Getting to the heart of the matter: long non-coding RNAs in cardiac development and disease

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1038/emboj.2013.134">http://dx.doi.org/10.1038/emboj.2013.134</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>Nature Publishing Group</td>
</tr>
<tr>
<td>Version</td>
<td>Author's final manuscript</td>
</tr>
<tr>
<td>Accessed</td>
<td>Sun Mar 31 19:37:44 EDT 2019</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/86043">http://hdl.handle.net/1721.1/86043</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Creative Commons Attribution-Noncommercial-Share Alike</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td><a href="http://creativecommons.org/licenses/by-nc-sa/4.0/">http://creativecommons.org/licenses/by-nc-sa/4.0/</a></td>
</tr>
</tbody>
</table>
Review

Getting to the heart of the matter: long non-coding RNAs in cardiac development and disease

Johanna C. Scheuermann and Laurie A. Boyer

Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge MA 02139, USA

1 These authors contributed equally to this work.
*To whom correspondence should be addressed:
Laurie A. Boyer
Email: lboyer@mit.edu
Tel: 617 324-3335
Fax: 617 253-8699

Key words
long non-coding RNA, cardiogenesis, congenital heart disease, cardiovascular disease, gene expression, heart development
Summary

Cardiogenesis in mammals requires exquisite control of gene expression and faulty regulation of transcriptional programs underpins congenital heart disease (CHD), the most common defect among live births. Similarly, many adult cardiac diseases involve transcriptional changes and sometimes have a developmental basis. Long non-coding RNAs (lncRNAs) are a novel class of transcripts that regulate cellular processes by controlling gene expression; however, detailed insights into their biological and mechanistic functions are only beginning to emerge. Here, we discuss recent findings suggesting that lncRNAs are important factors in regulation of mammalian cardiogenesis and in the pathogenesis of CHD as well as adult cardiac disease. We also outline potential methodological and conceptual considerations for future studies of lncRNAs in the heart and other contexts.
Introduction

The human genome project has opened the door for understanding development and disease at an unprecedented level. Once referred to as dark matter or “junk DNA”, it is now thought that up to 90% of the human genome is actively transcribed and produces many different types of transcripts including protein-coding and non-coding RNA (Carninci et al, 2005; Kapranov et al, 2007; ENCODE Project Consortium et al, 2012). Non-coding RNAs have been broadly classified according to transcript length as small and long non-coding RNAs, where long non-coding RNAs (lncRNAs) are considered to be greater than 200 nucleotides and can comprise up to thousands of nucleotides. The biological and mechanistic functions of different small non-coding RNA species have been extensively reviewed (Matera et al, 2007; Carthew & Sontheimer, 2009; Malone & Hannon, 2009; Ghildiyal & Zamore, 2009; Castel & Martienssen, 2013). Much less is known, however, about the functions of lncRNAs, an apparently heterogeneous class of RNA molecules with emerging biological functions.

LncRNAs are pervasively transcribed throughout the genome and the resulting transcripts display remarkable similarities to classical mRNAs in that they are transcribed by RNA Polymerase II (RNAP2) and are generally, but not always, alternatively spliced, 5’-capped, and polyadenylated (Derrien et al, 2012). In contrast to protein-coding genes, lncRNAs have limited coding potential as indicated by the lack of protein domains or significant open reading frames (ORFs). Moreover, lncRNAs display random codon usage and no significant bias towards silent nucleotide substitutions underscoring their low protein-coding potential; and these transcripts are rarely translated despite their engagement with ribosomes in some cases (Cabili et al, 2011; Lin et al, 2011; Ulitsky et al, 2011; Bánfai et al, 2012). Based largely on these criteria, a large number of lncRNA have been identified across eukaryotes, thousands of them in mouse and human (Ulitsky et al, 2011; Derrien et al, 2012). Remarkably, lncRNAs appear to be rapidly evolving and generally display low levels of sequence conservation. About one-third of all human lncRNAs have arisen only within the primate lineage (Derrien et al, 2012), while only a small subgroup of lncRNAs appear to be maintained throughout a range of species with conservation being generally most evident in their promoter regions (Chodroff et al, 2010; Ulitsky et al, 2011). Together, these data suggest that regulation of lncRNA expression patterns is important for the function of this class of transcripts.

Based on their genomic location, lncRNAs can be further grouped into different classes of transcripts (Ulitsky et al, 2011; Derrien et al, 2012) (Figure 1A). For example, long intergenic (or
Intervening) non-coding RNAs (lincRNAs) are located between coding or non-coding genes, and do not overlap the exons of other genes. IncRNA loci can also reside within the introns of protein-coding genes. Natural antisense transcripts (NATs) are produced from the opposite strand of a coding (or non-coding gene) but their transcription start site is located downstream relative to that of the coding gene, and these transcripts often display at least partial overlap with the coding sequence of the corresponding mRNA. LincRNAs are highly versatile in that they can partially base-pair with other RNA templates to form duplexes or with DNA which can lead to the formation of triplex structures (Figure 1B). Moreover, these transcripts can interact with a diverse repertoire of proteins. Thus, lincRNAs are thought to possess tremendous regulatory potential.

In general, lincRNAs can act in cis to regulate neighboring genes or in trans to modulate expression of their target genes by employing a wide range of molecular mechanisms. For example, a subclass of lincRNAs with apparent enhancer-like activity, termed ncRNA-activating (ncRNA-a), has recently been found to activate neighboring genes in cis using a mechanism involving DNA-looping between the lincRNA and its target gene (Orom et al, 2010; Lai et al, 2013). Evidence also indicates that a subset of lincRNAs regulates gene expression by acting in trans as recruiters or decoys for chromatin modifiers and transcription factors to activate or silence genes (Rinn et al, 2007; Khalil et al, 2009; Tsai et al, 2010; Ng et al, 2012; Klattenhoff et al, 2013; Grote et al, 2013). Notably, despite the lack of sequence conservation among lincRNAs, interactions with RNA binding proteins including Polycomb and Trithorax group members have been widely conserved between mouse and human (Guttman & Rinn, 2012). Alternatively, some lincRNAs have been reported to function as microRNA sponges, titrating these small transcripts away from their respective mRNA target (Cesana et al, 2011; Wang et al, 2013). LincRNAs have also been reported to influence mRNA splicing, translation, or degradation by binding to mRNAs or protein components of RNP complexes (Tripathi et al, 2010; Yoon et al, 2012; Gong & Maquat, 2011). Thus, lincRNAs have roles in transcriptional and post-transcriptional gene regulatory events. Consequently, the biological functions of lincRNAs as well as their mechanisms of action are expected to be diverse and will require a great deal of functional sub-classification in the future.

While most predicted lincRNAs await functional characterization, there are clear examples demonstrating prominent roles for these transcripts in a variety of cellular processes including dosage compensation (e.g. X chromosome inactivation), imprinting, regulation of cell cycle, and
apoptosis (Lee & Bartolomei, 2013; Rinn & Chang, 2012). Moreover, IncRNAs have been shown to play roles in somatic cell differentiation programs as well as maintenance of cell fate (Guttman et al., 2011; Hu et al., 2011; Kretz et al., 2012; 2013). Thus, IncRNAs may represent a new layer of regulation in differentiation and lineage commitment. Consistent with these broad roles, this class of transcripts has also been implicated as contributing factors to diseases with developmental components such as cancer and neurological disorders (Ponting et al., 2009; Wapinski & Chang, 2011; Mercer & Mattick, 2013; Batista & Chang, 2013; Ng et al., 2013). The emerging links between IncRNAs and disease as well as their tissue-specific expression patterns indicate that IncRNAs comprise a core transcriptional regulatory circuitry with master regulators and further suggest that they represent new molecules for targeted therapy.

Mammalian heart development is a tightly regulated process requiring exquisite control of transcriptional programs. Consistent with this idea, disruption of transcriptional networks underpins congenital heart disease (CHD) and certain forms of adult cardiac disease (Bruneau, 2008; Srivastava, 2006a). In fact, heart disease is the leading cause of morbidity and mortality worldwide (WHO 2011) with dramatic effects on the life quality of patients as well as on the health care system. Thus, dissecting the transcriptional regulatory principles that govern heart development and tissue homeostasis in the adult heart is of great interest to developmental and molecular biologists as well as clinicians. While it is known that the activities of DNA binding transcription factors, chromatin regulators, and signaling molecules converge to control tissue specific gene expression programs during heart development (Olson, 2006; Srivastava, 2006b; Chang & Bruneau, 2012; Bruneau, 2013), it is likely that non-coding transcripts also contribute to this highly orchestrated process. For example, members of the class of small non-coding RNAs such as miRNAs have critical roles in fine-tuning gene expression patterns during heart development (Cordes & Srivastava, 2009; Liu & Olson, 2010). Of particular interest is the recent discovery of IncRNAs that function in cardiac lineage commitment and heart development, revealing an additional layer of complexity (Klattenhoff et al., 2013; Grote et al., 2013). Although a general picture is emerging, we are only beginning to understand the implications of IncRNA regulation in heart development and cardiac-related disease. Here, we discuss the newly emerging roles of IncRNAs and cite specific examples in the context of heart development and cardiac disease as well as present considerations for the identification of additional IncRNA regulators of this process from a methodological and conceptual point of view.

**Heart development is regulated by tight control of gene expression patterns**
Heart development requires the concurrent differentiation of multiple cell types that must organize into a complex structure. This process necessitates tight control of gene expression patterns in a temporal and spatial manner (Srivastava, 2006b). In most species, the primitive heart is established by concordant expression of a highly conserved core cardiac transcription factor network (Olson, 2006; McCulley & Black, 2012). However, more complex structures such as those that comprise the multi-chambered mammalian heart may require additional species-specific regulatory factors. For example, the Drosophila heart is shaped as a relatively simple linear tube consisting of three sections that transport the hemolymph along the anterior-posterior body axis: heart proper, posterior aorta, and anterior aorta, with the beating heart proper being separated from the aorta by a cardiovascular valve (Medioni et al., 2009; Seyres et al., 2012) (Figure 2A). Mammalian heart development begins with formation of an analogous structure in embryogenesis, yet the developing heart undergoes a series of complex movements, rotations and subsequent refinement, resulting in a four-chambered heart with distinct in- and outflow tracts, cardiac valves separating the different compartments, and a mature conduction system (Harvey, 2002; Srivastava, 2006b; Vincent & Buckingham, 2010) (Figure 2B).

Despite the distinct structural differences between the simple Drosophila heart and the more complex mammalian heart, the core transcription factor network is highly conserved between both organisms (Reim & Frasch, 2010) (Figure 2C). In fact, Drosophila heart development has been used as a model to identify novel gene interactions leading to human heart disease (Qian et al., 2011; Qian & Bodmer, 2012). Mutations affecting cardiac transcription factors are often causative for congenital heart defects (Bruneau, 2008; McCulley & Black, 2012). For example, mutations in GATA4 or NKX2.5, members of the core cardiac transcription factor network, lead to atrial and ventricular septum defects and Tetralogy of Fallot. While the binding of sequence specific transcription factors to gene regulatory elements such as promoters and distal enhancers drive heart development, it has become increasingly clear that additional mechanisms contribute to fine-tuning the cardiac regulatory network. For example, microRNAs (e.g. miR-1, -126, -138, -143, -145) have roles in cardiogenic processes including angiogenesis, establishment of cardiac cell polarity, development of the cardiac conduction system, cardiac patterning, or smooth muscle cell differentiation, respectively, and in many cases these miRNAs regulate and interact with the core cardiac transcriptional network (Cordes & Srivastava, 2009; Liu & Olson, 2010). Notably, some miRNAs have been shown to play analogous roles in Drosophila heart development (Nguyen & Frasch, 2006). Despite the similarities between lower
eukaryotes such as *Drosophila* and more complex organisms, the mechanisms that give rise to the additional complexity of heart formation and function in mammals are not well understood. Given the recent identification of thousands of IncRNAs and their unique representation and tissue specific expression across organisms, these transcripts may signify a new class of regulatory molecules in cardiac development, and their functions may represent key distinctions that give rise to the more complex sub-structures of the mammalian heart.

**LncRNAs function to regulate developmental transitions in cardiac commitment**

LncRNAs have emerged as potent regulators of gene expression and may represent an important part of what distinguishes higher organisms from simpler eukaryotes. While unicellular organisms dedicate the majority of their genome to protein-coding sequence (~73 percent in *S. cerevisiae*), this fraction decreases significantly in multicellular organisms such as *Drosophila* (~18 percent), and coding sequences represent only a minor fraction of DNA in mammals (e.g. ~2-3 percent in humans) (Taft *et al*, 2007) (Figure 2D). Thus, it has been hypothesized that the ratio of non-coding compared to protein-coding transcripts in the genome, rather than the overall number of protein-coding genes, underpins organismal complexity. While this idea remains to be tested, the total number of IncRNAs transcripts increases dramatically from *Drosophila* to human (Figure 2E). For example, while mice and humans have thousands of putative IncRNA genes (Ponjavic *et al*, 2007; Ulitsky *et al*, 2011; Derrien *et al*, 2012; Sigova *et al*, 2013), recent estimates suggest that there are 17 (Tupy *et al*, 2005) and possibly upwards of ~1,000 IncRNA transcripts in *Drosophila* (Young *et al*, 2012), ~170 loci in *C. elegans* that specify ~272 lincRNAs (Nam & Bartel, 2012), and ~600 loci that produce ~700 transcripts in zebrafish (Ulitsky *et al*, 2011; Pauli *et al*, 2012). There are several caveats to this analysis however, as numbers of putative IncRNAs or subclasses such as lincRNAs can vary widely among studies given that different criteria are used to identify these transcripts in each case. Furthermore, the transcriptomes of lower eukaryotes have not been as thoroughly analyzed compared to mouse and human. Nevertheless, IncRNAs across multi-cellular organisms often display expression patterns that are highly tissue-specific, suggesting that at least some of these transcripts have roles in developmental processes (Cabili *et al*, 2011; Derrien *et al*, 2012). For example, many IncRNAs are expressed dynamically at specific developmental stages during cardiomyocyte differentiation (Wamstad *et al*, 2012). Indeed, two recent studies identified IncRNAs in mouse with functions in commitment to the cardiac lineage and heart development (Klattenhoff *et al*, 2013; Grote *et al*, 2013), opening the door to the possibility that IncRNAs represent new modes of developmental regulation.
Braveheart, a lncRNA required for the specification of a common cardiac progenitor

The lncRNA Braveheart (Bvht, AK143260) was discovered in mouse based on its unique expression pattern. Bvht is expressed at early developmental stages in mouse embryonic stem cells (mESCs) and also abundantly in the adult heart relative to other differentiated tissues suggesting that this lncRNA may be important for specification of the cardiac lineage. Consistent with this idea, depletion of Bvht in mESCs impaired formation of cardiomyocytes in multiple in vitro differentiation assays (Klattenhoff et al, 2013). Heart development involves the specification of cardiac progenitor cells within the lateral plate mesoderm. MESP1, an essential transcription factor that is conserved in vertebrates and some non-vertebrate chordate species (Saga et al, 1996; 1999; 2000; Satou et al, 2004; Kriegmair et al, 2013), marks the earliest known cardiac population during development as well as tissues that contribute to head mesenchyme (Bondue et al, 2008; Lindsley et al, 2008; David et al, 2008). MESP1 progenitors have the capacity to specify all cell types of the heart, including cardiomyocytes, endothelial cells, and cardiac smooth muscle cells. Notably, using an in vitro cardiomyocyte differentiation system, the investigators found that Bvht is required for induction of MesP1 and its downstream targets including the core cardiac transcription factors Gata4, Gata6, Hand1, Hand2, Tbx2, and Nkx2.5, among others. In contrast, mesoderm markers such as Brachury and Eomes were expressed normally at early stages of differentiation and remained expressed upon loss of Bvht. Notably, both BRACHYURY and EOMES are necessary for proper induction of MesP1 (Costello et al, 2011; David et al, 2011). Together, these data suggested that Bvht is necessary for the transition from nascent to cardiac mesoderm.

LncRNAs can function in trans by interacting with chromatin modifiers to mediate changes in gene expression (Guttman & Rinn, 2012). Notably, Bvht was found to interact with SUZ12, a core component of the Polycomb repressive complex 2 (PRC2) and loss of the lncRNA resulted in a failure to activate the cardiac gene expression program. Many of the transcription factors in the core cardiac network are targets of PRC2 mediated repression in ESCs and differentiation toward specific lineages requires the selective loss of PRC2 at subsets of these genes (Boyer et al, 2006). Upon Bvht depletion, PRC2 and its associated repressive modification H3K27me3 remained enriched at the promoters of critical genes in the cardiovascular network, including MesP1, Gata6, Hand1, Hand2, and Nkx2.5. Thus, Bvht may function as a molecular decoy to regulate expression of the core cardiac network (Figure 3A). However, whether Bvht regulates expression of the core cardiac network directly via its PRC2 binding activity or if it employs
additional molecular mechanisms to promote cardiac commitment remains an open question.

While *in vitro* ESC-based cardiomyocyte differentiation recapitulates many of the stages of gastrulation and specification of the cardiac lineage (Kattman *et al*, 2007; 2011; Wamstad *et al*, 2012), it will also be critical to determine the function of *Bvht* during development on an organismal level. A conserved *Bvht* transcript was not identified in rat or human suggesting that it is a rapidly evolving transcript (Klattenhoff *et al*, 2013). Although evidence of a transcript in rat and human is lacking, there is some DNA sequence conservation at syntenic sites among the three organisms making *Bvht* a particularly interesting example to study in terms of genomic evolution.

*Fendrr, a lncRNA necessary for heart and body wall development in mice.*
The lncRNA *Fendrr* (Foxf1 adjacent non-coding developmental regulatory RNA; ENSMUSG00000097336) was recently identified in mouse as a potential regulator of heart development by virtue of its specific expression in the lateral plate mesoderm (Grote *et al*, 2013). The lateral plate mesoderm gives rise to the heart and structures of the ventral body wall. Loss of *Fendrr* resulted in embryonic lethality in mice (~E13.5) and null embryos displayed open ventral body wall defects and hypoplastic ventricles, resulting in impaired heart function. Expression of a subset of cardiac transcription factors, *Nkx2.5* and *Gata6*, was increased in *Fendrr* loss of function hearts accompanied by corresponding changes in H3K4me3 levels at their promoters, whereas other members of the core cardiac network, such as *Gata4* and *Tbx5*, showed no changes. The *Fendrr* gene is proximal to Foxf1a, a gene that codes for a transcription factor involved in mesoderm formation. Consistent with a partial *cis* regulatory role for *Fendrr*, Foxf1a was also ectopically expressed in null animals. *Fendrr* interacted with PRC2 components as well as WDR5, a member of TrxG/MLL complex, to regulate mesoderm specific genes. These results suggested that *Fendrr* regulates the balance between repressive and activating marks at key genes during development (Figure 3B), although it is not clear if both complexes simultaneously interact with the non-coding transcript. The authors suggest that targeting these complexes to specific genomic sites is in part mediated through interactions between predicted unstructured regions of *Fendrr* and DNA; however, this idea needs to be further experimentally tested. Interestingly, a syntenic transcript exists in the human genome (ENSG00000268388) suggesting a conserved role for *Fendrr* in human heart development.
The interaction of *Fendrr* with multiple chromatin modifier complexes appears to be an emerging theme in lncRNA biology. For example, the lncRNA HOTAIR interacts with the PRC2 complex via its 5’ end and with the H3K4 demethylase LSD1 via its 3’ end, the Kcnq1ot1 lncRNA binds to both PRC2 and G9a (catalyzing the repressive H3K9 methylation mark), and the ANRIL lncRNA interacts with PRC1 and PRC2 (Mercer & Mattick, 2013). These observations are consistent with suggestions that lncRNAs act as dynamic modular scaffolds in a range of species (Guttman & Rinn, 2012; Spitale et al, 2011), potentially even adapting their binding capacities through conformation switches despite their lack of sequence conservation. In the case of *Fendrr* it will be important to test whether both the PRC2 as well as TrxG/MLL binding activities contribute directly to target gene regulation and how these functions are controlled molecularly. On a biological level, it will be necessary to dissect the developmental pathways regulated by *Fendrr*, since it seems to affect specific and distinct sub-populations within the lateral plate mesoderm.

There are a hundreds, if not thousands, of additional putative lncRNAs that are expressed during cardiogenesis, and in many cases even in a cell type-specific manner (Wamstad et al, 2012). While *Bvht* and *Fendrr* appear to function through epigenetic regulation of developmental gene expression programs, further detailed analyses of these individual candidates is expected to lead to the identification of additional lncRNAs with diverse functions in cardiovascular development. Dissecting how the expression of lncRNAs is regulated in a tissue-specific manner will be necessary in order to integrate these non-coding transcripts into the transcriptional regulatory circuitry that governs cardiogenesis. Similarly to protein-coding genes, lncRNAs appear to be regulated by cell type specific transcription factors. For example, in mESCs, a large subset of expressed lncRNAs is bound at their promoters by the key pluripotency regulators OCT4, SOX2, and NANOG (Guttman et al, 2011). Thus, it will also be of considerable interest to identify the set of transcription factors that regulate cardiac-specific lncRNAs in order to integrate this new class of regulators into the transcriptional regulatory circuitry of heart development.

**Cardiac disease and lncRNAs**

Mutations in key core cardiac transcription factors are causative for congenital heart disease and some adult cardiac-related diseases such as those that affect the heart muscle as well as the electrical circuits required for proper conduction. Given that lncRNAs appear to contribute to the regulation of cardiac networks, these transcripts are also expected to contribute to cardiac-related pathologies. Because many cardiac-related conditions are heritable, recent efforts to
identify potential new disease loci for cardiovascular diseases have relied in part on genome wide association studies (GWAS). The principle of GWAS is to analyze variations in nucleotide sequence, referred to as single nucleotide polymorphisms (SNPs), within a population of individuals that are afflicted with a particular condition as a means to identify new disease loci (Kathiresan & Srivastava, 2012). Notably, 93% of GWAS hits fall outside protein-coding regions and emerging evidence indicates that non-coding DNA, including distal regulatory elements as well as IncRNA genes that do not overlap known protein-coding genes, is enriched for disease SNPs (Maurano et al, 2012). In support of this idea, a number of IncRNAs have been implicated in adult cardiac disease by analysis of genetic variation among individuals with cardiac traits. Other examples of IncRNAs implicated in cardiac disease include natural antisense transcripts (NATs) that are transcribed in the opposite direction of critical heart development and structural genes suggesting that these NATs can impact the expression of key cardiac genes.

*MIAT, a IncRNA associated with myocardial infarction*

MIAT (myocardial infarction associated transcript) or Gomafu/RNCR2 was identified by GWAS as a risk factor associated with patients having suffered myocardial infarction (Ishii et al, 2006). Several variants were identified as significantly associated with higher susceptibility to myocardial infarction compared to controls. In fact, a particular SNP was associated with increased transcription of MIAT. MIAT accumulates in the nucleus in specific nuclear bodies and displays high expression levels in the central nervous system and lower levels in other tissues (Ishii et al, 2006; Sone et al, 2007; Tsuiji et al, 2011). It has also been implicated in retinal cell specification in the mouse (Rapicavoli et al, 2010) and may have a role in splicing regulation (Tsuiji et al, 2011), however MIAT's molecular role in myocardial infarction remains unknown. In some cases of cardiac disease, such as primary cardiomyopathy, the heart is directly affected, while in other cases cardiac disease results indirectly from conditions such as diabetes and inflammation, which increase the risk for developing atherosclerosis and coronary artery disease and eventually myocardial infarction. Consequently, the identification and functional validation of IncRNAs with roles in complex traits will be an added challenge.

*SRA, a bi-functional transcript implicated in dilated cardiomyopathy*

The steroid receptor RNA activator 1 (SRA1) gene generates both steroid receptor RNA activator protein (SRAP) as well as several noncoding SRA transcripts, depending on alternative transcription start site usage and alternative splicing (Cooper et al, 2011; Colley & Leedman, 2011). SRA1 non-coding transcripts act as co-activators of nuclear receptor signaling
in a ligand-dependent manner and are involved in regulating skeletal muscle differentiation by co-regulation of the muscle development gene MyoD (Caretti et al, 2006). The identification of genome wide significant SNPs coupled with linkage disequilibrium mapping implicated three co-segregating genes including HBEFG, IK, and SRA1 as determinants of human dilated cardiomyopathy (Friedrichs et al, 2009). Consistent with this finding, depletion of any one of these three genes in zebrafish led to contractile defects in the animals. While there is a clear function for the SRA1 protein in several cellular processes, the contributions of the alternatively spliced non-coding transcripts must be independently addressed, as a function for the putative IncRNA has not been established.

Natural antisense transcripts in cardiac disease

Natural antisense transcripts (NATs) are non-coding RNAs transcribed on the opposite strand of a given protein or non-coding gene and often partially overlap with its exon sequence distinguishing this class from lincRNAs (see Figure 1A). While antisense transcription is a widely employed mechanism for regulating gene expression in eukaryotes from plants, to fungi, to mammals (Derrien et al, 2012; Zhang et al, 2012), we are only beginning to understand the functions of these transcripts. NATs often, but not always, regulate the expression of their corresponding sense RNA and may employ different molecular mechanisms to do so. This mode of regulation is particularly important given that gene dosage is critical for proper heart development and function. In some cases, the act of transcription rather than the IncRNA itself may be necessary to exert the functional consequences. Here, we consider the following examples of NAT IncRNAs that have been implicated in aspects of cardiac disease.

The INK4/ARF locus comprises three tumor suppressor genes; INK4A, ARF, and INK4B, that have important roles in cell cycle regulation. The INK4 locus is subject to Polycomb-mediated regulation under normal conditions, however, how Polycomb complexes are recruited to this locus was not known. This question is of particular interest because expression of these genes is disrupted in many human cancers. ANRIL (antisense noncoding RNA in the INK4 locus, also P15 antisense RNA or CDKN2B antisense RNA) is expressed from the opposite strand and antisense to INK4B. Notably, ANRIL appears to interact with SUZ12, a core subunit of the PRC2 complex and with the CBX7 component of the PRC1 complex and mediates epigenetic silencing of INK4 in cis (Kotake et al, 2011; Yap et al, 2010) (Figure 3C). ANRIL is expressed in immune cells, smooth muscle cells, and endothelium. A risk haplotype is associated with the region encompassing ANRIL for coronary disease, stroke, type 2 diabetes, as well as some
cancers (Burd et al., 2010; Pasmant et al., 2011). In fact, some of the SNPs identified by GWAS appear to affect splicing of ANIRL transcripts (Burd et al., 2010). While it is not yet clear how regulation of INK4a/INK4b or its antisense transcript ANRIL contributes to risk of cardiovascular disease, isoforms containing exons proximal to the INK4/ARF locus correlated with disease risk. Thus, understanding how ANRIL is regulated under normal and disease conditions will be an important next step toward understanding its function in controlling the expression of critical genes in the INK4 locus.

Congenital heart disease as well as certain types of adult cardiac diseases can result from defects in the structural components of the heart, which are important for contraction and conduction functions. Alpha- and beta cardiac myosin heavy chains (MYH6 and MYH7, respectively) are part of the contractile machinery of the cardiac sarcomere. Notably, the ratio of MYH6 to MYH7 expression may constitute a developmentally regulated switch that correlates with heart maturation and cardiac performance (Miyata et al., 2000; Pandya & Smithies, 2011). For example, in mouse MYH7 levels are higher in the fetal heart whereas the MYH6/MYH7 ratio is higher in adult cardiomyocytes. Interestingly, the proportion of these two genes is also regulated by certain pathophysiological stress conditions that lead to hypertrophy where the ratio is more similar to the fetal heart (Hang et al., 2010; Pandya et al., 2008). Moreover, mutations in MYH6 and MYH7 are both associated with hypertrophic cardiomyopathy (Granados-Riveron et al., 2010). The Myh6 and Myh7 genes are juxtaposed in the genome in a head to tail manner, and the switch in their levels is partly regulated by expression of Myh7 antisense RNA in mouse (Haddad et al., 2003) (Figure 3D). The antisense IncRNA is transcribed across the Myh7 locus from the opposite strand and its expression negatively correlates with MYH7 protein abundance. One model is that antisense transcription may block elongation of Myh7 by RNA Polymerase II as has recently been described for the Airn IncRNA (Latos et al., 2012). Thus, the act of transcription may be important while the transcript itself has no function. Notably, a corresponding syntenic non-coding transcript appears to exist in the human genome (although the ratio of MYH6/MYH7 is opposite in human) suggesting that antisense transcription also has a conserved role in regulating this switch and that aberrant expression or mutation of the antisense transcript is also associated with hypertrophic conditions in human disease. Thus, it will be of interest to carefully dissect the role of MYH7 antisense transcription in heart development and cardiac hypertrophy in response to stress or injury.
The atrial myosin light chain gene (Alc1), which is important for sarcomere function, seems to be regulated by its antisense transcript at the translational level by forming RNA duplexes. Notably, Alc1-antisense is increased in hypertrophic ventricles in patients with Tetralogy of Fallot, a form of congenital heart disease, resulting in low ALC1 protein levels (Ritter et al., 1999). However, a direct role for Alc1-antisense transcription in regulating ALC1 levels requires further experimental validation. Cardiac troponin I (cTNI) is also essential for normal sarcomere function in the adult cardiomyocyte and its expression also appears to be regulated at the translational level by formation of cTNI sense-antisense duplexes (Podlowski et al., 2002). While elevated cTNI levels are correlated with ischemia and risk of heart failure, it should be noted that the role of the antisense transcript in disease has not yet been evaluated. In both of these cases, antisense transcripts may be important for regulating gene expression through formation of RNA duplexes that function as substrates for recruitment of factors that degrade the mRNA or by physically blocking translation of the message (Figure 3E). Notably, mRNA and lncRNA molecules in the cytoplasm can form imperfect base-pairs through regions of homology such as ALU sequences, a common repeat element found in the human genome. This type of interaction triggers messenger mediated decay through recruitment of the dsRNA binding protein STAU1 (Gong & Maquat, 2011). Alternatively, mRNA-lncRNA interactions can stabilize the expression of protein-coding transcripts. For example, the lncRNA TINCR interacts with a range of mRNAs important for human epidermal differentiation through a 25 base pair motif call the TINCR box (Kretz et al., 2013). In this case, TINCR-STAU1 interactions mediate stabilization of the message rather than decay. Consistent with this idea, other key proteins required for mRNA decay (i.e. UPF1 and UPF2) do not appear to play a role. Thus, it is possible that in some cases antisense transcripts contain regions of homology by overlapping with exon sequences of the corresponding mRNA to mediate post-transcriptional regulation of a protein-coding gene. Given the potential mechanisms of action of antisense lncRNAs, manipulation of sense transcripts associated with cardiac disease may be particularly amenable to therapeutic intervention by small transcripts such as siRNAs or antisense oligos.

While the discovery of lncRNAs in cardiac biology is only in its infancy, these examples provide a rationale for undertaking broad investigations to identify additional lncRNAs with roles in the cardiovascular system. One theme that is emerging is that the expression levels of lncRNAs must be tightly controlled. Similar to examples represented here, aberrant expression of several lncRNAs has been implicated in disease pathogenesis such as cancer progression. For
example, increased HOTAIR levels correlate with metastatic potential in a range of cancers including those of the breast, prostate, and pancreas (Wapinski & Chang, 2011). To this end, mutations or aberrant expression of IncRNAs is expected to reveal new disease pathways and possible therapeutic targets. Clearly, the link between IncRNAs and cardiac disease susceptibility and pathogenesis will require intensive efforts. Even in the case of cardiac-associated IncRNAs that display no or limited function, these transcripts could also represent a new class of biomarkers for diagnostics based on their specific expression patterns in normal and pathological conditions such as response to injury or stress.

Discovering new IncRNAs with functions in cardiac biology

Given the mounting evidence that IncRNAs play key roles in many cellular processes including cellular differentiation, it is intriguing that this class of transcripts has largely eluded identification in classical genetic screens. This result might be explained by the fact that genomes of model organisms generally used for genetic screens appear to contain fewer IncRNAs compared to mice and humans (see Figure 2E), and that these transcripts are poorly conserved across species. Moreover, until recently, IncRNAs have been poorly annotated so such hits in genetic screens would have been discarded in many cases. Furthermore, mammals display high levels of genetic redundancy, which often masks mutational phenotypes and may also apply in the case of IncRNAs. While IncRNAs are generally poorly conserved among species, functional redundancy may arise from transcripts that appear unrelated on a sequence level since similar secondary structures can be achieved by different combinations of nucleotides. Because of the lack of conservation, randomly occurring mutations may also be less likely to affect IncRNA function than mutations in protein-coding genes, making it more difficult to identify mutants in loss of function genetic screens. Also, given the significant structural differences in the heart between vertebrates and non-vertebrate species, there may be key limitations to using conventional genetic screens in lower eukaryotes for identifying heart-associated IncRNAs. Nevertheless, more recent targeted approaches using RNAi to specifically deplete IncRNAs have been successfully used to screen for specific phenotypes of a large number of these non-coding transcripts (Guttman et al, 2011; Chakraborty et al, 2012).

The overall low sequence conservation of IncRNAs between species suggests that IncRNAs are an extremely fast evolving family of regulatory molecules (Cabili et al, 2011; Ulitsky et al, 2011; Derrien et al, 2012). Relatively few IncRNAs have sequence homologues in other species; however, their promoter sequences are generally conserved significantly more than their exonic
sequences (Chodroff et al, 2010; Ulitsky et al, 2011), suggesting that at least in some cases the tissue-specific expression of a given conserved lncRNA is important. High-throughput sequencing efforts across different tissues, developmental stages, and pathological conditions are expected to reveal potential new regulators of heart development and disease. For example, lncRNAs that show the same stage and development specific expression patterns in a given species may regulate similar processes. Specifically, lncRNAs whose expression patterns cluster with other regulators of heart development (i.e. guilt by association) such as key transcription factors and signaling molecules may reveal classes of lncRNAs with functions in particular biological pathways. Furthermore, comparing transcriptional profiles of similar developmental time points among mammals may also reveal lncRNAs with conserved functions despite a lack of sequence conservation.

Given that lncRNAs appear to be regulated by sets of transcription factors similarly to mRNA genes, analysis of transcriptional programs in loss of function studies may also contribute to the identification of lncRNAs with particular roles that are downstream of key transcription factor genes. Alternatively, lncRNAs whose expression changes dramatically in response to environmental or developmental cues might reveal new effector lncRNAs. For example, induced expression of MesP1, a master regulator of cardiac commitment, leads to the concomitant up-regulation of several lncRNAs during cardiomyocyte differentiation in mouse including Fendrr (Klattenhoff, Scheuermann, and Boyer, unpublished), suggesting that Fendrr is a downstream effector of MESP1.

Along these lines, it will be of interest to analyze those lncRNAs that show changes in expression under stress conditions or in diseased hearts in human patients. Comparisons of lncRNA expression patterns may reveal new biomarkers for cardiac disease since lncRNAs show considerable changes in expression in some human cancers and in neurological function and diseases (Tsai et al, 2011; Ng et al, 2013; Mattick, 2011). In some cases, functional analysis of candidate transcripts involved in cardiac disease identified through profiling of human tissues may be possible by reprogramming patient-specific cells to induced pluripotent stem cells (iPS) followed by in vitro differentiation into cardiac lineages. This method has been used to successfully to model complex neurological and metabolic diseases (Bellin et al, 2012). Together these studies, combined with the wealth of GWAS data available for cardiac-related diseases, will likely lead to the identification of additional lncRNAs that represent new loci for studying disease pathology as well as novel targets for therapeutic intervention.
Dissecting how lncRNAs exert their functions on a molecular level will require precise *in vivo* models as well as *in vitro* cell differentiation systems that will enable biochemical and biophysical studies. Given the low level of sequence conservation and protein-coding potential, it is possible that some lncRNA function may depend on specific secondary structures. In contrast, unstructured regions may be important for interaction with other nucleic acids and this function may be critical for targeting lncRNAs to specific genomic sites as has recently been suggested for *Fendrr* (Grote *et al.*, 2013). To test this notion and to determine the function of individual RNA domains, experimental analyses of secondary structures within lncRNAs across different species by physical and chemical methods will be necessary. These studies will provide templates for mutational analyses and for studying the interactions between lncRNAs and proteins or other nucleic acids. It is also possible that more detailed functional studies will facilitate the design of small molecule drugs for therapeutic intervention.

**Conclusions**

While several thousands of putative lncRNAs have been identified in mammals, only relatively few have been studied in any detail. Thus, we anticipate that during the next few years the field will witness the discoveries of many more developmental, cellular, and molecular processes that are regulated by lncRNAs. The identification of lncRNAs in cardiac biology will open new doors to dissecting the complex gene regulatory mechanisms that drive organogenesis as well as tissue homeostasis, and for understanding how failure to properly regulate developmental gene expression programs can lead to cardiac-related diseases. Knowledge of the genetic basis of heart development and cardiac-related diseases is of substantial value to the medical field and can lead to better genetic tests for disease susceptibility and identification of candidate genes for therapeutic interventions. Given their emerging roles, we also expect that lncRNAs will feature prominently in devising new strategies for stem cell-based regenerative therapies.

**Acknowledgements**

We thank members of the Boyer lab especially Gizem Rizki for insightful discussions. J.C.S. is supported by an EMBO long-term fellowship. L.A.B. is a Pew Scholar in the Biomedical Sciences. This work was also supported in part by an Innovative Research Grant 13IRG14680029 from the American Heart Association (L.A.B.).
References


Castel SE & Martienssen RA (2013) RNA interference in the nucleus: roles for small RNAs in


expressed in the heart precursor cells and required for the formation of a single heart tube.
Development 126: 3437–3447

Development 131: 2533–2541

Seyres D, Röder L & Perrin L (2012) Genes and networks regulating cardiac development and

Sigova AA, Mullen AC, Molinie B, Gupta S, Orlando DA, Guenther MG, Almada AE, Lin C,
Sharp PA, Giallourakis CC & Young RA (2013) Divergent transcription of long noncoding

120: 2498–2506

Spitale RC, Tsai M-C & Chang HY (2011) RNA templating the epigenome: long noncoding
RNAs as molecular scaffolds. Epigenetics 6: 539–543

Rev Pathol 1: 199–213

Srivastava D (2006b) Making or breaking the heart: from lineage determination to
morphogenesis. Cell 126: 1037–1048

eukaryotic complexity. Bioessays 29: 288–299

Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, Watt AT, Freier SM, Bennett CF, Sharma A,
RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation.
Mol. Cell 39: 925–938

329: 689–693

progression. Cancer Res. 71: 3–7

between a noncoding exon and introns: Gomafu contains tandem UACUAAC repeats and
associates with splicing factor-1. Genes Cells 16: 479–490

(2005) Identification of putative noncoding polyadenylated transcripts in Drosophila

Utitzky I, Shkumatava A, Jan CH, Sive H & Bartel DP (2011) Conserved function of lincRNAs in
vertebrate embryonic development despite rapid sequence evolution. Cell 147: 1537–1550


Figure legends

Figure 1. LncRNAs are a heterogeneous class of transcripts and function to regulate gene expression by diverse mechanisms. A. Representative classes of long noncoding RNAs based on genomic location. LncRNAs can be located and transcribed within introns of protein-coding genes (left), as intervening genes known as long intergenic or intervening noncoding RNAs (lincRNAs) that do not overlap with the exons of other genes (middle), or they can be located on the opposite strand of a coding or noncoding gene and transcribed in the antisense direction (right). B. Global mechanisms of lncRNA function. LncRNAs can function as molecular scaffolds by interacting with proteins such as transcription factors or components of chromatin modifying complexes to affect positive or negative regulation of gene expression (left
Proposed mechanisms of action include targeting proteins to specific genomic sites such as promoter regions by complementary interactions with DNA. Alternatively, interactions with DNA may prevent the binding of specific factors to the DNA template (middle). LncRNAs can also base-pair with other RNA molecules such as mRNAs or may act as a sponge for miRNAs.

This scenario is thought to lead to post-transcriptional gene silencing (right panel).

**Figure 2. The developmental complexity of heart development varies among organisms despite conserved core cardiac TF network.**

**A.** Drosophila heart development begins during embryonic stage 11 by specification of two contralateral rows of cardiogenic mesoderm and formation of cardioblasts. Cardioblasts migrate towards the midline at stage 13-14 and form a simple linear closed tube with a central lumen by stage 16-17, subsequently differentiating into more mature cardiomyocytes. **B.** The first steps of mammalian heart development proceed in a very similar manner, yet the mature heart is considerably more complex with two atrial and two ventricular chambers, connecting the systemic and pulmonary circuits via four valves and in- and outgoing vessels. The earliest step of mammalian cardiogenesis involve the bilateral specification of cardiac progenitor cell populations from the first heart field (FHF) in the anterior lateral plate mesoderm, which condense into two lateral heart primordia (mouse E7.5, human day 15) to form the cardiac crescent. The secondary heart field (SHF) constitutes a separate cell population at the medial sides of the two processes of the cardiac crescent. The two processes of the cardiac crescent fuse to form a beating primitive linear heart tube (mouse E8.5, human day 21), which then undergoes rightward looping, resulting in formation of the early chambers (mouse E9, human day 28). During later stages, the mature shape of the heart is generated by differentiation of cardiac cell populations and extensive remodeling of the heart, resulting in four-chambered heart with distinct in- and outflow tracts, cardiac valves separating the different compartments, and a mature conduction system. **C.** The core transcription factor network necessary for specification of the cardiovascular lineages are conserved between *Drosophila* and mammals. **D.** The percentage of noncoding to protein-coding sequence increases with developmental complexity. Whereas *S. cerevisiae* dedicates most of its genome to protein-coding genes, only a small fraction of the genome codes for proteins in human. **E.** The total number of putative lncRNA transcripts is predicted to be significantly higher in mouse (~3,000 lincRNA transcripts as determined by Ponjavic *et al*, 2007; Sigova *et al*, 2013) and Human (~15,000 as defined by Derrien *et al*, 2012) as compared to lower eukaryotes such as *Drosophila* (17 based on stringent criteria in Tupy *et al.*, 2005 to greater than 1,000 based on
low stringency estimates from Young et al, 2012), *C. elegans* (262 from Nam & Bartel, 2012), and zebrafish (~700 transcripts predicted from Ultisky et al, 2011 and Pauli et al, 2012). The number of lncRNAs varies among studies as different criteria were used to define lncRNA transcripts.

Figure 3. Mechanisms of lncRNA function in heart development and cardiac disease.
A. *Braveheart* is necessary for commitment to the cardiac lineage in mouse. *Bvht* appears to function in *trans* through interaction with the epigenetic silencing complex PRC2 and may act as a decoy to antagonize its recruitment to key developmental genes during cardiomyocyte differentiation. Alternatively, *Bvht* may recruit PRC2 to gene(s) that repress the cardiac program. In either case, loss of *Bvht* leads to a failure to activate the core cardiac gene network that includes many TFs implicated in heart development and disease. B. *Fendrr* is expressed in the lateral plate mesoderm in mouse from which precursors for the heart and body wall are derived. *Fendrr* is proposed to function partly in *cis* to regulate its neighboring gene Foxf1a. *Fendrr* also functions in *trans* to regulate the expression of additional genes important for heart development. *Fendrr* interacted with PRC2 components as well as WDR5, a member of TrxG/MLL complex suggesting that *Fendrr* regulates the balance between repressive and activating marks at key genes during development. Thus, *Bvht* and *Fendrr* may represent examples of lncRNAs that regulate gene expression through epigenetic mechanisms. C-E. lncRNAs can also function as natural antisense transcripts (NATs) to affect gene expression at the transcriptional and post-transcriptional level. *ANRIL* was identified as a risk factor for coronary disease by GWAS. *ANRIL* is expressed in the opposite direction to INK4B/P15 in the INK4 locus. The antisense transcript appears to recruit PRC1 and PRC2 to mediate repression of the INK4a/INK4b tumor suppressor locus through an epigenetic silencing mechanism (C). The ratio of two important sarcomere components MYH6 and MYH7 vary during development and in stress-induced pathological conditions. Myh6 and Myh7 genes are juxtaposed in the mouse genome in a head to tail fashion. An antisense lncRNA (*Myh7-as*) is transcribed across the Myh7 locus and negatively correlates with MYH7 abundance. Thus, *Myh7-as* transcription may regulate the ratio of *Myh6* and *Myh7* (D). Some antisense transcripts are predicted to form RNA duplexes with their mRNA counterpart leading to post-transcriptional regulation of the target message. For example, antisense transcripts to Alc1 and cNTI, two genes that code for important sarcomere components in cardiac muscle, form RNA duplexes with the respective protein-coding transcript. *Alc1*-antisense is increased in hypertrophic ventricles in patients with
Tetralogy of Fallot, whereas elevated cTNI levels are correlated with ischemia and risk of heart failure. In both of these cases, antisense transcripts may be important for regulating gene expression through formation of RNA duplexes that are substrates for recruitment of factors that degrade the mRNA or that physically block translation of the message. RNA-RNA interactions can also stabilize the mRNA in some cases (E).
Figure 1.
Figure 2.
Figure 3.