

# **High-throughput Analysis of Gene Expression and Alternative Splicing in Cardiac Hypertrophy**

Elizabeth Grace Ames  
B.S. Genetics, Cell Biology, and Development,  
University of Minnesota, 2007

A Dissertation presented to the Graduate Faculty  
of the University of Virginia in Candidacy for the Degree of  
Doctor of Philosophy

Department of Molecular Physiology and Biological Physics  
University of Virginia

June 13, 2013

Thesis advisor:	Jeffrey W. Holmes, MD PhD
Examination committee:	Jessica J. Connelly, PhD
	Brant E. Isakson, PhD
	Aaron J. Mackey, PhD
	Robert K. Nakamoto, PhD
	Mark J. Yeager, MD PhD (chair)

## **Abstract**

Cardiac hypertrophy affects millions of Americans and occurs as a result of common disease states including high blood pressure and valvular disease. On the most basic level, cardiac hypertrophy is an enlargement of individual cardiomyocytes that predisposes the heart to myocardial infarction, arrhythmias, and heart failure. An intervention that could stop this remodeling process before significant impairment or pathology develops has been sought for decades [1,2]. Identifying novel mechanisms that regulate hypertrophy and would be amenable to drug targeting could potentially prevent costly and significant cardiovascular events. This proposal uses high-throughput techniques to explore two possible mechanisms underlying the process of cardiac hypertrophy.

Cardiomyocytes respond to a myriad of paracrine, hormonal, and mechanical signals. While many of the hormonal signaling pathways have been identified [2], a clear mechanotransduction pathway by which changes in cardiac workload produce changes in myocyte size and shape remains elusive [3]. Our previous studies have shown that the amount of shortening that individual cardiomyocytes experience may be an important mechanical cue in cardiac remodeling, particularly in determining myocyte shape [4,5]. Therefore, in aim 1, we will use an *in vitro* muscle culture system to control these mechanical cues with the goal of identifying genes that are responsive to changes in shortening. These genes may provide clues to how cardiomyocyte regulate cell shape during hypertrophy.

Over two decades ago, certain fetal genes involved in contraction and metabolism were identified as being re-expressed in cardiac hypertrophy. Collectively these genes were referred to as the 'fetal gene program'. More recently, alternative splicing has been shown to be an important aspect of heart development [6–8]. Specifically, alternative mRNA isoforms

undergo switches that are highly regulated throughout fetal and post-natal heart development. The second aim of this proposal will use whole-transcriptome analysis to characterize the similarities in alternative splicing between fetal and hypertrophied hearts in a rat model of hypertrophy.

In fetal hearts, these developmentally-regulated alternative splicing events are regulated by a handful of RNA-binding proteins (RBP), similar to the way transcription factors control gene expression. In the third aim, we will use similar computational approaches to identify which exons undergo alternative splice and which RBPs could potentially be mediating the re-expression of fetal splice variants in cardiac hypertrophy. Once candidate RBPs have been identified we will statistically test whether the presence of these motifs can explain the changes in exon inclusion we observed in both fetal and TAC hearts.

*The overall goal of these studies is to provide new understanding of the mechanisms of myocyte shape regulation, alternative splicing, and cardiac hypertrophy.*

**Specific Aim 1: Identify genes responsive to changes in cardiomyocyte shortening as candidate regulators of cell shape.** We hypothesize that only a few genes will be specifically regulated by changes in shortening. We will vary the amount of cardiomyocyte shortening *in vitro*, measure global gene expression, and identify shortening-responsive proteins and pathways as candidate regulators of cell shape. We will validate our experimental *in vitro* data with comparisons to genes that have similar patterns of expression in *in vivo* models of cardiac hypertrophy.

**Specific Aim 2: Test the hypothesis that alternative splicing patterns present during fetal heart development are re-expressed during cardiac hypertrophy.** We hypothesize that some alternative splicing isoforms that are expressed in fetal hearts will be expressed in cardiac hypertrophy, but not in normal adult hearts. We will measure alternative splicing using whole-transcriptome high-throughput sequencing (RNAseq) and compare mRNA isoform expression between cardiac hypertrophy, sham-operated adults, and fetal hearts to quantify amount of overlap of splicing patterns between hypertrophic and fetal hearts.

**Specific Aim 3: Identify candidate RNA-binding proteins that mediate the common regulation of splicing in heart development and cardiac hypertrophy.** We hypothesize that splicing changes seen in fetal and hypertrophied hearts are regulated by RNA-binding proteins (RBPs). We will identify alternative exons and search for short motifs that are over-represented in the regions flanking these exons to identify candidate RBPs that are responsible for alternative splicing in cardiac hypertrophy. To test the significance of the RBPs identified, we will specifically ask if the rates of exon inclusion for the exons that are flanked by a specific motif deviate from what would be expected if each exon was regulated independently.

## **Table of Contents**

High-throughput Analysis of Gene Expression and Alternative Splicing in Cardiac Hypertrophy .....	i
Abstract .....	ii
Table of Contents .....	v
Acknowledgments.....	x
Chapter 1. Introduction.....	1
Features and clinical relevance of cardiac hypertrophy.....	1
Cardiac hypertrophy and the fetal gene program .....	3
Introduction to alternative splicing .....	4
Alternative splicing in heart development.....	6
<i>Cardiac troponin T as a case study of alternative splicing in heart development</i> .....	7
<i>SRSF2 controls alternative splicing of CaMKII<math>\delta</math> in the developing heart</i> .....	9
<i>Titin: a case study in exquisite splicing diversity</i> .....	11
Alternative splicing in hypertrophic and failing hearts.....	14
<i>Alternative splicing changes in the transition from hypertrophy to heart failure</i> .....	14
<i>Specific isoforms of enigma homolog promote or prevent hypertrophy in vitro</i> .....	14
Conclusions.....	16
Chapter 2. Effect of shortening and stretch on gene expression in intact myocardium .....	18
1. Introduction.....	20
2. Materials and methods.....	23
2.1. <i>Papillary muscle isolation</i> .....	23
2.2. <i>Papillary muscle culture set up</i> .....	24
2.3. <i>Experimental design and mechanical inputs</i> .....	26
2.4. <i>Measurement and control of muscle mechanics</i> .....	26
2.5. <i>RNA isolation, preparation, and microarray assays</i> .....	27
2.6. <i>Microarray data analysis</i> .....	28
3. Results .....	29
3.1. <i>Differences in muscle culture mechanical inputs created four distinct groups</i> .....	29
3.2. <i>Distinct gene expression changes with decrease stretch and decreased shortening</i> .....	30

3.3. Comparison of papillary muscles gene expression to pressure overload and LVAD treatment.....	31
4. Discussion .....	32
5. Conclusions .....	35
Appendix A. Tables and figures .....	37
Table 1. Genes with concordant shortening-dependent expression at both physiologic (high) and sub-physiologic (low) mean stretch .....	37
Table 2. Genes with concordant stretch-dependent expression at both physiologic (high) and sub-physiologic (low) amounts of shortening.....	38
Table 3. Over-represented ontology terms among genes with significantly altered expression in response to reduced shortening in vitro and in TAC .....	39
Table 4. Over-represented ontology terms among genes with significantly altered expression in response to reduced mean stretch in vitro and LVAD treatment in failing human left ventricles .....	40
Figure 1. Two-way mechanical inputs for 12h culture of papillary muscles.....	41
Figure 2. Overlapping gene expression by individual mechanical inputs .....	42
Figure 3. Comparisons of in vitro mechanical changes to in vivo studies of pressure overload and left ventricular assist device treatment .....	43
Appendix B. Supplementary data.....	44
Supplementary Table 1. Genes regulated concordantly in response to reduced cyclic shortening in papillary muscles and TAC in mice .....	44
Supplementary Table 2. Genes regulated concordantly in response to reduced mean stretch in vitro and clinical LVAD therapy .....	50
Chapter 3. Sequencing of mRNA identifies re-expression of fetal splice variants in hypertrophy .....	54
1. Introduction.....	56
2. Materials and methods.....	58
2.1. Animals .....	58
2.2. Tissue harvest from fetal and adult rats.....	58
2.3. Minimally-invasive approach to transverse aortic constriction.....	59
2.4. RNAseq data generation and analysis.....	59
2.5. PCR validation .....	60
3. Results .....	60
3.1. Successful induction of hypertrophy in adult rats.....	60

3.2. Similarities in gene expression between hypertrophy and heart development .....	61
3.3. Changes in gene and isoform expression are independent in both fetal and hypertrophic hearts .....	62
3.4. Comparison of splicing events in hypertrophy and heart development identifies significant overlap .....	62
3.5. Differences between splicing patterns observed in fetal heart development and TAC.....	64
3.6.1. PCR validation of significant isoform identifies both developmentally- and hypertrophy-related alternative splicing .....	64
3.6.2. PCR validation of RNAseq results identifies known splice variants that are regulated in heart development.....	65
3.7. Potential upstream regulators of isoform changes seen in heart development, hypertrophy, or both processes .....	66
4. Discussion .....	67
5. Conclusions .....	72
Appendix A. Tables and figures .....	74
Table 1. Summary of sequencing results for each experimental group.....	74
Table 2. Gene ontology (GO) analysis of biological processes associated with up- and down-regulated isoforms seen in both hypertrophy and development .....	75
Figure 1. Induction of hypertrophy with a minimally-invasive approach to TAC .....	76
Figure 2. Comparison of gene expression changes during hypertrophy and heart development .....	77
Figure 3. Changes in gene and isoform expression are distinct in both fetal and hypertrophic hearts .....	78
Figure 4. Significant overlap of splicing patterns in heart development and hypertrophy .....	79
Figure 5. Similarities and differences between isoform expression observed in fetal heart development and TAC.....	80
Figure 6. Identification and validation of known developmental splicing events .....	81
Appendix B. Supplementary methods .....	82
Appendix C. Supplementary data .....	88
Supplementary Table 1. Top shared isoforms among TAC and fetal group.....	88
Supplementary Table 2. PCR primer sequences.....	90
Supplementary Table 3. Gene Ontology (GO) analysis of biological processes associated with isoforms specifically seen in heart development or hypertrophy.....	91

<i>Supplementary Table 4. Potential upstream regulators of changes in isoform expression in hypertrophy, heart development, and both processes</i> .....	92
<i>Supplementary Figure 1. Hypothesis testing and hypergeometric probability density functions</i> .....	94
<i>Supplementary Figure 2. Random downsampling of results shows that the trends seen in full data set are independent of sequencing read counts</i> .....	95
Chapter 4. Computational search for RNA-binding protein recognition motifs identifies candidate regulators of fetal splicing program in cardiac hypertrophy.....	96
1. Introduction.....	97
2. Materials and methods.....	100
2.1. Topology search to identify alternatively spliced exons .....	100
2.2. Gene ontology (GO) enrichment analysis.....	101
2.3. RBP motif search using a first-order Markov model.....	101
2.4. Motif enrichment heat map.....	103
2.5. Statistical testing of the relationship between a RBP motif and a predicted change in PSI .....	103
3. Results .....	105
3.1. Search for alternatively spliced exons identifies exons that are fetal-specific, TAC-specific, regulated in both groups, or not regulated in either group.....	105
3.2. Significant pentamers in regulatory region correspond to RBP recognition motifs .....	106
3.3. RBP motifs are enriched, rather than depleted, in introns flanking alternative splicing events.....	107
3.4. The presence of specific RBP motifs significantly influences exon inclusion.....	107
4. Discussion .....	108
5. Conclusions .....	113
Appendix A. Tables and figures .....	115
Table 1. Significant motifs with predictive effect on exon inclusion .....	115
Figure 1. Alternatively spliced exon represented as a directed acyclic graph .....	116
Figure 2. Sequence regions used as input for RBP motif search.....	117
Figure 3. Significant motifs for overlap exons .....	118
Figure 4. Motif enrichment for significant motifs across overlap, fetal specific, TAC-specific, null, and single exon genes .....	119
Figure 5. Summary of significant motifs and their effect on splicing.....	120

Appendix B. Supplementary data.....	121
<i>Supplementary Table 1. Gene ontology (GO) analysis of biological processes associated with alternative exons significant in both heart development and hypertrophy and exons that were not significantly different in TAC, fetal, or sham hearts</i> .....	121
<i>Supplementary Table 2. Comparison of significant motifs in murine fetal heart, galline heart development, and overlap group of hypertrophy and rat fetal heart development</i> .....	122
<i>Supplementary Table 3. Significant motifs used in hypergeometric probability density function probabilities at different PSI cut off criteria</i> .....	124
<i>Supplementary Figure 1. Exon size distribution among alternative and constitutive exons</i> .....	126
<i>Supplementary Figure 2. Pentamers enriched in fetal-specific exons</i> .....	127
<i>Supplementary Figure 3. Pentamers enriched in TAC-specific exons</i> .....	128
<i>Supplementary Figure 4. Pentamers enriched in single exon genes</i> .....	129
<i>Supplementary Figure 5. Expanded motif enrichment heat maps</i> .....	130
<i>Supplementary Figure 6. Hypothesis testing and hypergeometric probability density function</i> .....	131
Appendix C. Selected Perl scripts.....	132
<i>Code to construct DAGs and topology searches for alternative cassette exons</i> .....	132
<i>Code to count observed pentamers and calculated expected frequencies using a first-order Markov model</i> .....	136
Chapter 5. Future directions .....	139
Polypyrimidine tract binding protein 1 mediates the re-expression of fetal splice variants seen in cardiac hypertrophy .....	139
<i>Figure 1. Proposed mechanism of PTB regulation and re-expression of fetal splice variants in hypertrophy</i> .....	143
Computation approaches to determine tissue-specific elements that regulate cardiac hypertrophy and implications for public health .....	144
Physical coordination of RNA processing in hypertrophy .....	145
Changes in extracellular matrix and cell-cell contacts in response to mechanical input mediates cell shape differences in concentric and eccentric cardiac hypertrophy .....	146
Bibliography .....	148

## Acknowledgments

The work presented in this dissertation would not have been possible without the support and guidance of many people. I would like to thank Jeffrey W. Holmes, who served as my thesis advisor. His willingness to let me develop my own project that took his lab in a new direction is something that was challenging and immensely satisfying. I'm grateful that I was able to develop a project that I am excited to talk about every day. I also appreciated his patience and his probing questions to make sure I understood things beyond the "default settings". The lessons I have learned from him certainly have made me a better scientist.

I would also like to thank the members of the Cardiac Biomechanics Group that I had the pleasure of working with including Charles Haggart Ph.D., Katherine Parker, Andrew Rouillard Ph.D., Greg Fomovsky, Ph.D., Christian Moyer, Will Richardson Ph.D., Jennifer Bartell, M.S., and Samantha Clarke. I would especially like to thank Tracy Burcin, who taught me so many practical lessons about how to fix things, how to be patient, and how to laugh at myself.

The other members of my thesis committee, Aaron Mackey, Jessica Connelly, Robert Nakamoto, Brant Isakson, and Mark Yeager have all provided helpful feedback, guidance, and enthusiastic support. I want to specifically acknowledge Aaron Mackey, who has taught me so many skills and life lessons. His enthusiasm for computational biology, teaching, and science has truly inspired me. He taught me how to write computer code and by doing this, taught me that with enough hard work, I can do things that I thought were impossible. This is something I will always appreciate.

I was lucky enough to start working in a research lab in high school. I was even luckier that the people I worked with were so enthusiastic and passionate about their work. They instilled in me a deep love for genetics that I appreciate to this day. James Mickelson and Stephanie Valberg remain trusted advisors and role models for me today.

The members of my family (both two- and four-legged) have always supported me and their faith in me when I doubted myself helped me keep going during difficult times. At an early age my parents taught to be curious and ask questions, I have called upon this love of learning often and I'm so grateful to have such wonderful parents.

None of this work would have been possible without the constant love, support, and friendship of my husband. Aside from being my best friend, he is my most trusted colleague. I have been so lucky to be able to talk about genetics, this project, and science in general with him on a daily basis. His amazing intelligence and ability to ask great questions have helped me when I had given up on a problem. He has taught me so many things and made me a better person.

## **Chapter 1. Introduction**

### Features and clinical relevance of cardiac hypertrophy

Cardiac hypertrophy is a biological process that is defined by an increase in mass within cardiomyocytes. There are two types of hypertrophy that are differentiated by how the additional mass is distributed. In patients with pressure overload states such as chronic hypertension, individual cardiomyocytes and the left ventricular wall become thicker, which is known as concentric hypertrophy. These cell shape changes create a relatively stable clinical picture; heart failure typically develops only after decades of chronic hypertension in elderly patients [9,10]. In volume overload states such as mitral regurgitation, myocytes become longer and the left ventricle dilates, which is referred to as eccentric hypertrophy. When volume overload occurs, there is a much shorter time until the development of heart failure. For example, within 10 years of diagnosis of severe mitral regurgitation, 90% of patients demonstrated early signs of heart failure [11].

Indeed, these differences in clinical outcomes based on the type of hypertrophy were noted in the early 19<sup>th</sup> century. Jean-Nicolas Corvisart noted the profound morphological and clinical differences in cardiac hypertrophy. He described concentric hypertrophy as “a lively reaction of the muscular fibres” in response to high blood pressure. He also described dilatation of the heart and its severely impaired ability to function by noting “This pure and simple distention induces [the heart’s] attenuation, and loss of elasticity and contractility” [12]. Although physicians and anatomists have long known that different types of hypertrophy have different clinical outcomes, efforts to understand the mechanistic differences between these two conditions continues to this day.

The public health importance of cardiac hypertrophy cannot be emphasized enough. As of 2012, one in three Americans has high blood pressure (or hypertension, HTN) where HTN is defined as systolic blood pressure  $\geq 140$  mm Hg, diastolic blood pressure  $\geq 90$  mm Hg, taking an anti-hypertensive medicine, or having been told twice by a health professional that one has HTN [13]. Another 6% of Americans have undiagnosed HTN [14], which historically has been cited as a major risk factor for the development of left ventricular hypertrophy [15,16]. When left untreated, hypertrophy has significant adverse outcomes including myocardial infarction, arrhythmias, sudden cardiac death, and heart failure. Currently 5 million people are in heart failure in the US, where the estimated direct cost for heart failure in 2012 was \$32 billion. This cost estimate is projected to increase 120% to \$70 billion by 2030 [13]. Although a small fraction of cardiac hypertrophy is the result of genetic mutations, most cases of preventable left ventricular hypertrophy are easily treatable with antihypertensive medications as shown by multiple meta-analyses [17–19]. Yet as patients become more likely to survive severe heart attacks or valvular diseases, the resultant ventricular remodeling and hypertrophy are not readily treatable. We believe that the work presented here could potentially provide insights for the unmet need of additional therapies to treat and possibly prevent more refractory forms of cardiac hypertrophy. The cellular remodeling that occurs in the transition from hypertrophy to heart failure is a life-threatening and irreversible process [20]. If further mechanisms of hypertrophy are identified this could provide important pharmacologic targets to prevent heart failure and reduce a huge burden on the US health care system.

### Cardiac hypertrophy and the fetal gene program

For over twenty years, researchers have known that hearts undergoing concentric hypertrophy revert to a gene expression pattern normally seen only in fetal heart development [2,21]. Collectively this is referred to as the fetal gene program and is believed to be beneficial in short-term cardiac remodeling. Many of the genes cited as part of the fetal gene program are critical components of the sarcomere or involved in metabolism. Hallmarks include upregulation of  $\beta$ -myosin heavy chain, skeletal  $\alpha$ -actinin, atrial and brain natriuretic peptides, and downregulation of metabolic genes such as glucose transporter GLUT4 and glycogen synthase, as well as other metabolic enzymes [21]. Re-expression of the fetal gene program is believed to be a protective mechanism for the heart. As heart mass increases in hypertrophy, the environment surrounding individual cells becomes more hypoxic and acidic. This environment is similar to the relatively hypoxic environment of the fetal heart. During hypertrophy, the heart shifts from an oxygen-rich environment with fatty acids available for energy to an oxygen-poor environment with glucose as the main energy source, and expresses fetal genes that are better suited to this different metabolic environment in terms of utilizing available energy sources, differences in mitochondrial activity, and mechanical performance of the heart.

There are many transcription factors implicated in the development of cardiac hypertrophy including the HAND family [22,23], GATA4 [24–26], Nkx2-5 [27–29], the MEF2 family of transcription factors [30–32], c-Jun [2,33], c-myc [2,33], STAT [34], NF- $\kappa$ B [35], and SRF [36]. The NFAT family of transcription factors is most widely recognized as mediating the gene expression changes [37,38] in hypertrophy and upregulating the expression of fetal gene program members such as ANP, BNP, skeletal  $\alpha$ -actin, and  $\beta$ -myosin heavy chain

[32]. More recently other post-transcriptional and epigenetic mechanisms of gene expression regulation has been shown to be critical for cardiac hypertrophy [39–41].

### Introduction to alternative splicing

As transcription occurs, stretches of the nascent RNA strand known as introns (non-coding sequences) are removed while retaining exons (coding sequences) in the final messenger RNA (mRNA) transcript. This was first observed when viral DNA was hybridized with the corresponding mRNA molecules and loops of DNA without complementary RNA sequences were seen along the transcript [42]. Philip Sharp and Richard Roberts were awarded the Nobel Prize for Medicine or Physiology in 1993 for this discovery. The process of alternative splicing selectively removes introns and some exons from pre-messenger RNA molecules. The composition of mature mRNA transcripts from one gene can therefore have markedly different sequences based upon differences in alternative splicing [43]. After the human genome was sequenced, scientists were surprised by the relatively small number of protein-encoding genes. It was expected that cells, let alone organisms, would need an enormous number of genes to encode all the proteins and enzymes required for different specialized tissues. Given the surprisingly small number of genes in mammalian genomes, alternative splicing fulfills the demands of proteome diversity [44,45]. In fact, more than 95% of genes are alternatively spliced [46]. The prevalence of alternative splicing increases in more complex organisms that have multiple cell types with specialized functions, such as mammals [47–49]. An early example of how one gene can give rise to two very different tissue-specific proteins via alternative splicing is the calcitonin gene. The calcitonin gene contains 6 exons, where the first three exons are constitutively expressed in all transcripts. In thyroid “C” cells, the fourth exon is included and the calcitonin protein functions as an important regulator of calcium homeostasis. However, when the calcitonin gene is

expressed in neurons, exons 5 and 6 are included and the transcript is translated into calcitonin gene-related peptide (CGRP) [50]. CGRP functions as a potent vasodilator [51] and is centrally involved in migraine headache pathophysiology [52]. To emphasize the importance of alternative splicing, this process is highly conserved and similar tissue-specific examples have been shown to occur in invertebrates that mediate important functions including *Drosophila* sex determination and courtship behavior [53,54].

Although the exact amount of evolutionary conservation of alternative splicing events varies with tissue and temporal expression, most splicing events are species-specific [55,56]. Conservation estimates of alternative splicing events between mouse and human range from 11-83% in the literature. Two main conclusions can be drawn from this lack of conservation; 1) diversity in exon expression is a process that is subject to selective pressure and 2) when conservation occurs between mammalian species, it likely has an important function. One fascinating example of the functional consequences of alternative splicing and how evolutionary selective pressure can shape which exons are included in a final transcript is the TRPV1 gene in vampire bats. When compared to other fruit bats, vampire bats express a truncated form of the TRPV1 protein in their trigeminal ganglia. This truncation lowers the thermal activation threshold of the TRPV1 channel to 30°C, which allows these obligate blood feeders to detect infrared emitted warm-blooded prey [57].

The splicing of individual mRNA transcripts does not occur in isolation or on a case-by-case basis. RNA-binding proteins (RBPs) play a critical role in organizing alternative splicing of functionally-related mRNA transcripts. These RBPs recognize short sequences in the nascent mRNA transcript within the exon and introns. Each exon is defined by the presence of consensus splice sites within the mRNA transcript. These sequences are necessary for constitutively expressed exons, exons that are always present. When the

splice site sequences deviate from the consensus sequences recognized by the splicing machinery, other *cis*-regulatory sequences can either promote or repress the inclusion of these exons. These are referred to as exonic splicing enhancers (ESE) or exonic splicing silencers (ESS) if the sequence is located in the exon and intronic splicing enhancers (ISE) or intronic splicing silencers (ISS) if they are located in the introns flanking the alternative exon. An intron is physically removed from mRNA transcript via a two-step transesterification reaction, where the intervening intron forms a lariat structure at the specified branch point nucleotide, while the two adjacent exons are brought into proximity by other spliceosomal proteins and then joined together [45]. This process is now recognized to occur co-transcriptionally and is greatly influenced by chromatin marks and nucleosome positioning [58].

#### Alternative splicing in heart development

Just as gene expression programs exist during development of multiple tissues [59–62], coordinated programs of alternative splicing have been shown to participate in general embryonic development [63,64]. Changes in splicing are usually coordinated by a combination of tissue- and stage-specific RNA-binding proteins. Most of these protein-RNA interactions have been identified by rigorous *in vitro* and *in vivo* testing or more recently, by bioinformatic analysis of RNA-binding recognition motifs. Deletion studies of a few known splicing factors and RNA-binding proteins result in embryonically lethal phenotypes [65–69], strongly emphasizing their function in embryonic development. When other RNA-binding proteins are deleted, developmental defects are seen in only post-mitotic tissues, such as muscle, motor neurons, eye, and heart, suggesting that alternative splicing is important to terminal differentiation [70–73]. It is important to note that the differential changes in splicing

during development are separate from differential changes in gene expression [63,74]. RNA-binding proteins and alternative splicing create spatiotemporal control of gene expression that is not accomplished with transcriptional regulation alone [75,76]. To date, most studies of alternative splicing have been highly focused to only measure levels of known splice variants by polymerase chain reaction or exon-specific microarrays. With the advent of new technologies such as RNAseq and CLIPseq [77,78], our understanding of alternative splicing is greatly increasing.

#### *Cardiac troponin T as a case study of alternative splicing in heart development*

Cardiac troponin T exemplifies one of the most well-characterized cardiac development isoform switches. Knowledge of its gene structure, functionality of splice variants, and elucidation of splicing regulation that orchestrates the developmentally-regulated isoform switch has shown how alternative splicing can be employed in remodeling the heart during cardiac development. Troponin T is one of three troponin proteins that regulate thin filament activity within the sarcomere. Troponin T binds tropomyosin and manipulates tropomyosin's position to allow cross bridge cycling when calcium is bound by troponin C. There are three TnT isoforms, each specific to a different muscle type (fast, slow, and cardiac). A given type-specific isoform is more conserved between vertebrate species than all the muscle-type isoforms are within one species. This suggests that each isoform has an evolutionarily important and specialized role unique to that muscle type. Among all isoforms, the N-terminal region of TnT is the most diverse part of the protein [79]. This region does not serve a specific binding role or act as a single protein interaction domain, but functions in modulating the overall molecular conformation of TnT [79]. One major difference between TnT isoforms is the amount of negative charge within this region. Embryonic isoforms of TnT have more negative residues and are more tolerant to acidosis

than adult isoforms. In addition, it has been shown that more negative charge in this region increases the calcium sensitivity of myosin ATPase and force development [80,81]. In a more physiologic set of experiments, Nassar et al. overexpressed the human embryonic cTnT in mice. They did not observe any differences in gross morphology, sarcomere structure or physiological variables such as amount of calcium needed for half-maximal force development. Their most significant findings related to cardiac function on the organ level. Human embryonic cTnT decreased the rate of maximal left ventricular pressure and isovolumic relaxation [82]. These studies suggest that the developmentally-regulated alternative splicing has *in vivo* implications on left ventricular function.

Cardiac TnT undergoes changes in splicing over the course of development. This was first demonstrated twenty-five years ago in chicken hearts [83]. During the course of post-natal cardiac development, exon 5 is excluded from adult isoforms of cardiac troponin T. This developmentally-regulated isoform switch is conserved in rat, rabbit, mice, and humans [84–86]. An elegant explanation for this phenomenon was determined by TA Cooper's group. The exclusion or inclusion of exon 5 is determined by the expression levels and activity of two types of RNA-binding proteins, those that enhance exon inclusion (CELF; CUG binding protein and embryonic lethal abnormal vision type RNA binding protein 3 (ETR-3)-like factor) and those that repress exon inclusion (MBNL; muscle-blind like and PTB; poly-pyrimidine tract binding protein). During embryonic development CELF proteins, CUG-BP1 and CUG-BP2 dominate splicing decisions by binding poly-UG motifs in the intron downstream of exon 5 to promote exon inclusion in the final mRNA transcript [87]. MBNL and PTB proteins bind RNA motifs in the introns both up- and down-stream of exon 5 and favor exclusion of the alternative exon [88]. As might be expected when cardiomyocytes were depleted of functional CELF proteins, exon 5 inclusion significantly decreased,

whereas when PTB function was inhibited, exon 5 inclusion increased. This antagonistic relationship was further confirmed by measuring protein levels during cardiac development. As cardiac development progresses, nuclear CELF protein levels decrease while PTB and MBNL protein levels remain constant in the nucleus throughout development. The MBNL1 knockout mouse expressed cTnT with exon 5 present, suggesting that a very low level CELF protein expression remains in adulthood [71] and a dynamic balance between the antagonistic proteins truly determines exon inclusion/exclusion. From the initial report describing the requirement a balance between CELF and MBNL proteins to determine alternative splicing in cardiac development, one question remained. The CELF protein levels decreased with development, yet the mRNA levels of CELF family members remained high. It was recently shown through the use of a cardiac-specific Dicer knockout that microRNA 23a/b is necessary for the downregulation of CELF proteins during development [89]. Again this added layer of complexity suggests that there is a highly orchestrated program controlling gene expression beyond transcription in development.

A similar developmental isoform switch of skeletal muscle TnT occurs postnatally where the cardiac troponin T (including exon 5) is transiently expressed before birth. After birth, skeletal muscle begins to express the skeletal muscle gene for troponin T. The synchrony of these isoform switches on both the transcriptional and post-transcriptional levels strongly suggests that changes in alternative splicing after birth are not a cardiac-only response to adult circulation and increased workload, but are part of a systemic, coordinated developmental program [79].

#### *SRSF2 controls alternative splicing of CaMKII $\delta$ in the developing heart*

Several mechanistic papers have studied the role of the specific splicing regulatory proteins SC35, SRp38, and SRSF2 and their role in cardiac physiology [70,90,91]. Each

protein is required for embryonic development [65–67], yet when conditional knock-out technology is used to delete splicing regulators after cardiac organogenesis, each protein has unique and specific effects on cardiomyocyte function. Among these studies, Xu X et al. best characterized the role of SR protein (serine-arginine-rich) SRSF2 (also known as ASF/SF2) in cardiac development.

SR proteins bind specific motifs in exonic sequences and are required for constitutive splicing [45]. It also been shown that SR proteins participate in mRNA export [92–95], mRNA stability [96], and protein translation [97]. When SRSF2, a prototypical SR protein, is deleted only a small fraction of known cardiac splicing events are disrupted in knockout mice. All of the affected genes undergo developmentally-regulated isoform switches, including cardiac troponin T (cTnT), Cypher (in mice, or ZASP in humans), and Ca<sup>2+</sup>/calmodulin-dependent kinase II $\delta$  (CaMKII $\delta$ ). Unlike the isoform switches of cTnT and Cypher, CaMKII $\delta$  splicing in the heart was previously unknown. As the cardiac workload increases from adolescence to adulthood (around 4 weeks, Arber S et al., 1997), a unique phenotype was observed by Xu et al. among the SRSF2 knockout mice that was not seen in other SR protein knockout mice. Grossly, the heart chambers of SRSF2<sup>-/-</sup> mice were dilated with thin walls. Histopathology showed cardiomyocyte hypertrophy, fibrosis, and myofibril disarray. Electron microscopy revealed hypercontracted sarcomeres and disorganized Z disks. This hypertrophic, hypercontracted phenotype suggested abnormal calcium handling, which was confirmed by measurements of calcium handling during systole (peak Ca<sup>2+</sup> and caffeine-induced). A number of key Ca<sup>2+</sup> handling proteins were screened by Western blotting analysis to determine the cause of the abnormal Ca<sup>2+</sup> handling. Aberrant expression of the neuronal-specific CaMKII $\delta$ A ( $\delta$ A) isoform is localized in the nucleus, resulting in less phospholamban phosphorylation, less inhibition of SERCA2a, and therefore less calcium

reuptake into the sarcoplasmic reticulum. Around postnatal day 20, wild type cardiomyocytes begin to express CaMKII $\delta$ B and  $\delta$ C isoforms in the nucleus and cytoplasm, respectively. The location of the  $\delta$ C isoforms effectively increases calcium reuptake into the sarcoplasmic reticulum. This isoform switch does not occur in the absence of SRSF2. Instead, SRSF2<sup>-/-</sup> mice continue to express  $\delta$ A isoforms that are inappropriately targeted to the t-tubules. To confirm this finding, the authors generated transgenic mice that over-expressed CaMKII $\delta$ A isoform under regulation of a cardiac-specific promoter. The transgenic mice phenocopied the SRSF2<sup>-/-</sup> mice with regards to gross cardiac hypertrophy, decreases in fractional shortening, and calcium handling, albeit on a delayed timeline. This article did not specifically explore the actual mechanism by which SRSF2 would interact with CaMKII $\delta$  mRNA transcripts. It is possible that a secondary mediator could change the splicing pattern of CaMKII $\delta$ . This article still makes a strong case that general splicing regulators play specific and critical roles during cardiac development. This article is also important because it highlights that alternative splicing does not just control protein composition, but has critical roles in protein localization.

*Titin: a case study in exquisite splicing diversity*

Titin is the third most common protein in muscle cells by mass, behind actin and myosin. The average adult is composed of 500 grams of titin alone [98]. An individual molecule of titin spans half a sarcomere, from the Z disk to the center of the sarcomere, which is known as the M line [99]. Titin has many roles in myocytes, including generating passive tension through its many elastic immunoglobulin (Ig) domains that function as a molecular spring. With its prominent and constant position in the sarcomere, titin serves to measure and maintain appropriate sarcomere dimensions, giving titin the nickname

“molecular ruler” [100]. More recently, the signaling and scaffolding potential of titin has been explored in its putative serine/threonine kinase domain within the M line region [101].

The titin gene is located on the long arm of human chromosome 2 and is composed of 363 exons spanning 281,435 base pairs [100]. The protein itself is composed of several protein domains named in relationship to its sarcomere location. The M band region, at the carboxyl end of the protein, is known to have unique splicing patterns dependent on tissue type [102–104], species [105], and developmental stage [106]. The study that examined the developmentally-regulated splicing of the M band region only measured isoform expression in rat skeletal muscle. The investigators did not measure isoform expression in the heart during development [106], thus it would be interesting to determine if this isoform switch is specific to skeletal muscle or if the patterns of M band region splicing also occur in the heart and represent a striated-muscle specific developmental isoform switch. The Z disk region of titin contains several 45 amino acid repeats, known as Z repeats that are interspersed between 4 Ig domains. The number of Z repeats correlates with Z disk thickness and is unique to different muscle types [107]. The I-band region of titin is responsible for generating passive tension and has the most complex splicing pattern of all the titin regions, and perhaps in all mammalian genes. This region is composed of four separate structural regions: 1 - Ig-domains, 2 - PEVK segment (named for the segment’s enrichment in proline (P), glutamic acid (E), valine (V), and lysine (K) amino acid residues), 3 - the cardiac-specific N2B region, and 4 - the N2A region [108]. The two main splicing patterns of I-band titin are dependent on the inclusion of the N2B or N2A regions. In the heart, the N2B region is always expressed. The N2A region is always expressed in skeletal muscle and transiently expressed during cardiac development [109]. When both the N2A and N2B regions are present, this is referred to as N2BA isoform of titin. In the embryonic heart, the longer, more

compliant N2BA isoform is the predominant titin splice variant. N2BA titin includes a variable number of exons from exons 49-50 (N2B region) to exons 102-111 (N2A region). As embryonic development continues, the level of N2BA titin decreases while the N2B isoform, which is shorter and stiffer than N2BA, gradually increases in expression. This isoform switch is directly related to the amount of passive tension within the cardiomyocyte [108,110]. Experiments that measured passive tension in fetal and adult rat hearts definitively showed a large difference in passive tension at all sarcomere lengths between adult and fetal samples as a result of titin alternative splicing [108]. The PEVK segment is located between the N2B and N2A regions and also has dramatic variation in splicing patterns. Studies that cloned individual mRNA transcripts from human hearts have found that one heart can contain 10 or more splice variants in this region [111]. If one assumes that alternative splicing of each titin exon within the PEVK segment is an independent event, this could potentially generate more than 274 billion different isoforms [112]. The functionality of this extraordinary degree of isoform diversity is still unknown, but recent studies of different human cardiac pathologies suggest that these splicing patterns are responsive to physiological inputs and may play a role in disease pathogenesis. Human patients with diastolic heart failure have reduced N2BA:N2B ratios, meaning their hearts express less N2BA titin and are stiffer than the average adult heart [113]. In comparison, congestive heart failure patients have a 20% increase in their N2BA titin levels compared to normal patients [114]. A naturally occurring mutation in rats that dramatically delays the N2BA to N2B isoform switch was recently reported [115], and further experiments have determined that the RNA-binding protein and splicing repressor Rbm20 is responsible for this change in alternative splicing. Most interesting, the authors were able to show that

several cases of idiopathic dilated cardiomyopathy were caused by mutations that disrupted Rbm20 function [116,117].

### Alternative splicing in hypertrophic and failing hearts

While changes in alternative splicing in multiple developmental contexts have been thoroughly described in the literature, changes in alternative splicing during disease states pose more questions than answers. One of the major questions is what comes first, do changes in alternative splicing cause disease or does alternative splicing change as part of the response to disease pathogenesis. Here we describe what is known about alternative splicing in hypertrophy and the progression to heart failure.

#### *Alternative splicing changes in the transition from hypertrophy to heart failure*

Studies such as these illustrate the importance of understanding how alternative splicing influences cardiac hypertrophy and clinical progression into heart failure. Indeed, a recent study of alternative splicing in human heart failure patients identified specific isoforms that were diagnostic of heart failure in humans with 98% accuracy [118]. With the exception of a few highly focused studies as detailed in this chapter, the regulation and extent of splicing in cardiac hypertrophy is unknown. Recently the isoform expression in pressure-overload hypertrophy was compared with heart failure. The isoform expression seen in heart failure represented a more dramatic program that encompassed many of the isoforms expressed in hypertrophy [119]. Indeed, more isoforms were shared between hypertrophy and heart failure (454 isoforms) than were unique to hypertrophy (284 isoforms).

#### *Specific isoforms of enigma homolog promote or prevent hypertrophy in vitro*

Enigma homolog (ENH) is a member of the PDZ-LIM protein family. These proteins are comprised of two protein interaction domains, the PDZ domain and the LIM domain. The

PDZ domain was named after the proteins PSD-95, DLG, ZO-1 and these proteins were initially characterized as playing a critical role in epithelial and neuronal polarity. PDZ proteins undergo alternative splicing as a means to diversify protein scaffold interactions [120]. The LIM domain (named for proteins Lin11, Isl-1, Mec-3) contains 55 amino acid residues, mostly cysteine and histidine with 8 specific residues being highly conserved among all proteins. The LIM domain forms a tandem zinc finger structure that functions in a wide variety of cellular processes including gene regulation, cytoskeletal organization, signal transduction, cell motility, and cell adhesion [121]. Both PDZ and LIM domains are found throughout the eukaryotic kingdom and have a wide variety of functions [120,121]. In skeletal and cardiac muscle, proteins that contain both these domains are able to bind to the sarcomere and participate in various signaling pathways, making them top candidates for the signal transducers of mechanical forces from the Z disk to the nucleus. Within the PDZ-LIM protein family, ENH is a member of a subfamily that contains one N-terminal PDZ domain and three consecutive LIM domains at the C-terminus [101]. ENH was initially discovered as a protein kinase C binding partner [122]. Since then, ENH has been shown to localize to the Z disk [123] and interact with other signaling molecules, including protein kinase D1 and a subunit of the L-type voltage-gated calcium channel [124]. When ENH is deleted, other ENH binding partners, calsarcin and Cypher (ZASP in humans) are lost from the Z disk and dilated cardiomyopathy ensues [125]. Given the function of ENH as a scaffolding protein, it is not surprising that there are multiple splicing isoforms to accommodate multiple combinations of protein binding. Another member of this subfamily, ZASP has also been shown to be highly spliced as well [126].

Recently, Yamazaki and colleagues examined the specific functions of the 4 different ENH splice variants (ENH1-4). They found that during embryonic development ENH1 is the

predominant mRNA species, but after birth the pattern of splicing changes, increasing the level of isoforms ENH2, 3, and 4 to be the major species in heart. ENH1 contains three C-terminal LIM domains whereas the adult isoforms do not. Expression of the embryonic isoform ENH1 could be induced by hypertrophic stimuli including transverse aortic constriction (TAC) and pharmacologic agents. To determine the utility of this isoform switch, ENH1 was overexpressed and silenced *in vitro*. When the embryonic isoform ENH1 was overexpressed, neonatal rat cardiomyocytes demonstrated a dramatic hypertrophic response including increases in cellular volume (as measured by electrical capacitance) and upregulation of hypertrophy markers (ANP, BNP, and  $\beta$ -MHC). When ENH1 was knocked down by siRNA, the cells were unable to generate a hypertrophic response. Intriguingly, the same inability to upregulate markers of hypertrophy was observed when the adult isoform ENH4 was overexpressed. The authors proposed that the adult isoform ENH4 acts as a dominant-negative inhibitor of the embryonic, pro-hypertrophy ENH1. Both hemodynamic (TAC) and pharmacologic agents (phenylephrine, endothelin, and aldosterone) could induce expression of the embryonic isoform as well as promote hypertrophic gene expression, suggesting that ENH is downstream of the initial sensors for mechanical stress or hormonal signaling [40]. These findings strongly suggest that a specific developmental program of splicing is present in cardiomyocytes and is able to be turned on again when similar physiological cues are present. These findings prompt more questions. How are mechanical or pharmacologic signals transduced to change splicing patterns? Are these pathways amenable to pharmacologic intervention?

### Conclusions

In the following chapters, I will explore gene expression regulation in hypertrophy on multiple levels including transcription and alternative splicing. In chapter 2, I will present our

work to identify candidate genes for cardiomyocyte shape regulation in response to changing mechanics. Chapter 3 compares alternative splicing in heart development and pressure-overload hypertrophy. We have found strong similarities between these two processes and in chapter 4, we look at what might be controlling this process using computational and statistical models. Finally in chapter 5, I summarize the work completed in this dissertation and suggest future experiments.

**Full title: Effects of stretch and shortening on gene expression in intact myocardium**

Authors: Haggart CR (PhD)<sup>a,1,†</sup>, Ames EG (BS)<sup>a,c,†</sup>, Lee JK (PhD)<sup>d</sup>, Holmes JW (MD PhD)<sup>a,b,e</sup>

† these authors contributed equally to this work

Affiliations: <sup>a</sup> Department of Biomedical Engineering, University of Virginia (Box 800759, Charlottesville, VA, USA 22908)

<sup>b</sup> Robert M. Berne Cardiovascular Research Center, University of Virginia (Robert M. Berne Cardiovascular Research Center, 415 Lane Rd, Charlottesville, VA, USA 22908)

<sup>c</sup> Department of Molecular Physiology and Biological Physics, University of Virginia (Jordan Hall, PO Box 800736, Charlottesville, VA USA 22908)

<sup>d</sup> Center for Public Health Genomics, Department of Public Health Sciences, University of Virginia (P.O. Box 800717, Charlottesville, VA USA 22908)

<sup>e</sup> Department of Medicine, University of Virginia (PO Box 800793, Charlottesville, VA USA 22908)

Present addresses: <sup>1</sup>U.S. Food and Drug Administration, (10903 New Hampshire Ave, Silver Spring, MD 20993)

Email addresses: Ames EG (ega2d@virginia.edu), Haggart CR (charles.haggart@gmail.com), Lee JK (jaeklee@virginia.edu), Holmes JW (holmes@virginia.edu)

Correspondence to Jeffrey W. Holmes, Department of Biomedical Engineering, University of Virginia, Box 800759, Charlottesville, VA 22908. Fax: 434-982-3870 Telephone: 434-243-6321 Email: holmes@virginia.edu

Data availability: Microarrays are publically available from the NCBI's Gene Expression Omnibus website (series accession ID: GSE45250).

## Abstract

Multiple cues have been suggested as the mechanical stimulus for the heart's hypertrophic response. Our work has previously suggested that the amount of cyclic shortening in cardiomyocytes controls myocyte shape and that the amount of stretch controls overall myocyte size. To identify specific gene expression changes that occur in response to each of these mechanical perturbations, we used microarray analysis of RNA isolated from papillary muscles cultured for 12 hours at physiologic or reduced levels of cyclic shortening and physiologic or reduced overall mean stretch. Overall, genes related to extracellular matrix were surprisingly prominent in our analysis. Connective tissue growth factor (CGTF) was among a small group of genes regulated by the amount of cyclic shortening regardless of the level of mean stretch, and many more extracellular matrix genes were regulated by shortening at one of the levels of stretch. When we compared our results to gene expression data from an *in vivo* model of pressure overload (PO) – which also decreases myocyte shortening – we found the genes that were commonly regulated in PO and our decreased shortening groups were most significantly enriched for ontology terms related to the extracellular matrix, followed by genes associated with mechanosensing and the cytoskeleton. The list of genes commonly regulated in PO and our decreased shortening groups also included many of the genes known to change early in hypertrophy, such as myosin heavy chain (Myh7), natriuretic peptide (Nppb), and myosin binding protein C (Mybpc2). We conclude that in intact myocardium, the amount of cyclic shortening may be an important regulator not only of myocyte genes classically associated with hypertrophy but also of extracellular matrix genes.

## 1. Introduction

Hypertension, valve disease, myocardial infarction, and even pregnancy cause hypertrophy, defined as an increase in myocyte mass. The risk of progression to heart failure in these conditions, however, depends primarily on changes in myocyte shape. In patients with pressure overload (PO) states such as chronic hypertension, individual cardiomyocytes increase in cross-sectional area with the addition of sarcomeres added in parallel and subsequently the entire left ventricular wall become thicker; this process is known as concentric hypertrophy. These cell shape changes create a relatively stable clinical picture. Heart failure develops only after decades of chronic untreated hypertension in elderly patients [9,10] as the adverse effects of hypertrophy develop including interstitial fibrosis that leads to arrhythmias and myocardial infarctions as the heart outgrows its blood supply. In PO states, the ejection fraction and the amount of cardiomyocyte shortening are reduced. In contrast, volume overload (VO) states such as mitral regurgitation are characterized by a much shorter time until the development of heart failure. Within 10 years of diagnosis of severe mitral regurgitation, 90% of patients demonstrated early signs of heart failure [11]. When VO occurs, the heart is required to pump an increased stroke volume and individual cells are required to increase the amount they shorten each heartbeat. During this process, myocytes undergo eccentric hypertrophy, becoming longer and thinner. We showed that the amount of shortening correlates with myocyte shape across a range of experimental hypertrophy models in rats [5], and then demonstrated directly that the amount of shortening regulates cardiomyocyte shape in cultured rat right ventricular papillary muscles [4]. Understanding how cell shape is controlled by amount of shortening could enable new therapeutic approaches to direct hypertrophy and prevent or reverse heart failure.

Many investigators have proposed hypotheses regarding which mechanical inputs induce hypertrophy. The observation that patients with high blood pressure had hearts with concentric hypertrophy led Grossman et al. to hypothesize that increased systolic stress (more force per area) in hypertensive patients caused individual heart cells to add sarcomeres in parallel to increase the cell's diameter and cross-sectional area. This hypothesis suggests that by increasing the cross-sectional area, the higher systolic forces are distributed over a larger area, effectively normalizing the increased systolic stress each cardiomyocyte experiences [15]. Despite the fact that this hypothesis has been widely accepted for more than 30 years, there has been no adequate proof that systolic stress controls cell shape. While attractive, Grossman's hypothesis does not fully address the issue of cardiac hypertrophy. The authors provide no similar hypothesis for the regulation of eccentric hypertrophy. In addition, experiments to directly test the relationship between systolic stress and cell shape have proven to be difficult.

Unlike many other proposed mechanisms, our hypothesis that as the shortening amplitude increases, the length:width ratio of cardiomyocytes increases (cells get longer and skinnier) provides an explanation for both types of hypertrophy. In seeking mechanical signals that might trigger changes in cardiomyocyte shape, we began by comparing changes in a long list of candidate mechanical stimuli during experimental VO and PO [5]. One clear difference between these two states was their effect on stroke volume, ejection fraction, and related measures such as regional systolic strain and shortening. We therefore hypothesized that the amount of cyclic shortening experienced by myocytes controls myocyte shape, and demonstrated directly that myocytes from muscles cultured at a reduced level of systolic shortening to simulate PO get shorter and wider over 36 hours [4] compared to muscles cultured at physiologic levels (15%) of shortening. There is evidence

of this in volume overload, when cells are required to shorten more to pump the increased stroke volume, cells undergo eccentric hypertrophy. In pressure overload, when ejection fraction is reduced and the shortening amplitude of individual myocytes decrease, concentric hypertrophy occurs.

Although a host of other potential mechanical and biochemical stimuli for cell shape changes have been proposed and studied over the past century, it has been experimentally challenging to definitively identify the mechanisms directly responsible for cell shape regulation. *In vitro* cell stretching experiments allow precise manipulation of mechanical and chemical inputs, yet this experimental system is unphysiologic for myocytes: the cells are plated in two dimensions on artificial substrates, often devoid of appropriate connections to the extracellular matrix, neighboring myocytes, and cardiac fibroblasts, all of which have been shown to modulate cardiac hypertrophy [127,128]. *In vivo* experiments are confounded by neuroendocrine signaling. Hormones such as angiotensin II and norepinephrine are known to induce hypertrophy without mechanical perturbation; efforts to pharmacologically block these responses in order to isolate mechanical from chemical effects on hypertrophy during *in vivo* overload have produced few clear insights. *In vitro* culture of papillary muscles offers a convenient middle ground between cell culture and *in vivo* preparations, allowing excellent control of mechanical and biochemical inputs while maintaining cell-cell and cell-ECM connections [4].

The primary goal of the present study is to identify candidate genes that may play a role in regulating myocyte shape by identifying genes whose expression is specifically responsive to changes in myocyte shortening. In addition to changes in myocyte shape, myocyte size increases rapidly following the imposition of experimental PO or VO [5], and regresses rapidly following mechanical unloading with left ventricular assist devices (LVADs)

[129]. Based on the strong correlation between myocyte size and average chamber volume (or related local quantities such as myocyte stretch and sarcomere length) across these different states, we have hypothesized that average sarcomere length or myocyte stretch may trigger changes in gene expression that regulate myocyte mass. Therefore, in the present study we identify genes whose expression is specifically responsive to changes in average myocyte stretch and then compare these *in vitro* gene expression changes to other *in vivo* gene expression studies of pressure overload and LVAD treatment.

## **2. Materials and methods**

### *2.1. Papillary muscle isolation*

All experiments were conducted in accordance with the Guide for the Care and Use of Laboratory Animals [130] and approved by the University of Virginia's Institutional Animal Care and Use Committee. 12 adult male LBN-F1 rats (Harlan; Indianapolis, IN),  $281 \pm 13$  g (mean  $\pm$  SD), were euthanized by 1 ml intraperitoneal injection of 50 mg/ml pentobarbital sodium (Ovation Pharmaceuticals, Inc.; Deerfield, IL). Following intracardiac heparinization, the heart was rapidly dissected, transferred to a microdissection dish, and retrograde perfused via the ascending aortic stump with a solution to prevent cross-bridge cycling (4°C Krebs-Henseleit (K-H) Buffer Modified containing (in mM) 118 NaCl, 4.7 KCl, 1.2 MgSO<sub>4</sub>, 1.2 KH<sub>2</sub>PO<sub>4</sub>, 20 NaHCO<sub>3</sub>, 11 glucose, 0.25 CaCl<sub>2</sub>, 30 2,3-butanedione monoxime (BDM) and 20 IU/l insulin in equilibrium with 95% O<sub>2</sub>-5% CO<sub>2</sub>). Next, the right ventricle (RV) was opened by cutting along the posterior and superior edges of the free wall, followed by careful dissection of appropriately sized and shaped RV papillary muscles with the following dimensions (mean  $\pm$  SD): major radius,  $318 \pm 71$   $\mu$ m; minor radius,  $218 \pm 47$   $\mu$ m; length

(between the two mounting pins),  $2.44 \pm 0.57$  mm. To excise each papillary muscle, we first dissected all chordae tendinae attachments and subsequently cut around the septal insertion of the muscle. The excess tissue from the septal region was used to handle each muscle without damaging to the tissue. Dissected muscles were stored in oxygenated BDM K-H buffer on ice until the desired number of suitable muscles were obtained for the given culture experiment (typically four).

## *2.2. Papillary muscle culture set up*

We cultured each RV papillary muscle at 37°C for 12 h in a sterile culture system, modified from methods described by Janssen et al. [131] by the addition of a programmable servomotor to control muscle length. Muscles were mounted at slack length – such that a 50  $\mu$ m stretch yielded a measurable increase in passive force – via pins connected to a force transducer at one end and the servomotor at the other. After mounting each muscle in chilled BDM K-H buffer, we applied sterile TiO<sub>2</sub>-PBS (1:1 wt/wt) to each of four corners of the muscle's central region, creating two longitudinal pairs of markers. These visual markers allowed us to track muscle stretch and shortening throughout the culture experiment. The baths were then sealed and a superfusion of 95%O<sub>2</sub>-5%CO<sub>2</sub> was initiated to flow through each bath. The temperature of the culture system was increased until the media reached 37°C. We then exchanged the BDM K-H buffer for BDM-free K-H buffer with the same low concentration of Ca<sup>2+</sup> (0.25 mM Ca<sup>2+</sup>). Once the BDM K-H buffer had been thoroughly flushed from the culture system, we exchanged the low-Ca<sup>2+</sup> BDM-free solution for Medium 199 culture media (M199, 1.85 mM Ca<sup>2+</sup>; Mediatech; Manassas, VA) supplemented with the following: 21.7 mM NaHCO<sub>3</sub>, 2.0 mM L-carnitine, 4.4 mM creatine, 5.0 mM taurine, 2.0 mM L-glutamine, 20 IU/l insulin, 105 IU/ml penicillin, 100  $\mu$ g/ml streptomycin, and 1  $\mu$ g/ml amphotericin B. After the media exchange process and throughout the duration of the

culture period, muscles were electrically stimulated end to end by 5 ms asymmetric pulses at 1Hz frequency, 30-50% above the contraction-stimulating threshold voltage (2-4 V). The muscles were cultured and stimulated for 30-60 minutes to allow the muscles to reach steady state before beginning a pre-stretch protocol. Each muscle was stretched by 0.05 mm increments every 60-90 seconds until the central region of the muscle reached a pre-determined amount of stretch from slack (5 or 15%), which we confirmed by measuring the increased spread between both TiO<sub>2</sub> marker pairs during diastole. Following this pre-stretch protocol, we programmed the servomotors using one of four waveforms to impose a distinct combination of time-averaged stretch and cyclic shortening upon each cultured muscle throughout the 12 h culture period, as described below.

We used a customized data acquisition system (LabVIEW, National Instruments; Natick, MA) to continuously measure and record muscle force, control servomotor position, and periodically acquire video of muscle contraction-relaxation. From this video feed, we measured diastolic stretch (compared to slack length at time = 0) and adapted a marker tracking algorithm to calculate time-averaged stretch and cyclic shortening between both longitudinal pairs of TiO<sub>2</sub> markers. To preserve identical culture conditions for each sample, M199 and 95%O<sub>2</sub>-5%CO<sub>2</sub> superfusate gas were exchanged only at the midpoint (6h) of each experiment. Following measurements of force, stretch, and shortening at 12h, we terminated the electrical stimulus and any length control protocol, relaxed each muscle in cold BDM K-H buffer for 2-3 minutes while trimming the ends and measuring muscle dimensions, and flash froze each muscle in liquid nitrogen prior to storage at -80°C until subsequent RNA isolation.

### 2.3. Experimental design and mechanical inputs

We employed a factorial design [132] as the framework for defining our four experimental groups: two mechanical factors (mean stretch and cyclic shortening) at two distinct levels that represent “normal” and “reduced” states (*high* and *low*, respectively). In preliminary studies, we observed that too little or too much stretch leads to muscle contracture. These target values were chosen to maximize the difference between the groups while minimizing intra-group variability and risk of muscle contracture. We only accepted muscles that contributed to a group mean that was two times greater or less than the counterpart mechanical group input. We defined *high* target values for time-averaged stretch to be 16% above slack length and 4% for the *low* stretch group. We similarly defined the target values for *high* cyclic shortening to be 16% of slack length and 4% for the *low* shortening group.

### 2.4. Measurement and control of muscle mechanics

We adapted a marker tracking algorithm developed for measuring strain in mechanically tested tissues [133] to track the TiO<sub>2</sub> markers throughout each contractile cycle and quantify mean stretch and cyclic shortening in each muscle’s central region. The central region of contracting papillary muscles can move independently of the prescribed movement of the ends. By tracking the movement of the central region, we were able to carefully adjust the amount of prescribed shortening to ensure the muscle’s central region experienced the desired amount of cyclic shortening. Three groups (n = 3 each) were pre-stretched 16% from slack length (*high-ΔL/high-L*, *high-ΔL/low-L*, and *low-ΔL/high-L*), while the fourth group (n = 3) was pre-stretched 5% from slack (*low-ΔL/low-L*). In the case of low shortening muscles, we prescribed up to 5% cyclic lengthening when necessary to counteract substantial shortening in the muscle’s central region, especially at the beginning

of the culture period. In the case of high-shortening muscles, we prescribed initial shortening slightly above the target mean value to account for typical decreases over the course of the experiment. To achieve different amounts of time-averaged stretch between the *high- $\Delta L$*  groups, the low-stretch muscles were held at systolic length for a considerably longer duration than the high-stretch muscles (700 vs. 50 ms, respectively). At equal intervals throughout the experiment (0, 3, 6, 9, and 12 h) we measured muscle stretch and cyclic shortening. This allowed us to make frequent adjustments of length control parameters in order to achieve time-varying length target values. These pre-stretch and servomotor length control protocols yielded the following muscle groups: *high- $\Delta L$ /high-L*, normal shortening, normal stretch; *high- $\Delta L$ /low-L*, normal shortening, reduced stretch; *low- $\Delta L$ /low-L*, reduced shortening, reduced stretch; *low- $\Delta L$ /high-L*, reduced shortening, normal stretch (Fig. 1A).

#### 2.5. RNA isolation, preparation, and microarray assays

Three samples from each of the four mechanical input groups were removed from -80°C and partially submerged in a bath of liquid nitrogen within their own 1.5 ml tubes. Each frozen muscle was then pulverized with a liquid nitrogen-cooled pellet pestle (Kontes®, Thermo Fisher Scientific, Waltham, MA). We performed standard RNA extraction from this point, using 1 ml of Trizol (Invitrogen, Life Technologies, Carlsbad, CA), shearing genomic DNA with a 25-gauge needle, and adding glycogen during RNA precipitation. The total RNA pellet was resuspended in 9  $\mu$ l of RNase-free H<sub>2</sub>O and warmed for 10 minutes (60°C) to promote dissolution. Each sample was treated with DNase I (Invitrogen, Life Technologies; Carlsbad, CA) according to the manufacturer's protocol to remove any remaining genomic DNA contamination. Total RNA concentrations were measured using a small volume spectrophotometer (NanoDrop, Thermo Fisher Scientific; Waltham, MA). The samples were then diluted the samples to a concentration of 33.3 ng/ $\mu$ l. Subsequent processing and

microarray hybridization were performed by the Biomolecular Research Facility (School of Medicine, University of Virginia). Prior to microarray target preparation, each selected RNA sample was run on the Agilent Bioanalyzer 2100 with an RNA 6000 Nano Chip (Agilent; Santa Clara, CA) to assess sample quality and confirm total RNA and mRNA concentrations. Global gene expression of all samples was assayed using GeneChip® Rat Genome 230 2.0 arrays (Affymetrix; Santa Clara, CA). All biotin-labeled, amplified RNA for hybridization was prepared using the standard Affymetrix kit and protocol. Following hybridization, the Biomolecular Research Facility returned the raw expression files for bioinformatics pre-processing and analysis.

## 2.6. *Microarray data analysis*

We processed the 12 CEL files by loading them and the appropriate Affymetrix CDF library annotation file into R using the Bioconductor package ([www.R-project.org](http://www.R-project.org); [www.bioconductor.org](http://www.bioconductor.org)) to process the raw data with the Robust Multichip Average (RMA) algorithm [134,135]. We removed probe-set data from all arrays if the RMA-normalized  $\log_2$  signal of a probe-set was below  $\log_2(50)$  in at least 50% of the samples within each group to reduce noise created by low signal probes [136,137]. This signal-based filter reduced the number of probe-sets for subsequent analysis from 31,042 to 16,391.

To maximize our ability to detect inter-group expression differences between groups, we used the local-pooled-error (LPE) method (downloadable R package from Bioconductor) [138]. LPE estimates within-probe-set variance by pooling the variances of all probe-sets at similar expression intensities, across all samples within a given experimental group and corrects the variance for each probe-set to a level consistent with its overall expression intensity [138]. We computed LPE-based *Z*-scores for each probe-set comparison to identify the transcriptional effects of reduced cyclic shortening under physiologic mean stretch (*high-*

$\Delta L$ /*high-L* vs. *low- $\Delta L$ /*high-L**) or sub-physiologic mean stretch (*high- $\Delta L$ /*low-L** vs. *low- $\Delta L$ /*low-L**), and the effects of reduced mean stretch under physiologic cyclic shortening (*high- $\Delta L$ /*high-L** vs. *high- $\Delta L$ /*low-L**) or sub-physiologic cyclic shortening (*low- $\Delta L$ /*high-L** vs. *low- $\Delta L$ /*low-L**). For each comparison, we used a false discovery rate (FDR) cutoff of 0.3 to identify the most significantly regulated probe-sets. We then used DAVID (the Database for Annotation, Visualization, and Integrated Discovery) [139,140] to perform biological annotation and functional analysis of the resulting lists of regulated genes. We used the 'Functional Annotation Clustering' tool within DAVID to identify statistically significant overrepresentation of biological annotation terms associated with our significant gene lists.

### 3. Results

#### 3.1. Differences in muscle culture mechanical inputs created four distinct groups

Through different combinations of stretch and shortening protocols, we successfully cultured four distinct groups of mean stretch and cyclic shortening, as confirmed by our marker-tracking protocol (Figure 1). The time-averaged stretch values for each group were calculated to be (means  $\pm$  SD): *high- $\Delta L$ /*high-L**,  $1.157 \pm 0.012$ ; *low- $\Delta L$ /*high-L**,  $1.165 \pm 0.015$ ; *high- $\Delta L$ /*low-L**,  $1.033 \pm 0.015$ ; *low- $\Delta L$ /*low-L**,  $1.061 \pm 0.008$  (Fig. 1B). Average cyclic shortening values were (means  $\pm$  SD): *high- $\Delta L$ /*high-L**,  $0.165 \pm 0.009$ ; *low- $\Delta L$ /*high-L**,  $0.036 \pm 0.002$ ; *high- $\Delta L$ /*low-L**,  $0.164 \pm 0.005$ ; *low- $\Delta L$ /*low-L**,  $0.022 \pm 0.003$  (Fig. 1C). We recorded mechanical force continuously throughout the culture period and normalized this by muscle cross-sectional area to estimate stresses. Time-averaged diastolic stresses (means  $\pm$  SD) were not statistically different between any groups using one-way ANOVA. Due to their consistently low active contractile stress, mean systolic stress in *low- $\Delta L$ /*low-L** muscles ( $-0.18 \pm 1.08$  mN/mm<sup>2</sup>) was observed to trend lower than the other three groups (*high-*

$\Delta L$ /*high-L* muscles,  $2.28 \pm 0.64$  mN/mm<sup>2</sup>; *high-ΔL/low-L*,  $2.16 \pm 0.66$  mN/mm<sup>2</sup>; *low-ΔL/high-L* muscles,  $5.69 \pm 4.10$  mN/mm<sup>2</sup>), however this did not yield a statistically significant variance between groups by one-way ANOVA.

### 3.2. Distinct gene expression changes with decrease stretch and decreased shortening

In order to identify genes whose expression responded specifically to the amount of shortening, we analyzed the transcriptional effects of reduced cyclic shortening at *high* stretch (*low-ΔL/high-L* vs. *high-ΔL/high-L*; right oval, Fig. 2A) and at *low* stretch (*low-ΔL/low-L* vs. *high-ΔL/low-L*; left oval, Figure 2A). Using LPE (described in Materials and Methods), these two comparisons identified 50 and 6,214 differentially expressed genes, respectively. Of these, 36 annotated genes were significantly regulated within both *high-ΔL* vs. *low-ΔL* comparisons (Fig. 2A, oval overlap). Fourteen of these genes were concordantly regulated in the *low-L* and *high-L* comparisons (Table 1), while the expression of 22 genes changed in opposite directions in the *low* and *high* stretch comparisons. To sort the significance of each gene's overall differential expression, we calculated the geometric mean of each pair of LPE-determined *Z*-scores resulting from both reduced shortening comparisons. A greater mean *Z*-score reflects consistent shortening-dependent differential expression of a gene under both *high* and *low* stretch.

Conversely, to identify genes whose expression varied with stretch, but not shortening, we used LPE to compare the transcriptional effects of reduced stretch under *high* shortening (*high-ΔL/low-L* vs. *high-ΔL/high-L*) and under *low* shortening (*low-ΔL/low-L* vs. *low-ΔL/high-L*). We identified 20 genes that exhibited differential expression with a reduction in stretch under *high-ΔL* (Fig. 2B, upper oval) and 5,880 genes differentially expressed with reduced stretch under *low-ΔL* (Fig. 2B, lower oval). The overlap between

these two comparisons comprises of 6 annotated genes with concordant changes in expression (Table 2).

### 3.3. Comparison of papillary muscles gene expression to pressure overload and LVAD treatment

To compare these transcriptional changes to analogous *in vivo* and clinical results, we searched for published microarray studies that had experimental models similar to the relative alterations in myocardial mechanics in our muscle culture experiments. We identified a 7-day transverse aortic constriction (TAC) study performed in male wild-type mice [141] and a study exploring the global transcriptional effects of left ventricular assist device (LVAD) treatment of failing human hearts [142]. For each study, we used the CEL files from the original studies and performed the same data processing, low-signal filtering, and LPE analysis steps as was done for the cultured papillary muscle groups.

Pressure overload induced by TAC results in an acute reduction in cyclic shortening of the myocardium [143,144], a reduction we simulated in our *low- $\Delta L$*  muscle groups in the present study (Fig. 3A). An LPE comparison of the 7-day TAC and sham wild-type mice revealed 857 significantly regulated genes (FDR < 0.1). With this data, we aimed to identify genes that were regulated in both TAC and at least one of our shortening comparisons. To do this, we first eliminated any probe-sets that were discordantly regulated between our two shortening comparisons and filtered out any probe-sets with FDR  $\geq$  0.1. The overlap between the papillary muscle data and the TAC vs. sham data yielded 169 genes that were up-regulated in both pressure-overloaded mouse hearts and reduced cyclic shortening rat papillary muscles and 20 genes down-regulated in both comparisons (Supplementary Table S1). The gene ontology analysis of these 189 overlapping genes revealed 40 significantly enriched gene ontology clusters (enrichment score > 1.3), of which the top 12 are shown in

Table 3 including the most significant clusters “extracellular matrix”, “response to mechanical stimulus”, and blood vessel development.

Mechanical unloading induced by LVAD treatment of a failing human heart reduces the time-averaged ventricular chamber volume, a three-dimensional analogue of mean stretch [145]. Thus, post- vs. pre-LVAD data represents a change in stretch similar to the *high-L* vs. *low-L* muscle group comparison in our study (Fig. 3B). Again using LPE with an FDR cutoff of 0.1, we identified 482 genes significantly regulated following LVAD treatment of failing hearts. Using the same approach as described above, we identified 69 genes uniformly up-regulated in response to a reduced mean stretch in both cultured papillary muscles and LVAD treated human hearts and 33 genes down-regulated in both comparisons (Supplementary Table S2). Gene ontology analysis identified 11 significantly enriched functional annotation term clusters, which are listed in Table 4. The top clusters included “response to endogenous stimulus”, “negative regulation of cell differentiation”, and “negative regulation of apoptosis”.

#### **4. Discussion**

We previously hypothesized that the amount of cyclic shortening experienced by myocytes in the heart wall regulates myocyte shape, while the average stretch imposed on those myocytes regulates overall myocyte size. In order to identify genes specifically responsive to changes in cyclic shortening or time-averaged stretch, we cultured rat right ventricular papillary muscles for 12 hours at physiologic or reduced levels of shortening and stretch using a 2 x 2 factorial design, then used microarrays to assess changes in gene expression. We also examined the relationship between changes in gene expression in our cultured muscles and those reported during two *in vivo* interventions that alter shortening

and stretch – transverse aortic constriction (TAC) to induce pressure-overload hypertrophy and left ventricular assist device (LVAD) support of failing human hearts.

Based on our hypothesis that cyclic stretch regulates myocyte shape, we expected that genes specifically responsive to changes in cyclic stretch would be enriched for those encoding sarcomeric and cytoskeletal proteins. Instead, we found a relatively small set of genes that responded similarly to changes in shortening regardless of the level of underlying mean stretch (Table 1), including connective tissue growth factor (CTGF), heat shock protein 90 kDa  $\beta$ 1 (HSP90AB1), and nuclear factor, interleukin 3 regulated (NFIL3, also known as E4BP4). CTGF expression is induced by TGF- $\beta$  and is strongly up-regulated in heart failure [146] and in cultured myocytes exposed to pro-hypertrophic stimuli [147,148]. While its potential to cause cardiac fibrosis is still debated, CTGF was recently shown to be an important mediator of cardiac hypertrophy [149]. The HSP90 family has been shown to be up-regulated in cardiac stress and is involved in protein quality control of crucial cardiac proteins such as actin, tubulin, and TGF- $\beta$  [150]. NFIL3 was recently shown to be an important transcription factor for cell survival in myocytes [151].

Based on the fact that transverse aortic constriction (TAC) reduces myocyte shortening and dramatically alters myocyte shape, we also expected that genes encoding sarcomeric and cytoskeletal proteins would have largest changes in expression in the list of genes that were both regulated by reduced shortening in our papillary muscles and altered by TAC *in vivo*. The list of expression changes shared between these two datasets was enriched for cytoskeletal genes as expected (Table 3), and also included many known key components of the hypertrophic response such as Nppb, Ankrd1, Mybpc2, Myh7, Ctgf, Myl9, Myocd (see Supplementary Table 1 for full list). However, the degree of overlap in genes associated with extracellular matrix and ECM-receptor interaction was equally

striking. Gene ontology results emphasizes that the genes seen in both PO and our decreased shortening comparisons are over-represented for terms associated with cardiomyocyte remodeling including 'extracellular matrix', 'actin cytoskeleton', and 'embryonic development'. Indeed, given the similarities in gene expression between PO and papillary muscles with decreased shortening, the enrichment of ECM genes may represent a mechanical signal for fibrosis.

One striking result is the disparity between the numbers of differentially expressed genes within each pair of *high vs. low* shortening and *high vs. low* stretch comparisons. We saw the largest number of gene expression changes when we compared low shortening, low stretch group (*low- $\Delta L$ /low-L*) to either the low shortening, high stretch muscle group (*low- $\Delta L$ /high-L*; 6,214 regulated genes) or to the high shortening, low stretch group (*high- $\Delta L$ /low-L*; 5,880 genes). This result is particularly interesting, given that the mechanics of this group (low stretch, low shortening) were most similar to those of myocytes cultured on a rigid substrate. Indeed we now know that multiple cell types including fibroblasts and cardiomyocytes behave very differently in two-dimensional tissue culture models that lack appropriate ECM contacts [152,153], cellular, or electrical contacts with other cells [154,155] compared to those within three-dimensional tissue constructs and native organ structures. Our results suggest that mechanically perturbing cultured myocytes from an initial low-stretch, low-shortening background may generate a large number of changes in gene expression that would not be induced by a similar perturbation of intact myocardium cultured at a more physiologic initial baseline.

One potential source of error in our study is that we only examined gene expression at one 12h-time point during our stretching experiments. Although our previous observation that myocyte size changes significantly by 36 hours in this model suggested that relevant

gene expression changes must be present quite early, this time point is certainly extremely early compared to published *in vivo* experiments of cardiac overload, including those we used for comparison here. We chose an early time point to avoid any confounding effects of gradual degeneration of the cultured muscles at longer times, but that choice restricted the pool of shortening- and stretch-responsive genes we could detect to those that respond rapidly to perturbations. Another source of error in cell and organ culture experiments is the choice of media. Some components in our media are known to alter gene expression; insulin in particular is critical for energetics of contraction, but also functions as a pro-hypertrophic factor [156,157]. Because all papillary muscle gene expression comparisons were relative to other papillary muscle groups cultured in identical media, the primary impact of this error was likely that we could have missed a shortening- or stretch-dependent response of genes that are also regulated by insulin.

## **5. Conclusions**

We conclude that careful control of mechanical inputs to cultured papillary muscles allows analysis of early gene expression changes in response to specific mechanical inputs thought to play a role in hypertrophy, such as the amount of cyclic shortening and overall mean stretch. The degree of overlap between gene expression changes in papillary muscles cultured with decreased shortening and those measured in pressure overload suggests that decreased shortening may be important for triggering concentric hypertrophy *in vivo*. Extracellular matrix genes were also prominent in the group of genes regulated by decreased shortening *in vitro* and pressure overload *in vivo*, suggesting that decreased shortening may also be an important stimulus for fibrosis in response to pressure overload.

**Acknowledgments**

We gratefully acknowledge the technical assistance of T. Burcin and the University of Virginia Biomolecular Research Facility.

**Sources of funding**

This work was funded by NIH R01 HL-075639 (JWH), NIH T32 GM-007267-34 (EGA), and NIH T32 GM-008136-27 (EGA).

**Disclosures**

None declared.

## Appendix A. Tables and figures

*Table 1. Genes with concordant shortening-dependent expression at both physiologic (high) and sub-physiologic (low) mean stretch*

Gene Name	Gene Symbol	Expression Change	Mean Z-score
prostate transmembrane protein, androgen induced 1	Pmepa1	↑	5.74
nuclear factor, interleukin 3 regulated	Nfil3	↑	4.04
connective tissue growth factor	Ctgf	↑	3.84
activating transcription factor 3	Atf3	↑	3.64
stearoyl-CoA desaturase (delta-9-desaturase)	Scd	↑	3.35
myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, <i>Drosophila</i> ); translocated to, 3	Mllt3	↑	3.33
similar to RNA binding motif, single stranded interacting protein 3 isoform 1	LOC680726	↑	2.65
heat shock protein 90kDa alpha (cytosolic), class B member 1	Hsp90ab1	↑	2.61
cytoplasmic polyadenylation element binding protein 2	Cpeb2	↑	2.45
solute carrier family 41, member 3	Slc41a3	↓	4.95
interleukin 6	Il6	↓	4.59
fin bud initiation factor homolog (zebrafish)	Fibin	↓	4.16
thymic stromal lymphopoietin	Tslp	↓	3.71
chemokine (C-C motif) ligand 11	Ccl11	↓	3.27

*Table 2. Genes with concordant stretch-dependent expression at both physiologic (high) and sub-physiologic (low) amounts of shortening*

Gene Name	Gene Symbol	Expression Change	Mean Z-score
Epstein-Barr virus induced 3	Ebi3	↑	3.55
interleukin 6	Il6	↑	2.98
similar to RIKEN cDNA 9530077C05	RGD1561444	↓	3.64
lymphatic vessel endothelial hyaluronan receptor 1	Lyve1	↓	3.50
thrombospondin 4	Thbs4	↓	3.33
collagen, type VI, alpha 1	Col6a1	↓	2.34

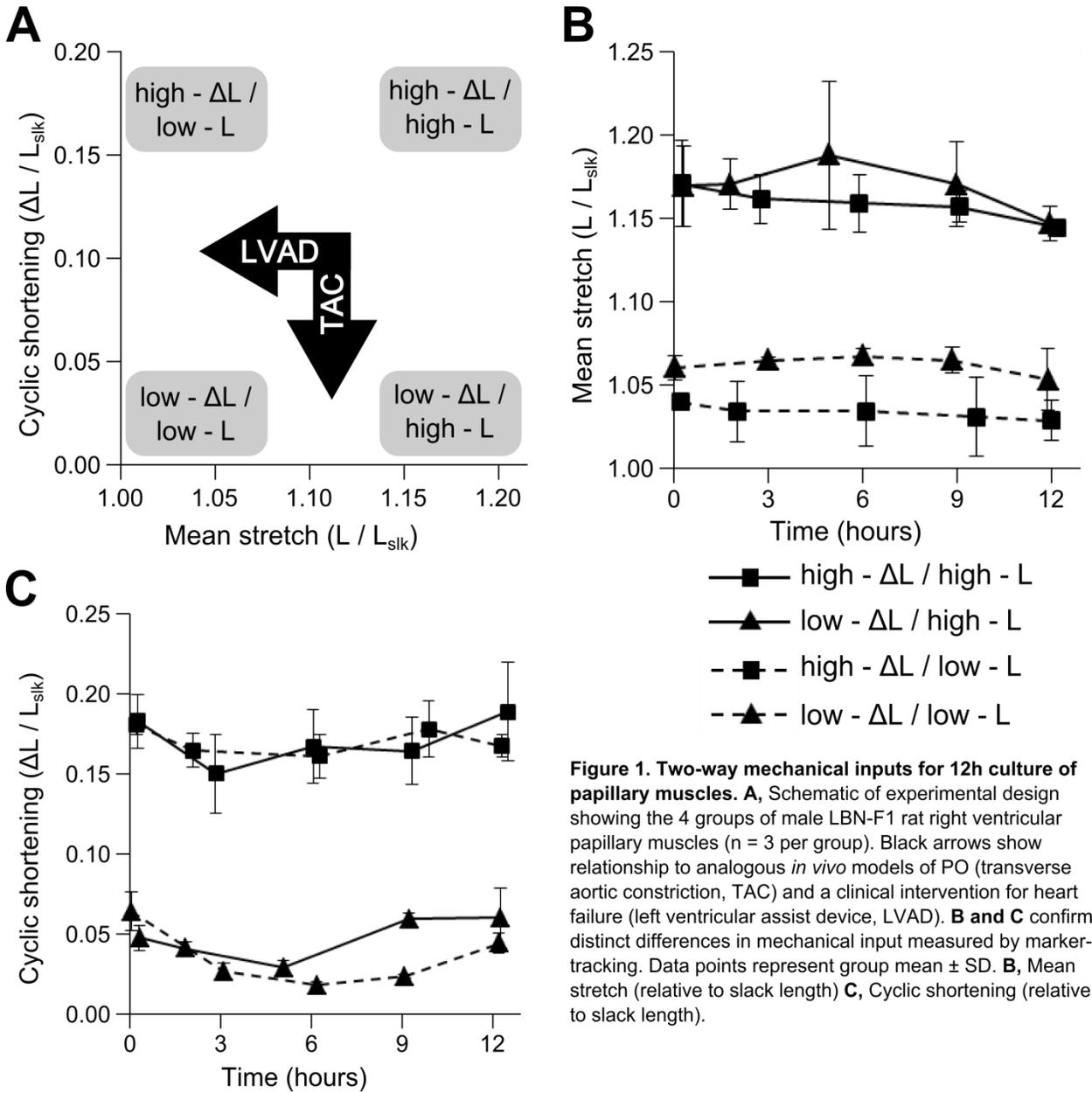
*Table 3. Over-represented ontology terms among genes with significantly altered expression in response to reduced shortening in vitro and in TAC*

Gene Ontology Term	<i>P</i> value	# of Genes	Fold Enrichment
extracellular matrix	1.11E-11	22	6.63
response to mechanical stimulus	5.77E-06	9	9.12
blood vessel development	1.36E-06	15	5.09
actin cytoskeleton organization	2.33E-06	13	5.80
regulation of actin cytoskeleton	4.48E-05	13	4.19
chordate embryonic development	2.37E-05	17	3.53
response to calcium ion	1.40E-06	9	10.97
cell migration	2.64E-05	14	4.22
extracellular region	1.97E-10	46	2.73
cell adhesion	8.37E-05	18	3.03
response to glucose stimulus	3.03E-04	7	7.58
ECM-receptor interaction	2.13E-07	11	9.15
identical protein binding	9.22E-06	21	3.14
regulation of protein complex assembly	3.11E-06	9	9.89
muscle organ development	5.92E-05	11	5.08
<b>25 additional enriched clusters</b>			

*Table 4. Over-represented ontology terms among genes with significantly altered expression in response to reduced mean stretch in vitro and LVAD treatment in failing human left ventricles*

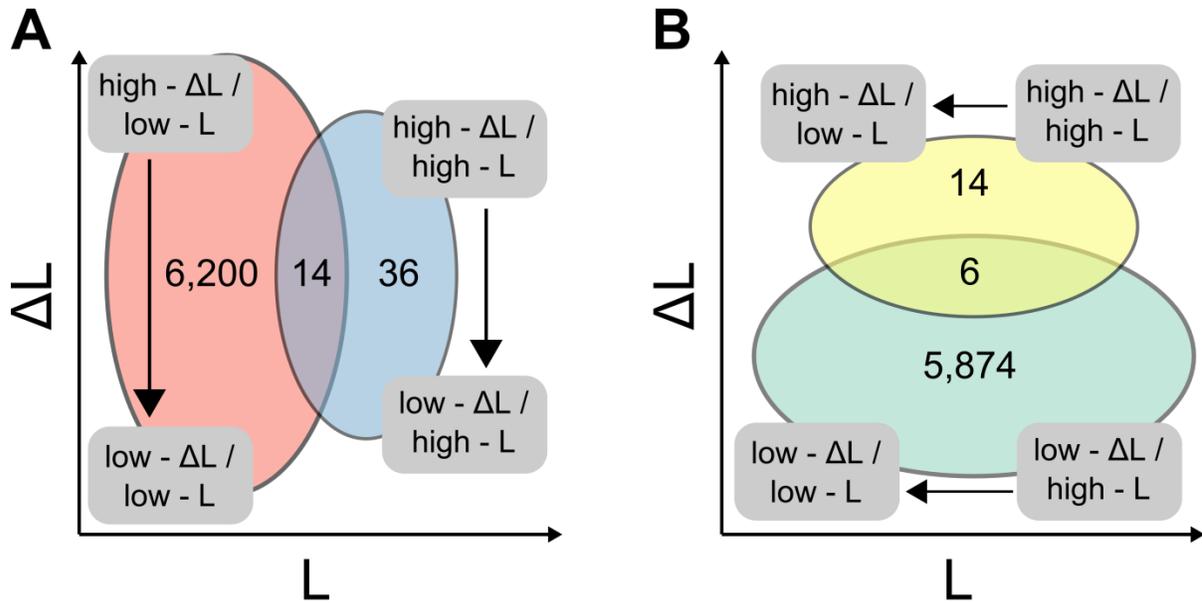
Gene Ontology Term	<i>P</i> value	# of Genes	Fold Enrichment
response to endogenous stimulus	3.62E-05	15	3.72
negative regulation of cell differentiation	3.53E-03	7	4.70
negative regulation of apoptosis	2.15E-03	9	3.84
regulation of cell proliferation	6.31E-04	14	2.99
extracellular matrix part	3.55E-04	6	9.66
collagen	3.36E-03	4	13.01
response to insulin stimulus	6.52E-02	4	4.28
response to glucose stimulus	1.37E-02	4	7.90
aging	8.64E-02	4	3.79
positive regulation of cellular biosynthetic process	5.98E-03	12	2.57
7 additional enriched clusters			

Figure 1. Two-way mechanical inputs for 12h culture of papillary muscles



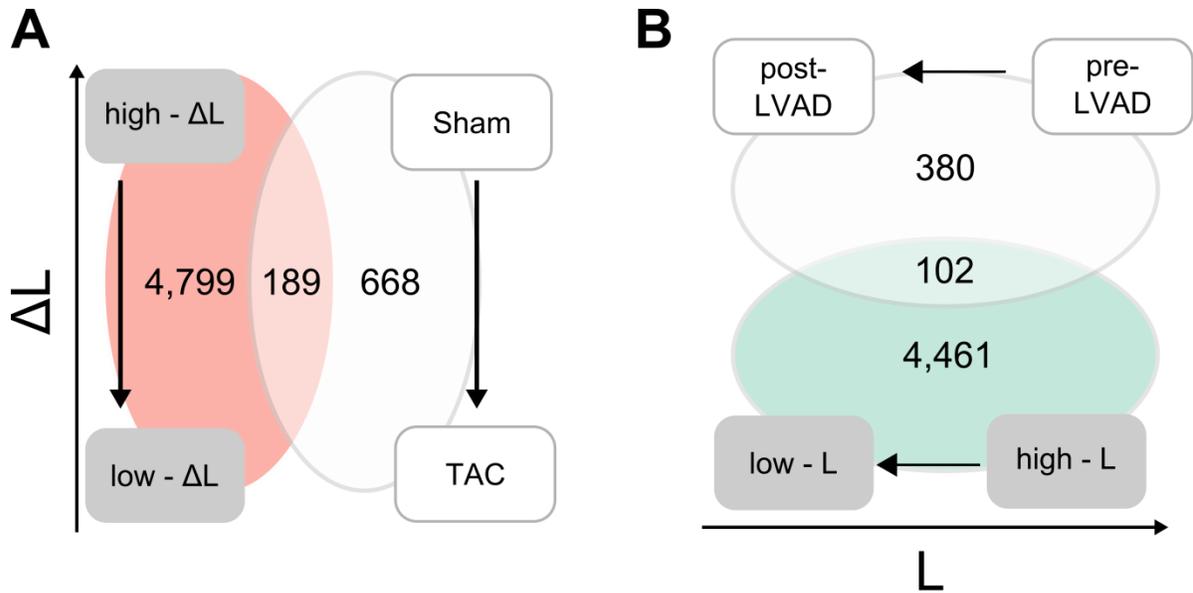
**Figure 1. Two-way mechanical inputs for 12h culture of papillary muscles.** **A**, Schematic of experimental design showing the 4 groups of male LBN-F1 rat right ventricular papillary muscles ( $n = 3$  per group). Black arrows show relationship to analogous *in vivo* models of PO (transverse aortic constriction, TAC) and a clinical intervention for heart failure (left ventricular assist device, LVAD). **B and C** confirm distinct differences in mechanical input measured by marker-tracking. Data points represent group mean  $\pm$  SD. **B**, Mean stretch (relative to slack length) **C**, Cyclic shortening (relative to slack length).

Figure 2. Overlapping gene expression by individual mechanical inputs



**Figure 2. Overlapping gene expression by individual mechanical inputs.** Numbers represent the number of genes that changed significantly concordantly in each comparison. **A**, Decreased cyclic shortening at low and high mean stretch. **B**, Mean stretch at low and high shortening.

Figure 3. Comparisons of in vitro mechanical changes to in vivo studies of pressure overload and left ventricular assist device treatment



**Figure 3. Comparisons of in vitro mechanical changes to in vivo studies of pressure overload and left ventricular assist device treatment (LVAD).** **A**, Decreased shortening in cultured papillary muscles and a 7-day TAC study in mice. **B**, Decreased mean stretch and LVAD. Numbers within each circle represent the number of significant genes.

## Appendix B. Supplementary data

*Supplementary Table 1. Genes regulated concordantly in response to reduced cyclic shortening in papillary muscles and TAC in mice*

Gene name	Gene symbol	Expression change	Mean Z-score
natriuretic peptide precursor B	Nppb	↑	10.50
collagen, type III, alpha 1	Col3a1	↑	10.12
actin, beta	Actb	↑	9.18
thymosin beta 4, X-linked	Tmsb4x	↑	8.66
ankyrin repeat domain 1 (cardiac muscle)	Ankrd1	↑	8.58
TIMP metalloproteinase inhibitor 1	Timp1	↑	8.54
actin, alpha 1, skeletal muscle	Acta1	↑	8.52
tropomyosin 4	Tpm4	↑	8.08
myosin binding protein C, fast-type	Mybpc2	↑	7.95
secreted protein, acidic, cysteine-rich (osteonectin)	Sparc	↑	7.91
retinol binding protein 1, cellular	Rbp1	↑	7.85
immediate early response 3	Ier3	↑	7.73
serine (or cysteine) peptidase inhibitor, clade E, member 2	Serpine2	↑	7.38
fatty acid binding protein 4, adipocyte	Fabp4	↑	7.30
biglycan	Bgn	↑	7.29
S100 calcium binding protein A11 (calizzarin)	S100a11	↑	7.26
mortality factor 4 like 2	Morf4l2	↑	7.26
collagen, type I, alpha 2	Col1a2	↑	7.24
interferon induced transmembrane protein 2	Ifitm2	↑	7.19
interferon induced transmembrane protein 3	Ifitm3	↑	7.02
thymosin, beta 10	Tmsb10	↑	6.99
B-cell translocation gene 2, anti-proliferative	Btg2	↑	6.67
Cd44 molecule	Cd44	↑	6.54
CD9 molecule	Cd9	↑	6.49
myosin, heavy chain 7, cardiac muscle, beta	Myh7	↑	6.45
plastin 3 (T-isoform)	Pls3	↑	6.37
solute carrier family 44, member 2	Slc44a2	↑	6.31
procollagen, type VI, alpha 3	Col6a3	↑	6.31
collagen, type I, alpha 1	Col1a1	↑	6.11
interferon-induced protein with tetratricopeptide repeats 2	Ifit2	↑	6.07

*Supplementary Table 1 (continued)*

Gene name	Gene symbol	Expression pattern	Mean Z-score
regulator of calcineurin 1	Rcan1	↑	6.02
fibrinogen-like 2	Fgl2	↑	6.01
nidogen 1	Nid1	↑	6.00
vascular cell adhesion molecule 1	Vcam1	↑	5.95
peptidylprolyl isomerase C	Ppic	↑	5.92
lectin, galactoside-binding, soluble, 3 binding protein	Lgals3bp	↑	5.87
tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, eta polypeptide	Ywhah	↑	5.75
actin related protein 2/3 complex, subunit 2	Arpc2	↑	5.69
ADP-ribosylation factor 6	Arf6	↑	5.67
connective tissue growth factor	Ctgf	↑	5.63
collagen, type VI, alpha 2	Col6a2	↑	5.53
complement component 1, q subcomponent, beta polypeptide	C1qb	↑	5.53
integrin beta 1 (fibronectin receptor beta)	Itgb1	↑	5.52
melanoma cell adhesion molecule	Mcam	↑	5.52
Versican	Vcan	↑	5.52
glutamate-ammonia ligase (glutamine synthetase)	Glul	↑	5.49
heat shock protein family, member 7 (cardiovascular)	Hspb7	↑	5.39
Tyro protein tyrosine kinase binding protein	Tyrobp	↑	5.32
caveolin 1, caveolae protein	Cav1	↑	5.31
cell division cycle 42 (GTP binding protein)	Cdc42	↑	5.31
calcium regulated heat stable protein 1	Carhsp1	↑	5.29
cyclin D1	Ccnd1	↑	5.23
glycoprotein m6b	Gpm6b	↑	5.22
PRELI domain containing 1	Prelid1	↑	5.21
septin 11	Sep11	↑	5.19
cysteine and glycine-rich protein 2	Csrp2	↑	5.14
matrix Gla protein	Mgp	↑	5.12
Fc fragment of IgE, high affinity I, receptor for; gamma polypeptide	Fcer1g	↑	5.10
Moesin	Msn	↑	5.09
tubulin, beta 6	Tubb6	↑	5.06
suppressor of cytokine signaling 3	Socs3	↑	4.97
eukaryotic translation initiation factor 4A1	Eif4a1	↑	4.97

*Supplementary Table 1 (continued)*

Gene name	Gene symbol	Expression pattern	Mean Z-score
serine (or cysteine) peptidase inhibitor, clade H, member 1	Serpinh1	↑	4.92
lymphocyte antigen 6 complex, locus E	Ly6e	↑	4.91
peptidylprolyl isomerase A (cyclophilin A)	Ppia	↑	4.89
guanine nucleotide binding protein (G protein), alpha inhibiting 2	Gnai2	↑	4.88
hematological and neurological expressed 1	Hn1	↑	4.87
cathepsin Z	Ctsz	↑	4.84
eukaryotic translation initiation factor 4E binding protein 1	Eif4ebp1	↑	4.80
tubulin, beta 5	Tubb5	↑	4.75
family with sequence similarity 46, member A	Fam46a	↑	4.73
calmodulin 1	Calm1	↑	4.68
tetraspanin 4	Tspan4	↑	4.65
plasminogen activator, tissue	Plat	↑	4.64
laminin, gamma 1	Lamc1	↑	4.62
mannan-binding lectin serine peptidase 1	Masp1	↑	4.59
collagen, type XV, alpha 1	Col15a1	↑	4.56
myosin, heavy chain 9, non-muscle	Myh9	↑	4.56
tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide	Ywhaz	↑	4.51
mesenchyme homeobox 1	Meox1	↑	4.50
dynein light chain LC8-type 1	Dynll1	↑	4.48
polypyrimidine tract binding protein 1	Ptbp1	↑	4.45
procollagen C-endopeptidase enhancer	Pcolce	↑	4.44
FXVD domain-containing ion transport regulator 5	Fxyd5	↑	4.42
fibronectin type III domain containing 1	Fndc1	↑	4.42
H2A histone family, member X	H2afx	↑	4.42
dpy-19-like 1 (C. elegans)	Dpy19l1	↑	4.41
collagen, type IV, alpha 1	Col4a1	↑	4.40
immediate early response 5	Ier5	↑	4.36
lectin, galactoside-binding, soluble, 1	Lgals1	↑	4.28
cytochrome b-245, alpha polypeptide	Cyba	↑	4.27
transmembrane protein 176B	Tmem176b	↑	4.24
interleukin 2 receptor, gamma	Il2rg	↑	4.23
elastin microfibril interfacer 1	Emilin1	↑	4.07

*Supplementary Table 1 (continued)*

Gene name	Gene symbol	Expression pattern	Mean Z-score
CD302 molecule	Cd302	↑	4.07
growth factor receptor bound protein 10	Grb10	↑	4.04
complement component 1, q subcomponent, C chain	C1qc	↑	4.03
complement component 1, q subcomponent, alpha polypeptide	C1qa	↑	4.02
transgelin 2	Tagln2	↑	4.02
serine carboxypeptidase 1	Scpep1	↑	4.01
serine (or cysteine) peptidase inhibitor, clade F, member 1	Serpinf1	↑	4.00
EF-hand domain family, member D2	Efhd2	↑	3.98
interferon gamma inducible protein 30	Ifi30	↑	3.97
S100 calcium binding protein A6	S100a6	↑	3.96
H2A histone family, member Z	H2afz	↑	3.95
capping protein (actin filament), gelsolin-like	Capg	↑	3.95
CUG triplet repeat, RNA binding protein 2	Cugbp2	↑	3.89
FK506 binding protein 1a	Fkbp1a	↑	3.88
epithelial membrane protein 3	Emp3	↑	3.86
SH3 domain binding glutamic acid-rich protein-like 3	Sh3bgrl3	↑	3.86
eukaryotic translation elongation factor 1 alpha 1	Eef1a1	↑	3.85
integrin alpha 9	Itga9	↑	3.84
S100 calcium binding protein A10	S100a10	↑	3.83
fibulin 2	Fbln2	↑	3.78
translocator protein	Tspo	↑	3.77
ADAM metallopeptidase domain 10	Adam10	↑	3.76
transglutaminase 2, C polypeptide	Tgm2	↑	3.66
cathepsin C	Ctsc	↑	3.66
cofilin 1, non-muscle	Cfl1	↑	3.64
EGF-containing fibulin-like extracellular matrix protein 2	Efemp2	↑	3.64
legumain	Lgmn	↑	3.62
annexin A4	Anxa4	↑	3.61
phosphoprotein enriched in astrocytes 15A	Pea15a	↑	3.61
selenoprotein	Sep15	↑	3.61
interferon-induced protein with tetratricopeptide repeats 3	Ifit3	↑	3.60
four and a half LIM domains 1	Fhl1	↑	3.60

Supplementary Table 1 (continued)

Gene name	Gene symbol	Expression pattern	Mean Z-score
caspace 8	Casp8	↑	3.60
GTPase, IMAP family member 4	Gimap4	↑	3.60
receptor (G protein-coupled) activity modifying protein 2	Ramp2	↑	3.56
SMAD family member 1	Smad1	↑	3.55
ras homolog gene family, member C	Rhoc	↑	3.53
actin, gamma 1	Actg1	↑	3.53
disabled homolog 2, mitogen-responsive phosphoprotein (Drosophila)	Dab2	↑	3.52
transketolase	Tkt	↑	3.51
transforming growth factor, beta receptor II	Tgfr2	↑	3.50
poly(A) binding protein, cytoplasmic 1	Pabpc1	↑	3.43
platelet/endothelial cell adhesion molecule 1	Pecam1	↑	3.42
dual specificity phosphatase 3	Dusp3	↑	3.40
abl-interactor 1	Abi1	↑	3.34
uncoupling protein 2 (mitochondrial, proton carrier)	Ucp2	↑	3.33
myotrophin	Mtpn	↑	3.30
granulin	Grn	↑	3.27
protein disulfide isomerase family A, member 3	Pdia3	↑	3.23
DNA (cytosine-5-)-methyltransferase 1	Dnmt1	↑	3.22
stem-loop binding protein	Slbp	↑	3.18
CAP, adenylate cyclase-associated protein 1 (yeast)	Cap1	↑	3.17
Harvey rat sarcoma virus oncogene, subgroup R	Rras	↑	3.16
TIMP metalloproteinase inhibitor 2	Timp2	↑	3.16
TIMP metalloproteinase inhibitor 3	Timp3	↑	3.14
cytochrome b5 reductase 3	Cyb5r3	↑	3.10
heparin-binding EGF-like growth factor	Hbegf	↑	3.02
AE binding protein 1	Aebp1	↑	3.02
collagen, type IV, alpha 2	Col4a2	↑	2.98
transmembrane 9 superfamily member 3	Tm9sf3	↑	2.95
cortactin	Ctnn	↑	2.94
tetraspanin 14	Tspan14	↑	2.91
C1q and tumor necrosis factor related protein 6	C1qtnf6	↑	2.91
transcription factor 4	Tcf4	↑	2.87

*Supplementary Table 1 (continued)*

Gene name	Gene symbol	Expression pattern	Mean Z-score
growth arrest and DNA-damage-inducible, gamma	Gadd45g	↑	2.87
aldehyde dehydrogenase 1 family, member A1	Aldh1a1	↑	2.85
enhancer of rudimentary homolog (Drosophila)	Erh	↑	2.76
growth arrest specific 5	Gas5	↑	2.76
filamin, beta	Flnb	↑	2.71
platelet factor 4	Pf4	↑	2.70
tetraspanin 2	Tspan2	↑	2.67
metadherin	Mtdh	↑	2.65
myosin, light chain 9, regulatory	Myl9	↑	2.58
myosin IE	Myo1e	↑	2.48
tweety homolog 2 (Drosophila)	Ttyh2	↑	2.42
solute carrier family 41, member 3	Slc41a3	↓	6.17
peroxisomal biogenesis factor 11 alpha	Pex11a	↓	5.42
t-complex 11 (mouse) like 2	Tcp11l2	↓	4.74
desmocollin 2	Dsc2	↓	4.46
microphthalmia-associated transcription factor	Mitf	↓	4.22
tubulin, alpha 8	Tuba8	↓	4.11
ankyrin repeat and SOCS box-containing 11	Asb11	↓	3.67
solute carrier family 7 (cationic amino acid transporter, y+ system), member 1	Slc7a1	↓	3.55
Eph receptor A4	Epha4	↓	3.52
branched chain keto acid dehydrogenase E1, beta polypeptide	Bckdhb	↓	3.52
X-linked myotubular myopathy gene 1	Mtm1	↓	3.44
cyclin-dependent kinase inhibitor 1C (P57)	Cdkn1c	↓	3.43
fragile X mental retardation, autosomal homolog 1	Fxr1	↓	3.14
ryanodine receptor 2, cardiac	Ryr2	↓	3.11
LUC7-like 2 (S. cerevisiae)	Luc7l2	↓	3.07
transcription factor myocardin	Myocd	↓	3.07
gap junction protein, alpha 1	Gja1	↓	2.83
alcohol dehydrogenase, iron containing, 1	Adhfe1	↓	2.79
pyruvate dehydrogenase phosphatase isoenzyme 2	Pdp2	↓	2.66
reticulon 4 interacting protein 1	Rtn4ip1	↓	2.40

*Supplementary Table 2. Genes regulated concordantly in response to reduced mean stretch in vitro and clinical LVAD therapy*

Gene Name	Gene Symbol	Expression Change	Mean Z-score
DNA-damage-inducible transcript 4	Ddit4	↑	9.764
nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha	Nfkbia	↑	7.984
B-cell translocation gene 2, anti-proliferative	Btg2	↑	7.375
CCAAT/enhancer binding protein (C/EBP), beta	Cebpb	↑	7.315
glycoprotein m6b	Gpm6b	↑	6.996
natriuretic peptide precursor B	Nppb	↑	6.601
phosphoinositide-3-kinase, regulatory subunit 1 (alpha)	Pik3r1	↑	6.138
secreted protein, acidic, cysteine-rich (osteonectin)	Sparc	↑	6.025
choline phosphotransferase 1	Chpt1	↑	5.937
myosin, heavy chain 6, cardiac muscle, alpha	Myh6	↑	5.845
metallothionein 2A	Mt2A	↑	5.715
collagen, type III, alpha 1	Col3a1	↑	5.677
cytochrome P450, family 4, subfamily b, polypeptide 1	Cyp4b1	↑	5.509
phosphatidic acid phosphatase type 2B	Ppap2b	↑	5.501
DnaJ (Hsp40) homolog, subfamily B, member 5	Dnajb5	↑	5.448
ras homolog gene family, member B	Rhob	↑	5.206
mitochondrial ribosomal protein S25	Mrps25	↑	5.075
cyclin-dependent kinase inhibitor 1A (p21, Cip1)	Cdkn1a	↑	5.025
zinc finger protein 36	Zfp36	↑	4.951
v-maf musculoaponeurotic fibrosarcoma oncogene homolog F (avian)	Maff	↑	4.904
microtubule-associated protein 4	Map4	↑	4.802
glutamate-ammonia ligase (glutamine synthetase)	Glul	↑	4.697
HOP homeobox	Hopx	↑	4.630
nuclear factor I/B	Nfib	↑	4.617
histidine triad nucleotide binding protein 3	Hint3	↑	4.523
GRAM domain containing 3	Gramd3	↑	4.396
TATA box binding protein-like 1	Tbpl1	↑	4.356
fibronectin type III domain containing 1	Fndc1	↑	4.345
LSM12 homolog ( <i>S. cerevisiae</i> )	Lsm12	↑	4.298
Kv channel-interacting protein 2	Kcnip2	↑	4.274
ligand of numb-protein X 1	Lnx1	↑	4.267

*Supplementary Table 2 (continued)*

Gene name	Gene symbol	Expression pattern	Mean Z-score
methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 2, methenyltetrahydrofolate cyclohydrolase	Mthfd2	↑	4.230
Janus kinase 2	Jak2	↑	4.209
homeodomain interacting protein kinase 2	Hipk2	↑	4.208
complement component 1, q subcomponent, C chain	C1qc	↑	4.153
procollagen, type VI, alpha 3	Col6a3	↑	4.133
septin 11	Sep11	↑	4.060
phosphatase and actin regulator 1	Phactr1	↑	4.046
dedicator of cytokinesis 9	Dock9	↑	4.013
RRN3 RNA polymerase I transcription factor homolog ( <i>S. cerevisiae</i> )	Rrn3	↑	4.012
complement component 1, q subcomponent, beta polypeptide	C1qb	↑	3.813
mitochondrial tumor suppressor 1	Mtus1	↑	3.761
thioredoxin interacting protein	Txnip	↑	3.755
cytoplasmic polyadenylation element binding protein 4	Cpeb4	↑	3.736
triadin	Trdn	↑	3.722
aquaporin 1	Aqp1	↑	3.672
mitogen-activated protein kinase kinase kinase 6	Map3k6	↑	3.619
nicotinamide phosphoribosyltransferase	Nampt	↑	3.560
cysteine-serine-rich nuclear protein 1	Csrnp1	↑	3.560
pleiotrophin	Ptn	↑	3.546
collagen, type IV, alpha 1	Col4a1	↑	3.523
serine/threonine kinase 17b	Stk17b	↑	3.518
Kruppel-like factor 15	Klf15	↑	3.514
transmembrane protein 30A	Tmem30a	↑	3.505
collagen, type XV, alpha 1	Col15a1	↑	3.393
insulin receptor substrate 2	Irs2	↑	3.334
TIMP metalloproteinase inhibitor 3	Timp3	↑	3.232
growth arrest-specific 2 like 1	Gas2l1	↑	3.091
phosphorylase, glycogen, muscle	Pygm	↑	3.071
lipin 1	Lpin1	↑	3.051
Cd47 molecule	Cd47	↑	3.012
cartilage associated protein	Crtap	↑	2.952
nuclear factor I/A	Nfia	↑	2.923

Supplementary Table 2 (continued)

Gene name	Gene symbol	Expression pattern	Mean Z-score
spondin 1, extracellular matrix protein	Spon1	↑	2.907
transducer of ERBB2, 2	Tob2	↑	2.859
phosphatase and tensin homolog	Pten	↑	2.833
ankyrin repeat domain 23	Ankrd23	↑	2.745
TSC22 domain family, member 1	Tsc22d1	↑	2.675
AE binding protein 1	Aebp1	↑	2.542
basic helix-loop-helix family, member e40	Bhlhe40	↓	8.871
natriuretic peptide precursor A	Nppa	↓	5.954
vascular endothelial growth factor C	Vegfc	↓	5.481
nuclear receptor subfamily 4, group A, member 3	Nr4a3	↓	5.282
protein phosphatase 1, regulatory (inhibitor) subunit 3C	Ppp1r3c	↓	5.154
solute carrier family 39 (iron-regulated transporter), member 1	Slc40a1	↓	4.811
LIM and senescent cell antigen-like domains 1	Lims1	↓	4.651
SERTA domain containing 4	Sertad4	↓	4.323
prostaglandin I2 (prostacyclin) synthase	Ptgis	↓	4.187
cytidine monophosphate (UMP-CMP) kinase 2, mitochondrial	Cmpk2	↓	3.933
protocadherin 18	Pcdh18	↓	3.834
solute carrier family 5 (inositol transporters), member 3	Slc5a3	↓	3.817
phosphodiesterase 4D, cAMP-specific (phosphodiesterase E3 dunce homolog, <i>Drosophila</i> )	Pde4d	↓	3.732
myosin Va	Myo5a	↓	3.648
latexin	Lxn	↓	3.644
polo-like kinase 2 ( <i>Drosophila</i> )	Plk2	↓	3.623
mitogen-activated protein kinase 1 interacting protein 1-like	Mapk1ip1l	↓	3.610
beta-1,3-N-acetylgalactosaminyltransferase 1	B3galnt1	↓	3.608
solute carrier family 7 (cationic amino acid transporter, y+ system), member 1	Slc7a1	↓	3.574
ganglioside-induced differentiation-associated-protein 2	Gdap2	↓	3.459
ribosomal protein S6 kinase polypeptide 3	Rps6ka3	↓	3.450
frizzled homolog 6 ( <i>Drosophila</i> )	Fzd6	↓	3.430
Cd55 molecule	Cd55	↓	3.370
radical S-adenosyl methionine domain containing 2	Rsad2	↓	3.316
lipopolysaccharide-induced TNF factor	Litaf	↓	3.277

*Supplementary Table 2 (continued)*

Gene name	Gene symbol	Expression pattern	Mean Z-score
prostaglandin E receptor 4 (subtype EP4)	Ptger4	↓	3.273
WW domain containing transcription regulator 1	Wwtr1	↓	2.875
RELT-like 1	Rel1	↓	2.821
zinc finger, CCHC domain containing 24	Zcchc24	↓	2.806
zinc finger protein 521	Znf521	↓	2.790
gap junction protein, alpha 1	Gja1	↓	2.788
BCL2/adenovirus E1B 19 kDa-interacting protein 3	Bnip3	↓	2.783
ADAM metallopeptidase with thrombospondin type 1 motif, 1	Adamts1	↓	2.510

Chapter 3 Title Page - accepted for publication in the Journal of Molecular and Cellular Cardiology

**Full title: Sequencing of mRNA identifies re-expression of fetal splice variants in cardiac hypertrophy**

Authors: Ames EG (BS)<sup>a,c</sup>, Lawson MJ (PhD)<sup>d,1</sup>, Mackey AJ (PhD)<sup>d</sup>, Holmes JW (MD PhD)<sup>a,b,e</sup>

Affiliations: <sup>a</sup> Department of Biomedical Engineering, University of Virginia (Box 800759, Charlottesville, VA, USA 22908)

<sup>b</sup> Robert M. Berne Cardiovascular Research Center, University of Virginia (Robert M. Berne Cardiovascular Research Center, 415 Lane Rd, Charlottesville, VA, USA 22908)

<sup>c</sup> Department of Molecular Physiology and Biological Physics, University of Virginia (Jordan Hall, PO Box 800736, Charlottesville, VA USA 22908)

<sup>d</sup> Center for Public Health Genomics, Department of Public Health Sciences, University of Virginia (P.O. Box 800717, Charlottesville, VA USA 22908)

<sup>e</sup> Department of Medicine, University of Virginia (PO Box 800793, Charlottesville, VA USA 22908)

Present addresses: <sup>1</sup> HemoShear LLC, (1115 5th Street Southwest Charlottesville, VA USA 22902)

Email addresses: Ames EG (ega2d@virginia.edu), Lawson MJ (mlawsonvt09@gmail.com), Mackey AJ (amackey@virginia.edu), Holmes JW (holmes@virginia.edu)

Correspondence to Jeffrey W. Holmes, Department of Biomedical Engineering, University of Virginia, Box 800759, Charlottesville, VA 22908. Fax: 434-982-3870 Telephone: 434-243-6321 Email: holmes@virginia.edu

Data availability: Readers are able to access raw and processed sequencing files from Gene Expression Omnibus on May 14, 2013 (Accession ID: GSE42411).

## Abstract

Cardiac hypertrophy has been well-characterized at the level of transcription. During cardiac hypertrophy, genes normally expressed primarily during fetal heart development are re-expressed, and this fetal gene program is believed to be a critical component of the hypertrophic process. Recently, alternative splicing of mRNA transcripts has been shown to be temporally regulated during heart development, leading us to consider whether fetal patterns of splicing also reappear during hypertrophy. We hypothesized that patterns of alternative splicing occurring during heart development are recapitulated during cardiac hypertrophy. Here we present a study of isoform expression during pressure-overload cardiac hypertrophy induced by 10 days of transverse aortic constriction (TAC) in rats and in developing fetal rat hearts compared to sham-operated adult rat hearts, using high-throughput sequencing of poly(A) tail mRNA. We find a striking degree of overlap between the isoforms expressed differentially in fetal and pressure-overloaded hearts compared to control: forty-four percent of the isoforms with significantly altered expression in TAC hearts are also expressed at significantly different levels in fetal hearts compared to control ( $P < 0.001$ ). The isoforms that are shared between hypertrophy and fetal heart development are significantly enriched for genes involved in cytoskeletal organization, RNA processing, developmental processes, and metabolic enzymes. Our data strongly support the concept that mRNA splicing patterns normally associated with heart development recur as part of the hypertrophic response to pressure overload. These findings suggest that cardiac hypertrophy shares post-transcriptional as well as transcriptional regulatory mechanisms with fetal heart development.

Keywords: heart development; cardiac hypertrophy; alternative splicing; RNAseq

## 1. Introduction

Over twenty years ago, hearts undergoing hypertrophy were found to revert to a gene expression pattern normally associated with fetal heart development [2,21]. Collectively this is known as the fetal gene program. Many of the genes considered part of the fetal gene program encode critical components of the sarcomere or enzymes involved in metabolism. Hallmarks of the fetal gene program include upregulation of atrial natriuretic peptide [33],  $\beta$ -myosin heavy chain [158–160], and skeletal  $\alpha$ -actin [161], as well as downregulation of metabolic genes such as glucose transporter GLUT4 (relative to GLUT1 expression [162]), and enzymes critical for fatty acid oxidation [21,163]. The re-expression of the fetal gene program is thought to be a protective mechanism for the heart. During hypertrophy, the heart shifts from an oxygen-rich environment with fatty acids available for energy to an oxygen-poor environment with glucose as the main energy source, and expresses fetal genes that are better suited to function in this different metabolic environment [21]. The extent and regulation of this phenomenon has been studied and reviewed extensively [2,21,164,165]. While re-expression of the fetal gene program has been observed and documented for years, more recently other post-transcriptional and epigenetic mechanisms have also been shown to be critical for cardiac hypertrophy [39,166–169]. With this new knowledge, it seems plausible that the changes observed in cardiomyocytes during hypertrophy arise from orchestrated phenotypic switching on multiple levels of regulation.

The process of removing introns from a premature mRNA transcript occurs in most eukaryotes, but not every transcript is processed the same way. The selective inclusion or exclusion of specific exons or parts of exons is referred to as alternative splicing. Alternative splicing has the ability to change gene expression, coding sequence, translational efficiency,

and mRNA localization [170]. Sequencing the human genome revealed that there are a surprisingly small number of protein-encoding genes given the diversity of proteins and enzymes required for the many specialized functions observed in different tissues. It is now clear that some of this diversity is produced by the process of alternative splicing [45,171]. In fact, more than 95% of human genes are alternatively spliced [46]. Splicing has been documented to change in a wide variety of developmental, physiological, and disease processes [8,43]. More recently alternative splicing has been shown to play a critical role in heart development [70–72,90,91] and to be altered in cardiac pathologies including hypertrophy [166,172,173] and heart failure [114,116,118].

Throughout organogenesis and after birth, the heart must adapt to an increased workload caused by changing pressures and resistances within the adult vascular system. In 2008, Kalsotra et al. showed that cardiac development involves a specific and coordinated program of alternative splicing events, greater than 60% of which were conserved between mouse and chicken embryonic hearts. To put this number in context, less than 20% of all alternative exon (cassette-type) splicing events are conserved between human and mouse. This high degree of conservation strongly suggests a functional role for alternative splicing [7]. In addition, deletion of specific splicing factors that control alternative splicing in heart development has shown that proper regulation of splicing is needed for normal heart development [70–72,90,91].

Alternative splicing has been shown to change critical properties of cardiomyocytes including compliance [108,110], protein-protein interactions [166], calcium handling, and contractility [90]. A more thorough understanding of how alternative splicing contributes to cardiac hypertrophy could not only increase our understanding of how RNA processing changes in response to physiological cues, but could also potentially contribute to identifying

novel drug targets to modulate cardiac remodeling. Accordingly, the main goal of this study was to test the hypothesis that specific fetal splicing patterns are re-expressed in surgically-induced pressure-overload hypertrophy in rats using high-throughput sequencing (HTS or RNAseq) to quantify poly(A) tail mRNA species. We compared mRNA isolated from ventricles of hearts undergoing pressure-overload hypertrophy, sham-operated hearts, and fetal hearts and identified the splicing events that occur in both fetal and hypertrophied hearts, but not sham-operated adult hearts.

## **2. Materials and methods**

Detailed methods for all procedures and analyses are provided on pages 81-86.

### *2.1. Animals*

All studies were performed in accordance with the Guide for the Care and Use of Laboratory Animals [130] and approved by the University of Virginia's Institutional Animal Care and Use Committee. A total of 23 Sprague-Dawley rats were used for these studies.

### *2.2. Tissue harvest from fetal and adult rats*

Five rats were timed, pregnant females; fetuses were harvested by caesarean section on gestational day 18 and the dams were sacrificed immediately afterwards. Fetal hearts from each litter were pooled; three of these litters were used for RNAseq experiments. Four unoperated adult male rats ( $365 \pm 15$  g) served as adult controls in splicing validation assays.

### *2.3. Minimally-invasive approach to transverse aortic constriction*

Seven males ( $308 \pm 35$  g) underwent the transverse aortic constriction (TAC) procedure and seven males ( $306 \pm 28$  g) were used as sham-operated controls. Three rats from the TAC and sham groups were used for high-throughput sequencing. Three separate rats from both TAC and sham groups were used for histology. We adapted a minimally-invasive approach to transverse aortic constriction originally described by Pu et al. in mice [174]. Adult male Sprague-Dawley rats were subjected to TAC using a 4-0 silk suture tied around a 19-gauge needle [175]. Ten days after surgery, rats were euthanized using pentobarbital and left ventricular tissue was harvested; this time point was chosen because previous work has shown that this time point after aortic banding has the largest number of gene expression changes [176]. In addition to heart weight-to-body weight ratio, the extent of hypertrophy was confirmed on the cellular level by measuring myocyte cross-sectional area in approximately 1000 myocytes per animal.

### *2.4. RNAseq data generation and analysis*

Five micrograms of high-quality poly(A) tail mRNA was used as starting material for the Illumina mRNA seq library preparation kit and was prepared to manufacturer's directions (Illumina). Libraries were sequenced on the Illumina Genome Analyzer IIx as 42 or 63 base pair paired-end reads. Tophat v1.3.1 [177] was used to align all reads including junction-spanning reads back to the rat genome (Ensembl RGSC3.4). Cufflinks v1.0.3 [178] was used to identify differential splicing, promoter usage, and gene expression changes between experimental groups. Using similar criteria to Lee et al. [172], we defined statistical significance in expression as  $> 1.5$  absolute fold-change,  $q$ -value (an adjusted  $p$ -value for multiple testing)  $< 0.05$ , and Fragments Per Kilobase of exon per Million reads mapped (FPKM)  $> 3$ ; FPKM is a measure of expression used in high-throughput sequencing data

that is normalized for both transcript length and total number of reads sequenced. All bioinformatics analyses and comparisons were implemented and performed using in-house scripts written in Unix, R, Python, or Perl.

### *2.5. PCR validation*

We validated our RNAseq results against a previous study of alternative splicing in heart development [7]. Splice variants in fetal compared to adult hearts identified both by Kalsotra et al. [7] and by our analysis were validated by one step PCR starting with 500 ng RNA and amplified for 25 cycles (Invitrogen). Custom primers (Integrated DNA Technologies) were designed to flank the alternatively spliced region. PCR products were electrophoresed and visualized with ethidium bromide-stained 2-5% Agarose-1000 gels (Invitrogen). Gel band intensities were measured using BioRad ImageLab software. The percentage inclusion of the alternative region was calculated as [inclusion band intensity / (inclusion band intensity + exclusion band intensity)] x 100. Percentage inclusion values were arcsine-transformed for statistical testing [179]. For validation of previously reported splicing differences between fetal and adult samples, two-tailed unpaired *t*-tests were used. Comparison of inclusion levels across TAC, fetal, and sham groups employed analysis of variance with Newman-Keuls post-tests where appropriate.

## **3. Results**

### *3.1. Successful induction of hypertrophy in adult rats*

The minimally-invasive approach to transverse aortic constriction [174] employed here induced significant hypertrophy in rats by post-operative day 10, as demonstrated by a significantly higher heart weight to body weight (HW:BW) ratio (Fig. 1A) and significantly

larger cardiomyocyte cross-sectional area (Fig. 1B) in TAC animals compared to sham-operated controls. The average TAC HW:BW per 1000g of tissue was greater ( $4.29 \pm 0.56$ ) than the average for the sham group ( $3.70 \pm 0.16$ ). The average cross-sectional area of the TAC group was larger ( $336.7 \pm 3.8$  pixels) than the sham group ( $293.2 \pm 3.5$  pixels).

### *3.2. Similarities in gene expression between hypertrophy and heart development*

In total, our data set included 267,727,027 reads sequenced across our three experimental groups (Table 1). The concept that a fetal gene program is re-expressed during cardiac hypertrophy suggests that our analysis should identify a number of genes whose expression is increased or decreased in both TAC and fetal hearts relative to sham-operated adults. We evaluated 18 genes that are commonly mentioned as members of the fetal gene program and found that 14 of the 18 known members of the fetal gene program [21,33] changed in the same direction in both TAC and fetal compared to sham, with the fetal changes typically of greater magnitude (Fig. 2A). Plotting fold changes in fetal hearts against those induced by TAC for all significantly regulated genes revealed as expected that a large number of genes differentially regulated in fetal hearts are not altered by TAC (Fig. 2B, dark blue clusters along y-axis). A relatively small number of genes are significantly regulated in TAC alone (Fig. 2B, points along x-axis), while a substantial number of genes change in the same direction in both groups (Fig. 2B, points in upper right and lower left quadrants). Of 221 genes significantly upregulated in TAC compared to sham, almost 25% (54 genes) were also upregulated in fetal hearts compared to sham (Fig. 2C). 26,640 genomic locations were expressed above 3 FPKM in at least one of our experimental groups. These genomic locations include unannotated genes, noncoding RNAs, and microRNA genes. If each of these genes were independently regulated, then the upregulation of 9.1% (2,423/26,640) of those genes in fetal hearts and 0.83% (221/26,640)

of those genes in TAC hearts we observed would suggest that roughly 20 genes should be upregulated in both groups ( $0.091 \times 0.0083 \times 26,640$ ). In fact, we observed more than twice as much overlap. As detailed in Supplementary Fig. 1, the probability of observing 54 or more commonly upregulated genes in the absence of shared regulatory mechanisms would be  $P = 1.16 \times 10^{-11}$ , while the probability of the degree of shared downregulation observed in these two groups would be  $3.5 \times 10^{-90}$ .

### *3.3. Changes in gene and isoform expression are independent in both fetal and hypertrophic hearts*

Our RNAseq analysis can detect significant changes in expression of an isoform due to alternative splicing, or due to changes in overall gene expression levels for that gene. We therefore examined what fraction of detected changes in isoform expression was at least partly explained by changes in gene expression (Fig. 3). In comparisons of fetal and sham hearts, we detected 10,553 significant changes in isoform expression, 53.5% of which were associated with significant, concordant changes in gene expression (i.e., isoform and gene expression levels both increased or decreased); 46.5% of changes in isoform expression occurred in the absence of detected changes in gene expression (Fig. 3A). Comparisons of TAC vs. sham hearts were notable for the much smaller number of both isoform and gene expression changes, and for a much higher percentage (77.5%) of isolated changes in isoform expression (Fig. 3B).

### *3.4. Comparison of splicing events in hypertrophy and heart development identifies significant overlap*

The Euler diagram of significantly altered isoform expression in Fig. 4A reveals a striking overlap between TAC and fetal groups; 44% of isoforms significantly altered in TAC showed concordant changes in fetal hearts. When we exclude all single isoform genes, we

measured 23,190 isoforms that were expressed above 3 FPKM in at least one of our experimental groups. If each of these isoforms were independently regulated, then downregulation of 22.3% (5,164/23,190) of those isoforms in fetal hearts and 8.1% (1,885/23,190) of those isoforms in TAC hearts would suggest that approximately 421 isoforms should be downregulated in both groups ( $0.08 \times 0.22 \times 23,190$ ). The amount of overlap that we observed was more than double this number. The probability of observing 934 or more commonly downregulated isoforms in the absence of shared regulatory mechanisms would be  $P = 3.42 \times 10^{-164}$ , while the probability of the degree of shared upregulation observed in these two groups would be  $8.08 \times 10^{-25}$  (see Supplementary Fig. 1 for calculations). The hexagonal histogram of isoform fold changes in Fig. 4B shows the distribution of significant splicing events and similarities between the TAC and fetal groups. Here we see that many of the isoforms differentially expressed in either group relative to sham are located in the lower left or upper right quadrants, demonstrating concordant changes in expression. Gene Ontology (GO) analysis revealed that these groups of shared up- and down-regulated isoforms are enriched for GO terms associated with sarcomeric and cytoskeletal organization, RNA processing and metabolism, and muscle contraction (Table 2). All significant isoforms of genes within the 'cytoskeleton organization' cluster identified by DAVID (FPKM > 3, q-value < 0.05, and log<sub>2</sub> fold change > 1.5) are plotted as heat map in Fig. 5A to show similarities in expression. Indeed, 72% of isoforms significantly altered in TAC are concordantly altered in fetal hearts. The shared isoforms with the most significant changes included: Ablim1, Ctnnb1, and Cryab (listed in Fig. 5A); Ptpb1, Hnrnpk, and Rbm39 (RNA processing); and Fxyd1 and Myh6 (muscle contraction). Many of these terms have previously been identified as associated with hypertrophic growth or as targets of alternative splicing [7,64,173].

### *3.5. Differences between splicing patterns observed in fetal heart development and TAC*

Despite the striking degree of overlap between splicing patterns observed in TAC and fetal hearts relative to sham, half the changes observed in TAC hearts were unique. To gain insight into the potential functional impact of shared and unique splicing patterns, we separated our significant isoform changes into those that occurred in both TAC and fetal hearts, those that only occurred in TAC, and those that only occurred in development. We then used gene ontology analysis to look for differences between the groups. We found that the isoforms specific to development or TAC had unique features not seen in the overlap group. Most notably, isoform expression changes that occurred only in TAC hearts were enriched for components of signal transduction pathways, including regulation of small GTPase-mediated signal transduction and positive regulation of MAP kinase kinase kinase (Supplementary Table 3); some of the most significantly altered isoforms unique to TAC included *Arfgap2*, *Cyth2*, and *Rapgef2* (others listed in Fig. 5B). Most of the isoforms in this cluster are significant in the TAC group only. By contrast, fetal-specific isoform changes were enriched in cell cycle and DNA maintenance (i.e. response to DNA damage, chromosome segregation, and chromatin organization; Supplementary Table 3). Some of the most significantly altered isoforms unique to fetal hearts included *Dlg1*, *Fbxo5*, and *Gmnn* (others listed in Fig. 5C). The isoform expression changes of the top group-specific gene ontology clusters are depicted as heat maps in Fig. 5.

#### *3.6.1. PCR validation of significant isoform identifies both developmentally- and hypertrophy-related alternative splicing*

One advantage of RNAseq is the ability to identify multiple complex splicing patterns within a single gene. As an individual example of multiple splicing patterns, the tropomyosin

3 gene (Tpm3) contains mutually exclusive exons 5a or 5b and three alternative terminal exons (Fig. 4C). Previous literature has shown that the mutually exclusive exon 5 is developmentally regulated, i.e. fetal and adult tissues express different exons. Our RNAseq analysis detected both developmental regulation of exon 5 (higher levels of exon 5a in both adult groups compared to fetal), in agreement with previous reports [180], and hypertrophy-associated reversion to a fetal pattern of terminal exon splicing (lower levels of terminal exon A in TAC and fetal groups compared to sham). We confirmed both changes by PCR (Fig. 4D). Ninety percent and 96% of Tpm3 in TAC and sham hearts contain exon 5a, respectively, while fetal hearts express exon 5a in 73% of Tpm3 transcripts (Fig. 4D). The predominant terminal exon used in the sham group is terminal exon A, referred to as TermA in Fig. 4D which makes up  $51.85 \pm 8.66\%$  of all terminal exons expressed in the sham group. This terminal exon A is not the predominant terminal exon in TAC and fetal groups and is expressed at a significantly lower level ( $22.81 \pm 7.45\%$  in fetal hearts and  $23.36 \pm 10.73\%$  in TAC hearts).

### *3.6.2. PCR validation of RNAseq results identifies known splice variants that are regulated in heart development*

To further validate our RNAseq findings, we compared differences in alternative splicing between fetal and adult hearts detected by our analysis to published reports on alternative splicing in heart development, and confirmed individual alternative splicing events by PCR. Kalsotra et al. identified developmental isoform switches that occur during murine heart development by measuring percentage alternative exon inclusion on embryonic days 14 and 18 (E14 and E18, respectively), postnatal day 1, and adulthood. Of the 48 isoform switches that had percent inclusion changes  $>20\%$ , 33 splicing events changed more dramatically from E18 to adulthood than from E14 to E18. The fetal hearts in

our study were harvested at embryonic day 18 as well, which allowed us to directly compare significant differences between fetal and adult rat hearts identified by our analysis to those identified by Kalsotra using exon-specific microarrays. Of the 33 events identified by Kalsotra with large changes between E18 and adulthood, 17 events in our dataset had similar developmental trends in exon inclusion percentage as estimated from FPKM values. Twelve of these events were detected and reported as statistically significant in our bioinformatics analysis. PCR validation confirmed statistically significant changes in percent inclusion for all 12 of these splicing events (Fig. 6).

### *3.7. Potential upstream regulators of isoform changes seen in heart development, hypertrophy, or both processes*

We used Ingenuity Pathway Analysis (IPA; Ingenuity Systems, Inc.; Redwood City, CA) to identify potential upstream regulators of the significant isoform expression changes (Supplementary Table 4). This analysis incorporates information on signaling transduction pathways, protein-protein interactions, and transcriptional regulation to determine molecules that are upstream of a given list of gene names. We used this feature to identify candidate signaling pathways that might be controlling the changes in isoform expression that we observed in TAC and fetal hearts. There was a large amount of overlap between the upstream regulators of hypertrophy and fetal heart development including insulin receptor (INSR), insulin growth factor receptor (IGFR1), mediator of RNA polymerase II transcription subunit 30 (MED30), transcription factor GATA4, titin (TTN), and growth hormone (GH1). There were several upstream regulators that were specific to TAC including both of the p38 MAPK subunits (MAPK11 and MAPK14), aldosterone receptor (NR3C2), and MURF3, a protein that links sarcomeres to microtubules (also known as tripartite motif-containing 54, TRIM54) [99]. There were more upstream regulators that were unique to the isoform

expression changes seen in fetal hearts including E2F transcription factor 6 (E2F6), MAP kinase kinases 3 and 6 (MAP2K3, MAP2K3), histone deacetylase 5 (HDAC5), retinoblastoma protein (RB1), myocardin (MYOCD), thrombospondin receptor (CD36), and myocyte-specific enhancer factor 2D (MEF2D).

#### **4. Discussion**

The changes in individual mRNA isoforms we identified during TAC displayed a striking degree of overlap with changes observed in fetal hearts relative to sham; nearly half of all significant splicing changes we observed in TAC were present and concordant in fetal hearts. These shared isoforms are enriched in gene ontology groups known to be targets of alternative splicing, including sarcomeric and metabolic proteins [7,64,173]. Interestingly, whereas many splicing events identified in fetal hearts were associated with altered overall expression levels of the associated gene, a much greater fraction of the splicing events we identified in TAC samples occurred in the absence of gene expression changes. These observations suggest that alternative splicing plays a relatively greater role in determining mRNA changes during hypertrophy, and relies at least in part on regulatory mechanisms also active during development. Other groups have examined the changes in alternative splicing in murine models of hypertrophy and how these events change with the development of heart failure [172]. Indeed it is important to note that alternative splicing does change in heart failure patients [118] and in response to drugs used to treat heart failure [181]. In addition, recent *in vitro* experiments have shown that alternative splicing in skeletal muscle cell culture is sensitive to changes in mechanical cues [182].

A wide array of studies has now measured gene expression during heart development, cardiac hypertrophy, and heart failure using microarray analysis. Recently,

Dewey et al. combined data from 478 microarray studies and used coexpression network analysis to determine how similar gene expression was during heart development compared to hypertrophied or failing hearts. Given the small number of coexpression modules, the authors questioned the extent and importance of the fetal gene program [183]. Although the focus of our study was alternative splicing, changes in gene expression levels computed from our analysis are consistent with the concept that some developmental patterns of gene expression do recur during hypertrophy. Not surprisingly, many more genes were significantly up- or down-regulated in fetal hearts than in TAC hearts relative to sham-operated adult controls; therefore, the fraction of fetal gene expression changes shared by TAC hearts was small. However, the fraction of the genes significantly altered in TAC that were coordinately regulated in fetal hearts was much higher than would be expected in the absence of shared regulatory mechanisms, suggesting that hypertrophying hearts do re-use some of the same regulatory mechanisms active during heart development.

While this study revealed substantial overlap between splicing patterns in TAC and fetal hearts, both TAC and fetal hearts also displayed many unique splicing differences compared to sham-operated adult hearts. In order to gain preliminary insight into upstream regulatory mechanisms that might explain the similarities and differences we observed between TAC and fetal hearts, we utilized Ingenuity Pathway Analysis. The RNA-binding proteins that control alternative splicing are regulated by canonical signaling cascades and common post-translational modifications [184,185] that would be identified by IPA. The upstream regulatory pathway most strongly associated with changes in isoform expression unique to TAC was p38 MAPK, while shared isoform changes were most strongly associated with insulin signaling. In addition, the upstream regulators of fetal-specific isoforms are known to be critical for heart development, such as myocardin and MEF2D.

However, more work is certainly needed to identify RNA binding proteins that bind to significantly altered isoforms in our dataset and understand their role in regulating the changes we observed. In order to begin to identify the functional impact of the splicing patterns reported here, we used gene ontology (GO) analysis to identify functional groups overrepresented among isoforms altered in TAC alone, fetal alone, or both fetal and TAC groups. Not surprisingly, changes in fetal hearts, which are still undergoing cell division [186], included significant upregulation of many more isoforms related to cell cycle. In contrast, isoforms involved in some signaling pathways that have been shown to be essential for cardiac hypertrophy [157,187] were more likely to be altered in TAC hearts.

As an individual example of multiple splicing patterns within a single gene, our analysis showed that of two splicing changes associated with development in the tropomyosin 3 gene (Tpm3), one was re-expressed following 10 days of TAC, while the other was not. The tropomyosin protein family demonstrates tremendous isoform diversity which is driven by four separate genes, multiple promoters, and alternative splicing [180]. While the expression of Tpm3 isoforms in neural tissues and skeletal muscle has been intensely studied, there is considerably less known about the different Tpm3 isoforms in the heart [188]. In our analysis, the mutually exclusive exon 5 expression differed by age, with fetal hearts expressing different levels of exon 5a relative to either adult groups (TAC or sham). In contrast, the terminal exon splicing of both the TAC and fetal groups were more similar to one another than to sham group, suggesting recurrence of fetal splicing patterns in the terminal exon during hypertrophy.

The specific functional consequences of alternative splicing on the function of individual proteins are known for only a relatively small subset of the altered isoforms we identified. One example is Myo1b, a gene that displayed the same alternative splicing event

in both fetal and TAC groups in our study. Myo1b is known to function between the cell membrane and actin cytoskeleton as a mechanotransduction sensor [189]. The percentage inclusion of Myo1b exon 11 is higher in both TAC and fetal hearts compared to the sham group; the inclusion of this exon has been shown to increase the sensitivity of Myo1b in sensing tension[190]. An example of a TAC-specific splicing event we detected involves TEA domain family member 1 (Tead1). Tead1 has been shown to be upregulated in hypertrophy [191] and is responsible for the re-expression of components of the fetal gene program [192,193]. We observed a decrease in exon 4 inclusion in TAC hearts compared to the fetal or sham groups. Exclusion of exon 4 does not disrupt the reading frame, but the loss of this exon has been shown to occur during transformation of multiple cell types [194]. Exon 4 is only 12 nucleotides in length and is located in the highly-conserved TEA domain that is responsible for physical interaction with other essential transcription factors for the hypertrophic response, SRF [195] and MEF2 [196]. More work is needed to follow up on the consequences of this TAC-specific isoform and its role in the development of hypertrophy.

Validation is a critical component of any high-throughput study, and we therefore validated the bioinformatics analysis presented here using multiple comparisons to both published data and PCR. First, gene expression levels computed from our analysis confirmed changes known to occur during hypertrophy, including concordant changes in TAC and fetal hearts of 14 of 18 genes frequently identified in the literature as components of the fetal gene program (Fig. 2A). Second, we confirmed that our analysis identified splicing events that were previously reported during fetal heart development. We compared our RNAseq analysis of E18 and adult rat hearts to a prior study using custom microarrays to detect alternative splicing in E18 and adult mouse hearts. Given the different species and methodology, it is not surprising that only about half of the previously reported splicing

events were apparent in our sequencing data (17 of 33 events that altered exon percent inclusion by at least 20% were present, with 12 changes reaching statistical significance). More meaningful as a validation of our overall approach to identifying changes in alternative splicing is the fact that all 12 of the events that were reported as statistically significant by our analysis were also significantly different by PCR. As a final validation, in the specific case of tropomyosin 3 even a complex pattern of alternative splicing changes among the fetal, TAC, and sham groups identified by our analysis was confirmed by PCR.

The most common source of error in high-throughput analyses is the potential for false positives. In the present study, we attempted to limit false discovery rates by adjusting for multiple comparisons and using a cutoff of the adjusted  $p$ -value, or  $q$ -value  $<0.05$ . We also required expression levels  $>3$  FPKM and an absolute fold change  $>1.5x$  as criteria for recognizing changes in gene or isoform expression. As suggested by our validation against previously reported developmental splicing changes, this conservative approach may result in missing some events (we identified 12 significant events out of the 33 reported by Kalsotra et al.), but provides confidence that reported changes are substantive (all 12 were confirmed on PCR). One potential source of error in our analysis was the higher number of total sequencing reads in the fetal group compared to our TAC and sham groups. In order to test whether this difference in the number of total reads could have affected our conclusions, we randomly downsampled each sample to match the smallest number of reads. Random downsampling did diminish the total number of genes identified as differentially expressed in fetal vs. sham hearts, but the trend of more significant genes and isoforms in the fetal group compared to the TAC group remained (Supplementary Fig. 2), as did the conclusion that a much higher fraction of isoform expression changes in TAC are concordant in fetal hearts than would be expected in the absence of common regulatory mechanisms. Thus the

primary results of the present study cannot be attributed to differences in sequencing depth. One other limitation of the present study is that we conducted the analysis at a single time point, 10 days following TAC. We chose this time point because microarray studies had previously identified it as having a large number of gene expression changes [176]. However, there are certainly transient changes in gene expression known to occur prior to 10 days after TAC that would be missed by our analysis; there may also be transient early splicing changes that were similarly missed.

## **5. Conclusions**

We tested the hypothesis that patterns of alternative splicing occurring during heart development recur during cardiac hypertrophy in response to transverse aortic constriction using high-throughput sequencing of mRNA isolated from fetal rats, adult sham-operated rats, and adult rats subjected to 10 days of pressure overload. We found a striking degree of overlap between changes in both gene and isoform expression in TAC and fetal hearts relative to sham. Our data strongly support the concept that mRNA splicing patterns normally associated with heart development recur as part of the hypertrophic response to pressure overload and suggest that cardiac hypertrophy shares post-transcriptional as well as transcriptional regulatory mechanisms with fetal heart development. Our results suggest that a deeper understanding of alternative splicing and its role in cardiac hypertrophy may reveal new mechanistic insight and therapeutic opportunities.

## **Acknowledgments**

We acknowledge K. Janes for valuable discussions and technical support from T. Burcin, the University of Virginia Biomolecular Research Facility, and the Robert M. Berne Cardiovascular Research Center Surgery Core.

## **Sources of funding**

This work was funded by NIH R01 HL-075639 (JWH), NIH T32 GM-007267-34 (EGA), and NIH T32 GM-008136-27 (EGA).

**Disclosures**

None declared.

## Appendix A. Tables and figures

*Table 1. Summary of sequencing results for each experimental group*

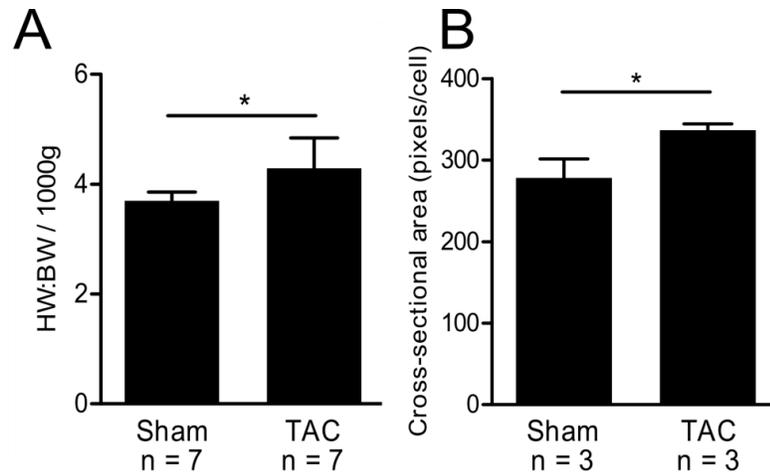
Sample	Sequenced fragments	Aligned fragments	Percent Aligned (%)
TAC1	28.9 M	18.3 M	63.28
TAC2	24.2 M	17.4 M	71.51
TAC3	44.9 M	32.1 M	71.59
Sham1	54.0 M	33.8 M	62.65
Sham2	46.1 M	32.7 M	71.04
Sham3	85.3 M	55.6 M	65.19
Fetal1	73.4 M	54.3 M	73.98
Fetal2	76.8 M	52.4 M	68.16
Fetal3	102 M	74.2 M	72.92

*Table 2. Gene ontology (GO) analysis of biological processes associated with up- and down-regulated isoforms seen in both hypertrophy and development*

Gene ontology term	No. of genes	Fold enrichment	<i>P</i> -value
Cytoskeleton organization (GO:0007010)	35	2.13	4.4E-05
Messenger RNA processing (GO:0006397)	28	2.58	1.1E-05
Muscle contraction (GO:0006936)	12	2.70	4.6E-03
Intracellular transport (GO:0046907)	49	1.93	1.5E-05
Response to reactive oxygen species (GO:0000302)	13	2.73	2.6E-03
Microtubule-based transport (GO:0010970)	5	4.61	2.1E-02
Endocytosis (GO:0006897)	16	2.01	1.3E-02
Cellular protein complex assembly (GO:0043623)	13	2.19	1.5E-02
Regulation of cell size (GO:0008361)	18	1.88	1.5E-02
Response to estradiol stimulus (GO:0032355)	12	2.80	3.5E-03
Glucose metabolic process (GO:0006006)	19	2.09	4.3E-03
Regulation of actin filament polymerization (GO:0030833)	9	4.46	7.1E-04
Muscle organ development (GO:0007517)	20	2.29	1.1E-03

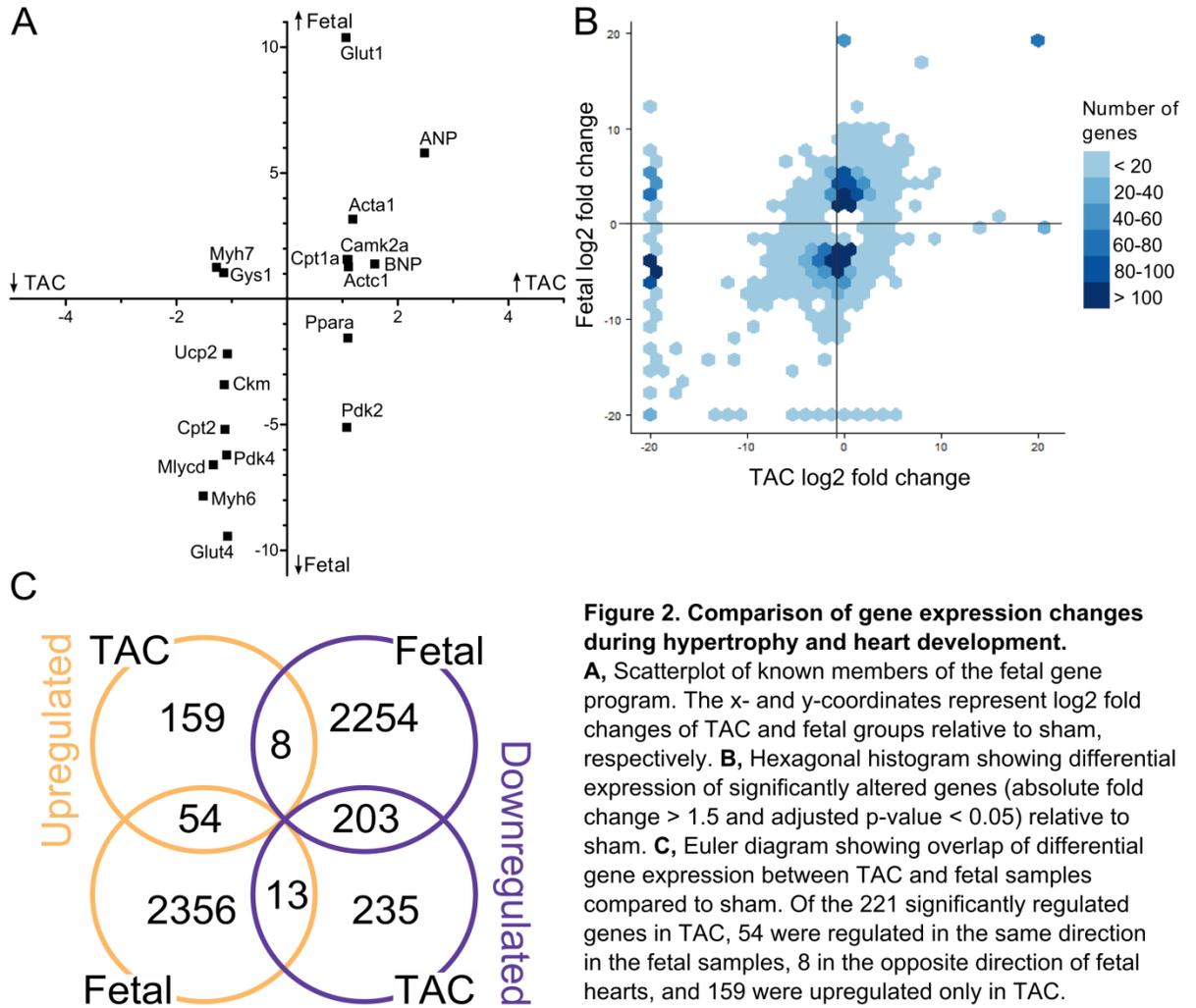
Representative GO terms from significant GO term clusters from the 453 upregulated isoforms and the 934 isoforms downregulated in both TAC and fetal hearts. No. of genes = number of genes associated with that GO term in the dataset. Fold enrichment quantifies the extent that each GO term is overrepresented in the dataset compared to the rat genome. A *P*-value signifies the degree of statistical significance of the degree of enrichment.

Figure 1. Induction of hypertrophy with a minimally-invasive approach to TAC



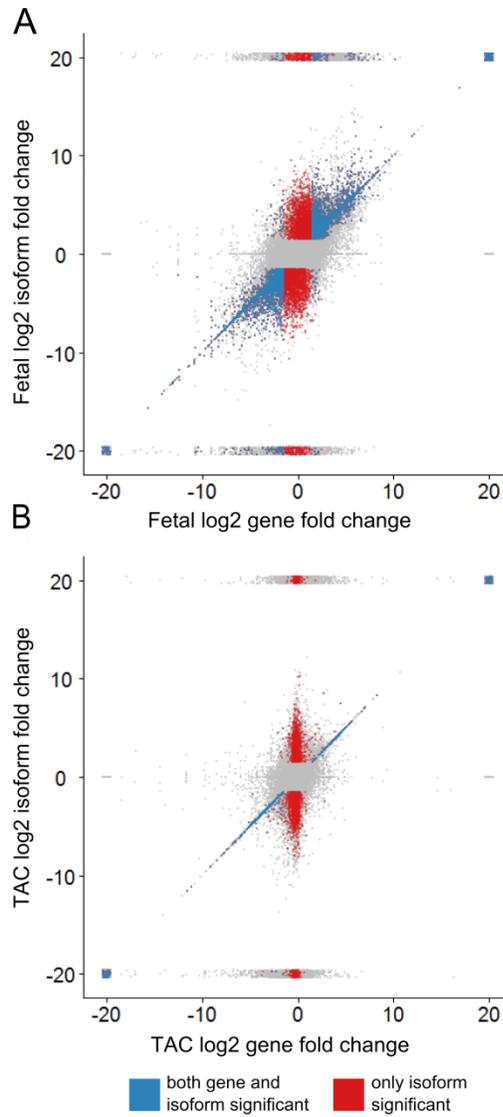
**Figure 1. Induction of hypertrophy with a minimally-invasive approach to TAC.** **A**, TAC rats had increased heart weight to body weight ratio per 1000 g of tissue compared to sham-operated rats 10 days after surgery. **B**, TAC cardiomyocytes had larger cross-sectional area compared to sham-operated cardiomyocytes (approx. 1000 cell measurements per animal). \* indicates  $P < 0.05$

Figure 2. Comparison of gene expression changes during hypertrophy and heart development



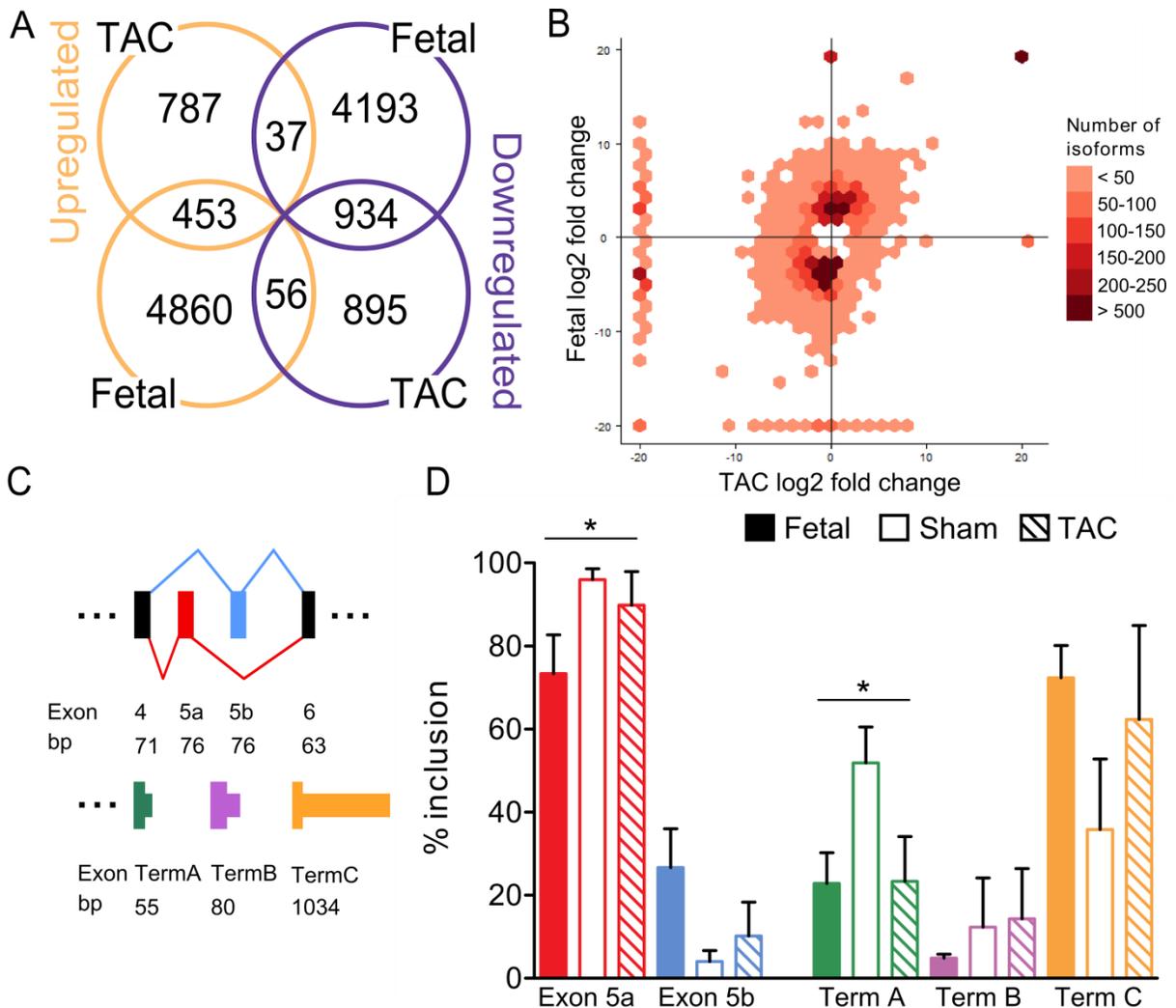
**Figure 2. Comparison of gene expression changes during hypertrophy and heart development.** **A**, Scatterplot of known members of the fetal gene program. The x- and y-coordinates represent log2 fold changes of TAC and fetal groups relative to sham, respectively. **B**, Hexagonal histogram showing differential expression of significantly altered genes (absolute fold change > 1.5 and adjusted p-value < 0.05) relative to sham. **C**, Euler diagram showing overlap of differential gene expression between TAC and fetal samples compared to sham. Of the 221 significantly regulated genes in TAC, 54 were regulated in the same direction in the fetal samples, 8 in the opposite direction of fetal hearts, and 159 were upregulated only in TAC.

Figure 3. Changes in gene and isoform expression are distinct in both fetal and hypertrophic hearts



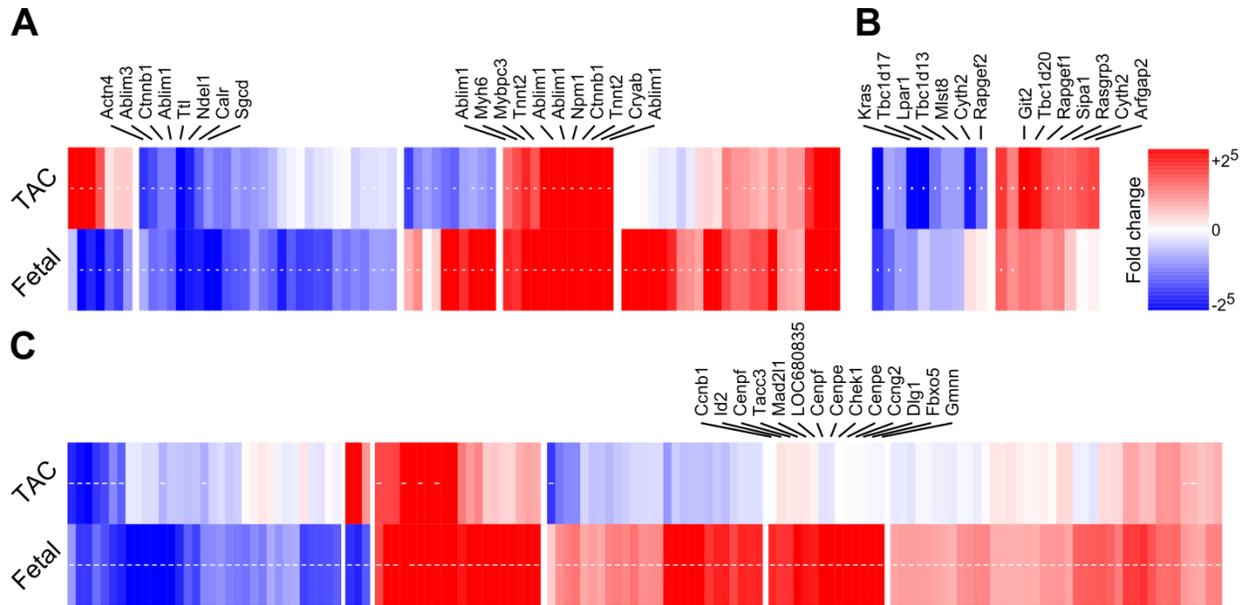
**Figure 3. Changes in gene and isoform expression are distinct in both fetal and hypertrophic hearts.** **A**, Each isoform expressed in the fetal or sham group above 3 FPKM is plotted as a gray dot, where the x-coordinate is the gene expression fold change relative to sham and the y-coordinate is the isoform fold change (relative to sham). Blue dots represent a transcript where both the isoform and gene expression are significantly different from sham (absolute fold change > 1.5 and an adjusted p-value < 0.05). Red dots represent transcripts where only isoform expression is significantly altered. **B**, TAC gene and isoform expression plotted in a similar fashion as **A**.

Figure 4. Significant overlap of splicing patterns in heart development and hypertrophy



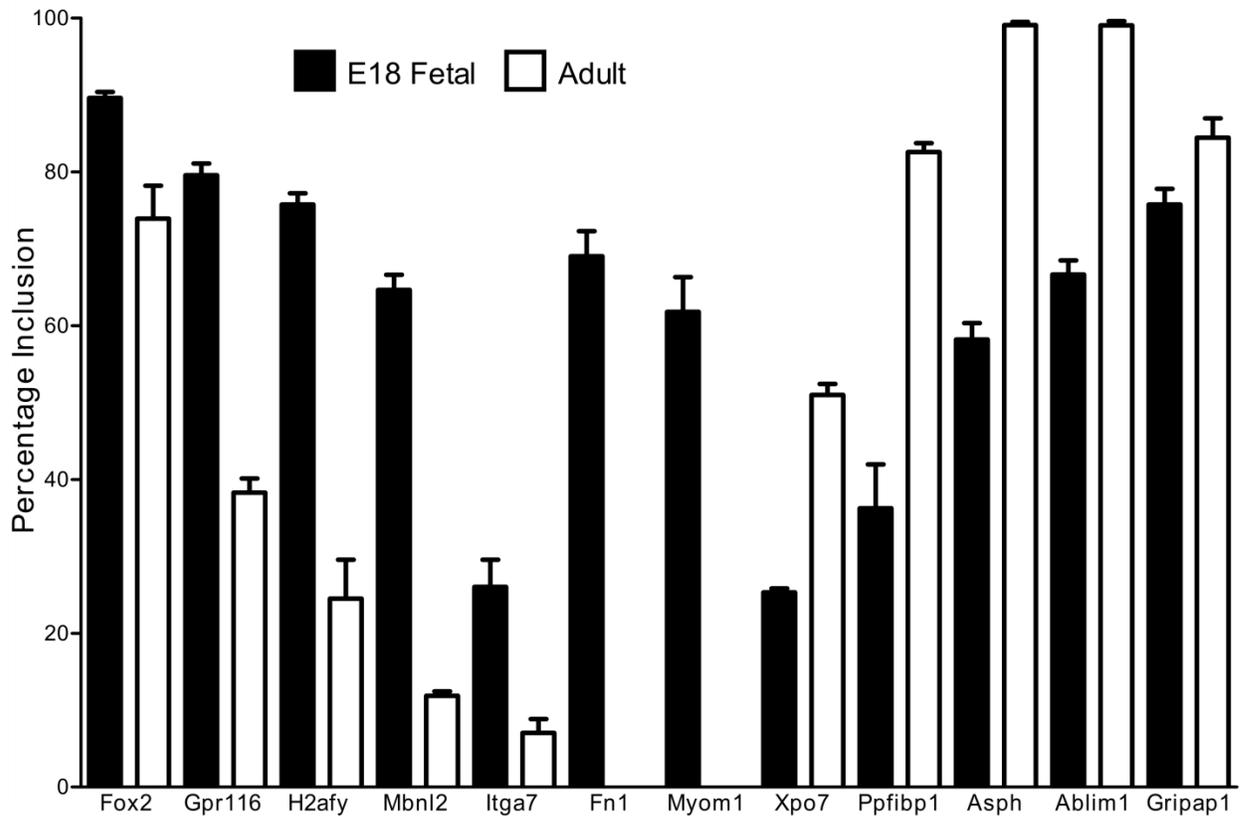
**Figure 4. Significant overlap of splicing patterns in heart development and hypertrophy.** **A**, Overlap of changes in isoform expression between TAC and fetal samples compared to sham as shown in Euler diagram. Four hundred fifty-three transcripts are concordantly upregulated in both TAC and fetal hearts, while 934 isoforms are downregulated in both groups. **B**, The hexagonal histogram showing expression of each significant isoform (absolute fold change > 1.5 and an adjusted  $p$ -value < 0.05) relative to sham. **C**, Exon diagram of tropomyosin 3 (Tpm3) demonstrates multiple splicing patterns; a mutually exclusive exon (5a - red or 5b - blue) and three alternative terminal exons (TermA - green, TermB - purple, TermC - yellow). **D**, The bar chart shows percentage of all Tpm3 transcripts contain each splicing event based on band intensity from gel electrophoresis ( $n = 4$ ). The expression of mutually exclusive exon 5a is statistically greater in both TAC and sham groups compared to the fetal group. In contrast, the expression of terminal exon TermA is significantly greater in sham than compared to both TAC and fetal groups, where \* represents  $P < 0.05$ .

Figure 5. Similarities and differences between isoform expression observed in fetal heart development and TAC



**Figure 5. Similarities and differences between isoform expression observed in fetal heart development and TAC.** **A**, Top gene ontology term "cytoskeleton organization" (GO:0007010) for isoforms that were commonly regulated in both TAC and fetal hearts. **B**, TAC-specific most significant gene ontology term "regulation of small GTPase-mediated signal transduction" (GO:0051056). **C**, Fetal-specific most significant gene ontology term "cell cycle" (GO:0007049). All significant isoforms of genes within a gene ontology cluster identified by DAVID (FPKM > 3, adjusted p-value < 0.05, and absolute log<sub>2</sub> fold change > 1.5) are plotted as a heat map to show similarities and differences between TAC and fetal isoform expression. Gene symbols of significant isoforms with distinct expression are listed above the heat map. White dots within heat map bars denote statistical significance in isoform expression relative to sham.

Figure 6. Identification and validation of known developmental splicing events



**Figure 6. Identification and validation of known developmental splicing events.** A comparison of our fetal and sham group identified 12 out of 33 events reported by Kalsotra et al. as significantly regulated in the same pattern in our data. PCR validation confirmed all 12 events are significantly different between E18 and adult hearts ( $n = 4$  for each group and  $P < 0.05$  for all events).

## **Appendix B. Supplementary methods**

### *Surgical procedures*

The institutional animal care and use committee approved all studies. A minimally-invasive transverse aortic constriction (TAC) procedure was adapted from [174] for rats. Briefly, 14 male Sprague-Dawley rats ( $307 \pm 30$  g) were anesthetized with isoflurane and a topical depilatory agent was applied to the neck and chest of each rat. Body temperature was maintained at  $37^{\circ}\text{C}$  with a warming pad. The surgical site was cleaned and sterilized with betadine and alcohol. A longitudinal skin incision was made above the suprasternal notch. The mediastinum was entered at the point of insertion of the first two ribs on the right side of the sternum and the thymus was retracted to allow visualization of the aortic arch. A wire with a snare on the end was then passed under the aorta between the origin of the right innominate and left common carotid arteries. The wire was threaded with a 4-0 silk suture and pulled back around the aorta. A bent 19-gauge needle was then placed next to the aortic arch and the suture was snugly tied around the needle and the aorta. The size of the rat, type of suture, and needle gauge were adapted from a procedure in rats that involved open-chest surgery [175]. After the suture was tied, the needle was quickly removed. The skin was closed and the rat was allowed to recover on a warming pad until fully awake, at which time buprenorphine HCl was administered for analgesia. The sham procedure was identical to the TAC procedure except the suture was not tied. A pilot study of our TAC procedure in 5 rats showed that the three rats that underwent TAC had a significantly higher pressure gradient between the right and left carotid artery ( $40 \pm 6$  mmHg) compared to two sham-operated rats ( $-7 \pm 22$  mmHg). On post-operative day 7, TAC rats were imaged with the Vevo 2100 system to visually confirm an aortic constriction (VisualSonics). At 10 days after surgery, rats were euthanized with an overdose of pentobarbital (150 mg/kg)

administered intraperitoneally. The 10-day time point was chosen based on a previous report in mice that this day is the time point with the maximal number of gene expression changes [176]. Before the hearts were removed, an intraventricular injection of heparin (0.2 mL) was used to prevent excessive blood clots in the tissue sample. Hearts were quickly removed, rinsed in cold phosphate buffered saline, weighed, and cut in appropriately sized pieces for preservation in RNAlater solution (Qiagen, Valencia, CA).

Five pregnant rats were anesthetized with isoflurane at day 18-20 of gestation. Under deep anesthesia, the ventral midline was opened and the uterus with embryos exteriorized. Fetuses were removed, decapitated, hearts quickly dissected out, and placed in RNAlater. Dams were euthanized while under anesthesia by exsanguination. After 24 hours incubation in RNAlater, atria were dissected from ventricles and discarded. RNA was isolated from only ventricular fetal tissues.

#### *Cell cross-sectional area measurements*

Five-micron sections from short axis slices of flash frozen left ventricle at the level of the papillary muscle from the TAC and sham rats were used to measure cross-sectional area of cardiomyocytes. Sections were stained with an anti-basement membrane antibody (rabbit anti-entactin, Ab14511, Abcam, Cambridge, MA) and fluorescently labeled with polyclonal chicken anti-rabbit IgG conjugated with Alexa647 (A21443, Molecular Probes, Eugene, OR). Images were taken at 20x near the endocardial wall. A universal ImageJ macro was used to analyze all images and consisted of color enhancement (saturated = 0.6%), splitting color channels to only analyze the blue channel, smoothing by Gaussian blur ( $\sigma = 2$ ), background subtraction (12 pixels radius), thresholding to the same level (20), converting to a binary image (with 3 commands: Close-, Skeletonize, Dilate), and then

inverting. Cross-sectional area measurements were only collected from cells that were not touching the border of the image and had a circularity  $> 0.6$  (circularity =  $4\pi(\text{area}/\text{perimeter}^2)$ ) to exclude oblique myocytes. Three or more images were collected from each animal to generate an average cross-sectional area for each animal. These averages were used in the student's *t*-test to determine significance.

#### *Library preparation and validation*

Total RNA isolation was completed following instructions provided with Qiagen RNeasy Fibrous Tissue Kit including DNase I (Qiagen) treatment step. RNA isolation purity and quantity was measured with Agilent RNA chip (Agilent). Only samples with an RNA integrity score  $> 9$  (range: 0 - 10) were used in library preparation. A small amount of cDNA was synthesized and tested for genomic DNA contamination by PCR with primers specific to genomic DNA. No RNA sample with evidence of genomic DNA contamination was used in library preparation. Five micrograms of total RNA were used as starting material in the Illumina mRNA seq library preparation kit and was prepared exactly to manufacturer's directions (Illumina). Briefly, poly(A) tail mRNA was isolated using oligo-d(T) magnetic beads. Messenger RNA was then fragmented, cDNA synthesized, adapters ligated onto cDNA, libraries were selected for an insert size of approximately 300 bp, and finally libraries were amplified by 15 cycles of PCR. RNAseq libraries were validated by several methods including Agilent High Sensitivity DNA chip (Agilent) to assess quality and approximate quantity and sequencing cloned libraries. Briefly, one microliter of prepared libraries were cloned into blunt end plasmids (Invitrogen Zero Blunt TOPO kit; Carlsbad, CA). Ten colonies were picked and amplified by colony PCR using primers specific to the cloning plasmid and then separated by agarose gel electrophoresis. PCR products were gel extracted using

Qiagen Gel Extraction kit (Qiagen) and submitted for sequencing. Only libraries with more than 7 of 10 sequenced inserts containing known genes were candidates for RNAseq. Lastly before sequencing on Illumina Genome Analyzer Ix, library concentration was measured with real-time PCR. Libraries were sequenced as paired-end reads of 42 or 63 base pairs in length.

### *Bioinformatics pipeline*

Tophat v1.3.1 [177] was used to align all reads including junction-spanning reads back to the rat genome (Ensembl RGSC3.4). Cufflinks v1.0.3 [178] was used to identify differential splicing, promoter usage, and gene expression changes between experimental groups. Optional correction methods for cDNA synthesis bias [197] and for reads aligning to multiple locations in the genome were implemented for each sample and reads were normalized by the upper-quartile of fragments mapping to individual loci rather than the total number of reads mapped. In addition, gene transfer format (GTF) files were created using the reference annotation based transcript assembly method [197,198]. Using similar criteria to the recently published methods paper using Cufflinks in cardiac hypertrophy [172], we defined statistical significance in expression as  $> 1.5$  fold-change, adjusted p-value  $< 0.05$ , and FPKM  $> 3$ . All bioinformatics analyses and comparisons were performed using in-house scripts written in Unix or Perl. The most significantly up- and down-regulated isoforms are listed in Supplementary Table 1.

### *Gene ontology (GO) enrichment analysis*

The isoforms that were significantly up- or down-regulated in both TAC and fetal groups were used for gene ontology analysis using DAVID functional annotation clustering [140,199]. Clusters were grouped using the 'GOTERM\_BP\_FAT' option.

#### *Heat maps of significant GO terms*

All isoforms of genes contained within a GO term of interest were filtered for those that had a significant adjusted  $p$ -value in either TAC vs. sham or fetal vs. sham comparisons. The log<sub>2</sub> fold change of these isoforms (compared to sham) were then used to generate a heat map in R using gplots package [200].

#### *PCR confirmation of splice variants*

Using previously published methods [7], we validated our isoform expression differences using the Invitrogen One-Step Reverse Transcription PCR kit (Carlsbad, CA). Splice variants in fetal compared to adult hearts identified both by [7] and by our analysis were validated by one step PCR starting with 500 ng RNA and amplified for 25 cycles (Invitrogen). Uniform RNA quantities across all samples and reactions were confirmed by measuring RNA concentration with spectrophotometry (NanoDrop) and the Agilent Bioanalyzer, as well as running a one-step PCR with 18S RNA primers to visually confirm uniform RNA amplification. Custom primers (Integrated DNA Technologies) were designed to flank the alternatively spliced region. PCR products were electrophoresed and visualized with ethidium bromide-stained 2-5% Agarose-1000 gels (Invitrogen). The percentage inclusion of the alternative region was calculated with the following equation: [inclusion band

intensity / (inclusion band intensity + exclusion band intensity)] x 100. Gel band intensities were measured using BioRad ImageLab software. A similar primer design strategy was employed for Tpm3 exons 5a and 5b. To amplify the terminal exon splice variants, one universal forward primer was designed in the most 3' exon that each of the 3 splice variants (TermA, TermB, and TermC) shared. The reverse primers were designed to uniquely target each of three terminal exons (see Supplementary Table 2 for primer sequences). The gel band intensity values were used to calculate the percentage of each terminal exon present. For example, the percentage of all Tpm3 transcripts with 'TermA' was calculated as  $[\text{TermA}/(\text{TermA} + \text{TermB} + \text{TermC})] \times 100$ . Since percentages are limited to values between 0 and 100 and do not follow a normal distribution, we transformed our percentage inclusion rates into values that would follow a normal distribution using the arcsine of the square root of the percentage prior to performing statistical tests [179].

## Appendix C. Supplementary data

Supplementary Table 1. Top shared isoforms among TAC and fetal group

Isoforms were selected by largest differences between TAC and sham FPKM, then sorted by TAC *q*-value

### Upregulated isoforms

Nearest Ensembl transcript ID	Gene Name	Locus	TAC log2 (fold change)	TAC <i>q</i> -value	Fetal log2 (fold change)	Fetal <i>q</i> -value	Fetal FPKM	TAC FPKM	Sham FPKM
ENSRNOT0000017941	Cmip	19:47364257-47574027	3.62	-02	3.91	5.96E-05	86700.90	70959.90	5753.67
ENSRNOT0000032264	Aldoa	1:185961307-185974615	>20	-06	>20	7.91E-06	1585.85	1754.60	0
ENSRNOT0000059127	Cryab	8:54107471-54118847	>20	-06	>20	1.35E-04	1543.96	1248.02	0
ENSRNOT0000047682	Tnnt2	13:48867869-48885815	5.24	0.00E+00	6.41	0.00E+00	2026.84	905.68	23.88
ENSRNOT0000047682	Tnnt2	13:48867869-48885815	4.28	1.11E-04	6.29	0.00E+00	3502.95	869.04	44.63
ENSRNOT0000003611	Nme2	10:82582764-82589980	2.12	1.77E-04	2.79	2.47E-07	1074.63	676.24	155.39
ENSRNOT0000014407	Cox6c	7:71465789-71478657	6.51	1.25E-11	5.20	3.53E-07	230.23	571.22	6.26
ENSRNOT0000019311	Csrp3	1:98601676-98620600	9.29	1.42E-08	8.35	0.00E+00	291.63	557.16	0.89
ENSRNOT0000003611	Nme2	10:82582764-82589980	2.86	6.85E-07	1.65	1.85E-02	197.21	455.70	62.80
ENSRNOT0000051911	LOC686074	18:23893499-24032157	2.22	3.69E-03	2.24	4.22E-03	373.90	367.86	78.93
ENSRNOT0000002116	Atp5j	11:24316737-24352754	>20	8.53E-07	>20	2.78E-04	84.55	233.88	0
ENSRNOT0000019335	Myl7	14:86567233-86570720	1.72	1.01E-02	3.27	2.64E-09	628.16	214.37	65.20
ENSRNOT0000028624	Fxyd1	1:86095253-86099434	10.25	9.64E-04	9.55	2.61E-03	124.47	201.87	0.17
ENSRNOT0000001500	Pebp1	12:40491334-40495594	3.60	2.83E-11	3.56	4.94E-11	166.27	170.40	14.07
ENSRNOT0000010545	Pdhb	15:18733732-18743392	1.64	1.17E-03	1.79	3.10E-04	146.70	132.00	42.43

### Downregulated isoforms

Nearest Ensembl transcript ID	Gene Name	Locus	TAC log2 (fold change)	TAC <i>q</i> -value	Fetal log2 (fold change)	Fetal <i>q</i> -value	Fetal FPKM	TAC FPKM	Sham FPKM
ENSRNOT0000001490	Tesc	12:39574222-39703686	<-20	3.36E-04	-3.35	2.00E-05	8.25	0	84.18
ENSRNOT0000061542	ENSRNOG0000040030	19:9122839-9123847	<-20	1.40E-04	<-20	1.40E-04	0	0	56.28
ENSRNOT0000018691	Gpx4	7:11162516-11165497	<-20	2.84E-05	<-20	2.84E-05	0	0	52.01

*Supplementary Table 1 (continued)*

ENSRNOT0		3:170295329-		1.40E		2.56E				
0000016947	Eef1a2	170304828	<-20	-05	-2.34	-05	9.99	0	0	50.74
ENSRNOT0		18:71380658-		2.64E		2.64E				
0000019512	Myo5b	71582298	<-20	-69	<-20	-69	0	0	0	47.69
ENSRNOT0		8:112861951-		4.21E		4.21E				
0000064436	Gnai2	112882663	<-20	-05	<-20	-05	0	0	0	42.67
ENSRNOT0		12:42315763-		1.30E		0.00E				
0000001518	Arbp	42319050	<-20	-05	-7.31	+00	0.26	0	0	40.74
ENSRNOT0		4:122326742-		1.28E		1.28E				
0000017323	Eefsec	122539572	<-20	-02	<-20	-02	0	0	0	29.85
ENSRNOT0		1:263380574-		2.23E		1.81E				
0000067257	Abli1	263608779	<-20	-04	-10.68	-04	0.02	0	0	28.71
ENSRNOT0		5:157980852-		2.36E		2.36E				
0000066151	Capzb	158080678	<-20	-03	<-20	-03	0	0	0	24.07
ENSRNOT0		2:182078050-		7.92E		7.92E				
0000023605	Tpm3	182325292	<-20	-04	<-20	-04	0	0	0	23.24
ENSRNOT0		1:190987538-		0.00E		0.00E				
0000027999	Acadsb	191028970	<-20	+00	<-20	+00	0	0	0	22.67
ENSRNOT0		4:13448840-		1.36E		1.36E				
0000066224	LOC685953	13670274	<-20	-05	<-20	-05	0	0	0	19.68
ENSRNOT0		12:10584699-		3.06E		3.06E				
0000001336	Bri3	10836977	<-20	-02	<-20	-02	0	0	0	17.85
ENSRNOT0		2:182078050-		1.50E		1.50E				
0000066527	LOC361985	182325292	<-20	-03	<-20	-03	0	0	0	17.25

Supplementary Table 2. PCR primer sequences

	Gene	Forward primer	Reverse primer
Figure 4	Tpm3 e5a	GAGCTCCAGGAAATCCAGCTGAAGGAA	CTCAGACACTTCAGGTTCTGGTCC AT
	Tpm3 e5b	CAGGAAATCCAGCTGAAGGAAGCA	TCCGCCTGAGCCTCAAGAGATTT
	Tpm3 e10a	GGCTGAGTTTGCTGAAAGATCGGT	TCCAGAACAGAGCAGAAAACGGTGA
	Tpm3 e10b	GGCTGAGTTTGCTGAAAGATCGGT	CAGTCACAAAGACCATGCAGCGTT
	Tpm3 e10c	GGCTGAGTTTGCTGAAAGATCGGT	TCTGGGAAGAGGCTGACACAGTTT
Figure 5	Abli1 e9	GCAGCAGATGCAACCAGATGTTCA	TTTGGTTGGAGTTCACCAGACAGG
	Asph e4	CACGTGGTTCATGGTCATTGCGTT	CTTCAGTCTTGCTAGGCACAGCTT
	Fn1 e25	TGGTTCATGCCGATCAGAGTTCCCT	AGTAGCGCACCAAGAGGTTGGTTA
	Fox2 e6	AGTAAATCCACGCCCAAGCGACTA	TGGGCTTAACTCCAGCCATTTGC
	Gpr116 e4,5	TACTCAAGGCAGATGGAACCCA	AAGCTACTCACATCGCTGAGGCAA
	Gripap1 e12	AGGAAGCTGATCTGAAAGCCCAGT	AGTTGAGGTCGTCTTTCTGTCCCA
	H2afy e6	ACGGCTTTACTGTCCTCTCTACCA	TCATCAGCTAGAGCCAGGCAGTTT
	Itga7 e5	ACTGCTACAGAGTGGACATCGACA	TGTACACATACACGGCACCTCCAA
	Mbnl2 e8	TCTTGCACTACCAGCAGGCTTTGA	GTCTTTGAACATCGCAGCTGGTCT
	Myom1 e17	AGTGAACGCAGCTGGACTTAGTGA GGCTAATACAAGAGATCAATGATTTGAG GC	CCTTGATGTTGGCTTCCCTCCATT
	Ppfibp1 e4		TCGATCGAGGATTTTCGACCC
	Xpo7 e5	ATATGCCAGGATCACAAAGCTGGG	ATGAGTCTCGGAAGGAAGAGGCTA

Supplementary Table 3. Gene Ontology (GO) analysis of biological processes associated with isoforms specifically seen in heart development or hypertrophy

<i>Fetal-specific isoforms</i>			
Gene ontology term	No. of genes	Fold enrichment	<i>P</i> -value
Cell cycle (GO:0007049)	151	2.14	1.83E-21
Response to DNA damage stimulus (GO:0006974)	75	1.89	2.54E-08
Phosphorylation (GO:0016310)	164	1.50	3.74E-08
Ubiquitin-dependent protein catabolic process (GO:0006511)	58	2.28	8.46E-10
Cytoskeleton organization (GO:0007010)	93	1.91	3.09E-10
Protein complex assembly (GO:0006461)	104	1.63	2.10E-07
Regulation of microtubule cytoskeleton organization (GO:0070507)	19	3.10	7.70E-06
Regulation of cell cycle (GO:0051726)	73	1.93	1.89E-08
Chromosome segregation (GO:0007059)	28	3.44	1.42E-09
Chromatin organization (GO:0006325)	73	1.90	3.90E-08
<i>TAC-specific isoforms</i>			
Gene ontology term	No. of genes	Fold enrichment	<i>P</i> -value
Regulation of small GTPase mediated signal transduction (GO:0051056)	14	3.12	5.66E-04
In utero embryonic development (GO:0001701)	17	3.12	1.14E-04
Actin cytoskeleton organization (GO:0030036)	11	2.69	7.95E-03
Protein transport (GO:0015031)	25	2.28	8.46E-10
Tetrapyrrole metabolic process (GO:0033013)	5	7.37	4.34E-03
Positive regulation of MAPKKK cascade (GO:0043410)	6	1.63	2.10E-07

Significant GO terms of isoforms that were uniquely up- or down-regulated in TAC and fetal hearts. No. of genes = number of genes associated with that GO term in the dataset. Fold enrichment quantifies the extent that each GO term is overrepresented in the dataset compared to the rat genome. The *P*-value signifies the degree of statistical significance of the fold enrichment of each term.

*Supplementary Table 4. Potential upstream regulators of changes in isoform expression in hypertrophy, heart development, and both processes*

*Shared upstream regulators in both hypertrophy and heart development*

Upstream regulator	Fetal Overlap <i>p</i> -value	TAC Overlap <i>p</i> -value	Target molecules in dataset
INSR	7.02E-12	5.67E-06	TAC: ACAA2, ACADVL, ACO2, ACOT13, ATP5G1, ATP5H, ATP5J, ATP5L, COX4L1, COX5B, CYC1, DES, ECI1, FLNC, HSPD1, IDH3A, MYOM2, NDUFA3, NDUFA9, NDUFB4, NOS3, NPPB, OGDH, PDHA1; Fetal: ACAA2, ACADL, ACADM, ACADS, ACADVL, ACOT13, ACOT2, ACSL1, ACSL4, ACSL5, ACTA1, ACTN4, ATP5A1, ATP5D, ATP5G1, ATP5G3, ATP5H, ATP5J, ATP5L, COX4L1, COX6A2, COX8B, CPT2, CRAT, DECR1, DES, ECI1, ESRRA, ETFA, ETFB, FLNC, GOT2, HADHB, HSD17B10, HSPD1, IDH3A, MYH7, NDUFA1, NDUFA3, NDUFA9, NDUFB4, NDUFV2, NRF1, PDHA1, PDK4, PECR, SCP2, SIRT1, SLC25A20, SLC2A1, SLC2A4, TNNI1, TPM2, UQCRC1, UQCRC2
IGF1R	1.89E-09	1.11E-03	TAC: ATP5G1, ATP5H, ATP5J, ATP5L, COX4L1, COX5B, DES, FLNC, MYOM2, NDUFA3, NDUFA9, NDUFB4; Fetal: ACADM, ACADS, ACSL1, ACTA1, ACTN4, ATP5A1, ATP5D, ATP5G1, ATP5G3, ATP5H, ATP5J, ATP5L, COX4L1, COX6A2, COX8B, CPT2, DES, FLNC, MYH7, NDUFA1, NDUFA3, NDUFA9, NDUFB4, NDUFV2, PECR, SCP2, SLC25A20, TNNI1, TPM2, UQCRC1, UQCRC2
MED30	3.13E-03	1.37E-03	TAC: COX5B, NDUFA10, NDUFB3, NDUFB7, NDUFS2, NDUFS7, PPARGC1A; Fetal: COX6A2, ESRRA, LPL, NDUFA10, NDUFB3, NDUFB7, NDUFS2, NDUFS7, PPARGC1A, SDHB, SOD2
GATA4	4.66E-02	1.77E-02	TAC: HSPA5, NPPB, TNNC1, TNNI3; Fetal: ACTA1, BCL2, HSPA5, MEF2C, MYH7, TNNI3
TTN	3.72E-03	3.30E-02	TAC: FHL2, FLNC; Fetal: ACTA1, ATP2A2, CNN1, GATA4, GJA5, MEF2C, MYH6, MYH7, SCN5A, TNNI1, TNNT2
GH1	2.47E-02	1.92E-02	TAC: APOE, CLU, HBB, IGF1, PRDX3, TSC22D1, UQCRH; Fetal: ACTB, CLU, COL3A1, DECR1, FN1, HBB, IGF1, LPL, NDRG2, SLC2A1, TSC22D1, TUBA1A, UQCRH

*TAC-specific upstream regulators*

Upstream regulator	TAC Overlap <i>p</i> -value	Target molecules in dataset
MAPK11	1.12E-02	CALR, MYH6, NPPB
MAPK14	3.30E-02	CALR, MYH6, NPPB
NR3C2	5.61E-03	ABCC9, CMYA5, COL1A1, DES, GHR, HSPA9, IDH3B, NFIA, PPIB, RAPGEF4, RRAD
TRIM54	1.21E-02	FHL2, FLNC
ESR2	3.36E-02	LMNA, MYH6
DVL1	3.36E-02	CTNNB1, NPPB

*Fetal-specific upstream regulators*

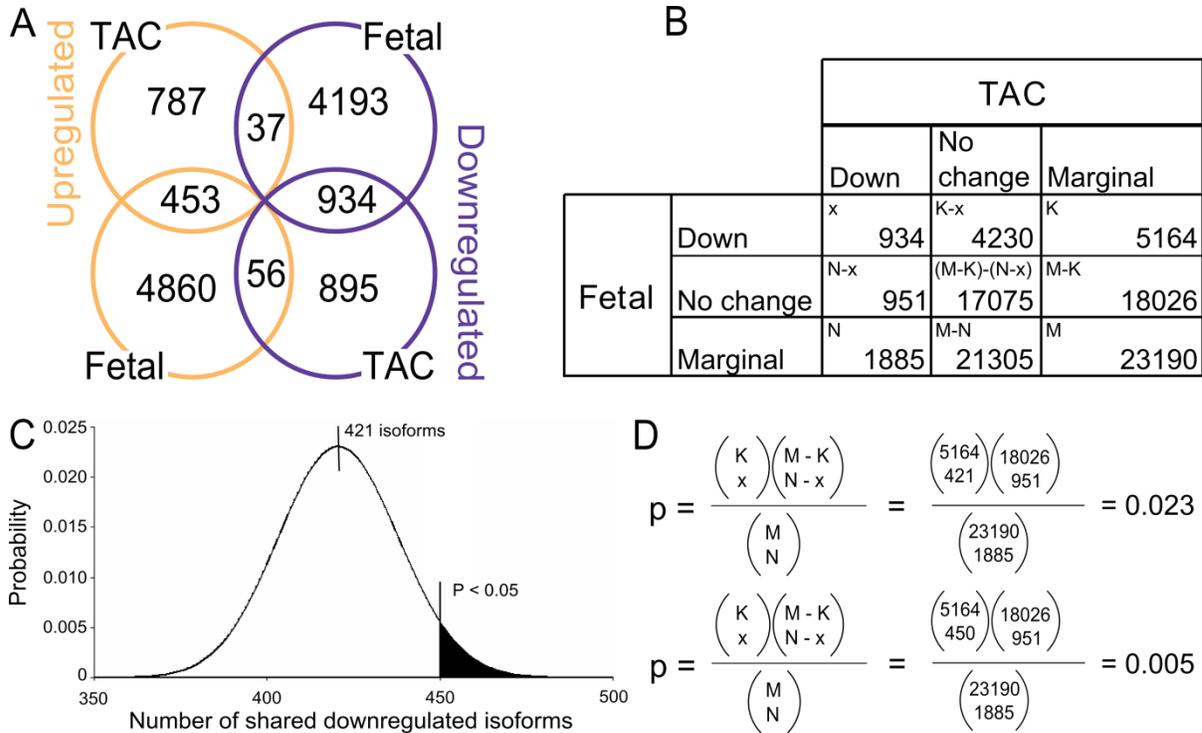
Upstream regulator	Fetal Overlap <i>p</i> -value	Target molecules in dataset
ESRRA	9.95E-05	ACADM, ACADVL, ACOX1, ACSL1, ATP2A2, CD36, CHKA, COX8B, CYB5A, ESRRA, FABP3, GOT1, HMGB2, LIPA, LPL, MTCH2, MYH6, MYH7, PDK4, PLN, SCD2, SMARCA4

*Supplemental Table 4 (continued)*

---

TFAM	6.47E-04	ACADL, ACADM, ACADS, ACOX1, CPT2, ECH1, FABP3, FABP4, PFKP, PGK1, SLC25A20
E2F6	2.39E-03	CBX5, CDC6, GJA1, RFC4, UNG
MAP2K6	3.72E-03	ACTA1, ATP2A2, CRYAB, MYH6, MYH7, PLN
HDAC5	4.20E-03	ACADL, ACADM, ACSL1, CPT2, HK2, PPARGC1A, PYGM
RBL2	8.00E-03	AURKB, CDK1, E2F1, MYC
MAP2K3	1.08E-02	ACTA1, ATP2A2, MYH6, MYH7, PLN

Supplementary Figure 1. Hypothesis testing and hypergeometric probability density functions

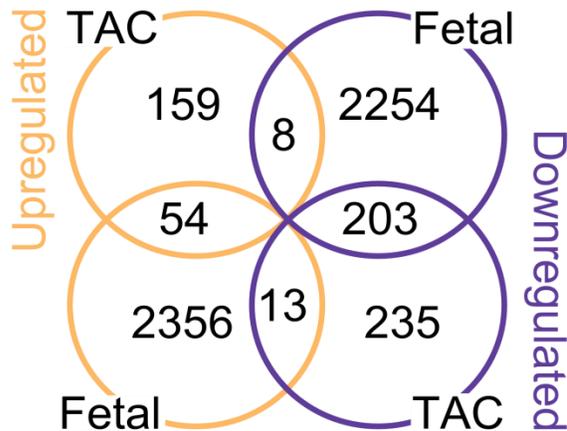


**Supplementary Figure 1. Hypothesis testing and hypergeometric probability density functions. A**, Overlap of isoform expression between TAC and fetal samples compared to sham as shown in an Euler diagram. Four hundred fifty-three transcripts are concordantly upregulated in both TAC and fetal hearts, while 934 isoforms are downregulated in both groups. **B**, Example of a contingency table used to test the null hypothesis that the amount of overlap seen in TAC and occurred independently.  $N$  represent the total number of significantly downregulated isoforms in TAC and  $K$  is equal to the number of significantly downregulated fetal isoforms, while  $x$  is the number of downregulated isoforms shared between the two groups. Since 8.1% of all isoforms are downregulated in TAC vs. sham ( $N$ , 1,885 / 23,190) and 22.3% of all isoforms are downregulated in fetal vs. sham ( $K$ , 5,164 / 23,190), multiplying the two percentages times the total number of isoforms ( $M$ , 23,190) would predict an expected overlap of 421 isoforms ( $0.081 \times 0.223 \times 23,190$ ) if the expression of each isoform were independently regulated. **C**, Graphical representation of probability for each possible number of isoforms shared between TAC and fetal groups. The most likely number of shared isoforms is 421 ( $P = 0.023$  of observing exactly 421 shared isoforms). The probability of observing increasing numbers of shared isoforms decreases to  $P = 0.05$  at 450 isoforms, more importantly, the sum of probabilities of observing all possible numbers of shared isoforms  $\geq 450$  is  $< 0.05$  (shown by shaded area), indicating that 450 is a significantly higher number of shared isoforms than would be expected if the expression each isoform was independently regulated. The fact that we saw 934 shared isoforms strandy suggests that there is some common regulatory mechanism. **D**, The probability of seeing any specific amount of overlap was calculated using a hypergeometric probability density function (calculations shown for 421 and 450 overlapping isoforms). The probability of seeing at least the degree of overlap  $x$  was computed by summing probabilities for all values  $\geq x$  using MATLAB with the following command 'sum(hygepdf(x:N,M,K,N))'.

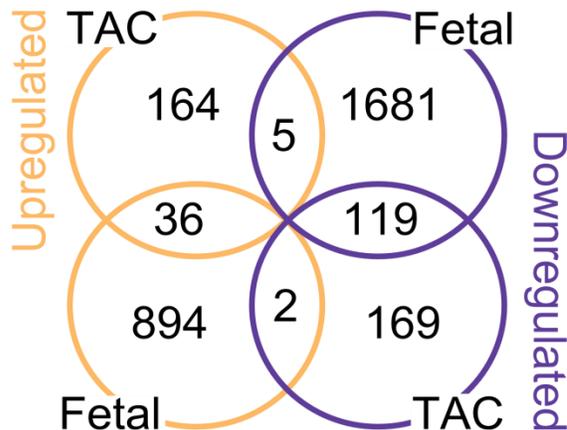
Supplementary Figure 2. Random downsampling of results shows that the trends seen in full data set are independent of sequencing read counts

### Gene expression

Full dataset



Downsampled dataset



Probability of overlap in upregulated genes:

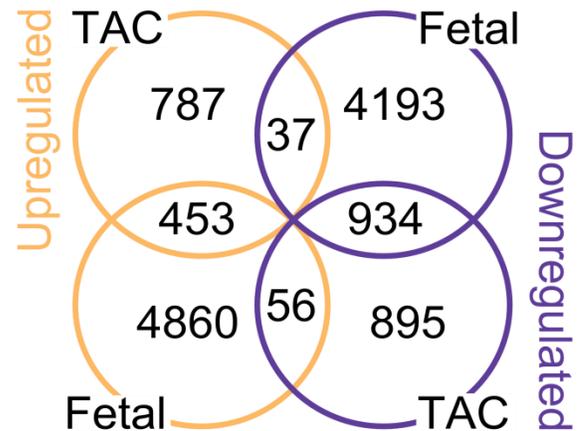
0.049

Probability of overlap in downregulated genes:

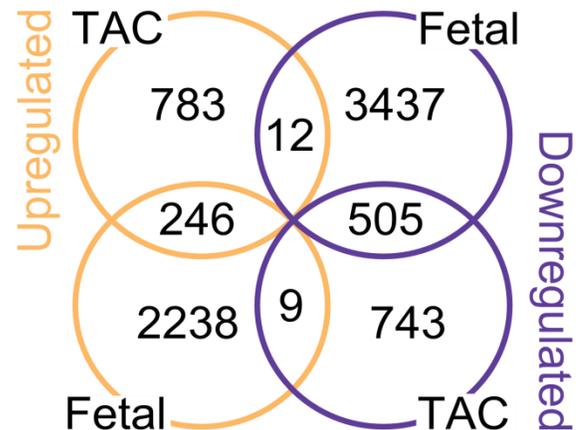
$2.1 \times 10^{-27}$

### Isoform expression

Full dataset



Downsampled dataset



Probability of overlap in upregulated isoforms:

$2.3 \times 10^{-4}$

Probability of overlap in downregulated isoforms:

$4.9 \times 10^{-14}$

**Supplementary Figure 2. Random downsampling of results shows that trends seen in full data set are independent of sequencing read counts.** Sequences from each sample were randomly downsampled to match the read count of the smallest sample. Results show that significant trends are maintained and that the gene and isoform expression in fetal comparisons are much larger than the TAC vs. sham comparisons.

Chapter 4 Title Page

**Full title: Computational search for RNA-binding protein recognition motifs identifies candidate regulators of fetal splicing program in cardiac hypertrophy**

Authors: Ames EG (BS)<sup>a,c</sup>, Mackey AJ (PhD)<sup>d</sup>, Holmes JW (MD PhD)<sup>a,b,e</sup>

Affiliations: <sup>a</sup> Department of Biomedical Engineering, University of Virginia (Box 800759, Charlottesville, VA, USA 22908)

<sup>b</sup> Robert M. Berne Cardiovascular Research Center, University of Virginia (Robert M. Berne Cardiovascular Research Center, 415 Lane Rd, Charlottesville, VA, USA 22908)

<sup>c</sup> Department of Molecular Physiology and Biological Physics, University of Virginia (Jordan Hall, PO Box 800736, Charlottesville, VA USA 22908)

<sup>d</sup> Center for Public Health Genomics, Department of Public Health Sciences, University of Virginia (P.O. Box 800717, Charlottesville, VA USA 22908)

<sup>e</sup> Department of Medicine, University of Virginia (PO Box 800793, Charlottesville, VA USA 22908)

Email addresses: Ames EG (ega2d@virginia.edu), Mackey AJ (amackey@virginia.edu), Holmes JW (holmes@virginia.edu)

Correspondence to Jeffrey W. Holmes, Department of Biomedical Engineering, University of Virginia, Box 800759, Charlottesville, VA 22908. Fax: 434-982-3870 Telephone: 434-243-6321 Email: holmes@virginia.edu

## 1. Introduction

Messenger RNA transcripts undergo significant changes via alternative splicing in many physiological transitions including organ development [7,201,202], cell differentiation [64,74,203–207], and in the pathogenesis of a wide variety of diseases such as obesity [208], cancer [209], and autism [210]. Notably during heart development there are a series of tightly regulated alternative splicing events that occur before and after birth. Like other instances of alternative splicing, these changes are mediated by only a few RNA-binding proteins (RBPs). In Chapter 3, we showed that a specific subset of isoforms expressed during heart development is re-expressed in cardiac hypertrophy [211]. The goal of the work presented here is to begin to identify regulatory mechanisms responsible for the re-expression of fetal splicing variants in hypertrophy. From the developmentally-regulated isoforms that are re-expressed in hypertrophy, we have now identified changes in individual alternative cassette exon inclusion that are shared between heart development and hypertrophy. In the regulatory regions surrounding these exons, we also identified several significantly enriched 5-base pair sequences that correspond to the binding motifs of known RBPs involved in alternative splicing. We then show that the presence of a few of these motifs is highly predictive of changes in alternative splicing seen in both heart development and hypertrophy.

Pre-mRNA transcripts undergo RNA splicing, which is the process of removing stretches of non-coding RNA (introns) from the original transcript. Some exons are always present in the final transcript, while other exons or parts of exons are alternatively included or excluded. The process of alternative splicing is highly conserved across the eukaryotic kingdom [45] and has been shown to occur in more than 90% of human genes [46]. As the RNA transcript emerges from the RNA polymerase complex, RNA-binding proteins (RBPs)

immediately recognize short, degenerate motifs (5-8 nucleotides) in the primary sequence or the RNA's developing secondary structure. These RBPs function in the early stages of splicing while the spliceosomal machinery that mediates intron removal assembles on the transcript. Ultimately RBPs influence which exons are present in the final mRNA transcript [212]. Similar to transcription factors that bind specific DNA elements to promote transcription, RBPs act as *trans*-factors and recognize *cis*-elements within the pre-mRNA transcript in the introns flanking the alternative exon as well as the exon itself. These motifs function to either promote or suppress the inclusion of the alternative exons and are referred to as exon/intron splicing enhancers (E/ISEs) and exon/intron splicing suppressors (E/ISS), respectively [55,212]. There are two main groups of RBPs that promote or repress exon inclusion via recognition of these *cis*-elements within the pre-mRNA transcript. The ability of these RBPs to influence exon inclusion not only depends on recognition of their cognate motifs, but also the location of the motifs within the primary transcript relative to the alternative exon and the sequence surrounding the motif. SR proteins (or serine/arginine-rich proteins) contain an RNA-binding domain and an RS domain that is rich in serine and arginine residues. Classically, SR proteins mainly function to promote exon inclusion by binding E/ISEs via its RNA-binding domain, which serve to demarcate the exon boundaries while the RS domain acts to facilitate interactions with the splicing machinery [213,214]. Proteins within the larger family of heterogeneous nuclear ribonucleoproteins (hnRNPs) have one or more RNA-binding domains. The function of hnRNPs are much more varied in alternative splicing; hnRNPs are able to both promote and repress splicing [215].

Many RBPs are tissue-specific and likely contribute to tissue-specific splicing patterns. The orchestration of these RBPs to act together and coordinate multiple splicing events can be thought of as a “post-transcriptional operon” [75,76]. Operons were originally

described as groups of functionally-related genes in bacteria that were physically located adjacent to one another and transcribed together [216]. Rather than physical proximity, RNA transcripts are regulated and functionally coordinated by cohorts of RBPs. RNA-binding proteins control multiple levels of tissue-specific mRNA processing including splicing, the addition of the 5' cap, and poly-adenylation. These same RBPs are also active outside the nucleus and are involved in regulating properties during translation such as transcript localization and translational efficiency. In fact there is considerable evidence that multiple RBPs act in concert and in competition to determine splicing patterns [217]. The regulation of approximately half of all alternative splicing events that occur immediately before and after birth in the heart are controlled by two RBPs, CELF and MBNL [7]. CELF proteins promote fetal splicing patterns and dominate splicing decisions in fetal hearts. After birth the protein levels of CELF decline in response to miRNA regulation [89]. This allows MBNL to dominate and promote adult splicing patterns. Here, we sought to identify RBPs that play a role in regulating re-expression of fetal isoforms during hypertrophy. From our RNAseq data, we isolated RNA sequences from five region in and flanking alternative exons that were shared between TAC and fetal hearts, but not in the sham-operated controls. We searched for RBP recognition motifs in these regions that were statistically unlikely to occur compared to a random intronic sequences from rat mRNA transcripts expressed in the heart. Using this approach also allows us to focus on single alternative cassette exons that are better understood in terms of regulatory mechanisms compared to more complex alternative splicing. We will also be able to leverage this knowledge to identify candidate RBPs that are controlling the re-expression of fetal splice variants and compare our results to the RBPs that control alternative splicing in murine heart development [7].

## 2. Materials and methods

### 2.1. Topology search to identify alternatively spliced exons

Using an approach similar to that proposed by Sammeth and Rogers et al. [218,219], we employed a topology search to identify splicing events. As an input, we used the gene names of significantly regulated isoforms from our previous analysis [211] (specifically from the Euler diagram in Chapter 3, Fig. 4A). Briefly, we converted each gene's GTF (gene transfer format) file into a directed, acyclic graph (DAG). This file contains basic information about the start and end coordinates of the gene itself, every transcript of the gene, the start and stop codons, the coding sequence, and every exon. We then created a DAG where each node represented a chromosomal location (a start or end site) and each edge represented a connection to that node that was present in the GTF (Fig. 1). The input GTF file was generated by Cufflinks [177,178] and as described in Chapter 3. The Cufflinks-generated GTF contained some transcripts that were likely an artifact of splice site identification, i.e. an exon-exon junction 2-3 nucleotides away from an exon-exon junction present in the reference genome. To correct for this artifact, we treated nodes that were less than 10 nucleotides away from another as one collective node in the construction of individual gene's DAG. We then used the Graph::Directed package in Perl [220] to search each DAG for loops that explicitly contained one edge that was also covered by three edges (Fig. 1B). This topology search was then used to identify alternatively spliced cassette exons and output basic information about the splicing event including transcript identifiers of transcripts that either contained the exon or skipped it, as well as coordinates of the upstream, alternative, and downstream exons and transcript IDs for both the exon inclusion and exclusion event. The transcript identifiers were then used to calculate the percent-spliced-in (PSI) from the FPKM calculated by Cufflinks, where  $PSI = \frac{\text{sum of transcripts' FPKM}}{\text{sum of transcripts' FPKM}}$

that contained the alternative exon / (the sum of the transcripts' FPKM that contained and excluded the alternative exon) ] \* 100. The Perl code implementing the construction of DAGs and searching for alternative exons is presented in Appendix C. To avoid more complicated splicing events (including intron retention), we employed a size cut off to only examine alternative exons that were longer than 20 nucleotides but less than 1000 nucleotides in subsequent analyses.

## *2.2. Gene ontology (GO) enrichment analysis*

To determine if different groups of alternatively spliced exons differed in biological function, we performed gene ontology analysis on all of the gene names of exons that were shared between TAC and fetal hearts compared to the gene names of alternative exons that were not different between all 3 groups using DAVID functional annotation clustering [140,199]. Clusters were grouped using to 'GOTERM\_BP\_FAT' option.

## *2.3. RBP motif search using a first-order Markov model*

The start and end coordinates of the alternative exons we identified in our topology search were used to derive the coordinates of intronic regions of interest to use as input for the RBP motif search. Fig. 2 shows the 5 regions that were analyzed in this search, hereafter referred in relation to the alternative exon as distal upstream, proximal upstream, exon, proximal downstream, and distal downstream. Previously, RBPs were thought to recognize and bind motifs near the alternative exon. However, the size of these regulatory regions were shown to extend further into introns (about 300 nucleotide) than previously expected [221] and accordingly, we included these larger regions in our analysis.

The distal upstream region spans 10 nucleotides downstream to 300 nucleotides downstream of the exon preceding the alternatively spliced exon. The proximal upstream

region is 300 nucleotides to 35 nucleotides upstream of the alternative exon. It should be noted that this region avoids the polypyrimidine tract and branchpoint nucleotide that are both essential elements for constitutive splicing of all exons and would confound our search for enriched sequences. The exon region contains the entire exonic sequence identified in the topology search. The proximal downstream region was 10 to 300 nucleotides past the alternative exon. The distal downstream region was 300 to 35 nucleotides upstream of the downstream exon. Introns less than 35 bp were excluded from the analysis. Introns that were shorter than 300 bp were used as both the proximal and distal regions.

In order to accurately compare our results to previous work to identify enriched motifs involved in alternative splicing during heart development, we followed the approach and methods of Kalsotra et al. [7]. Briefly, once each region had been defined for each alternative exon, we divided the sequences into 10 bins based on GC content to account for background sequence biases. A first-order Markov model calculated the relative single nucleotide and dinucleotide frequencies within a given GC bin. The sequences were then analyzed to count all observed pentamers within the sequences. The expected frequency of each pentamer was calculated by multiplying the probabilities of the following frequencies: the frequency of the first nucleotide, the frequency of the dinucleotide pair of the first and second nucleotides divided by the sum of all possible dinucleotide frequencies that start with the first nucleotide, the frequency of the dinucleotide pair of the second and third nucleotides divided by the sum of all possible dinucleotide frequencies that start with the second nucleotide, the dinucleotide frequency of the third and fourth nucleotides divided by the sum of all possible dinucleotide frequencies that start with the third nucleotide, and the dinucleotide frequency of the fourth and fifth nucleotides in the pentamer divided by the sum of all possible dinucleotide frequencies that start with the fourth nucleotide. The probability

of a given pentamer can be calculated with the following equation where  $X_n$  represents each nucleotide in a given pentamer:

$$\begin{aligned}
 & \text{probability of } X_1X_2X_3X_4X_5 \\
 &= \left( \frac{X_1}{X_A + X_C + X_G + X_U} \right) \times \left( \frac{X_1X_2}{X_1X_A + X_1X_C + X_1X_G + X_1X_U} \right) \\
 &\times \left( \frac{X_2X_3}{X_2X_A + X_2X_C + X_2X_G + X_2X_U} \right) \times \left( \frac{X_3X_4}{X_3X_A + X_3X_C + X_3X_G + X_3X_U} \right) \\
 &\times \left( \frac{X_4X_5}{X_4X_A + X_4X_C + X_4X_G + X_4X_U} \right)
 \end{aligned}$$

The Perl code to implement the first order Markov model is provided in Appendix C. Statistical significance of each pentamer count was determined by using a binomial test of the observed to expected ratio which was implemented in R ([www.R-project.org](http://www.R-project.org)). This  $p$ -value was corrected by the false discovery rate.

#### 2.4. Motif enrichment heat map

The fold enrichment of individual pentamers of interest was calculated as follows:  $\log_2(\text{the sum of observed counts of pentamer } X \text{ across all GC bins} / \text{the sum of expected counts of pentamer } X \text{ across all GC bins})$ . These  $\log_2$  enrichment factors were then used to generate a heat map in R using `gplots` package [200]. Any motif that had an enrichment  $p$ -value  $< 0.05$  was included on these graphs.

#### 2.5. Statistical testing of the relationship between a RBP motif and a predicted change in PSI

We used a cumulative hypergeometric probability density function to calculate the probability of seeing  $x$  or more exons with a significant change in PSI that also had a

specific motif in a given region assuming that each exon was regulated independently with the following equation:

$$P(x) = \sum_{i=x}^N \frac{\binom{K}{x} \binom{M-K}{N-x}}{\binom{M}{N}}$$

where  $K$  is the number of exons with the motif,  $x$  is the number of exons that had both the motif and a significant change in PSI,  $N$  is the number of exons that had a significant change in PSI,  $M$  is the total number of exons compared,  $i$  is the summation index. Individual calculations were performed using `hygepdf` function in MATLAB (Natick, MA) [222,223]. To further explore the significance of positive relationships between a motif and a significant change in PSI, we tested the relationship at multiple PSI cut offs (10%, 15%, and 20% difference between TAC minus sham and fetal minus sham). If the  $p$ -value for two of the three comparisons was  $< 0.05$ , we considered this motif to have a significant effect on exon inclusion. To calculate the proportion of the exons that could be potentially explained by the presence of this motif, we divided the difference between the number of observed exons and the number of expected exons by the total number of exons that met the PSI cut off criteria. We also know that pairs of RBPs can act together to influence splicing decisions in two main arrangements: 1) when one RBP binds in the proximal upstream region and the other RBP binds in the proximal downstream region or 2) when one RBP binds in the proximal downstream region and the other RBP binds in the distal downstream region [224]. We tested the motifs that significantly predicted a change in exon inclusion (listed in Table 1) with the other overlap-specific motifs in the regions known to be involved in pairwise regulation of alternative splicing to determine if any combination of two RBP motifs could increase the percent of exons explained by the presence of either motif.

### 3. Results

#### *3.1. Search for alternatively spliced exons identifies exons that are fetal-specific, TAC-specific, regulated in both groups, or not regulated in either group*

Our search for alternative exons identified 1,365 alternative cassette-type exons. We used the gene names of the significantly regulated isoforms that we previously identified [211] and in Chapter 3, Figure 4A as the input for our topology search and then separated these exons into four groups of interest. We divided the cassette exons into the groups using the percent-spliced-in (PSI) of each exon to determine differential expression. Nearly one-fifth (258 exons or 18.9% of identified alternative exons) are alternatively spliced, but not significantly different among TAC, sham, and fetal hearts. We considered this group of exons to be our negative control, or “null” group of alternative exons. The remaining exons were either classified as fetal-specific (256 exons, 18.7%), TAC-specific (36 exons, 2.6%), or exons that were shared between fetal and TAC hearts (815 exons, 59.7%). As an additional control group, we also identified single exon genes (472 genes) that would presumably have no motifs for alternative or constitutive splicing factors. A histogram showing the size distribution of these exons suggests that the exons we identified are smaller than the constitutive exons in our data set (Supplementary Fig.1), where constitutive exons were defined as an exon that was present in all three groups and no significant difference in PSI (< 20% difference between both TAC and fetal compared to sham). In addition, we used gene ontology analysis to identify differences in biological functions for genes undergoing alternative splicing selectively in TAC and fetal hearts compared to alternative exons that were not significantly different in any group. The gene ontology group uniquely enriched in our overlap group contained specialized terms similar to our previous work [211] and to results published on alternative splicing in heart development [7] including

'cytoskeleton organization' and 'muscle cell differentiation' (full gene ontology results shown in Supplementary Table 1). In contrast, there were less significant functional clusters for the null exons and the clusters were enriched for terms associated with basic cellular functions such as 'tRNA aminoacylation' and 'intracellular transport'.

### *3.2. Significant pentamers in regulatory region correspond to RBP recognition motifs*

Our search for significant motifs identified several pentamers in each of the regions flanking our alternative exons (as shown in Fig. 2). We calculated the statistical significance of all 1,024 possible pentamers in both the overlap and null groups and plotted the negative log<sub>10</sub> of the adjusted *p*-value in Figure 3. While almost all of the motifs are degenerate sequences, it is important to note that these motifs correspond to known recognition motifs of RBPs. The significant motifs that agree with previous pentamers identified as being enriched in developmentally-spliced alternative exons in the heart [7] are denoted by an asterisk next to pentamer in Fig. 3. The significant pentamers for the overlap group include: distal up - AAAAA\*, ACACA\*, AGAGA, CACAC\*, CCUCC, CUCUC\*, GAGAG, GUGUG, UAUAU, UCUCU, UGUGU, UUUUAU, UUUUU\*; proximal up - AAAAA, ACACA, AGAGA\*, CACAC, CUCUC\*, GAGAG, GUGUG\*, UCUCU\*, UGUGU\*, UUUUU\*; exon - ACACA, CACAC, CUGCU, GCUGC, UCUUC, UGCUG, UUCUU, UUUUU; proximal down - AAAAA, AAGAA, ACACA, AGAGA, AUAUA\*, CACAC, CUCUC, GAGAG, UAUAU, UCUCU, UCUGU, UGUCU, UUUUU\*; and distal down - GUGUG\*, UGUGU\*, UUUUU. We also compared our motif analysis results to the pentamers that were identified as significant in murine and galline fetal heart development. The results are listed in Supplementary Table 2 and also list the RBP that recognizes each motif, if known. The significant motif plots of the other alternative exon groups are shown as supplementary figures (fetal-specific exons in Supplementary Fig. 2, TAC-specific exons in Supplementary Fig. 3, and single-exon genes

in Supplementary Fig. 4). Single exon genes that would be expected to have no recognition motifs for alternative splicing were enriched for pentamers not seen in any of the alternative splicing exons including highly GC-rich motifs CGCGC, GCGCG, CUUGG, CGCCU, and GGGGG.

### *3.3. RBP motifs are enriched, rather than depleted, in introns flanking alternative splicing events*

The negative log<sub>10</sub> of each pentamer's *p*-value only represents the probability that the observed count deviated from what our model predicted and not whether the motif is more common than the expected frequency (enriched) or less common than expected (depleted). When we compared motif enrichment values for the motifs that were identified as significant, we found that most of them are enriched rather than depleted as is illustrated by the abundance of red in the heat maps in Fig. 4 and Supplementary Fig. 5. There were several motifs that were significantly over-represented and enriched in only the overlap group. These overlap-specific motifs include: distal up - AGAGA, CCUCC, CUCUC, UCUCU, UUUAU; proximal up - UGUGU, UCUCU, GAGAG, AGAGA, CUCUC; exon - GCUGC, UGCUG, CUGCU, UUUUU; proximal down - AAGAA, CUCUC, UGUCU, UCUCU, UCUGU, AUAUA, UAUAU; and distal down - UGUGU, UUUUU.

### *3.4. The presence of specific RBP motifs significantly influences exon inclusion*

If we assumed as a null hypothesis that the regulation of each alternative exon is controlled independently and that the presence of RBP motifs has no effect on this outcome, we can calculate the number of exon inclusion events we would expect to be shared between TAC and fetal hearts, but not sham hearts. To test the relative effect that each motif's presence might have on exon inclusion, we used a cumulative hypergeometric probability density function to calculate the probability that a motif's presence in a given

region significantly promoted or repressed exon inclusion. Of the overlap-specific enriched pentamers that we identified in section 3.3, only a few motifs (distal up - AAAAA, CCUCC, UUUUAU; proximal up - ACACAC, CUCUCU) were significantly predictive of an effect on exon inclusion; these results are shown in Table 1, Figure 5, and Supplementary Table 3. Figure 5 shows that most of the motifs we identified as predictive were previously implicated to have a role in alternative splicing during heart development (denoted with an asterisk). When we looked for pairwise interactions that predicted a significant change in PSI, we found that the presence of the 'CUCUCU' motif in the proximal upstream region and the 'UCUGU' motif in the proximal downstream region was able to explain more of the exons with a decrease in exon inclusion than either motif alone.

#### **4. Discussion**

The re-expression of fetal isoforms in hypertrophy is an exciting finding and in order to identify potential RBPs that could be mediating these changes in alternative splicing we looked for RBP recognition motifs in the regions flanking alternative cassette exons. Our computational methods identified significantly enriched pentamers flanking alternatively spliced cassette exons. These pentamers correspond to RNA-binding proteins that are known to mediate alternative splicing events and their predicted effects on splicing are shown in Figure 4. Overall, our analysis identifies strong candidates for the RBPs that are mediating the re-expression of fetal splice variants in hypertrophy.

First we identified alternative cassette exons within isoforms significantly regulated in both TAC and fetal hearts. When we examined the alternative exon sizes, we found that there is a clear trend toward smaller exons compared to constitutive exons. The average length of all alternatively spliced exons in the human and mouse genomes is 107 bp [218],

which is smaller than the average mammalian exon (173 bp, an average across human, mouse, dog, and chimp genomes) [225]. The smaller size of the identified exons suggests that they are candidates for alternative splicing. When we compared the gene ontology terms associated with the overlap and null groups, we saw a striking similarity between the enriched terms in our data and the terms associated with genes that undergo alternative splicing in heart development [7]. In contrast, we saw no similarities between previously published work and our null group of exons. The GO terms associated with the null exons are basic cellular functions and possibly represent alternative splicing events that occur in all hearts regardless of development or disease state. The majority of the significant motifs we identified were enriched rather than depleted, suggesting that the presence of these motifs is functionally important for alternative splicing. Similar to the fact that a gene requires a specific transcription factor binding site to be regulated by that transcription factor, alternative splicing requires the presence of RBP recognition motifs. The presence of these significantly enriched RBP motifs strongly suggests that their presence is functionally important for regulation of alternative splicing.

The polypyrimidine-tract binding protein (PTB) family of RNA-binding proteins recognizes RNA motifs rich in cytidine and uridine residues as shown by multiple experimental approaches including electrophoretic mobility shift assays, fluorescence methods, UV cross-linking, and NMR [226–228]. The pentamers ‘CUCUC’ and ‘UCUCU’ were significantly enriched in the proximal upstream region of exons that were alternative spliced in both fetal heart development and hypertrophy. Like other members of the hnRNP family, it has previously been shown that when PTB binds upstream of an exon it promotes exon exclusion [229], and our statistical testing shows that PTB recognition motifs in this region are associated with more exon exclusion than would be expected if each exon was

independently regulated. We further discuss the mechanism of how PTB may be regulated in hypertrophy in Chapter 5.

The pentamer 'AAAAA' was enriched in the distal upstream region and was statistically associated with exon inclusion. There are multiple RBPs that have been experimentally demonstrated to bind A-rich sequences by homopolymer assays. Among the RBPs that could be binding A-rich motifs, two notable candidates are the Mex3 family and polyadenylate-binding proteins. The Mex3 (muscle excess 3) family of RBPs was first identified in *C. elegans* as causing excessive body wall muscles when deleted. While the human homologs of these proteins are not known to mediate alternative splicing, they are known to bind mRNA and influence translation (see Chapter 5 for a further discussion of the Mex3 family). Similarly, polyadenylate-binding proteins are also involved in multiple components of mRNA processing including 3' end processing [230], polyadenylation [231], cap-dependent translation [232], and nonsense mediated decay [233]. The potential interaction between RBPs that are involved in translation regulation and transcripts that are alternatively spliced in heart development and hypertrophy may suggest that Mex3 or PABP are part of a 'hypertrophic' post-transcriptional operon.

The 'ACACAC' motif that we identified in our analysis corresponds to the recognition motif for heterogeneous nuclear ribonucleoprotein L (hnRNP L). Of the common simple-sequence repeats in the murine and human genomes, 'AC' repeats are the most frequent repeat, occurring 62.1 and 19.4 times per 1 megabase of sequence, respectively [234]. hnRNP L has been shown to regulate multiple steps in mRNA processing including alternative splicing in multiple tissues [235] and has been suggested to be a global splicing regulator [236]. Aside from its role in alternative splicing, hnRNP L binds 'CA' repeats in single exon genes to promote nuclear export [237]. Given its high abundance in mammalian

genomes, it is interesting to note that the 'ACACAC' motif was significant in multiple other regions and groups, including the single exon genes (Supplementary Figure 4). Despite the fact that it was significantly enriched in other groups, the presence of this motif in the proximal upstream region was overlap-specific and significantly predictive of exon inclusion in both TAC and fetal hearts. In addition, the 'ACACAC' motif was shown to be significantly enriched in other regions flanking exons that were alternatively spliced in fetal heart development [7], suggesting that hnRNP L is a strong candidate regulator of alternative splicing in the heart. While the function of hnRNP L has not been specifically studied in the heart, given its ability to promote and repress exon inclusion [238] in various tissues [239–241] and the fact that it is known to physically interact with PTBP1 [242], it would be interesting to explore the role of hnRNP L in the heart more closely.

When we tested if there were any combinations of two RBP recognition motifs that increased the percentage of alternatively spliced exons that were accounted for by the presence of motifs, the combination of the PTB motif in the proximal upstream region and the 'UCUGU' motif in the proximal downstream region had nearly 50% more exons than expected with a decreased PSI difference between TAC minus sham and fetal minus sham greater than 20%. The specific RBP that recognizes the 'UCUGU' motif is not known, yet it is very similar to the CUG- and UG-rich motifs recognized by CELF proteins [243,244]. While the PTB motif significantly predicted decreased exon inclusion, the 'UCUGU' motif was not predictive alone. It is interesting to note that the CELF (CUG-binding protein and ETR3-like factors) family has previously been shown to mediate the re-expression of fetal splicing variants as a gain-of-function mutation in myotonic dystrophy and fetal heart development [7,245]. In applying these findings, it will be important to establish the protein levels and cellular locations of these RBPs during hypertrophy.

Historically, alternative splicing is generally viewed as single exon events, i.e. the inclusion/exclusion of one exon, the retention of one intron, or the use of one alternative splice site. With high-throughput sequencing and more sophisticated computational approaches, we now know that there are more than 24,000 alternative splicing events in humans and a large fraction ( > 27%) involve more than two exons [218]. We are also beginning to understand that the input parameters into splicing decisions are extremely complex and include many variables such as the expression and activity of RBPs, underlying nucleosome spacing, chromatin marks, speed of transcription, the strength of individual splice sites, and the presence of pseudo-splice sites that are littered throughout introns [212]. Pseudo-splice sites are sequences within introns that match the consensus splice site as well or better than the true splice site, but are not used unless the actual splice is inactivated by a mutation [246]. With a system as complex as this, multiple groups have tried to assemble the basic rules or patterns that are able to predict splicing decisions. The most complete example to date examined splicing patterns of 3,665 cassette-type exons in 27 murine tissues (including whole embryos at multiple time points) and used machine learning to identify the 200 most important features for exon inclusion/exclusion. To increase their power to identify patterns, these tissues were combined into four groups (nervous system, muscle, digestive system and embryonic tissues). Ultimately, they found that tissue-specific factors (i.e. RBPs) are critical for splicing decisions [221]. As we accumulate more sequencing data from RNA from more specific tissues and different species, our ability to predict tissue-specific splicing rules will likely increase. While we appreciate the immense variation in splicing regulation and transcript composition, we chose to focus on what potential RBPs may be controlling single exon splicing events that were observed in both rat fetal heart development and a surgical model of cardiac hypertrophy in rats. Since this type

of splicing is better understood compared to more complex patterns, this is the logical first step to pursue the mechanism regulating the patterns of alternative splicing that we have previously identified.

During our process to identify alternative splicing events, significantly enriched motifs surrounding them, and what effect those motifs may have on splicing, we made several simplifications and assumptions that should be recognized. First and foremost, we chose to focus on single alternative cassette exons. While we know that splicing events are much more complex [218], much of the experimental knowledge about how RBPs control splicing pertains to individually spliced exons. By identifying candidate regulators of these splicing events, we wanted to ensure that we could compare our computational results to previously published work that determined the relationships between RBPs and alternative splicing in fetal heart development. When calculating the effect that individual RBP recognition motifs may have on splicing decisions, we used an estimate of PSI derived from FPKM calculations from Cufflinks. Recently some groups have questioned this approach, recommending instead that exon-exon junction reads should be used in calculating PSI [247]. By focusing on the simplest case of single cassette exons, our methods were sufficient to determine the patterns that we were interested in, i.e. differences in exon inclusion in TAC or fetal hearts compared to sham.

## **5. Conclusions**

We have identified pentamers that are significantly enriched in the regions flanking alternative exons regulated in both heart development and hypertrophy. These pentameric motifs correspond to RBPs that are known to regulate alternative splicing. The presence of

several of these RBP recognition motifs significantly predicts instances of exon inclusion and exclusion. The main RBP identified in this work, PTB, is known to regulate developmental splicing in the heart and represent a strong target to validate their role in controlling alternative splicing in cardiac hypertrophy and the re-emergence of fetal splicing patterns.

### **Sources of funding**

This work was funded by NIH R01 HL-075639 (JWH), NIH T32 GM-007267-34 (EGA), and NIH T32 GM-008136-27 (EGA).

### **Disclosures**

None declared.

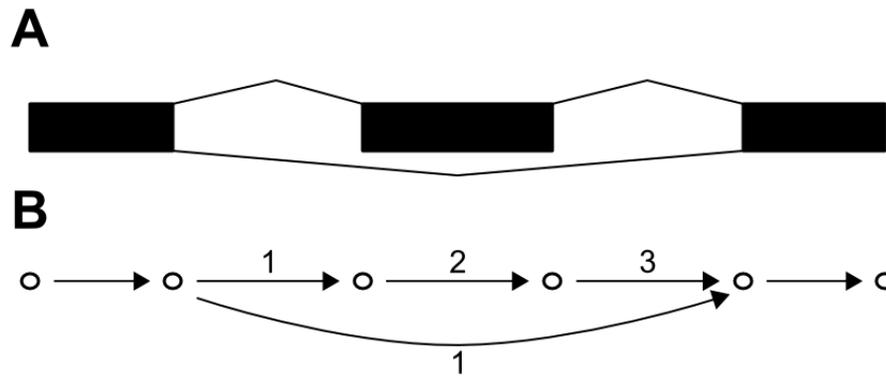
## Appendix A. Tables and figures

Table 1. Significant motifs with predictive effect on exon inclusion

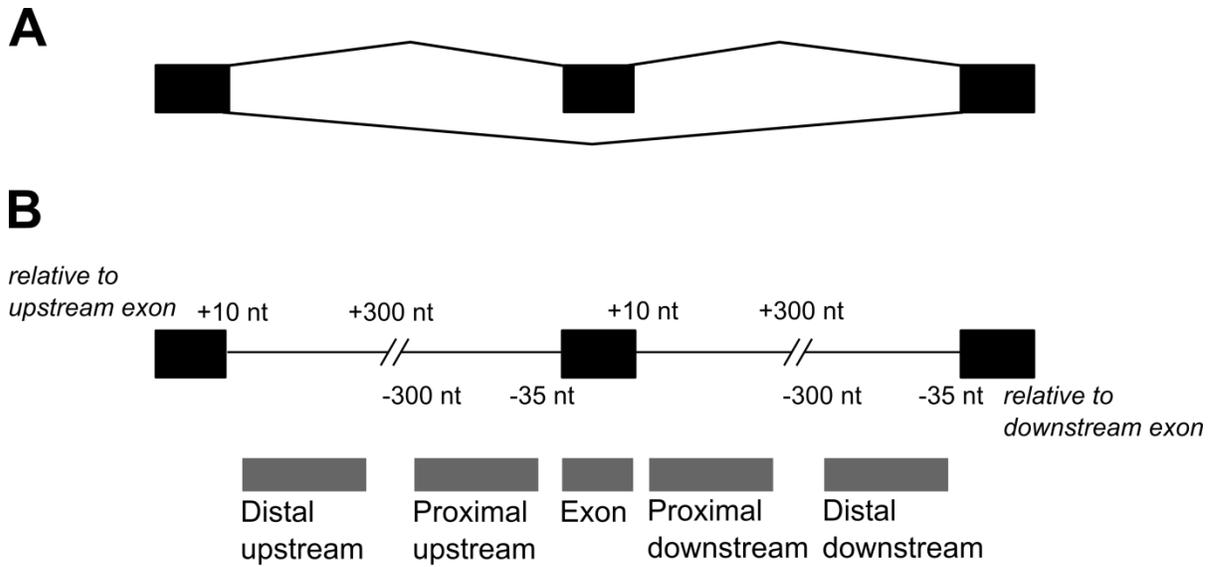
Motif	Location	Effect on splicing	P-value	Percent increase above expected
AAAAA	Distal upstream	Increased inclusion	**	40%
CCUCC	Distal upstream	Decreased inclusion	*	24.5%
UUUAU	Distal upstream	Increased inclusion	**	42.9%
CUCUCU	Proximal upstream	Decreased inclusion	*	31%
ACACAC	Proximal upstream	Increased inclusion	*	27%
CUCUCU / UCUGU	Proximal upstream / proximal downstream	Decreased inclusion	**	47.9%

P-value was calculated by summing the probability of seeing X or more exons regulated in a given data set using a hypergeometric probability function. Percent explained refers to the percent of exons explained by the presence of a given motif. This percent was calculated by (observed number of exons - expected number of exons) / total number of exons that met PSI cut off criteria of > 20%. The expected number of exons was determined by the maximum of the probability density function. \* denotes  $p < 0.05$ , and \*\* denotes  $p < 0.01$ .

Figure 1. Alternatively spliced exon represented as a directed acyclic graph



**Figure 1. Alternatively spliced exon represented as a directed acyclic graph.** **A**, Standard representation of an alternative exon. **B**, Start and end coordinates converted to nodes (depicted as open circles) and edges represent the connections as written in GTF file (lines with arrowheads).



**Figure 2. Sequence regions used as input for RBP motif search.** **A**, Standard representation of an alternative exon. **B**, Grey boxes show regions and their names used in RBP motif search. These regions, including the alternative exon are the most common sites for RBPs to bind and regulate splicing decisions.

*Figure 2. Sequence regions used as input for RBP motif search*



Figure 4. Motif enrichment for significant motifs across overlap, fetal specific, TAC-specific, null, and single exon genes

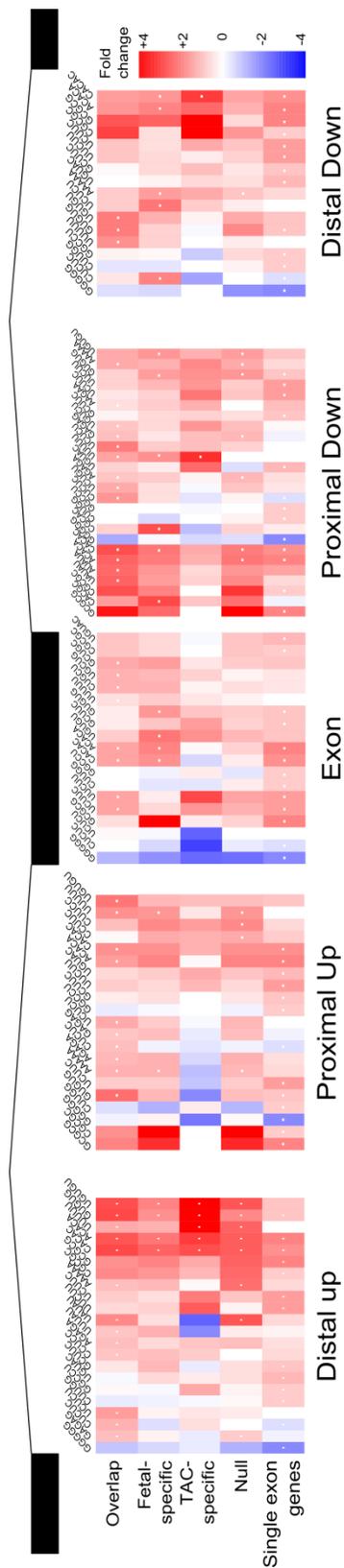
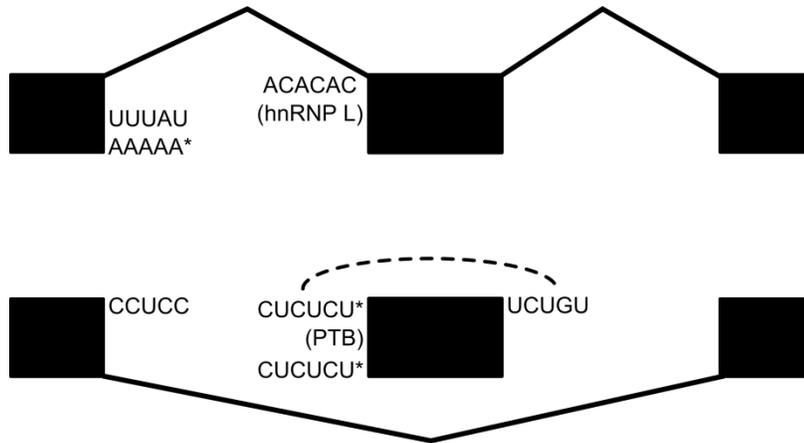


Figure 4. Motif enrichment for significant motifs across overlap, fetal-specific, TAC-specific, null, and single exon genes. Each motif that was significant in one group is listed above the heat map. Fold enrichment of each listed motif was calculated from the log2 (observed / expected) for each pentamer. White dots denote the motif was statistically significant for that group and region.

Figure 5. Summary of significant motifs and their effect on splicing



**Figure 5. Summary of significant motifs and their effect on splicing.** Motifs are listed in the region where they were found to be enriched and had a significant effect on splicing. \* denotes motif that was significantly enriched in fetal heart development splicing [7]. Dashed line signifies that the pair of motifs significantly predicted an effect on exon inclusion.

## Appendix B. Supplementary data

*Supplementary Table 1. Gene ontology (GO) analysis of biological processes associated with alternative exons significant in both heart development and hypertrophy and exons that were not significantly different in TAC, fetal, or sham hearts*

Gene ontology term	No. of genes	Fold enrichment	P-value
<i>Overlap exons</i>			
Cytoskeleton organization (GO:0007010)	19	3.01	6.1E-05
Regulation of actin filament-based process (GO:0032970)	8	6.02	3.4E-04
Negative regulation of gene expression (GO:0010629)	21	2.17	2.9E-03
Phosphorylation (GO:0016310)	22	1.55	4.5E-02
Membrane organization (GO:0016044)	18	3.13	6.4E-05
Muscle cell differentiation (GO:0042692)	7	2.99	3.0E-02
Intracellular transport (GO:0046907)	23	2.36	3.0E-04
<i>Null exons</i>			
tRNA aminoacylation for protein translation (GO:0006418)	5	11.64	8.4E-04
Intracellular transport (GO:0046907)	6	2.78	3.6E-03
Vesicle-mediated transport (GO:0016192)	9	2.34	3.7E-02

Representative GO terms from significant GO term clusters from the 815 overlap exons and the 258 null exons. No. of genes = number of genes associated with that GO term in the dataset. Fold enrichment quantifies the extent that each GO term is overrepresented in the dataset compared to the rat genome. A P-value signifies statistical significance of the degree of enrichment.

*Supplementary Table 2. Comparison of significant motifs in murine fetal heart, galline heart development, and overlap group of hypertrophy and rat fetal heart development*

<b>Distal upstream</b>	Mouse heart development	Chicken heart development	Rat heart development + hypertrophy	Recognition motif for
AAAAA	•		•	
ACACA	•		•	hnRNP L
AGAGA			•	
CACAC	•		•	hnRNP L
CCUCC			•	
CGUGG	•	•		
CUCUC	•		•	PTB
CUGGG	•			
GACAG	•			
GAGAG			•	
GCGUC	•			
GGGAA	•			
GUGUG			•	CELF
UAUUAU			•	
UCUCU			•	PTB
UGUGU			•	CELF
UUUAU			•	
UUUUU	•	•	•	

<b>Proximal upstream</b>	Mouse heart development	Chicken heart development	Rat heart development + hypertrophy	Recognition motif for
AAAAA			•	
AAUAU	•			
ACACA			•	hnRNP L
AGAGA	•	•	•	
AUAGA	•			
CACAC			•	hnRNP L
CCUUU	•			
CUCUC	•	•	•	PTB
GAGAG			•	
GCGGG	•			
GUGUG	•	•	•	CELF
UCUCU	•	•	•	PTB
UGUGU	•	•	•	CELF
UUUCU	•	•		
UUUUU			•	

Supplementary Table 2 (cont.)

<b>Proximal downstream</b>	Mouse heart development	Chicken heart development	Rat heart development + hypertrophy	Recognition motif for
AAAAA			•	
AAGAA			•	
ACACA	•		•	hnRNP L
AGAGA			•	
AUAUA	•	•	•	
CACAC			•	hnRNP L
CUCUC			•	
GAGAG			•	
GAGGG	•			
GCAUG	•	•		FOX
GCGUG	•			CELF
GGUUU	•			
GUGUG	•	•		CELF
UAAAG	•			
UAACC	•			
UAUUA			•	
UCUCU			•	
UCUGU			•	
UGUCU			•	
UGUGU	•	•		CELF
UUUCU	•	•		PTB
UUUUU	•	•	•	
<b>Distal downstream</b>	Mouse heart development	Chicken heart development	Rat heart development + hypertrophy	Recognition motif for
AAUUA	•			
AGAAA	•			
CGUGU	•			CELF
CUCUC	•			PTB
GGAAA	•	•		
GUGUC	•			CELF
GUGUG	•	•	•	CELF
UCUCU	•			PTB
UGUGU	•	•	•	CELF
UUUUU			•	

Data for pentamer enrichment for developmentally-regulated exons in mouse and chicken hearts from Kalsotra et al. [7]. Dots indicate pentamer was significant in analysis. Shading reflects the level of agreement between the three comparisons; purple - motif was significant in all three analyses, green - motif was significant in mouse and rat, blue - motif was significant in mouse and chicken hearts.

*Supplementary Table 3. Significant motifs used in hypergeometric probability density function probabilities at different PSI cut off criteria*

Motif	Location	Effect on splicing	10% PSI cut off			
			Observed	Expected	<i>P</i> -value	Percent increase above expected
CCUCC	Distal upstream	Decreased inclusion	54	48	0.142	11.4%
AAAAA	Distal upstream	Increased inclusion	39	29	0.021	24.4%
UUUAU	Distal upstream	Increased inclusion	34	23	0.005	32.9%
CUCUCU	Proximal upstream	Decreased inclusion	48	47	0.495	1.1%
ACACAC	Proximal upstream	Increased inclusion	55	46	0.047	15.5%
CUCUCU / UCUGU	Proximal upstream / proximal downstream	Decreased inclusion	34	27	0.062	21.6%

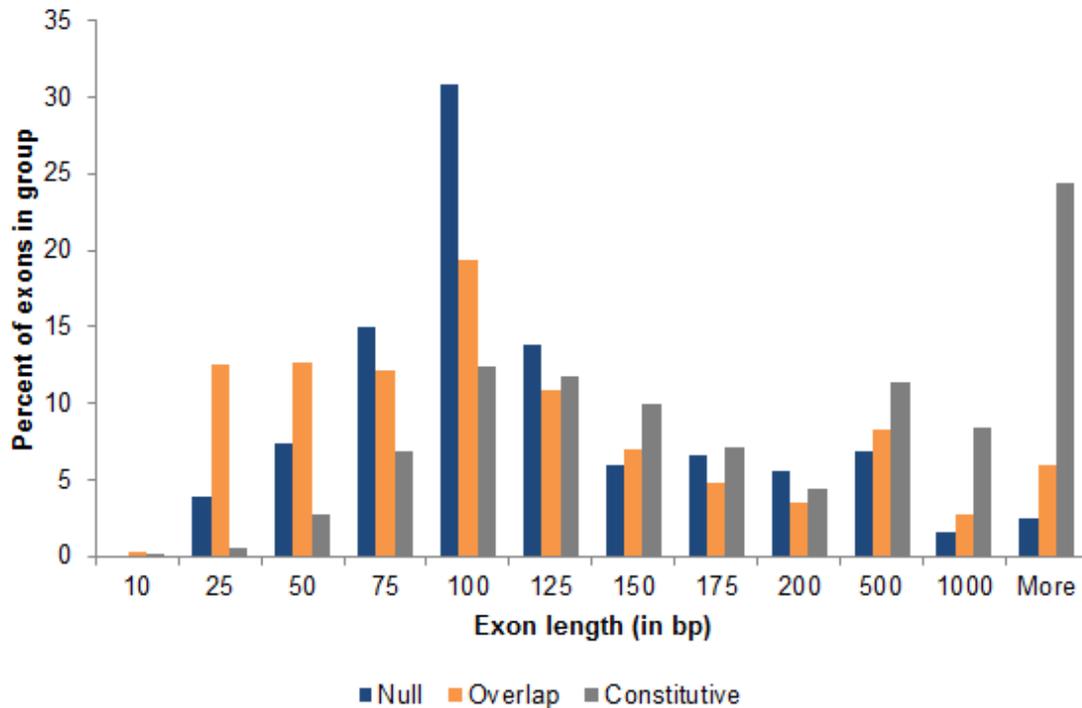
Motif	Location	Effect on splicing	15% PSI cut off			
			Observed	Expected	<i>P</i> -value	Percent increase above expected
CCUCC	Distal upstream	Decreased inclusion	40	32	0.0462	20.0%
AAAAA	Distal upstream	Increased inclusion	29	19	0.0048	34.6%
UUUAU	Distal upstream	Increased inclusion	24	15	0.0051	38.9%
CUCUCU	Proximal upstream	Decreased inclusion	40	32	0.0409	20.6%
ACACAC	Proximal upstream	Increased inclusion	36	30	0.0761	17.0%
CUCUCU / UCUGU	Proximal upstream / proximal downstream	Decreased inclusion	30	18	0.0012	40.5%

Supplementary Table 3 (cont.)

Motif	Location	Effect on splicing	20% PSI cut off			
			Observed	Expected	P-value	Percent increase above expected
CCUCC	Distal upstream	Decreased inclusion	26	20	0.0486	24.5%
AAAAA	Distal upstream	Increased inclusion	21	13	0.0044	40.2%
UUUAU	Distal upstream	Increased inclusion	17	10	0.0082	42.9%
CUCUCU	Proximal upstream	Decreased inclusion	28	20	0.0121	30.5%
ACACAC	Proximal upstream	Increased inclusion	27	20	0.0182	26.8%
CUCUCU / UCUGU	Proximal upstream / proximal downstream	Decreased inclusion	21	11	0.0011	47.9%

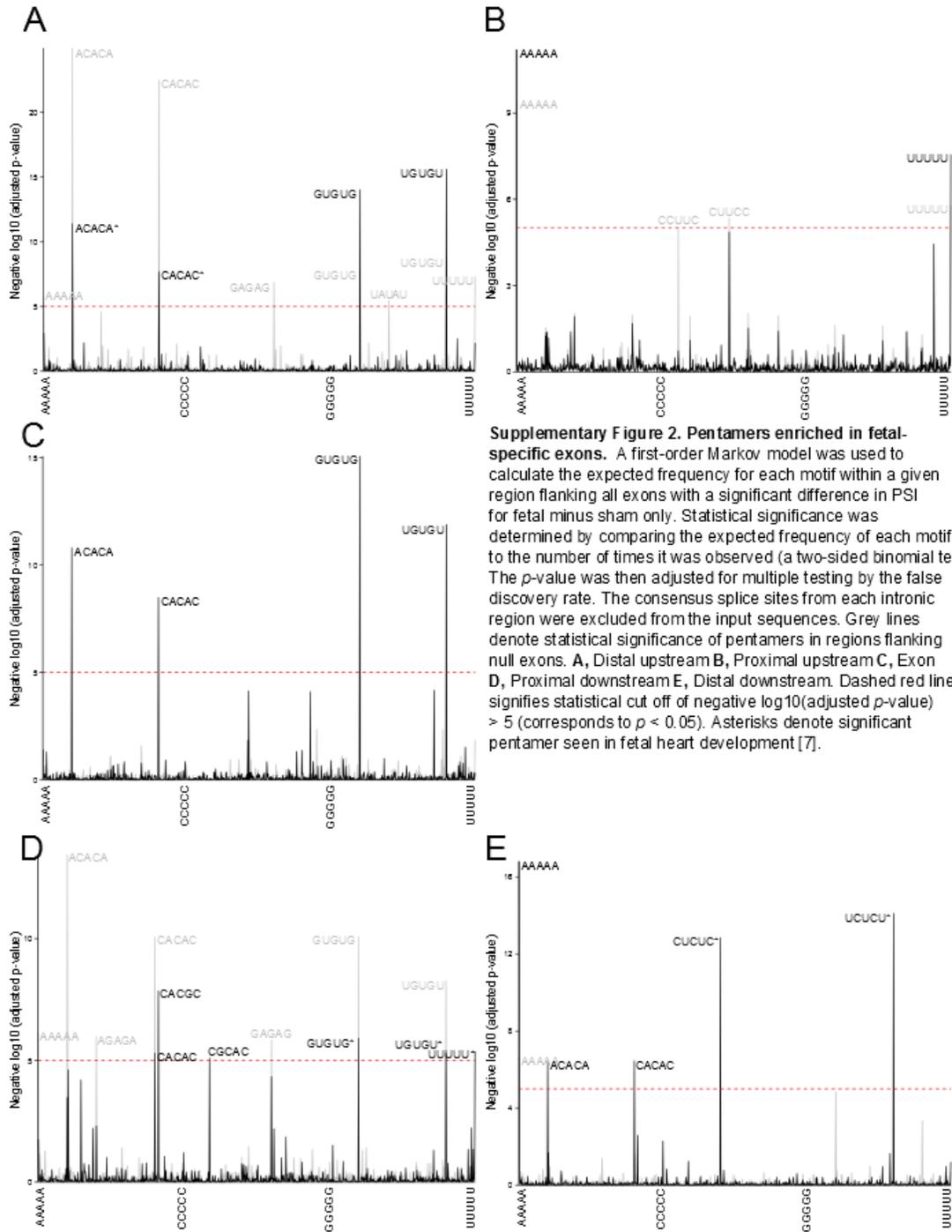
Percent PSI cut off refers to the minimal difference in percent-spliced-in (PSI) between both TAC minus sham and fetal minus sham. Observed refers to the number of exons that had the motif and met the PSI cut off criteria. Expected was calculated from the product of percentage of total exons that had the motif, the percentage that met the PSI cut off criteria, and the total number of exons. P-value was calculated by summing the probability of seeing x or more exons regulated in a given data set using a hypergeometric probability function. The percent of exons explained by motif was calculated by (observed number of exons - expected number of exons) / total number of exons that met PSI cut off criteria.

Supplementary Figure 1. Exon size distribution among alternative and constitutive exons



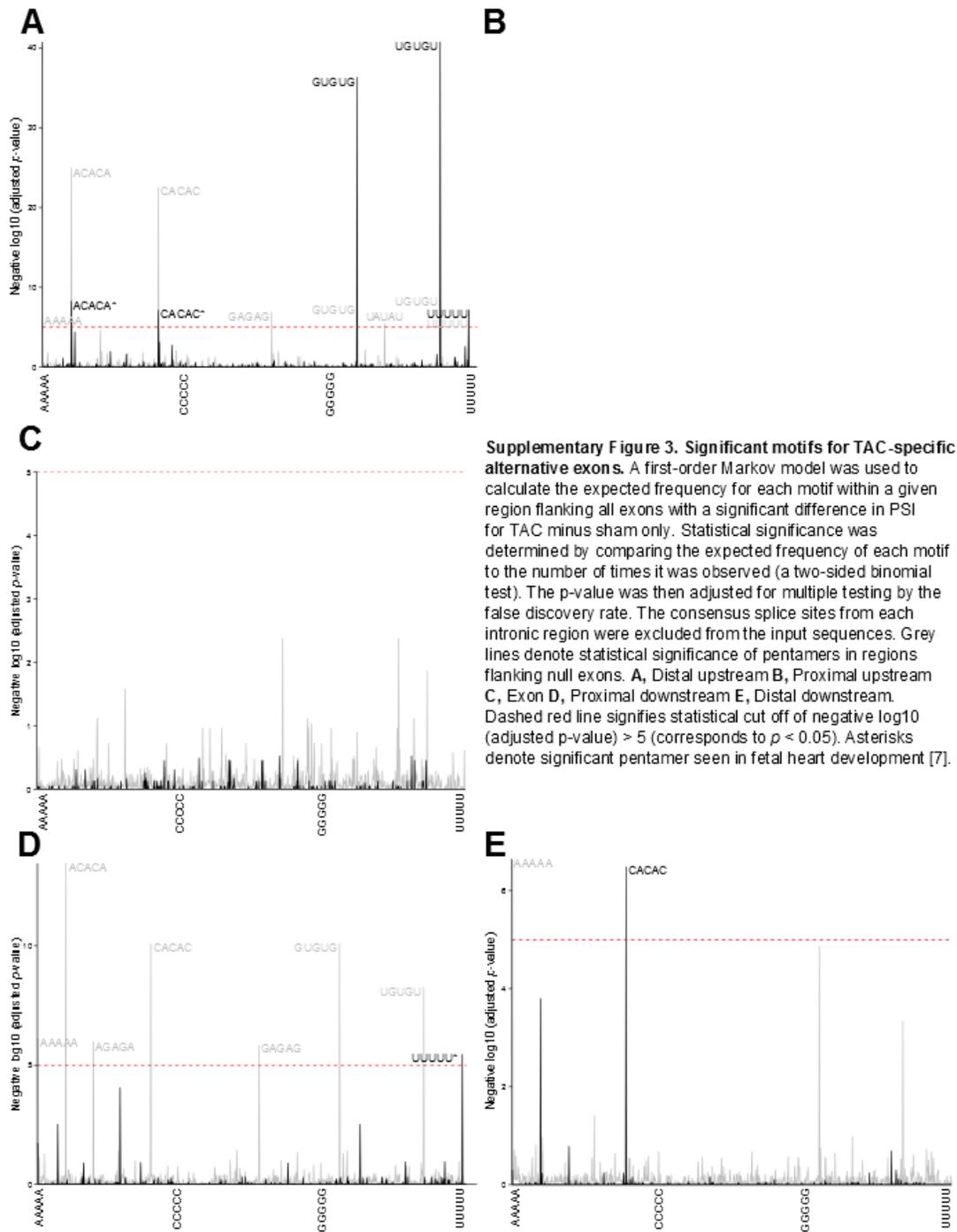
**Supplementary Figure 1. Exon size distribution among alternative and constitutive exons.** Exons that were identified by topology search as alternative cassette exons were either categorized as part of the overlap group (isoform significant in both TAC and fetal comparisons to sham,  $n = 815$ ) or the null group (no significant difference between all three groups,  $n = 258$ ). Exons that were present in each transcript of a gene were assigned to the constitutive group ( $n = 8,118$ ).

Supplementary Figure 2. Pentamers enriched in fetal-specific exons

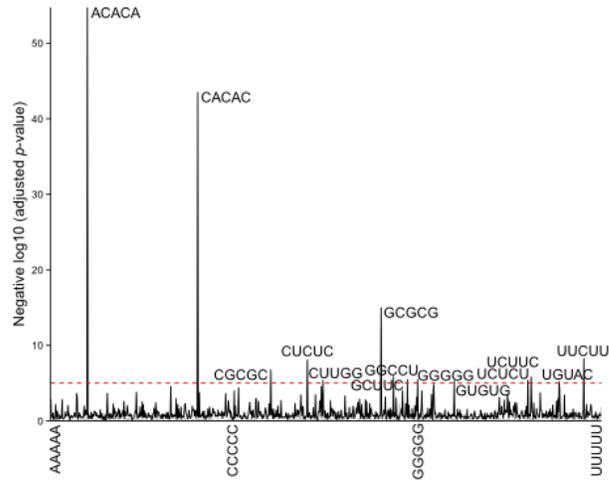


**Supplementary Figure 2. Pentamers enriched in fetal-specific exons.** A first-order Markov model was used to calculate the expected frequency for each motif within a given region flanking all exons with a significant difference in PSI for fetal minus sham only. Statistical significance was determined by comparing the expected frequency of each motif to the number of times it was observed (a two-sided binomial test). The  $p$ -value was then adjusted for multiple testing by the false discovery rate. The consensus splice sites from each intronic region were excluded from the input sequences. Grey lines denote statistical significance of pentamers in regions flanking null exons. A, Distal upstream B, Proximal upstream C, Exon D, Proximal downstream E, Distal downstream. Dashed red line signifies statistical cut off of negative  $\log_{10}(\text{adjusted } p\text{-value}) > 5$  (corresponds to  $p < 0.05$ ). Asterisks denote significant pentamer seen in fetal heart development [7].

Supplementary Figure 3. Pentamers enriched in TAC-specific exons

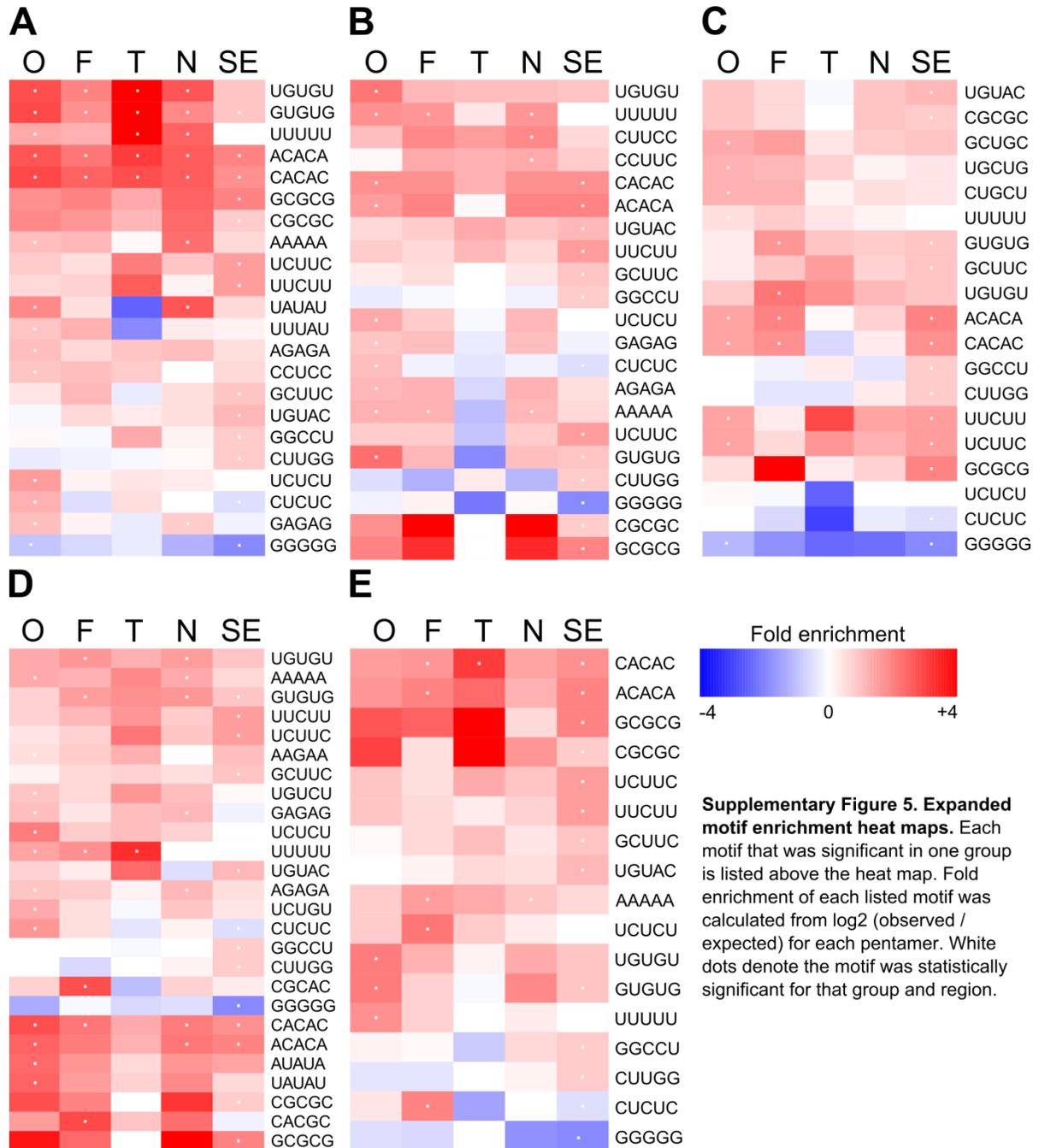


Supplementary Figure 4. Pentamers enriched in single exon genes

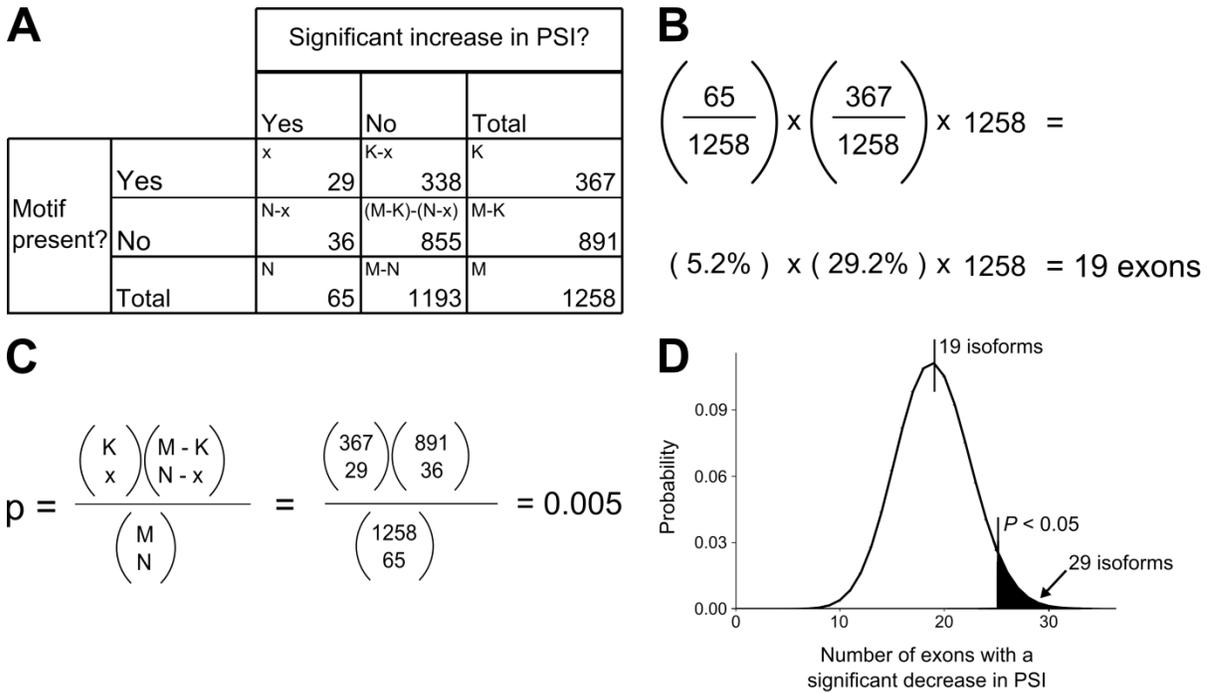


**Supplementary Figure 4. Pentamers enriched in single exon genes.** A first-order Markov model was used to calculate the expected frequency for each motif within single exon genes. Statistical significance was determined by comparing the expected frequency of each motif to the number of times it was observed (a two-sided binomial test). The  $p$ -value was then adjusted for multiple testing by the false discovery rate. The consensus splice sites from each intronic region were excluded from the input sequences. Dashed red line signifies statistical cut off of negative log10 (adjusted  $p$ -value) > 5, which corresponds to  $p < 0.05$ .

Supplementary Figure 5. Expanded motif enrichment heat maps



Supplementary Figure 6. Hypothesis testing and hypergeometric probability density function



**Supplementary Figure 6. Hypothesis testing and hypergeometric probability density functions.** **A**, Example of a contingency table used to test the null hypothesis that the number of exons upregulated with 15% or greater increase in PSI for both TAC and fetal compared to sham occurred independently.  $N$  represent the total number of significantly upregulated exons in both groups and  $K$  is equal to the number of distal upstream regions that contained the 'AAAAA' motif, while  $x$  is the number of exons that were significantly regulated in both groups and that had the 'AAAAA' motif in the distal upstream region. **B**, Since 5.2% of all exons a significant increase in PSI in both TAC and fetal hearts ( $N$ , 65 / 1,258) and 29.2% of all distal upstream regions contained the 'AAAAA' motif ( $K$ , 367 / 1,258), multiplying the two percentages times the total number of isoforms ( $M$ , 1,258) would predict 19 exons to have a significant increase in PSI ( $0.052 \times 0.292 \times 1,258$ ) if each exon was regulated independent of the 'AAAAA' motif. **C**, The probability of seeing any specific number of upregulated exons was calculated using a hypergeometric probability density function (calculations shown for 29 exons). The probability of seeing at least the degree of upregulated exons  $x$  was computed by summing probabilities for all values  $\geq x$  using MATLAB with the following command 'sum(hygepdf(x:N,M,K,N))'. **D**, Graphical representation of probability for each possible number of exons upregulated in TAC and fetal groups. The most likely number of shared isoforms is 19, as derived in B. The probability of observing increasing numbers of upregulated exons decreases to  $P = 0.05$  at 24 exons. More importantly, the sum of probabilities of observing all possible numbers of upregulated exons  $\geq 24$  is  $< 0.05$  (shown by shaded area), indicating that 24 is a significantly higher number of upregulated exons than would be expected if the inclusion of each exon was independently regulated.

## Appendix C. Selected Perl scripts

### *Code to construct DAGs and topology searches for alternative cassette exons*

```
#!/usr/bin/perl
#written on 1/18/13
#create splice graphs for genes of interest, identify alternative exons
#and the transcripts that contain or skip that exon by cufflinks TCONS ID
use strict; use warnings;

#add ~/lib dir to @INC
use lib '/home/ega2d/lib/';

#INPUTS REQUIRED: FILE OF GENE NAMES (ONE PER LINE) AND GTF
#specify input file name with --file space then name of file with significant genes

use Getopt::Long;
use Graph::Directed;
my ($file, $gtf);

GetOptions(
    'file=s'=>\$file,
    'gtf=s'=>\$gtf);

#initialize variables
my %genes;
my %splicedGenes;
my ($tcons, $geneName, $exonNumber, $chr, $start, $end, $strand);

open (GTF, "<$gtf") or die ("Could not open $gtf for input\n");
open (GENES, "<$file") or die ("Could not open $file for input\n");

#create hash of significant gene names
while (my $splicedGene = <GENES>) {
    chomp($splicedGene);
    $splicedGenes{$splicedGene}++;
} #close GENE read in

#read gtf, look for genes of interest
while (my $line = <GTF>){
    chomp($line);
    my @data = split /\t/, $line;
    ($chr,$start,$end,$strand) = ($data[0],$data[3],$data[4],$data[6]);
    # capture gene name from last field
    if ($data[8] =~ /gene_name."?([^\;]+)"/){
        $geneName = $1;
    } #gene name
    # capture tcons ID from last field
    if ($data[8] =~ /transcript_id."?([^\;]+)"/){
        $tcons = $1;
    } #tcons
    # capture exon number from last field
    if ($data[8] =~ /exon_number."?([^\;]+)"/){
        $exonNumber = $1;
    } #exon #number
    #store desired data in %genes
```

```

next unless $splicedGenes{$geneName};
$genes{$geneName}->{chromosome} = $chr;
$genes{$geneName}->{strand} = $strand;

#create hash of genes that contain a hash of transcripts that contain an array of arrays
#where the array of exons contains an array of start, end coords

push @{$genes{$geneName}->{transcripts}->{tcons}->{exons} }, [$start, $end];
} #close GTF read in
close GTF;
close GENES;

#iterate thru hash of genes

GENE: while (my ($gene, $geneData) = each %genes) {
#create variables about gene of interest
#scalars for chromosome and strand and a reference to a hash of transcripts (that contains an AoA of
#exons/start,end coords)

my ($chromosome, $strand, $transcripts) = @$geneData{qw(chromosome strand transcripts)};
#create new graphObject for each gene of interest
my $graph = Graph::Directed->new;
#iterate thru hash of transcripts for one gene
while ( my ($transcript, $transcriptData) = each %$transcripts ) {
#create an array of exons where each array contains a two-element array of start,
#end coords
my @exons = @{$transcriptData -> {exons} };
for (my $i=1; $i < @exons; $i++) {
#create variables needed for edges
my $prevEnd = $exons[$i-1][1];
my $start = $exons[$i][0];
my $end = $exons[$i][1];

#add egde to graph for previous intron
$graph->add_edge($prevEnd, $start);
$graph->set_edge_attribute($prevEnd, $start, $transcript, 1);

#add egde to graph for current exon
$graph->add_edge($start, $end);
$graph->set_edge_attribute($start, $end, $transcript, 1);
} #for adding exon to graph object
} #for hash of transcripts, graph object should be complete
my $cyclicCheck = $graph->is_cyclic();

#check for odd behaviors that would create edges going backwards
if ($cyclicCheck) {
next GENE;
} #for cyclicCheck

##initialize variables to determine type of AS##

my @out;
#array of vertices with out degree > 2
my @in; #array of vertices with in degree > 2

#number of paths between exon end and alternative splice site
my $altExon = 3; #number of paths between exon end and skipped exon
my $uniqueNodes;

```

```

##open output file##
open (ALTEX, ">>AlternativeExonsAttributes.txt") or die ("can't open altEx output\n");

##find alignment errors by looking for vertices that are within 10bp of one another##
my @allVertices = $graph->vertices();

#create array to check for alignment errors
my @sortedVertices = sort { $a <=> $b } @allVertices;
for (my $i=1; $i < @sortedVertices; $i++) {
    my $coordDiff = ($sortedVertices[$i]) - ($sortedVertices[$i-1]);
    if ($coordDiff < 10) {
        my $vertexToDelete = $sortedVertices[$i-1];
        $graph->delete_vertex($vertexToDelete);
    } #if loop to delete alignment error vertex
} #allVertices loop

##create arrays for vertices with in/out degrees > 1##
my @vertices = $graph->vertices();

#create array of all vertices in graph
NODE: foreach my $vertex (@vertices) {

#iterate thru each vertex to find those that have more than 1 out edge
my $outDegree = $graph->out_degree($vertex);
if ($outDegree > 1) {
    push (@out, $vertex);
} #for outDegree
my $inDegree = $graph->in_degree($vertex);
my $sinkVertexCheck = $graph->is_sink_vertex($vertex);
if ($inDegree > 1) {
    if ($sinkVertexCheck) {
        next NODE;
    } #skips sink vertices
} else{
    push (@in, $vertex);
} #write push $inVertex
} #for inDegree
} #finishes finding in and out vertices

##skip nodes that are terminal exon start sites (one edge removed from sink vertex)##
my $terminalExonStart;

#immediate predecessor of sink vertices

my @sinkVertices = $graph->sink_vertices();
foreach my $sinkVertex (@sinkVertices) {
    my @nextToSink = $graph->predecessors($sinkVertex);
    foreach my $secondToLast (@nextToSink) {
        $graph->set_vertex_attribute($secondToLast, $terminalExonStart, 1);
    } #for set attribute
} #for @sinkVertex

##iterate thru out vertices##
OUTVERTEX: foreach my $outVertex (@out) {

##check to see if out vertex is terminal exon start by attribute##
my $terminalExonStartCheck = $graph->has_vertex_attribute($outVertex, $terminalExonStart);
if ($terminalExonStartCheck) {
    next OUTVERTEX;
}
}

```

```

    }

##create subgraphs based on number of edges spanned by splicing event##
my $altExSubgraph = $graph->subgraph_by_radius($outVertex, $altExon);

##iterate thru in vertices to look for closure of splice graphs##
    foreach my $inVertex (@in) {

##look for alternative exons##
        my $altExMatch = $altExSubgraph->has_vertex($inVertex);
        if ($altExMatch) {
            my $shortestPath = $altExSubgraph->APSP_Floyd_Warshall();
            my $shortestPathLength = $shortestPath->path_length($outVertex, $inVertex);
            if ($shortestPathLength == 1) {

#skipping event should only have one edge between out and in vertices

##iterate thru successors to find path that starts w/ outVertex and 3 edges later ends w/ inVertex##
                my @altStarts = $altExSubgraph->successors($outVertex);
                foreach my $potentialAltStart (@altStarts) {
                    if ($potentialAltStart != $inVertex) {
                        my @altEnds = $altExSubgraph->successors($potentialAltStart);
                        foreach my $potentialAltEnd (@altEnds) {
                            my @dnStart = $altExSubgraph->successors($potentialAltEnd);

                            foreach my $potentialDnStart (@dnStart) {
                                if ($potentialDnStart == $inVertex) {
                                    my $skipTcons = $graph->get_edge_attributes($outVertex, $inVertex);
                                    my $inclTcons = $graph->get_edge_attributes($potentialAltStart,$potentialAltEnd);

my @skipList = keys %{$skipTcons};
my @inclList = keys %{$inclTcons};
print ALTEX $gene, "\t", $chromosome, "\t", $strand, "\t", $outVertex, "\t",
    $potentialAltStart, "\t", $potentialAltEnd, "\t", $inVertex, "\n";

print ALTEX $gene, " TCONS that skip exon ", "@skipList\n";
print ALTEX $gene, " TCONS that include exon ", "@inclList\n";
                                } #potentialDnStart
                            } #foreach DnStart
                        } #foreach $potentialAltEnd
                    } #if potentialAltStart
                } #foreach potentialAltStart
                print ALTEX $gene, "\t", $chromosome, "\t", $outVertex, "\t", $inVertex, "\t", $shortestPathLength, "\n";
                print ALTEX $gene, "\t", $chromosome, "\t", $altExSubgraph, "\n";
                print ALTEX $gene, "\t", $graph, "\n";
            } #closes shortest path loop
        } #closes alternative exon search
    } #inVertex
} #outvertex
} #for hash of genes
close ALTEX;

```

*Code to count observed pentamers and calculated expected frequencies using a first-order Markov model*

```
#!/usr/bin/perl
#written on 1/24/13
#read in sequence, count pentamers, create a first order Markov model,
#output observed and expected frequencies
use strict; use warnings;
use Getopt::Long;
no warnings 'uninitialized';
#REQUIRED INPUT: TAB-DELIMITED FASTA FILE WITH ADDITIONAL FIELD FOR GC %

#initialize variables
my $file;
my $output;
my @sequences;
my @seqLength;
my %unicounts;
my %dicounts;
my @seqPentCounts;
my @seqPents;

GetOptions(
    'output=s'=>\$output,
    'file=s'=>\$file);

open(SEQ, "<$file" or die "Cannot open sequence file\n";
#parse fasta file
while (my $line = <SEQ>) {
    chomp ($line);
    my @data = split /\t/, $line;
    my ($name, $gc, $fasta) = ($data[0], $data[1], $data[2]);
    push (@sequences, $fasta);
} #close SEQ read in
close SEQ;

print "@seqLength", "\n";

for my $sequence (@sequences) {
    while ($sequence =~ m/\G(.)(.)(.)(.)(.)/g) {
        $unicounts{$1}->{count}++;
        $dicounts{"$1$2"}->{count}++;
        pos($sequence) = pos($sequence) - 1; #moves search backward 1 base
    } #close while count learning loop
} #for each in @sequences

#calculate frequencies
my $basetotal = 0;
for my $base (keys %unicounts) {
    $basetotal += $unicounts{$base}->{count}; #calculates total length of sequence
} #close base total loop

for my $base (keys %unicounts) {
    $unicounts{$base}->{prob} = $unicounts{$base}->{count}/$basetotal; #for each bp creates probability
} #close uncount probability

#calculate dinucleotide probabilities
```

```

my $dinucleototal = 0;
for my $dibase (keys %dicounts) {
    $dinucleototal += $dicounts{$dibase}->{count};
} #close dinucleototal loop

for my $dibase (keys %dicounts) {
    $dicounts{$dibase}->{prob} = $dicounts{$dibase}->{count}/$dinucleototal;
}

#check outputs
my ($totalDinucleoProb, $totalSingleNucleoProb);
my $upperThreshold = 1.01;
my $lowerThreshold = 0.99;

for my $base (keys %unicounts) {
    $totalSingleNucleoProb += $unicounts{$base}->{prob};
}
for my $dibase (keys %dicounts) {
    $totalDinucleoProb += $dicounts{$dibase}->{prob};
}
if ($totalSingleNucleoProb < $lowerThreshold || $totalSingleNucleoProb > $upperThreshold) {
    print "ERROR in single nt probability calculations\n";
}
if ($totalDinucleoProb < $lowerThreshold || $totalDinucleoProb > $upperThreshold) {
    print "ERROR in dinucleotide probability calculations\n";
}

#enumerate the observed pentamers:
my %pentamers;
my $penttotal=0;

for my $sequence (@sequences) {
    PENT: while ($sequence =~ m/AG(.....)/g) {
        if ($1 =~ m/N/g) {
            next PENT;
        }
        else {
            $pentamers{$1}->{count}++;
            $penttotal++;
            pos($sequence) = pos($sequence) - 4;
        }
    } #close pentamer enumeration
} #for each sequence

#calculated expected pentamer frequencies
for my $pentamer (keys %pentamers) {
    my @bases = split("", $pentamer);
    my $first = $bases[0];
    my $second = $bases[1];
    my $third = $bases[2];
    my $fourth = $bases[3];
    my $last = $bases[4];
    my $pentProb = $unicounts{$first}->{prob};
    my $ZOprob = $pentProb*$unicounts{$second}->{prob}*$unicounts{$third}->{prob}*$unicounts{$fourth}->{prob}*$unicounts{$last}->{prob};

    for my $nextbase (@bases[1..4]) {
        my $pentProbDenominator;
        if(exists $dicounts{"$first$nextbase"}) {

```

```

        for my $nt (qw(A C G T)) {
            $pentProbDenominator += $dicounts{"$first$nt"}->{prob};
        }
        $pentProb *= ($dicounts{"$first$nextbase"}->{prob} / $pentProbDenominator);
        $first = $nextbase;
    } #close check for existence of dinucleo
else {
    print "$first$nextbase", " not defined\n";
    for my $nt (qw(A C G T)) {
        $pentProbDenominator += $dicounts{"$first$nt"}->{prob};
    }
    $pentProb *= ($dicounts{"$first$nextbase"}->{prob} / $pentProbDenominator);
    $first = $nextbase;
    } #else undefined
} #close next base loop
$pentamers{$pentamer}->{FOexp} = $pentProb;
$pentamers{$pentamer}->{ZOexp} = $ZOprob;

open(OUT, ">>$output") or die "Cannot open output file\n";

    if ($pentProb != 0) {
        print OUT "$pentamer\t", "$pentamers{$pentamer}->{count}\t",
"$pentamers{$pentamer}->{FOexp}\t",
"$pentamers{$pentamer}->{ZOexp}\t", "$penttotal\n";
    } #close if loop

} #close pentamer loop

close OUT;

sub genMer {
    my $ct = shift;
    my @return;
    if ($ct > 1) {
        my @mers = genMer($ct-1);
        for my $nt (qw(A C G T)) {
            push @return, map { $nt . $_ } @mers;
        }
    }
    else {
        @return = qw(A C G T);
    }
    return @return;
}

```

## **Chapter 5 - Future directions**

Our findings suggest a number of different follow-up studies to determine the specific mechanisms by which the RNA-binding proteins (RBPs) we identified regulate splicing, as well as to better understand the significance of changes in extracellular matrix gene expression we measured in response to changes in mechanics.

### **Role of polypyrimidine tract binding protein 1 mediates the re-expression of fetal splice variants seen in cardiac hypertrophy**

The role of alternative splicing in hypertrophy [172,173] and heart failure [118] is only beginning to be characterized. We believe that determining the mechanism that causes alternative splicing to revert back to fetal splicing patterns in response to pressure overload will improve our understanding of hypertrophy and possibly identify new therapeutic targets.

Our initial motif search suggests that only a few RNA-binding proteins (RBPs) might be mediating the re-expression of fetal splice variants in hypertrophy. One of the most significant motifs was the 'CUCUCU' motif that is recognized by the polypyrimidine tract binding (PTB) protein family of RBPs composed of two main proteins, PTBP1 and PTBP2. PTBP1 is widely expressed [248,249] whereas PTBP2 is only expressed in post-mitotic neurons [250]. Recently PTBP1 was shown to be developmentally regulated in the heart [251]. In fetal hearts, PTBP1 acts as a splicing repressor resulting in exon exclusion of several key cardiac genes. As the developing heart shifts towards a post-natal circulatory system, histone deacetylase 5 (HDAC5) decreases in the heart and this reduction causes decreased levels of the caspase inhibitor cellular FLICE-like inhibitory protein (cFLIP). As

the level of cFLIP decreases, the inhibition on caspases is released, allowing PTBP1 to be degraded by these caspases. In adulthood when there are lower levels of PTBP1, the exon inclusion increases for known targets of PTBP1 [251]. In hypertrophy, class II histone deacetylases (HDACs 5 and 9) are phosphorylated on multiple serine residues, exported from the nucleus, and bound by 14-3-3 proteins in the cytoplasm. This phosphorylation and subsequent sequestration of HDAC5 allows transcription factors to activate genes that drive the hypertrophic response [252].

Recently it was shown that PTBP1 levels increase in adult hearts in response to calcineurin overexpression [253]. The phosphatase calcineurin is activated by calcium signaling and is a critical mediator of cardiac hypertrophy. Calcineurin dephosphorylates the transcription factor NFAT, which is then translocated into the nucleus to activate pro-hypertrophic genes [38]. Further, when cultured cardiomyocytes were treated with the hypertrophic agonist phenylephrine, PTBP1 levels also increased [253].

We hypothesize that in hypertrophy, as hypertrophic signaling cascades including calcineurin increase, PTBP1 protein levels increase, and the splicing pattern seen in fetal heart development is re-expressed. This proposed mechanism is shown in cartoon format in Figure 1. It is also interesting to note that PTBP1 generally functions as a splicing repressor, i.e. causes more exon exclusion, and we identified more exon exclusion events in our TAC and fetal groups compared to the sham-operated hearts. This is another piece of evidence suggesting that PTBP1 may be one of the key mediators of the fetal splicing program. To further explore the relationship between PTBP1 in hypertrophy and the re-expression of fetal splice variants in hypertrophy we propose the following experiments:

**Hypothesis 1: PTBP1 protein levels increase in pressure overload hypertrophy.**

Increased PTBP1 activity could be caused by one or more mechanisms. As described earlier, probing the relationship between calcineurin signaling and PTBP1 in hypertrophy is one of our top priorities. First, we would propose to measure protein localization and quantities of PTBP1 using an *in vivo* model of hypertrophy. In terms of alternative mechanisms for how PTBP1 activity may be regulated, it is also important to note that like many other RBPs, PTBP1 itself is alternatively spliced, specifically during heart development [88,254]. This isoform shift may correspond to the inclusion or exclusion of a binding site for Raver1. Raver1 is a protein that has been shown to be a co-repressor with PTBP1 of alternative splicing regulation [255] and is also shuttled between the nucleus and the Z-line of sarcomeres [256]. While our main focus for this aim will be elucidating the relationship between hypertrophic signaling and PTBP1, we realize that there are other ways that PTBP1 activity could be increased in hypertrophy.

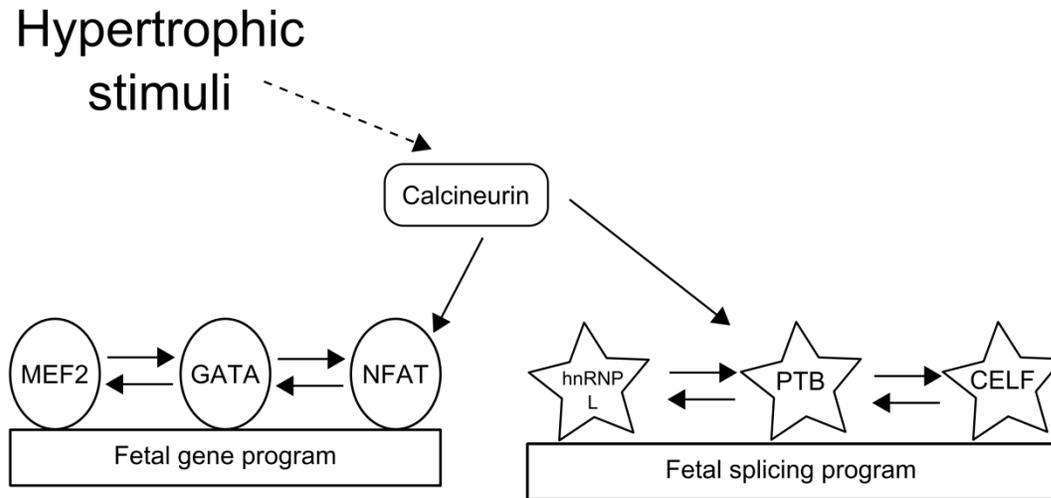
**Hypothesis 2: PTBP1 is necessary for the re-expression of fetal splice variants during hypertrophy.**

As a first step to study alternative splicing in hypertrophy, it is important to establish if the re-expression of fetal splice variants occurs *in vitro*. We would treat adult cardiomyocytes with the hypertrophic agonist phenylephrine and measure alternative splicing events we identified in Chapter 4 by PCR. We would then use adenovirus to knockdown PTBP1. If the splicing events we detected in hypertrophy and in the *in vitro* system do not occur with PTBP1 knockdown, it would strongly suggest that PTBP1 is mediating the splicing changes we saw. As further proof that PTBP1 mediates fetal splicing patterns in hypertrophy, we could also over-express PTBP1 in untreated adult cardiomyocytes and measure alternative splicing to determine if PTBP1 alone is sufficient to

cause the changes in isoform expression that we observed in pressure overload hypertrophy.

**Hypothesis 3: The re-expression of fetal splice variants is sufficient for hypertrophy.** It remains unknown if alternative splicing changes are a result or a cause of various physiological transitions such as hypertrophy. In order to determine if the alternative splicing we observed in TAC is sufficient to cause hypertrophy, we propose to study this process *in vivo* by creating a cardiac-specific inducible overexpression system for PTBP1 in a mouse model of hypertrophy. We would determine the cardiac phenotype of these mice with and without pressure overload to see if overexpression of PTBP1 causes hypertrophy or a larger degree of hypertrophy when subjected to pressure overload than wild-type controls. The final piece of evidence that PTBP1 and the re-expression of fetal splicing patterns are sufficient to drive pressure-overload hypertrophy would be to dissect changes in alternative splicing from the transcriptional program in hypertrophy. We would do this by crossing the PTBP1 transgenic mouse with MEF2D-null mice. MEF2D-null mice are resistant to cardiac hypertrophy, lack cardiac fibrosis, and most importantly, are unable to upregulate components of the fetal gene program in response to pressure overload. These double transgenic mice would be lacking the ability to activate the fetal gene program, but if we saw hypertrophy would suggest that alternative splicing is capable of causing hypertrophy [257].

Figure 1. Proposed mechanism of PTB regulation and re-expression of fetal splice variants in hypertrophy



**Figure 1. Proposed mechanism of PTB regulation and re-expression of fetal splice variants in hypertrophy.** Hormonal and mechanical hypertrophic stimuli cause signaling cascades via calcineurin that activate the transcription factor NFAT. NFAT, together with the additional transcription factor families GATA and MEF2 promotes expression of the fetal gene program. We hypothesize that in hypertrophy, PTB levels increase in response to calcineurin signaling and that along with other RNA-binding proteins, CELF and hnRNP L mediate the re-expression of fetal splice variants in hypertrophy.

## **Computational approaches to determine tissue-specific elements that regulate cardiac hypertrophy and implications for public health**

There are a few additional bioinformatics analyses that could be done with our RNAseq data. Our motif search looked for over-represented pentamers flanking the alternative cassette exons. Other groups have shown that the regulatory *cis*-elements that determine alternative splicing are far more complex than short motifs, therefore I would recommend forming a collaboration with Brendan Frey's group to use their more complex computational model to identify additional regulatory elements that could predict alternative splicing patterns seen in our data set. They use machine-learning techniques to prioritize which features out of 1,000 in their database most accurately predict splicing events observed in a given set of alternative exons [221]. As has been mentioned before, tissue-specificity is an important determinant of alternative splicing, and combining our heart-specific RNAseq data with their more thorough analysis may allow us to identify other candidate RBPs that are involved in coordinating alternative splicing in heart development and hypertrophy. We could also leverage our other groups of alternative exons (null, fetal-specific, TAC-specific) to determine if the regulatory elements that the model predicts are unique to hypertrophy.

While our preliminary analysis focused on alternative splicing of cassette exons, there are many other forms of alternative splicing. It would be interesting to analyze the intronic sequences flanking alternative splice sites and intron retention events to determine if the same RBPs that we identified in Chapter 4 are also involved in these splicing events or if other RBPs may have a role in hypertrophy-associated alternative splicing.

We would also like to determine the functional consequences of alternative splicing seen in hypertrophy. Do changes in alternative splicing result in changes in the amino acid sequence of proteins? Recently it has been shown that microRNAs play a significant role in

the response to pressure overload [258,259]. Do changes in alternative polyadenylation sites cause differences in microRNA binding sites? And lastly, does alternative splicing contribute to post-transcriptional gene regulation via insertion of premature stop codons?

In terms of another large-scale bioinformatics analysis, it would be interesting to utilize the wealth of data generated by genome wide association studies to determine if the regions that contain RBP recognition motifs are associated with differences in cardiovascular disease and specifically with left ventricular mass. There is a growing body of knowledge that mutations in splicing regulatory sequences play a distinct role in disease susceptibility [260,261].

### **The role of physical coordination of RNA processing in hypertrophy**

Several of the motifs identified by our search in Chapter 4 correspond to RBPs that have multiple functions aside from alternative splicing. These RBPs include CELF, hnRNP L, PTB, and possibly the Mex3 and poly-adenylate binding protein families of RBPs that potentially recognize A-rich motifs. In other systems, cell stressors such as heat or hypoxia have been shown to change the physical organization of the nucleus [262]. During the development of cardiac hypertrophy in response to pressure overload, several heat shock proteins increase in expression similar to other cell stress models [150], suggesting that cardiac hypertrophy may prompt stress granule formation, but the physical changes in hypertrophic nuclear organization remain unknown. These RNA stress granules are located within the cytoplasm and nucleus and function to delay and sequester mRNA at different processing steps including alternative splicing [263] and translation [264]. The physical interactions of these RBPs with their target mRNA sequences could represent components of a hypertrophic post-transcriptional operon. Using immunofluorescence to investigate the

nuclear localization of our RBPs or interest and co-localization with other known members of RNA stress granules could shed light on the physical coordination of alternative splicing in a physiological transition. Specifically, I would propose to use antibodies against the RNA-binding protein TIA-1 to mark cytosolic stress granules [265] associated with environmental stressors [266], and antibodies against hnRNP A1 interacting protein HAP to mark nuclear stress granules [267].

### **Changes in extracellular matrix and cell-cell contacts in response to mechanical input and potential implications on cell shape differences in concentric and eccentric cardiac hypertrophy**

One of the most surprising findings from our work exploring gene expression changes in response to decreased shortening in Chapter 2 was the number of extracellular matrix genes identified. With extensive data about signaling networks and transcriptional responses to changes in cardiomyocyte mechanics, there is less evidence about how the extracellular matrix changes in response to mechanical cues in the heart.

In Chapter 2 we proposed that in order to undergo cell shape changes in hypertrophy, cardiomyocytes must remodel their cell-cell and cell-ECM connections. It would be important to establish a few important facts about this hypothesis. The ECM is an incredibly complex and diverse set of proteins and glycosylated proteins that provide a scaffold for cells and attachments between cells [268,269]. Our microarrays measured gene expression from all cells present in the papillary muscles, and some of the gene expression changes for ECM genes may be from fibroblasts rather than cardiomyocytes. The fact that fibroblasts would be re-organizing the ECM in response to changes in mechanics is logical, given what our group and others have shown in the past [127,270,271]. I would propose doing preliminary experiments looking at ultrastructural changes in the extracellular matrix in

response to changes in mechanics utilizing our muscle bath experimental system to determine if there are visible differences between normal and pressure overload-like mechanics at early time points. The muscle bath system would allow us to tightly control the changes in mechanics and then quickly preserve the tissue under these perturbed mechanics as opposed to *in vivo* experiments that might be confounded during the process of tissue harvesting.

When we compared the gene expression changes in response to decreased shortening to an analogous *in vivo* model of pressure overload, we expected that we would see expression changes in sarcomeric genes. We hypothesized that in order to remodel a cardiomyocyte, the cell must rearrange and synthesize more elements of the contractile apparatus. However, what we found instead was that many of the significant gene expression changes that occurred in both our decreased shortening group and in the TAC data set included genes that are components of the extracellular matrix. By culturing intact myocardium, we were also able to include the contribution of fibroblast gene expression at early time points of mechanical changes. This provides an intriguing new hypothesis for cardiomyocyte remodeling: perhaps the extracellular connections between cardiomyocytes are integral for determining cell shape. One could imagine a mechanism when cardiomyocytes undergo eccentric hypertrophy, where individual cells become longer in response to loss of lateral connections on the long axis of the myocyte. In contrast, concentric hypertrophy could be the result of reinforced lateral connections that predispose the cardiomyocyte to be pulled by other lateral connections and become wider. Future experiments into how the cardiac extracellular matrix and basement membrane respond to mechanical cues and how they influence cardiomyocyte shape could be incredibly important for our understanding and ability to prevent adverse cell shape changes.

## Bibliography

- [1] Frey N, Katus HA, Olson EN, Hill JA. Hypertrophy of the heart: a new therapeutic target? *Circulation* 2004; 109: 1580–9.
- [2] Frey N, Olson EN. Cardiac hypertrophy: the good, the bad, and the ugly. *Annu Rev Physiol* 2003; 65: 45–79.
- [3] Hoshijima M. Mechanical stress-strain sensors embedded in cardiac cytoskeleton: Z disk, titin, and associated structures. *Am J Physiol-Heart C* 2006; 290: H1313–25.
- [4] Guterl K a, Haggart CR, Janssen PM, Holmes JW. Isometric contraction induces rapid myocyte remodeling in cultured rat right ventricular papillary muscles. *Am J Physiol-Heart C* 2007; 293: H3707–12.
- [5] Holmes JW. Candidate mechanical stimuli for hypertrophy during volume overload. *J Appl Physiol* 2004; 97: 1453–60.
- [6] Cooper TA. Alternative splicing regulation impacts heart development. *Cell* 2005; 120: 1–2.
- [7] Kalsotra A, Xiao X, Ward AJ, Castle JC, Johnson JM, Burge CB, et al. A postnatal switch of CELF and MBNL proteins reprograms alternative splicing in the developing heart. *Proc Natl Acad Sci USA* 2008; 105: 20333–20338.
- [8] Kalsotra A, Cooper TA. Functional consequences of developmentally regulated alternative splicing. *Nat Rev Genet* 2011; 12: 715–729.
- [9] Dorn GW. The fuzzy logic of physiological cardiac hypertrophy. *Hypertension* 2007; 49: 962–70.
- [10] Gottdiener JS, Arnold AM, Aurigemma GP, Polak JF, Tracy RP, Kitzman DW, et al. Predictors of congestive heart failure in the elderly: the Cardiovascular Health Study. *J Am Coll Cardiol* 2000; 35: 1628–37.
- [11] Ling LH, Enriquez-Sarano M, Seward JB, Tajik AJ, Schaff H V, Bailey KR, et al. Clinical outcome of mitral regurgitation due to flail leaflet. *New Engl J Med* 1996; 335: 1417–23.
- [12] Corvisart J. *An Essay on the Organic Diseases and Lesions of the Heart and Great Vessels.*, Bradford & Read, Boston, 1812;
- [13] 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: e6–e245.

- [14] Li C, Balluz LS, Ford ES, Okoro CA, Zhao G, Pierannunzi C. A comparison of prevalence estimates for selected health indicators and chronic diseases or conditions from the Behavioral Risk Factor Surveillance System, the National Health Interview Survey, and the National Health and Nutrition Examination Survey, 200. *Prev Med* 2012; 54: 381–7.
- [15] Grossman W, Jones D, McLaurin LP. Wall stress and patterns of hypertrophy in the human left ventricle. *J Clin Invest* 1975; 56: 56–64.
- [16] Badeer HS. The stimulus to hypertrophy of the myocardium. *Circulation* 1964; 30: 128–36.
- [17] Dahlöf B, Pennert K, Hansson L. Reversal of left ventricular hypertrophy in hypertensive patients. A metaanalysis of 109 treatment studies. *Am J Hypertens* 1992; 5: 95–110.
- [18] Klingbeil AU, Schneider M, Martus P, Messerli FH, Schmieder RE. A meta-analysis of the effects of treatment on left ventricular mass in essential hypertension. *Am J Med* 2003; 115: 41–6.
- [19] Schmieder RE, Martus P, Klingbeil A. Reversal of left ventricular hypertrophy in essential hypertension. A meta-analysis of randomized double-blind studies. *JAMA* 1996; 275: 1507–13.
- [20] Russell B, Motlagh D, Ashley WW. Form follows function: how muscle shape is regulated by work. *J Appl Physiol* 2000; 88: 1127–32.
- [21] Rajabi M, Kassiotis C, Razeghi P, Taegtmeyer H. Return to the fetal gene program protects the stressed heart: a strong hypothesis. *Heart Fail Rev* 2007; 12: 331–343.
- [22] Thattaliyath BD, Livi CB, Steinhilber ME, Toney GM, Firulli AB. HAND1 and HAND2 are expressed in the adult-rodent heart and are modulated during cardiac hypertrophy. *Biochem Biophys Res Commun* 2002; 297: 870–5.
- [23] Bär H, Kreuzer J, Cojoc A, Jahn L. Upregulation of embryonic transcription factors in right ventricular hypertrophy. *Basic Res Cardiol* 2003; 98: 285–94.
- [24] Hautala N, Tokola H, Luodonpää M, Puhakka J, Romppanen H, Vuolteenaho O, et al. Pressure overload increases GATA4 binding activity via endothelin-1. *Circulation* 2001; 103: 730–5.
- [25] Herzig TC, Jobe SM, Aoki H, Molkentin JD, Cowley AW, Izumo S, et al. Angiotensin II type1a receptor gene expression in the heart: AP-1 and GATA-4 participate in the response to pressure overload. *Proc Natl Acad Sci USA* 1997; 94: 7543–8.

- [26] Liang Q, De Windt LJ, Witt SA, Kimball TR, Markham BE, Molkenin JD. The transcription factors GATA4 and GATA6 regulate cardiomyocyte hypertrophy in vitro and in vivo. *J Biol Chem* 2001; 276: 30245–53.
- [27] Thompson JT, Rackley MS, O'Brien TX. Upregulation of the cardiac homeobox gene Nkx2-5 (CSX) in feline right ventricular pressure overload. *Am J Physiol* 1998; 274: H1569–73.
- [28] Saadane N, Alpert L, Chalifour LE. Expression of immediate early genes, GATA-4, and Nkx-2.5 in adrenergic-induced cardiac hypertrophy and during regression in adult mice. *Brit J Pharmacol* 1999; 127: 1165–76.
- [29] Takimoto E, Mizuno T, Terasaki F, Shimoyama M, Honda H, Shiojima I, et al. Up-regulation of natriuretic peptides in the ventricle of Csx/Nkx2-5 transgenic mice. *Biochem Bioph Res Co* 2000; 270: 1074–9.
- [30] Molkenin JD, Markham BE. Myocyte-specific enhancer-binding factor (MEF-2) regulates alpha-cardiac myosin heavy chain gene expression in vitro and in vivo. *J Biol Chem* 1993; 268: 19512–20.
- [31] Xu J, Gong NL, Bodi I, Aronow BJ, Backx PH, Molkenin JD. Myocyte enhancer factors 2A and 2C induce dilated cardiomyopathy in transgenic mice. *J Biol Chem* 2006; 281: 9152–62.
- [32] Maillet M, Van Berlo JH, Molkenin JD. Molecular basis of physiological heart growth: fundamental concepts and new players. *Nat Rev Mol Cell Biol* 2013; 14: 38–48.
- [33] Izumo S, Nadal-Ginard B, Mahdavi V. Protooncogene induction and reprogramming of cardiac gene expression produced by pressure overload. *Proc Natl Acad Sci USA* 1988; 85: 339–343.
- [34] Uozumi H, Hiroi Y, Zou Y, Takimoto E, Toko H, Niu P, et al. gp130 plays a critical role in pressure overload-induced cardiac hypertrophy. *J Biol Chem* 2001; 276: 23115–9.
- [35] Kawano S, Kubota T, Monden Y, Kawamura N, Tsutsui H, Takeshita A, et al. Blockade of NF-kappaB ameliorates myocardial hypertrophy in response to chronic infusion of angiotensin II. *Cardiovasc Res* 2005; 67: 689–98.
- [36] Zhang X, Azhar G, Chai J, Sheridan P, Nagano K, Brown T, et al. Cardiomyopathy in transgenic mice with cardiac-specific overexpression of serum response factor. *Am J Physiol-Heart C* 2001; 280: H1782–92.
- [37] Wilkins BJ, Molkenin JD. Calcium-calcineurin signaling in the regulation of cardiac hypertrophy. *Biochem Bioph Res Co* 2004; 322: 1178–91.

- [38] Molkentin JD, Lu JR, Antos CL, Markham B, Richardson J, Robbins J, et al. A calcineurin-dependent transcriptional pathway for cardiac hypertrophy. *Cell* 1998; 93: 215–28.
- [39] Hang CT, Yang J, Han P, Cheng H-L, Shang C, Ashley E, et al. Chromatin regulation by Brg1 underlies heart muscle development and disease. *Nature* 2010; 466: 62–67.
- [40] Yamazaki T, Walchli S, Fujita T, Ryser S, Hoshijima M, Schlegel W, et al. Splice variants of enigma homolog, differentially expressed during heart development, promote or prevent hypertrophy. *Cardiovasc Res* 2010; 86: 374–382.
- [41] Rooij E Van, Olson EN. Science in medicine MicroRNAs : powerful new regulators of heart disease and provocative therapeutic targets. *Proc Natl Acad Sci USA* 2007; 117:
- [42] Berget SM, Moore C, Sharp PA. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci USA* 1977; 74: 3171–5.
- [43] Wang GS, Cooper TA. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet* 2007; 8: 749–761.
- [44] Maniatis T, Tasic B. Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature* 2002; 418: 236–243.
- [45] Black DL. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* 2003; 72: 291–336.
- [46] Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* 2008; 40: 1413–1415.
- [47] Blencowe BJ. Alternative splicing: new insights from global analyses. *Cell* 2006; 126: 37–47.
- [48] Lareau LF, Green RE, Bhatnagar RS, Brenner SE. The evolving roles of alternative splicing. *Curr Opin Struct Biol* 2004; 14: 273–282.
- [49] Lee C, Roy M. Analysis of alternative splicing with microarrays: successes and challenges. *Genome Biol* 2004; 5: 231.
- [50] Rosenfeld MG, Mermod JJ, Amara SG, Swanson LW, Sawchenko PE, Rivier J, et al. Production of a novel neuropeptide encoded by the calcitonin gene via tissue-specific RNA processing. *Nature* n.d. 304: 129–35.
- [51] Brain SD, Williams TJ, Tippins JR, Morris HR, MacIntyre I. Calcitonin gene-related peptide is a potent vasodilator. *Nature* n.d. 313: 54–6.

- [52] Durham PL. Calcitonin gene-related peptide (CGRP) and migraine. *Headache* 2006; 46 Suppl 1: S3–8.
- [53] Forch P, Valcarcel J. Splicing regulation in *Drosophila* sex determination. *Prog Mol Subcell Biol* 2003; 31: 127–151.
- [54] Dulac C. Sex and the single splice. *Cell* 2005; 121: 664–666.
- [55] Irimia M, Blencowe BJ. Alternative splicing: decoding an expansive regulatory layer. *Curr Opin Cell Biol* 2012; 24: 323–32.
- [56] Irimia M, Rukov JL, Roy SW, Vinther J, Garcia-Fernandez J. Quantitative regulation of alternative splicing in evolution and development. *Bioessays* 2009; 31: 40–50.
- [57] Gracheva EO, Cordero-Morales JF, González-Carcacía JA, Ingolia NT, Manno C, Aranguren CI, et al. Ganglion-specific splicing of TRPV1 underlies infrared sensation in vampire bats. *Nature* 2011; 476: 88–91.
- [58] Kornblihtt AR, Schor IE, Alló M, Dujardin G, Petrillo E, Muñoz MJ. Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat Rev Mol Cell Biol* 2013; 14: 153–65.
- [59] Latif S, Masino A, Garry DJ. Transcriptional pathways direct cardiac development and regeneration. *Trends Cardiovasc Med* 2006; 16: 234–240.
- [60] De VS, Black BL. Transcriptional control of endothelial cell development. *Dev Cell* 2009; 16: 180–195.
- [61] Gill PS, Rosenblum ND. Control of murine kidney development by sonic hedgehog and its GLI effectors. *Cell Cycle* 2006; 5: 1426–1430.
- [62] Puelles E. Genetic control of basal midbrain development. *J Neurosci Res* 2007; 85: 3530–3534.
- [63] Revil T, Gaffney D, Dias C, Majewski J, Jerome-Majewska LA. Alternative splicing is frequent during early embryonic development in mouse. *BMC Genomics* 2010; 11: 399.
- [64] Bland CS, Wang ET, Vu A, David MP, Castle JC, Johnson JM, et al. Global regulation of alternative splicing during myogenic differentiation. *Nucleic Acids Res* 2010; 38: 7651–7664.
- [65] Wang J, Takagaki Y, Manley JL. Targeted disruption of an essential vertebrate gene: ASF/SF2 is required for cell viability. *Genes Dev* 1996; 10: 2588–2599.
- [66] Jumaa H, Wei G, Nielsen PJ. Blastocyst formation is blocked in mouse embryos lacking the splicing factor SRp20. *Curr Biol* 1999; 9: 899–902.

- [67] Wang HY, Xu X, Ding JH, Bermingham Jr. JR, Fu XD. SC35 plays a role in T cell development and alternative splicing of CD45. *Mol Cell* 2001; 7: 331–342.
- [68] Shibayama M, Ohno S, Osaka T, Sakamoto R, Tokunaga A, Nakatake Y, et al. Polypyrimidine tract-binding protein is essential for early mouse development and embryonic stem cell proliferation. *FEBS J* 2009; 276: 6658–6668.
- [69] Graziotto JJ, Inglehearn CF, Pack MA, Pierce EA. Decreased levels of the RNA splicing factor Prpf3 in mice and zebrafish do not cause photoreceptor degeneration. *Invest Ophthalmol Vis Sci* 2008; 49: 3830–3838.
- [70] Feng Y, Valley MT, Lazar J, Yang AL, Bronson RT, Firestein S, et al. SRp38 regulates alternative splicing and is required for Ca<sup>2+</sup> handling in the embryonic heart. *Dev Cell* 2009; 16: 528–538.
- [71] Kanadia RN, Johnstone KA, Mankodi A, Lungu C, Thornton CA, Esson D, et al. A muscleblind knockout model for myotonic dystrophy. *Science* 2003; 302: 1978–1980.
- [72] Ladd AN, Taffet G, Hartley C, Kearney DL, Cooper TA. Cardiac tissue-specific repression of CELF activity disrupts alternative splicing and causes cardiomyopathy. *Mol Cell Biol* 2005; 25: 6267–6278.
- [73] Ruggiu M, Herbst R, Kim N, Jevsek M, Fak JJ, Mann MA, et al. Rescuing Z<sup>+</sup> agrin splicing in Nova null mice restores synapse formation and unmask a physiologic defect in motor neuron firing. *Proc Natl Acad Sci USA* 2009; 106: 3513–3518.
- [74] Ip JY, Tong A, Pan Q, Topp JD, Blencowe BJ, Lynch KW. Global analysis of alternative splicing during T-cell activation. *RNA* 2007; 13: 563–572.
- [75] Keene JD, Lager PJ. Post-transcriptional operons and regulons co-ordinating gene expression. *Chromosome Res* 2005; 13: 327–337.
- [76] Keene JD. RNA regulons: coordination of post-transcriptional events. *Nat Rev Genet* 2007; 8: 533–543.
- [77] Matkovich SJ, Zhang Y, Van Booven DJ, Dorn GW. Deep mRNA sequencing for in vivo functional analysis of cardiac transcriptional regulators: application to Galphaq. *Circ Res* 2010; 106: 1459–1467.
- [78] Sanford JR, Wang X, Mort M, Vanduy N, Cooper DN, Mooney SD, et al. Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. *Genome Res* 2009; 19: 381–394.
- [79] Wei B, Jin JP. Troponin T isoforms and posttranscriptional modifications: Evolution, regulation and function. *Arch Biochem Biophys* 2010;

- [80] Gomes A V, Guzman G, Zhao J, Potter JD. Cardiac troponin T isoforms affect the Ca<sup>2+</sup> sensitivity and inhibition of force development. Insights into the role of troponin T isoforms in the heart. *J Biol Chem* 2002; 277: 35341–35349.
- [81] McAuliffe JJ, Gao LZ, Solaro RJ. Changes in myofibrillar activation and troponin C Ca<sup>2+</sup> binding associated with troponin T isoform switching in developing rabbit heart. *Circ Res* 1990; 66: 1204–1216.
- [82] Nassar R, Malouf NN, Mao L, Rockman HA, Oakeley AE, Frye JR, et al. cTnT1, a cardiac troponin T isoform, decreases myofilament tension and affects the left ventricular pressure waveform. *Am J Physiol-Cell Ph* 2005; 288: H1147–H1156.
- [83] Cooper TA, Ordahl CP. A single cardiac troponin T gene generates embryonic and adult isoforms via developmentally regulated alternate splicing. *J Biol Chem* 1985; 260: 11140–11148.
- [84] Jin JP, Lin JJ. Isolation and characterization of cDNA clones encoding embryonic and adult isoforms of rat cardiac troponin T. *J Biol Chem* 1989; 264: 14471–14477.
- [85] Anderson PA, Moore GE, Nassar RN. Developmental changes in the expression of rabbit left ventricular troponin T. *Circ Res* 1988; 63: 742–747.
- [86] Townsend PJ, Barton PJ, Yacoub MH, Farza H. Molecular cloning of human cardiac troponin T isoforms: expression in developing and failing heart. *J Mol Cell Cardiol* 1995; 27: 2223–2236.
- [87] Charlet B, Logan P, Singh G, Cooper TA. Dynamic antagonism between ETR-3 and PTB regulates cell type-specific alternative splicing. *Mol Cell* 2002; 9: 649–658.
- [88] Ladd AN, Stenberg MG, Swanson MS, Cooper TA. Dynamic balance between activation and repression regulates pre-mRNA alternative splicing during heart development. *Dev Dyn* 2005; 233: 783–793.
- [89] Kalsotra A, Wang K, Li PF, Cooper TA. MicroRNAs coordinate an alternative splicing network during mouse postnatal heart development. *Genes Dev* 2010; 24: 653–658.
- [90] Xu X, Yang D, Ding JH, Wang W, Chu PH, Dalton ND, et al. ASF/SF2-regulated CaMKII $\delta$  alternative splicing temporally reprograms excitation-contraction coupling in cardiac muscle. *Cell* 2005; 120: 59–72.
- [91] Ding J-H, Xu X, Yang D, Chu P-H, Dalton ND, Ye Z, et al. Dilated cardiomyopathy caused by tissue-specific ablation of SC35 in the heart. *EMBO J* 2004; 23: 885–896.
- [92] Huang Y, Steitz JA. Splicing factors SRp20 and 9G8 promote the nucleocytoplasmic export of mRNA. *Mol Cell* 2001; 7: 899–905.

- [93] Huang Y, Gattoni R, Stevenin J, Steitz JA. SR splicing factors serve as adapter proteins for TAP-dependent mRNA export. *Mol Cell* 2003; 11: 837–843.
- [94] Lai MC, Tarn WY. Hypophosphorylated ASF/SF2 binds TAP and is present in messenger ribonucleoproteins. *J Biol Chem* 2004; 279: 31745–31749.
- [95] Tarn W-Y. Cellular signals modulate alternative splicing. *J Biomed Sci* 2007; 14: 517–22.
- [96] Lemaire R, Prasad J, Kashima T, Gustafson J, Manley JL, Lafyatis R. Stability of a PKCI-1-related mRNA is controlled by the splicing factor ASF/SF2: a novel function for SR proteins. *Genes Dev* 2002; 16: 594–607.
- [97] Sanford JR, Gray NK, Beckmann K, Caceres JF. A novel role for shuttling SR proteins in mRNA translation. *Genes Dev* 2004; 18: 755–768.
- [98] Labeit S, Kolmerer B, Linke WA. The giant protein titin. Emerging roles in physiology and pathophysiology. *Circ Res* 1997; 80: 290–294.
- [99] Gregorio CC, Perry CN, McElhinny AS. Functional properties of the titin/connectin-associated proteins, the muscle-specific RING finger proteins (MURFs), in striated muscle. *J Muscle Res Cell M* 2005; 26: 389–400.
- [100] Kontogianni-Konstantopoulos A, Ackermann MA, Bowman AL, Yap S V, Bloch RJ. Muscle giants: molecular scaffolds in sarcomerogenesis. *Physiol Rev* 2009; 89: 1217–1267.
- [101] Hoshijima M. Mechanical stress-strain sensors embedded in cardiac cytoskeleton: Z disk, titin, and associated structures. *Am J Physiol-Cell Ph* 2006; 290: H1313–H1325.
- [102] Sjoström M, Squire JM. Cryo-ultramicrotomy and myofibrillar fine structure: a review. *J Microsc* 1977; 111: 239–278.
- [103] Edman AC, Squire JM, Sjoström M. Fine structure of the A-band in cryo-sections. Diversity of M-band structure in chicken breast muscle. *J Ultrastruct Mol Struct Res* 1988; 100: 1–12.
- [104] Podlubnaya ZA, Shpagina MD, Lednev V V. Manifestation of the stripes of minor proteins location in A-bands of rabbit cardiac myofibrils. *J Mol Biol* 1989; 210: 655–658.
- [105] Pask HT, Jones KL, Luther PK, Squire JM. M-band structure, M-bridge interactions and contraction speed in vertebrate cardiac muscles. *J Muscle Res Cell Motil* 1994; 15: 633–645.
- [106] Carlsson E, Thornell LE. Diversification of the myofibrillar M-band in rat skeletal muscle during postnatal development. *Cell Tissue Res* 1987; 248: 169–180.

- [107] Gautel M, Goulding D, Bullard B, Weber K, Furst DO. The central Z-disk region of titin is assembled from a novel repeat in variable copy numbers. *J Cell Sci* 1996; 109 ( Pt 1: 2747–2754.
- [108] Opitz CA, Leake MC, Makarenko I, Benes V, Linke WA. Developmentally regulated switching of titin size alters myofibrillar stiffness in the perinatal heart. *Circ Res* 2004; 94: 967–975.
- [109] Freiburg A, Trombitas K, Hell W, Cazorla O, Fougerousse F, Centner T, et al. Series of exon-skipping events in the elastic spring region of titin as the structural basis for myofibrillar elastic diversity. *Circ Res* 2000; 86: 1114–1121.
- [110] Lahmers S, Wu Y, Call DR, Labeit S, Granzier H. Developmental control of titin isoform expression and passive stiffness in fetal and neonatal myocardium. *Circ Res* 2004; 94: 505–13.
- [111] Greaser ML, Krzesinski PR, Warren CM, Kirkpatrick B, Campbell KS, Moss RL. Developmental changes in rat cardiac titin/connectin: transitions in normal animals and in mutants with a delayed pattern of isoform transition. *J Muscle Res Cell M* 2005; 26: 325–32.
- [112] Guo W, Bharmal SJ, Esbona K, Greaser ML. Titin diversity--alternative splicing gone wild. *J Biomed Biotechnol* 2010; 2010: 753675.
- [113] VanHeerebeek L, Borbely A, Niessen HW, Bronzwaer JG, Van V D, Stienen GJ, et al. Myocardial structure and function differ in systolic and diastolic heart failure. *Circulation* 2006; 113: 1966–1973.
- [114] Neagoe C, Kulke M, Del Monte F, Gwathmey JK, De Tombe PP, Hajjar RJ, et al. Titin isoform switch in ischemic human heart disease. *Circulation* 2002; 106: 1333–41.
- [115] Greaser ML, Warren CM, Esbona K, Guo W, Duan Y, Parrish AM, et al. Mutation that dramatically alters rat titin isoform expression and cardiomyocyte passive tension. *J Mol Cell Cardiol* 2008; 44: 983–991.
- [116] Guo W, Schafer S, Greaser ML, Radke MH, Liss M, Govindarajan T, et al. RBM20, a gene for hereditary cardiomyopathy, regulates titin splicing. *Nat Med* 2012; 18: 766–773.
- [117] Li S, Guo W, Dewey CN, Greaser ML. Rbm20 regulates titin alternative splicing as a splicing repressor. *Nucleic Acids Res* 2013; 41: 2659–72.
- [118] Kong SW, Hu YW, Ho JW, Ikeda S, Polster S, John R, et al. Heart failure-associated changes in RNA splicing of sarcomere genes. *Circ Cardiovasc Genet* 2010; 3: 138–146.

- [119] Lee J-H, 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;
- [120] Sierralta J, Mendoza C. PDZ-containing proteins: alternative splicing as a source of functional diversity. *Brain Res Rev* 2004; 47: 105–115.
- [121] Kadrmaz JL, Beckerle MC. The LIM domain: from the cytoskeleton to the nucleus. *Nat Rev Mol Cell Biol* 2004; 5: 920–931.
- [122] Kuroda S, Tokunaga C, Kiyohara Y, Higuchi O, Konishi H, Mizuno K, et al. Protein-protein interaction of zinc finger LIM domains with protein kinase C. *J Biol Chem* 1996; 271: 31029–31032.
- [123] Nakagawa N, Hoshijima M, Oyasu M, Saito N, Tanizawa K, Kuroda S. ENH, containing PDZ and LIM domains, heart/skeletal muscle-specific protein, associates with cytoskeletal proteins through the PDZ domain. *Biochem Bioph Res Co* 2000; 272: 505–512.
- [124] Maturana AD, Walchli S, Iwata M, Ryser S, Van LJ, Hoshijima M, et al. Enigma homolog 1 scaffolds protein kinase D1 to regulate the activity of the cardiac L-type voltage-gated calcium channel. *Cardiovasc Res* 2008; 78: 458–465.
- [125] Cheng H, Kimura K, Peter AK, Cui L, Ouyang K, Shen T, et al. Loss of enigma homolog protein results in dilated cardiomyopathy. *Circ Res* 2010; 107: 348–356.
- [126] Faulkner G, Pallavicini A, Formentin E, Comelli A, Ievolella C, Trevisan S, et al. ZASP: a new Z-band alternatively spliced PDZ-motif protein. *J Cell Biol* 1999; 146: 465–475.
- [127] Berk B, Fujiwara K, Lehoux S. ECM remodeling in hypertensive heart disease. *J Clin Invest* 2007; 117:
- [128] Ruwhof C, Van Wamel AE, Van der Valk LJ, Schrier PI, Van der Laarse A. Direct, autocrine and paracrine effects of cyclic stretch on growth of myocytes and fibroblasts isolated from neonatal rat ventricles. *Arch Physiol Biochem* 2001; 109: 10–7.
- [129] Barbone A, Oz MC, Burkhoff D, Holmes JW. Normalized diastolic properties after left ventricular assist result from reverse remodeling of chamber geometry. *Circulation* 2001; 104: I229–32.
- [130] Life Sciences CI. *Guide for the Care and Use of Laboratory Animals*, National Academy Press, 1996;
- [131] Janssen PM, Lehnart SE, Prestle J, Lynker JC, Salfeld P, Just H, et al. The trabecula culture system: a novel technique to study contractile parameters over a multiday time period. *Am J Physiol* 1998; 274: H1481–8.

- [132] Pavlidis P. Using ANOVA for gene selection from microarray studies of the nervous system. *Methods* 2003; 31: 282–289.
- [133] Grashow JS, Sacks MS, Liao J, Yoganathan AP. Planar biaxial creep and stress relaxation of the mitral valve anterior leaflet. *Ann Biomed Eng* 2006; 34: 1509–18.
- [134] Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 2003; 19: 185–93.
- [135] Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* 2003; 31: e15.
- [136] Modlich O, Prisack H-B, Munnes M, Audretsch W, Bojar H. Immediate gene expression changes after the first course of neoadjuvant chemotherapy in patients with primary breast cancer disease. *Clin Cancer Res* 2004; 10: 6418–31.
- [137] Calza S, Raffelsberger W, Ploner A, Sahel J, Leveillard T, Pawitan Y. Filtering genes to improve sensitivity in oligonucleotide microarray data analysis. *Nucleic Acids Res* 2007; 35: e102.
- [138] Jain N, Thatte J, Braciale T, Ley K, O’Connell M, Lee JK. Local-pooled-error test for identifying differentially expressed genes with a small number of replicated microarrays. *Bioinformatics* 2003; 19: 1945–51.
- [139] Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009; 4: 44–57.
- [140] Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 2009; 37: 1–13.
- [141] Bisping E, Ikeda S, Kong SW, Tarnavski O, Bodyak N, McMullen JR, et al. Gata4 is required for maintenance of postnatal cardiac function and protection from pressure overload-induced heart failure. *Proc Natl Acad Sci USA* 2006; 103: 14471–6.
- [142] Schwientek P, Ellinghaus P, Steppan S, D’Urso D, Seewald M, Kassner A, et al. Global gene expression analysis in nonfailing and failing myocardium pre- and postpulsatile and nonpulsatile ventricular assist device support. *Physiol Genomics* 2010; 42: 397–405.
- [143] Nakamura A, Rokosh DG, Paccanaro M, Yee RR, Simpson PC, Grossman W, et al. LV systolic performance improves with development of hypertrophy after transverse aortic constriction in mice. *Am J Physiol-Heart C* 2001; 281: H1104–12.

- [144] Esposito G, Rapacciuolo A, Naga Prasad S V, Takaoka H, Thomas SA, Koch WJ, et al. Genetic alterations that inhibit in vivo pressure-overload hypertrophy prevent cardiac dysfunction despite increased wall stress. *Circulation* 2002; 105: 85–92.
- [145] Barbone A, Holmes JW, Heerdt PM, The' AH, Naka Y, Joshi N, et al. Comparison of right and left ventricular responses to left ventricular assist device support in patients with severe heart failure: a primary role of mechanical unloading underlying reverse remodeling. *Circulation* 2001; 104: 670–5.
- [146] Koitabashi N, Arai M, Niwano K, Watanabe A, Endoh M, Suguta M, et al. Plasma connective tissue growth factor is a novel potential biomarker of cardiac dysfunction in patients with chronic heart failure. *Eur J Heart Fail* 2008; 10: 373–9.
- [147] Matsui Y, Sadoshima J. Rapid upregulation of CTGF in cardiac myocytes by hypertrophic stimuli: implication for cardiac fibrosis and hypertrophy. *J Mol Cell Cardiol* 2004; 37: 477–81.
- [148] He Z, Way KJ, Arikawa E, Chou E, Opland DM, Clermont A, et al. Differential regulation of angiotensin II-induced expression of connective tissue growth factor by protein kinase C isoforms in the myocardium. *J Biol Chem* 2005; 280: 15719–26.
- [149] Panek AN, Posch MG, Alenina N, Ghadge SK, Erdmann B, Popova E, et al. Connective tissue growth factor overexpression in cardiomyocytes promotes cardiac hypertrophy and protection against pressure overload. *PLoS One* 2009; 4: e6743.
- [150] Willis MS, Patterson C. Hold me tight: Role of the heat shock protein family of chaperones in cardiac disease. *Circulation* 2010; 122: 1740–51.
- [151] Weng Y-J, Hsieh DJ-Y, Kuo W-W, Lai T-Y, Hsu H-H, Tsai C-H, et al. E4BP4 is a cardiac survival factor and essential for embryonic heart development. *Mol Cell Biochem* 2010; 340: 187–94.
- [152] Baker BM, Chen CS. Deconstructing the third dimension: how 3D culture microenvironments alter cellular cues. *J Cell Sci* 2012; 125: 3015–24.
- [153] Banerjee I, Fuseler JW, Price RL, Borg TK, Baudino T a. Determination of cell types and numbers during cardiac development in the neonatal and adult rat and mouse. *Am J Physiol-Heart C* 2007; 293: H1883–91.
- [154] LaFramboise WA, Scalise D, Stoodley P, Graner SR, Guthrie RD, Magovern JA, et al. Cardiac fibroblasts influence cardiomyocyte phenotype in vitro. *Am J Physiol-Cell Ph* 2007; 292: C1799–808.
- [155] Gaudesius G, Miragoli M, Thomas SP, Rohr S. Coupling of cardiac electrical activity over extended distances by fibroblasts of cardiac origin. *Circ Res* 2003; 93: 421–8.

- [156] Belke DD, Betuing S, Tuttle MJ, Graveleau C, Young ME, Pham M, et al. Insulin signaling coordinately regulates cardiac size, metabolism, and contractile protein isoform expression. *J Clin Invest* 2002; 109: 629–39.
- [157] Heineke J, Molkentin JD. Regulation of cardiac hypertrophy by intracellular signalling pathways. *Nat Rev Mol Cell Biol* 2006; 7: 589–600.
- [158] Gorza L, Pauletto P, Pessina AC, Sartore S, Schiaffino S. Isomyosin distribution in normal and pressure-overloaded rat ventricular myocardium. An immunohistochemical study. *Circ Res* 1981; 49: 1003–1009.
- [159] Mahdavi V, Lompre AM, Chambers AP, Nadal-Ginard B. Cardiac myosin heavy chain isozymic transitions during development and under pathological conditions are regulated at the level of mRNA availability. *Eur Heart J* 1984; 5 Suppl F: 181–191.
- [160] Mercadier JJ, Lompré AM, Wisnewsky C, Samuel JL, Bercovici J, Swynghedauw B, et al. Myosin isoenzyme changes in several models of rat cardiac hypertrophy. *Circ Res* 1981; 49: 525–532.
- [161] Schwartz K, De la Bastie D, Bouveret P, Oliviero P, Alonso S, Buckingham M. Alpha-skeletal muscle actin mRNA's accumulate in hypertrophied adult rat hearts. *Circ Res* 1986; 59: 551–555.
- [162] Weinberg E, Thienelt C, Lorell B. Pretranslational regulation of glucose-transporter isoform expression in hearts with pressure-overload left ventricular hypertrophy. *Circulation* 1995; 92: I–385(Abstr.).
- [163] Sack MN, Disch DL, Rockman HA, Kelly DP. A role for Sp and nuclear receptor transcription factors in a cardiac hypertrophic growth program. *Proc Natl Acad Sci USA* 1997; 94: 6438–6443.
- [164] Barry SP, Davidson SM, Townsend PA. Molecular regulation of cardiac hypertrophy. *Int J Biochem Cell B* 2008; 40: 2023–2039.
- [165] Taegtmeyer H, Sen S, Vela D. Return to the fetal gene program: a suggested metabolic link to gene expression in the heart. *Ann NY Acad Sci* 2010; 1188: 191–198.
- [166] Yamazaki T, Wälchli S, Fujita T, Ryser S, Hoshijima M, Schlegel W, et al. Splice variants of Enigma homolog, differentially expressed during heart development, promote or prevent hypertrophy. *Cardiovasc Res* 2010; 86: 374–82.
- [167] Kong Y, Tannous P, Lu G, Berenji K, Rothermel BA, Olson EN, et al. Suppression of class I and II histone deacetylases blunts pressure-overload cardiac hypertrophy. *Circulation* 2006; 113: 2579–2588.

- [168] Van Rooij E, Marshall WS, Olson EN. Toward microRNA-based therapeutics for heart disease: the sense in antisense. *Circ Res* 2008; 103: 919–928.
- [169] Delgado-Olguín P, Huang Y, Li X, Christodoulou D, Seidman CE, Seidman JG, et al. Epigenetic repression of cardiac progenitor gene expression by Ezh2 is required for postnatal cardiac homeostasis. *Nat Genet* 2012; 44: 343–347.
- [170] Keene JD. RNA regulons: coordination of post-transcriptional events. *Nat Rev Genet* 2007; 8: 533–543.
- [171] Nilsen TW, Graveley BR. Expansion of the eukaryotic proteome by alternative splicing. *Nature* 2010; 463: 457–463.
- [172] Lee J-H, 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: 1332–1341.
- [173] Park JY, Li W, Zheng D, Zhai P, Zhao Y, Matsuda T, et al. Comparative analysis of mRNA isoform expression in cardiac hypertrophy and development reveals multiple post-transcriptional regulatory modules. *PLoS One* 2011; 6: e22391.
- [174] Hu P, Zhang D, Swenson L, Chakrabarti G, Abel ED, Litwin SE. Minimally invasive aortic banding in mice: effects of altered cardiomyocyte insulin signaling during pressure overload. *Am J Physiol-Heart C* 2003; 285: H1261–1269.
- [175] Freeling J, Wattier K, LaCroix C, Li Y-F. Neostigmine and pilocarpine attenuated tumour necrosis factor alpha expression and cardiac hypertrophy in the heart with pressure overload. *Exp Physiol* 2008; 93: 75–82.
- [176] Zhao M, Chow A, Powers J, Fajardo G, Bernstein D. Microarray analysis of gene expression after transverse aortic constriction in mice. *Physiol Genomics* 2004; 19: 93–105.
- [177] Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 2009; 25: 1105–1111.
- [178] Trapnell C, Williams B, Pertea G, Mortazavi A, Kwan G, Van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 2010; 28: 511–515.
- [179] Software G. InStat guide to choosing and interpreting statistical tests, GraphPad Software, Inc., San Diego California USA, 2001;
- [180] Gunning PW, Schevzov G, Kee AJ, Hardeman EC. Tropomyosin isoforms: diving rods for actin cytoskeleton function. *Trends Cell Biol* 2005; 15: 333–41.

- [181] Anderson ES, Lin C-H, Xiao X, Stoilov P, Burge CB, Black DL. The cardiotoxic steroid digitoxin regulates alternative splicing through depletion of the splicing factors SRSF3 and TRA2B. *RNA* 2012; 18: 1041–1049.
- [182] Schilder RJ, Kimball SR, Jefferson LS. Cell-autonomous regulation of fast troponin T pre-mRNA alternative splicing in response to mechanical stretch. *Am J Physiol-Cell Ph* 2012; 303: C298–307.
- [183] Dewey FE, Perez M V, Wheeler MT, Watt C, Spin J, Langfelder P, et al. Gene coexpression network topology of cardiac development, hypertrophy, and failure. *Circ Cardiovasc Genet* 2011; 4: 26–35.
- [184] Heyd F, Lynch KW. Degrade, move, regroup: signaling control of splicing proteins. *Trends Biochem Sci* 2011; 36: 397–404.
- [185] Blaustein M, Pelisch F, Srebrow A. Signals, pathways and splicing regulation. *Int J Biochem Cell B* 2007; 39: 2031–48.
- [186] Ahuja P, Sdek P, MacLellan WR. Cardiac myocyte cell cycle control in development, disease, and regeneration. *Physiol Rev* 2007; 87: 521–44.
- [187] Lezoualc'h F, Métrich M, Hmitou I, Duquesnes N, Morel E. Small GTP-binding proteins and their regulators in cardiac hypertrophy. *J Mol Cell Cardiol* 2008; 44: 623–32.
- [188] Wang C-LA, Coluccio LM. New insights into the regulation of the actin cytoskeleton by tropomyosin. *Int Rev Cel Mol Bio* 2010; 281: 91–128.
- [189] Laakso JM, Lewis JH, Shuman H, Ostap EM. Myosin I can act as a molecular force sensor. *Science* 2008; 321: 133–6.
- [190] Laakso JM, Lewis JH, Shuman H, Ostap EM. Control of myosin-I force sensing by alternative splicing. *Proc Natl Acad Sci USA* 2010; 107: 698–702.
- [191] Azakie A, Fineman JR, He Y. Myocardial transcription factors are modulated during pathologic cardiac hypertrophy in vivo. *J Thorac Cardiovasc Sur* 2006; 132: 1262–71.
- [192] Stewart AF, Suzow J, Kubota T, Ueyama T, Chen HH. Transcription factor RTEF-1 mediates alpha1-adrenergic reactivation of the fetal gene program in cardiac myocytes. *Circ Res* 1998; 83: 43–9.
- [193] Tsika RW, Ma L, Kehat I, Schramm C, Simmer G, Morgan B, et al. TEAD-1 overexpression in the mouse heart promotes an age-dependent heart dysfunction. *J Biol Chem* 2010; 285: 13721–35.

- [194] Zuzarte PC, Farrance IK, Simpson PC, Wildeman AG. Tumor cell splice variants of the transcription factor TEF-1 induced by SV40 T-antigen transformation. *Biochim Biophys Acta* 2000; 1517: 82–90.
- [195] Gupta M, Kogut P, Davis FJ, Belaguli NS, Schwartz RJ, Gupta MP. Physical interaction between the MADS box of serum response factor and the TEA/ATTS DNA-binding domain of transcription enhancer factor-1. *J Biol Chem* 2001; 276: 10413–22.
- [196] Maeda T, Gupta MP, Stewart AFR. TEF-1 and MEF2 transcription factors interact to regulate muscle-specific promoters. *Biochem Biophys Res Co* 2002; 294: 791–7.
- [197] Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L. Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biol* 2011; 12: R22.
- [198] Roberts A, Pimentel H, Trapnell C, Pachter L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* 2011; 1–5.
- [199] Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009; 4: 44–57.
- [200] Warnes, GR. Includes R source code and/or documentation contributed by: Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, Lumley T, Maechler M, Magnusson A, Moeller S, Schwartz M VB. *gplots: Various R programming tools for plotting data.* 2012;
- [201] Fagnani M, Barash Y, Ip JY, Misquitta C, Pan Q, Saltzman AL, et al. Functional coordination of alternative splicing in the mammalian central nervous system. *Genome Biol* 2007; 8: R108.
- [202] Peng L, Yoo B, Gunewardena SS, Lu H, Klaassen CD, Zhong X-B. RNA sequencing reveals dynamic changes of mRNA abundance of cytochromes P450 and their alternative transcripts during mouse liver development. *Drug Metab Dispos* 2012; 40: 1198–209.
- [203] Tomczak KK, Marinescu VD, Ramoni MF, Sanoudou D, Montanaro F, Han M, et al. Expression profiling and identification of novel genes involved in myogenic differentiation. *FASEB J* 2004; 18: 403–5.
- [204] LaGamba D, Nawshad A, Hay ED. Microarray analysis of gene expression during epithelial-mesenchymal transformation. *Dev Dyn* 2005; 234: 132–42.
- [205] Salomonis N, Nelson B, Vranizan K, Pico AR, Hanspers K, Kuchinsky A, et al. Alternative splicing in the differentiation of human embryonic stem cells into cardiac precursors. *PLoS Comput Biol* 2009; 5: e1000553.

- [206] Warzecha CC, Jiang P, Amirikian K, Dittmar KA, Lu H, Shen S, et al. An ESRP-regulated splicing programme is abrogated during the epithelial-mesenchymal transition. *EMBO J* 2010; 29: 3286–300.
- [207] Liu P, Barb J, Woodhouse K, Taylor JG, Munson PJ, Raghavachari N. Transcriptome profiling and sequencing of differentiated human hematopoietic stem cells reveal lineage-specific expression and alternative splicing of genes. *Physiol Genomics* 2011; 43: 1117–34.
- [208] Pihlajamäki J, Lerin C, Itkonen P, Boes T, Floss T, Schroeder J, et al. Expression of the splicing factor gene *SFRS10* is reduced in human obesity and contributes to enhanced lipogenesis. *Cell Metab* 2011; 14: 208–18.
- [209] David CJ, Manley JL. Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes Dev* 2010; 24: 2343–64.
- [210] Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S, et al. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* 2011; 474: 380–4.
- [211] Ames EG, Lawson MJ, Mackey AJ, Holmes JW. Sequencing of mRNA identifies re-expression of fetal splice variants in cardiac hypertrophy. *J Mol Cell Cardiol* 2013;
- [212] Braunschweig U, Gueroussov S, Plocik AM, Graveley BR, Blencowe BJ. Dynamic Integration of Splicing within Gene Regulatory Pathways. *Cell* 2013; 152: 1252–1269.
- [213] Shepard PJ, Hertel KJ. The SR protein family. *Genome Biol* 2009; 10: 242.
- [214] Shen H, Green MR. RS domains contact splicing signals and promote splicing by a common mechanism in yeast through humans. *Genes Dev* 2006; 20: 1755–65.
- [215] Han SP, Tang YH, Smith R. Functional diversity of the hnRNPs: past, present and perspectives. *Biochem J* 2010; 430: 379–92.
- [216] Jacob F, Perrin D, Sanchez C, Monod J. Operon: a group of genes with the expression coordinated by an operator. *C R Hebd Seances Acad Sci* 1960; 250: 1727–1729.
- [217] Chen M, Manley JL. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol* 2009; 10: 741–54.
- [218] Sammeth M. Complete alternative splicing events are bubbles in splicing graphs. *J Comput Biol* 2009; 16: 1117–40.
- [219] Rogers MF, Thomas J, Reddy AS, Ben-Hur A. SpliceGrapher: detecting patterns of alternative splicing from RNA-Seq data in the context of gene models and EST data. *Genome Biol* 2012; 13: R4.

- [220] Hietaniemi J. *Graph::Directed*. 2010;
- [221] Barash Y, Calarco JA, Gao W, Pan Q, Wang X, Shai O, et al. Deciphering the splicing code. *Nature* 2010; 465: 53–59.
- [222] Suyama M, Harrington ED, Vinokourova S, Von Knebel Doeberitz M, Ohara O, Bork P. A network of conserved co-occurring motifs for the regulation of alternative splicing. *Nucleic Acids Res* 2010; 38: 7916–26.
- [223] Sudarsanam P, Pilpel Y, Church GM. Genome-wide co-occurrence of promoter elements reveals a cis-regulatory cassette of rRNA transcription motifs in *Saccharomyces cerevisiae*. *Genome Res* 2002; 12: 1723–31.
- [224] Ule J, Stefani G, Mele A, Ruggiu M, Wang X, Taneri B, et al. An RNA map predicting Nova-dependent splicing regulation. *Nature* 2006; 444: 580–6.
- [225] Koralewski TE, Krutovsky K V. Evolution of exon-intron structure and alternative splicing. *PLoS One* 2011; 6: e18055.
- [226] Lamichhane R, Daubner GM, Thomas-Crusells J, Auweter SD, Manatschal C, Austin KS, et al. RNA looping by PTB: Evidence using FRET and NMR spectroscopy for a role in splicing repression. *Proc Natl Acad Sci USA* 2010; 107: 4105–10.
- [227] Oberstrass FC, Auweter SD, Erat M, Hargous Y, Henning A, Wenter P, et al. Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science* 2005; 309: 2054–7.
- [228] Yuan X, Davydova N, Conte MR, Curry S, Matthews S. Chemical shift mapping of RNA interactions with the polypyrimidine tract binding protein. *Nucleic Acids Res* 2002; 30: 456–62.
- [229] Llorian M, Schwartz S, Clark TA, Hollander D, Tan L-Y, Spellman R, et al. Position-dependent alternative splicing activity revealed by global profiling of alternative splicing events regulated by PTB. *Nat Struct Mol Biol* 2010; 17: 1114–23.
- [230] Shi Y, Di Giammartino DC, Taylor D, Sarkeshik A, Rice WJ, Yates JR, et al. Molecular architecture of the human pre-mRNA 3' processing complex. *Mol Cell* 2009; 33: 365–76.
- [231] Wahle E. A novel poly(A)-binding protein acts as a specificity factor in the second phase of messenger RNA polyadenylation. *Cell* 1991; 66: 759–68.
- [232] Otero LJ, Ashe MP, Sachs AB. The yeast poly(A)-binding protein Pab1p stimulates in vitro poly(A)-dependent and cap-dependent translation by distinct mechanisms. *EMBO J* 1999; 18: 3153–63.

- [233] Ishigaki Y, Li X, Serin G, Maquat LE. Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. *Cell* 2001; 106: 607–17.
- [234] Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, et al. Initial sequencing and comparative analysis of the mouse genome. *Nature* 2002; 420: 520–62.
- [235] Rossbach O, Hung L-H, Schreiner S, Grishina I, Heiner M, Hui J, et al. Auto- and cross-regulation of the hnRNP L proteins by alternative splicing. *Mol Cell Biol* 2009; 29: 1442–51.
- [236] Hui J, Hung L-H, Heiner M, Schreiner S, Neumüller N, Reither G, et al. Intronic CA-repeat and CA-rich elements: a new class of regulators of mammalian alternative splicing. *EMBO J* 2005; 24: 1988–98.
- [237] Guang S, Felthouser AM, Mertz JE. Binding of hnRNP L to the pre-mRNA processing enhancer of the herpes simplex virus thymidine kinase gene enhances both polyadenylation and nucleocytoplasmic export of intronless mRNAs. *Mol Cell Biol* 2005; 25: 6303–13.
- [238] Motta-Mena LB, Heyd F, Lynch KW. Context-dependent regulatory mechanism of the splicing factor hnRNP L. *Mol Cell* 2010; 37: 223–34.
- [239] Goehe RW, Shultz JC, Murudkar C, Usanovic S, Lamour NF, Massey DH, et al. hnRNP L regulates the tumorigenic capacity of lung cancer xenografts in mice via caspase-9 pre-mRNA processing. *J Clin Invest* 2010; 120: 3923–39.
- [240] Li J, Guo W, Li F, He J, Yu Q, Wu X, et al. HnRNPL as a key factor in spermatogenesis: Lesson from functional proteomic studies of azoospermia patients with sertoli cell only syndrome. *J Proteomics* 2012; 75: 2879–91.
- [241] Liu G, Razanau A, Hai Y, Yu J, Sohail M, Lobo VG, et al. A conserved serine of heterogeneous nuclear ribonucleoprotein L (hnRNP L) mediates depolarization-regulated alternative splicing of potassium channels. *J Biol Chem* 2012; 287: 22709–16.
- [242] Hahm B, Cho OH, Kim JE, Kim YK, Kim JH, Oh YL, et al. Polypyrimidine tract-binding protein interacts with HnRNP L. *FEBS Lett* 1998; 425: 401–6.
- [243] Mori D, Sasagawa N, Kino Y, Ishiura S. Quantitative analysis of CUG-BP1 binding to RNA repeats. *J Biochem* 2008; 143: 377–83.
- [244] Edwards J, Malaurie E, Kondrashov A, Long J, De Moor CH, Searle MS, et al. Sequence determinants for the tandem recognition of UGU and CUG rich RNA elements by the two N-terminal RRM of CELF1. *Nucleic Acids Res* 2011; 39: 8638–50.

- [245] Philips A V, Timchenko LT, Cooper TA. Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. *Science* 1998; 280: 737–41.
- [246] Roca X, Krainer A, Eperon I. Pick one, but be quick: 5' splice sites and the problems of too many choices. *Genes Dev* 2013; 129–144.
- [247] Kakaradov B, Xiong HY, Lee LJ, Jojic N, Frey BJ. Challenges in estimating percent inclusion of alternatively spliced junctions from RNA-seq data. *BMC Bioinformatics* 2012; 13 Suppl 6: S11.
- [248] Chou MY, Underwood JG, Nikolic J, Luu MH, Black DL. Multisite RNA binding and release of polypyrimidine tract binding protein during the regulation of c-src neural-specific splicing. *Mol Cell* 2000; 5: 949–57.
- [249] Wagner EJ, Garcia-Blanco MA. Polypyrimidine tract binding protein antagonizes exon definition. *Mol Cell Biol* 2001; 21: 3281–8.
- [250] Boutz PL, Stoilov P, Li Q, Lin C-H, Chawla G, Ostrow K, et al. A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. *Genes Dev* 2007; 21: 1636–52.
- [251] Ye J, Llorian M, Cardona M, Rongvaux A, Moubarak RS, Comella JX, et al. A pathway involving HDAC5, cFLIP and caspases regulates expression of the splicing regulator polypyrimidine tract binding protein in the heart. *J Cell Sci* 2013; 126: 1682–91.
- [252] Zhang CL, McKinsey TA, Chang S, Antos CL, Hill JA, Olson EN. Class II histone deacetylases act as signal-responsive repressors of cardiac hypertrophy. *Cell* 2002; 110: 479–88.
- [253] Ye J, Cardona M, Llovera M, Comella JX, Sanchis D. Translation of Myocyte Enhancer Factor-2 is induced by hypertrophic stimuli in cardiomyocytes through a Calcineurin-dependent pathway. *J Mol Cell Cardiol* 2012; 53: 578–87.
- [254] Wollerton MC, Gooding C, Wagner EJ, Garcia-Blanco MA, Smith CWJ. Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. *Mol Cell* 2004; 13: 91–100.
- [255] Rideau AP, Gooding C, Simpson PJ, Monie TP, Lorenz M, Hüttelmaier S, et al. A peptide motif in Raver1 mediates splicing repression by interaction with the PTB RRM2 domain. *Nat Struct Mol Biol* 2006; 13: 839–48.
- [256] Zieseniss A, Schroeder U, Buchmeier S, Schoenenberger C-A, Van den Heuvel J, Jockusch BM, et al. Raver1 is an integral component of muscle contractile elements. *Cell Tissue Res* 2007; 327: 583–94.

- [257] Kim Y, Phan D, Van Rooij E, Wang D-Z, McAnally J, Qi X, et al. The MEF2D transcription factor mediates stress-dependent cardiac remodeling in mice. *J Clin Invest* 2008; 118: 124–32.
- [258] Van Rooij E, Sutherland LB, Qi X, Richardson J a, Hill J, Olson EN. Control of stress-dependent cardiac growth and gene expression by a microRNA. *Science* 2007; 316: 575–9.
- [259] Tatsuguchi M, Seok HY, Callis TE, Thomson JM, Chen J-F, Newman M, et al. Expression of microRNAs is dynamically regulated during cardiomyocyte hypertrophy. *J Mol Cell Cardiol* 2007; 42: 1137–41.
- [260] Sterne-Weiler T, Howard J, Mort M, Cooper DN, Sanford JR. Loss of exon identity is a common mechanism of human inherited disease. *Genome Res* 2011; 21: 1563–71.
- [261] Krawczak M, Thomas NST, Hundrieser B, Mort M, Wittig M, Hampe J, et al. Single base-pair substitutions in exon-intron junctions of human genes: nature, distribution, and consequences for mRNA splicing. *Human Mutat* 2007; 28: 150–8.
- [262] Biamonti G, Vourc'h C. Nuclear stress bodies. *Cold Spring Harb Perspect Biol* 2010; 2: a000695.
- [263] Biamonti G, Caceres JF. Cellular stress and RNA splicing. *Trends Biochem Sci* 2009; 34: 146–53.
- [264] Balagopal V, Parker R. Polysomes, P bodies and stress granules: states and fates of eukaryotic mRNAs. *Curr Opin Cell Biol* 2009; 21: 403–8.
- [265] Gilks N, Kedersha N, Ayodele M, Shen L, Stoecklin G, Dember LM, et al. Stress granule assembly is mediated by prion-like aggregation of TIA-1. *Mol Biol Cell* 2004; 15: 5383–98.
- [266] Anderson P, Kedersha N. Stressful initiations. *J Cell Sci* 2002; 115: 3227–34.
- [267] Denegri M, Chiodi I, Corioni M, Cobianchi F, Riva S, Biamonti G. Stress-induced nuclear bodies are sites of accumulation of pre-mRNA processing factors. *Mol Biol Cell* 2001; 12: 3502–14.
- [268] Fomovsky GM, Thomopoulos S, Holmes JW. Contribution of extracellular matrix to the mechanical properties of the heart. *J Mol Cell Cardiol* 2010; 48: 490–6.
- [269] Geiger B, Bershadsky A, Pankov R, Yamada KM. Transmembrane crosstalk between the extracellular matrix--cytoskeleton crosstalk. *Nat Rev Mol Cell Biol* 2001; 2: 793–805.
- [270] Rouillard AD, Holmes JW. Mechanical regulation of fibroblast migration and collagen remodelling in healing myocardial infarcts. *J Physiol* 2012; 590: 4585–602.

[271] Conway SJ, Molkenin JD. Periostin as a heterofunctional regulator of cardiac development and disease. *Curr Genomics* 2008; 9: 548–55.