Submission ID #: 68707
Scriptwriter Name: Poornima G
Project Page Link: https://review.jove.com/account/file-uploader?src=20955683

Title: Navigating the Mass Spectrometry-Based Proteomic Data Using Free Computational Tools

Authors and Affiliations:
Shijie He[1], Fangfei Zhang[2]

[1]School of Life Sciences, Westlake University
[2]Molemuse Biotech Studio

Corresponding Authors:
Fangfei Zhang pushomix@molemuse.ac.cn

Email Addresses for All Authors:
Shijie He heshijie@westlake.edu.cn
Fangfei Zhang pushomix@molemuse.ac.cn

Author Questionnaire1. We have marked your project as author-provided footage, meaning you film the video yourself and provide JoVE with the footage to edit. JoVE will not send the videographer. Please confirm that this is correct.

√ Correct

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. . **Proposed filming date:** To help JoVE process and publish your video in a timely manner, please indicate the proposed date that your group will film the **interviews** here: **MM/DD/YYYY**

When you are ready to submit your video files, please contact our China Location Producer, Yuan Yue.

Current Protocol Length
Number of Steps: 17
Number of Shots: 36 (35 SC)

Introduction

1.1. **Shijie He:** Our work provides a guide for biologists to analyze complex proteomic data using free tools. We aim to empower them to validate findings and explore public datasets.

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* 68707_intro_1.1.mp4

What research gap are you addressing with your protocol?

1.2. **Shijie He:** Many biologists can't use public proteomic data due to a lack of clear guides. Our protocol bridges this gap, enabling validation without new wet-lab experiments.

1.2.1. INTERVIEW: Named talent says the statement above in an interview-style shot, looking slightly off-camera. *Suggested B-roll: 4.2.1* 68707_intro_1.2.mp4

What advantage does your protocol offer compared to other techniques?

1.3. **Shijie He:** Our protocol uses free, state-of-the-art software that is vendor-agnostic. These tools are faster, more sensitive, and provide more precise proteome discovery than many traditional approaches.

1.3.1. INTERVIEW: Named talent says the statement above in an interview-style shot, looking slightly off-camera. *Suggested B-roll: 4.3.1* 68707_intro_1.3.mp4

Protocol

2. Data-Independent Acquisition (DIA) Data Analysis Using DIA-NN

Demonstrator: Shijie He

2.1. To begin, download the reference proteome FASTA *(fasta)* file from the UniProt database **[1]**.

2.1.1. Lab media: 68707_Screenshot_2.1.

2.2. Click on the **Add FASTA** button to load the reference proteome file into the DIA-NN software **[1]**.

2.2.1. SCREEN: 68707_screenshot_2.2.mp4 00:00–00:20.

2.3. Select the two options **FASTA digest for library-free search library generation** and **Deep learning-based spectra, RTs and IMs prediction** under the Precursor ion generation section **[1]**. Then, click on the **Run** button to generate a predicted spectral library **[2]**.

2.3.1. SCREEN: 68707_screenshot_2.3.mp4 00:00–00:08.

2.3.2. SCREEN: 68707_screenshot_2.3.mp4 00:09–00:24.

2.4. Unselect the two options under the **Precursor ion generation** section **[1]**. Click on the **Type** button that corresponds to the file format under the Input section to load the DIA data **[2]**.

2.4.1. SCREEN: 68707_screenshot_2.4-(1).mp4 00:00–00:07.

2.4.2. SCREEN: 68707_screenshot_2.4-(1).mp4 00:20–00:38.

2.5. Now, set both Mass Accuracy and MS1 Accuracy to 0 parts per million under the Algorithm section **[1]**.

2.5.1. SCREEN: 68707_screenshot_2.5.mp4 00:00–00:16.

2.6. Adjust the precursor and fragment mass range settings under the Precursor ion generation section according to the experimental setup **[1]**.

2.6.1. SCREEN: 68707_screenshot_2.6.mp4 00:05–00:25.

2.7. Keep the other software settings unchanged **[1]**.

2.7.1. SCREEN: 68707_screenshot_2.7.mp4 00:00–00:10.

2.8. Next, click on the **Run** button **[1]**. Wait until **Finished** is displayed on the operation interface, indicating the analysis is complete **[2]**.

2.8.1. SCREEN: 68707_screenshot_2.8.mp4 00:00–00:07.

2.8.2. SCREEN: 68707_screenshot_2.8.mp4 00:08–00:17.

3. Data-Dependent Acquisition (DDA) Data Analysis Using FragPipe

3.1. Click on the **FragPipe** *(frag-pipe)* icon located in bin folder after installation **[1]**. Navigate to the **Config** tab to view all dependent settings **[2]**. Check whether MS-Fragger, Ion-Quant, dia-Tracer, DIA-NN, and Python modules are available on your system **[3]**. If any modules are missing, click on **Download/Update** *(download update)* or **Download** to retrieve them **[4]**.

3.1.1. SCREEN: 68707_screenshot_3.1.mp4 00:00–00:15.

3.1.2. SCREEN: 68707_screenshot_3.1.mp4 00:22-00:40.

3.1.3. SCREEN: 68707_screenshot_3.1.mp4 00:47-00:50.

3.1.4. SCREEN: 68707_screenshot_3.1.mp4 00:50-01:00.

3.2. Now, switch to the **Workflow** tab **[1]**. Select **Default** from the workflow dropdown and click on **Load workflow [2]**. Then, click on **Add files** to input the file paths **[3]**. Assign experiment name and biological replicate number under Assign files, or leave it blank **[4]**.

3.2.1. SCREEN: 68707_screenshot_3.2.mp4 00:00–00:05.

3.2.2. SCREEN: 68707_screenshot_3.2.mp4 00:06–00:15.

3.2.3. SCREEN: 68707_screenshot_3.2.mp4 00:25–00:40.

3.2.4. SCREEN: 68707_screenshot_3.2.mp4 00:55–01:05.

3.3. Next, click on the **Database** tab to switch to it **[1]**. Load a FASTA file from disk or download one that corresponds to the sample species **[2]**. During download, select the

options **Reviewed sequences only**, **Add decoys**, and **Add common contaminants** for a simple run **[3]**.

3.3.1. SCREEN: 68707_screenshot_3.3.mp4 00:00–00:05.

3.3.2. SCREEN: 68707_screenshot_3.3.mp4 00:15–00:23.

3.3.3. SCREEN: 68707_screenshot_3.3.mp4 00:24–00:36.

3.4. Click on the **MSFragger** *(M-S-fragger)* tab to change the view **[1]**, Select **Closed Search default config** and click on **Load [2]**.

3.4.1. SCREEN: 68707_screenshot_3.4.mp4 00:03–00:07.

3.4.2. SCREEN: 68707_screenshot_3.4.mp4 00:08–00:30.

3.5. Under Peak Matching settings, retain all default values **[1]**. For both Calibration and Optimization, select **None** to reduce processing time **[2]**.

3.5.1. SCREEN: 68707_screenshot_3.5.mp4 00:00–00:07.

3.5.2. SCREEN: 68707_screenshot_3.5.mp4 00:08–00:30.

3.6. For protein digestion, adjust the parameters based on your experiment's requirements **[1]** and maintain the remaining default settings **[2]**.

3.6.1. SCREEN: 68707_screenshot_3.6.mp4 00:18–00:32.

3.6.2. SCREEN: 68707_screenshot_3.6.mp4 00:33–00:36.

3.7. Now, switch to the **Validation** tab **[1]**. Uncheck **Predict RT** and **Predict spectra**, as these options are intended for data-independent acquisition workflows **[2]**.

3.7.1. SCREEN: 68707_screenshot_3.7.mp4 00:00–00:05.

3.7.2. SCREEN: 68707_screenshot_3.7.mp4 00:06–00:15.

3.8. Click on the **Quant (MS1)** *(quant-MS-1)* tab **[1]**. Select **Run MS1 quant** and then click on **Load Quant defaults [2]**. Choose **IonQuant** and leave all other settings at default values **[3]**.

3.8.1. SCREEN: 68707_screenshot_3.8.mp4 00:00–00:05.

3.8.2. SCREEN: 68707_screenshot_3.8.mp4 00:06–00:09.

3.8.3. SCREEN: 68707_screenshot_3.8.mp4 00:10–00:20.

3.9. Finally, click on the **Run** tab to proceed **[1]**. Select the desired output directory **[2]** and click on **RUN** to begin analyzing the data **[3]**.

3.9.1. SCREEN: 68707_screenshot_3.9.mp4 00:00-00:05.

3.9.2. SCREEN: 68707_screenshot_3.9.mp4 00:06-00:10.

3.9.3. SCREEN: 68707_screenshot_3.9.mp4 00:11-00:27.

Results

4. Results

4.1. In patients with pancreatic ductal adenocarcinoma or PDAC *(P-D-A-C)*, SERPINA5 *(serpin-A-5)* and HPSE *(H-P-S-E)* showed significantly reduced expression **[1]**, while FGB displayed increased expression in serum compared to normal individuals **[2]**.

4.1.1. LAB MEDIA: Figure 7A. *Video editor: Highlight the DATA labels "SERPINA5" and "HPSE" on the left side of the volcano plot*.

4.1.2. LAB MEDIA: Figure 7A. *Video editor: Highlight the data label "FGB" on the right side of the volcano plot*.

4.2. In hepatocellular carcinoma tumor samples, ENO3 *(eno-3)*, PLS3 *(P-L-S-3)*, MTAP *(M-tap)*, SERPINB9 *(serpin-B-9)*, and ITPR2 *(I-T-P-R-2)* showed reduced expression relative to paired tissues **[1]**, whereas ME1 *(M-E-1)*, CYP27A1 *(C-Y-P-27-A-1)*, RPS16 *(R-P-S-16)*, and ATP5PF *(A-T-P-5-P-F)* were significantly increased **[2]**.

4.2.1. LAB MEDIA: Figure 7B. *Video editor: Highlight the data labels "ENO3", "PLS3", "MTAP", "SERPINB9", and "ITPR2" on the left side of the volcano plot*.

4.2.2. LAB MEDIA: Figure 7B. *Video editor: Highlight the data labels "ME1", "CYP27A1", "RPS16", and "ATP5PF" on the right side of the volcano plot.*

4.3. Heatmap visualization revealed consistently elevated protein expression in the serum of PDAC patients compared to normal individuals **[1]**.

4.3.1. LAB MEDIA: Figure 8A. *Video editor: Highlight the right three columns labeled "PDAC_Serum_8696," "PDAC_Serum_8568," and "PDAC_Serum_8526" .*

4.4. Gene Ontology enrichment analysis of PDAC serum revealed significant upregulation of processes related to coagulation and hemostasis **[1]**.

4.4.1. LAB MEDIA: Figure 8C. *Video editor: Highlight the "regulation of blood coagulation," " regulation of hemostasis," and " regulation of coagulation" on the y-axis.*

4.5. GO analysis of hepatocellular carcinoma tumors identified enrichment in nucleotide and metabolic processes, including purine nucleotide metabolism and NAD metabolic pathways **[1]**.

4.5.1. LAB MEDIA: Figure 8D. *Video editor: Highlight the "purine nucleotide metabolic process," "nucleotide metabolic process," and "NAD metabolic process*."

4.6. KEGG *(Keg)* pathway enrichment analysis of PDAC serum revealed significant activation of the complement and coagulation cascades pathway, along with glycosaminoglycan degradation **[1]**.

4.6.1. LAB MEDIA: Figure 8E. *Video editor: Highlight the "Complement and coagulation cascades" and "Glycosaminoglycan degradation."*

4.7. KEGG analysis in hepatocellular carcinoma tumor samples identified enrichment in PPAR *(P-P-A-R)* signaling, carbon metabolism, and neurodegenerative disease pathways, though with lower statistical significance **[1]**.

4.7.1. LAB MEDIA: Figure 8F. *Video editor: Highlight the lower right data points labeled "PPAR signaling pathway," "Carbon metabolism," "Parkinson disease", "Huntington Disease" "prion disease"" "Alzheimer disease*."

4.8. The protein-protein interaction network for upregulated PDAC serum proteins revealed a central cluster involving coagulation factor XI *(eleven)*, fibrinogen beta chain, and plasma serine protease inhibitor **[1]**, as well as several isolated proteins including HPSE, CD5 antigen-like, and CRISP3 **[2]**.

4.8.1. LAB MEDIA: Figure 8G. *Video editor: Emphasize the central triangle formed by nodes "P03951," "P02675," and "P05154."*

4.8.2. LAB MEDIA: Figure 8G. *Video editor: Highlight the circles "Q9Y251," "O43866," and "P54108" placed away from the central cluster*.

Pronunciation guide

1. **Proteome**
   Pronunciation link:
   https://www.merriam-webster.com/dictionary/proteome
   IPA: /ˈprōtiˌōm/
   Phonetic Spelling: proh-tee-ohm
2. **FASTA**
   Pronunciation link:
   https://www.merriam-webster.com/dictionary/FASTA

IPA: /ˈfæstə/
Phonetic Spelling: fas-tuh

3. **DIA-NN**
No confirmed link found
IPA: /ˌdiː.aɪˈɛnˈɛn/
Phonetic Spelling: dee-eye-en-en
4. **Precursor**
Pronunciation link:
https://www.merriam-webster.com/dictionary/precursor
IPA: /ˈpriˌkɜrsər/
Phonetic Spelling: pree-kur-ser
5. **Spectra**
Pronunciation link:
https://www.merriam-webster.com/dictionary/spectra
IPA: /ˈspɛktrə/
Phonetic Spelling: spek-truh
6. **Hepatocellular**
Pronunciation link:
https://www.merriam-webster.com/dictionary/hepatocellular
IPA: /ˌhɛpətoʊˈsɛljələr/
Phonetic Spelling: hep-uh-toh-sel-yuh-ler
7. **Carcinoma**
Pronunciation link:
https://www.merriam-webster.com/dictionary/carcinoma
IPA: /ˌkɑrsəˈnoʊmə/
Phonetic Spelling: kar-suh-noh-muh
8. **MSFragger**
No confirmed link found
IPA: /ˌɛmˌɛsˈfrægər/
Phonetic Spelling: em-ess-frag-er
9. **IonQuant**
No confirmed link found
IPA: /ˈaɪənˌkwɑnt/
Phonetic Spelling: eye-on-kwahnt
10. **Decoy**
Pronunciation link:
https://www.merriam-webster.com/dictionary/decoy
IPA: /ˈdiˌkɔɪ/
Phonetic Spelling: dee-koy
11. **SERPINA5**
No confirmed link found
IPA: /ˈsɜrpɪnˌeɪˌfaɪv/
Phonetic Spelling: ser-pin-ay-five
12. **HPSE**
No confirmed link found

IPA: /ˌeɪtʃˌpiːˌɛsˌi/
Phonetic Spelling: H-P-S-E

13. **FGB**
No confirmed link found
IPA: /ˌɛfˌʤiːˌbiː/
Phonetic Spelling: F-G-B
14. **PLS3**
No confirmed link found
IPA: /ˌpiːˌɛlˌɛsˈθriː/
Phonetic Spelling: P-L-S-three
15. **MTAP**
No confirmed link found
IPA: /ˈɛmˌtæp/
Phonetic Spelling: M-tap
16. **SERPINB9**
No confirmed link found
IPA: /ˈsɜrpɪnˌbiːˌnaɪn/
Phonetic Spelling: ser-pin-bee-nine
17. **ITPR2**
No confirmed link found
IPA: /ˌaɪˌtiːˌpiːˌɑrˈtuː/
Phonetic Spelling: I-T-P-R-two
18. **CYP27A1**
No confirmed link found
IPA: /ˌsiːˌwaɪˌpiːˌtwɛntiˈsɛvənˌeɪˈwʌn/
Phonetic Spelling: C-Y-P-twenty-seven-A-one
19. **ATP5PF**
No confirmed link found
IPA: /ˌeɪˌtiːˌpiːˌfaɪvˌpiːˌɛf/
Phonetic Spelling: A-T-P-five-P-F
20. **Ontology**
Pronunciation link:
https://www.merriam-webster.com/dictionary/ontology
IPA: /ɑnˈtɑlədʒi/
Phonetic Spelling: on-tol-uh-jee
21. **Hemostasis**
Pronunciation link:
https://www.merriam-webster.com/dictionary/hemostasis
IPA: /ˌhiːmoʊˈsteɪsɪs/
Phonetic Spelling: hee-moh-stay-sis
22. **Nucleotide**
Pronunciation link:
https://www.merriam-webster.com/dictionary/nucleotide
IPA: /ˈnuːkliəˌtaɪd/
Phonetic Spelling: noo-klee-uh-tide

23. **Glycosaminoglycan**
    Pronunciation link:
    https://www.merriam-webster.com/dictionary/glycosaminoglycan
    IPA: /ˌglaɪkoʊsəˌmɪnoʊˈglaɪˌkæn/
    Phonetic Spelling: gly-koh-suh-min-oh-gly-kan
24. **KEGG**
    No confirmed link found
    IPA: /kɛg/
    Phonetic Spelling: keg
25. **PPAR**
    No confirmed link found
    IPA: /ˌpiːˌpiːˌeɪˈɑr/
    Phonetic Spelling: P-P-A-R