Long noncoding RNAs in biology and hematopoiesis

Vikram R. Paralkar¹ and Mitchell J. Weiss²*

Running title: LncRNAs in hematopoiesis

1. Division of Hematology/Oncology, The Perelman School of Medicine at The University of Pennsylvania, Philadelphia, Pennsylvania

2. Division of Hematology, The Children’s Hospital of Philadelphia, Philadelphia, PA

*Corresponding author:
Mitchell J. Weiss
The Children’s Hospital of Philadelphia
316B ARC
3615 Civic Center Blvd.
Philadelphia PA, 19104
weissmi@email.chop.edu
phone: 215 590-0565
Abstract

Genome and transcriptome sequencing have revealed a rich assortment of noncoding RNAs in eukaryote cells, including long noncoding RNAs (lncRNAs), which regulate gene expression independent of protein coding potential. LncRNAs modulate protein coding gene expression in many cell types by regulating multiple processes including epigenetic control of transcription, splicing and mRNA stability. While little is known about lncRNAs in hematopoiesis, they are likely to exert widespread roles in this process.
New layers of genetic regulation through non-coding RNAs

A remarkable discovery of whole-genome sequencing studies initiated at the turn of this century was that protein coding exons account for less than 2% of mammalian DNA.1 Subsequent large-scale transcriptome studies have established that at least two-thirds of the genome is nonetheless transcribed into RNA, exposing novel and exciting layers of genetic regulation.2 Throughout eukaryotic evolution, in yeast, nematodes, flies and mammals, genomic regions previously believed to be “junk DNA” are now known to be rich with transcripts encoding a large and diverse population of RNAs without protein-coding potential, referred to as noncoding RNAs (ncRNAs). Studies of these RNAs have provided new insights into the development of virtually all mammalian tissues and opened up new fields of biology. It is now known that one class of ncRNAs, termed microRNAs, exert diverse roles in normal and pathological hematopoiesis.3 The same is likely to be true for long noncoding RNAs (lncRNAs), the topic of this Blood Spotlight review.

Classification of noncoding RNAs

The earliest recognized ncRNAs - ribosomal RNAs and transfer RNAs – comprise more than 95% of cellular RNA content. Other ncRNAs are broadly classified on the basis of length. Short ncRNAs are less than 200 nucleotides and include microRNAs, small nucleolar RNAs, piwi-interacting RNAs, and others. By arbitrary definition, lncRNAs are greater than 200 nucleotides and represent a diverse group with many functions. This class of RNAs was appreciated more than 20 years ago, but is now rapidly gaining prominence as a pervasive player in gene
regulation. Here we provide a succinct summary of lncRNAs for Blood readers, with a focus on potential roles in hematopoiesis. Space limitations prevent us from including many important details and references. For more in-depth information on lncRNAs, we refer readers to several excellent review articles.4 5 6

What are lncRNAs?

The concept that long RNAs can regulate gene expression by non-classical mechanisms has been known for some time. For example, Xist, a lncRNA that mediates X-chromosome inactivation, was discovered in 1991.6 This lncRNA and a handful of others were considered to be relatively infrequent, until large-scale transcriptome sequencing identified thousands more.2 Now, candidate lncRNAs are recognized by computational algorithms that identify transcripts without protein coding potential, as evidenced by their lack of long open reading frames, conserved codons, or homology with protein databases. Follow up studies have revealed that some of these RNAs have noncoding functions mediated through direct interaction with effector proteins, DNA and/or protein coding RNAs.

Long noncoding RNAs range in size from 200 nucleotides to beyond 10 kilobases. They can be polyadenylated or not, nuclear or cytoplasmic, and are generally expressed at lower levels compared to protein coding mRNAs. Like protein-coding genes, many lncRNA genes are bound by essential cell type specific nuclear factors, transcribed from what appear to be conventional promoters, and spliced.7 Other lncRNAs, both polyadenylated and nonpolyadenylated, arise from enhancers.8 It is estimated that many thousands of lncRNAs are encoded in the human genome.
and are expressed in exquisitely tissue-specific patterns. Thus, it is likely that deep sequencing of specific cell types, including normal and diseased blood lineages, will discover new lncRNAs.

**LncRNAs regulate gene expression through diverse mechanisms**

Long noncoding RNAs are versatile molecules that can interact physically and functionally with DNA, other RNAs and proteins, either through nucleotide base pairing or via formation of structural domains generated by RNA folding. These properties endow lncRNAs with a versatile range of capabilities that are only beginning to be appreciated. The best-characterized role of lncRNAs is in epigenetic regulation. For example, in female (XX) cells, one X chromosome produces the prototypical lncRNA Xist, which coats that chromosome and recruits repressive chromatin complexes to condense and silence it in a process termed “Lyonization”. The lncRNA HOTAIR is transcribed from the HOXC locus, and silences the HOXD locus genes by recruiting repressive complexes. LncRNAs were thought initially to act only on neighboring genes (in cis) and HOTAIR was the first lncRNA shown to act on another chromosome (in trans). Numerous cis- and trans-acting lncRNAs have now been described. In addition to their role in recruiting chromatin modifying proteins, lncRNAs have been shown to interfere with protein-DNA binding, organize nuclear architecture, regulate mRNA stability and translation, modulate mRNA levels by competing for microRNA binding, and directly alter protein function (Figure 1). LncRNAs are therefore capable of acting at multiple levels in the hierarchy of gene expression.
LncRNAs regulate normal biological processes, including hematopoiesis

LncRNAs modulate cell survival, division and differentiation to facilitate the development of many tissues. For example: RNAi knockdown of 150 mouse embryonic stem cell lncRNAs identified over 20 that are required for maintenance of pluripotency. Cyrano is a zebrafish lncRNA required for normal embryonic development, and its depletion produces CNS and notocord defects. HOTTIP, an lncRNA transcribed adjacent to the HOXA locus, is required for the normal expression of multiple genes therein. Depletion of HOTTIP in the embryonic chick forelimb produces shortened, abnormal limb bones. All together, functional studies have identified about a hundred lncRNAs that regulate various aspects of vertebrate cell biology and development, indicating that this class of molecules exerts broad critical functions.

Only a few hematopoietic lncRNAs have been studied closely. LincRNA-EPS is a mouse nuclear lncRNA that is upregulated during terminal erythropoiesis and represses Pycard, a pro-apoptotic gene. RNAi knockdown of LincRNA-EPS in erythroblasts de-represses Pycard, causing apoptosis. Conversely, in vitro overexpression of LincRNA-EPS protects erythroblasts from apoptosis caused by erythropoietin deprivation. The lncRNA EGO regulates eosinophil granule protein expression, and HOTAIR1, a lncRNA in the HOXA cluster, is upregulated during myeloid development and required for normal induction of certain HOXA and myeloid differentiation genes. Several abstracts presented at the 2012 American Society of Hematology Annual Meeting identified and categorized lncRNAs in hematopoietic stem cells, myeloid cells, erythroblasts and megakaryocytes. We anticipate that functions for some of these lncRNAs will emerge over the next few years.
LncRNAs and human disease

A handful of lncRNAs have been shown to influence malignant transformation and progression. Elevated expression of HOTAIR in breast cancer cells worsens survival by promoting tissue invasion and metastases.\textsuperscript{19} MALAT1 lncRNA is highly expressed in various malignancies and disruption of its genomic locus impairs the metastatic potential of a human lung cancer cell line in animal models.\textsuperscript{20} The lncRNA p15AS is antisense to the tumor-suppressor gene p15, which it suppresses by inducing local chromatin condensation. The expression of p15AS is elevated and associated with p15 silencing in AML and ALL blasts compared to normal hematopoietic cells.\textsuperscript{21} Deletion of Xist in the hematopoietic compartment of female (but not male) mice produces a highly penetrant and aggressive form of myeloproliferative/myelodysplastic syndrome and leukemia through de-repression of numerous genes on the second X-chromosome, a finding that may explain why human malignancies sometimes show supernumary X chromosomes.\textsuperscript{22} Most recently, RNA-Seq of Mycosis fungoides and Sezary syndrome cells from affected patients identified 12 unannotated lncRNAs that are differentially expressed in normal and malignant T-lymphocytes.\textsuperscript{23}

Dysregulation of lncRNAs can also contribute to inherited diseases. Fascioscapulohumeral muscular dystrophy is a rare autosomal dominant disorder associated with deletions on chromosome 4q35. These deletions appear to activate transcription of DBE-T, a chromosome 4 lncRNA that recruits chromatin-activating complexes and abnormally activates neighboring genes that mediate the disease phenotype.\textsuperscript{24} Germline mutations that interfere with lncRNA
function can also cause diseases. Two families with Brachydactyly Type E (shortened metacarpals and metatarsals) have translocations of chromosome 12, leading to the displacement of lncRNA DA125942 from its location near *PTHLH* (an important chondrocyte gene) to another chromosome. The normal role of DA125942 is to enhance *PTHLH* expression in cis, but upon translocation the lncRNA is displaced to another chromosome, causing reduced *PTHLH* expression and abnormal cartilage development.\(^{25}\) LncRNAs are also implicated in the pathogenesis of Prader-Willi and HELLP syndromes.\(^ {25}\) Other than the examples listed above, limited lncRNA data is available for human hematopoietic abnormalities, though this is likely to change as noncoding transcripts are extensively mapped and DNA sequence analysis of congenital and acquired hematological disorders interrogates noncoding regions of the genome.

**Challenges, controversies and future directions**

Bioinformatic methods have identified thousands of lncRNAs, and an interesting debate has emerged regarding how many of these actually exert important biological functions.\(^ {26}\) For example, enhancers frequently produce lncRNA transcripts. Some of these may simply represent non-functional by-products of transcription factor chromatin occupancy,\(^ {27}\) while others are apparently required in cis for the regulatory activity of the enhancer.\(^ {17}\) To date, only a small fraction of lncRNAs has been studied mechanistically and publications are probably biased towards lncRNAs with detectable functions. Systematic functional studies on large numbers of lncRNAs will help to address their general significance. However, screening studies must be interpreted with caution. Manipulation of the lncRNAs MALAT1 and NEAT1 produce distinct effects in tissue culture cells, yet the corresponding knockout mice exhibit no obvious
phenotype. Further testing may reveal abnormalities in these mutant mice, analogous to many microRNAs, in which in vivo loss-of-function phenotypes are only apparent under specific stresses.

Understanding the functions of IncRNAs is also complicated by their interesting and unusual evolutionary properties. For example, IncRNA promoters and loci show greater sequence conservation than background DNA, though it appears that many IncRNAs detected in one species might not even be transcribed in another species. Whether this implies lack of function or an ability to gain and lose function more rapidly during evolution than protein-coding genes requires further study. Supporting the latter possibility, a IncRNA required for cardiovascular development in mice is undetectable in rats and humans. Thus, failure to exhibit cross-species conservation in expression does not necessarily indicate lack of IncRNA function. In addition, some IncRNAs exhibit little overall sequence homology, yet seem to share common functions. For example, although the human or mouse orthologs of Cyrano exhibit low sequence similarity to the zebrafish gene, they can partially substitute for its loss. In this and other cases, evolutionary constraint of IncRNAs may occur at secondary and tertiary structural levels and selectively on small domains, all of which will be challenging to define.

It is likely that future studies will uncover prominent roles for IncRNAs in hematopoiesis and associated disorders. The discovery of LincRNA-EPS, which promotes survival of murine erythroblasts, may help to identify novel pathways that govern red blood cell production and potentially be manipulated for therapeutic purposes. Similarly, multiple other IncRNAs are likely to promote the specification, maintenance and function of other hematopoietic lineages.
Given their established roles in gene regulation, lncRNAs are also likely to influence hematological malignancies by modulating the expression of oncogenes, tumor suppressors, cell cycle regulators and proteins that regulate apoptosis via multiple diverse mechanisms outlined in Figure 1. For example, a canonical function of lncRNAs is to modulate epigenetic landscapes. Since mutations or altered expression of genes encoding epigenetic modifier proteins are known to promote various leukemias, derangements of currently unidentified lncRNA genes may produce similar effects (Figure 2). LncRNAs could also promote malignant transformation through their known abilities to regulate splicing, a process that has been implicated recently in myelodysplastic syndromes and myeloproliferative neoplasms. To investigate these possibilities, it is important for future genome and transcriptome sequencing studies of malignant hematopoietic disorders to examine not only protein encoding genes, but also lncRNAs. In this regard, lncRNAs exhibit highly cell type specific expression and have already been shown to have differential expression profiles in cutaneous T-cell lymphomas. It will be of interest to determine whether lncRNA expression patterns are useful for sub-stratifying hematopoietic malignancies with respect to underlying mutational profiles, prognosis and treatment sensitivities (Figure 2C).

Long noncoding RNAs may eventually themselves become pharmacologic targets that enhance our ability to modulate pathological gene expression. For example, many (but not all) lncRNAs can be suppressed by RNA interference. In addition, lncRNAs are likely to act by adopting specific folding conformations, which could present surfaces that are amenable to manipulation by peptides, other RNAs or small molecules. In many ways, the lncRNA field today is where the
microRNA field was a decade ago and there is considerable enthusiasm for the potential to mine biological insights from its study. We predict that the information unearthed by these studies will advance science and medicine in exciting and unexpected ways.

Acknowledgements

We thank Harvey Lodish, Gerd Blobel, Douglas Higgs, Wenqian Hu, Juan Alvarez and Ana Marques for comments on the manuscript. Noncoding RNA studies in our laboratory are funded by NIH R01 DK092318, DK065806 and The Roche Foundation for Anemia Research. V.R.P. is a recipient of the American Society of Hematology Research Training Award for Fellows and the ASH Fellow Scholar Award.

Author Contributions

V.R.P and M.J.W. wrote the manuscript.

Conflict of Interest Disclosures

The authors have no conflicts of interest to declare.
References

Figure Legends

Figure 1. In numerous tissues, lncRNAs (indicated in green) have been shown to regulate gene expression at multiple levels – chromatin, transcription, mRNA, translation, and protein. Hematopoietic lncRNAs may act at any of these levels. “MicroRNA sponge” refers to the ability of lncRNAs to sequester cellular microRNAs and prevent them from binding mRNA targets.14

Professional illustration by Debra T. Dartez.

Figure 2. Potential functions for lncRNAs in normal and malignant hematopoiesis.

A) Based on their known activities in many cell types (see Figure 1), lncRNAs may function during normal hematopoiesis to recruit transcription factors (TFs), co-activators, and activating chromatin factors to lineage-specific genes, thereby modulating gene expression to facilitate the differentiation and maturation of blood lineages. LncRNAs are indicated in green.

B) Absence or inappropriate expression of relevant lncRNAs could alter transcription to block normal differentiation and activate leukemic pathways.

C) Hematological malignancy subtypes may express distinct lncRNA expression profiles that reflect cell type of origin, mechanism of transformation, treatment sensitivities and prognosis.

Professional illustration by Debra T. Dartez.
Long noncoding RNAs in biology and hematopoiesis

Vikram R. Paralkar and Mitchell J. Weiss

Information about reproducing this article in parts or in its entirety may be found online at: http://www.bloodjournal.org/site/misc/rights.xhtml#repub_requests

Information about ordering reprints may be found online at: http://www.bloodjournal.org/site/misc/rights.xhtml#reprints

Information about subscriptions and ASH membership may be found online at: http://www.bloodjournal.org/site/subscriptions/index.xhtml

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include digital object identifier (DOIs) and date of initial publication.