FINAL SCRIPT: APPROVED FOR FILMING

Submission ID #: 69400
Scriptwriter Name: Pallavi Sharm
Project Page Link: https://review.jove.com/account/file-uploader?src=21167048

Title: A Computational Pipeline for Intergenic/Intragenic Enhancer RNA Quantification in Mouse Embryonic Stem Cells

Authors and Affiliations:
Myunggeun Oh[*], Seunghwa Jeong[*], Seung-Kyoon Kim

Department of Convergent Bioscience and Informatics and Graduate School of Biological Sciences, Chungnam National University (CNU), Daejeon 34134, Korea

[*]These authors contributed equally

Corresponding Authors:
Seung-Kyoon Kim sk.kim@cnu.ac.kr

Email Addresses for All Authors:
Myunggeun Oh omg3327@naver.com
Seunghwa Jeong tmdghk04031@naver.com

Author Questionnaire

1. **Microscopy**: Does your protocol require the use of a dissecting or stereomicroscope for performing a complex dissection, microinjection technique, or something similar? **No**

2. **Software:** Does the part of your protocol being filmed include step-by-step descriptions of software usage? **Yes, all done**

3. **Filming location:** Will the filming need to take place in multiple locations? **No**

4. **Testimonials (optional):** Would you be open to filming two short testimonial statements **live during your JoVE shoot**? These will **not appear in your JoVE video** but may be used in JoVE's promotional materials. **No**

Current Protocol Length
Number of Steps: 16
Number of Shots: 20

Introduction

Videographer: Obtain headshots for all authors available at the filming location.

INTRODUCTION:

1.1. **Seung-Kyoon Kim:** Our group investigates enhancer epigenetics and epitranscriptomics, focusing on developing accurate methods to quantify enhancer RNAs as indicators of enhancer activation.

1.1.1. INTERVIEW: Named talent says the statement above in an interview-style shot, looking slightly off-camera. *Suggested B roll: Figure 1*

~~What are the current experimental challenges?~~

1.2. **Myunggeun Oh:** Because intragenic enhancers overlap with their host gene transcripts, distinguishing enhancer-derived signals and quantifying them precisely is challenging.

1.2.1. INTERVIEW: Named talent says the statement above in an interview-style shot, looking slightly off-camera.

CONCLUSION:

~~What significant findings have you established in your field?~~

1.3. **Myunggeun Oh:** We assembled a strand-aware eRNA quantification pipeline, validated by activity-dependent transcription in intergenic or intragenic contexts.

1.3.1. INTERVIEW: Named talent says the statement above in an interview-style shot, looking slightly off-camera. *Suggested B roll: 2.12.1.*

~~What research gap are you addressing with your protocol?~~

1.4. **Myunggeun Oh:** We address the lack of standardized, accessible eRNA quantification, especially for intragenic enhancers confounded by host-gene overlap

1.4.1. INTERVIEW: Named talent says the statement above in an interview-style shot, looking slightly off-camera.

~~What questions will future research focus on?~~

1.5. **Myunggeun Oh:** We will refine sense-derived quantification, separate overlapping transcripts from both strands, and benchmark analysis pipelines across genome assemblies, experimental protocols, and cell types.

Videographer: Obtain headshots for all authors available at the filming location.

Protocol

2. Enhancer Transcript Quantification Pipeline

Demonstrators: Seunghwa Jeong, Myunggeun Oh

2.1. To begin, open the terminal **[1]** and type the command to prepare the necessary files and references for enhancer identification **[2]**.

2.1.1. WIDE: Talent in front of the computer opening the terminal.

2.1.2. SCREEN: 69400_screenshot_1.mp4**:** 00:00-00:23 Video editor: Highlight the command *bash Step12_E_identification_material.sh > Step12_log.txt 2>&1 at 00:20*

2.2. Type the command in the terminal to enter the working directory **[1]**. Then, run the command that copies the enhancer-identification scripts into the current directory **[2-TXT]**.

2.2.1. SCREEN: 69400_screenshot_2.mp4**:** 00:00-00:06

2.2.2. SCREEN: 69400_screenshot_2.mp4: 00:06-00:22 **TXT: cp ../scripts/Step{13..20}_*.sh ./**

2.3. Then run the corresponding command to generate BED *(B-E-D)* files for promoter regions, gene bodies, and protein-coding genes using GENCODE *(Gen-Code)* annotation **[1]**.

2.3.1. SCREEN: 69400_screenshot_3.mp4: 00:00-00:24

2.4. To prepare the ATAC-seq (*A-T-A-C-Seq*) and histone ChIP-seq *(Chip-Seq)* peak files for enhancer identification, enter the appropriate preprocessing command in the terminal **[1]**.

2.4.1. SCREEN: 69400_screenshot_4.mp4: 00:00-00:18 *Video editor: Highlight bash Step14_ATAC_ChIP-seq_processing.sh > Step14_log.txt 2>&1 at 00:16*

2.5. Next, run the command to define and classify enhancers using chromatin peak data **[1]**.

2.5.1. SCREEN: 69400_screenshot_5.mp4: 00:10-00:17 *Video editor: Highlight "bash*

Step15_inter_intragenic_E_sets_identification.sh > Step15_log.txt 2>&1" at 00:16

2.6. Type the command to assign temporary strand information to the intragenic-enhancer BED file **[1]**.

2.6.1. SCREEN: 69400_screenshot_6.mp4: 00:03-00:16 *Video editor: Highlight bash Step16_assign_temp_strand_from_gene_overlap.sh > Step16_log.txt 2>&1 at 00:16*

2.7. Next, enter the command to determine the correct strand direction for intragenic enhancers that overlap with genes located on both strands **[1]**.

2.7.1. SCREEN: 69400_screenshot_7.mp4: 00:00-00:18 *Video editor: Highlight the bash Step17_initial_strand_assignment_for_both_strand_enhancers_PCG_based.sh > Step17_log.txt 2>&1 at 00:16*

2.8. Afterward, calculate strand-specific RPKM *(R-P-K-M)* for genes overlapping enhancers on the same strand **[1]**.

2.8.1. SCREEN: 69400_screenshot_8.mp4: 00:00-00:18

2.9. Then, enter the command to finalize strand assignment for intragenic enhancers **[1]**.

2.9.1. SCREEN: 69400_screenshot_9.mp4: 00:10-00:14 *Video editor: Highlight bash Step19_second_strand_assignment_by_RPKM.sh > Step19_log.txt 2>&1 at 00:14*

2.10. Assign the final strand information to all intragenic enhancers and their corresponding summit regions **[1]**.

2.10.1. SCREEN: 69400_screenshot_10.mp4: 00:05-00:17 *Video editor: Highlight bash Step20_strand_assignment_for_intragenicE_and_summits.sh > Step20_log.txt 2>&1 at 00:16*

2.11. Type the command in the terminal to move to the pipeline root directory **[1].** Then, use the appropriate command to copy the script for preparing downstream analysis **[2].**

2.11.1. SCREEN: 69400_screenshot_11.mp4: 00:02-00:07

2.11.2. SCREEN: 69400_screenshot_11.mp4: 00:15-00:20 *Video editor: Highlight cp scripts/Step21_preparing_quantification_and_visualization.sh ./ at 00:16*

2.12. Using the appropriate command, prepare all necessary files for enhancer aggregation, GRO-seq signal processing, and enhancer RNA quantification **[1]**.

2.12.1. SCREEN: 69400_screenshot_12.mp4: 00:05-00:20 *Video editor: Highlight bash Step21_preparing_quantification_and_visualization.sh > Step21_log.txt 2>&1 at 00:20*

2.13. Next, type the script to enter the working directory **[1]** and then, copy the necessary scripts for downstream analysis **[2-TXT]**.

2.13.1. SCREEN: 69400_screenshot_13.mp4: 00:00-00:08

2.13.2. SCREEN: 69400_screenshot_13.mp4: 00:09-00:26

TXT: cp ../scripts/Step{22..24}_*.* ./

2.14. To create aggregation plots showing chromatin signal patterns around each type of enhancer summit, type the corresponding command in the terminal **[1]**.

2.14.1. SCREEN: 69400_screenshot_14.mp4: 00:13-00:25 *Video editor: Highlight bash Step22_generation_of_aggregation_plot.sh > Step22_log.txt 2>&1 at 00:23*

2.15. Once the prompt returns, enter the command to quantify enhancer RNA expression levels from GRO-seq using featureCounts **[1]**.

2.15.1. SCREEN: 69400_screenshot_15.mp4: 00:10-00:27 *Video editor: Highlight bash Step23_quantifing_eRNA_RPKM.sh > Step23_log.txt 2>&1 in the terminal at 00:25*

2.16. Finally, run the script to visualize and compare enhancer RNA expression levels across enhancer groups using R **[1]**.

2.16.1. SCREEN: 69400_screenshot_16.mp4: 00:00-00:22 *Video editor: Highlight Rscript Step24_visualization_of_enhancer_transcript.R > Step24_log.txt 2>&1* at 00:20

Results

3. Results

3.1. In the GRO-seq dataset, poly-nucleotide tails were present before trimming and successfully removed after applying the trimming parameters **[1]**. In paired-end ATAC-seq datasets, Nextera adapter sequences were detected in both read pairs before trimming and were removed after preprocessing **[2].**

3.1.1. LAB MEDIA: Figure 2A and 3A. *Video editor: Highlight the "After" plot.*

3.1.2. LAB MEDIA: Figure 2B and 3B**.** *Video editor: Highlight the even quality scores in both reads after trimming in the "After" plots*.

3.2. ChIP-seq datasets for H3K27ac *(H-Three-K-twenty-Seven-Acetylation)* showed minimal adapter contamination, requiring only default trimming **[1].**

3.2.1. LAB MEDIA: Figure 2C and 3C**.** *Video editor: Show both "Before" and "After" quality score plots for Read1 and Read2.*

3.3. Filtering of ATAC-seq data yielded 71,165 non-promoter open chromatin regions, which included 47,317 enhancer regions **[1],** of which 23,147 were classified as active enhancers and 24,170 as non-active enhancers based on overlap with histone marks **[2].**

3.3.1. LAB MEDIA: Figure 4A. *Video editor: Highlight the green outer circle labeled "All Enhancer" within the large circle of "Non-promoter Open Chromatin Regions".*

3.3.2. LAB MEDIA: Figure 4A. *Video editor: Highlight the yellow circle labeled "Non-active Enhancer" and red circle labeled "Active Enhancer" nested inside the green "All Enhancer" circle.*

3.4. Among all enhancer regions, 45.2% were intergenic and 54.8% were intragenic **[1].** This distribution was similar across non-active **[2]** and active enhancer categories **[3].**

3.4.1. LAB MEDIA: Figure 4B (top). *Video editor: Highlight the pie chart showing the 45.2% intergenic and 54.8% intragenic split for "All Enhancers".*

3.4.2. LAB MEDIA: Figure 4B (bottom left). *Video editor: Highlight the pie chart showing the 45.5% intergenic and 54.5% intragenic split for "Non-active Enhancers".*

3.4.3. LAB MEDIA: Figure 4B (bottom right). *Video editor: Highlight the pie chart showing the 45.0% intergenic and 55.0% intragenic split for "Active Enhancers".*

3.5. An intergenic enhancer upstream of the Nanog gene showed strong enrichment for ATAC-seq, H3K4me1 *(H-Three-K-Four-Mono-methylation)*, H3K27ac *(H-Three-K-twenty-Seven-Acetylation)*, and GRO-seq signals **[1].**

3.5.1. LAB MEDIA: Figure 5A. *Video editor: Highlight the boxed region labeled "Active Enhancer" and show the aligned peaks in the ATAC-seq, H3K27ac, H3K4me1, and GRO-seq tracks.*

3.6. An intragenic enhancer near the Chd2 *(C-H-D-Two)* gene displayed similar chromatin and transcriptional features as the Nanog enhancer, with clear signals in all four datasets **[1].**

3.6.1. LAB MEDIA: Figure 5B**.** *Video editor: Highlight the "Active Enhancer" region within the Chd2 gene and show aligned peaks in the ATAC-seq, H3K27ac, H3K4me1, and GRO-seq tracks*.

3.7. Aggregation plots showed that active enhancers had coordinated enrichment of ATAC-seq, H3K4me1, and H3K27ac signals **[1],** while non-active enhancers lacked H3K27ac enrichment but retained strong ATAC-seq and H3K4me1 signals **[2].**

3.7.1. LAB MEDIA: Figure 5C (right column). *Video editor: Highlight the peaks in all three tracks (ATAC, H3K4me1, H3K27ac) in the "Active" panel for both intergenic and intragenic enhancers.*

3.7.2. LAB MEDIA: Figure 5C (center column). *Video editor: Highlight the missing H3K27ac signal in the "Non-active" panels while showing retention of ATAC and H3K4me1 peaks.*

3.8. Active enhancers showed significantly higher GRO-seq–derived transcription than non-active enhancers in both intergenic **[1]** and intragenic regions **[2].**

3.8.1. LAB MEDIA: Figure 6A (left). *Video editor: Highlight the higher GRO-seq signal distribution for "active E" compared to "non-active E" in the intergenic enhancer violin plot.*

3.8.2. LAB MEDIA: Figure 6A (right). *Video editor: Highlight the higher GRO-seq signal distribution for "active E" compared to "non-active E" in the intragenic enhancer violin plot.*

1. **BED (file format)**
   Pronunciation link: https://en.wikipedia.org/wiki/BED_%28file_format%29 (Wikipedia)
   IPA: /bɛd/
   Phonetic Spelling: BED
2. **ATAC-seq** (Assay for Transposase-Accessible Chromatin using sequencing)
   Pronunciation link: https://nbis-workshop-epigenomics.readthedocs.io/en/latest/content/tutorials/ATACseq/lab-atacseq-bulk.html (nbis-workshop-epigenomics.readthedocs.io)
   IPA: /ˌeɪtiːˌeɪˈsiː siːˈkwɛns/
   Phonetic Spelling: A-T-A-C-seq
3. **ChIP-seq** (Chromatin ImmunoPrecipitation sequencing)
   Pronunciation link: (general abbreviation) No single standard dictionary page found
   IPA: /ˌtʃɪp siːˈkwɛns/
   Phonetic Spelling: CHIP-seq
4. **RPKM** (Reads Per Kilobase of transcript, per Million mapped reads)
   Pronunciation link: (abbreviation; no standard dictionary entry)
   IPA: /ˌɑːr piː keɪ ˈɛm/
   Phonetic Spelling: R-P-K-M
5. **strand-specific**
   Pronunciation link: (no single dictionary page)
   IPA: /strænd-spəˈsɪfɪk/
   Phonetic Spelling: STRAND-spe-SIF-ik
6. **enhancer**
   Pronunciation link: https://www.merriam-webster.com/dictionary/enhancer
   IPA: /ɛnˈhænsər/
   Phonetic Spelling: en-HAN-ser
7. **intragenic**
   Pronunciation link: (no standard dictionary entry)
   IPA: /ˌɪntrəˈdʒɛnɪk/
   Phonetic Spelling: in-truh-GEN-ik
8. **intergenic**
   Pronunciation link: (no standard dictionary entry)
   IPA: /ˌɪntərˈdʒɛnɪk/
   Phonetic Spelling: in-ter-GEN-ik
9. **aggregation (plot)**
   Pronunciation link: https://www.merriam-webster.com/dictionary/aggregation
   IPA: /ˌægrəˈgeɪʃən/
   Phonetic Spelling: ag-ruh-GAY-shun
10. **quantification**
    Pronunciation link: https://www.merriam-webster.com/dictionary/quantification
    IPA: /ˌkwɑːntɪfɪˈkeɪʃən/
    Phonetic Spelling: kwan-ti-fi-KAY-shun

11. **transcriptome**
    Pronunciation link: https://www.merriam-webster.com/dictionary/transcriptome
    IPA: /trænsˈkrɪptoʊm/
    Phonetic Spelling: trans-KRIP-tohm
12. **summit region**
    – **summit**: https://www.merriam-webster.com/dictionary/summit
    IPA: /ˈsʌmɪt/
    Phonetic Spelling: SUM-it
    – **region**: https://www.merriam-webster.com/dictionary/region
    IPA: /ˈriːdʒən/
    Phonetic Spelling: REE-juhn
    Combined: SUM-it REE-juhn