color of its corresponding module. Modules which lack significant GO enrichments are removed for visibility.

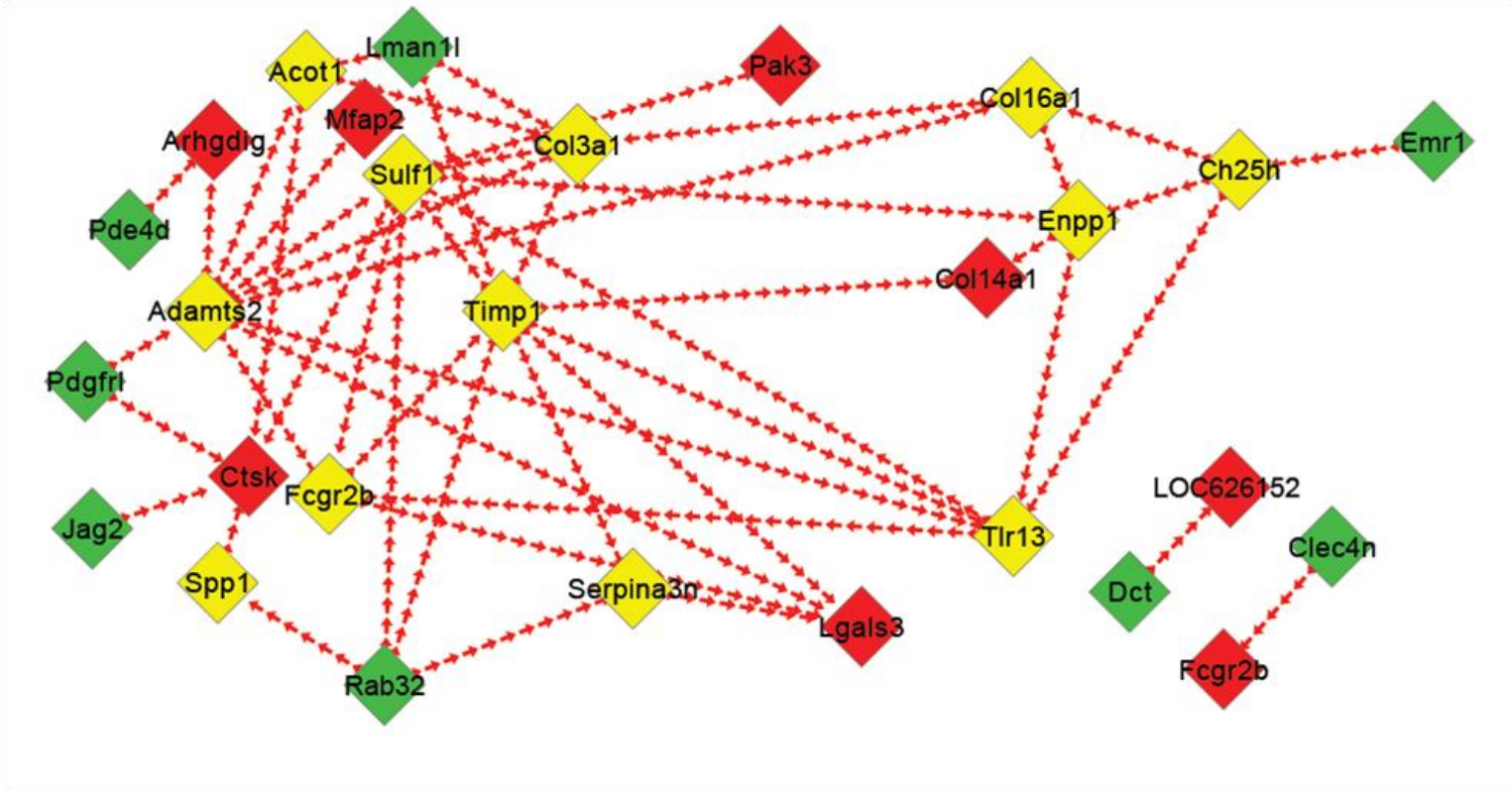

Figure 5.5 **Near Edge Orientation reveals underlying directionality in module 5.** The NEO algorithm was used to determine the directionality of each edge in module 5. Undirected edges and genes which did not have any directed edges are removed for visibility. Genes are color coded as either drivers (green, having no other genes within module 5 as inputs), receivers (red, having no outputs to other genes in module 5) or intermediates (yellow, having both inputs and outputs within module 5.

Table 5.1. eQTL hotspots which regulate at least 5% of the expressed cells in the heart. **Bold** genes are significantly correlated with heart weights, while underlined genes contain non-synonymous mutations.

Chromosome	Start of Window	End of Window	eQTLs Regulated at Locus	Percent of Expressed Genes Regulated	Candidate Gene(s) at Locus
Control Mice					
6	17500000	18500000	846	6.431015	**Cav2**
9	112000000	121000000	1437	10.9236	**Tgfbr2** **Cmtm7**
15	12000000	18000000	1490	11.32649	Drosha
Treated Mice					
No Hotspots Found					
Ratio of Treated to Untreated Expression					
3	79500000	88500000	430	5.291656	**Tgfb3**
11	110000000	117000000	410	5.045533	**Sox9** **Slc39a11**
12	77500000	78500000	925	11.38321	**Akap5**
12	104000000	106000000	790	9.72188	**Serpina3n**

Table 5.2. Fourteen genes in Module 5 are associated with Cardiovascular Disease and play a role in key GO processes

Implicated in Cardiovascular Disease			Not Yet Implicated in Cardiovascular Disease				
Adamts2	GpnmB	Spp1	Acot1	Ctsk	Lman1l	Rfx2	2610028H24Rik
Col3a1	Lgals3	Serpina3n	Arhgdig	Dct	LOC100048556	Rpp25	
Col14a1	Timp1		Ch25h	Dmkn	LOC626152	Srpx2	
Col16a1	Pak3		Clec4d	Enpp1	Mfap2	Sulf1	
Emr1	Pde4d		Clec4n	Fcgr2b	Pdgfrl	Tlr13	
Enpp1	Slc1a2		Comp	Jag2	Rab32	Tmem82	
Signaling		Glycoproteins	Extracellular Matrix			Other/Unknown	

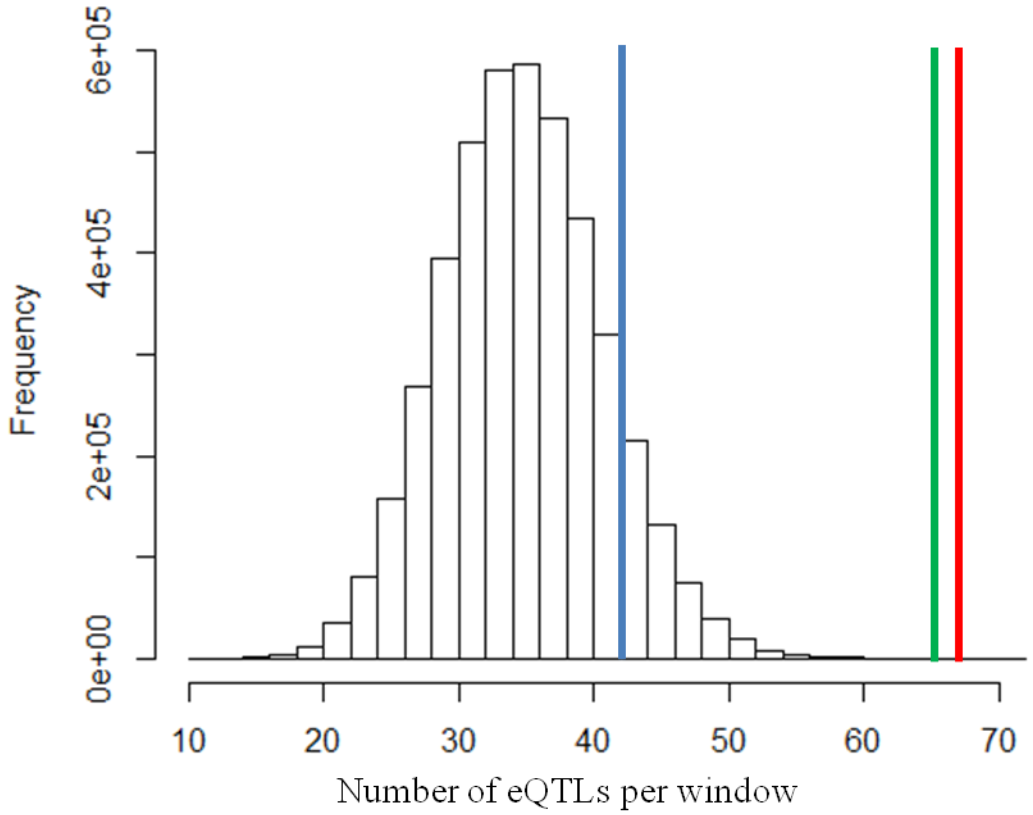

Figure 5.6. Thresholds for eQTL hotspot significance P-values were drawn at random to form 5,000,000 windows and the number of significant eQTLs in each window was calculated. The Blue line indicates the number of eQTLs within a window necessary for nominal (P=0.05) significance, red indicates the number necessary to pass a Bonforroni-corrected p-value for either control or treated eQTL windows, and green indicates the number necessary for the windows for the ratio of treated to control expression.

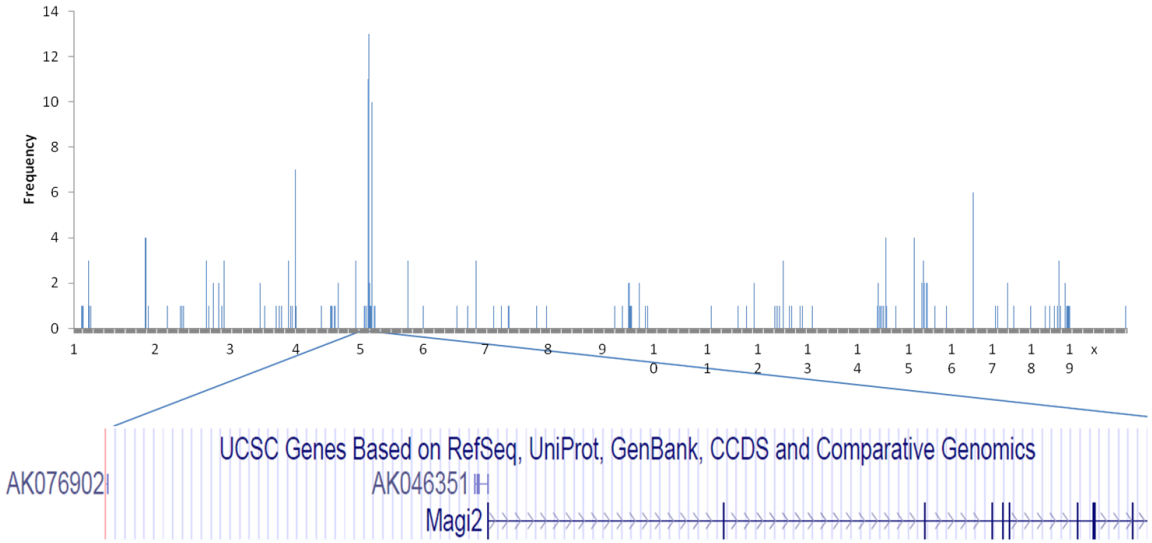

Figure 5.7. eQTL hotspot analysis of genes in module 5 reveals a locus on chromosome 5 which contains *Magi2*

Table 5.3. All Significant eQTL Hotspots

Chromosome	Start of Window	End of Window	# of Significant eQTLs in Locus	P-value
		Control		
15	1.50E+07	15500000	1490	<1E-7
9	114500000	1.15E+08	1437	<1E-7
15	13500000	1.40E+07	1367	<1E-7
15	15500000	1.60E+07	1192	<1E-7
15	1.40E+07	14500000	1088	<1E-7
15	16500000	1.70E+07	1041	<1E-7
15	1.30E+07	13500000	1009	<1E-7
6	17500000	1.80E+07	846	<1E-7
15	2.30E+07	23500000	780	<1E-7
15	1.60E+07	16500000	708	<1E-7
15	1.80E+07	18500000	703	<1E-7
15	2.50E+07	25500000	692	<1E-7
9	1.19E+08	119500000	580	<1E-7
6	1.30E+07	13500000	543	<1E-7
15	12500000	1.30E+07	542	<1E-7
9	1.15E+08	115500000	514	<1E-7
11	1.14E+08	114500000	514	<1E-7
11	113500000	1.14E+08	487	<1E-7
9	1.16E+08	116500000	475	<1E-7
15	2.40E+07	24500000	473	<1E-7
15	25500000	2.60E+07	470	<1E-7
11	112500000	1.13E+08	469	<1E-7
9	1.14E+08	114500000	468	<1E-7
15	1.70E+07	17500000	445	<1E-7
15	17500000	1.80E+07	440	<1E-7
15	23500000	2.40E+07	440	<1E-7
9	115500000	1.16E+08	436	<1E-7
9	1.17E+08	117500000	428	<1E-7
11	1.12E+08	112500000	421	<1E-7
11	115500000	1.16E+08	405	<1E-7
11	1.13E+08	113500000	388	<1E-7
11	111500000	1.12E+08	341	<1E-7
15	26500000	2.70E+07	339	<1E-7
9	117500000	1.18E+08	330	<1E-7

15	70500000	7.10E+07	328	<1E-7
11	103500000	1.04E+08	311	<1E-7
12	77500000	7.80E+07	310	<1E-7
11	104500000	1.05E+08	308	<1E-7
9	119500000	1.20E+08	305	<1E-7
11	6.00E+06	6500000	292	<1E-7
4	98500000	9.90E+07	277	<1E-7
9	1.12E+08	112500000	276	<1E-7
15	2.70E+07	27500000	269	<1E-7
9	118500000	1.19E+08	255	<1E-7
11	1.04E+08	104500000	254	<1E-7
15	1.00E+07	10500000	252	<1E-7
11	1.11E+08	111500000	248	<1E-7
4	97500000	9.80E+07	230	<1E-7
9	1.20E+08	120500000	227	<1E-7
4	5.60E+07	56500000	226	<1E-7
9	120500000	1.21E+08	225	<1E-7
15	1.20E+07	12500000	221	<1E-7
12	3.70E+07	37500000	204	<1E-7
11	1.10E+07	11500000	203	<1E-7
1	81500000	8.20E+07	199	<1E-7
12	7.00E+07	70500000	194	<1E-7
1	75500000	7.60E+07	193	<1E-7
15	7.10E+07	71500000	188	<1E-7
9	113500000	1.14E+08	186	<1E-7
11	110500000	1.11E+08	184	<1E-7
17	5.10E+07	51500000	176	<1E-7
11	9.00E+06	9500000	175	<1E-7
12	3.80E+07	38500000	175	<1E-7
9	116500000	1.17E+08	174	<1E-7
17	7.20E+07	72500000	174	<1E-7
12	69500000	7.00E+07	172	<1E-7
15	3.10E+07	31500000	172	<1E-7
4	9.70E+07	97500000	170	<1E-7
11	116500000	1.17E+08	167	<1E-7
11	114500000	1.15E+08	161	<1E-7
13	56500000	5.70E+07	161	<1E-7
11	7.30E+07	73500000	160	<1E-7
11	1.15E+08	115500000	158	<1E-7
11	16500000	1.70E+07	157	<1E-7

11	8.00E+06	8500000	155	<1E-7
11	9500000	1.00E+07	155	<1E-7
11	105500000	1.06E+08	155	<1E-7
3	1.05E+08	105500000	153	<1E-7
4	6.30E+07	63500000	153	<1E-7
18	63500000	6.40E+07	151	<1E-7
15	42500000	4.30E+07	149	<1E-7
11	9.40E+07	94500000	148	<1E-7
15	6.10E+07	61500000	144	<1E-7
11	5500000	6.00E+06	143	<1E-7
13	5.70E+07	57500000	139	<1E-7
12	7.30E+07	73500000	138	<1E-7
8	130500000	1.31E+08	136	<1E-7
12	6.90E+07	69500000	135	<1E-7
12	3.40E+07	34500000	134	<1E-7
1	7.80E+07	78500000	133	<1E-7
1	8.20E+07	82500000	133	<1E-7
4	5.80E+07	58500000	133	<1E-7
1	7.70E+07	77500000	132	<1E-7
3	1.04E+08	104500000	132	<1E-7
11	7.00E+06	7500000	132	<1E-7
11	97500000	9.80E+07	131	<1E-7
17	5.00E+07	50500000	131	<1E-7
2	1.42E+08	142500000	130	<1E-7
8	1.27E+08	127500000	129	<1E-7
2	1.29E+08	129500000	127	<1E-7
12	73500000	7.40E+07	127	<1E-7
15	27500000	2.80E+07	125	<1E-7
11	1.16E+08	116500000	122	<1E-7
12	72500000	7.30E+07	122	<1E-7
5	1.30E+08	130500000	121	<1E-7
8	9.20E+07	92500000	120	<1E-7
11	6500000	7.00E+06	120	<1E-7
3	38500000	3.90E+07	119	<1E-7
11	1.40E+07	14500000	119	<1E-7
5	1.06E+08	106500000	118	<1E-7
11	13500000	1.40E+07	118	<1E-7
15	71500000	7.20E+07	118	<1E-7
2	127500000	1.28E+08	117	<1E-7
2	1.43E+08	143500000	117	<1E-7

11	1.03E+08	103500000	117	<1E-7
1	7.40E+07	74500000	116	<1E-7
4	63500000	6.40E+07	116	<1E-7
5	1.29E+08	129500000	116	<1E-7
5	129500000	1.30E+08	116	<1E-7
8	90500000	9.10E+07	116	<1E-7
17	1.20E+07	12500000	116	<1E-7
2	179500000	1.80E+08	115	<1E-7
1	78500000	7.90E+07	113	<1E-7
17	44500000	4.50E+07	113	<1E-7
4	43500000	4.40E+07	112	<1E-7
8	91500000	9.20E+07	111	<1E-7
11	9.70E+07	97500000	111	<1E-7
6	119500000	1.20E+08	110	<1E-7
9	1.18E+08	118500000	110	<1E-7
13	6.00E+07	60500000	110	<1E-7
7	1.33E+08	133500000	109	<1E-7
9	86500000	8.70E+07	109	<1E-7
12	2.60E+07	26500000	109	<1E-7
15	31500000	3.20E+07	109	<1E-7
1	7.50E+07	75500000	108	<1E-7
12	6.50E+07	65500000	108	<1E-7
13	57500000	5.80E+07	108	<1E-7
2	142500000	1.43E+08	107	<1E-7
1	124500000	1.25E+08	106	<1E-7
2	141500000	1.42E+08	106	<1E-7
6	1.80E+07	18500000	106	<1E-7
6	7.70E+07	77500000	106	<1E-7
9	8.70E+07	87500000	106	<1E-7
12	3.30E+07	33500000	106	<1E-7
6	8.20E+07	82500000	105	<1E-7
6	83500000	8.40E+07	105	<1E-7
1	79500000	8.00E+07	104	<1E-7
1	80500000	8.10E+07	104	<1E-7
9	68500000	6.90E+07	104	<1E-7
11	10500000	1.10E+07	104	<1E-7
11	93500000	9.40E+07	104	<1E-7
15	37500000	3.80E+07	104	<1E-7
4	8.90E+07	89500000	103	<1E-7
8	125500000	1.26E+08	103	<1E-7

11	8500000	9.00E+06	103	<1E-7
3	8.50E+07	85500000	102	<1E-7
13	64500000	6.50E+07	102	<1E-7
15	4.40E+07	44500000	102	<1E-7
17	8.60E+07	86500000	102	<1E-7
2	119500000	1.20E+08	101	<1E-7
2	1.52E+08	152500000	101	<1E-7
11	7500000	8.00E+06	101	<1E-7
11	1.30E+07	13500000	101	<1E-7
12	32500000	3.30E+07	101	<1E-7
12	36500000	3.70E+07	101	<1E-7
17	4.50E+07	45500000	101	<1E-7
18	54500000	5.50E+07	101	<1E-7
9	112500000	1.13E+08	100	<1E-7
11	1.08E+08	108500000	100	<1E-7
3	104500000	1.05E+08	99	<1E-7
3	105500000	1.06E+08	99	<1E-7
12	33500000	3.40E+07	99	<1E-7
12	74500000	7.50E+07	99	<1E-7
4	1.26E+08	126500000	98	<1E-7
8	9.10E+07	91500000	98	<1E-7
12	7.40E+07	74500000	98	<1E-7
14	8.50E+07	85500000	98	<1E-7
17	8.70E+07	87500000	98	<1E-7
1	8.10E+07	81500000	97	<1E-7
5	1.31E+08	131500000	97	<1E-7
11	1.10E+08	110500000	97	<1E-7
5	7500000	8.00E+06	96	<1E-7
9	6.70E+07	67500000	96	<1E-7
17	48500000	4.90E+07	96	<1E-7
15	4.30E+07	43500000	95	<1E-7
13	6.40E+07	64500000	94	<1E-7
17	49500000	5.00E+07	94	<1E-7
4	99500000	1.00E+08	93	<1E-7
11	1.70E+07	17500000	93	<1E-7
3	83500000	8.40E+07	92	<1E-7
6	125500000	1.26E+08	92	<1E-7
1	74500000	7.50E+07	91	<1E-7
1	77500000	7.80E+07	91	<1E-7
2	125500000	1.26E+08	91	<1E-7

3	3.10E+07	31500000	91	<1E-7
11	96500000	9.70E+07	91	<1E-7
15	6.80E+07	68500000	91	<1E-7
1	76500000	7.70E+07	90	<1E-7
3	85500000	8.60E+07	90	<1E-7
4	141500000	1.42E+08	90	<1E-7
17	4.10E+07	41500000	90	<1E-7
3	1.03E+08	103500000	89	<1E-7
4	142500000	1.43E+08	89	<1E-7
5	1.20E+07	12500000	89	<1E-7
12	37500000	3.80E+07	89	<1E-7
4	9.90E+07	99500000	88	<1E-7
8	92500000	9.30E+07	88	<1E-7
11	12500000	1.30E+07	88	<1E-7
17	5.70E+07	57500000	88	<1E-7
1	176500000	1.77E+08	87	<1E-7
6	50500000	5.10E+07	87	<1E-7
11	7.40E+07	74500000	87	<1E-7
11	92500000	9.30E+07	87	<1E-7
15	82500000	8.30E+07	87	<1E-7
18	7.00E+07	70500000	87	<1E-7
3	6.00E+07	60500000	86	<1E-7
3	8.30E+07	83500000	86	<1E-7
3	122500000	1.23E+08	86	<1E-7
6	138500000	1.39E+08	86	<1E-7
7	8.30E+07	83500000	86	<1E-7
8	79500000	8.00E+07	86	<1E-7
15	6.00E+07	60500000	86	<1E-7
18	6.50E+07	65500000	86	<1E-7
3	81500000	8.20E+07	85	<1E-7
3	1.43E+08	143500000	85	<1E-7
9	44500000	4.50E+07	85	<1E-7
11	1.60E+07	16500000	85	<1E-7
13	59500000	6.00E+07	85	<1E-7
13	7.00E+07	70500000	85	<1E-7
4	127500000	1.28E+08	84	<1E-7
6	80500000	8.10E+07	84	<1E-7
7	19500000	2.00E+07	84	<1E-7
11	11500000	1.20E+07	84	<1E-7
11	44500000	4.50E+07	84	<1E-7

12	26500000	2.70E+07	84	<1E-7
12	7.80E+07	78500000	84	<1E-7
13	60500000	6.10E+07	84	<1E-7
13	7.20E+07	72500000	84	<1E-7
18	61500000	6.20E+07	84	<1E-7
19	17500000	1.80E+07	84	<1E-7
7	10500000	1.10E+07	83	<1E-7
8	126500000	1.27E+08	83	<1E-7
9	90500000	9.10E+07	83	<1E-7
11	72500000	7.30E+07	83	<1E-7
12	7500000	8.00E+06	83	<1E-7
15	10500000	1.10E+07	83	<1E-7
15	59500000	6.00E+07	83	<1E-7
15	8.60E+07	86500000	83	<1E-7
17	12500000	1.30E+07	83	<1E-7
17	50500000	5.10E+07	83	<1E-7
18	68500000	6.90E+07	83	<1E-7
2	152500000	1.53E+08	82	<1E-7
3	7.10E+07	71500000	82	<1E-7
5	1.30E+07	13500000	82	<1E-7
8	1.30E+08	130500000	82	<1E-7
11	5.10E+07	51500000	82	<1E-7
11	6.20E+07	62500000	82	<1E-7
15	8.00E+07	80500000	82	<1E-7
15	85500000	8.60E+07	82	<1E-7
18	8.50E+07	85500000	82	<1E-7
1	9.30E+07	93500000	81	<1E-7
2	1.28E+08	128500000	81	<1E-7
3	1.29E+08	129500000	81	<1E-7
5	5.10E+07	51500000	81	<1E-7
11	24500000	2.50E+07	81	<1E-7
17	4.90E+07	49500000	81	<1E-7
2	120500000	1.21E+08	80	<1E-7
2	1.44E+08	144500000	80	<1E-7
2	158500000	1.59E+08	80	<1E-7
3	9.60E+07	96500000	80	<1E-7
4	54500000	5.50E+07	80	<1E-7
5	150500000	1.51E+08	80	<1E-7
7	1.05E+08	105500000	80	<1E-7
11	15500000	1.60E+07	80	<1E-7

17	9500000	1.00E+07	80	<1E-7
1	3.40E+07	34500000	79	<1E-7
2	1.65E+08	165500000	79	<1E-7
3	103500000	1.04E+08	79	<1E-7
7	2.70E+07	27500000	79	<1E-7
11	1.00E+07	10500000	79	<1E-7
12	1.20E+07	12500000	79	<1E-7
17	3.70E+07	37500000	79	<1E-7
17	43500000	4.40E+07	79	<1E-7
17	55500000	5.60E+07	79	<1E-7
18	6.30E+07	63500000	79	<1E-7
18	64500000	6.50E+07	79	<1E-7
1	1.61E+08	161500000	78	<1E-7
1	190500000	1.91E+08	78	<1E-7
4	128500000	1.29E+08	78	<1E-7
7	1.70E+07	17500000	78	<1E-7
11	1.02E+08	102500000	78	<1E-7
12	47500000	4.80E+07	78	<1E-7
14	8.60E+07	86500000	78	<1E-7
2	129500000	1.30E+08	77	<1E-7
2	151500000	1.52E+08	77	<1E-7
2	1.66E+08	166500000	77	<1E-7
3	59500000	6.00E+07	77	<1E-7
3	6.30E+07	63500000	77	<1E-7
3	8.20E+07	82500000	77	<1E-7
6	49500000	5.00E+07	77	<1E-7
7	17500000	1.80E+07	77	<1E-7
7	29500000	3.00E+07	77	<1E-7
11	1.20E+07	12500000	77	<1E-7
12	27500000	2.80E+07	77	<1E-7
12	53500000	5.40E+07	77	<1E-7
2	1.20E+08	120500000	76	<1E-7
2	121500000	1.22E+08	76	<1E-7
4	35500000	3.60E+07	76	<1E-7
4	69500000	7.00E+07	76	<1E-7
5	1.49E+08	149500000	76	<1E-7
12	34500000	3.50E+07	76	<1E-7
15	9.00E+06	9500000	76	<1E-7
15	101500000	1.02E+08	76	<1E-7
18	65500000	6.60E+07	76	<1E-7

5	2.40E+07	24500000	75	<1E-7
7	1.06E+08	106500000	75	<1E-7
8	7500000	8.00E+06	75	<1E-7
9	9.50E+07	95500000	75	<1E-7
11	7.10E+07	71500000	75	<1E-7
17	33500000	3.40E+07	75	<1E-7
17	7.90E+07	79500000	75	<1E-7
18	7.80E+07	78500000	75	<1E-7
3	30500000	3.10E+07	74	<1E-7
3	32500000	3.30E+07	74	<1E-7
3	7.60E+07	76500000	74	<1E-7
6	55500000	5.60E+07	74	<1E-7
6	126500000	1.27E+08	74	<1E-7
12	40500000	4.10E+07	74	<1E-7
15	35500000	3.60E+07	74	<1E-7
17	36500000	3.70E+07	74	<1E-7
18	88500000	8.90E+07	74	<1E-7
1	1.25E+08	125500000	73	<1E-7
2	1.48E+08	148500000	73	<1E-7
4	129500000	1.30E+08	73	<1E-7
6	82500000	8.30E+07	73	<1E-7
11	9.60E+07	96500000	73	<1E-7
12	38500000	3.90E+07	73	<1E-7
12	51500000	5.20E+07	73	<1E-7
14	85500000	8.60E+07	73	<1E-7
15	14500000	1.50E+07	73	<1E-7
17	4.40E+07	44500000	73	<1E-7
1	92500000	9.30E+07	72	<1E-7
3	33500000	3.40E+07	72	<1E-7
3	60500000	6.10E+07	72	<1E-7
3	8.70E+07	87500000	72	<1E-7
4	5.40E+07	54500000	72	<1E-7
4	1.00E+08	100500000	72	<1E-7
7	124500000	1.25E+08	72	<1E-7
11	26500000	2.70E+07	72	<1E-7
11	7.20E+07	72500000	72	<1E-7
12	2.80E+07	28500000	72	<1E-7
12	60500000	6.10E+07	72	<1E-7
15	43500000	4.40E+07	72	<1E-7
15	58500000	5.90E+07	72	<1E-7

15	8.50E+07	85500000	72	<1E-7
15	8.90E+07	89500000	72	<1E-7
17	25500000	2.60E+07	72	<1E-7
17	51500000	5.20E+07	72	<1E-7
17	5.20E+07	52500000	72	<1E-7
3	69500000	7.00E+07	71	1.00E-07
3	1.27E+08	127500000	71	1.00E-07
4	4.40E+07	44500000	71	1.00E-07
4	95500000	9.60E+07	71	1.00E-07
5	8.20E+07	82500000	71	1.00E-07
6	56500000	5.70E+07	71	1.00E-07
6	5.70E+07	57500000	71	1.00E-07
8	1.26E+08	126500000	71	1.00E-07
9	85500000	8.60E+07	71	1.00E-07
11	6.90E+07	69500000	71	1.00E-07
12	6.60E+07	66500000	71	1.00E-07
13	71500000	7.20E+07	71	1.00E-07
15	3.20E+07	32500000	71	1.00E-07
17	47500000	4.80E+07	71	1.00E-07
18	6.40E+07	64500000	71	1.00E-07
1	4.00E+07	40500000	70	1.00E-07
2	1.17E+08	117500000	70	1.00E-07
2	1.37E+08	137500000	70	1.00E-07
3	8.40E+07	84500000	70	1.00E-07
4	55500000	5.60E+07	70	1.00E-07
5	7.00E+06	7500000	70	1.00E-07
5	28500000	2.90E+07	70	1.00E-07
8	1.31E+08	131500000	70	1.00E-07
9	6.20E+07	62500000	70	1.00E-07
9	67500000	6.80E+07	70	1.00E-07
11	3.20E+07	32500000	70	1.00E-07
11	8.30E+07	83500000	70	1.00E-07
12	6.10E+07	61500000	70	1.00E-07
2	106500000	1.07E+08	69	1.00E-07
2	134500000	1.35E+08	69	1.00E-07
3	52500000	5.30E+07	69	1.00E-07
3	6.50E+07	65500000	69	1.00E-07
3	1.02E+08	102500000	69	1.00E-07
3	143500000	1.44E+08	69	1.00E-07
4	96500000	9.70E+07	69	1.00E-07

4	1.41E+08	141500000	69	1.00E-07
5	128500000	1.29E+08	69	1.00E-07
9	9.10E+07	91500000	69	1.00E-07
9	1.13E+08	113500000	69	1.00E-07
11	102500000	1.03E+08	69	1.00E-07
12	65500000	6.60E+07	69	1.00E-07
16	1.20E+07	12500000	69	1.00E-07
17	24500000	2.50E+07	69	1.00E-07
18	8.40E+07	84500000	69	1.00E-07
1	87500000	8.80E+07	68	2.00E-07
2	1.21E+08	121500000	68	2.00E-07
2	1.27E+08	127500000	68	2.00E-07
2	137500000	1.38E+08	68	2.00E-07
3	1.23E+08	123500000	68	2.00E-07
7	108500000	1.09E+08	68	2.00E-07
8	1.25E+08	125500000	68	2.00E-07
11	4500000	5.00E+06	68	2.00E-07
12	54500000	5.50E+07	68	2.00E-07
13	63500000	6.40E+07	68	2.00E-07
13	85500000	8.60E+07	68	2.00E-07
16	55500000	5.60E+07	68	2.00E-07
17	1.00E+07	10500000	68	2.00E-07
2	115500000	1.16E+08	67	2.00E-07
2	1.22E+08	122500000	67	2.00E-07
3	7.30E+07	73500000	67	2.00E-07
4	1.49E+08	149500000	67	2.00E-07
5	4.10E+07	41500000	67	2.00E-07
7	18500000	1.90E+07	67	2.00E-07
7	8.10E+07	81500000	67	2.00E-07
7	105500000	1.06E+08	67	2.00E-07
9	8.50E+07	85500000	67	2.00E-07
10	1.22E+08	122500000	67	2.00E-07
11	9.90E+07	99500000	67	2.00E-07
11	99500000	1.00E+08	67	2.00E-07
12	3.50E+07	35500000	67	2.00E-07
14	4.60E+07	46500000	67	2.00E-07
15	6.90E+07	69500000	67	2.00E-07
15	8.40E+07	84500000	67	2.00E-07
16	72500000	7.30E+07	67	2.00E-07
17	4.80E+07	48500000	67	2.00E-07

19	21500000	2.20E+07	67	2.00E-07
		Treated		
3	28000000	28500000	289	<1E-7
9	67000000	67500000	240	<1E-7
9	67500000	68000000	188	<1E-7
9	66500000	67000000	180	<1E-7
9	65500000	66000000	173	<1E-7
3	29500000	30000000	161	<1E-7
2	94500000	95000000	156	<1E-7
3	28500000	29000000	152	<1E-7
9	65000000	65500000	145	<1E-7
2	129000000	129500000	144	<1E-7
9	68000000	68500000	142	<1E-7
9	29000000	29500000	139	<1E-7
3	29000000	29500000	138	<1E-7
3	30000000	30500000	135	<1E-7
2	77000000	77500000	129	<1E-7
2	119500000	120000000	128	<1E-7
9	116500000	117000000	128	<1E-7
7	53500000	54000000	127	<1E-7
6	80500000	81000000	122	<1E-7
12	73000000	73500000	119	<1E-7
6	77000000	77500000	117	<1E-7
11	116500000	117000000	117	<1E-7
11	4500000	5000000	116	<1E-7
2	120000000	120500000	114	<1E-7
9	107500000	108000000	112	<1E-7
9	114500000	115000000	109	<1E-7
9	68500000	69000000	108	<1E-7
3	152500000	153000000	107	<1E-7
2	114000000	114500000	104	<1E-7
17	41000000	41500000	102	<1E-7
19	40000000	40500000	100	<1E-7
2	123500000	124000000	99	<1E-7
2	129500000	130000000	99	<1E-7
6	78500000	79000000	99	<1E-7
19	26500000	27000000	99	<1E-7
2	127500000	128000000	97	<1E-7
4	141000000	141500000	95	<1E-7
9	66000000	66500000	95	<1E-7

8	91000000	91500000	94	<1E-7
9	115000000	115500000	94	<1E-7
1	187000000	187500000	93	<1E-7
3	52500000	53000000	93	<1E-7
5	106500000	107000000	93	<1E-7
7	35500000	36000000	93	<1E-7
8	92000000	92500000	93	<1E-7
17	48500000	49000000	93	<1E-7
7	106000000	106500000	92	<1E-7
2	119000000	119500000	91	<1E-7
13	37000000	37500000	91	<1E-7
17	45000000	45500000	91	<1E-7
1	92500000	93000000	90	<1E-7
6	78000000	78500000	90	<1E-7
9	87000000	87500000	90	<1E-7
11	44500000	45000000	90	<1E-7
2	93500000	94000000	88	<1E-7
4	98500000	99000000	88	<1E-7
9	118500000	119000000	88	<1E-7
12	73500000	74000000	88	<1E-7
16	45500000	46000000	88	<1E-7
3	104000000	104500000	87	<1E-7
9	117500000	118000000	87	<1E-7
13	7000000	7500000	87	<1E-7
18	70000000	70500000	87	<1E-7
6	82000000	82500000	86	<1E-7
13	100000000	100500000	86	<1E-7
7	133000000	133500000	85	<1E-7
8	90500000	91000000	85	<1E-7
11	73000000	73500000	85	<1E-7
17	44500000	45000000	85	<1E-7
2	120500000	121000000	84	<1E-7
3	130000000	130500000	84	<1E-7
4	142500000	143000000	84	<1E-7
7	47500000	48000000	84	<1E-7
7	105000000	105500000	84	<1E-7
9	103000000	103500000	84	<1E-7
2	128000000	128500000	83	<1E-7
4	141500000	142000000	83	<1E-7
5	110500000	111000000	83	<1E-7

12	15500000	16000000	83	<1E-7
13	9000000	9500000	83	<1E-7
14	22000000	22500000	83	<1E-7
19	38500000	39000000	83	<1E-7
2	124500000	125000000	82	<1E-7
5	7500000	8000000	82	<1E-7
9	117000000	117500000	82	<1E-7
6	50500000	51000000	81	<1E-7
9	111000000	111500000	81	<1E-7
17	44000000	44500000	81	<1E-7
1	3000000	3500000	80	<1E-7
1	111500000	112000000	80	<1E-7
1	154000000	154500000	80	<1E-7
2	127000000	127500000	80	<1E-7
3	103500000	104000000	80	<1E-7
4	128500000	129000000	80	<1E-7
11	116000000	116500000	80	<1E-7
12	74000000	74500000	80	<1E-7
12	100500000	101000000	80	<1E-7
2	123000000	123500000	79	<1E-7
18	63500000	64000000	79	<1E-7
2	117000000	117500000	78	<1E-7
3	138500000	139000000	78	<1E-7
13	37500000	38000000	78	<1E-7
7	52500000	53000000	77	<1E-7
7	108500000	109000000	77	<1E-7
15	100500000	101000000	77	<1E-7
18	81000000	81500000	77	<1E-7
2	76500000	77000000	76	<1E-7
2	125500000	126000000	76	<1E-7
2	153500000	154000000	76	<1E-7
5	51500000	52000000	76	<1E-7
7	105500000	106000000	76	<1E-7
9	71000000	71500000	76	<1E-7
13	11500000	12000000	76	<1E-7
17	12000000	12500000	76	<1E-7
19	26000000	26500000	76	<1E-7
4	142000000	142500000	75	<1E-7
17	49500000	50000000	75	<1E-7
17	75500000	76000000	75	<1E-7

18	68500000	69000000	75	<1E-7
5	109000000	109500000	74	<1E-7
7	107000000	107500000	74	<1E-7
7	108000000	108500000	74	<1E-7
8	125500000	126000000	74	<1E-7
18	65000000	65500000	74	<1E-7
1	78000000	78500000	73	<1E-7
2	122000000	122500000	73	<1E-7
3	101500000	102000000	73	<1E-7
6	49500000	50000000	73	<1E-7
7	106500000	107000000	73	<1E-7
8	92500000	93000000	73	<1E-7
15	37500000	38000000	73	<1E-7
16	95000000	95500000	73	<1E-7
18	80500000	81000000	73	<1E-7
3	63000000	63500000	72	<1E-7
3	85000000	85500000	72	<1E-7
3	143000000	143500000	72	<1E-7
8	91500000	92000000	72	<1E-7
13	8000000	8500000	72	<1E-7
17	31000000	31500000	72	<1E-7
17	78500000	79000000	72	<1E-7
18	12000000	12500000	72	<1E-7
18	64500000	65000000	72	<1E-7
5	51000000	51500000	71	1.00E-07
5	111000000	111500000	71	1.00E-07
5	131000000	131500000	71	1.00E-07
7	53000000	53500000	71	1.00E-07
9	44500000	45000000	71	1.00E-07
13	3500000	4000000	71	1.00E-07
17	37000000	37500000	71	1.00E-07
5	129500000	130000000	70	1.00E-07
7	103500000	104000000	70	1.00E-07
8	94500000	95000000	70	1.00E-07
9	86500000	87000000	70	1.00E-07
13	7500000	8000000	70	1.00E-07
13	23000000	23500000	70	1.00E-07
5	128500000	129000000	69	1.00E-07
9	106500000	107000000	69	1.00E-07
11	69000000	69500000	69	1.00E-07

13	6500000	7000000	69	1.00E-07
18	30000000	30500000	69	1.00E-07
1	190500000	191000000	68	2.00E-07
2	113500000	114000000	68	2.00E-07
3	150500000	151000000	68	2.00E-07
4	129500000	130000000	68	2.00E-07
6	126500000	127000000	68	2.00E-07
7	104500000	105000000	68	2.00E-07
7	107500000	108000000	68	2.00E-07
9	50000000	50500000	68	2.00E-07
11	45000000	45500000	68	2.00E-07
11	64500000	65000000	68	2.00E-07
11	70500000	71000000	68	2.00E-07
11	71000000	71500000	68	2.00E-07
13	4500000	5000000	68	2.00E-07
13	108500000	109000000	68	2.00E-07
17	43500000	44000000	68	2.00E-07
17	49000000	49500000	68	2.00E-07
19	25000000	25500000	68	2.00E-07
Ratio of Treated to Untreated Expression				
12	77500000	7.80E+07	925	<1E-7
12	103500000	1.04E+08	790	<1E-7
3	8.50E+07	85500000	430	<1E-7
12	8.10E+07	81500000	427	<1E-7
11	110500000	1.11E+08	410	<1E-7
11	112500000	1.13E+08	393	<1E-7
2	1.44E+08	144500000	369	<1E-7
2	143500000	1.44E+08	368	<1E-7
2	144500000	1.45E+08	366	<1E-7
2	152500000	1.53E+08	361	<1E-7
4	1.26E+08	126500000	342	<1E-7
3	8.20E+07	82500000	340	<1E-7
14	4.60E+07	46500000	337	<1E-7
14	35500000	3.60E+07	335	<1E-7
2	6.10E+07	61500000	330	<1E-7
11	1.12E+08	112500000	330	<1E-7
13	9.30E+07	93500000	320	<1E-7
15	1.50E+07	15500000	307	<1E-7
12	81500000	8.20E+07	303	<1E-7
4	125500000	1.26E+08	300	<1E-7

15	2.10E+07	21500000	293	<1E-7
3	8.30E+07	83500000	290	<1E-7
15	1.30E+07	13500000	290	<1E-7
11	113500000	1.14E+08	286	<1E-7
11	111500000	1.12E+08	279	<1E-7
11	1.13E+08	113500000	275	<1E-7
3	81500000	8.20E+07	273	<1E-7
15	20500000	2.10E+07	271	<1E-7
12	7.80E+07	78500000	267	<1E-7
15	16500000	1.70E+07	267	<1E-7
2	1.53E+08	153500000	261	<1E-7
11	1.14E+08	114500000	260	<1E-7
13	92500000	9.30E+07	256	<1E-7
15	11500000	1.20E+07	256	<1E-7
3	8.40E+07	84500000	251	<1E-7
15	13500000	1.40E+07	250	<1E-7
2	5.40E+07	54500000	248	<1E-7
3	85500000	8.60E+07	246	<1E-7
3	83500000	8.40E+07	244	<1E-7
3	90500000	9.10E+07	243	<1E-7
2	1.52E+08	152500000	242	<1E-7
15	1.40E+07	14500000	237	<1E-7
3	86500000	8.70E+07	234	<1E-7
2	158500000	1.59E+08	230	<1E-7
3	98500000	9.90E+07	229	<1E-7
11	116500000	1.17E+08	228	<1E-7
12	3.80E+07	38500000	224	<1E-7
11	1.10E+08	110500000	217	<1E-7
15	15500000	1.60E+07	216	<1E-7
6	17500000	1.80E+07	214	<1E-7
13	89500000	9.00E+07	211	<1E-7
12	108500000	1.09E+08	210	<1E-7
3	1.05E+08	105500000	209	<1E-7
2	1.43E+08	143500000	204	<1E-7
12	1.09E+08	109500000	202	<1E-7
12	7.50E+07	75500000	195	<1E-7
13	9.40E+07	94500000	194	<1E-7
13	86500000	8.70E+07	192	<1E-7
13	8.90E+07	89500000	191	<1E-7
2	157500000	1.58E+08	190	<1E-7

3	105500000	1.06E+08	190	<1E-7
3	8.70E+07	87500000	182	<1E-7
3	58500000	5.90E+07	180	<1E-7
8	1.23E+08	123500000	178	<1E-7
3	6.00E+07	60500000	177	<1E-7
2	1.42E+08	142500000	176	<1E-7
15	12500000	1.30E+07	175	<1E-7
15	63500000	6.40E+07	174	<1E-7
13	9.00E+07	90500000	172	<1E-7
2	1.57E+08	157500000	169	<1E-7
3	5.80E+07	58500000	169	<1E-7
2	141500000	1.42E+08	168	<1E-7
2	142500000	1.43E+08	166	<1E-7
3	57500000	5.80E+07	165	<1E-7
15	2.30E+07	23500000	164	<1E-7
3	59500000	6.00E+07	161	<1E-7
13	9.50E+07	95500000	161	<1E-7
2	1.58E+08	158500000	159	<1E-7
2	72500000	7.30E+07	156	<1E-7
13	8.60E+07	86500000	154	<1E-7
15	6.70E+07	67500000	153	<1E-7
13	84500000	8.50E+07	150	<1E-7
13	90500000	9.10E+07	150	<1E-7
3	56500000	5.70E+07	149	<1E-7
13	95500000	9.60E+07	149	<1E-7
13	8.80E+07	88500000	148	<1E-7
13	88500000	8.90E+07	146	<1E-7
15	1.10E+07	11500000	146	<1E-7
13	93500000	9.40E+07	144	<1E-7
6	1.30E+07	13500000	143	<1E-7
15	1.00E+07	10500000	141	<1E-7
4	1.25E+08	125500000	138	<1E-7
15	25500000	2.60E+07	137	<1E-7
3	6.10E+07	61500000	133	<1E-7
3	1.04E+08	104500000	133	<1E-7
13	83500000	8.40E+07	131	<1E-7
2	6.20E+07	62500000	130	<1E-7
3	79500000	8.00E+07	128	<1E-7
15	39500000	4.00E+07	128	<1E-7
3	104500000	1.05E+08	127	<1E-7

13	9.10E+07	91500000	126	<1E-7
15	22500000	2.30E+07	123	<1E-7
4	1.51E+08	151500000	120	<1E-7
12	72500000	7.30E+07	119	<1E-7
2	147500000	1.48E+08	116	<1E-7
15	18500000	1.90E+07	114	<1E-7
16	97500000	9.80E+07	110	<1E-7
15	6.40E+07	64500000	109	<1E-7
15	6.60E+07	66500000	109	<1E-7
12	3.70E+07	37500000	108	<1E-7
12	7.30E+07	73500000	107	<1E-7
15	2.00E+07	20500000	106	<1E-7
2	153500000	1.54E+08	104	<1E-7
11	115500000	1.16E+08	104	<1E-7
13	94500000	9.50E+07	104	<1E-7
6	7.70E+07	77500000	100	<1E-7
12	9.50E+07	95500000	100	<1E-7
14	2.80E+07	28500000	100	<1E-7
13	85500000	8.60E+07	98	<1E-7
1	20500000	2.10E+07	97	<1E-7
3	5.90E+07	59500000	97	<1E-7
6	7.80E+07	78500000	96	<1E-7
3	3.00E+07	30500000	94	<1E-7
13	82500000	8.30E+07	94	<1E-7
15	9.00E+06	9500000	93	<1E-7
1	4.40E+07	44500000	92	<1E-7
3	2.80E+07	28500000	92	<1E-7
13	91500000	9.20E+07	92	<1E-7
3	84500000	8.50E+07	91	<1E-7
7	127500000	1.28E+08	91	<1E-7
2	9.30E+07	93500000	89	<1E-7
15	64500000	6.50E+07	88	<1E-7
16	3.50E+07	35500000	88	<1E-7
4	1.53E+08	153500000	87	<1E-7
3	5.70E+07	57500000	86	<1E-7
11	6.00E+06	6500000	86	<1E-7
13	8.70E+07	87500000	86	<1E-7
3	87500000	8.80E+07	85	<1E-7
15	8500000	9.00E+06	85	<1E-7
7	1.27E+08	127500000	84	<1E-7

15	6.50E+07	65500000	84	<1E-7
15	65500000	6.60E+07	83	<1E-7
3	60500000	6.10E+07	82	<1E-7
3	8.80E+07	88500000	81	<1E-7
15	66500000	6.70E+07	81	<1E-7
12	3.30E+07	33500000	80	<1E-7
4	124500000	1.25E+08	78	<1E-7
12	7.40E+07	74500000	78	<1E-7
15	1.60E+07	16500000	78	<1E-7
15	5.30E+07	53500000	78	<1E-7
16	9.80E+07	98500000	78	<1E-7
12	59500000	6.00E+07	77	<1E-7
17	5.00E+06	5500000	75	<1E-7
9	5.00E+06	5500000	74	<1E-7
13	8.20E+07	82500000	74	<1E-7
2	179500000	1.80E+08	71	1.00E-07
12	73500000	7.40E+07	71	1.00E-07
18	6.50E+07	65500000	71	1.00E-07
4	35500000	3.60E+07	70	1.00E-07
15	72500000	7.30E+07	70	1.00E-07
16	35500000	3.60E+07	70	1.00E-07
8	1.26E+08	126500000	69	1.00E-07
12	74500000	7.50E+07	69	1.00E-07
14	28500000	2.90E+07	69	1.00E-07
15	1.20E+07	12500000	69	1.00E-07
15	1.80E+07	18500000	69	1.00E-07
1	92500000	9.30E+07	67	2.00E-07
2	92500000	9.30E+07	66	6.00E-07
6	119500000	1.20E+08	66	6.00E-07
8	126500000	1.27E+08	66	6.00E-07
11	1.16E+08	116500000	66	6.00E-07
12	8.70E+07	87500000	66	6.00E-07

Table 5.4. Weighted Eigengenes show high correlation to eigengenes at various hard thresholds

Eigengene at % Cutoff	Correlation to Weighted Eigengene
10	0.99565614
15	0.99622204
20	0.991846504
25	0.98249414
30	0.977296708
35	0.968713871
40	0.959325472
45	0.948212019
50	0.926094597
55	0.87246181
60	0.838856973
65	0.81198492
70	0.776717497
75	0.786895364
80	0.739061392
85	0.867267286
90	0.861934555

Table 5.5. GO Enrichments for Module 5.

Term	# of genes in module	% of genes in module	Genes in Category	Fold Enrichment	Benjamini-corrected P-value
glycoprotein	24	64.86	3600	3.400762	2.92E-07
signal	21	56.75	2970	3.606869	1.76E-06
extracellular matrix	8	21.62	213	19.15922	4.72E-06
short sequence motif:Cell attachment site	6	16.21	76	36.13759	3.78E-05
disulfide bond	17	45.94	2469	3.51233	8.76E-05
Secreted	13	35.13	1420	4.67006	1.43E-04
proteinaceous extracellular matrix	8	21.62	297	10.52525	1.43E-04
disulfide bond	16	43.24	2379	3.078556	0.002175
protein complex binding	4	10.81	78	21.9818	0.041418

Table 5.6. Significant Near Edge Orientation (NEO) values linking genes in module 5 to HF-related Phenotypes. A Leo.AtoB or BtoA score of greater than 0.75 and a p-value of less than <0.05 are considered significant.

Model	leo.nb.AtoB	leo.nb.BtoA	mlogp.M.AtoB	mlogp.M.BtoA
Causal Genes				
SNP51771 to Dct to Right Ventricle	1.21	-5.92	0.0069	5.93
SNP28253 to Dct to Left Ventricle	1.33	-6.57	0.00741	6.58
SNP508278 to Dct to Total Heart	0.862	-2.8	0.0223	2.83
SNP3161235 to 2610028H24Rik to Left Ventricle	0.812	-3.42	0.0421	3.46
SNP3436153 to Dct to Right Ventricle	0.968	-3.96	0.045	4.01
SNP3436325 to Dct to Right Ventricle	0.968	-3.96	0.045	4.01
SNP3589676 to Dct to Left Ventricle	1.11	-4.68	0.0452	4.72
SNP552781 to Dct to Right Ventricle	0.776	-2.89	0.046	2.94
Reactive Genes				
SNP3110538 to Col14a1 to Right Ventricle	-4.57	0.753	4.57	0.00193
SNP3110537 to Col14a1 to Right Ventricle	-4.57	0.753	4.57	0.00193
SNP2693574 to Col14a1 to Total Heart	-2.64	0.843	2.66	0.0175
SNP2693574 to Col14a1 to Left Ventricle	-2.55	0.831	2.58	0.0302
SNP2693574 to Clecsf8 to Right Ventricle	-4.63	0.765	4.68	0.0426
SNP2693574 to Col14a1 to Right Ventricle	-4.59	0.814	4.64	0.0471

References

1. Mudd, J. O. & Kass, D. A. Tackling heart failure in the twenty-first century. *Nature* **451,** 919–28 (2008).

2. Go, A. S. *et al.* Heart disease and stroke statistics--2013 update: a report from the American Heart Association. *Circulation* **127,** e6–e245 (2013).

3. Villard, E. *et al.* A genome-wide association study identifies two loci associated with heart failure due to dilated cardiomyopathy. *European heart journal* **32,** 1065–76 (2011).

4. Parsa, A. *et al.* Hypertrophy-associated polymorphisms ascertained in a founder cohort applied to heart failure risk and mortality. *Clinical and translational science* **4,** 17–23 (2011).

5. Ellinor, P. T. *et al.* A novel locus for dilated cardiomyopathy, diffuse myocardial fibrosis, and sudden death on chromosome 10q25-26. *Journal of the American College of Cardiology* **48,** 106–11 (2006).

6. Ghazalpour, A. *et al.* Hybrid mouse diversity panel: a panel of inbred mouse strains suitable for analysis of complex genetic traits. *Mammalian genome : official journal of the International Mammalian Genome Society* **23,** 680–92 (2012).

7. Bennett, B. J. *et al.* A high-resolution association mapping panel for the dissection of complex traits in mice. *Genome research* **20,** 281–90 (2010).

8. Park, C. C. *et al.* Gene networks associated with conditional fear in mice identified using a systems genetics approach. *BMC Systems Biology* **5,** 43 (2011).

9. Farber, C. R. *et al.* Mouse genome-wide association and systems genetics identify Asxl2 as a regulator of bone mineral density and osteoclastogenesis. *PLoS genetics* **7,** e1002038 (2011).

10. Farber, C. R. Systems-level analysis of genome-wide association data. *G3 (Bethesda, Md.)* **3,** 119–29 (2013).

11. Calabrese, G. *et al.* Systems genetic analysis of osteoblast-lineage cells. *PLoS genetics* **8,** e1003150 (2012).

12. Dorn, G. W. & Liggett, S. B. Pharmacogenomics of beta-adrenergic receptors and their accessory signaling proteins in heart failure. *Clinical and translational science* **1,** 255–62 (2008).

13. Kumaran, K. S. & Prince, P. S. M. Caffeic acid protects rat heart mitochondria against isoproterenol-induced oxidative damage. *Cell stress & chaperones* **15,** 791–806 (2010).

14. Brittsan, a G. *et al.* The effect of isoproterenol on phospholamban-deficient mouse hearts with altered thyroid conditions. *Journal of molecular and cellular cardiology* **31,** 1725–37 (1999).

15. Zheng, J., Shen, H., Xiong, Y., Yang, X. & He, J. The beta1-adrenergic receptor mediates extracellular signal-regulated kinase activation via Galphas. *Amino acids* **38,** 75–84 (2010).

16. Breckenridge, R. Heart failure and mouse models. *Disease models & mechanisms* **3,** 138–43 (2010).

17. Rau, C. D. *et al.* Maximal information component analysis: a novel non-linear network analysis method. *Frontiers in genetics* **4,** 28 (2013).

18. De Boer, R. a, Yu, L. & van Veldhuisen, D. J. Galectin-3 in cardiac remodeling and heart failure. *Current heart failure reports* **7,** 1–8 (2010).

19. Barton, P. J. R. *et al.* Increased expression of extracellular matrix regulators TIMP1 and MMP1 in deteriorating heart failure. *The Journal of heart and lung transplantation the official publication of the International Society for Heart Transplantation* **22,** 738–44 (2003).

20. Ikonomidis, J. S. *et al.* Accelerated LV remodeling after myocardial infarction in TIMP-1-deficient mice: effects of exogenous MMP inhibition. *American journal of physiology Heart and circulatory physiology* **288,** H149–58 (2005).

21. Xu, J. *et al.* beta 1-adrenergic receptor association with the synaptic scaffolding protein membrane-associated guanylate kinase inverted-2 (MAGI-2). Differential regulation of receptor internalization by MAGI-2 and PSD-95. *The Journal of biological chemistry* **276,** 41310–7 (2001).

22. Kang, H. M. *et al.* Efficient control of population structure in model organism association mapping. *Genetics* **178,** 1709–23 (2008).

23. Orozco, L. D. *et al.* Unraveling Inflammatory Responses using Systems Genetics and Gene-Environment Interactions in Macrophages. *Cell* **151,** 658–670 (2012).

24. Yalcin, B. *et al.* Sequence-based characterization of structural variation in the mouse genome. *Nature* **477,** 326–9 (2011).

25. Humphreys, D. T. *et al.* Complexity of murine cardiomyocyte miRNA biogenesis, sequence variant expression and function. *PloS one* **7,** e30933 (2012).

26. Mishra, P. K., Tyagi, N., Kumar, M. & Tyagi, S. C. MicroRNAs as a therapeutic target for cardiovascular diseases. *Journal of cellular and molecular medicine* **13,** 778–89 (2009).

27. Bauersachs, J. & Thum, T. Biogenesis and regulation of cardiovascular microRNAs. *Circulation research* **109,** 334–47 (2011).

28. Kashishian, A. *et al.* AKAP79 inhibits calcineurin through a site distinct from the immunophilin-binding region. *The Journal of Biological Chemistry* **273,** 27412–27419 (1998).

29. Nichols, C. B. *et al.* Sympathetic Stimulation of Adult Cardiomyocytes Requires Association of AKAP5 With a Subpopulation of L-Type Calcium Channels. *Circulation Research* **107,** 747–56 (2010).

30. Faul, C., Dhume, A., Schecter, A. D. & Mundel, P. Protein kinase A, Ca2+/calmodulin-dependent kinase II, and calcineurin regulate the intracellular trafficking of myopodin between the Z-disc and the nucleus of cardiac myocytes. *Molecular and cellular biology* **27,** 8215–27 (2007).

31. Slowik, A. *et al.* Alpha1-antichymotrypsin gene (SERPINA3) A/T polymorphism as a risk factor for aneurysmal subarachnoid hemorrhage. *Stroke: A Journal of Cerebral Circulation* **36,** 737–740 (2005).

32. Liu, W., Zhu, Y., Ge, M., Pang, Q. & Yu, Y. Polymorphism rs4934 of SERPINA3 and sporadic intracranial aneurysms in the Chinese population. *Cerebrovascular diseases Basel Switzerland* **29,** 68–72 (2010).

33. Asakura, M. & Kitakaze, M. Global gene expression profiling in the failing myocardium. *Circulation journal : official journal of the Japanese Circulation Society* **73,** 1568–76 (2009).

34. Farber, C. R. *et al.* Genetic dissection of a major mouse obesity QTL (Carfhg2): integration of gene expression and causality modeling. *Physiological genomics* **37,** 294–302 (2009).

35. Tritchler, D., Parkhomenko, E. & Beyene, J. Filtering genes for cluster and network analysis. *BMC bioinformatics* **10,** 193 (2009).

36. Dennis, G. *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome biology* **4,** P3 (2003).

37. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4,** 44–57 (2009).

38. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics* **9,** 559 (2008).

39. Weiss, J. N. *et al.* "Good enough solutions" and the genetics of complex diseases. *Circulation research* **111,** 493–504 (2012).

40. Lehnart, S. E. *et al.* Phosphodiesterase 4D deficiency in the ryanodine-receptor complex promotes heart failure and arrhythmias. *Cell* **123,** 25–35 (2005).

41. Beca, S. *et al.* Phosphodiesterase 4D Regulates Baseline Sarcoplasmic Reticulum Ca2+ Release and Cardiac Contractility, Independently of L-Type Ca2+ Current. *Circ Res* **109,** 1024–1030 (2011).

42. Aten, J. E., Fuller, T. F., Lusis, A. J. & Horvath, S. Using genetic markers to orient the edges in quantitative trait networks: the NEO software. *BMC systems biology* **2,** 34 (2008).

43. Wu, C. *et al.* BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome biology* **10,** R130 (2009).

44. Iida, J. *et al.* Synaptic scaffolding molecule alpha is a scaffold to mediate N-methyl-D-aspartate receptor-dependent RhoA activation in dendrites. *Molecular and cellular biology* **27,** 4388–405 (2007).

45. Krumholz, H. M. *et al.* Predictors of readmission among elderly survivors of admission with heart failure. *American heart journal* **139,** 72–7 (2000).

46. Morrison, A. C. *et al.* Genomic variation associated with mortality among adults of European and African ancestry with heart failure: the cohorts for heart and aging research in genomic epidemiology consortium. *Circulation. Cardiovascular genetics* **3,** 248–55 (2010).

47. Smith, N. L. *et al.* Association of genome-wide variation with the risk of incident heart failure in adults of European and African ancestry: a prospective meta-analysis from the cohorts for heart and aging research in genomic epidemiology (CHARGE) consortium. *Circulation. Cardiovascular genetics* **3,** 256–66 (2010).

48. Dorn, G. W. Adrenergic signaling polymorphisms and their impact on cardiovascular disease. *Physiological reviews* **90,** 1013–62 (2010).

49. Satwani, S., Dec, G. W. & Narula, J. Beta-adrenergic blockers in heart failure: review of mechanisms of action and clinical outcomes. *Journal of cardiovascular pharmacology and therapeutics* **9,** 243–55 (2004).

50. Salathe, M., Weiss, P. & Ritz, R. Rapid reversal of heart failure in a patient with phaeochromocytoma and catecholamine-induced cardiomyopathy who was treated with captopril. *British heart journal* **68,** 527–8 (1992).

51. Xiao, Y. *et al.* Discovering Dysfunction of Multiple MicroRNAs Cooperation in Disease by a Conserved MicroRNA Co-Expression Network. *PloS one* **7,** e32201 (2012).

52. Jentzsch, C. *et al.* A phenotypic screen to identify hypertrophy-modulating microRNAs in primary cardiomyocytes. *Journal of molecular and cellular cardiology* **52,** 13–20 (2012).

53. Coronado, M. J. *et al.* Testosterone and interleukin-1β increase cardiac remodeling during coxsackievirus B3 myocarditis via serpin A 3n. *American journal of physiology. Heart and circulatory physiology* **302,** H1726–36 (2012).

54. Samal, R. *et al.* OMICS-based exploration of the molecular phenotype of resident cardiac progenitor cells from adult murine heart. *Journal of proteomics* 1–12 (2012). doi:10.1016/j.jprot.2012.06.010

55. Fermin, D. R. *et al.* Sex and age dimorphism of myocardial gene expression in nonischemic human heart failure. *Circulation. Cardiovascular genetics* **1,** 117–25 (2008).

56. Guo, F.-J. *et al.* Expression and functional characterization of platelet-derived growth factor receptor-like gene. *World Journal of Gastroenterology* **16,** 1465–1472 (2010).

57. Smyth, G. K. in *Bioinformatics and Computational Biology Solutions using R and Bioconductor* (Gentleman, R., Irizarry, W. & Huber, W.) 397–420 (Springer, 2005).

58. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics (Oxford, England)* **8,** 118–27 (2007).

59. Parks, B. W. *et al.* Genetic control of obesity and gut microbiota composition in response to high-fat, high-sucrose diet in mice. *Cell metabolism* **17,** 141–52 (2013).

60. Reshef, D. N. *et al.* Detecting Novel Associations in Large Data Sets. *Science* **334,** 1518–1524 (2011).

61. Parkkinen, J. a & Kaski, S. Searching for functional gene modules with interaction component models. *BMC systems biology* **4,** 4 (2010).

62. Chessel, D., Dufour, A. B. & Thioulouse, J. The ade4 package - I : One-table methods. **4,** 5–10 (2004).

Chapter 6

Conclusions and Implications

for Future Studies

I have always been interested in solving complex, open-ended problems. My first exposure to a systems-level biology problem occurred during my undergraduate education when my summer mentor handed me the first microarrays she, or anyone else at my college, had ever performed and told me to figure out how to analyze it and tell her what it meant. That summer experience and other experiences after it ignited a passion for the fields of genetics and genomics and how mathematics and computer science can help researchers determine what regulates any number of biological processes. My dissertation research has focused on genome wide association studies (GWAS) and network biology, both exciting advances in genome research which leverage the notion that simply by observing phenotypes, genotypes and gene expression over a large number of individuals it is possible to specifically identify the genes which may underlie any number of conditions. As I quickly came to realize, however, the process of actually determining what phenotypes to query and how to identify the relevant genes or variations underlying a GWAS locus was not nearly so simple.

Advances in the past decade have allowed researchers unprecedented insight into the genetic mechanisms that underlie complex phenotypes such as heart failure(HF). Despite these advances and the identification of causal genes for hundreds of different disease traits[1], human studies have been stymied by ethical and technical issues, particularly in the study of diseases of the elderly such as HF. This has resulted in very few novel genes[2–5] relevant to HF being identified using GWAS or network biology in human studies. These results prompted our lab to look to model organisms as a potential means of analyzing HF without the limitations associated with human research.

The Hybrid Mouse Diversity Panel

In Chapter 2, I described the resource that our lab has developed for the study of complex traits in mice: the Hybrid Mouse Diversity Panel (HMDP). The HMDP offers several major advantages over other means of performing systems-level studies. In contrast with human studies, the HMDP consists of model organisms, and we are able to carefully control the environment, induce responses with physical and pharmacological interventions, collect tissues and perform invasive analyses which would not be feasible in human populations. Compared to other systems-level methods in mice, a major advantage of the HMDP is that it uses inbred lines of mice as opposed to F2 mice. This allows researchers to examine a phenotype in multiple mice of a given strain, further reducing the noise introduced by environmental confounders. It also offers researchers the ability to do true case/control studies where the treated and untreated animals are genetically identical to one another, an experiment which is impossible in non-inbred resources. Another advantage over previous systems is the high resolution we observe in the HMDP, which is often an order of magnitude improved compared to prior QTL studies in mice[6,7] We have previously shown that the HMDP can be used to study a number of clinical traits relevant to humans including atherosclerosis[7], diet-induced obesity[8], bone mineral density[9], and others. We now add heart failure to this list, using a case-control model that would be incredibly challenging to perform without the HMDP resource.

Genome Wide Association Studies for Heart Failure in Mice

Chapter 3 describes the results of a GWAS performed on the HMDP to identify genes involved in HF. Unlike a number of other common disorders, GWAS analyses in humans for HF have resulted in only modest success, returning only two loci which reached the genome-wide significance threshold of 1E-8 despite studies involving tens of thousands of individuals[3]. We set out to examine whether we could identify candidate genes for HF using the HMDP. We stimulated the beta adrenergic pathway with isoproterenol (ISO), a synthetic non-specific β-adrenergic agonist for three weeks. Similarly to previous studies using the HMDP, we observed significant variation in the phenotypes we collected. Furthermore, we noted that some strains appeared to show a much greater response to ISO than others. After surveying over 700 mice from 105 strains, we identified 19 genome-wide significant loci for cardiac hypertrophy or surrogate measures of HF. Several of these loci corresponded to well-studied genes known to be involved in HF, including *Ppp3ca*[10], *Sgcd*[11] and *Pln*[12,13], although our study represents the first time any of these genes have been identified in a HF GWAS. Additional genes identified in our study have previously been described as implicated in rare mendelian forms of cardiomyopathy[14,15], but natural variation in the expression of these genes have not been linked to common forms of HF. Still other genes identified by our study have never been implicated in HF. One of these genes, the long non-coding RNA *Miat,* was examined *in vitro*. We observe in isolated cardiomyocytes that knockdown of this gene using siRNA attenuates the hypertrophic response of the cardiomyocytes to catecholamine stimulation.

The identification of loci for HF using the HMDP represents only the beginning of the analysis necessary to identify the underlying genetics within each locus. Although we have putatively identified candidate genes at each locus based on a combination of literature searches, *cis*-eQTLs, correlations of gene expression to phenotypes and identification of non-synonymous SNPs present within the locus, there were often several genes which met at least one of these criteria. Future research should focus on each of these loci in turn, systematically progressing through candidates until one (or more) genes which influence a HF-related phenotype are indentified in each. We have currently begun this process for several loci, beginning with *in vitro* analyses similar to that performed on *Miat.* We have also begun to generate knockout and transgenic lines in mice and/or morpholino knockouts in zebrafish to further confirm a role for these genes in HF. Eventually, research will focus on the mechanisms by which each HF-related gene identified in this study drives its related HF trait, leading to greater understanding of the disease and possible treatment options.

Genome Wide Association of Cardiac Fibrosis

Chapter 3 also describes a GWAS on cardiac fibrosis that was carried out in parallel with the work on cardiac hypertrophy. Cardiac fibrosis is an incredibly difficult phenotype to analyze in humans and is usually only examined post-mortem, making an animal model such as the HF HMDP the preferred resource for studying this trait on a population-based level. To our surprise, no QTL studies for cardiac fibrosis have been reported in either mouse or rat[16], making our study the first to report GWAS loci for

cardiac fibrosis. Several of our loci contain interesting candidate genes, including *Snrpn*, a gene which resides within the Praeder-Willi imprinting locus and which demonstrates altered allele-specific expression in non-ischemic heart failure[17]. We also identified a novel role in ISO-induced fibrosis for *Abcc6*, a gene which causes the human disorder pseudoxanthoma elasticum[18,19] We observed *in vivo* using both *Abcc6* knockouts and transgenics that knockout or rescue of the gene has a direct effect on cardiac fibrosis after catecholamine stimulation, proving that it is a causal gene at its locus.

Similar to the work that is necessary for the HF phenotypes, additional focus on the candidate genes within the fibrosis loci are necessary using *in vitro* models such as cardiomyocyte cell culture as well as *in vivo* models such as morpholino zebrafish or knockout/transgenic mice. In addition to further analyzing the cardiac fibrosis loci, there are a number of other HF-related phenotypes in the HMDP HF study which have not yet been carefully analyzed. These include traits relating to changes in lipid levels in the panel after ISO stimulation, changes to cardiomyocyte cell size and changes to the weight of the adrenal glands. Perhaps our most intriguing trait for further research is an ISO-induced, strain-specific sudden-death response for which we have identified a single significant locus and identified a possible candidate gene, *Zfp277*.

MICA: A Novel Method for Network Analysis

In chapter 4, I describe a new method for examining gene networks called Maximal Information Component Analysis (MICA). MICA attempts to address two erroneous assumptions made by many commonly used network analysis methods. The

first improvement it offers over other methods is that it will consider nonlinear interactions between genes through the use of maximal information rather than Pearson's correlation. Even a simple two-gene system where the genes act to repress one another in a negative feedback loop results in a nonlinear relationship between these two genes. This nonlinearity cannot be recovered using Pearson's correlation, but is observable with maximal information. The addition of additional genes to the system, not to mention experimental and technical noise, would only add to this nonlinearity. Using a measure that can incorporate noise and nonlinearity into the system makes the resulting edges of the network graph more accurate reflect the true underlying genetic network. MICA's second improvement is the use of component modeling to allow genes to belong to more than a single module. There are many examples of genes, such as *Akt*, which play multiple roles in the cell[20–22]. By forcing genes to belong to a single module, other methods distort the modules and any summary statistics based on which genes belong to which module by not allowing all genes to belong to all modules with which they interact. We have demonstrated on several datasets from the HMDP that MICA either matches or outperforms WGCNA, a well-regarded network analysis method. We observe improvements specifically in the case of case/control studies, such as the HF HMDP[23].

Despite these advances over other network construction algorithms, there are several aspects of MICA which could benefit from continued development. One possible improvement would be to incorporate the strength of the interaction between two genes into the analysis to better inform the module detection process. This incorporation of

interaction strength or edge weight has been demonstrated to improve the results of other methods[24], leading us to hypothesize that the development of a weighted version of MICA might do the same and lead to more accurate analysis of gene networks. Furthermore, it is important to develop MICA from a set of 'add-on' functions to WGCNA, a common and well-regarded analysis method, into a standalone R Package. This will improve the usefulness of the method to the scientific community as a whole.

MICA Applied to the HF HMDP Study

Chapter 5 describes the application of MICA, developed in chapter 4, to the HF HMDP study described in chapter 3. It also describes several overall features of the transcriptome data from the project using eQTL hotspot analysis. Understanding the underlying architecture of the genetic machinery that leads some individuals to be predisposed to HF and others to be resistant is particularly important to future studies of the disease. We first looked at genome-wide regulation of gene expression by using eQTLs to identify 'hotspots' where a single locus controls the expression of a significant fraction of expressed genes. We observed several *trans*-eQTL hotspots in control mice, including three loci, each significantly regulating over 10% of the expressed genes in the heart. We observed few eQTL hotspots in the ISO-treated gene expression arrays, possibly due to the substantial disruption that ISO represents to normal heart signaling. We did observe several large eQTL hotspots when analyzing the change of gene expression, suggesting that the changes to gene expression after ISO are regulated by a few key regions of the genome. Our data implicates the miRNA editing pathway as

being very important to the overall regulation of gene expression in the heart, and suggests as well that *Serpina3n* plays an important role in the response of the heart to ISO stimulation. Application of MICA to the HF transcriptomes revealed several modules which are highly correlated with HF-relevant phenotypes, including one module of about 40 genes. Using mathematical modeling techniques and eQTL analysis, we have developed a proposed model for the underlying structure of this module. Our model suggests that *Adamts2* and *Serpina3n* play important roles in the overall functionality of the module, and presumably play a role in the cardiac remodeling response to ISO stimulation as well. eQTL hotspot analysis on module 5 has revealed *Magi2,* a gene which regulates the turnover of the beta1 adrenergic receptor in the brain[25], appears to be a master regulator of module 5, acting perhaps through *Pdgfrl*.

Like a GWAS, the identification of a gene module which is related to a given phenotype is only the first step in a process that culminates in the validation and analysis of individual genes which influence that phenotype. Gene modules must be validated, regulators confirmed and analysis of the effects of altering those regulators examined. For the analysis of module 5, we are interested in using *in vitro* systems to first query the role of in-module regulators such as *Adamts2* and *Serpina3n* in the module. If we knockdown gene expression of these two regulators, will we disrupt the entire module structure, and possibly abrogate the ISO response? We will then explore *in vivo* the role of these genes and *Magi2*, which is not expressed in myocytes. We have also identified other modules which show significant correlation to HF-related phenotypes in the

network that have not been analyzed in detail. These should be carefully examined in the same manner as the module containing the genes described above.

Concluding Remarks

Heart failure is a complex disorder that is difficult to treat in humans, in part due to the large number of etiologies which can give rise to the syndrome. Attempts to identify genome-wide significant regulators of HF using systems-level analyses are easily confounded when there is a lack of understanding as to the origins and exacerbating factors of each patient's HF. Preliminary analysis of some of our candidate genes and pathways in mice with HF induced through other methods, such as angiotensin stimulation, which activates the alpha-adrenergic receptors or trans-aortic constriction, which is a purely physical inhibition of blood flow suggest that while gross morphological parameters such as change in heart weight (hypertrophy) after treatment remain similar between these models, some of the genes implicated in ISO mice are influenced in a very different manner by angiotensin or TAC challenge. In addition to underlying changes in the genes leading to HF, human studies are further complicated by additional variations between individuals in terms of diet, exercise, climate and other personal choices made over the course of a subject's life. The combination of the activation of different gene pathways for different etiologies of the disease, and the strong influence of environmental factors on HF phenotypes means that HF proves to be a difficult disease to examine in humans.

In this dissertation, I have demonstrated that it is possible to examine HF in mice using the HMDP by carefully controlling the means by which HF is induced and controlling the environment in which the animals live. This study resulted in the identification of nearly twenty loci for HF and cardiac hypertrophy, including several novel genes that appear to regulate HF *in vitro.* I was also able to perform, to our knowledge, the first ever GWAS analysis of cardiac fibrosis in humans, mice or rats. This study identified over a dozen candidates for the control of cardiac fibrosis, including *Abcc6,* a gene whose role in ISO-induced fibrosis was confirmed using *in vivo* models. Additionally, through the use of a novel network analysis method, I was able to identify a small module of ~40 genes which appears to play an important role in the control of HF phenotypes as well as identify possible regulators of both that module as well as the cardiac response to ISO stimulation as a whole.

Future research on the results of this dissertation will focus on the confirmation of the work already performed, but will also seek to examine additional phenotypes of this disorder using the HMDP and increasingly sophisticated analysis methods. One of the great benefits of the HMDP is its ability to allow comparisons between different levels of regulation in the organism through repeated querying of mice from the same strains. In this study we have examined the genome (DNA), transcriptome (mRNA) and portions of the phenome (phenotypes) as well as the interaction between these three levels using GWAS and network analysis. There are many additional " 'omes " to be analyzed, including the spliceosome (RNA), miRNAome (miRNA), proteome (proteins), interactome (proteins) and metabolome (circulating factors in the blood). By further

analysis using these additional 'omes and the interactions between them, HF studies using the HMDP will yield additional insights into the disease and will lead to large improvements in the treatment of this deadly syndrome.

References

1. Hindorff, L. *et al.* A Catalog of Published Genome-Wide Association Studies. at < www.genome.gov/gwastudies>

2. Parsa, A. *et al.* Hypertrophy-associated polymorphisms ascertained in a founder cohort applied to heart failure risk and mortality. *Clinical and translational science* **4,** 17–23 (2011).

3. Villard, E. *et al.* A genome-wide association study identifies two loci associated with heart failure due to dilated cardiomyopathy. *European heart journal* **32,** 1065–76 (2011).

4. Ellinor, P. T. *et al.* A novel locus for dilated cardiomyopathy, diffuse myocardial fibrosis, and sudden death on chromosome 10q25-26. *Journal of the American College of Cardiology* **48,** 106–11 (2006).

5. Morrison, A. C. *et al.* Genomic variation associated with mortality among adults of European and African ancestry with heart failure: the cohorts for heart and aging research in genomic epidemiology consortium. *Circulation. Cardiovascular genetics* **3,** 248–55 (2010).

6. Ghazalpour, A. *et al.* Hybrid mouse diversity panel: a panel of inbred mouse strains suitable for analysis of complex genetic traits. *Mammalian genome : official journal of the International Mammalian Genome Society* **23,** 680–92 (2012).

7. Bennett, B. J. *et al.* A high-resolution association mapping panel for the dissection of complex traits in mice. *Genome research* **20,** 281–90 (2010).

8. Parks, B. W. *et al.* Genetic control of obesity and gut microbiota composition in response to high-fat, high-sucrose diet in mice. *Cell metabolism* **17,** 141–52 (2013).

9. Farber, C. R. *et al.* Mouse genome-wide association and systems genetics identify Asxl2 as a regulator of bone mineral density and osteoclastogenesis. *PLoS genetics* **7,** e1002038 (2011).

10. Lu, M. *et al.* The function of calcineurin and ERK1/2 signal in the antihypertrophic effects of kappa-opioid receptor stimulation on myocardial hypertrophy induced by isoprenaline. *Die Pharmazie* **67,** 182–6 (2012).

11. Goehringer, C. *et al.* Prevention of cardiomyopathy in delta-sarcoglycan knockout mice after systemic transfer of targeted adeno-associated viral vectors. *Cardiovascular research* **82,** 404–10 (2009).

12. Chu, G. & Kranias, E. G. Phospholamban as a therapeutic modality in heart failure. *Novartis Foundation symposium* **274,** 156–71; discussion 172–5, 272–6 (2006).

13. Eijgelsheim, M. *et al.* Genome-wide association analysis identifies multiple loci related to resting heart rate. *Human molecular genetics* **19,** 3885–94 (2010).

14. Arad, M. *et al.* Constitutively active AMP kinase mutations cause glycogen storage disease mimicking hypertrophic cardiomyopathy. *Journal of Clinical Investigation* **109,** 357–362 (2002).

15. Pall, G. S. *et al.* A novel transmembrane MSP-containing protein that plays a role in right ventricle development. *Genomics* **84,** 1051–9 (2004).

16. Laulederkind, S. J. F. *et al.* The Rat Genome Database 2013--data, tools and users. *Briefings in bioinformatics* **14,** 520–6 (2013).

17. Fermin, D. R. *et al.* Sex and age dimorphism of myocardial gene expression in nonischemic human heart failure. *Circulation. Cardiovascular genetics* **1,** 117–25 (2008).

18. Finger, R. P. *et al.* Pseudoxanthoma elasticum: genetics, clinical manifestations and therapeutic approaches. *Survey of ophthalmology* **54,** 272–85

19. Meng, H. *et al.* Identification of Abcc6 as the major causal gene for dystrophic cardiac calcification in mice through integrative genomics. *Proceedings of the National Academy of Sciences of the United States of America* **104,** 4530–5 (2007).

20. Beaulieu, J.-M. *et al.* A beta-arrestin 2 signaling complex mediates lithium action on behavior. *Cell* **132,** 125–36 (2008).

21. Winbanks, C. E. *et al.* Follistatin-mediated skeletal muscle hypertrophy is regulated by Smad3 and mTOR independently of myostatin. *The Journal of cell biology* **197,** 997–1008 (2012).

22. Banerjee, S. K., McGaffin, K. R., Huang, X. N. & Ahmad, F. Activation of cardiac hypertrophic signaling pathways in a transgenic mouse with the human PRKAG2 Thr400Asn mutation. *Biochimica et biophysica acta* **1802,** 284–91 (2010).

23. Rau, C. D. *et al.* Maximal information component analysis: a novel non-linear network analysis method. *Frontiers in genetics* **4,** 28 (2013).

24. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Statistical applications in genetics and molecular biology* **4,** Article17 (2005).

25. Xu, J. *et al.* beta 1-adrenergic receptor association with the synaptic scaffolding protein membrane-associated guanylate kinase inverted-2 (MAGI-2). Differential regulation of receptor internalization by MAGI-2 and PSD-95. *The Journal of biological chemistry* **276,** 41310–7 (2001).