MIT Open Access Articles

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

Review

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

Johanna C. Scheuermann[1] and Laurie A. Boyer[1][*]

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. lncRNA 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. LncRNAs 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, lncRNAs are thought to possess tremendous regulatory potential.

In general, lncRNAs 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 lncRNAs 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 lncRNA and its target gene (Orom *et al*, 2010; Lai *et al*, 2013). Evidence also indicates that a subset of lncRNAs 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 lncRNAs, interactions with RNA binding proteins including Polycomb and Trithorax group members have been widely conserved between mouse and human (Guttman & Rinn, 2012). Alternatively, some lncRNAs 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). LncRNAs 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, lncRNAs have roles in transcriptional and post-transcriptional gene regulatory events. Consequently, the biological functions of lncRNAs 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 lncRNAs 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, lncRNAs 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, lncRNAs 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 lncRNAs and disease as well as their tissue-specific expression patterns indicate that lncRNAs 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 lncRNAs 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 lncRNA regulation in heart development and cardiac-related disease. Here, we discuss the newly emerging roles of lncRNAs and cite specific examples in the context of heart development and cardiac disease as well as present considerations for the identification of additional lncRNA 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 lncRNAs 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 lncRNAs transcripts increases dramatically from *Drosophila* to human (Figure 2E). For example, while mice and humans have thousands of putative lncRNA 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 lncRNA 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 lncRNAs 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, lncRNAs 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 lncRNAs are expressed dynamically at specific developmental stages during cardiomyocyte differentiation (Wamstad *et al*, 2012). Indeed, two recent studies identified lncRNAs 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 lncRNAs 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 *Nkx*2.5, among others. In contrast, mesoderm markers such as *Brachyury* 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 lncRNA 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 lncRNAs have been implicated in adult cardiac disease by analysis of genetic variation among individuals with cardiac traits. Other examples of lncRNAs 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 lncRNA 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 lncRNAs 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 lncRNA 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 lncRNA itself may be necessary to exert the functional consequences. Here, we consider the following examples of NAT lncRNAs 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 lncRNA 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* lncRNA (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 *Alc*1-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 lincRNA 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 lncRNAs is expected to reveal new disease pathways and possible therapeutic targets. Clearly, the link between lncRNAs and cardiac disease susceptibility and pathogenesis will require intensive efforts. Even in the case of cardiac-associated lncRNAs 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 lncRNAs with functions in cardiac biology
Given the mounting evidence that lncRNAs 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 lncRNAs compared to mice and humans (see Figure 2E), and that these transcripts are poorly conserved across species. Moreover, until recently, lncRNAs 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 lncRNAs. While lncRNAs 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 lncRNA 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 lncRNAs. Nevertheless, more recent targeted approaches using RNAi to specifically deplete lncRNAs 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 lncRNAs between species suggests that lncRNAs are an extremely fast evolving family of regulatory molecules (Cabili *et al*, 2011; Ulitsky *et al*, 2011; Derrien *et al*, 2012). Relatively few lncRNAs have sequence homologues in other species; however, their promoter sequences are generally conserved significantly more than their exonic

15

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

References

Batista PJ & Chang HY (2013) Long Noncoding RNAs: Cellular Address Codes in Development and Disease. Cell 152: 1298–1307

Bánfai B, Jia H, Khatun J, Wood E, Risk B, Gundling WE, Kundaje A, Gunawardena HP, Yu Y, Xie L, Krajewski K, Strahl BD, Chen X, Bickel P, Giddings MC, Brown JB & Lipovich L (2012) Long noncoding RNAs are rarely translated in two human cell lines. Genome Res. 22: 1646–1657

Bellin M, Marchetto MC, Gage FH & Mummery CL (2012) Induced pluripotent stem cells: the new patient? Nat. Rev. Mol. Cell Biol. 13: 713–726

Bondue A, Lapouge G, Paulissen C, Semeraro C, Iacovino M, Kyba M & Blanpain C (2008) Mesp1 acts as a master regulator of multipotent cardiovascular progenitor specification. Cell Stem Cell 3: 69–84

Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, Lee TI, Levine SS, Wernig M, Tajonar A, Ray MK, Bell GW, Otte AP, Vidal M, Gifford DK, Young RA & Jaenisch R (2006) Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature 441: 349–353

Bruneau BG (2008) The developmental genetics of congenital heart disease. Nature 451: 943–948

Bruneau BG (2013) Signaling and transcriptional networks in heart development and regeneration. Cold Spring Harb Perspect Biol 5: a008292

Burd CE, Jeck WR, Liu Y, Sanoff HK, Wang Z & Sharpless NE (2010) Expression of linear and novel circular forms of an INK4/ARF-associated non-coding RNA correlates with atherosclerosis risk. PLoS Genet. 6: e1001233

Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A & Rinn JL (2011) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. 25: 1915–1927

Caretti G, Schiltz RL, Dilworth FJ, Di Padova M, Zhao P, Ogryzko V, Fuller-Pace FV, Hoffman EP, Tapscott SJ & Sartorelli V (2006) The RNA helicases p68/p72 and the noncoding RNA SRA are coregulators of MyoD and skeletal muscle differentiation. Dev. Cell 11: 547–560

Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, Ravasi T, Lenhard B, Wells C, Kodzius R, Shimokawa K, Bajic VB, Brenner SE, Batalov S, Forrest ARR, Zavolan M, Davis MJ, Wilming LG, Aidinis V, et al (2005) The transcriptional landscape of the mammalian genome. Science 309: 1559–1563

Carthew RW & Sontheimer EJ (2009) Origins and Mechanisms of miRNAs and siRNAs. Cell 136: 642–655

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

transcription, epigenetics and beyond. Nat. Rev. Genet. 14: 100–112

Cesana M, Cacchiarelli D, Legnini I, Santini T, Sthandier O, Chinappi M, Tramontano A & Bozzoni I (2011) A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. Cell 147: 358–369

Chakraborty D, Kappei D, Theis M, Nitzsche A, Ding L, Paszkowski-Rogacz M, Surendranath V, Berger N, Schulz H, Saar K, Hubner N, & Buchholz F (2012) Combined RNAi and localization for functionally dissecting long noncoding RNAs. Nat Methods 9: 360-362

Chang C-P & Bruneau BG (2012) Epigenetics and cardiovascular development. Annu. Rev. Physiol. 74: 41–68

Chodroff RA, Goodstadt L, Sirey TM, Oliver PL, Davies KE, Green ED, Molnár Z & Ponting CP (2010) Long noncoding RNA genes: conservation of sequence and brain expression among diverse amniotes. Genome Biol. 11: R72

Colley SM & Leedman PJ (2011) Steroid Receptor RNA Activator - A nuclear receptor coregulator with multiple partners: Insights and challenges. Biochimie 93: 1966–1972

Cooper C, Vincett D, Yan Y, Hamedani MK, Myal Y & Leygue E (2011) Steroid Receptor RNA Activator bi-faceted genetic system: Heads or Tails? Biochimie 93: 1973–1980

Cordes KR & Srivastava D (2009) MicroRNA regulation of cardiovascular development. Circ. Res. 104: 724–732

Costello I, Pimeisl I-M, Dräger S, Bikoff EK, Robertson EJ & Arnold SJ (2011) The T-box transcription factor Eomesodermin acts upstream of Mesp1 to specify cardiac mesoderm during mouse gastrulation. Nat. Cell Biol. 13: 1084–1091

David R, Brenner C, Stieber J, Schwarz F, Brunner S, Vollmer M, Mentele E, Müller-Höcker J, Kitajima S, Lickert H, Rupp R & Franz W-M (2008) MesP1 drives vertebrate cardiovascular differentiation through Dkk-1-mediated blockade of Wnt-signalling. Nat. Cell Biol. 10: 338–345

David R, Jarsch VB, Schwarz F, Nathan P, Gegg M, Lickert H & Franz W-M (2011) Induction of MesP1 by Brachyury(T) generates the common multipotent cardiovascular stem cell. Cardiovasc. Res. 92: 115–122

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, Lagarde J, Veeravalli L, Ruan X, Ruan Y, Lassmann T, Carninci P, Brown JB, Lipovich L, Gonzalez JM, Thomas M, et al (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. 22: 1775–1789

ENCODE Project Consortium, Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, Epstein CB, Frietze S, Harrow J, Kaul R, Khatun J, Lajoie BR, Landt SG, Lee B-K, Pauli F, Rosenbloom KR, Sabo P, Safi A, Sanyal A, et al (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489: 57–74

Friedrichs F, Zugck C, Rauch G-J, Ivandic B, Weichenhan D, Müller-Bardorff M, Meder B, Mokhtari El NE, Regitz-Zagrosek V, Hetzer R, Schäfer A, Schreiber S, Chen J, Neuhaus I, Ji R,

Siemers NO, Frey N, Rottbauer W, Katus HA & Stoll M (2009) HBEGF, SRA1, and IK: Three cosegregating genes as determinants of cardiomyopathy. Genome Res. 19: 395–403

Ghildiyal M & Zamore PD (2009) Small silencing RNAs: an expanding universe. Nat. Rev. Genet. 10: 94–108

Gong C & Maquat LE (2011) lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. Nature 470: 284–288

Granados-Riveron JT, Ghosh TK, Pope M, Bu'Lock F, Thornborough C, Eason J, Kirk EP, Fatkin D, Feneley MP, Harvey RP, Armour JAL & David Brook J (2010) Alpha-cardiac myosin heavy chain (MYH6) mutations affecting myofibril formation are associated with congenital heart defects. Hum. Mol. Genet. 19: 4007–4016

Grote P, Wittler L, Hendrix D, Koch F, Währisch S, Beisaw A, Macura K, Bläss G, Kellis M, Werber M & Herrmann BG (2013) The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. Dev. Cell 24: 206–214

Guttman M & Rinn JL (2012) Modular regulatory principles of large non-coding RNAs. Nature 482: 339–346

Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, Young G, Lucas AB, Ach R, Bruhn L, Yang X, Amit I, Meissner A, Regev A, Rinn JL, Root DE & Lander ES (2011) lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature 477: 295–300

Haddad F, Bodell PW, Qin AX, Giger JM & Baldwin KM (2003) Role of antisense RNA in coordinating cardiac myosin heavy chain gene switching. J. Biol. Chem. 278: 37132–37138

Hang CT, Yang J, Han P, Cheng H-L, Shang C, Ashley E, Zhou B & Chang C-P (2010) Chromatin regulation by Brg1 underlies heart muscle development and disease. Nature 466: 62–67

Harvey RP (2002) Patterning the vertebrate heart. Nat. Rev. Genet. 3: 544–556

Hu W, Yuan B, Flygare J & Lodish HF (2011) Long noncoding RNA-mediated anti-apoptotic activity in murine erythroid terminal differentiation. Genes Dev. 25: 2573–2578

Ishii N, Ozaki K, Sato H, Mizuno H, Saito S, Takahashi A, Miyamoto Y, Ikegawa S, Kamatani N, Hori M, Saito S, Nakamura Y & Tanaka T (2006) Identification of a novel non-coding RNA, MIAT, that confers risk of myocardial infarction. J. Hum. Genet. 51: 1087–1099

Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, Stadler PF, Hertel J,

Hackermüller J, Hofacker IL, Bell I, Cheung E, Drenkow J, Dumais E, Patel S, Helt G,

Ganesh M, Ghosh S, Piccolboni A, Sementchenko V, et al (2007) RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science 316: 1484–1488

Kathiresan S & Srivastava D (2012) Genetics of human cardiovascular disease. Cell 148: 1242–1257

Kattman SJ, Adler ED & Keller GM (2007) Specification of multipotential cardiovascular

progenitor cells during embryonic stem cell differentiation and embryonic development. Trends Cardiovasc. Med. 17: 240–246

Kattman SJ, Witty AD, Gagliardi M, Dubois NC, Niapour M, Hotta A, Ellis J & Keller G (2011) Stage-specific optimization of activin/nodal and BMP signaling promotes cardiac differentiation of mouse and human pluripotent stem cell lines. Cell Stem Cell 8: 228–240

Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, Thomas K, Presser A, Bernstein BE, van Oudenaarden A, Regev A, Lander ES & Rinn JL (2009) Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc. Natl. Acad. Sci. U.S.A. 106: 11667–11672

Klattenhoff CA, Scheuermann JC, Surface LE, Bradley RK, Fields PA, Steinhauser ML, Ding H, Butty VL, Torrey L, Haas S, Abo R, Tabebordbar M, Lee RT, Burge CB & Boyer LA (2013) Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. Cell 152: 570–583

Kotake Y, Nakagawa T, Kitagawa K, Suzuki S, Liu N, Kitagawa M & Xiong Y (2011) Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. Oncogene 30: 1956–1962

Kretz M, Siprashvili Z, Chu C, Webster DE, Zehnder A, Qu K, Lee CS, Flockhart RJ, Groff AF, Chow J, Johnston D, Kim GE, Spitale RC, Flynn RA, Zheng GXY, Aiyer S, Raj A, Rinn JL, Chang HY & Khavari PA (2013) Control of somatic tissue differentiation by the long non-coding RNA TINCR. Nature 493: 231–235

Kretz M, Webster DE, Flockhart RJ, Lee CS, Zehnder A, Lopez-Pajares V, Qu K, Zheng GXY, Chow J, Kim GE, Rinn JL, Chang HY, Siprashvili Z & Khavari PA (2012) Suppression of progenitor differentiation requires the long noncoding RNA ANCR. Genes Dev. 26: 338–343

Kriegmair MCM, Frenz S, Dusl M, Franz W-M, David R & Rupp RAW (2013) Cardiac differentiation in Xenopus is initiated by mespa. Cardiovasc. Res. 97: 454–463

Lai F, Orom UA, Cesaroni M, Beringer M, Taatjes DJ, Blobel GA & Shiekhattar R (2013) Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature 494: 497–501

Latos PA, Pauler FM, Koerner MV, Şenergin HB, Hudson QJ, Stocsits RR, Allhoff W, Stricker SH, Klement RM, Warczok KE, Aumayr K, Pasierbek P & Barlow DP (2012) Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. Science 338: 1469–1472

Lee JT & Bartolomei MS (2013) X-inactivation, imprinting, and long noncoding RNAs in health and disease. Cell 152: 1308–1323

Lin MF, Jungreis I & Kellis M (2011) PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. Bioinformatics 27: i275–82

Lindsley RC, Gill JG, Murphy TL, Langer EM, Cai M, Mashayekhi M, Wang W, Niwa N, Nerbonne JM, Kyba M & Murphy KM (2008) Mesp1 coordinately regulates cardiovascular fate restriction and epithelial-mesenchymal transition in differentiating ESCs. Cell Stem Cell 3: 55–68

Liu N & Olson EN (2010) MicroRNA regulatory networks in cardiovascular development. Dev. Cell 18: 510–525

Malone CD & Hannon GJ (2009) Small RNAs as guardians of the genome. Cell 136: 656–668

Matera AG, Terns RM & Terns MP (2007) Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. Nat. Rev. Mol. Cell Biol. 8: 209–220

Mattick JS (2011) The central role of RNA in human development and cognition. FEBS Lett. 585: 1600–1616

Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, Shafer A, Neri F, Lee K, Kutyavin T, Stehling-Sun S, Johnson AK, Canfield TK, Giste E, Diegel M, Bates D, et al (2012) Systematic localization of common disease-associated variation in regulatory DNA. Science 337: 1190–1195

McCulley DJ & Black BL (2012) Transcription factor pathways and congenital heart disease. Curr. Top. Dev. Biol. 100: 253–277

Medioni C, Sénatore S, Salmand P-A, Lalevée N, Perrin L & Sémériva M (2009) The fabulous destiny of the Drosophila heart. Curr. Opin. Genet. Dev. 19: 518–525

Mercer TR & Mattick JS (2013) Structure and function of long noncoding RNAs in epigenetic regulation. Nat. Struct. Mol. Biol. 20: 300–307

Miyata S, Minobe W, Bristow MR & Leinwand LA (2000) Myosin heavy chain isoform expression in the failing and nonfailing human heart. Circ. Res. 86: 386–390

Nam J-W & Bartel DP (2012) Long noncoding RNAs in C. elegans. Genome Res. 22: 2529–2540

Ng S-Y, Johnson R & Stanton LW (2012) Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. EMBO J. 31: 522–533

Ng S-Y, Lin L, Soh BS & Stanton LW (2013) Long noncoding RNAs in development and disease of the central nervous system. Trends Genet.

Nguyen HT & Frasch M (2006) MicroRNAs in muscle differentiation: lessons from Drosophila and beyond. Curr. Opin. Genet. Dev. 16: 533–539

Olson EN (2006) Gene regulatory networks in the evolution and development of the heart. Science 313: 1922–1927

Orom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, Lai F, Zytnicki M, Notredame C, Huang Q, Guigó R & Shiekhattar R (2010) Long noncoding RNAs with enhancer-like function in human cells. Cell 143: 46–58

Pandya K & Smithies O (2011) β-MyHC and cardiac hypertrophy: size does matter. Circ. Res. 109: 609–610

Pandya K, Cowhig J, Brackhan J, Kim HS, Hagaman J, Rojas M, Carter CW, Mao L, Rockman

HA, Maeda N & Smithies O (2008) Discordant on/off switching of gene expression in myocytes during cardiac hypertrophy in vivo. Proc. Natl. Acad. Sci. U.S.A. 105: 13063–13068

Pasmant E, Sabbagh A, Vidaud M & Bièche I (2011) ANRIL, a long, noncoding RNA, is an unexpected major hotspot in GWAS. FASEB J. 25: 444–448

Pauli A, Valen E, Lin MF, Garber M, Vastenhouw NL, Levin JZ, Fan L, Sandelin A, Rinn JL, Regev A & Schier AF (2012) Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. Genome Res. 22: 577–591

Podlowski S, Bramlage P, Baumann G, Morano I & Luther HP (2002) Cardiac troponin I sense-antisense RNA duplexes in the myocardium. J. Cell. Biochem. 85: 198–207

Ponjavic J, Ponting CP & Lunter G (2007) Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. Genome Res. 17: 556–565

Ponting CP, Oliver PL & Reik W (2009) Evolution and functions of long noncoding RNAs. Cell 136: 629–641

Qian L & Bodmer R (2012) Probing the polygenic basis of cardiomyopathies in Drosophila. J. Cell. Mol. Med. 16: 972–977

Qian L, Wythe JD, Liu J, Cartry J, Vogler G, Mohapatra B, Otway RT, Huang Y, King IN, Maillet M, Zheng Y, Crawley T, Taghli-Lamallem O, Semsarian C, Dunwoodie S, Winlaw D, Harvey RP, Fatkin D, Towbin JA, Molkentin JD, et al (2011) Tinman/Nkx2-5 acts via miR-1 and upstream of Cdc42 to regulate heart function across species. J. Cell Biol. 193: 1181–1196

Rapicavoli NA, Poth EM & Blackshaw S (2010) The long noncoding RNA RNCR2 directs mouse retinal cell specification. BMC Dev. Biol. 10: 49

Reim I & Frasch M (2010) Genetic and genomic dissection of cardiogenesis in the Drosophila model. Pediatr Cardiol 31: 325–334

Rinn JL & Chang HY (2012) Genome regulation by long noncoding RNAs. Annu. Rev. Biochem. 81: 145–166

Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E & Chang HY (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 129: 1311–1323

Ritter O, Haase H, Schulte HD, Lange PE & Morano I (1999) Remodeling of the hypertrophied human myocardium by cardiac bHLH transcription factors. J. Cell. Biochem. 74: 551–561

Saga Y, Hata N, Kobayashi S, Magnuson T, Seldin MF & Taketo MM (1996) MesP1: a novel basic helix-loop-helix protein expressed in the nascent mesodermal cells during mouse gastrulation. Development 122: 2769–2778

Saga Y, Kitajima S & Miyagawa-Tomita S (2000) Mesp1 expression is the earliest sign of cardiovascular development. Trends Cardiovasc. Med. 10: 345–352

Saga Y, Miyagawa-Tomita S, Takagi A, Kitajima S, Miyazaki JI & Inoue T (1999) MesP1 is

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

Satou Y, Imai KS & Satoh N (2004) The ascidian Mesp gene specifies heart precursor cells. Development 131: 2533–2541

Seyres D, Röder L & Perrin L (2012) Genes and networks regulating cardiac development and function in flies: genetic and functional genomic approaches. Brief Funct Genomics 11: 366–374

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 RNA/mRNA gene pairs in embryonic stem cells. Proc. Natl. Acad. Sci. U.S.A. 110: 2876–2881

Sone M, Hayashi T, Tarui H, Agata K, Takeichi M & Nakagawa S (2007) The mRNA-like noncoding RNA Gomafu constitutes a novel nuclear domain in a subset of neurons. J. Cell. Sci. 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

Srivastava D (2006a) Genetic regulation of cardiogenesis and congenital heart disease. Annu Rev Pathol 1: 199–213

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

Taft RJ, Pheasant M & Mattick JS (2007) The relationship between non-protein-coding DNA and eukaryotic complexity. Bioessays 29: 288–299

Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, Watt AT, Freier SM, Bennett CF, Sharma A, Bubulya PA, Blencowe BJ, Prasanth SG & Prasanth KV (2010) The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Mol. Cell 39: 925–938

Tsai M-C, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F, Shi Y, Segal E & Chang HY (2010) Long noncoding RNA as modular scaffold of histone modification complexes. Science 329: 689–693

Tsai M-C, Spitale RC & Chang HY (2011) Long intergenic noncoding RNAs: new links in cancer progression. Cancer Res. 71: 3–7

Tsuiji H, Yoshimoto R, Hasegawa Y, Furuno M, Yoshida M & Nakagawa S (2011) Competition between a noncoding exon and introns: Gomafu contains tandem UACUAAC repeats and associates with splicing factor-1. Genes Cells 16: 479–490

Tupy JL, Bailey AM, Dailey G, Evans-Holm M, Siebel CW, Misra S, Celniker SE & Rubin GM (2005) Identification of putative noncoding polyadenylated transcripts in Drosophila melanogaster. Proc. Natl. Acad. Sci. U.S.A. 102: 5495–5500

Ulitsky 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

Vincent SD & Buckingham ME (2010) How to make a heart: the origin and regulation of cardiac progenitor cells. Curr. Top. Dev. Biol. 90: 1–41

Wamstad JA, Alexander JM, Truty RM, Shrikumar A, Li F, Eilertson KE, Ding H, Wylie JN, Pico AR, Capra JA, Erwin G, Kattman SJ, Keller GM, Srivastava D, Levine SS, Pollard KS, Holloway AK, Boyer LA & Bruneau BG (2012) Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. Cell 151: 206–220

Wang Y, Xu Z, Jiang J, Xu C, Kang J, Xiao L, Wu M, Xiong J, Guo X & Liu H (2013) Endogenous miRNA Sponge lincRNA-RoR Regulates Oct4, Nanog, and Sox2 in Human Embryonic Stem Cell Self-Renewal. Dev. Cell 25: 69–80

Wapinski O & Chang HY (2011) Long noncoding RNAs and human disease. Trends Cell Biol. 21: 354–361

Yap KL, Li S, Muñoz-Cabello AM, Raguz S, Zeng L, Mujtaba S, Gil J, Walsh MJ & Zhou M-M (2010) Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. Mol. Cell 38: 662–674

Yoon J-H, Abdelmohsen K, Srikantan S, Yang X, Martindale JL, De S, Huarte M, Zhan M, Becker KG & Gorospe M (2012) LincRNA-p21 suppresses target mRNA translation. Mol. Cell 47: 648–655

Young RS, Marques AC, Tibbit C, Haerty W, Bassett AR, Liu J-L & Ponting CP (2012) Identification and properties of 1,119 candidate lincRNA loci in the Drosophila melanogaster genome. Genome Biol Evol 4: 427–442

Zhang X, Xia J, Lii YE, Barrera-Figueroa BE, Zhou X, Gao S, Lu L, Niu D, Chen Z, Leung C, Wong T, Zhang H, Guo J, Li Y, Liu R, Liang W, Zhu J-K, Zhang W & Jin H (2012) Genome-wide analysis of plant nat-siRNAs reveals insights into their distribution, biogenesis and function. Genome Biol. 13: R20

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

panel). 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)**.
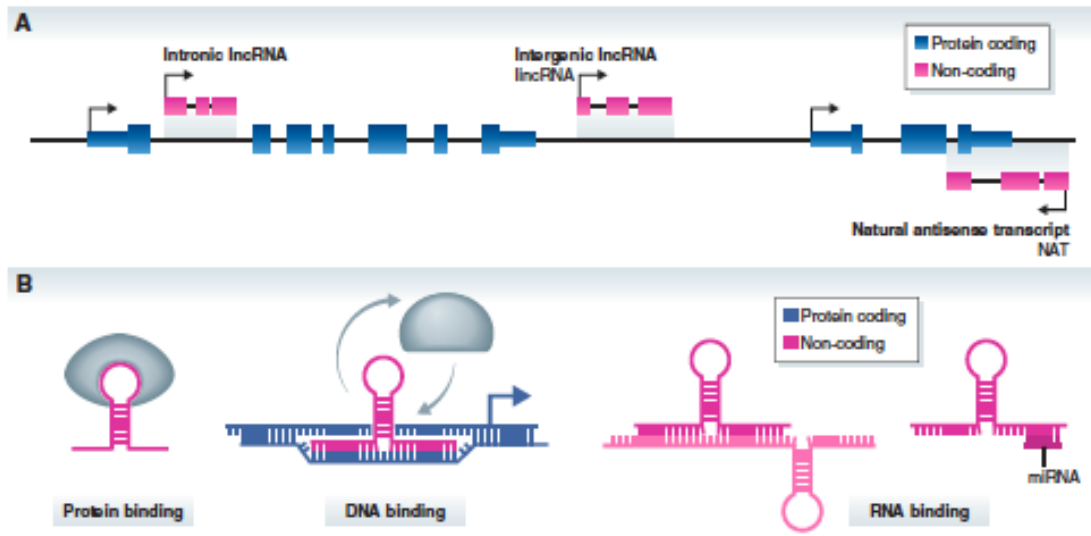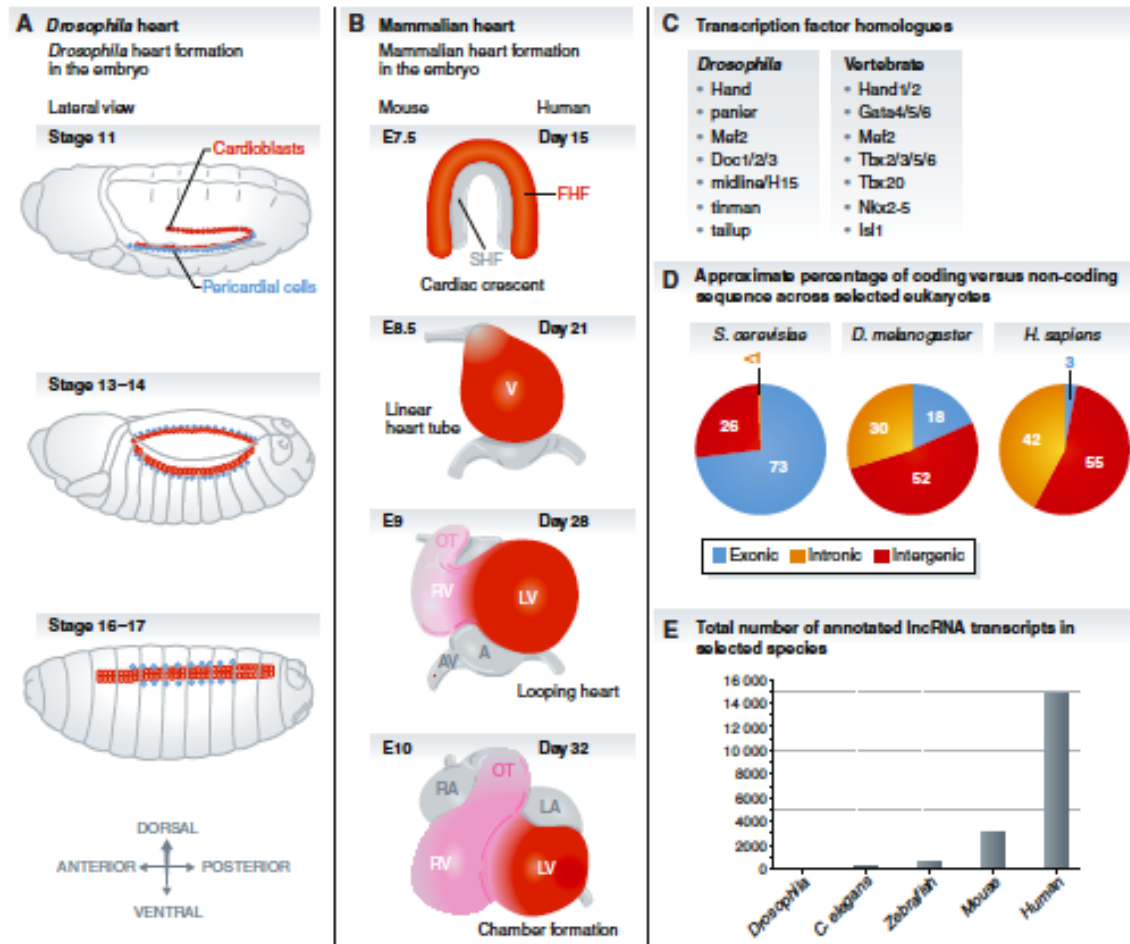
Figure 1.

Figure 2.

A *Drosophila* heart

Drosophila heart formation in the embryo

Lateral view

Stage 11

Cardioblasts
Pericardial cells

Stage 13–14

Stage 16–17

DORSAL
ANTERIOR ← → POSTERIOR
VENTRAL

B Mammalian heart

Mammalian heart formation in the embryo

Mouse — Human

E7.5 — Day 15
FHF
SHF
Cardiac crescent

E8.5 — Day 21
V
Linear heart tube

E9 — Day 28
OT
RV
LV
AV A
Looping heart

E10 — Day 32
OT
RA
LA
RV
LV
Chamber formation

C Transcription factor homologues

Drosophila	Vertebrate
• Hand	• Hand1/2
• panier	• Gata4/5/6
• Mef2	• Mef2
• Doc1/2/3	• Tbx2/3/5/6
• midline/H15	• Tbx20
• tinman	• Nkx2-5
• tailup	• Isl1

D Approximate percentage of coding versus non-coding sequence across selected eukaryotes

S. cerevisiae — *D. melanogaster* — *H. sapiens*

<1
26
73

30
18
52

3
42
55

■ Exonic ■ Intronic ■ Intergenic

E Total number of annotated lncRNA transcripts in selected species

16 000
14 000
12 000
10 000
8000
6000
4000
2000
0

Drosophila | *C. elegans* | Zebrafish | Mouse | Human

Figure 3.