New Lncs to mesendoderm specification

Michael Alexanian, Thierry Pedrazzini, Samir Ounzain

Experimental Cardiology Unit, Department of Medicine, University of Lausanne Medical School, Lausanne, Switzerland

Contributions: (I) Conception and design: All authors; (II) Administrative support: None; (III) Provision of study materials or patients: None; (IV) Collection and assembly of data: None; (V) Data analysis and interpretation: None; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Dr. Samir Ounzain. Experimental Cardiology Unit, Department of Medicine, University of Lausanne Medical School, CH-1011 Lausanne, Switzerland. Email: samir.ounzain@chuv.ch.

Abstract: Mammalian genomes are pervasively transcribed generating thousands of long noncoding RNAs (lncRNAs) with emergent regulatory roles. Many of these lncRNAs exhibit highly specialised expression patterns during development and typically flank and regulate key developmental factors. In this review, we discuss and summarise the latest advances in our understanding of the roles of lncRNAs during mesendoderm (ME) specification, a key step during gastrulation and the formation of the primitive streak (PS).

Keywords: Long noncoding RNAs (lncRNAs)

Translating embryology to embryonic stem cells (ESCs): mesendoderm (ME) specification

A number of diseases that represent a major cause of morbidity and mortality in the developed world are, at their root, diseases of cellular deficiency. Most of the cells that build our tissues and organs share a common origin during the early events of embryogenesis: they all derive from a transient structure known as the primitive streak (PS). The specification of the distinct PS populations is tightly regulated in a spatial and temporal manner, giving rise to both endoderm and mesoderm cells. ESCs have the potential to produce any of the three germ layer-derived cells and represent an ideal and relevant *in vitro* system to recapitulate the early embryological events. During their *in vitro* differentiation into mesoderm and endoderm, ESCs transition an intermediate stage called ME, which is equivalent to the PS (1-3). The gene regulatory networks (GRNs) that dictate ME specification are under the control of distinct combinations of transcription factors (TFs) that are necessary to elicit cell fate and lineage determination. Emerging evidence suggests that GRN activity is, in addition to TFs, under the control of complex regulatory non-coding RNA networks. A better understanding of the non-coding RNA networks implicated in ME specification is needed in order to provide important insights into the regulatory mechanisms operating in the early embryo and provide novel approaches to improve stem cell commitment for regenerative medicine.

Long noncoding RNAs (lncRNAs)

Among all the classes of non-coding RNAs, lncRNAs are operationally defined as RNA molecules that are longer than 200 nucleotides in length and exhibit minimal protein coding potential. Currently, the best characterized lncRNAs are typically polymerase II (Pol II) transcribed, multi-exonic, 5' capped, alternatively spliced and polyadenylated. LncRNAs are usually expressed at a lower level than mRNAs but importantly are characterized by a greater tissue and cell specificity (4-7). Recent studies have shown that lncRNAs regulate various biological processes such as chromosome inactivation (8), imprinting (9), splicing (10) and transcriptional control (11). LncRNAs represent a very dynamic class of molecules that are emerging as key regulators of every aspect of GRN activity, including

transcriptional control, post-transcriptional processing and chromatin remodeling (12). Recently, lncRNAs have been shown to be involved in the fine-tuning of GRNs controlling cell-fate and lineage specification. Here, we review the lncRNAs that have been implicated in regulating ME specification *in vitro* and/or *in vivo*.

lncRNAs in ME specification

The first lncRNA shown to be required for controlling ME specification, specifically into the cardiac fate, was *Braveheart* (*Bvht*) (13). Depletion of *Bvht* in mouse ESCs decreases their ability to activate a core GRN directing cardiovascular cell fate. This lncRNA acts upstream of the TF mesoderm posterior protein 1 (MESP1), master regulator of cardiac mesoderm specification. *Bvht* interacts with the epigenetic silencer polycomb repressive complex 2 (PCR2), possibly displacing PRC2 from the *MesP1* locus to activate cardiac mesoderm-specific genes. Very recently, the mechanism of action of *Bvht* has been characterized in detail. The secondary structure of this lncRNA was determined and the 5' asymmetric G-rich internal loop (AGIL) was shown to be essential for *Bvht* function (14). Indeed, mouse ESCs carrying a specific deletion of AGIL showed impaired cardiogenic differentiation. Importantly, Bvht interacts through AGIL with CNBP, a zing finger TF which acts as negative regulator of the cardiac gene program. Our group has characterized a lncRNA associated with the same locus controlling *Bvht* in mouse. This novel lncRNA called *CARMEN*, (Car)diac (M)esoderm (E)nhancer-associated (N)oncoding RNA, was identified as one of the most up-regulated lncRNAs during the cardiac differentiation of human cardiac progenitor cells (CPCs) isolated from the fetal human heart (15). As opposed to *Bvht*, *Carmen* is conserved in human and is associated to an active super-enhancer in both the fetal and human heart. *Carmen* loss of function (LoF) by shRNAi in mouse P19CL6s resulted in a decreased ability of the cells to differentiate into the cardiac lineage. Interestingly, *Carmen* depletion resulted in the down-regulation of *Eomesodermin* (*Eomes*), a crucial TF known to dictate mesoderm and definitive endoderm (DE) specification. Thus, *Carmen* acts upstream of *Eomes* in the GRNs driving early ME specification.

LncRNA that are transcribed on the opposite strand to their neighboring protein coding gene are named divergent lncRNAs and comprise ~20% of the total lncRNAs in the mammalian genome. The first example of a divergent lncRNA implicated in early development was *Fendrr*, transcribed divergently from the TF forkhead box F1 (FOXF) (16). *Fendrr* is specifically expressed in the lateral plate mesoderm of the developing embryo and interacts with both activating (TrxG/MLL) and repressive (PCR2) chromatin modifying complexes to modulate expression of specific TFs implicated in the development of the heart. *Fendrr* KO resulted in embryonic lethality at embryonic day 13.75 with abdominal wall defects and pooling of blood in the right atrium. An important feature of *Fendrr* is its long-term effect. Epigenetic signatures, like the ones established via *Fendrr* activity, persist through several stages of differentiation. *Fendrr* is therefore a modulator of the epigenetic landscape during early development. The crosstalk between *Fendrr* and *Foxf1* represents an example of an intriguing transcriptional and functional coupling of a divergent lncRNA with its adjacent PCG, both of which are essential for development of the same embryonic tissue. Another example of a divergent lncRNA implicated in early ME specification is *Evx1as*, transcribed on the opposite strand of its nearby gene Even-Skipped Homeobox 1 (*Evx1*) (17). EVX1 is an important homeodomain TF that promotes ME specification during mouse ESC differentiation and its expression is highly correlated with its divergent lncRNA *Evx1as*. Both *Evx1as* genomic deletion and post-transcriptional knockdown significantly repressed *Evx1* expression. To evaluate whether *Evx1as* induced *Evx1* transcription via a *cis*-regulatory mechanism, the authors artificially activated *Evx1as* expression using a CRISPR-ON system. Consistent with a cis-regulatory mechanism, activation of *Evx1as* induced *Evx1* expression. Finally, using an elegant dCas9-mediated approach to tether the *Evx1as* RNA to the promoter of *Evx1as/Evx1*, the authors demonstrated a direct role for *Evx1as* RNA in controlling *Evx1* transcription through *cis* regulation. Very interestingly, the depletion of *Evx1* had a lesser impact on ME specification than the depletion of its divergent lncRNA, suggesting the possibility that *Evx1as* acts also in *trans* independently of *Evx1*. *DIGIT* represents an alternative interesting antisense lncRNA, divergent to *Goosecoid* (*GSC*), an important regulator of ME specification. Knock-down of *DIGIT* in human ESCs impacted upon DE differentiation and expression of important developmental protein coding genes including *SRY-Box 17* (*SOX17*) and *Forkhead Box A2* (*FOXA2*). *DIGIT* has an ortholog in mouse expressed in early development at E6.5, a stage corresponding to the formation of the anterior PS. Importantly, *Digit* expression peaks at E.7.5, when the DE is formed. Depletion of *DIGIT* via post-transcriptional approaches (shRNAs and LNAs)

and the insertion of a poly(A) termination sequences in human ESCs demonstrated that *DIGIT* RNA rather than the act of transcription at the *DIGIT* locus is essential for *GSC* expression. Importantly, ectopic re-expression of the *DIGIT* transcript in *DIGIT*-deficient cells induced *GSC* expression, as well as expression of other ME markers such as *FOXA2* and *SOX17*. This suggests that *DIGIT* regulates *GSC* expression via a *trans*-regulatory mechanism as it does not need to be transcribed in close proximity to *GSC* to elicit its regulatory role.

Another lncRNA implicated in early development is *Alien*. *Alien* starts to be expressed at the PS stage, showing maximal expression in the lateral plate mesoderm of E8.5 mouse embryos (18). This lncRNA is proximal to *Foxa2* and is conserved in human. *Alien* expression correlates with crucial ME genes such as *Brachyury* (*T*) and *Mix Paired-Like Homeobox* (*Mixl1*). LoF of *Alien* is associated with severe defects in multiple anatomic structures, revealing its importance during ME specification. In another study, the same lncRNA (on this occasion named *DEANR1*) was shown to be an important regulator of DE specification. Depletion of *DEANR1* in human ESCs affects endoderm differentiation but is not required for maintaining the stemness features of ESCs. Interestingly, *DEANR1* regulates *FOXA2* expression by facilitating SMAD2/3 binding at the *FOXA2* promoter.

Deng and colleagues identified and characterized *HoxBlinc*, a lncRNA transcribed in the Homeobox B (*hoxb*) gene locus upstream of the *hoxb4* gene (19). *HoxBlinc* depletion during mouse ESCs differentiation affects *hoxb* genes and other crucial regulators of ME specification such as *Eomes* and *Mesp1*. Using RNA immune precipitation (RIP) and chromatin isolation by RNA purification (ChIRP), the authors demonstrated that the *HoxBlinc* transcript binds to the promoters of the *hoxb1-hoxb6* genes and recruits Set1/MLL1 complexes, driving H3K4me3 to the promoters of the *hoxb* genes. Importantly, *HoxBlinc* RNA regulates the chromatin looping of the *hoxb* gene cluster, locking the *hoxb* locus in a permissive conformation for being actively transcribed and drive ME specification into the hematopoietic lineages.

Conclusions

The assumption that the GRNs controlling cell fate specification were primarily protein-based regulatory systems has been somewhat premature. We are now beginning to appreciate and decipher the complex roles of lncRNAs in contributing to the regulation of GRNs dictating ME specification. In the near future, we envisage that the manipulation of this new class of regulatory molecules may facilitate the development of more efficient cell differentiation protocols for regenerative medicine. Finally, understanding the roles of lncRNAs in development and cell differentiation is crucial to develop therapeutic approaches to treat human diseases for which lncRNA dysregulation is reported.

Acknowledgements

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

1. Yu P, Pan G, Yu J, et al. FGF2 sustains NANOG and switches the outcome of BMP4-induced human embryonic stem cell differentiation. Cell Stem Cell 2011;8:326-34.

2. Gouon-Evans V, Boussemart L, Gadue P, et al. BMP-4 is required for hepatic specification of mouse embryonic stem cell-derived definitive endoderm. Nat Biotechnol 2006;24:1402-11.

3. Kunisada Y, Tsubooka-Yamazoe N, Shoji M, et al. Small molecules induce efficient differentiation into insulin-producing cells from human induced pluripotent stem cells. Stem Cell Res 2012;8:274-84.

4. Cabili MN, Trapnell C, Goff L, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev 2011;25:1915-27.

5. Derrien T, Johnson R, Bussotti G, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res 2012;22:1775-89.

6. Morán I, Akerman I, van de Bunt M, et al. Human beta cell transcriptome analysis uncovers lncRNAs that are tissue-specific, dynamically regulated, and abnormally expressed in type 2 diabetes. Cell Metab 2012;16:435-48.

7. Ramos AD, Diaz A, Nellore A, et al. Integration of genome-wide approaches identifies lncRNAs of adult neural stem cells and their progeny in vivo. Cell Stem Cell 2013;12:616-28.

8. Zhao J, Sun BK, Erwin JA, et al. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science 2008;322:750-6.

9. Latos PA, Pauler FM, Koerner MV, et al. Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. Science 2012;338:1469-72.

10. Tripathi V, Ellis JD, Shen Z, et al. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Mol cell 2010;39:925-38.

11. Tsai MC, Manor O, Wan Y, et al. Long noncoding RNA as modular scaffold of histone modification complexes. Science 2010;329:689-93.

12. Ounzain S, Crippa S, Pedrazzini T. Small and long non-coding RNAs in cardiac homeostasis and regeneration. Biochim Biophys Acta 2013;1833:923-33.

13. Klattenhoff CA, Scheuermann JC, Surface LE, et al. Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. Cell 2013;152:570-83.

14. Xue Z, Hennelly S, Doyle B, et al. A G-Rich Motif in the lncRNA Braveheart Interacts with a Zinc-Finger Transcription Factor to Specify the Cardiovascular Lineage. Mol Cell 2016;64:37-50.

15. Ounzain S, Micheletti R, Arnan C, et al. CARMEN, a human super enhancer-associated long noncoding RNA controlling cardiac specification, differentiation and homeostasis. J Mol Cell Cardiol 2015;89:98-112.

16. Grote P, Wittler L, Hendrix D, et al. The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. Dev Cell 2013;24:206-14.

17. Luo S, Lu JY, Liu L, et al. Divergent lncRNAs Regulate Gene Expression and Lineage Differentiation in Pluripotent Cells. Cell Stem Cell 2016;18:637-52.

18. Kurian L, Aguirre A, Sancho-Martinez I, et al. Identification of novel long noncoding RNAs underlying vertebrate cardiovascular development. Circulation 2015;131:1278-90.

19. Deng C, Li Y, Zhou L, et al. HoxBlinc RNA Recruits Set1/MLL Complexes to Activate Hox Gene Expression Patterns and Mesoderm Lineage Development. Cell Rep 2016;14:103-14.