FINAL SCRIPT: APPROVED FOR FILMING

Submission ID #: 67266
Scriptwriter Name: Poornima G
Project Page Link: https://review.jove.com/account/file-uploader?src=20516178

Title: Using R, Seurat, and CellChat to Analyze a Single-Cell Transcriptomics Dataset of Mouse Skin Wound Healing

Authors and Affiliations:
Shalyn Keiser, Nichole Botello, Ethan Cruz, Mateusz S. Wietecha

Department of Oral Biology, College of Dentistry, University of Illinois Chicago

Corresponding Authors:
Mateusz S. Wietecha mwietec2@uic.edu

Email Addresses for All Authors:
Shalyn Keiser skeiser2@uic.edu
Nichole Botello nbote2@uic.edu
Ethan Cruz ecruz45@uic.edu
Mateusz S. Wietecha mwietec2@uic.edu

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**

Current Protocol Length
Number of Steps: 25
Number of Shots: 57 (56 SC)

Introduction

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

1.1. **Mateusz Wietecha:** In our lab, we use emerging tools such as single-cell and spatial transcriptomics with systems biology and bioinformatics approaches to investigate the spatio-temporal cellular dynamics of differential healing outcomes.

1.1.1. INTERVIEW: Named talent says the statement above in an interview-style shot, looking slightly off-camera. *Suggested B-roll: 2.3.1*

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

1.2. **Mateusz Wietecha:** In recent years, we have seen a rapid adoption of single-cell transcriptomics to the study of wound healing in humans and in model organisms such as mice.

1.2.1. INTERVIEW: Named talent says the statement above in an interview-style shot, looking slightly off-camera. *Suggested B-roll: 2.5.1*

What are the current experimental challenges?

1.3. **Mateusz Wietecha:** The full analysis of single-cell datasets is prohibitive for bench scientists with little to no experience with bioinformatics. This means that, all too often, single-cell datasets are underutilized by scientists in the field of wound healing.

1.3.1. INTERVIEW: Named talent says the statement above in an interview-style shot, looking slightly off-camera. *Suggested B-roll: 3.5.1*

What research gap are you addressing with your protocol?

1.4. **Mateusz Wietecha:** This is the first comprehensive protocol that assumes no prior experience with bioinformatics, which takes a user all the way from dataset download to the output of relevant analyses in the context of wound healing research.

1.4.1. INTERVIEW: Named talent says the statement above in an interview-style shot, looking slightly off-camera. *Suggested B-roll: 3.12.1*

How will your findings advance research in your field?

1.5. **Mateusz Wietecha:** Our protocol should serve as a template for wound healing researchers to more fully analyze their own single-cell datasets and be able to extract new insights from publicly available datasets.

1.5.1. INTERVIEW: Named talent says the statement above in an interview-style shot, looking slightly off-camera. *Suggested B-roll: 4.1.1*

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

Testimonial Questions (OPTIONAL):

Videographer: Please capture all testimonial shots in a wide-angle format with sufficient headspace, as the final videos will be rendered in a 1:1 aspect ratio. Testimonial statements will be presented live by the authors, sharing their spontaneous perspectives.

How do you think publishing with JoVE will enhance the visibility and impact of your research?

1.6. **Mateusz Wietecha, DMD, PhD**. **Assistant Professor at the Department of Oral Biology, University of Illinois Chicago College of Dentistry**: Publishing with JoVE will make single-cell transcriptomics and bioinformatics methods more accessible to the wound healing research community, which will also enhance the impact of our lab's research in this field since it will enable more scientists to make use of our datasets and tools in their own labs and projects.

1.6.1. INTERVIEW: Named talent says the statement above in an interview-style shot, looking slightly off-camera. *Suggested B-roll: 2.3.1*

Protocol

2. Loading a Single-Cell Wound Healing Dataset and Performing Quality Control

Demonstrator: Shalyn Keiser

2.1. To begin, navigate to the dataset files from the Gene Expression Omnibus repository using the accession number GSE204777 *(G-S-E-2-0-4-7-7-7)* **[1-TXT]**. Click on the first dataset titled GSM6190913 *(G-S-M-6-1-9-0-9-1-3)* **[2]**.

2.1.1. WIDE: Talent taking the seat at the computer. **TXT: Download the JoVE_Rscript.R for easy execution**

2.1.2. SCREEN: 67266_ScreenCapture_Steps2.1-2.10.mp4. 0:00:01-0:00:13

2.2. Scroll to the bottom of the GSM6190913 page and download the three listed files using either the **ftp** or **html** links **[1]**. Using the computer's file explorer, move the downloaded files into a directory named **b1**, ensuring it is located within the working directory **[2]**.

2.2.1. SCREEN: 67266_ScreenCapture_Steps2.1-2.10.mp4. 0:00:13-0:00:32

2.2.2. SCREEN: 67266_ScreenCapture_Steps2.1-2.10.mp4. 0:00:33-0:00:42

2.3. Retrieve the directory path information for the single-cell sequencing files that were downloaded **[1]**.

2.3.1. SCREEN: 67266_ScreenCapture_Steps2.1-2.10.mp4. 0:00:44-0:00:51

2.4. Now, load the single-cell sequencing files into the working environment **[1]**. Then, separate the gene expression and multiplexing HTO data from the working dataset **[2]**.

2.4.1. SCREEN: 67266_ScreenCapture_Steps2.1-2.10.mp4. 0:00:52-0:01:10

2.4.2. SCREEN: 67266_ScreenCapture_Steps2.1-2.10.mp4. 0:01:29-0:01:39

2.4.3. SCREEN: 67266_ScreenCapture_Steps2.1-2.10.mp4. 0:01:29-0:01:39

2.5. Create a Seurat object using the gene expression data while filtering out genes detected in fewer than 5 cells and cells with fewer than 200 genes **[1]**. For datasets lacking HTO data, create the Seurat object with the same filtering parameters **[2]** and switch to the gene

expression assay **[3]**.

2.5.1. SCREEN: 67266_ScreenCapture_Steps2.1-2.10.mp4. 0:01:40-0:02:00

2.5.2. SCREEN: 67266_ScreenCapture_Steps2.1-2.10.mp4. 0:03:00-0:03:16

2.6. Calculate the mitochondrial gene percentage in each cell and assign this value as a metadata variable **[1]**.

2.6.1. SCREEN: 67266_ScreenCapture_Steps2.1-2.10.mp4. 0:03:22-0:03:28

2.6.2. SCREEN: 67266_ScreenCapture_Steps2.1-2.10.mp4. 0:03:30-0:03:34

2.6.3. SCREEN: 67266_ScreenCapture_Steps2.1-2.10.mp4. 0:03:44-0:03:48

2.7. Visualize the distribution of gene counts, total RNA, and mitochondrial gene percentage across all cells **[1]**.

2.7.1. SCREEN: 67266_ScreenCapture_Steps2.1-2.10.mp4. 0:03:49-0:04:01

2.8. Remove cells with mitochondrial content exceeding 25 percent using a threshold **[1]** and visualize the updated distributions after filtering out these low-quality cells **[2]**.

2.8.1. SCREEN: 67266_ScreenCapture_Steps2.1-2.10.mp4. 0:04:07-0:04:13

2.8.2. SCREEN: 67266_ScreenCapture_Steps2.1-2.10.mp4. 0:04:14-0:04:28

2.9. Detect likely doublets using the scDblFinder *(S-C-Doublet finder)* method. Run the scDblFinder pipeline using the commands and assign the resulting doublet scores as a new metadata variable **[1]**.

2.9.1. SCREEN:67266_ScreenCapture_Steps2.1-2.10.mp4 0:04:30-0:04:55

2.10. Now, visualize the distribution of doublet scores across all cells **[1]**. Remove all cells with doublet scores above 0.25 **[2]** and save the cleaned Seurat object as an RDS file in the working directory **[3]**.

2.10.1. SCREEN: 67266_ScreenCapture_Steps2.1-2.10.mp4. 0:04:57-0:05:05

2.10.2. SCREEN: 67266_ScreenCapture_Steps2.1-2.10.mp4. 0:05:08–0:05:12

2.10.3. SCREEN: 67266_ScreenCapture_Steps2.1-2.10.mp4. 0:05:13–0:05:17

3. Analyzing a Single-Cell Wound Healing Dataset Using Seurat

3.1. Perform data normalization, scaling and principal component analysis **[1]**. Visualize the variance contribution across the first 50 principal components **[2]**.

3.1.1. SCREEN: 67266_ScreenCapture_Steps3.1-3.15.mp4. 0:00:01-0:00:15

3.1.2. SCREEN: 67266_ScreenCapture_Steps3.1-3.15.mp4. 0:00:16-0:00:26

3.2. Cluster the cells using the first 13 principal components and a clustering resolution of 0.1 **[1]**.

3.2.1. SCREEN: 67266_ScreenCapture_Steps3.1-3.15.mp4. 0:00:27-0:00:36

3.3. Perform Uniform Manifold Approximation and Projection or UMAP *(U-map)* reduction and neighbor analysis using the first 13 principal components and set the seed number to 123 **[1]**.

3.3.1. SCREEN: 67266_ScreenCapture_Steps3.1-3.15.mp4. 0:00:37-0:00:40 and 0:01:00-0:01:07

3.4. Now, visualize cell clustering on a UMAP plot **[1]**, followed by wound time and space annotations on a UMAP plot **[2]**.

3.4.1. SCREEN: 67266_ScreenCapture_Steps3.1-3.15.mp4. 0:01:08-0:01:22

3.4.2. SCREEN: 67266_ScreenCapture_Steps3.1-3.15.mp4. 0:01:23-0:01:35

3.5. Then, generate a table associating cell clusters with wound time and space annotations **[1]**.

3.5.1. SCREEN: 67266_ScreenCapture_Steps3.1-3.15.mp4. 0:01:36-0:01:44

3.6. Determine major cell type identities after calculating differentially expressed genes between all clusters **[1]** and assign the resulting DEG lists to a variable before saving them as a delimited text file in the working directory **[2]**.

3.6.1. SCREEN: 67266_ScreenCapture_Steps3.1-3.15.mp4. 0:01:44-0:01:59

3.6.2. SCREEN: 67266_ScreenCapture_Steps3.1-3.15.mp4. 0:02:00-0:02:04

3.7. Now, open the **dataset_cluster_markers.txt** *(dataset cluster markers)* file in a spreadsheet application **[1]**. Use the **Text Import Wizard** to set the comma as a delimiter and

format gene name columns as **text** to prevent automatic conversion of gene names into dates **[2]**.

3.7.1. SCREEN: 67266_ScreenCapture_Steps3.1-3.15.mp4. 0:02:05-0:02:14

3.7.2. SCREEN: 67266_ScreenCapture_Steps3.1-3.15.mp4. 0:02:14-0:02:34

3.8. In a spreadsheet, rank the **avg_log2FC** *(average log 2 F-C)* column from largest to smallest to order the rows by decreasing log2 fold change values **[1]**, followed by the **cluster** column from smallest to largest to order the rows by increasing Seurat cluster numbers **[2]**. Filter the **avg_log2FC** column to include only values greater than or equal to 2.5 **[3]** and then filter the **pct.1** *(P-C-T-1)* column to include values greater than or equal to 0.4 **[4]**. Next, filter the **pct.2** *(P-C-T-2)* column to include values less than or equal to 0.2 **[5]**. Finally, filter the p_val_adj *(P value adjusted)* column to include values less than or equal to 0.01 **[6]**.

3.8.1. SCREEN: 67266_ScreenCapture_Steps3.1-3.15.mp4. 0:02:36-0:02:51

3.8.2. SCREEN: 67266_ScreenCapture_Steps3.1-3.15.mp4. 0:02:52-0:03:02

3.8.3. SCREEN: 67266_ScreenCapture_Steps3.1-3.15.mp4. 0:03:03-0:03:22

3.8.4. SCREEN: 67266_ScreenCapture_Steps3.1-3.15.mp4. 0:03:22-0:03:36

3.8.5. SCREEN: 67266_ScreenCapture_Steps3.1-3.15.mp4. 0:03:36-0:03:53

3.8.6. SCREEN: 67266_ScreenCapture_Steps3.1-3.15.mp4. 0:03:53-0:04:05

3.9. Now, open the EnrichR *(enrich R)* web-based enrichment analysis tool **[1]**. For each cluster, copy the list of differentially expressed genes into a separate EnrichR window and click **Analyze [2]**.

3.9.1. SCREEN: 67266_ScreenCapture_Steps3.1-3.15.mp4. 0:04:07-0:04:18

3.9.2. SCREEN: 67266_ScreenCapture_Steps3.1-3.15.mp4. 0:04:20-0:04:31

3.10. Then, click the **Cell Types** tab above the analysis output **[1]** and focus on the top 5 enrichments within the three curated cell marker databases **[2]**.

3.10.1. SCREEN: 67266_ScreenCapture_Steps3.1-3.15.mp4. 0:04:32-0:04:36

3.10.2. SCREEN: 67266_ScreenCapture_Steps3.1-3.15.mp4. 0:04:36-0:04:42

3.11. Based on the top enrichments from the EnrichR analysis, assign likely identities to the 8 clusters **[1]**. Combine clusters 2 and 6 into a single annotation labeled fibroblasts **[2]** and assign these annotations as a new metadata variable named **cell_types** *(cell types)* **[3]**.

3.11.1. SCREEN: 67266_ScreenCapture_Steps3.1-3.15.mp4. 0:04:44-0:04:53

3.11.2. SCREEN: 67266_ScreenCapture_Steps3.1-3.15.mp4. 0:04:53-0:04:56

3.11.3. SCREEN: 67266_ScreenCapture_Steps3.1-3.15.mp4. 0:04:56-0:05:01

3.12. Then, visualize the annotated cell types on a UMAP plot **[1]** and display the localization of the bolded top cluster marker genes on a series of UMAP plots **[2]**.

3.12.1. SCREEN: 67266_ScreenCapture_Steps3.1-3.15.mp4. 0:05:02-0:05:16

3.12.2. SCREEN: 67266_ScreenCapture_Steps3.1-3.15.mp4. 0:05:27-0:05:32; 0:05:45-0:05:53

3.13. Visualize the top cluster marker differentially expressed genes on a dot plot grouped by original cluster numbers **[1]** and then the top marker genes again in a dot plot, this time grouped by annotated cell types **[2]**.

3.13.1. SCREEN: 67266_ScreenCapture_Steps3.1-3.15.mp4. 0:06:10-0:06:22

3.13.2. SCREEN: 67266_ScreenCapture_Steps3.1-3.15.mp4. 0:06:24-0:06:33

3.14. To prepare for time-series analysis, remove spatial annotation and simplify the dataset **[1]**. Reassign the wound time and space metadata into a new variable called **DPW** for days post-wounding **[2]**. Visualize the new **DPW** time-course groupings on a UMAP plot **[3]** and generate tables showing the number of cells of each type within each **DPW** group **[4]**.

3.14.1. SCREEN: 67266_ScreenCapture_Steps3.1-3.15.mp4. 0:06:35-0:06:47

3.14.2. SCREEN: 67266_ScreenCapture_Steps3.1-3.15.mp4. 0:06:35-0:06:47

3.14.3. SCREEN: 67266_ScreenCapture_Steps3.1-3.15.mp4. 0:06:49-0:07:03

3.14.4. SCREEN: 67266_ScreenCapture_Steps3.1-3.15.mp4. 0:07:04-0:07:18

3.15. Next, convert the cell counts to proportions to assess relative changes in cell type composition during healing **[1]** and visualize the proportion of each **DPW** category within each cell type **[2]**. Finally, visualize the proportion of each cell type within each **DPW** group **[3]** and save the final Seurat object containing all annotations and filters as an RDS file into the working directory **[4]**.

3.15.1. SCREEN: 67266_ScreenCapture_Steps3.1-3.15.mp4. 0:07:21-0:07:30

3.15.2. SCREEN: 67266_ScreenCapture_Steps3.1-3.15.mp4. 0:07:31-0:07:37

3.15.3. SCREEN: 67266_ScreenCapture_Steps3.1-3.15.mp4. 0:07:44-0:07:57

3.15.4. SCREEN: 67266_ScreenCapture_Steps3.1-3.15.mp4. 0:08:00-0:08:05

Results

4. Results

4.1. All cells in the dataset clustered distinctly into major color-coded cell types on the UMAP plot, confirming successful annotation based on enriched cell type signatures **[1]**.

4.1.1. LAB MEDIA: Figure 4. *Video editor: Highlight each color-coded region one by one along with the corresponding cell type labels on the right.*

4.2. High expression of top cluster marker genes was localized within their respective cell type clusters on the UMAP plots **[1]**.

4.2.1. LAB MEDIA: Figure 5.

4.3. Dot plot visualization confirmed the highest expression levels of cluster marker genes were restricted to their annotated major cell types **[1]**.

4.3.1. LAB MEDIA: Figure 6. *Video editor: sequentially highlight the big red dots* .

4.4. The stacked bar plot revealed that neutrophils and macrophages were dominant at day 1 post-wounding **[1]**, while fibroblasts, epithelial, and endothelial cells became more prevalent at later time points, reflecting the known cellular cascade of wound healing **[2]**.

4.4.1. LAB MEDIA: Figure 7. *Video editor: Highlight the sections of "neutrophils" and "macrophages" in bar "D1".*

4.4.2. LAB MEDIA: Figure 7. *Video editor: Emphasize the coloured sections for "fibroblasts, epithelial cells, and endothelial cells" in the bars "D7" and "D14".*

Pronunciation guide:

---

1. Seurat

- **Pronunciation link**: https://www.youtube.com/watch?v=GBJiNc7ov2o
- **IPA**: /səˈrɑː/
- **Phonetic Spelling**: suh-RAH(youtube.com)

2. UMAP (Uniform Manifold Approximation and Projection)

- **Pronunciation link**: https://www.howtopronounce.com/umap
- **IPA**: /ˈjuːmæp/
- **Phonetic Spelling**: YOO-map(howtopronounce.com, howtosay.co.in)

3. scDblFinder

- **Pronunciation link**: No confirmed link found
- **IPA**: /ɛs siː ˈdʌbəl ˈfaɪndər/
- **Phonetic Spelling**: ess-see DUB-uhl FIND-er

4. EnrichR

- **Pronunciation link**: No confirmed link found
- **IPA**: /ɛnˈrɪtʃɑːr/
- **Phonetic Spelling**: en-RICH-ar(r-packages.io, dictionary.cambridge.org, merriam-webster.com)