UNIVERSITY OF CALIFORNIA

Los Angeles

Heart Failure Genetics in Mice and Men

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Human Genetics

by

Jessica Jen-Chu Wang

2014

ABSTRACT OF THE DISSERTATION

Heart Failure Genetics in Mice and Men

by

Jessica Jen-Chu Wang

Doctor of Philosophy in Human Genetics

University of California, Los Angeles, 2014

Professor Aldons Jake Lusis, Chair

The genetics of heart failure is complex. In familial cases of cardiomyopathy, where mutations of large effects predominate in theory, genetic testing using a gene panel of up to 76 genes returned negative results in about half of the cases. In common forms of heart failure (HF), where a large number of genes with small to modest effects are expected to modify disease, only a few candidate genomic loci have been identified through genome-wide association (GWA) analyses in humans. We aimed to use exome sequencing to rapidly identify rare causal mutations in familial cardiomyopathy cases and effectively filter and classify the variants based on family pedigree and family member samples. We identified a number of existing variants and novel genes with great potential to be disease causing. On the other hand, we also set out to understand genetic factors that predispose to common late-onset forms of heart failure by performing GWA in the isoproterenol-induced HF model across the Hybrid Mouse Diversity Panel (HMDP) of 105 strains of mice. We performed fine phenotyping using serial echocardiograms and controlled for environmental heterogeneity in the

experimental setting. As a result, we achieved high heritability estimates of 64% to 84% for all cardiac traits and had superior power for mapping HF-related trait, compared to human studies. Association analyses of cardiac traits, corrected for population structure and multiple comparisons, revealed genome-wide significant loci across the spectrum of cardiac traits. Cardiac tissue gene expression profiling, expression quantitative trait loci, expression-phenotype correlation, and coding sequence variation analyses were performed to prioritize candidate genes and to generate hypotheses for downstream mechanistic studies. Future directions will be to further bridge the gap of understanding between rare Mendelian and common cardiovascular diseases, which together will have wide spread therapeutic implications in delaying or reversing HF progression in human populations.

The dissertation of Jessica Jen-Chu Wang is approved.

Rita Cantor

Karen Reue

Yibin Wang

Aldons Jake Lusis, Committee Chair

University of California, Los Angeles

2014

Dedication

To my mother, Mei-Tsu, who has been a source of encouragement and inspiration to me throughout my life. To my father, Ta-Jan, who has been a steady moral compass and provider to help me find and realize my potentials. To my sisters Sandra and Karen, for being so supportive. To my dear husband, Richard, for always cheering me on. Most of all thanks to God who continues to make the impossible possible.

Table of Contents

Table of Tables

Table of Figures

Acknowledgement

The author is greatly indebted to members of her thesis and advisory committees. The author wishes to acknowledge ongoing support and guidance from the research faculty members: Professors Aldons Jake Lusis, Yibin Wang, Thomas Vondriska, James Weiss, Karen Reue, and Rita Cantor; and clinical faculty members: Arnold Baas, Eugene DePasquale, Tamara Horwich, Ali Nsair, Will Suh, Martin Cadeiras, and Mario Deng.

The author also wishes to acknowledge contributions from: Cheri Silverstein, Hane Lee, Michael Fishbein, Melissa Spencer, Ekaterina Mokhonova, Stanley Nelson (Chapter 2); Rozeta Avetisyan, Christoph Rau (Chapter 3); Milagros Romay, Richard Davis (Chapter 4).

This research was supported in part by the National Heart, Lung, and Blood Institute (NHLBI) of the National Institute of Health (NIH) under grants HL123295, HL110667 and HL28481 to Professor Aldons Jake Lusis and HL114437, HL123295, HL110667 to Professor Yibin Wang. The author was supported the NIH training fellowship T32HL007895. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institute of Health. In addition, the author was supported by the Specialty Training and Advanced Research (STAR) program at the David Geffen School of Medicine at UCLA, the Tibor Fabian Research Award, the Department of Medicine at UCLA.

Vita

Education

1996-2000	Massachusetts Institute of Technology	BS in Biology
2000-2004	University of Pennsylvania	MD

Training

2004-2007	UT Southwestern Medical Center	Internal Medicine
2007-2008	Baylor University Medical Center	Hospitalist
2008-2013	University of California, Los Angeles	Cardiovascular Disease

Employment

2013-2014	University of California, Los Angeles	Clinical Instructor in Medicine

Publications

Hendren SK, **Wang J**, Gorman J, Peacock T, Hershock DM, Rosato EF. Esophagectomy and splenectomy in a patient with osteopetrosis. *J Thorac Cardiovasc Surg*. 2005;129:1457-1458

Wang JJ, Reimold SC. Chest pain resulting from histoplasmosis pericarditis: A brief report and review of the literature. *Cardiol Rev*. 2006;14:223-226

Sadek H, Hannack B, Choe E, **Wang J**, Latif S, Garry MG, Garry DJ, Longgood J,

Frantz DE, Olson EN, Hsieh J, Schneider JW. Cardiogenic small molecules that enhance myocardial repair by stem cells. *Proc Natl Acad Sci U S A*. 2008;105:6063-6068

Ghazalpour A, Rau CD, Farber CR, Bennett BJ, Orozco LD, van Nas A, Pan C, Allayee H, Beaven SW, Civelek M, Davis RC, Drake TA, Friedman RA, Furlotte N, Hui ST, Jentsch JD, Kostem E, Kang HM, Kang EY, Joo JW, Korshunov VA, Laughlin RE, Martin LJ, Ohmen JD, Parks BW, Pellegrini M, Reue K, Smith DJ, Tetradis S, **Wang J**, Wang Y, Weiss JN, Kirchgessner T, Gargalovic PS, Eskin E, Lusis AJ, LeBoeuf RC. Hybrid mouse diversity panel: A panel of inbred mouse strains suitable for analysis of complex genetic traits. *Mamm Genome*. 2012;23:680-692

Weiss JN, Karma A, MacLellan WR, Deng M, Rau CD, Rees CM, **Wang J**, Wisniewski N, Eskin E, Horvath S, Qu Z, Wang Y, Lusis AJ. "Good enough solutions" and the genetics of complex diseases. *Circ Res*. 2012;111:493-504

1 Introduction

Heart failure (HF) is a complex clinical syndrome that results from impaired ventricular filling or ejection of blood. Its cardinal features are dyspnea, fatigue, and fluid retention. HF may be associated with a wide spectrum of left ventricular (LV) functional abnormalities, ranging from normal LV size and preserved ejection fraction (EF) to severe dilatation and/or markedly reduced EF. HF with reduced EF (HFrEF) is defined as the clinical diagnosis of HF and EF≤40%. HF with preserved EF (HFpEF) is defined as the clinical diagnosis of HF, EF>40%, and abnormal LV diastolic dysfunction, after the exclusion of non-cardiac causes[1, 2]. HF*r*EF and HF*p*EF each make up about half of the overall HF burden[3]. Risk factors for HFrEF include a family history of cardiomyopathy, long-standing hypertension, previous myocardial infarction (MI), cardiotoxic agents, valvular abnormalities, congenital heart lesions, arrhythmias, and others. Current therapies for HFrEF range from prevention, non-pharmacological interventions, pharmacological treatment, device therapy, to inotropic support, mechanical circulatory support and cardiac transplantation at the end-stage. Risk factors for HFpEF include hypertension, obesity, coronary artery disease (CAD), diabetes mellitus, atrial fibrillation (AF), and hyperlipidemia[1]. Effective pharmacological treatments for HFrEF have generally been disappointing when used in patients with HFpEF[4]. The lifetime risk of developing HF is 20% for Americans ≥40 years of age. The total cost of HF care in the United States exceeds $30 billion annually, with over half of these costs spent on hospitalizations[5]. HF is the primary diagnosis in >1 million hospitalizations annually. A subset of patients with chronic HF will continue to progress

and develop persistently severe symptoms that are refractory to maximum therapy. The absolute mortality rates for HF remain approximately 50% within 5 years of diagnosis[1].

In the past 25 years, through linkage studies of familial cardiomyopathies and arrhythmias, a number of genes associated with early-onset HF have been identified, each harboring unique amino acid substitution, splice site, and truncation mutations[6-9]. Definitive causal link of some of these mutations to disease have been established in mouse models[10-12]. Allele-specific silencing of the mutant allele demonstrated therapeutic benefit[13]. In the past decade, next-generation sequencing has enabled the massive and parallel sequencing of many genes at once, which brought the cost of sequencing down to a level similar to other routine medical tests. Genetic testing for many disorders can now be ordered commercially from any medical office outside of the academic setting, including gene panels for specific forms of familial cardiomyopathies and arrhythmias. Because familial cardiomyopathies and arrhythmias are clinically overlapping and genetically heterogeneous, comprehensive cardiomyopathy and arrhythmia panel tests have also been made available. Genetic test panels are designed to effectively and accurately capture sequence variants in the protein-coding regions of genes known to be associated with disease. A more global approach is whole exome sequencing, where in theory all protein-coding variants of the genome can be evaluated. Genetic polymorphisms are frequently observed in any given exome. Our results showed that each exome contains approximately 22,000 single nucleotide polymorphisms (SNPs) and 1-10 bp insertion-deletions (indels) (Table 2-3). In other words, one variant would be found in each gene sequenced on average. In the absence

of family member samples, prioritization of variants for disease significance generally relies on the following criteria: 1) previous report in literature, 2) allele frequency in the reference population, and 3) *in silico* pathogenicity algorithms, based on the following assumptions: 1) causal mutation lies in protein coding regions of the genes or exome, 2) mutations reported in literature are unbiased and truly causal, 3) causal mutations are rare, 4) the referenced controls are representative of the sequenced individual's genetic background, and 5) *in silico* pathogenicity algorithms give biologically meaningful predictions. In practice, the assignment of variant classes as disease-causing, likely disease-causing, and of unknown significance is complex for both gene panel and whole exome testing. As technologies and analytical methods for genome sequencing continue to advance, current limitations in the assumptions will likely be overcome and variant interpretation will improve.

The majority of HF encountered in the community is not explained by a single gene mutation. Rather, there is a complex interplay between genetic susceptibility and environmental factors that results in the phenotype of clinical HF and LV structural or functional abnormalities that are known to predispose to HF. Genome-wide association (GWA) of common genetic variants offers an alternative, high-throughput and unbiased approach to identify risk loci contributing to HF. A number of HF-related GWA has been performed to date, each revealing a small number of loci and candidate genes. For example, two SNPs, one at 58.8kb from USP3 (ubiquitin-specific protease) gene in individuals of European ancestry and the other at 6.3 kb from the LRIG3 (leucine-rich, immunoglobulin like domain) gene in individuals of African ancestry, have been

associated with the development of HF[14]. A third SNP located in the intron of CMTM7 (CKLF-like MARVEL transmembrane domain containing 7) gene among individuals of European ancestry was associated with HF mortality[15]. In a European consortium study on DCM, two SNPs were associated with sporadic DCM and replicated in independent samples. One of these SNPs was located near HSPB7 (heat shock 27kDa protein family, member 7). Another was a non-synonymous SNP located in BAG3 (BCL2-associated athanogene 3). Sequencing of BAG3 exons in 168 independent index cases with familial DCM identified four heterozygous truncating and two missense mutations, which were present in all genotyped relatives affected by disease and absent in a control group of 347 healthy individuals[16]. Lastly, GWA of cardiac structure and systolic function among African Americans associated 4 genetic loci near UBE2V2 (ubiquitin-conjugating enzyme E2 variant 2), WIPI1 (WD repeat domain, phosphoinositide interacting 1), PPAPDC1A (phosphatidic acid phosphatase type 2 domain containing 1A) and KLF5 (Kruppel-like factor 5) with left ventricular mass, interventricular septal wall thickness, left ventricular internal diameter and ejection fraction, respectively

Base on the Framingham Heart Study, the genetic basis of HF in human is modest. The hazard ratio for HF in those with parental HF was 1.70 [95% confidence interval (CI) 1.11-2.60][17] and the estimated heritability of LV mass (LVM) was 0.24-0.32[18]. One twin study demonstrated interaclass correlation coefficients of 0.69 and 0.32 for monozygotic and dizygotic twins, respectively[19], consistent with significant environmental influences on HF traits, thus limiting the power of signal detection in a GWA. Complementary to the studies in humans, we conducted a large-scale survey of

cardiac structure and function in 105 inbred mouse strains, termed the Hybrid Mouse Diversity Panel (HMDP), at baseline and in each subsequent week during a 3-week isoproterenol infusion to investigate genomic loci controlling variations in traits relevant to HF. We observed striking variations among the mouse strains and deteremined that cardiac structure and function were highly heritable at baseline and in response to isoproterenol, with heritability estimates between 0.64 and 0.84. Using a linear mixed model algorithm, which accounts for relatedness among the strains, GWA was performed on the observed traits. A number of loci contributing to heart weight, fibrosis, cardiac structure and function were identified. We prioritized genes in each associated genomic locus using expression quantitative trait locus (eQTL) analysis and coding sequence variations in the Wellcome Trust Mouse Genomes Project. We validated one of the association loci candidate genes *Abcc6* as causal for the variations in isoproterenol-induced fibrosis. We were able to identify more significantly associated loci than in the human studies, because we were able to fully control for environmental confounders in the laboratory setting, which contributed to improved heritability estimates and significantly improved mapping power.

Through gene panel testing and exome sequencing in human subjects and HF GWA in mice, we have implicated a number of novel genes and genetic variants in HF-related traits, which need to be replicated or validated experimentally, to establish definitive causal role in disease. Given the number of genes and variants to test, a high-throughput pipeline is required to screen associated disease genes and variants experimentally. We have explored two high-throughput model systems for gene

validation thus far, one using neonatal rat ventricular cardiomyocytes (NRVMs) and one using zebrafish. Additional gene validation technique include adeno-associated virus 9 (AAV9) to examine the role of the candidate gene in the mouse heart. To validate genetic variants, we plan to use the clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 system to generate targeted coding sequence variants in zebrafish and mice. Finally, we will also explore the role of allele-specific silencing of disease-causing mutation using RNA interference and CRISPR to be carried out in patient-derived induced pluripotent stem (iPS) cells. Once the causal role of a gene or genetic variant is establish, genetically stable mouse models will be generated for further functional and downstream mechanistic studies.

The results from our GWA and those of others showed that the majority of the genomic regions controlling susceptibility to HF are located in intergenic regions, which highlights the importance of gene regulation rather than protein-coding sequence variations in the mechanisms affecting complex traits such as HF. A number of mechanisms have been proposed to link non-protein-coding sequence variations to gene regulation, including promoter region variations, methylation changes, non-coding RNA, and altered chromatin states among others. Using Reduced Representation Bisulfite Sequencing (RRBS) we studied the methylation patterns in the left ventricular tissues from 2 mouse strains, susceptible and resistant to ventricular dilation and decreased ejection fraction after isoproterenol treatment. We found that the differences in methylation patterns were greater between strains than between treatment groups. We plan to examine the role of methylation patterns and their relationships to SNP

variations and HF traits across the HMDP. In the coming years, the use of fine and deep phenotyping, large-scale whole genome sequencing, and evermore sophisticated mapping strategies, will work together to improve our current understanding of heart health genetics even more. This new understanding will translate into preventive and treatment options for those who have no effective therapies currently, such as familial cardiomyopathy and HFpEF.

Bibliography

1. Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE, Jr., Drazner MH, Fonarow GC, Geraci SA, Horwich T, Januzzi JL, Johnson MR, Kasper EK, Levy WC, Masoudi FA, McBride PE, McMurray JJ, Mitchell JE, Peterson PN, Riegel B, Sam F, Stevenson LW, Tang WH, Tsai EJ, Wilkoff BL. 2013 accf/aha guideline for the management of heart failure: A report of the american college of cardiology foundation/american heart association task force on practice guidelines. *J Am Coll Cardiol*. 2013;62:e147-239

2. Garcia-Pavia P, Cobo-Marcos M, Guzzo-Merello G, Gomez-Bueno M, Bornstein B, Lara-Pezzi E, Segovia J, Alonso-Pulpon L. Genetics in dilated cardiomyopathy. *Biomark Med*. 2013;7:517-533

3. Owan TE, Redfield MM. Epidemiology of diastolic heart failure. *Prog Cardiovasc Dis*. 2005;47:320-332

4. Edelmann F, Wachter R, Schmidt AG, Kraigher-Krainer E, Colantonio C, Kamke W, Duvinage A, Stahrenberg R, Durstewitz K, Loffler M, Dungen HD, Tschope C, Herrmann-Lingen C, Halle M, Hasenfuss G, Gelbrich G, Pieske B. Effect of spironolactone on diastolic function and exercise capacity in patients with heart failure with preserved ejection fraction: The aldo-dhf randomized controlled trial. *JAMA*. 2013;309:781-791

5. Go AS, Mozaffarian D, Roger VL, Benjamin EJ, Berry JD, Borden WB, Bravata DM, Dai S, Ford ES, Fox CS, Franco S, Fullerton HJ, Gillespie C, Hailpern SM, Heit JA, Howard VJ, Huffman MD, Kissela BM, Kittner SJ, Lackland DT, Lichtman JH, Lisabeth LD, Magid D, Marcus GM, Marelli A, Matchar DB, McGuire DK, Mohler ER, Moy CS, Mussolino ME, Nichol G, Paynter NP, Schreiner PJ, Sorlie PD, Stein J, Turan TN, Virani SS, Wong ND, Woo D, Turner MB. Heart disease and stroke statistics--2013 update: A report from the american heart association. *Circulation*. 2013;127:e6-e245

6. Marian AJ, Roberts R. Molecular genetics of hypertrophic cardiomyopathy. *Annu Rev Med*. 1995;46:213-222

7. Maron BJ, Maron MS. Hypertrophic cardiomyopathy. *Lancet*. 2013;381:242-255

8. Callis TE, Jensen BC, Weck KE, Willis MS. Evolving molecular diagnostics for familial cardiomyopathies: At the heart of it all. *Expert Rev Mol Diagn*. 2010;10:329-351

9. Nava A, Bauce B, Basso C, Muriago M, Rampazzo A, Villanova C, Daliento L, Buja G, Corrado D, Danieli GA, Thiene G. Clinical profile and long-term follow-up of 37 families with arrhythmogenic right ventricular cardiomyopathy. *J Am Coll Cardiol*. 2000;36:2226-2233

10. Barefield D, Kumar M, de Tombe PP, Sadayappan S. Contractile dysfunction in a mouse model expressing a heterozygous mybpc3 mutation associated with hypertrophic cardiomyopathy. *Am J Physiol Heart Circ Physiol*. 2014;306:H807-815

11. Blankenburg R, Hackert K, Wurster S, Deenen R, Seidman JG, Seidman C, Lohse MJ, Schmitt JP. Beta-myosin heavy chain variant met606val causes very mild hypertrophic cardiomyopathy in mice, but exacerbates hcm phenotypes in mice carrying other hcm mutations. *Circ Res*. 2014

12. Tardiff JC, Hewett TE, Palmer BM, Olsson C, Factor SM, Moore RL, Robbins J, Leinwand LA. Cardiac troponin t mutations result in allele-specific phenotypes in a mouse model for hypertrophic cardiomyopathy. *J Clin Invest*. 1999;104:469-481

13. Jiang J, Wakimoto H, Seidman JG, Seidman CE. Allele-specific silencing of mutant myh6 transcripts in mice suppresses hypertrophic cardiomyopathy. *Science*. 2013;342:111-114

14. Smith NL, Felix JF, Morrison AC, Demissie S, Glazer NL, Loehr LR, Cupples LA, Dehghan A, Lumley T, Rosamond WD, Lieb W, Rivadeneira F, Bis JC, Folsom

AR, Benjamin E, Aulchenko YS, Haritunians T, Couper D, Murabito J, Wang YA, Stricker BH, Gottdiener JS, Chang PP, Wang TJ, Rice KM, Hofman A, Heckbert SR, Fox ER, O'Donnell CJ, Uitterlinden AG, Rotter JI, Willerson JT, Levy D, van Duijn CM, Psaty BM, Witteman JCM, Boerwinkle E, Vasan RS. 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*. 2010;3:256-266

15. Morrison AC, Felix JF, Cupples LA, Glazer NL, Loehr LR, Dehghan A, Demissie S, Bis JC, Rosamond WD, Aulchenko YS, Wang YA, Haritunians T, Folsom AR, Rivadeneira F, Benjamin EJ, Lumley T, Couper D, Stricker BH, O'Donnell CJ, Rice KM, Chang PP, Hofman A, Levy D, Rotter JI, Fox ER, Uitterlinden AG, Wang TJ, Psaty BM, Willerson JT, van Duijn CM, Boerwinkle E, Witteman JCM, Vasan RS, Smith NL. 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*. 2010;3:248-255

16. Villard E, Perret C, Gary F, Proust C, Dilanian G, Hengstenberg C, Ruppert V, Arbustini E, Wichter T, Germain M, Dubourg O, Tavazzi L, Aumont MC, DeGroote P, Fauchier L, Trochu JN, Gibelin P, Aupetit JF, Stark K, Erdmann J, Hetzer R, Roberts AM, Barton PJ, Regitz-Zagrosek V, Aslam U, Duboscq-Bidot L, Meyborg M, Maisch B, Madeira H, Waldenstrom A, Galve E, Cleland JG, Dorent R, Roizes G, Zeller T, Blankenberg S, Goodall AH, Cook S, Tregouet DA, Tiret L, Isnard R, Komajda M, Charron P, Cambien F. A genome-wide association study identifies two loci associated with heart failure due to dilated cardiomyopathy. *Eur Heart J*. 2011;32:1065-1076

17. Lee DS, Pencina MJ, Benjamin EJ, Wang TJ, Levy D, O'Donnell CJ, Nam B-H, Larson MG, D'Agostino RB, Vasan RS. Association of parental heart failure with

risk of heart failure in offspring. *New England Journal of Medicine*. 2006;355:138-147

18. Post WS, Larson MG, Myers RH, Galderisi M, Levy D. Heritability of left ventricular mass: The framingham heart study. *Hypertension*. 1997;30:1025-1028

19. Swan L, Birnie DH, Padmanabhan S, Inglis G, Connell JM, Hillis WS. The genetic determination of left ventricular mass in healthy adults. *Eur Heart J*. 2003;24:577-582

2 Exome Sequencing Identifies a Novel Candidate Gene *ZBTB17* for Arrhythmogenic Right Ventricular Cardiomyopathy.

Background

Family history is an important risk factor for heart disease, due to shared environment and genetic inheritance. Over the past 25 years, through family-based linkage analyses, an increasing number of genes have been implicated in inherited cardiomyopathies, arrhythmias and congenital heart disease[1]. In many cases, these genes harbor rare missense, frameshift and nonsense mutations in the protein-coding regions that are unique to individual families. Familial cardiomyopathies are clinically and genetically heterogenous. The correct identification of the disease-causing mutation is not only important for prognostication in family members but also critical in providing a novel entry point for disease mechanism investigation and therapeutic designs. The study of familial cardiomyopathies can reveal important insights in disease causation and open opportunities for novel therapies, as there is currently no effective treatments to reverse, stall, or prevent disease among patients and their family members.

In the last decade, the cost of sequencing has decreased dramatically. With increasing awareness of genomic data and acceptance by the public and insurance companies, genetic testing for familial diseases has moved into the clinic, most commonly in the form of commercial gene panel testing. The reported diagnostic yield

for HCM and DCM/LVNC gene panel testing were reported by commercial genetic testing company to be as high as 60% and 50%, respectively. Without detailed pedigree information and additional DNA samples from family member to demonstrate genetic segregation with disease, variants detected from commercial gene panel testing are routinely evaluated solely based on variants previously reported in literature, allele frequency in the reference population, and *in silico* predicted pathogenicity, all of which have potential pitfalls and drawbacks.

To address the bias inherent in gene panel testing, an alternative diagnostic approach to gene panel testing is available in the form of whole exome sequencing, where in theory all protein-coding variants of the genome can be evaluated at once for disease-causing mutations and novel gene and mutation discovery. A major challenge of using exome sequencing data is to find the novel disease genes among the background of rare non-pathogenic variants and sequencing errors. Exome sequencing identifies about 20,000 to 24,000 single nucleotide variants (SNVs) on average. While about 95% of the variants that are known common polymorphisms in the human population can be quickly excluded, strategies to further narrow down on the variant list are to: 1) Exclude variants outside of identity by descent (IBD) regions (if known for affected members of the family), 2) exclude variants conflicting with mode of inheritance, 3) exclude common variants present in the reference population, 4) exclude synonymous variants, 5) exclude variants predicted not to be deleterious, 6) select variants in genes which have physical protein-protein interaction, share same biological pathway, or are found in literature to be associated with other previously identified

genes of the disease. Among these some of the most effective strategies to narrow down on the list of candidate variants are based on evidence for shared ancestral polymorphisms and/or genetic linkage co-segregation. Other strategies include *in silico* prediction of pathogenicity similar to those employed in gene panel testing[2].

Traditional gene mapping approaches to gene discovery require the collection of families or large family pedigrees, in order to finely map the region of interest to a manageable size for targeted resequencing. One of the attractive features of exome sequencing in facilitating disease-causing gene or mutation discovery is that the exome, while constituting only approximately 1% of the human genome, is enriched with 85% of the mutations with large effects on disease-related traits. By employing filtering criteria based on mode of inheritance and allele frequency and additional genotyping information from select few family members, it is possible to quickly narrow in on the putative causal genetic variant[3].

The purpose of this study is 1) to estimate the rate and quality of genetic variant detection based on commercial gene panel testing sent from the UCLA Cardiomyopathy Center, 2) to identify the pitfalls and promises of genetic variant testing using whole exome sequencing, 3) to identify possible causal genes and mutations in a highly consanguineous family where a few family samples significantly improved variant filtering and variant classification.

Methods

Family pedigree construction. Familial cases of cardiomyopathy were collected from the Ahmanson-UCLA Cardiomyopathy Center. Health history, electrocardiogram, echocardiogram, stress test, and cardiac MRI results from the proband and extended family members (whenever available) were collected through detailed family interviews. Detailed family pedigrees were constructed to the extent possible. Genetic counseling was provided to the patients and their families prior to genetic testing.

Genetic testing strategies and sample collection. Patients with insurance coverage for genetic testing were referred to gene panel testing by the commercial company GeneDx. In addition, patients hospitalized at the Ronald Reagan UCLA Medical Center with a particularly strong family history were considered for UCLA Clinical Exome Sequencing testing. Follow-up patient family member saliva samples were collected via standard procedures with informed consent under the UCLA Office of the Human Research Protection Program Institutional Review Board for segregation analysis.

Gene Panel Testing. Genomic DNA was extracted from blood (2-5 mL in EDTA) using standard methods and sequencing was performed using a novel solid-state sequencing-by-synthesis process that allows sequencing a large number of amplicons in parallel[4]. DNA sequence was assembled and compared to the published genomic reference sequences. The presence of any potentially disease-associated sequence variant(s) was confirmed by dideoxy DNA sequence analysis. A reference library of

almost 800 alleles was used to evaluate the allele frequency of novel sequence variants if indicated. If appropriate, testing of one affected relative or, if not available, of both biological parents, was performed to clarify variants of unknown significance at no additional charge. HCM panel: Approximately 150 exons and their splice junctions of 18 genes were sequenced. DCM/LVNC panel: Approximately 620 exons of 38 genes including their splice junctions are sequenced. ARVC panel: The entire coding region of seven genes and their splice junctions are sequenced. CCM panel: Approximately 1300 coding exons and their flanking splice junctions of 76 genes are sequenced. Concurrently, targeted array CGH analysis with exon-level resolution is performed to evaluate for a deletion or duplication of one or more exons in 61 nuclear genes of the 76 genes included on the panel. Deletion/duplication testing does not include *FKTN*, *GATAD1* and the 14 mitochondrial genes. The presence of a potentially disease-associated deletion/duplication mutation is confirmed by quantitative PCR, repeat aCGH, or by another appropriate method. The genes tested in each panel are listed in Table 2-1.

Whole Exome Sequencing. DNA was extracted from the specimen using standard methods. Exome sequencing was performed using the Agilent SureSelect Human All Exon 50 Mb kit and an Illumina HiSeq2000. The raw sequencing data was mapped to and analyzed in comparison with the published human genome build UCSC hg19 reference sequence. The quality of the sequencing data was evaluated based on genomic variation quality parameters. Candidate variants, including single nucleotide variants or insertions or deletions (in/dels), were identified. Variants were filtered

according to mode of inheritance, allele frequency, and deleteriousness based on protein prediction algorithms. Any mutations/copy number changes found in the affected individual and thought to cause the clinical condition was confirmed by a second independent method such as dideoxy sequence analysis, or another appropriate method. Variants of uncertain clinical significance, which were unrelated to the primary clinical concern(s), as well as any incidental findings, which were not medically actionable, were not be reported. The UCLA Genomic Data Board, an interdisciplinary team of physicians, pathologists, clinical geneticists, laboratory directors, genetic counselors, and informatics specialists, reviewed the results and rendered the interpretation (Figure 2-1).

Results

Gene panel testing is unable to capture the causal genetic variant in more than half of HCM and DCM cases

The UCLA Cardiomyopathy Center began sending blood samples of familial cardiomyopathy patients to the commercial gene testing company GeneDx starting in March 2011 for next-generation sequencing-based genetic panel testing. Depending on the patient's clinical history, samples were directed towards genetic test panels for hypertrophic cardiomyopathy (HCM), dilated cardiomyopathy/left ventricular noncompaction (DCM/LVNC), or arrhythmogenic right ventricular cardiomyopathy (ARVC). Starting from November 2013, a comprehensive cardiomyopathy (CCM) gene panel consisting of 76 genes previously associated with any familial cardiomyopathy also became available (Table 2-1). Based on our review of the reports, variants identified in the genes tested were prioritized by GeneDx for pathogenicity by the following criteria: 1) previously reported in literature with strong evidence for its association with disease, 2) absence or rare in presumably healthy control samples, and 3) *in silico* predicted to be pathogenic based on vicinity to previously reported pathogenic amino acid substitution, deleterious effects on protein structure and function by prediction algorithms SIFT and Poly-Phen2, and conservation across species. Suspected pathogenic variants were reported by following classifications: disease-causing (DC), likely disease-causing (LDC), or variant of unknown significance (VUS).

We compiled results from all 21 HCM and 9 DCM patients tested between March 2011 and March 2014 (Table 2-2). We found that one patient from each group had 2 variants reported. In the HCM patient, one DC mutation in the *TPM1* gene and one LDC mutation in the *TNNI3* gene were identified. In the DCM patient, one LDC mutation in the *TTN* gene and one VUS in the *TNNT2* gene were identified. Overall, in 38% of HCM and in 33% of DCM cases, a definite disease-causing variant was reported. In 57% of HCM and 56% of DCM cases, no variant with evidence stronger than that at the level of VUS was identified. The distribution variants among the variant classes indicated that the current state of gene panel testing is unable to capture the causal genetic variant in more than half of the cases (Figure 2-2).

Exome sequencing relies on primary gene list in variant prioritization to identify disease-causing mutations in genes previously associated with disease

To examine the performance of whole exome sequencing in our patient population, we selected 3 patients to refer to the UCLA Clinical Exome Sequencing service for exome sequencing. Exome 1 was performed on a 33-year-old Hispanic male with a history of restrictive cardiomyopathy, who underwent heart transplant due to refractory symptoms of congestion. His father died at the age of 32 from heart failure. He has 2 brothers and 2 sons with heart failure, and a nephew who died of heart failure at the age of 13 (Figure 2-3A). Exome 2 was performed on a 70-year-old Filipino woman with a history of sudden cardiac death and dilated cardiomyopathy, who received an automatic implantable cardioverter defibrillator (AICD) and LV assist device.

Her father died suddenly at the age of 55. One of her brothers died of sudden cardiac arrest at the age of 49. Another brother underwent a heart transplant at the age of 60. One of her sisters has ventricular tachycardia and AICD. Another sister takes medications for her heart and has a daughter who underwent cardiac ablation for arrhythmia (Figure 2-3B). Exome 3 was performed on a 32-year-old woman with familial dilated cardiomyopathy who presented with frequent ventricular tachycardia requiring a heart transplant. Both of her brothers were diagnosed with dilated cardiomyopathy at the age of 18 and received heart transplants. Both she and her sister were diagnosed with dilated cardiomyopathy in their 30s and received heart transplant. Both her mother and her grandfather died in their 30s of unknown causes. Both grandmothers are alive and well. Of note, she and her siblings are products of a consanguineous union from a small village in Mexico (Figure 2-3C).

Whole exome capture and sequencing was performed using the Agilent SureSelect Human All Exon 50 Mb kit and an Illumina HiSeq2000. The raw sequencing data was mapped to and analyzed in comparison with the published human genome build UCSC hg19 reference sequence. The quality of the sequencing data were evaluated based on genomic variation quality parameters. Candidate variants, including single nucleotide variants or insertions or deletions (in/dels), were identified (Table 2-3). Variant filtration was performed by the exclusion of common SNPs based on dbSNPs and nonsynonymous SNPs. At this point each variant is referenced for presence in the human gene mutation database (HGMD), allele frequency based on the ~1,300 exome samples at UCLA, predicted pathogenicity based on protein prediction algorithms such

as SIFT, Poly-Phen2 and Condel, gene expression level in different tissues, and other supplementary information. In addition, a primary gene (PG) list was created for each patient, based on the clinical features of disease and associated genes, which is similar but more extensive than the cardiomyopathy gene panel offered by GeneDx. Based on the mode of inheritance, variants that intersect the PG list and are rare (<1%) were given first priority for consideration. Any mutations/copy number changes found in the affected individual and thought to cause the clinical condition was confirmed by a second independent method such as dideoxy sequence analysis, or another appropriate method. Variants of uncertain clinical significance, which were unrelated to the primary clinical concern(s), as well as any incidental findings, which were not medically actionable, were not reported. The UCLA Genomic Data Board, an interdisciplinary team of physicians, pathologists, clinical geneticists, laboratory directors, genetic counselors, and informatics specialists, reviewed the results.

In exome 1, six heterozygous variants in the primary gene list remained after filtering for allele frequency of < 1%. Only of the variants c.1544 T>C p.Met515Thr in exon 15 of 40 *MYH7* was predicted by SIFT, PolyPhen-2, and Condel to be deleterious. The same variant was found in the HGMD database to have been previously identified in an Indian female who had asymmetrical septal hypertrophy with septal thickness of 19 mm and died of sudden cardiac arrest[5]. Targeted sequencing confirmed the same *MYH7* mutation in his 2 affected children.

In exome 2, twenty-four heterozygous variants in the primary gene list remained after filtering for allele frequency of < 1%, six of which were considered deleterious by protein prediction algorithms. Among the genes intersected by these 6 variants, S*CN5A* and *LMNA* only were directly associated with cardiomyopathy and arrhythmia. The *SCN5A* variant c.1567 C>T p.Arg523Cys found in exon 12 of 28 had previously been identified to be associated with long QT syndrome in a Norwegian individual[6]. The *LMNA* variant c.646 C>T p.Arg216Cys in exon 4 of 12 has not been previously reported and was classified as a variant of unknown significance. Incidentally, patient was also noted to have homozygous disease-associated variant c.109G>A p.Val37Ile in exon 2 of 2 of *GJB2* and a heterozygous c.3374 T>C p.Val1125Ala variant in the *APC* gene based on a survey of the HGMD results. The *GJB2* variant represents 7.8% of deafness-associated *GJB2* variants and is the most common variant among Asians[7]. The patient was subsequently confirmed to have a history of childhood-onset hearing impairmentThe *APC* variant has previously been associated with one individual with non-familial adenomatous polyposis colorectal adenoma[8]. All of these results were reported.

Exome sequencing relies on family segregation to rapidly identifies disease-causing variant

Prior to exome sequencing, gross and histopathological examination of the hearts from the proband and her brother revealed characteristic fibrofatty replacement changes of ARVC, involving predominantly the left ventricle, which was not suspected

prior to these findings (Figure 2-4). Due to the lack of adequate health history on the exome 3 proband's mother and grandfather, the pattern of inheritance could not be definitively established but appears to be dominant. However, given the highly consanguineous pedigree, pseudo-dominant pattern of inheritance, in which an autosomal recessive condition is present in individuals in two or more generations of a family, cannot be excluded. Similar to the prior approach in exome 1 and exome 2, variants filtered according to the dominant model were examined (Table 2-4). After filtering for rare (<1%) heterozygous variants found in the primary gene list, no variants remained. There was an uncommon heterozygous variant in *DSG2,* a gene previously associated with ARVC; however, this was likely benign, as it was observed in 3% of the population. In addition, the patient was compound heterozygous in *TTN* gene with allele frequency of 46% (c.49942A>G p.Ile16648Val) and 1% (c.21241A>G p.Lys7081Glu). Neither of these variants were predicted by SIFT, PolyPhen-2, and Condel to be deleterious. HGMD database flagged a heterozygous c.835C>T p.Arg279Cys variant of *NEXN* to have been previously reported in a Chinese family with hypertrophic cardiomyopathy. However, the allele frequency for this variant is 29% in our reference database. None of these variants seemed to be consistent with the severe phenotypes we observed in this family. Next, variant prioritization was performed on the homozygous variants, based on a recessive mode of inheritance. No rare homozygous mutations was found within the primary gene list. However, a homozygous truncation mutation in *MYOM2* (c.1366C>T p.Arg456X), with an allele frequency 1.9% in the reference population, was found. Given that this is a truncation mutation, this mutation

was deemed to be a causal variant. The mutation was confirmed by Sanger sequencing in the proband.

MYOM2 is a 1465-amino-acid sarcomere structural protein localized to the M-line. If this variant were responsible for disease, the homozygous Arg456* truncation may be expected to nullify the function MYOM2 and disrupt M-line ultrastructure. We performed electromicrograph (EM) to evaluate the sarcomeric structure in the proband. Surprisingly, EM of the proband cardiac sarcomere demonstrated normal M-line pattern (Figure 2-5). To evaluate whether the MYOM2 mutation segregated with disease, we obtained saliva samples from available family member, including the proband's father, living siblings as well as the explanted heart tissue samples from the deceased brother. However, the PCR genotyping results did not confirm segregation of homozygosity for the *MYOM2* c.1366C>T variant with the phenotype (Table 2-5).

To further investigate the exome variants according to the autosomal recessive model, the Illumina HumanOmniExpress BeadChip assay was performed on the proband and her living affected siblings (Figure 2-6). Three variants were found to be homozygous, rare (<1%) and shared among the three affected siblings (Table 2-4). Among the genes affected by these candidate variants, *ZBTB17* resides in a region that was previously implicated in dilated cardiomyopathy in two select SNP association studies (one in Europeans and one in Han Chinese) and one genome-wide association study in Europeans[9-11]. The exome variant c.535+2T>A in *ZBTB17* is an essential splice

site variant not previously identified in more than 1,300 control samples in the UCLA exome database.

Discussion

In spite of the number of genes included in genetic panel(s) for cardiomyopathy, only 38% and 33% of the patients received a definitive disease-causing mutation diagnosis. There remains a substantial gap in the understanding genetic causes of familial cardiomyopathies. While some of the causal mutations for familial cardiomyopathy may lie in non-coding regions of known genes, there are likely still a significant number genes whose role in cardiomyopathy have not yet been discovered. Gene panel testing is cheap, simple and easy for the average cardiologist to interpret; however, its major drawback is the failure to detect potentially causal genetic variants in more than half of the cases. Whole exome sequencing serves as an important alternative to close the gap of understanding in cardiomyopathy genetics.

Both gene panel testing and exome sequencing employ variant filtering strategies based on previous reports in literature, allele frequency and *in silico* prediction of pathogenicity to classify variants, with the following underlying assumptions: 1) Disease-causing mutations reported in literature are supported with good evidence. 2) Disease-causing mutation is rare in the population. 3) *In silico* prediction of missense variant pathogenicity is biological meaningful. In practice, disease-causing mutations in the reported literature demonstrate variable levels of

evidence. Unfortunately, some of the reported mutations in literature were identified using targeted sequencing in isolated cases without segregation data in the proband's family. Therefore, previously reported disease-causing mutations may in fact be benign genetic polymorphisms, especially if the cases were of non-Caucasian genetic backgrounds and harbored variants that appeared to be novel and rare at the time of identification. The apparently familial disease may be due to two or three different mutations, each with an allele frequency above the threshold for being called rare. The reference population for variant allele frequency calculation may be inappropriate and result in false allele frequency calculation. For example, using a predominantly Caucasian reference population to estimate the allele frequency of a variant found commonly in the Asian population may falsely underestimate the allele frequency for some and misclassify them as rare variants, while in fact the variant is common in the Asian population. Finally, *in silico* prediction of missense variant pathogenicity by the most commonly used prediction algorithms, SIFT, Poly-Phen2, and Condel, may not agree with each other or represent biologically meaningful results. Finally, disease causal variants may not be in protein-coding regions, even in Mendelian cases, which will be missed by the current gene panel and exome sequencing strategies by design.

Whole exome sequencing demonstrates great promise for more than half of the cases tested negative on genetic test panels. In the case of exome 3, we illustrated the importance of family pedigree analysis and additional family member sample testing. The identification of a relatively rare homozygous truncation mutation in *MYOM2*, predicted to be deleterious to protein function, led us down the path of studying muscle

M-line structure by EM, which turned out to be normal. Targeted genotyping in family members verified that the candidate variant in *MYOM2* did not segregate with disease. The presence of a rare variant with predicted deleterious effect thus presents as a difficult diagnostic dilemma in real life. It also highlights the fact that exome sequencing results can be very difficult to interpret in isolation. But, a well-defined pedigree and a few additional family member samples can be rapidly filtered down the variant list to a handful of candidates.

ZBTB17 (zinc finger and BTB domain containing 17), also known as Miz-1, is located less than 40 kb from HSPB7 in a region previously demonstrated by select SNP association and genome-wide association (GWA) to be associated with dilated cardiomyopathy[9, 11]. Last year, the genotype of SNP rs10927875 in *ZBTB17* (OR=5.19, 95% CI =1.00 to 27.03, P=0.05) was associated with DCM in a Han Chinese population and there was no difference in genotype or allele frequencies in *HSPB7* between DCM patients and control subjects[10]. Interestingly, of the two replicated loci reported from the DCM GWA, one has already been validated in familial DCM human subjects. Thus, the candidate essential splice-site variant in *ZBTB17* found on the second locus presents *ZBTB17* as a very interesting gene to validate experimentally or in human cohorts.

There is significant potential in variant and gene discovery using exome sequencing in the clinical population, especially in those with clearly demonstrated Mendelian pattern of inheritance and adequate number of family samples. A systematic approach to screen patients and families, collect deep pedigree information, deep

cardiac phenotyping, patient and family sample collection, sequence, filter, and validate candidate variants will be critical to the rapid identification of causal genetic variants in familial cardiomyopathies and will build an important phenotype-genotype correlation resource that will guide future prognostic and treatment strategies.

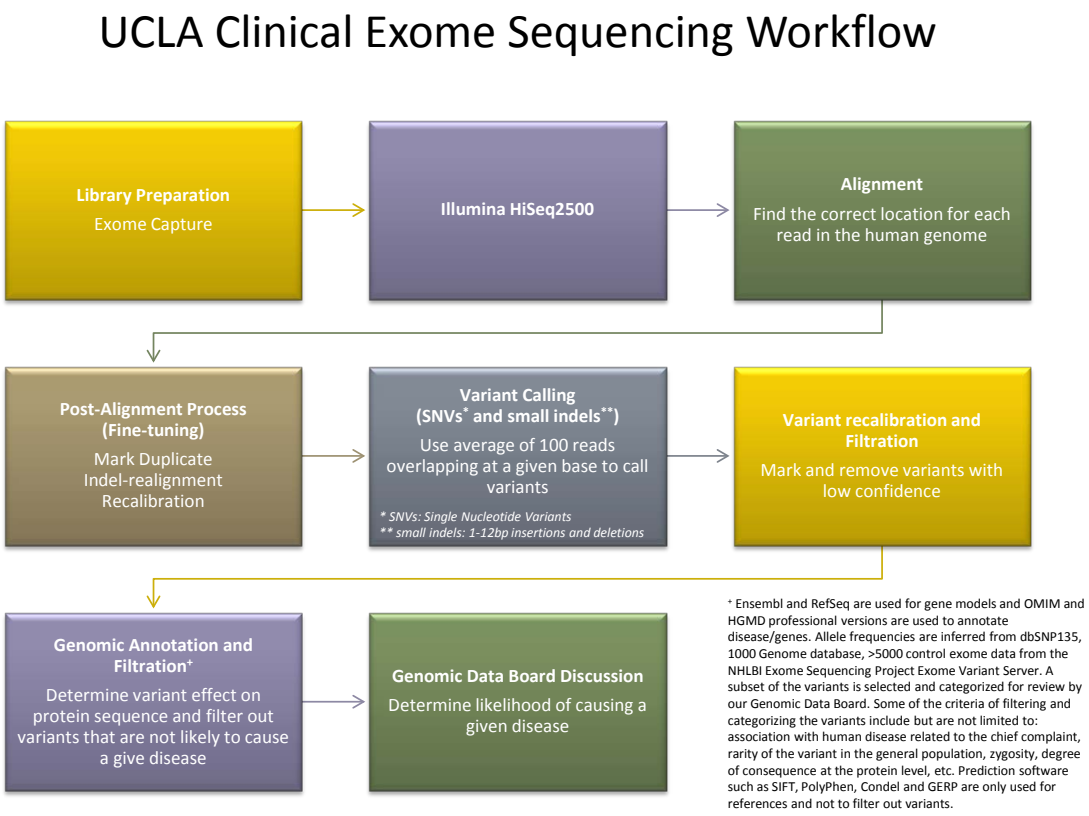
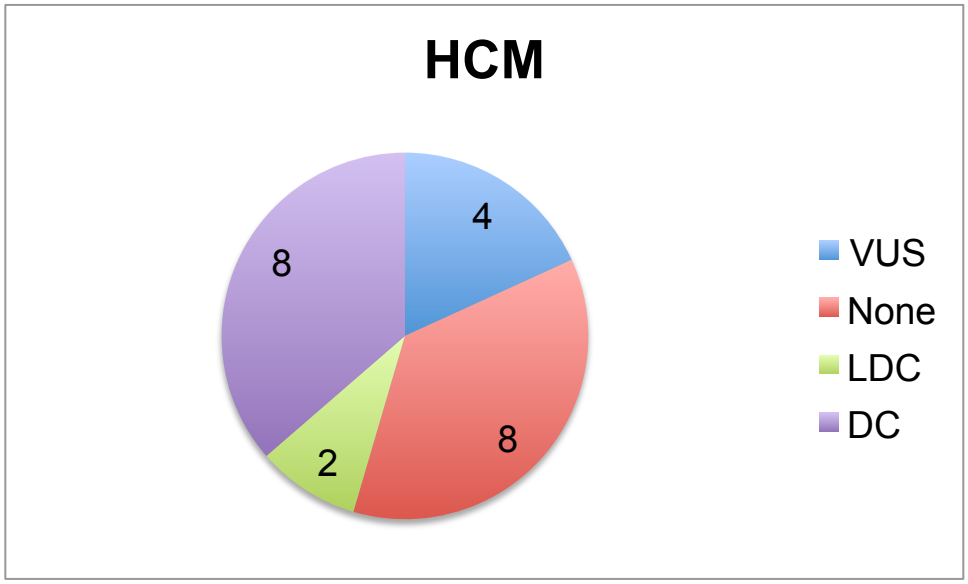


Figure 2-1. UCLA Clinical Exome Sequencing workflow.
Ensembl and RefSeq were used for gene models. OMIM and HGMD professional versions were used to annotate disease/genes. Allele frequencies were inferred from dbSNP135, 1000 Genome database, > 5000 control exome data from the NHLBI Exome Sequencing Project Exome Variant Server. A subset of the variants was selected and categorized for review by the Genomic Data Board. Some of the criteria of filtering and categorizing the variants included but were not limited to: association with human disease related to the chief complaint, rarity of the variant in the general population, zygosity, degree of consequence at the protein level, etc. Prediction software such as SIFT, PolyPhen, and Condel and GERP were only used for references and not to filter out variants.

A.



B.

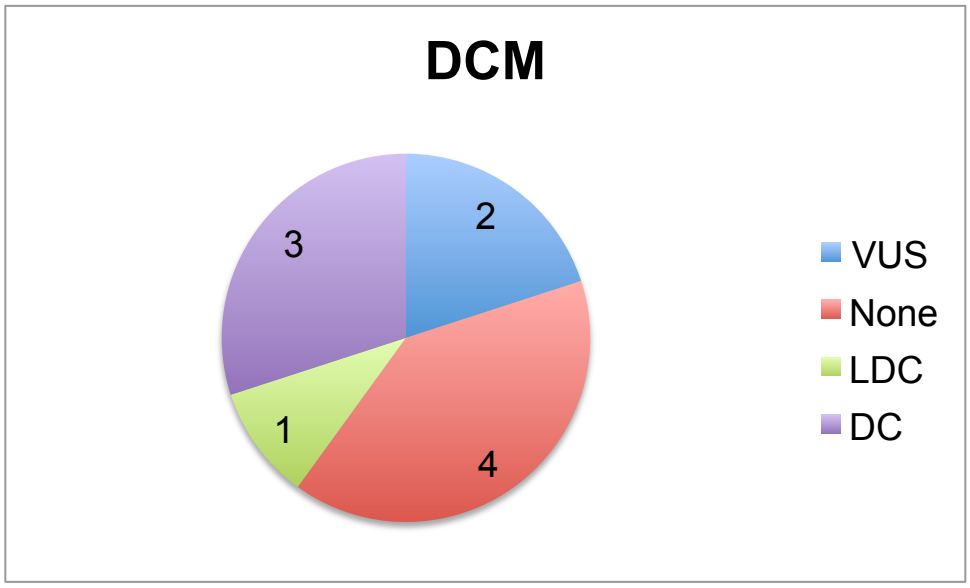
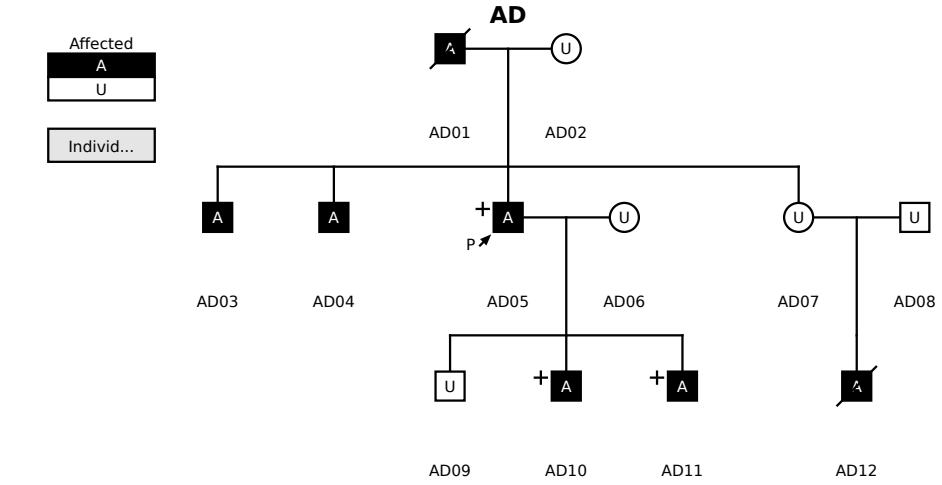
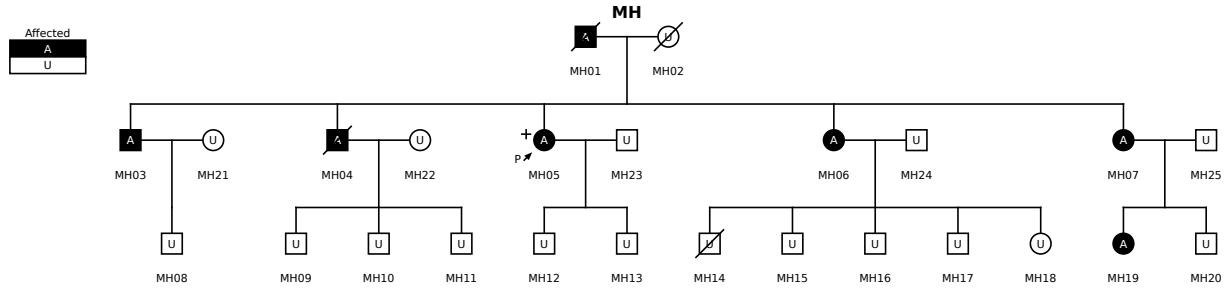


Figure 2-2. GeneDx genetic test results by variant classifications for familial cardiomyopathy patients.
A. Hypertrophic cardiomyopathy (HCM) patients. B. Dilated cardiomyopathy (DCM) patients. DC represents disease-causing mutation. LDC represents likely disease-causing mutation. VUS represents variant of unknown significance.



A.

B.



C.

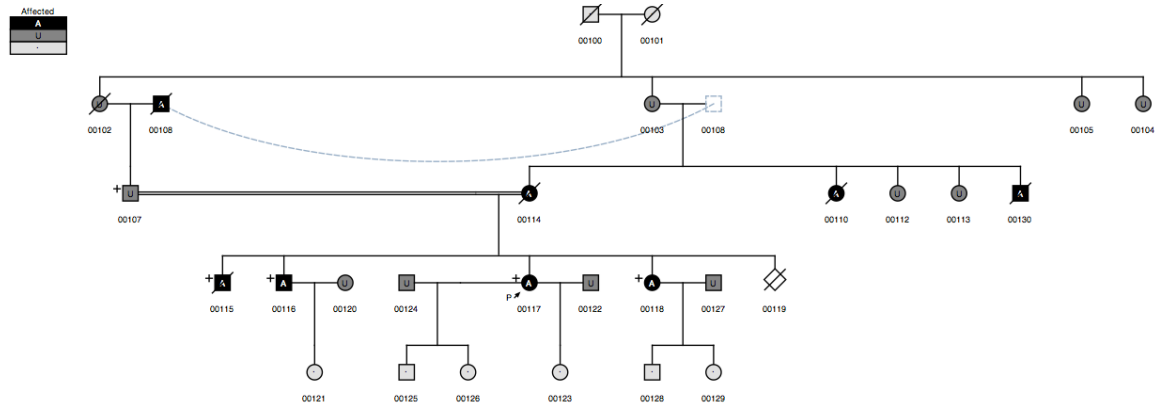


Figure 2-3. Family pedigrees for exome sequencing patients. A. Family pedigree for Exome 1. B. Family pedigree for Exome 2. C. Family pedigree for Exome 3. P represents the proband on whom exome sequencing was performed. + represents DNA sampled individuals.

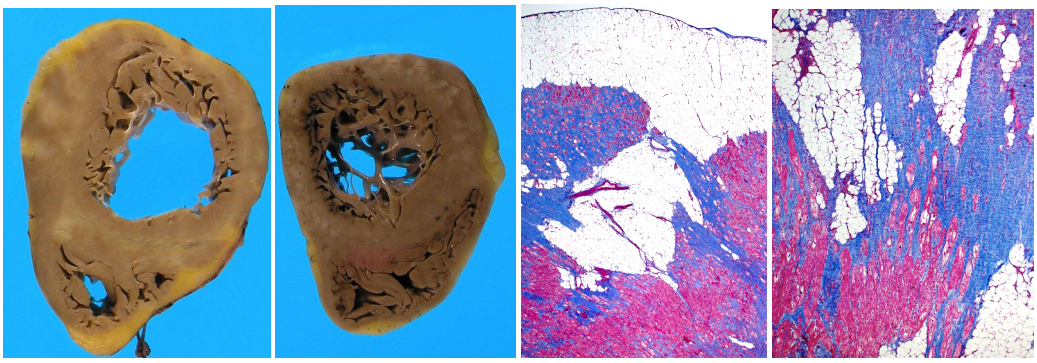


Figure 2-4. Gross and histopathologic examination of the exome 3 proband and family member are consistent with arrhythmogenic right ventricular cardiomyopathy (ARVC). The gross appearance of the proband's (left most) and proband's brother's (left) heart. Masson's trichrome stained sections of the proband's heart in the longitudinal (right) and the cross-sectional (right most) views.

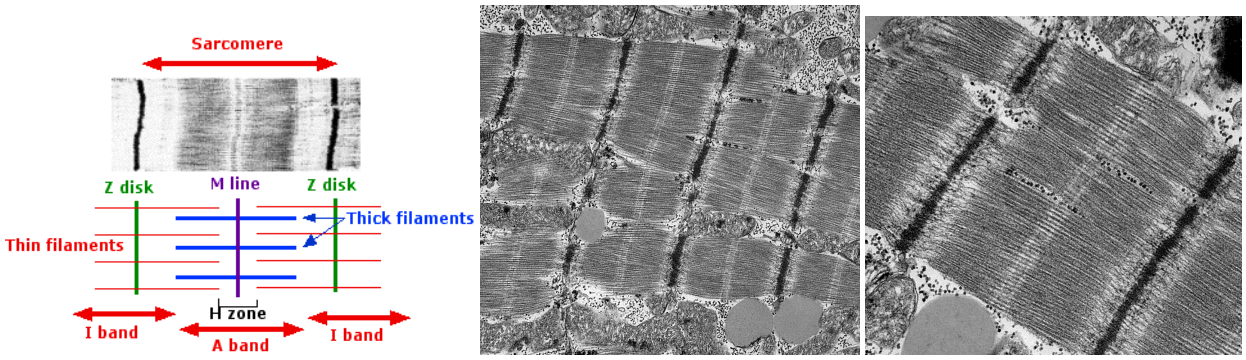


Figure 2-5. The M-line ultrastructure of exome 3 proband is intact. Normal electromicrograph (EM) and diagram (left), low magnification EM (middle) and high magnification EM (right) of the exome 3 proband's heart tissue.

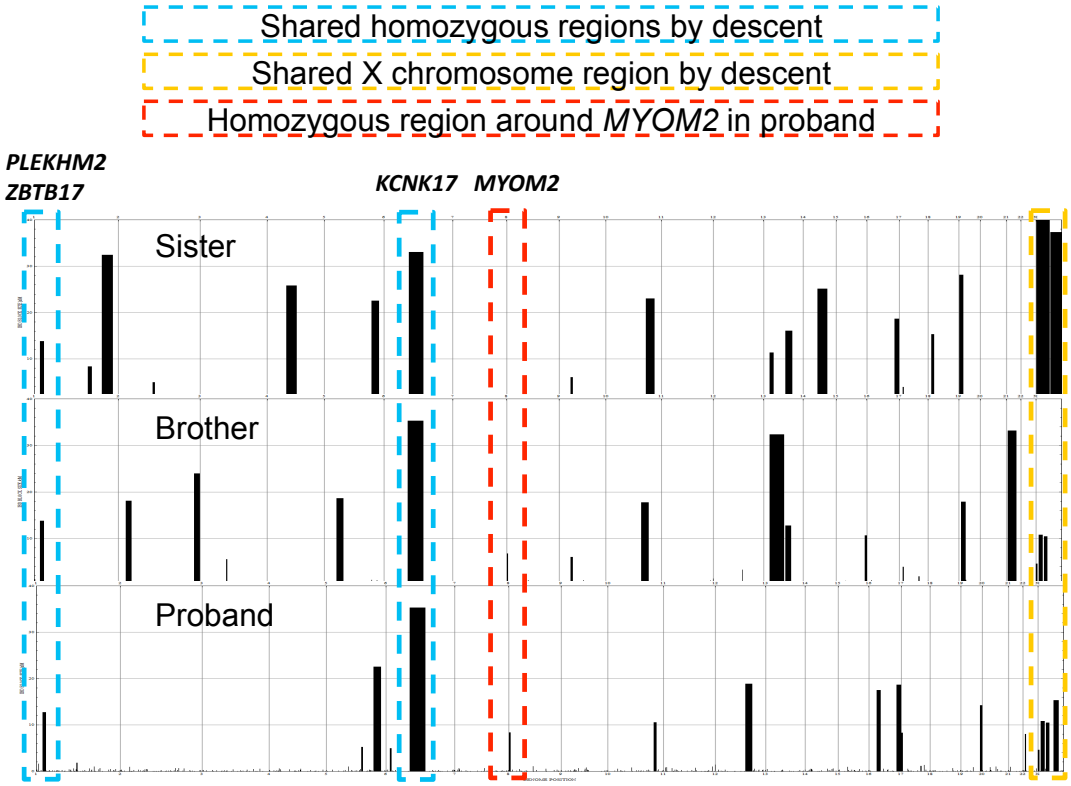


Figure 2-6. Homozygosity by descent analysis.
The Illumina HumanOmniExpress BeadChip assay was performed on the proband and her two living affected siblings. Homozygous identical by descent regions are marked by the black bars. Of the two homozygous identical by descent regions shared by all three siblings, three variants in *PLEKHM2*, *ZBTB17*, and *KCNK17* were found to be rare.

Table 2-1. Genes in the GeneDx cardiac genetic test panels.
HCM represents hypertrophic cardiomyopathy, DCM represents dilated cardiomyopathy, ARVC represents arrhythmogenic right ventricular cardiomyopathy, and CCM represents comprehensive cardiomyopathy genetic test panel.

HCM	DCM	ARVC	CCM (since Nov 2013)	
ACTC (ACTC1)	*ACTC (ACTC1)*	*DSC2*	*ABCC9*	*MTTL1*
CAV3	*ACTN2*	*DSG2*	*ACTC (ACTC1)*	*MTTL2*
GLA	*ANKRD1*	*DSP*	*ACTN2*	*MTTM*
LAMP2	*CSRP3*	*JUP*	*ANKRD1*	*MTTQ*
MTTG	*DES*	*PKP2*	*BAG3*	*MTTS1*
MTTI	*EMD*	*RYR2*	*BRAF*	*MTTS2*
MTTK	*LAMP2*	*TMEM43*	*CAV3*	*MYBPC3*
MTTQ	*LDB3 (ZASP)*		*CRYAB*	*MYH7*
MYBPC3	*LMNA*		*CSRP3*	*MYL2*
MYH7	*MTND1*		*DES*	*MYL3*
MYL2	*MTND5*		*DMD*	*MYLK2*
MYL3	*MTND6*		*DSC2*	*MYOZ2*
PRKAG2	*MTTD*		*DSG2*	*MYPN*
TNNC1	*MTTH*		*DSP*	*NEBL*
TNNI3	*MTTI*		*DTNA*	*NEXN*
TNNT2	*MTTK*		*EMD*	*NRAS*
TPM1	*MTTL1*		*FKTN*	*PDLIM3*
TTR	*MTTL2*		*GATAD1*	*PKP2*
	MTTM		*GLA*	*PLN*
	MTTQ		*HRAS*	*PRKAG2*
	MTTS1		*ILK*	*PTPN11*
	MTTS2		*JPH2*	*RAF1*
	MYBPC3		*JUP*	*RBM20*
	MYH7		*KRAS*	*RYR2*
	NEXN		*LAMA4*	*SCN5A*
	PLN		*LAMP2*	*SGCD*
	RBM20		*LDB3 (ZASP)*	*SOS1*
	SCN5A		*LMNA*	*TAZ*
	SGCD		*MAP2K1*	*TCAP*
	TAZ		*MAP2K2*	*TMEM43*
	TCAP		*MTND1*	*TMPO*
	TNNC1		*MTND5*	*TNNC1*
	TNNI3		*MTND6*	*TNNI3*
	TNNT2		*MTTD*	*TNNT2*
	TPM1		*MTTG*	*TPM1*
	TTN		*MTTH*	*TTN*
	TTR		*MTTI*	*TTR*
	VCL		*MTTK*	*VCL*

Table 2-2. Reported genes from gene panel testing by variant classification. DC represents disease-causing mutation. LDC represents likely disease-causing mutation. VUS represents variant of unknown significance. § represent that the variants are from the same hypertrophic cardiomyopathy (HCM) patient. ¶ represent that the variants are from the same dilated cardiomyopathy (DCM) patient.

A. Reported genes by variant classification in HCM cases.

Gene	DC	LDC	VUS
BAG3			1
TNNI3	1	1§	
TPM1	1§		1
MYH7	1	1	2
MYBPC3	5		

B. Reported genes by variant classification in DCM cases.

Gene	DC	LDC	VUS
LMNA	1		
MT-TH			1
MYH7	1		
SCN5A	1		
TNNT2			1¶
TTN		1¶	

Table 2-3. Exome sequencing summary data.
Het represents heterozygous variants, Hom represents homozygous variants. Greater than 95% of the variants found within the primary gene list had > 10X sequencing coverage.

	Exome 1	Exome 2	Exome 3
Ethnicity	Hispanic	Asian	Hispanic
Mode of Inheritance	Dominant	Dominant	Dominant*
Phenotype	DCM	DCM	DCM
Mapped Reads (10^9)	15.2	14.6	15.2
Coverage	165X	157X	163X
Total Variants	21,956	22,171	22,369
SNV	20,845	21,055	21,267
Indel (1-10bp)	1,111	1,116	1,102
Common & synonymous excluded	659	883	604
Het SNV	635	851	579
Het Indel	24	32	25
Hom SNV	26	16	37
Hom Indel	1	3	0
MT	2	1	3
Primary gene list	376	655	176
Het	13	40	7
< 1%	6	24	0
Compound Het	0	2	2
Hom	0	1	0
MT	0	0	0
Gene	*MYH7*	*SCN5A*	NA
DNA	c.1544T>C	c.1567C>T	
Protein	p.Met515Thr	p.Arg523Cys	

Table 2-4. Select genetic variants from exome 3.

Gene	Zygosity	DNA	Protein	Allele Frequency
DSG2	Heterozygous	c.877A>G	p.Ille293Val	3%
TTN	Compound	c.49942A>G	p.Ile16648Val	46%
	Heterozygous	c.21241A>G	p.Lys7081Glu	1%
NEXN	Heterozygous	c.835C>T	p.Arg279Cys	29%
MYOM2	Homozygous	c.1366C>T	p.Arg456X	1.9%
PLEKHM2	Homozygous	c.2594T>G	p.Met865Arg	0%
ZBTB17	Homozygous	c.535+2T>A		0%
KCNK17	Homozygous	c.949C>T	p.Gln317X	0%

Table 2-5. PCR genotyping results of candidate *MYOM2* variant.

Reference	Proband	Father	Sister	Brother	Brother
C/C	T/T	C/T	T/T	C/T	C/T

Bibliography

1. Seidman CE, Seidman JG. Identifying sarcomere gene mutations in hypertrophic cardiomyopathy: A personal history. *Circ Res*. 2011;108:743-750

2. Bamshad M, Ng S, Bigham A, Tabor H, Emond M, Nickerson D, Shendure J. Exome sequencing as a tool for mendelian disease gene discovery. *Nature reviews. Genetics*. 2011;12:745-800

3. Choi M, Scholl U, Ji W, Liu T, Tikhonova I, Zumbo P, Nayir A, Bakkalo*f*ülu Aü, Ozen S, Sanjad S, Nelson-Williams C, Farhi A, Mane S, Lifton R. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America*. 2009;106:19096-19197

4. Bennett S. Solexa ltd. *Pharmacogenomics*. 2004;5:433-438

5. Rai TS, Ahmad S, Bahl A, Ahuja M, Ahluwalia TS, Singh B, Talwar KK, Khullar M. Genotype phenotype correlations of cardiac beta-myosin heavy chain mutations in indian patients with hypertrophic and dilated cardiomyopathy. *Mol Cell Biochem*. 2009;321:189-196

6. Berge KE, Haugaa KH, Fruh A, Anfinsen OG, Gjesdal K, Siem G, Oyen N, Greve G, Carlsson A, Rognum TO, Hallerud M, Kongsgard E, Amlie JP, Leren TP. Molecular genetic analysis of long qt syndrome in norway indicating a high prevalence of heterozygous mutation carriers. *Scand J Clin Lab Invest*. 2008;68:362-368

7. Putcha GV, Bejjani BA, Bleoo S, Booker JK, Carey JC, Carson N, Das S, Dempsey MA, Gastier-Foster JM, Greinwald JH, Jr., Hoffmann ML, Jeng LJ, Kenna MA, Khababa I, Lilley M, Mao R, Muralidharan K, Otani IM, Rehm HL, Schaefer F, Seltzer WK, Spector EB, Springer MA, Weck KE, Wenstrup RJ, Withrow S, Wu BL, Zariwala MA, Schrijver I. A multicenter study of the frequency

and distribution of gjb2 and gjb6 mutations in a large north american cohort. *Genet Med*. 2007;9:413-426

8. Azzopardi D, Dallosso AR, Eliason K, Hendrickson BC, Jones N, Rawstorne E, Colley J, Moskvina V, Frye C, Sampson JR, Wenstrup R, Scholl T, Cheadle JP. Multiple rare nonsynonymous variants in the adenomatous polyposis coli gene predispose to colorectal adenomas. *Cancer Res*. 2008;68:358-363

9. Stark K, Esslinger UB, Reinhard W, Petrov G, Winkler T, Komajda M, Isnard R, Charron P, Villard E, Cambien F, Tiret L, Aumont MC, Dubourg O, Trochu JN, Fauchier L, Degroote P, Richter A, Maisch B, Wichter T, Zollbrecht C, Grassl M, Schunkert H, Linsel-Nitschke P, Erdmann J, Baumert J, Illig T, Klopp N, Wichmann HE, Meisinger C, Koenig W, Lichtner P, Meitinger T, Schillert A, Konig IR, Hetzer R, Heid IM, Regitz-Zagrosek V, Hengstenberg C. Genetic association study identifies hspb7 as a risk gene for idiopathic dilated cardiomyopathy. *PLoS Genet*. 2010;6:e1001167

10. Li X, Luo R, Mo X, Jiang R, Kong H, Hua W, Wu X. Polymorphism of zbtb17 gene is associated with idiopathic dilated cardiomyopathy: A case control study in a han chinese population. *Eur J Med Res*. 2013;18:10

11. Villard E, Perret C, Gary F, Proust C, Dilanian G, Hengstenberg C, Ruppert V, Arbustini E, Wichter T, Germain M, Dubourg O, Tavazzi L, Aumont MC, DeGroote P, Fauchier L, Trochu JN, Gibelin P, Aupetit JF, Stark K, Erdmann J, Hetzer R, Roberts AM, Barton PJ, Regitz-Zagrosek V, Aslam U, Duboscq-Bidot L, Meyborg M, Maisch B, Madeira H, Waldenstrom A, Galve E, Cleland JG, Dorent R, Roizes G, Zeller T, Blankenberg S, Goodall AH, Cook S, Tregouet DA, Tiret L, Isnard R, Komajda M, Charron P, Cambien F. A genome-wide association study identifies two loci associated with heart failure due to dilated cardiomyopathy. *Eur Heart J*. 2011;32:1065-1076

3 Genetic Contributions to Cardiac Hypertrophy and Fibrosis Induced by Beta-adrenergic Stimulation in Mice

Abstract

Chronic heart failure exhibits both a wide range in severity and a high degree of heterogeneity in clinical manifestation in human patients. These pathological variations are the result of complex interactions between pathological stressors and underlying genetic variations. Although numerous factors have been established as key pathological stressors for heart failure, the genetic impact on the pathological outcome of these stimuli remains poorly understood. Using the β-adrenergic agonist isoproterenol as a specific pathological stressor to circumvent the problem of etiological heterogeneity, we performed a Genome Wide Association Study (GWAS) for genes influencing cardiac hypertrophy and fibrosis in a large panel of inbred mice. Seven significant and 22 suggestive loci, containing an average of 14 genes were found to affect cardiac hypertrophy, fibrosis and other pathological traits associated with heart failure. Several loci contained candidate genes, which have already been implicated in familial cardiomyopathies in humans or pathological cardiac remodeling in experimental models. Other loci contained genes with previously uncharacterized roles in heart failure. In particular, we identified Abcc6 as a novel gene regulating isoproterenol-induced cardiac fibrosis and validated that an allele with a splice mutation of Abcc6 promoted fibrosis accumulation. In conclusion, we find that genetic variants significantly contribute to the phenotypic heterogeneity of stress-induced cardiomyopathy. Systems

genetics is an effective approach to identify genes and pathways underlying the specific pathological features of cardiomyopathies. Abcc6 is among many previously unrecognized players in the development of stress-induced cardiac hypertrophy, fibrosis and other pathological features of heart failure.

Significance

Heart failure (HF) is a common cause of mortality and morbidity, with a lifetime risk of 1 in 9 in first world countries. Due to its complex etiology, genome-wide association studies (GWAS) of HF in humans have been only modestly successful in determining the genetic bases of HF. We approach HF GWAS using a novel population of mice, the Hybrid Mouse Diversity Panel. GWAS analysis revealed known genes for cardiomyopathy and cardiac pathology as well as previously uncharacterized players affecting the process. In particular, we identified and validated Abcc6 as a previously unknown regulator of stress-induced cardiac fibrosis. The application of GWAS in a genetically diverse mouse population is a robust approach to dissect the genetic basis of heart failure.

Background

Heart failure (HF) is a common cause of death with a lifetime risk of at least one in 9 for both men and women in developed countries[1]. Heart failure is a complicated

syndrome, characterized by a large number of pathological changes, such as contractile dysfunction, cardiomyocyte hypertrophy, edema and myocardial fibrosis[2-4]. The onset and severity of these pathological manifestations are highly heterogeneous among HF patients, likely due to complex interactions between the genetic variants and the pathological stressors including mechanical overload and humoral overstimulation. Indeed, a number of humoral factors such catecholamines and angiotensin II are known to play key roles in triggering HF, however the genetic variations underlying the pathological outcome in response to these stressors remains elusive. Dissecting the genetic contributions to specific pathological changes in the failing heart would provide important insights for the future development of personalized diagnoses and targeted therapies.

In contrast to many other common disorders, GWAS of HF have had modest success in elucidating the genetics underlying this complex disease. Only two heart-failure related loci[5] have reached accepted levels of genome-wide significance, despite meta-analyses of tens of thousands of patients[6, 7]. The challenge of performing GWAS in human 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 metabolic disorders, each of which are complex traits with significant environmental confounders[1]. Attempts to dissect the genetics of HF traits in rodents have been only modestly successful; although a number of loci for hypertrophy and fibrosis have been identified, the poor mapping resolution of traditional linkage analyses has complicated the identification of the underlying genes[8-11]. The development of a method to perform

high resolution, association-based mapping of complex traits in mice[12] provided an opportunity to identify genetic factors contributing to common forms of HF under defined stress conditions.

In this study, we have conducted a comprehensive phenotypic characterization in a large panel of densely genotyped inbred mice from the Hybrid Mouse Diversity Panel (HMDP)[12] following chronic treatment with a β-adrenergic agonist, isoproterenol (ISO). A wide spectrum of phenotypic changes was observed among the HMDP mice in ISO induced cardiac hypertrophy, fibrosis and peripheral edema. Using GWAS, we uncovered 7 significant loci and 22 suggestive loci, each containing an average of 14 genes. Several of these loci included genes with established causal roles in familial cardiomyopathies in humans or heart failure phenotypes in experimental models. In addition, we identified Abcc6 as a previously unrecognized regulator of ISO-induced cardiac fibrosis. Therefore, our study demonstrates clear evidence that genetic variants have a significant contribution to the phenotypic heterogeneity of stress-induced cardiomyopathy. Systems genetics is a potent approach to reveal genes and pathways underlying the specific pathological features of cardiomyopathies. The genetic information and the phenotypic spectra established by this study will be a valuable resource for future heart failure studies.

Methods

Ethics Statement

All animal experiments were conducted following guidelines established and approved by the University of California, Los Angeles Institutional Animal Care and Use Committee. All surgery and echocardiography was performed under isoflurane anesthesia, and every effort was made to minimize suffering.

Online database

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

Mice and isoproterenol treatment

The mouse strains listed Table 3-1 were obtained from The Jackson Laboratory and then bred in our colony. All mice have been previously genotyped at over 130,000 locations. Isoproterenol (30 mg per kg body weight per day, Sigma) was administered for 21 d in 8- to 10- week-old female mice using ALZET osmotic minipumps, which were surgically implanted intra- peritoneal. Abcc6 KO and transgenic mice[13, 14] 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 and body weight 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 48 hrs 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 for fibrosis and Alizarin Red for calcification. Sections were analyzed using a Nikon Eclipse, TE2000-U microscope 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) or red (for calcification) to the total tissue area. As expected[4, 15], we observed strong correlations between cardiac fibrosis and total heart weight (P=1.1E-07). Our results compare favorably to prior quantifications of fibrosis in a limited number of strains[16].

Association Analysis

We performed the association testing of each SNP with a linear mixed model, which accounts for the population structure among the n animals using the following model[17]:

$$y = 1_n + m + xb + u + e,$$

where m is the mean, b is the allele effect of the SNP, x is the (n x 1) vector of observed genotypes of the SNP, u is the random effects due to genetic relatedness with $\mathrm{var}(u) = \sigma_u^2 IuK$, and e is the random noise with $\mathrm{var}(e) = \sigma_e^2 I$. K denotes the identity-by-state kinship matrix estimated from all the SNPs, I denotes the (n x n) identity matrix, and 1_n is the (n x 1) vector of ones. We estimated σ_u^2 and σ_e^2 using restricted maximum

likelihood (REML) and computed p values using the standard F test to test the null hypothesis b=0. Genome-wide significance threshold and genome-wide association mapping are determined as the family-wise error rate as the probability of observing one or more false positives across all SNPs for phenotype. Prior work[17] has demonstrated that for the HMDP, 4.1E-6 is the correct significance threshold for a single trait. We conservatively estimated that we were observing ten independent phenotypes in our data and determined our final significance threshold, 4.1E-7, by Bonferroni correction. LD was determined by calculated pairwise r^2 SNP correlations for each chromosome. Approximate LD boundaries were determined by visualizing r^2 > 0.8 correlations in MATLAB (MathWorks).

Locus Overlap with Other Studies

Gwas.gov was queried for all human GWAS loci for the terms 'heart failure' or 'cardiac hypertrophy.' All loci with p-value < 5E-7 were selected. The NCBI homology maps (http://www.ncbi.nlm.nih.gov/projects/homology/maps/) were used to find syntenic location of the 5 Mb region surrounding the peak SNP of the human HF locus. If a mapped syntenic region overlapped with the LD block of a suggestive locus from our study, it was considered a positive hit. A significance p value was obtained by permutation testing, in which all suggestive loci were randomly placed across the genome and the number of overlaps measured. Final significance was calculated as the number of permutations that surpassed the observed number of overlaps.

Microarray and eQTL analysis

Following homogenization of LV 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[18] and batch effects addressed through the use of COMbat[19]. eQTLs were then calculated using EMMA, as described above. Significance thresholds were calculated as in Parks et al.[20] 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. and described above[17].

Results

Pathological Analysis of ISO Induced Cardiomyopathy in HMDP Mice

Although rarely an initial impetus for HF, β-adrenergic stimulation is considered a common and critical driving force behind ongoing hypertrophy and progression to heart failure[21]. We treated mice chronically with isoproterenol (ISO), a synthetic non-selective β- adrenergic agonist[22, 23]. 748 mice from 105 different strains of the Hybrid Mouse Diversity Panel (HMDP) were divided into control (average 2.2 per strain) and treated (average 4.1 per strain) cohorts (Table 3-1). Treated mice were implanted with an Alzet micropump and given 30 mg/kg/day of ISO for three weeks, at which point all mice were sacrificed. We characterized a variety of phenotypes to capture specific portions of the complex heart failure syndrome. For this report, cardiac hypertrophy and pulmonary and liver edema were assessed by measuring the weights of the four cardiac chambers, the

lungs and the liver. Cardiac fibrosis was measured by quantifying fibrotic tissue area as a percentage of all tissue area in LV sections stained using Masson Trichrome.

As shown in Figure 3-1, we observed striking differences in cardiac hypertrophy, fibrosis, and degrees of pulmonary and hepatic edema among the strains. Our results are consistent with another report from a more limited strain survey[24]. 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, whereas none of the control cohort died. We did not observe a significant correlation between the mice that died prior to end of protocol and any of our analyzed phenotypes or any of our expressed genes. This leads us to believe that the cause of the ISO-induced death is independent from other pathological phenotypes, but is clearly determined by genetic background (Figure 3-2).

GWAS for Pathological Traits

Association analysis was performed using ~132,000 SNPs across the genome with the EMMA algorithm[17] to correct for population structure. In addition to the absolute tissue 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. Prior work with EMMA and the HMDP, employing simulation and permutation, has suggested that an appropriate genome-wide significance threshold for a single trait is 4.1E-06[12]. This is approximately equivalent to a Bonferroni correction[12]. To correct for multiple comparisons, we have chosen the threshold of 4.1E-07 and a

minimum minor allele frequency (MAF) of 7.5% for our study. Given that the traits are correlated, this threshold (10-fold lower than the genome- wide significance level for a single trait) is conservative. Using these thresholds, we have identified 7 significant loci and 22 additional loci, which matched the nominal significance threshold of 4.1E-06 (Table 3-2 and Table 3-3). While linkage analysis in mice typically exhibits a resolution of tens of Mb[8, 25], the loci identified in this study averaged 1-2 Mb in size, based on linkage disequilibrium, with the majority being less than 1 Mb.

The right and LV weight (RVW, LVW) variations mirrored each other closely, with associations being somewhat stronger for RVW (Figure 3-1b). In total we observed three significant and five suggestive loci corresponding to treated RVW (Figure 3-3a), one significant and one suggestive locus for the ratio of treated to untreated RVW (Figure 3-3b), and a single suggestive 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 (Figure 3-1c and Figure 3-1d). Lung weight in particular showed a robust increase with ISO treatment. We observed one significant and four suggestive loci corresponding to ISO-treated lung weights (Figure 3-3c), and one significant and one suggestive locus corresponding to ISO-treated liver weights (Figure 3-3d). Cardiac fibrosis also varied significantly in both baseline and treated mice, with the extent of fibrosis being much greater in treated mice (Figure 3-1e). We observed a total of seven suggestive loci for cardiac fibrosis in untreated animals (Figure 3-3e), and one significant and five suggestive loci in treated animals (Figure 3-3f).

eQTL Analysis for Candidate Genes from ISO treated HMDP Mice

To help identify candidate genes at the heart failure associated loci, we carried out global expression analysis of LV heart tissue from 92 strains of ISO treated mice . The loci controlling gene expression levels were mapped using EMMA, and are referred to as expression quantitative trait loci (eQTL). eQTLs were termed 'cis' if the locus maps within 1Mb of the gene and otherwise were termed 'trans'. Overall, we observed 3093 cis eQTL (5% FDR p-value < 3.6E-3). Additionally, the Wellcome Trust Mouse Genomes Project sequencing database[26], 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. All significant and suggestive loci as well as gene expression data are available at http://systems.genetics.ucla.edu/data.

Using eQTL analysis combined with GWAS, we identified a causal gene in one of the cardiac hypertrophy loci on chromosome 3 (Figure 3-4). The peak SNP (P=1.9E-6), for the trait of treated-to-untreated right RVW ratio, lies between the second and third exons of Ppp3ca, encoding the alpha isozyme of calcineurin A, which is also the only gene contained within the Linkage Disequilibrium (LD) block surrounding the significantly associated SNPs. Calcineurin A is a known target of β-adrenergic signaling, with a well-described role in ISO-induced hypertrophy[34]. Calcineurin A is the only gene in LD with the peak SNP and has a significant cis-eQTLs (P=1.3E-3) for the ratio of treated to control calcineurin A expression (Figure 3-4a and Figure 3-4b). 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 strongly suggestive cis-eQTL (4.7E-3) as well as a minor allele with an insertion in a splice site in several strains of the HMDP. In addition to calcinerurin A, we identified several other genes with well-established roles in cardiac physiology and pathology within other disease-associated loci. These include the key calcium cycling regulator phospholamban[27] and structural protein sarcoglycan delta[28] as well as other genes which have previously been implicated in cardiac hypertrophy such as Prkag2[29], or cardiac malformation, such as Mospd3[30] (Table 3-2 and Table 3-3).

Conservation of Cardiomyopathy Loci Identified from HMDP

We explored whether the loci we identified overlap with human GWAS results by examining the top twelve previously identified significant and suggestive human loci[6, 7, 31]. The human loci were mapped onto the mouse genome using the NCBI Homologene resource and compared to a set of loci identified for the weight traits based on a slightly relaxed stringency (P<1E-05, MAF>5%) from the HMDP study. We observed six out of twelve human loci, including one of the genome-wide significant loci near USP3, replicating in our study (Table 3-4). We determined that this overlap is highly significant (P=3.5E-4) by permutation analysis. Our result supports the concept that genetic influences in β-adrenergic signaling significantly contribute to polygenic human HF.

Furthermore, sixteen loci for HF-related traits have previously been identified via linkage studies in mice for CHF-related traits. We further observed that seven of these sixteen QTLs overlap with weight loci identified in this study (P=1.2E-3 by permutation analysis) (Table 3-5), despite the fact that some of the linkage studies utilized different hypertrophy-inducing stressors such as a calsequestrin transgene[8], which likely influence distinct HF pathways.

Validation of Abcc6 as Causal Gene for ISO Induced Cardiac Fibrosis

We have identified a locus contributing to ISO-induced fibrosis on chromosome 7 (Figure 3-5, P=7.1E-7). One of the 28 genes within the LD block, Abcc6, has a splice site variation[45] resulting in a premature stop codon that is found in 19 of the strains we analyzed for cardiac fibrosis (KK/HiJ, C3H/HeJ, DBA/2J, 11 BxD, 2 BxH,3 CxB) in the HMDP. In untreated animals, we did not observe any significant difference (P=0.25) between the degrees of fibrosis present in the left ventricle among HMDP strains divided based on Abcc6 genotype. In contrast, we observed a marked increase in cardiac fibrosis in the mice containing the Abcc6 splice mutation allele in response to ISO treatment (P=1E-4) (Figure 3-5b). Abcc6 deficiency is the cause of pseudoxanthoma elasticum, a disorder characterized by progressive tissue calcification[14, 32], and a deficiency of Abcc6 has previously been linked to calcification phenotypes in aged mice by our laboratory[33]. However, the fibrosis phenotype observed in these studies is clearly distinct from that of calcification. In fact, we did not observe a significant association with heart calcification in our study under basal (P=0.8) or ISO

treatment condition (P=0.12) in mice with different Abcc6 genotypes, although we did observe some outlier strains such as KK/HlJ, that had markedly increased calcification after ISO stimulation (Figure 3-5c). Therefore, Abcc6 is a likely candidate gene contributing to ISO induced fibrosis in heart.

To validate the role of Abcc6 in cardiac fibrosis, we studied a previously described Abcc6 knockout mouse carrying a targeted mutation in a C57BL/6J strain background (KO) as well as the wildtype C57BL/6J(Control) mice[14]. We previously reported that the Abcc6 KO mice exhibited increased cardiac calcification beginning from 6 months of age[13]. At three months of age, neither the wild type nor the KO mice exhibited significant differences in calcification as judged by Alizarin Red staining or cardiac fibrosis based on Masson Trichrome staining in the absence of ISO treatment (Figure 3-5d). Consistent with our observations of the entire panel, the ISO treatment significantly increased fibrosis levels in the Abcc6 KO animals without significantly increasing calcification levels in these mice. This result suggests that the age-associated calcification phenotype observed in Abcc6 KO is distinct from the ISO-induced cardiac fibrosis phenotype.

To further establish Abcc6 as a causal gene for ISO induced cardiac fibrosis, we studied transgenic mice carrying the Abcc6 wildtype locus from a C57BL/6J bacterial artificial chromosome (BAC) on the background of the fibrosis-susceptible C3H/HeJ strain. Strain C3H/HeJ mice lack functional Abcc6 due to a splice variation[34]. In the absence of ISO neither C3H/HeJ mice nor C3H/HeJ mice carrying the Abcc6 BAC-

transgene exhibited significant calcification or fibrosis in the heart, whereas after ISO treatment the C3H/HeJ mice but not the Abcc6-BAC transgenic mice exhibited substantial fibrosis and calcification (Figure 3-5e). The differing results between the C57BL/6J KO mice and the naturally mutant C3H/HeJ mice suggests that additional modifier genes are necessary to induce cardiac calcification in Abcc6 KO mice after ISO stimulation, whereas Abcc6 KO is sufficient to cause ISO-induced cardiac fibrosis. The mechanisms by which Abcc6 promotes fibrosis is unknown, but DAVID analysis of genes significantly correlated with Abcc6 expression in heart showed highly significant enrichment for mitochondrial genes (Table 3-6). Systemic factors are clearly involved in the calcification phenotype of Abcc6 deficiency[34].

Discussion

In conclusion, we have used a strategy involving GWAS in a large panel of inbred mice to perform fine mapping of loci contributing to specific pathological features of cardiomyopathy following treatment with ISO[22, 23]. We have combined this strategy with global gene expression analysis in the heart to help identify candidate genes. A significant number of genetic loci revealed from this study are replicated in human GWAS analysis, supporting a conserved genetic network contributing to human heart failure. A number of the loci contain genes known to be involved in cardiomyopathy based on previous biochemical or genetic studies, supporting the validity of this approach to uncover important mechanisms and pathways related to the onset of heart failure. Finally, we validated Abcc6 a candidate GWAS hit, as a novel player in stress-

induced cardiac fibrosis. These findings should complement human studies to identify genes and pathways contributing to this very common and poorly understood disorder.

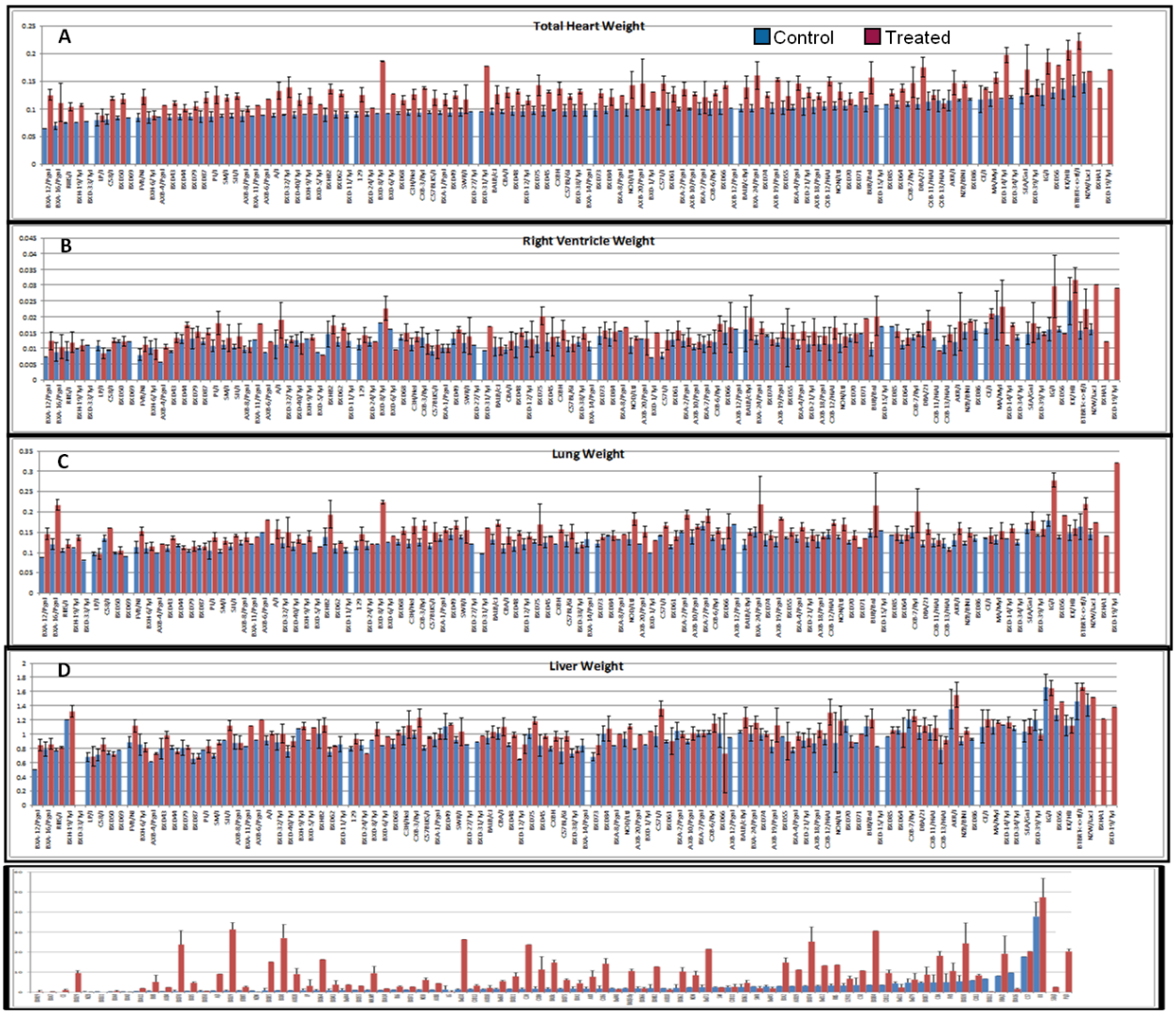


Figure 3-1. Wide variation of heart failure traits between the strains of the HMDP. A) Total Heart Weight B) Right Ventricular Weight C) Lung Weight D) Liver Weight. E) Cardiac Fibrosis. A-D are organized by the untreated heart weight of the strain and display mean +/- standard deviation.

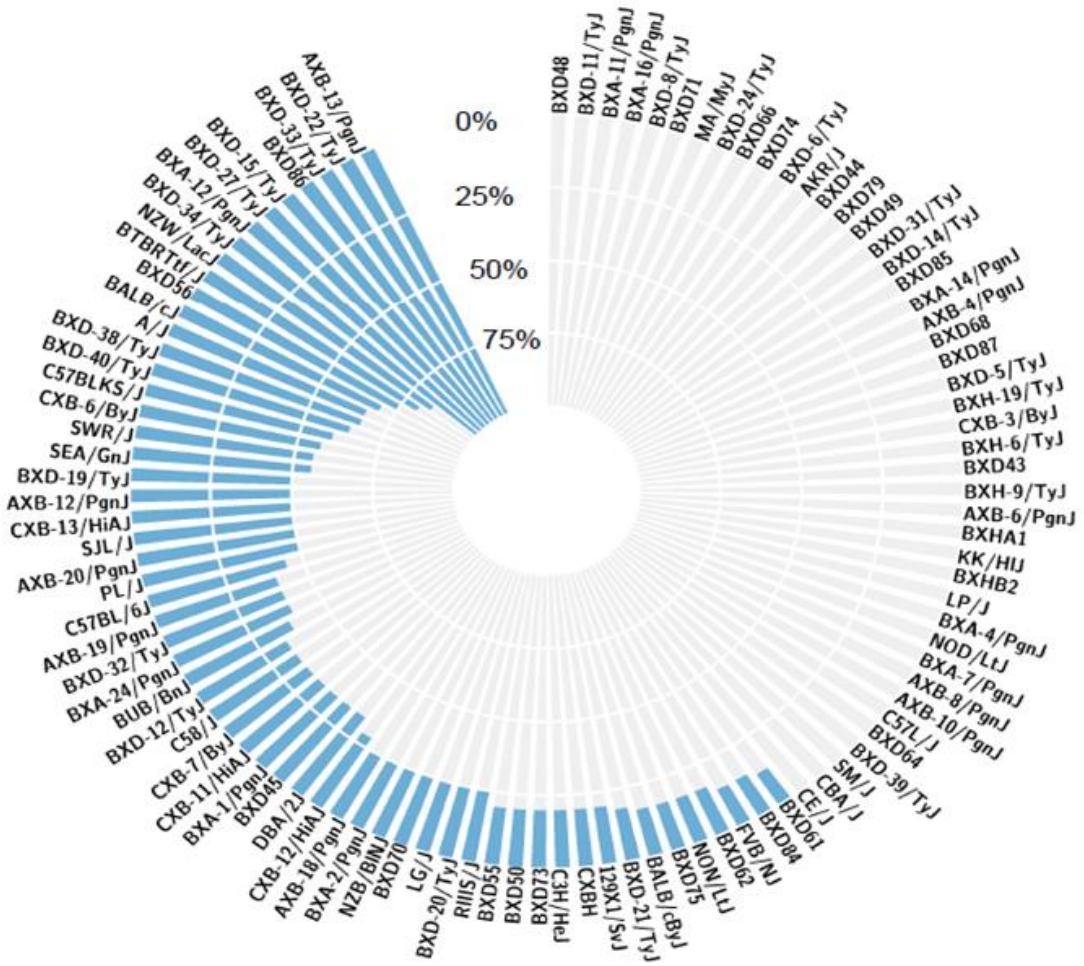


Figure 3-2. Isoproterenol-induced lethality.
The HMDP strains varied in the percent survival in response to ISO. In some strains, survival was 100% even when large numbers (>10) of mice were studied while, in other strains, 100% of mice died, usually within the first week.

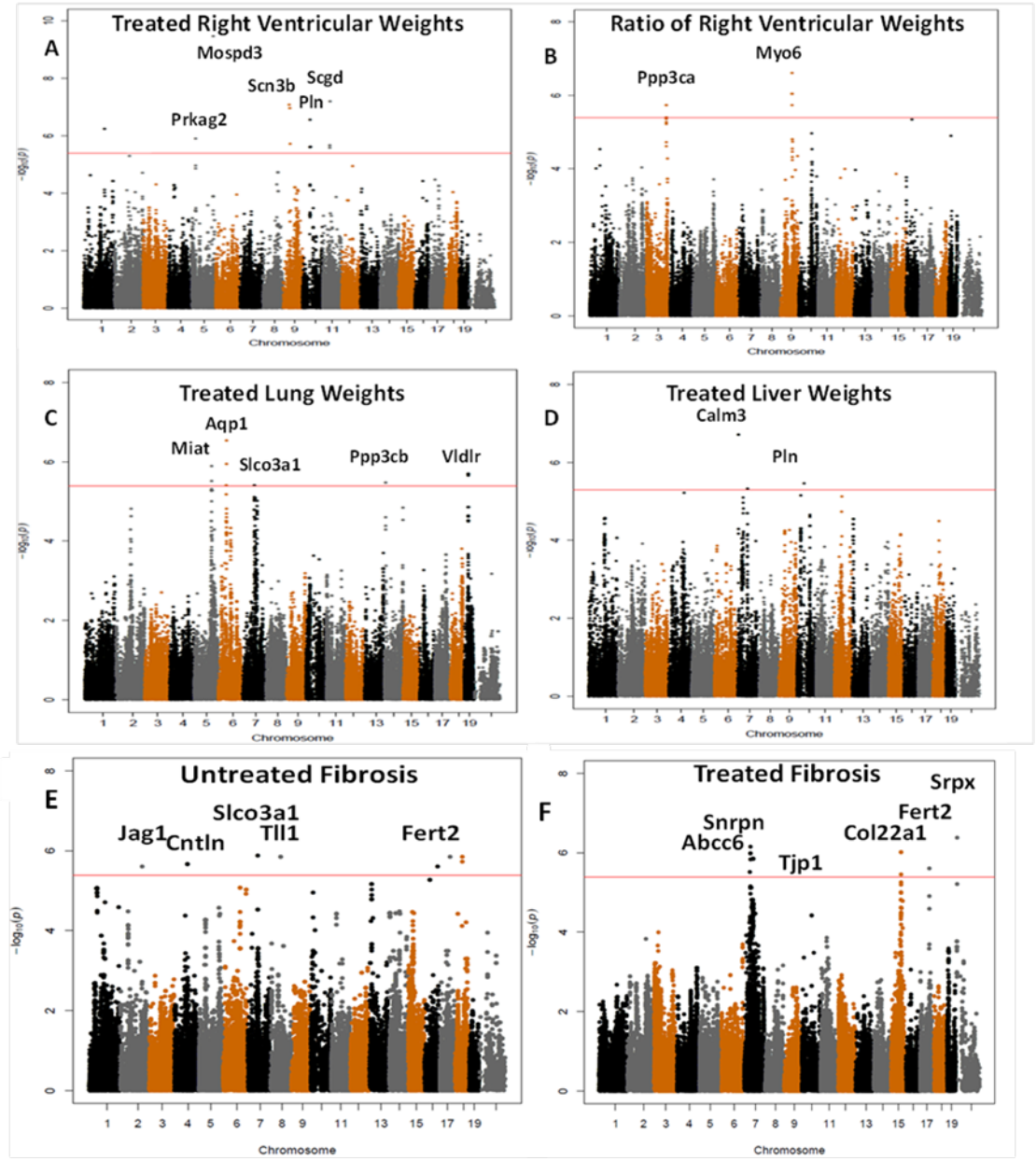


Figure 3-3. Manhattan plots of heart failure traits. A) treated right ventricular weights B) Ratio of treated to untreated right ventricular weights C) treated lung weights D) treated liver weights. E) baseline cardiac fibrosis and F) treated cardiac fibrosis The red line indicates the threshold for suggestive association between a SNP and a phenotype. Proposed candidate genes are indicated by gene symbols above peaks.

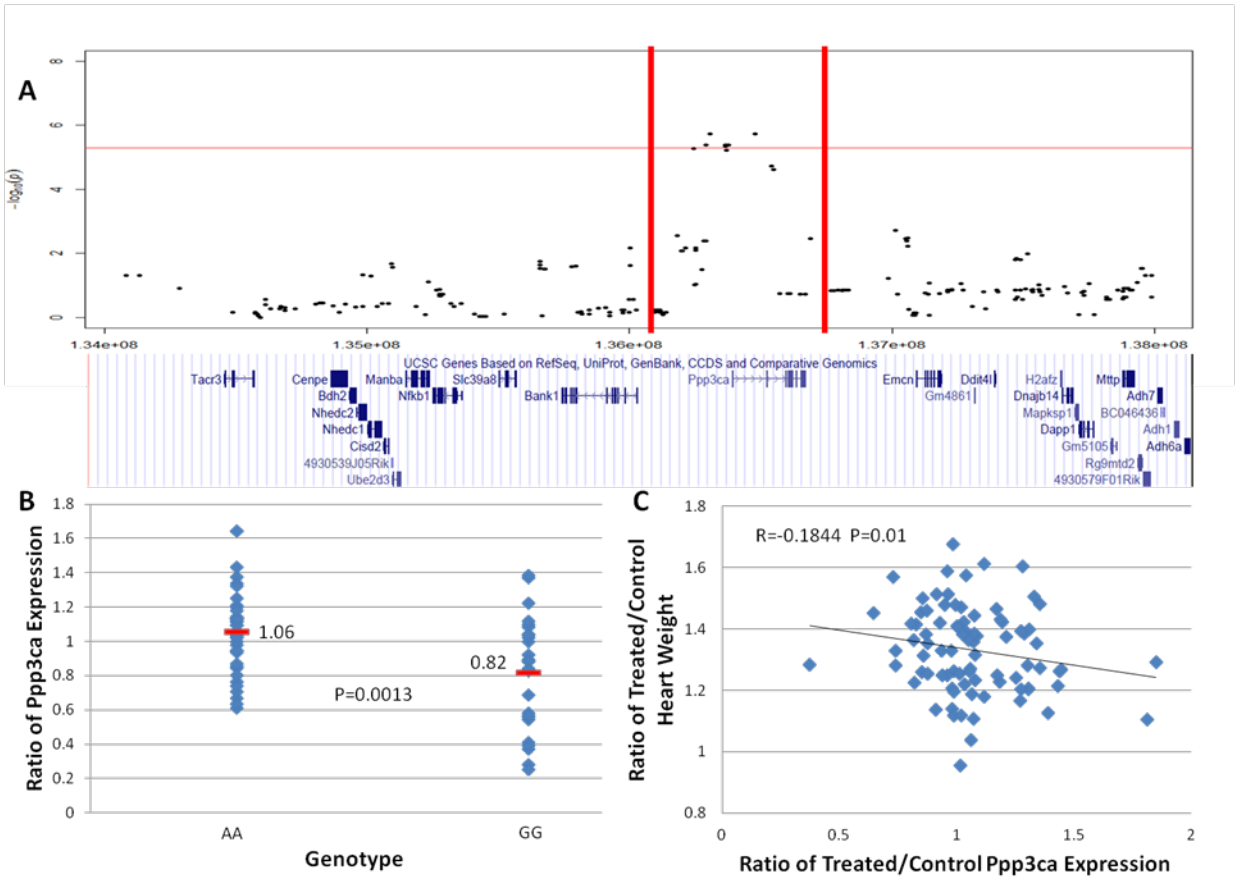


Figure 3-4. Ppp3ca (calcineurin A) is a candidate gene at the chromosome 3 right ventricular weight ratio locus. A) Ppp3ca is the only gene within the locus. The red horizontal line represents the significance threshold, while the red vertical lines indicate the limits of the Linkage disequilibrium block. The dots represent SNPs, plotted in bp along the chromosome with –log10 (p value) on the Y axis. Below are the locations of genes from a genome browser. B) Ppp3ca has a significant cis-eQTL (P=1.3E-3) 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 significant (P=0.01) negative correlation with the ratio of treated to untreated heart weights.

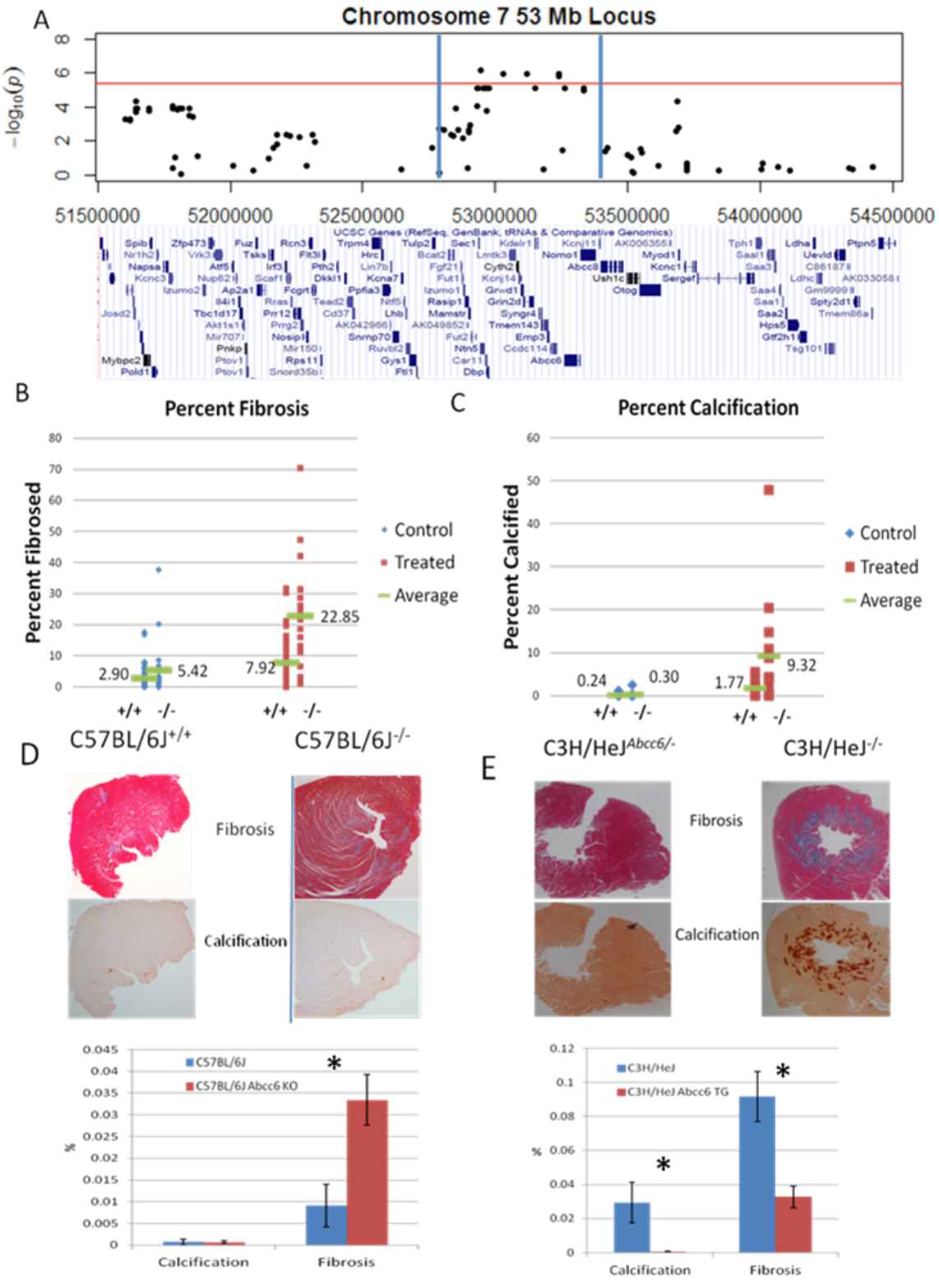

A
Chromosome 7 53 Mb Locus
$-\log_{10}(p)$
0
2
4
6
8
51500000
52000000
52500000
53000000
53500000
54000000
54500000
UCSC Genes (RefSeq, GenBank, tRNAs & Comparative Genomics)
B
Percent Fibrosis
Percent Fibrosed
80
70
60
50
40
30
20
10
0
Control
Treated
Average
22.85
2.90
5.42
7.92
+/+
-/-
+/+
-/-
C57BL/6J+/+
C57BL/6J-/-
C
Percent Calcification
Percent Calcified
60
50
40
30
20
10
0
Control
Treated
Average
0.24
0.30
1.77
9.32
+/+
-/-
+/+
-/-
C3H/HeJAbcc6/-
C3H/HeJ-/-
D
Fibrosis
Calcification
0.045
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
%
C57BL/6J
C57BL/6J Abcc6 KO
*
Calcification
Fibrosis
E
Fibrosis
Calcification
0.12
0.1
0.08
0.06
0.04
0.02
0
%
C3H/HeJ
C3H/HeJ Abcc6 TG
*
*
Calcification
Fibrosis

Figure 3-5. 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)

Table 3-1. List of mouse strains in the study.

Strain	Ctrl	Iso	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 230	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/BINJ	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

Table 3-2. Significant heart failure trait loci identified in HMDP GWA.
The significance threshold is defined as p-value < 4.1E-07. RV, Liver, and Lung represents isoproterenol-treated right ventricular, liver and lung weights at week 3, respectively. RV* represents the ratio of isoproterenol-treated versus control RV weight at week 3. Fibrosis represents isoproterenol-treated LV fibrosis at week 3. For each locus, the peak SNP location, given by chromosome (Chr) and base pair position (Bp) in the NCBI-build-37 assembly, and association p-value are reported, along with the number of genes (N) within the estimated LD block (LD) surrounding the peak SNP and the top candidate gene (Gene). Bold entries represent genes which contain nonsynonymous mutations within the HMDP as reported by the Wellcome Trust Mouse Genome Project, while underlined entries contain significant cis-eQTLs.

Phenotype	Chr	Bp	P-value	LD	N	Gene
Hypertrophic Loci						
RV	5	137934905	3.49E-10	137.93-138.15	11	*Mospd3*
RV	9	40202022	8.41E-08	39.77-40.52	15	*Scn3b*
RV	10	49818583	2.80E-07	48.19- 54.24	22	*Pln*
RV*	9	80542295	2.94E-07	80.00-80.99	2	*Myo6*
Fluid Retention Loci						
Liver	7	15251391	1.93E-07	15.13-18.75	57	*Calm3*
Lung	6	53975816	2.90E-07	53.88-55.57	17	*Aqp1*
Fibrosis Loci						
Fibrosis	X	10277028	4.10E-07	5-12.5	127	*Srpx*

Table 3-3. Suggestive heart failure trait loci identified in HMDP GWA.
The suggestive threshold is defined as p-value < 4.1E-06. RV, Liver, and Lung represents isoproterenol-treated right ventricular, liver and lung weights at week 3, respectively. RV* represents the ratio of isoproterenol-treated versus control RV weight at week 3. Fibrosis represents either control or isoproterenol-treated LV fibrosis at week 3. For each locus, the peak SNP location, given by chromosome (Chr) and base pair position (Bp) in the NCBI-build-37 assembly, and association p-value are reported, along with the number of genes (N) within the estimated LD block (LD) surrounding the peak SNP and the top candidate gene (Gene). Bold entries represent genes which contain nonsynonymous mutations within the HMDP as reported by the Wellcome Trust Mouse Genome Project, while underlined entries contain significant cis-eQTLs.

Phenotype	Chr	Bp	P-value	LD	N	Gene
Hypertrophic Loci						
RV	1	134467906	5.75E-07	133.78-134.53	14	-
RV	5	23873494	1.23E-06	23.82-24.47	20	Prkag2
RV	11	47181489	2.15E-06	46.18-49.3	41	Sgcd
RV*	3	136305887	7.83E-07	136.04-136.79	1	Ppp3ca
RV*	7	142011844	1.41E-06	141.50-144.81	15	Mgmt
Fluid Retention Loci						
Liver	10	49468021	3.48E-06	48.19-54.24	22	Pln
Lung	5	111867706	1.28E-06	110.87-112.87	22	Miat
Lung	7	81841621	3.88E-06	79.8-82.2	6	Slco3a1 Iqgap1
Lung	14	14941056	3.34E-06	8.5-21.5	50	Ppp3cb
Lung	19	27061190	2.01E-06	26.68-27.43	1	Vldlr
Baseline Fibrosis Loci						
Fibrosis	2	139163425	2.51E-06	13.7-14.0	6	Jag1
Fibrosis	4	84420058	2.20E-06	84-85	2	*Cntln*
Fibrosis	7	73365047	1.31E-06	72.3-74.3	7	*Tjp1*
Fibrosis	8	64382692	1.43E-06	63-66	12	*Tll1*
Fibrosis	16	83682440	2.52E-06	82.5-84.25	0	-
Fibrosis	17	64376735	1.43E-06	64.2-65.8	4	*Fert2*
Fibrosis	18	47786513	1.87E-06	47.2-48.2	3	-
Isoproterenol Treated Fibrosis Loci						
Fibrosis	7	52946331	7.11E-07	52.85-53.42	28	*Abcc6*
Fibrosis	7	68593223	1.40E-06	60.5-69.5	18	*Snrpn*
Fibrosis	7	73365047	1.40E-06	72.3-74.3	7	*Tjp1*
Fibrosis	15	69907056	9.60E-07	68.4-71.4	3	*Col22a1*
Fibrosis	17	64833212	2.47E-06	64.2-65.8	4	*Fert2*

Table 3-4. Significant overlap of HMDP and human GWA loci for heart failure 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 in human are matched in the HMDP study. DCM represents dilated cardiomyopathy, LVM represents Left ventricular, HF represents heart failure, and Death represents mortality among HF patients. Syntenic region represents the mouse region that is syntenic to the loci reported in the human studies[6, 7, 31, 35]. The genomic locations are notated by chromosome and bp position in Mb. RV represent right ventricular weight, TH represents total heart weight, RV* represents the difference in RV weight between isoproterenol-treated and control animals at week 3.

Human					HMDP		
Study	Trait	Peak SNP	P-value	Syntenic Region	Trait	P-value	SNP Location
Ellinor	DCM	-	LOD 8.2	chr1:187-193	RV	1.56E-10	chr1:186.73
Parsa	LVM	rs12757165	1.00E-07	chr1:189.7	TH	1.08E-06	chr1:186.73
Parsa	LVM	rs16830359	1.00E-07	chr4:120	RV*	8.99E-07	chr4:122.0
Parsa	LVM	rs17636733	2.00E-07	chr7:66.5	RV*	9.57E-07	chr7:64.06
Smith	HF	rs10519210	1.00E-08	chr9:66.25	RV*	4.93E-06	chr9:66.33
Morrison	Death	rs12638540	3.00E-07	chr9:114.5	RV*	9.97E-06	chr9:116.8

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. HR represents heart rate, HW represents heart weight, DCM represents dilated cardiomyopathy, and HF represents heart faiure. LV represents the LV weight in controls, TH' represents isoproterenol-treated total heart weight at week 3, RA represents right atrial weight in controls, and RA* represents the difference in RA between isoproterenol-treated and control animals at week 3.

Linkage			HMDP		
QTL	Phenotype	Location	Phenotype	P-value	Location
Hrq3	HR	chr1:90.5-170	LV	8.47E-06	chr1:154.47
Hrtq1	HW	chr2:64.5-141.8	TH'	6.45E-06	chr2:79.63
Hrq1	HR	chr2:64.6-159.4	TH'	6.45E-06	chr2:79.63
Cmn1	DCM	chr9:110.4-118.2	RA*	9.97E-06	chr9:116.8
Hrtq3	HW	chr10:26-88.5	LV	1.85E-06	chr10:41.11
Hrtfm6	HF	chr13:108.1-134.2	RA	6.75E-07	chr13:117.73
Hrtfm4	HF	chr18:29.6-65.3	RA*	2.46E-07	chr18:47.78

Table 3-6. DAVID enrichment of all genes significantly correlated with Abcc6 expression in ISO treated mouse heart.

No. represents the number of genes in each term category. Enrichment represents fold enrichment by DAVID. Adj. p-value represents the Benjamini-Hochberg adjusted p-value.

Term	Count	%	No.	Enrichment	Adj. p-value
mitochondrion	97	18.3	1322	2.4	4.65E-15
transit peptide	40	7.5	457	3.1	1.36E-07
acetylation	104	19.6	2325	1.6	1.51E-04
transferase	65	12.3	1385	1.7	0.005
mitochondrial envelope	27	5.1	391	2.3	0.006
ribosome	16	3	192	2.8	0.02
coenzyme metabolic process	16	3	143	4	0.02
cytoplasm	114	21.5	3029	1.3	0.04

Bibliography

1. Mudd JO, Kass DA. Tackling heart failure in the twenty-first century. *Nature*. 2008;451:919-928

2. Beltrami CA, Finato N, Rocco M, Feruglio GA, Puricelli C, Cigola E, Quaini F, Sonnenblick EH, Olivetti G, Anversa P. Structural basis of end-stage failure in ischemic cardiomyopathy in humans. *Circulation*. 1994;89:151-163

3. Frangogiannis NG. Regulation of the inflammatory response in cardiac repair. *Circ Res*. 2012;110:159-173

4. Kong P, Christia P, Frangogiannis NG. The pathogenesis of cardiac fibrosis. *Cell Mol Life Sci*. 2014;71:549-574

5. Villard E, Perret C, Gary F, Proust C, Dilanian G, Hengstenberg C, Ruppert V, Arbustini E, Wichter T, Germain M, Dubourg O, Tavazzi L, Aumont MC, DeGroote P, Fauchier L, Trochu JN, Gibelin P, Aupetit JF, Stark K, Erdmann J, Hetzer R, Roberts AM, Barton PJ, Regitz-Zagrosek V, Aslam U, Duboscq-Bidot L, Meyborg M, Maisch B, Madeira H, Waldenstrom A, Galve E, Cleland JG, Dorent R, Roizes G, Zeller T, Blankenberg S, Goodall AH, Cook S, Tregouet DA, Tiret L, Isnard R, Komajda M, Charron P, Cambien F. A genome-wide association study identifies two loci associated with heart failure due to dilated cardiomyopathy. *Eur Heart J*. 2011;32:1065-1076

6. Morrison AC, Felix JF, Cupples LA, Glazer NL, Loehr LR, Dehghan A, Demissie S, Bis JC, Rosamond WD, Aulchenko YS, Wang YA, Haritunians T, Folsom AR, Rivadeneira F, Benjamin EJ, Lumley T, Couper D, Stricker BH, O'Donnell CJ, Rice KM, Chang PP, Hofman A, Levy D, Rotter JI, Fox ER, Uitterlinden AG, Wang TJ, Psaty BM, Willerson JT, van Duijn CM, Boerwinkle E, Witteman JCM, Vasan RS, Smith NL. 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*. 2010;3:248-255

7. Parsa A, Chang YP, Kelly RJ, Corretti MC, Ryan KA, Robinson SW, Gottlieb SS, Kardia SL, Shuldiner AR, Liggett SB. Hypertrophy-associated polymorphisms ascertained in a founder cohort applied to heart failure risk and mortality. *Clin Transl Sci*. 2011;4:17-23

8. Le Corvoisier P, Park HY, Carlson KM, Marchuk DA, Rockman HA. Multiple quantitative trait loci modify the heart failure phenotype in murine cardiomyopathy. *Hum Mol Genet*. 2003;12:3097-3107

9. McDermott-Roe C, Ye J, Ahmed R, Sun XM, Serafin A, Ware J, Bottolo L, Muckett P, Canas X, Zhang J, Rowe GC, Buchan R, Lu H, Braithwaite A, Mancini M, Hauton D, Marti R, Garcia-Arumi E, Hubner N, Jacob H, Serikawa T, Zidek V, Papousek F, Kolar F, Cardona M, Ruiz-Meana M, Garcia-Dorado D, Comella JX, Felkin LE, Barton PJ, Arany Z, Pravenec M, Petretto E, Sanchis D, Cook SA. Endonuclease g is a novel determinant of cardiac hypertrophy and mitochondrial function. *Nature*. 2011;478:114-118

10. Wheeler FC, Fernandez L, Carlson KM, Wolf MJ, Rockman HA, Marchuk DA. Qtl mapping in a mouse model of cardiomyopathy reveals an ancestral modifier allele affecting heart function and survival. *Mamm Genome*. 2005;16:414-423

11. Wheeler FC, Tang H, Marks OA, Hadnott TN, Chu PL, Mao L, Rockman HA, Marchuk DA. Tnni3k modifies disease progression in murine models of cardiomyopathy. *PLoS Genet*. 2009;5:e1000647

12. Bennett BJ, Farber CR, Orozco L, Min Kang H, 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. A high-resolution association mapping panel for the dissection of complex traits in mice. *Genome Research*. 2010;20:281-290

13. Martin LJ, Lau E, Singh H, Vergnes L, Tarling EJ, Mehrabian M, Mungrue I, Xiao S, Shih D, Castellani L, Ping P, Reue K, Stefani E, Drake TA, Bostrom K, Lusis AJ. Abcc6 localizes to the mitochondria-associated membrane. *Circ Res*. 2012;111:516-520

14. Mungrue IN, Zhao P, Yao Y, Meng H, Rau C, Havel JV, Gorgels TG, Bergen AA, MacLellan WR, Drake TA, Bostrom KI, Lusis AJ. Abcc6 deficiency causes increased infarct size and apoptosis in a mouse cardiac ischemia-reperfusion model. *Arterioscler Thromb Vasc Biol*. 2011;31:2806-2812

15. Berk BC, Fujiwara K, Lehoux S. Ecm remodeling in hypertensive heart disease. *J Clin Invest*. 2007;117:568-575

16. Bronson RT. Cross sectional pathology of aging rodents. *Genetic effects of aging*. Caldwell, New Jersey: The Telford Press; 1990:289-358.

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

18. Smyth GK. Limma: Linear models for microarray data. In: Gentleman R, ed. *Bioinformatics and computational biology solutions using r and bioconductor*. New York: Springer Science+Business Media; 2005:xix, 473 p.

19. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical bayes methods. *Biostatistics*. 2007;8:118-127

20. Parks BW, Nam E, Org E, Kostem E, Norheim F, Hui ST, Pan C, Civelek M, Rau CD, Bennett BJ, Mehrabian M, Ursell LK, He A, Castellani LW, Zinker B, Kirby M, Drake TA, Drevon CA, Knight R, Gargalovic P, Kirchgessner T, Eskin E, Lusis AJ. Genetic control of obesity and gut microbiota composition in response to high-fat, high-sucrose diet in mice. *Cell Metab*. 2013;17:141-152

21. Dorn GW, 2nd, Liggett SB. Pharmacogenomics of beta-adrenergic receptors and their accessory signaling proteins in heart failure. *Clin Transl Sci*. 2008;1:255-262

22. Breckenridge R. Heart failure and mouse models. *Dis Model Mech*. 2010;3:138-143

23. Zhang X, Szeto C, Gao E, Tang M, Jin J, Fu Q, Makarewich C, Ai X, Li Y, Tang A, Wang J, Gao H, Wang F, Ge XJ, Kunapuli SP, Zhou L, Zeng C, Xiang KY, Chen X. Cardiotoxic and cardioprotective features of chronic beta-adrenergic signaling. *Circ Res*. 2013;112:498-509

24. Berthonneche C, Peter B, Schupfer F, Hayoz P, Kutalik Z, Abriel H, Pedrazzini T, Beckmann JS, Bergmann S, Maurer F. Cardiovascular response to beta-adrenergic blockade or activation in 23 inbred mouse strains. *PLoS One*. 2009;4:e6610

25. Flint J, Eskin E. Genome-wide association studies in mice. *Nat Rev Genet*. 2012;13:807-817

26. Yalcin B, Wong K, Agam A, Goodson M, Keane TM, Gan X, Nellaker C, Goodstadt L, Nicod J, Bhomra A, Hernandez-Pliego P, Whitley H, Cleak J, Dutton R, Janowitz D, Mott R, Adams DJ, Flint J. Sequence-based characterization of structural variation in the mouse genome. *Nature*. 2011;477:326-329

27. Schonberger J, Seidman CE. Many roads lead to a broken heart: The genetics of dilated cardiomyopathy. *Am J Hum Genet*. 2001;69:249-260

28. Bauer R, Macgowan GA, Blain A, Bushby K, Straub V. Steroid treatment causes deterioration of myocardial function in the delta-sarcoglycan-deficient mouse model for dilated cardiomyopathy. *Cardiovasc Res*. 2008;79:652-661

29. Banerjee SK, McGaffin KR, Huang XN, Ahmad F. Activation of cardiac hypertrophic signaling pathways in a transgenic mouse with the human prkag2 thr400asn mutation. *Biochim Biophys Acta*. 2010;1802:284-291

30. Pall GS, Wallis J, Axton R, Brownstein DG, Gautier P, Buerger K, Mulford C, Mullins JJ, Forrester LM. A novel transmembrane msp-containing protein that plays a role in right ventricle development. *Genomics*. 2004;84:1051-1059

31. Ellinor PT, Sasse-Klaassen S, Probst S, Gerull B, Shin JT, Toeppel A, Heuser A, Michely B, Yoerger DM, Song BS, Pilz B, Krings G, Coplin B, Lange PE, Dec GW, Hennies HC, Thierfelder L, MacRae CA. A novel locus for dilated cardiomyopathy, diffuse myocardial fibrosis, and sudden death on chromosome 10q25-26. *J Am Coll Cardiol*. 2006;48:106-111

32. Campens L, Vanakker OM, Trachet B, Segers P, Leroy BP, De Zaeytijd J, Voet D, De Paepe A, De Backer T, De Backer J. Characterization of cardiovascular involvement in pseudoxanthoma elasticum families. *Arterioscler Thromb Vasc Biol*. 2013;33:2646-2652

33. Meng H, Vera I, Che N, Wang X, Wang SS, Ingram-Drake L, Schadt EE, Drake TA, Lusis AJ. Identification of abcc6 as the major causal gene for dystrophic cardiac calcification in mice through integrative genomics. *Proc Natl Acad Sci U S A*. 2007;104:4530-4535

34. Jiang Q, Oldenburg R, Otsuru S, Grand-Pierre AE, Horwitz EM, Uitto J. Parabiotic heterogenetic pairing of abcc6-/-/rag1-/- mice and their wild-type counterparts halts ectopic mineralization in a murine model of pseudoxanthoma elasticum. *Am J Pathol*. 2010;176:1855-1862

35. Smith NL, Felix JF, Morrison AC, Demissie S, Glazer NL, Loehr LR, Cupples LA, Dehghan A, Lumley T, Rosamond WD, Lieb W, Rivadeneira F, Bis JC, Folsom AR, Benjamin E, Aulchenko YS, Haritunians T, Couper D, Murabito J, Wang YA, Stricker BH, Gottdiener JS, Chang PP, Wang TJ, Rice KM, Hofman A, Heckbert SR, Fox ER, O'Donnell CJ, Uitterlinden AG, Rotter JI, Willerson JT, Levy D, van Duijn CM, Psaty BM, Witteman JCM, Boerwinkle E, Vasan RS. 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*. 2010;3:256-266

4 Genetic Dissection of Cardiac Remodeling in an Isoproterenol-induced Heart Failure Mouse Model

Background

Heart failure (HF) affects about 5.1 million people in the United States and 23 million people worldwide[1, 2]. Although current therapies have been demonstrated to slow down HF progression and improve mortality, HF remains a lethal condition with 5- and 10-year survival rates of less than 50% and 30% from the time of diagnosis[3-6]. A variety of etiologic factors, including coronary artery disease, hypertension, valvular disease, alcohol, chemotherapy and others, contribute to HF. Studies of rare familial cases have identified dozens of genes that can cause HF but whether more subtle mutations of these and other genes contribute to HF progression is unknown. Irrespective of the primary insult, compensatory sympathetic and renin-angiotensin activation augment heart rate, contractility and fluid retention to maintain adequate cardiac output and preserve organ function, while leading to chronic maladaptive cellular growth and irreversible myocardial injury, furthering HF progression[7]. Understanding how common genetic variations modify HF progression will provide insights into the management and the design of viable therapeutics to improve lifestyle and survival of all HF patients.

The genetic basis of the HF spectrum is complex. The unbiased genome-wide association study (GWAS) design is well suited to detect the effects of genetic variations on complex traits[8]. However, a number of human HF GWAS performed to-

date have had limited success. For example, a HF GWAS in humans was performed by meta-analyses of 4 community-based cohorts involving nearly 24,000 subjects. In spite of its scale, only two loci were identified to be significantly associated with incident HF, explaining a very small fraction of the variance[9]. Villard et al. reported the first GWAS based on sporadic dilated cardiomyopathy, which included 1179 cases and 1108 controls in several European populations, and identified only two associated loci[10]. Recently, a genome-wide association study of cardiac structure and function in 6,765 African Americans identified 4 loci related to left ventricular mass, interventricular septal wall thickness, LV internal diastolic diameter, and ejection fraction[11]. In spite of large amounts of resources channeled into human HF GWAS, the extent of phenotypic, environmental and genetic heterogeneity as well as the limitations in phenotypic and endophenotypic data have hampered the detection of meaningful signals that are broadly applicable to diverse populations or informative about the underlying mechanisms.

Genetic studies with animal models provide a means of overcoming some of the challenges in humans. Particularly informative have been reverse genetic studies of candidate genes and pathways, utilizing engineered mouse models[12]. Also, studies of natural variations of mice and rats, some involving sensitized models, have resulted in the identification of novel pathways contributing to HF[13-16]. A major difficulty of the latter studies, however, has been the poor resolution of classical linkage, making the identification of underlying causal genes a very difficult and laborious process[17]. Recently, with the development of high-density genotyping and sequencing in rodents,

relatively high-resolution association mapping approaches, analogous to human GWAS, have become feasible[17, 18]. Our group has pioneered a resource termed the Hybrid Mouse Diversity Panel (HMDP), a panel of 100+ strains of inbred mice that have either been sequenced or densely genotyped and display natural inter-strain genetic variation, allowing a mapping resolution more than an order of magnitude higher than traditional crosses[19]. The method combines the use of classic inbred strains for mapping resolution and recombinant inbred strains for power, and has been used to successfully identify many genes and loci involved in obesity, lipid, bone, and behavioral traits[18, 20-22]. Because the HMDP strains are renewable, the resource is well suited to the application of systems genetics approaches, involving the integration of high throughput molecular phenotypes, such as expression array data, with clinical phenotypes.

Cardiac remodeling in the setting of HF involves structural changes, such as hypertrophy, fibrosis and dilatation, and multiple abnormalities of cellular and molecular function. Most HF therapies with long-term benefit result in reverse remodeling. Thus, cardiac remodeling traits are not only important prognostic indicators but likely therapeutic targets[23]. Echocardiography has emerged as a powerful and noninvasive tool to serially monitor cardiac structure and function in murine injury models[24, 25]. Isoproterenol, a non-selective β-adrenergic agonist, has been used widely in laboratories to mimic the heart failure state in experimental animals[26, 27]. We now report the genetic analysis of isoproterenol-induced cardiac remodeling as measured by echocardiography in the HMDP resource. We demonstrate high heritability of a number of cardiac remodeling traits. Using association analysis, we identify 3 genome-wide

significant and 13 suggestive loci with high resolution (many loci less than a Mb in size). We also perform expression array profiling of LV tissues, both before and after isoproterenol treatment, from the entire panel to understand the genetic control of gene expression, gene expression correlation to phenotype, and to prioritize candidate genes.

Methods

Mice. Breeding pairs from the Hybrid Mouse Diversity Panel (HMDP) inbred strains were obtained from the Jackson Laboratory (Bar Harbor, ME, USA). Eight- to ten-week-old female offspring from the following 105 mouse strains were used, including 30 classical inbred strains (129X1/SvJ, A/J, AKR/J, BALB/cByJ, BALB/cJ, BTBRT<+>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/LtJ, NON/LtJ, NZB/BINJ, NZW/LacJ, PL/J, RIIIS/J, SEA/GnJ, SJL/J, SM/J, SWR/J) and 75 recombinant inbred lines [RI (number of strains) – AXB (9), BXA (10), BXD (44), 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.

Chronic β-adrenergic stimulation and tissue collection. Isoproterenol (30 μg per g body weight per day) was administered for 21 days using intra-abdominally implanted osmotic minipumps from ALZET (Cupertino, CA, USA) in approximately 4 mice per

strain. At the end of the protocol, LV tissues were harvested and frozen immediately in liquid nitrogen.

Echocardiography. Transthoracic echocardiograms were performed using the Vevo 770 ultrasound system (VisualSonics, Inc., Toronto, ON, Canada). Inhaled isoflurane (1.25% during induction and 1% during maintenance) was administered to ensure adequate sedation while maintaining heart rate above 450 beats per minute. A parasternal long-axis B-mode image was obtained. The maximal long-axis of the left ventricle (LV) was positioned perpendicular to the ultrasound beam. A 90° rotation of the ultrasound probe at the papillary muscle level was performed to obtain a parasternal short-axis view of the LV. A M-mode image was captured to document LV dimensions. Then a semi-apical long-axis view of the LV was obtained. The LV ejection time, E and A wave velocities were obtained from this view using pulse wave Doppler. Images were saved for analysis at a later time point using the Vevo 770 cardiac analysis package. In summary, a baseline echocardiogram was performed on all of the mice. Among control mice, a second echocardiogram was performed in 70 mouse strains at week 3. In isoproterenol-treated mice, serial echocardiograms were performed at 1, 2, and 3 weeks. A single operator, who followed a standard operating protocol detailed above, performed all of the echocardiograms. Saved images were analyzed at a later time point by a single observer who was blinded to mouse strains.

Heritability calculation. One-way analysis of variance was used to calculate inter-strain and intra-strain variances for each phenotype and time point. Heritability H^2 was calculated as the following:

$$H^2 = \frac{Var(G)}{Var(G) + Var(E)}$$

where Var(G) is the genetic (i.e. the inter-strain) variance and Var(E) is the environmental (i.e. the intra-strain) variance of the phenotype.

RNA extraction and expression array profiling. Frozen LV tissues were homogenized in QIAzol Lysis Reagent prior to RNA isolation using RNeasy columns (QIAGEN, Valencia, CA, USA). RNA quality was assessed using the Bioanalyzer RNA kits (Agilent Technologies, Santa Clara, CA, USA). Expression profiling of samples with RIN ≥ 7.0 was performed using Illumina MouseRef-8 v2.0 Expression BeadChip arrays (Illumina, Inc., San Diego, CA, USA). Raw data were deposited in the Gene Expression Omnibus (GEO) online database http://www.ncbi.nlm.nih.gov/geo/ (Accession GSE48760). To minimize artifacts due to single nucleotide polymorphisms (SNPs), we excluded probes that aligned to sequences containing known SNPs. Background correction and quantile normalization of the image data was performed using the neqc method from the R package limma[28]. Hierarchical clustering of samples was performed to exclude outlier samples using the R package WGCNA[29]. Expression profiles of biological replicates were averaged by strain and treatment. In total expression profiles for 90 control and 91 isoproterenol strains (including 82 strains with matching control and isoproterenol samples) were included in downstream analyses.

Differential gene expression. Using the R package limma, moderated t-statistics and the associated p values were calculated. Multiple testing was corrected by controlling for false discovery rate using the Benjamini-Hochberg procedure[30]. Probes with $\log_2$-fold change > 0.2 and adjusted p-value < 0.05 were considered significantly differentially expressed. Functional analysis of probe lists was performed using DAVID to identify pathways and cellular processes enriched in genes differentially regulated by isoproterenol stimulation[31, 32].

Genotypes. Classical and recombinant inbred mouse strains were genotyped at the Jackson Laboratory using the JAX Mouse Diveristy Genotyping Array[33]. To select for informative and high quality SNPs, each SNP was filtered for > 5% minor allele frequency and < 10% missing values among the strains using plink[34].

Genome-wide significance threshold. Random phenotypes with a correlation structure derived from the genetic background of the 105 strains were generated using the R package emmaPowerSim and the genome-wide significant threshold of association at the 5% α level was estimated by simulation[35].

Association mapping. We used Factored Spectrally Transformed Linear Mixed Models (FaST-LMM) to test for association while accounting for the population structure and genetic relatedness among strains as previously described[36]. Briefly, we applied the following linear mixed model:

$$y = \mu + x\beta + u + e,$$

where μ represents mean, x represents SNP, β represents the SNP effect, e represents error, and u represents random effects due to genetic relatedness with Var(u) = $\sigma_g^2 K$ and Var(e) = σ_e^2, where K represents identity-by-state matrix across all genotypes in the HMDP panel. A restricted maximum likelihood estimate for σ_g^2 and σ_e^2 was computed and the association mapping was performed based on the estimated variance component with a standard F test to test β. Association mapping was performed for clinical traits, such as raw and baseline body weight adjusted echocardiographic measurements at each time point and the change in clinical traits compared to baseline, and for gene expression traits, under baseline and isoproterenol-treated conditions, to define clinical quantitative trait loci (cQTL) and expression quantitative trait loci (eQTL), respectively.

Linkage disequilibrium. Genomic boundaries around peak association SNPs were chosen based on flanking SNPs with p-values < 1×10^{-5} that were no more than 2 Mb apart between nearest consecutive pairs. Peak SNPs without any neighboring SNPs at p-value < 1×10^{-5} were excluded. Linkage disequilibrium (LD) between peak SNPs and flanking SNPs were calculated and visualized by plotting regional plots using LocusZoom[37].

Candidate gene prioritization. Genes within LD ($r^2 > 0.8$ or $r^2 > 0.9$) of the peak SNPs were examined for coding sequence and splice region variations using the Wellcome Trust Mouse Genomes Project sequencing database[38]. SIFT score, a functional annotation score for non-synonymous variants, was noted whenever available[39]. The

expression profiles of genes within LD or nearby the peak SNPs were further examined for the presence of cis-expression quantitative trait loci (cis-eQTL). When a transcript's cis-eQTL coincides with the clinical trait locus and is correlated with the trait, a causal relationship between the locus, the transcript and the trait may be inferred.

Statistical analysis. The standard R package was used to performed t-test, correlation, analysis of variance, and hierarchical clustering. The R package WGCNA was used to calculate bicor and associated p-value[40].

Results

Study Design

We characterized 105 classical and recombinant inbred mouse strains by directly obtaining multiple weight and echocardiographic measurements under the baseline condition and in response to chronic administration of isoproterenol for 3 weeks. Briefly, isoproterenol was administered at a dose of 30 mg/kg body weight/day for 21 days using intra-abdominally implanted minipumps in four mice, 8-10 weeks of age, from each strain. In addition, 2-4 control mice were examined. A baseline echocardiogram was performed on all the mice. In isoproterenol-treated mice, serial echocardiograms were performed at 1, 2, and 3 weeks of treatment. In control mice, a repeat echocardiogram was performed at week 3 as an internal control. At the end of the protocol, relevant tissues were weighed. Global gene expression profiling of LV tissues

from control and isoproterenol-treated mice was performed to identify genes whose expression was correlated to HF traits and to identify expression quantitative trait loci (eQTL) for purposes of prioritizing candidate genes. This manuscript focuses on the analyses pertaining to echocardiographic measures of clinical significance, including interventricular septal wall thickness at end diastole (IVSd), LV internal diameter at end diastole (LVIDd), fractional shortening (FS), and LV mass (LVM).

Cardiac structure and function across the HMDP

To minimize inter-operator and inter-observer variations, all echocardiograms were performed by a single operator and interpreted by the same observer who was blinded to strain name and treatment assignment. Inhaled isoflurane was titrated to achieve adequate sedation, while maintaining target heart rate above 475 bpm[41]. The reproducibility of our echocardiographic measurements was assessed by performing echocardiograms at baseline and at week 3 in control mice from 70 strains. IVSd, LVIDd, and LVM were not statistically different between the 2 time points, although FS was 2.3% higher at week 3 compared to baseline (p=0.00176), possibly due a more optimized sedation dosing at week 3 due to operator learning curve or the effects of co-housing with isoproterenol-treated animals (Table 4-1). Importantly, LVM as calculated using echocardiographic measures was significantly correlated to independently measured LV weight (r=0.80 in control and r=0.78 in isoproterenol mice at week 3), validating our echocardiographic measures externally. Pearson correlation of baseline body weight adjusted residuals for IVSd, LVIDd, FS, and LVM revealed significant

correlations among some of the echocardiographic measures (Figure 4-1). Study sample characteristics across the mouse panel are reported in Table 4-2.Study sample characteristics.. In short, chronic isoproterenol infusion resulted in early IVSd and FS increases at week 1 and progressive LVIDd and LVM increases throughout the 3-week time course. In addition, we observed striking variations in the measures of cardiac structure and function among the mouse strains (Figure 4-2). One-way analyses of variance demonstrated significant inter-strain versus intra-strain variances, indicating that cardiac structure and function have a strong genetic component, with heritability calculated to be from 64 to 84% (Table 4-3). Our heritability estimates are much higher than the estimates found in human studies, emphasizing the importance of genetics as well as gene-by-environment interactions in controlling cardiac traits[42].

Gene expression profiles across the HMDP

We carried out expression array profiling of LV tissues from the mice and obtained control samples for 90 strains and isoproterenol samples for 91 strains with matching control and isoproterenol samples in 82 strains. In total 1,502 of the 18,335 probes (8.2%) were differentially expressed at > 15% change and an adjusted p-value of 0.05, including 840 up-regulated and 662 down-regulated probes. We found that chronic administration of isoproterenol induced a number of changes in gene expression downstream of the β-adrenergic receptor signaling. Notably, there was a 17% reduction in β1-adrenoceptor expression (*Adrb1*, adjusted p-value = 2.56×10^{-5}) and a 16% increase in the expression of the inhibitory G-protein G_i (*Gnai2*, adjusted p-

value = 4.47 x 10^{-7}), reminiscent of cardiomyopathic hearts with diminished β-adrenergic responsiveness secondary to chronic β-adrenergic overdrive. Gene ontology analysis of the up- and down-regulated probes was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID)[31]. The up-regulated probes were most enriched for secreted signal glycoprotein, proteinaceous extracellular matrix (ECM), angiogenesis, polysaccharide binding, actin cytoskeleton, vacuole, response to wounding, chemokine signaling pathway, and epidermal growth factor (EGF)-like calcium-binding, prenylation and growth factor binding. The down-regulated probes were most enriched in mitochondrial matrix, mitochondrial inner membrane, and flavoprotein. These findings provide strong evidence in support of isoproterenol-induced cellular and ECM remodeling.

To identify transcripts and processes that are important for measures of cardiac remodeling, phenotype to expression correlations were performed under baseline and week 3 time points. DAVID gene ontology analysis was performed on the top 1000 correlated transcripts for each phenotype-expression pair. Under the baseline condition, transcripts correlated with LVIDd were modestly enriched for actin-binding, cell junction and synapse. At week 3 of isoproterenol, transcripts correlated with LVID and LVM were highly enriched for proteinaceous ECM, secreted signal glycoprotein, actin cytoskeleton, collagen, ECM receptor interaction, EGF, polysaccharide binding, and mitochondrial membrane. These findings suggest that isoproterenol-induced cardiac remodeling as measured by echocardiography corresponds to cellular and extracellular changes found in HF.

Association mapping of cardiac remodeling

Association analyses were performed using ~190,000 single nucleotide polymorphisms (SNPs) across the genome with the Factored Spectrally Transformed Linear Mixed Models (FaST-LMM) algorithm to correct for population structure[36]. The threshold for genome-wide significance of 4.1×10^{-6} was determined by simulation and permutation of strain genotypes, as previously described[21]. Because we examined multiple traits and time points, we increased our significance threshold to 4×10^{-7} and considered loci exceeding 4×10^{-6} suggestive. Since many of the traits we examined are related and correlated, we consider this a conservative adjustment. The boundaries of the associated loci were calculated based on the presence of nearby supportive level SNPs with p-values less than 1×10^{-5}. A total of 3 genome-wide significant loci and 13 genome-wide suggestive loci were identified, of which all but one were less than 2 Mb in size (Table 4-4). One of these loci was previously associated with sudden cardiac death in humans by GWAS. Nine of these loci were replicated either at a different time point or by another independently measured echocardiographic or weight trait in this study. We further performed eQTL analysis using FaST-LMM to identify loci that control gene expression levels. We used the results of eQTL analysis and the Wellcome Trust Mouse Genomes Project (MGP) sequencing database to prioritize our candidate gene list. A few of the associated loci are highlighted below.

The most statistically significant SNP was rs40560913 with a p-value of 3.17 x 10^{-9} on chromosome 7 for the change in LVM by week 3 (Figure 4-3A). The AA genotype at rs40560913 conferred a LVM increase of 43 mg compared to the GG genotype with a LVM increase of 28 mg. The peak SNP is located in the intron of *Izumo2* and is in linkage disequilibrium ($r^2 > 0.8$) with SNPs spanning 2 other genes, *Myh14* and *2310016G11Rik* (Figure 4-3B). *2310016G11Rik* has no known function and does not exhibit structural variations among 10 strains available through the MGP sequencing database. *Izumo2* has 3 splice region SNPs rs45784669 (A -> G), rs50422221 (C -> T), and rs48802363 (T -> C) and encodes a sperm-specific protein[43]. *Myh14* has 2 splice region SNPs rs47454934 (A -> G) and rs255254297 (T -> C). *Myh14*, myosin heavy polypeptide 14, encodes the nonmuscle myosin (NM) heavy chain II-C protein. While NM II-C ablated mice survive to adulthood and showed no obvious defects compared with wild-type littermates, NM II-C ablated mice expressed only 12% of wild-type amounts of NM II-B and developed marked increase in cardiac myocyte hypertrophy compared with NM II-B hypomorphic mice. In addition, the NM II-C ablated hearts developed interstitial fibrosis associated with diffuse N-cadherin and β-catenin localization at the intercalated discs, where both NM II-B and II-C are normally concentrated[44]. *Myh14* is a likely causal gene at this locus responsible for the variations in the change of LVM by week 3.

The second most significant association SNPs were rs13480288 and rs29940243 with a p-value of 1.15 x 10^{-8} on chromosome 9 for the change of IVSd by week 1 (Figure 4-4A). The CC and AA genotypes at rs13480288 and rs29940243

conferred an IVSd increase of 0.18 mm compared to 0.09 mm associated with the alternate genotypes. The peak SNPs were located in the intergenic region between *5730403I07Rik* and *Lrrc1* and in an intron of *Lrrc1*, where nearby SNPs in LD of $r^2 > 0.8$ spanned a 4 Mb region. The SNPs in LD of $r^2 > 0.9$ spanned *Lrrc1* (Figure 4-4B). The MGP suggests that *Lrrc1* contains a splice region variant. Expression QTL analysis of *Lrrc1* showed that baseline *Lrrc1* expression was controlled in cis by the same locus controlling week 1 change of IVSd. The most significantly associated cis-eQTL SNP rs49772635 resided in an intron of *Lrrc1* and had an association p-value of 1.96×10^{-10} (Figure 4-4C). The AA and GG genotypes at rs49772635 conferred a baseline log2 *Lrrc1* expression of 6.63 and 6.15 respectively. Baseline *Lrrc1* expression level was negatively correlated with the week 1 change in IVSd (bicor = -0.29, p-value = 0.01). In addition, *Lrrc1* expression in the heart decreased by 15% after 3 weeks of isoproterenol (p-value = 6.1×10^{-6}). *Lrrc1*, leucine rich repeat containing 1, is a gene of unknown function conserved in human, chimpanzee, Rhesus monkey, dog, cow, rat, chicken and zebrafish. Our findings suggest that *Lrrc1* may be a repressor of early compensatory hypertrophy and itself down-regulated after chronic β-adrenergic stimulation. *Lrrc1* is a likely causal gene responsible for the variations in the early response of IVSd to isoproterenol.

The third most significant association SNP was rs48791248 with a p-value of 2.18×10^{-7} on chromosome 15 for FS at week 1 (Figure 4-5A). The AA genotype conferred a week 1 FS of 40% compared to 48% by the GG genotype. The region in LD with the peak SNP spanned an intergenic region between *Lrp12* and *Zfpm2* (Figure

4-5B). Both *Lrp12* and *Zfpm2* were expressed in the mouse heart. *Zfpm2*, zinc finger protein multitype 2, also known as the transcriptional regulator friend of Gata2 (FOG2), is known to regulate adult heart function and coronary angiogenesis[45]. *Zfpm2* expression level did not have a significant cis-eQTL, was not correlated with week 1 FS and did not change with isoproterenol treatment. *Lrp12*, low density lipoprotein receptor-related protein (LRP) 12, is a member of the low-density lipoprotein receptor (LDLR) superfamily, whose expression is particularly abundant in heart and skeletal muscle[46]. Its cytoplasmic domain contains several motifs implicated in endocytosis and signal transduction and interacts with three proteins involved in signal transduction and endocytosis, i.e. activated protein C kinase (RACK1), muscle integrin binding protein (MIBP) and SMAD anchor for receptor activation (SARA)[47]. Baseline *Lrp12* expression was negatively correlated with week 1 FS (bicor = -0.36, p-value = 0.0014). *Lrp12* expression level was increased by 14% after 3 weeks of isoproterenol (p-value = 1.37×10^{-6}). Interestingly, the syntenic intergenic region between *LRP12* and *ZFPM2* has previously been implicated in vascular endothelial growth factor (VEGF) levels (p-value = 5×10^{-23}) and sudden cardiac arrest in patients with coronary artery disease (p-value = 1×10^{-6}) based on human GWAS[48, 49]. Further investigations are needed to define the regulatory elements and mechanisms responsible for the variations in week 1 FS due to this locus.

A suggestive association locus was located around SNPs rs27816235 and rs27816170 in the intergenic region between *Xpa* and *Foxe1* on chromosome 4 with a p-value of 2.12×10^{-6} for week 2 LVM (Figure 4-6A). The GG and GG genotypes at

SNPs rs27816235 and rs27816170 conferred a LVM of 115 mg compared to 129 mg with the alternate TT and AA genotypes. The region in LD ($r^2 > 0.9$) with the peak SNPs spanned a region containing 10 genes, but the SNPs with the highest p-values spanned a smaller subset of 5 genes (Figure 4-6B). Among these genes, *Xpa* expression level was the highest, was correlated with and mapped to the same locus as week 2 LVM, and was increased with isoproterenol treatment by 12% ($p=1.86 \times 10^{-8}$; Figure 4-6C). The GG genotype at SNPs rs27816235 and rs27816170 conferred a high level of *Xpa* expression and a low level of week 2 LVM. *Xpa*, xeroderma pigmentosum, complementation group A, is a zinc finger protein involved in DNA excision repair. The encoded protein is part of the nucleotide excision repair complex, which is responsible for repair of UV radiation-induced photoproducts and DNA adducts induced by chemical carcinogens[50]. Our findings provide evidence that *Xpa* is upregulated in response to isoproterenol and may play a role in the cardiac response to protect against adverse cardiac remodeling.

A suggestive association locus was located around SNP rs27811538 in the intergenic region between *Tmem38b* and *Zfp462* on chromosome 4 with a p-value of 1.48×10^{-6} for week 3 change in LVM (Figure 4-3A). The GG and TT genotypes at SNP rs27811538 conferred a week 3 LVM change of 30 mg and 46 mg respectively. The region in LD ($r^2 > 0.8$) with the peak SNP spanned a region containing 4 genes (Figure 4-7A). Isoproterenol-treated *Klf4* expression levels had a significant cis-eQTL at around SNP rs27794497 (p-value = 5.9×10^{-5}) in the same locus as week 3 change in LVM (Figure 4-7B). In addition, control *Klf4* expression was negatively correlated with week 3

change in LVM (bicor = -0.3, p-value = 0.009). *Klf4*, Kruppel-like factor 4, is a repressor of hypertrophic gene program in cardiomyocytes and is one of the four Yamanaka factors, that are sufficient to reprogram differentiated cells to an embryonic-like state designated induced pluripotent stem cells (iPSCs). Our results suggest that *Klf4* is a likely causal gene at this locus responsible for the variations in the week 3 change in LVM.

A suggestive association locus was located around SNPs rs27851114 in the intergenic region between *Svep1* and *Musk* on chromosome 4 with a p-value of 1.75 x 10^{-6} for week 1 change in LVIDd. The region in LD ($r^2 > 0.8$) with the peak SNP spanned a region containing *Svep1* and *Musk*. The AA and GG genotypes at SNP rs27851114 conferred a week 1 change in LVIDd of 0.24 mm and -0.07 mm. *Svep1* harbors a missense variant with a SIFT score of 0.01 that results in the change of amino acid 1149 from T to I. *Svep1*, sushi, von Willebrand factor type A, EGF and pentraxin domain containing 1, is a cell adhesion molecule (CAM) specifically expressed in activated satellite cells prior to their determination to the myogenic lineage, whose expression declines as the myotube matures[51]. Our results suggest that coding sequence variations in *Svep1* may play role in the variations observed in week 1 LVIDd change in the mouse panel.

A suggestive association locus was located around SNP rs33896682 on chromosome 9 with a p-value of 2.28 x 10^{-6} for week 2 change in FS (Figure 4-8A). The CC genotype conferred a week 2 FS increase of 6% compared to 0% by the TT

genotype. The peak SNP rs33896682 resided in an intron of *Arpp19* and the region in LD with the peak SNP spanned an approximately 1.5 Mb region containing many genes (Figure 4-8B). *Arpp19* expression was not significantly altered by isoproterenol treatment. Both baseline and isoproterenol-treated *Arpp19* expression levels were strongly controlled by the same locus as the week 2 change in FS (Figure 4-8C). The CC and TT genotypes conferred a log2 intensity of about 8.5 and 9.2 respectively. In addition, *Arpp19* expression was negatively correlated with the change in week 2 FS (bicor = -0.32, p-value = 0.0041). *Arpp19*, cAMP-regulated phosphoprotein 19, is a protein kinase A (PKA) substrate that is ubiquitously expressed[52]. When phosphorylated by the kinase Greatwall/*Mastl*, *Arpp19* becomes a potent PP2A inhibitor to allow mitosis entry[53]. The role of *Arpp19* in the heart has not been previously described. Our findings suggest that the CC genotype at SNP rs33896682 conferred a lower level of *Arpp19* expression, which appeared to protect the heart against isoproterenol-induced FS deterioration by week 2.

Discussion

We have studied how common genetic variation among inbred strains of mice can influence cardiac remodeling in response to isoproterenol challenge. Precise echocardiographic measurements were taken over the three-week study time course and these measurements were complemented with global LV gene expression profiling in both treated and control mice. A number of conclusions have emerged from the study. First, we established the relationships of the echocardiographic measurements

across time points and to each other and that cardiac remodeling was highly heritable. Second, we examined isoproterenol-induced changes in gene expression and functional enrichment in genes correlated with cardiac remodeling. Third, we mapped genetic loci contributing to the variations using high-resolution association analysis and prioritized our candidate genes using cis-acting expression variations, correlation to phenotype and coding sequence variations. We identified 3 significant and 13 suggestive loci and provided evidence in support of causal candidate genes. Our results provide a powerful resource for the identification of novel pathways and gene-by-environment interactions contributing to cardiac remodeling.

Our study has several advantages over human GWAS. First, we were able to fully control the age, environment, severity and timing of cardiac injury, to accurately assess cardiac structure and functional changes with minimal confounding factors. We were able to establish that cardiac structure and function were highly heritable both at baseline and after cardiac injury. Our heritability estimates were similar to the heritability estimate of 0.69 based on healthy adult monozygotic twins and were significantly higher than the estimates of 0.24-0.32 based on the Framingham Heart Study population[54, 55]. The relatively high heritability estimate in our panel provided us significant power to discover novel and interesting loci compared to the human studies. We also had the benefit of having access to LV tissue expression profiles to examine the relationships between gene transcripts and phenotypes as well as genetic loci to uncover likely causal genes, which overcomes at least in part the challenges of working with relatively long LD structure in mice compared to humans. Finally, the MGP provided us a map of

coding variations among 10 inbred mouse strains in our panel, which allowed us to identify functionally relevant variants that segregated with the association loci SNPs. This combined approach allowed us to quickly prioritize the lists of candidate genes to manageable sizes for the majority of the loci in an unbiased manner without a prior knowledge about the genes in the loci. Our study is a proof of principle study, whose results have the potential of being extended to humans in future investigations.

As with any study in mice, there are several disadvantages of our approach compared to human studies. We have identified several likely candidates that point to genes and pathways relevant to the isoproterenol-induced heart failure model. While isoproterenoal injury mimics the HF state, artificial off-target effects of isoproterenol may potential affect our study results and make them less relevant to human HF. For example, the variations in phenotype may be due to variations in isoproterenol metabolism; however, there is no indication based on our candidate gene lists to suggest that differences in isoproterenol metabolism played a major role. The variations in phenotype may also be confounded by adrenergic receptor variations. It has been hypothesized that differences in β-adrenergic receptor (β-AR) density and/or coupling between A/J and C57BL6/J mice contribute to the differences in cardiac remodeling upon isoproterenol stimulation[56]. Based on the MGP sequences, there were 2 missense variants in the coding sequence of *Adrb3* and no sequence variations in *Adrb1* and *Adrb2*. Based on our expression profiles, only control *Adrb1* was regulated in cis. We did not map any of our phenotypes to the adrenergic receptors to indicate that adrenoceptor status was a major contributor to variations observed in our study. This

finding is not dissimilar to findings in a study of LVM in humans, which demonstrated that LVM variations were not significantly associated with genetic polymorphisms in β1-adrenoceptor[54]. The final disadvantage of a study in the mouse is that, although the identified variants, genes and pathways may be relevant and important in human disease pathology, the specific genetic variants will most likely not be directly translatable to humans.

Our study also highlights the strength of a systems approach to studying common variations relevant to HF. By overlaying genetic, expression and phenotypic information, we were able to begin to deduce potential mechanisms of action and generate hypothesis driven studies to further understand genetic underpinnings of phenotypic variations. For example, isoproterenol treatment resulted in the overall decrease of *Lrrc1* expression. At the same time, *Lrrc1* expression under control and isoproterenol treated conditions was significantly controlled by the genotype at the same locus that controlled the week 1 change in IVSd, or early septal wall hypertrophy, suggesting the *Lrrc1* played a regulatory role in early hypertrophic response to isoproterenol. Mice with the lower *Lrrc1* expression genotype had a higher degree of IVSd hypertrophy at week 1 and maintained an elevated FS at week 2, which suggests that *Lrrc1* expression may be a negative effector of compensatory hypertrophy and cardiac contractility.

Cardiac remodeling is one of the most important prognostic determinants of clinical HF. The laboratory mouse with its fully sequenced and annotated genome,

targeted germline modification and many inbred strains, is an essential tool in biomedical research that complements the strengths of human studies. Our study is the first to survey isoproterenol-induced cardiac remodeling in a panel of 100+ laboratory inbred mouse strains. Our study provided strong evidence for the contribution of common genetic variations to cardiac structure and function in normal and diseased heart and that HF traits are complex, influenced by many genes in addition to various environmental factors. Future advances in understanding how common genetic variations in a population modify HF progression are needed to provide insights in genetic risk profiling, gene-gene and gene-environment interactions as well as the design of personalized therapies for all HF patients.

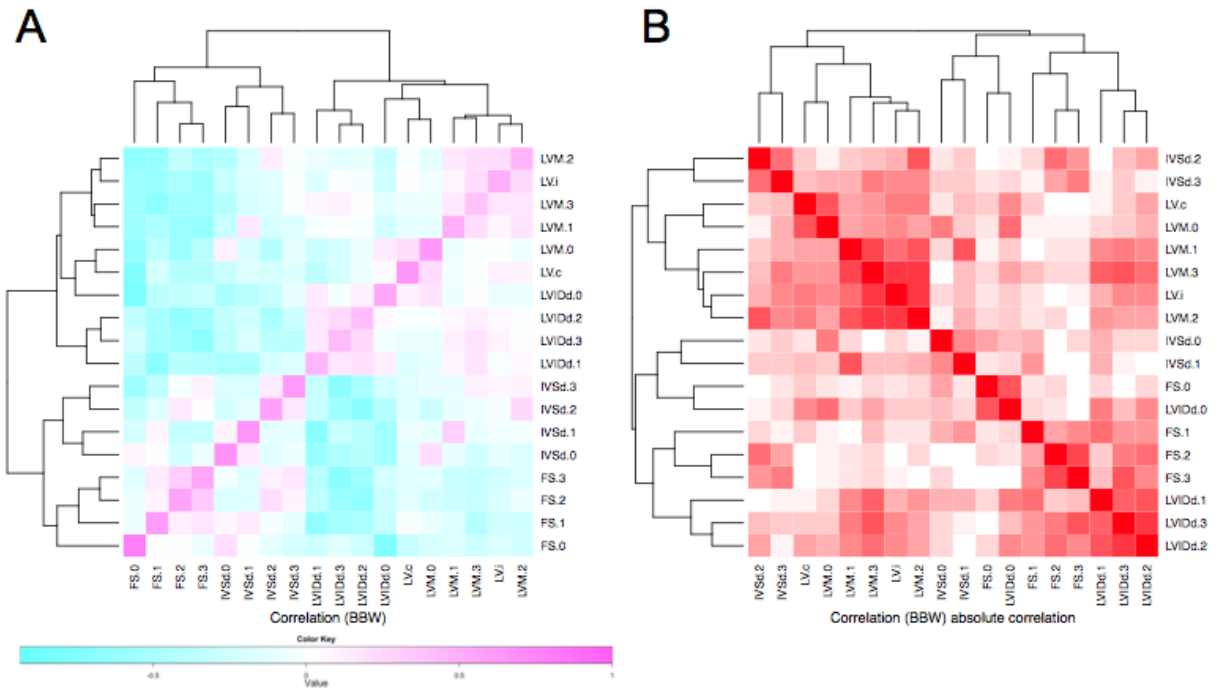


Figure 4-1. Correlation among echocardiographic measures and LV (LV) weights. Signed (A) and absolute (B) correlation and hierarchical clustering of baseline body weight adjusted echocardiographic measures, including interventricular septal wall thickness at end diastole (IVSd), left ventricular internal diameter at end diastole (LVIDd), fractional shortening (FS), and left ventricular mass (LVM), and left ventricular weights (LV). Suffixes denote control (c) and week 3 isoproterenol-treated (i) LV and baseline (0), week 1 (1), 2 (2), and 3 (3) isoproterenol-treated echocardiographic measures.

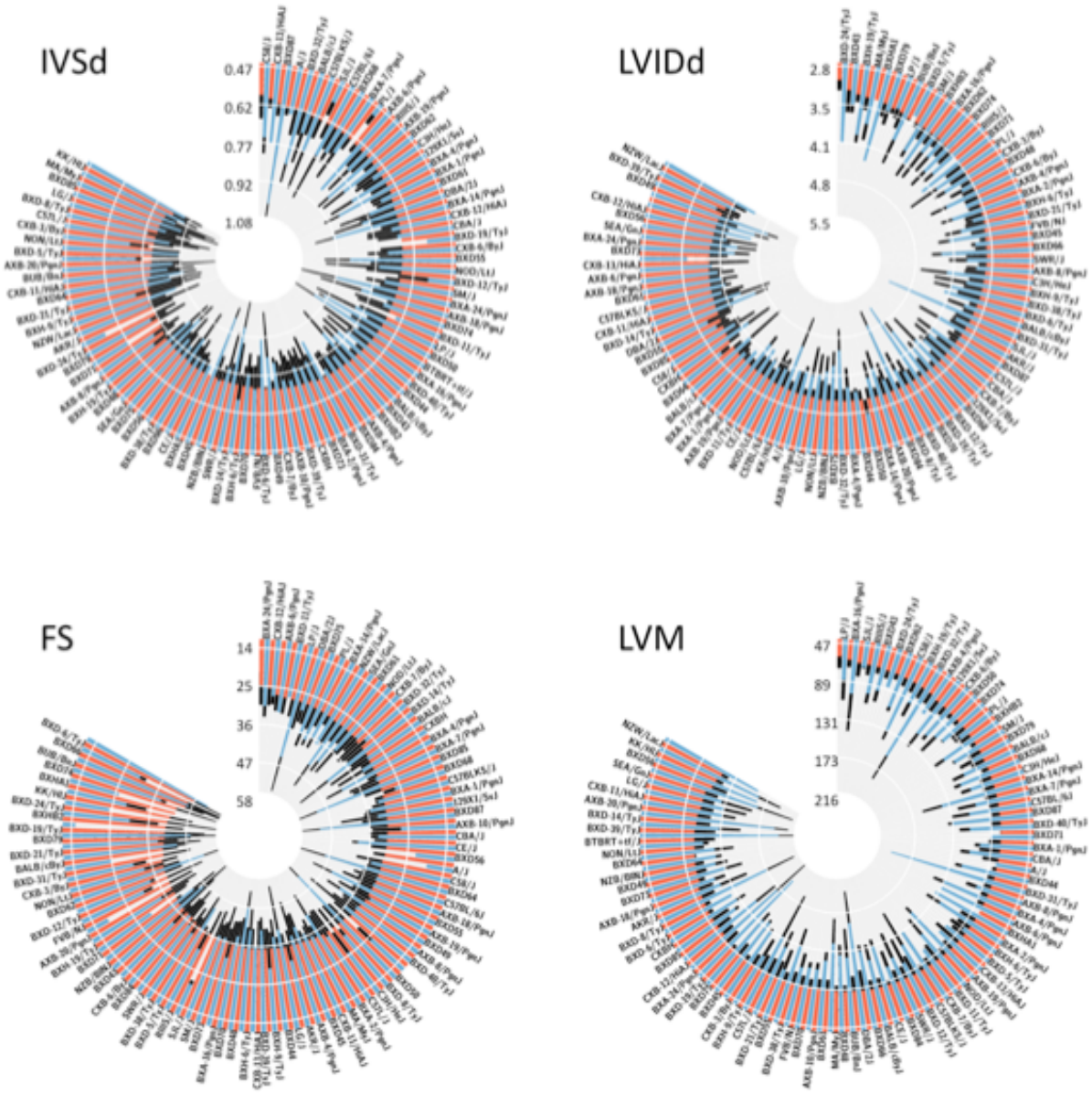


Figure 4-2. Cardiac structural and functional variations in the HMDP. Circos plots represent IVSd (mm), LVIDd (mm), FS (%), and LVM (mg) at baseline (red bars in ranked order) and after 3 weeks of isoproterenol (blue bars). Black bars represent measurements within 1 standard deviation.

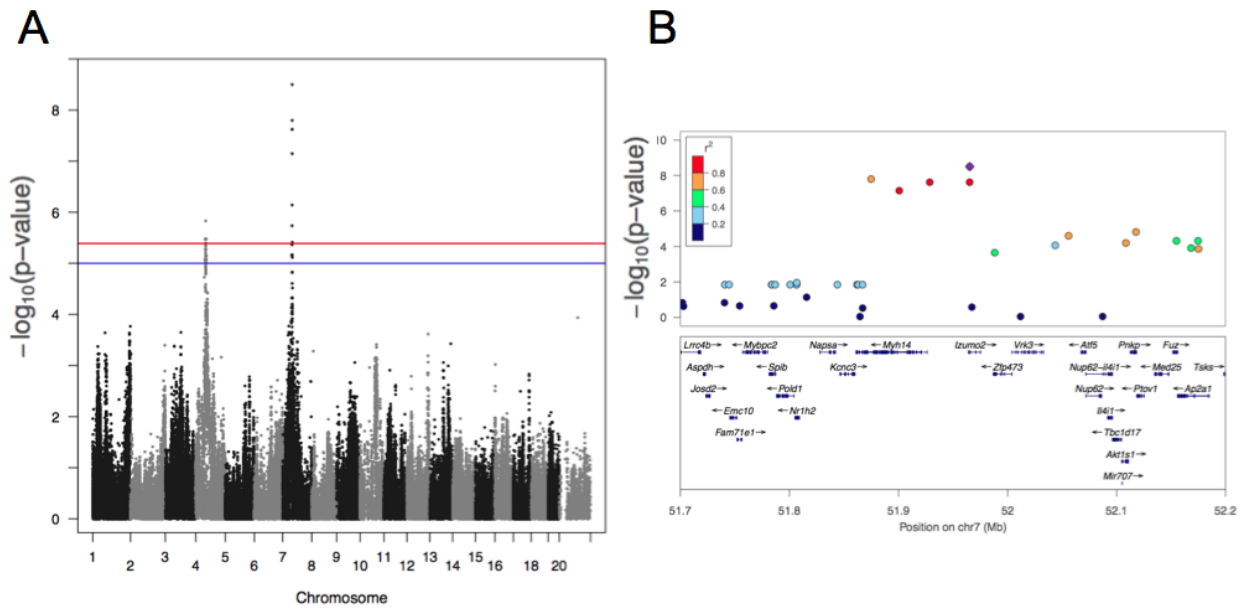


Figure 4-3. Fine mapping of the change in week 3 LVM chromosome 7 locus. A. Manhattan plot for the change in week 3 LVM. Blue line indicates p-value threshold of 1 x 10^{-5}. Red line indicates p-value threshold of 4.1 x 10^{-6}. B. Regional plot for the change in week 3 LVM around peak single nucleotide polymorphism (SNP) rs40560913 (purple). The strengths of pairwise linkage disequilibrium (LD) measured in r^2 between the peak SNP and surrounding SNPs are denoted by corresponding colors.

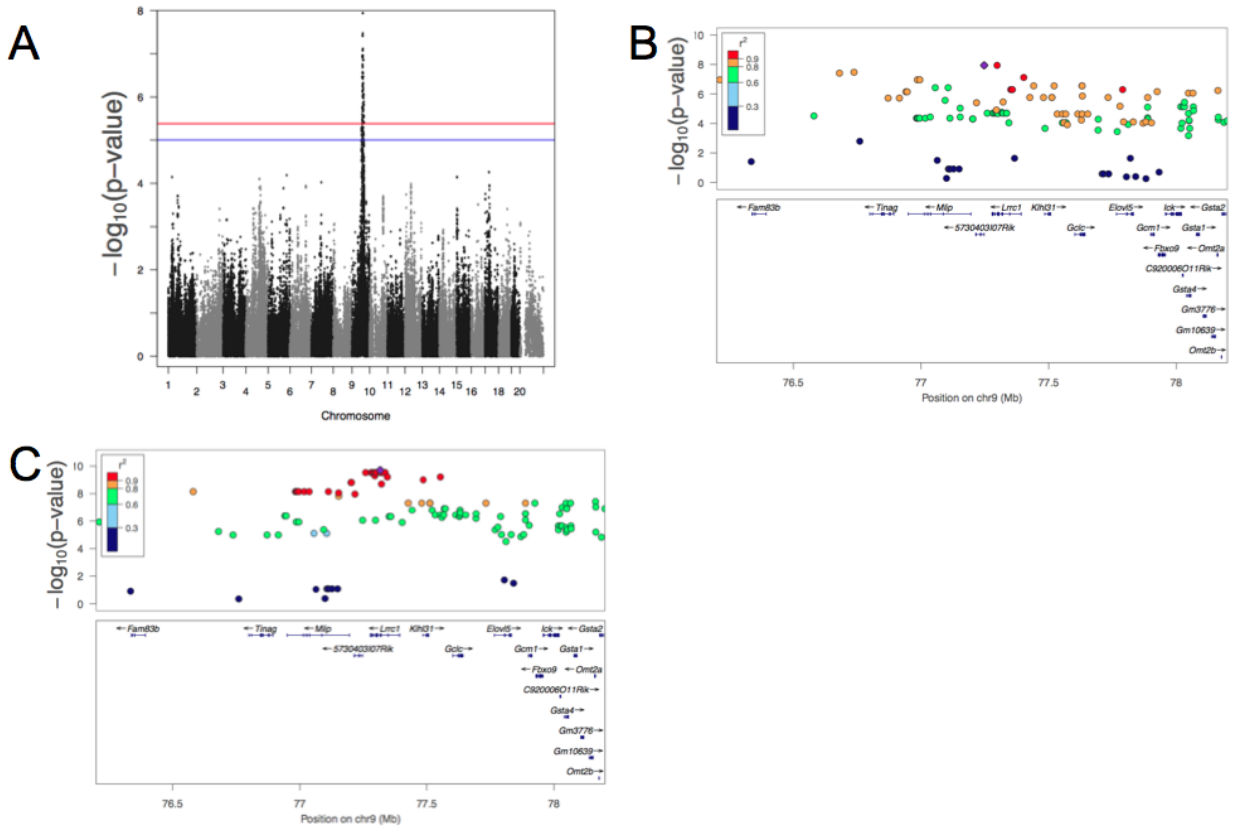


Figure 4-4. Fine mapping of the change in week 1 IVSd chromosome 9 locus. A. Manhattan plot for the change in week 1 IVSd. B. Regional plot for the change in week 1 IVSd around peak SNPs rs13480288 (purple) and rs29940243. C. Regional plot for the control *Lrrc1* expression (ILMN_2868457) around SNP rs49772635 (purple).

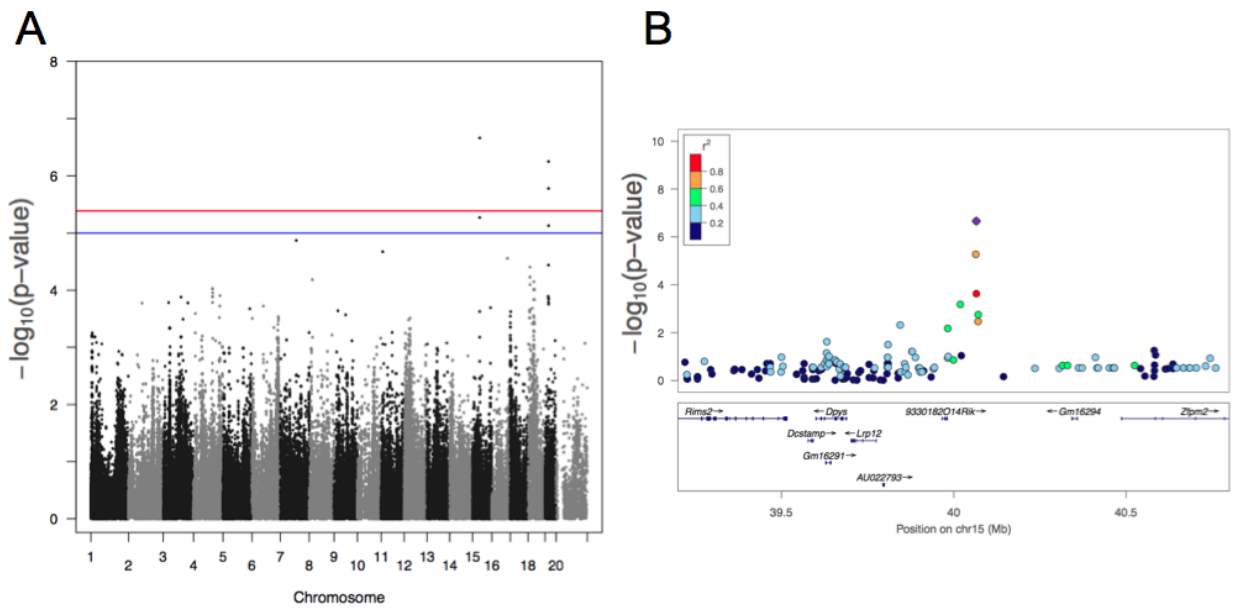


Figure 4-5. Fine mapping of week 1 FS chromosome 15 locus.A. Manhattan plot for week 1 FS. B. Regional plot for week 1 FS around peak SNP rs48791248 (purple).

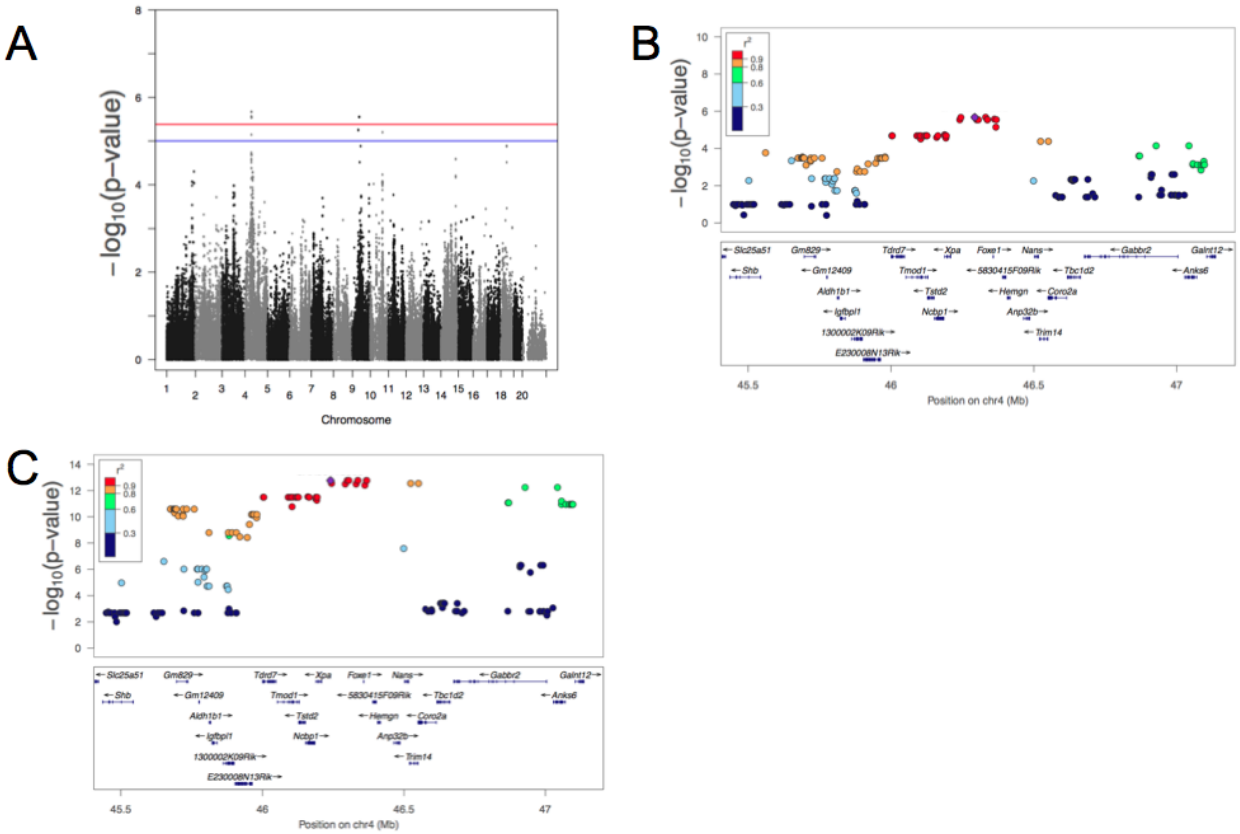


Figure 4-6. Fine mapping of week 2 LVM chromosome 4 locus. A. Manhattan plot for week 2 LVM. B. Regional plot for week 2 LVM around peak SNPs rs27816235 (purple) and rs27816170. C. Regional plot for the control *Xpa* expression (ILMN_2968515) around SNP rs27816412 (purple).

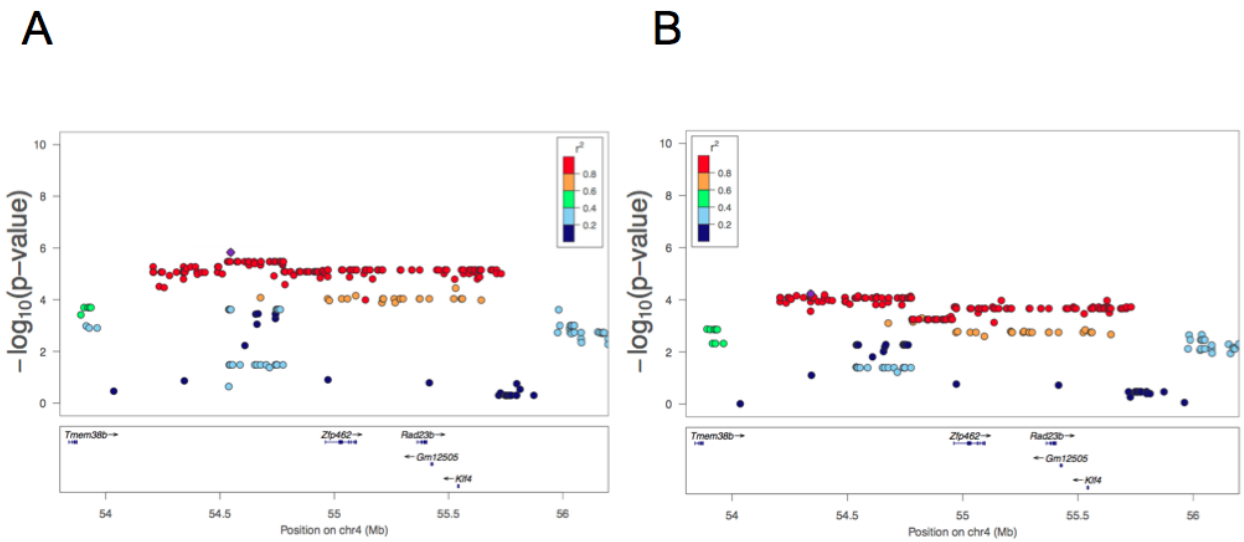


Figure 4-7. Fine mapping of the change in week 3 LVM chromosome 4 locus. A. Regional plot for the change in week 3 LVM around peak SNP rs27811538 (purple). B. Regional plot for the isoproterenol-treated *Klf4* expression (ILMN_1221264) around SNP rs27794497 (purple).

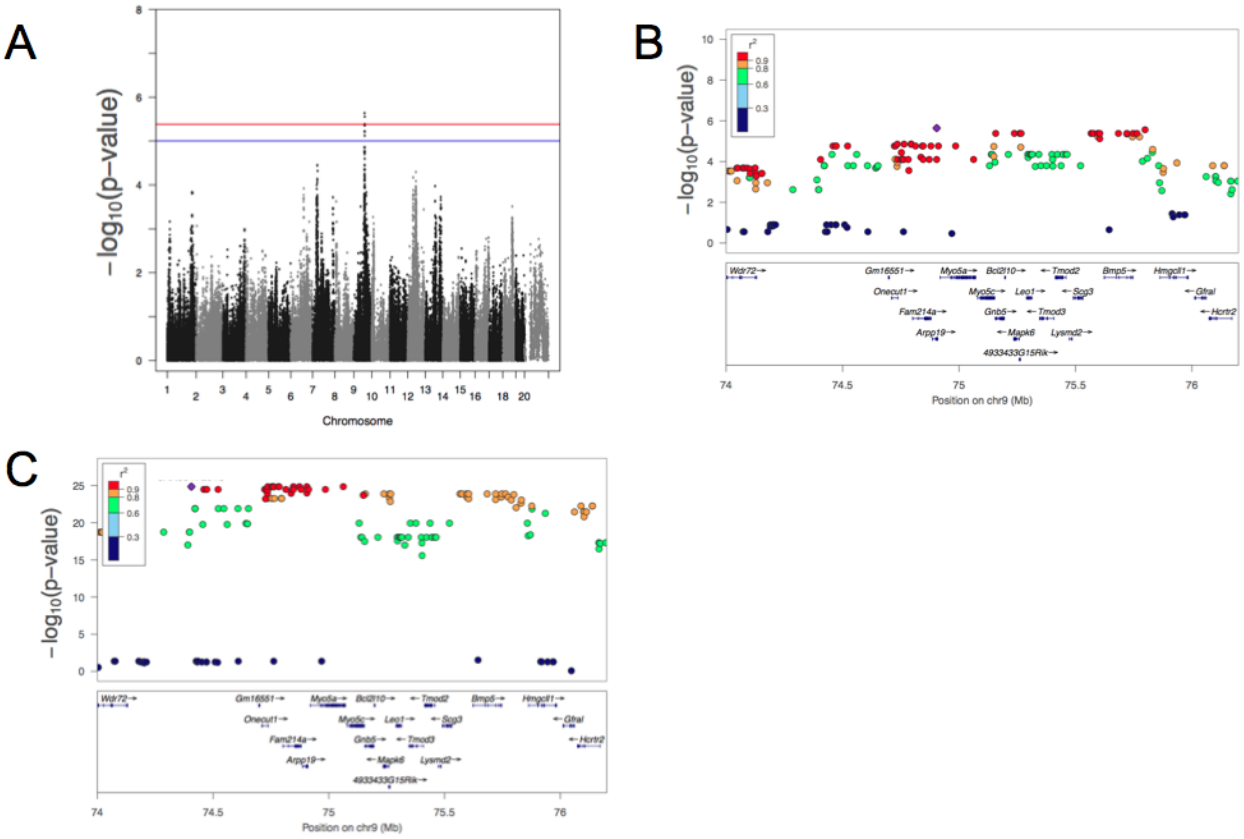


Figure 4-8. Fine mapping of the change in week 2 FS chromosome 9 locus. A. Manhattan plot for the change in week 2 FS. B. Regional plot for the change in week 2 FS around peak SNP rs33896682 (purple). C. Regional plot for the control Arpp19 expression (ILMN_2674407) around SNP rs29739331 (purple).

Table 4-1. Echocardiographic traits in control mice measured at 2 time points.

	Baseline	Week 3	p-value
IVSd (mm)	0.73	0.75	0.725
LVIDd (mm)	3.75	3.68	0.0192
FS (%)	36.5	38.8	0.00176
LVM (mg)	92.3	93.5	0.345

Table 4-2. Study sample characteristics.

Dataare represented as population mean ± standard deviation.

	Baseline	Week 1	Week 2	Week 3
IVSd (mm)	0.73 ± 0.08	0.85 ± 0.12	0.80 ± 0.11	0.80 ± 0.10
LVIDd (mm)	3.7 ± 0.2	3.9 ± 0.4	4.1 ± 0.3	4.2 ± 0.4
FS (%)	37 ± 5	41 ± 7	37 ± 7	38 ± 8
LVM (mg)	90 ± 14	121 ± 23	124 ± 24	127 ± 26

Table 4-3. Heritability estimates of cardiac structure and function in the HMDP. All heritability values exceed ANOVA p-value of <0.001.

	Control	Week 1	Week 2	Week 3
IVSd	0.72	0.71	0.66	0.64
LVIDd	0.81	0.76	0.75	0.76
FS	0.79	0.74	0.71	0.76
LVM	0.84	0.77	0.72	0.79

Table 4-4. Genome-wide significant and suggestive cardiac remodeling loci.

A. Significant Loci

Analysis	Trait	SNP	P-value	MAF	Location (mm9)	SNP Type	LD block (Mb)	n.	Candidate Gene(s)	Exp	eQTL	Cor	SV
Week 3 Delta	LVM*[1]	rs40560913	3.17E-09	0.40	7:51965267	I	51.9-52.0	3	Myh14	-	-	-	S
									2310016G11Rik	-	-	-	-
									Izumo2[†]	-	-	-	S
Week 1 Delta	IVSd[2]	rs13480288	1.15E-08	0.38	9:77247379	IG	75.8-80.2	M	Mlip	-	-	-	MS
		rs29940243			9:77297835	I			5730403I07Rik	-	-	-	-
									Lrrc1[†]	6[d]	Yes	Yes	S
Week 1	FS	rs48791248	2.18E-07	0.10	15:40065358	IG	40.0-40.2[‡]	0	Lrp12	7[d]	No	Yes	-
									Zfpm2	8	No	No	-

B. Suggestive Loci

Analysis	Trait	SNP	P-value	MAF	Location (mm9)	SNP Type	LD block (Mb)	n.	Candidate Gene(s)	Exp	eQTL	Cor	SV
Week 2	LVM[3]	rs27816235	2.12E-06	0.49	4:46292922	IG	45.5-46.6	M	Xpa	9[d]	Yes	Yes	M
		rs27816170			4:46331009				Foxe1	4	No	No	-
									5830415F09Rik	5	No	No	-
									Hemgn	-	-	-	-
									Anp32b	4	No	No	-
Week 3 Delta	LVM*	rs27811538	1.48E-06	0.36	4:54546700	IG	54.2-55.8	4	Tmem38b	10[d]	No	No	-
									Zfp462	4	No	No	MS
									Rad23b	-	-	-	M
									Klf4	7	Yes	Yes	-
Week 1 Delta	LVIDd	rs27851114	1.75E-06	0.06	4:58277624	IG	58.1-58.5	2	Svep1	-	-	-	MS[§]
									Musk	4	No	No	-
Week 3	LVM[‖]	rs3656076	6.10E-07	0.44	4:63785765	I	63.7-64.0	2	Tnc	4	No	No	-
		rs27880677			4:63812861	IG			AK033210[†]	-	-	-	-
Week 3 Delta	IVSd	rs49724804	2.03E-06	0.1	5:108887309	IG	107.7-109.2	M	Mfsd7a	-	-	-	M
									Pcgf3	-	-	-	S
Week 3	IVSd	rs37578592	1.13E-06	0.1	6:113357105	I	112.7-113.5	M	Ttll3[†]	-	-	-	M
Week 2	LVM[4]	rs37851745	2.77E-06	0.06	9:49025659	IG	48.9-49.3	0	Zw10	7	Yes	Yes	M
		rs37535934			9:49025682								
		rs37836074			9:49025884								
Week 2 Delta	FS*[5]	rs33896682	2.28E-06	0.40	9:74904030	I	74.5-76.0	M	Arpp19[†]	8	Yes	Yes	-
Week 1 Delta	IVSd*	rs36266287	8.62E-07	0.38	9:84204400	IG	84.0-84.6	4	Bckdhb	-	-	-	-
Week 3	LVIDd[6]	rs29350578	1.76E-06	0.49	10:86893551	IG	86.0-87.8	M	1700113H08Rik	-	-	-	-

Analysis	Trait	SNP	P-value	MAF	Location (mm9)	SNP Type	LD block (Mb)	n.	Candidate Gene(s)	Exp	eQTL	Cor	SV
Delta		rs29357973			10:86906901				Ascl1	4	No	No	-
		rs6258490			10:86918146				Pah	4	Yes	No	-
		rs29328181			10:86919534								
		rs49299504			10:86938274								
Week 3	LVIDd[7]	rs30582427	3.75E-06	0.42	10:106475587	I	106.0-107.0	4	Acss3†	-	-	-	S
Delta		rs13480753			10:106486522	E			Lin7a	-	-	-	-
									Myf5	4	No	No	-
									Myf6	-	-	-	-
Week 3	LVIDd*[8]	rs4181997	4.90E-07	0.43	16:45814479	I	45.0-46.0	1	Slc35a5	5	Yes	Yes	-
Delta									Cd200	5	Yes	Yes	-
									Phldb2†	4	No	No	-
Week 1	FS*[9]	rs50977365	5.61E-07	0.06	19:21000692	I	20.9-21.5	2	Tmc1†	-	-	-	-
									Zfand5	-	-	-	-

Delta denotes change from baseline. Numerical superscript indicates replication at a different time point or by an independently measured echocardiographic or weight trait in the same population. MAF denotes minor allele frequency. SNP types are coded I (intronic), IG (intergenic), and E (exonic). Column n. denotes the number of genes in LD (r^2>0.8) with peak SNP, where M denotes many genes. Candidate genes are short listed by LD cutoffs of r^2>0.9 or 0.95, if applicable. Exp denotes approximate average log2 intensity of expression, annotated by ^{d} for up- and $_{d}$ for down-regulation by isoproterenol. eQTL denotes the presence of cis-eQTL (p-value < 3.63 x 10^{-3}). Cor denotes the presence of correlation between gene expression and clinical trait (p-value < 0.05). SV denotes structural variations, including M for missense (SIFT score if predicted deleterious) and S for splice region variants.

[1]Week 3 Vold; [2]Week 1 Delta IVSs, Week 2 Delta FS/EF, Ctrl BW/TH/LV/RV/Liver, Iso RV; [3]Week 1 A/E, Week 1 MNSER, Week 1 Vcf; [4]Ctrl TH/LV/RV; [5]Week 1 Delta IVSd/IVSs, Week 2 Delta EF, Ctrl BW/TH/LV/RV/Liver, Iso RV; [6]Week 3 Delta Vold, Iso LV; [7]Week 3 Delta RWTd; [8]Week 3 Delta Vold, Week 3 Delta LVIDs/Vols, Week 3 FS/EF/Vcf; [9]Week 1 Vcf/MNSER.

*Replicated in baseline body weight (BBW) adjusted residuals; †Gene containing peak SNP; ‡Locus for sudden cardiac arrest in coronary artery disease human GWAS (PMID:21658281); §Deleterious by SIFT (rs27853431, T1149I, SIFT=0.01); ||Unique for BBW adjusted residuals

Abbreviations: A/E (A vs. E velocity ratio), BW (body weight), Ctrl (control), EF (ejection fraction), Iso (isoproterenol-treated), IVSs (IVS at end systole), Liver (liver weight), LV (LV weight), LVIDs (LVID at end systole), MNSER (mean normalized systolic ejection rate), RV (right ventricular weight), RWTd (relative wall thickness at end diastole), TH (total heart weight), Vcf (velocity of circumferential shortening), Vold (LV volume at end diastole) and Vols (LV volume at end systole).

Bibliography

1. Go AS, Mozaffarian D, Roger VL, Benjamin EJ, Berry JD, Borden WB, Bravata DM, Dai S, Ford ES, Fox CS, Franco S, Fullerton HJ, Gillespie C, Hailpern SM, Heit JA, Howard VJ, Huffman MD, Kissela BM, Kittner SJ, Lackland DT, Lichtman JH, Lisabeth LD, Magid D, Marcus GM, Marelli A, Matchar DB, McGuire DK, Mohler ER, Moy CS, Mussolino ME, Nichol G, Paynter NP, Schreiner PJ, Sorlie PD, Stein J, Turan TN, Virani SS, Wong ND, Woo D, Turner MB. Heart disease and stroke statistics--2013 update: A report from the american heart association. *Circulation*. 2013;127:e6-e245

2. McMurray JJ, Petrie MC, Murdoch DR, Davie AP. Clinical epidemiology of heart failure: Public and private health burden. *Eur Heart J*. 1998;19 Suppl P:P9-16

3. Kannel W, Ho K, Thom T. Changing epidemiological features of cardiac failure. *British heart journal*. 1994;72:9

4. Levy D, Kenchaiah S, Larson MG, Benjamin EJ, Kupka MJ, Ho KK, Murabito JM, Vasan RS. Long-term trends in the incidence of and survival with heart failure. *N Engl J Med*. 2002;347:1397-1402

5. Taylor CJ, Roalfe AK, Iles R, Hobbs FD. Ten-year prognosis of heart failure in the community: Follow-up data from the echocardiographic heart of england screening (echoes) study. *Eur J Heart Fail*. 2012;14:176-184

6. Go AS, Mozaffarian D, Roger VL, Benjamin EJ, Berry JD, Blaha MJ, Dai S, Ford ES, Fox CS, Franco S, Fullerton HJ, Gillespie C, Hailpern SM, Heit JA, Howard VJ, Huffman MD, Judd SE, Kissela BM, Kittner SJ, Lackland DT, Lichtman JH, Lisabeth LD, Mackey RH, Magid DJ, Marcus GM, Marelli A, Matchar DB, McGuire DK, Mohler ER, 3rd, Moy CS, Mussolino ME, Neumar RW, Nichol G, Pandey DK, Paynter NP, Reeves MJ, Sorlie PD, Stein J, Towfighi A, Turan TN,

Virani SS, Wong ND, Woo D, Turner MB. Heart disease and stroke statistics--2014 update: A report from the american heart association. *Circulation*. 2014;129:e28-e292

7. Lymperopoulos A, Rengo G, Koch WJ. Adrenergic nervous system in heart failure: Pathophysiology and therapy. *Circ Res*. 2013;113:739-753

8. Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. *Science*. 2008;322:881-888

9. Smith NL, Felix JF, Morrison AC, Demissie S, Glazer NL, Loehr LR, Cupples LA, Dehghan A, Lumley T, Rosamond WD, Lieb W, Rivadeneira F, Bis JC, Folsom AR, Benjamin E, Aulchenko YS, Haritunians T, Couper D, Murabito J, Wang YA, Stricker BH, Gottdiener JS, Chang PP, Wang TJ, Rice KM, Hofman A, Heckbert SR, Fox ER, O'Donnell CJ, Uitterlinden AG, Rotter JI, Willerson JT, Levy D, van Duijn CM, Psaty BM, Witteman JCM, Boerwinkle E, Vasan RS. 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*. 2010;3:256-266

10. Villard E, Perret C, Gary F, Proust C, Dilanian G, Hengstenberg C, Ruppert V, Arbustini E, Wichter T, Germain M, Dubourg O, Tavazzi L, Aumont MC, DeGroote P, Fauchier L, Trochu JN, Gibelin P, Aupetit JF, Stark K, Erdmann J, Hetzer R, Roberts AM, Barton PJ, Regitz-Zagrosek V, Aslam U, Duboscq-Bidot L, Meyborg M, Maisch B, Madeira H, Waldenstrom A, Galve E, Cleland JG, Dorent R, Roizes G, Zeller T, Blankenberg S, Goodall AH, Cook S, Tregouet DA, Tiret L, Isnard R, Komajda M, Charron P, Cambien F. A genome-wide association study identifies two loci associated with heart failure due to dilated cardiomyopathy. *Eur Heart J*. 2011;32:1065-1076

11. Fox ER, Musani SK, Barbalic M, Lin H, Yu B, Ogunyankin KO, Smith NL, Kutlar A, Glazer NL, Post WS, Paltoo DN, Dries DL, Farlow DN, Duarte CW, Kardia SL,

Meyers KJ, Sun YV, Arnett DK, Patki AA, Sha J, Cui X, Samdarshi TE, Penman AD, Bibbins-Domingo K, Buzkova P, Benjamin EJ, Bluemke DA, Morrison AC, Heiss G, Carr JJ, Tracy RP, Mosley TH, Taylor HA, Psaty BM, Heckbert SR, Cappola TP, Vasan RS. Genome-wide association study of cardiac structure and systolic function in african americans: The candidate gene association resource (care) study. *Circ Cardiovasc Genet*. 2013;6:37-46

12. Nakada K, Sato A, Hayashi J. Reverse genetic studies of mitochondrial DNA-based diseases using a mouse model. *Proc Jpn Acad Ser B Phys Biol Sci*. 2008;84:155-165

13. Le Corvoisier P, Park HY, Carlson KM, Marchuk DA, Rockman HA. Multiple quantitative trait loci modify the heart failure phenotype in murine cardiomyopathy. *Hum Mol Genet*. 2003;12:3097-3107

14. Wheeler FC, Fernandez L, Carlson KM, Wolf MJ, Rockman HA, Marchuk DA. Qtl mapping in a mouse model of cardiomyopathy reveals an ancestral modifier allele affecting heart function and survival. *Mamm Genome*. 2005;16:414-423

15. Suzuki M, Carlson KM, Marchuk DA, Rockman HA. Genetic modifier loci affecting survival and cardiac function in murine dilated cardiomyopathy. *Circulation*. 2002;105:1824-1829

16. Tsujita Y, Iwai N, Tamaki S, Nakamura Y, Nishimura M, Kinoshita M. Genetic mapping of quantitative trait loci influencing left ventricular mass in rats. *Am J Physiol Heart Circ Physiol*. 2000;279:H2062-2067

17. Flint J, Valdar W, Shifman S, Mott R. Strategies for mapping and cloning quantitative trait genes in rodents. *Nat Rev Genet*. 2005;6:271-286

18. Bennett BJ, Farber CR, Orozco L, Min Kang H, 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. A high-resolution association mapping

panel for the dissection of complex traits in mice. *Genome Research*. 2010;20:281-290

19. Ghazalpour A, Rau CD, Farber CR, Bennett BJ, Orozco LD, van Nas A, Pan C, Allayee H, Beaven SW, Civelek M. Hybrid mouse diversity panel: A panel of inbred mouse strains suitable for analysis of complex genetic traits. *Mammalian Genome*. 2012:1-13

20. Parks BW, Nam E, Org E, Kostem E, Norheim F, Hui ST, Pan C, Civelek M, Rau CD, Bennett BJ, Mehrabian M, Ursell LK, He A, Castellani LW, Zinker B, Kirby M, Drake TA, Drevon CA, Knight R, Gargalovic P, Kirchgessner T, Eskin E, Lusis AJ. Genetic control of obesity and gut microbiota composition in response to high-fat, high-sucrose diet in mice. *Cell Metab*. 2013;17:141-152

21. 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. Mouse genome-wide association and systems genetics identify asxl2 as a regulator of bone mineral density and osteoclastogenesis. *PLoS Genet*. 2011;7:e1002038

22. 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. Gene networks associated with conditional fear in mice identified using a systems genetics approach. *BMC Syst Biol*. 2011;5:43

23. Koitabashi N, Kass D. Reverse remodeling in heart failure--mechanisms and therapeutic opportunities. *Nature reviews. Cardiology*. 2012;9:147-204

24. Liu J, Rigel DF. Echocardiographic examination in rats and mice. *Methods Mol Biol*. 2009;573:139-155

25. Moran CM, Thomson AJ, Rog-Zielinska E, Gray GA. High-resolution echocardiography in the assessment of cardiac physiology and disease in preclinical models. *Exp Physiol*. 2013;98:629-644

26. Galindo CL, Skinner MA, Errami M, Olson LD, Watson DA, Li J, McCormick JF, McIver LJ, Kumar NM, Pham TQ, Garner HR. Transcriptional profile of isoproterenol-induced cardiomyopathy and comparison to exercise-induced cardiac hypertrophy and human cardiac failure. *BMC Physiology*. 2009;9:23

27. Molojavyi A, Lindecke A, Raupach A, Moellendorf S, Kohrer K, Godecke A. Myoglobin-deficient mice activate a distinct cardiac gene expression program in response to isoproterenol-induced hypertrophy. *Physiol Genomics*. 2010;41:137-145

28. Smyth GK. Limma: Linear models for microarray data. In: Gentleman R, ed. *Bioinformatics and computational biology solutions using r and bioconductor*. New York: Springer Science+Business Media; 2005:xix, 473 p.

29. Langfelder P, Horvath S. Wgcna: An r package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559

30. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*. 2004;3:Article3

31. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using david bioinformatics resources. *Nat Protoc*. 2009;4:44-57

32. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*. 2009;37:1-13

33. Yang H, Ding Y, Hutchins LN, Szatkiewicz J, Bell TA, Paigen BJ, Graber JH, de Villena FP, Churchill GA. A customized and versatile high-density genotyping array for the mouse. *Nat Methods*. 2009;6:663-666

34. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. Plink: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559-575

35. 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. Fine mapping in 94 inbred mouse strains using a high-density haplotype resource. *Genetics*. 2010;185:1081-1095

36. Lippert C, Listgarten J, Liu Y, Kadie CM, Davidson RI, Heckerman D. Fast linear mixed models for genome-wide association studies. *Nat Methods*. 2011;8:833-835

37. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ. Locuszoom: Regional visualization of genome-wide association scan results. *Bioinformatics*. 2010;26:2336-2337

38. 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. Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature*. 2011;477:289-294

39. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the sift algorithm. *Nat Protoc*. 2009;4:1073-1081

40. Langfelder P, Horvath S. Fast r functions for robust correlations and hierarchical clustering. *J Stat Softw*. 2012;46

41. Wu J, Bu L, Gong H, Jiang G, Li L, Ma H, Zhou N, Lin L, Chen Z, Ye Y, Niu Y, Sun A, Ge J, Zou Y. Effects of heart rate and anesthetic timing on high-resolution echocardiographic assessment under isoflurane anesthesia in mice. *Journal of ultrasound in medicine : official journal of the American Institute of Ultrasound in Medicine*. 2010;29:1771-1778

42. Jin Y, Kuznetsova T, Bochud M, Richart T, Thijs L, Cusi D, Fagard R, Staessen JA. Heritability of left ventricular structure and function in caucasian families. *Eur J Echocardiogr*. 2011;12:326-332

43. Clark S, Naz RK. Presence and incidence of izumo antibodies in sera of immunoinfertile women and men. *Am J Reprod Immunol*. 2013;69:256-263

44. Ma X, Jana SS, Conti MA, Kawamoto S, Claycomb WC, Adelstein RS. Ablation of nonmuscle myosin ii-b and ii-c reveals a role for nonmuscle myosin ii in cardiac myocyte karyokinesis. *Mol Biol Cell*. 2010;21:3952-3962

45. Zhou B, Ma Q, Kong SW, Hu Y, Campbell PH, McGowan FX, Ackerman KG, Wu B, Tevosian SG, Pu WT. Fog2 is critical for cardiac function and maintenance of coronary vasculature in the adult mouse heart. *J Clin Invest*. 2009;119:1462-1476

46. Qing J, Wei D, Maher VM, McCormick JJ. Cloning and characterization of a novel gene encoding a putative transmembrane protein with altered expression in some human transformed and tumor-derived cell lines. *Oncogene*. 1999;18:335-342

47. Battle MA, Maher VM, McCormick JJ. St7 is a novel low-density lipoprotein receptor-related protein (lrp) with a cytoplasmic tail that interacts with proteins related to signal transduction pathways. *Biochemistry*. 2003;42:7270-7282

48. Debette S, Visvikis-Siest S, Chen MH, Ndiaye NC, Song C, Destefano A, Safa R, Azimi Nezhad M, Sawyer D, Marteau JB, Xanthakis V, Siest G, Sullivan L, Pfister M, Smith H, Choi SH, Lamont J, Lind L, Yang Q, Fitzgerald P, Ingelsson E, Vasan RS, Seshadri S. Identification of cis- and trans-acting genetic variants explaining up to half the variation in circulating vascular endothelial growth factor levels. *Circ Res*. 2011;109:554-563

49. Aouizerat BE, Vittinghoff E, Musone SL, Pawlikowska L, Kwok PY, Olgin JE, Tseng ZH. Gwas for discovery and replication of genetic loci associated with

sudden cardiac arrest in patients with coronary artery disease. *BMC Cardiovasc Disord*. 2011;11:29

50. Nakane H, Takeuchi S, Yuba S, Saijo M, Nakatsu Y, Murai H, Nakatsuru Y, Ishikawa T, Hirota S, Kitamura Y, et al. High incidence of ultraviolet-b-or chemical-carcinogen-induced skin tumours in mice lacking the xeroderma pigmentosum group a gene. *Nature*. 1995;377:165-168

51. Shefer G, Benayahu D. Svep1 is a novel marker of activated pre-determined skeletal muscle satellite cells. *Stem cell reviews*. 2010;6:42-51

52. Dulubova I, Horiuchi A, Snyder GL, Girault JA, Czernik AJ, Shao L, Ramabhadran R, Greengard P, Nairn AC. Arpp-16/arpp-19: A highly conserved family of camp-regulated phosphoproteins. *J Neurochem*. 2001;77:229-238

53. Haccard O, Jessus C. Greatwall kinase, arpp-19 and protein phosphatase 2a: Shifting the mitosis paradigm. *Results Probl Cell Differ*. 2011;53:219-234

54. Swan L, Birnie DH, Padmanabhan S, Inglis G, Connell JM, Hillis WS. The genetic determination of left ventricular mass in healthy adults. *Eur Heart J*. 2003;24:577-582

55. Post WS, Larson MG, Myers RH, Galderisi M, Levy D. Heritability of left ventricular mass: The framingham heart study. *Hypertension*. 1997;30:1025-1028

56. Faulx MD, Ernsberger P, Vatner D, Hoffman RD, Lewis W, Strachan R, Hoit BD. Strain-dependent beta-adrenergic receptor function influences myocardial responses to isoproterenol stimulation in mice. *Am J Physiol Heart Circ Physiol*. 2005;289:H30-36

5 Future Directions

Genetic studies have the potential to reveal insights into the mechanisms of disease causation, progression, prognosis and potential points of therapeutic intervention. From gene panel tests in human patients, a number of genetic variants were implicated in known disease-causing genes with varying levels of evidence. From exome sequencing of familial cardiomyopathy cases, we have identified a novel candidate gene that is highly likely to be causal for disease. Genome-wide association (GWA) in the isoproterenol-induced heart failure (HF) model in mice identified dozens of significant and suggestive loci controlling cardiac structure, function and remodeling. Understanding whether the identified genes and variants are indeed causal for disease will inform the relevance of genes and variants in HF development and any future mechanistic studies. In the case of human patients, knowing definitively that the variant reported to be disease-causing or likely disease-causing not only helps the patient understand his/her disease process but also inform post-genetic counseling with patients and their family members in understanding the risk for disease. Given the number of mouse and human candidate genes and variants to screen, we need a high-throughput method of functional characterization protocol and candidate gene and variant screening. We have conducted preliminary studies in a cell-based model using neonatal rat ventricular cardiomyocytes (NRVMs) as well as in a whole organism model the zebrafish. We will explore additional model systems for effective site mutagenesis to fill the need for gene variant screening.

The choice of isolated NRVMs has several advantages, such as the ability to sample cells from different areas of the heart including the atria, left and right ventricle. Imaging of isolated cells is trivial and is well suited for experiments aimed at visualizing cellular structure and the precise localization of intracellular molecules, which is especially important for our studies aiming to observe cardiomyocyte hypertrophy. Isolated cardiomyocytes are also routinely used for studies examining intracellular Ca2+ homeostasis, cellular mechanics, and protein biochemistry, and can be easily infected or transfected for gene transfer studies[1]. We transfected NRVMs using two siRNA oligos against the *Myh14* gene, one of the top candidates for the change of left ventricular mass, and induced hypertrophy with isoproterenol and phenylephrine for 24 hours. The *Myh14* siRNA treated NRVMs failed to mount a hypertrophic response after isoproterenol and phenylephrine stimulation as in the siRNA scramble control group (Figure 5-1). In future experiments, cardiomyocyte cell size will be measured. The expression levels of cardiac hypertrophy markers, atrial natriuretic peptide (*Nppa*), brain natriuretic peptide (*Nppb*), alpha-myosin heavy chain (*Myh6*) and beta-myosin heavy chain (*Myh7*) will be analyzed using quantitative RT-RCR. Replication-deficient adenoviruses for each siRNA screened positive candidate gene will be constructed using the AdEasy adenoviral system. Cell growth and marker gene induction will be evaluated as above to establish the role of each gene in cardiomyocyte hypertrophy.

In a semi-hierarchical approach, genes screened positive in NRVMs or are expected to have a phenotype on the whole heart level will be evaluated in the zebrafish system to assess for their role in the intact heart. We will inactivate candidate genes by

injecting morpholinos into fertilized eggs, targeted to the translation initiation site or splicing site of the genes, at the 1-2 cell stage and visualize the heart using the cardiac myosin light chain promoter (cmlc)-drive green fluorescent protein (GFP) transgenic zebrafish. Similar to the NRVMs, zebrafish embryo and larvae are well suited for experiments aimed at visualizing cardiac function in the GFP tagged hearts. This is especially important for the functional characterization of cardiac contraction. Zebrafish are also routinely used for studies examining intracellular Ca2+ homeostasis and can be easily injected with capped mRNA for transgene expression and anti-sense morpholino[2]. We injected an anti-sense morpholino against Tnnt2, cardiac troponin T, and demonstrated successful arrest of the heart, pericardial fluid accumulation (congestion), and cardiac compression (Figure 5-2). We do not expect the genes that we screen will have such a dramatic effect but rather have effect on ventricular size and function.

To evaluate individual genetic variants, we will explore the role of clustered regularly interspaced short palindromic repeats (CRISPR) technologies in the zebrafish and mice, to characterize the functional effects of the candidate variant in a high-throughput manner[3, 4]. We will establish cardiac functional characterization protocol and methods to measure electrical conduction in the zebrafish through collaboration. We will also establish detailed cardiac functional characterization protocol, including measures of diastolic function to measure subtle changes from the candidate genetic variants, and methods to measure electrical conduction to fully assess the status of the heart. Finally, the role of using patient-derived iPS cells to validate candidate variant,

personalized therapies based on pharmacogenomic, and personalized allele-specific knock-down technologies may be explored.

The results from genome-wide association analyses suggest that many of the genomic region controlling susceptibility to heart failure are located in gene poor or intergenic regions, which suggests that gene regulation rather than protein coding sequence variations play a major role in mechanisms affecting complex traits. A number of mechanisms have been proposed to effect non-coding sequence variations on gene regulation, including promoter region variations, methylation changes, non-coding RNA, and chromatin states among others. In addition to coding-sequence variations, understanding of gene regulation in genes important in heart failure susceptibility may provide clues to lifestyle factors that may promote heart health and prevent disease, in spite of non-modifiable genetic backgrounds.

The future of genetics research in heart failure holds great promise for heart failure prevention. Improved disease risk prediction informed by genomics data will help guide counseling of patients and installment of programs to prevent disease, such as diastolic dysfunction, for which there are no effective treatments. Improved fine phenotyping of traits, large-scale sequencing for rare variants, and ever more sophisticating mapping strategies will work together to improve the understanding of heart health genetics. Improve translational research to understand how genes predispose to disease will guide drug design to address common pathological

processes shared by heart failure subtypes to improve cardiac health across the general population.

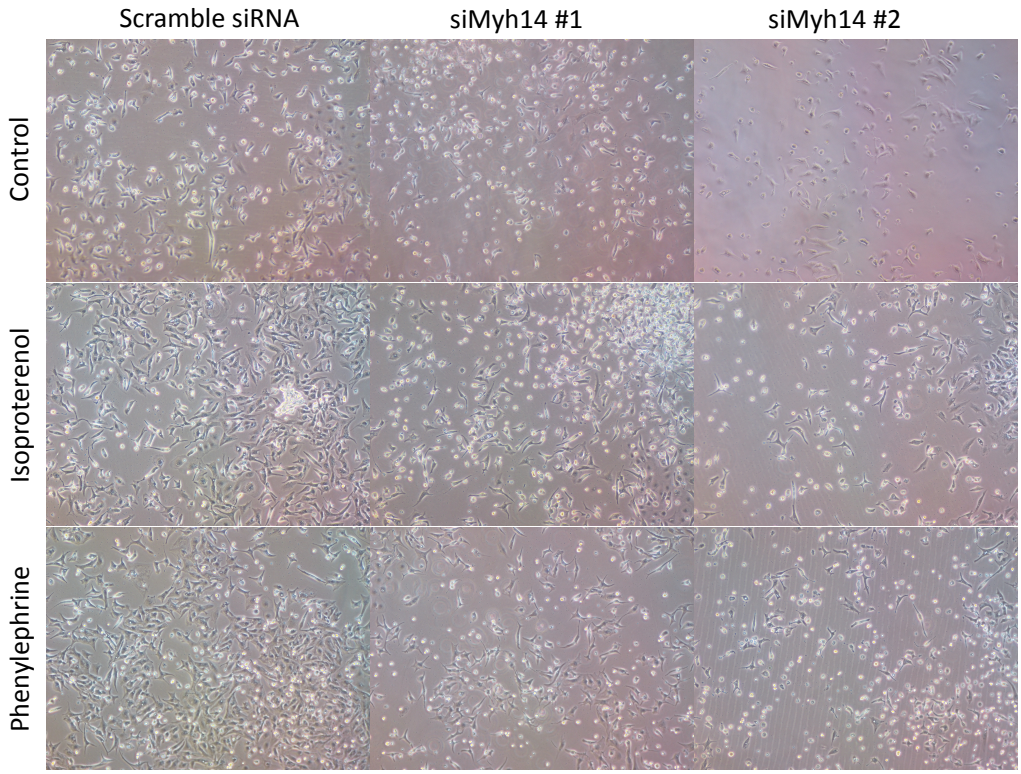


Figure 5-1. siRNA knock-down of Myh14 in neonatal rat ventricular cardiomyocytes abrogates hypertrophic effects of isoproterenol and phenylephrine.
siMyh14#1 represents the first *Myh14* siRNA. siMyh14#2 represents the second *Myh14* siRNA.

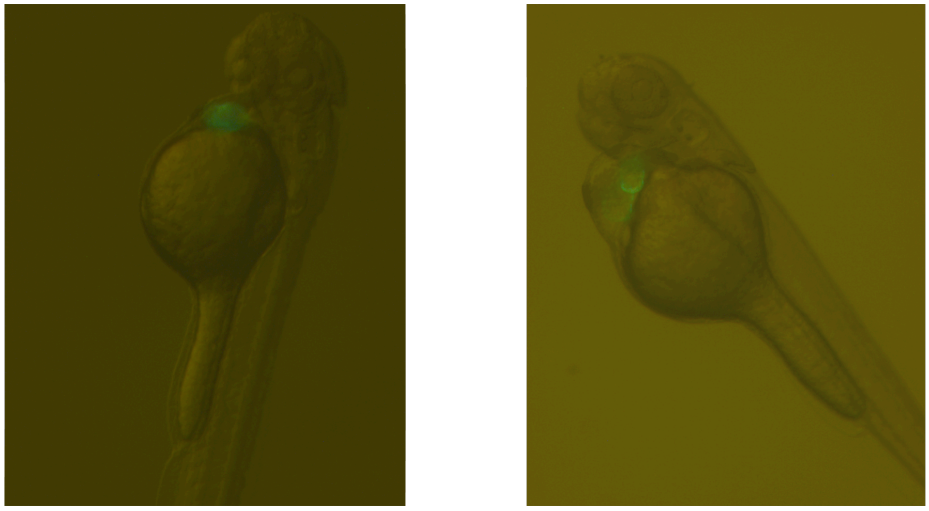


Figure 5-2. Transgenic zebrafish strain with cmlc2::EGFP. Control (left) and Tnnt2 (right) morpholino treated embryos at 54 hours post fertilization (hpf).

Bibliography

1. Louch WE, Sheehan KA, Wolska BM. Methods in cardiomyocyte isolation, culture, and gene transfer. *J Mol Cell Cardiol*. 2011;51:288-298

2. Bakkers J. Zebrafish as a model to study cardiac development and human cardiac disease. *Cardiovascular research*. 2011;91:279-367

3. Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, Jaenisch R. One-step generation of mice carrying mutations in multiple genes by crispr/cas-mediated genome engineering. *Cell*. 2013;153:910-918

4. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. Rna-guided human genome engineering via cas9. *Science*. 2013;339:823-826