Submission ID #: 66840
Scriptwriter Name: Debopriya Sadhukhan
Project Page Link: https://review.jove.com/account/file-uploader?src=20386248

Title: A Web-Based Workflow for Selecting Gene- and Tissue-Specific Enhancers

Authors and Affiliations:

Jooa Kwon [1,2], George Z. He[3,4], Mirana Ramialison[1,2,3,4,5], Hieu T. Nim [1,2,3,4]

[1]Department of Paediatrics, Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne
[2]Australian Regenerative Medicine Institute, Monash University
[3]Stem Cell Medicine Department, Murdoch Children's Research Institute, The Royal Children's Hospital
[4]The Novo Nordisk Foundation Center for Stem Cell Medicine, reNEW Melbourne, Murdoch Children's Research Institute
[5]System Biology Institute (SBI) Australia

Corresponding Authors:

Mirana Ramialison (mirana.ramialison@mcri.edu.au)
Hieu T. Nim (hieu.nim@mcri.edu.au)

Email Addresses for All Authors:

Jooa Kwon (jooa.kwon@mcri.edu.au)
George Z. He (george.he@mcri.edu.au)
Mirana Ramialison (mirana.ramialison@mcri.edu.au)
Hieu T. Nim (hieu.nim@mcri.edu.au)

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**

3. **Filming location:** Will the filming need to take place in multiple locations? **N/A**

Current Protocol Length
Number of Steps: 15
Number of Shots: 35

Introduction

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

1.1. **George Z. He**: We are interested in decoding the non-coding genome, to understand how genes are regulated to be expressed at the right time and at the right place. We develop user-friendly protocols using web-based genomics tools to make enhancer discovery accessible to all biologists.

1.1.1. INTERVIEW: Named talent says the statement above in an interview-style shot, looking slightly off-camera. *Suggested B roll: 4.5.3*

What are the most recent developments in your field of research?

1.2. **George Z. He:** Wide availability of multi-modal data that can help narrow down the location of enhancers, and the easy access of this data through open-source web-interfaces. This enables comprehensive enhancer identification without requiring programming expertise from researchers.

1.2.1. INTERVIEW: Named talent says the statement above in an interview-style shot, looking slightly off-camera. *Suggested B roll: 3.1.3*

What research gap are you addressing with your protocol?

1.3. **George Z. He:** This protocol provides a foundation for anyone without a background in enhancer biology to start navigating publicly available datasets to retrieve enhancers for their genes of interest.

1.3.1. INTERVIEW: Named talent says the statement above in an interview-style shot, looking slightly off-camera. *Suggested B roll: 2.3*

What new scientific questions have your results paved the way for?

1.4. **George Z. He:** We used TBX5, a well-known player in heart biology, as a case study. Our workflow identified 21 enhancers, which may shed light on heart development mechanisms and congenital heart disease.

1.4.1. INTERVIEW: Named talent says the statement above in an interview-style shot, looking slightly off-camera. *Suggested B roll: figure 5*

What research questions will your laboratory focus on in the future?

1.5. **George Z. He:** We'll expand our protocol with new data sources, including spatial genomics, while developing additional user-friendly tools for biologists.
   1.5.1. INTERVIEW: Named talent says the statement above in an interview-style shot, looking slightly off-camera.

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

Protocol

2. Locating the Gene of Interest and Defining the Enhancer Detection Region

Demonstrator: George Z. He

2.1. To begin, open the EnsEMBL *(En-sem-bull)* genome browser **[1]**. Select the genome assembly that matches the species and version of interest **[2]**. Enter the gene of interest in the search field and click **Go [3]**. From the results, click on the appropriate EnsEMBL gene ID. Then, scroll to the summary section and click the **Region in detail** link to visualize the surrounding region of the gene of interest **[5]**.

2.1.1. WIDE: 66840_2.1-3.4.2_t7.mp4: 00:10-00:16

2.1.2. SCREEN: 66840_2.1-3.4.2_t7.mp4: 00:18-00:24

2.1.3. SCREEN: 66840_2.1-3.4.2_t7.mp4: 00:24-00:27

2.1.4. SCREEN: 66840_2.1-3.4.2_t7.mp4: 00:33-00:45

2.2. Now, search for the two genes flanking the gene of interest using the Gene Legend track. These genes appear as visual elements representing merged EnsEMBL and Havana annotations within the Basic Gene Annotations from the GENCODE track **[1]**. Determine the directionality of the gene of interest by observing the greater-than or less-than signs beside the gene names **[2]**. Click and drag the cursor across the intergenic region between these genes, then click **Jump to region** in the pop-up box to view the selected region **[4-TXT]**.

2.2.1. SCREEN: 66840_2.1-3.4.2_t7.mp4: 00:53-01:06

2.2.2. SCREEN: 66840_2.1-3.4.2_t7.mp4: 01:13-01:20

2.2.3. SCREEN: 66840_2.1-3.4.2_t7.mp4: 02:02-02:17. **TXT: Redefine the region of interest at any time by repeating this step**

2.3. Customize the display by clicking **Add/remove tracks** *(Add or remove tracks)* at the top of the track viewer **[1]**. Use the zoom and navigation controls to adjust the view for enhanced visualization of the region **[2]**.

2.3.1. SCREEN: 66840_2.1-3.4.2_t7.mp4: 02:23-02:25

2.3.2. SCREEN: 66840_2.1-3.4.2_t7.mp4: 02:26-02:32

3. Histone Mark Analysis

3.1. In the **Region in detail** tab viewer, click **Configure this page** in the sidebar **[1]**. In the **Configure Region Image** sidebar, under **Regulation**, select **Activity by Cell/ Tissue** *(Activity-By-Cell-Or-Tissue)* **[2]**. Use the **Cell/Tissue** search bar to find and select the desired tissues or use the alphabetical index below the bar **[3]**. Click on the **Experiments** tab adjacent to the **Cell/ Tissue** option, choose **H3K4me1** *(H-three-K-four-M-E-one)* and **H3K27ac** *(H-three-K-two-seven-A-C)* as a marker for enhancers and **H3K4me3** as a marker for promoters, and click **Configure track display [4]**.

3.1.1. SCREEN: 66840_2.1-3.4.2_t7.mp4: 02:43-02:56

3.1.2. SCREEN: 66840_2.1-3.4.2_t7.mp4: 03:08-03:16

3.1.3. SCREEN: 66840_2.1-3.4.2_t7.mp4: 03:19-03:42

3.2. Then, select **View tracks** to visualize regions marked by **H3K4me1** within the enhancer detection region and regions marked by **H3K4me3** upstream of the gene of interest **[5]**.

3.2.1. SCREEN: 66840_2.1-3.4.2_t7.mp4: 03:53-04:00, 04:14-04:15

3.3. Click on the colored visual or box elements in the H3K4me1 track to retrieve the genomic coordinates of regions marked within the defined detection region. This opens the **Hists & Pols** *(Hists and Pols)* pop-up, which displays the element's genomic location in base pairs **[1]**.

3.3.1. SCREEN: 66840_2.1-3.4.2_t7.mp4: 04:27-.04:52

3.4. Alternatively, manually define the regions of interest for each genomic feature by clicking and dragging on the track to enclose graph peaks under the H3K4me1 or H3K27ac tracks **[1]**. Then, copy the genomic location coordinates into a text file and save the file in **.bed** *(B-E-D)* format **[2-TXT]**.

3.4.1. SCREEN: 66840_2.1-3.4.2_t7.mp4: 05:11-05:37

3.4.2. SCREEN: 66840_2.1-3.4.2_t7.mp4: 06:07-06:12 **TXT: Repeat for promoter regions using the H3K4me3 track upstream of the GoI**

4. Chromatin Conformation Capture (Hi-C) Analysis

4.1. Access the 4DN *(Four-D-N)* data portal **[1]**. On the homepage, ensure that **Experiment Sets** is selected as the Y-axis of the main stacked bar, **Experiment Type** is selected as the X-axis, and the plot is grouped by **Organism [2]**.

4.1.1. SCREEN: 66840_4.1-4.7_t3.mp4: 00:04-00:09

4.1.2. SCREEN:66840_4.1-4.7_t3.mp4: 00:10-00:00:22

4.2. Locate the ***in situ*** **Hi-C** *(In-Situ-Hi-C)* bar along the X-axis and click on the portion representing **Human Experiment Sets [1]**. In the pop-up, click the **Browse** button and select cardiac muscle myoblasts and our gene of interest, TBX5 associated with heart organogenesis **[2].**

4.2.1. SCREEN: 66840_4.1-4.7_t3.mp4: 00:28-00:38

4.2.2. SCREEN: 66840_4.1-4.7_t3.mp4: 00:39-00:48, 00:56-01:02

4.3. Click on the link in the **Title** column of the relevant biosample corresponding to the tissue of interest **[1]**. Then, in the **Processed Files** tab, click **Explore Data** to examine the Hi-C dataset in more detail **[2]**.

4.3.1. SCREEN: 66840_4.1-4.7_t3.mp4: 01:14-01:27

4.3.2. SCREEN66840_4.1-4.7_t3.mp4: 01:28-01:34

4.4. Enter the coordinates of the identified promoter in the tissue of interest and right-click the heatmap to mark the region horizontally. These lines allow the promoter region to be visually tracked across the heatmap **[1].** If required, adjust the view by dragging it vertically until the upper and lower bounds in the Y-coordinates of the search bar align with the boundaries of the region of interest **[2]**. To remove any unwanted lines, right-click the line and select **horizontal/vertical rule** *(horizontal or vertical rule)*, then click **Close Series [3]**.

4.4.1. SCREEN: 66840_4.1-4.7_t3,mp4: 01:43-02:05

4.4.2. SCREEN: 66840_4.1-4.7_t3.mp4: 02:58-03:07

4.4.3. SCREEN: 66840_4.1-4.7_t3.mp4: 02:08-02:16

4.5. Enter the coordinates of all experimentally validated control enhancers to calculate the interaction threshold based on their minimum interaction values **[1]**. Display the three control enhancers from the literature alongside the promoter region from EnsEMBL using the pre-prepared view **[2]**. Then, define the promoter-enhancer threshold using these control enhancers by selecting the lowest non-zero interaction score, as indicated by the color key on the right side of the heatmap matrix **[3]**.

4.5.1. SCREEN: 66840_4.1-4.7_t3.mp4: 03:39-03:43.

4.5.2. SCREEN: 66840_4.1-4.7_t3.mp4: 03:58-04:09

4.5.3. SCREEN: 66840_4.1-4.7_t3.mp4: 04:19-04:23.

4.6. Enter the genomic coordinates of all H3K4me1-associated enhancer regions and mark

vertically on the Hi-C heatmap **[1]**.

4.6.1. SCREEN: : 66840_4.1-4.7_t3.mp4: 04:36-04:42

4.7. Filter out weakly interacting regions by comparing the interaction scores of H3K4me3-marked regions against the interaction threshold **[1]**. Select the genomic coordinates showing interaction frequencies above this threshold, which appear as more concentrated or darker signals on the heatmap, and save them in BED *(B-E-D)* format **[2]**.

4.7.1. SCREEN: 66840_4.1-4.7_t3.mp4: 04:58-05:02, 05:10-05:35

4.7.2. SCREEN: 66840_4.1-4.7_t3.mp4: 05:40-05:50

Results

5. Results

5.1. Three of the four cardiac enhancers identified in the VISTA Cardiac Enhancer Browser—hs2329 *(H-S-Two-Three-Two-NIne)*, mm1282 *(M-M-One-Two-Eight-Two)*, and mm370 *(M-M-Three-Seventy)*—overlapped with regions predicted by the web-based enhancer detection protocol **[1].**

5.1.1. LAB MEDIA: Figure 5. *Video editor: Highlight the brown bar labeled "hs2329", "mm1282" and "mm370"* .

5.2. Enhancer 2 overlapped with both a region predicted by the enhancer detection protocol and an experimentally validated enhancer **[1].** Enhancer 16 showed overlap with both predicted enhancer regions and prior experimental data **[2].**

5.2.1. LAB MEDIA: Figure 5. *Video editor: Highlight the boxed region labeled "Enhancer 2"*

5.2.2. LAB MEDIA: Figure 5. *Video editor: Highlight the boxed region labeled "Enhancer 16"*

5.3. Enhancer 9 did not overlap with any predicted enhancers from the pipeline but showed partial H3K4me1 signal enrichment **[1].**

5.3.1. LAB MEDIA: Figure 5. *Video editor: Highlight the boxed region labeled "Enhancer 9"*

1. **EnsEMBL**
   - Pronunciation link: https://www.ensembl.org (no direct dictionary entry, but the project website uses "En-sembl")

- IPA: /ˈɛn sɛmbl/
- Phonetic: EN-semm-bull

2. **GENCODE**
   - Pronunciation link: No confirmed dictionary listing
   - IPA: /dʒiːˈnoʊd/
   - Phonetic: jee-NOHD
3. **H3K4me1**
   - (This is a formula / code, pronounced “H three K four me one”)
   - IPA: /eɪtʃ θri keɪ fɔr mi wʌn/
   - Phonetic: H-three K-four me-one
4. **H3K27ac**
   - (“H three K twenty-seven ac”)
   - IPA: /eɪtʃ θri keɪ ˈtwɛnti ˈsɛvən eɪ ˈsi/
   - Phonetic: H-three K-twen-ty seven ac
5. **H3K4me3**
   - (“H three K four me three”)
   - IPA: /eɪtʃ θri keɪ fɔr mi θri/
   - Phonetic: H-three K-four me-three
6. **intergenic**
   - Pronunciation link: https://www.merriam-webster.com/dictionary/intergenic
   - IPA: /ˌɪn.təˈdʒɛn.ɪk/
   - Phonetic: in-tuh-JEN-ik
7. **promoter**
   - Pronunciation link: https://www.merriam-webster.com/dictionary/promoter
   - IPA: /prəˈmoʊtər/
   - Phonetic: pruh-MOH-tər
8. **enhancer**
   - Pronunciation link: https://www.merriam-webster.com/dictionary/enhancer
   - IPA: /ɛnˈhænsər/
   - Phonetic: en-HAN-sir

9. **Hi-C**
    - Pronounced “H I C” or “High-C” (context determines)
    - IPA (as acronym): /eɪtʃ aɪ si/
    - Phonetic: H-I-C
10. **in situ**
    - Pronunciation link: https://www.merriam-webster.com/dictionary/in%20situ
    - IPA: /ɪn ˈsaɪtu/
    - Phonetic: in SY-too
11. **biosample**
    - No major dictionary; compound “bio + sample”
    - IPA: /ˌbaɪ.oʊˈsæmpəl/
    - Phonetic: bye-oh-SAMP-ul
12. **genomic**
    - Pronunciation link: https://www.merriam-webster.com/dictionary/genomic
    - IPA: /dʒəˈnoʊmɪk/
    - Phonetic: juh-NOH-mik