Submission ID #: 68892
Scriptwriter Name: Pallavi Sharma
Project Page Link: https://review.jove.com/account/file-uploader?src=21012108

Title: Mining Spatial Transcriptomics Datasets using DeepSpaceDB

Authors and Affiliations:
Nupura Prabhune[1,2], Yilin Du[1,3], Afeefa Zainab[4], Satoru Ebihara[3], Shinji Takeoka[2], Shinpei Kawaoka[1,5], Alexis Vandenbon[4,6]

[1]Department of Integrative Bioanalytics, Institute of Development, Aging and Cancer, Tohoku University
[2]Department of Life Science and Medical Bioscience, Graduate School of Advanced Science and Engineering, Waseda University
[3]Department of Rehabilitation Medicine, Tohoku University Graduate School of Medicine
[4]Institute for Life and Medical Sciences, Kyoto University
[5]Inter-Organ Communication Research Team, Institute for Life and Medical Sciences, Kyoto University
[6]Institute for Liberal Arts and Sciences, Kyoto University

Corresponding Authors:
Shinpei Kawaoka shinpei.kawaoka.c1@tohoku.ac.jp
Alexis Vandenbon alexisvdb@infront.kyoto-u.ac.jp

Email Addresses for All Authors:
Shinpei Kawaoka shinpei.kawaoka.c1@tohoku.ac.jp
Alexis Vandenbon alexisvdb@infront.kyoto-u.ac.jp
Nupura Prabhune nupuraprabhune@fuji.waseda.jp
Yilin Du du.yilin.q2@dc.tohoku.ac.jp
Afeefa Zainab zainab.afeefa.5r@kyoto-u.ac.jp
Satoru Ebihara satoru.ebihara.c4@tohoku.ac.jp
Shinji Takeoka takeoka@waseda.jp

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? **No**

Current Protocol Length
Number of Steps: 22
Number of Shots: 45

Introduction

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

1.1. **Alexis Vandenbon:** We are building a spatial transcriptomics database, DeepSpaceDB, with the goal of making spatial transcriptomics data more accessible for biologists and bioinformaticians.

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 most recent developments in your field of research?

1.2. **Alexis Vandenbon:** Several spatial transcriptomics platforms have been developed. They allow researchers to study gene expression patterns within tissue samples. But this technology is expensive, and the data analysis requires high-level bioinformatics expertise.

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

What technologies are currently used to advance research in your field?

1.3. **Yilin Du:** We have been using the Visium and Xenium spatial platforms in our cancer cachexia research. These platforms allow us to dissect tumors and surrounding or even distant host tissues within the same organ context, and to resolve changes in gene expression and cellular composition in each compartment separately.

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

What are the current experimental challenges?

1.4. **Nupura Prabhune**: One major challenge for biologists is data analysis. Many researchers still lack sufficient programming or computational expertise, which makes it difficult to fully explore and interpret the increasingly large number of spatial datasets now available.

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

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

1.5. **Nupura Prabhune:** By making spatial data more easily accessible, this database can be used to generate new hypotheses about mechanisms underlying diseases. For example, what genes are induced in tumor microenvironments, etc.

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. Analysis of a Mouse Brain Sample

Demonstrators: Nupura Prabhune and Alexis Vandenbon

2.1. To begin, click on the **Database** tab **[1]** and select the organism as **mouse**, the organ as **brain**, and the source as **Zenodo [2]**. Scroll through the resulting samples and select the sample labeled DSID001557 *(D-S-I-D-Zero-Zero-One-Five-Five-Seven)* **[3]**.

2.1.1. WIDE: Talent in front of the computer screen, clicking on the Database tab. NOTE: Two takes with both talent

2.1.2. SCREEN: 68892_screenshot_1.mp4: 00:05-00:15

2.1.3. SCREEN: 68892_screenshot_1.mp4: 00: 21-00:25

2.2. Then, click on the selected sample and confirm that the description reads **2 × 10^6 cells in 100 µL saline-NK cell** *(2 million cells in 100 microliters saline-N-K cell)* **[1]**.

2.2.1. SCREEN: 68892_screenshot_2.mp4: 00:07-00:10

2.3. Click on the **Quality** tab to evaluate the sample quality **[1]**. From the **Quality Measures** drop-down menu, select options such as **Detected Genes**, **Read Count**, and **Mito** to view the respective parameter distributions across the sample slice **[2]**.

2.3.1. SCREEN: 68892_screenshot_3.mp4: 00:00-00:03

2.3.2. SCREEN: 68892_screenshot_3.mp4: 00:05-00:19

2.4. Now, navigate to the **Image Annotation** tab to identify different regions in the sample slice **[1]**.

2.4.1. SCREEN: 68892_screenshot_4.mp4: 00:00-00:02

2.5. Move the mouse cursor over the **sample slice** to display annotations. View the grid-based annotations generated by a large language model that show anatomical features and associated conditions **[1-TXT]**.

2.5.1. SCREEN: 68892_screenshot_4.mp4: 00:02-00:19

TXT: These are AI-generated predictions; some samples include expert annotations

2.6. Then, navigate to the **Clusters** tab to examine the cell type clusters in the sample slice **[1]**. View the two-dimensional embedding of the clusters and the corresponding color-coded representation across spots on the sample slice **[2]**.

2.6.1. SCREEN: 68892_screenshot_5.mp4: 00:00-00:03

2.6.2. SCREEN: : 68892_screenshot_5.mp4: 00:03-00:14

2.7. Next, navigate to the **Genes** tab to examine the spatially variable genes in the sample **[1]**.

2.7.1. SCREEN: : 68892_screenshot_6.mp4: 00:00-00:02

2.8. Click on some of the top genes in the list to generate spatial plots of their expression across the tissue slice. Observe the color-coded expression patterns, which clearly show distinct spatial distributions for the highest-scoring genes **[1]**.

2.8.1. SCREEN: : 68892_screenshot_6.mp4: 00:02-00:18

2.9. Then, navigate to the **Pathways** tab to examine the activity of gene sets associated with common biological pathways **[1]**. View the list of spatially variable pathways, with pathway activity estimated based on the expression levels of related genes **[2]**.

2.9.1. SCREEN: : 68892_screenshot_7.mp4: 00:00-00:03

2.9.2. SCREEN: : 68892_screenshot_7.mp4: 00:03-00:13

2.10. Click on some of the top pathways in the list to generate spatial plots of their activity across the tissue slice. Observe the color-coded patterns of pathway activity across different tissue regions **[1]**.

2.10.1. SCREEN: SCREEN: : 68892_screenshot_7.mp4: 00:13-00:18

2.11. Now, go to the **Tissue Explorer** tab, which allows users to freely select regions of interest and compare gene expression patterns between them. Ensure **Manual Selection** is activated **[1]**. Using the mouse cursor, select the spots in the hippocampal region on the left side of the mouse brain slice **[2]**. Click on **Set 1** and then **Add to Set** to highlight the selected spots on the right panel **[3]**.

2.11.1. SCREEN: 68892_screenshot_8.mp4: 00:00-00:06

2.11.2. SCREEN: 68892_screenshot_8.mp4: 00:06-00:16

2.11.3. SCREEN: 68892_screenshot_8.mp4: 00:16-00:22

2.12. Then, click on **Set 2** and use the mouse cursor to select the spots in the hypothalamic region of the brain slice **[1].** Click on **Add to Set** to highlight these selected spots on the right side **[2]**.

2.12.1. SCREEN: 68892_screenshot_8.mp4: 00:30-00:32, 00:22-00:29

2.12.2. SCREEN: 68892_screenshot_8.mp4: 00:33-00:38

2.13. After completing the spot selection, click on the **Compare Gene Expression** button **[1]**. This generates a table displaying the average gene expression values for each selected region, along with a scatterplot representation **[2]**. Move the cursor over individual points on the scatterplot to confirm the gene names and the average expression values in both regions **[3]**.

2.13.1. SCREEN: 68892_screenshot_9.mp4: 00:00-00:05

2.13.2. SCREEN: 68892_screenshot_9.mp4: 00:14-00:18

2.13.3. SCREEN: 68892_screenshot_10.mp4

2.14. ~~Based on the comparison results, identify differentially expressed genes **[1]**.~~ Navigate back to the **Genes** tab and visualize the expression of these genes across the tissue slice **[1]**. NOTE: The VO is edited for the deleted shot

2.14.1. ~~SCREEN: Talent reviewing the scatterplot or table to identify differentially expressed genes. NOTE: Not filmed Redundant shot~~

2.14.2. SCREEN: 68892_screenshot_11.mp4

3. Identification of Differentially Expressed Genes in Metastatic Regions of Colorectal Origin in Mouse Livers

3.1. Click on the **Database** tab and use the filter to select the organism as **mouse**, the organ as **liver**, and the condition as **cancer [1]**. From the resulting sample list, select sample DSID001005 *(D-S-I-D-Zero-Zero-One-Zero-Zero-Five)* **[2]**. Click on the selected sample and confirm that the description indicates that **the sample is from a mouse liver containing metastasis of colorectal cancer origin [3]**.

3.1.1. SCREEN: 68892_screenshot_12.mp4: 00:00-00:15

3.1.2. SCREEN: 68892_screenshot_12.mp4: 00:15-00:21

3.1.3. SCREEN: 68892_screenshot_12.mp4: 00:21-00:31

3.2. Then, navigate to the **Tissue Explorer** tab and activate **Manual Selection** mode **[1]**. Using the mouse cursor, select the spots corresponding to the tumor region, identified by positive expression of the Epcam *(Ep-Cam)* marker in sample DSID001005 **[2]**. Click on **Set 1**, then select **Add to Set** to highlight the selected tumor spots on the right side **[3]**.

3.2.1. SCREEN: 68892_screenshot_13.mp4: 00:00-00:04

3.2.2. SCREEN: 68892_screenshot_13.mp4: 00:04-00:12

3.2.3. SCREEN: 68892_screenshot_13.mp4: 00:13-00:22

3.3. Now, click on **Set 2** and use the cursor to select the spots in the distant non-tumor region of the liver sample **[1]**. Click on **Add to Set** to highlight the selected non-tumor spots on the right side of the display **[2]**.

3.3.1. SCREEN: 68892_screenshot_13.mp4: 00:33-00:34, 00:28-00:32

3.3.2. SCREEN: 68892_screenshot_13.mp4: 00:34-00:38

3.4. To perform further analysis of gene expression data, click on the **Download CSV** option, generating a Comma-Separated Values file of the gene expression data for the two regions of the sample **[1]**.

3.4.1. SCREEN: 68892_screenshot_14.mp4

3.5. After repeating the database navigation steps for DSID001007, confirm that the description states it is another slice from a mouse liver containing metastases of colorectal cancer origin **[1]**.

3.5.1. SCREEN: 68892_screenshot_15.mp4:

3.6. Next, confirm that two CSV files have been generated, one each from samples DSID001005 and DSID001007, containing two columns representing average gene expression in tumor and non-tumor regions **[1]**.

3.6.1. SCREEN: 68892_screenshot_16.mp4

3.7. Load both CSV files into the R programming environment **[1]**. Merge the datasets to perform downstream analysis using two replicates per condition **[2].**

3.7.1. SCREEN: 68892_screenshot_17.mp4: 00:00-00:09

3.7.2. SCREEN: 68892_screenshot_17.mp4: 00:09-00:19

3.8. In R, use the limma package to perform differential gene expression analysis on the merged dataset. Assign the colorectal metastases regions from both samples to the cancer group, and the distant healthy regions to the control group **[1]**. Filter the results to identify upregulated genes with a log fold change greater than 0.5 and an adjusted p-value less than 0.05. Similarly, extract downregulated genes with a log fold change less than minus 0.5 and an adjusted p-value less than 0.05 **[2]**. NOTE: 3.8.1-3.8.2 and 3.8.3-3.8.4 shots and narration are merged

3.8.1. SCREEN: 68892_screenshot_18.mp4: 00:00-00:16

3.8.2. SCREEN: 68892_screenshot_18.mp4: 00:16-00:33

Results

4. Results

4.1. A distinct low-quality region was observed on the left side of the mouse brain sample, characterized by a reduced number of detected genes and a lower read count **[1].** The sample showed an average of approximately 4000 genes detected per spot, aligning well with the distribution of other samples in the database **[2].**

4.1.1. LAB MEDIA: Figure 1 *Video editor: Highlight 1A, B and C*

4.1.2. LAB MEDIA: Figure 1D**.** *Video editor: Highlight the red dashed line and the surrounding bar at the 4000-gene mark in the center of the histogram.*

4.2. Fifteen spatial clusters were identified across the mouse brain sample, with distinct boundaries representing anatomical differences **[1].**

4.2.1. LAB MEDIA: Figure 1E.

4.3. The gene *Nrgn (N-R-G-N)*, *Slc17a7 (S-L-C-Seventeen-A-Seven),* and Ddn *(D-D-N)* showed strong expression in the hippocampal region **[1].** In contrast, Ly6h *(L-Y-Six-H)* expression was localized in the cortical regions, particularly the lower-left and right outer edges of the slice **[2].**

4.3.1. LAB MEDIA: Figure 2A, B and D *Highlight the red colored, pale yellow, and greyish regions in the upper left and central areas of 2A, B and D, respectively*

4.3.2. LAB MEDIA: Figure 2C. *Video editor: Highlight the outer red-toned bands around the lower periphery of the brain slice.*

4.4. Neuropeptide signaling activity was notably increased in the lower cortical regions of the sample slice **[1].** Regulation of synaptic plasticity was activated across the hippocampal region, particularly in the upper-middle zones **[2].** Neurotransmitter transport activity was elevated across the mid and upper-right sections of the hippocampus **[3].**

4.4.1. LAB MEDIA: Figure 3A. *Video editor: Highlight the red-tinged area in the lower portion of the brain image.*

4.4.2. LAB MEDIA: Figure 3B. *Video editor: Highlight the bright red areas in the upper-middle part of the brain slice*.

4.4.3. LAB MEDIA: Figure 3C**.** *Video editor: Highlight the orange-to-red regions*

spanning the middle to upper-right areas.

4.5. The genes *Cldn7 (C-L-D-N-Seven)*, *Cldn4 (C-L-D-N-Four)*, and *Actg1 (A-C-T-G-One)* exhibited clear upregulation at the tumor region with colorectal metastasis in liver sample DSID001005 **[1].** In contrast, the expression of *Cldn7*, *Cldn4*, and *Actg1* were notably lower in the distant, healthy liver tissue of sample DSID001007 **[2].**

4.5.1. LAB MEDIA: Figure 7A, B, and C**.** *Video editor: Highlight the red-orange cluster on the left-central region of the tissue*.

4.5.2. LAB MEDIA: Figure 7D, E and F.

Words and Pronunciations

1. **DSID001557**
   - Pronunciation link: *No confirmed link found*
   - IPA: /diː-ɛs-aɪ-diː-zɪəroʊ-zɪəroʊ-wʌn-fʌɪv-fʌɪv-sɛvən/
   - Phonetic spelling: dee-ess-eye-dee-zero-zero-one-five-five-seven
2. **µL** (microliter)
   - Pronunciation link: https://www.merriam-webster.com/dictionary/microliter
   - IPA: /ˈmaɪkroʊˌliːtər/
   - Phonetic spelling: my-kroh-lee-ter
3. **mito** (short for mitochondrial, or as used in “Mito”)
   - Pronunciation link: https://www.merriam-webster.com/dictionary/mitochondrial (for root)
   - IPA: /mɪˈtoʊ/ or /ˈmɪtoʊ/ (as a short form)
   - Phonetic spelling: mi-toh
4. **Epcam**
   - Pronunciation link: *No confirmed link found*
   - IPA: /ˈɛpˌkæm/
   - Phonetic spelling: ep-kam
5. **limma**
   - Pronunciation link: https://www.merriam-webster.com/dictionary/lemma

(similar root)

- IPA: /ˈlɪmə/
- Phonetic spelling: lim-ah

6. **log fold change**
   - Pronunciation link: *No confirmed link found*
   - IPA: /lɔg foʊld tʃeɪndʒ/
   - Phonetic spelling: log fold chaynge
7. **adjusted p-value**
   - Pronunciation link: *No confirmed link found*
   - IPA: /əˈdʒʌstɪd piː ˈvæljuː/
   - Phonetic spelling: uh-just-id pee val-yoo
8. **Cldn7**, **Cldn4**, **Actg1**
   - These are gene symbols, typically spelled out letter-by-letter or as “C-L-D-N seven,” etc.
   - Phonetic spelling: see below:
     - C-L-D-N-7: see-el-dee-en seven
     - C-L-D-N-4: see-el-dee-en four
     - A-C-T-G-1: ay-cee-tee-gee one