Ferrata Storti Foundation

Haematologica 2019
Volume 104(5):894-906

# Long noncoding RNAs of single hematopoietic stem and progenitor cells in healthy and dysplastic human bone marrow

**Zhijie Wu,[1]* Shouguo Gao,[1]* Xin Zhao,[1] Jinguo Chen,[2] Keyvan Keyvanfar,[1] Xingmin Feng,[1] Sachiko Kajigaya[1] and Neal S. Young[1]**

[1]Hematology Branch, National Heart, Lung, and Blood Institute, National Institutes of Health and [2]Trans-NIH Center for Human Immunology, Autoimmunity, and Inflammation, National Institutes of Health, Bethesda, MD, USA

*ZW and SG contributed equally to this work.*

## ABSTRACT

Long noncoding RNAs (lncRNAs) are regulators of cell differentiation and development. The lncRNA transcriptome in human hematopoietic stem and progenitor cells is not comprehensively defined. We investigated lncRNAs in 979 human bone marrow-derived CD34$^+$ cells by single cell RNA sequencing followed by *de novo* transcriptome reconstruction. We identified 3,173 lncRNAs in total, among which 2,365 were previously unknown, and we characterized lncRNA stem, differentiation, and maturation signatures. lncRNA expression exhibited high cell-to-cell variation, which was only apparent in single cell analysis. lncRNA expression followed a lineage-specific and highly dynamic pattern during early hematopoiesis. lncRNAs in hematopoietic cells closely correlated with protein-coding genes of known functions in the regulation of hematopoiesis and cell fate decisions, and the potential regulatory roles of lncRNAs in hematopoiesis were imputed by projection from protein-coding genes with a "guilt-by-association" approach. We characterized lncRNAs preferentially expressed in hematopoietic stem cells and in various downstream differentiated lineage progenitors. We also profiled lncRNA expression in single cells from patients with myelodysplastic syndromes and in aneuploid cells in particular. Our study provides a global view of lncRNAs in human hematopoietic stem and progenitor cells. We observed a highly ordered pattern of lncRNA expression and participation in regulation of early hematopoiesis, and coordinate aberrant messenger RNA and lncRNA transcriptomes in dysplastic hematopoiesis. (*Registered at clinicaltrials.gov with identifiers: 00001620, 00001397*)

## Introduction

Long noncoding RNAs (lncRNAs), which are defined as a subclass of noncoding RNAs, are longer than 200 nucleotides and lack protein coding capacity. lncRNAs are newly recognized as regulators of gene expression, transcriptionally and post-transcriptionally.[1-3] Unlike messenger RNAs (mRNAs), which localize specifically to the cytoplasm, lncRNAs can occupy various nuclear compartments and/or the cytoplasm. lncRNAs function via RNA-DNA, RNA-RNA, and RNA-protein interactions.[2-6] As a result, they affect multiple stages of gene regulation, including placement of chromatin marks, mRNA biogenesis, and signaling pathways.

lncRNA expression is tissue- and cell type-specific[5,7-9] but less conserved across species than is mRNA expression.[10,11] lncRNAs have been linked to the development of several lineages in hematopoiesis and in the immune response. Some lncRNAs were found to be enriched in hematopoietic stem cells (HSCs)[12] or dynamically expressed during erythropoiesis.[13,14] RNA interference studies have revealed that lncRNAs control HSC self-renewal and differentiation,[12] erythroid precursor maturation,[14] and granulocytic differentiation of hematopoietic stem and progenitor cells (HSPCs).[15] Intergenic lncRNA signatures exhibit subset-specificity in T and B lymphocytes.[16-18] lincR-Ccr2-5'AS, together with GATA3, is essential in the regulation

**Correspondence:**

*ZHIJIE WU*
zhijie.wu@nih.gov

Received: October 12, 2018.

Accepted: November 22, 2018.

Pre-published: December 13, 2018.

doi:10.3324/haematol.2018.208926

*Check the online version for the most updated information on this article, online supplements, and information on authorship & disclosures: www.haematologica.org/content/104/5/894*

of gene expression and migration of Th2 cells.[16] Downregulation of linc-MAF-4 skews T-cell differentiation towards the Th2 phenotype.[17] TMEVPG1, a Th1-specific intergenic lncRNA, controls the expression of interferon-γ together with the Th1-specific transcription factor T-bet, and is critical in modulating susceptibility to infection with Theiler virus.[19,20] Expression of lncRNAs in pro-B and mature B cells is regulated by PAX5, a transcriptional factor required to specify B-cell lineage.[18] Despite these many examples of specific functions for either stem cells or differentiated lineages, the repertoire of lncRNAs in human HSPCs has not been fully described.

Whole transcriptome sequencing allows large scale profiling of lncRNAs in tissues and diseases and, therefore, enables the identification of many putative lncRNAs.[5,21,22] lncRNAs in general are expressed at much lower levels[3,4,23,24] but are more cell type-specific than are mRNAs.[9,25] Until recently, lncRNA expression was assessed by averaging transcriptomes of bulk RNA extracted from mixed cell populations, which limits the sensitivity to detect lncRNA expression in small cell populations and thus to resolve diversity within a cell type. With recent advances in single cell transcriptome profiling methods, many seemingly homogeneous cell populations have shown unexpected variability in gene expression. Recently published studies profiling lncRNAs at the single cell level have revealed the cell-specific expression of these RNAs.[5,26-30]

In the current work, we performed single cell RNA sequencing (scRNA-seq) of 979 freshly isolated bone marrow-derived human CD34$^+$ cells from both healthy donors and patients with myelodysplastic syndrome (MDS). Using *de novo* transcriptome reconstruction, we identified a total of 3,173 lncRNAs, including 2,365 potential novel lncRNAs not reported in public databases. We further characterized the features and expression patterns of lncRNAs in CD34$^+$ cells, revealing stage- and lineage-specificity of lncRNA expression and putative functions in normal hematopoiesis. Expression and lineage-specificity of almost 40 lncRNAs, including those novel lncRNAs, were validated by quantitative real-time polymerase chain reaction (RT-PCR). We also profiled lncRNAs in MDS cells, and aneuploid cells in particular. Our study provides a global assessment of lncRNA biology in early human hematopoiesis.

## Methods

### Subjects and samples

Bone marrow samples from seven healthy donors and five MDS patients were obtained after written informed consent in accordance with the Declaration of Helsinki and under protocols (*www.clinicaltrials.gov NCT00001620 and NCT00001397*) approved by the Institutional Review Boards of the National Heart, Lung, and Blood Institute. Of the five patients with MDS, patients 1, 2, and 5 had evolved to MDS from aplastic anemia while patients 3 and 4 had *de novo* MDS. Fluorescence activated cell sorting (FACS) was performed using the FACSAria II Cell Sorter (BD Biosciences) after isolation of bone marrow mononuclear cells. The gating strategies are shown in *Online Supplementary Figure S1A*. CD34$^+$CD38$^-$ and CD34$^+$CD38$^+$ cells from four healthy donors and patient 4 were sequenced separately, while only the CD34$^+$ populations of patients 1, 2, 3, and 5 were sequenced due to limited cell numbers (*Online Supplementary Figure S1B*). The clinical characteristics of these patients have been published.[31] Another set of bone marrow cells from a further three healthy donors was used for quantitative RT-PCR (*Online Supplementary Figure S2*).

### Single cell RNA sequencing

The C1 Single-cell Auto Prep System (Fluidigm) was employed to perform SMARTer (Clontech) whole transcriptome amplification on as many as 96 individual cells, according to the manufacturer's protocols (www.fluidigm.com). Whole transcriptome amplification products were converted to Illumina sequencing libraries using the Nextera XT DNA Sample Preparation Kit (Illumina). Final cDNA libraries were quantified using High Sensitivity DNA Kits (Agilent) and sequenced on a HiSeq 2500 or 3000 (Illumina), using the paired-end 75-bp protocol, as described previously.[31] RNA-seq data from this study have been deposited at the National Center for Biotechnology Information Gene Expression Omnibus (accession number GSE99095), and updated with intermediate and result files from the lncRNA analysis. Aliquots of whole transcriptome amplification products were used for quantitative RT-PCR analysis.

### Bioinformatic analysis

Total reads were mapped to the reference genome (hg19) with RSubreader and gene-level read counts were calculated using featureCounts.[32] Only data from high-quality cells with captured genes were utilized further. The schematic pipeline has been published.[31] Aneuploidy was evaluated by three independent methods, including a sliding window analysis of copy number variations, chromosome relative expression value distribution, and analysis of the degree of loss of heterozygosity.

### Identification and classification of long noncoding RNAs

After filtering computationally for quality,[31] single cells were used to define lncRNAs with a pipeline adopted from published methods of identifying high-confidence gene models.[13,14,16,17,28] Fastq files of cells from each subject were merged. Reads were mapped to human genome hg19 with Tophat2 and assembled using Cufflinks packages.[33] The assembled transcripts from all subjects were merged with Cuffmerge[33] before removing genes with <200 nucleotides or containing single exons in order to obtain long transcripts. Assembled genes overlapping with known protein-coding genes were excluded, and we removed those with low expression (FPKM<2) to improve the reliability of the model. We investigated the coding potential of the remaining genes using three independent algorithms: (i) protein database homology with BlastX and Pfam 31.0 (hmmer2.0); (ii) codon potential assessment with CPAT;[34] and (iii) presence of long open reading frames >100 amino acids with EMBOSS GetORF.[35] Defined lncRNAs were compared with annotated databases from Ensembl, University of California Santa Cruz (UCSC) Genome Browser, and GENCODE:[36] overlapping lncRNAs were defined as "annotated lncRNAs" and the others as putative "novel lncRNAs". If supported by cap analysis of gene expression (CAGE) data,[37] lncRNA transcripts obtained by the same filtering pipeline, but with medium expression levels (FPKM 0.1-2) were also defined to be expressed in human CD34$^+$ cells (*Online Supplementary Methods and Results*).

## Results

### Identification and characterization of long noncoding RNAs in human CD34$^+$ hematopoietic cells

To assess lncRNA expression in human HSPCs, we purified CD34$^+$ cells from the marrow of four healthy donors and five MDS patients. We then analyzed polyadenylated

RNA by scRNA-seq. After filtering, 391 cells from healthy donors and 588 cells from MDS patients were retained for analysis, with over 9.1 billion 75 bp paired-end mapped reads in total and 7.7 million reads per cell on average. Using a published strategy,[31] a total of 10,791 protein-coding genes were captured, 3,777 per cell on average.

To obtain reliable models of lncRNA expression, we followed a *de novo* transcript assembly pipeline (Figure 1A), in which "high-confidence" transcriptomes[13,14,16,17,28] from CD34⁺ single cells of all nine subjects were merged in order to undergo multi-step filtering for: (i) overlap with known mRNA exon annotations, (ii) size and multiexonic selection, (iii) known protein domains, (iv) low levels of expression, and (v) predicted coding potential. Using this conservative multilayered analysis, we identified a total of 2,892 lncRNAs across 979 single human CD34⁺ cells. To assign lncRNAs to specific classes, we examined their overlap with annotated noncoding genes present in public databases: 808 lncRNAs were previously annotated and 2,084 were putative novel lncRNAs (Figure 1B and *Online Supplementary File 1*). In addition, transcripts that were expressed at medium levels and supported by CAGE data[37] were also defined to be lncRNAs (n=281) expressed in human CD34⁺ cells (*Online Supplementary File 2*). Defined lncRNAs exhibited similarly low protein-coding potential (relative to protein-coding genes) as had previously annotated lncRNAs in the GENCODE database (Figure 1C). Such defined lncRNAs in single human CD34⁺ cells were distributed across all chromosomes, at much lower average abundance than were protein-coding transcripts. Compared with protein-coding genes, lncRNA-encoding genes had fewer exons, were shorter and less well conserved. In general, lncRNA-encoding genes were enriched in 4-kb regions around the transcriptional start sites of their neighboring protein-coding genes, in agreement with previous work,[38] suggesting that they share promoter regions [lncRNA-encoding genes show higher co-expression with protein-coding neighbors than do pro-

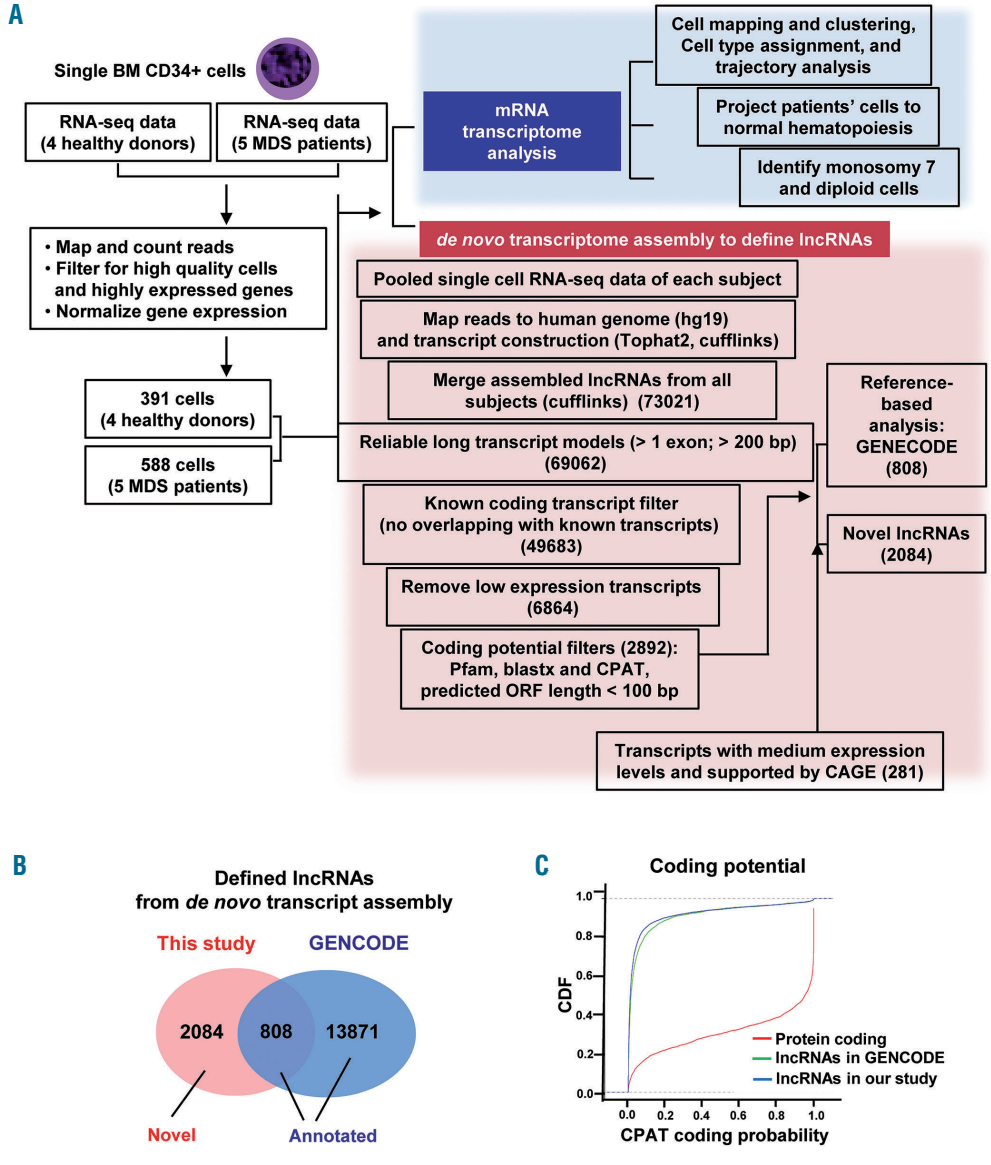



**Figure 1. Identification of long noncoding RNAs expressed in single human CD34⁺ cells.** (A) Bioinformatics pipeline for identification of long noncoding RNAs (lncRNAs). Single cell RNA-sequencing (scRNA-seq) data from nine subjects were processed and filtered before further analysis of messenger RNA (mRNA) and lncRNA expression. mRNA transcriptome analysis including cell clustering, cell type assignment, and identification of monosomy 7 cells was described[32] and employed to analyze gene expression patterns among cell types, functional imputation of lncRNAs, and differentiation trajectory analysis in the current study. scRNA-seq data were processed by *de novo* genome-based transcriptome reconstruction for the quantification of lncRNAs expressed in human CD34⁺ cells through the multi-step filtering bioinformatic pipeline. Numbers of remaining transcripts after each filtering step are indicated. (B) By comparing defined lncRNA transcripts in *de novo* transcript assembly with transcripts in the GENCODE database, 808 lncRNAs were previously annotated while 2,084 were classified as potential novel lncRNAs. (C) Comparison of coding potential among previously annotated lnRNAs, novel lncRNAs, and mRNAs. x axis, coding probability calculated with CPAT; y axis, cumulative distribution function (CDF).

tein-coding gene pairs (see *Online Supplementary Results* "Characterization of lncRNAs defined in human CD34+ hematopoietic cells"; *Online Supplementary Figure S3*)].

### Detection of long noncoding RNAs with single cell RNA-sequencing

Expression of lncRNAs showed more variation among single cells than did the expression of coding transcripts (Figure 2A). Across all percentiles of gene expression levels, lncRNAs were expressed in smaller proportions of cells than were mRNAs (Figure 2B). Low overall expression of lncRNAs in bulk samples was likely partly attributable to limited but high expression of lncRNAs in a minority of cells or in small cell populations. Seven bulk samples of the CD34+ population from the nine individuals studied were sequenced in parallel with single cells. We sought to compare the maximum abundance of mRNAs or lncRNAs *versus* housekeeping genes in bulk samples and individual cells,[28] to quantify the power of gene expression detection by these different technical approaches. mRNAs were detected at a similar ratio to housekeeping genes in both bulk samples and single cells,

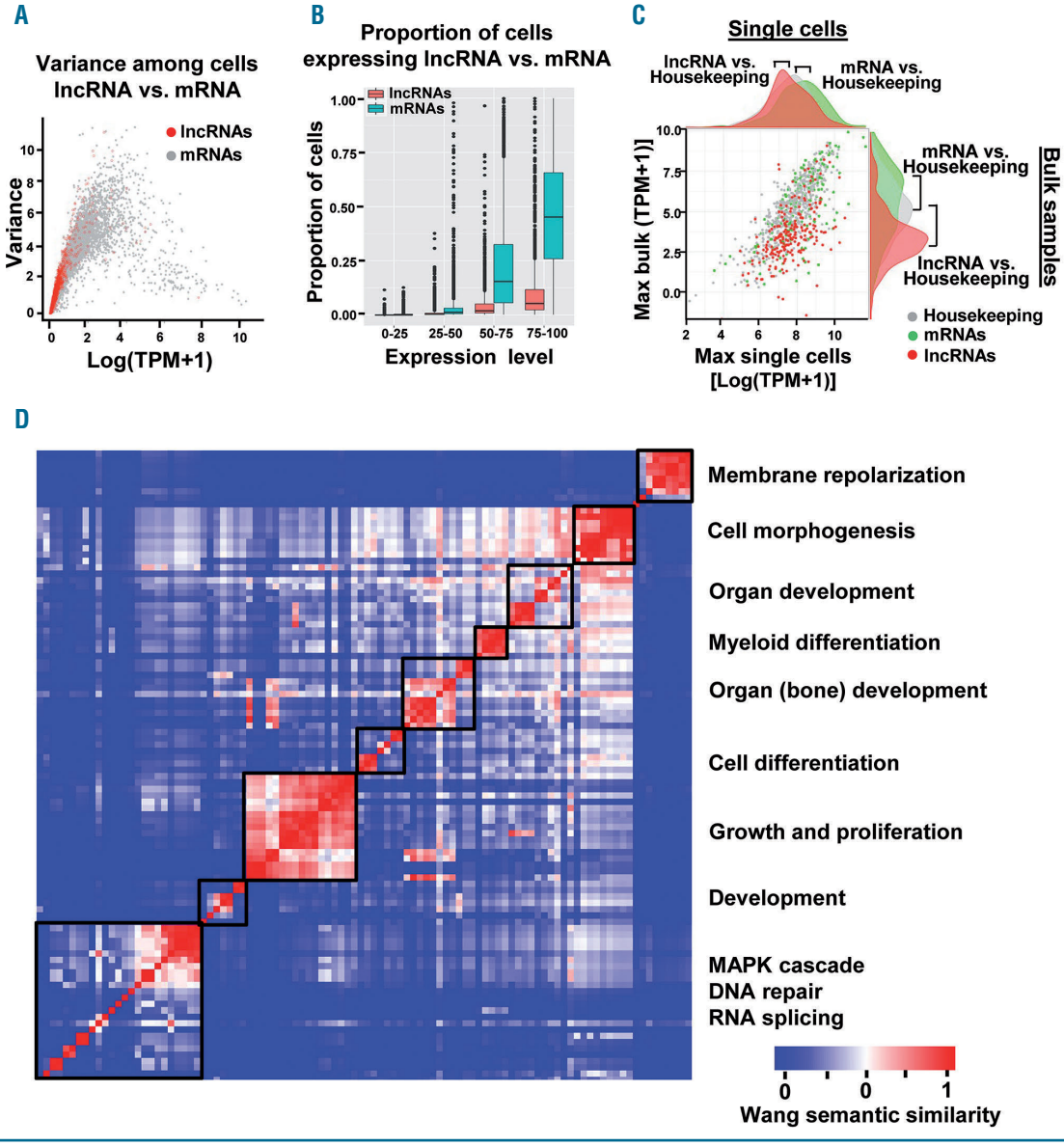



**Figure 2. Detection of long noncoding RNAs by single cell RNA sequencing.** (A) Variance of long noncoding RNA (lncRNA) and messenger RNA (mRNA) expression among single cells. x axis, Log (TPM+1); y axis, variance. (B) Proportion of CD34+ cells (individual dots) that express individual lncRNAs (blue) and mRNAs (red), separated by expression quantile of the set of all transcripts (lncRNAs and mRNAs combined). x axis, average expression level quantiles; y axis, proportion of cells. (C) Comparison of single cell and bulk tissue maximum expression levels of mRNAs and lncRNAs. Gray, housekeeping genes; green, mRNAs; red, lncRNAs. Projected density plots summarized expression levels of scatter plots along the single-cell (horizontal) and bulk tissue (vertical) axes. Short lines noted alongside the histogram plots represent the difference of the median expression of lncRNAs or mRNAs to the median expression of housekeeping genes in single cell or bulk tissue RNA-seq. (D) Gene-ontology semantic similarity matrix of protein-coding genes defined by a guilt-by-association approach of lncRNAs in human CD34+ cells. Gene ontology terms involved in a similar functional matrix were adjacent and formed a block with Pearson R values ranging from -1 to 1. Terms noted on the right side depict common biological processes of the block of gene-ontology terms.

but the ratio of maximum expression of lncRNAs relative to housekeeping genes was about 4-fold higher in single cells than in bulk samples. By scRNA-seq, the maximum expression of lncRNAs was similar to that of both mRNAs and housekeeping genes (Figure 2C). Genes with high variance tended to be captured by the single cell analysis rather than by the bulk approach (*Online Supplementary Figure S4*). Thus, lncRNA expression appeared to be better detected among single cells due to an expression pattern of high cell-to-cell variation and cell-specificity.

We then sought to infer putative functions of defined lncRNAs in hematopoiesis by a comprehensive "guilt by association" approach *(Online Supplementary Methods and Results*), correlating expression of lncRNAs with protein-coding genes of known functions.[4,15,39-41] Associated protein-coding genes of defined lncRNAs across CD34$^+$ cells were enriched in gene ontology (GO) terms related to myeloid cell differentiation, cell growth, and cellular functions including DNA repair, mRNA splicing, gene expression, and epigenetic regulation (Figure 2D), implicating lncRNAs in the regulation of human hematopoiesis and associated cellular functions.

## Stage- and lineage-specific expression of long noncoding RNAs in normal hematopoiesis

To obtain a profile of lncRNA expression in normal human hematopoiesis, we assessed lncRNA expression in 391 CD34$^+$ cells from healthy donors. We first studied whether a lncRNA signature separated CD38$^-$ and CD38$^+$ cell populations. lncRNAs detected with 20 reads in at least 20 cells were retained, and highly variable lncRNAs were used for stage-specific analysis (*Online Supplementary Figure S5A*). The method of t-distributed stochastic neighbor embedding (t-SNE) was adopted for non-linear dimension reduction based solely on batch-corrected (by Combat/SVA) lncRNA expression (*Online Supplementary Figure S5B*). In an unsupervised t-SNE plot, sorted CD38$^-$ cells formed a cluster distinct from CD38$^+$ cells, while CD38$^+$ cells were more dispersed (Figure 3A). To determine stage specificity, we performed pair-wise comparison of lncRNA expression in CD38$^-$ cells relative to expression in CD38$^+$ cells. lncRNA expression exhibited substantial differences in two stages (*Online Supplementary Table S3*); heatmaps of differentially expressed mRNAs and lncRNAs of CD38$^-$ and CD38$^+$ populations are shown in Figure 3B.

We previously assigned single CD34$^+$ cells to a cell type according to their protein-coding transcriptome profiles, based on gene expression data from flow cytometrically-sorted cell populations.[42] The cell types to which the single cells were assigned included HSC, multilymphoid progenitor (MLP), megakaryocyte-erythroid progenitor (MEP), granulocyte-monocyte progenitor (GMP), pro-B cell (ProB), and earliest thymic progenitor (ETP).[31] We applied weighted gene co-expression network analysis[43] to assess the potential functions of lncRNAs in CD38$^+$ and CD38$^-$ cells. When protein-coding and lncRNA-encoding genes were simultaneously analyzed, they clustered into seven unsupervised modules (*Online Supplementary Table S4*), and genes in individual modules were analyzed for GO term enrichment (Figure 3C). Genes in module 1 showed high enrichment of lymphocyte activation pathway genes, and their expression levels were higher in ProB and ETP than in other cell types. Genes in module 6 were enriched in the heme metabolic process, and they showed higher expression in MEP. These data suggest roles of lncRNAs in hematopoiesis and lineage specificity of lncRNA expression.

By t-SNE, cells tended to cluster according to cell types (Figure 4A, right) and were coincident with the pattern of hematopoietic differentiation based on mRNA expression in pseudotime ordering (Figure 4A, middle).[31] Thus lncRNAs appeared as powerful as their protein-coding counterparts in resolving subtypes of CD34$^+$ cells. We then analyzed cell-type specificity of gene expression by cell-type variance (Figure 4B) and assessed a Jensen-Shannon score[8] (JScore) (Figure 4C). lncRNA expression showed higher cell-type specificity than did mRNA expression (JScore, $P$=1x10$^{-16}$). There was more cell-to-cell variation in lncRNA expression than in mRNA expression, even within the same cell type (*Online Supplementary Figure S6*). We investigated our dataset for lncRNA signatures in various lineages, using difference in expression in a lineage, relative to expression in all other subsets, by pairwise comparisons, at a threshold $P$<0.05 (Figure 4E and *Online Supplementary Table S5*). Heatmaps revealed that MLP had signatures of both mRNAs and lncRNAs similar to those of HSC, in contrast to distinctive gene expression patterns in other lineages. These data were congruent with those of earlier studies,[31,42,44] and indicated that HSC and MLP defined by a transcriptome signature were enriched in a phenotypically characterized CD34$^+$CD38$^-$ population, while the other lineages comprised the more heterogeneous CD34$^+$CD38$^+$ population. We examined overlap of lncRNA and mRNA expression among lineages: 94.8% of mRNAs were shared by at least five out of six lineages, but only 62.2% of lncRNAs were so widely expressed (Figure 4D, top panel); conversely, 81.4% of lineage-signature mRNAs were specific to only one lineage, while 92.2% of lncRNAs were equivalently specific (Figure 4D, bottom panel). Again, lncRNA expression appeared more lineage-restricted than did the counterpart coding gene expression. In summary, we found lncRNA expression to be highly stage- and lineage-specific during early hematopoiesis.

To confirm our findings of potential novel lncRNAs and lineage-specific expression patterns of lncRNAs, we compared our results with a publicly available dataset.[44] This scRNA-seq study was conducted with human HSPCs sorted based on cell surface antigens (GSE75478). Lineage-specific lncRNAs (and mRNAs) defined in the current study were also detected and showed consistent lineage-specific expression in the two datasets (*Online Supplementary Results* and *Online Supplementary Figures S7* and *S8*). We then assessed 39 lncRNAs and 14 mRNAs by quantitative RT-PCR of aliquots of whole transcriptome amplification from those 391 single CD34$^+$ cells and another set of flow cytometry-sorted bulk samples (*Online Supplementary Methods and Results*; *Online Supplementary Table S6*). All 39 signature lncRNAs, including 20 novel lncRNAs, were detectable in single cells and bulk samples by quantitative RT-PCR, indicating expression in human CD34$^+$ cells. We confirmed cell type assignment of single cells by expression of well-recognized mRNAs (*Online Supplementary Figure S9C*) and confirmed lineage-specific expression for 35 out of 39 lineage signature lncRNAs in single cells. Moreover, their lineage-specific expression patterns in single cells were reproducible in independent sorted bulk samples (*Online Supplementary Figure S9A,B*). Expression of these lineage-specific lncRNAs in hematopoietic differentiation, by scRNA-seq and quantitative RT-PCR, is illustrated in Figure 4F.

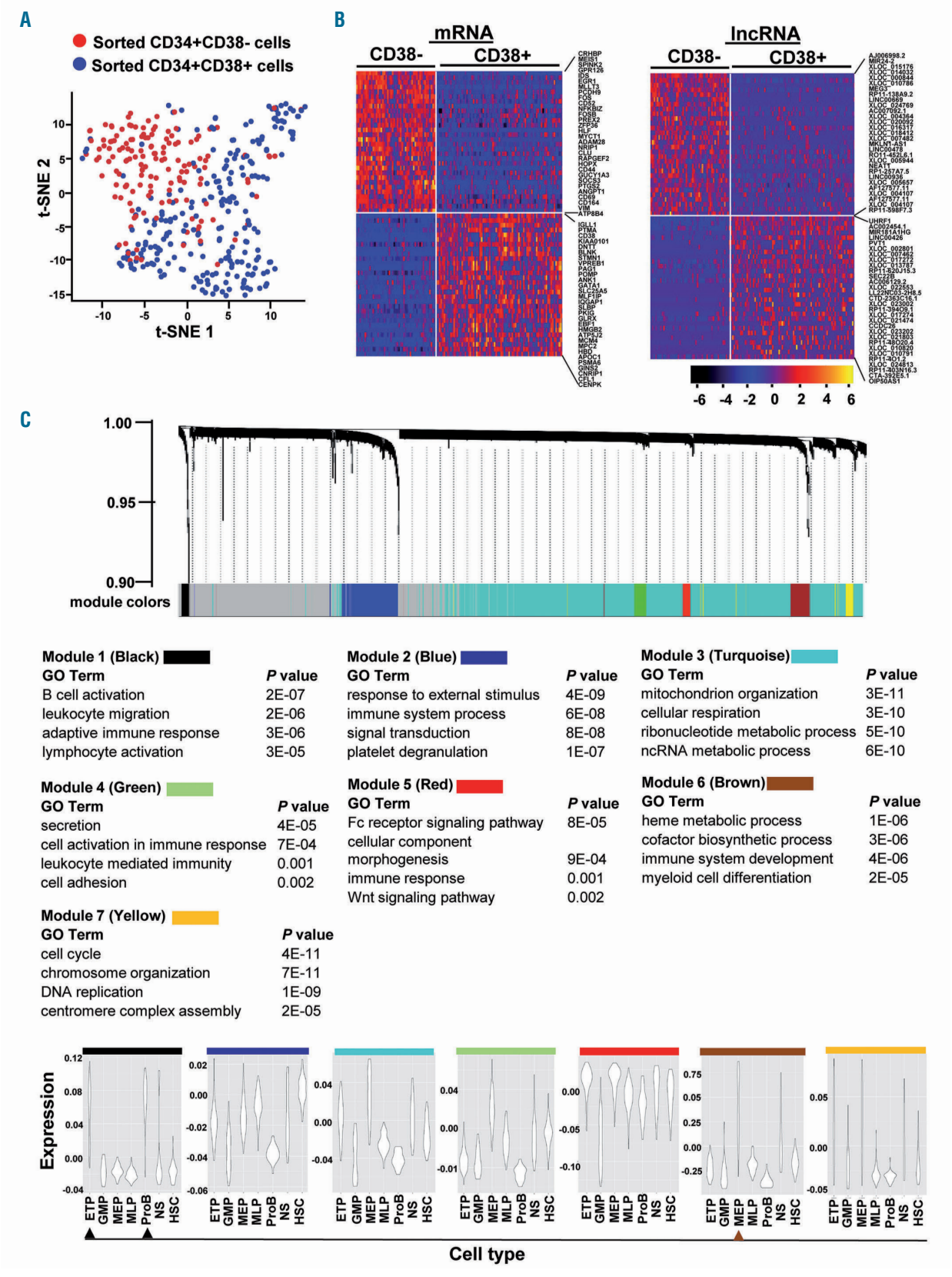



**Figure 3. Long noncoding RNA expression exhibited a cell stage-specific pattern.** (A) Single cell RNA sequencing (scRNA-seq) data of 391 cells with merely long non-coding RNA (lncRNA) expression were clustered using t-SNE in the Seurate package to obtain nonlinear dimension reduction and visualization in two dimensions (t-SNE1 and t-SNE2). scRNA-seq data of two different cell stages ($CD34^+CD38^-$ and $CD34^+CD38^+$) sorted by FACS were plotted in red and blue, respectively. (B) Heatmaps of messenger RNA (mRNA) (left) and lncRNA (right) expression in $CD34^+CD38^-$ and $CD34^+CD38^+$ cells. (C) Modules of protein-coding and lncRNA-encoding gene expression across single cells identified through weighted gene co-expression network analysis. Gene co-expression modules including both lncRNAs and mRNAs based on expression quantity and seven unsupervised modules are distinguished by colors (top panel); gene ontology (GO) terms for each module of genes identified in the co-expression matrix (middle panel); expression levels of individual modules of genes in different cell types (bottom panel). Detailed information on individual gene modules is presented in *Online Supplementary Table S4*.

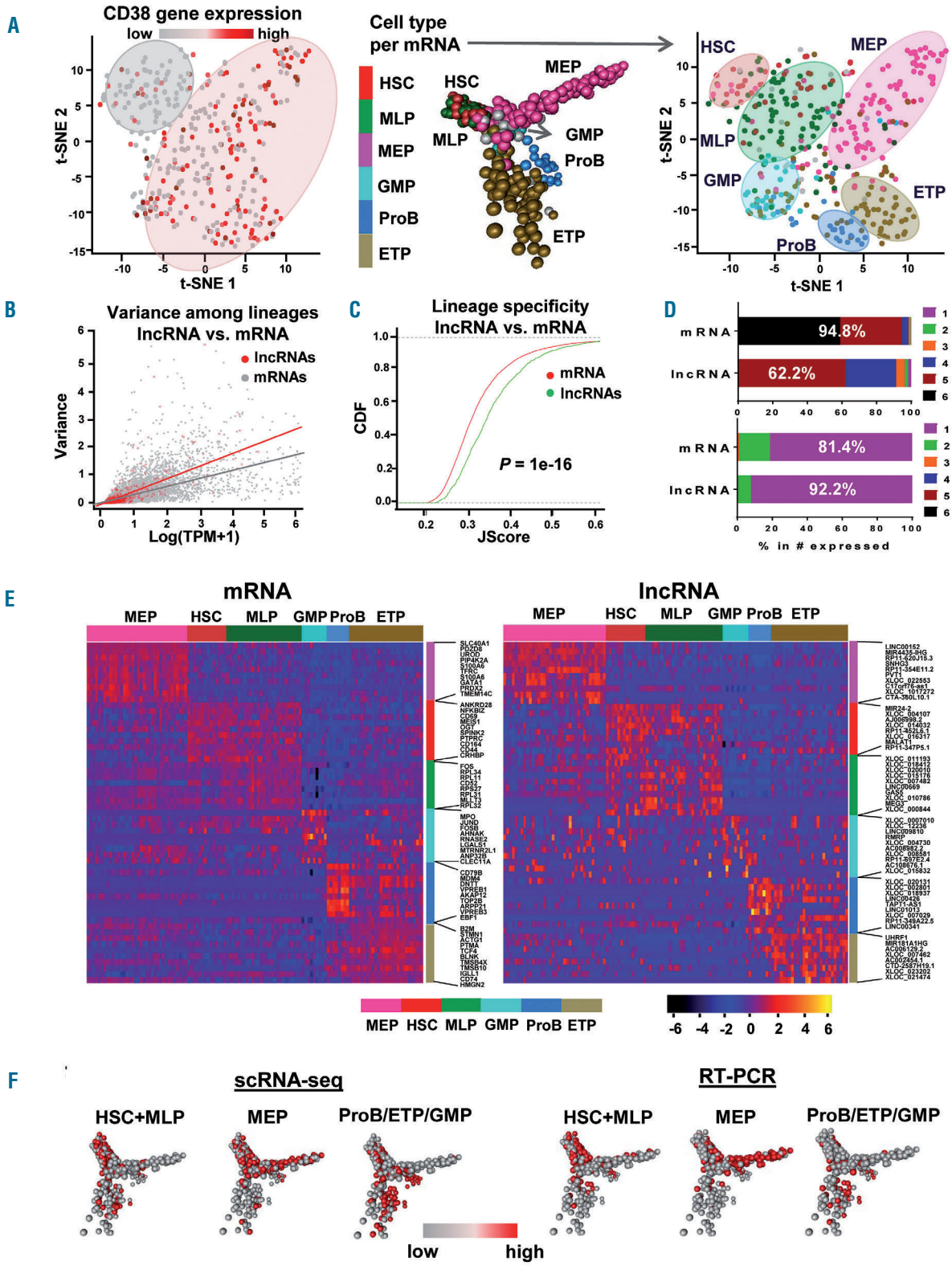



**Figure 4. Long noncoding RNA expression exhibited cell lineage-specific patterns.** (A) The same t-SNE plot as in Figure 3A, highlighted single cells with a CD38 expression level (left); cell types were assigned to single cells using messenger RNA (mRNA) expression information, followed by differentiation tree reconstruction using a pseudotime ordering method (middle); single cells are colored according to their corresponding cell types, and gray circles indicate clustering of the same cell type (right). (B) Variance of long noncoding RNA (lncRNA) versus. mRNA expression among different lineages. x axis, Log(TPM+1); y axis, variance. (C) JScore to assess lineage specificity of lncRNA and mRNA expression. x axis, JScore; y axis, cumulative distribution function (CDF). (D) Percentages of mRNAs and lncRNAs defined (top) or preferentially expressed (bottom) in various numbers of cell types. HSC: hematopoietic stem cell; MLP: multilymphoid progenitor; MEP: megakaryocyte-erythroid progenitor; GMP: granulocyte-monocyte progenitor, ProB: pro-B cell; ETP: earliest thymic progenitor. (E) Heatmaps of mRNA (left) and lncRNA (right) expression in different lineages. (F) Expression of a group of lineage-specific lncRNAs for HSC/MLP, MEP, ProB, and ETP along the differentiation tree, measured by single cell RNA sequencing (scRNA-seq) (left) and quantitative reverse transcriptase polymerase chain reaction (RT-PCR) analysis (right). Expression (shown as a mean expression level) is presented as a relative quantity in one lineage relative to expression in all the others.

### Coordinated activation and suppression of signature messenger RNAs and long noncoding RNAs during hematopoiesis

To systematically assess expression of lncRNAs that might be activated or suppressed during hematopoiesis, we focused on dynamic changes of the mRNA and lncRNA transcriptomes along differentiation trajectories defined by pseudotime ordering of HSC/MLP into MEP and GM/L (granulocyte/monocyte/lymphocyte progenitors) (Figure 5 and *Online Supplementary Tables S7* and *S8*). Sequentially upregulated/downregulated mRNAs and lncRNAs along the two trajectories were analyzed and gene expression was visualized in heatmaps (MEP trajectory in Figure 5A and GM/L trajectory in Figure 5B). Common downregulated mRNAs in MEP and GM/L trajectories (Figure 5C) were involved in signaling pathways related to stemness, including NRF2, AP-1, ATF-2, C-MYB, HIF-1, and IL-6 signaling. Downregulated genes specifically in the MEP differentiation pathway were mostly enriched in T cells and for broad immune response; enrichment in the EPO signaling pathway was observed only among GM/L downregulated genes. Frequently upregulated genes were involved in DNA replication, cell cycle, and cell proliferation; genes specifically upregulated in GM/L were enriched in B- and T-cell signaling and immune response (Figure 5D, right); hemoglobin synthesis and androgen receptors were enriched only among MEP upregulated genes (Figure 5D, left). lncRNA expression along the two differentiation trajectories was synchronously coordinated with lineage-specific coding genes and interrelated in functional pathways of stemness, megakaryocyte/erythrocyte development, and granulocyte/monocyte/lymphocyte development. Collectively, these data suggest the ordered expression of lncRNAs in hematopoietic differentiation and involvement in the regulation of hematopoiesis.

### Long noncoding RNAs are bound by lineage-specific transcription factors and might be regulated by epigenetic mechanisms

Transcription factors are critical in cell fate decisions and thus in the regulation of lineage-specific gene expression. Given the observation of highly ordered expression patterns of lncRNAs during hematopoiesis and co-expression with lineage-specific transcription factors, we investigated roles of lineage-specific transcription factors in regulating lncRNA expression during hematopoiesis. The transcription factor GATA1 regulates erythrocyte and megakaryocyte differentiation,[45,46] and indeed its expression was sequentially increased as HSC differentiate into MEP (Figure 5A). Using data obtained by chromatin immunoprecipitation sequencing (ChIP-seq) for GATA1 binding (Encode Ref# ENCSR000EFT), we found that GATA1 binding to promoters was higher in lncRNA-encoding (Figure 6A, top) as well as protein-coding genes (Figure 6A, bottom) preferentially expressed in MEP than for other cell types. lncRNA-encoding genes preferentially expressed in MEP, such as SNHG3 and RP11-620J15.3 (Figure 6B), bound to GATA1 and had high read coverage of active histone marks (H3K27Ac, H3K79me2, and H3K4me2) and low coverage of repressive histone marks (H3K27me3) in erythroid cells. Our analysis, together with published data,[8,13,14,16,18,39,47] indicated that cell fate decisions were controlled by critical lineage-specific transcription factors, as evidenced by expression of both lineage-specific mRNAs and lncRNAs bound and regulated by corresponding transcription factors, probably involving epigenetic modification.

### Long noncoding RNAs exhibit aberrant expression in aneuploid cells from patients with myelodysplastic syndromes

Gene expression of 588 single $CD34^+$ cells from five MDS patients was compared with that of cells from four healthy donors. lncRNAs were differentially expressed in MDS cells compared with those from healthy donors ($P<0.05$): 372 and 590 lncRNAs were upregulated and downregulated, respectively (Figure 7A and *Online Supplementary Table S10*). By guilt-by-association, downregulated lncRNAs were associated with gene sets involved in immune response, cellular response, and gene expression and DNA damage response; upregulated lncRNAs were involved in cell metabolism and cell signaling (Figure 7B,C).

We adopted three bioinformatics methods to distinguish cells with abnormal karyotypes from diploid cells.[31] We observed that 200 and 56 lncRNAs were downregulated and upregulated, respectively, in monosomy 7 cells, compared to diploid cells ($P<0.05$) (Figure 7D and *Online Supplementary Table S11*). By guilt-by-association, downregulated lncRNAs were associated with genes involved in immune response, cell apoptosis and cell death, and DNA modification; upregulated lncRNAs displayed involvement in Ras signaling, Wnt signaling, and interleukin-8 production (Figure 7E,F).

## Discussion

In the current study, we profiled the repertoire of lncRNAs in human bone marrow-derived $CD34^+$ cells, with the goal of understanding lncRNA biology in early human hematopoiesis. The majority of the human genome is transcribed but only a small proportion of transcripts encode proteins,[4,48,49] and thus the number of lncRNA genes is predicted to be very large. Deep RNA sequencing followed by *de novo* transcriptome reconstruction was adopted for genome-wide annotation and functional characterization of novel lncRNAs.[12-14,16-18] Moreover, by scRNA-seq, we and others observed higher cell-to-cell variation of lncRNA expression compared to mRNA expression.[26,28,30,50] The validation of defined lncRNAs, including potential novel ones, with quantitative RT-PCR in single cells and a new set of sorted bulk samples proved the validity of scRNA-seq and bioinformatic analysis in defining lncRNAs in the current study. Our strategy of single cell deep sequencing in combination with *de novo* transcript assembly could be adopted to further facilitate annotation of the complete lncRNA repertoire.

The very large number of both annotated and novel lncRNAs presents a challenge to functional validation. Based on earlier studies,[4,15,39-41] we adopted a systematic, computational guilt-by-association method, from which we could confirm defined lncRNAs in human HSPCs to be likely involved in hematopoietic differentiation and anticipated cell functions. Conventional functional validation of the many hundreds of known and new lncRNAs would not only be prohibitively costly and time-consuming, but the choice of assays and conditions of testing is not obvious, nor is there an established statistic by which to judge

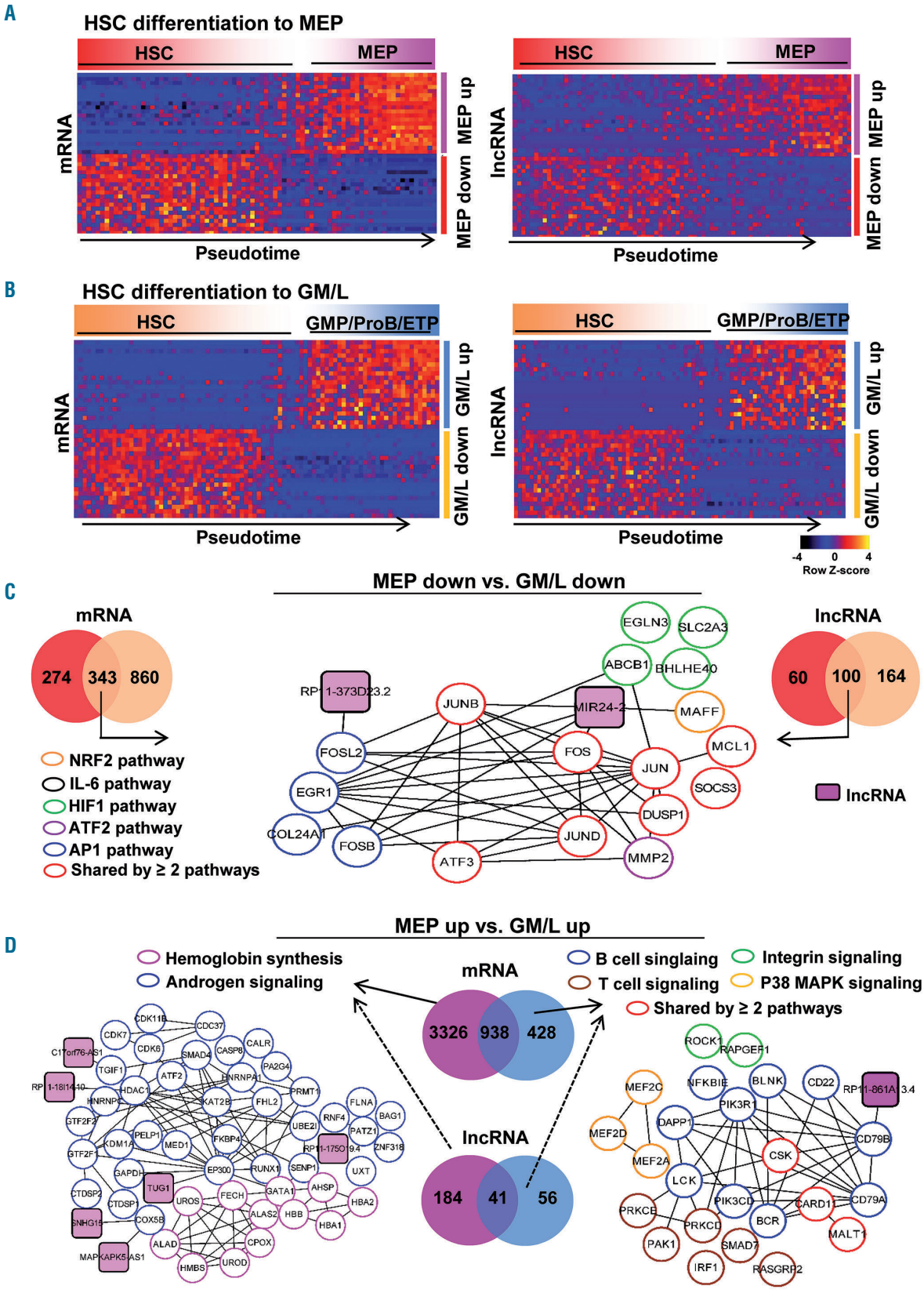



**Figure 5. Dynamically expressed long noncoding RNAs in differentiation.** Expression of sequentially upregulated/downregulated messenger RNAs (mRNAs) (left) and long noncoding RNAs (lncRNAs) (right) from HSC to MEP (A), and to GMP/ProB/ETP (B). MEP downregulated genes (red), MEP upregulated genes (pink), GM/L downregulated genes (orange), and GM/L upregulated genes (blue). (C) A network of commonly downregulated mRNAs and lncRNAs in NRF2, IL-6, HIF1, ATF2, and AP1 signaling pathways. (D) A network of mRNAs and lncRNAs specifically upregulated in MEP in hemoglobin synthesis and androgen signaling pathways (left); and a network of mRNAs and lncRNAs specifically upregulated in GM/L in B-cell, T-cell, and integrin signaling pathways (right). HSC: hematopoietic stem cell; MLP: multilymphoid progenitor; MEP: megakaryocyte-erythroid progenitor; GMP: granulocyte-monocyte progenitor, ProB: pro-B cell; ETP: earliest thymic progenitor; GM/L: granulocyte/monocyte/lymphocyte progenitor.

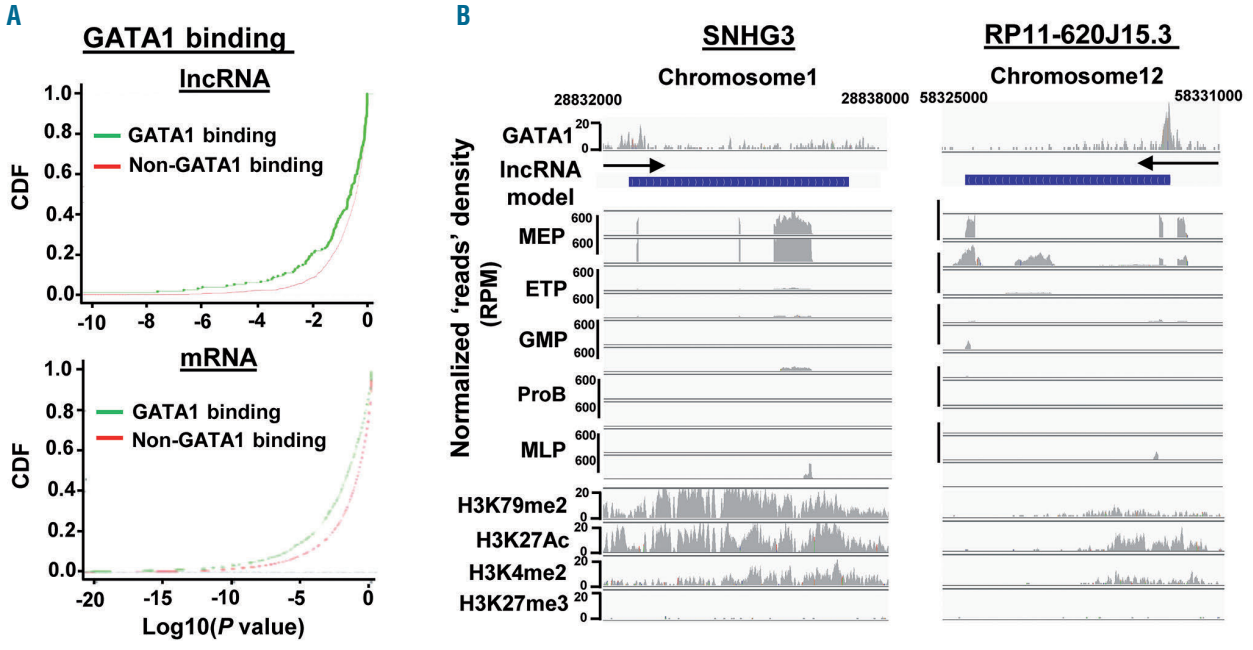


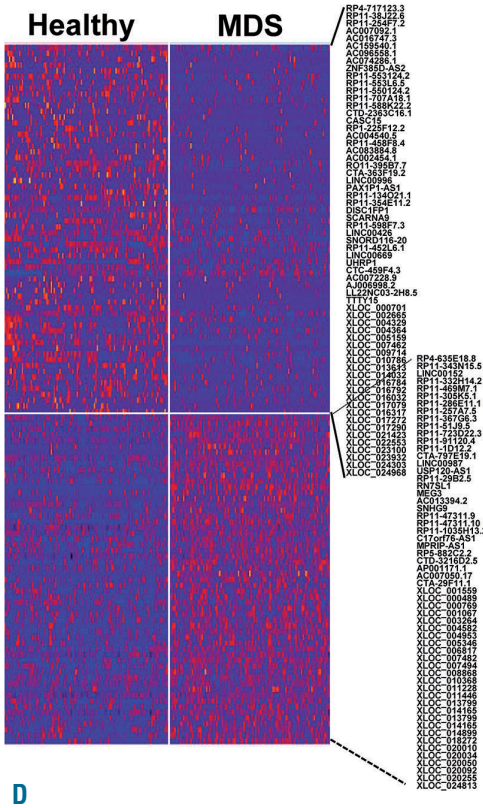

**Figure 6. Transcription factor occupancy and epigenetic modification of long noncoding RNAs during hematopoietic differentiation.** (A) Cumulative distribution of genes [long noncoding RNA (lncRNA)-encoding, up; protein-coding, down] with or without GATA1 binding at promoters. x axis, $\log_{10}$(*P* value), indicating the significance of gene expression in MEP *versus* non-MEP cells; y axis, a cumulative distribution function (CDF) of lncRNAs (%) or messenger RNAs (mRNAs) (%). For both lncRNAs and mRNAs, the lower $\log_{10}$(*P* value), which means the higher significance of preferential gene expression in MEP cells *versus* non-MEP cells, indicated the higher GATA1 binding CDF. (B) Distribution of single cell RNA sequencing (scRNA-seq) reads across two MEP-specific lncRNAs (SHG3 and RP11-620J15.3) in MEP and other cell types, and the histone modification marks in the same region. Top tracks are images from the IGV Browser depicting scRNA-seq signals as the density of mapped scRNA-seq reads, and chromatin immunoprecipitation sequencing (ChIP-seq) signals as the density of processed signal enrichment of GATA1. Track 2 shows a lncRNA transcript model. Tracks 3 to 7 represent scRNA-seq signals of two MEP-specific lncRNAs (SHG3 and RP11-620J15.3) in two single cells of each cell type including MEP and others. Tracks 8 to 11 depict the ChIP-sequ signal for active histone modification marks (H3K79me2, H3K27Ac, and H3K4me2) and repressive histone modification mark H3K27me3 in a human erythroleukemia cell line, K562. MEP: megakaryocyte-erythroid progenitor; ETP: earliest thymic progenitor; GMP: granulocyte-monocyte progenitor, ProB: pro-B cell; MLP: multilymphoid progenitor.

correlation. We attempted to computationally distinguish lncRNA roles as primary and possibly regulatory from secondary and "epiphenomonal". To this end, we first determined whether lncRNAs were preferentially expressed in specific cell types; if so, their functions were postulated to relate to lineage-specific protein-coding genes. We then applied pseudotime ordering to reconstruct hematopoietic differentiation in order to examine dynamic gene expression. HSCs are assumed to lose "stemness" and to progressively gain restricted lineage commitment gene expression during differentiation. Indeed, we observed repression of stemness genes and activation of the cell proliferation/metabolism gene program, accompanied by activation of specific-lineage genes and repression of alternative pathway of differentiation genes. By this analysis, we defined lncRNAs that are coordinately expressed in those gene modules and thus have a greater probability of regulatory roles in lineage specification. Our data should assist in narrowing the scope of future efforts including *in vitro* perturbation and *in vivo* experiments to study functions of individual lncRNAs in hematopoiesis.

The highly ordered expression pattern of lncRNAs during hematopoiesis implies regulatory constraint. Our analysis and earlier studies[8,39,47] indicated that lncRNAs are likely regulated by cell-type specific transcription factors.[13,14,16] The observation that lncRNAs exhibited higher expression variability than did mRNAs in the same regulatory program suggests more diverse and active expression of lncRNAs. lncRNAs exert regulatory roles transcriptionally and post-transcriptionally by a variety of mechanisms.[1-6] These features of lncRNAs would make them more dynamic participants in cell states and biological processes, facilitating prompt adaptive responses to stimuli or perturbations, and add another layer of complexity in gene expression regulation and cell fate decision.

Our data indicated considerable stage- and lineage-specificity of lncRNAs in human HSPCs and potential engagement in early priming of cell fate, consistent with tissue- and cell type-specificity observed in previous studies.[5,7-9, 13-18] This conclusion was confirmed by extension to an external independent scRNA-seq study of 1,034 sorted single human HSPCs,[45] and the reproducible lineage-specificity of 35 lncRNAs in both single cells and sorted bulk samples by quantitative RT-PCR. lncRNAs often form secondary structures and there are sensitive, rapid, low-cost methods readily available for lncRNA quantification, all of which make lncRNAs promising biomarkers for disease detection, diagnosis, and prognosis. One study based on microarray assay of bone marrow mononuclear cells from 176 adult patients with MDS established a four-lncRNA risk-scoring system that correlated with distinctive clinical features, and was an independent prognostic factor for survival and leukemia transformation.[51] We also found lncRNAs to be dysregulated in MDS cells, but due to the limited number of patients, lncRNA signatures of MDS patients in the current study should be interpreted with

A

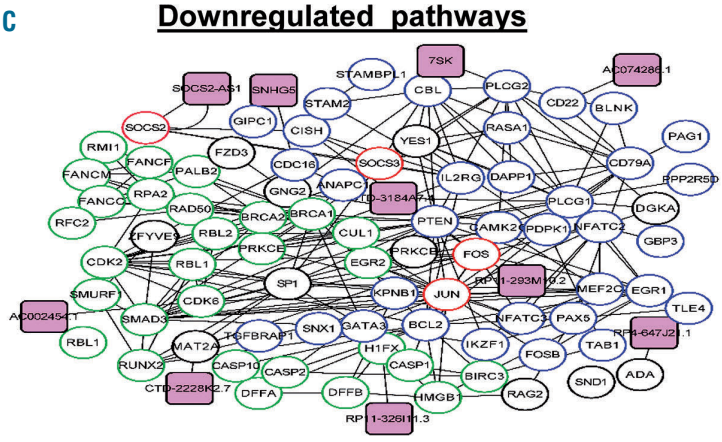



B

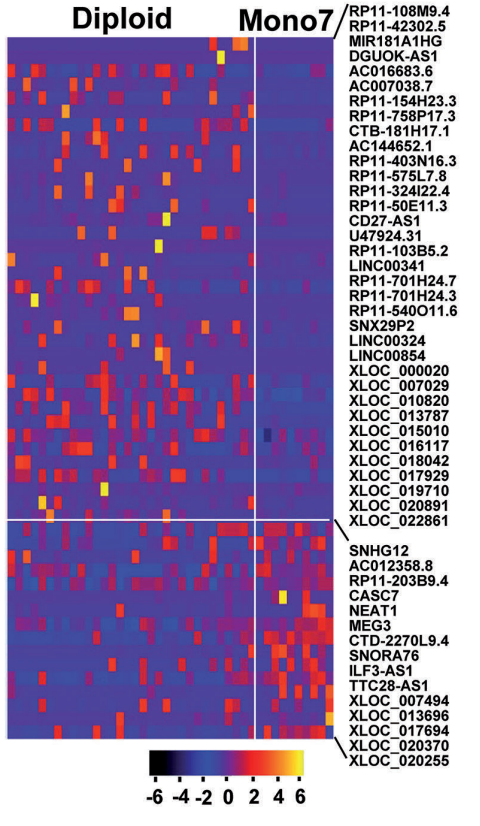
| Downregulated | | |
|---|---|---|
| **Immune response** | **Cellular response and gene expression** | **DNA damage response** |
| Immue response of phagocytosis | Unfolded protein response | DNA damage response |
| B cell activation | Stress-activated protein kinase signaling | G2 DNA damage checkpoint |
| Immunoglobulin production | Inflammatory response | |
| Antigen receptor-mediated signaling | Gene expression | |
| Leukocyte migration and proliferation | DNA methylation | |
| IL-12 production | Histone modificatoin | |

| Upregulated | |
|---|---|
| **Cell metablism** | **Cell signaling** |
| Protein synthesis | Response to IL-4 |
| Translantional inititation | Interferon gamma sinagling |
| Cellular respiration | NF-kappaB siganling |
| Lysosome organization | Cell communication |

C




D




E

| Downregulated | | |
|---|---|---|
| **Immune response** | **Cell apoptosis and death** | **DNA modificatoin** |
| B cell differentiation | Programmed cell death | DNA alkylation |
| Innate immune response | Cell cycle | DNA methylation |
| Lymphocyte mediated immunity | DNA replication | |
| Leukocyte proliferation | | |
| T cell migration | | |

| Upregulated |
|---|
| Ras signaling |
| Wnt signaling |
| IL-8 production |

F

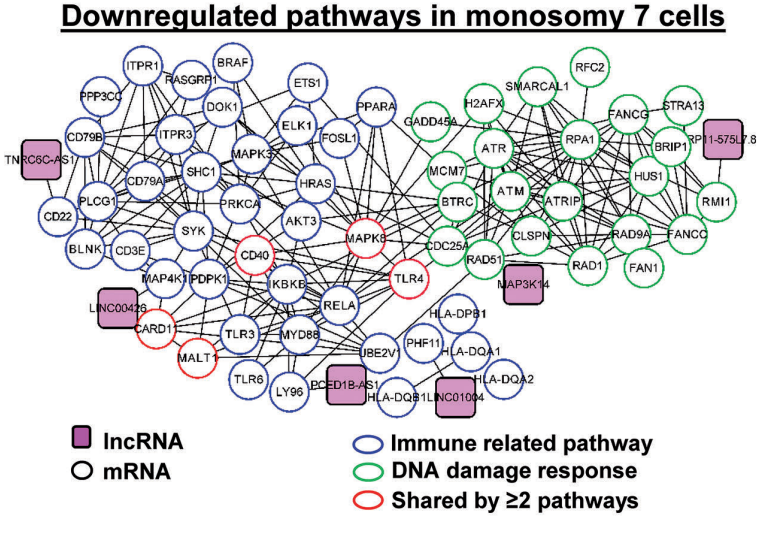



**Figure 7. Long noncoding RNAs are differentially expressed in myelodysplastic syndrome cells and aneuploid cells.** (A) A heatmap of long noncoding RNAs (lncRNAs) differentially expressed in myelodysplastic syndrome (MDS) and healthy cells. (B) Pathway analysis of downregulated and upregulated lncRNAs. (C) A network of downregulated lncRNAs with associated messenger RNAs (mRNAs) in different pathways. (D) A heatmap of lncRNAs differentially expressed in aneuploid cells compared with diploid cells. (E) Pathway analysis of downregulated and upregulated lncRNAs in aneuploid cells. (F) A network of downregulated lncRNAs with associated mRNAs in immune-related and DNA damage response pathways. Mono7: monosomy 7.

caution. Nevertheless, our results were in agreement with reported microarray data from 183 MDS patients, which related abnormal lncRNAs with gene expression, cancer, and malignancy.[52] Also, differentially expressed lncRNAs in monosomy 7 cells were involved in similar pathways as their mRNA counterparts in our previous study.[31]

Our results are not a complete profile of lncRNAs due to several limitations, especially the use of only polyA-enriched RNAs,[8] and the limited cell numbers from a few individuals due to the high cost of scRNA-seq. Additionally, annotation of novel lncRNAs is context dependent. We adopted commonly used pipelines,[12-14,16-18] but annotation might vary using different algorithms. Nevertheless, our work creates a model for future profiling of the repertoire of lncRNAs in other cell types. Lineage signatures of lncRNAs are comparison-based, and thus may vary when such comparisons are made among different subsets. Others have categorized HSCs *versus* cells of specific lineages and among differentiated cells or distinct subsets.[12-18] In contrast, we defined lncRNA signatures by making comparisons among subsets within a relatively homogeneous HSPC population, which may compromise our power to detect differences. Furthermore, pseudotime ordering reconstructs the hematopoietic hierarchy based on bioinformatic analysis of transcriptome similarity, and it has demonstrated high agreement with purified cell compartments;[44] however, dynamic gene expression in hematopoiesis might be preferably assessed in purified cell populations obtained after physical sorting based on membrane proteins, including after induction of differentiation or other *in vitro* perturbations. Given the high cell-type specificity of lncRNAs, signature lncRNAs may be superior to mRNAs in discriminating and differentiating cell subsets or new cell types that cannot be easily distinguished based on cell surface markers. We did not compare the efficacy of lncRNAs and mRNAs in defining cell types due to a lack of detailed surface marker information for single cells. Future studies with larger cell numbers, complete surface marker characterization, and whole transcriptome expression data should be of great interest in defining new cells/subtypes.

Rapid evolution and low species conservation are features of lncRNAs,[10,11] making a human catalog a prerequisite to successful, clinically relevant lncRNA studies. Based on next-generation sequencing and single cell technology, we provide a global database that should be foundational for future studies of lncRNA biology in human HSPCs.

***Acknowledgments***

*The authors acknowledge the support of the Trans-NIH Center for Human Immunology, Autoimmunity, and Inflammation (National Institute of Allergy and Infectious Diseases, NIH, Bethesda, MD, USA). We thank patients and healthy volunteers who donated bone marrow. Sequencing and technical support were provided by the DNA Sequencing and Genomics Core of NHLBI. FACS sorting was performed by Keyvan Keyvanfar and the Flow Cytometry Core of NHLBI. This research was supported by an Intramural Research Program of the National Heart, Lung, and Blood Institute.*

## References

1. Alvarez-Dominguez JR, Lodish HF. Emerging mechanisms of long noncoding RNA function during normal and malignant hematopoiesis. Blood. 2017;130(18):1965-1975.
2. Satpathy AT, Chang HY. Long noncoding RNA in hematopoiesis and immunity. Immunity. 2015;42(5):792-804.
3. Engreitz JM, Ollikainen N, Guttman M. Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. Nat Rev Mol Cell Biol. 2016;17(12):756-770.
4. 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(9):1775-1789.
5. Cabili MN, Dunagin MC, McClanahan PD, et al. Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. Genome Biol. 2015; 16:20.
6. Djebali S, Davis CA, Merkel A, et al. Landscape of transcription in human cells. Nature. 2012;489(7414):101-108.
7. Ulitsky I, Bartel DP. lincRNAs: genomics, evolution, and mechanisms. Cell. 2013;154 (1):26-46.
8. 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(18):1915-1927.
9. Washietl S, Kellis M, Garber M. Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals. Genome Res. 2014;24(4):616-628.
10. Pang KC, Frith MC, Mattick JS. Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. Trends Genet. 2006;22(1):1-5.
11. Wang J, Zhang J, Zheng H, et al. Mouse transcriptome: neutral evolution of 'non-coding' complementary DNAs. Nature. 2004;431 (7010):1 p following 757; discussion following 757.
12. Luo M, Jeong M, Sun D, et al. Long non-coding RNAs control hematopoietic stem cell function. Cell Stem Cell. 2015;16(4):426-438.
13. Alvarez-Dominguez JR, Hu W, Yuan B, et al. Global discovery of erythroid long noncoding RNAs reveals novel regulators of red cell maturation. Blood. 2014;123(4):570-581.
14. Paralkar VR, Mishra T, Luan J, et al. Lineage and species-specific long noncoding RNAs during erythro-megakaryocytic development. Blood. 2014;123(12):1927-1937.
15. Schwarzer A, Emmrich S, Schmidt F, et al. The non-coding RNA landscape of human hematopoiesis and leukemia. Nat Commun. 2017;8(1):218.
16. Hu G, Tang Q, Sharma S, et al. Expression and regulation of intergenic long noncoding RNAs during T cell development and differentiation. Nat Immunol. 2013;14(11):1190-1198.
17. Ranzani V, Rossetti G, Panzeri I, et al. The long intergenic noncoding RNA landscape of human lymphocytes highlights the regulation of T cell differentiation by linc-MAF-4. Nat Immunol. 2015;16(3):318-325.
18. Brazão TF, Johnson JS, Müller J, et al. Long noncoding RNAs in B-cell development and activation. Blood. 2016;128(7):e10-19.
19. Collier SP, Collins PL, Williams CL, et al. Cutting edge: influence of Tmevpg1, a long intergenic noncoding RNA, on the expression of Ifng by Th1 cells. J Immunol. 2012;189(5):2084-2088.
20. Vigneau S, Rohrlich PS, Brahic M, et al. Tmevpg1, a candidate gene for the control of Theiler's virus persistence, could be implicated in the regulation of gamma interferon. J Virol. 2003;77(10):5632-5638.
21. Iyer MK, Niknafs YS, Malik R, et al. The landscape of long noncoding RNAs in the human transcriptome. Nat Genet. 2015;47 (3):199-208.
22. Lei L, Xia S, Liu D, Li X, et al. Genome-wide characterization of lncRNAs in acute myeloid leukemia. Brief Bioinform. 2018. 19(4):627-635.
23. Heward JA, Lindsay MA. Long non-coding RNAs in the regulation of the immune response. Trends Immunol. 2014;35(9):408-419.
24. Ulitsky I, Bartel DP. lincRNAs: genomics, evolution, and mechanisms. Cell. 2013;154 (1):26-46.
25. 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(18):1915-1927.
26. Wang J, Roy B. Single-cell RNA-seq reveals lincRNA expression differences in Hela-S3 cells. Biotechnol Lett. 2017;39(3):359-366.
27. Kim DH, Marinov GK, Pepke S, et al. Single-

cell transcriptome analysis reveals dynamic changes in lncRNA expression during reprogramming. Cell Stem Cell. 2015;16(1):88-101.
28. Liu SJ, Nowakowski TJ, Pollen AA, et al. Single-cell analysis of long non-coding RNAs in the developing human neocortex. Genome Biol. 2016;17:67.
29. Gawronski KAB, Kim J. Single cell transcriptomics of noncoding RNAs and their cell-specificity. Wiley Interdiscip Rev RNA. 2017;8(6).
30. Hu W, Wang T, Yang Y, et al. Tumor heterogeneity uncovered by dynamic expression of long noncoding RNA at single-cell resolution. Cancer Genet. 2015;208(12):581-586.
31. Zhao X, Gao S, Wu Z, et al. Single-cell RNA-seq reveals a distinct transcriptome signature of aneuploid hematopoietic cells. Blood. 2017;130(25):2762-2773.
32. Liao Y, Smyth GK, Shi W, et al. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923-930.
33. Trapnell C, Roberts A, Goff L, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 2012;7(3):562-578.
34. Wang L, Park HJ, Dasari S, et al. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. Nucleic Acids Res. 2013;41(6):e74.
35. Rice P, Longden I, Bleasby A. et al. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 2000;16(6):276-277.
36. Harrow J, Frankish A, Gonzalez JM, et al. GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. 2012;22(9):1760-1774.
37. Hon CC, Ramilowski JA, Harshbarger J, et al. An atlas of human long non-coding RNAs with accurate 5' ends. Nature. 2017;543:199-204.
38. Zhang K, Huang K, Luo Y, et al. Identification and functional analysis of long non-coding RNAs in mouse cleavage stage embryonic development based on single cell transcriptome data. BMC Genomics. 2014;15:845.
39. Guttman M, Amit I, Garber M, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature. 2009;458(7235):223-227.
40. Huarte M, Guttman M, Feldser D, et al. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. Cell. 2010;142(3):409-419.
41. Yan X, Hu Z, Feng Y, et al. Comprehensive genomic characterization of long non-coding RNAs across human cancers. Cancer Cell. 2015;28(4):529-540.
42. Laurenti E, Doulatov S, Zandi S, et al. The transcriptional architecture of early human hematopoiesis identifies multilevel control of lymphoid commitment. Nat Immunol. 2013;14(7):756-763.
43. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.
44. Velten L, Haas SF, Raffel S, et al. Human haematopoietic stem cell lineage commitment is a continuous process. Nat Cell Biol. 2017;19(4):271-281.
45. Stachura DL, Chou ST, Weiss MJ. Early block to erythromegakaryocytic development conferred by loss of transcription factor GATA-1. Blood. 2006;107(1):87-97.
46. Shivdasani RA, Fujiwara Y, McDevitt MA, et al. A lineage-selective knockout establishes the critical role of transcription factor GATA-1 in megakaryocyte growth and platelet development. EMBO J. 1997;16(13):3965-3973.
47. Guttman M, Donaghey J, Carey BW, et al. lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature. 2011;477(7364):295-300.
48. Ezkurdia I, Juan D, Rodriguez JM, et al. Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. Hum Mol Genet. 2014;23(22):5866-5878.
49. Mattick JS, Rinn JL. Discovery and annotation of long noncoding RNAs. Nat Struct Mol Biol. 2015;22(1):5-7.
50. Zhou F, Li X, Wang W, et al. Tracing haematopietic stem cell formation at single-cell resolution. Nature. 2016;533(7604):487-492.
51. Yao CY, Chen CH, Huang HH, et al. A 4-lncRNA scoring system for prognostication of adult myelodysplastic syndromes. Blood Adv. 2017;1(19):1505-1516.
52. Liu K, Beck D, Thoms JAI, et al. Annotating function to differentially expressed LincRNAs in myelodysplastic syndrome using a network-based method. Bioinformatics. 2017;33(17):2622-2630.