

Multiple types of long non-coding RNAs regulate red blood cell development

Harvey Lodish

Whitehead Institute for Biomedical Research Cambridge, USA

Red blood cell development is a highly coordinated process that initiates from multipotent stem cells residing in the fetal liver or the adult bone marrow via cell lineage specification, proliferation, and differentiation. Coordination of this process requires dynamic and precise control of gene expression, and disruption of erythroid transcriptional networks leads to disease. Thus, identifying novel modulators of erythrocyte production is crucial to better understand the regulatory circuitry underlying red blood cell development and to find novel opportunities for treatment of anemias and other erythroid disorders. Mammalian genomes encode thousands of long non-coding RNAs (lncRNAs), transcripts longer than 200 nucleotides lacking functional coding capacity, but knowledge of their importance for cellular development is limited. Recently we catalogued the entire repertoire of lncRNAs expressed during mouse fetal red blood cell development *in vivo*. We integrated genome-wide surveys of expression, chromatin states, and transcription factor occupancy to uncover global and subclass-specific features of lncRNA biogenesis, regulation, and tissue and developmental specificity. We identified multiple classes of erythroid lncRNAs: intergenic, antisense to other genes, intron-overlapping with protein-coding genes, small RNA (sRNA) hosting, enhancer-derived, and pseudogene-derived.

Importantly, we discovered that binding of key transcription factors GATA1 and TAL1 at lncRNA and mRNA promoters is associated with H3K4me2 deposition along with transcriptional activation during erythroid differentiation. We further focused on lncRNAs abundantly induced during differentiation, including novel erythroid-specific lncRNAs conserved in humans that are nuclear-localized. Depleting 10 of these candidates impaired proper maturation of enucleated erythroblasts. One of them, alncRNA-EC7, is specifically needed for activation of the neighboring gene encoding BAND 3/SLC4A1, a major component of the erythrocyte membrane. Thus, diverse types of lncRNAs participate in the regulatory circuitry underlying lineage-specific development.

lodish@wi.mit.edu