

UCLA

UCLA Electronic Theses and Dissertations

Title

Control of Cardiac Gene Expression by Chromatin Architectural Proteins: Mechanisms at the Level of Chromatin Fiber, Transcriptome Remodeling and Cellular Phenotype

Permalink

<https://escholarship.org/uc/item/1qd782bx>

Author

Monte, Emma Marie

Publication Date

2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Control of Cardiac Gene Expression by Chromatin Architectural Proteins: Mechanisms at the
Level of Chromatin Fiber, Transcriptome Remodeling and Cellular Phenotype

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy in
Molecular, Cellular, and Integrative Physiology

by

Emma Marie Monte

2015

ABSTRACT OF THE DISSERTATION

Control of Cardiac Gene Expression by Chromatin Architectural Proteins: Mechanisms at the Level of Chromatin Fiber, Transcriptome Remodeling and Cellular Phenotype

by

Emma Marie Monte

Doctor of Philosophy in Molecular, Cellular, and Integrative Physiology

University of California, Los Angeles, 2015

Professor Thomas M. Vondriska, Chair

When faced with chronic stress, the heart enters a compensatory hypertrophic stage; without intervention it eventually succumbs to decompensation marked by a dilated left ventricular chamber and decreased ejection fraction. While the morphological cardiac remodeling that occurs during the progression of heart failure is well characterized, the exact molecular cause for this gradual switch to failure is not known. In addition to the numerous alterations in signaling pathways, a conserved switch in the transcriptome, known as the fetal gene program, occurs during hypertrophy as a protective effort to sustain contractility by reverting to fetal isoforms of metabolic, contractile and calcium handling genes. We hypothesize that the reproducible, coordinated reprogramming of gene expression is orchestrated by a change in chromatin structure that enables pathologic gene expression.

To determine the proteins involved in repackaging chromatin during cardiac pathology, we performed quantitative proteomic analyses of nuclear proteins in a mouse model of pressure overload hypertrophy and failure. Among the hundreds of proteins we measured on chromatin, my subsequent analyses have focused on two candidates that had the potential to alter gene

expression by directly affecting chromatin packing. The first was Nucleolin, a major component of the nucleolus where it mediates ribosomal biogenesis. Using isolated myocytes and the developing zebrafish embryo, we uncovered a role for Nucleolin to regulate cardiac looping, with its effect on hypertrophy context dependent, such that in isolated myocytes knockdown can promote pathologic gene expression, but loss of Nucleolin during development does not alter myocyte size, instead affecting differentiation along the cardiac lineage.

The second protein I functionally validated was High mobility group protein B2 (HMGB2), a non-histone chromatin structural protein that increases 3-fold in our proteomic analyses. We show that HMGB2 is necessary for ribosomal RNA transcription and is enriched in the nucleolus in hypertrophy; however, overexpression of HMGB2 shuts down transcription globally by compacting DNA. Furthermore, we find HMGB2 knockdown alters the chromatin environment of individual gene promoters in the same manner as hypertrophic agonist signaling in isolated myocytes. Finally, we find that the effect of HMGB2 abundance on the expression of individual genes can be partially explained by the chromatin context, and specifically identify a novel relationship between HMGB2 and CTCF. These studies add to the growing body of work characterizing chromatin remodeling in hypertrophy, and demonstrate that this remodeling extends outside of gene bodies and promoters. Finally, this work begins to uncover what features of chromatin are responsible for tailoring the effects of ubiquitous chromatin proteins toward a cell-type specific outcome.

The dissertation of Emma Marie Monte is approved:

Jeff Abramson

Sarah Franklin

Siavash Kurdistanani

Yibin Wang

Thomas M. Vondriska, Committee Chair

University of California, Los Angeles

2015

TABLE OF CONTENTS

Abstract of the Dissertation	ii
Committee Members	iv
Table of Contents	v
List of Figures	vi
List of Tables	viii
Acknowledgments.	ix
Biographical Sketch	xi
Preface	1
Preface: References	5
Chapter 1: Epigenomes: the missing heritability in human cardiovascular disease?	7
Chapter 1: References	23
Chapter 2: Regulation of chromatin structure in the cardiovascular system	30
Chapter 2: References	51
Chapter 3: Systems proteomics of cardiac chromatin identifies nucleolin as a regulator of growth and cellular plasticity in cardiomyocytes	63
Chapter 3: References	95
Chapter 4: Reciprocal regulation of cardiac chromatin by HMGB and CTCF: implications for transcriptional regulation in pathologic hypertrophy	100
Chapter 4: References	143
Future Directions	151
Future Directions: References	155

LIST OF FIGURES

Preface Figure. Multiple tiers of chromatin organization	4
Figure 1-1. Chromatin variation and disease susceptibility	20
Figure 1-2. Physiological differences in molecular networks affect patient response to therapy	21
Figure 2-1. Mechanisms of chromatin remodeling	47
Figure 2-2. Methodologies and applications	48
Figure 2-3. Multiple levels of structural organization regulate gene expression in an interphase nucleus	49
Figure 3-1. Proteomic quantification of chromatin proteins in murine heart	84-86
Figure 3-2. Identification of Nucleolin as a candidate regulator of cardiac growth	87-88
Figure 3-3. Nucleolin expression in the developing zebrafish embryo	89
Figure 3-4. Loss of <i>nucleolin</i> in zebrafish results in cardiac morphological defects	90
Figure 3-5. Nucleolin is essential for proper cardiac function and myocyte differentiation	91
Figure 3-6. Nucleolin over-expression causes defects in heart chamber looping and dorsal ventral axis formation	92-93
Figure 3-7. Nucleolin regulates growth and plasticity in cardiomyocytes	94
Figure 4-1. HMGB2 and CTCF abundances have opposite relationships to cardiac size	129
Figure 4-2. CTCF and HMGB2 are co-regulated in the mouse heart	130
Figure 4-3. CTCF and HMGB2 can occupy the same loci, but not coincidentally	131
Figure 4-4. HMGB2 and CTCF down-regulate each other	132
Figure 4-5. HMGB2 overexpression strongly represses transcription	133
Figure 4-6. HMGB2 and CTCF knockdown down-regulate nucleolar transcription	134-135

Figure 4-7. HMGB2 binds ribosomal DNA regulating phenotype in a cell-type specific manner	136
Figure 4-8. Chromatin environment partially explains opposing effects of HMGB2 on transcription	137
Figure 4-9. HMGB2 and CTCF regulate cardiac gene accessibility and expression	138
Supplemental Figure 4-1. Basal HMGB2 and CTCF levels predict different disease phenotypes	139
Supplemental Figure 4-2. BALB/cJ and BUB/BnJ respond differently to ISO	139
Supplemental Figure 4-3. CTCF and HMGB2 are not co-regulated in all tissues	139
Supplemental Figure 4-4. Immunohistochemistry on cardiac tissue	140
Supplemental Figure 4-5. CTCF enrichment at HMGB2 binding sites most pronounced in intergenic regions	140
Supplemental Figure 4-6. CTCF and HMGB2 do not co-localize	140
Supplemental Figure 4-7. HMGB2 siRNA control	140
Supplemental Figure 4-8. 5'Fluorouridine labels newly transcribed RNA	141
Supplemental Figure 4-9. HMGB3 knockdown induces hypertrophy	141
Supplemental Figure 4-10. PHE model of hypertrophy	141
Supplemental Figure 4-11. The effect of HMGB2 knockdown on gene expression does not discriminate between genes regulated by different cardiac transcription factors	142

LIST OF TABLES

Table 1-1. Open questions regarding epigenomics as a regulator of heart disease	
heritability	22
Table 2-1. Current state of understanding of various chromatin structural regulators in the	
cardiovascular system	50

ACKNOWLEDGMENTS

I am extremely grateful to have been able to work in Dr. Vondriska's lab. He is selfless in the opportunities and time he gives to train us on every facet of academic research. He has taught me through his example that the two traits necessary for successfully designing and executing a project are insight, which is a reflection of his own intelligence, and a huge amount of hard work. It is very daunting to work with him, and others in our lab and at UCLA and see just how far I have to go to develop my mind to their level. But their example also means that no matter how much I may focus on the things I could have done better in lab on any give day, there is more to grad school than just getting my experiments to work. Being in this lab has meant the opportunity to learn from amazing scientists.

I need to thank all of the members of the Vondriska Lab for working together with me on these projects, including the striking Douglas Chapski, Shanxi Jiang, Elaheh Karbassi, Todd Kimball, Rachel Lopez, Manuel Rosa Garrido, and Irene Shih. The results presented here represent an intersection of projects bringing together the work of Manuel, Elaheh, Rachel, Haodong Chen and Sarah Franklin. I especially need to thank Sarah for guiding me through grad school, Rachel for her enthusiasm and commitment to our project, Manuel for always helping everyone else in the lab push their projects forward and Elaheh for writing two songs for Haodong with me. We worked closely with Kevin Mouillesseaux in Jau-Nian Chen's lab on the zebrafish experiments and with Yibin Wang's lab for the murine disease modeling in the Nucleolin project (Chapter 3). Haodong Chen carried out the HMGB2 ChIP-seq experiments working with Stanley Nelson, and our analysis of chromatin structural protein expression was based on the work of Christoph Rau and Jessica Wang in the labs of Yibin Wang and Jake Lusic (Chapter 4).

I am very appreciative of my committee members Jeff Abramson, Sarah Franklin, Siavash Kurdistani and Yibin Wang. While their expertise is intimidating, their insight, and

willingness to teach me has helped me learn how to frame my experiments around a defined goal.

I would also like to acknowledge Wiley, Japanese Circulation Society and American Physiological Society for permission to use our published work for this thesis. I would like to thank the American Heart Association for providing me a predoctoral fellowship. Research in the Vondriska Lab is supported by the American Heart Association, the National Institutes of Health and the David Geffen School of Medicine at UCLA.

BIOGRAPHICAL SKETCH

Institution and Location	Degree	Year	Field of Study
University of California, Berkeley, CA	BS	2005-2009	Biology
University of California, Los Angeles, CA	PhD student	2010-present	Physiology

A. Honors and Awards

- 2014 Best Poster Award (Gordon Research Conference, Cardiac Regulatory Mechanisms)
- 2012 Research Travel Grant (UCLA, Graduate Division)
- 2012-2013 Certification of Distinction in Teaching. Awarded for my performance as a Teaching Assistant for Life Sciences 2. (UCLA, Life Sciences Division)
- 2011-2012 Certification of Distinction in Teaching. Awarded for my performance as a Teaching Assistant for Life Sciences 2/3 Laboratory. (UCLA, Life Sciences Division)
- 2011 Jennifer S. Buchwald Graduate Fellowship in Physiology (UCLA, Department of Physiology)

B. Publications Manuscripts

- **Monte E**, Vondriska TM. Epigenomes: the missing heritability in human cardiovascular disease? *Proteomics Clin Appl*. 2014; 8(7-8): 480-7.
- Chapski DJ, **Monte E**, Vondriska TM. Positive feedback in cardioprotection: can more mechanism lead to translation? *Circ Res*. 2014; 114(8): 1225-7.
- **Monte E***, Mouillesseaux K*, Chen H, Ren S, Wang Y, Chen JN, Vondriska TM, Franklin S. Systems proteomics of cardiac chromatin identifies Nucleolin as a regulator of growth and cellular plasticity in cardiomyocytes. *Am J Physiol Heart Circ Physiol*. 2013; 305(11): H1624-38. *Denotes equal contribution.
- Rosa-Garrido M, Karbassi E, **Monte E**, Vondriska TM. Regulation of chromatin structure in the cardiovascular system. *Circ J*. 2013; 77(6): 1389-98.
- Chen H, **Monte E**, Parvatiyar MS, Rosa-Garrido M, Franklin S, Vondriska TM. Structural considerations for chromatin state models with transcription as a functional readout. *FEBS Lett*. 2012; 586(20): 3548-54.
- **Monte E**, Chen H, Kolmakova M, Parvatiyar M, Vondriska TM, Franklin S. Quantitative analysis of chromatin proteomes in disease. *JoVE*. 2012; (70).
- Chen H, **Monte E**, Vondriska TM, Franklin S. Systems proteomics of healthy and diseased chromatin. *Methods in Molecular Biology*. Humana Press. 2013.

- Huo YX, Cho KM, Rivera JG, **Monte E**, Shen CR, Yan Y, Liao JC. Conversion of proteins into biofuels by engineering nitrogen flux. *Nat Biotechnol.* 2011;4:346-351. PMID:21378968

Abstracts

- **Monte E**, Kim J, Rau CD, Lopez R, Weiss JN, Lusic AJ, Wang Y, Vondriska TM. Distinct histone proteomes underlie genetic differences in heart failure susceptibility. In: Human Proteome Organization (HUPO) Conference; 2015 Mar 16; Tempe, AZ.
- **Monte E**, Rosa-Garrido M, Karbassi E, Lopez R, Chen H, Vondriska TM. HMGB knockdown promotes hypertrophy and widespread euchromatin abundance in isolated myocytes. In: Gordon Research Conference, Cardiac Regulatory Mechanisms; 2014 June 8-13; New London, NH.
- **Monte E**, Mouillesseaux K, Chen H, Ren S, Wang Y, Chen JN, Vondriska TM, Franklin S. Systems proteomics of cardiac chromatin identifies Nucleolin as a regulator of growth and cellular plasticity in cardiomyocytes. In: Molecular Biology Institute Retreat; 2012 Oct 19-21; Los Angeles, CA.
- Franklin S, Chen H, Karbassi E, **Monte E**, Vondriska TM. Systems epigenomics of chromatin structure in the heart. In: American Heart Association Conference, BCVS; 2012 July 23-26; New Orleans, LA
- **Monte E**, Mouillesseaux K, Chen H, Ren S, Wang Y, Chen JN, Vondriska TM, Franklin SF. Nucleolar proteins controlling cardiac phenotype in rat myocytes and zebrafish revealed by chromatin proteomics. In: American Heart Association Conference, BCVS; 2012 July 23-26; New Orleans, LA.
- **Monte E**, Mouillesseaux K, Kimball T, Vondriska TM, Chen JN, Franklin S. In vivo screening of chromatin remodelers in zebrafish reveals proteins governing cardiac growth. In: Experimental Biology Conference; 2012 April 21-25; San Diego, CA.
- Franklin S, **Monte E**, Mouillesseaux K, Ren S, Wang Y, Chen JN, Vondriska TM. Systems analysis of cardiac chromatin reveals novel roles of chromatin structural proteins in cardiac development and disease. In: Keystone Symposium, Chromatin Dynamics; 2012 Jan 17-22; Keystone, CO.

C. Research Support

High Mobility Group Proteins: Remodeling Chromatin for Heart Failure

American Heart Association Predoctoral Fellowship, 01/2013-12/2014

Preface

There are ~24,000 genes in both the human and mouse genomes. Of those, only approximately one third are expressed in the cardiomyocyte. We know that cell-type specific gene regulation is established during differentiation and involves transcription factors and alterations to the packaging of the DNA, which, in brief, involves the binding of proteins and RNAs. Disease—specifically cardiac hypertrophy and failure—involves changes in expression of ~5% of the genes in the cardiomyocyte. We hypothesize that altered transcription factor abundance, while important, is insufficient to allow for the global reprogramming during cardiac hypertrophy. Rather, there also must be a change in the chromatin landscape to enable an efficient response to the transcription factor regulation.

In addition to changes in specific genes associated with cardiac pathology, cardiac hypertrophy also necessitates up-regulation of housekeeping genes to support the enlarged cell, including those involved in ribosomal biogenesis. In mammals, the ribosome is made up of ~85 proteins (30-50 for 40S subunit and 40-50 for 60S subunit) and 4 ribosomal RNAs (rRNAs; 18S, 5S, 5.8S, 25S)[1]. The 18S, 28S and 5.8S rRNAs all come from the same precursor, whose gene exists in clusters of tandem repeats (>100 copies/genome[2]) across the short ends of multiple acrocentric chromosomes (the specific set of chromosomes depends on the species)[3]. These sequences of DNA colocalize within the nucleus, organized into NORs (nucleolar organizing regions) surrounding the nucleolus[4]. The nucleolus in turn is a sophisticated transcription factory, specialized for ribosomal biogenesis. The nucleolus is composed of a large granular component within which reside multiple clusters of fibrillar centers surrounded by dense fibrillar components[4]. It is at the boundary of the fibrillar center and dense fibrillar component that rRNA transcription occurs. As the rRNA is processed it leaves the dense fibrillar components and moves into the granular component[4]. Like mRNA genes, rRNA gene expression is regulated by histone post-translational modifications and DNA

methylation[5]. Due to their high copy number, 50-90% of rRNA genes are silenced in the basal state of most cells[6].

The nucleolus is responsible for sensing and transmitting multiple different stress signals[7] and plays a major role in cardiac disease[8]. In humans, increased nucleolar size of myocytes is an early indicator of cardiac hypertrophy[9] and also is found in patients with ischemic and dilated cardiomyopathy[10]. However, other reports show that the abundance of agNORs (silver labeled nucleolar organizing regions), while increased in cardiomyocytes of patients with ischemic heart disease, are decreased in patients with ischemic heart disease who also have severe heart failure, potentially due to metabolic limitations on protein synthesis (and thus the need for ribosome biogenesis) in heart failure[11].

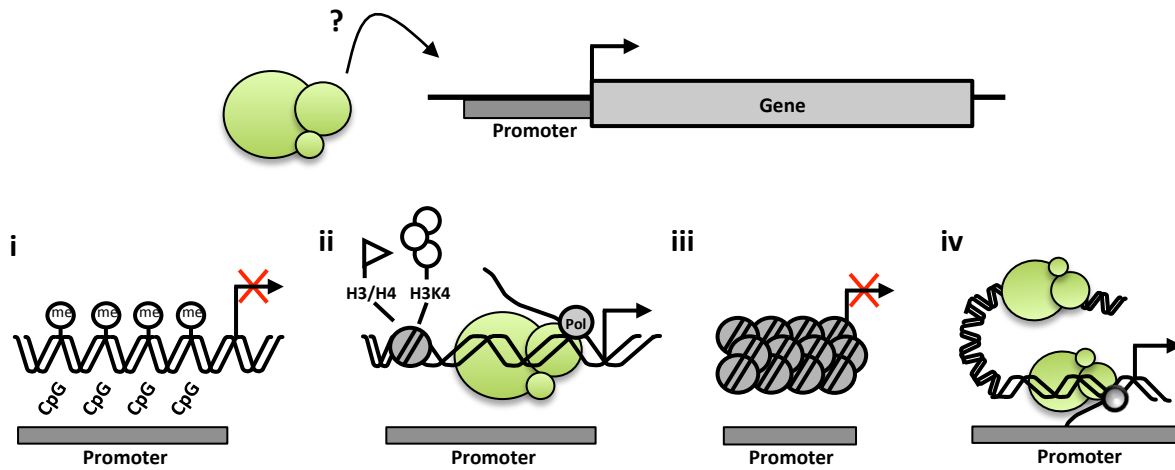
Beyond the fundamental importance to cellular growth, ribosomal biogenesis and nucleolar function also highlight the importance of structural organization for efficient gene expression. The technological advances in DNA sequencing have spurred the development of techniques for measuring novel parameters of chromatin structure as well as generated a huge number of datasets measuring localization of the same chromatin features across multiple models. As a result, the field has created a multi-tiered framework for understanding chromatin structure at the level of individual bases, nucleosomes, genes and regulatory regions, and even large spans of DNA interacting with specific nuclear localization (laminin associated domains [regions associated with nuclear membrane], topological associated domains [regions of DNA in close proximity to each other], chromosome territories [segregated localization of distinct chromosomes]) (Preface Figure and Chapter 2).

These tiers can be co-regulated. DNA methyltransferases can be recruited by histone methyltransferases targeting lysine 9, while blocked by the presence of histone trimethylation on lysine 4 [12]. Methyl-CpG-binding protein 2 (MeCP2) binds to methylated DNA and recruits histone deacetylase[13]. Histone modification can in turn regulate the organization of

nucleosomes through recruitment of chromatin remodeling complexes and structural proteins[14].

Furthermore, we have the ability to extrapolate from a chromatin feature to a functional effect on gene expression. However, a deficiency in our understanding that is common to all tiers of chromatin organization is identification of the DNA sequence and/or chromatin structural cues that faithfully establish the chromatin features at the necessary loci to enable appropriate gene expression. Adding to the complexity is the fact that these conserved mechanisms must build 100s of different chromatin environments on the same genome depending on the cell-type and physiological state of that cell.

My project is focused on chromatin architectural proteins that are present in multiple cell-types, but that we know from proteomic experiments are specifically up-regulated in a mouse model of cardiac hypertrophy[15, 16]. These are proteins that we hypothesized acted across tiers of chromatin scale (i.e. above the level of a gene, but within the confines of larger domains), to regulate fundamental processes involved in chromatin remodeling during disease. In addition to studying their contribution to ribosomal biogenesis, we uncovered roles for these chromatin structural proteins in cardiac development and physiology (Nucleolin) and for cooperation with other chromatin structural regulators in the heart (HMGB2). In Chapter 1, I outline the clinical relevance of chromatin structure from the perspective of heritability in common forms of cardiovascular disease. In Chapter 2, I provide background on the major features of chromatin structure and their role in cardiac development and disease. In Chapters 3 and 4, I present our findings investigating the role of Nucleolin and HMGB2 in regulating cardiac gene expression and phenotype.



Preface Figure. Multiple tiers of chromatin organization. Gene expression is coordinated through multiple scales of chromatin structure. Depicted here are different features which can regulate a gene promoter by recruiting or excluding transcriptional machinery (green complex). **i.** At the level of the nucleotide, DNA methylation on cytosine in the promoter is associated with repressed transcription (red cross through the transcription start site [black arrow]). **ii.** DNA is organized around a complex of histone proteins termed the nucleosome (gray ball). The histone tails can have repressive or activating post-translational modifications (shown here as H3/H4 lysine acetylation and H3 lysine 4 trimethylation). **iii** Compaction of multiple nucleosomes can exclude transcriptional machinery by sequestering DNA binding motifs. **iv.** DNA sequences in three dimensional proximity to the promoter can increase local concentration of activating machinery, increasing its association with the promoter.

Preface: References

1. Nazar, R.N., *Ribosomal RNA processing and ribosome biogenesis in eukaryotes*. IUBMB life, 2004. **56**(8): p. 457-65.
2. Grummt, I., *Life on a planet of its own: regulation of RNA polymerase I transcription in the nucleolus*. Genes & development, 2003. **17**(14): p. 1691-702.
3. Kusnadi, E.P., et al., *Regulation of rDNA transcription in response to growth factors, nutrients and energy*. Gene, 2015. **556**(1): p. 27-34.
4. Farley, K.I., et al., *Determinants of mammalian nucleolar architecture*. Chromosoma, 2015.
5. Bartova, E., et al., *Structure and epigenetics of nucleoli in comparison with non-nucleolar compartments*. The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society, 2010. **58**(5): p. 391-403.
6. Raska, I., et al., *The nucleolus and transcription of ribosomal genes*. Biology of the cell / under the auspices of the European Cell Biology Organization, 2004. **96**(8): p. 579-94.
7. Grummt, I., *The nucleolus-guardian of cellular homeostasis and genome integrity*. Chromosoma, 2013. **122**(6): p. 487-97.
8. Hariharan, N. and M.A. Sussman, *Stressing on the nucleolus in cardiovascular disease*. Biochimica et biophysica acta, 2014. **1842**(6): p. 798-801.
9. Neuburger, M., et al., *Change in size, number and morphology of the nucleoli in human hearts as a result of hyperfunction*. Pathology, research and practice, 1998. **194**(6): p. 385-9.
10. Rosello-Lleti, E., et al., *Influence of heart failure on nucleolar organization and protein expression in human hearts*. Biochemical and biophysical research communications, 2012. **418**(2): p. 222-8.

11. Mamaev, N.N., A.Y. Gudkova, and K.K. Amineva, *AgNORs in the myocardium in ischaemic heart disease complicated by heart failure: a postmortem study*. *Molecular pathology* : MP, 1998. **51**(2): p. 102-4.
12. Cedar, H. and Y. Bergman, *Linking DNA methylation and histone modification: patterns and paradigms*. *Nature reviews. Genetics*, 2009. **10**(5): p. 295-304.
13. Nan, X., et al., *Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex*. *Nature*, 1998. **393**(6683): p. 386-9.
14. Bannister, A.J. and T. Kouzarides, *Regulation of chromatin by histone modifications*. *Cell research*, 2011. **21**(3): p. 381-95.
15. Franklin, S., et al., *Quantitative analysis of the chromatin proteome in disease reveals remodeling principles and identifies high mobility group protein B2 as a regulator of hypertrophic growth*. *Molecular & cellular proteomics* : MCP, 2012. **11**(6): p. M111 014258.
16. Monte, E., et al., *Systems proteomics of cardiac chromatin identifies nucleolin as a regulator of growth and cellular plasticity in cardiomyocytes*. *American journal of physiology. Heart and circulatory physiology*, 2013. **305**(11): p. H1624-38.

Chapter 1: Epigenomes: the missing heritability in human cardiovascular disease?

Emma Monte and Thomas M. Vondriska

[This research was originally published in PROTEOMICS – Clinical Applications by Monte et al. Epigenomes: the missing heritability in human cardiovascular disease? *Proteomics Clin Appl.* 2014; 8(7-8): 480-7. © Wiley.]

Abstract

Cardiovascular disease is a tremendous burden on human health and results from malfunction of various networks of biological molecules in the context of environmental stress. Despite strong evidence of heritability, many common forms of heart disease (heart failure in particular) have not yielded to genome-wide association studies to identify causative mutations acting via the disruption of individual molecules. Increasing evidence suggests, however, that genetic variation in non-coding regions is strongly linked to disease susceptibility. We hypothesize that epigenomic variation may engender different chromatin environments in the absence of (or in parallel with) changes in protein or mRNA sequence and abundance. In this manner, distinct—genetically encoded—chromatin environments can exhibit distinct responses to environmental stresses that cause heart failure, explaining a significant portion of the altered susceptibility that is observed in human disease.

Introduction

One of the most important problems in biology is how the genome is read in different cells. Somatic mutations aside, the sequence of the genome is invariant between cells in eukaryotes. Of course every other tier of biological information varies greatly—from the transcriptome, to the proteome and metabolome and obviously the phenotype—between cell type and amongst copies of the same cell. The developmental basis for this variation has been studied extensively, with detailed transcriptional pathways elucidated to underlie the various mammalian lineages. Indeed it was in this context that the hypothesis of an epigenetic

landscape was put forth by Waddington, wherein gene interactions determine unidirectional cell fate progression [1]. A contemporary definition of epigenetics is more appropriately *heritable (through mitosis and, in rare cases, meiosis) changes that do not involve alteration of the DNA sequence*, although colloquial usage of the term has come to refer to *anything that involves the nucleus and gene expression*. For this paper, we consider epigenetics to refer to persistent changes in chromatin features that establish and/or maintain cellular phenotype.

Cardiovascular disease is the leading cause of death worldwide. The major forms of cardiovascular disease, including myocardial infarction, hypertension, atherosclerosis, heart failure and arrhythmias have all been shown to have a significant heritable component, yet genome-wide association studies (GWAS) have had only mixed success in identifying causative loci in genes [2-4]. One interpretation of this observation is that rather than a few rare variants having a large effect, the genetic component of common diseases like cardiovascular disease arises from multiple common sequence variants acting alone or in a combinatorial manner. Heart failure is the end result of many forms of cardiovascular disease and affects 5-6 million Americans [5]. Discerning the cellular mechanisms underlying the onset and progression of polygenic diseases like heart failure, and moreover how to diagnose and treat them, is a major challenge in clinical medicine. The purpose of this article is to discuss the increasingly recognized role of the epigenome as a point of integration between genetics and environment in the context of complex, common diseases. While the existence of epigenetic mechanisms has been known for decades, we are recently becoming aware of the central role that chromatin structure plays in shaping cellular phenotype and disease response.

Heritability of Cardiovascular Disease: Role of Common Genetic Variation

Cardiovascular disease accounts for 31.9% of deaths in the United States and is the leading cause of death in this country and worldwide [5]. Cardiovascular disease involves a spectrum of inter-related and coincident conditions affecting the heart and vasculature, including

hypertension (high blood pressure), atherosclerosis (lipid deposition and inflammation of the vasculature), myocardial infarction (death of cardiac myocytes), arrhythmias (abnormal cardiac conduction) and heart failure (inability of the heart to supply the oxygen demands of the body). Metabolic syndrome and type II diabetes, which are prevalent in developed societies, are strongly linked with increased risk for cardiovascular diseases [6]. Current clinical management of heart failure involves fluid balance modulation, blood pressure control and regulation of cardiac inotropy, which have powerful effects to mitigate the symptoms of the disease but fail to prevent its progression and have been relatively unchanged for decades. Regenerative therapies remain on the horizon but the advent of low cost genomic sequencing makes personalized heart failure therapy a near term possibility. However, its success is dependent on a greater understanding of the cellular networks through which genetic variation impacts disease.

One of the best-characterized genetic causes of heart disease is hypertrophic cardiomyopathy, caused by mutations in a class of genes mostly encoding sarcomeric proteins. As the name implies, this disease involves increased cardiac myocyte size, increased muscle mass without chamber dilation and a predisposition to arrhythmias and sudden death. Although the effects on the individual patient are devastating, the incidence of hypertrophic cardiomyopathy due to these mutations is estimated at 1 in 500 [7], compared to the approximately 5-6 million Americans known to suffer from heart failure (about 1 in 55). In most cases described clinically, the mutations causing hypertrophic cardiomyopathy occur in the coding regions of sarcomeric genes, disrupting the expression level and/or function of the protein [8]. This is one example where genetic screening is playing a key role in clinical treatment: families known to harbor hypertrophic cardiomyopathy mutations are frequently screened to identify at risk individuals [9]. The study of heart rhythm disorders (e.g. long QT syndrome) has also revealed these diseases to have strong genetic components, often due to mutations in cardiac ion channels [10, 11].

GWAS for coronary artery disease, a frequent precursor to heart failure, have yielded insights into the genetic contributors to this disease [12-15]. A more recent study in dilated cardiomyopathy—a condition not associated with ischemia or obvious vascular lesions—revealed two loci, near the genes *HSPB7* and *BAG3* (both of which encode heat shock/stress response proteins), associated with increased incidence of the disease [16]. Interestingly, when variants in the *BAG3* sequence were further explored it was found that nonsynonymous mutations, which were rare, were linked with monogenic familial forms of the disease (that is, the mutation in *BAG3* caused the disease). Meanwhile, more commonly occurring variation was also observed in the *BAG3* sequence, with these SNPs associated with sporadic (i.e. non-familial) incidence of dilated cardiomyopathy [16]. These findings demonstrate both the heritable and somatic genetic insults that cause disease can do so by converging on the same genomic regions. Dilated cardiomyopathy, like hypertrophic cardiomyopathy, has also been linked to a series of mutations in a small subset of genes, screening of which has made an impact clinically [17]. In a 2013 report, GWAS was used to identify a novel risk locus for dilated cardiomyopathy after which the investigators used expression quantitative trait loci (eQTL) analyses, which investigates how variation in a region of the genome affects the transcriptome, to demonstrate that the locus controls inflammatory gene expression [18].

For common forms of heart failure—as distinguished from familial forms and accounting for the major portion of the disease burden [5]—there is still strong evidence of heritability, although the mechanisms are unclear. In a Framingham cohort of ~1500 patients, longitudinal analyses revealed that having a parent with heart failure increased the risk of the disease to an incidence of 2.72% (versus 1.62% in patients without a parent with heart failure), an ~70% increase in risk [2]. The original study implicating *HSPB7* came from subgenomic analyses of heart failure patients (in a case-control format), showing that SNPs in this gene (as well as in the gene *FRMD4B*) indicate increased risk for heart failure. The association was not observed in an African American sub-cohort of patients [19], but was found to be equally associated with

ischemic and non-ischemic disease, suggesting *HSPB7* is a bona fide risk indicator for heart failure, as opposed to acting only as a marker for one of its clinical precursors like coronary artery disease [20]. Subsequent discovery of *HSPB7* as a risk allele for dilated cardiomyopathy [16] also supports this interpretation.

Despite this progress, known genetic variation explains only a small portion of the observed heritability of heart failure in humans. There are several possible explanations for this (aside from the real possibility that the heritability of the disease is inflated in the first place [21]) including incomplete sampling of individual populations, differences in racial lineages and localization of casual variants near other variants with larger effects. The heritability can also come from variation other than SNPs, including copy number variation (although this was shown not to be the case for early onset myocardial infarction [12]) and chromosomal translocations. An alternative explanation is that the genetic risk reveals itself only through a combination of SNPs, which is in agreement with the common interpretation of GWAS findings in cardiovascular disease as indicating many common SNPs each with small contribution to the phenotype. Studies from the non-cardiovascular arena have shown the utility of such multi-SNP approaches, including for body mass index [22] and type II diabetes [23]. But what are the cellular mechanisms through which these multiple SNPs, many of which reside in non-protein coding regions [24], interact to cause complex common disease? Recent evidence suggests that chromatin may be a regulatory hub through which common genetic variation controls phenotype.

Chromatin Regulation in the Heart

In considering how the epigenome may serve to integrate genetic variation and environmental stimuli, it can be instructive to reflect on what is known about the structural hierarchy of chromatin. DNA alone [25] or in complex with the nucleosome heteromultimer [26] is structurally resolved with atomic resolution. Also characterized *in vitro* are the fibers of DNA-

protein that form when nucleosomes are reconstituted on DNA in a cell free system, which range in scale from 30 to 100s of nanometers and are affected by non-nucleosomal chromatin structural proteins [27, 28]. Beyond this scale, however, our understanding of structure-function relationships within eukaryotic genomes, outside of the context of mitosis, is lacking. This issue has been discussed in other recent papers [29, 30] and is not the focus of this work. It is critical to consider, however, the manner in which chromatin structural changes can affect phenotype and how these changes may come about due to variation in non-coding regions of the genome and/or epigenomic features that integrate genetics and environment.

It is increasingly recognized, that many SNPs associated with cardiovascular and other common diseases lie outside of coding regions, suggesting these variants are acting in trans to regulate gene expression. The commonly assumed mechanism is one in which the SNP falls in a distal enhancer, disrupting its ability to bind a protein responsible for transcriptional regulation and thereby exerting a biological effect. A second, and to our knowledge less explored mechanism involves the disruption of local chromatin structure, presumably by changing a consensus motif that alters the binding of chromatin structural proteins and/or lncRNAs, the function of which have been shown to exert chromatin regulatory effects in the heart [31, 32] and other organs [33] (Figure 1-1).

Many of these mechanisms have been understood for decades—what has changed is the ability to (a) interrogate them in a global manner with locus specificity based on advances in nucleotide sequencing [34]; (b) reveal the extent of the protein [35] and modification [36] landscape based on advances in mass spectrometry-based proteomics; and (c) to reveal combinatorial behavior that confers biological function (that is, groups of proteins/histone PTMs that when localized together affect DNA in a predictable manner [37]). The heart presents several interesting biological challenges to our understanding of how chromatin dynamics regulate phenotype. Genetic engineering in mice has revealed a role for multiple chromatin-regulating enzymes in cardiac and vascular development [38] and developmental mechanisms

may hold insights into novel approaches for cardiac regeneration [39]. How chromatin is regulated during mammalian heart development has been recently explored with next generation sequencing approaches combined with immunoprecipitation of well characterized chromatin marks (ChIP-seq) to reveal the genome-wide patterns of epigenetic modification during lineage commitment in mice [40] and humans [41]. These studies reveal that combinations of chromatin marks distinguish different regions of the genome throughout differentiation, predisposing genes to appropriate activation at the correct developmental stage. These combinations of chromatin marks also discriminate between genes that are drivers of cardiac differentiation and known to participate in disease. Chromatin states—that is, combinations of histone post-translational modifications at a given locus that correspond to a transcriptional behavior—are predictive of gene expression and are dynamically regulated developmentally. For example, proteins with powerful effects on gene expression (like transcription factors) have a unique chromatin signature contrasted with other genes, in which the transcription factors and signaling molecules controlling cardiac development exhibit high H3K27me3 signal in the pluripotent state (which is lost with commitment to the cardiac lineage, commensurate with acquisition of H3K4me3 and H3K36me3). In contrast, genes for structural proteins, which are equally cell type specific but which lack the ability to robustly alter phenotype that is possessed by transcription factors, acquire the activating marks with differentiation but lack a strong silencing mark in the primitive state [41]. Intriguingly, very recent studies explored chromatin modifications in the setting of cardiac hypertrophy [42] using an *in vivo* model of pressure overload, revealing that, similar to development, patterns of histone post-translational modifications appear to specify distinct gene expression regimes. These approaches have been powerful for the identification of new master regulators of differentiation and have revealed the general principles for how different chromatin features affect transcriptional outcomes.

Despite these important advances, much is still unknown about how chromatin is coordinated on a genome-wide scale in the adult myocyte. How genetic variability between

individuals alters the genomic structure/function *in vivo* to alter the transcriptome is unknown in the cardiovascular system, as is at what scale do changes in chromatin structure matter for function (SNPs—histone modifications—chromatin binding proteins—higher order structure). In the realm of phenotypic plasticity, two recent observations suggest potentially therapeutic roles for chromatin remodeling in cardiac rejuvenation: fibroblasts can be directly reprogrammed to cardiac myocytes *in vivo* [43, 44], and the efficiency of reprogramming of fibroblasts to pluripotency is hindered by extensive deposition of silencing marks on chromatin [45]. Could strategies to selectively loosen regions of chromatin necessary for cardiac gene expression complement the current transcription factor-based reprogramming strategies, enhancing efficacy and repair?

Epigenomic Regulation of Adult Phenotypes: Heritability of Diseased Proteomes

It has recently become apparent that proteome diversity is a heritable trait, independent of transcript diversity. A study in humans using a cohort of individuals from the HapMap project demonstrated that proteome variation could be used to define protein quantitative trait loci, or regions of genetic variation that explain proteome level differences between individuals [46]. This level of variation was not captured in the transcriptome, illustrating that these tiers of information vary independently. Previous studies in model organisms had demonstrated this same phenomenon. A study precisely monitoring the abundance of yeast ribosomal genes and proteins demonstrated that genetic variation of proteins was distinct from that of their cognate transcripts [47]. A study in mice also observed low correlation between the loci controlling transcript or protein levels in the liver [48]. Together, these studies support the now widespread observation of a small correlation between abundances of mRNAs and their respective proteins—modern methods of measuring protein abundance have put to rest the concerns that this lack of correlation between mRNA and protein arises from poor quantitation at the level of protein measurements. Transcript networks are related to protein networks through non-linear

relationships and both of these networks can be independently controlled by genetics. An argument for the importance of these findings is that they prompt a need to explain the mechanisms for how these networks interact to transfer the genetic information residing in the genome...bringing the discussion back to chromatin.

The Encyclopedia of DNA Elements (ENCODE) Consortium systematically examined the human epigenome for histone post-translational modifications, measures of DNA accessibility and nucleosome occupancy, transcription factor association and production of RNA. The most recent reports from this study have been very high profile and purported to upend long held dogma regarding the portion of the genome that is functional. ENCODE investigators reported that the vast majority, ~80% according to some estimates, of the genome is 'functional', with large stretches of DNA once thought to be silent now shown to be transcribed at some level [49]. The implications of ENCODE findings have been discussed in detail in other venues and have not escaped vociferous criticism [50]—what these studies serve to highlight is a general appreciation for the importance of chromatin for interpreting genomic function (there are now countless papers demonstrating differences in chromatin marks as predictive of development and disease, as discussed above). Perhaps the best way to determine a causal role for chromatin features is in the context of inheritance. For example, GWAS studies from multiple organs and disease states have shown that 88% of trait associated SNPs reside in introns (45%) or intergenic DNA (43%), that is, outside of protein-coding genes [51]. Many of these disease-associated SNPs live in DNase I hypersensitive sites [24], supporting the concept that their effects may involve changing chromatin structure. Regardless of what these non-coding regions are doing (whether one chooses to call them 'junk DNA' or 'functional elements', for example, is inconsequential), the fact that genetic variation outside of a gene affects the expression of genes and proteins implicates the 3D structure of the genome in the mechanism of action, be it alteration in enhancer function, changes in nucleosome occupancy or disruption of other types of RNAs, like long noncoding RNAs, that may impact

gene expression by modulating chromatin structure/function.

So far, we have discussed how SNPs in non-coding regions can mediate heritability to common diseases by regulating genomic structure to affect transcription. Because the majority of SNPs associated with these diseases are in fact in non-coding regions, this offers an important, and under-explored area of research into genetic predisposition. However, there is also evidence for chromatin to regulate disease susceptibility in a heritable fashion without resulting from genetic mutation. There is a school of thought [52, 53] that various environmental factors can impinge on chromatin in a manner that is heritable, through modulation of DNA methylation or even protein-based chromatin marks such as histone modifications, with recent examples including the role of paternal diet in the offspring's metabolic make-up [54, 55]. But whether epigenetic changes are heritable independent of genetic change can be very hard to prove. As has been described adroitly elsewhere, epigenetic changes like histone modifications and DNA methylation can influence DNA sequence: for example by affecting mismatch repair, which can happen directly in the case of DNA methylation, or by altering the accessibility of DNA repair genes through nucleosome features, thereby affecting DNA sequence in trans. These epigenetic changes throughout life may be selectable, regardless of whether they affect DNA sequence directly, so long as they change the phenotype [56]: there can be a direct causal link between environmental stress, epigenetic change, phenotypic change and natural selection that need not involve changes in DNA sequence. In this manner, chromatin can integrate environmental stress independent from the genome, thereby serving as a substrate for affecting natural selection. To be clear, in this model heritability is still through a conventional, Darwinian genetic mechanism, even if selection (also Darwinian) occurs independent of genetic variation.

A recent multi-center study of genetic cardiomyopathies resulting from de novo mutations made the interesting observation that a preponderance of these disease-inducing new variants occurs in histone modifying enzymes[57]. While most of the mutations affect the gene product—and therefore the pathogenic mechanisms presumably result from altered

regulation of genes regulated by histone modifications deposited by these enzymes—these human genetic studies provide additional evidence for chromatin as a point of convergence for cardiac disease mechanisms.

Epigenomics for Breakthroughs in Cardiovascular Medicine?

In closing, we reflect on how advances in our understanding of multi-genic diseases and chromatin biology together may influence the study and treatment of cardiovascular disease (Table 1-1). As we have discussed, much of the heritable component of cardiovascular disease has not been linked to single gene mutations through GWAS or other studies. How, then, do these multi-genic diseases ‘work’ at the molecular level? While recent evidence has implicated global chromatin remodeling in the response to heart disease [42], we suggest that transformative advances in the diagnosis and treatment of cardiovascular disease will come from systematic analysis of epigenomic features in the context of heritability. That is, how does genetic variation lead to altered chromatin structure, which in turn influences disease susceptibility?

Although many studies in cardiovascular medicine perform multivariable adjustments to correct for comorbid conditions (e.g. the Framingham Offspring study found an increased risk of heart failure in offspring of those with the disease even after correcting for age, sex, height, weight, systolic blood pressure, valve disease, diabetes, prior myocardial infarction and treatment for hypertension [2]), most of these studies do not account for other environmental factors that are often shared within families, including dietary habits, use of recreational drugs, exposure to air pollution, and physical activity as examples. Future studies must take environmental factors into account when studying human populations, including how these environmental factors can impact chromatin to change cellular phenotype. Furthermore, this limitation highlights the importance of model organisms, wherein strict environmental control can be implemented in parallel with the study of genetic effects. A terrific example of this principle is

seen with the hybrid mouse diversity panel: ~100 strains of densely genotyped mice allowing strong genetic differences in phenotype to be explored using modern 'omics techniques [58]. The HMDP has proven to be a powerful model for revealing the genetic basis of common human traits including bone density [59], blood cell parameters [60], inflammation [61] and obesity [62]. Because all the animals in a given strain are genetically identical and all mice in the experiments are exposed to the same environmental conditions, the underlying genetic contribution to complex disease, including the manner in which this manifests through gene, transcript, protein and metabolite networks, can be measured—ruling out a 'reverse causality' situation in which an environmental factor that selects for a given phenotype is incorrectly interpreted as a genetic predisposition.

It is likely that networks of genes, transcripts and proteins are all independently under genetic control and all independently contribute to disease susceptibility and progression. What remains unclear is at what level of resolution we need to understand these networks to target them in disease (Figure 1-2). Chromatin is a perfect example of the intersection between networks of different biological molecules. Determining at what scale of genomic structure chromatin modifications influence pathological transcription will be critical and will involve studies into how altered modification of individual bases (DNA methylation) or chromatin structural proteins (such as histone post-translational modifications) can affect global gene expression and phenotype. How are these marks integrated and coordinated across the genome? What gives rise to myocyte specific chromatin structure and how is this maintained in the terminally differentiated myocyte? Further complicating analysis of cardiovascular disease is that it is a multi-systems disease, in which heritability can come from multiple epigenomes. The same SNP may confer a positive benefit in the context of one organ's epigenome while at the same time being deleterious to cellular function in another organ's epigenome.

A solution moving forward is large human cohort studies that link complete genomic sequencing with deep phenotyping and extensive, quantitative multi-'omics profiling. What

would arise from these studies is a molecular knowledgebase for human cardiovascular health. To accomplish this goal, one would collect multiple non-biased datasets from large human populations (on the scale of 50-100,000) with diverse ethnic background and with/without pre-existing cardiovascular disease, including genomics, proteomics, epigenomics, transcriptomics and metabolomics. This data would be matched with patient electronic medical records to reveal genetic features that control susceptibility to and progression of cardiovascular disease. As a longitudinal study, this project would follow these patients for at least the next decade but preferably for the rest of their lives. One could then investigate the relationship between molecular genetic features and cardiovascular disease traits, to reveal new diagnostic insights and therapeutic approaches. These studies would enable development of new risk prediction algorithms that would enhance decision making for the physician caregiver and improve patient outcomes.

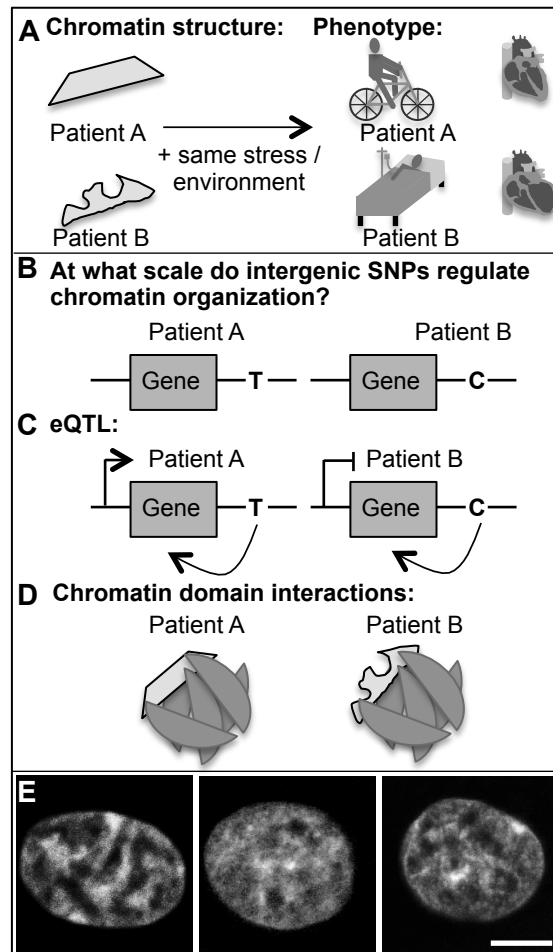


Figure 1-1. Chromatin variation and disease susceptibility. **A.** Two patients will manifest different phenotypes under the same stress due to differences in their epigenomic landscapes, which result in different proteomes. **B.** GWAS hits often reside in intergenic regions. **C.** The majority of validation experiments probe for simple eQTL relationships between a single SNP and a single gene. However, the relative contribution of these simple relationships to phenotype, as compared to more complex interactions between chromatin domains, is unknown (**D**). **E.** A challenge is to distinguish between chromatin packaging differences with functional consequences versus those that do not to alter cellular phenotype. Shown is an example of three DAPI-stained neonatal rat cardiomyocyte nuclei all from a single isolation (note differences in relative DAPI intensity and distinct relative abundance of heterochromatin), representing variability in global chromatin structure within a range ostensibly tolerated by the cell. Scale bar is 5 μm . Images acquired on a Nikon A1 confocal microscopy with a 63x oil immersion lens and converted to black and white in Photoshop.

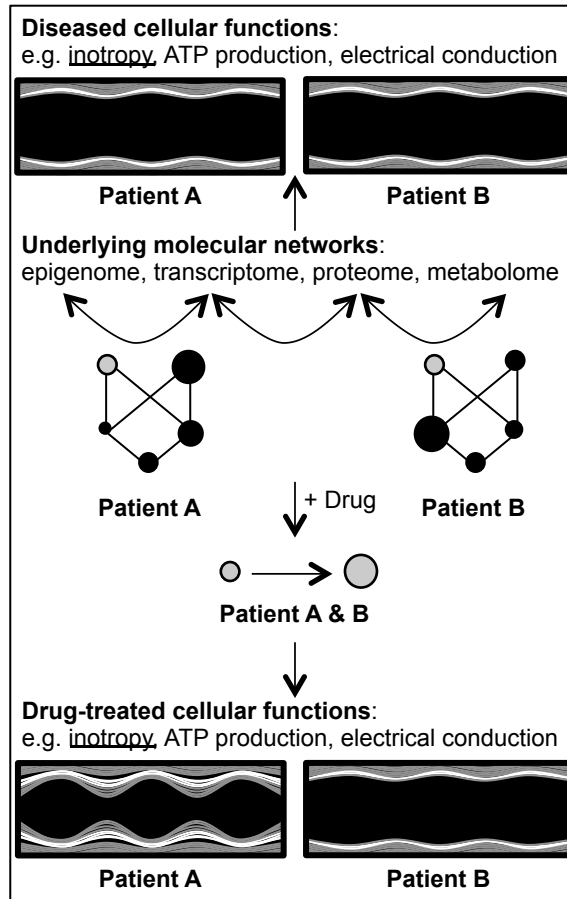


Figure 1-2. Physiological differences in molecular networks affect patient response to therapy. Multiple molecular networks interact to control cell phenotype. In disease, two individuals can present with a similar phenotype (i.e. symptoms), such as impaired cardiac function, but the manner in which the cellular networks are altered to produce the phenotype is different. Here, node size represents molecular abundance. A drug targeting this cellular function may produce the same effect in both individuals at an individual node (in this example, increased expression/activity of the grey node) but the ability of the drug to alter clinical phenotype is dependent on the other components in the networks of the two individuals. Therefore, in only one patient does treatment correct the phenotype, due to the interaction of the drug-induced change with the rest of the network that controls inotropy. Furthermore, the abundance of a given network component can be influenced in non-intuitive ways by genetic variation that impacts chromatin structure directly and transcript/protein abundance through secondary, nonlinear relationships.

Open Questions Regarding Epigenomics as a Regulator of Heart Disease Heritability

- i. How do multi-genic diseases ‘work’ at the molecular level? What is the role of network level interactions and at what level of resolution do we need to understand these networks to target them therapeutically?
- ii. How does genetic variation lead to altered chromatin structure?
- iii. How are chromatin marks integrated and coordinated across the genome? Beyond observing correlations amongst marks, we must determine the logic for how they encode phenotype.
- iv. What gives rise to myocyte-specific chromatin structure and how is this maintained in the terminally differentiated myocyte?
- v. What are the contributions from non-CV epigenomes—that is, from the epigenomes in other organs (e.g. brain, adrenal gland, kidneys)—that together influence function of CV system?

Table 1-1.

Chapter 1: References

- [1] Waddington, C. H., *The Strategy of the Genes*, The Macmillan Company, New York 1957.
- [2] Lee, D. S., Pencina, M. J., Benjamin, E. J., Wang, T. J., *et al.*, Association of parental heart failure with risk of heart failure in offspring. *N Engl J Med* 2006, *355*, 138-147.
- [3] Ryabikov, A., Malyutina, S., Ryabikov, M., Kuznetsova, T., *et al.*, Intrafamilial correlations of carotid intima-media thickness and flow-mediated dilation in a Siberian population. *American journal of hypertension* 2007, *20*, 248-254.
- [4] Snieder, H., Boomsma, D. I., Van Doornen, L. J., De Geus, E. J., Heritability of respiratory sinus arrhythmia: dependency on task and respiration rate. *Psychophysiology* 1997, *34*, 317-328.
- [5] Go, A. S., Mozaffarian, D., Roger, V. L., Benjamin, E. J., *et al.*, Heart Disease and Stroke Statistics--2014 Update: A Report From the American Heart Association. *Circulation* 2013.
- [6] Mottillo, S., Filion, K. B., Genest, J., Joseph, L., *et al.*, The metabolic syndrome and cardiovascular risk a systematic review and meta-analysis. *J Am Coll Cardiol* 2010, *56*, 1113-1132.
- [7] Force, T., Bonow, R. O., Houser, S. R., Solaro, R. J., *et al.*, Research priorities in hypertrophic cardiomyopathy: report of a Working Group of the National Heart, Lung, and Blood Institute. *Circulation* 2010, *122*, 1130-1133.
- [8] Alcalai, R., Seidman, J. G., Seidman, C. E., Genetic basis of hypertrophic cardiomyopathy: from bench to the clinics. *Journal of cardiovascular electrophysiology* 2008, *19*, 104-110.
- [9] Maron, B. J., McKenna, W. J., Danielson, G. K., Kappenberger, L. J., *et al.*, American College of Cardiology/European Society of Cardiology clinical expert consensus document on hypertrophic cardiomyopathy. A report of the American College of Cardiology Foundation Task Force on Clinical Expert Consensus Documents and the European Society of Cardiology Committee for Practice Guidelines. *J Am Coll Cardiol* 2003, *42*, 1687-1713.

- [10] Ackerman, M. J., Priori, S. G., Willems, S., Berul, C., *et al.*, HRS/EHRA expert consensus statement on the state of genetic testing for the channelopathies and cardiomyopathies this document was developed as a partnership between the Heart Rhythm Society (HRS) and the European Heart Rhythm Association (EHRA). *Heart rhythm : the official journal of the Heart Rhythm Society* 2011, 8, 1308-1339.
- [11] Schwartz, P. J., Ackerman, M. J., George, A. L., Jr., Wilde, A. A., Impact of genetics on the clinical management of channelopathies. *J Am Coll Cardiol* 2013, 62, 169-180.
- [12] Kathiresan, S., Voight, B. F., Purcell, S., Musunuru, K., *et al.*, Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet* 2009, 41, 334-341.
- [13] McPherson, R., Pertsemlidis, A., Kavaslar, N., Stewart, A., *et al.*, A common allele on chromosome 9 associated with coronary heart disease. *Science* 2007, 316, 1488-1491.
- [14] Helgadottir, A., Thorleifsson, G., Manolescu, A., Gretarsdottir, S., *et al.*, A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science* 2007, 316, 1491-1493.
- [15] Samani, N. J., Erdmann, J., Hall, A. S., Hengstenberg, C., *et al.*, Genomewide association analysis of coronary artery disease. *N Engl J Med* 2007, 357, 443-453.
- [16] Villard, E., Perret, C., Gary, F., Proust, C., *et al.*, A genome-wide association study identifies two loci associated with heart failure due to dilated cardiomyopathy. *European heart journal* 2011, 32, 1065-1076.
- [17] Moretti, M., Merlo, M., Barbati, G., Di Lenarda, A., *et al.*, Prognostic impact of familial screening in dilated cardiomyopathy. *European journal of heart failure* 2010, 12, 922-927.
- [18] Meder, B., Ruhle, F., Weis, T., Homuth, G., *et al.*, A genome-wide association study identifies 6p21 as novel risk locus for dilated cardiomyopathy. *European heart journal* 2013.
- [19] Cappola, T. P., Li, M., He, J., Ky, B., *et al.*, Common variants in HSPB7 and FRMD4B associated with advanced heart failure. *Circulation. Cardiovascular genetics* 2010, 3, 147-154.

- [20] Dorn, G. W., 2nd, Genetics of common forms of heart failure. *Current opinion in cardiology* 2011, 26, 204-208.
- [21] Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., *et al.*, Finding the missing heritability of complex diseases. *Nature* 2009, 461, 747-753.
- [22] Ehret, G. B., Lamparter, D., Hoggart, C. J., Whittaker, J. C., *et al.*, A multi-SNP locus-association method reveals a substantial fraction of the missing heritability. *American journal of human genetics* 2012, 91, 863-871.
- [23] Kang, C., Yu, H., Yi, G. S., Finding type 2 diabetes causal single nucleotide polymorphism combinations and functional modules from genome-wide association data. *BMC medical informatics and decision making* 2013, 13 Suppl 1, S3.
- [24] Maurano, M. T., Humbert, R., Rynes, E., Thurman, R. E., *et al.*, Systematic localization of common disease-associated variation in regulatory DNA. *Science* 2012, 337, 1190-1195.
- [25] Watson, J. D., Crick, F. H., Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 1953, 171, 737-738.
- [26] Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., Richmond, T. J., Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 1997, 389, 251-260.
- [27] Ghirlando, R., Felsenfeld, G., Chromatin structure outside and inside the nucleus. *Biopolymers* 2013, 99, 225-232.
- [28] Song, F., Chen, P., Sun, D., Wang, M., *et al.*, Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science* 2014, 344, 376-380.
- [29] Chen, H., Monte, E., Parvatiyar, M. S., Rosa-Garrido, M., *et al.*, Structural considerations for chromatin state models with transcription as a functional readout. *FEBS letters* 2012, 586, 3548-3554.
- [30] Sanyal, A., Bau, D., Marti-Renom, M. A., Dekker, J., Chromatin globules: a common motif of higher order chromosome structure? *Current opinion in cell biology* 2011, 23, 325-331.

- [31] Klattenhoff, C. A., Scheuermann, J. C., Surface, L. E., Bradley, R. K., *et al.*, Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. *Cell* 2013, 152, 570-583.
- [32] Grote, P., Wittler, L., Hendrix, D., Koch, F., *et al.*, The Tissue-Specific lncRNA Fendrr Is an Essential Regulator of Heart and Body Wall Development in the Mouse. *Developmental cell* 2013, 24, 206-214.
- [33] Lee, J. T., Epigenetic regulation by long noncoding RNAs. *Science* 2012, 338, 1435-1439.
- [34] Maunakea, A. K., Chepelev, I., Zhao, K., Epigenome mapping in normal and disease States. *Circ Res* 2010, 107, 327-339.
- [35] Monte, E., Chen, H., Kolmakova, M., Parvatiyar, M., *et al.*, Quantitative analysis of chromatin proteomes in disease. *J Vis Exp* 2012.
- [36] Karch, K. R., Denizio, J. E., Black, B. E., Garcia, B. A., Identification and interrogation of combinatorial histone modifications. *Frontiers in genetics* 2013, 4, 264.
- [37] Ernst, J., Kellis, M., Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nature biotechnology* 2011, 28, 817-825.
- [38] Chang, C. P., Bruneau, B. G., Epigenetics and cardiovascular development. *Annual review of physiology* 2012, 74, 41-68.
- [39] Mercola, M., Ruiz-Lozano, P., Schneider, M. D., Cardiac muscle regeneration: lessons from development. *Genes Dev* 2011, 25, 299-309.
- [40] Wamstad, J. A., Alexander, J. M., Truty, R. M., Shrikumar, A., *et al.*, Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. *Cell* 2012, 151, 206-220.
- [41] Paige, S. L., Thomas, S., Stoick-Cooper, C. L., Wang, H., *et al.*, A temporal chromatin signature in human embryonic stem cells identifies regulators of cardiac development. *Cell* 2012, 151, 221-232.

- [42] Papait, R., Cattaneo, P., Kunderfranco, P., Greco, C., *et al.*, Genome-wide analysis of histone marks identifying an epigenetic signature of promoters and enhancers underlying cardiac hypertrophy. *Proc Natl Acad Sci U S A* 2013, *110*, 20164-20169.
- [43] Song, K., Nam, Y. J., Luo, X., Qi, X., *et al.*, Heart repair by reprogramming non-myocytes with cardiac transcription factors. *Nature* 2012, *485*, 599-604.
- [44] Qian, L., Huang, Y., Spencer, C. I., Foley, A., *et al.*, In vivo reprogramming of murine cardiac fibroblasts into induced cardiomyocytes. *Nature* 2012, *485*, 593-598.
- [45] Soufi, A., Donahue, G., Zaret, K. S., Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* 2012, *151*, 994-1004.
- [46] Wu, L., Candille, S. I., Choi, Y., Xie, D., *et al.*, Variation and genetic control of protein abundance in humans. *Nature* 2013, *499*, 79-82.
- [47] Foss, E. J., Radulovic, D., Shaffer, S. A., Goodlett, D. R., *et al.*, Genetic variation shapes protein networks mainly through non-transcriptional mechanisms. *PLoS biology* 2011, *9*, e1001144.
- [48] Ghazalpour, A., Bennett, B., Petyuk, V. A., Orozco, L., *et al.*, Comparative analysis of proteome and transcriptome variation in mouse. *PLoS genetics* 2011, *7*, e1001393.
- [49] Bernstein, B. E., Birney, E., Dunham, I., Green, E. D., *et al.*, An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012, *489*, 57-74.
- [50] Graur, D., Zheng, Y., Price, N., Azevedo, R. B., *et al.*, On the immortality of television sets: "function" in the human genome according to the evolution-free gospel of ENCODE. *Genome biology and evolution* 2013, *5*, 578-590.
- [51] Hindorff, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., *et al.*, Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* 2009, *106*, 9362-9367.
- [52] Jirtle, R. L., Skinner, M. K., Environmental epigenomics and disease susceptibility. *Nat Rev Genet* 2007, *8*, 253-262.

- [53] Morgan, D. K., Whitelaw, E., The case for transgenerational epigenetic inheritance in humans. *Mammalian genome : official journal of the International Mammalian Genome Society* 2008, 19, 394-397.
- [54] Ng, S. F., Lin, R. C., Laybutt, D. R., Barres, R., *et al.*, Chronic high-fat diet in fathers programs beta-cell dysfunction in female rat offspring. *Nature* 2010, 467, 963-966.
- [55] Carone, B. R., Fauquier, L., Habib, N., Shea, J. M., *et al.*, Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals. *Cell* 2010, 143, 1084-1096.
- [56] Turner, B. M., Environmental sensing by chromatin: An epigenetic contribution to evolutionary change. *FEBS letters* 2011.
- [57] Zaidi, S., Choi, M., Wakimoto, H., Ma, L., *et al.*, De novo mutations in histone-modifying genes in congenital heart disease. *Nature* 2013, 498, 220-223.
- [58] Ghazalpour, A., Rau, C. D., Farber, C. R., Bennett, B. J., *et al.*, Hybrid mouse diversity panel: a panel of inbred mouse strains suitable for analysis of complex genetic traits. *Mammalian genome : official journal of the International Mammalian Genome Society* 2012, 23, 680-692.
- [59] Farber, C. R., Bennett, B. J., Orozco, L., Zou, W., *et al.*, Mouse genome-wide association and systems genetics identify *Asxl2* as a regulator of bone mineral density and osteoclastogenesis. *PLoS genetics* 2011, 7, e1002038.
- [60] Davis, R. C., van Nas, A., Bennett, B., Orozco, L., *et al.*, Genome-wide association mapping of blood cell traits in mice. *Mammalian genome : official journal of the International Mammalian Genome Society* 2013, 24, 105-118.
- [61] Orozco, L. D., Bennett, B. J., Farber, C. R., Ghazalpour, A., *et al.*, Unraveling inflammatory responses using systems genetics and gene-environment interactions in macrophages. *Cell* 2012, 151, 658-670.

[62] Parks, B. W., Nam, E., Org, E., Kostem, E., *et al.*, Genetic control of obesity and gut microbiota composition in response to high-fat, high-sucrose diet in mice. *Cell metabolism* 2013, 17, 141-152.

Chapter 2: Regulation of chromatin structure in the cardiovascular system

Manuel Rosa-Garrido, Elaheh Karbassi, Emma Monte, Thomas M Vondriska

[This research was originally published in *Circulation Journal* by Rosa Garrido et al. Regulation of chromatin structure in the cardiovascular system. *Circ J.* 2013; 77(6): 1389-98. © Japanese Circulation Society.]

Abstract

It has been appreciated for some time that cardiovascular disease involves large-scale transcriptional changes in various cell types. What has become increasingly clear only in the last few years, however, is the role of chromatin remodeling in cardiovascular phenotypes in normal physiology as well as in development and disease. This review summarizes the state of the chromatin field in terms of distinct mechanisms to regulate chromatin structure *in vivo*, identifying when these modes of regulation have been demonstrated in cardiovascular tissues. We describe areas in which a better understanding of chromatin structure is leading to new insights into the fundamental biology of cardiovascular disease.

Introduction

The ordered packing in the nucleus of six billion base pairs of DNA, along with thousands of proteins and RNAs, is one of the most remarkable feats of molecular organization in biology. Genomic structure is highly dynamic *in vivo*: mitotic reorganization of the chromosomes is exquisitely choreographed and, when not replicating itself, the genome can scaffold rapid and large-scale changes in transcription. Genomic function is inherently dependent on non-DNA factors and is likewise quite plastic: the same DNA holds the information to create every cell in a multicellular eukaryote which, when read by the appropriate protein and RNA machinery, can lead to normal differentiation, multipotency, or disease.

The functional unit of chromatin is the nucleosome, a heteromultimeric protein complex comprised of 8 proteins and bound by ~150 bp of DNA, the structure of which is known at atomic resolution.(1) The precise structural features of the endogenous genome are unknown but there are several key properties that influence the likelihood of a region of DNA to be available for transcription, including: histone variants (and their post-translational modification, PTM), histone modifying enzymes and ATP-dependent chromatin remodeling proteins, non-nucleosomal chromatin structural proteins, and non-protein-coding RNAs. The DNA sequence itself also has a major role in genomic structure, and modifications to this sequence (e.g. cytosine methylation) participate in local accessibility and transcription. These DNA and non-DNA factors together endow global features alternatively associated with gene expression, referred to as euchromatin, or transcriptional repression, labeled heterochromatin (which can further be divided into the constitutive or facultative variety, depending on whether it is universal or cell type-specific, respectively; Figures 2-1 and 2-3). To understand transitions in genomic structure that underlie changes in phenotype of the cell, the logic through which these multiple factors influence genomic structure must be revealed.

Cardiovascular disease is a scourge on the developed world, where it is the leading cause of morbidity and mortality. Despite the development of new therapeutic approaches,(2) heart failure is a particularly debilitating form of cardiovascular disease, with a ~50% death rate 5 years after diagnosis.(3) The last two decades have seen remarkable progress in our understanding of the molecular events that mediate heart failure, including its antecedent, cardiac hypertrophy. The role of transcriptional circuits has been extensively studied in cardiac development and disease; in contrast, the role of endogenous chromatin structure and its remodeling to facilitate transcriptional responses (thereby underpinning cell type-specific proteomes, and phenotypes) in the cardiovascular system is a recently emerging field. Our purpose herein is to succinctly summarize the current understanding of chromatin biology in the cardiovascular system, with a particular emphasis on the heart. We also aim to highlight

concepts and techniques (Figure 2-2) from non-cardiovascular systems that reveal new avenues of investigation to understand cardiovascular biology and disease.

Histone Modifications and Histone Variants

Although histones have been known for some time to be regulated by PTM, recent advances in proteomic mass spectrometry over the past decade have advanced our understanding of the extent of this regulation, and dramatically increased the rote number of modifications that have been *identified* (>100 for some cell types).(4) An advantage of mass spectrometry is the ability to study combinations of PTMs on a single peptide or even intact protein.(5) One example of this approach comes from analysis of histone H3.2 PTMs in HeLa cells where differential PTM resulted in 150 different forms, or “protein species”, of H3.2.(6) In addition to combinatorial PTM identification, intact proteins can also be useful to study splice variants, and histone variants with high sequence similarity, where there are few unique tryptic peptides.

Functionally *characterizing* these modifications, on the other hand, remains a low throughput endeavor, and thus there are a handful of modifications for which specific transcriptional phenotypes have been ascribed and subsequently validated in multiple cell types. Commonly studied silencing marks, which by definition are associated with DNA regions with lower expression, include H3K9me3 (constitutive heterochromatin)(7) and H3K27me3 (facultative heterochromatin),(8) where constitutive heterochromatin is considered more permanently silenced. Commonly studied euchromatin marks include H3K4me3, which marks promoters and transcriptional elongation,(9) H3K4me1, which marks enhancers,(10) H3K27ac, which marks active enhancers,(11) and H3K36me3, which marks actively transcribed gene bodies.(12) In general, histone acetylation makes the DNA more euchromatic.(13) Less well-studied PTMs include phosphorylation, ubiquitination and SUMOylation, although all these modifications have been shown to exist on chromatin.

Nucleosomes can also be modified by the histone variants they contain. Histone variants have diverse roles, including association with transcriptional elongation in active gene bodies, kinetochore formation at the centromere, RNA Pol II recruitment at promoters, and chromatin remodeler recruitment after DNA breaks, as just some examples.(14) Finally, linker histone H1 is another class of histone, with its own panel of variants that associate peripherally with nucleosomes, organizing them into higher-order structure.(15)

Chromatin immunoprecipitation followed by next generation DNA sequencing (ChIP-seq) has become a widely used tool for studying the distribution of specific histone PTMs and variants (as well as other chromatin-bound proteins) across the genome. Integration of data sets from multiple ChIP-seq experiments enables localization of PTMs which may cooperate together to form a histone code,(16) or which may, even in the absence of existing on the same nucleosome, define chromatin domains.(17, 18)

Perhaps the most mature field of chromatin research in the heart is the regulation of histone modifications (Figure 2-1), acetylation in particular, which is governed by HDACs (histone deacetylases) and HATs (histone acetyltransferases).(19) Despite important insights from many cardiac studies on HDAC and HAT isoforms—primarily executed using pharmacologic and/or genetic gain/loss of function of individual isoforms or protein families—a critical unresolved question is which residues on the histone tails are modulated during disease, and how do the HDACs/HATs (and other modifying enzymes, such as histone methyltransferases) target the correct nucleosomes to bind and modify in their target genes. It is likely that a combination of histone variant expression, histone PTMs, DNA sequence and non-nucleosomal chromatin binding proteins together specify a disease-specific genomic structure and transcriptional response in cardiomyocytes (thus explaining how global inhibition of HDACs and HATs can have apparently specific effects), but a combination of proteomics to dissect HDAC/HAT complexes in the heart and genomics to map the localization of these molecules would be needed to test this conjecture.

One salient study in which ChIP-seq revealed a global principle of gene regulation in the heart has come in the area of development. During the process of differentiating human embryonic stem cells (hESCs) into cardiomyocytes, H3K27me3 decreases, while H3K36me3 and H3K4me3 increase at the promoters of genes that regulate cardiac differentiation.(20) This differs from cardiac structural proteins that also gain H3K36me3 and H3K4me3 during differentiation but lack inhibitory H3K27me3 even in less differentiated states (and even when these genes are silenced).(20) This suggests that repressive marks are not necessary to block transcription, but rather, are employed as an extra precaution (in addition to the lack of activating marks) to keep genes that promote differentiation tightly silenced in precursor lineages, since aberrant expression of a proteins whose expression directly induces the expression of other proteins (e.g. a developmental transcription factor) would be more detrimental than aberrant expression of a single structural protein.(20) A similar study performing ChIP-seq and RNA-seq during the differentiation of mouse ESCs into cardiomyocytes found that multiple genes showing a similar expression pattern during differentiation could be subdivided by chromatin marks to separate out genes coding for different gene ontology analysis-defined pathways.(21) They then used the presence of H3K4me1 alone, or H3K4me1 and H3K27ac together, to identify poised or active cardiac enhancers, respectively.(21) Interestingly, many poised enhancers in earlier lineages never became active in the cardiac linages, which the authors attribute to the plasticity of the precursor cells (i.e., enhancers poised for activation in other lineages).(21) Furthermore, the group of active enhancers was highly cell type-specific.(21) H3K27ac levels at active enhancers help explain the convergence of chromatin structural features and transcription factors to coordinate large-scale gene expression profiles. This study found that loci with high levels of H3K27ac surrounding a small dip in H3K27ac signal represented active enhancers which were open to transcription factor (TF) binding (hence the dip).(21) Taken together, these studies show how the chromatin patterns at promoters and enhancers progress during development to

modulate the multiple large-scale gene expression patterns being driven by the—simultaneous, developmentally speaking—changing abundances of distinct transcription factors.

ChIP-seq for H3K4me3, H3K4me1 and H3K27ac was performed on adult mouse hearts and embryonic mouse hearts, day E14.5 and used, in conjunction with ChIP-seq for other chromatin proteins, to identify promoters and enhancers.(22) Typically, the enhancer is matched to the closest TSS or to the TSSs not separated from the enhancer by a CTCF binding event. However, this study compared the chromatin state at each enhancer to the presence of RNA Pol II at every promoter on the chromosome, to find the cis- and trans-regulated promoters whose activities are best correlated with that of the enhancer.(22) Rather than explain developmental or pathological processes specifically, this data offers insight into the structural (domains of interacting loci) and functional organization of the cardiac genome.

ChIP followed by microarray analysis of myocytes isolated from healthy or failing hearts of Dahl salt-sensitive rats shows that global distribution of H4K20me3, H3K27me2, H3ac, H4ac, H3K4me2 and H3K9me2 are significantly correlated between healthy and failing hearts.(23) However, H3K4me3 (as opposed to H3K4me2) and H3K9me3 (as opposed to H3K9me2) show a significant lack of correlation between healthy and disease states.(23) ChIP-seq for H3K4me3 and H3K9me3 were repeated on human heart samples from patients with valvular disease and preserved ejection fraction or patients with dilated cardiomyopathy and depressed ejection fraction.(23) Loci of 1kb with multiple reads mapped to it were deemed “high clusters” of H3K4me3 or H3K9me3, and were found to be enriched in cardiac specific genes.(23) In fact, even high clusters only found in one or the other of the two disease patient groups were part of pathways which overlapped between the two groups, even if the specific genes with the altered H3 PTMs did not.(23) This suggests that there is enrichment of H3K9me3 and H3K4me3 high-density regions at cell-specific genes, as opposed to, for example, predominately being found at developmentally silenced and housekeeping genes respectively, and as such, it is these PTMs that show the greatest change during disease.

Features of the “histone code”, if it exists, have been described to include writers, readers and erasers that respectively add, recognize or remove modifications from histones. There are two general approaches to studying the relationship amongst these types of chromatin modifiers: single gene-based approaches, which examine the behavior around well-characterized loci, drawing broad conclusions, and genome-wide or proteome-wide approaches, that are loci independent but that, for reasons of scale, lack extensive traditional analysis of individual target genes. One recent study(24) in the latter category determined the proteins associated with known trimethylation marks (five in total: H3K4me3, H3K36me3, H3K9me3, H3K27me3, and H4K20me3) as a discovery-based approach to map the proteomes of molecules ostensibly reading these histone PTMs.

Finally, histone variants, in addition to PTMs, play a role in heart disease. Quantitative proteomic analysis of proteins isolated from mouse cardiac nuclei by acid extraction found 54 histone protein variants in the heart, with a global decrease in the ratio of linker histone to core nucleosome in cardiac hypertrophy.(25) Additionally, histone H2A.Z was found to increase in a mouse model of hypertrophy, with shRNA experiments in isolated cells demonstrating its necessity in this process.(26) Proteins peripheral to the nucleosome and yet fundamentally involved in chromatin structure, including the H1 family (so-called linker histones(25)) and high mobility group, HMGA1(27) and HMGB2(25), specifically, have all been implicated in pathological cardiac growth. The role of other histone variants and non-nucleosomal chromatin structural proteins, and how they target specific nucleosomes in the heart, remains to be elucidated.

DNA Modifications

In mammals, DNA methylation occurs on cytosine residues with an adjacent guanine residue on the 3' end, so called CpG sites, forming 5-methylcytosine.(28) The majority of CpG sites are methylated in vertebrates, however, there are regions of DNA, called CpG islands,

which are enriched for CpG and are hypomethylated.(29) Analysis of multiple vertebrate genomes shows CpG islands enriched at the 5' end of housekeeping genes, as well as some occurring at the 5' or 3' end of tissue-specific genes.(30) It was thought that disease-specific changes in methylation occurred at CpG islands and gene promoters, however a study on human colon cancer found that the majority of the changes were occurring in a new region, coined "CpG island shores," which are regions within 2kb of CpG islands with less dense CpG nucleotides.(31)

Methyl CpG binding protein 2 (MeCP2) is a chromatin protein that can bind to 5-methylcytosine and facilitate gene compaction and silencing.(32) It is considered a general rule that cytosine methylation silences transcription, including repressing transposon expression.(33) However, there are nuances. Highly expressed genes tend to have hypomethylation in the promoter, but hypermethylation in the gene body. It is thus not surprising that the methylation patterns vary between cell types. Methylation array analysis of human heart, lung, and kidney found tissue-specific methylation patterns.(34)

DNA (cytosine-5-)-methyltransferase 1 (Dnmt1), acts on hemimethylated DNA to maintain methylation,(35) while Dnmt3A and Dnmt3B are responsible for *de novo* methylation,(36) with some exceptions.(37) DNA methylation can be maintained between cell divisions, making it truly epigenetic.(38) However, DNA demethylation can also occur during cell division (passively), and it has recently become appreciated that demethylation can also be an active process involving hydroxymethylation; hydroxymethyl-cytosine is enriched at enhancers and active promoters and exons in ESCs and may poise chromatin for transcription.(39) Additionally, 5-methylcytosine is enriched at nucleosomes and in exons.(40)

The most widely applied method for genome-wide analysis of DNA methylation is bisulfite sequencing, in which treatment of DNA with bisulfite converts unmethylated cytosines to uracil.(41) Following comparison with a reference genome, the mismatched bases are inferred to be unmethylated cytosines. Other methods that include targeted approaches using

designer probes or methylation sensitive restriction enzyme digestion have advantages of increased coverage of select regions of the genome and of better specificity in terms of single base resolution and less bias for size of sequenced fragments, respectively. A final sequencing-based method is a variation of ChIP-seq, where antibody to methylated cytosine is used, so called methylated DNA immunoprecipitation (MeDIP).(42) MeDIP analysis is biased toward loci with high methylation,(42) and since the base methylation status is not read, single nucleotide resolution is not obtained.

There is an increasing recognition that specific genes may be differentially methylated in heart and vascular disease and, with the advent of the aforementioned genome-wide technologies, the extent of this regulation is becoming more evident. Some recent examples include a study on ventricular septal defects, the most common form of congenital heart disease, which found hypermethylation in the NOX5 gene, which plays a role in septum development.(43) Another study analyzed human left ventricular samples by MeDIP-chip (MeDIP followed by analysis on a microarray chip containing CpG island and promoter sequences) followed by bisulfite treatment and PCR for confirmation.(44) This study found altered methylation in disease corresponding to altered expression at 3 genes involved in angiogenesis.(44) Differential methylation can also affect intergenic regions. One study found human heart samples from end-stage cardiomyopathic patients were significantly hypomethylated in satellite repeat regions and these regions were up-regulated 27-fold in expression over control hearts.(45) In a separate study, pregnant mice exposed to the endocrine disrupting compound diethylstilbestrol birthed offspring which exhibited altered contractility when exercised, in part due to increased expression of Dnmt3a and increased methylation in the promoter of calsequestrin 2.(46) A study of selenium deficiency, associated with diastolic dysfunction and myocardial fibrosis in mice, found that low doses of selenium supplementation were capable of altering DNA methylation due to changes in the methionine-

homocysteine cycle.(47) This suggests that methylation in the heart can be quite dynamic and potentially sensitive to a range of stimuli.

MeDIP followed by sequencing analysis of healthy and diseased human left ventricular samples showed significant differences in global methylation of CpG islands of gene promoters (cardiomyopathy samples hypomethylated versus control) and gene bodies (cardiomyopathy samples hypermethylated versus control), but not in 3'UTRs, enhancers, or intergenic regions.(48) The occurrence in disease of increased hypomethylation in promoters, increased hypermethylation in gene bodies, and lack of alteration in promoters of down-regulated genes together suggest that there is a global change in methylation during disease which helps up-regulate a subset of genes. If this overall mechanism is more universal, and how the specific pattern of methylation varies depending on the type of cardiovascular disease, remains unknown.

DNA methylation is also an important part of cellular reprogramming and differentiation. Multiple studies have evaluated the effectiveness of using 5-azacytidine (a nucleoside analogue which inhibits DNA methylation) on driving mesenchymal stem cells toward a cardiomyogenic lineage,(49) including a proteomic study identifying the cytoskeletal and metabolic proteins whose expression is altered after treatment with the compound.(50) Additionally, MeDIP analysis coupled with CpG island arrays on mouse brain, heart, liver and testis tissue taken at multiple developmental time periods (E15 to adult) shows that the majority of methylation differences are due to tissue specificity.(51) Thus, methylation is an important part of endowing cell specificity during differentiation and a deeper understanding of this regulation can be used to manipulate cellular reprogramming.

Noncoding RNA

Advances in genomic technologies have enabled investigation of the complete transcriptome of a cell. These analyses have reinforced the concept that RNA is not only an

intermediary molecule between DNA and proteins, instead executing various tasks related to gene expression, macromolecular complex formation, mRNA translation and chromatin structure (Figure 2-1). In fact, around 98% of all transcriptional output in humans is non-protein-coding RNA,(52) with microRNAs being particularly well-studied in the molecular basis of cardiac development and disease(53, 54) including as potential biomarkers.(55) Of the two major techniques used to analyze the regulatory role of noncoding RNAs (ncRNA), the main advantages of RNA-seq over RNA microarrays are that the former does not require the target genomic sequence on the array and produces lower background levels.

Several recent investigations using RNA-seq in mouse models of heart disease have shown a widespread role for other forms of RNA processing, including splicing(56, 57), although the extent to which RNAs function to directly affect chromatin structure and gene expression in cardiovascular disease, is unknown.

High-throughput techniques have shown evidence for production of large amounts of long non-coding RNAs from the human genome.(58) These lncRNAs are molecules bigger than 200 bp and lack protein-coding potential. lncRNAs have been shown to have many interesting actions(59) including gene silencing, enhancer activity, alternative splicing and as scaffolds for complexes of proteins bound to the genome (Figure 2-1). Two very recent studies have demonstrated the existence of tissue-specific lncRNAs involved in cardiac function, including the mesoderm-specific *Fendrr*, loss of which is embryonic lethal due to cardiac hypoplasia and functional defects,(60) and *Braveheart*, required for commitment of embryonic stem cells to the cardiac lineage.(61) The extent to which these or other ncRNAs participate in normal cardiac function in adults, including their roles in disease, remain to be determined.

Chromatin Remodelers

ATP-dependent chromatin remodelers are a class of chromatin proteins important for dynamic reorganization and maintenance of chromatin structure (Figure 2-1). There are 5 main

families of chromatin remodeling complexes: ISWI and SWI/SNF (which are the best studied families), INO80, NURD/Mi-2/CHD, and SWRI.(62) These complexes can slide/space(63) or eject nucleosomes(64) to alter access to the DNA by transcription factors. Additionally, they can remove the H2A-H2B dimer(65) and exchange dimers to incorporate histone variants(66) as revealed by single molecule methods.(67) This set of nucleosome manipulations enables chromatin remodelers to participate in a variety of biological processes.(62) For example, the SWI/SNF family is involved in RNA polymerase regulation,(68) as well as alternative splicing(69) and transcriptional elongation.(70) The ISWI family also regulates elongation.(71) The NURD/Mi-2/CHD family can facilitate gene silencing, with Mi-2 able to associate simultaneously with HDACs and methylated DNA binding proteins, linking DNA methylation to histone deacetylation and transcriptional repression.(72) Outside of affecting transcription, the SWRI proteins are important for DNA repair,(73) and the ISWI family(74) and the SWI/SNF family(75) are important for cohesion between sister chromatids during DNA replication.

The role of chromatin remodelers in the heart has been reviewed elsewhere,(76) as has been the role of the SWI/SNF complex in cardiac progenitor cells.(77) Brg1, one of two ATPase subunits of the SWI/SNF family (BAF family in vertebrates), is critical for trabeculation, as well as the formation of the right ventricle, outflow tract, and septa during development, in part through its regulation of Bmp10 and Adamts1 expression as shown by mouse knockout studies.(78) Baf180 and Baf60C knockdown are both embryonic lethal resulting in underdeveloped ventricles, with Baf60C serving as a heart-specific Baf subunit that helps target the complex to cardiac enhancers.(79, 80) In disease, Brg1 is up-regulated following pressure overload and acts with HDACs and poly (ADP ribose) polymerase (PARPs) to coordinate the change in myosin heavy chain expression.(81) Inducible knockout of Brg1 before pressure overload attenuates both hypertrophy and fibrosis. Thus, like in development, the Baf complex can respond to environmental cues to regulate gene expression and cardiac morphology.

Members of the other chromatin remodeling complexes have also been analyzed, including by genetic manipulation in animal models. In zebrafish, a gain of function mutation in *reptin*, an ATPase in the INO80 family, results in cardiomyocyte hyperplasia during development after formation of the heart tube, while knockdown of *Pontin*, another ATPase in the complex, also results in hyperplasia.(82) Like with *Brg1*, *Chd7*, a member of the CHD family, has also been shown to interact with Tbx transcription factors, with heterozygous loss of function mutant mice having malformed aortic arches.(83) Taken together, these studies support the observations in the SWI/SNF (Baf) complex for different protein constituents to be able to tailor the function of the complex, for the complexes to modulate transcription factor activity, and for the complexes to participate in large-scale global transcriptional reprogramming.

As with the histone-modifying enzymes, it is important to better understand how these remodeling complexes are targeted to specific nucleosomes in different cell types, developmental stages, and disease states, with some evidence already existing for how histone PTMs facilitate this process.(80, 84) It remains to be determined how the exchangeable protein components of these complexes can modulate their function and localization. Because many of the studies on molecular actions of ATP-dependent remodelers are performed on reconstituted chromatin *in vitro*, it remains unknown what effect the presence of other chromatin binding proteins and modifications have on the actions of these remodelers when they target the genome *in vivo*.

Global Structural Features of Genomes

What the genome looks like *in vivo* is unknown. All of the foregoing chromatin features (Figure 2-2) combine to package the genome into the nucleus in a non-random, dynamic structure (Figure 2-3); approaches to measure this structure directly are a key frontier in chromatin biology. Microscopy studies have provided insight into the regulation of nuclear size and morphology in cardiovascular disease,(85, 86) providing the rationale to investigate in a

locus-specific, yet genome-wide manner, the properties of DNA packing in heart and vascular cells. The development of different Chromosome Conformation Capture (3C) variants during the last decade and the combination of these techniques with the new high-throughput sequencing methods has provided a global view of how chromatin architecture is involved in gene regulation by allowing the formation of active or repressive “microenvironments”.(87) These techniques share the advantage of much higher resolution, to the single base level,(88) in contrast with microscopy. All of the 3C variants have a common workflow of cell fixation, DNA digestion (using restriction enzymes) and DNA re-ligation (under diluted conditions to promote intramolecular ligations between cross-linked fragments) followed by sequencing of the resultant DNA library to examine the frequency of interactions between different loci.

Although this is the early days for application of these techniques to the cardiovascular system in a genome-wide manner, some very interesting studies have emerged in the area of heart failure. Previous studies had demonstrated that the transcription factor Nkx2-5 mediates formation of a chromatin hub in a regulatory element upstream of the atrial natriuretic factor (ANF) gene,(89) a known indicator of hypertrophy. Using a 3C approach to map this chromatin hub, it was subsequently revealed how this hub ensures ANF structural availability and expression during heart development, but is not necessary in the setting of pressure-overload hypertrophy and failure, when ANF is reactivated. New enhancers can be identified by analyzing the gene expression microarray databases generated from previous publications(90) but also by using 3C technology, although this approach has not been applied to the cardiovascular system to our knowledge. ChIP-seq has been applied for this endpoint in the heart,(91)—for the transcriptional co-activator normally present in enhancers, p300, and five cardiac transcription factors essential for initiating or maintaining cardiac gene expression: GATA4, NKX2-5, TBX5, SRF and MEF2A. Genetic variation in several of these transcription factors has recently been reported in humans with congenital heart disease,(92) highlighting the potential translational impact of studies to understand their regulation at the genomic level.

A recent study examining global dynamics of chromatin function involved ChIP-seq-based determination of changes in RNA POL II binding and H3K9-ac occupancy across the genome following pressure overload in mice.(93) Four patterns of change were revealed 4 days after stress, determined by RNA POL II behavior but also showing characteristic features for H3K9ac: promoter clearance of POL II (i.e. a paused polymerase begins transcribing), *de novo* recruitment of POL II across the promoter and gene body, a combination of the previous two, and decreased POL II binding. These modes of RNA POL II behavior were used by the authors to distinguish between transcripts regulated at the transcriptional or post-transcriptional level, although these studies also hold the potential to discriminate functionally different genomic regions based on their occupancy by RNA POL II and/or H3K9ac, regardless of the known annotation of the DNA (e.g. as coding or noncoding).

Several other approaches for examining chromatin functionality and/or nucleosome positioning (which itself can influence gene expression and genomic packaging(94)) have been developed and applied in a genome-wide manner in recent years. These include Formaldehyde-Assisted Isolation of Regulatory Elements, followed by sequencing (FAIRE-seq), which is used to identify regulatory elements of actively transcribed genes.(95) DNase-seq takes advantage of DNase I endonuclease to cut at DNase I sensitive sites, usually accessible DNA not protected by packaging, followed by sequencing, to detect regulatory elements and linker DNA.(96) Unlike DNase I digestion, micrococcal nuclease (MNase) digestion is less restrained (not limited by accessibility) and cuts nucleosome-depleted linker DNA.(97) DNA adenine methyltransferases identification (DamID) is another widely used method, in which a protein of interest is tagged with *E. coli* adenine methyltransferase (not normally expressed or present in eukaryotic cells),(98) to understand DNA-protein interactions. Any regions of the genome that interact with the tagged protein also are methylated, enriched by pull down of antibody recognizing methylated sites or enzymatic digestion (which eliminates potential variability due to antibody affinity), and then mapped and analyzed.(99) DamID has been used to map HP1 to

heterochromatin associated with transposons.(100) This technique has also been used to assess DNA regions that associate with the nuclear lamina and mechanism of laminar recruitment of chromatin. One such study has identified a GAGA sequence motif, common to DNA regions associated with lamina, as a potential sequence-dependent mechanism for recruitment of DNA to the nuclear membrane in mouse fibroblasts.(101) Together, these genome-wide studies of nucleosome positioning are revealing the principles for endogenous chromatin structure from the standpoint of DNA sequence specification.(94, 102)

Future Perspectives

The rapid advance of the study of chromatin (Figure 2-2), with the advent of next generation sequencing technologies, proteomics and computational biology, have driven many investigators to propose models of chromatin states(17, 18) which, in our opinion,(103) should include direct measurements of structural features that can be correlated with function (often transcription, although there are other functions(104) the genome serves in the cell) of super-genic regions of the genome. Such efforts to codify the behavior of chromatin and to understand intermediate structural elements—between the scale of the nucleosome and the entire nucleus—are a critical step to understanding how the same genome encodes multiple cell types.

In the cardiovascular realm, a major challenge is to determine what aspects of the biology of chromatin understood from simpler organisms are relevant in the human heart and vasculature (Table 2-1; Table references:(25, 48, 105-108)). Key aspects of chromatin biology discussed in this review that remain unknown include: What is the role of DNA methylation in higher organisms and is this modification truly stable in differentiated cells?(109) How do the numerous histone variants, not present in lower organisms but known to be expressed in mammals,(25) contribute to genome specification amongst different cell types? What is the logic that governs interaction between chromatin remodelers and transcription factors, and how is this

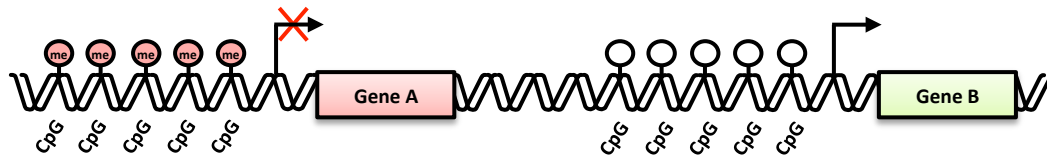
temporally regulated in disease to activate genome-wide changes in transcription? What are the unique structural adaptations that the multi-nucleated, largely heterochromatic, adult cardiomyocyte carries out to package its genome? We believe that investigating new areas of chromatin biology in cardiovascular system will be most effective if executed in two parallel approaches that differ in conception: we should try to “confirm” observations from lower organisms and non-cardiovascular cells in the heart and vasculature; however, a uniquely fruitful approach, we contend, is to explore anew in an unbiased manner (not eschewing what is known in lower organisms) the behavior of chromatin in various stages of cardiovascular development, normal physiology and disease. The latter approach hypothesizes that multicellular eukaryotes have evolved different, albeit not necessarily more complex or more effective, mechanisms to solve the problem of genome packaging and regulation.

Chromatin States

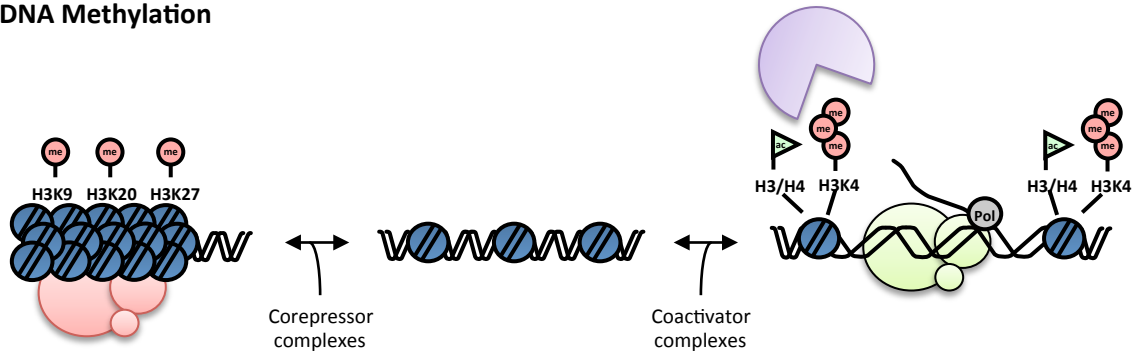


Hetero- and euchromatin

Chromatin Remodelers



DNA Methylation



Histone Modifications and Chromatin Remodelers

Non-coding RNAs

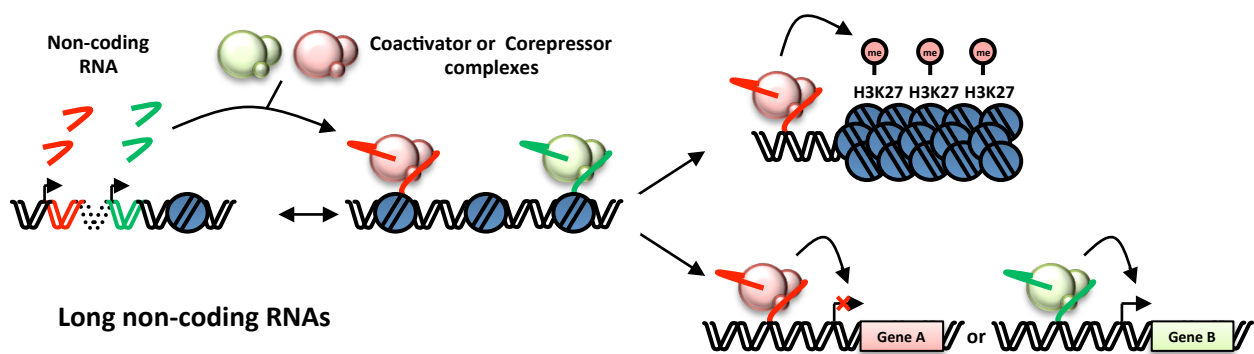


Figure 2-1. Mechanisms of chromatin remodeling. (Top Panel) DNA can be packaged as heterochromatin (tightly packaged and generally inaccessible to proteins such as transcriptional machinery) or as euchromatin (accessible to transcription). (Middle Panel) Chromatin remodeling proteins can alter DNA methylation or histone PTMs, the combination of which can be recognized by “reader proteins” (purple) to confer chromatin state. (Bottom Panel) Non coding RNAs can interact with these chromatin remodelers to target them to specific loci, leading alternatively to gene activation or silencing.

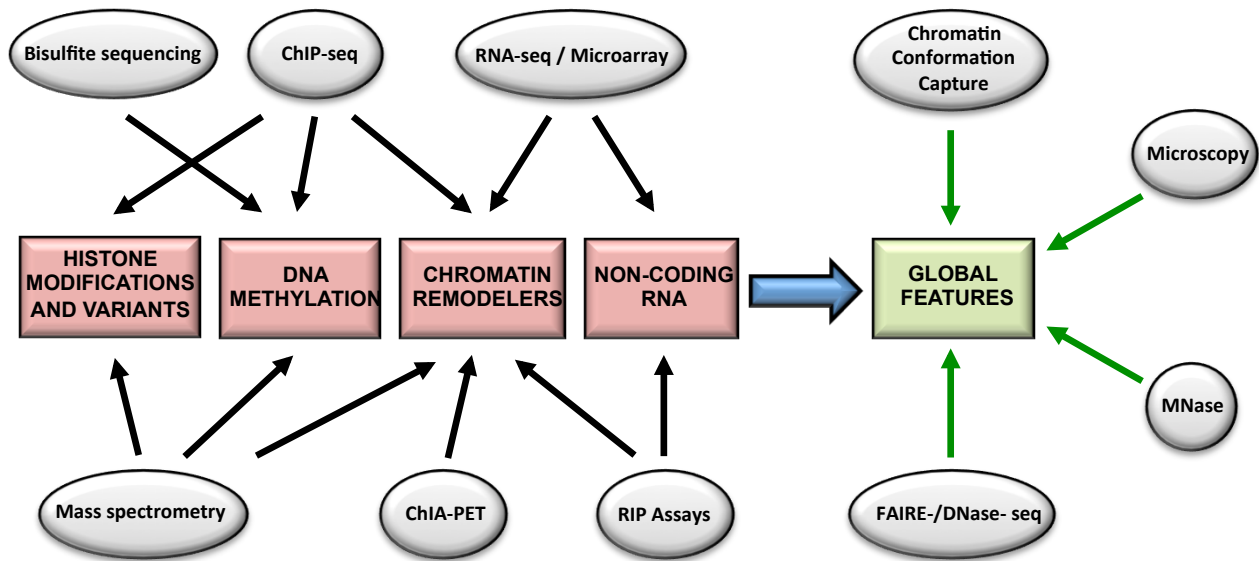


Figure 2-2. Methodologies and applications. Various techniques currently utilized, alone or in combination, to analyze chromatin structure and regulation.

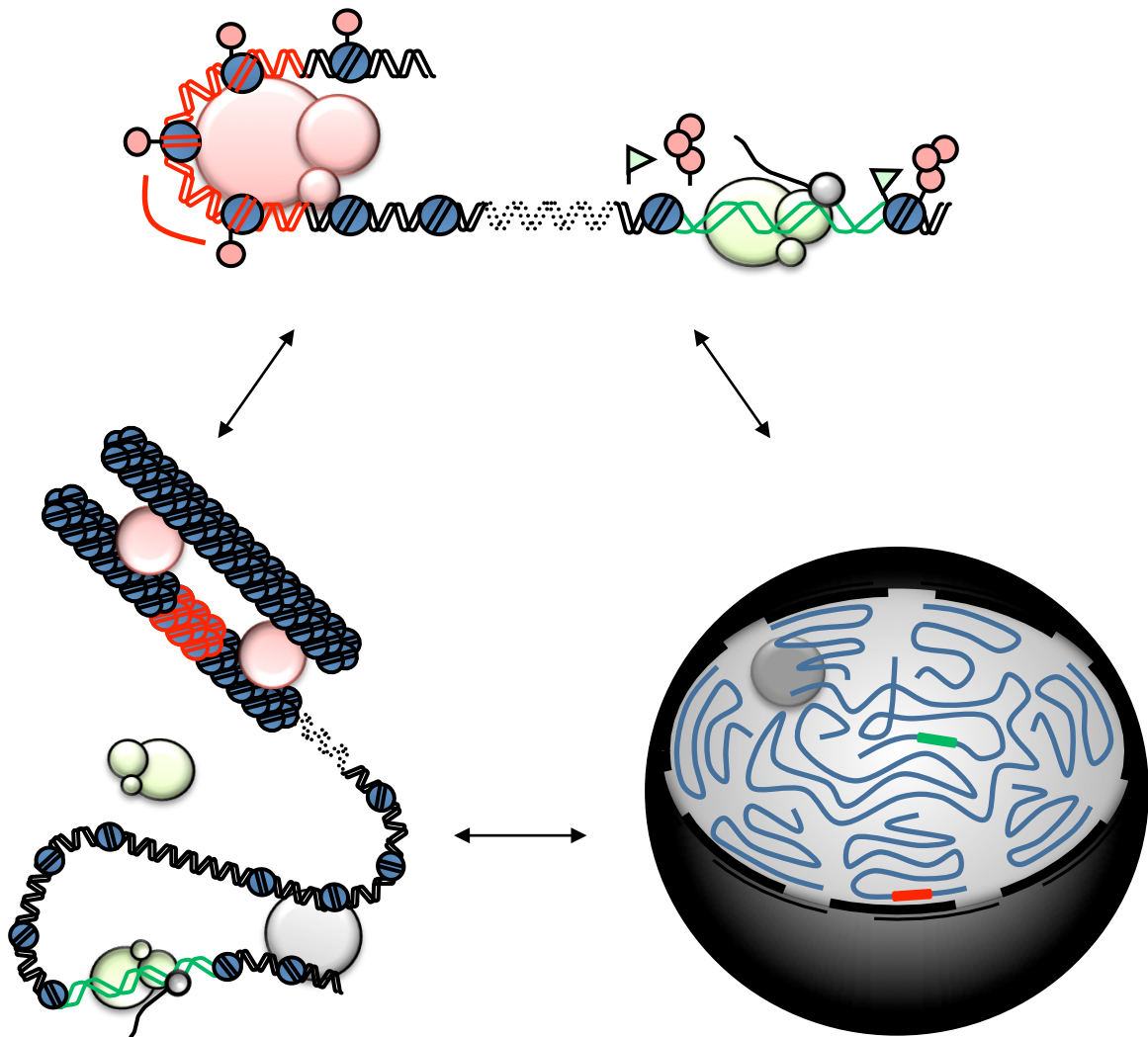


Figure 2-3. Multiple levels of structural organization regulate gene expression in an interphase nucleus. (Top) An actively transcribed (green) and repressed (red) segment of DNA occur on distinct regions of the same chromosome, directly regulated by chromatin remodelers and noncoding RNA. (Bottom Left) The higher-order chromatin structure surrounding these segments endows another level of regulation, with the cumulative effect of local modifications of nearby segments and neighboring chromosomes combining to create distinct repressive and activating environments. (Bottom Right) The position of the segments in relation to the global features of the nucleus imparts yet another level of regulation. All three levels regulate each other interdependently.

	Studied in CV Development	Studied in CV Disease	Well-studied Players in CV System	Comments
Histone Modifying Enzymes	Extensive	Extensive	HDAC: Class I,II HAT: p300, CBP HMT: PRC2,, MLL HDM: Jumonji domain-containing proteins	Extensive KD/OE and pharmacological studies (105) Limited understanding for genomic distribution of specific modifications and target loci
Histone Variants	Limited	Moderate	H2A.z	Variants expressed in the heart identified (25) but only a few understood mechanistically (26)
Histone Post-translational Modifications	Moderate	Moderate	Acetylation, Methylation	Global positioning analyzed by ChIP-seq (20, 21, 23) Proteomic analyses of PTM abundance, PTM combinations, and PTM interactions with “reader” proteins are still necessary
Non-nucleosomal Structural Proteins	Limited	Moderate	HMGA, HMGB	Some structural proteins have been well-studied in other systems, but their functions have yet to be addressed in the heart (25, 27)
ATP-dependent Chromatin Remodelers	Moderate	Moderate	Brg1, BAF complex	Members of major families studied in the heart (79, 80), though disease studies in this field are still incomplete
DNA Methylation	Moderate	Moderate	meC, Dnmt	Some loci specific studies (106) Few global studies (48)
Noncoding RNA	Extensive	Extensive	miRNAs	Specific miRNAs studied extensively (53, 54) for transcriptional regulation, but the role of noncoding RNA for regulating chromatin structure is needed (60, 61, 107)
Nucleosome Positioning	Limited	Limited	--	Emerging field in chromatin biology not yet explored in the heart
Chromatin Conformation	Limited	Limited	Enhancers and chromatin loops	Studies in the heart of this type are recent and include measuring long-range interactions (89) and nuclear organization by microscopy (108)

Table 2-1. Current state of understanding of various chromatin structural regulators in the cardiovascular system.

Chapter 2: References

1. Richmond TJ, Davey CA. The structure of DNA in the nucleosome core. *Nature*. 2003;423(6936):145-50.
2. Nair AP, Timoh T, Fuster V. Contemporary medical management of systolic heart failure. *Circulation journal : official journal of the Japanese Circulation Society*. 2012;76(2):268-77.
3. Go AS, Mozaffarian D, Roger VL, Benjamin EJ, Berry JD, Borden WB, et al. Heart Disease and Stroke Statistics--2013 Update: A Report From the American Heart Association. *Circulation*. 2013;127(1):e6-e245.
4. Young NL, DiMaggio PA, Plazas-Mayorca MD, Baliban RC, Floudas CA, Garcia BA. High throughput characterization of combinatorial histone codes. *Mol Cell Proteomics*. 2009;8(10):2266-84.
5. Siuti N, Kelleher NL. Decoding protein modifications using top-down mass spectrometry. *Nat Methods*. 2007;4(10):817-21.
6. Garcia BA, Pesavento JJ, Mizzen CA, Kelleher NL. Pervasive combinatorial modification of histone H3 in human cells. *Nat Methods*. 2007;4(6):487-9.
7. Nakayama J, Rice JC, Strahl BD, Allis CD, Grewal SI. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science*. 2001;292(5514):1110-3.
8. Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, et al. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science*. 2002;298(5595):1039-43.
9. Strahl BD, Ohba R, Cook RG, Allis CD. Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in *Tetrahymena*. *Proc Natl Acad Sci U S A*. 1999;96(26):14967-72.
10. Koch CM, Andrews RM, Flicek P, Dillon SC, Karaoz U, Clelland GK, et al. The landscape of histone modifications across 1% of the human genome in five human cell lines. *Genome research*. 2007;17(6):691-707.

11. Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America*. 2010;107(50):21931-6.
12. Kolasinska-Zwierz P, Down T, Latorre I, Liu T, Liu XS, Ahringer J. Differential chromatin marking of introns and expressed exons by H3K36me3. *Nature genetics*. 2009;41(3):376-81.
13. Struhl K. Histone acetylation and transcriptional regulatory mechanisms. *Genes & development*. 1998;12(5):599-606.
14. Talbert PB, Henikoff S. Histone variants--ancient wrap artists of the epigenome. *Nature reviews Molecular cell biology*. 2010;11(4):264-75.
15. Ramakrishnan V. Histone H1 and chromatin higher-order structure. *Critical reviews in eukaryotic gene expression*. 1997;7(3):215-30.
16. Strahl BD, Allis CD. The language of covalent histone modifications. *Nature*. 2000;403(6765):41-5.
17. Ernst J, Kheradpour P, Mikkelson TS, Shores N, Ward LD, Epstein CB, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*. 2011;473(7345):43-9.
18. Fillion GJ, van Bommel JG, Braunschweig U, Talhout W, Kind J, Ward LD, et al. Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell*. 2010;143(2):212-24.
19. Haberland M, Montgomery RL, Olson EN. The many roles of histone deacetylases in development and physiology: implications for disease and therapy. *Nat Rev Genet*. 2009;10(1):32-42.
20. Paige SL, Thomas S, Stoick-Cooper CL, Wang H, Maves L, Sandstrom R, et al. A temporal chromatin signature in human embryonic stem cells identifies regulators of cardiac development. *Cell*. 2012;151(1):221-32.

21. Wamstad JA, Alexander JM, Truty RM, Shrikumar A, Li F, Eilertson KE, et al. Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. *Cell*. 2012;151(1):206-20.
22. Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, et al. A map of the cis-regulatory sequences in the mouse genome. *Nature*. 2012;488(7409):116-20.
23. Kaneda R, Takada S, Yamashita Y, Choi YL, Nonaka-Sarukawa M, Soda M, et al. Genome-wide histone methylation profile for heart failure. *Genes Cells*. 2009;14(1):69-77.
24. Vermeulen M, Eberl HC, Matarese F, Marks H, Denissov S, Butter F, et al. Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell*. 2010;142(6):967-80.
25. Franklin S, Chen H, Mitchell-Jordan S, Ren S, Wang Y, Vondriska TM. Quantitative analysis of the chromatin proteome in disease reveals remodeling principles and identifies high mobility group protein B2 as a regulator of hypertrophic growth. *Mol Cell Proteomics*. 2012;11(6):M111 014258.
26. Chen IY, Lypowy J, Pain J, Sayed D, Grinberg S, Alcendor RR, et al. Histone H2A.z is essential for cardiac myocyte hypertrophy but opposed by silent information regulator 2alpha. *The Journal of biological chemistry*. 2006;281(28):19369-77.
27. Fedele M, Fidanza V, Battista S, Pentimalli F, Klein-Szanto AJ, Visone R, et al. Haploinsufficiency of the Hmga1 gene causes cardiac hypertrophy and myelolymphoproliferative disorders in mice. *Cancer Res*. 2006;66(5):2536-43.
28. Cooper DN, Taggart MH, Bird AP. Unmethylated domains in vertebrate DNA. *Nucleic acids research*. 1983;11(3):647-58.
29. Bird A, Taggart M, Frommer M, Miller OJ, Macleod D. A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell*. 1985;40(1):91-9.
30. Gardiner-Garden M, Frommer M. CpG islands in vertebrate genomes. *Journal of molecular biology*. 1987;196(2):261-82.

31. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, et al. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nature genetics*. 2009;41(2):178-86.
32. Nikitina T, Shi X, Ghosh RP, Horowitz-Scherer RA, Hansen JC, Woodcock CL. Multiple modes of interaction between the methylated DNA binding protein MeCP2 and chromatin. *Molecular and cellular biology*. 2007;27(3):864-77.
33. Newell-Price J, Clark AJ, King P. DNA methylation and silencing of gene expression. *Trends Endocrinol Metab*. 2000;11(4):142-8.
34. Xie L, Weichel B, Ohm JE, Zhang K. An integrative analysis of DNA methylation and RNA-Seq data for human heart, kidney and liver. *BMC Syst Biol*. 2011;5 Suppl 3:S4.
35. Yoder JA, Soman NS, Verdine GL, Bestor TH. DNA (cytosine-5)-methyltransferases in mouse cells and tissues. Studies with a mechanism-based probe. *Journal of molecular biology*. 1997;270(3):385-95.
36. Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*. 1999;99(3):247-57.
37. Bestor TH. The DNA methyltransferases of mammals. *Human molecular genetics*. 2000;9(16):2395-402.
38. Johnson R, Richter N, Bogu GK, Bhinge A, Teng SW, Choo SH, et al. A genome-wide screen for genetic variants that modify the recruitment of REST to its target genes. *PLoS genetics*. 2012;8(4):e1002624.
39. Irier HA, Jin P. Dynamics of DNA methylation in aging and Alzheimer's disease. *DNA Cell Biol*. 2012;31 Suppl 1:S42-8.
40. Chodavarapu RK, Feng S, Bernatavichute YV, Chen PY, Stroud H, Yu Y, et al. Relationship between nucleosome positioning and DNA methylation. *Nature*. 2010;466(7304):388-92.

41. Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, et al. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature*. 2008;452(7184):215-9.
42. Mohn F, Weber M, Schubeler D, Roloff TC. Methylated DNA immunoprecipitation (MeDIP). *Methods in molecular biology*. 2009;507:55-64.
43. Zhu C, Yu ZB, Chen XH, Ji CB, Qian LM, Han SP. DNA hypermethylation of the NOX5 gene in fetal ventricular septal defect. *Exp Ther Med*. 2011;2(5):1011-5.
44. Movassagh M, Choy MK, Goddard M, Bennett MR, Down TA, Foo RS. Differential DNA methylation correlates with differential expression of angiogenic factors in human heart failure. *PloS one*. 2010;5(1):e8564.
45. Haider S, Cordeddu L, Robinson E, Movassagh M, Siggins L, Vujic A, et al. The landscape of DNA repeat elements in human heart failure. *Genome biology*. 2012;13(10):R90.
46. Haddad R, Kasneci A, Mephram K, Sebag IA, Chalifour LE. Gestational exposure to diethylstilbestrol alters cardiac structure/function, protein expression and DNA methylation in adult male mice progeny. *Toxicol Appl Pharmacol*. 2013;266(1):27-37.
47. Metes-Kosik N, Luptak I, Dibello PM, Handy DE, Tang SS, Zhi H, et al. Both selenium deficiency and modest selenium supplementation lead to myocardial fibrosis in mice via effects on redox-methylation balance. *Mol Nutr Food Res*. 2012;56(12):1812-24.
48. Movassagh M, Choy MK, Knowles DA, Cordeddu L, Haider S, Down T, et al. Distinct epigenomic features in end-stage failing human hearts. *Circulation*. 2011;124(22):2411-22.
49. Antonitsis P, Ioannidou-Papagiannaki E, Kaidoglou A, Papakonstantinou C. In vitro cardiomyogenic differentiation of adult human bone marrow mesenchymal stem cells. The role of 5-azacytidine. *Interact Cardiovasc Thorac Surg*. 2007;6(5):593-7.
50. Ye NS, Chen J, Luo GA, Zhang RL, Zhao YF, Wang YM. Proteomic profiling of rat bone marrow mesenchymal stem cells induced by 5-azacytidine. *Stem Cells Dev*. 2006;15(5):665-76.

51. Liang P, Song F, Ghosh S, Morien E, Qin M, Mahmood S, et al. Genome-wide survey reveals dynamic widespread tissue-specific changes in DNA methylation during development. *BMC genomics*. 2011;12(1):231.
52. Mattick JS. Non-coding RNAs: the architects of eukaryotic complexity. *EMBO reports*. 2001;2(11):986-91.
53. Dorn GW, 2nd. Decoding the cardiac message: the 2011 Thomas W. Smith Memorial Lecture. *Circ Res*. 2012;110(5):755-63.
54. van Rooij E, Olson EN. MicroRNA therapeutics for cardiovascular disease: opportunities and obstacles. *Nature reviews Drug discovery*. 2012;11(11):860-72.
55. Kawashima T, Shioi T. MicroRNA, emerging role as a biomarker of heart failure. *Circulation journal : official journal of the Japanese Circulation Society*. 2011;75(2):268-9.
56. Song HK, Hong SE, Kim T, Kim do H. Deep RNA sequencing reveals novel cardiac transcriptomic signatures for physiological and pathological hypertrophy. *PLoS One*. 2012;7(4):e35552.
57. Lee JH, Gao C, Peng G, Greer C, Ren S, Wang Y, et al. Analysis of transcriptome complexity through RNA sequencing in normal and failing murine hearts. *Circ Res*. 2011;109(12):1332-41.
58. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes & development*. 2011;25(18):1915-27.
59. Lee JT. Epigenetic regulation by long noncoding RNAs. *Science*. 2012;338(6113):1435-9.
60. Grote P, Wittler L, Hendrix D, Koch F, Wahrisch S, Beisaw A, et al. The Tissue-Specific lncRNA Fendrr Is an Essential Regulator of Heart and Body Wall Development in the Mouse. *Developmental cell*. 2013;24(2):206-14.

61. Klattenhoff CA, Scheuermann JC, Surface LE, Bradley RK, Fields PA, Steinhauser ML, et al. Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. *Cell*. 2013;152(3):570-83.
62. Saha A, Wittmeyer J, Cairns BR. Chromatin remodelling: the industrial revolution of DNA around histones. *Nature reviews Molecular cell biology*. 2006;7(6):437-47.
63. Fazio TG, Tsukiyama T. Chromatin remodeling in vivo: evidence for a nucleosome sliding mechanism. *Mol Cell*. 2003;12(5):1333-40.
64. Boeger H, Griesenbeck J, Strattan JS, Kornberg RD. Removal of promoter nucleosomes by disassembly rather than sliding in vivo. *Mol Cell*. 2004;14(5):667-73.
65. Bruno M, Flaus A, Stockdale C, Rencurel C, Ferreira H, Owen-Hughes T. Histone H2A/H2B dimer exchange by ATP-dependent chromatin remodeling activities. *Mol Cell*. 2003;12(6):1599-606.
66. Mizuguchi G, Shen X, Landry J, Wu WH, Sen S, Wu C. ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science*. 2004;303(5656):343-8.
67. Cairns BR. Chromatin remodeling: insights and intrigue from single-molecule studies. *Nat Struct Mol Biol*. 2007;14(11):989-96.
68. Armstrong JA, Papoulas O, Daubresse G, Sperling AS, Lis JT, Scott MP, et al. The *Drosophila* BRM complex facilitates global transcription by RNA polymerase II. *Embo J*. 2002;21(19):5245-54.
69. Batsche E, Yaniv M, Muchardt C. The human SWI/SNF subunit Brm is a regulator of alternative splicing. *Nat Struct Mol Biol*. 2006;13(1):22-9.
70. Corey LL, Weirich CS, Benjamin IJ, Kingston RE. Localized recruitment of a chromatin-remodeling activity by an activator in vivo drives transcriptional elongation. *Genes Dev*. 2003;17(11):1392-401.

71. Morillon A, Karabetsou N, O'Sullivan J, Kent N, Proudfoot N, Mellor J. Isw1 chromatin remodeling ATPase coordinates transcription elongation and termination by RNA polymerase II. *Cell*. 2003;115(4):425-35.
72. Wade PA, Geggion A, Jones PL, Ballestar E, Aubry F, Wolffe AP. Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. *Nat Genet*. 1999;23(1):62-6.
73. Kusch T, Florens L, Macdonald WH, Swanson SK, Glaser RL, Yates JR, 3rd, et al. Acetylation by Tip60 is required for selective histone variant exchange at DNA lesions. *Science*. 2004;306(5704):2084-7.
74. Hakimi MA, Bochar DA, Schmiesing JA, Dong Y, Barak OG, Speicher DW, et al. A chromatin remodelling complex that loads cohesin onto human chromosomes. *Nature*. 2002;418(6901):994-8.
75. Huang J, Hsu JM, Laurent BC. The RSC nucleosome-remodeling complex is required for Cohesin's association with chromosome arms. *Mol Cell*. 2004;13(5):739-50.
76. Han P, Hang CT, Yang J, Chang CP. Chromatin remodeling in cardiovascular development and physiology. *Circ Res*. 2011;108(3):378-96.
77. Lei I, Gao X, Sham MH, Wang Z. SWI/SNF protein component BAF250a regulates cardiac progenitor cell differentiation by modulating chromatin accessibility during second heart field development. *J Biol Chem*. 2012;287(29):24255-62.
78. Stankunas K, Hang CT, Tsun ZY, Chen H, Lee NV, Wu JI, et al. Endocardial Brg1 represses ADAMTS1 to maintain the microenvironment for myocardial morphogenesis. *Developmental cell*. 2008;14(2):298-311.
79. Lickert H, Takeuchi JK, Von Both I, Walls JR, McAuliffe F, Adamson SL, et al. Baf60c is essential for function of BAF chromatin remodelling complexes in heart development. *Nature*. 2004;432(7013):107-12.

80. Wang Z, Zhai W, Richardson JA, Olson EN, Meneses JJ, Firpo MT, et al. Polybromo protein BAF180 functions in mammalian cardiac chamber maturation. *Genes Dev.* 2004;18(24):3106-16.
81. Hang CT, Yang J, Han P, Cheng HL, Shang C, Ashley E, et al. Chromatin regulation by Brg1 underlies heart muscle development and disease. *Nature.* 2010;466(7302):62-7.
82. Rottbauer W, Saurin AJ, Lickert H, Shen X, Burns CG, Wo ZG, et al. Reptin and pontin antagonistically regulate heart growth in zebrafish embryos. *Cell.* 2002;111(5):661-72.
83. Hurd EA, Capers PL, Blauwkamp MN, Adams ME, Raphael Y, Poucher HK, et al. Loss of Chd7 function in gene-trapped reporter mice is embryonic lethal and associated with severe defects in multiple developing tissues. *Mammalian genome : official journal of the International Mammalian Genome Society.* 2007;18(2):94-104.
84. Lange M, Kaynak B, Forster UB, Tonjes M, Fischer JJ, Grimm C, et al. Regulation of muscle development by DPF3, a novel histone acetylation and methylation reader of the BAF chromatin remodeling complex. *Genes Dev.* 2008;22(17):2370-84.
85. Koda M, Takemura G, Okada H, Kanoh M, Maruyama R, Esaki M, et al. Nuclear hypertrophy reflects increased biosynthetic activities in myocytes of human hypertrophic hearts. *Circulation journal : official journal of the Japanese Circulation Society.* 2006;70(6):710-8.
86. Nikolova V, Leimena C, McMahon AC, Tan JC, Chandar S, Jogia D, et al. Defects in nuclear structure and function promote dilated cardiomyopathy in lamin A/C-deficient mice. *J Clin Invest.* 2004;113(3):357-69.
87. van Steensel B, Dekker J. Genomics tools for unraveling chromosome architecture. *Nature biotechnology.* 2010;28(10):1089-95.
88. Simonis M, Kooren J, de Laat W. An evaluation of 3C-based methods to capture DNA interactions. *Nature methods.* 2007;4(11):895-901.

89. Warren SA, Terada R, Briggs LE, Cole-Jeffrey CT, Chien WM, Seki T, et al. Differential role of Nkx2-5 in activation of the atrial natriuretic factor gene in the developing versus failing heart. *Molecular and cellular biology*. 2011;31(22):4633-45.
90. Narlikar L, Sakabe NJ, Blanski AA, Arimura FE, Westlund JM, Nobrega MA, et al. Genome-wide discovery of human heart enhancers. *Genome research*. 2010;20(3):381-92.
91. He A, Kong SW, Ma Q, Pu WT. Co-occupancy by multiple cardiac transcription factors identifies transcriptional enhancers active in heart. *Proc Natl Acad Sci U S A*. 2011;108(14):5632-7.
92. Kodo K, Nishizawa T, Furutani M, Arai S, Ishihara K, Oda M, et al. Genetic analysis of essential cardiac transcription factors in 256 patients with non-syndromic congenital heart defects. *Circulation journal : official journal of the Japanese Circulation Society*. 2012;76(7):1703-11.
93. Sayed D, He M, Yang Z, Lin L, Abdellatif M. Transcriptional Regulation Patterns Revealed by High Resolution Chromatin Immunoprecipitation during Cardiac Hypertrophy. *J Biol Chem*. 2013;288(4):2546-58.
94. Segal E, Widom J. What controls nucleosome positions? *Trends Genet*. 2009;25(8):335-43.
95. Giresi PG, Kim J, McDaniel RM, Iyer VR, Lieb JD. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome research*. 2007;17(6):877-85.
96. Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, et al. High-resolution mapping and characterization of open chromatin across the genome. *Cell*. 2008;132(2):311-22.
97. Cui K, Zhao K. Genome-wide approaches to determining nucleosome occupancy in metazoans using MNase-Seq. *Methods Mol Biol*. 2012;833:413-9.
98. van Steensel B, Henikoff S. Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. *Nature biotechnology*. 2000;18(4):424-8.

99. Xiao R, Moore DD. DamIP: using mutant DNA adenine methyltransferase to study DNA-protein interactions in vivo. *Current protocols in molecular biology* / edited by Frederick M Ausubel [et al]. 2011;Chapter 21:Unit21
100. de Wit E, Greil F, van Steensel B. Genome-wide HP1 binding in *Drosophila*: developmental plasticity and genomic targeting signals. *Genome research*. 2005;15(9):1265-73.
101. Zullo JM, Demarco IA, Pique-Regi R, Gaffney DJ, Epstein CB, Spooner CJ, et al. DNA sequence-dependent compartmentalization and silencing of chromatin at the nuclear lamina. *Cell*. 2012;149(7):1474-87.
102. Zhang Z, Wippo CJ, Wal M, Ward E, Korber P, Pugh BF. A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome. *Science*. 2011;332(6032):977-80.
103. Chen H, Monte E, Parvatiyar MS, Rosa-Garrido M, Franklin S, Vondriska TM. Structural considerations for chromatin state models with transcription as a functional readout. *FEBS letters*. 2012;586(20):3548-54.
104. Solovei I, Kreysing M, Lanctot C, Kosem S, Peichl L, Cremer T, et al. Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution. *Cell*. 2009;137(2):356-68.
105. Ohtani K, Dimmeler S. Epigenetic regulation of cardiovascular differentiation. *Cardiovasc Res*. 2011;90(3):404-12.
106. Udali S, Guarini P, Moruzzi S, Choi SW, Friso S. Cardiovascular epigenetics: From DNA methylation to microRNAs. *Molecular aspects of medicine*. 2012.
107. Schonrock N, Harvey RP, Mattick JS. Long noncoding RNAs in cardiac development and pathophysiology. *Circ Res*. 2012;111(10):1349-62.
108. Mitchell-Jordan S, Chen H, Franklin S, Stefani E, Bentolila LA, Vondriska TM. Features of endogenous cardiomyocyte chromatin revealed by super-resolution STED microscopy. *J Mol Cell Cardiol*. 2012;53(4):552-8.

109. Guo JU, Ma DK, Mo H, Ball MP, Jang MH, Bonaguidi MA, et al. Neuronal activity modifies the DNA methylation landscape in the adult brain. *Nature neuroscience*. 2011;14(10):1345-51.

Chapter 3: Systems proteomics of cardiac chromatin identifies nucleolin as a regulator of growth and cellular plasticity in cardiomyocytes

Emma Monte*, Kevin Mouillesseaux*, Haodong Chen, Todd Kimball, Shuxun Ren, Yibin Wang, Jau-Nian Chen, Thomas M. Vondriska, Sarah Franklin

**Indicates equal contribution*

[This research was originally published in American Journal of Physiology by Monte et al. Systems proteomics of cardiac chromatin identifies Nucleolin as a regulator of growth and cellular plasticity in cardiomyocytes. *Am J Physiol Heart Circ Physiol*. 2013; 305(11): H1624-38. © American Physiological Society.]

Abstract

Myocyte hypertrophy antecedent to heart failure involves changes in global gene expression, although the preceding mechanisms to coordinate DNA accessibility on a genomic scale are unknown. Chromatin-associated proteins can alter chromatin structure by changing their association with DNA, thereby altering the gene expression profile. Little is known about the global changes in chromatin sub-proteomes that accompany heart failure, and the mechanisms by which these proteins alter chromatin structure. The present study tests the fundamental hypothesis that cardiac growth and plasticity in the setting of disease recapitulates conserved developmental chromatin remodeling events. We used quantitative proteomics to identify chromatin-associated proteins extracted via detergent and to quantify changes in their abundance during disease. Our study identified 321 proteins in this sub-proteome, demonstrating it to have modest conservation (37%) with that revealed using strong acid. Of these proteins, 176 exhibited altered expression during cardiac hypertrophy and failure; we conducted extensive functional characterization of one of these proteins, Nucleolin. Morpholino-based knockdown of *nucleolin* nearly abolished protein expression but surprisingly had little impact on gross morphological development. However, hearts of fish lacking Nucleolin

displayed severe developmental impairment, abnormal chamber patterning and functional deficits, ostensibly due to defects in cardiac looping and myocyte differentiation. The mechanisms underlying these defects involve perturbed bone morphogenetic protein 4 expression, decreased rRNA transcription and a shift to more heterochromatic chromatin. This study reports the quantitative analysis of a new chromatin sub-proteome in the normal and diseased mouse heart. Validation studies in the complementary model system of zebrafish examine the role of Nucleolin to orchestrate genomic reprogramming events shared between development and disease.

Introduction

During the development of hypertrophy and failure in the mammalian heart, adult cardiomyocytes undergo extensive transcriptional reprogramming, becoming more plastic in response to stress. Myocytes increase their mass and adopt a gene expression profile, as well as phenotypic features, normally associated with fetal myocardium.(36) These gross changes in transcription may initially promote functional stability in the heart by allowing cells to compensate for an increased workload, however studies from humans and animal models have consistently shown that hypertrophic remodeling leads to cardiac dysfunction and failure. While the transcription factors responsible for cardiac hypertrophy have been extensively studied,(10, 35) the global chromatin remodeling events are less well understood. In the past few years, the field has gained a new appreciation for the global changes in chromatin features responsible for cardiovascular phenotypes in development and disease,(3, 31) including from studies on chromatin remodeling complexes.(20) However, large changes in gene expression must be preceded by coordinated alterations in chromatin structure to allow or deny accessibility of transcriptional machinery to specific genomic regions. This global process has local

ramifications, in which the functional unit of chromatin—the nucleosome (an obligate DNA-protein complex of two copies each of four histone proteins)—alters its accessibility between structurally compact and transcriptionally inactive heterochromatin and structurally loose and transcriptionally active euchromatin. Therefore, knowledge of the constitutive protein occupants of chromatin, as well as the remodelers and other structural proteins that alter their association with the genome during development and disease, is an important step towards understanding how global changes in chromatin accessibility are coordinated.

To address this question, we sought to quantify a sub-proteome of chromatin during the development of hypertrophy and failure following pressure overload in the mouse. Of the 321 (698 total, 321 by 2 or more peptides) proteins measured, 176 (338 total, 176 by 2 or more peptides) changed with heart disease in the mouse. We then employed a complementary model system (zebrafish development) to investigate the role of one chromatin-bound protein, Nucleolin.

Nucleolin is commonly characterized as a nucleolar protein, although it has been observed throughout the nucleus, in cytoplasm and on plasma membrane. Nucleolin is highly expressed in proliferating cells and has been proposed to regulate cell growth and apoptosis in non-cardiac systems. It is also involved in several aspects of gene expression including chromatin remodeling, RNA transcription, rRNA processing, nucleo-cytoplasmic transport and mRNA stabilization.(11, 17-19, 22) Much research has focused on Nucleolin's ability to regulate ribosome biogenesis, because protein synthesis is a key element for controlling growth, making this process a key readout for endogenous Nucleolin function. Nucleolin has also been shown to modulate the expression of p53 both negatively and positively, depending on the cell type. Robust cardiac Nucleolin expression has been observed during murine embryogenesis, followed by a precipitous decline during the first week of life, concomitant with decreased myocyte growth and cell cycle exit, with relatively low levels of Nucleolin detected in the normal adult heart.(2) However, the functional role of Nucleolin in cardiac phenotype during

development and disease is unknown in any species. Our findings implicate Nucleolin in cardiomyocyte differentiation and heart formation, providing evidence for this protein in regulation of chromatin structure and ribosome biogenesis.

Materials & Methods

Mouse model of cardiac hypertrophy and failure and echocardiographic determination of cardiac function

All protocols involving animals conform to the NIH Guide for the Care and Use of Laboratory Animals and were approved by the UCLA Animal Research Committee. Adult male balb/c mice aged 8-12 weeks (Charles River Laboratories) were used for this study. The murine model of transverse aortic banding (TAC)-induced cardiac hypertrophy was performed as described previously.(13, 32) SHAM operated mice underwent the same procedure without placement of aortic clamp. Echocardiography (ECHO) was used to determine cardiac parameters in live mice as described,(13, 32) including the following indices: left ventricular (LV) size (end-diastolic and end-systolic dimension), wall thickness (intra-ventricular septum and posterior wall thickness), ventricular mass, ventricular function (ejection fraction [EF]), and blood flow. All mice underwent ECHO analyses once before TAC or SHAM surgery, once a day after and then once every 5 days for the duration of the study. Animals were considered hypertrophic when their LV mass was greater than the mean of the control animal with no depression of LV function as measured by EF; animals were considered in heart failure when the EF was significantly decreased below the mean of the control animals. These phenotypes, hypertrophy and failure, corresponded to ~2 and ~4 weeks after TAC surgery, although animals were sacrificed based on echo parameters and not solely on time.

Nuclear isolation and fractionation

All buffers used for cell isolation or fractionation in this study contained the following protease, phosphatase and deacetylase inhibitors, respectively: 0.1mM phenylmethanesulfonylflouride, protease inhibitor cocktail pellet (Roche), 0.2mM sodium orthovanadate, 0.1 mM sodium flouride and 10mM sodium butyrate. Cardiac nuclei were isolated as previously described.(14) We consistently achieve $\geq 80\%$ purity of nuclei with this method as observed by electron microscopy and western blotting analysis. Following isolation of nuclei, further fractionation was carried out to separate nucleoplasm from chromatin using detergent extraction. Briefly, isolated nuclei were resuspended in buffer (20mM HEPES [pH 7.6], 7.5 mM $MgCl_2$, 30 mM NaCl, 1M Urea, 1% NP-40) to solubilize the nuclear membrane and extract soluble proteins in the nucleoplasm. After solubilization samples were centrifuged at 13,000g for 10 min to pellet the insoluble chromatin and remove the nucleoplasm fraction. The chromatin pellet was washed with PBS, solubilized in 50mM Tris (pH 8), 10 mM EDTA, 1% SDS, sonicated to shear the DNA and centrifuged at 13,000g to extract proteins (referred to as detergent-extracted fraction). This method is distinct from the low pH method of protein extraction (acid-extraction) and as demonstrated throughout the current manuscript (see Figure 3-1), reveals a biologically distinct sub-proteome of molecules.

Enzyme digestion

Detergent-extracted proteins isolated from chromatin were separated by SDS-PAGE. Each gel lane was cut into 25 slices (~2mm each) for protein identification by mass spec. Gel plugs were dehydrated in acetonitrile and dried in a speedvac. Samples were reduced and alkylated with 10 mM dithiotreitol and 10mM tris(2-carboxyethyl)phosphine solution in 50 mM ammonium bicarbonate (30 min at 56°C) and 100 mM iodoacetamide (25 min in dark), respectively. Gels were washed with 50 mM ammonium bicarbonate, dehydrated with acetonitrile, and dried in a speedvac. Gel pieces were then swollen in digestion buffer containing 50mM ammonium bicarbonate, and 20.0 ng/ μL of trypsin (37°C, overnight). Peptides

were extracted with 0.1% formic acid in 50% acetonitrile solution, dried down and resuspended in 0.1% formic acid, 2% acetonitrile. For each condition (basal, hypertrophy and failure) three biological (*de novo* preparation of samples from different animals) and two technical (multiple LC/MS/MS experiments on the same preparation) replicates were analyzed by mass spectrometry.

Mass spectrometry analyses and database searching

Extracted peptides were analyzed by nano-flow LC/MS/MS on a Thermo Orbitrap with dedicated Eksigent nanopump using a reversed phase column (75 μm i.d. 10 cm, BioBasic C18 5 μm particle size, New Objective) and a flow rate of 200 nL/min. For peptide separation a linear gradient was utilized from 95% Buffer A (0.1% formic acid, 2% acetonitrile) and 5% Buffer B (0.1% formic acid, 20% water in acetonitrile) to 50% Buffer A and 50% Buffer B over 60 minutes. Spectra were acquired in data-dependent mode with dynamic exclusion where the instrument selects the top six most abundant ions in the parent spectra for fragmentation. Data were searched against the Uniprot database (version 03.2011) using the SEQUEST algorithm in the BioWorks software program version 3.3.1 SP1 and through the Rosetta Elucidator software (Microsoft). False discovery rate, which was calculated on several independent datasets within this study by reverse database searching, ranged from 1.4-1.7%. All spectra used for identification had $\Delta\text{CN} > 0.1$, consensus score ≥ 20 and met the following Xcorr criteria: > 3 (+2), > 4 (+3), and > 5 (+4). Searches required full tryptic cleavage, ≤ 3 missed cleavages and were performed with the differential modifications of carbamidomethylation on cysteine and methionine oxidation. Mass tolerance was 0.5 Da for precursor and 1 Da for product ions. All proteins were identified on the basis of two or more unique peptides.

Bioinformatics and protein annotation

Label-free quantitation of peptide/protein expression was accomplished using the Rosetta Elucidator software (Microsoft). For Elucidator analyses, peptides across the entire chromatographic run for each sample were aligned between mass spec runs and between conditions (basal, hypertrophy, failure). The peak intensity for each eluting peptide was calculated as area under the extracted ion chromatographic curve. To determine protein abundance, intensity data for all peptides mapping to a protein were combined and data from 3 biological and 2 technical replicates was averaged for each of the three conditions. Proteins whose intensity changed ≥ 2 fold between conditions with a p-value ≤ 0.01 were considered to be statistically significant. To identify modules of proteins with similar expression behavior, intensity values were converted to Z-scores. Proteins were then clustered using a self-organizing map with x and y nodes of 3,3 and cosine correlation. Intensity data was coupled to peptide identification which was determined using the SEQUEST algorithm described above.

Protein expression plots were generated as described.(27) Redundancy in proteins was eliminated at the primary sequence level by manual inspection using CLUSTAL to compare the sequences in UniProt. For genome analysis (i.e. for determining from where in the genome mRNAs for the detected proteins were transcribed) UniProt IDs, from MS data, were converted into Ensembl gene IDs using the UniProt ID Mapping tool (<http://www.uniprot.org/?tab=mapping>), and gene annotations were obtained from Ensembl data (version 65). GO annotation enrichment analysis was performed using the DAVID Bioinformatics Resource (v6.7) developed by the NIAID (NIH). The Interpro and KEGG analysis functions of DAVID were utilized to determine enrichment in protein domains and functional pathways, respectively.

GO term enrichment analyses of proteins differentially regulated in hypertrophy or failure chromatin were compared to GO terms of proteins differentially regulated in published datasets from more or less differentiated cells or cancer or healthy tissue to determine if biological processes involved in heart disease are more similar to either cancer or development. To

compare these sub-proteomes, we examined the “biological process” term set, with that grouping being the broadest categorization; we subdivided other more discriminating terms into “intermediate” or “specific” as informed by their hierarchical level in the ontological tree.

Zebrafish studies

Zebrafish colonies were cared for and bred under standard conditions.(7, 30) Developmental stages of embryos were determined using standard morphological features of fish raised at 28.5 °C .(30)

Three of the proteins identified from our mass spectrometry data (with known or hypothesized roles in regulating chromatin structure and gene expression) were initially screened in zebrafish. Nucleolin was one of these three proteins and gave the most promising phenotype after knock-down and over-expression; we therefore chose to examine it in further detail.

For over-expression experiments, the full-length cDNA for zebrafish *nucleolin* was amplified with KOD polymerase (Novagen) and cloned into pCS2 + 3XFLAG. Plasmids were cut with Sall and SP6 RNA polymerase was used to generate mRNA for injection. Nucleolin mRNA was injected alone or coinjected with a p53 MO. Control injections performed using p53 MO alone have no affect on zebrafish development (data not shown). For Morpholino-based knockdown experiments a Morpholino oligonucleotide targeting the translation initiation site of *nucleolin* (Ncl MO) was purchased from Gene Tools, LLC. Nucleolin MO was co-injected with a p53 Morpholino to prevent non-specific cell death. The sequences of the Morpholinos used are as follows (from 5' to 3'): Nucleolin MO: TAGCTGCCTTAGCGAGCTTTACCAT; p53 MO: GCGCCATTGCTTTGCAAGAATTG. Morpholino efficacy was tested by western blotting to detect levels of Nucleolin protein in 24 and 72 hpf embryo lysates. Embryos were deyolked in calcium-free Ringer’s solution with 1mM EDTA and then lysed in buffer containing 50mM Tris (pH 8), 10 mM EDTA, and 1% SDS, with protease (0.1mM phenylmethanesulfonylflouride,

protease inhibitor cocktail pellet [Roche]), phosphatase (0.2mM sodium orthovanadate, 0.1 mM sodium fluoride) and deacetylase (10mM sodium butyrate) inhibitors, sonicated to shear the DNA and centrifuged at 13,000g to extract proteins.

Whole mount *in situ* hybridization

Embryos for *in situ* hybridization were raised in embryo medium supplemented with 0.2 mM 1-phenyl-2 thiourea to maintain optical transparency.(30) Whole-mount *in situ* hybridization was performed as described previously.(4) The antisense RNA probes used in this study include *nucleolin (ncl)*, *ventricular myosin heavy chain (vmhc)*, *atrial myosin heavy chain (amhc, also known as myh6)*, *bone morphogenetic protein 4 (bmp4)*, *versican*.

Whole mount immunohistochemistry and cell size analysis

Embryos expressing mCherry protein from the *cmlc2* promoter (also known as *myl7*) were injected with Nucleolin Morpholino at the one-cell stage and assessed at 48 hpf for a cardiac phenotype. Uninjected embryos were also maintained for control. At 72 hpf, embryos were fixed overnight in 4% paraformaldehyde in phosphate buffered saline (PBS) at 4°C. Embryos were then rinsed two times in phosphate buffered saline with 0.1% Tween 20 (PBST) and left in methanol overnight at 4°C. The pericardia were manually removed and the embryos left in acetone at -20°C for 15 min. They were then rinsed two times in water and once in PBS and then washed two times, 5 minutes each, in PBST. They were blocked one hour in blocking buffer (0.2g/ml Roche Blocking Reagent [Cat.# 11 096 176 001]), and 5% goat serum in PBST) at room temperature and left overnight at 4°C in a 1:100 primary antibody dilution in blocking buffer (anti-Neurolin Zn8 [Zn-8-s, Developmental Studies Hybridoma Bank]). Embryos were washed 4 times 15 minutes each in PBST and left overnight at 4°C in a 1:150 secondary antibody dilution in blocking buffer (Goat anti-mouse FITC [1070-02, Southern Biotech]). Embryos were again washed 4 times, 15 minutes each, in PBST and mounted using low-melt agarose. Imaging was performed on a Leica TCS-SP1 upright confocal microscope (in the CNSI Advanced Microscopy/Spectroscopy Core at UCLA) using a water-immersion objective (HCX

APO L 40x 0.8W). FITC was visualized with an argon laser (488 nm), and mCherry with a diode laser (561 nm). The ventricle was optically sectioned into 3.012 μm sections and the maximum intensity projection used for analysis in Photoshop. Anti-Zn8 antibody demarcates the cell boundary by labeling the cell membrane protein Neurolin. Laser power was optimized for each fish to enable discrimination of the cell borders in the maximum intensity projection. Any cell with its complete border clearly visible was traced in Photoshop, with roughly 12 cells analyzed per fish.

For imaging Nucleolin's localization in the zebrafish heart, whole mount immunohistochemistry was performed on uninjected, 48 hpf embryos expressing *cm/c2*-driven mCherry following the same protocol as for cell size analysis with the exception of the antibodies (primary: anti-Nucleolin [ab22758, abcam], secondary: Goat anti-rabbit FITC [4050-02, Southern Biotech and A11008, Invitrogen]). Embryos were soaked overnight in PBS containing DAPI and rinsed three times in PBST prior to imaging. Images were acquired on a Zeis LSM 510 with a water immersion objective (63x/0.9W Achroplan).

Cardiac output

End-diastolic volume (EDV), end-systolic volume (ESV), stroke volume and cardiac output (product of heart rate and stroke volume) were determined at 48 hpf from time-lapse recordings of a lateral view of the beating zebrafish heart. To calculate stroke volume we separately averaged EDV and then ESV for a given fish across three consecutive beats. Volume was calculated using the Simpson method of stacked discs.(8, 37) Briefly, two-dimensional cross sectional images of the ventricle were obtained (end diastole and end systole determined as the largest and smallest 2D sections, respectively) and the ventricular space was divided into a series of bands representing the three dimensional chamber. We set the thickness of each disc (or band) to be one pixel (1.2 μm) and measured the length of the band (diameter of the disc) in pixels as well. We assumed that the height of the disc is the same as

length and from there calculated volume. Finally, the volumes of each disc were summed to give chamber volume at end diastole or end systole.

Isolated rat cardiomyocytes

Neonatal rat ventricular myocytes (NRVMs) were obtained by enzymatic dissociation from 1 day old litters and plated in DMEM media (Invitrogen, #11965) containing 1% penicillin, 1% streptomycin, 1% insulin-transferrin-sodium selenite supplement and 10% fetal bovine serum for the first 24 hr after which the cells are cultured in serum- and antibiotic-free media. NRVMs were treated with 50nM siRNA targeted to *nucleolin* (25nM each SI00252476 and SI03107776; Qiagen) or scrambled siRNA (Cat.# 1027280; Qiagen) for 72 hrs. Transfections were performed with Lipofectamine RNAiMax Transfection Reagent (Invitrogen). To induce hypertrophy in isolated NRVMs, cells were treated with one of the following three hypertrophic agonists: isoproterenol (ISO, 1 μ M), phenylephrine (PE, 10 μ M) or endothelin-1(ET-1, 1nM) for 48 hours.

Electrophoresis and Western blotting

Proteins were separated by standard SDS-PAGE using Laemmli buffer. Gels (12%) were stained with Oriole (BioRad). For western blotting, proteins were transferred to nitrocellulose, membranes blocked with milk and protein signals detected by enzyme linked chemiluminescence (GE Biosciences). Ponceau staining of membranes was used to confirm transfer and protein loading. Antibodies used in this study were as follows, including sources: Histone H3 (Abcam, ab1791, 1:10,000 dilution); Gapdh (Santa Cruz, sc-20357, 1:1000 dilution); Histone H3-trimethylated-K9 (Abcam, ab8898, 1:500 dilution); Nucleolin (Abcam, ab22758, 1:1000 dilution); β -Actin (Sigma, A1978, 1:1000 dilution); p53 (Santa Cruz, sc1313, 1:500 dilution[NRVMs] and Abcam, ab77813, 1:200 dilution[zebrafish]); FLAG (Sigma, F1804, 1:1000 dilution).

Quantitative Real Time PCR Analysis

Total RNA was isolated from the left ventricle of the heart, from cultured NRVMs and from zebrafish embryos using TRIzol (Invitrogen) according to the manufacturer's protocol. Total RNA was transcribed using SuperScript First-Strand Synthesis system for RT-PCR (Invitrogen) according to the manufacturer's protocol to produce cDNA. cDNA transcripts were amplified on the iCycler iQ real-time PCR detection system with iQ SYBR Green Supermix (Bio-Rad). Expression levels were analyzed using the iQ5 Optical Systems software v2.0 and normalized against GAPDH by subtracting the mean cycle number for each experimental group from the mean cycle number for GAPDH from the same group. The normalized means were then applied to the formula $((2^{(\text{cycle\#})})^{-1})$ to calculate fold change. Primers used in this study are as follows: GAPDH F-5'- CCCACTAACATCAAATGGGG-3' R-5'- CCTTCCACA ATGCCAAAGTT-3'; ANF F-5'-CTGATGGATTTCAAGAACCTGCT-3' R-5'-CTCTGGGCTCCAATCCTGTC-3'; SERCA2a F-5'-CCTTCTACCAGCTGAGTCATTT-3' R5'- CAGATGGAGCCCACGACCCA-3'; α -MHC F-5'-GAACAGCTGGGAGAAGGGGG-3' R-5'-GCCTCTGAGGCTATTCTATTGG-3'; and β -MHC F-5'-CTCAACTGGGAAGAGCATCCA-3' R-5'-CCTTCAGCAAACCTCTGGAGGC-3' Ncl-Mouse F-5'-AAGCAGCACCTGGAAAACG-3' R-5'-TCTGAGCCTTCTACTTTCTGTTTCTTG-3'; 18S-Mouse F-5'-CGAGCCGCCTGGATAACC-3' R-5'-CATGGCCTCAGTTCCGAAA-3'; ITS-Mouse F-5'-TCCGTGTCTACGAGGGGCGG-3' R-5'-GGGTGCCGGGAGAGCAAAGC-3'; 18S-Rat F-5'-CGAGCCGCCTGGATAACC-3' R-5'-CATGGCCTCAGTTCCGAAA-3'; ITS-Rat F-5'-GGCGGAGGGGGTTTC-3' R-5'-GAGCGAGAAAACGGAGGAG-3'; 18S-Zebrafish F-5'-GATTGATAGCTCTTTCTCGATTCTG-3' R-5'-GTA ACTATTTAGCATGCCGGAGTCT-3'; ITS-Zebrafish F-5'-GTTCAAAGACCTTCCCGTCTC-3' R-5'-CGACACCACAAAGAGGATGTT-3'; Gapdh-Zebrafish F-5'-TGTGATGGGAGTCAACCAGGACAA-3' R-5'-TTAGCCAGAGGAGCCAAGCAGTTA-3'.

Results

Measurement of chromatin proteins with altered expression during disease.

To identify proteins regulating gene expression changes in the mouse heart during disease, we isolated chromatin-binding proteins from healthy animals as well as those in stages of cardiac hypertrophy or failure (Figure 3-1A). In contrast to our recent report(13) using this same disease model, the present study used detergent to extract proteins, thereby selecting for a more loosely associated chromatin sub-proteome, as compared to the low pH extraction (acid-extraction) used in our previous paper. This technical change in the experimental workflow has biological implications, as shown in the Venn diagram comparing the two studies in Figure 3-1E (this study verses (13)).

Proteins were separated by 1D electrophoresis, in-gel digested and analyzed on an Orbri-trap mass spectrometer (Figure 3-1A). Label-free quantitation of identified peptides was performed to elucidate abundance changes. In total, 2,068 peptides were identified and mapped to 698 proteins (377 identified by 1 peptide, 321 identified by 2 or more peptides), 338 of which displayed altered abundance by 2 fold or greater with a p-value ≤ 0.01 (Figure 3-1B). To confirm the reproducibility of the abundance changes, we performed statistical analysis (ANOVA) on all peptides identified as well as principal component analysis (PCA), which shows clustering of both biological and technical replicates (Figure 3-1C). The 338 proteins whose abundance was modulated during disease (176 quantified by 2 or more peptides) were mapped to their chromosomal location to identify genomic regions altered during cardiac dysfunction in the mouse (Figure 3-1D). We also performed unsupervised clustering of these 338 proteins to identify self-organizing trends, which revealed 9 modules of proteins, the members of which have similar behavior in terms of abundance on chromatin during disease progression (Figure 3-1F to 3-1H). Gene ontology (GO) analysis of the proteins in each module revealed the molecular functions and biological processes enriched in each group (Figure 3-1I). Proteins

identified in this study were compared to our previous analysis of *acid*-extracted chromatin, which primarily enriches for histone proteins. This comparison revealed only 37% of the proteins in the *detergent*-extracted procedure (present study) were also present in the pool revealed following *acid*-extraction (previous study; Figure 3-1E), with some proteins behaving differently during disease in the distinct fractions (Figure 3-1J). We also performed GO term enrichment analysis in which proteins differentially regulated in hypertrophy or failure chromatin were compared to proteins differentially regulated in published datasets from more or less differentiated cells or cancer or healthy tissue to determine if biological processes involved in heart disease are more similar to either cancer or development. Grouping the biological terms by hierarchical level in the GO analysis tree revealed that the processes enriched in either hypertrophy or failure had greater overlap with those enriched in healthy, as opposed to those enriched in cancer, at all levels of the hierarchical tree (broad, intermediate and specific) (Figure 3-1K, right panel). However, when comparing the most specific processes of hypertrophy and failure to differentiation (red boxes), we observe distinct behavior: hypertrophy shares more processes with less differentiated cells, while failure better matches more differentiated cells (Figure 3-1K, left panels).

Nucleolin knockdown in mammalian cardiomyocytes promotes fetal gene expression

To examine the hypothesis that cardiac growth and plasticity in disease recapitulates conserved aspects of development, we explored the role of one protein—Nucleolin—whose abundance on chromatin increased in the mouse heart during hypertrophy and failure (Figure 3-2A to 3-2D). Mass spec based quantitation of one nucleolin peptide (Figure 3-2A to 3-2C), performed by determining the area under the curve for all 18 chromatographic peaks, identified by mass spectrometry from mouse hearts under basal, hypertrophy or failure conditions (six replicates per condition) was confirmed by western blotting (Figure 3-2D). Although nuclear

Nucleolin was found to increase in this model, total cellular (nuclear and cytoplasmic) levels of Nucleolin, in mice sacrificed in stages of hypertrophy or failure (Figure 3-2E), decreased in the hypertrophic myocardium (Figure 3-2G), accompanied by changes in pre-(ITS) and mature-(18S) rRNA (Figure 3-2F to 3-2G).

Having observed alterations of chromatin-associated Nucleolin in pressure overload hypertrophy we sought to explore the actions of Nucleolin in mammalian cardiomyocytes. Knockdown was carried out in isolated neonatal rat ventricular myocytes (NRVMs) using siRNA, leading to an ~80% decrease in Nucleolin protein level, concomitant with an increase in the heterochromatin mark, H3K9Me3 (Figure 3-2H) supporting Nucleolin's role in decondensing chromatin. While Nucleolin KD did not have an effect on cell size under basal conditions (Figure 3-2I), the addition of hypertrophic agonist in the presence of Ncl siRNA was lethal to these cells (data not shown). To evaluate the effect of Nucleolin knockdown on gene transcription we quantified levels of pre-rRNA (ITS) and mature rRNA (18S) in the absence or presence of siRNA (Figure 3-2F). Nucleolin KD in NRVMs resulted in decreased rRNA transcription (ITS) and subsequent maturation (18S) (Figure 3-2J) providing a functional read-out for Nucleolin knockdown and suggesting that Nucleolin plays a role in regulating ribosomal transcription in the cardiomyocyte. Surprisingly, we observed increases in the fetal isoform (beta) of myosin heavy chain, ANF and the calcium handling protein SERCA (Figure 3-2K), suggesting that Nucleolin may regulate transcription of other (non-ribosomal) genes in the cardiomyocyte (although whether this regulation involves direct binding of Nucleolin to these genes remains to be determined). Interestingly, NRVMs treated with hypertrophic agonists (ET-1, ISO or PE) displayed differential regulation of Nucleolin protein in an agonist-dependent manner with increases in protein observed after ISO and PE treatment and a decrease in protein after ET-1 treatment (Figure 3-2L).

Nucleolin knockdown and over-expression in zebrafish leads to cardiac dysfunction

To further evaluate our hypothesis we employed a new model system to examine development in the zebrafish (*Danio rerio*), whose translucent and relatively porous viscera respectively facilitate microscopic observation and studies of cardiac development (because morphogenic abnormalities in the heart are often not embryonically lethal, in that the animal can absorb oxygen directly from the water through early maturation). Nucleolin is a highly conserved protein (Figure 3-3A) whose mRNA is ubiquitously expressed during development (Figure 3-3B), including in both chambers of the heart (Figure 3-3C to 3-3D) where it is most highly expressed in the nucleus (Figure 3-3E to 3-3F).

Because of its high conservation, ubiquitous (and early, relative to fertilization) expression and known role in ribosomal biogenesis, we anticipated that loss of Nucleolin from the single cell stage would have drastic effects on multiple aspects of the zebrafish body plan. While MO-induced knockdown was very effective in reducing Nucleolin protein levels (Figure 3-4A) and increasing the heterochromatic mark histone H3 lysine 9 trimethylation (H3K9Me3) and p53 (Figure 3-4B), the morphant embryos appeared grossly normal, with the exception of some thoracic edema (Figures 3-4C). In some cell types Nucleolin knockdown (KD) has been shown to increase *p53* expression, so co-injection of *nucleolin* and *p53* MO was performed to eliminate cell death specific phenotypes resulting from increased p53. While all *nucleolin* MO based experiments were performed in the presence and absence of *p53* MO, no differences were observed between these two conditions therefore data obtained with *p53* MO is presented as it is the more rigorous control.

Although the hearts in Nucleolin MO injected zebrafish appeared normal at 24 hpf, by 48 hpf a significant proportion exhibited defects in chamber morphology and patterning. Epifluorescence imaging of the heart revealed a number of severe cardiac abnormalities, including small ventricles (2%) and no (15%) or reverse (5%) looping (Figures 3-4D to 3-4F). Overall, we observed a cardiac phenotype in ~25% of injected embryos (Figure 3-4G), consistent with the penetrance commonly observed in Morpholino-based screens.(9, 23, 28)

Interestingly the most common abnormality was an inability of the linear heart tube to loop, a process that normally begins by ~36 hpf and is complete by 48 hpf (Figure 3-4F), leading to an underdeveloped cardiac structure in these animals. These defects in chamber looping, as well as the presence of reverse looped hearts suggests abnormal left-right asymmetry, most likely resulting from the disruption of proper signaling responses.(24) To evaluate the effect of Nucleolin knockdown on gene transcription we quantified levels of pre-rRNA (ITS) and mature rRNA (18S) in the absence or presence of *nucleolin* MO injection (Figure 3-4H). Loss of Nucleolin lead to a ~40% reduction in rRNA transcription and maturation in these fish, suggesting that Nucleolin plays a role in regulating ribosomal transcription in zebrafish as it does in mammals. No change in cardiomyocyte size was detected in zebrafish displaying a cardiac phenotype after loss of Nucleolin (Figure 3-4I).

To characterize the effects of Nucleolin knockdown on cardiac parameters, we analyzed live videos of zebrafish hearts imaged by virtue of *cmhc2* driven cardiac-specific expression of the mCherry fluorescent protein (Figure 3-5A). Line scanning software enabled quantitation of heart rate and fractional shortening in MO-injected and control fish.(34) The heart rate of embryos exhibiting edema or no looping decreased ~10% compared to uninjected controls, while the fractional shortening increased 15-30% and decreased 10% in fish with no looping/small ventricles and reverse-looped hearts, respectively (Figure 3-5B). To assess whether changes in fractional shortening were compensated by inverse changes in heart rate, we calculated the cardiac output (Figure 3-5C). Zebrafish with small ventricular chambers or edema displayed a significant decrease in overall cardiac output (Figure 3-5D).

Analysis of chamber specification and differentiation in zebrafish hearts was accomplished by *in situ* hybridization using markers of the atrium (*amhc*), the ventricles (*vmhc*) and the A/V boundary (*versican* and *bmp4*) in 48 hpf embryos (Figure 3-5E to 3-5F). Chamber specific expression of atrial (*amhc*) and ventricular (*vmhc*) myosin heavy chain appeared normal after Nucleolin MO injection (Figure 3-5E). However, analysis of the myocardial A/V boundary

marker *bmp4* (myocardial specific) displayed aberrant staining throughout the ventricular chamber in Nucleolin deficient animals, in contrast to the A/V boundary restriction observed with the endocardial marker *versican* (Figure 3-5F), which was indistinguishable in localization between normal and MO-treated animals. While ventricular expression of *bmp4* is normal in 24 hpf embryos, these cells typically silence *bmp4* transcription prior to 48 hpf as they continue differentiating into cardiomyocytes, suggesting that Nucleolin knockdown inhibits differentiation of ventricular cells. In addition, abnormal signaling by *bmp4*, a known regulator of left-right cardiac asymmetry during development,(5, 6, 26) provides a mechanistic explanation for the morphological defects in chamber looping.

To examine the effects of Nucleolin over-expression, flag-tagged *nucleolin* RNA was injected into one cell-stage embryos. Western blotting for total (Nucleolin antibody) and exogenous Nucleolin (FLAG antibody) in 24 hpf embryos confirmed an increase of ~3 fold in injected animals (Figure 3-6A) and increased p53. Embryos began exhibiting gross morphological defects at 24 hpf, with ~60% of embryos displaying severe defects at 48 hpf (Figures 3-6B to 3-6C) including edema and abnormal dorsal ventral patterning indicative of a bent or truncated body axis (short, curled tail).(15, 39) Analysis of the hearts in these fish revealed defects in chamber morphology and patterning similar to those observed after Nucleolin KD, which were mildly attenuated with co-injection of the p53 MO (Figure 3-6D to 3-6E). Staining for *bmp4* in embryos with elevated Nucleolin levels revealed aberrant expression seen by intense ventricular staining in contrast to the A/V restricted expression in uninjected embryos (Figure 3-6F). These impairments in dorsal ventral axis formation (Figure 3-6G) and cardiac looping (Figure 3-6H) upon Nucleolin over-expression mirror the phenotypes observed after perturbation of BMP4 signaling in the developing embryo as reported previously.(5, 6, 12, 15, 26, 38, 39)

Taken together, these results demonstrate that Nucleolin decondenses chromatin and facilitates normal rRNA transcription, processing and ribosomal biogenesis in the heart. Loss of

Nucleolin perturbs gene expression (BMP4 and β -MHC) and impairs cardiac chamber looping and function, whereas upregulation affects dorsal ventral axis formation (Figure 3-7).

Discussion

The hypothesis that cardiac hypertrophy and failure involves recapitulation of fetal gene expression is widely accepted, however, it lacks a fundamental mechanism to explain how such a genome wide change in DNA accessibility would occur. Genes whose expression is altered in heart disease are structurally distributed across the genome, so a global act of chromatin remodeling, that enables rapid, systematic and reproducible reprogramming of genomic structure would be a desirable feature of the above-referenced “mechanism” for fetal gene reprogramming. To address this knowledge gap, detergent-extracted chromatin proteins were identified and quantified in the healthy mouse heart and comparison made with the pressure overloaded heart in stages of hypertrophy and failure. The first component to this study was the determination of the chromatin-associated proteins in the diseased heart, which enhances our understanding of the raw materials for genome packing and modification. The second major component of the present work involved testing the hypothesis that chromatin-associated proteins activated during disease are responsible for gene expression and phenotypic events normally operative in cardiac development.

To test this hypothesis, we performed gain- and loss-of-function studies in zebrafish, demonstrating that Nucleolin regulates chamber patterning and heart function. While additional studies will be necessary to elucidate the roles of all the individual proteins measured in the first phase of this study, our findings demonstrate the power of systems proteomics to investigate the relationship between chromatin sub-proteomes in the diseased and developing heart by combining unbiased screening and hypothesis-driven experiments.

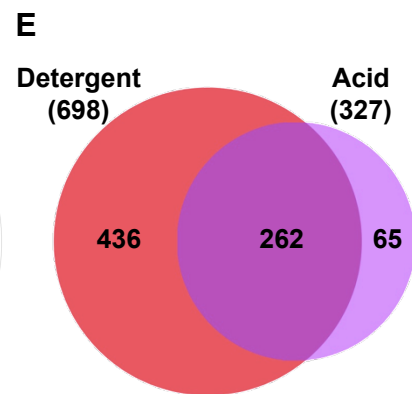
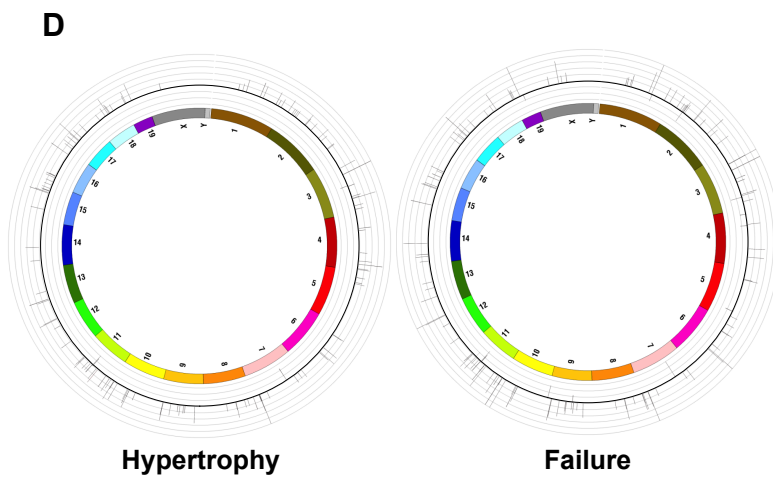
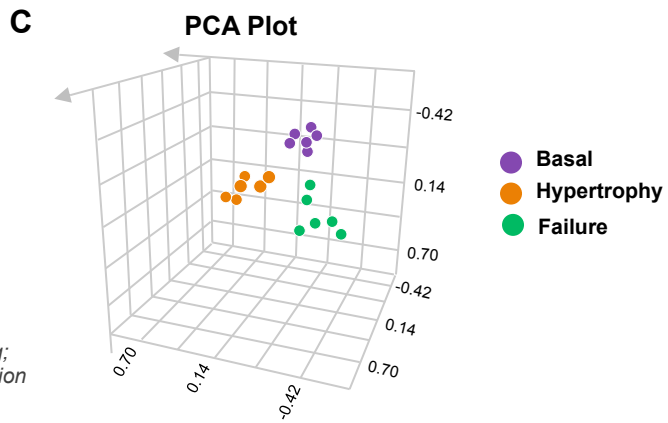
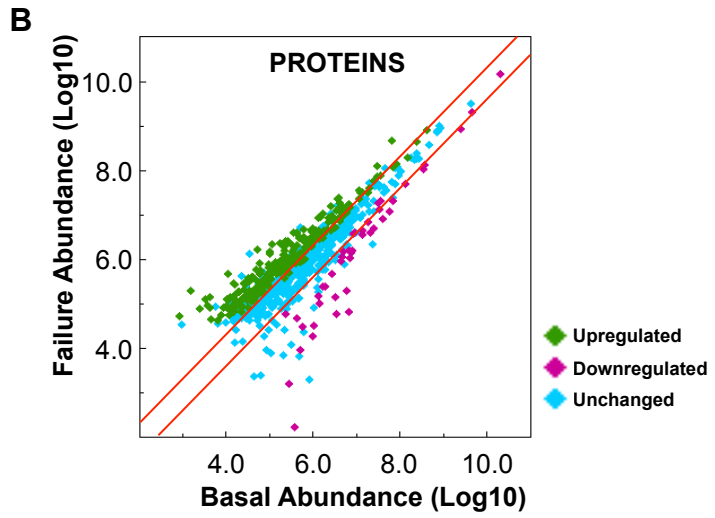
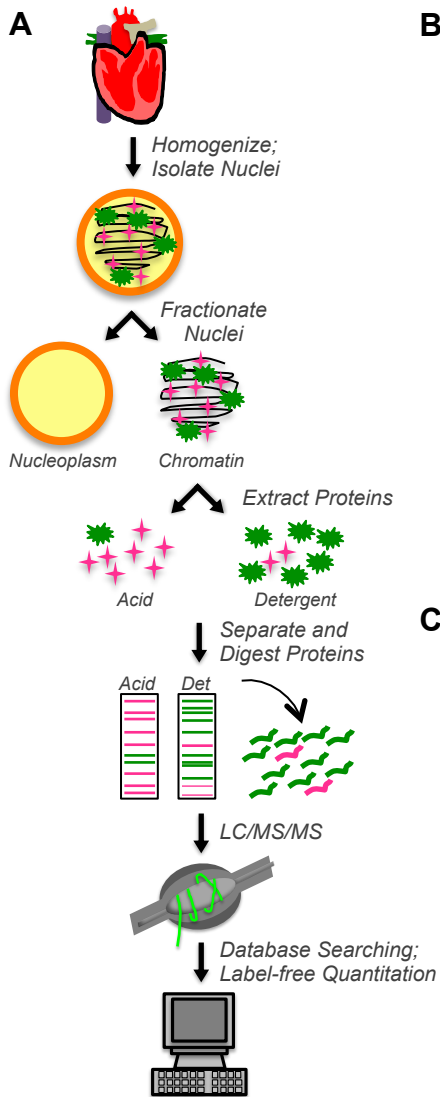
Previous studies had implied a role for nucleolar stress in cardiac disease,(1, 2) although the role of individual proteins, including Nucleolin, was unknown. While nuclear Nucleolin was found to increase its association with chromatin during hypertrophy, total cellular levels of Nucleolin fluctuated (increased or decreased) in the presence of different agonists. This observation is consistent with previous reports which have shown that cytoplasmic Nucleolin is differentially expressed compared to its nuclear counterpart, each having substantially different half lives of 45 minutes(21) and >24 hours,(25) respectively. Consistent with this, Hovanessian *et al.* reported reduced levels of Nucleolin mRNA was accompanied by markedly diminished cytoplasmic but not nuclear Nucleolin protein.(21)

Our results indicate that Nucleolin is necessary for normal cardiomyocyte differentiation during development and that loss of Nucleolin promotes the expression of fetal markers in zebrafish (*Bmp4*) and isolated cardiomyocytes (ANF, β -MHC). This phenotype is similar to that reported(40) by Yang and colleagues who observed abnormal differentiation of embryonic stem cells (ESCs) after Nucleolin knockdown. Specifically, ESCs displayed properties of differentiation but maintained the expression of pluripotency genes after loss of Nucleolin in their study. These effects were partially attributed to p53 which was found to be upregulated after Nucleolin knockdown, consistent with our observations in isolated myocytes.

The disruption in *nucleolin* expression involved misregulation of *bmp4* expression and ultimately lead to defects in patterning, specifically left-right cardiac asymmetry and dorsal ventral axis formation. These results are consistent with perturbations in BMP signaling during development which have shown that BMP4 is required for left-right patterning in the mouse(16, 26), chick(33) and zebrafish(5, 6) embryos. In addition, inhibition of BMP-4 signaling in *Xenopus*(15, 39) and zebrafish(12, 38) embryos leads to duplication, curving or truncation of the dorsal ventral body axis resulting in embryos with short curled tails (similar to our results after Nucleolin over-expression). In these models, inhibition of BMP signaling altered the proliferation and differentiation of cells in the developing embryos. These data suggest that Nucleolin may

influence myocyte plasticity via BMP4 signaling and provides a mechanistic explanation for the defects in chamber morphology and patterning, given BMP4's role in regulating muscle development, left-right asymmetry and dorsal-ventral specification.

Overall, these data support the hypothesis that some regulators of chromatin structure and gene accessibility in the developing heart are also active during heart disease to recapitulate conserved aspects of chromatin remodeling and gene expression.



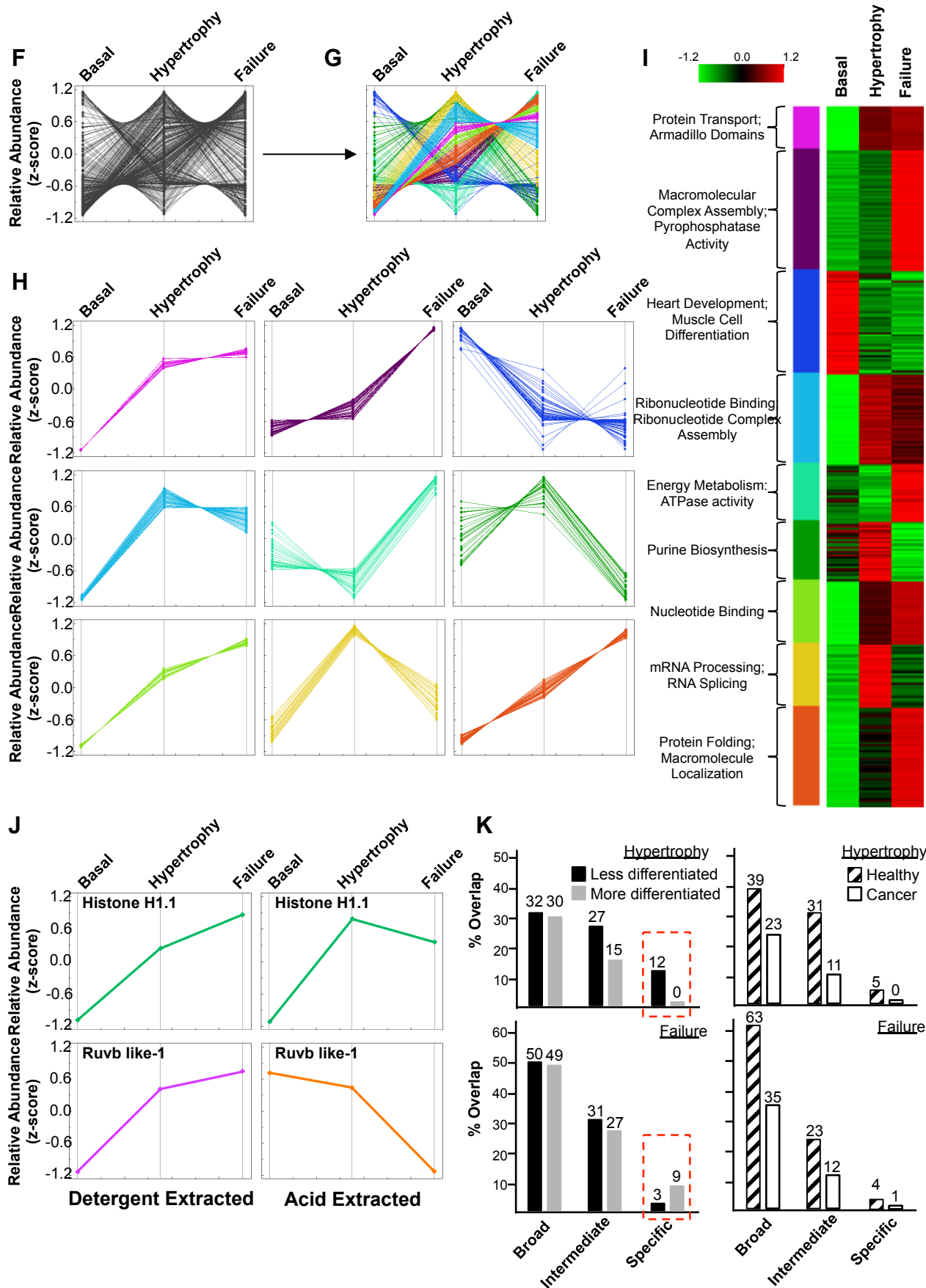


Figure 3-1: Proteomic quantification of chromatin proteins in murine heart. (A) Schematic workflow of mass spectrometric identification of mouse chromatin proteins and label-free quantitation. In this study, loosely associated chromatin proteins were isolated using detergent (as opposed to tightly bound proteins which can be isolated only in the presence of low pH, referred to as “acid-extracted proteins” in the figure and text) to investigate proteins capable of transient regulation of the genome during stages of cardiac hypertrophy and failure. (B) Peptides identified by mass spectrometry were mapped to proteins and relative quantitation determined using the Elucidator software program. Proteins increasing in abundance during the failure stage are shown as green diamonds, those decreasing as purple diamonds and those unchanged (or not statistically significant) as blue diamonds. Red lines indicate two fold change. (C) The reproducibility of peptide abundance changes from basal, hypertrophied and failing hearts was calculated using ANOVA on the 6 replicates (in each of the three conditions) and principal component analysis (PCA) performed. (D) All proteins whose abundance was found to change in stages of hypertrophy or failure by 2 fold or greater were mapped to their chromosomal location, with relative change in abundance indicated by the inflection of the lines (towards the center being downregulated, and toward the outside being upregulated), each of which corresponds to a single protein. (E) All proteins identified in this study were compared to proteins identified from our recent analysis of acid-extracted chromatin(13) using a Venn diagram to display overlap in the two datasets; note that the majority of proteins identified in the present study are distinct from our previous analysis, supporting this fractionation approach as having revealed a biologically distinct pool of proteins. (F) To identify groups of proteins with similar changes in abundance, we converted relative abundance values to Z-scores, with each protein displayed as a single line, and (G) performed unsupervised clustering based on similar quantitative behavior during disease. (H) The nine resulting modules contain proteins with corresponding changes in abundance. (I) Gene ontology (GO) analysis of each module—rendered in this panel as a heatmap—highlights biological processes and molecular functions enriched in each module. (J) Changes in protein abundance across disease states were compared between the two chromatin compartments (detergent [this study] vs acid-extraction(13), with some proteins (Histone H1.1) showing the same pattern in both fractions, while others (Ruvb like-1) behaved differently. (K) We also performed GO term enrichment analysis in which proteins differentially regulated in hypertrophy or failure chromatin were compared to proteins differentially regulated in published datasets from more or less differentiated cells or cancer or healthy tissue to determine if biological processes involved in heart disease are more similar to either cancer or development (Supplemental Table 3-4). To compare these sub-proteomes, we examined the “biological process” term set, with that grouping being the broadest categorization; we subdivided other more discriminating terms into “intermediate” or “specific” as informed by their tier in the ontological tree. While processes enriched in either hypertrophic or failing chromatin had greater overlap with healthy processes, as opposed to cancer (right panels), when comparing the most specific processes (red boxes), we observed that hypertrophy shares more common processes with less differentiated cells while failure better matches more differentiated cells (left panels).

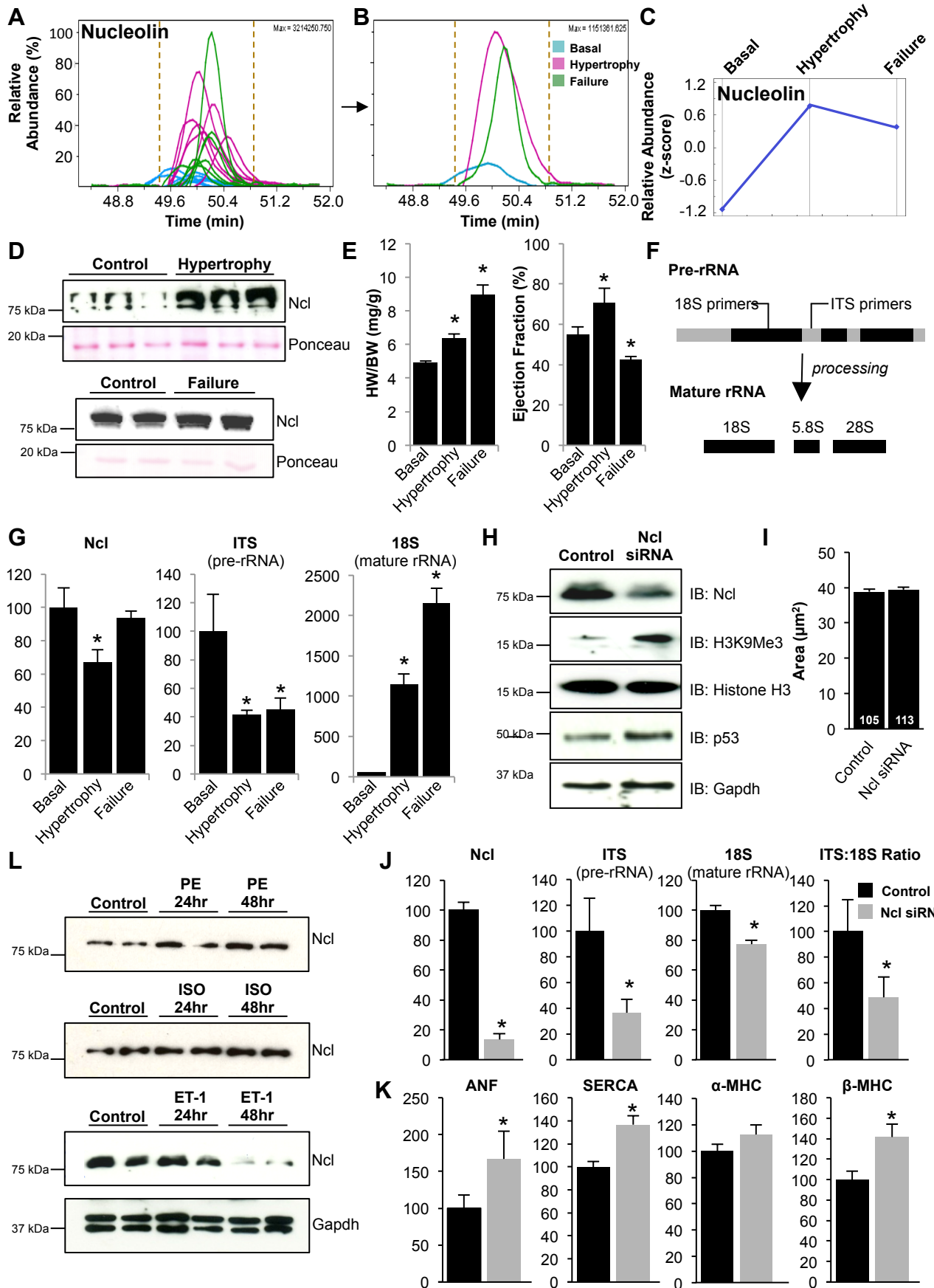


Figure 3-2: Identification of Nucleolin as a candidate regulator of cardiac growth. (A) The abundance of chromatin-bound Nucleolin was quantified by calculating the area under the curve (from 18 replicates: 6 replicates in each of three conditions) for each of two Nucleolin peptides used for identification and quantitation. Shown are the extracted ion chromatograms of the Nucleolin peptide VEGSEPTTPFNLFIGNLNPNK. x-axis indicates chromatographic time. **(B)** Replicates from the three biological conditions (basal, hypertrophy and failure) were averaged to give average abundance values for each condition which, as shown in the trend plot (**panel C**; this plot includes mean value of two peptides used for detection/quantitation), demonstrate an increase in Nucleolin during hypertrophy that persists in the failing heart. **(D)** Increased levels of chromatin-bound Nucleolin from mice in cardiac hypertrophy and failure were confirmed via western blotting, supporting mass spectrometry results. Total cellular (nuclear and cytoplasmic) levels of Nucleolin was also examined in **(E)** mice sacrificed in stages of hypertrophy or failure, as determined by heart weight (HW) to body weight (BW) ratio and ejection fraction (ECHO), by qPCR analysis showing a decrease in the hypertrophic myocardium, accompanied by changes in pre-(ITS) and mature- (18S) rRNA using primers that uniquely detect these species **(F&G)**. These results suggest that while chromatin bound Nucleolin increases during cardiac hypertrophy, total Nucleolin mRNA levels in the cell decrease. To examine the role of Nucleolin in the cardiomyocyte isolated NRVMs were treated with Nucleolin siRNA or lipofectamine control. Western blot analysis confirmed that **(H)** loss of Nucleolin results in an increase in the heterochromatin mark H3K9Me3 (blots are indicative of 3 independent experiments, N=3 in each) but **(I)** had no effect on cell size under basal conditions when quantified via phalloidin staining. Number of cells measured is indicated in each bar. **(F)** rRNA transcription and maturation was examined using primers that detect pre-rRNA (ITS) and mature rRNA (18S) by qRT-PCR. **(J)** Loss of Nucleolin in NRVMs also led to a decrease in both rRNA transcription and maturation. N=3/group. Asterisks indicate a p-value less than 0.05. **(K)** Strikingly, loss of Nucleolin in rat ventricular myocytes also led to an increase in the expression of β -MHC, ANF and Serca2a, suggesting that Nucleolin regulates some of the mRNA-encoding genes known to control pathologic cardiac growth. **(L)** Additionally, while the fraction of Nucleolin bound to chromatin increases in a mouse model of pressure overload hypertrophy, total cellular levels of Nucleolin were altered in an agonist-dependent manner in the setting of hypertrophy in NRVMs (N=3/group). Error bars indicate SEM.

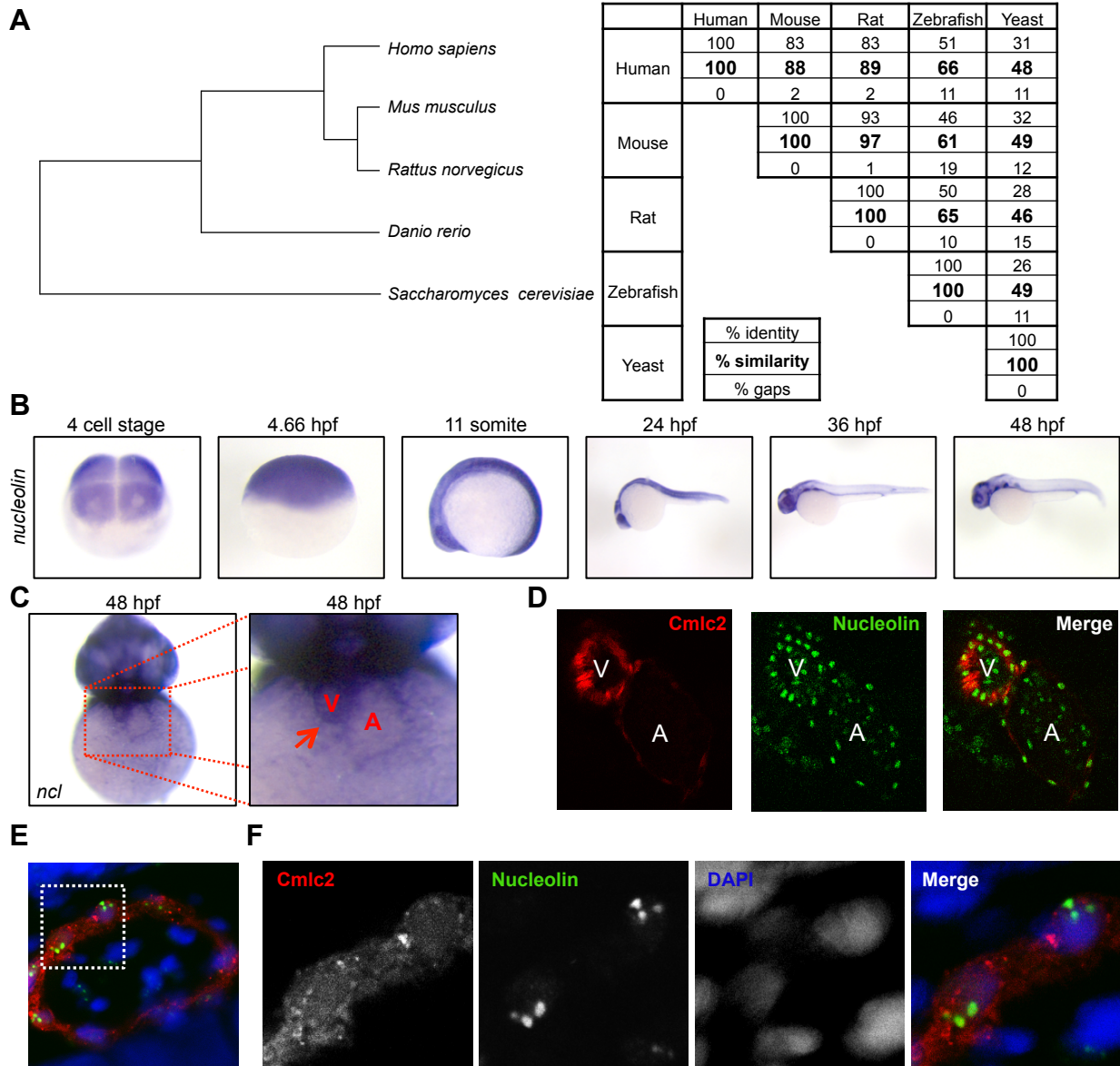


Figure 3-3: Nucleolin expression in the developing zebrafish embryo. (A) (Left) Phylogenetic analysis of Nucleolin protein sequence demonstrates degree of conservation amongst vertebrates and between model animals used in this study and a representative of the kingdom fungi (source blastP, MultiAlin, GeneBee). (Right) Numbers highlight the percent similarity between two species (source: blastP and EBI). (B) To explore the role of Nucleolin in cardiac function, endogenous *nucleolin* expression was visualized by *in situ* hybridization at the indicated stages following fertilization (hours post fertilization, hpf) in the developing zebrafish embryo. (C) At 48 hpf the zebrafish ventricle (indicated by red arrow, V) expresses higher levels of *nucleolin* than the atria (A). (D) Embryos expressing the mCherry fluorescent protein under the *cmlc2* promoter to demarcate cardiac cells (red) were fixed at 48 hpf and stained via whole mount immunohistochemistry to visualize Nucleolin protein expression (green) in the ventricle and atria. (E) The nuclear marker DAPI (blue) confirms nuclear localization of Nucleolin (green) in cardiac cells expressing *cmlc2* driven mCherry (red). White box indicates area enlarged in (F).

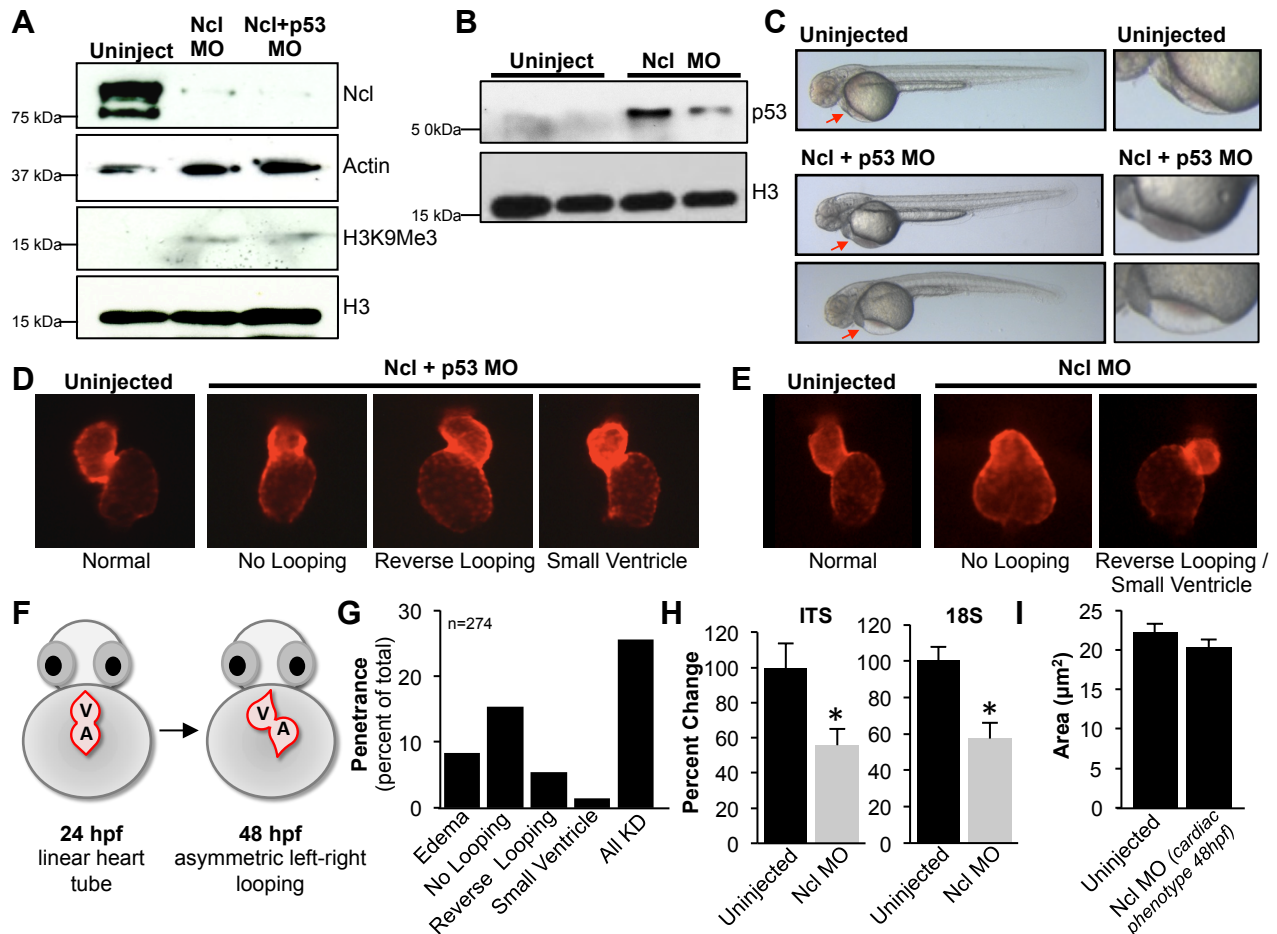


Figure 3-4: Loss of nucleolin in zebrafish results in cardiac morphological defects. Zebrafish embryos were injected at the one cell stage with 3ng of Nucleolin (Ncl) Morpholino (MO) in the presence or absence of p53 MO (3ng). Co-injection with p53 is a control to rule out generalized cell death phenotypes induced by the Nucleolin MO; in all experiments performed, the co-injected Ncl+p53 was indistinguishable from the Ncl alone, and so the former is shown as this is the more rigorous control. **(A,B)** Western blot analysis of 24 hpf embryos reveals near complete loss of Nucleolin protein, concomitant with significant increases in Histone H3 trimethylation on lysine K9 (H3K9Me3) and p53. **(C)** Depletion of Nucleolin protein had minimal effects on gross development in zebrafish, as observed at 48 hpf (a time point at which Nucleolin protein levels are still significantly diminished), with the exception of the heart, in which a population of embryos displayed blood pooling (middle panels) and/or edema (fluid accumulation in pericardial space; bottom panels). Red arrows highlight regions of defect. **(D,E)** Embryos expressing mCherry fluorescent protein under control of the *cm1c2* promoter enabled live imaging of heart morphology (images taken at 48hpf). **(F)** Normal cardiac development involves looping of the linear heart tube between 24 and 36 hpf, eventually exhibiting proper orientation by 48 hpf. **(G)** Three prominent defects in chamber patterning and morphology were observed after Nucleolin knockdown: no looping, reverse looping and decreased ventricular size (small ventricle). **(H)** rRNA transcription and maturation was examined using primers that detect pre-rRNA (ITS) and mature rRNA (18S) by qRT-PCR, demonstrating that loss of Nucleolin impaired both rRNA transcription and processing (N=3/group. Asterisks indicate a p-value less than 0.01). **(I)** 72hpf embryos were fixed and stained against Zn8, a cell membrane marker (green). Cell size was quantified in embryos with normal or reduced Nucleolin levels (displaying a cardiac phenotype at 48hpf).

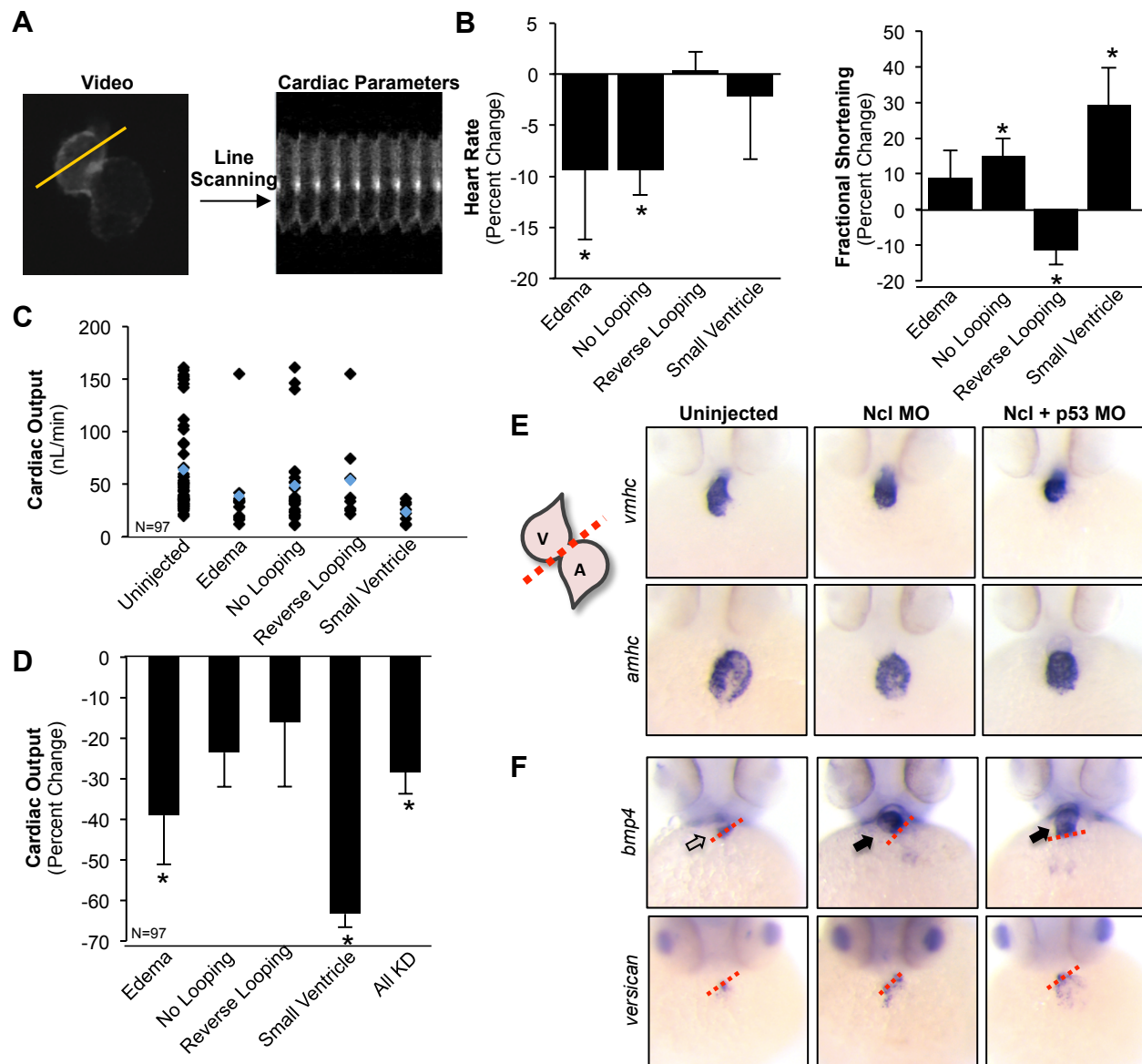


Figure 3-5: Nucleolin is essential for proper cardiac function and myocyte differentiation. (A) Embryos were injected with Nucleolin MO and heart muscle recorded at 48 hpf via fluorescence video illumination of the cardiac-specific mCherry fluorescent protein. Line scanning software was used to calculate heart rate and fractional shortening (see methods and (29)). (B) Animals with pericardial edema or no heart looping displayed decreased heart rate, while significant changes in fractional shortening were observed in no looping, reverse looping and small ventricle fish. (C) We calculated cardiac output (left panel, black diamond indicates individual fish, blue indicates group average) and (D) the percent change in the cardiac output compared to uninjected controls. (E) *In situ* hybridization was performed in 48 hpf embryos in the presence or absence of Nucleolin MO to examine markers of chamber specification and myocyte differentiation; schematic illustrates normal zebrafish heart chambers at 48 hpf. Expression and distribution of atrial (*amhc*; top panels) and ventricular (*vmhc*, bottom panels) specific myosin heavy chain were normal in Nucleolin KD embryos. (F) *Bmp4* (top panels), a myocardial lineage marker normally restricted to the AV boundary (red dashed line), displayed intense expression throughout the ventricle of hearts lacking *nucleolin*, while A/V boundary expression of *versican* (bottom panels), which is an endocardial marker, appeared normal.

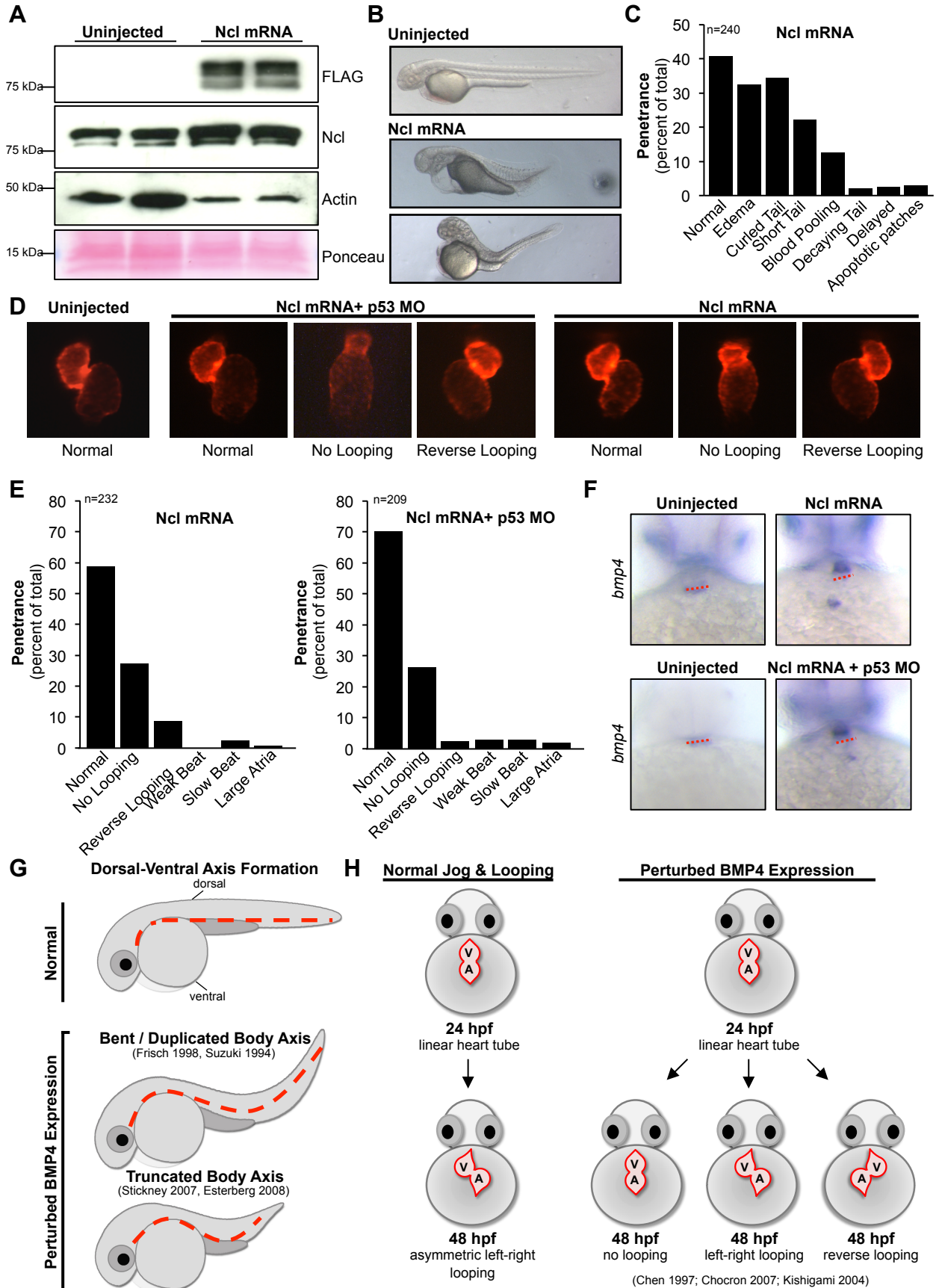


Figure 3-6: Nucleolin over-expression causes defects in heart chamber looping and dorsal ventral axis formation. Embryos were dechorinated and injected at the one cell stage with 300pg of flag-tagged Nucleolin RNA in the presence or absence of p53 MO. **(A)** Western blot analysis of 24 hpf embryos confirms modest over-expression of Nucleolin (Actin and ponceau shown as controls; we observed a consistent increase in Actin levels concomitant with Nucleolin over-expression, the origin of which is unknown at this time) and increased p53 (Histone H3 shown as control). **(B)** Analysis of gross morphology 48 hpf demonstrates severe developmental abnormalities indicative of a truncated or bent body axis. **(C)** Sixty percent of embryos displayed at least one defect, including edema, shortened yolk extension and curled tail. **(D)** The myocardia were visualized in these embryos at 48 hpf by imaging the *cmlc2* driven mCherry fluorescent protein. **(E)** Quantification of cardiac phenotypes seen after Nucleolin over-expression; greater penetrance was observed in the embryos lacking the p53 MO. **(F)** *In situ* hybridization for *bmp4* in 48 hpf embryos in the presence of Ncl mRNA (with or without p53 MO) revealed ventricular expression as opposed to the AV boundary (red dashed line) restricted expression in uninjected controls. **(G)** Defects in dorsal ventral axis formation (dotted red line) and **(H)** cardiac chamber looping observed in zebrafish embryos after Nucleolin over-expression mimic results from other investigators after perturbation of *bmp4* expression during development suggesting that Nucleolin regulates BMP4 signaling.

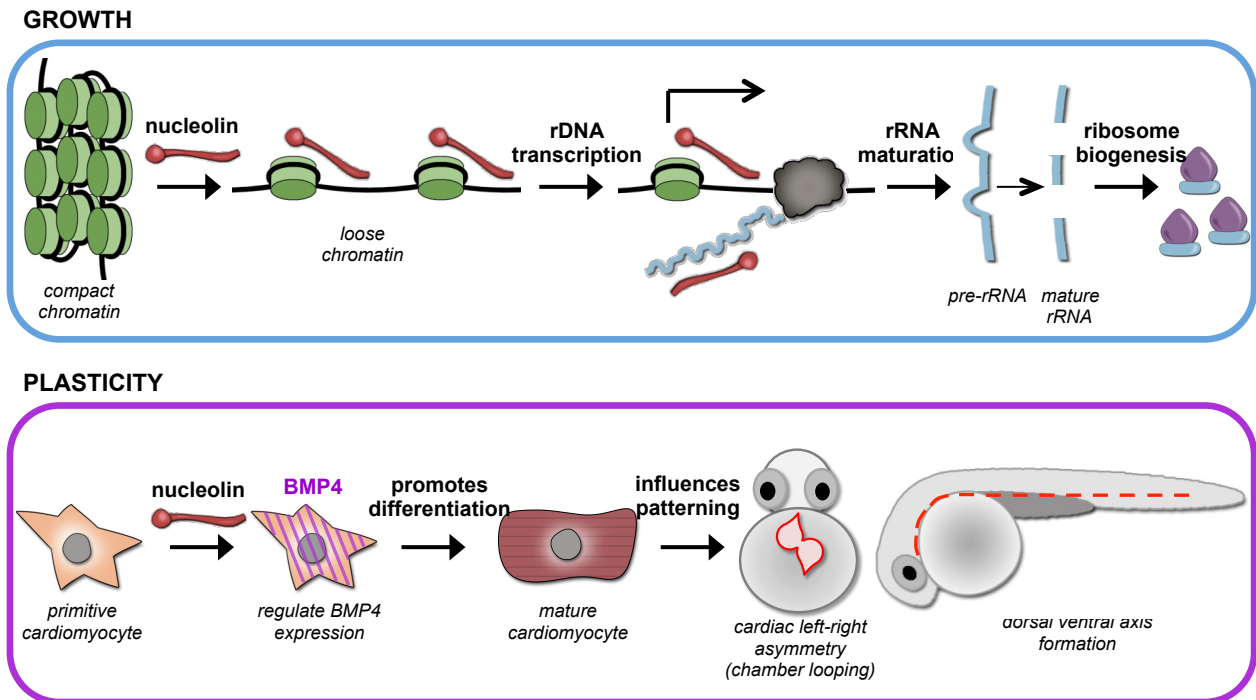


Figure 3-7: Nucleolin regulates growth and plasticity in cardiomyocytes. (Top Panel) Nucleolin binds to and decondenses chromatin in the heart, thereby enhancing rDNA transcription and maturation to support hypertrophic growth in the cardiomyocyte during disease (loss of Nucleolin in NRVMs counteracts this process). (Bottom Panel) Additionally, Nucleolin regulates BMP4 expression (and other fetal genes in mice) necessary for proper cardiomyocyte differentiation and patterning (cardiac looping and axis formation). The molecular mechanisms by which this regulation occurs and whether this is a property of nuclear or non-nuclear Nucleolin is unknown.

Chapter 3: References

1. Avitabile D, Bailey B, Cottage CT, Sundararaman B, Joyo A, McGregor M, Gude N, Truffa S, Zarrabi A, Konstandin M, Khan M, Mohsin S, Volkens M, Toko H, Mason M, Cheng Z, Din S, Alvarez R, Jr., Fischer K, and Sussman MA. Nucleolar stress is an early response to myocardial damage involving nucleolar proteins nucleostemin and nucleophosmin. *Proceedings of the National Academy of Sciences of the United States of America* 108: 6145-6150, 2011.
2. Bicknell K, Brooks G, Kaiser P, Chen H, Dove BK, and Hiscox JA. Nucleolin is regulated both at the level of transcription and translation. *Biochem Biophys Res Commun* 332: 817-822, 2005.
3. Bruneau BG. Chromatin remodeling in heart development. *Curr Opin Genet Dev* 20: 505-511, 2010.
4. Chen JN, and Fishman MC. Zebrafish tinman homolog demarcates the heart field and initiates myocardial differentiation. *Development* 122: 3809-3816, 1996.
5. Chen JN, van Eeden FJ, Warren KS, Chin A, Nusslein-Volhard C, Haffter P, and Fishman MC. Left-right pattern of cardiac BMP4 may drive asymmetry of the heart in zebrafish. *Development* 124: 4373-4382, 1997.
6. Chocron S, Verhoeven MC, Rentzsch F, Hammerschmidt M, and Bakkers J. Zebrafish Bmp4 regulates left-right asymmetry at two distinct developmental time points. *Developmental biology* 305: 577-588, 2007.
7. Choi J, Mouillesseaux K, Wang Z, Fiji HD, Kinderman SS, Otto GW, Geisler R, Kwon O, and Chen JN. Aplexone targets the HMG-CoA reductase pathway and differentially regulates arteriovenous angiogenesis. *Development* 138: 1173-1181, 2011.
8. Coucelo J, Joaquim N, and Coucelo J. Calculation of volumes and systolic indices of heart ventricle from *Halobatrachus didactylus*: echocardiographic noninvasive method. *The Journal of experimental zoology* 286: 585-595, 2000.

9. Delous M, Yin C, Shin D, Ninov N, Debrito Carten J, Pan L, Ma TP, Farber SA, Moens CB, and Stainier DY. *sox9b* Is a Key Regulator of Pancreaticobiliary Ductal System Development. *PLoS genetics* 8: e1002754, 2012.
10. Dorn GW, 2nd, and Molkentin JD. Manipulating cardiac contractility in heart failure: data from mice and men. *Circulation* 109: 150-158, 2004.
11. Erard MS, Belenguer P, Caizergues-Ferrer M, Pantaloni A, and Amalric F. A major nucleolar protein, nucleolin, induces chromatin decondensation by binding to histone H1. *Eur J Biochem* 175: 525-530, 1988.
12. Esterberg R, Delalande JM, and Fritz A. Tailbud-derived *Bmp4* drives proliferation and inhibits maturation of zebrafish chordamesoderm. *Development* 135: 3891-3901, 2008.
13. Franklin S, Chen H, Mitchell-Jordan SA, Ren S, Wang Y, and Vondriska TM. Quantitative analysis of chromatin proteome reveals remodeling principles and identifies HMGB2 as a regulator of hypertrophic growth. *Mol Cell Proteomics* 2012.
14. Franklin S, Zhang MJ, Chen H, Paulsson AK, Mitchell-Jordan SA, Li Y, Ping P, and Vondriska TM. Specialized compartments of cardiac nuclei exhibit distinct proteomic anatomy. *Mol Cell Proteomics* 10: 703, 2011.
15. Frisch A, and Wright CV. XBMPRII, a novel *Xenopus* type II receptor mediating BMP signaling in embryonic tissues. *Development* 125: 431-442, 1998.
16. Fujiwara T, Dehart DB, Sulik KK, and Hogan BL. Distinct requirements for extra-embryonic and embryonic bone morphogenetic protein 4 in the formation of the node and primitive streak and coordination of left-right asymmetry in the mouse. *Development* 129: 4685-4696, 2002.
17. Ghisolfi-Nieto L, Joseph G, Puvion-Dutilleul F, Amalric F, and Bouvet P. Nucleolin is a sequence-specific RNA-binding protein: characterization of targets on pre-ribosomal RNA. *J Mol Biol* 260: 34-53, 1996.

18. Ginisty H, Amalric F, and Bouvet P. Nucleolin functions in the first step of ribosomal RNA processing. *EMBO J* 17: 1476-1486, 1998.
19. Goyette M, Petropoulos CJ, Shank PR, and Fausto N. Regulated transcription of c-Ki-ras and c-myc during compensatory growth of rat liver. *Mol Cell Biol* 4: 1493-1498, 1984.
20. Han P, Hang CT, Yang J, and Chang CP. Chromatin remodeling in cardiovascular development and physiology. *Circ Res* 108: 378-396, 2010.
21. Hovanesian AG, Soundaramourty C, El Khoury D, Nondier I, Svab J, and Krust B. Surface expressed nucleolin is constantly induced in tumor cells to mediate calcium-dependent ligand internalization. *PLoS One* 5: e15787, 2010.
22. Jordan G. At the heart of the nucleolus. *Nature* 329: 489-490, 1987.
23. Just S, Meder B, Berger IM, Etard C, Trano N, Patzel E, Hassel D, Marquart S, Dahme T, Vogel B, Fishman MC, Katus HA, Strahle U, and Rottbauer W. The myosin-interacting protein SMYD1 is essential for sarcomere organization. *Journal of cell science* 124: 3127-3136, 2011.
24. Kathiriya IS, and Srivastava D. Left-right asymmetry and cardiac looping: implications for cardiac development and congenital heart disease. *American journal of medical genetics* 97: 271-279, 2000.
25. Kim SK, and Srivastava M. Stability of Nucleolin protein as the basis for the differential expression of Nucleolin mRNA and protein during serum starvation. *DNA Cell Biol* 22: 171-178, 2003.
26. Kishigami S, Yoshikawa S, Castranio T, Okazaki K, Furuta Y, and Mishina Y. BMP signaling through ACVRI is required for left-right patterning in the early mouse embryo. *Developmental biology* 276: 185-193, 2004.
27. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, and Marra MA. Circos: an information aesthetic for comparative genomics. *Genome Res* 19: 1639-1645, 2009.

28. Langenbacher AD, Nguyen CT, Cavanaugh AM, Huang J, Lu F, and Chen JN. The PAF1 complex differentially regulates cardiomyocyte specification. *Developmental biology* 353: 19-28, 2011.
29. Lu G, Ren S, Korge P, Choi J, Dong Y, Weiss J, Koehler C, Chen JN, and Wang Y. A novel mitochondrial matrix serine/threonine protein phosphatase regulates the mitochondria permeability transition pore and is essential for cellular survival and development. *Genes Dev* 21: 784-796, 2007.
30. M W. *The zebrafish book*. The University of Oregon Press, 2000.
31. Maunakea AK, Chepelev I, and Zhao K. Epigenome mapping in normal and disease States. *Circ Res* 107: 327-339, 2010.
32. Mitchell-Jordan SA, Holopainen T, Ren S, Wang S, Warburton S, Zhang MJ, Alitalo K, Wang Y, and Vondriska TM. Loss of Bmx nonreceptor tyrosine kinase prevents pressure overload-induced cardiac hypertrophy. *Circ Res* 103: 1359-1362, 2008.
33. Monsoro-Burq A, and Le Douarin NM. BMP4 plays a key role in left-right patterning in chick embryos by maintaining Sonic Hedgehog asymmetry. *Molecular cell* 7: 789-799, 2001.
34. Nguyen CT, Lu Q, Wang Y, and Chen JN. Zebrafish as a model for cardiovascular development and disease. *Drug discovery today Disease models* 5: 135-140, 2008.
35. Olson EN. A decade of discoveries in cardiac biology. *Nat Med* 10: 467-474, 2004.
36. Razeghi P, Young ME, Alcorn JL, Moravec CS, Frazier OH, and Taegtmeier H. Metabolic gene expression in fetal and failing human heart. *Circulation* 104: 2923-2931, 2001.
37. Schiller NB, Shah PM, Crawford M, DeMaria A, Devereux R, Feigenbaum H, Gutgesell H, Reichek N, Sahn D, Schnittger I, and et al. Recommendations for quantitation of the left ventricle by two-dimensional echocardiography. American Society of Echocardiography Committee on Standards, Subcommittee on Quantitation of Two-Dimensional Echocardiograms. *Journal of the American Society of Echocardiography : official publication of the American Society of Echocardiography* 2: 358-367, 1989.

38. Stickney HL, Imai Y, Draper B, Moens C, and Talbot WS. Zebrafish *bmp4* functions during late gastrulation to specify ventroposterior cell fates. *Developmental biology* 310: 71-84, 2007.
39. Suzuki A, Thies RS, Yamaji N, Song JJ, Wozney JM, Murakami K, and Ueno N. A truncated bone morphogenetic protein receptor affects dorsal-ventral patterning in the early *Xenopus* embryo. *Proc Natl Acad Sci U S A* 91: 10255-10259, 1994.
40. Yang A, Shi G, Zhou C, Lu R, Li H, Sun L, and Jin Y. Nucleolin maintains embryonic stem cell self-renewal by suppression of p53 protein-dependent pathway. *J Biol Chem* 286: 43370-43382, 2011.

Chapter 4: Reciprocal regulation of cardiac chromatin by HMGB and CTCF: implications for transcriptional regulation in pathologic hypertrophy

Emma M. Monte, Manuel Rosa-Garrido, Elaheh Karbassi, Haodong Chen, Christoph Rau, Jessica Wang, Yong Wu, Enrico Stefani, James N. Weiss, Aldons J. Lusis, Yibin Wang, Siavash K. Kurdistani, Sarah Franklin, Thomas M. Vondriska

Abstract

Transcriptome remodeling in heart failure is coordinated through changes in the abundance of transcription factors, histone modifications and other chromatin features localized at these genes. However, it remains unknown if, and to what extent, chromatin reorganization on a genome-wide scale throughout the nucleus contributes to the gene expression remodeling seen in the pathologic stress response in the heart. In this study, we examined the roles of two chromatin structural proteins, CTCF and HMGB2, to regulate pathologic gene expression. Our data demonstrate a reciprocal relationship between HMGB2 and CTCF in regulating myocyte hypertrophy and many aspects of chromatin structure and gene expression. HMGB2, but not CTCF, is positively correlated with heart mass across a panel of inbred mouse strains in the basal setting—in contrast, the response of HMGB2 (but not CTCF) expression to hypertrophic stress is strongly influenced by common genetic variation. Both proteins regulate each other's expression as well as the transcription of fetal genes and ribosomal RNA in cardiac myocytes: however, only HMGB2 does so in a manner that involves targeted reprogramming of chromatin accessibility. To our knowledge, these studies (using a combination of micrococcal nuclease digestion, chromatin fragment analysis and PCR) are the first to directly measure chromatin structure in cardiac cells. Lastly, while both proteins share gene targets, HMGB2 and CTCF neither bind these genes simultaneously nor do they physically co-localize in myocyte nuclei. Our study uncovers a previously unknown relationship between these two ubiquitous chromatin proteins and provides a mechanistic explanation for how HMGB2 regulates gene expression and cellular phenotype. Furthermore, we demonstrate direct evidence for hierarchical

remodeling of chromatin (at a scale above the level of a single gene) on a genome-wide scale in the setting of cardiac disease.

Introduction

Transcriptome remodeling during pathologic stress to the heart has been well documented, as has been the role of DNA and histone-modifying proteins in this process. The role of chromatin structure in gene expression during cardiovascular disease remains poorly understood. Chromatin is comprised of nucleosomes around which DNA is organized in the nucleus. Different cell types ought to have different chromatin structure underpinning their different transcriptomes, although most of the insights we have into endogenous chromatin structure[1-3] and the non-nucleosome proteins that regulate it[4, 5] comes from non-cardiac cells. There is clear evidence of chromatin regulating gene expression in cardiac hypertrophy and failure[6]. In a transverse aortic constriction (TAC) model of heart failure in mice, alterations to cis-acting histone post-translational modifications coordinate the expression changes of 325 genes[7]. In this case, the measured histone modification(s) matched the predicted expression of the gene to which it localized[8]. However, chromatin regulation of cardiac disease phenotypes extends beyond this well-defined relationship between gene expression and local histone modifications. At the level of the whole nucleus, we see a decrease in trimethylation of lysine 9 on histone H3 (H3K9me3), a marker of constitutively silenced DNA[9], and an increase in H3K4me3 abundance, a marker of active expression[10], in failing hearts after TAC[11]. Similarly, decreased H3K9me2 and increased H3K4me2 in the heart was seen in a mouse model of diabetes with glomerulosclerosis, a condition which can lead to heart disease in humans and induces hypertrophy of the cardiomyocytes in the mice[12]. This suggests a more plastic chromatin environment *on a genome wide scale* underlies gene expression remodeling during heart failure. Nuclear organization in a larger context is critical for cardiomyocyte function, with evidence from LMNA mutation-induced cardiomyopathy[13] and more recently,

studies of high mobility group nucleosome-binding domain-containing protein 5 (HMGN5) showing that the ability of the nuclear organization to withstand the forces of myocyte contraction is dependent on heterochromatin at the nuclear periphery[14]. The molecular mechanisms directly responsible for global reorganization of chromatin during pathologic gene expression in heart failure remain unknown.

CCCTC-binding factor (CTCF) is an eleven zinc-finger protein that organizes higher-order chromatin structure by one or more of the following actions: insulating genes from their enhancers[15], orchestrating DNA looping to bring together genes and their regulatory elements[16] and localizing to the boundaries between heterochromatin (compact and silenced DNA) and euchromatin (loosely-packed and accessible DNA) to prevent heterochromatin spreading[17]. Despite the well-established role for CTCF in genome organization, virtually nothing is known about its role in the normal or diseased cardiomyocyte.

High mobility group protein B2 (HMGB2) is a non-nucleosome chromatin structural protein, which through bending DNA can facilitate higher order chromatin structure[18]. The high mobility group (HMG) superfamily is composed of HMGA (recognizes DNA, regulates transcription, component of enhancesomes), HMGN (recognizes histones, “opens up” chromatin, binds nucleosomes) and HMGB[19]. HMGB binds DNA via two HMG box domains, bending DNA and unwinding the supercoiling at the region of association between the minor groove and HMG box domain[20]. HMGB has a high affinity for bent, non B-type DNA structure (including four-way junctions, and cisplatin-damaged DNA) and plays a role in several different DNA repair pathways by recognizing the damaged DNA and interacting with other proteins in the complex[21]. Similarly HMGB is involved in V(D)J recombination[22].

Evidence for HMGB's interaction with chromatin remodelers, transcription factors and transcriptional repressors has also been observed[18]. One example is a complex formed between YY1, Nucleolin (a proteins whose major role is regulating ribosomal biogenesis) and HMGB2 in human cells[23]. The authors demonstrated that in skeletal muscle, this complex

forms in a non-coding region that is deleted in patients with facioscapulohumeral muscular dystrophy and sits 37kb proximal to genes which are silenced by the presence of this complex, but become active in diseased patients[23]. This illustrates two key points. Firstly, HMGB2 can have specific effects on gene expression without localizing within the gene or promoter by altering the chromatin environment in the region (the authors found a distance-dependent repression of gene expression versus proximity to the YY1/Nucleolin/HMGB2 complex[23]). Secondly, HMGB2 can interact with Nucleolin outside of the nucleolus. We present data here showing HMGB2 regulates nucleolar transcription and affects Nucleolin levels in cardiomyocytes.

However, while HMGB facilitates complex formation its association with DNA is often dynamic. In mitosis, HMGB1 and 2 are enriched on condensed chromosomes, but there is continuous exchange between the DNA-bound and cytoplasmic pools[24]. FRAP analysis of 19 chromatin proteins in interphase mammalian nuclei revealed that most of the proteins exist in two pools of DNA-bound protein, those that dissociate quickly from the DNA (~3-11s) and those that dissociate slowly (~14-73s for non-histones) with the unbound fraction being less than 10% for all proteins examined[25]. This was true for proteins like HMGN1, HP1 β , BRG1 and Myc, however the exception was Histone H2B, Histone H1 and FUSE-binding protein which all only had a slowly dissociating pool and HMGB1 which only had a quickly dissociating pool (HMGB1 had a residence time of 4.7s, with 95% of molecules being DNA-bound at any given time)[25]. The authors propose a model where chromatin proteins bind transiently to the DNA, and then diffuse for an even shorter period of time (H1 diffuses ~200-400ms) before binding somewhere new, with the majority of molecules being DNA-bound at any give time[25]. Under this model, the chromatin state at an individual locus has the potential to turn over every several seconds by changes in local protein concentrations of the currently bound molecule and competitor proteins[25]. HMGB1 stands out in the model as being the only protein to lack a more stably bound population, which suggests its functionality does not require stable complex formation.

However, FRET analysis of HMGB1 with Glucocorticoid Receptor (GR) shows that HMGB1 can be enriched at specific loci by interaction with GR in a manner which requires both proteins be bound to the DNA[26]. Similarly, loss of HMGB1 in MEFs slowed GR mobility, and this was furthered by ATP depletion[26]. However, the reverse was not true. ATP depletion (in the absence of GR stimulation) affected the mobility of only 2% of the HMGB1 population[26]. This suggests that the effect of HMGB binding can last longer than the residence time of HMGB binding and poses a constraint on the ability to dissect the actions of HMGB at specific loci by co-immunoprecipitation. It further demonstrates that binding by one chromatin protein can greatly increase the likelihood for binding by a second chromatin protein, such that HMGB2 can act as a catalyst for chromatin binding events allowing them to occur at sub-optimal protein concentrations of the recruited protein.

Importantly, HMGBs do more than just facilitate complex formation. By destabilizing DNA histone interactions (by pulling on the linker DNA) they can promote access to nucleosome-bound DNA allowing chromatin remodeling complexes to act at otherwise inaccessible loci[27]. Furthermore, by increasing the flexibility of DNA, HMGB allows DNA to become more tightly packaged[28], with HMGB able to regulate global levels of DNA compaction in competition with linker Histone H1[29, 30]. Studies in yeast show deletion of the HMGB homologue results in hypersensitivity to digestion of DNA by micrococcal nuclease[31] providing further evidence to this point.

Despite being ubiquitous across cell types, HMG proteins have cell-type specific effects on gene regulation through mechanisms that remain poorly characterized. ChIP-seq global mapping of HMGN localization in human T-cells reveals a nonrandom distribution of HMGN1 to transcriptionally active regions[4]. HMGA1 knockout mice develop cardiac hypertrophy starting at 2 months of age, with enlarged myocytes and no change in ejection fraction or blood pressure[32]. In neonatal rat ventricular myocytes, HMGA1 knockdown also induces hypertrophy[32]. At 12 months, HMGA1 knockout mice also show abnormal proliferation of

lymphoid cells[32]. This demonstrates the role of HMG proteins in the heart, though the mechanism by which loss of HMGA causes hypertrophy is unclear, and also shows the diverse effects of the knockout depending on the cell type. Furthermore analysis of gene expression changes induced by HMGA1 knockout in embryonic stem cells was compared to gene expression changes induced by knockout in MEFs, and adult heart, revealing that for some genes, including Laminin α 1, the effect of loss of HMGA1 on expression is opposite in the heart than it is in stem cells[33].

As compared to the other HMGBs, HMGB1 is considered the most ubiquitous in the adult, in that its expression level is not down-regulated during development to the same extent. HMGB1 knockout mice die within their first day of life from hypoglycemia[34]. Remarkably, they appear mostly-normal at birth at the organ level but lack the ability to respond to glucocorticoid receptor signaling[34]. This very specific defect in transcription highlights a cell-type specific cooperation between HMGB1 and transcription factors. HMGB2 knockout males have reduced fertility and exhibit morphological defects in their testes and of spermatozoa[35]. HMGB3 knockout mice show decreased differentiation of hematopoietic stem cells into common lymphoid and myeloid progenitor cells, altering the balance of self-renewal and differentiation[36, 37]. HMGB4 is the most newly discovered member of the family and remains poorly characterized[38]. It lacks an acidic tail and is most highly expressed in the testis[38].

Our lab has found HMGB1 and 2 to be increased in the DNA-bound fraction of the cell in the TAC model of hypertrophy and failure[11]. HMGB1 is up-regulated by isoproterenol and phenylephrine in NRVMs, whereas HMGB2 displays antithetical behavior[11]. Separate from its role in the nucleus, HMGB1 that is secreted (from immune cells) or released (from necrotic cells) is an important mediator of innate immunity and may also have roles in adaptive immunity[39]. HMGB1 is up-regulated locally in atherosclerosis and promotes immune cell recruitment and adhesion as well as smooth muscle cell migration into the intimal layer[39]. After myocardial infarction, HMGB1 is up-regulated in the cardiac tissue and local white blood

cells and animal models suggest suppressing HMGB1 in the acute period is protective[39]. In humans, serum HMGB1 levels at 2-4hrs after hospital admission for myocardial infarction is inversely correlated with residual ejection fraction at 6 months post-infarction[40]. However, in other animal models, administration of HMGB1 in the hours or weeks following the infarct is beneficial[39], possibly due to its role in angiogenesis, as shown in the setting of cancer[41] and in regeneration. HMGB1 administered 4hrs after infarction in mice was shown to promote c-kit+ cells to proliferate and differentiate into cardiomyocytes[42] by stimulating cardiac fibroblasts to release paracrine factors that act on the progenitor population[43]. In its nuclear capacity in the heart, one study found HMGB2 as one of ten genes up-regulated in exercised versus sedentary rat hearts[44]. Additionally, it has been suggested that HMGB2 plays a role in growth, with levels of HMGB2 directly correlated to the proliferative potential of organs[45] and HMGB2 knockdown blocking transition from G1 to S phase in COS-1 cells[46]. However, nothing is known about the role of HMGB to package chromatin in the normal or hypertrophied heart.

We previously found that HMGB2 knockdown in isolated rat cardiomyocytes induces hypertrophy, alters the ratio of heterochromatic and euchromatic histone post-translational modifications and results in up-regulation of 101 genes and down-regulation of 99 genes as measured by microarray[11]. However, we lacked a mechanism to explain why individual genes responded differently to HMGB2 levels and how HMGB2 coordinates a phenotypic change specific to cardiomyocytes (hypertrophy), because it is not cardiac-specific and lacks sequence-specificity for DNA binding. It is these two questions that we sought to answer with this project. While we tested the possibility that HMGB2 actions could be explained by HMGB2 regulating a core set of genes known to drive hypertrophy, and in a manner targeted by cardiac-specific factors, this was not the case. Instead, we found HMGB2 can act at scales higher than individual genes with its effect on these loci determined by the chromatin environment within which it is acting.

Herein we uncover a previously unknown relationship between HMGB2 and CTCF, and use this relationship to explore the role of chromatin architectural proteins in regulating cardiac gene expression in disease.

Materials and Methods

Analysis of the hybrid mouse diversity panel (HMDP)

Microarray (RNA isolated from whole heart) and phenotypic data from 86 classical inbred and recombinant strains of mice in the basal state or after treatment with isoproterenol (ISO)[47] were analyzed. Transcript abundance was correlated with individual phenotypes to determine an r-squared value, which was converted to a p-value.

Isoproterenol treatment of mice

Adult (aged 8-10 weeks), female mice were surgically implanted with an Alzet microosmotic pump releasing 30mg/kg/day of isoproterenol for one week[47]. Control mice were untreated. Mice were monitored by echocardiography (Vevo 770) and heart weight to body weight ratio assessed at the end of one week.

Cell culture

HEK 293T and HeLa cell lines were grown in DMEM (Gibco, 11965-092) with 10% FBS. Primary neonatal rat ventricular myocytes (NRVMs) from one day old rat pups were isolated via enzymatic digestion and plated (1X penicillin-streptomycin-glutamine/10% horse serum/5% newborn calf serum/1.68% M199 salts in DMEM) for 24 hours and transferred to DMEM media containing 0.1% insulin-transferrin-sodium selenite supplement.

Knockdown was performed with 50nM total of two siRNAs per mRNA target (Qiagen: HMGB2 mouse: SI01067773, SI01067759; HMGB2 rat SI02877252, SI02877266; HMGB3: SI05428570, SI05428563; CTCF: SI01503187, SI01503208) suspended in lipofectamine (Invitrogen, NRVMs: 13778-075; cell lines: 11668-027) at time zero. CTCF siRNA treatment was repeated at 24 hours. Cells were assayed at 72 hours. Control cells were treated with

lipofectamine alone to control for toxicity. Scrambled siRNA (Qiagen, 1027281) control gave a similar result (Supplemental Figure 4-8).

Overexpression in NRVMs was performed using adenovirus (HMGB2: Vector BioLabs Adv-290952 or CTCF: Vector BioLabs, Adv-206223, 50 MOI) and assayed at 24 hours. In cell lines, plasmid constructs with GFP tagged protein or GFP alone (HMGB2: ProSpec, PRO-888, CTCF and GFP: pEGFP-CTCF and pEGFP-C2 were the generous gift of Dr. María Dolores Delgado Villar at Instituto de Biomedicina de Cantabria) were administered via lipofectamine 2000, and cells were assayed at 24 hours.

To model hypertrophy, NRVMs were treated with 1 μ M ISO (Sigma, I5627) at time zero and 24 hours or with 10 μ M phenylephrine (PHE, Sigma, P-6126) at time zero, and cells were assayed at 48 hours[11, 48]. To visualize cell density, cells were submerged in crystal violet (EMD-Millipore, 192-12) that was diluted (1% in methanol) for two minutes, and then gently rinsed.

Western blotting

Acid extraction on whole heart was performed following published protocol[49]. Isolated cells were lysed (50mM Tris pH 7.4/10mM ethylenediaminetetraacetic acid [EDTA]/1% Sodium dodecyl sulfate [SDS]/0.1mM phenylmethanesulfonylflouride/protease inhibitor cocktail pellet (Roche)/0.2mM sodium orthovanadate/0.1mM sodium fluoride/10mM sodium butyrate), sonicated, and separated via SDS-PAGE using Laemmli buffer. Detection was performed on the LI-COR odyssey. Antibodies were as follows: CTCF 1:1000 (Active Motif, 61311), HMGB2 1:1000 (Abcam, ab67282), Histone H1 1:1000 (Abcam, ab4269), H3K27me3 1:1000 (Abcam, ab6002), Histone H3 1:10,000 (Abcam, ab1791), GAPDH 1:1000 (Santa Cruz Biotechnology, sc20357), Actin 1:1000 (Santa Cruz Biotechnology, sc1616), secondaries 1:1000 (LI-COR, IRDye conjugated).

Quantitative PCR

Cells were lysed in Trizol (Ambion, 15596018). cDNA was synthesized using iScript cDNA Synthesis Kit (Bio-Rad, 170-8891). qPCR was performed using SsoFast EvaGreen Supermix (Bio-Rad, 172-5201) on a BioRad, C1000 thermocycler. Primers were as follows: CTCF F:5'-CCCAGAGTGGTACCATGAAG-3' R:5'-ACAGCATCACAGTAGCGACA-3', HMGB1 F:5'-CCAAAAGCGTGAGCTTAAA-3' R:5'-GAAGGGCAAAGGCATTACAG-3', HMGB2 F:5'-AAGCCGCGGGGCAAGATGTC-3' R:5'-TGCCCTTGGCACGGTATGCA-3', HMGB3 F:5'-TGGCTAAAGGTGATCCCAAG-3' R:5'-CTTGGCCATTTCTGTCAAAC-3', GAPDH F:5'-CCCACTAACATCAAATGGGG-3' R:5'-CCTTCCACAATGCCAAAGTT-3', SERCA2a F:5'-CCTTCTACCAGCTGAGTCATTT-3' R:5'-CAGATGGAGCCCACGACCCA-3', α -MHC: F:5'-GAACAGCTGGGAGAAGGGGG-3' R:5'-GCCTCTGAGGCTATTCTATTGG-3', β -MHC F:5'-CTCAACTGGGAAGAGCATCCA-3' R:5'-CCTTCAGCAAACCTCTGGAGGC-3', c-Myc F:5'-CGAGCTGAAGCGTAGCTTTT-3' R:5'-CTCGCGTTTCCTCAGTAAG-3', GLUT1 F:5'-AGACTTGTGGCCTCTATGTCTC-3' R:5'-CATCGTCGTTGGGATCTTA-3', GLUT4 F:5'-CCCCGATACCTCTACAT-3' R:5'-CTGGGCTGATGGTGTCTGATGC-3', 18S: F:5'-CGAGCCGCCTGGATACC-3' R:5'-CATGGCCTCAGTTCCGAAAA-3', ITS: F:5'-GGCGGAGGGGGTTTC-3' R:5'-GAGCGAGAAAACGGAGGAG-3'.

HMGB2 ChIP-seq and bioinformatics analysis

HMGB2 was analyzed by chromatin immunoprecipitation followed by massively parallel DNA sequencing (ChIP-seq). NRVMs were fixed (1% formaldehyde, 10 minutes), lysed in lysis buffer (50mM HEPES pH7.5/150mM NaCl/1mM EDTA pH8/1% Triton X-100/0.1% sodium deoxycholate/0.1% SDS, protease inhibitor cocktail tablet [Roche]), sonicated to fragments of 500 bp and diluted in RIPA buffer. DNA-bound protein was immunoprecipitated using anti-HMGB2 (Abcam, ab67282) and precipitated with protein A conjugated magnetic beads (Millipore, LSKMAGA10). Beads were washed (twice in wash buffer: 0.1%SDS/1%Triton X-100/2mM EDTA pH8/150mM NaCl/20mM Tris-HCl pH8; once in 500mM NaCl in wash buffer). DNA was eluted (1%SDS/100mM NaHCO₃, 15 minutes, 30°C) and phenol:chloroform purified.

Samples were ligated to sequencing adapters with Illumina Paired-End sample prep kit and sequenced on Illumina Genome Analyzer Iix using paired-end sequencing. Reads were aligned to the rat reference genome (rn4) using Bowtie (0.12.7)[50], with a maximum of two allowable mismatches in the seed region (first 28 nucleotides). Randomized reads of the same length and number as the HMGB2 data set served as control. MACS 1.4.1[51] was used for peak calling, with a p-value cutoff of 10^{-5} . Promoters were defined as 2kb upstream of the transcription start site (TSS). LiftOver, from UCSC genome browser, was used to convert to mouse and human genomes.

HMGB2 data was compared with the following data sets: CTCF ChIP-seq in human CD4⁺ cells, gene expression omnibus accession GSM325895; CTCF ChIP-seq in adult mouse heart, UCSC accession wgEncodeEM001684; CTCF ChIP-seq in mouse ES cells, gene expression omnibus accession GSM699165; CTCF ChIP-seq in rat liver[52]; cardiac transcription factor ChIP-seq and GATA4 knockdown microarray in HL-1[53]; HMGB2 knockdown microarray in NRVMs [11]; gene clusters based on similar expression during differentiation from mouse ES cell to cardiomyocyte[54]. Overlap of genes regulated by HMGB2 and cardiac transcription factors was determined normalizing the fraction of genes regulated by HMGB2 and also associated with the transcription factor to the fraction of all genes associated with that transcription factor.

HMGB2 ChIP-seq and microarray data was compared with Hi-C chromatin conformation capture data[3], to identify genes regulated by HMGB2 that fell in the same topological domain. Genes regulated by HMGB2 (microarray or microarray with HMGB2 promoter binding [ChIP-seq]) in rat were found in the genomes of the Hi-C data sets (mm9, hg18). For domains with more than one gene, each gene was compared to every other gene to determine if they had matching transcriptional responses to HMGB2 knockdown (either both up-regulated or both down-regulated). As control, all RefSeq genes were randomly sampled to create a mock list of genes of the same size as our HMGB2-regulated list, which were then randomly designated as

up or down-regulated, and used to determine the percentage of inter-domain comparisons that matched. This was repeated for a total of 70 times. In all cases, the average percentage that matched plateaued (~50%) after 5-30 random samples.

Genes regulated by HMGB2 (microarray) were found in the mouse genome and compared to ChIP-seq data in the adult mouse heart from ENCODE datasets: ENCFF001LJP, ENCFF001LKH, ENCFF001LKL, ENCFF001KHJ, ENCFF001KHD, ENCFF001KHV. Alignment of ChIP-seq data across regulated genes was performed using SeqPlots with the following parameters: anchored features, 10bp bins, extend targets 1kb up and downstream. As control, alignment across all genes was determined using RefSeq genes and gene predictions for mm9 downloaded from UCSC Genome Browser.

ChIP-PCR and ChIP-reChIP

ChIP was performed[55] on 30 million NRVMs fixed (1% formaldehyde, 10 minutes), lysed (50mM Tris-HCl pH 8/10mM EDTA/1% SDS/protease inhibitor cocktail Set I CALBIOCHEM I) and sonicated using an EpiShear™ Multi-Sample Sonicator (Active Motif), leading to fragments between 300 and 1000 bp. ChIP was performed using Pierce Magnetic ChIP Kit (Thermo, 26157). DNA-bound protein was immunoprecipitated using anti-HMGB2 (Abcam, ab67282), anti-CTCF (Active Motif, 61311; Abcam, ab70303) or IgG (Santa Cruz, sc2027). Results from both anti-CTCF immunoprecipitations were averaged.

For ChIP-reChIP experiments[56] manufacturer guidelines for the ChIP kit were followed up until elution. Elution was performed in elution buffer shaken at 65°C for 30 minutes. The second immunoprecipitation was performed on elutant diluted in IP dilution buffer. ChIP-reChIP experiments used anti-CTCF (Abcam, ab70303) and anti-HMGB2 (Abcam, ab67282) antibody except when the second immunoprecipitation was for the same protein as the first, in which case, the second immunoprecipitation used anti-CTCF (Active Motif, 61311) or anti-HMGB2 (Abcam, ab55169). Primers were as follows (primers target the promoter of the indicated genes; negative control determined from HMGB2 ChIP-seq): Nfkb2 F:5'-

CTGAACCGGGCCGAAGCCAA-3' R:5'-ACCCACTCCCCACACACC-3', Acta1 F:5'-
 CGCTTGCTCTGGGCCTCGTC-3', R:5'-GTTCTCCCTGCCCCCTGGCT-3', Hmgn2 F:5'-
 TGCGCGACTGGGCACATC-3' R:5'-GCCAGGCCTCGCAAACCCCT-3', Cacna1c F:5'-
 ACTGACGCCAGTCAGCCACCA-3' R:5'-ACCTGAGGGCTTTGTCCTGGCTC-3', Brd2 F:5'-
 GCGCGTCCCTGAGCTCCCTT-3' R:5'-CCGAGGCAGAGCCTCCAGCA-3', Nppb F:5'-
 ACCAGAGTGCCCGGAAGTGGTG-3' R:5'-AGGCCCTGCCCGGCTACCAA-3', Fgf16 F:5'-
 CCCCTTAAGCGCTCCCACCCT-3' R:5'-TCCCCTAGTCCCACTCCCCACC-3', Adra1b F:5'-
 GAAGATGCGCCCCAGCACCC-3' R:5'-CTGCGGACGCCACCAACTAC-3', Casp2 F:5'-
 AAGGGGCTGATGGCGGCTGA-3' R:5'-CGCGGGACCAGGCCAAGAAG-3', Mmp14 F:5'-
 AAGGAGGGCATTGGGGCGGG-3' R:5'-CGGCGAACTGAGTTGGAAGCCC-3', Tnni3 F:5'-
 AACCCGTGGCCCAGAGAGGG-3' R:5'-AGCGACGTCCGACAGGAGCA-3', Ndufb4 F:5'-
 CGGAGACCCGAAAGGCGCAG-3' R:5'-CACGATGTGCGTTTGGCGTTTGG-3', Por F:5'-
 CCCGCGGTCTGTAGGTCTCTG-3' R:5'-CCGCAGCCTTCTGGTCGGTG-3', Tgfb3 F:5'-
 CGCGATCCTGGCAGCGGTT-3' R:5'-CAGAGGGCACCCCTCGGCCTT-3', Mapk7 F:5'-
 CGGAGTGGACGGGTTCTGCCT-3' R:5'-TCCCCTCTCGCCTGCTGCTC-3', Cabin1 F:5'-
 CCTGAGCGCGACGGACCAAC-3' R:5'-TGCGCGCCAGACACACACAG-3', Parp1 F:5'-
 CTGCGGCACGAGAGGGAGGA-3' R:5'-TGCGGAGCGAGTCCTTGGGG-3', Tuba4a F:5'-
 TGGCTCAGGAGGGGGTGCTG-3' R:5'-GCGCGGGTTGGTGTTAGGGG-3', Ldha F:5'-
 CTGGGGTGGAGGTGCAGGGT-3' R:5'-CAGGCCCGCCATCCCCCTAA-3', Scn3b F:5'-
 CGTTGTCCCTGGGGGAGGCA-3' R:5'-AGTAGCCTTCCCGCCCACCC-3', Negative control 14
 F:5'-TGACAATGATGGCCCTAACA-3' R:5'-AACCGGGAACACATCATCTC-3'.

Immunohistochemistry

Hearts from BALB/c mice (8-10weeks) were fixed with formalin and paraffin embedded. Coronal sections (4µm thickness) were deparaffinized with serial washes: xylene (2x5 minutes), absolute ethanol (3x3 minutes), 95% ethanol (2x3 minutes), 70% ethanol (1x3 minutes) and distilled water (several rinses and 5 minutes). For work in isolated cells, samples were fixed with

formalin for 10 minutes. For immunostaining, samples were washed with PBS three times for 5 minutes and blocked with 5% BSA in PBS for 30 minutes. Samples were incubated with primary antibodies overnight at 4°C (1:100 in 2.5% BSA/PBS; HMGB2, Abcam ab67282 or Abcam ab55169; CTCF, BD 612148), washed with PBS, and incubated with respective secondary antibodies (confocal: 1:100 in PBS; Alexa Fluor conjugated, Life Technologies; STED: 1:100 in PBS; Atto 647N, Sigma for CTCF and Oregon Green, Life Technologies for HMGB2). DAPI (1:100) was used to demarcate the nucleus, phalloidin (1:100) was used for cell size analysis and wheat-germ agglutinin (1:100) was used to label the cell membranes. Samples were mounted with Prolong Gold.

5'Fluorouridine transcriptional run-on assay

Cells were treated with 4mM 5'fluorouridine (Sigma, F5130) for 30 minutes at 37°C [57], rinsed with 1X HEPES wash buffer (65mM PIPES/30mM HEPES/2mM MgCl₂-6H₂O/10mM EGTA) and fixed (3.7% formaldehyde/1X HEPES/0.05% Triton X-100, 10 minutes [NRVM] or 15 minutes [293T]). Cells were washed with 5 minutes washes (1X HEPES twice, PBS and 0.05% Tween/PBS) and incubated with primary antibody (BrdU 1:50, Sigma, B8434) for 2 hours at 37°C. Coverslips were washed with 0.05% Tween/PBS twice and PBS, incubated with secondary antibody (1:100 Alexa Fluor conjugated, Life Technologies) for 1 hour at room temperature, washed with PBS (3x5 minutes) and mounted with Prolong Gold. DAPI (1:100) was used to demarcate the nucleus, and Nucleolin (1:100, Abcam, ab22758) was used to mark the nucleolus. To inhibit RNA Polymerase I, cells were treated with 2uM CX5461 (Selleckchem, S2684) for 20 minutes prior to 5'FU labeling, and during the 30 minute treatment with 5'FU. P-values are based on Mann-Whitney test.

Microscopy

Images were acquired on a Nikon A1R confocal microscope and analyzed in ImageJ. For colocalization analysis, super-resolution was achieved using dual-color stimulated emission depletion (STED) microscopy on a STED instrument developed at UCLA.

MNase digestion

NRVMs were lysed (10mM Tris-HCl pH7.4/10mM NaCl/2mM MgCl₂/0.5% NP-40) and centrifuged at max speed 5 minutes. Nuclei were washed and resuspended in MNase digestion buffer (10mM Tris-HCl pH7.4/15mM NaCl/60mM KCl/1mM CaCl₂) and treated with 0.001U micrococcal nuclease (MNase, Worthington, LS004798) for 5 minutes at 37°C. Digestion was stopped with 240µL MNase digestion buffer, 60ul MNase stop buffer (100mM EDTA/10mM EGTA pH7.5), 30µL 20% SDS and 9µL proteinase K (25mg/ml). The sample was vortexed, left overnight at 37°C, and phenol:chloroform purified.

5µg of digested DNA was loaded per lane on a 1.5% agarose gel, and separated for ~6 hours at ~50 volts at 4°C. DNA was excised from the gel as follows: 2kb-20kb (most compact), 700bp-1.5kb (intermediate) and 500-650bp + 300-500bp + 100-200bp (least compact) (boundaries captured all prominent bands). DNA was purified using QIAquick Gel Extraction Kit (Qiagen, 28706). Equal volumes of DNA were analyzed by quantitative PCR. Primers are the same as listed for the ChIP-PCR protocol with the addition of: H42.1 gene body F:5'-GACGGAATGAGTGTGTGTGG-3' R:5'-CTTGCCTGTCACCCTCTCA-3'[56], Dhhrs7c promoter F:5'-TAAGACAGGCAGGACCCAAC-3' R:5'-ATCAGTGGTTTCCGATGGTC-3', Nppa promoter F:5'-CAGCTGAGATGCAAGCAGAG-3' R:5'-CCTCAGCTGCAAGAGTCACA-3'. Expression data for genes whose promoters were analyzed came from microarray for HMGB2 knockdown in NRVMs[11] or PHE treatment in NRVMs[58].

Results

To uncover the contribution of CTCF and HMGB2 to cardiac phenotype, we analyzed microarray data[47] from the hearts of 86 classical inbred and recombinant strains of mice in the basal state and after three weeks of treatment with isoproterenol (ISO), a beta-adrenergic agonist which increases the inotropy of the heart and is used to induce hypertrophy or failure in animal models[59]. Expression data and phenotypic data revealed an inverse relationship

between CTCF mRNA abundance and cardiac size after ISO treatment, as measured by total heart weight normalized to body weight and normalized right ventricular mass (Figure 4-1, left columns). Conversely, we saw a direct relationship between HMGB2 and cardiac size in the basal condition (Figure 4-1, right columns). Basal HMGB2 level also displayed a direct correlation with left and right ventricular mass after ISO, whereas basal CTCF levels were inversely correlated with right ventricular mass after ISO (Supplemental Figure 4-1A). Furthermore, CTCF expression was strongly regulated by the ISO treatment in a manner that was independent of genetics, in that the majority of strains displayed decreased CTCF after ISO (Figure 4-2A). This observation was confirmed in neonatal rat ventricular myocytes (NRVMs) treated with ISO or phenylephrine (PHE), another hypertrophic agonist, both of which down-regulated CTCF (Figure 4-2D). In contrast, regulation of HMGB2 mRNA levels was highly strain-specific (Figure 4-2B), indicating a strong effect of genetics. Western blotting from acid-extracted cardiac chromatin from two strains of mice highlights the distinct response of HMGB2 in different strains. In this case, HMGB2 is up-regulated in BUB/BnJ, a mouse which exhibits hypertrophy but not failure when treated with ISO (Supplemental Figure 4-2 and[47]) and down-regulated in BALB/cJ (Figure 4-2E), a mouse which has a depressed ejection fraction after ISO[47].

In the basal condition, HMGB2 and CTCF mRNA levels are inversely correlated and this relationship is weakened, but still significant, after ISO (Figure 4-2C). HMGB2 mRNA abundance had no significant correlation with either HMGB1 or HMGB3 in the basal, ISO, or change with ISO comparisons (Figure 4-2C and Supplemental Figure 4-1B). In support of this observation having functional significance in the heart, microarray data from the same panel of mice taken from other tissues (73 strains analyzed for liver[60], and 98 strains analyzed for macrophages[61]) showed that the relationship between HMGB2 and CTCF levels is organ-dependent (Supplemental Figure 4-3). Immunohistochemistry to label for HMGB2 and CTCF in

mouse cardiac tissue sections confirmed that both proteins are expressed in the nuclei of adult myocytes (Figure 4-2F; Supplemental Figure 4-4 shows omission of primary antibody controls).

We next examined available CTCF chromatin immunoprecipitation followed by massively parallel sequencing (ChIP-seq) data from human CD4⁺ cells (gene expression omnibus accession GSM325895), adult mouse heart (UCSC accession wgEncodeEM001684), mouse embryonic stem cells (gene expression omnibus accession GSM69916) and rat liver[52]. We performed ChIP-seq for HMGB2 in NRVMs and used liftOver to compare HMGB2 binding to CTCF peaks from the other four samples. In all four comparisons, HMGB2 reads were strongly enriched around CTCF binding peaks as compared to a randomized set of reads of similar length and number (Figure 4-3A). We also compared CTCF ChIP-seq data in the heart to HMGB2 binding peaks separated by whether they fell in genes, promoters or intergenic regions, and found the greatest enrichment in intergenic regions (Supplemental Figure 4-5). For comparison, we mapped HMGB2 enrichment around peaks for Nkx2.5, a cardiac transcription factor, in HL1 cells[53], an atrial myocyte cell line (Figure 4-3A); HMGB2 enrichment at these peaks was much less pronounced. Together, these suggest (1) HMGB2 and CTCF can bind the same regions of the genome and (2) that a portion of HMGB2 peaks may be cell-type independent.

We validated the HMGB2 ChIP-seq by ChIP-PCR in NRVMs at 20 peaks that fell within 2kb upstream of the TSS of a gene (Figure 4-3B). In our analysis, 19 of the 20 peaks showed enriched pull-down over an IgG control. ChIP-PCR for CTCF showed CTCF binding 5 of the 19 promoter regions (Figure 4-3B). These five promoters also showed CTCF binding in the adult mouse heart (UCSC accession wgEncodeEM001684) (Figure 4-3C). We then examined whether HMGB2 and CTCF bound these five promoters at the same time by ChIP-reChIP (Figure 4-4D). We immunoprecipitated CTCF with one antibody, eluted the protein and DNA complex from the beads, and then reimmunoprecipitated for CTCF using a different antibody. There was some loss of DNA in this process, but in all cases, the ChIP-reChIP successfully

pulled-down the five promoter sequences. We then immunoprecipitated HMGB2 with one antibody, followed by immunoprecipitation with a second HMGB2 antibody. In this case, the loss in sample was greater, due to the poor utility of the second HMGB2 antibody for immunoprecipitation. Despite these limitations, we still achieved enrichment of one of the five HMGB2 targets. Finally, we immunoprecipitated for HMGB2 followed by immunoprecipitation for CTCF (and vice versa with the first immunoprecipitation for CTCF followed by immunoprecipitation for HMGB2). In these experiments, the better performing HMGB2 antibody was used. However, unlike the control experiments using different CTCF antibodies against the same protein, here we saw loss of enrichment when immunoprecipitating for both CTCF and HMGB2 on the same sample. Together these experiments indicate that both CTCF and HMGB2 bind these five regions, but not at the same time, hence the ability of CTCF and HMGB2 to pull-down non-overlapping pools of these DNA fragments. Immunolabeling for HMGB2 and CTCF in NRVM nuclei (Figure 4-3E) followed by super resolution microscopy, confirmed the lack of colocalization of HMGB2 and CTCF in cardiomyocytes. HMGB2 and CTCF also lack colocalization in 293T cells (Supplemental Figure 4-6).

We next sought to understand the functional consequences of HMGB2 and CTCF on local and global chromatin packing. We used adenoviruses to overexpress or siRNAs to knockdown HMGB2 and CTCF in NRVMs (Figures 4-4A and 4-4C). siRNA transfection was mediated by lipofectamine RNAiMax, with lipofectamine addition-only used as a control (a scrambled siRNA gave comparable response as lipofectamine alone as a control for our knockdown experiments; Supplemental Figure 4-7). HMGB2 overexpression with increasing amount of virus resulted in a dose-dependent reduction in CTCF mRNA level, whereas HMGB2 knockdown caused up-regulation of CTCF (Figure 4B). HMGB2 knockdown did not affect levels of Histone H1, another chromatin structural protein (Figure 4-4B). Likewise, CTCF knockdown up-regulated HMGB2 (Figure 4-4C), whereas CTCF overexpression down-regulated HMGB2 (Figure 4-4C). By microscopy, we observed a decrease in the overall abundance of HMGB2 in

nuclei depleted of CTCF (Figure 4-4D). Together these findings extend our observation of an endogenous inverse relationship between CTCF and HMGB2 levels in the mouse heart by demonstrating that this relationship is dynamic and responsive to experimental perturbation.

To characterize the phenotypic implications of disrupting the balance of HMGB2 or CTCF in myocytes, we analyzed the global effect of HMGB2 and CTCF on cardiac gene expression. Previous studies in MEFs and have shown that loss of CTCF increases nucleolar area, however in that cell type, nucleoli number decreased[56]. Furthermore CTCF depletion in ES cells causes a modest decrease in ribosomal RNA transcripts[56].

5'fluoruridine is a uracil analogue that incorporates into newly transcribed RNA when added to the media of living cells (Supplemental Figure 4-8). We overexpressed GFP alone or HMGB2 or CTCF tagged with GFP in 293T cells and incubated them in 5'fluorouridine for 30 minutes and then stopped the reaction, fixed the cells and used immunocytochemistry to detect the localization and intensity of 5'fluoruridine labeling. We found a stark absence of transcription in the nucleoli of HMGB2 overexpressing cells (Figure 4-5A). This effect was seen in neither the CTCF nor GFP-only overexpressing cells; cells in the same plate as the HMGB2 overexpression that did not actively express the plasmid and which therefore were not green also did not show transcriptional inhibition (Figure 4-5A). We quantified the top 250 cells with the most GFP expression. There was a greater than fifty percent reduction in the mean 5'fluoruridine intensity of HMGB2 overexpressing cells (Figure 4-5B), with an inverse relationship between 5'fluoruridine intensity and GFP intensity in the overexpressing cells (Figure 4-5B). Both of these observations were also true when analyzing all 329 HMGB2 overexpressing cells, rather than focusing on the cells with the greatest overexpression (reduction in median level of 5'FU by 49%, significant correlation between GFP and 5'FU in overexpressing cells, p-value <0.001). We repeated these analyses in NRVMs overexpressing HMGB2 and again saw a reduction in total 5'FU transcription (Figure 4-5C).

This global reduction in transcription fit with our previous finding that HMGB2 knockdown in NRVMs increased the abundance of methylation on lysine 4 of histone H3 (H3K4methylation)[11], a modification associated with active promoters and enhancers [10, 62, 63], whereas loss of HMGB2 decreased the abundance of H3K9me3[11], a marker of constitutively silenced DNA[9]. In the current study, we now show that HMGB2 knockdown decreased the abundance of H3K27me3 (Figure 4-5D), a marker of facultative heterochromatin[64] (heterochromatin that is more likely to be dynamically regulated over the lifetime of the cell). Together this shows that large changes in HMGB2 abundance affect global levels of heterochromatin and transcription.

We repeated these analyses in NRVMs after HMGB2 knockdown. HMGB2 knockdown decreased nucleolar transcription with no change in total transcription, increasing the ratio of nucleoplasmic transcription to nucleolar transcription (localization of transcription determined by co-staining for Nucleolin to label nucleoli; Figure 4-6A). To directly measure mRNA transcription, we inhibited RNA Polymerase I with CX5461 and found a decreased level of 5'FU signal (Figure 4-6A). We tested whether the loss of nucleolar transcription could be due to alterations in the nucleolar structure but found no significant difference in nuclear size, nucleolar size, or the percentage of nucleoli in each nuclei that had holes (a phenomenon that we see in a minority of cells; Figure 4-6B). However, we did find fewer nucleoli on average per nuclei in the knockdown cells (Figure 4-6B). Additionally, we found a reduction in the total levels of Nucleolin (Figure 4-6B). The ratio of Nucleolin localized within the nucleoli versus within the nucleoplasm did not change with knockdown, however, when we treated cells with CX5461 to inhibit RNA Polymerase I transcription, the ratio of Nucleolin in the nucleoplasm increased, suggesting loss of rRNA transcription depleted Nucleolin from the nucleoli (Figure 4-6B). This suggests that HMGB2 knockdown is disrupting rRNA transcription separate from destroying the gross morphology of the nucleoli. Whether the loss of Nucleolin levels with knockdown is the cause or

reactive (as we see in the case of decreased nucleolar Nucleolin with rRNA transcription inhibition) remains unanswered.

When we compared this to CTCF knockdown in NRVMs we also saw a decrease in nucleolar and nucleoplasmic transcription levels (Figure 4-6C). However, unlike HMGB2 knockdown, CTCF knockdown increased the number of nucleoli per nucleus and decreased the size of the nuclei, resulting in a greater proportion of the nuclear area being occupied by the nucleolus (Figure 4-6C).

We next evaluated the effect of CTCF levels on specific genes whose expression changes are hallmarks of hypertrophy. Treatment of NRVMs with ISO resulted in hypertrophy as well as down-regulation of sarcoendoplasmic reticulum Ca^{2+} -ATPase (SERCA) and α -myosin heavy chain (α -MHC) and up-regulation of atrial natriuretic factor (ANF) (Figure 4-6D). CTCF knockdown had the opposite effect, with both CTCF knockdown and ISO treatment together resulting in an attenuation of the ISO response (Figure 4-6D). Inversely, CTCF overexpression for increasing duration resulted in a time-dependent change in expression mimicking that seen with ISO (Figure 4-6D). We had previously shown that HMGB2 knockdown also disrupts expression of these genes[11]. As an aside—but for completeness—we also examined the effects of HMGB3 to regulate this panel of genes (Supplemental Figure 4-9C), based on our finding that HMGB3 knockdown in NRVMs can also induce hypertrophy (Supplemental Figure 4-9B) and is dynamically regulated in disease (Supplemental Figure 4-9D). This candidate was deemed of potentially less relevance *in vivo* (and thus was not functionally evaluated in this study) due to the low expression detected in the microarray analysis of the hearts from the panel of 86 mouse strains (data not shown).

We next sought to explain the observation that both HMGB2 knockdown and overexpression could decrease ribosomal transcription. We examined HMGB2 binding to ribosomal genes and found enrichment across rRNA gene bodies in NRVMs as compared to the entire genome (Figure 4-7A). We hypothesized that these differences could be explained by

concentration-dependent functionality of HMGB2, such that baseline levels of HMGB2 are necessary to promote rRNA transcription by facilitating the actions of an unknown complex, however, an exorbitant excess of HMGB2 overloads steady-state of HMGB2 association with rRNA genes and promotes non-specific chromatin condensation. Importantly, mammalian genomes contain >100 repeats of ribosomal DNA units[65]. In the rat, the 45S rRNA gene (the precursor for 18S, 5.8S, 28S) is clustered in repeats on chromosomes 3, 11 and 12 with ~35 copies per chromosome (NCBI and RefSeq 2012). The majority of gene copies are silenced in mammalian cells[66]. We partially digested chromatin from NRVMs with micrococcal nuclease, isolated heterochromatic, euchromatic and intermediately-compacted DNA based on the level of digestion, and used qPCR to amplify a region of the rRNA gene (designated H42.1[56]) (Figure 4-7B). We normalized the distribution to the level of the DNA sequence in the most heterochromatic fraction (which was the majority). We then compared how this ratio changed with HMGB2 knockdown or overexpression and found that HMGB2 overexpression had minimal effect, while phenylephrine treatment increased the ratio of euchromatic to heterochromatic DNA, which we would predict based on the increased rRNA transcription in hypertrophy (Figure 4-7B). HMGB2 knockdown caused an increase in the ratio of intermediately-compacted DNA as compared to heterochromatin (Figure 4-7B), illustrating that HMGB2 knockdown and overexpression do not have directly opposing effects on this set of genes.

We next sought to further explore the specific phenotypes regulated by HMGB2 acting on rRNA transcription. We hypothesized that the dramatic decrease in rRNA synthesis upon HMGB2 overexpression would disrupt cell growth. We labeled cells with crystal violet, and then gently washed the plate, removing dead cells. In 3T3 cells, HMGB2 overexpression was lethal, as predicted, however, this was not the case in HeLa cells, where HMGB2 knockdown but not overexpression resulted in a loss of cells (Figure 4-7C). CTCF knockdown was also lethal in HeLa cells (Figure 4-7C). In NRVMs, HMGB2 knockdown, HMGB2 overexpression, CTCF knockdown or CTCF overexpression all had no effect on cell death, possibly because the cells

are not dividing, and can therefore better withstand disrupted rRNA synthesis (Figure 4-7C). This is in line with published data. In COS-1 cells, HMGB2 knockdown suppresses cell division[46]. Furthermore, CTCF knockout is embryonic lethal in mice[67, 68]. CTCF depletion in isolated cells often affects cell division or cell death processes, but in a cell-type dependent manner, potentially due to the cell-type specific localization of CTCF and arrangement of higher-order structure coordinated by CTCF[69].

We hypothesized that in cardiomyocytes HMGB2 regulation of rRNA synthesis may impact hypertrophy without being involved in cell division or death. We know from our previous work that HMGB2 is down-regulated by phenylephrine in NRVMs[11], and we know that rRNA synthesis increases in hypertrophy[70], yet we see HMGB2 knockdown decreases rRNA synthesis. In our preliminary data, we treated NRVMs with phenylephrine for 15 minutes or 1 hour and labeled for HMGB2 and Nucleolin (Figure 4-7D). (As a control, we show phenylephrine can induce hypertrophy in NRVMs in our hands; Supplemental Figure 4-10.) As we saw previously, there is less HMGB2 with phenylephrine treatment, but we also see a time-dependent enrichment of the remaining HMGB2 in the nucleolus (Figures 4-7D & E). (These studies have not been repeated on a biological replicate yet.) Thus, we propose that one of the mechanisms by which HMGB2 levels regulate cardiac hypertrophy is through regulating expression of ribosomal RNA.

We next turned to individual genes, and asked why HMGB2 knockdown affects expression of different genes uniquely. One function of CTCF is to maintain large-scale genome organization through topological domains (units of DNA spanning mega bases with a high level of self-interaction that are largely conserved across species and cell types)[3, 71]. Based on our finding that HMGB2 and CTCF binding sites overlap, we asked if the domain a gene fell in determined how HMGB2 regulated it. We took the list of genes with altered expression after HMGB2 knockdown (microarray[11]), found their location in the mouse and human genome, and determined which domain they fell in based using four different published datasets for

topological domains (Hi-C data[3]). We asked whether genes in the same domain matched in the direction with which their expression changed. In all four Hi-C datasets, when analyzing all HMGB2-regulated genes (>1.5 fold change with knockdown) and those regulated and bound by HMGB2 (promoter peak), we saw more genes matched expression than didn't (Figure 4-8A). Random sampling generated an average percent match of 50%, though some random samples were greater. The actual HMGB2 data was in the 70th percentile or greater for percent matching in each data set as compared to random sampling (Figure 4-8A, number above bar indicates percentile). By contrast, analysis comparing HMGB2 gene targets to the targets of cardiac transcription factors (Supplemental Figure 4-11) suggests that HMGB2 does not regulate cardiac physiology by targeting a specific transcriptional program.

We then looked for other measures of local chromatin environment at HMGB2-regulated genes. HMGB2 occupancy was grossly similar between up-regulated and down-regulated genes (Figure 4-8B). ChIP-seq data from the adult mouse heart (ENCODE) was used to determine the percentage of HMGB2-regulated genes with binding peaks for various chromatin proteins (Figure 4-8C). For all chromatin proteins examined, a higher percentage of HMGB2-regulated genes had binding peaks than did all genes in the genome with the exception of H3K27me3 (Figure 4-8C). Compared to all genes and down-regulated genes, genes that are up-regulated by HMGB2 knockdown had greater levels of H3K4me3 and RNA Pol II, both associated with active regions, across their promoter (1kb upstream of TSS) (Figure 4-8D). By contrast, down-regulated genes had greater levels of H3K27me3 in their promoters, a mark associated with silenced DNA (Figure 4-8D). This suggests that removing HMGB2 potentiates the transcriptional effect of the basal chromatin environment.

Lastly, we sought to directly connect the chromatin packing actions of these proteins with their function to regulate transcription and cellular phenotype. To assay the chromatin environment at individual genes, we isolated nuclei from NRVMs and treated them with 0.001U of micrococcal nuclease (MNase), an enzyme that preferentially digests DNA that is easily

accessible, namely DNA that is nucleosome free. The digested genome of a control NRVM, when run on an agarose gel, gives bands of multiple sizes: the smallest bands migrate around 150-200 bp (the size of a mono-nucleosome); bands migrating at increasing molecular masses correspond to sections of chromatin which, endogenously, reside in states of greater compaction (Figure 4-9A). Regions higher to the top of the gel are thus interpreted as more heterochromatic; those at the bottom, more euchromatic. HMGB2 knockdown shifted the distribution of the genome towards more euchromatic DNA (Figure 4-9B), whereas CTCF knockdown or PHE treatment had minimal effect on the global pattern of DNA compaction (Figure 4-9B).

Next we aimed to investigate the specific genes residing in the fragments of chromatin at different levels of compaction from the previous experiment. To do this, we cut the agarose gel containing the digested genome into three sections representing DNA that had been in heterochromatic environments (most compact), euchromatic environments (least compact) or that came from an intermediate region of the genome, and analyzed by qPCR the distribution of specific promoter sequences between these three chromatin environments. The disruption to each promoter's distribution after treatment was locus specific, and not consistent for all loci treated with HMGB2 knockdown or with PHE when normalized to basal distribution (Figure 4-9C & D). However, for the majority of promoters, the particular distribution conferred by PHE at those loci was very similar to the effect of HMGB2 knockdown at those loci, though in some cases the effect of HMGB2 knockdown was less pronounced. This was true even for genes that had distinct transcriptional responses to PHE treatment[58] and HMGB2 knockdown[11] (Figure 4-9D). This observation suggests a partially transcriptional-independent effect on chromatin structure that is still locus specific and is shared in both HMGB2 knockdown-induced and PHE-induced hypertrophy. By contrast, HMGB2 overexpression tended to uniformly have either minimal effect or shift the promoter into a heterochromatic environment (Figure 4-9C & D).

Similarly, CTCF overexpression generally resulted in more heterochromatic packing of the promoter (Figure 4-9E).

Discussion

The goal of this study was to determine the mechanisms through which chromatin structural proteins regulate pathologic gene expression. Based on our experiments, we propose a model (Figure 4-9F) whereby CTCF, acting in an insulator capacity, serves as a boundary for heterochromatin spreading. In the absence of CTCF, heterochromatin can spread, silencing nearby regions, a phenomenon that is accompanied by increased presence of HMGB2, which maintains the compact environment. Inversely, increased abundance of HMGB2 can promote heterochromatin spreading and thereby evict CTCF from that specific locus. By disrupting the balance and boundaries of heterochromatin, HMGB2 and CTCF can affect multiple genes in a given region, depending on the extent of other modification at the locus.

We favor this model, as opposed to one in which HMGB2 preferentially targets (and differentially regulates) individual genes, because HMGB2 lacks DNA sequence specificity and is not a cardiac specific protein. These observations, coupled with our new understanding of the finely regulated balance between HMGB2 and CTCF, indicate that while the overall chromatin structure of the myocyte may not be regulated with per gene resolution, this structure is critical for regulating myocyte physiology in health and disease.

An open question when we began these studies was the molecular basis for how HMGB2 ostensibly promotes the transcription of some genes while inhibiting the expression of others. Previous studies have implicated HMGB2 in transcriptional activation or repression [72-74], attributing these actions to cooperation with distinct proteins. The predominate hypothesis for the role of HMGBs in gene expression is that they bend DNA to promote binding by other proteins; HMGBs may only transiently interact with these client proteins (if at all), thereby acting as a promiscuous chaperone at different loci [18]. Our data support this model, indicating that

HMGB2 has locus-specific effects on gene expression notwithstanding a conserved effect to compact chromatin on a genome-wide scale. Future studies using locus-specific proteomic analyses will be required to determine which HMGB2 binding partners encode, in a combinatorial manner, different transcriptional logic. While informative, this approach would obfuscate the issue of trans effects, particularly through non-coding regions of the genome, by arbitrarily examining HMGB2's function only within the unit of a gene.

The increased DNA flexibility conferred by HMGB2 binding and bending[28] that facilitates formation of locus-specific complexes can also more generally facilitate tighter packaging of DNA. Previous reports have shown HMGB1 is enriched in euchromatic regions of photoreceptor nuclei[75]. However, HMGB1 and HMGB2 are also bound to highly heterochromatic DNA formed during mitosis[24]. Compared to linker histone H1, HMGB1 also compacts DNA, although to a lesser degree[29] and can directly compete H1 off of linker DNA[30]. We show that HMGB2 knockdown disrupts global measures of chromatin (H3K27me3 abundance) and the chromatin environment at specific genes without altering global levels of H1. One potential explanation is that HMGB2 knockdown decreases nucleosome abundance. Others have shown that HMGB1 deficient mouse embryonic fibroblast cells (MEFs) have reduced nucleosome number, and yeast deficient in the HMGB1 homologue display a shift to euchromatin when assayed by MNase[76], similar to what we see with HMGB2 knockdown. Indeed, *in vitro*, HMGB1 can facilitate nucleosome deposition[77]. However, we see no difference in the abundance of Histone H3 with HMGB2 knockdown, though we did not directly measure nucleosome assembly. HMGB2 can also alter nucleosome distribution by facilitating nucleosome sliding via interaction with SWI/SNF ATP-dependent chromatin remodeling complexes, as shown *in vitro*[78].

Here we propose a model where HMGB2 targeting is partially regulated by the distribution of heterochromatin, such that CTCF mediates the boundaries of hetero- and euchromatin, and HMGB2 regulates the maintenance of heterochromatin. We reason that the

overlap between HMGB2 ChIP-seq reads from rat cardiomyocytes with CTCF ChIP-seq peaks in other species and tissues indicates cell-type independent functions of the proteins. Indeed, CTCF is enriched at[3] and critical for maintaining, topological domains[71], which are largely conserved between cell types and species[3]. Furthermore, unlike transcription factors, CTCF does not preferentially localize to genes that belong to a similar class[79]. Similarly, we find that which genes are regulated by HMGB2 cannot be explained by their shared regulation by other cardiac transcription factors, all of which support a level of cell-type independent functionality; yet, we see strong cell-type specific effects on the expression of individual genes. We show CTCF overexpression can induce hypertrophic gene expression in NRVMs and that CTCF is down-regulated by hypertrophic agonists in animal and cell models. Furthermore, the changes to chromatin accessibility at individual promoters, induced by HMGB2 knockdown, largely mimic the effects of PHE treatment.

We also find both CTCF and HMGB2 regulate nucleolar transcription in multiple cells type, having particular implications for cardiac cells. Nucleolar disruption occurs with cardiac stress[80], and rRNA synthesis is up-regulated in hypertrophy[70]. Indeed, previous findings identified HMGB3 as a component of the T-cell nucleolar proteome[81] and found evidence for interaction between HMGB2 and Nucleolin[23], a major protein component of the nucleolus.

The relationship between HMGB2 and CTCF suggests a mechanism by which both chromatin structural proteins are regulated (in abundance and localization) in part by the chromatin environment, as opposed to being targeted to DNA based on the functionality of the gene in that region. CTCF, unlike HMGB2, has DNA binding consensus motifs, and CTCF is co-regulated with HMGB2; perhaps this co-regulation involves sequestering of CTCF's binding sites into heterochromatin by HMGB2, although this will take additional experiments to unequivocally test. In this model, CTCF organizes the framework of the genome within which the cell-type specific chromatin factors operate. HMGB2 also acts within the boundaries of this model to maintain heterochromatic regions (with a high density of HMGB2 to allow for tight

packaging) and facilitate complex formation (with a low density of HMGB2 priming DNA for binding by other proteins) whose specific functions are dependent on the cell-type specific proteome. In support of this model, we observe a promoter-specific effect of HMGB2 knockdown on chromatin accessibility, but a uniform effect of HMGB2 to regulate chromatin at the genomic scale. Thus, in a cell-type dependent way, the nucleus regulates the regions established by CTCF. However, disruption to the balance of HMGB2 or CTCF dismantles the boundaries of heterochromatin undoing the cell-type specific silencing.

We have shown that cardiomyocytes regulate the balance between HMGB2 and CTCF, that the two proteins are correlated with cardiac size and that they regulate pathologic gene expression in a reciprocal fashion. We hypothesize that the disruption in the ratio of HMGB2 to CTCF that we observe in disease allows for a more plastic genome, supporting a model in which global genome organization is an important component to transcriptome remodeling occurring in cardiac hypertrophy and failure.

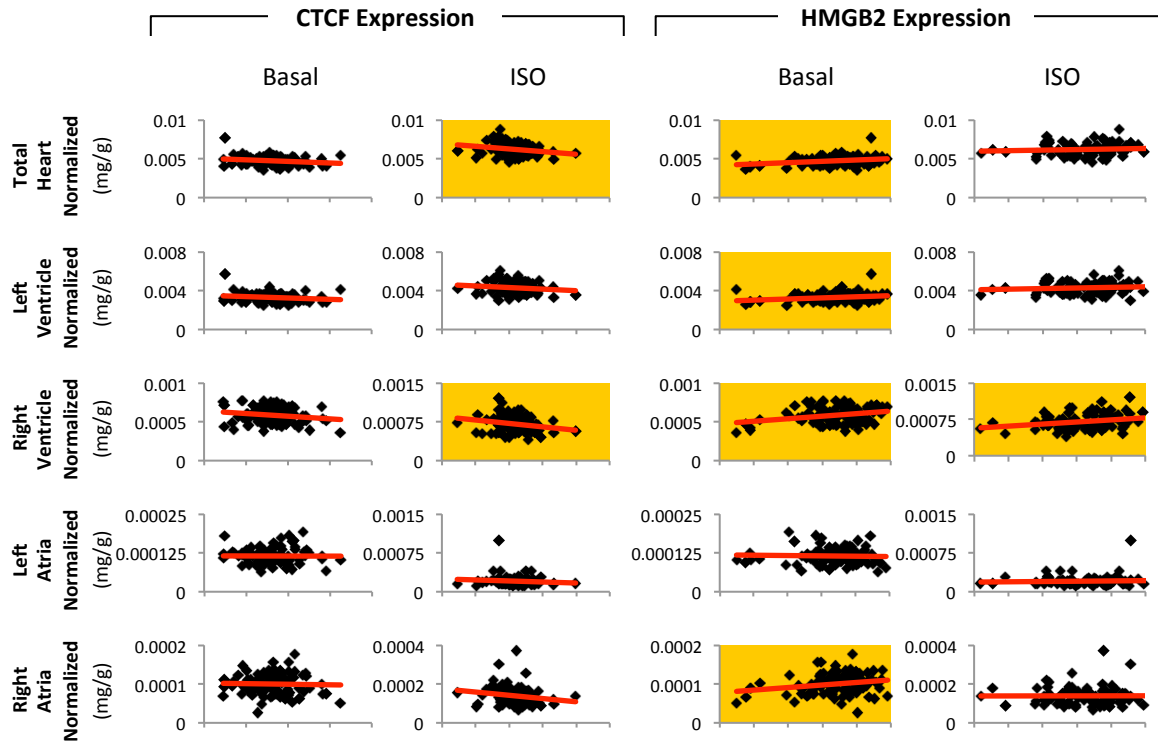


Figure 4-1. HMGB2 and CTCF abundances have opposite relationships to cardiac size. Phenotype data from 86 strains of mice[47] in the basal state or after treatment with ISO were plotted against CTCF or HMGB2 mRNA expression[47] in the respective condition. Total heart mass and right ventricular mass normalized to body weight showed significant negative correlation with CTCF expression in the ISO-treated hearts. HMGB2 exhibited a significant, direct correlation with heart size in the basal state (yellow indicates p -value < 0.05 , red line represents linear regression).

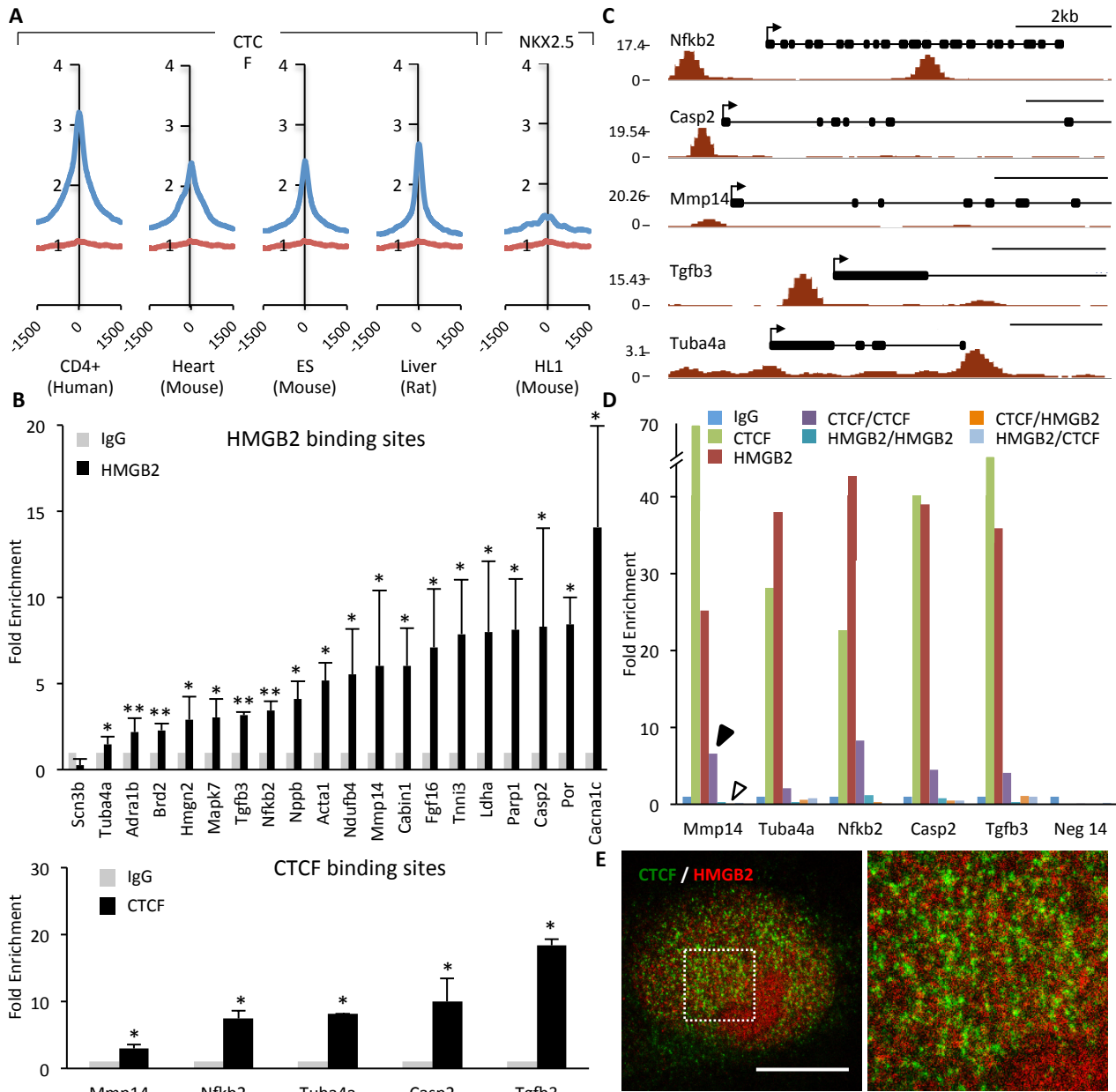


Figure 4-3. CTCF and HMGB2 can occupy the same loci, but not coincidentally. **A.** ChIP-seq data for HMGB2 in NRVMs was compared to published ChIP-seq data for CTCF in multiple tissues, including the mouse heart. HMGB2 reads (blue) are enriched around CTCF binding peaks (point 0 on x-axis), but not the cardiac transcription factor Nkx2.5. Randomly generated reads of the same number and size as the HMGB2 data-set show no enrichment. **B.** ChIP-PCR in NRVMs for HMGB2 or IgG confirmed HMGB2 binding at 19 of the 20 HMGB2 promoter peaks tested from the ChIP-seq data (* indicates p-value <0.05, ** indicates p-value < 0.005, error=standard deviation). ChIP-PCR for CTCF by two different antibodies at these same 20 loci revealed 5 (boxed above and plotted below) with significant enrichment over IgG (n=3 samples and 2 antibodies, * indicates p-value <0.05, error=standard deviation). **C.** UCSC genome browser view of CTCF ChIP-seq in the adult mouse heart confirms CTCF binding in the promoter of these genes. **D.** ChIP-reChIP experiments in NRVMs for CTCF and HMGB2 revealed loss of DNA-recovery at these five loci when immunoprecipitation for CTCF was followed by immunoprecipitation for HMGB2 (orange, white arrow), but not when followed by a second CTCF immunoprecipitation (purple, black arrow) as compared to CTCF (green) or HMGB2 (red) immunoprecipitation alone. This suggests CTCF does not bind these regions at the same time as HMGB2. No binding was found in a negative control region, chosen based on absence of reads in the HMGB2 ChIP-seq data (Neg 14). **E.** Immunolabeling for HMGB2 and CTCF in NRVMs was detected by STED microscopy (zoom in right panel) and confirmed a lack of colocalization of these two proteins (colocalization would appear as yellow). Scale is 5 μ m.

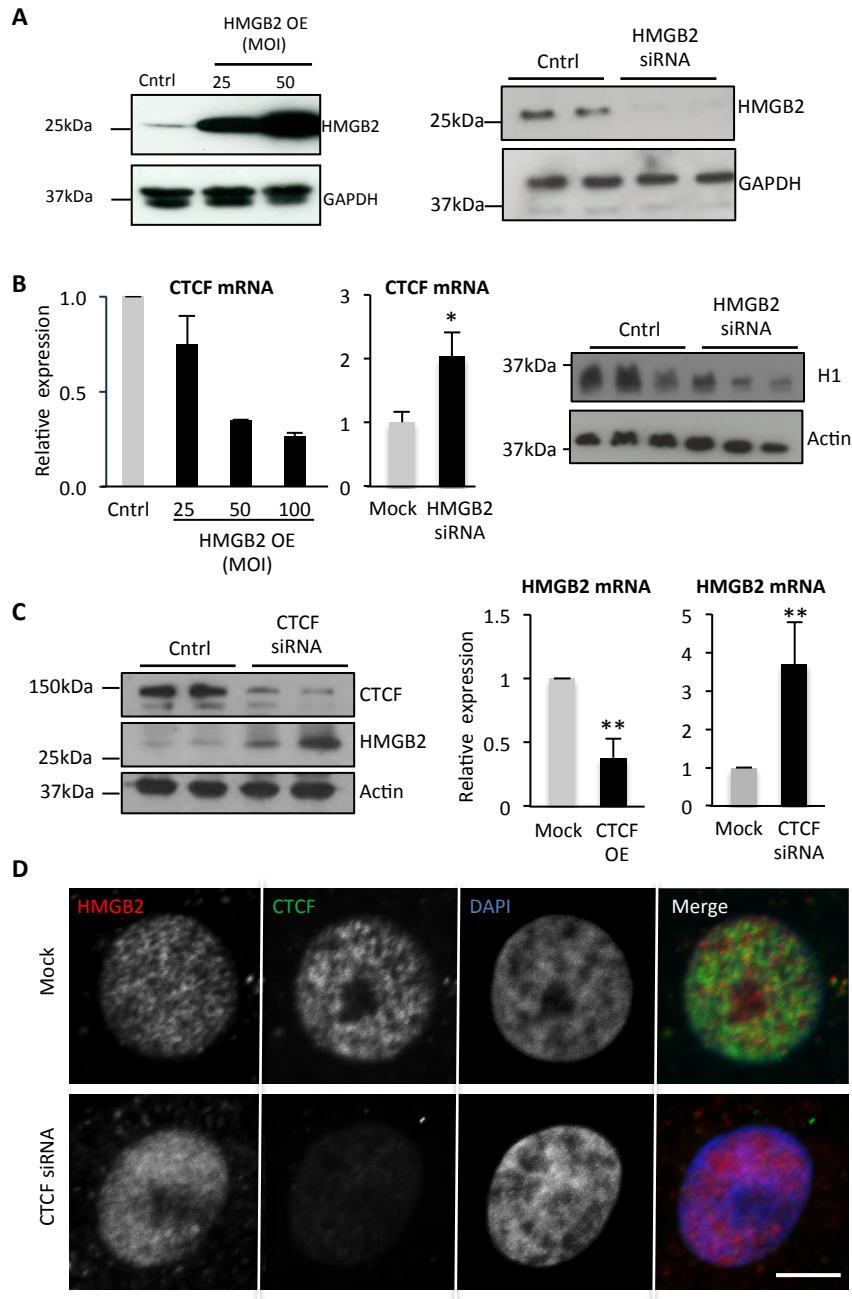


Figure 4-4. HMGB2 and CTCF down-regulate each other. **A.** HMGB2 overexpression (HMGB2 virus for 24hr) and knockdown (HMGB2 siRNA for 72hr) was carried out in NRVMs, and confirmed by Western blotting. **B.** qPCR revealed down-regulation of CTCF with HMGB2 overexpression, and up-regulation of CTCF with HMGB2 knockdown (n=3, * indicates p-value <0.05, error=standard deviation). HMGB2 knockdown no change in histone H1. **C.** Similarly, CTCF knockdown caused an up-regulation of HMGB2 at the protein level, and mRNA level, while CTCF overexpression down-regulated HMGB2 (n=5 overexpression, n=3 knockdown, ** indicates p-value <0.01, error=standard deviation). **D.** Immunolabeling for HMGB2 and CTCF in NRVMs confirmed a decrease in HMGB2 abundance after CTCF knockdown. Scale is 5 μ m.

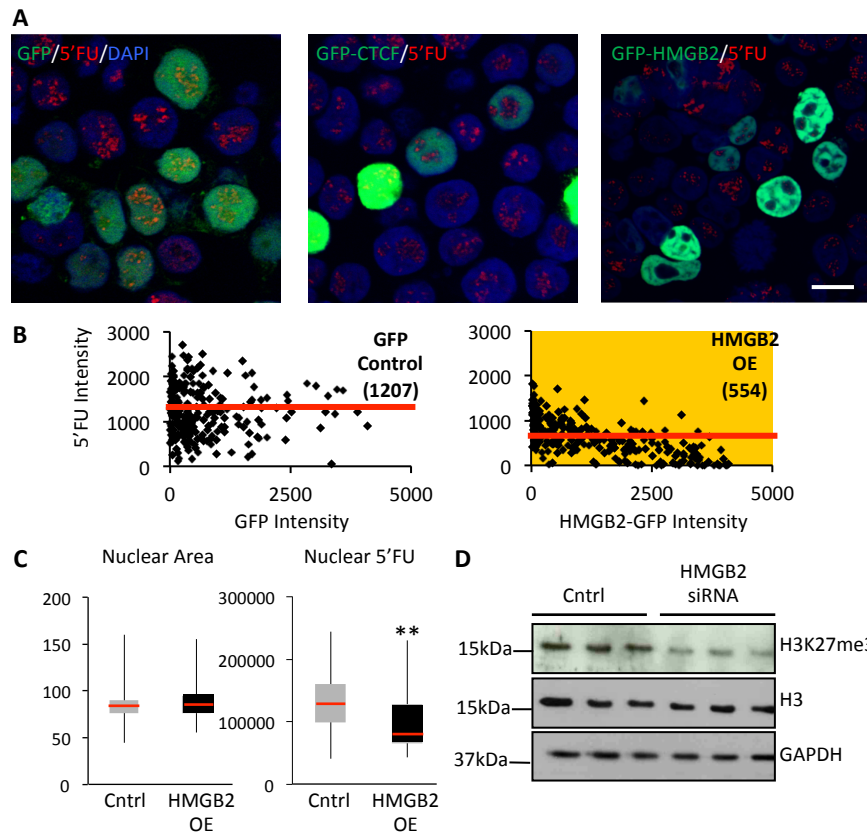
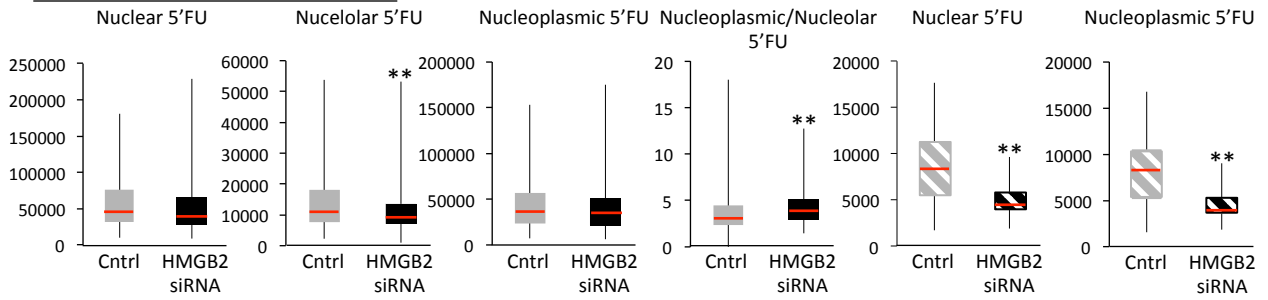
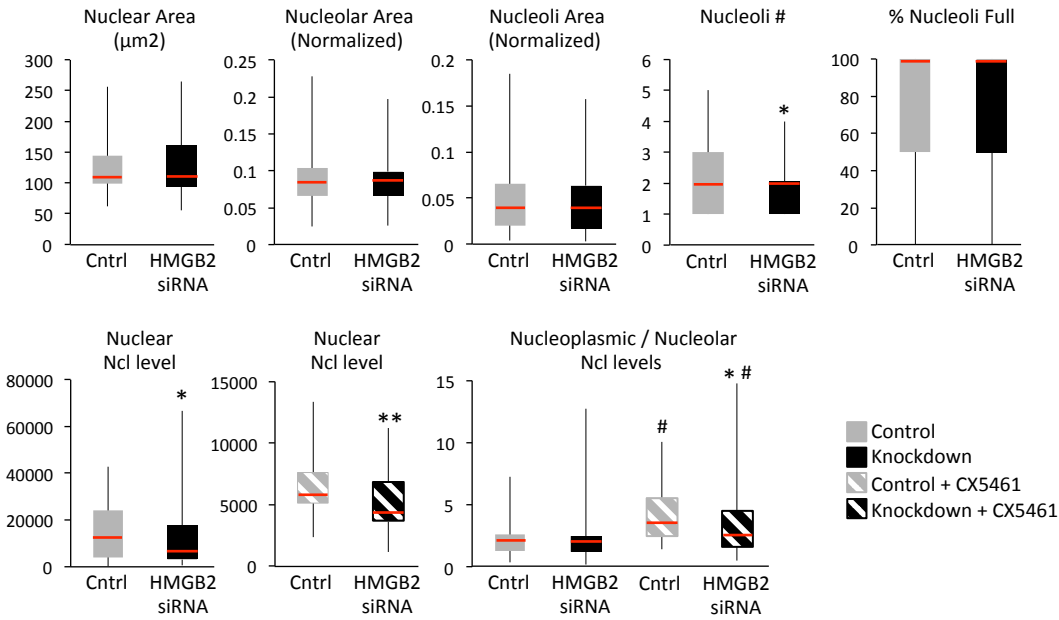


Figure 4-5. HMGB2 overexpression strongly represses transcription. **A.** Control, HMGB2-GFP overexpressing or CTCF-GFP overexpressing 293T cells were treated with 5'fluoruridine (5'FU) to label newly transcribed RNA. HMGB2 overexpressing cells exhibited a dramatic loss of nucleolar transcription (5'FU in red) in the nuclei overexpressing HMGB2 (green cells) that was seen with neither the CTCF overexpression nor GFP overexpression alone. Scale is 10 μ m. **B.** There was a significant (yellow indicates p-value <0.001) inverse relationship between the intensity of HMGB2-GFP in the nuclei, and the intensity of the 5'FU signal, which was not seen in the GFP-only treatment (n=top 250 cells with most GFP intensity; significance also true when compared to all 329 cells measured; median 5'FU intensity indicated by labeled red line). **C.** We repeated these analyses in NRVMs, confirming that HMGB2 overexpression decreases total 5'FU nuclear signal without affecting nuclear size (** indicates p-value <0.0001). **D.** Our previous work suggested HMGB2 knockdown may down-regulated heterochromatic histone post-translational modifications. Western blotting confirmed a decrease in tri-methylation of lysine 27, a mark of facultative heterochromatin, supporting a role for large-scale changes in HMGB2 concentration to modulate global transcriptional levels.

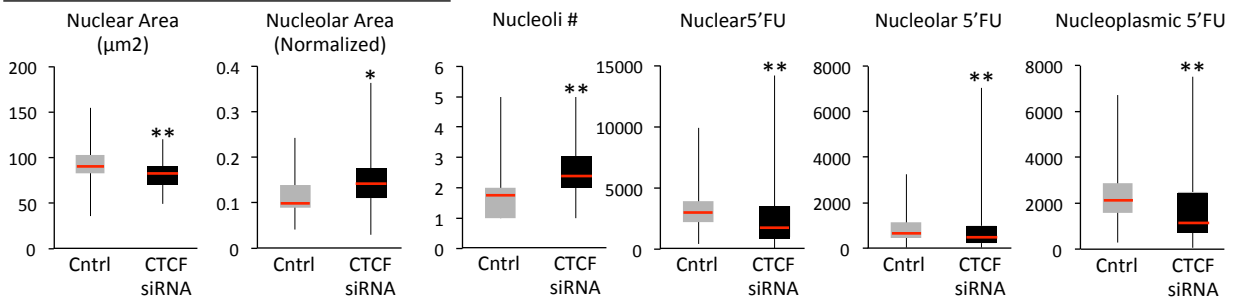
A HMGB2 knockdown – global transcription



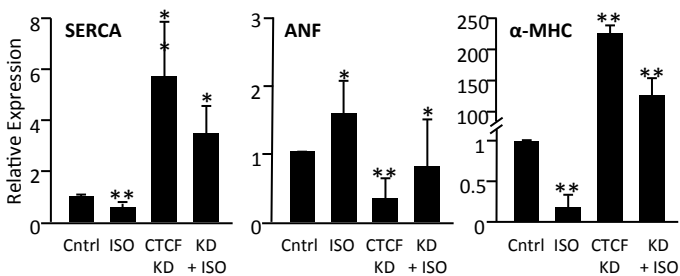
B HMGB2 knockdown - morphology



C CTCF knockdown - morphology, global transcription



D CTCF knockdown - transcription



CTCF overexpression - transcription

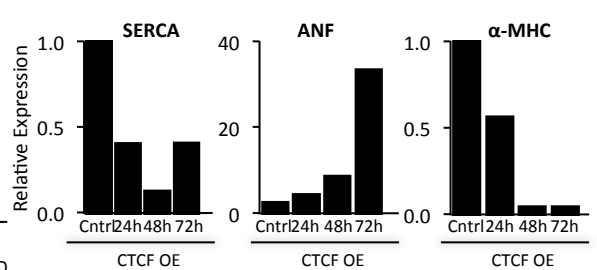


Figure 4-6. HMGB2 and CTCF knockdown down-regulate nucleolar transcription. **A.** Control or HMGB2 knockdown NRVMs were treated with 5'fluoruridine (5'FU) to label newly transcribed RNA. HMGB2 knockdown decreased nucleolar transcription, while increasing the ratio of nucleoplasmic to nucleolar transcription (n=186 control, 181 knockdown). Inhibition of RNA Polymerase I (which transcribes rRNA in the nucleoli) with CX5461 allowed us to more accurately quantify changes to mRNA transcription (n=47 control, n= 47 knockdown). **B.** The effect of HMGB2 on nucleolar transcription could not be explained by changes to nucleolar morphology (% full indicates the percentage of nucleoli within a single nuclei with even Nucleolin staining, as opposed to nucleoli with holes). However we found a decrease in the abundance of Nucleolin levels (* indicates $p<0.05$, ** indicates $p<0.01$, # indicates $p<0.05$ comparing between CX5461 treated or un-treated cells). **C.** CTCF knockdown in NRVMs also decreased nucleolar and total transcription, however, unlike HMGB2, the number of nucleoli/nucleus increased (the opposite was true with HMGB2 knockdown), and total nuclear size decreased (n=264 control, n=287 knockdown, * indicates $p<0.05$, ** indicates $p<0.001$). **D.** qPCR in NRVMs reveals CTCF knockdown attenuates pathologic gene expression induced by ISO, while CTCF overexpression has the opposite effect (* indicates $p<0.05$, ** indicates $p<0.001$, error=standard deviation).

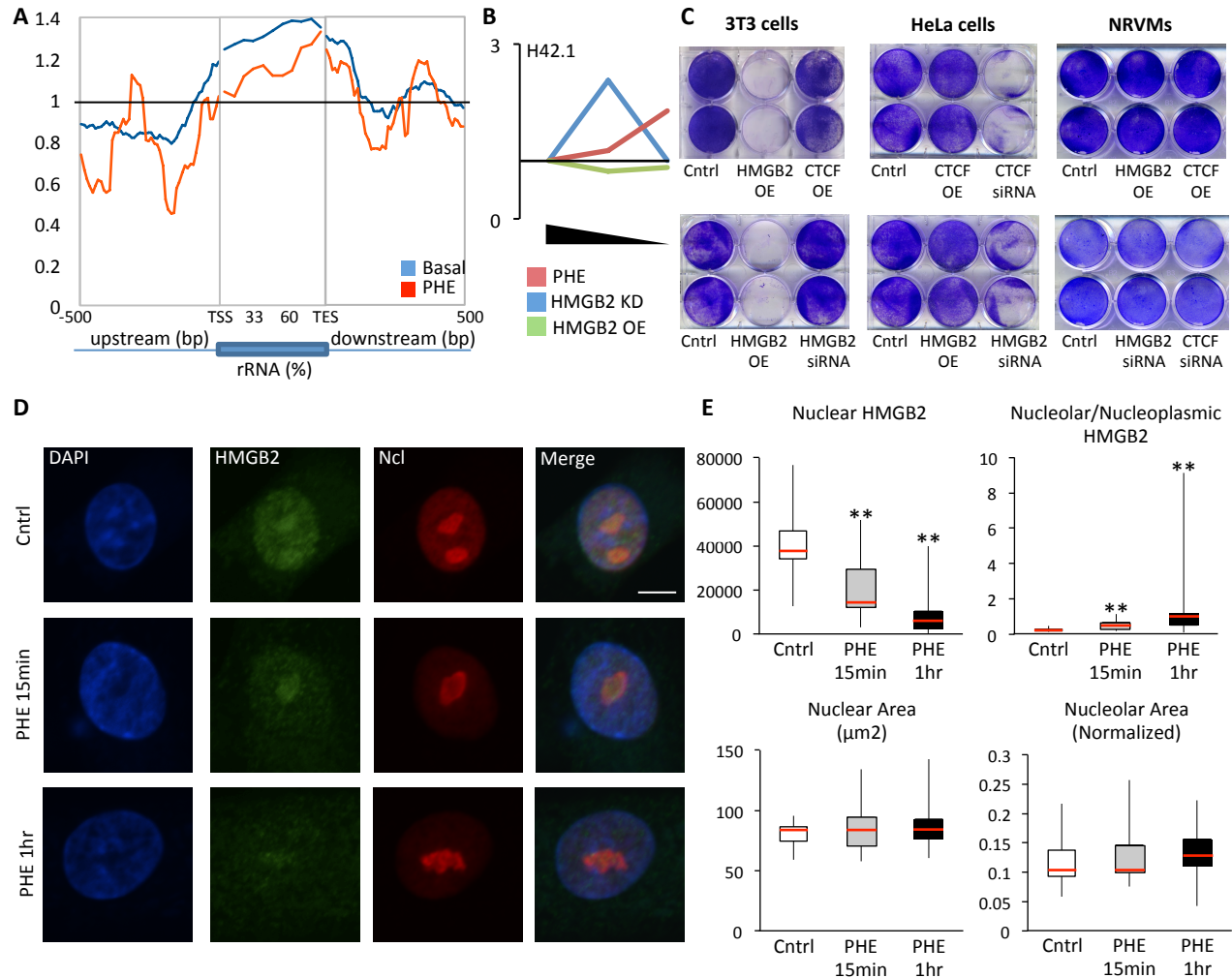


Figure 4-7. HMGB2 binds ribosomal DNA regulating phenotype in a cell-specific manner. **A.** HMGB2 ChIP-seq reads were aligned across ribosomal RNA genes showing HMGB2 is enriched at these loci. With 48 hours phenylephrine treatment, HMGB2 remains enriched, but to a lesser extent. However its localization upstream of rRNA genes increases. **B.** Partial chromatin digestion by micrococcal nuclease was used to isolate euchromatic and heterochromatic DNA in NRVMs followed by qPCR to determine the relative distribution of H42.1, a region of rRNA, between heterochromatic (left end of plot), intermediate (center of plot) and euchromatic (right end of plot) environments. Distribution was normalized by setting the abundance in the most heterochromatic fraction to one. The change in distribution after HMGB2 knockdown or overexpression is plotted as a fraction of the control distribution. In the control setting, the majority of H42.1 sequences were in the most heterochromatic fraction. HMGB2 overexpression had little effect on this (change close to 1 for all three chromatin states), however, HMGB2 knockdown increased the ratio of intermediately-packed to tightly packed DNA, despite transcription decreasing (Figure 4-7A). 48 hours of phenylephrine treatment shifted the rRNA to a more euchromatic environment, which is what we would expect if rRNA transcription increases in hypertrophy. **C.** Cell viability was assayed by labeling with crystal violet, which stains living cells. While the effect of overexpression of HMGB2 was minimal in NRVMs, in dividing 3T3 cells there was a dramatic loss of cells. In contrast, CTCF or HMGB2 knockdown resulted in cell death in HeLa cells. **D.** NRVMs were treated with phenylephrine for 15 minutes or 1 hour and labeled for HMGB2 and Nucleolin. Scale is 5 μm . **E.** Based on preliminary analysis, there is a time-dependent decrease in total HMGB2 levels with phenylephrine (confirming what we previously saw by Western[11]). However, there is also a time-dependent increase in enrichment of the remaining HMGB2 into the nucleoli, with no change in nucleolar area or nuclei size at these time points (n=28 control, n=26 15 minute PHE, n=24 1 hour PHE, ** indicates $p < 0.001$).

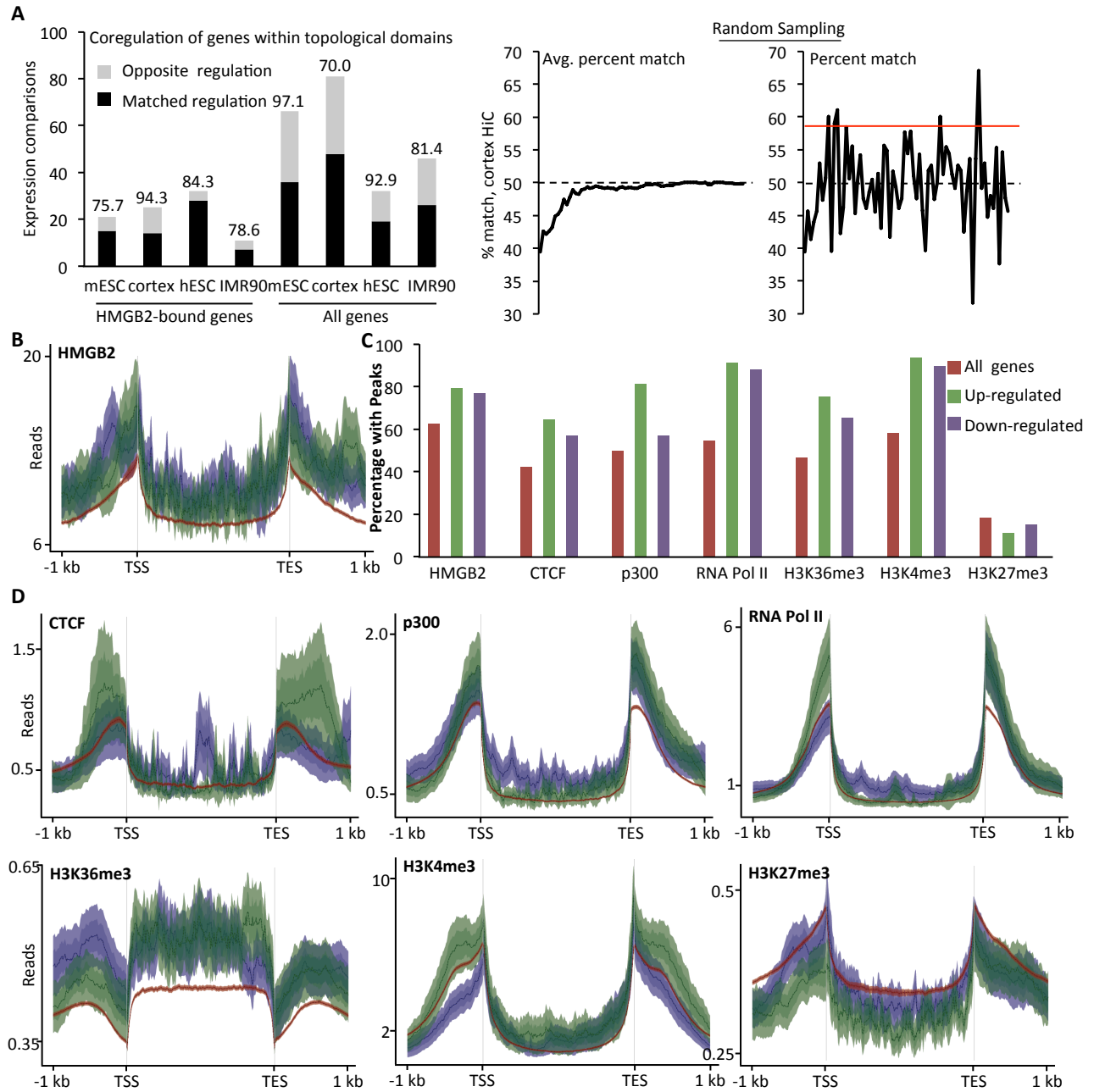


Figure 4-8. Chromatin environment partially explains opposing effects of HMGB2 on transcription. **A.** Genes regulated by HMGB2 (>1.5 fold change with HMGB2 knockdown; or >1.5 fold change and HMGB2 binding peak in promoter) were aligned to topological domains (Hi-C[3]). Genes within the same domain were compared against each other to determine if their expression change matched (both up-regulated or both down-regulated). Percent match of 70 random gene lists of the same size were determined for each Hi-C dataset. Red line indicates actual percent match of our HMGB2-regulated genes. Number over each column indicates percentile of the actual percent match over the random datasets. All random sampling plateaued at an average percent match of ~50%. **B.** HMGB2 reads across all up-regulated or down-regulated genes (>1.5 fold change), or all genes in rn4 genome suggest HMGB2 distribution is not the major determinant of whether a gene's expression will increase or decrease with HMGB2 knockdown. **C.** The percentage of up-regulated, down-regulated (>1.5 fold change) or all genes in the genome with ChIP-seq peaks for chromatin proteins in their gene body (HMGB2 rat[11] and ENCODE data, adult mouse heart). For all chromatin proteins measured, there was a greater percentage of HMGB2-regulated genes bound by the marks than compared to all genes, with the exception of H3K27me3 **D.** Distribution of ChIP-seq reads across HMGB2-regulated genes suggests genes that are up-regulated by knockdown are enriched in the activating marks H3K4me3 and RNA Pol II in the basal heart as compared to genes that are down-regulated by knockdown.

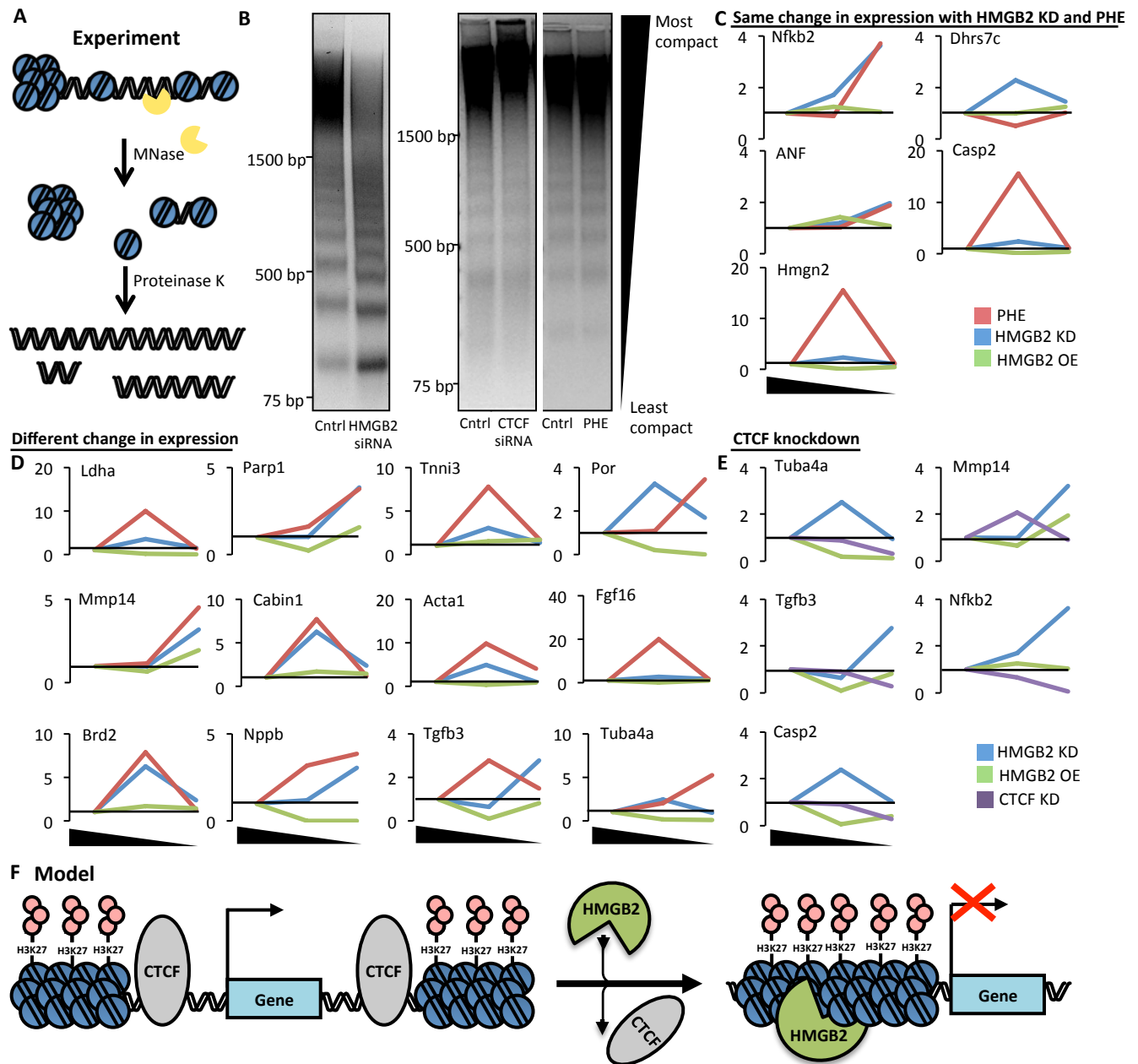
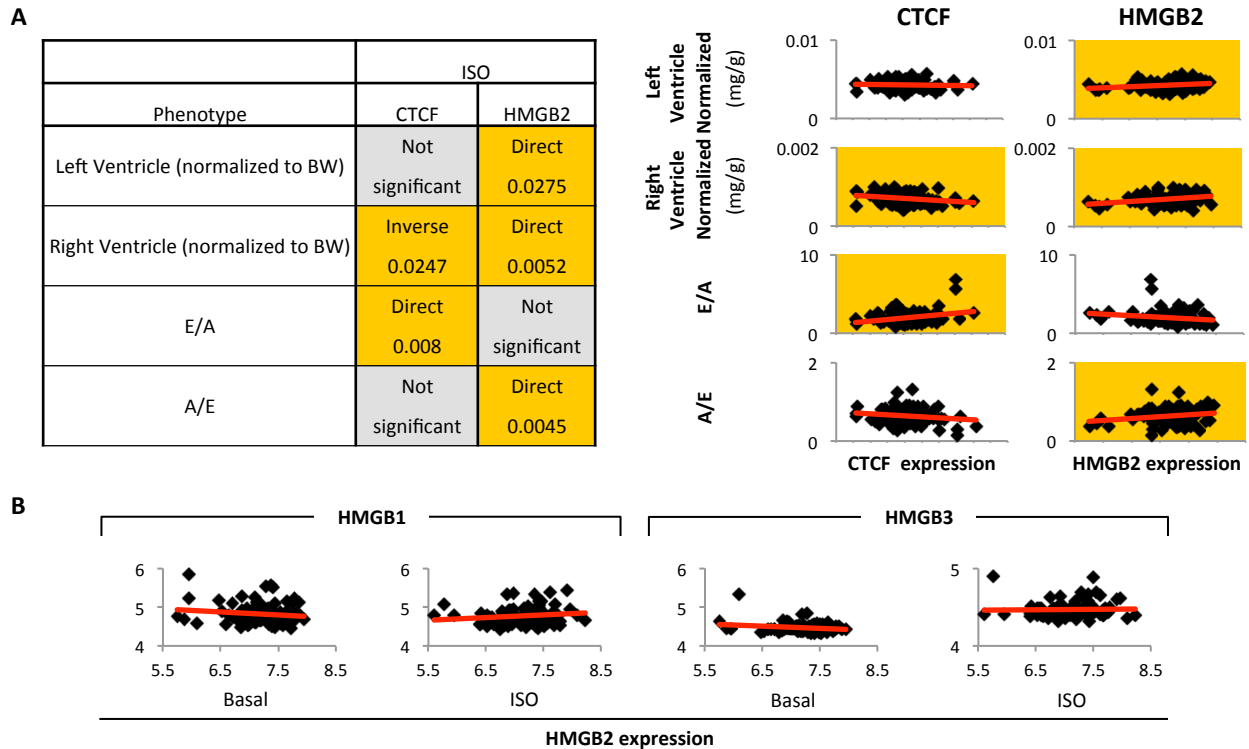
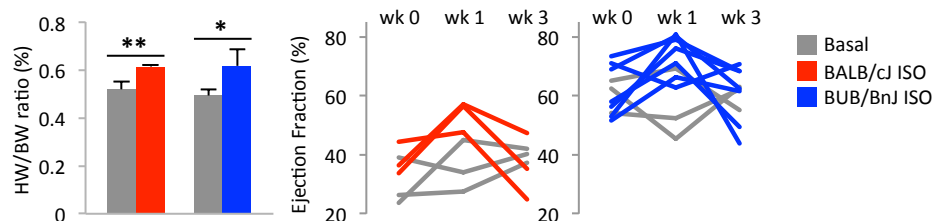


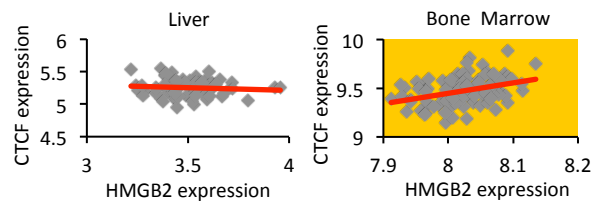
Figure 4-9. HMGB2 and CTCF regulate cardiac gene accessibility and expression. **A.** 0.001U of micrococcal nuclease (MNase) was used to treat control or HMGB2 knockdown NRVMs, resulting in partial digestion of the genome based on accessibility. **B.** HMGB2 knockdown increased the abundance of highly digested (small) fragments of DNA, while decreasing the abundance of less-digested (large) fragments, suggesting a global increase in DNA accessibility. Such dramatic change in accessibility occurred with neither CTCF knockdown nor PHE treatment. **C.** After MNase digestion, DNA fragments were cut from the top, middle, and bottom of the agarose gel and analyzed by qPCR to determine the relative distribution of individual promoter sequences between heterochromatic (large fragments) and euchromatic (small fragments) regions of the genome. Gene expression for HMGB2 knockdown and PHE from published microarray data in NRVMs[11,58] were used to distinguish between genes with similar expression changes induced by HMGB2 and PHE (**C**) or genes regulated differently by HMGB2 and PHE (**D**). Interestingly, HMGB2 knockdown and PHE showed similar trends for shifting promoter sequences between these categories even when they (HMGB2 knockdown or PHE) had different effects on the transcription of the gene. **E.** Promoter sequences for the genes bound by CTCF and HMGB2 by IP were also examined. In three of the five cases, CTCF knockdown and HMGB2 overexpression showed a similar trend to shift these sequences to more compact regions of the genome. **F.** Model for relationship between HMGB2 and CTCF. CTCF serves as a boundary preventing the spread of heterochromatin, while HMGB2 promotes heterochromatin formation.



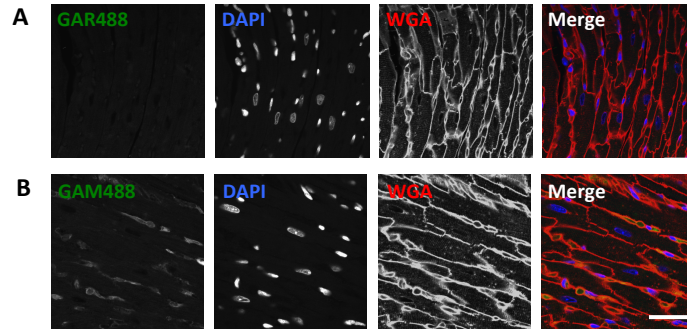
Supplemental Figure 4-1. Basal HMGB2 and CTCF levels predict different disease phenotypes. **A.** Phenotype data from 86 strains[47] of mice after treatment with ISO were compared against CTCF or HMGB2 mRNA expression taken from hearts in the basal condition. Right ventricular mass in disease showed significant negative correlation with basal CTCF expression, while HMGB2 expression had a significant, direct correlation with heart size in these mice (yellow indicates p -value < 0.05). **B.** Unlike CTCF, HMGB1 and HMGB3 show no correlation with HMGB2 expression.



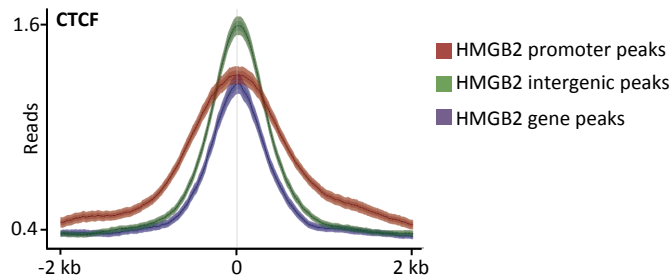
Supplemental Figure 4-2. BALB/cJ and BUB/BnJ respond differently to ISO. BALB/cJ and BUB/BnJ mice were treated with 30mg/kg/day of ISO for one week, and then sacrificed to measure heart weight to body weight ratio ($n=3$ mice/group, * indicates p -value < 0.05 , ** indicates p -value < 0.001 , error=standard deviation). A separate cohort were kept on ISO for 3 weeks, and monitored by echocardiography for ejection fraction. Both strains exhibit increased ejection fraction in the first week. By week 3, two of the three BALB/cJ mice have a lower ejection fraction than controls, while only two of seven BUB/BnJ have dropped below the controls.



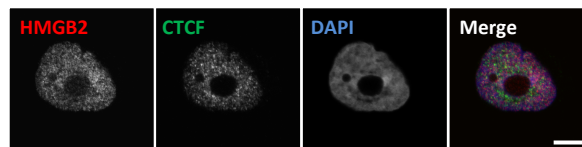
Supplemental Figure 4-3. CTCF and HMGB2 are not coregulated in all tissues. HMGB2 and CTCF expression was analyzed from microarray in the liver (73 mouse strains) or bone marrow (98 strains). Unlike the heart, which showed a significant inverse correlation between HMGB2 and CTCF abundance, the liver showed no correlation, and the bone marrow had a direct correlation (yellow indicates p -value < 0.001).



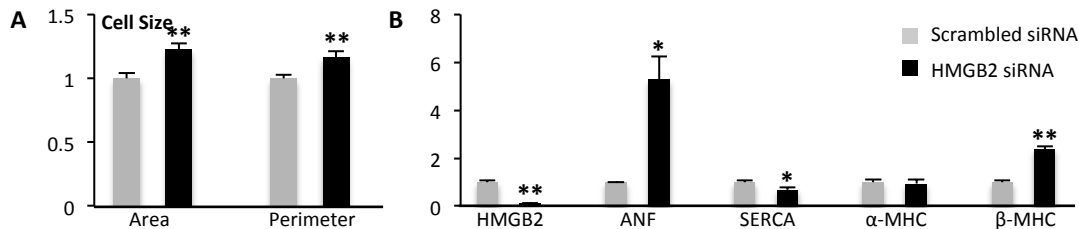
Supplemental Figure 4-4. Immunohistochemistry on cardiac tissue. **A.** As control for the immunohistochemistry experiments in Figure 2F, labeling was performed on heart sections omitting the primary antibody against HMGB2 or CTCF. Secondary for neither goat anti-rabbit nor **B** goat anti-mouse alone produced a signal. Scale is 25 μm .



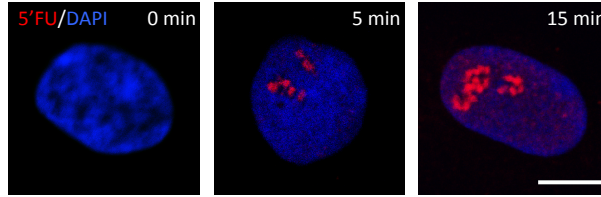
Supplemental Figure 4-5. CTCF enrichment at HMGB2 binding sites most pronounced in intergenic regions. To further dissect the relationship between CTCF and HMGB2 binding, HMGB2 ChIP-seq peaks were subset by whether they fell in gene bodies, promoters (-2kb, +500bp of TSS) or intergenic regions and converted to the mm9 genome using liftOver. CTCF ChIP-seq reads from the adult mouse heart were aligned across HMGB2 ChIP-seq peaks.



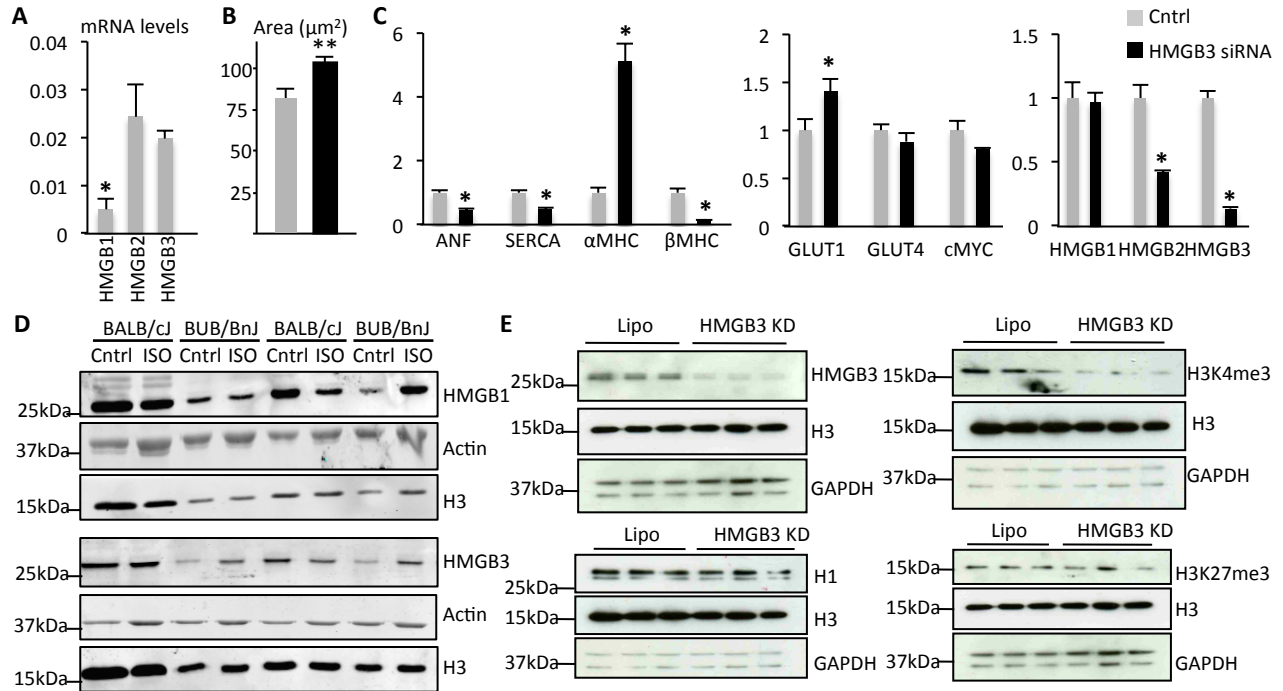
Supplemental Figure 4-6. CTCF and HMGB2 do not colocalize. Immunolabeling for HMGB2 and CTCF in 293T cells confirmed a lack of colocalization of these two proteins (colocalization would appear as yellow), just as seen in NRVMs. Scale is 5 μm .



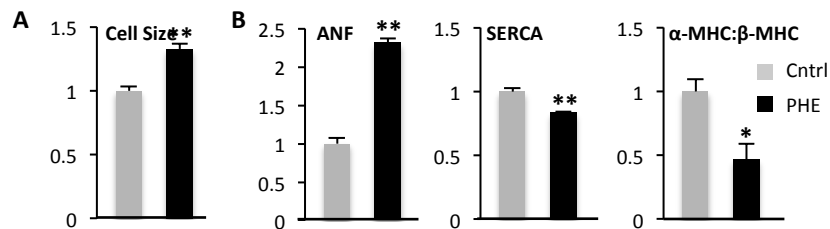
Supplemental Figure 4-7. HMGB2 siRNA control. NRVMs were treated with 50nM scrambled siRNA, or HMGB2 siRNA for 72hrs, using lipofectamine RNAiMAX, to ensure that the scrambled siRNA control gave a similar result to what we show with a lipofectamine-only control. **A.** HMGB2 knockdown cells were significantly larger (n=100 cells/group, ** indicates p-value < 0.01, error=standard error of the mean). **B.** qPCR confirmed activation of the fetal gene program (n=3, ** indicates p-value < 0.01, * indicates p-value < 0.05, error=standard error of the mean).



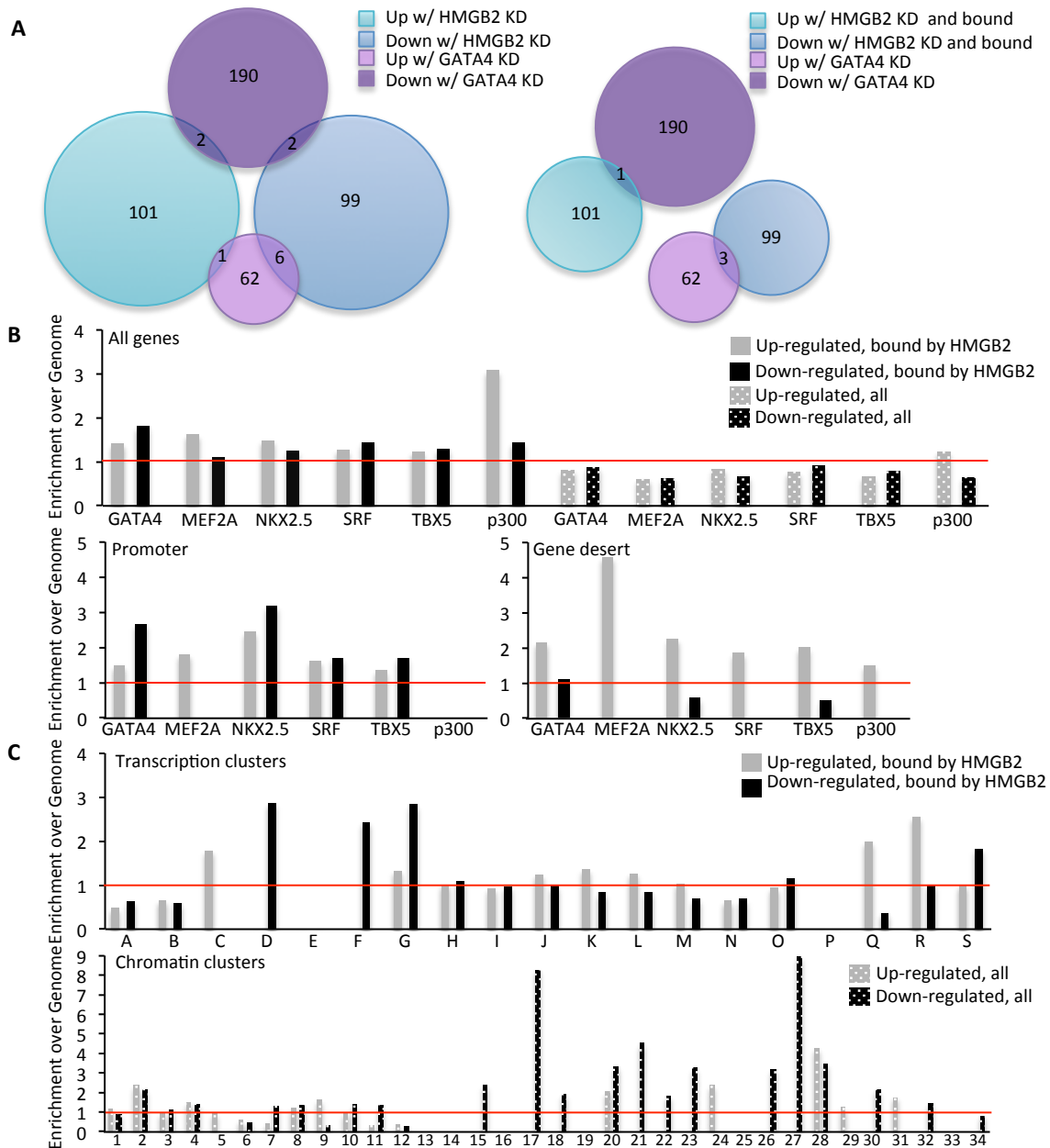
Supplemental Figure 4-8. 5'Fluorouridine labels newly transcribed RNA. NRVMs were treated with 5'fluorouridine (5'FU) to label newly transcribed RNA. We see a time dependent signal for 5'FU labeling. Scale is 5 μ m.



Supplemental Figure 4-9. HMGB3 knockdown induces hypertrophy. **A.** qPCR on control NRVMs showed HMGB1 is significantly less abundant than HMGB2 and HMGB3 (n=3, * indicates p-value < 0.05, error=standard error of the mean). **B.** Like HMGB2, HMGB3 knockdown in NRVMs increases cell size (n=78, ** indicates p-value < 0.01). **C.** However, there was only partial induction of the fetal gene program in the HMGB3 knockdown, as measured by qPCR (n=3, ** indicates p-value < 0.01, * indicates p-value < 0.05). **D.** Like HMGB2, HMGB3 and HMGB1 are down-regulated by ISO in the susceptible, BALB/cJ strain, but up-regulated in the resistant, BUB/BnJ strain. **E.** HMGB3 knockdown had no effect on global levels of Histone H1 or H3K27me3, but moderately down-regulated H3K4me3.



Supplemental Figure 4-10. PHE model of hypertrophy. Control NRVMs, or NRVMs treated with 10 μ M of PHE for 48hrs were analyzed. **A.** PHE-treated cells were significantly larger (n=132 cells/group, ** indicates p-value < 0.01, error=standard error of the mean). **B.** qPCR confirmed activation of the fetal gene program (n=3, ** indicates p-value < 0.01, * indicates p-value < 0.05, error=standard error of the mean).



Supplemental Figure 4-11. The effect of HMGB2 knockdown on gene expression does not discriminate between genes regulated by different cardiac transcription factors. **A.** We compared our published microarray data for HMGB2 knockdown in NRVMs to microarray data for GATA4 knockdown in HL1 cells[53], revealing minimal overlap between genes regulated by these two proteins (**top panel**). We repeated this comparison by limiting the group of genes that changed in expression with HMGB2 knockdown, to only those genes with HMGB2 binding peaks in their promoter (ChIP-seq, NRVMs), and still saw minimal overlap (**bottom panel**). **B.** Published transcription factor binding peaks in HL1 cells[53] listed the nearest gene to each peak. We asked if HMGB2-regulated genes were enriched in these lists and found minimal enrichment. We then narrowed our analysis to only genes with transcription factor peaks within 2kb upstream of the TSS (promoter group), and as control to genes identified due to a peak over 100kb away (gene desert group). HMGB2 targets are enriched for cardiac transcription factors in their promoter, with no difference between up-regulated and down-regulated targets. **C.** We next analyzed published gene clusters[54], grouped based on their similar expression (**top panel**) or chromatin signature (**bottom panel**) during cardiomyocyte differentiation. HMGB2 is found in transcription cluster I. We found minimal enrichment for HMGB2-regulated genes in any transcriptional or chromatin clusters. Together, this suggested that HMGB2 shares targets with cardiac transcription factors but that HMGB2's role to up or down-regulate genes may be through other chromatin cues.

Chapter 4: References

1. Lieberman-Aiden, E., et al., *Comprehensive mapping of long-range interactions reveals folding principles of the human genome*. Science, 2009. **326**(5950): p. 289-93.
2. Simonis, M., et al., *Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C)*. Nature genetics, 2006. **38**(11): p. 1348-54.
3. Dixon, J.R., et al., *Topological domains in mammalian genomes identified by analysis of chromatin interactions*. Nature, 2012. **485**(7398): p. 376-80.
4. Cuddapah, S., et al., *Genomic profiling of HMGN1 reveals an association with chromatin at regulatory regions*. Molecular and cellular biology, 2011. **31**(4): p. 700-9.
5. Kim, T.H., et al., *Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome*. Cell, 2007. **128**(6): p. 1231-45.
6. Olson, E.N., J. Backs, and T.A. McKinsey, *Control of cardiac hypertrophy and heart failure by histone acetylation/deacetylation*. Novartis Foundation symposium, 2006. **274**: p. 3-12; discussion 13-9, 152-5, 272-6.
7. Papait, R., et al., *Genome-wide analysis of histone marks identifying an epigenetic signature of promoters and enhancers underlying cardiac hypertrophy*. Proceedings of the National Academy of Sciences of the United States of America, 2013. **110**(50): p. 20164-9.
8. Strahl, B.D. and C.D. Allis, *The language of covalent histone modifications*. Nature, 2000. **403**(6765): p. 41-5.
9. Nakayama, J., et al., *Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly*. Science, 2001. **292**(5514): p. 110-3.
10. Guenther, M.G., et al., *A chromatin landmark and transcription initiation at most promoters in human cells*. Cell, 2007. **130**(1): p. 77-88.

11. Franklin, S., et al., *Quantitative analysis of the chromatin proteome in disease reveals remodeling principles and identifies high mobility group protein B2 as a regulator of hypertrophic growth*. Molecular & cellular proteomics : MCP, 2012. **11**(6): p. M111014258.
12. Gaikwad, A.B., et al., *Renal failure increases cardiac histone h3 acetylation, dimethylation, and phosphorylation and the induction of cardiomyopathy-related genes in type 2 diabetes*. The American journal of pathology, 2010. **176**(3): p. 1079-83.
13. Lu, J.T., et al., *LMNA cardiomyopathy: cell biology and genetics meet clinical medicine*. Disease models & mechanisms, 2011. **4**(5): p. 562-8.
14. Furusawa, T., et al., *Chromatin decompaction by the nucleosomal binding protein HMGN5 impairs nuclear sturdiness*. Nature communications, 2015. **6**: p. 6138.
15. Bell, A.C., A.G. West, and G. Felsenfeld, *The protein CTCF is required for the enhancer blocking activity of vertebrate insulators*. Cell, 1999. **98**(3): p. 387-96.
16. Handoko, L., et al., *CTCF-mediated functional chromatin interactome in pluripotent cells*. Nature genetics, 2011. **43**(7): p. 630-8.
17. Cuddapah, S., et al., *Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains*. Genome research, 2009. **19**(1): p. 24-32.
18. Agresti, A. and M.E. Bianchi, *HMGB proteins and gene expression*. Current opinion in genetics & development, 2003. **13**(2): p. 170-8.
19. Goodwin, G.H., C. Sanders, and E.W. Johns, *A new group of chromatin-associated proteins with a high content of acidic and basic amino acids*. European journal of biochemistry / FEBS, 1973. **38**(1): p. 14-9.
20. Travers, A.A., *Priming the nucleosome: a role for HMGB proteins?* EMBO reports, 2003. **4**(2): p. 131-6.

21. Lange, S.S. and K.M. Vasquez, *HMGB1: the jack-of-all-trades protein is a master DNA repair mechanic*. *Molecular carcinogenesis*, 2009. **48**(7): p. 571-80.
22. Kwon, J., et al., *Accessibility of nucleosomal DNA to V(D)J cleavage is modulated by RSS positioning and HMG1*. *Molecular cell*, 1998. **2**(6): p. 829-39.
23. Gabellini, D., M.R. Green, and R. Tupler, *Inappropriate gene activation in FSHD: a repressor complex binds a chromosomal repeat deleted in dystrophic muscle*. *Cell*, 2002. **110**(3): p. 339-48.
24. Pallier, C., et al., *Association of chromatin proteins high mobility group box (HMGB) 1 and HMGB2 with mitotic chromosomes*. *Molecular biology of the cell*, 2003. **14**(8): p. 3414-26.
25. Phair, R.D., et al., *Global nature of dynamic protein-chromatin interactions in vivo: three-dimensional genome scanning and dynamic interaction networks of chromatin proteins*. *Molecular and cellular biology*, 2004. **24**(14): p. 6393-402.
26. Agresti, A., et al., *GR and HMGB1 interact only within chromatin and influence each other's residence time*. *Molecular cell*, 2005. **18**(1): p. 109-21.
27. Bonaldi, T., et al., *The DNA chaperone HMGB1 facilitates ACF/CHRAC-dependent nucleosome sliding*. *The EMBO journal*, 2002. **21**(24): p. 6865-73.
28. Zhang, J., et al., *Mechanism of DNA flexibility enhancement by HMGB proteins*. *Nucleic acids research*, 2009. **37**(4): p. 1107-14.
29. Ner, S.S. and A.A. Travers, *HMG-D, the Drosophila melanogaster homologue of HMG 1 protein, is associated with early embryonic chromatin in the absence of histone H1*. *The EMBO journal*, 1994. **13**(8): p. 1817-22.
30. Cato, L., et al., *The interaction of HMGB1 and linker histones occurs through their acidic and basic tails*. *Journal of molecular biology*, 2008. **384**(5): p. 1262-72.
31. Lu, J., R. Kobayashi, and S.J. Brill, *Characterization of a high mobility group 1/2 homolog in yeast*. *The Journal of biological chemistry*, 1996. **271**(52): p. 33678-85.

32. Fedele, M., et al., *Haploinsufficiency of the Hmga1 gene causes cardiac hypertrophy and myelo-lymphoproliferative disorders in mice*. *Cancer research*, 2006. **66**(5): p. 2536-43.
33. Martinez Hoyos, J., et al., *Identification of the genes up- and down-regulated by the high mobility group A1 (HMGA1) proteins: tissue specificity of the HMGA1-dependent gene regulation*. *Cancer research*, 2004. **64**(16): p. 5728-35.
34. Calogero, S., et al., *The lack of chromosomal protein Hmg1 does not disrupt cell growth but causes lethal hypoglycaemia in newborn mice*. *Nature genetics*, 1999. **22**(3): p. 276-80.
35. Ronfani, L., et al., *Reduced fertility and spermatogenesis defects in mice lacking chromosomal protein Hmgb2*. *Development*, 2001. **128**(8): p. 1265-73.
36. Nemeth, M.J., M.R. Kirby, and D.M. Bodine, *Hmgb3 regulates the balance between hematopoietic stem cell self-renewal and differentiation*. *Proceedings of the National Academy of Sciences of the United States of America*, 2006. **103**(37): p. 13783-8.
37. Nemeth, M.J., et al., *Hmgb3 deficiency deregulates proliferation and differentiation of common lymphoid and myeloid progenitors*. *Blood*, 2005. **105**(2): p. 627-34.
38. Catena, R., et al., *HMGB4, a novel member of the HMGB family, is preferentially expressed in the mouse testis and localizes to the basal pole of elongating spermatids*. *Biology of reproduction*, 2009. **80**(2): p. 358-66.
39. de Souza, A.W., et al., *HMGB1 in vascular diseases: Its role in vascular inflammation and atherosclerosis*. *Autoimmunity reviews*, 2012. **11**(12): p. 909-17.
40. Andrassy, M., et al., *HMGB1 as a predictor of infarct transmuralty and functional recovery in patients with myocardial infarction*. *Journal of internal medicine*, 2011.
41. Schlueter, C., et al., *Angiogenetic signaling through hypoxia: HMGB1: an angiogenetic switch molecule*. *The American journal of pathology*, 2005. **166**(4): p. 1259-63.

42. Limana, F., et al., *Exogenous high-mobility group box 1 protein induces myocardial regeneration after infarction via enhanced cardiac C-kit+ cell proliferation and differentiation*. *Circulation research*, 2005. **97**(8): p. e73-83.
43. Rossini, A., et al., *HMGB1-stimulated human primary cardiac fibroblasts exert a paracrine action on human and murine cardiac stem cells*. *Journal of molecular and cellular cardiology*, 2008. **44**(4): p. 683-93.
44. Schweitzer, N.B., et al., *Access to exercise and its relation to cardiovascular health and gene expression in laboratory animals*. *Life sciences*, 2005. **77**(18): p. 2246-61.
45. Seyedin, S.M. and W.S. Kistler, *Levels of chromosomal protein high mobility group 2 parallel the proliferative activity of testis, skeletal muscle, and other organs*. *The Journal of biological chemistry*, 1979. **254**(22): p. 11264-71.
46. Yamazaki, F., et al., *Repression of cell cycle progression by antisense HMG2 RNA*. *Biochemical and biophysical research communications*, 1995. **210**(3): p. 1045-51.
47. Rau, C.D., et al., *Mapping genetic contributions to cardiac pathology induced by Beta-adrenergic stimulation in mice*. *Circulation. Cardiovascular genetics*, 2015. **8**(1): p. 40-9.
48. Mitchell-Jordan, S., et al., *Features of endogenous cardiomyocyte chromatin revealed by super-resolution STED microscopy*. *Journal of molecular and cellular cardiology*, 2012. **53**(4): p. 552-8.
49. Franklin, S., et al., *Specialized compartments of cardiac nuclei exhibit distinct proteomic anatomy*. *Molecular & cellular proteomics : MCP*, 2011. **10**(1): p. M110 000703.
50. Langmead, B., et al., *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome*. *Genome Biol*, 2009. **10**(3): p. R25.
51. Zhang, Y., et al., *Model-based analysis of ChIP-Seq (MACS)*. *Genome biology*, 2008. **9**(9): p. R137.
52. Schmidt, D., et al., *Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages*. *Cell*, 2012. **148**(1-2): p. 335-48.

53. He, A., et al., *Co-occupancy by multiple cardiac transcription factors identifies transcriptional enhancers active in heart*. Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**(14): p. 5632-7.
54. Wamstad, J.A., et al., *Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage*. Cell, 2012. **151**(1): p. 206-20.
55. Ubil, E., et al., *Mesenchymal-endothelial transition contributes to cardiac neovascularization*. Nature, 2014. **514**(7524): p. 585-90.
56. van de Nobelen, S., et al., *CTCF regulates the local epigenetic state of ribosomal DNA repeats*. Epigenetics & chromatin, 2010. **3**(1): p. 19.
57. Rosa-Garrido, M., et al., *A cell cycle role for the epigenetic factor CTCF-L/BORIS*. PLoS one, 2012. **7**(6): p. e39371.
58. Bush, E., et al., *A small molecular activator of cardiac hypertrophy uncovered in a chemical screen for modifiers of the calcineurin signaling pathway*. Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(9): p. 2870-5.
59. Yeager, J.C. and S.G. Iams, *Isoproterenol-induced cardiac failure in the spontaneously hypertensive rat*. Proceedings of the Society for Experimental Biology and Medicine. Society for Experimental Biology and Medicine, 1981. **168**(1): p. 137-42.
60. Ghazalpour, A., et al., *Genetic regulation of mouse liver metabolite levels*. Molecular systems biology, 2014. **10**: p. 730.
61. Farber, C.R., et al., *Mouse genome-wide association and systems genetics identify *Asxl2* as a regulator of bone mineral density and osteoclastogenesis*. PLoS genetics, 2011. **7**(4): p. e1002038.
62. Strahl, B.D., et al., *Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in *Tetrahymena**. Proc Natl Acad Sci U S A, 1999. **96**(26): p. 14967-72.

63. Koch, C.M., et al., *The landscape of histone modifications across 1% of the human genome in five human cell lines*. Genome research, 2007. **17**(6): p. 691-707.
64. Cao, R., et al., *Role of histone H3 lysine 27 methylation in Polycomb-group silencing*. Science, 2002. **298**(5595): p. 1039-43.
65. Grummt, I., *Life on a planet of its own: regulation of RNA polymerase I transcription in the nucleolus*. Genes & development, 2003. **17**(14): p. 1691-702.
66. Raska, I., et al., *The nucleolus and transcription of ribosomal genes*. Biology of the cell / under the auspices of the European Cell Biology Organization, 2004. **96**(8): p. 579-94.
67. Fedoriw, A.M., et al., *Transgenic RNAi reveals essential function for CTCF in H19 gene imprinting*. Science, 2004. **303**(5655): p. 238-40.
68. Moore, J.M., et al., *Loss of maternal CTCF is associated with peri-implantation lethality of Cctf null embryos*. PloS one, 2012. **7**(4): p. e34915.
69. Fiorentino, F.P. and A. Giordano, *The tumor suppressor role of CTCF*. Journal of cellular physiology, 2012. **227**(2): p. 479-92.
70. Brandenburger, Y., et al., *Cardiac hypertrophy in vivo is associated with increased expression of the ribosomal gene transcription factor UBF*. FEBS letters, 2003. **548**(1-3): p. 79-84.
71. Zuin, J., et al., *Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells*. Proceedings of the National Academy of Sciences of the United States of America, 2014. **111**(3): p. 996-1001.
72. Stros, M., et al., *HMGB1 and HMGB2 cell-specifically down-regulate the p53- and p73-dependent sequence-specific transactivation from the human Bax gene promoter*. The Journal of biological chemistry, 2002. **277**(9): p. 7157-64.
73. Zwilling, S., H. Konig, and T. Wirth, *High mobility group protein 2 functionally interacts with the POU domains of octamer transcription factors*. The EMBO journal, 1995. **14**(6): p. 1198-208.

74. Lehming, N., et al., *Chromatin components as part of a putative transcriptional repressing complex*. Proceedings of the National Academy of Sciences of the United States of America, 1998. **95**(13): p. 7322-6.
75. Hoppe, G., M.E. Rayborn, and J.E. Sears, *Diurnal rhythm of the chromatin protein Hmgb1 in rat photoreceptors is under circadian regulation*. The Journal of comparative neurology, 2007. **501**(2): p. 219-30.
76. Celona, B., et al., *Substantial histone reduction modulates genomewide nucleosomal occupancy and global transcriptional output*. PLoS biology, 2011. **9**(6): p. e1001086.
77. Bonne-Andrea, C., et al., *Rat liver HMG1: a physiological nucleosome assembly factor*. The EMBO journal, 1984. **3**(5): p. 1193-9.
78. Ugrinova, I., I.G. Pashev, and E.A. Pasheva, *Nucleosome binding properties and Co-remodeling activities of native and in vivo acetylated HMGB-1 and HMGB-2 proteins*. Biochemistry, 2009. **48**(27): p. 6502-7.
79. Phillips, J.E. and V.G. Corces, *CTCF: master weaver of the genome*. Cell, 2009. **137**(7): p. 1194-211.
80. Avitabile, D., et al., *Nucleolar stress is an early response to myocardial damage involving nucleolar proteins nucleostemin and nucleophosmin*. Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**(15): p. 6145-50.
81. Jarboui, M.A., et al., *Proteomic profiling of the human T-cell nucleolus*. Molecular immunology, 2011. **49**(3): p. 441-52.

Future Directions

One of the intriguing observations made by Sarah Franklin in the lab before I started on the project was that knockdown of HMGB2 in neonatal rat myocytes caused a global increase in H3K4me3 and a decrease in H3K9me3[1]. This observation fit with our hypothesis that the genome should be more plastic in the diseased state to allow for transcriptional reprogramming. We followed up on the implications of this by measuring global levels of transcription after HMGB2 knockdown. However, we never focused on identifying the mechanism for HMGB2 to regulate histone modifications. In fact, this would have been one of the few roles of HMGB2 well suited for dissection by co-immunoprecipitation experiments.

Analysis of individual genes demonstrated a locus-dependent effect of HMGB2 on chromatin and expression. Thus, interpretation of co-immunoprecipitation data for protein interactors of HMGB2 would be limited by not knowing the location of the complexes (that is: which binding partners interact with HMGB2 at which genomic loci), which is why we relied on ChIP-seq data for these questions. However, because the changes in H3K4me3 and H3K9me3 are widespread, there is less of a need for localization information when determining the proteins HMGB2 interacts with to cause this outcome. Co-immunoprecipitation experiments could be used to test if HMGB2 directly interacts with proteins responsible for adding or removing histone methylation. Such experiments should be done. Yet the experiment may not be successful: HMGB2 knockdown may induce the same phenotype by instead interfering with the boundaries of heterochromatin, which we began to examine in the context of the relationship between HMGB2 and CTCF in Chapter 4.

We still have not tested the hypothesis posed by these observations, which is that the genome is more plastic in cardiac hypertrophy. Specifically, this hypothesis asks if regions of DNA whose expression does not change with hypertrophy also undergo chromatin remodeling which makes them easier to activate. Secondly, do these changes precede transcriptional changes at the loci where transcription is different in disease (are chromatin changes the cause

or consequence of transcription changes)? We have most of the data we need to answer this by incorporating published RNA-seq and CHIP-seq datasets with our lab's analysis of chromatin structural proteins, DNA methylation, and nucleosome-free regulatory regions. I could begin these analyses now by testing for correlation between local chromatin features and gene expression. I would need a list of chromatin features that define "accessible/amenable to transcription" regions, ideally exploring several iterations of these lists by varying inclusion criteria. This is not a new concept. Others have used combinatorial CHIP-seq, genome annotation, and transcription to define chromatin states[2, 3]. The only difference is that we would be looking for the occurrence of chromatin state changes that should favor transcription but were independent of transcription. Another way: does chromatin structure serve any purpose other than transcription?

The crux of this hypothesis is the assumption that the nucleus in the basal state was perfectly chromatinized such that only genes that were supposed to be on could be on, and therefore changes to transcription require a change in said pattern of chromatinization. We know that the differentiation process involves cell type-specific chromatin silencing by H3K9me3 in the heart[4]. Further, akin to HMGB2 knockdown, the pressure overload-induced mouse model of heart failure is accompanied by a decrease in H3K9me3 (a repressive mark) with a concomitant increase in H3K9me3 (an activating mark) [1]. Together, these two observations support the concept that there must be an undoing of the basal chromatin to remodel the transcriptome for disease. This hypothesis also speculates that these changes involve higher-order structure and a loosening of chromatin, based on the assumption that sequence independent loosening would expose DNA binding motifs and therefore proceed sequence-specific remodeling. If this hypothesis is correct, the analysis I describe above would identify widespread chromatin changes associated with active transcription that proceed some of the transcriptional changes in disease. However, it is possible that the hypothesis is wrong, and chromatin remodeling within genes is reflective of the transcription of the gene, as seen in one

analysis of ChIP-seq performed on myocytes before and after transverse aortic constriction[5]. Importantly, this study does not analyze changes to intergenic regions, which is where we could most easily test this hypothesis without the variable of transcription. Yet it is interesting to note that network analysis suggests co-regulation of genes is tightened under disease stimuli in the heart (unpublished data from the Christoph Rau in the lab of Jake Lusis). To me, this suggests either A) a more restrictive chromatin environment that decreases variability, or B) an equally permissive chromatin environment that is being acted on by transcription factors that have become less promiscuous due to pathologic stress. Both of these possibilities would go against a more restrictive basal state, and still allow for the local chromatin and transcriptional observations we have made. Yet neither explain the large decrease in H3K9me3 in disease. To determine a definition of “plastic” and test whether the level of the plasticity of the genome is changing, genome-wide chromatin measurements from multiple time-points will need to be measured and analyzed within the context of transcription-dependent chromatin changes, and transcription-independent chromatin changes. As stated previously when interpreting a similar switch from H3K9me3 to H3K4me3 in the HMGB2 knockdown myocytes, one potential mechanism by which H3K9me3 levels could be altered in disease is via a disruption to the boundaries separating hetero- and euchromatin.

Chromatin conformation capture utilizing next-generation sequencing has been used to map megabase-scale domains of DNA with a high-level of self-interaction that segregate hetero- and euchromatin, and which are conserved between mouse and human and between stem cells and differentiated cells[6]. This scaffold is orchestrated in part by sequence-specific binding of CTCF, a chromatin structural protein.

The conservation of these domains has three ramifications to our understanding of proteins like HMGB2, which lack DNA recognition motifs, and are expressed across cell types. Firstly, like with H3K9me3 spreading, topological associated domains (TADs) could support

different concentrations of chromatin structural proteins, thereby influencing their targeting or their available interacting partners.

Secondly, the TADs do not function as enlarged, modified operons or transcription factories; rather, genes within TADs can have different expression profiles. However, housekeeping genes and marks of active transcription are enriched at the boundaries. We already knew that three-dimensional DNA structure was important, with examples such as DNA loops allowing enhancers to regulate their target gene. Yet, here we see genome organization operating at a scale above regulating individual genes. This opens a new role by which chromatin structural proteins can modulate the transcriptome. Additionally it provides a cut-off for the scale at which cell-type specific chromatin features operate and the higher scale at which cell-type independent chromatin features exist. However, the functional role of TADs is still an open question

Thirdly, these data suggest, although they do not prove, trans effects of chromatin structure on gene expression. By defining interacting domains, the chromatin conformation capture technique also provides all of the DNA regions with which a locus does not interact. We can then test if chromatin in one TAD can effect the expression of genes in another TAD, and if we find that there is cross-regulation we know it occurs in the absence of physical interaction. If we find evidence for such a relationship, it would open a new mechanism by which intergenic chromatin could regulate gene expression. Our own work provides preliminary evidence that trans chromatin effects could exist. Knockdown of HMGB2 alters the expression of some genes without changing the chromatin accessibility at their promoter, indicating that expression is being regulated by exterior factors. We know that globally there is a change in the accessibility of the genome under HMGB2 knockdown conditions. The hypothesis would be that altering accessibility of one region could affect the expression of a distant region by changing the properties of the genome.

Future Directions: References

1. Franklin, S., et al., *Quantitative analysis of the chromatin proteome in disease reveals remodeling principles and identifies high mobility group protein B2 as a regulator of hypertrophic growth*. Molecular & cellular proteomics : MCP, 2012. **11**(6): p. M111014258.
2. Ernst, J., et al., *Mapping and analysis of chromatin state dynamics in nine human cell types*. Nature, 2011. **473**(7345): p. 43-9.
3. Filion, G.J., et al., *Systematic protein location mapping reveals five principal chromatin types in Drosophila cells*. Cell, 2010. **143**(2): p. 212-24.
4. Sdek, P., et al., *Rb and p130 control cell cycle gene silencing to maintain the postmitotic phenotype in cardiac myocytes*. The Journal of cell biology, 2011. **194**(3): p. 407-23.
5. Papait, R., et al., *Genome-wide analysis of histone marks identifying an epigenetic signature of promoters and enhancers underlying cardiac hypertrophy*. Proceedings of the National Academy of Sciences of the United States of America, 2013. **110**(50): p. 20164-9.
6. Dixon, J.R., et al., *Topological domains in mammalian genomes identified by analysis of chromatin interactions*. Nature, 2012. **485**(7398): p. 376-80.