プロテオームデータベースjPOSTの挑戦

京大院薬 石濱 泰

&

jPOST TEAM
What is jPOST?

Supported by NBDC-JST since 2015

https://jpostdb.org

For Data Integration & Sharing in Life Science

Overview of jPOST: Japan Proteome Standard Repository/Database

Supported by NBDC-JST since 2015
MS raw data with metadata are stored in ProteomeXchange formats.
ProteomeXchange Consortium

repository in Asia-Oceania
July 2019

265 users from 28 countries
Registered Project Number

July 2019

Registered 476 projects
522 projects are registered. 323 are opened.

47615 files amount to 15.0 TB.

92 species.

Oct 2019

- Homo sapiens (Human)
- Mus musculus (Mouse)
- Escherichia coli
- Arabidopsis thaliana (Mouse-ear cress)
- Rattus rattus (Black rat)
- Other
jPOST Re-Analysis
Target-Decoy Approach for estimating FDR(%) 

False Discovery Rate = False Positive/(False Positive + True Positive)
Multiple search engines with FDR control

PSM at FDR=1% by different search engines

- HeLa global proteome dataset (JPST000200)

Target
72,279 PSM in total

Decoy
1,739 PSM in total

MaxQuant (49,084 PSM)
Comet (62,917 PSM)
Mascot (61,251 PSM)
X!Tandem (66,154 PSM)

FDR 2.41%
Universal index for annotated MSMS spectra

How can we merge the results from different sources?

UniScore

- based on peak annotation in MSMS
- search engine independent
- MS instrument independent
- search DB independent
- can be used as universal threshold for peptide identification
Relationship between different scores

Dataset: JPST000200 (5 files)
Tryptic peptides from HeLa cells, by Thermo Q-Exactive with 8h gradient

- Mascot
  \[ y = 3.0792x - 22.96 \]
  \[ R^2 = 0.7115 \]

- MaxQuant
  \[ y = 4.7915x - 0.7984 \]
  \[ R^2 = 0.6613 \]

- Comet
  \[ y = 0.0908x + 0.3914 \]
  \[ R^2 = 0.9208 \]

- X!Tandem
  \[ y = 1.5248x + 2.904 \]
  \[ R^2 = 0.8996 \]
Optimized UniScore to rank down the decoy hits

Dataset: JPST000200 (5 files)
Tryptic peptides from HeLa cells, by Thermo Q-Exactive with 8h gradient

4 engines merged results
Sorted by UniScore

- Mascot results
- X! tandem results
- MaxQuant results
- Comet results

# PSM
Optimized UniScore to rank down the decoy hits

Dataset: PXD001792
Phosphopeptides from mouse Hepa cells by Thermo Q-Exactive

Mascot results sorted by UniScore
X! tandem results sorted by UniScore
MaxQuant results sorted by UniScore
Comet results sorted by UniScore

4 engines merged results
Sorted by UniScore

FDR%

# PSM

# PSM
Re-analysis of large-scale HPP datasets

Mass-spectrometry-based draft of the human proteome

11204

- Shiromizu_JPR_2013
- Geiger_MCP_2012
- Nagaraj_MSB_2011
- Munoz_MSB_2011
- Mertins_NatMethods_2013
- Geiger_CancerRes_2012

- Blood platelet
- Ureter
- Urate
- Ovarian cysts
- Uterus
- Blood
- Basement membranes and muscle proteins
- Cervical mucus
- Ovary
- Prostate
- Blood platelet
- Ovarian cysts
- Uterus
- Blood
- Basement membranes and muscle proteins
- Cervical mucus
- Ovary
- Prostate
- Blood platelet
- Ureter
- Urate
- Ovarian cysts
CRC_iTRAQ (HCD)
- MaxQuant: 56,787
- UniScore: 72,733
- 55,204 peptides
- 1,583 peptides

mHCT116_iTRAQ (HCD)
- MaxQuant: 60,394
- UniScore: 76,854
- 58,809 peptides
- 1,585 peptides

CRC_phospho_iTRAQ (HCD)
- MaxQuant: 9,028
- UniScore: 11,773
- 8,662 peptides
- 366 peptides

mHCT116_phospho (HCD)
- MaxQuant: 15,524
- UniScore: 18,496
- 14,932 peptides
- 592 peptides

The diagram shows the overlap of peptide counts and scores between the different datasets.
The jPOST Environment

The jPOST environment: an integrated proteomics data repository and database

Moriya et al., *Nucleic Acids Res*, 2019 Jan 8;47(D1):D1218-D1224
jPOST では Slice (データセットグループ) 間で変動解析するための簡易的なツールを提供しており、有意な変動のあったタンパク質（遺伝子）リストを用いてシームレスに ChIP-Atlas でエンリッチメント解析を行うことが可能になりました。これまで提供していた KEGG, GO の他に ChIP-Atlas が追加されました。
ChIP-Atlas 連携

変動遺伝子リスト

ChIP-Atlasサイト

変動遺伝子の発現を制御している転写因子などのリスト
Crop Proteome Datasets in jPOST

### Search Result

<table>
<thead>
<tr>
<th>jPOST ID</th>
<th>PXID</th>
<th>Project title</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>JPST000358</td>
<td>PXD000000</td>
<td>Shotgun Phosphoproteomic analysis unravelled me ...</td>
<td></td>
</tr>
<tr>
<td>JPST000156</td>
<td>PXD004603</td>
<td>Rice (Oryza sativa) proteome</td>
<td></td>
</tr>
<tr>
<td>JPST000585</td>
<td>PXD013575</td>
<td>SK07_CatharanthusRoseus_Phosphoproteomics</td>
<td></td>
</tr>
<tr>
<td>JPST000374</td>
<td>PXD008603</td>
<td>TK03_Proteome_Soybean</td>
<td></td>
</tr>
</tbody>
</table>

### Free word

- **rice**

### Project type

- All
- Mass spectrometry
- Gel electrophoresis
- Antibody

### Search

- **Search**
- **Reset**
Crop proteome datasets in jPOST

Re-analysis

- jPOST と他の PX repositoryからのデータ抽出
- UniScoreによる再解析
- 再解析IDによるデータ登録
- データインテグレーション

Database

- プロテオーム発現データベース化と比較定量
- 各種ツールによるデータ可視化、ネットワーク解析
- Missing protein の同定
統計情報

- UniProt（ヒトはneXtProt）におけるタンパク質の同定レベル

Protein Evidence

![Protein Evidence Diagram]

Inferred from Homology

<table>
<thead>
<tr>
<th>Protein Name</th>
<th>ID</th>
<th>Accession</th>
<th>#Peptide</th>
<th>#PSM</th>
<th>Unique Peptide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arylacetamide deacetylase-like 4</td>
<td>ADCL4_HUMAN</td>
<td>Q5VUY2</td>
<td>1</td>
<td>2</td>
<td>yes</td>
</tr>
<tr>
<td>COBW domain-containing protein 6</td>
<td>CBWD6_HUMAN</td>
<td>Q4V339</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Cytochrome b-c1 complex subunit 6-like, mitochondrial</td>
<td>QCR6L_HUMAN</td>
<td>A0A095LP55</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Eukaryotic translation initiation factor 3 subunit C-like protein</td>
<td>EIFCL_HUMAN</td>
<td>B5ME19</td>
<td>6</td>
<td>141</td>
<td></td>
</tr>
<tr>
<td>Golgin subfamily A member 6-like protein 6</td>
<td>GG6L6_HUMAN</td>
<td>A8MZA4</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>HIG1 domain family member 1C</td>
<td>HIG1C_HUMAN</td>
<td>A8MV81</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Homeobox protein unc-4 homolog</td>
<td>UNC4_HUMAN</td>
<td>A6NJ70</td>
<td>2</td>
<td>2</td>
<td>yes</td>
</tr>
<tr>
<td>Immunoglobulin kappa variable 1-12</td>
<td>KV112_HUMAN</td>
<td>A0A04DH73</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Immunoglobulin kappa variable 3D-15</td>
<td>KVD15_HUMAN</td>
<td>A0A087WSY6</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Olfactory receptor 5AC2</td>
<td>O5AC2_HUMAN</td>
<td>Q9NZP5</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
キュレーションの深化と強化
〜Journal of Proteome Data and Methods (JPDM) 創刊〜

Journal of Proteome Data and Methods

About JPDM

Proteomics is a rapidly growing research area that produces large amounts of data, which has led to complex challenges around data management. Proteomics data and datasets need to become more interoperable, discoverable and trusted, and therefore better used, re-used and more valuable. To address these challenges and facilitate better proteome data management, the field needs to develop formal structures and procedures. To this end, the Japanese Proteomics Society (JProS) is proposing to launch a new journal – the Journal of Proteome Data and Methods.

Instructions to Authors

Guide to Reviewers

Ethics Policies

FAQ

投稿システム（ScholarOne）
https://mc.manuscriptcentral.com/jpdm

• 各種ドキュメントの整備
• ウェブサイト構築
• 投稿システム構築
• JST J-stage システムから提供予定
From Bench to Internet: Sharing Proteomics Data and Methods through the Open Access Journal

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0).

Proteomics is a rapidly growing research area relating to all of bioscience fields. Advances in proteomics technologies including mass spectrometry (MS) coupled with bioinformatics had opened the way to produce large amounts of data from the entire protein complement of a cell, tissue, or organism etc. under the several set of conditions. Now, these advances have given us complex challenges around data management. Proteomics data and datasets coupled with detailed metadata need to become more Findable, Accessible, Interoperable, and Re-usable to promote data sharing in the scientific community. To address these challenges and facilitate better proteome data management, the field needs to develop formal structures and procedures to be an applicable standard. To this end, the Japanese Proteomics Society (JPrOS) decided to launch a new data journal – the Journal of Proteome Data and Methods (JPDM).

JPDM is a peer-reviewed, fully open access and online only journal, provided via the J-stage platform. JPDM publishes four article types: “Data descriptors”, “Protocols”, “Data processing notes”, and “Reviews”. The main content of the journal is “Data descriptors,” in which pure descriptions of data, i.e., detailed metadata on the samples to be analyzed, the sample preparation methods, MS measurement and the data analysis are provided, not the functional data for attempting to address hypotheses or for providing novel scientific interpretation and insight. We ask authors to fill in a dataset summary table to provide the basic information on their data and to show why the data described in the JPDM article is of value to the greater scientific community. These requirements make datasets searchable with easy way and consequently more useful to others. JPDM tightly collaborates with Japan Proteome Standard Repository/Database (jPOST), one of the partner repositories in ProteomeXchange consortium (PX). When data producers deposit their datasets in jPOST repository then post detailed metadata in JPDM, the journal sends a feedback to the PX repository. JPDM contributes to add a value for proteome datasets by providing detailed metadata. Since JPDM is a data journal that accepts only metadata, the biological outputs can still be posted to ordinary journals.

In addition to Data descriptors, articles in other three categories (Protocols, Data processing notes, and Reviews) highlight and share the experimental protocols and the computational tools that may have never left the lab, or that normally goes unnoticed in supplementary files.

Now, JPDM can accelerate the visibility of all kinds of data, protocols and tools in the field of proteomics.

September 2019
Yasushi Ishihama
Editor-in-chief
Journal of Proteome Data and Methods
jPOSTメンバー2019