

# Journal of Visualized Experiments

## Identification of Key Factors Regulating Self-renewal and Differentiation in EML Hematopoietic Precursor Cells by RNA-sequencing Analysis --Manuscript Draft--

<b>Manuscript Number:</b>	JoVE52104R3
<b>Full Title:</b>	Identification of Key Factors Regulating Self-renewal and Differentiation in EML Hematopoietic Precursor Cells by RNA-sequencing Analysis
<b>Article Type:</b>	Methods Article - JoVE Produced Video
<b>Keywords:</b>	EML Cells; Self-renewal; differentiation; Hematopoietic precursor cell; RNA-Sequencing; Data analysis
<b>Manuscript Classifications:</b>	1.11.148.378: Hematopoietic Stem Cells (HSC); 5.5.393.760.319: High-Throughput Nucleotide Sequencing; 95.51: Life Sciences (General)
<b>Corresponding Author:</b>	Jiaqian Wu, Ph.D The University of Texas Medical School at Houston Houston, TX UNITED STATES
<b>Corresponding Author Secondary Information:</b>	
<b>Corresponding Author E-Mail:</b>	Jiaqian.Wu@uth.tmc.edu
<b>Corresponding Author's Institution:</b>	The University of Texas Medical School at Houston
<b>Corresponding Author's Secondary Institution:</b>	
<b>First Author:</b>	Shan Zong
<b>First Author Secondary Information:</b>	
<b>Other Authors:</b>	Shan Zong
	Shuyun Deng
	Kenian Chen
<b>Order of Authors Secondary Information:</b>	
<b>Abstract:</b>	Hematopoietic stem cells (HSCs) are used clinically for transplantation treatment to rebuild a patient's hematopoietic system in many diseases such as leukemia and lymphoma. Elucidating the mechanisms controlling HSCs self-renewal and differentiation is important for application of HSCs for research and clinical use. However, it is not possible to obtain large quantity of HSCs due to their inability to proliferate in vitro. To overcome this hurdle, we used a mouse bone marrow derived cell line, the EML (Erythroid, Myeloid, and Lymphocytic) cell line, as a model system for this study.
	RNA-sequencing (RNA-Seq) has been increasingly used to replace microarray for gene expression studies. We report here a detailed method of using RNA-Seq technology to investigate the potential key factors in the decision between self-renewal and differentiation in EML cells. The protocol provided in this paper is divided into three parts. The first part explains how to culture EML cells and separate Lin-CD34+ and Lin-CD34- cells. The second part of the protocol offers detailed procedures for total RNA preparation and the subsequent library construction for high-throughput sequencing. The last part describes the method for RNA-Seq data analysis and how to use the data to identify differentially expressed transcription factors between Lin-CD34+ and Lin-CD34- cells. The most significantly differentially expressed transcription factors were identified to be the potential key regulators controlling EML cell self-renewal and differentiation. In the discussion section of this paper, we highlight the key steps for successful performance of this experiment.
	In summary, this paper offers a method of using RNA-Seq technology to identify

	potential regulators of self-renewal and differentiation in EML cells. The key factors identified are subjected to downstream functional analysis in vitro and in vivo.
<b>Author Comments:</b>	
<b>Additional Information:</b>	
<b>Question</b>	<b>Response</b>
If this article needs to be "in-press" by a certain date to satisfy grant requirements, please indicate the date below and explain in your cover letter.	
If this article needs to be filmed by a certain date to due to author/equipment/lab availability, please indicate the date below and explain in your cover letter.	



**Medical School**

**Department of Neurosurgery**

March 10, 2014

Dr. Nandita Singh  
Senior Science Editor  
17 Sellers Street  
Cambridge, Massachusetts 02139  
United States

Dear Nandita:


Thank you for inviting us to publish a methods video-article in the Journal of Visualized Experiments. We are submitting the enclosed manuscript "Identification of Key Factors Regulating Self-renewal and Differentiation in EML Hematopoietic Precursor Cells by RNA-sequencing Analysis" for consideration.

RNA-sequencing (RNA-Seq) technology is a powerful method for global gene expression study. We report here a detailed method of using RNA-Seq technology to identify the potential key factors in the decision between self-renewal and differentiation in EML (Erythroid, Myeloid, and Lymphocytic) multipotential hematopoietic precursor cells. Murine bone marrow derived EML are ideal for studying the molecular control of early hematopoietic differentiation at a large scale. EML cells can grow in large culture and give rise to a mixture of self-renewing Lin-SCA+CD34+ cells and partially differentiated non-renewing Lin-SCA-CD34- cells in a cell autonomous fashion. We hypothesized that key transcription factors (TFs) control a transcriptional regulatory network that determines the choice between self-renewal and differentiation. In this paper, we first introduced the methods for EML cell culture and Lin-CD34+ and Lin-CD34- cell separation. Subsequently, we described the procedures of RNA sequencing sample preparation from Lin-CD34+ and Lin-CD34- cells. Next we gave a detailed description on the sequencing data analysis and the identification of significantly differentially expressed transcription factors that potentially play a role in the decision of EML cell self-renewal and differentiation. In this paper, we also highlighted the important steps for successful EML cell culture, RNA quality control, library preparation, and accurate estimation of gene expression levels for both known and novel transcripts. Overall, this paper gives detailed instructions on how to identify potential key transcriptional regulators for EML cell self-renewal/differentiation switch in the early hematopoietic development using RNA-sequencing technology. We believe that this

work will be of high interest to scientists in hematopoietic cell development and gene transcription field in general.

Potential reviewers include Zhengdong Zhang, Scott Tenenbaum, Gary Bader, Nicholas Luscombe, Paul Bertone and Huck Ng. We would prefer that Barbara Wold, Benjamin Blencowe, Peggy Farnham, George Stamatoyannopoulos, Stuart Orkin and Bing Ren do not review the manuscript because of competing interests. Thank you very much.

Sincerely,

A handwritten signature in cursive script that reads "Jiaqian Wu".

Jiaqian Wu  
Assistant Professor  
The Vivian L. Smith Department of Neurosurgery  
Center for Stem Cell and Regenerative Medicine  
Investigator, the Senator Lloyd & B.A. Bentsen Center for Stroke Research  
Investigator, NIH Pathway to Independence Award (K99/R00)  
The University of Texas Medical School at Houston

**TITLE:** Identification of Key Factors Regulating Self-renewal and Differentiation in EML Hematopoietic Precursor Cells by RNA-sequencing Analysis

**AUTHORS:** Shan Zong, Shuyun Deng, Kenian Chen

**AUTHOR AFFILIATION:**

Zong, Shan\*  
The Vivian L. Smith Department of Neurosurgery  
Center for Stem Cell and Regenerative Medicine  
University of Texas Health Science Center,  
Houston, USA  
[Shan.zong@uth.tmc.edu](mailto:Shan.zong@uth.tmc.edu)

Deng, Shuyun\*  
The Vivian L. Smith Department of Neurosurgery  
Center for Stem Cell and Regenerative Medicine  
University of Texas Health Science Center,  
Houston, USA  
[Shuyun.deng@uth.tmc.edu](mailto:Shuyun.deng@uth.tmc.edu)

Chen, Kenian  
The Vivian L. Smith Department of Neurosurgery  
Center for Stem Cell and Regenerative Medicine  
University of Texas Health Science Center,  
Houston, USA  
[Kenian.Chen@uth.tmc.edu](mailto:Kenian.Chen@uth.tmc.edu)

\*these authors contribute equally to this work.

**CORRESPONDING AUTHOR:**

Wu, Jiaqian  
The Vivian L. Smith Department of Neurosurgery  
Center for Stem Cell and Regenerative Medicine  
University of Texas Health Science Center,  
The University of Texas Graduate School of Biomedical Sciences at Houston,  
Houston, USA  
[jiaqian.wu@uth.tmc.edu](mailto:jiaqian.wu@uth.tmc.edu)

**KEYWORDS:**

EML Cells, Self-renewal, Differentiation, Hematopoietic precursor cell, RNA-Sequencing, Data analysis

**SHORT ABSTRACT:**

RNA-Sequencing and bioinformatics analyses were used to identify significantly and differentially expressed transcription factors in Lin-CD34<sup>+</sup> and Lin-CD34<sup>-</sup> subpopulations of mouse EML cells. These transcription factors might play important roles in determining the switch between self-renewing Lin-CD34<sup>+</sup> and partially differentiated Lin-CD34<sup>-</sup> cells.

**LONG ABSTRACT:**

Hematopoietic stem cells (HSCs) are used clinically for transplantation treatment to rebuild a patient's hematopoietic system in many diseases such as leukemia and lymphoma. Elucidating the mechanisms controlling HSCs self-renewal and differentiation is important for application of HSCs for research and clinical uses. However, it is not possible to obtain large quantity of HSCs due to their inability to proliferate *in vitro*. To overcome this hurdle, we used a mouse bone marrow derived cell line, the EML (Erythroid, Myeloid, and Lymphocytic) cell line, as a model system for this study.

RNA-sequencing (RNA-Seq) has been increasingly used to replace microarray for gene expression studies. We report here a detailed method of using RNA-Seq technology to investigate the potential key factors in regulation of EML cell self-renewal and differentiation. The protocol provided in this paper is divided into three parts. The first part explains how to culture EML cells and separate Lin-CD34<sup>+</sup> and Lin-CD34<sup>-</sup> cells. The second part of the protocol offers detailed procedures for total RNA preparation and the subsequent library construction for high-throughput sequencing. The last part describes the method for RNA-Seq data analysis and explains how to use the data to identify differentially expressed transcription factors between Lin-CD34<sup>+</sup> and Lin-CD34<sup>-</sup> cells. The most significantly differentially expressed transcription factors were identified to be the potential key regulators controlling EML cell self-renewal and differentiation. In the discussion section of this paper, we highlight the key steps for successful performance of this experiment.

In summary, this paper offers a method of using RNA-Seq technology to identify potential regulators of self-renewal and differentiation in EML cells. The key factors identified are subjected to downstream functional analysis *in vitro* and *in vivo*.

## INTRODUCTION:

Hematopoietic stem cells are rare blood cells that reside mainly in the adult bone marrow niche. They are responsible for the production of cells required to replenish the blood and the immune systems<sup>1</sup>. As a kind of stem cells, HSCs are capable of both self-renewal and differentiation. Elucidating mechanisms that control the fate decision of HSCs, toward either self-renewal or differentiation, will offer valuable guidance on the manipulation of HSCs for blood disease researches and clinical usage<sup>2</sup>. One problem faced by the researchers is that HSCs can be maintained and expanded *in vitro* to a very limited extent; the vast majority of their progeny are partially differentiated in culture<sup>2</sup>.

In order to identify key regulators that control the processes of self-renewal and differentiation at a genome-wide scale, we used a mouse primitive hematopoietic progenitor cell line EML as a model system. This cell line was derived from murine bone marrow<sup>3,4</sup>. When fed with different growth factors, EML cells can differentiate into erythroid, myeloid, and lymphoid cells *in vitro*<sup>5</sup>. Importantly, this cell line can be propagated in large quantity in culture medium containing stem cell factor (SCF) and still retaining their multipotentiality. EML cells can be separated into subpopulations of self-renewing Lin-SCA+CD34+ and partially differentiated Lin-SCA-CD34- cells based on surface markers CD34 and SCA<sup>6</sup>. Similar to short-term HSCs, SCA+CD34+ cells are able of self-renewal. When treated with SCF, Lin-SCA+CD34+ cells can rapidly regenerate a mixed population of Lin-SCA+CD34+ and Lin-SCA-CD34- cells and continue to proliferate<sup>6</sup>. Although the two populations are similar in morphology and have similar levels of c-kit mRNA and protein<sup>6</sup>, Lin-SCA-CD34- cells are capable of propagating in media containing IL-3 instead of SCF<sup>3</sup>. Unveiling the key regulators in the EML cell fate decision will offer better understanding of cellular and molecular mechanisms in early developmental transition during hematopoiesis.

In order to investigate the underlying molecular differences between the self-renewing Lin-SCA+CD34+ and partially differentiated Lin-SCA-CD34- cells, we used RNA-Seq to identify differentially expressed genes. In particular, we focus on transcription factors, as transcription factors are crucial in determining cell fate. RNA-Seq is a recently developed approach that utilizes the capabilities of next-generation sequencing (NGS) technologies to profile and quantify RNAs transcribed from genome<sup>7,8</sup>. In brief, total RNA is poly-A selected and fragmented as the initial template. The RNA template is then converted into cDNA using reverse transcriptase. In order to map full-length RNA transcripts, using intact, non-degraded RNA for constructing cDNA library is important. For the purpose of sequencing, specific adapter sequences are added to both ends of cDNA. Then, in most cases, cDNA molecules are amplified by PCR and each cDNA fragment is then sequenced in a high-throughput manner.

After sequencing, the resulting reads can be aligned to a reference genome and a transcriptome database. The number of reads that map to the reference gene is counted and this information can be used to estimate the gene expression level. The reads can also be assembled *de novo* without a reference genome, enabling the study of transcriptomes in non-model organisms<sup>9</sup>. RNA-seq technology has also been used to detect splice isoforms<sup>10-12</sup>, novel transcripts<sup>13</sup> and gene fusions<sup>14</sup>. In addition to the detection of protein-coding genes, RNA-Seq

can also be used to detect novel and analyze transcription level of non-coding RNAs, such as long non-coding RNA<sup>15,16</sup>, microRNA<sup>17</sup>, siRNA *etc.*<sup>18</sup>. Because of the accuracy of this method, it has been utilized for detection of single nucleotide variations<sup>19,20</sup>.

Before the advent of RNA-Seq technology, microarray was the main method used for analyzing gene expression profile. Pre-designed probes are synthesized and subsequently attached to a solid surface to form a microarray slide<sup>21</sup>. mRNA is extracted and converted to cDNA. During the reverse transcription process, fluorescently labeled nucleotides are incorporated into the cDNA and the cDNA can be hybridized onto the microarray slides. The intensity of the signal collected from a specific spot depends on the amount of cDNA binding to the specific probe on that spot<sup>21</sup>. Compared with RNA-Seq technology, microarray has several limitations. First, microarray relies on the pre-existing knowledge of gene annotation, while RNA-Seq technology is able to detect novel transcripts. In addition, owing to the hybridization process, microarray has a relative high background level, which limits its use when gene expression level is low. Besides, the RNA-Seq technology has much higher dynamic range of detection (8000 folds)<sup>7</sup>, whereas, due to background and saturation of signals, the accuracy of microarray is limited for both highly and lowly expressed genes<sup>7,22</sup>. Finally, microarray probes differ in their hybridization efficiencies, which make the results less reliable when comparing relative expression levels of different transcripts within one sample<sup>23</sup>. Although RNA-Seq has many advantages over microarray, its data analysis is complex. This is one of the reasons that many researchers still use microarray instead of RNA-Seq. Various bioinformatics tools are required for RNA-Seq data processing and analysis<sup>24</sup>.

Among several next-generation sequencing (NGS) platforms, 454, Illumina, SOLID and Ion Torrent are the most widely used ones. 454 was the first commercial NGS platform. In contrast to the other sequencing platforms such as Illumina and SOLID, the 454 platform generates longer read length (average 700 base reads)<sup>25</sup>. Longer reads are better for initial characterization of transcriptome due to their higher assemble efficiency<sup>25</sup>. The main disadvantage of the 454 platform is its high cost per megabase of sequence. The Illumina and SOLID platforms generate reads with increased numbers and short lengths. The cost per megabase of sequence is much lower than the 454 platform. Due to the large numbers of short reads for the Illumina and SOLID platforms, data analysis is much more computationally intensive. The price of the instrument and reagents for sequencing for the Ion Torrent platform is cheaper and the sequencing time is shorter<sup>25</sup>. However, the error rate and the cost per megabase of sequence are higher compared to the Illumina and SOLID platforms. Different platforms have their own advantages and disadvantages and require different methods for data analysis. The platform should be chosen based on the sequencing purpose and the availability of funding.

In this paper, we take Illumina RNA-Seq platform as an example. We used EML cell as a model system to investigate the key regulators in EML cell self-renewal and differentiation, and provided a detailed methods of RNA-Seq library construction and data analysis for expression level calculation and novel transcript detection. We have shown in our previous publication that RNA-seq study in EML model system<sup>2</sup>, when coupled with functional test (e.g. shRNA



knockdown) provide a powerful approach in understanding the molecular mechanism of the early stages of hematopoietic differentiation, and can serve as a model for the analysis of cell self-renewal and differentiation in general.

## PROTOCOL:

1. EML cell culture and separation of Lin-CD34<sup>+</sup> and Lin-CD34<sup>-</sup> cells using magnetic cell sorting system and fluorescence-activated cell sorting method.

1.1) Preparation of Baby hamster kidney (BHK) cell culture medium for stem cell factor collection.

1.1.1) Culture BHK cells in DMEM medium containing 10% FBS in 25 cm<sup>2</sup> flask (Table 1) at 37 °C, 5% CO<sub>2</sub> in a cell culture incubator.

1.1.2) When cells grow to 80-90% confluence, wash cells once with 10 ml of PBS. Add 5 ml of 0.25% trypsin-EDTA solution to the monolayer and incubate the cells for 1 to 5 minutes (min) at room temperature (RT) until the cells are detached.

1.1.3) Pipet the solution up and down gently to break up clumps of cells. Add 5 ml of complete DMEM to the flask to stop trypsin activity. Collect cells by centrifugation at 200 x g for 5 min at RT.

1.1.4) Remove the medium and resuspend the cell pellet in 10 ml of fresh BHK cell culture medium.

1.1.5) Transfer 2 ml of the cell suspension from the step 1.1.4 to a new 75 cm<sup>2</sup> flask and add 48 ml of fresh BHK cell culture medium to the flask.

1.1.6) Culture the BHK cells for two days and collect the culture medium. Passage the medium through a 0.45 µm filter. Store the medium in -20 °C until further use.

1.2) EML cell culture

1.2.1) Culture EML cells (in suspension) in EML basic medium containing BHK cell culture medium (Table 1) at 37 °C, 5% CO<sub>2</sub> in a cell culture incubator.

1.2.2) Maintain the EML cells at low cell density (0.5-5 x 10<sup>5</sup> cells/ml) with the peak density less than 6 x 10<sup>5</sup> cells/ml. Split the cells every 2 to 3 days at the ratio of 1:5. Passage EML cells gently and discard the culture after passaging for 10 generations.

1.3) Depletion of lineage positive cells.

1.3.1) Harvest the EML cells by centrifugation at 200 x g for 5 min and wash the cells once with PBS. Collect the cells by centrifugation at 200 x g for 5 min.

1.3.2) Resuspend the cells with PBS and count the cells with a hemocytometer. Determine the antibody concentration in the subsequent cell separation step according to the number of the cells (Please refer to the instructions offered by the provider of the cell isolation system).

1.3.3) Isolate the lineage negative (Lin-) cells using lineage antibody cocktail (Cocktail of biotin-conjugated monoclonal antibodies CD5, CD45R (B220), CD11b, Anti-Gr-1 (Ly-6G/C), 7-4 and Ter-119) and a magnetic activated cell sorting system according to manufacturer's instructions.

#### 1.4) Separation of Lin-CD34+ and Lin-CD34- cells

1.4.1) Spin down the Lin- cells from the step 1.3.3 at 200 x g for 5 min. Resuspend the cell pellet with PBS and count the cells with a hemocytometer.

1.4.2) Wash the cells twice with FACS buffer and pellet the cells at 200 x g for 5 min.

1.4.3) Label five 1.5 ml microcentrifuge tubes with the number 1, 2, 3, 4, 5 respectively. Resuspend the cells with 100 µl FACS buffer per 10<sup>6</sup> cells (10<sup>6</sup> cells per tube).

1.4.4) Add 1 µg of Anti-Mouse CD34 FITC antibody to tube 1 and tube 2 and mix the tubes gently.

1.4.5) Incubate all tubes at 4 °C for 1 hour in the dark.

1.4.6) Add 0.25 µg of PE-conjugated Anti-Sca1 antibody and 20 µl of APC-conjugated Lineage Cocktail antibodies to tube 1, 0.25 µg of PE-conjugated Anti-Sca1 antibody to tube 3, and 20 µl of APC-conjugated Lineage Cocktail antibodies to tube 4.

1.4.7) Mix all the tubes gently and incubate the cells on ice for additional 30 min in the dark.

1.4.8) Add 300 µl of FACS buffer to the cells and spin down the cells at 200 x g for 5 min.

1.4.9) Wash the cells with 500 µl of FACS buffer for three times.

1.4.10) Resuspend the cell pellet in 500 µl of FACS buffer.

1.4.11) Use the cells in tube 2,3,4 and 5 for setting up compensation. Isolate Lin-SCA+CD34+ and Lin-SCA-CD34- cells in tube 1 using FACS Aria.

## 2. RNA preparation and library construction for high-throughput sequencing

### 2.1) Isolation, quality analysis and quantification of RNA

2.1.1) Extract total RNA from Lin-CD34+ and Lin-CD34- cells respectively using TRIzol following the manufactures' protocol.

2.1.2) Remove the contaminated DNA using Deoxyribonuclease I following the manufacture's protocol. Optionally, store the RNA at -80 °C at this step for further use.

2.1.3) Assess the quality of total RNA using Bioanalyzer according to the instructions offered by the supplier. Use RNA sample with RNA Integrity Number (RIN) larger than 9.

## 2.2) Library Construction and high-throughput sequencing.

NOTE: This protocol describes RNA-Seq using Illumina platform. For other sequencing platforms, different library preparation methods are required.

2.2.1) Use 0.1-4 µg of high quality total RNA per sample for library preparation. Normally 2 µg total RNA can be extracted from  $10^5$  EML cells.

2.2.2) Use a RNA-sequencing sample preparation system for mRNA purification and fragmentation, first and second strand cDNA synthesis, end repair, 3' ends adenylation, adapter ligation and PCR amplification, following the detailed standard procedures from the provider's instructions.

2.2.2.1) Positively select PolyA mRNA positively using oligo-dT magnetic beads and fragment.

2.2.2.2) Perform reverse transcription using random primers to obtain the cDNA and subsequently synthesized second strand cDNA is to generate double strand cDNA.

2.2.2.3) Remove the 3' overhangs and fill the 5' overhangs by DNA polymerase. Adenylate 3' Ends to prevent cDNA fragments from ligating to one another.

2.2.2.4) Add multiplex indexing adapters to both ends of the ds cDNA. Perform PCR for the enrichment of DNA fragments.

2.2.3) Measure the A260/A280 to obtain information about the concentration of library using a spectrophotometer.

2.2.4) Assess the library quality and measure the size range of DNA fragments using a Bioanalyzer.

## 3 Data analysis (For reference of software used in this part, please see table 2)

### 3.1) Data file processing for downstream analysis.

3.1.1) Convert .bcl (base call file) file to .fastq file using CASAVA software (Illumina, version 1.8.2).

3.1.1.1) Fire up the 'Terminal' in linux system. Go to the data folder that contains the data file from an illumina HiSeq2000 sequencing machine. Suppose the result folder is 'NASboy1/JiaqianLabData/HiSeq\_RUN/2013\_07\_11/130627\_SN860\_0309\_A\_2013-166\_H0PW9ADXX/', type in the command in Figure S1A, and enter the data folder.

3.1.1.2) Install CASAVA 1.8.2 in the linux system. Suppose the output folder is 'Unaligned', use the command in Figure S1B to prepare the configuration file for converting. Use the option --fastq-cluster-count 0 to ensure only one .fastq file is created for each sample. The generated .fastq file is in .gz format. Unzip it for downstream analysis (Figure S1B).

3.1.1.3) After the 'Unaligned' folder has been generated, go to the 'Unaligned' folder (Figure S1C).

3.1.1.4) Use the command in Figure S1D to begin the converting process. The '-j' parameter supplies the cpu number that will be used.

3.1.1.5) After the system finished the converting process, go to the result folder under 'Unaligned' folder (Figure S1E).

3.1.1.6) Use the command in Figure S1F to decompress the .fastq.gz file into .fastq file under each sample folder.

## 3.2) Detect novel transcripts and evaluate the expression level using Tuxedo Suite<sup>26</sup>.

3.2.1) Map the paired-end RNA-Seq reads to the mouse reference genome (UCSC version mm9, obtained from <http://cufflinks.cbc.umd.edu/igenomes.html>) using Tophat software (version 1.3.3)<sup>27</sup>, which uses the Bowtie read mapper (version 0.12.7)<sup>28</sup>. Tophat is supplied with "--no-novel-juncs" option to improve estimation accuracy of expression level.

3.2.1.1) Put the .fastq files in a folder where the mapping process will be implemented. Suppose there are 2 .fastq files (rename to Example1.read1, Example1.read2) for a paired-end sequencing sample, use the command in Figure S2 to do the mapping (adjust the parameters according to the system setting). The "-p" parameter supplies the cpu number that will be used. The "-r" and "--mate-std-dev" parameters can be obtained from library QC or inferred from a subset of aligned reads (Figure S2).

3.2.2) Assemble the mapped reads into RNA transcripts using the Cufflinks software (version 1.3.0)<sup>29</sup>. Run Cufflinks using the annotation file of known genes (same .gtf file used by Tophat) and .bam file produced by Tophat.

3.2.2.1) After Tophat finished running, in the same folder, use the command in Figure S3A to run cufflinks to construct transcriptome and estimate transcript expression level. The 'mm9\_repeatMasker.gtf' and genome sequence files in the 'GenomeSeqMM9' folder can be obtained from UCSC Genome Browser.

3.2.2.2) The resulting genes.expr and transcripts.expr files contain the expression value of genes and transcripts (isoforms). Copy and paste the file contents to an Excel file and manipulate with spreadsheet application (Figure S3B).

3.2.2.3) Use the command in Figure S3C to compare the resulting 'transcripts.gtf' file to the reference 'mm9\_genes.gtf' file in order to identify novel transcripts.

3.2.2.4) The resulting .tmap file contains the comparison result. Copy and paste the file contents to an Excel file and manipulate with spreadsheet application. Transcripts with class code 'u' can be considered as 'novel' compared to the reference .gtf file provided (Figure S3D).

NOTE: For downstream analysis convenience, set the FPKM values to 0.1 if the values are under 0.1.

Step 3.2.3 to 3.2.6 is optional for who wishes to improve accuracy of novel transcripts' expression estimation. This will take much longer time, because mapping and transcriptome construction need to be run more than once.

3.2.3) Run Tophat using default parameters and then run cufflinks to generated .gtf file using the command in Figure S3E.

3.2.4) Compare the resulting .gtf file to the reference genome .gtf file using the command in Figure S3F.

3.2.5) Parse the resulted .tmap file as described in the step 3.2.2.4. Copy and paste the file contents to an Excel file and manipulate with spreadsheet application. Transcripts with class code 'u' can be considered as 'novel' compared to the reference .gtf file provided.

3.2.6) After the step 3.2.5, there is a .combined.gtf file in the folder which can be used as the reference .gtf file. A second run of Tophat and cufflinks can be performed as described in the step 3.2.1 and 3.2.2 to obtain a more accurate FPKM estimation of novel transcripts.

### 3.3) Detect differentially expressed genes using DESeq package<sup>30</sup>.

3.3.1) The input of DESeq is a raw read counts table. To obtain such a table, use the htseq-count script distributed with the HTSeq Python package which can be downloaded from HTSeq website (<http://www-huber.embl.de/users/anders/HTSeq/doc/count.html>).

3.3.1.1) Ensure that samtools, python, and htseq-count programs are installed in the system. Obtain raw read count numbers from tophat output by using the command in Figure S4A.

3.3.1.2) Prepare 'Raw\_Count\_Table.txt', 'ExperimentDesign.txt' files using Excel. Copy and save the content in .txt format for the DESeq R package (Figure S4B).

3.3.1.3) Install R program in the system. In the terminal, type 'R' and press ENTER. A screen message will appear as showed in Figure S4C.

3.3.1.4) Read 'Raw\_Count\_Table.txt', 'ExperimentDesign.txt' into R using the command in Figure S4D.

3.3.1.5) Load DESeq package using the command in Figure S4E.

3.3.1.6) Factorize conditions in R (Figure S4F).

3.3.1.7) Use the command in Figure S4G to run negative binominal test on the normalized count table.

3.3.1.8) Use the command in Figure S4H to output significant differential expressed genes in a .csv file.

3.4) Look up transcription factors' (TFs) FPKM values across samples using Excel. Intersect DE gene table and TFs table. Genes belong to both table are differentially expressed transcription factors.

3.4.1) Go to the website <http://www.bioguo.org/AnimalTFDB/download.php> and download the transcription factors. Then lookup the DE transcription factors in the Excel (Figure S5).

3.5) Generating .bigwig file for UCSC genome browser visualization.

3.5.1) Download 'bedtools' software package from the website <https://github.com/arq5x/bedtools2> and install the software in the system<sup>31</sup>. Download the UCSC tools 'bedGraphToBigWig' from the website <http://hgdownload.cse.ucsc.edu/admin/exe/> and install the software in the system.

3.5.2) In the folder containing the .bam file, use the command in Figure S6A to convert .bam file generated by tophat into .bed file.

3.5.3) After the .bed file is produced, use the command in Figure S6B to generate .bigwig file. The file 'ChromInfo.txt' can be obtained from following url <http://hgdownload.cse.ucsc.edu/goldenPath/mm9/database/chromInfo.txt.gz>.

3.5.4) Observe a custom track on UCSC Genome Browser. Refer to the website <http://genome.ucsc.edu/goldenPath/help/customTrack.html> on how to display a custom track using UCSC genome browser.

## REPRESENTATIVE RESULTS:

In order to analyze differentially expressed genes in Lin-CD34<sup>+</sup> and Lin-CD34<sup>-</sup> EML cells, we used RNA-Seq technology. Figure 1 shows the workflow of the procedures. After isolation of lineage negative cells by magnetic cell sorting, we separated Lin-SCA<sup>+</sup>CD34<sup>+</sup> and Lin-SCA<sup>-</sup>CD34<sup>-</sup> cells using FACS Aria. Lin-enriched EML cells were stained with anti-CD34, anti-Sca1 and lineage cocktail antibodies. Only Lin<sup>-</sup> cells were gated for analysis of Sca1 and CD34 expression. Two populations (SCA<sup>+</sup>CD34<sup>+</sup> and SCA<sup>-</sup>CD34<sup>-</sup> EML cells) could be observed by FACS analysis<sup>6</sup> (Figure 2<sup>6</sup>).

After cell separation, we extracted total RNA from CD34<sup>+</sup> and CD34<sup>-</sup> cells respectively and analyzed the quality of RNA. The accuracy of RNA-Seq data largely relies on the quality of RNA-Seq library and the quality of total RNA is vital for preparing a high quality library. High quality RNA sample should have an OD 260/280 value between 1.8 and 2.0. In addition to using the spectrophotometer, RNA quality was further assessed with greater accuracy by Bioanalyzer. Figure 3 shows a result of a high quality RNA sample with the RIN equal to 9.4. Only high quality total RNA sample with RIN value greater than 9 was used for mRNA extraction and subsequent library construction procedures.

Ribosomal RNA is the most abundant type of RNA in cell. Currently, two main strategies – depletion of rRNA or positively selection of polyadenylated mRNA (poly-A mRNA) are used for enrichment of target RNA before library construction. Non polyadenylated RNA species are lost during the selection of poly-A mRNA. In contrast, rRNA depletion methods such as RiboMinus could preserve non polyadenylated RNA species. The purpose of our study is to look for differentially expressed coding genes in two cell types, thus we used the poly-A mRNA selection method for enrichment of target RNAs before library construction. When library construction was finished, the size of DNA fragments in the library was checked before sequencing using Bioanalyzer. Figure 4 shows a good quality library with the fragment size peaks at about 300 bp.

In the subsequent step, the library was subjected to high-throughput sequencing. In principle, longer read length will be helpful for read mapping. It can reduce the probability that the read is mapped to multiple locations due to similarity among duplicate genes or gene family members. As the pair-end sequencing sequences are from both ends of the fragments, the read length chosen should be less than half of the average fragments length. If the main goal of the experiment is to measure the expression level instead of constructing transcript structure, single-end read (75 or 100bp) can reduce the cost without losing too much information. Paired-end sequencing is more useful for transcript structure construction and shorter read length can be used to reduce cost. Certainly, when sufficient funding is available, longer read length is preferred.

For differential expression analysis, there are many alternative algorithms other than DESeq. There is also one included in cufflinks package named cuffdiff<sup>32</sup>. DESeq is one of the most widely used count based DE gene analysis algorithms. DESeq method is based on a well characterized statistics model --- negative binomial distribution. In our experience, DESeq is



more stable compare to cuffdiff. Early versions of cuffdiff often give significantly different numbers of DE genes. Therefore we used DESeq for DE analysis here.

Because transcription factors are crucial for cell fate determination, we focused on the significantly differentially expressed transcription factors<sup>33</sup>. The TFs changed >1.5 fold between Lin-CD34+ and Lin-CD34- were found and are shown on the heatmap (Figure 5)<sup>2</sup>. Notably, the relative expression level of *Tcf7* in Lin-CD34+ cells is more than 100 fold higher than that in Lin-CD34- cells. Thus *Tcf7* was chosen for further ChIP-Sequencing (Chromatin Immunoprecipitation and sequencing) analysis and functional test to confirm *Tcf7*'s function in regulation of EML cell self-renewal and differentiation<sup>2</sup>.

## Figure Legends:

**Figure 1: Workflow of the procedures.** Lin-CD34<sup>+</sup> and Lin-CD34<sup>-</sup> cells were separated by magnetic cell separation system and fluorescence-activated cell sorting method. Total RNA was extracted followed by mRNA purification and library construction. After analysis of library quality, samples were subjected to high throughput sequencing. Data were analyzed and differentially expressed transcription factors were identified.

**Figure 2: Separation of Lin-CD34<sup>+</sup> and Lin-CD34<sup>-</sup> EML cells<sup>6</sup>.** Lin<sup>-</sup> EML cells were enriched by magnetic cell sorting. Lin<sup>-</sup> cells were stained with anti-CD34, anti-Sca1 and lineage mixture antibodies. Lin<sup>-</sup> cells were gated for expression of CD34 and Sca1. Lin-CD34<sup>+</sup>Sca1<sup>+</sup> and Lin-CD34<sup>-</sup>Sca1<sup>-</sup> EML cell populations were sorted.

**Figure 3: A representative of high-quality total RNA sample.** The quality of total RNA was assessed by Bioanalyzer. The RNA Integrity Number is 9.4 (FU, Fluorescence Units).

**Figure 4: Fragments size range of Paired-End library.** The DNA size distribution of the library was analyzed using Bioanalyzer. Most fragments are within the size range of 250-500 bp.

**Figure 5: Differentially expressed transcription factors (>1.5 fold) between Lin-CD34<sup>+</sup> cells and Lin-CD34<sup>-</sup> cells<sup>2</sup>.** For each cell type, two independent experiments were performed. Up-regulated genes are indicated as red color and down-regulated genes are indicated as green color.

**Table 1:** Buffers and Cell culture mediums

**Table 2:** List of software for data analysis

## DISCUSSION:

Mammalian transcriptome is very complex<sup>34-38</sup>. RNA-Seq technology plays an increasingly important role in the studies of transcriptome analysis, novel transcripts detection and single nucleotide variation discovery etc.. It has many advantages over other methods for gene expression analysis. As mentioned in the introduction, it overcomes the hybridization artifacts of microarray and can be used to identify novel transcripts *de novo*. One limitation of RNA-sequencing is relative short read length comparing to Sanger Sequencing. However, with the rapid improvement of sequencing technology, read length is increasing constantly. In this paper, we provide detailed methods of using this technology to identify potential key regulators in mouse EML cell self-renewal and differentiation.

The first key step for this protocol is EML cell culture. Although EML is a hematopoietic precursor cell line and it can be propagated in large quantity with SCF. The culturing condition of EML cells requires more attention than the usual immortalized cell lines. The cells should be fed and passaged at a regular basis with gentle operation; otherwise the cells could change in their properties of self-renewal and differentiation and undergo cell death. As the first step after collecting enough cells, we isolated lineage negative cells using a magnetic activated cell sorting system. Then we separated CD34+ and CD34- cells using fluorescence-activated cell sorting. The EML cells are normally passaged less than 10 generations before using for RNA extraction and the numbers of CD34+ and CD34- cells should be similar after separation. If the two populations vary greatly in cell number, it is advisable to discard the culture and re-thaw another tube of cell stock for culture.

After separation of CD34+ and CD34- cell, total RNA extraction was performed -another important step for this study. High quality RNA is the base for construction of a high quality library, which promises the accuracy of the sequencing data. In this critical step, any contact with RNase should be avoided. All reagents should be RNase free. It is important to wear gloves at all times while handling RNA. High quality RNA sample has an OD 260/280 value between 1.8 and 2.0. When collecting the aqueous phase containing RNA, be careful not to carry any organic phase with the RNA sample. Any residual organic solvents such as phenol or chloroform in the RNA would result in an OD260/280 value lower than 1.65. If the OD260/280 value is lower than 1.65, precipitate RNA again with ethanol. After washing with 75% ethanol, do not overdry RNA pellet. Drying RNA pellet completely will affect the solubility of RNA and lead to low yield of RNA.

The next key step for this protocol is library preparation. After total RNA extraction, a step of using DNase for removal of contaminated DNA is highly recommended, since DNA contamination might result in wrong estimation of the amount of total RNA used. It is recommended to perform the downstream procedure immediately after RNA isolation, since after long-term storage and freeze-thawing procedure, RNA will degrade to some degree. If the subsequent steps after RNA isolation can not be performed immediately, store the RNA in – 80 °C. Before total RNA is used for mRNA purification and cDNA synthesis, the quality should always be checked. Only high quality RNA can be used for library preparation. Using low quality

or degraded RNA might lead to over-representation of 3'-ends. Before sequencing, library quality was assessed to ensure maximum sequencing efficiency.

In the data analysis part, after performing a run of Cufflinks without a reference transcriptome, we combined the novel transcripts with known transcripts to form a reference .gtf file and run Tophat and cufflinks for the second time. This two-run procedure is recommended, since this provide more accurate FPKM estimation than running only once. After data analysis, the differentially expressed genes were identified. Downstream experiments can be performed to validate the function of genes *in vitro* and *in vivo*. In our previous publication<sup>2</sup>, we chose the significantly differentially expressed transcription factors and identified the genome binding site of these factors by performing chromatin immunoprecipitation and sequencing (ChIP-Seq). In addition, we applied shRNA knockdown assay to test the functional effect of *Tcf7*. We found that in *Tcf7* knockdown cells, up-regulated genes were the genes highly enriched in CD34- cells, while down-regulated genes were found to be significantly enriched in CD34+ cells. Therefore, the gene expression profile of *Tcf7* knockdown cells shifted toward a partially differentiated CD34- state. Overall, using EML cell as a model system coupled with RNA-Sequencing technology and functional assays, we identified and confirmed *Tcf7* as an important regulator of EML cell self-renewal and differentiation.

**ACKNOWLEDGMENTS:**

JQW, SZ, SD and KC are supported by grant from the National Institutes of Health and the Staman Ogilvie Fund—Memorial Hermann Foundation.

**DISCLOSURES:**

The authors have nothing to disclose.

## REFERENCES:

- 1 Chambers, S. M. & Goodell, M. A. Hematopoietic stem cell aging: wrinkles in stem cell potential. *Stem Cell Rev* **3**, 201-211, doi:SCR:3:3:201 [pii] (2007).
- 2 Wu, J. Q. *et al.* Tcf7 is an important regulator of the switch of self-renewal and differentiation in a multipotential hematopoietic cell line. *PLoS genetics* **8**, e1002565, doi:10.1371/journal.pgen.1002565 (2012).
- 3 Ye, Z. J. *et al.* Complex interactions in EML cell stimulation by stem cell factor and IL-3. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 4882-4887, doi:10.1073/pnas.1018002108 (2011).
- 4 Tsai, S., Bartelmez, S., Sitnicka, E. & Collins, S. Lymphohematopoietic progenitors immortalized by a retroviral vector harboring a dominant-negative retinoic acid receptor can recapitulate lymphoid, myeloid, and erythroid development. *Genes Dev* **8**, 2831-2841 (1994).
- 5 Weiler, S. R. *et al.* D3: a gene induced during myeloid cell differentiation of Linlo c-Kit+ Sca-1(+) progenitor cells. *Blood* **93**, 527-536 (1999).
- 6 Ye, Z. J., Kluger, Y., Lian, Z. & Weissman, S. M. Two types of precursor cells in a multipotential hematopoietic cell line. *Proc Natl Acad Sci U S A* **102**, 18461-18466, doi:0509314102 [pii] 10.1073/pnas.0509314102 (2005).
- 7 Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics* **10**, 57-63, doi:10.1038/nrg2484 (2009).
- 8 Chu, Y. & Corey, D. R. RNA sequencing: platform selection, experimental design, and data interpretation. *Nucleic acid therapeutics* **22**, 271-274, doi:10.1089/nat.2012.0367 (2012).
- 9 Hornett, E. A. & Wheat, C. W. Quantitative RNA-Seq analysis in non-model species: assessing transcriptome assemblies as a scaffold and the utility of evolutionary divergent genomic reference species. *BMC genomics* **13**, 361, doi:10.1186/1471-2164-13-361 (2012).
- 10 Eswaran, J. *et al.* RNA sequencing of cancer reveals novel splicing alterations. *Scientific reports* **3**, 1689, doi:10.1038/srep01689 (2013).
- 11 Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470-476, doi:10.1038/nature07509 (2008).
- 12 Wu, J. Q. *et al.* Dynamic transcriptomes during neural differentiation of human embryonic stem cells revealed by short, long, and paired-end sequencing. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 5254-5259, doi:10.1073/pnas.0914114107 (2010).
- 13 Loraine, A. E., McCormick, S., Estrada, A., Patel, K. & Qin, P. RNA-seq of Arabidopsis pollen uncovers novel transcription and alternative splicing. *Plant physiology* **162**, 1092-1109, doi:10.1104/pp.112.211441 (2013).
- 14 Edgren, H. *et al.* Identification of fusion genes in breast cancer by paired-end RNA-sequencing. *Genome biology* **12**, R6, doi:10.1186/gb-2011-12-1-r6 (2011).
- 15 Illott, N. E. & Ponting, C. P. Predicting long non-coding RNAs using RNA sequencing. *Methods* **63**, 50-59, doi:10.1016/j.ymeth.2013.03.019 (2013).
- 16 Sun, L. *et al.* Prediction of novel long non-coding RNAs based on RNA-Seq data of mouse Klf1 knockout study. *BMC bioinformatics* **13**, 331, doi:10.1186/1471-2105-13-331 (2012).
- 17 Luo, S. MicroRNA expression analysis using the Illumina microRNA-Seq Platform. *Methods in molecular biology* **822**, 183-188, doi:10.1007/978-1-61779-427-8\_12 (2012).

- 18 Bolduc, F., Hoareau, C., St-Pierre, P. & Perreault, J. P. In-depth sequencing of the siRNAs associated with peach latent mosaic viroid infection. *BMC molecular biology* **11**, 16, doi:10.1186/1471-2199-11-16 (2010).
- 19 Chepelev, I., Wei, G., Tang, Q. & Zhao, K. Detection of single nucleotide variations in expressed exons of the human genome using RNA-Seq. *Nucleic acids research* **37**, e106, doi:10.1093/nar/gkp507 (2009).
- 20 Djari, A. *et al.* Gene-based single nucleotide polymorphism discovery in bovine muscle using next-generation transcriptomic sequencing. *BMC genomics* **14**, 307, doi:10.1186/1471-2164-14-307 (2013).
- 21 Murphy, D. Gene expression studies using microarrays: principles, problems, and prospects. *Advances in physiology education* **26**, 256-270 (2002).
- 22 Chen, K. *et al.* RNA-seq characterization of spinal cord injury transcriptome in acute/subacute phases: a resource for understanding the pathology at the systems level. *PLoS one* **8**, e72567, doi:10.1371/journal.pone.0072567 (2013).
- 23 Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M. & Gilad, Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome research* **18**, 1509-1517, doi:10.1101/gr.079558.108 (2008).
- 24 Ramskold, D., Kavak, E. & Sandberg, R. How to analyze gene expression using RNA-sequencing data. *Methods in molecular biology* **802**, 259-274, doi:10.1007/978-1-61779-400-1\_17 (2012).
- 25 Glenn, T. C. Field guide to next-generation DNA sequencers. *Mol Ecol Resour* **11**, 759-769, doi:10.1111/j.1755-0998.2011.03024.x (2011).
- 26 Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols* **7**, 562-578, doi:10.1038/nprot.2012.016 (2012).
- 27 Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-1111, doi:10.1093/bioinformatics/btp120 (2009).
- 28 Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **10**, R25, doi:10.1186/gb-2009-10-3-r25 (2009).
- 29 Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology* **28**, 511-515, doi:10.1038/nbt.1621 (2010).
- 30 Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome biology* **11**, R106, doi:10.1186/gb-2010-11-10-r106 (2010).
- 31 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033  
btq033 [pii] (2010).
- 32 Cheranova, D. *et al.* RNA-seq analysis of transcriptomes in thrombin-treated and control human pulmonary microvascular endothelial cells. *J Vis Exp*, doi:10.3791/4393  
4393 [pii] (2013).
- 33 Zhang, H. M. *et al.* AnimalTFDB: a comprehensive animal transcription factor database. *Nucleic acids research* **40**, D144-149, doi:10.1093/nar/gkr965 (2012).

- 34 Wu, J. Q. *et al.* Systematic analysis of transcribed loci in ENCODE regions using RACE sequencing reveals extensive transcription in the human genome. *Genome Biol* **9**, R3, doi:10.1186/gb-2008-9-1-r3gb-2008-9-1-r3 [pii] (2008).
- 35 Wu, J. Q. *et al.* Large-scale RT-PCR recovery of full-length cDNA clones. *Biotechniques* **36**, 690-696, 698-700 (2004).
- 36 Wu, J. Q., Shteynberg, D., Arumugam, M., Gibbs, R. A. & Brent, M. R. Identification of rat genes by TWINSKAN gene prediction, RT-PCR, and direct sequencing. *Genome Res* **14**, 665-671, doi:10.1101/gr.195960414/4/665 [pii] (2004).
- 37 Dewey, C. *et al.* Accurate identification of novel human genes through simultaneous gene prediction in human, mouse, and rat. *Genome Res* **14**, 661-664, doi:10.1101/gr.193980414/4/661 [pii] (2004).
- 38 Wu, J. *Characterize Mammalian Transcriptome Complexity*. (LAP Lambert Academic Publishing, 2011).

Figure 1  
[Click here to download Figure: Figure 1.pdf](#)

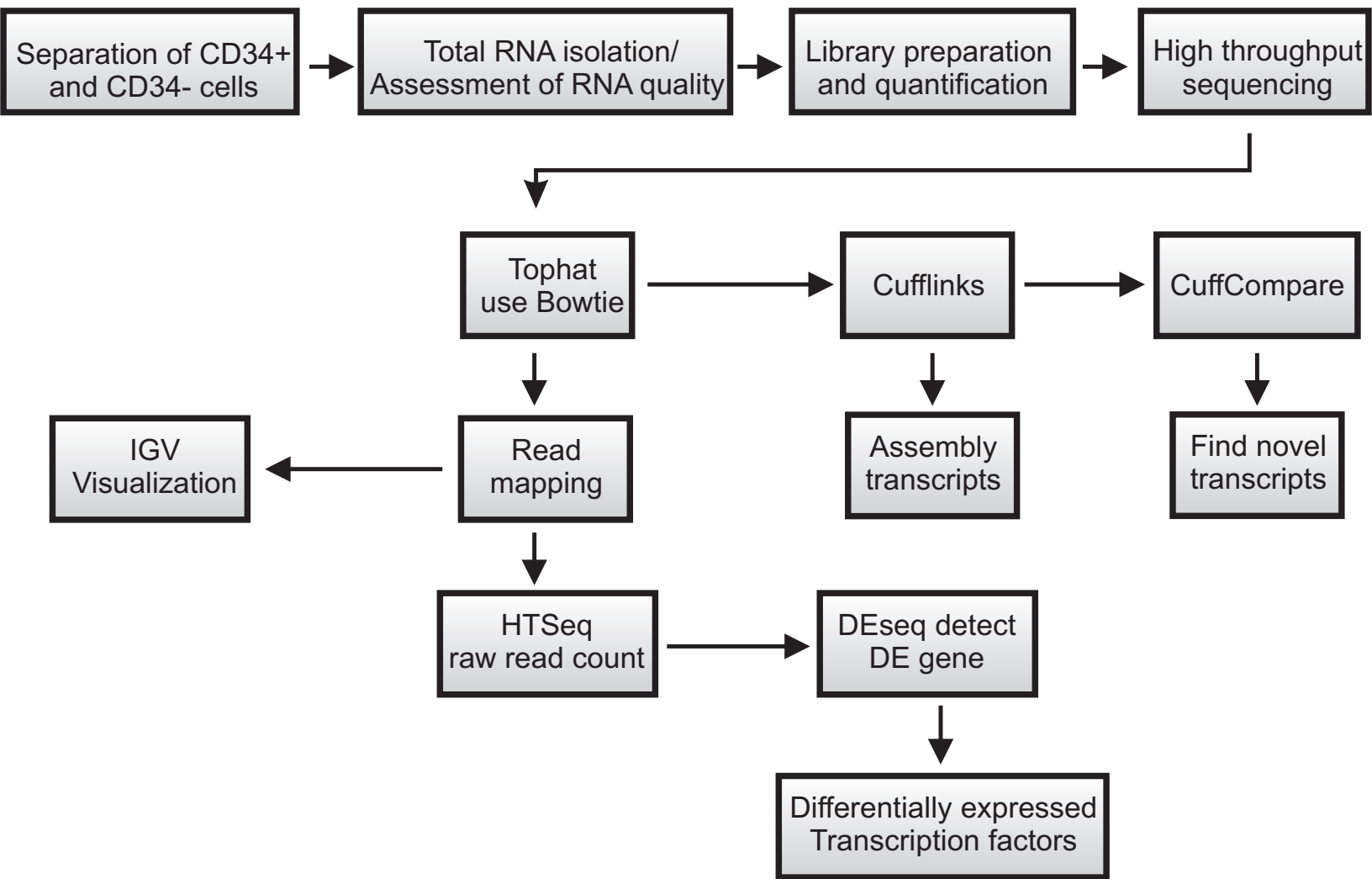




Figure 2  
[Click here to download Figure: Figure 2.pdf](#)

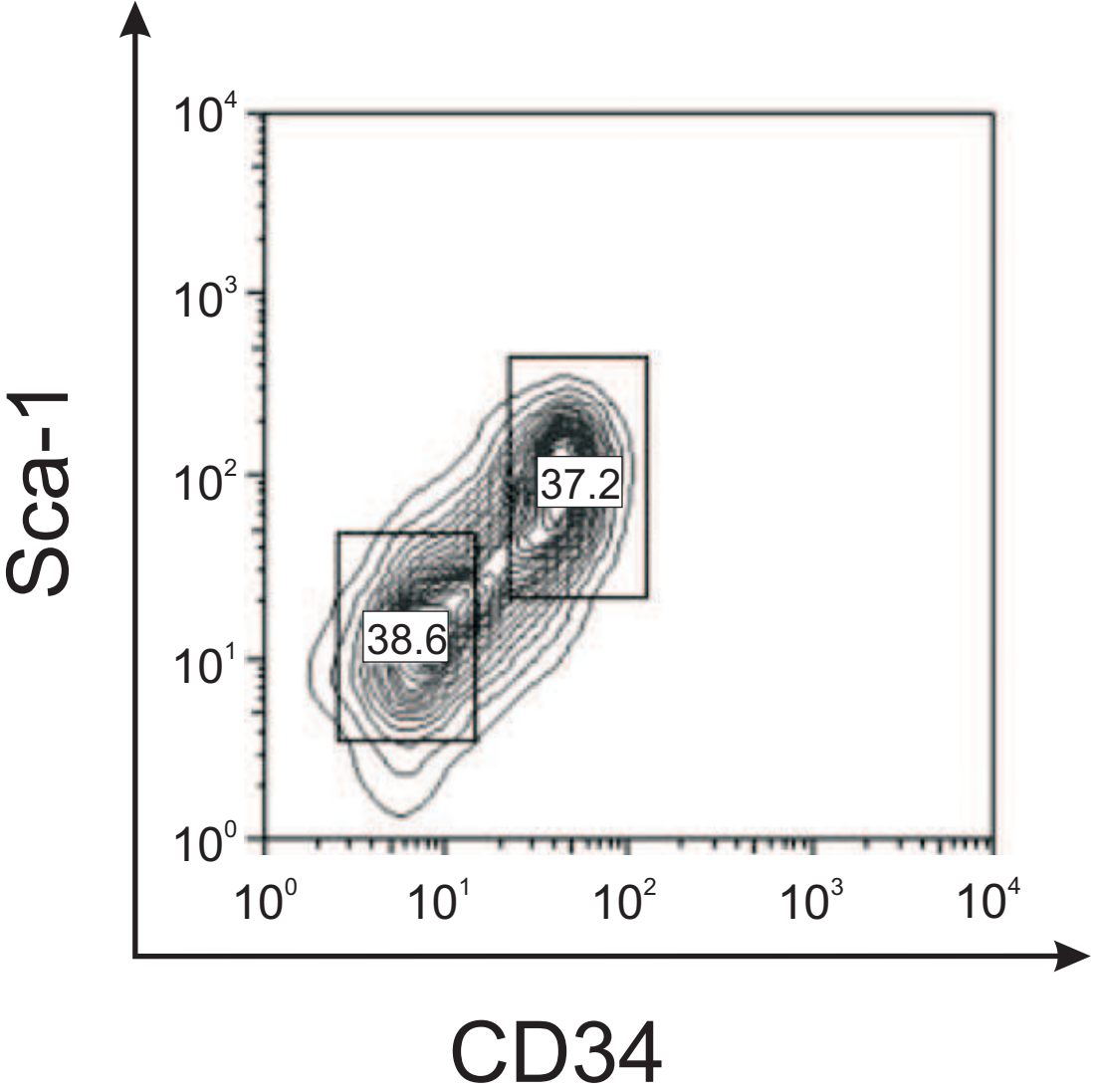


Figure 3  
[Click here to download Figure: Figure 3.pdf](#)

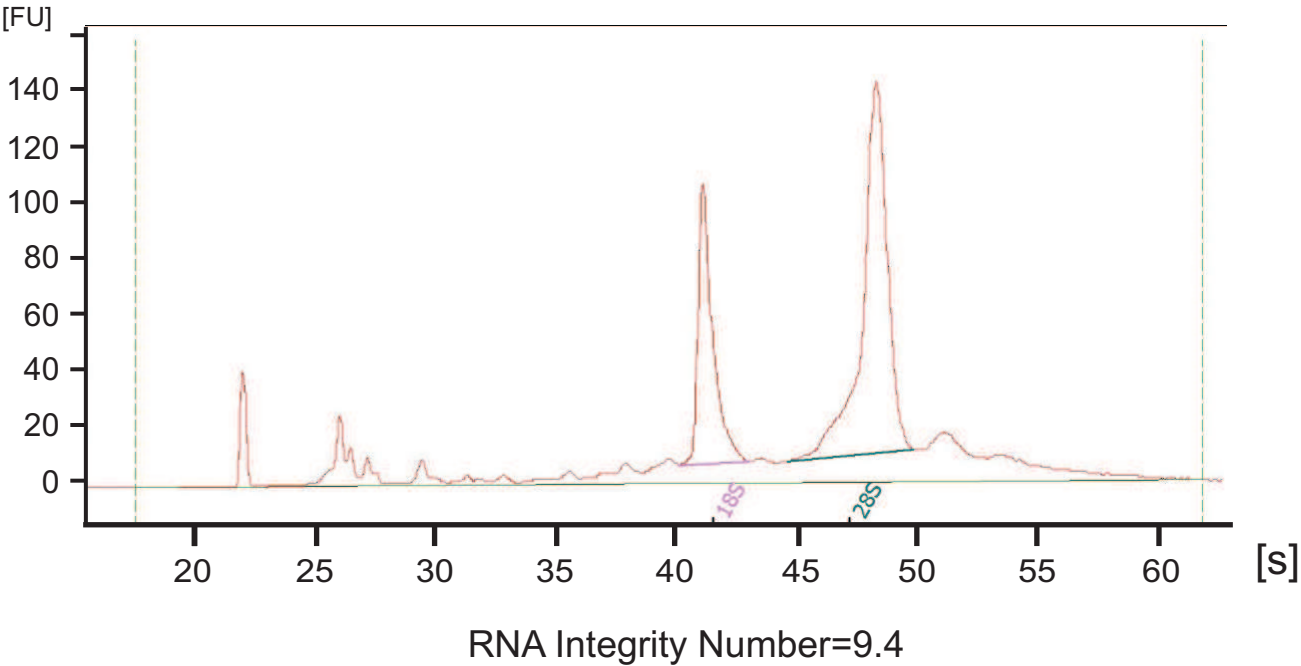


Figure 4  
[Click here to download Figure: Figure 4.pdf](#)

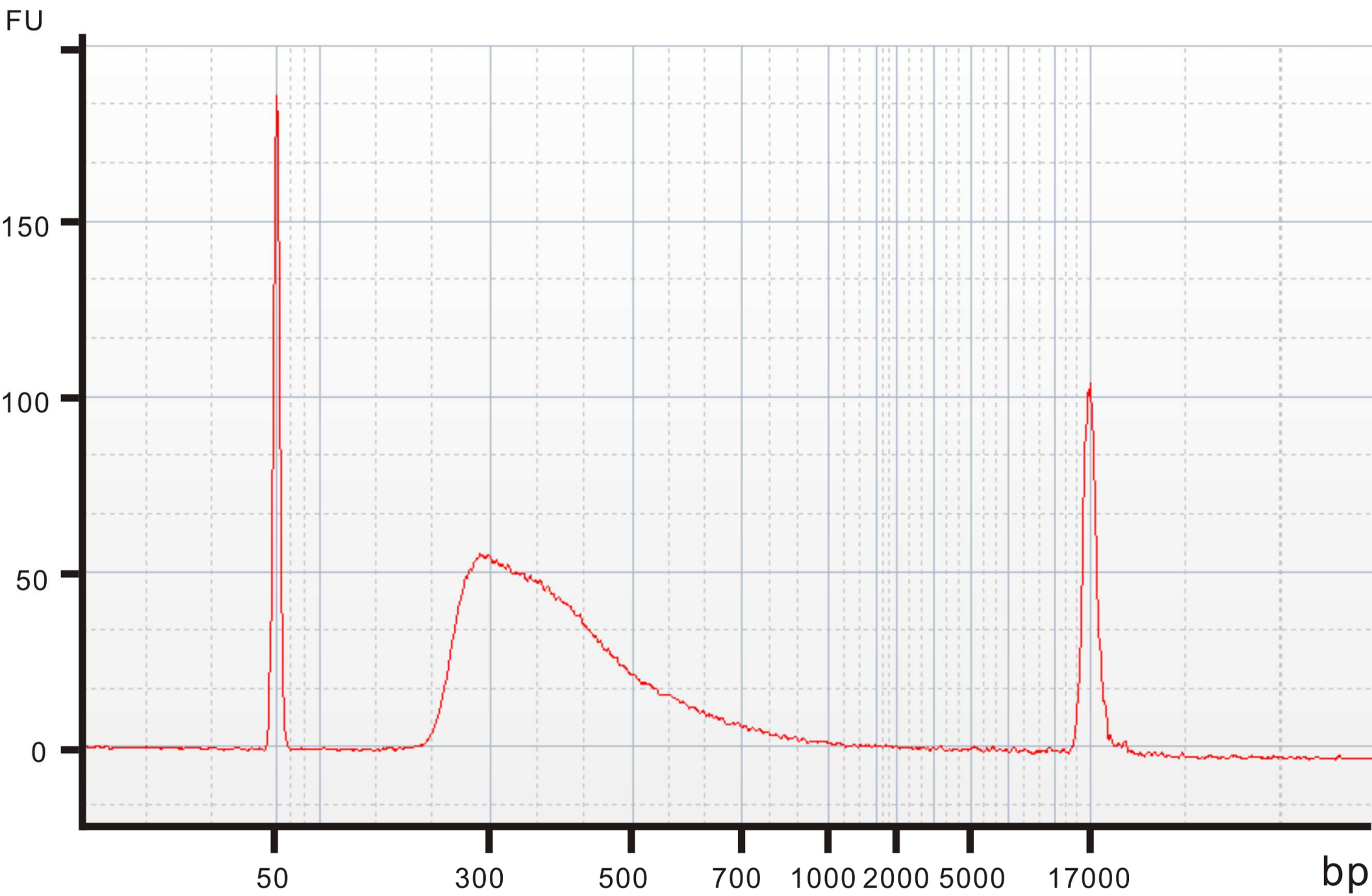


Figure 5  
[Click here to download Figure: Figure 5.pdf](#)

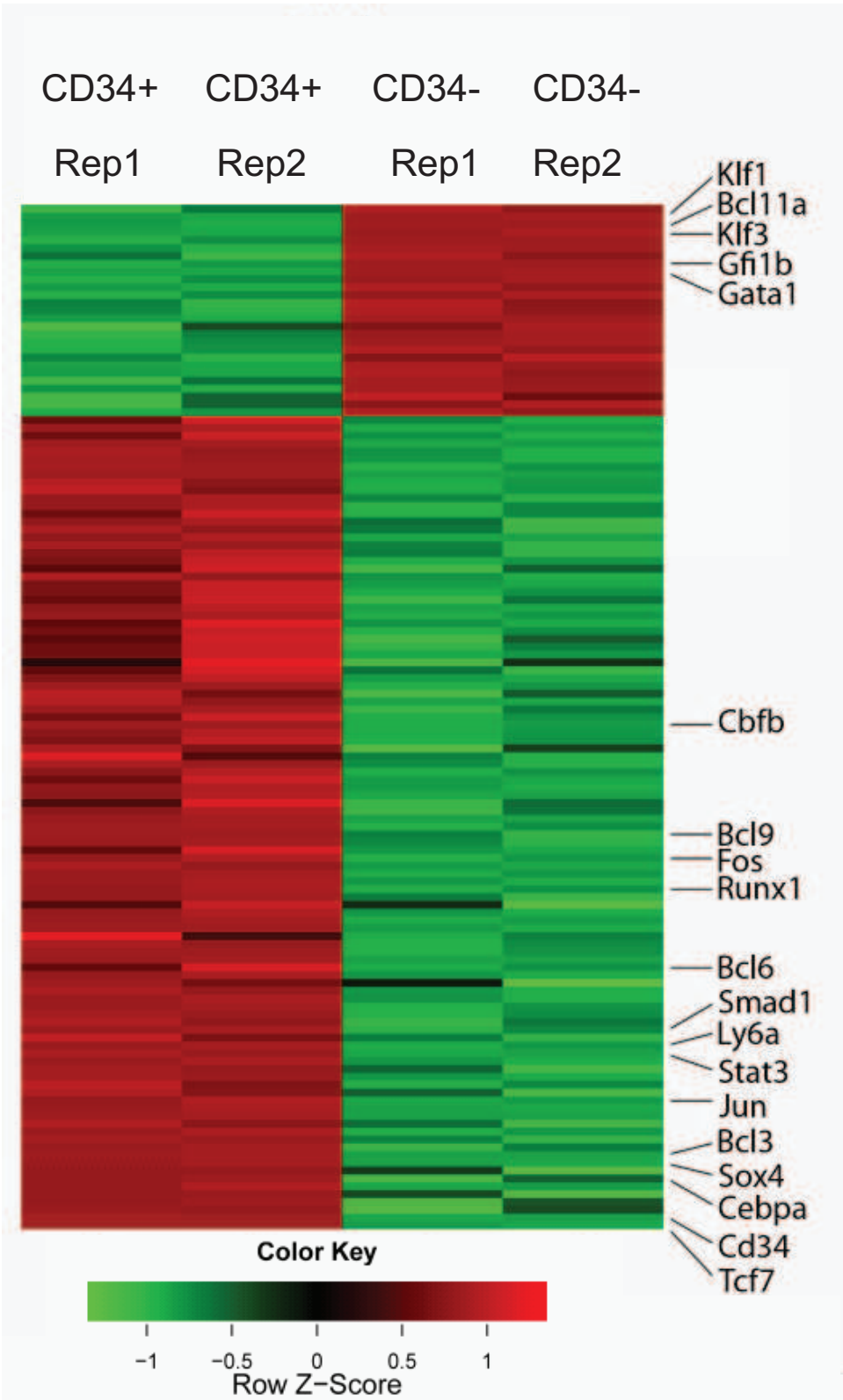


Table 1  
[Click here to download Table: Table 1.xlsx](#)

BHK medium	
100x Antibiotic-Antimycotic	10 ml
200 mM L-Glutamine	10 ml
FBS	100 ml
DMEM	880 ml
Total volume	1000 ml
EML Basic medium	
IMDM	390 ml
HI horse serum	100 ml
100 x Penicillin-Streptomycin	5 ml
200 mM L-Glutamine	5 ml
BHK medium	75 ml
Total Volume	575 ml
Filtrate through 0.45 µM filter	
FACS buffer	
BSA	0.50%
EDTA	1mM
Dissolved in PBS and filtrate through 0.45 µM filter	

Table 2  
[Click here to download Table: Table 2.xlsx](#)

Software	Usage
Bowtie 1.2.7	Used by Tophat for mapping
Tophat 1.3.3	Mapping reads back to reference genome
Cufflinks 1.3.0	Transcripts construction and expression level estimation
DESeq 1.16.0	Differential expression analysis
Bedtools 2.18	Convert .bam file into .bed file
bedGraphToBigWig	Convert .bed file to .bigwig file

Reference
[28]
[27]
[29]
[30]
[31]
<a href="http://genome.ucsc.edu/">http://genome.ucsc.edu/</a>

Name of Reagent/ Equipment	Company
Antibiotic-Antimycotic	Invitrogen
Anti-Mouse CD34 FITC	eBioscience
Anti-Mouse Ly-6A/E (Sca-1) PE	eBioscience
APC Mouse Lineage Antibody Cocktail	BD Biosciences
BD FACSAria Cell Sorter	BD Biosciences
Corning™ Cell Culture Treated Flasks 75cm <sup>2</sup>	Corning incorporated
Corning™ Cell Culture Treated Flasks 25cm <sup>2</sup>	Corning incorporated
Deoxyribonuclease I, Amplification Grade	Invitrogen
DMEM	Invitrogen
DPBS	Gibco
HI FBS	Invitrogen
Horse Serum	Invitrogen
IMDM	HyClone
L-Glutamine	Invitrogen
Lineage Cell Depletion Kit, mouse	Miltenyi Biotec
NanoVue Plus spectrophotometer	GE Healthcare
Thermo Scientific™ Napco™ 8000 Water-Jacketed CO2 Incubators	Thermo Scientific
Penicillin-Streptomycin	Invitrogen
TRIzol® Reagent	Invitrogen
TruSeq™ RNA Sample Prep Kit v2 -Set B (48rxn)	Illumina
2100 Electrophoresis Bioanalyzer Instrument	Agilent
0.25% Trypsin-EDTA	Gibco
0.45 µm Syringe Filters	Nalgene



<b>Catalog Number</b>	<b>Comments/Description</b>	<b>steps used in protocol</b>
15240-062	BHK cell culture	1.1
11-0341-81	FACS sorting	1.4
12-5981-81	FACS sorting	1.4
558074	FACS sorting	1.3
Special offer sysmtem	FACS sorting	1.4
430641	Cell culture	1.1/1.2
430639	Cell culture	1.1/1.2
18068-015	Library preparation	2.1
11965-092	BHK cell culture	1.1
14190	Cell culture	1.1/1.2
16140071	BHK cell culture	1.1
16050-122	EML cell culture	1.2
SH30228.02	EML cell culture	1.2
25030-081	Cell culture	1.1/1.2
130-090-858	Isolation of lineage negative cells	1.3
28-9569-62	Quality control	2.1
15-497-002	Cell culture	1.1/1.2
15140-122	EML cell culture	1.2
15596-018	RNA extraction	2.1
RS-122-2002	Library preparation	2.2
G2939AA	Quality control	2.1/2.2
25200	Cell culture	1.1/1.2
190-2545	Cell culture	1.1/1.2



1 Alewife Center #200  
Cambridge, MA 02140  
tel. 617.945.9051  
[www.jove.com](http://www.jove.com)

## ARTICLE AND VIDEO LICENSE AGREEMENT

Title of Article:

*Identification of key factors regulating self-renewal and differentiation in EML hematopoietic precursor cells by RNA-sequencing analysis*

Author(s):

*Shan Zong, Shuyan Deng, Kenan Chen, Jiaqian Wu*

Item 1 (check one box): The Author elects to have the Materials be made available (as described at <http://www.jove.com/publish>) via: ☒ Standard Access ☐ Open Access

Item 2 (check one box):

- ☒ The Author is NOT a United States government employee.
- ☐ The Author is a United States government employee and the Materials were prepared in the course of his or her duties as a United States government employee.
- ☐ The Author is a United States government employee but the Materials were NOT prepared in the course of his or her duties as a United States government employee.

### ARTICLE AND VIDEO LICENSE AGREEMENT

1. **Defined Terms.** As used in this Article and Video License Agreement, the following terms shall have the following meanings: "Agreement" means this Article and Video License Agreement; "Article" means the article specified on the last page of this Agreement, including any associated materials such as texts, figures, tables, artwork, abstracts, or summaries contained therein; "Author" means the author who is a signatory to this Agreement; "Collective Work" means a work, such as a periodical issue, anthology or encyclopedia, in which the Materials in their entirety in unmodified form, along with a number of other contributions, constituting separate and independent works in themselves, are assembled into a collective whole; "CRC License" means the Creative Commons Attribution-Non Commercial-No Derivs 3.0 Unported Agreement, the terms and conditions of which can be found at: <http://creativecommons.org/licenses/by-nc-nd/3.0/legalcode>; "Derivative Work" means a work based upon the Materials or upon the Materials and other pre-existing works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which the Materials may be recast, transformed, or adapted; "Institution" means the institution, listed on the last page of this Agreement, by which the Author was employed at the time of the creation of the Materials; "JoVE" means MyJoVE Corporation, a Massachusetts corporation and the publisher of *The Journal of Visualized Experiments*; "Materials" means the Article and / or the Video; "Parties" means the Author and JoVE; "Video" means any video(s) made by the Author, alone or in conjunction with any other parties, or by JoVE or its affiliates or agents, individually or in collaboration with the Author or any other parties, incorporating all or any portion of the Article, and in which the Author may or may not appear.

2. **Background.** The Author, who is the author of the Article, in order to ensure the dissemination and protection of the Article, desires to have the JoVE publish the Article and create and transmit videos based on the Article. In furtherance of such goals, the Parties desire to memorialize in this Agreement the respective rights of each Party in and to the Article and the Video.

3. **Grant of Rights in Article.** In consideration of JoVE agreeing to publish the Article, the Author hereby grants to JoVE, subject to Sections 4 and 7 below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Article in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Article into other languages, create adaptations, summaries or extracts of the Article or other Derivative Works (including, without limitation, the Video) or Collective Works based on all or any portion of the Article and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and (c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. If the "Open Access" box has been checked in Item 1 above, JoVE and the Author hereby grant to the public all such rights in the Article as provided in, but subject to all limitations and requirements set forth in, the CRC License.



## ARTICLE AND VIDEO LICENSE AGREEMENT

4. Retention of Rights in Article. Notwithstanding the exclusive license granted to JoVE in **Section 3** above, the Author shall, with respect to the Article, retain the non-exclusive right to use all or part of the Article for the non-commercial purpose of giving lectures, presentations or teaching classes, and to post a copy of the Article on the Institution's website or the Author's personal website, in each case provided that a link to the Article on the JoVE website is provided and notice of JoVE's copyright in the Article is included. All non-copyright intellectual property rights in and to the Article, such as patent rights, shall remain with the Author.

5. Grant of Rights in Video – Standard Access. This **Section 5** applies if the "Standard Access" box has been checked in **Item 1** above or if no box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby acknowledges and agrees that, Subject to **Section 7** below, JoVE is and shall be the sole and exclusive owner of all rights of any nature, including, without limitation, all copyrights, in and to the Video. To the extent that, by law, the Author is deemed, now or at any time in the future, to have any rights of any nature in or to the Video, the Author hereby disclaims all such rights and transfers all such rights to JoVE.

6. Grant of Rights in Video – Open Access. This **Section 6** applies only if the "Open Access" box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby grants to JoVE, subject to **Section 7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Video in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Video into other languages, create adaptations, summaries or extracts of the Video or other Derivative Works or Collective Works based on all or any portion of the Video and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and (c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. For any Video to which this Section 6 is applicable, JoVE and the Author hereby grant to the public all such rights in the Video as provided in, but subject to all limitations and requirements set forth in, the CRC License.

7. Government Employees. If the Author is a United States government employee and the Article was prepared in the course of his or her duties as a United States government employee, as indicated in **Item 2** above, and any of the licenses or grants granted by the Author hereunder exceed the scope of the 17 U.S.C. 403, then the rights granted hereunder shall be limited to the maximum rights permitted under such

statute. In such case, all provisions contained herein that are not in conflict with such statute shall remain in full force and effect, and all provisions contained herein that do so conflict shall be deemed to be amended so as to provide to JoVE the maximum rights permissible within such statute.

8. Likeness, Privacy, Personality. The Author hereby grants JoVE the right to use the Author's name, voice, likeness, picture, photograph, image, biography and performance in any way, commercial or otherwise, in connection with the Materials and the sale, promotion and distribution thereof. The Author hereby waives any and all rights he or she may have, relating to his or her appearance in the Video or otherwise relating to the Materials, under all applicable privacy, likeness, personality or similar laws.

9. Author Warranties. The Author represents and warrants that the Article is original, that it has not been published, that the copyright interest is owned by the Author (or, if more than one author is listed at the beginning of this Agreement, by such authors collectively) and has not been assigned, licensed, or otherwise transferred to any other party. The Author represents and warrants that the author(s) listed at the top of this Agreement are the only authors of the Materials. If more than one author is listed at the top of this Agreement and if any such author has not entered into a separate Article and Video License Agreement with JoVE relating to the Materials, the Author represents and warrants that the Author has been authorized by each of the other such authors to execute this Agreement on his or her behalf and to bind him or her with respect to the terms of this Agreement as if each of them had been a party hereto as an Author. The Author warrants that the use, reproduction, distribution, public or private performance or display, and/or modification of all or any portion of the Materials does not and will not violate, infringe and/or misappropriate the patent, trademark, intellectual property or other rights of any third party. The Author represents and warrants that it has and will continue to comply with all government, institutional and other regulations, including, without limitation all institutional, laboratory, hospital, ethical, human and animal treatment, privacy, and all other rules, regulations, laws, procedures or guidelines, applicable to the Materials, and that all research involving human and animal subjects has been approved by the Author's relevant institutional review board.

10. JoVE Discretion. If the Author requests the assistance of JoVE in producing the Video in the Author's facility, the Author shall ensure that the presence of JoVE employees, agents or independent contractors is in accordance with the relevant regulations of the Author's institution. If more than one author is listed at the beginning of this Agreement, JoVE may, in its sole discretion, elect not take any action with respect to the Article until such time as it has received complete, executed Article and Video License Agreements from each such author. JoVE reserves the right, in its absolute and sole discretion and without giving any reason therefore, to accept or decline any work submitted to JoVE. JoVE and its employees, agents and independent contractors shall have



## ARTICLE AND VIDEO LICENSE AGREEMENT

full, unfettered access to the facilities of the Author or of the Author's institution as necessary to make the Video, whether actually published or not. JoVE has sole discretion as to the method of making and publishing the Materials, including, without limitation, to all decisions regarding editing, lighting, filming, timing of publication, if any, length, quality, content and the like.

11. **Indemnification.** The Author agrees to indemnify JoVE and/or its successors and assigns from and against any and all claims, costs, and expenses, including attorney's fees, arising out of any breach of any warranty or other representations contained herein. The Author further agrees to indemnify and hold harmless JoVE from and against any and all claims, costs, and expenses, including attorney's fees, resulting from the breach by the Author of any representation or warranty contained herein or from allegations or instances of violation of intellectual property rights, damage to the Author's or the Author's institution's facilities, fraud, libel, defamation, research, equipment, experiments, property damage, personal injury, violations of institutional, laboratory, hospital, ethical, human and animal treatment, privacy or other rules, regulations, laws, procedures or guidelines, liabilities and other losses or damages related in any way to the submission of work to JoVE, making of videos by JoVE, or publication in JoVE or elsewhere by JoVE. The Author shall be responsible for, and shall hold JoVE harmless from, damages caused by lack of sterilization, lack of cleanliness or by contamination due to the making of a video by JoVE its employees, agents or independent contractors. All sterilization, cleanliness or decontamination procedures shall be solely the responsibility of the Author and shall be undertaken at the Author's

expense. All indemnifications provided herein shall include JoVE's attorney's fees and costs related to said losses or damages. Such indemnification and holding harmless shall include such losses or damages incurred by, or in connection with, acts or omissions of JoVE, its employees, agents or independent contractors.

12. **Fees.** To cover the cost incurred for publication, JoVE must receive payment before production and publication the Materials. Payment is due in 21 days of invoice. Should the Materials not be published due to an editorial or production decision, these funds will be returned to the Author. Withdrawal by the Author of any submitted Materials after final peer review approval will result in a US\$1,200 fee to cover pre-production expenses incurred by JoVE. If payment is not received by the completion of filming, production and publication of the Materials will be suspended until payment is received.

13. **Transfer, Governing Law.** This Agreement may be assigned by JoVE and shall inure to the benefits of any of JoVE's successors and assignees. This Agreement shall be governed and construed by the internal laws of the Commonwealth of Massachusetts without giving effect to any conflict of law provision thereunder. This Agreement may be executed in counterparts, each of which shall be deemed an original, but all of which together shall be deemed to me one and the same agreement. A signed copy of this Agreement delivered by facsimile, e-mail or other means of electronic transmission shall be deemed to have the same legal effect as delivery of an original signed copy of this Agreement.

A signed copy of this document must be sent with all new submissions. Only one Agreement required per submission.

### CORRESPONDING AUTHOR:

Name:

Jiaqian Wu

Department:

The Vivian L. Smith Department of Neurosurgery

Institution:

University of Texas Health Science Center

Article Title:

Identification of key factors regulating self-renewal and differentiation in EML hematopoietic precursor cells by RNA-sequencing analysis

Signature:

Jiaqian Wu

Date:

03/10/14

Please submit a signed and dated copy of this license by one of the following three methods:

- 1) Upload a scanned copy of the document as a pdf on the JoVE submission site;
- 2) Fax the document to +1.866.381.2236;
- 3) Mail the document to JoVE / Attn: JoVE Editorial / 1 Alewife Center #200 / Cambridge, MA 02139

For questions, please email [submissions@jove.com](mailto:submissions@jove.com) or call +1.617.945.9051

Dear editor and reviewers,

Thanks for all the comments and suggestions. The modifications that we made in the manuscript were changed to red color and were listed below:

#### **Editor's suggestions**

**1) The protocol is >3 pages and is not highlighted. The authors will need to highlight 2.75 pages of material to be filmed.**

Following the requirement of the journal, we now have highlighted 2.75 pages of material for filming in the revised manuscript.

**2) Figures 1-5 are not Results; these should be changed to Supplemental Figures, and Figs. 6-10 should be renumbered appropriately.**

As suggested, we moved figures 1-5 to supplemental figures and renumbered figures 6-10 accordingly.

**3) Please take this opportunity to thoroughly proofread your manuscript to ensure that there are no spelling or grammar issues. Your JoVE editor will not copy-edit your manuscript and any errors in your submitted revision may be present in the published version.**

Thanks for the suggestion, we have checked and corrected the spelling and grammar errors.

#### **Reviewer #1's suggestions**

##### **Major Concerns:**

**I did not find major flaws in the paper. It should be mentioned that DESeq is not the only application available for differential gene expression analyses. Cuffdiff could be explained. Alternative apps may be explained. The advantages of DESeq could be detailed some more here.**

We appreciate reviewer's comments on the manuscript. Following the reviewer's suggestion, we have explained different methods for differential gene expression analyses including DESeq and Cuffdiff in the result part of the revised manuscripts. In addition we also addressed the advantages of DESeq. It is true that DESeq is not the only application available for differential gene calling. We chose DESeq because it is a widely used count based method for DE gene calling from RNA-Seq data. In our experience, cuffdiff gives significantly different numbers of DE genes among different versions of software packages. The count based method DESeq is more stable in this sense.

##### **Minor Concerns:**

**1 The RNA manipulation steps (2.2.2) could be provided in more detail. The make of the machines used for the analyses should be provided.**

Following the reviewer's advice, the step 2.2.2 is provided in more detail. The machines used for analyses have already been listed out in the material part. It is not allowed to write any commercial names (e.g. NanoVue Plus spectrophotometer) of instruments in the text.

**2 The Lineage antibodies used for negative selection should be listed.**

Lineage antibodies used for negative selection are now listed in the step 1.3.3. The antibodies were contained in the Lineage Cell Depletion Kit and the kit has been mentioned in the material part.

**3 Fig 9 legend, delete FU, it is not applicable to base pair. Editor's note: Please write the legends in a new sentence.**

Thanks for the suggestion; FU has been deleted and a new legend has been written.

#### **Reviewer #2's suggestions**

##### **Major Concerns:**

**1) As is, the portion of the manuscript discussing culture of EML cells is not novel nor particularly useful beyond the study described. A brief discussion of the potential use of EML cells as a model for differentiation or gene function (e.g. shRNA knockdown) coupled with RNA-seq would broaden the utility of this protocol.**

Following the reviewer's suggestion, we added a discussion of using EML cells as a model for differentiation and gene function study coupled with RNA-sequencing technology in the introduction and the discussion.

**2) As the most useful part of the manuscript is the description of the RNAseq process, there should be more discussion about enrichment of mRNA from total RNA (Ribo minus vs. PolyA enrichment), how to decide what kinds of reads to generate (paired end vs. single end; 50nt vs. 75nt), and different sequencing platforms. In fact, the sequencing platform is not even discussed.**

Thanks for the reviewer's comments. We described and compared different sequencing platforms in the introduction part. In addition, we added discussions about different methods for enrichment of mRNA and the criteria of choosing different kinds of reads and read lengths in the result session.

**The protocol jumps from the cDNA library to uploading data files. It seems as this protocol is intended for Illumina based sequencing. But the readers should be made aware that other sequencers exist (and may be the only choice for the reader) and that they will require different adaptor kits and possibly different means of data analysis.**

We added the description and comparison of different sequencing platforms in the introduction. In addition we explained in the protocol session that the library preparation part described in the paper is for Illumina sequencing platform.

Minor Concerns:

**1) In the introduction, CD34+ cells are referred to as self-renewing population. However, in mice, the long-term reconstituting HSCs are Kit+, Sca+, CD150+, and CD34-. Does this population exist in EML cells? Moreover, it can be argued that Sca+, Kit+, CD34+ cells are multipotent progenitors with limited self-renewing potential.**

The reviewer is correct. The closest normal proximity of Lin-CD34+ cells are short-term (ST)-HSC or multipotent progenitors (MPP).

**The mechanisms of self-renewal in MPPs may be different than those of true, quiescent HSCs.**

We agree with the reviewer that the mechanisms of self-renewal in MPPs might be different from that of quiescent HSCs. To address this question, we are in the process of performing experiments to test the functions of candidate key transcription factors in human HSCs.

**The expression of c-Kit should also be briefly mentioned.**

Following reviewer's suggestion, we added the expression of c-Kit in EML cells in the introduction part of the manuscript.

**2) The citation for the derivation of the EML cell line is not the original paper describing the EML. Please reference Tsai, S., et al (1994) Genes & Development.**

A citation for the original paper has been added according to the reviewer's suggestion

**3) For culture of EML cells, is there variation between lots of horse serum used to culture the cells in terms of proliferation and differentiation?**

According to our experience, different lots of horse serum had no significant influence on EML cell proliferation and differentiation

**In the discussion, it is mentioned that the cells need to be treated with "gentle operation." This should be stated in the procedure.**

Following the reviewer's suggestion, we added "gentle operation" into the step 1.2.2.

**The warning that the ratio of CD34+ to CD34- is critical is too vague-- at what point should the user discard the cultures (e.g. >60% CD34-).**

The percentage at which EML cell should be discarded is still debatable. We normally discard the culture after passaging for 10 generations.

**Also, how many cells are required to get RNA in the 0.1-4ug range? Is this before or after mRNA enrichment?**

We usually use  $10^5$  cells for extracting total RNA and get about 1-2 ug RNA. The 0.1-4 ug RNA refers to the amount of total RNA before mRNA enrichment. We have already mentioned this information in the step 2.2.1.

**4) A table of the software needed would be useful for someone who has no prior experience with analyzing this kind of data.**

As suggested, we added a table (Table 2) of the software used in this manuscript.

**5) Would be useful if described how to generate a file to use as a custom track on the UCSC Genome Browser.**

Thanks for the reviewer's suggestion. We have added a description about how to generate .bigwig file in the protocol part. There is excellent information on how to load custom track on UCSC Genome Browser website. We added a reference link to the website, because the detailed information on this would be too long for the manuscript.

**6) Will the commands in the figures be legible? The font is very small in the proof.**

We have changed the commands to larger font size.

**Additional Comments to Authors:**

**There are many grammatical errors in the text. Please proofread carefully.**

We are sorry for the grammatical errors. We proofread the manuscript and made corrections.

## Supplemental Figure 1

[Click here to download Supplemental File \(as requested by JoVE\): Supplemental Figure 1.pdf](#)



## Supplemental Figure 2

[Click here to download Supplemental File \(as requested by JoVE\): Supplemental Figure 2.pdf](#)

## Supplemental Figure 3

[Click here to download Supplemental File \(as requested by JoVE\): Supplemental Figure 3.pdf](#)

## Supplemental Figure 4

[Click here to download Supplemental File \(as requested by JoVE\): Supplemental Figure 4.pdf](#)

## Supplemental Figure 5

[Click here to download Supplemental File \(as requested by JoVE\): Supplemental Figure 5.pdf](#)

## Supplemental Figure 6

[Click here to download Supplemental File \(as requested by JoVE\): Supplemental Figure 6.pdf](#)

