Long noncoding RNAs in hematopoiesis [version 1; peer review: 2 approved]

Xu Zhang, Wenqian Hu
Department of Biochemistry and Molecular Biology, Mayo Clinic, Rochester, MN, 55905, USA

Abstract
Mammalian development is under tight control to ensure precise gene expression. Recent studies reveal a new layer of regulation of gene expression mediated by long noncoding RNAs. These transcripts are longer than 200nt that do not have functional protein coding capacity. Interestingly, many of these long noncoding RNAs are expressed with high specificity in different types of cells, tissues, and developmental stages in mammals, suggesting that they may have functional roles in diverse biological processes. Here, we summarize recent findings of long noncoding RNAs in hematopoiesis, which is one of the best-characterized mammalian cell differentiation processes. Then we provide our own perspectives on future studies of long noncoding RNAs in this field.

Keywords
hematopoiesis, noncoding RNAs, gene expression
Introduction

One of the exciting findings from genomic studies over the past decade is the identification of a large number of long noncoding RNAs (lncRNAs) in mammalian transcriptomes. These transcripts are longer than 200 nucleotides in length, and though structurally similar to mRNAs, lncRNAs do not have functional protein-coding capacity. In addition, lncRNAs are usually expressed at lower levels compared with mRNAs, and most lncRNAs are not evolutionarily conserved in primary sequences. Interestingly, transcriptomic studies revealed that many lncRNAs are expressed in a manner highly specific to cell types, tissues, developmental stages, and pathological conditions in mammals, suggesting that these transcripts are involved in these biological processes. Consistent with this, the biological functions of many lncRNAs are being characterized in diverse biological and pathological settings. Mechanistically quite different from small RNAs, such as micro-RNAs that control gene expression predominantly at one particular step (mRNA translation and degradation), lncRNAs can regulate gene expression at multiple levels, from transcription and splicing to mRNA translation and degradation, in mammals. There are several outstanding reviews summarizing our current understandings of the mechanistic aspects of lncRNA-mediated regulation of gene expression and the hematopoiesis-related lncRNAs. Here, we focus on biological functions of lncRNAs in hematopoiesis.

For three reasons, the hematopoietic system is one of the best paradigms for studying cellular lineage determination and differentiation in mammals. First, hematopoiesis is well characterized at the cellular level in both humans and mice, and a single hematopoietic stem cell (HSC) can re-constitute all the cells in the hematopoietic system. Based on this, many lineage specification and differentiation processes of HSCs and their progenies are characterized. Second, almost all the multi-potential and lineage-determined progenitor cells in hematopoiesis can be enriched or isolated in an almost pure form by flow cytometry using combinations of cell surface markers. This makes the phenotypic studies of a specific cellular lineage relatively easy. Third, both in vitro and in vivo assays are well established for functional characterization of hematopoietic progenitor cells. Collectively, these features of the hematopoietic system greatly facilitate the exploration of both proteins and RNAs controlling cell fate determination and differentiation.

Here, we summarize recent studies on the biological functions of lncRNAs in hematopoiesis. Particularly, we focus on lncRNAs in the regulation of HSCs and differentiation of several major hematopoietic cell lineages. Then we provide our own perspectives on future studies of lncRNAs in hematopoiesis.

Long noncoding RNAs in hematopoietic stem cell maintenance and differentiation

HSCs are the best-characterized somatic stem cells in mammals. Recent observations suggest that lncRNAs play regulatory roles in HSC biology. For example, epigenetic controlling allele-specific expression of H19, an lncRNA involved in imprinting, is critical for maintaining “stemness” of HSCs. Specifically, using a mouse model with maternally-specific deletion of the differentially methylated region upstream of H19, a critical cis-element in controlling imprinting, the Li group found that the expression of H19 target Igf2 is de-repressed, which results in reduced adult HSC quiescence and compromise of HSC function. Although in this mouse model the H19 locus is not deleted, removing the critical cis-element inhibits the expression of H19.

### Table 1. Examples of long noncoding RNAs in hematopoiesis.

<table>
<thead>
<tr>
<th>Long noncoding RNA</th>
<th>Cell types</th>
<th>Assay</th>
<th>Observed phenotype</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>H19</td>
<td>HSC</td>
<td>Deletion of the upstream region of H19</td>
<td>Reduced adult HSC quiescence and compromise of HSC function</td>
<td>21</td>
</tr>
<tr>
<td>LncHSC-1/2</td>
<td>HSC</td>
<td>shRNA knockdown</td>
<td>Impacted HSC self-renewal and differentiation</td>
<td>23</td>
</tr>
<tr>
<td>lincRNA-EPS</td>
<td>Erythroid progenitors</td>
<td>shRNA knockdown and overexpression</td>
<td>Essential for maturation of red blood cells</td>
<td>26</td>
</tr>
<tr>
<td>alncRNA-EC7</td>
<td>Erythroid</td>
<td>shRNA knockdown</td>
<td>Depletion of BAND 3</td>
<td>24</td>
</tr>
<tr>
<td>HOTAIRM1</td>
<td>Myeloid progenitors</td>
<td>shRNA knockdown</td>
<td>Attenuates the maturation of granulocytes</td>
<td>29, 31</td>
</tr>
<tr>
<td>Eosinophil granule ontogeny</td>
<td>Eosinophils</td>
<td>shRNA knockdown</td>
<td>Compromises the expression of several proteins that are important for eosinophil development</td>
<td>32</td>
</tr>
<tr>
<td>Inc-DC</td>
<td>DCs</td>
<td>RNA interference and overexpression</td>
<td>Vital for DC differentiation in both humans and mice</td>
<td>42</td>
</tr>
<tr>
<td>linc-MAF-4</td>
<td>Th1 cell</td>
<td>siRNA-mediated knockdown</td>
<td>Increases the expression of MAF and skews T-cell differentiation toward the Th2 phenotype</td>
<td>40</td>
</tr>
</tbody>
</table>

DC, dendritic cell; HOTAIRM1, HOX antisense intergenic RNA myeloid 1; HSC, hematopoietic stem cell; shRNA, short hairpin RNA; siRNA, small interfering RNA.
also give rise to a microRNA named miR-675, which also represses the expression of the insulin-like growth factor 1 receptor (Igf1r). Thus, inhibiting H19 expression may have pleiotropic effects on mouse development. Nonetheless, the results from this genetically modified mouse indicate that proper epigenetic modifications of the lncRNA H19 locus are important for maintaining HSCs. To comprehensively identify and annotate lncRNAs specifically expressed in HSCs, the Goodell lab performed comprehensive transcriptomic profiling on highly purified mouse HSCs followed by a series of stringent filtering steps to exclude the protein-coding genes and the lowly expressed transcripts. By comparing the transcriptome of HSCs with those of lineage-committed hematopoietic cells, they identified 159 lncRNAs that are enriched in HSCs. Interestingly, short hairpin RNA (shRNA) knockdown of two of these HSC-enriched lncRNAs compromises HSC self-renewal and lineage commitment. Although shRNA-mediated approaches may have potential off-target effects, the results from this study argue that some of these HSC-enriched lncRNAs are functionally important for HSC biology. This study provides a solid foundation for future functional and mechanistic characterization of lncRNAs in HSCs.

### Long noncoding RNAs in erythropoiesis and megakaryopoiesis

Erythroid cells and megakaryocytes are thought to derive from a common progenitor named megakaryocyte-erythroid progenitor (MEP). Using mouse fetal liver erythroblasts, two groups independently performed RNA sequencing (RNA-seq) analysis on erythroid cells at different developmental stages. Computational analysis on these transcriptomic datasets revealed that about 650 lncRNAs are dynamically expressed during erythropoiesis. Interestingly, by using shRNA-mediated loss-of-function studies on lncRNAs that are significantly induced (more than threefold) during terminal erythroid differentiation, both of these groups show that several erythroid-specific lncRNAs have regulatory roles in generating enucleated reticulocytes, at least in vitro, arguing that these noncoding transcripts are new regulators in erythropoiesis. For example, LincRNA-EPS, an lncRNA specific to terminal differentiating erythroblasts, prevents cellular apoptosis during terminal erythroid differentiation by inhibiting the expression of Pycard, a pro-apoptotic gene. In addition, by integrating the transcriptomic datasets with the chromatin immunoprecipitation-sequencing datasets, they found that many erythroid-specific lncRNAs are regulated by Gata1 and Tal1, two erythroid-important transcription factors. These observations collectively indicate that lncRNAs are targets of transcriptional regulatory networks in erythroid cell differentiation.

In addition to surveying mouse erythropoiesis, several groups surveyed the transcriptomic dynamics in human erythroid cell differentiation by RNA-seq. Interestingly, comparing lncRNAs expressed in mouse erythroid cells with those from human erythroid cells, Paralkar et al. observed that for most lncRNAs, including those that are functionally involved in mouse erythropoiesis, they are poorly conserved in both primary sequence and expression pattern between mouse and human. This observation suggests that lncRNAs are under different evolutionary pressures compared with that of protein-coding genes.

Paralkar et al. also performed RNA-seq analysis in mouse megakaryocytes and MEP cells. Computational analysis on transcriptomic datasets from erythroid cells, megakaryocytes, and MEPs indicates that many lncRNAs are highly specific to one cell type versus the other two. Thus, it would be very interesting to identify and characterize megakaryocyte-specific lncRNAs, as these noncoding transcripts may potentially regulate megakaryopoiesis. Similarly, analyzing the expression patterns of those MEP-specific lncRNAs during differentiation and loss-of-function and gain-of-function studies of these lncRNAs may reveal how these transcripts modulate the biology of the MEP bi-potential progenitor cells, such as self-renewal and lineage specification.

### Long noncoding RNAs in myelopoiesis and innate immunity

The myeloid leukocytes, including eosinophilic granulocytes, basophilic granulocytes, neutrophilic granulocytes, and monocyes, are thought to derive from a common progenitor named granulocyte-macrophage progenitor in the hematopoietic system. Several observations indicate that lncRNAs are involved in modulating myelopoiesis. For example, a human myeloid lineage-specific lncRNA, HOX antisense intergenic RNA myeloid 1 (HOTAIRM1), is upregulated during granulocytic differentiation of myeloid progenitors. The genomic locus of this lncRNA is located within the HOXA loci, which encode several HOXA transcription factors that are important for myelopoiesis. Interestingly, shRNA knockdown of this lncRNA compromises the activation of HOXA1 and HOXA4 during myeloid differentiation and attenuates the maturation of granulocytes as determined by cell surface markers. HOTAIRM1 knockdown is further shown to influence cell cycle arrest at the G1/S transition to inhibit granulocytic maturation in NB4 cells. This observation strongly suggests that lncRNA-mediated regulation of HOXA gene expression is important during granulocyte differentiation. LncRNAs are also implicated in regulating eosinophil formation during myelopoiesis. An lncRNA, with primary sequence conserved among human, mouse, and chicken, named eosinophil granule ontology (EGO) was identified from the inositol triphosphate receptor type 1 gene locus. This lncRNA is highly expressed in mature eosinophils, and biochemical experiments indicate that this transcript is not associated with ribosomes and does not have conserved open reading frames (ORFs), strongly arguing that it is noncoding. Interestingly, shRNA-mediated loss-of-function studies revealed that reduction of EGO level compromises the expression of several proteins that are important for eosinophil development, suggesting the functional importance of EGO in eosinophilopoiesis. It would be very interesting to explore how this conserved lncRNA regulates gene expression.

The terminal differentiated effector cells, such as macrophages, from myelopoiesis play important roles in innate immunity. Interestingly, several groups observed that many lncRNAs are differentially expressed when macrophages are challenged by bacteria infection. The Fitzgerald group functionally characterized one such lncRNA named lincRNA-Cox2. This lncRNA is upregulated more than 20 fold by the Toll-like receptor signaling pathway in mouse bone marrow-derived macrophages. Importantly, biochemical analysis indicates that lincRNA-Cox2 is not
associated with ribosomes. Detailed functional and molecular mechanistic studies revealed that lncRNA-Cox2 interacts with heterogeneous nuclear ribonucleoproteins A/B and A2/B2 to modulate the expression of several immune response genes during inflammatory signaling.

Long noncoding RNAs in lymphoid lineages

Lymphopoiesis is the regulated generation of lymphoid cells, including T cells, B cells, natural killer cells, and dendritic cells (DCs) in the hematopoietic system. Transcriptomic studies from numerous groups identified a large number of lymphoid-specific lncRNAs that are differentially expressed during the maturation of these lymphoid cells\textsuperscript{11,20,37}. For example, during CD8\textsuperscript{+} lymphocyte differentiation, hundreds of lymphoid-specific lncRNAs were identified\textsuperscript{11}. More recently, comprehensive transcriptomic surveillances in human lymphoid cells identified over 3,000 IncRNAs, and many of these transcripts are highly lineage-specific or developmental stage-specific or both\textsuperscript{36}. These observations suggest that lncRNAs may play important roles in lymphoid cells. Indeed, the biological functions of several IncRNAs are characterized in these immune cells. For instance, Ranzani et al. identified an IncRNA, Linc-MAF-4, which is Th1 cell-specific and plays important roles in helper T-cell differentiation\textsuperscript{38}. Linc-MAF-4 can recruit chromatin modifiers LSD1 and EZH2 to the promoter to repress the expression of MAF, a Th2-associated transcription factor. Small interfering RNA (siRNA)-mediated knockdown of linc-MAF-4 increases the expression of MAF and skews T-cell differentiation toward the Th2 phenotype. This study reveals the functional importance of lncRNAs in T-cell differentiation \textit{in vitro}. T cells play important roles in modulating immune response during viral and bacterial infection. Interestingly, lncRNAs seem to have regulatory functions in these processes. For example, the IncRNA nettoie Salmonella pas Theiler’s (NeST) is specifically expressed in the mouse Th1 helper T cells. During viral and bacterial infections, this NeST regulates the expression of the cytokine interferon-gamma, thereby modulating the immune response mediated by the Th1 helper T cells\textsuperscript{39,40}. When lncRNA expression in various helper T-cell subsets was profiled, a Th2-specific IncRNA named LincR-Ccr2-5’AS was identified and characterized\textsuperscript{41}. Knockdown of this IncRNA by shRNA compromises the ability of the cells to migrate to the lungs \textit{in vivo}. Interestingly, the global gene expression changes of LincR-Ccr2-5’AS knockdown are similar to those of Gata3 depletion, suggesting that there may be a functional link between the lncRNA and the Gata3 transcription factor during T-cell differentiation and immune function.

In addition to finding T cells, Wang et al. found an lncRNA, Lnc-DC, which seems to be exclusively expressed in human DCs\textsuperscript{42}. This lncRNA is required for optimal DC differentiation. Functionally, this transcript is involved in regulating DC-mediated T-cell activation, and mechanistically, it achieves this by direct binding to the signal transducer and activator of transcription 3 (STAT3) in the cytoplasm, which promotes STAT3 phosphorylation on tyrosine-705, thereby facilitating STAT3 activation. Collectively, these observations indicate that lncRNAs are involved in both lymphoid cell differentiation and modulating lymphoid cell-mediated immune responses.

Future directions

Thanks to the widely used high-throughput transcriptomic profiling techniques, such as RNA-seq, the number of annotated lncRNAs exploded over the past few years. Loss-of-function and gain-of-function studies have revealed and likely will continue to reveal the regulatory roles of lncRNAs in hematopoiesis and other biological processes. Here, we discuss from our own perspectives some of the outstanding challenges and opportunities for future studies of IncRNA in hematopoiesis.

Coding versus noncoding

Most lncRNAs were identified by computational analysis on transcriptomic datasets. These transcripts are annotated as non-coding by bioinformatics approaches. Though powerful, these computational algorithms only predict, but by no means demonstrate, the absence of coding ability/potential of those ORFs present on lncRNAs. Interestingly, however, recent observations indicate that some of the ORFs on lncRNAs are used by the translational apparatus, ribosome, to generate small peptides (reviewed in \textsuperscript{43}). Critically, the small peptides resulting from annotated lncRNAs can be biologically functional\textsuperscript{44-46}. Thus, an outstanding question is how many IncRNAs are truly noncoding and how many of these computationally annotated lncRNAs are actually coding transcripts? One approach to address this question is to test the RNA and ribosome association by using polysome analysis for specific transcripts or ribosome profiling for the whole transcriptome\textsuperscript{47}. Discriminating coding versus noncoding of the computationally annotated IncRNAs will provide important insights into whether the function of the transcript is mediated by the RNA itself or the polypeptides it encodes.

Testing long noncoding RNA functions \textit{in vivo}

Studies from hematopoietic cell lines and cultured primary cells revealed the regulatory roles of several lncRNAs in hematopoiesis. Whether the phenotypes from these loss-of-function studies \textit{in vitro} can be recapitulated \textit{in vivo}, however, is unknown. Thus, it is important to use knockout mouse models to determine the \textit{in vivo} functions of lncRNAs. In addition, knockout mouse models can solve some technical challenges of \textit{in vitro} IncRNA studies. For example, some lncRNAs, particularly those localized in the nucleus, cannot be efficiently knocked down by current shRNA/siRNA methods. Therefore, absence of phenotypes in these loss-of-function studies is uninterpretable. In addition, shRNA/siRNA approaches may have off-target effects. Thus, complete depletion of lncRNA locus using a mouse model can provide unambiguous information for establishing roles of lncRNAs in hematopoiesis.

Importantly, several lncRNA knockout mouse models have been generated to address this important question. Encouragingly, knocking out Xist in female mice, which comprises maintenance of X-inactivation in females, severely inhibits HSC maturation and results in highly aggressive myeloproliferative neoplasm and myelodysplastic syndrome with full penetrance\textsuperscript{48}. Similarly, knocking out Dlue2, an lncRNA frequently deleted in lymphocytic leukemia 2, in mouse results in dys-regulation of cell cycle progression and apoptosis of B cells. Interestingly, the mouse
develops a chronic lymphocytic leukemia (CLL) that is similar to human CLL. These two IncRNA in vivo studies clearly indicate that proper expression of IncRNA is required for normal hematopoiesis. An important issue associated with the IncRNA mouse model is to discriminate whether the phenotype is caused by the RNA or by the DNA sequence that can potentially function as regulatory cis-elements. Bassett et al. provided an insightful discussion of this caveat10. The rapid development and application of targeted genome editing technologies, such as the CRISPR-Cas9 system, greatly facilitate the generation of targeted gene-altered mouse models. We believe that these technical advances will be of great help in functional characterizing IncRNAs in vivo.

Molecular mechanisms of long noncoding RNA-mediated regulation of gene expression

IncRNA can regulate gene expression via diverse mechanisms10–14. One common theme of IncRNA-mediated regulation of gene expression is that IncRNA recruits protein partners to exert its biological effect(s). Although not all protein-binding events on IncRNA will result in functional consequence10, identifying the protein(s) that IncRNAs specifically associated with is an important first step to characterize the molecular mechanisms. To this end, an expanding number of biochemical methods for RNA purification coupled with mass spectrometry have been established11–15. Usually, many proteins that specifically bind target IncRNA can be identified from the mass spectrometry analysis. The next important question is to test the functional importance of the identified RNA-protein interactions. This can be achieved by structure-function mapping to reveal critical regions on the RNA that are important for recruiting the protein partner(s), and then mutant IncRNA can be generated to test the functional consequence of disrupting the RNA-protein interaction. One caveat of the biochemical approaches is that they usually require a large amount of material. This can be challenging for some lowly expressed IncRNAs. Thus, one complementary approach is to use single-molecular RNA fluorescent in situ hybridization to directly visualize IncRNA localization in the cell. This can provide important insights into the molecular mechanisms of IncRNA. For instance, this cell biology approach can determine whether the target IncRNA is localized to functional subcellular compartments, such as paraspeckles that are involved in splicing, certain chromatin regions, and co-localization with candidate proteins. In combination, the biochemical approaches and cell biology approaches can greatly facilitate the characterization of molecular mechanisms of IncRNA-mediated regulation of gene expression in hematopoiesis.

Abbreviations

CLL, chronic lymphocytic leukemia; DC, dendritic cell; EGO, eosinophil granule ontology; HSC, hematopoietic stem cell; IncRNA, long noncoding RNA; MEP, megakaryocyte-erythroid progenitor; NeST, nettoie Salmonella pas Theiler’s; ORF, open reading frame; RNA-seq, RNA sequencing; shRNA, short hairpin RNA; siRNA, small interfering RNA; STAT3, signal transducer and activator of transcription 3.

Competing interests

The authors declare that they have no competing interests.

Grant information

This work was supported by start-up funding from the Mayo Clinic Foundation for Research and Education and a grant from the National Heart, Lung, and Blood Institute (R00HL118157) to Wenqian Hu.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

Open Peer Review

Current Peer Review Status: ✔ ✔

Editorial Note on the Review Process

Faculty Reviews are review articles written by the prestigious Members of Faculty Opinions. The articles are commissioned and peer reviewed before publication to ensure that the final, published version is comprehensive and accessible. The reviewers who approved the final version are listed with their names and affiliations.

The reviewers who approved this article are:

1. Chris P. Ponting
   University of Edinburgh Western General Hospital, Edinburgh, UK
   Competing Interests: No competing interests were disclosed.

2. Vinod Scaria
   CSIR-IGIB, New Dehli, India
   Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com