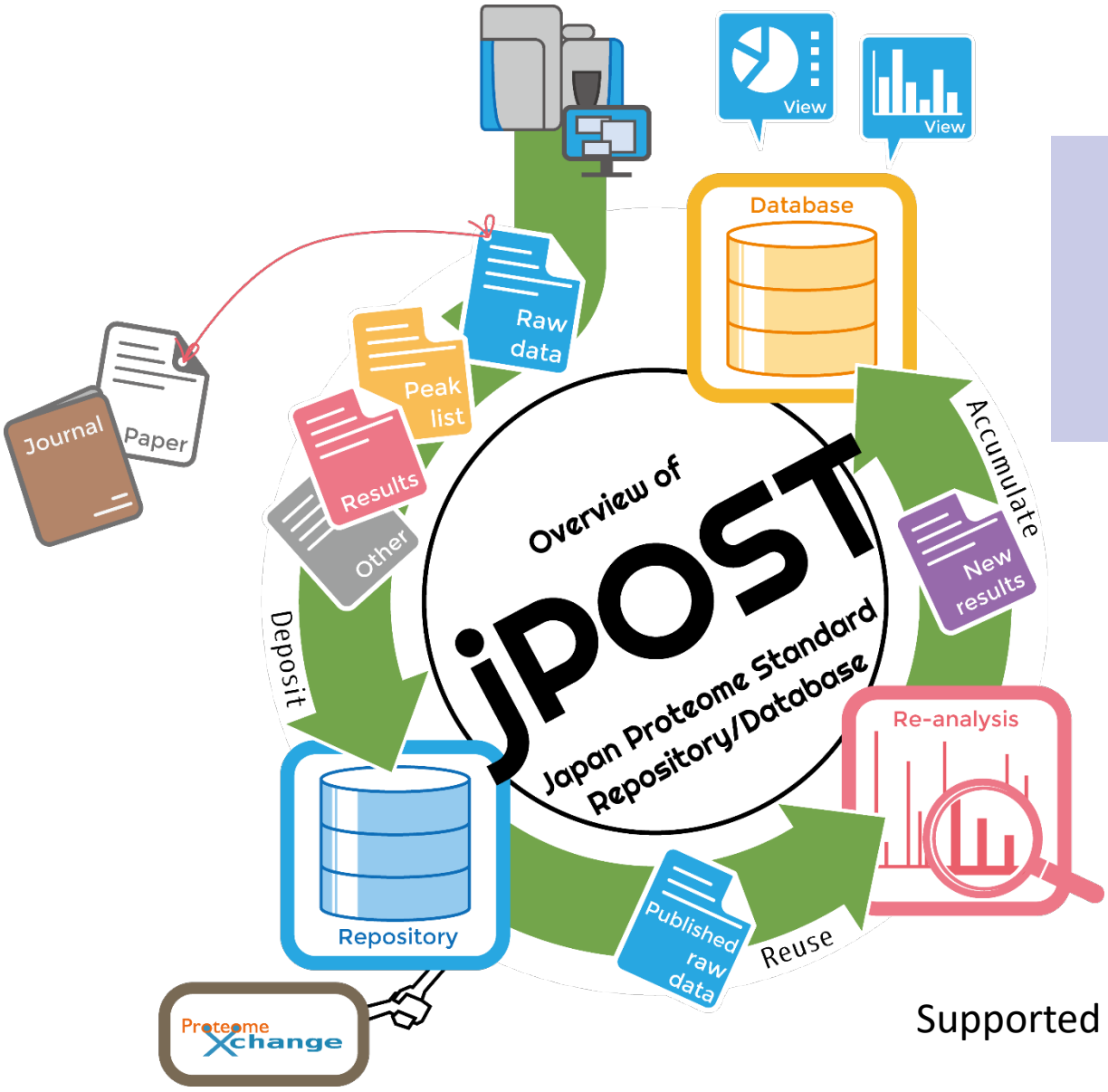
プロテオームデータベースjPOSTの挑戦

京大院薬　石濱 泰
&
jPOST TEAM

Licensed under CC-BY 4.0 ©2019 Yasushi Ishihama (Kyoto Univ)

What is jPOST?

https://jpostdb.org

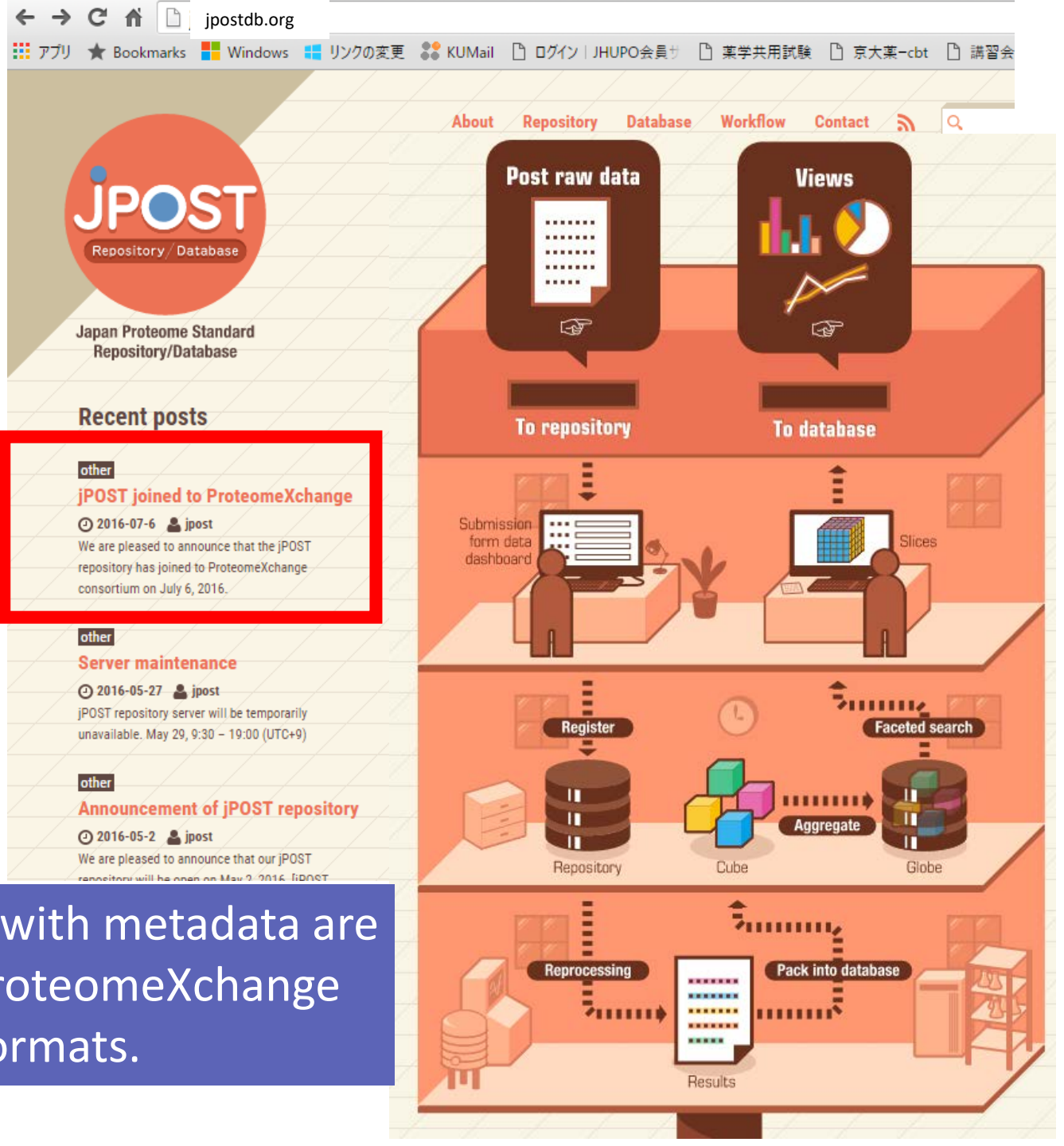


For Data Integration

& Sharing

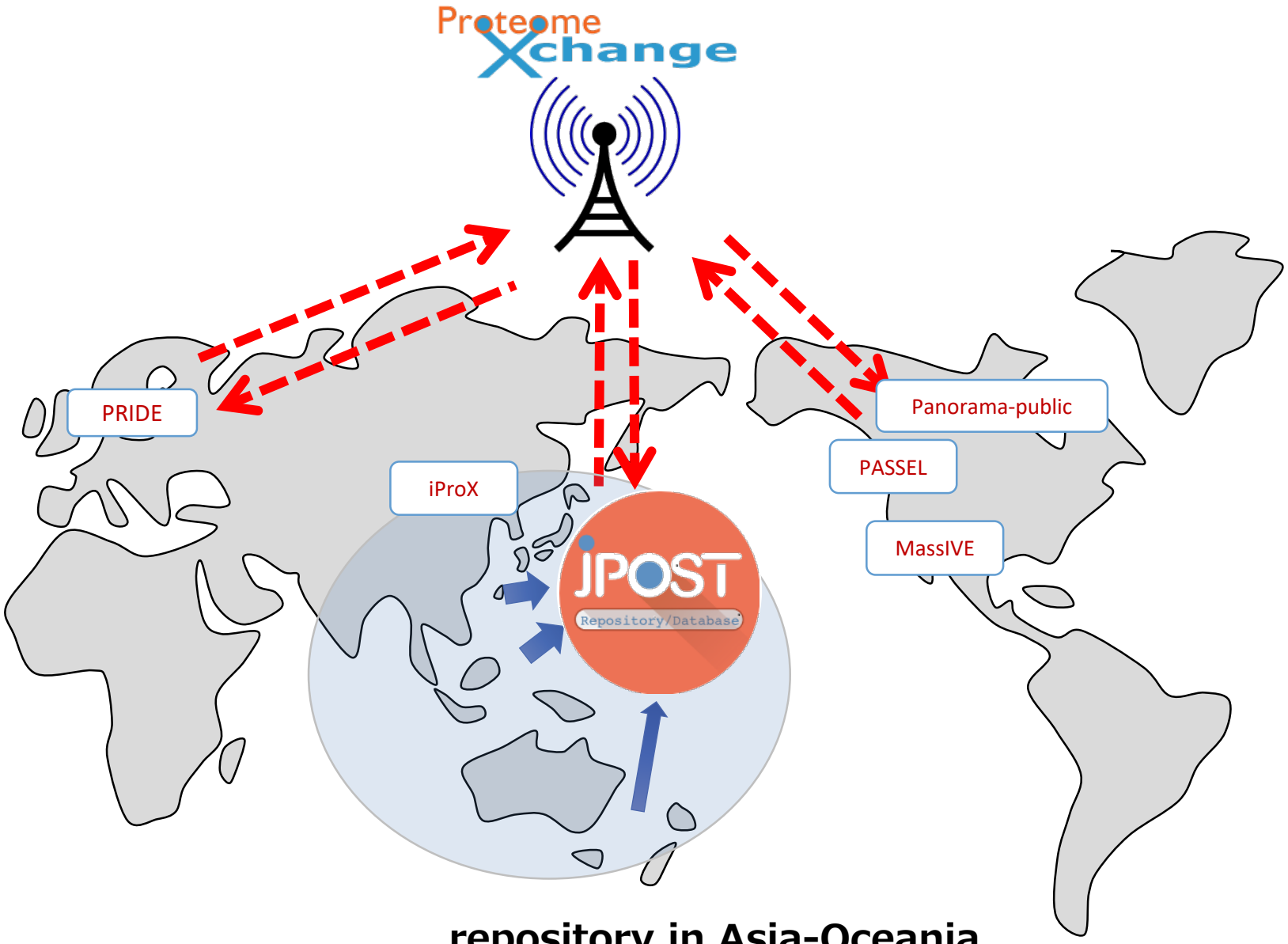
in Life Science

Supported by NBDC-JST since 2015



MS raw data with metadata are stored in ProteomeXchange formats.

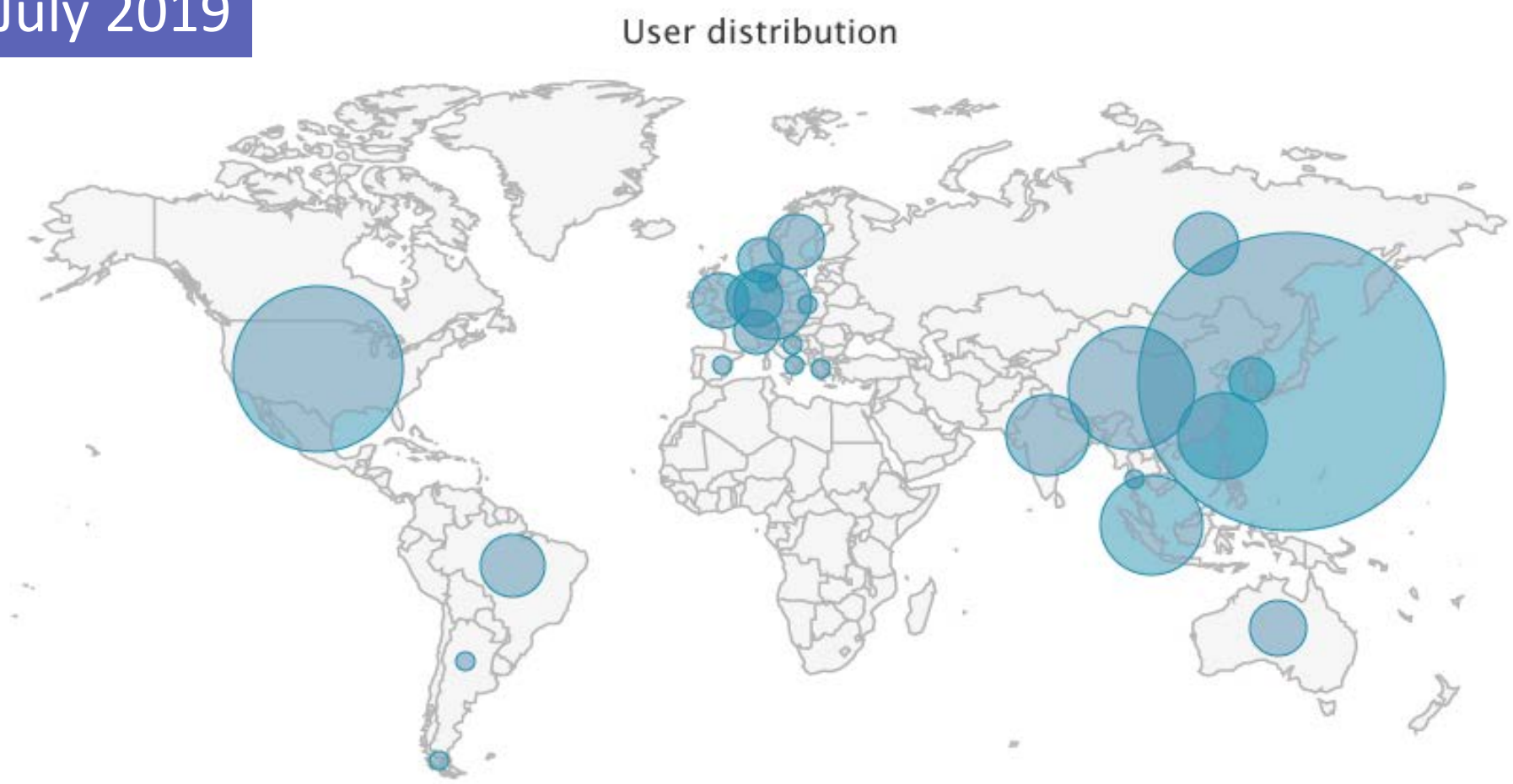
ProteomeXchange Consortium



repository in Asia-Oceania

jPOST User Distribution

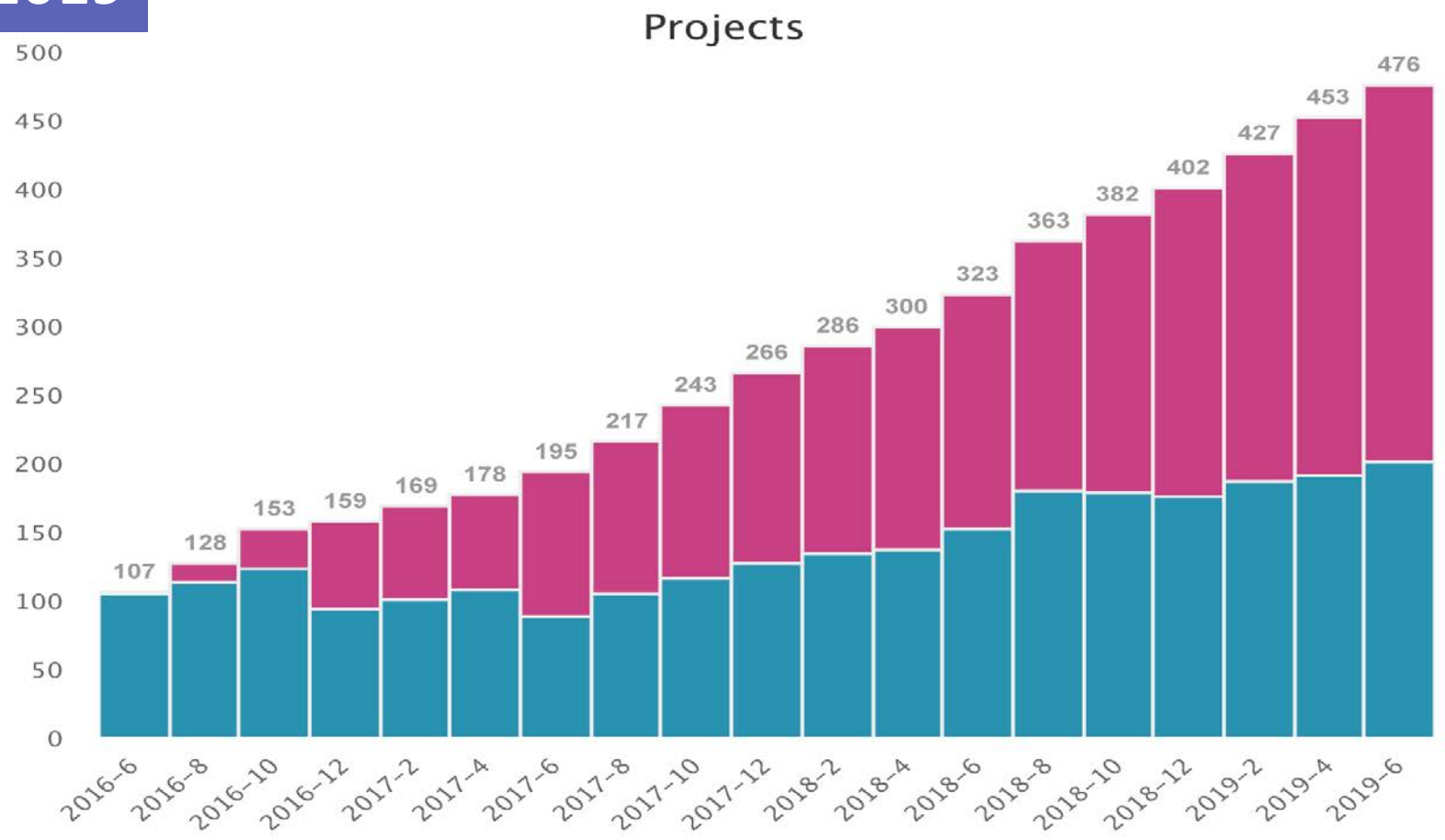
July 2019



265 users from 28 countries

Registered Project Number

July 2019



Registered 476 projects

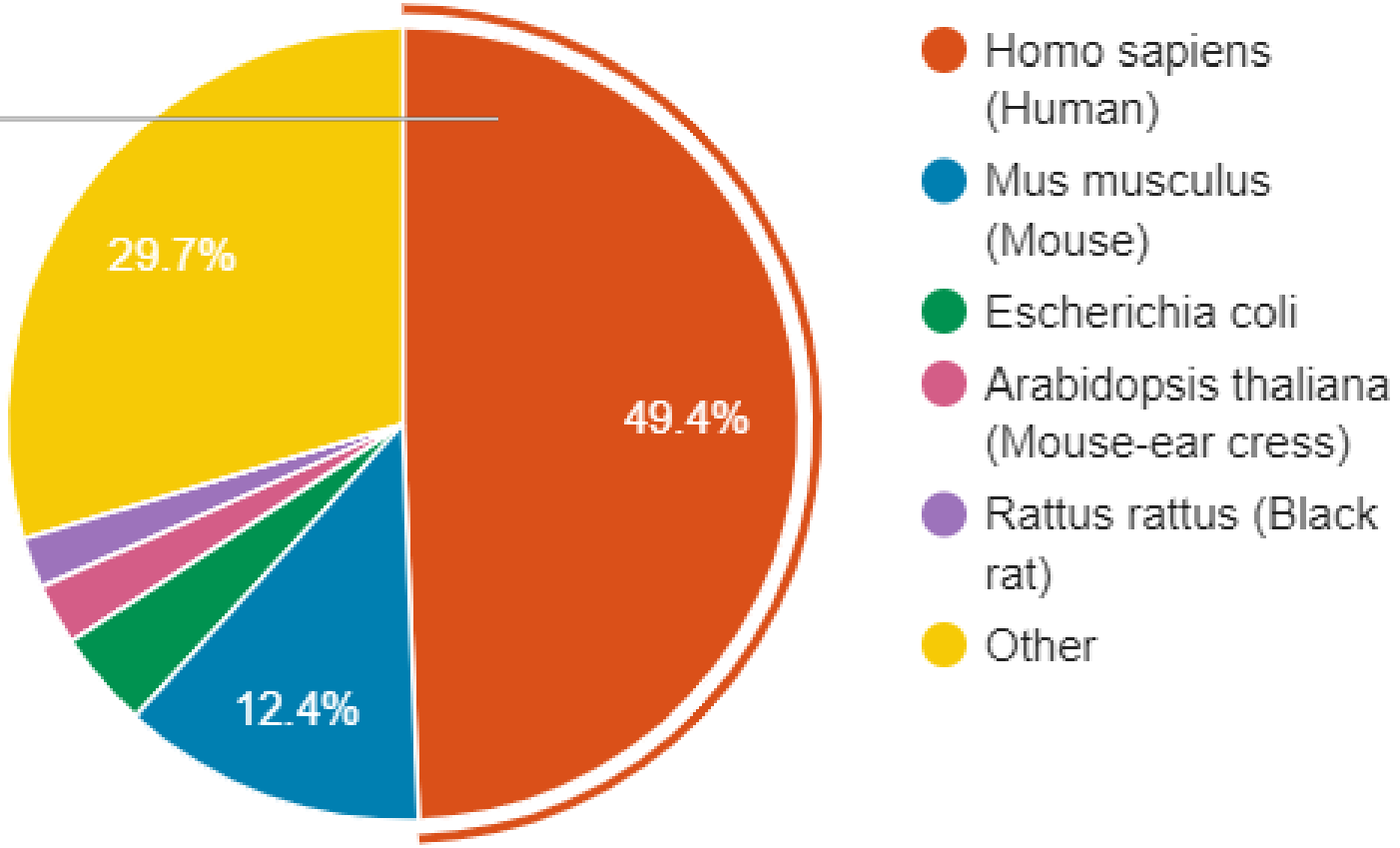
Statistics

522 projects are registered. **323** are opened.

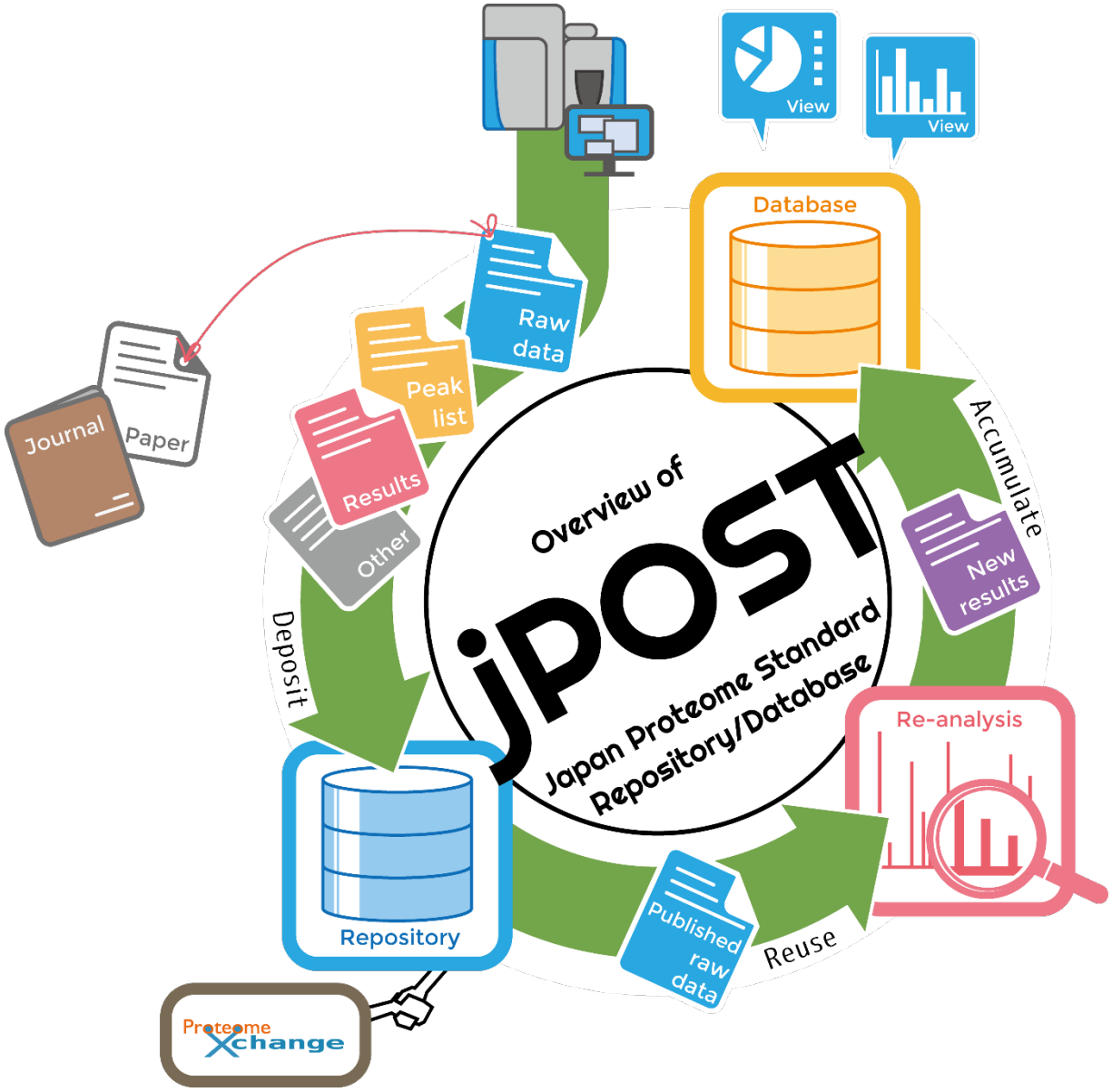
47615 files amount to **15.0 TB**.

92 species.

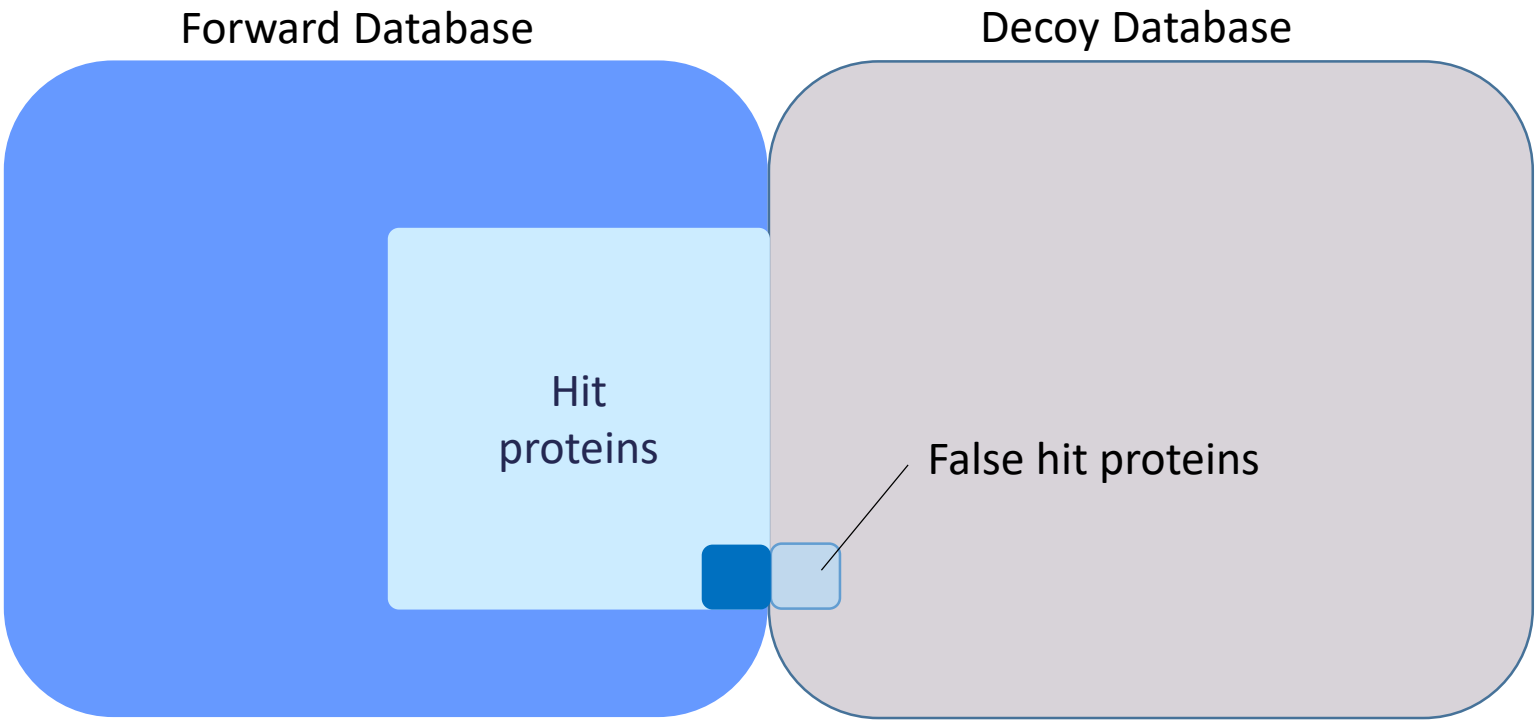
Oct 2019



jPOST Re-Analysis



Target-Decoy Approach for estimating FDR(%)



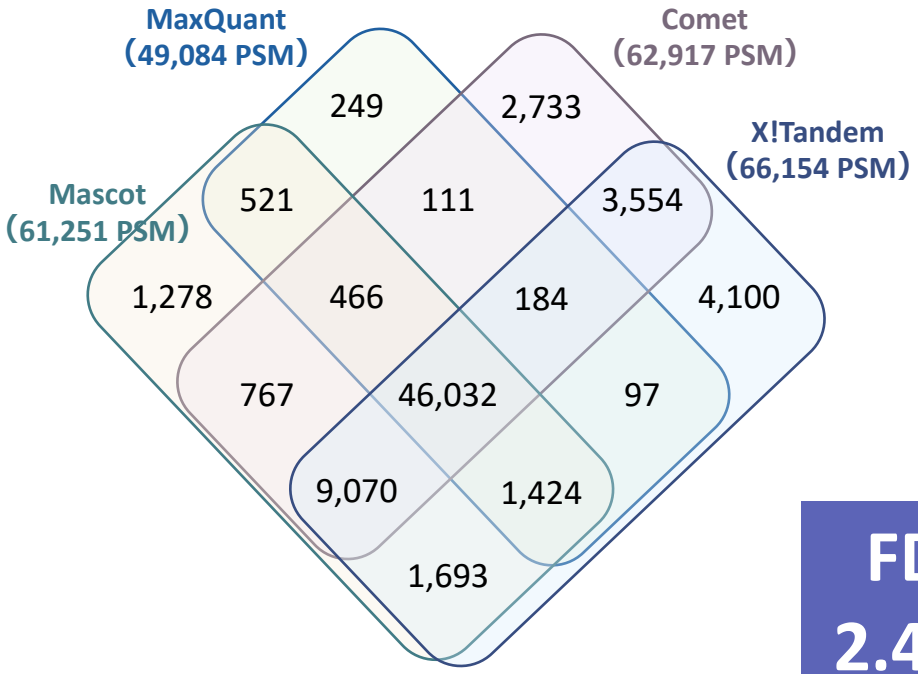
False Discovery Rate = False Positive/(False Positive + True Positive)

Multiple search engines with FDR control

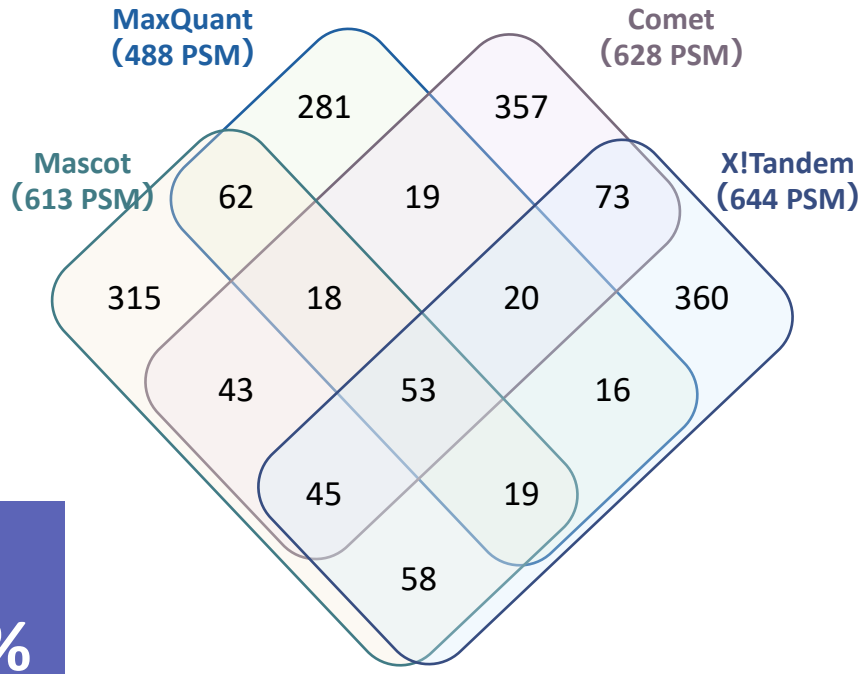
PSM at **FDR=1%** by different search engines

- HeLa global proteome dataset (JPST000200)

Target
72,279 PSM in total



Decoy
1,739 PSM in total



FDR 2.41%

Universal index for annotated MSMS spectra

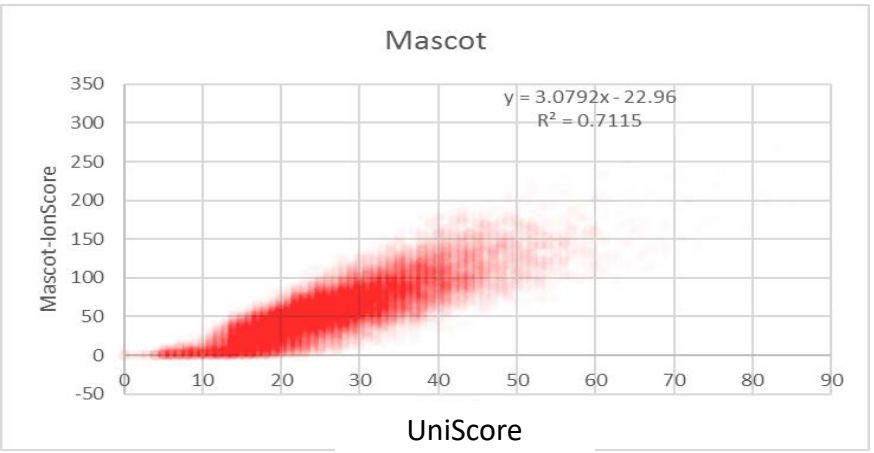
How can we merge the results from different sources?

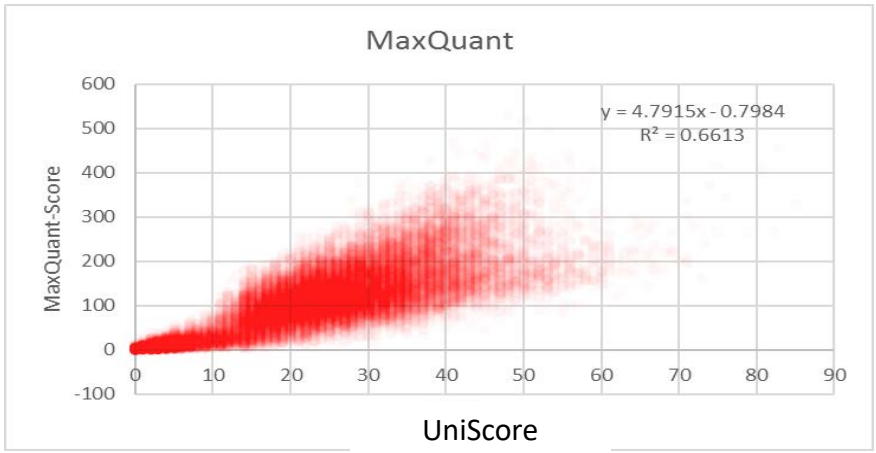
UniScore

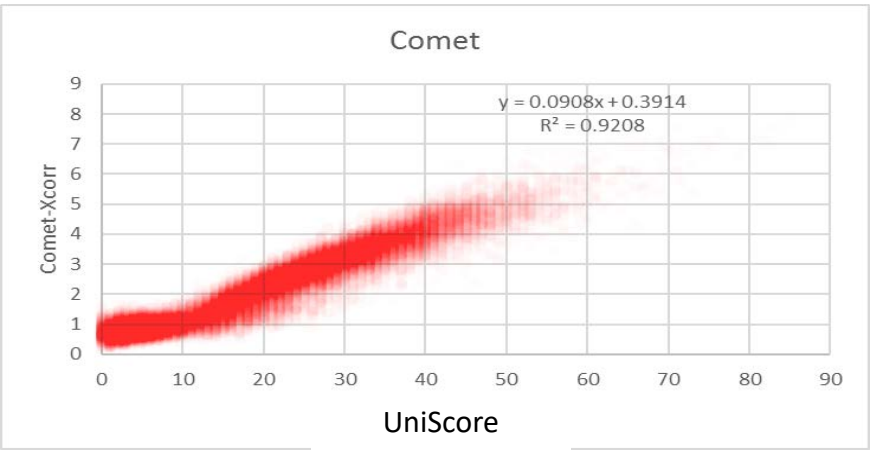
- based on peak annotation in MSMS
- search engine independent
- MS instrument independent
- search DB independent
- can be used as universal threshold for peptide identification

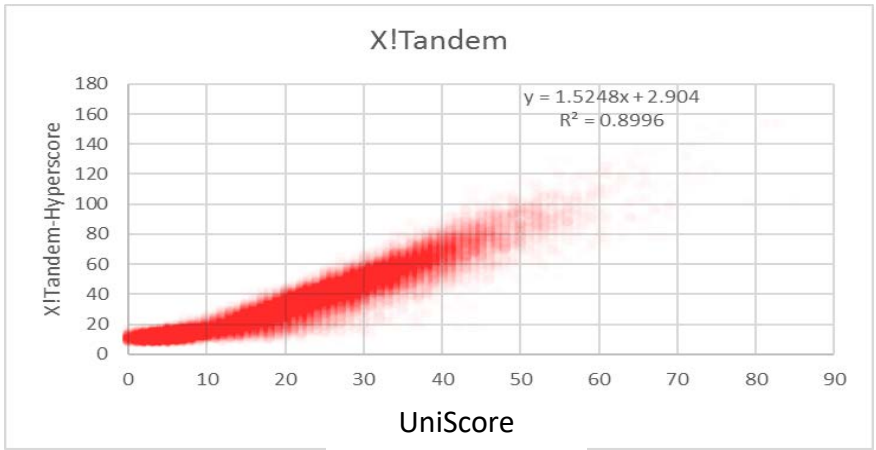
Relationship between different scores

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



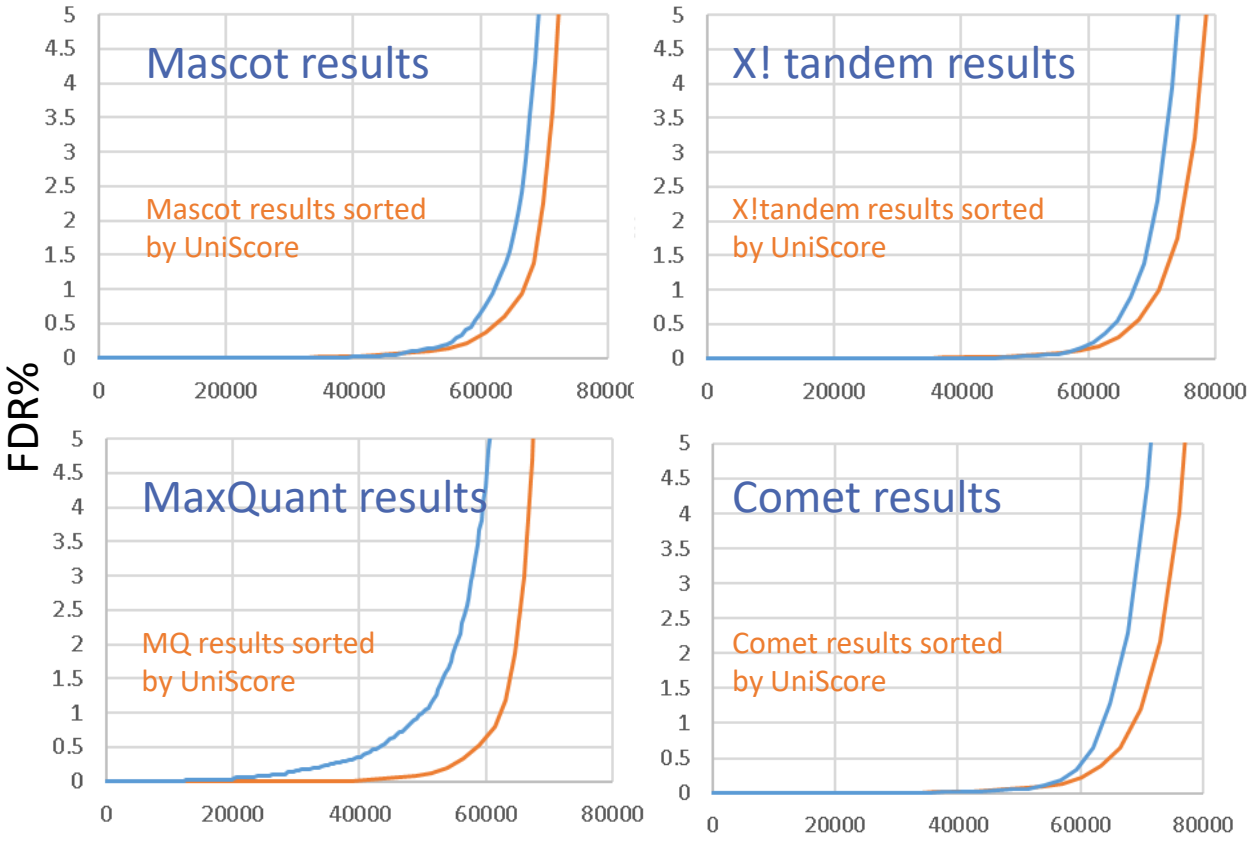


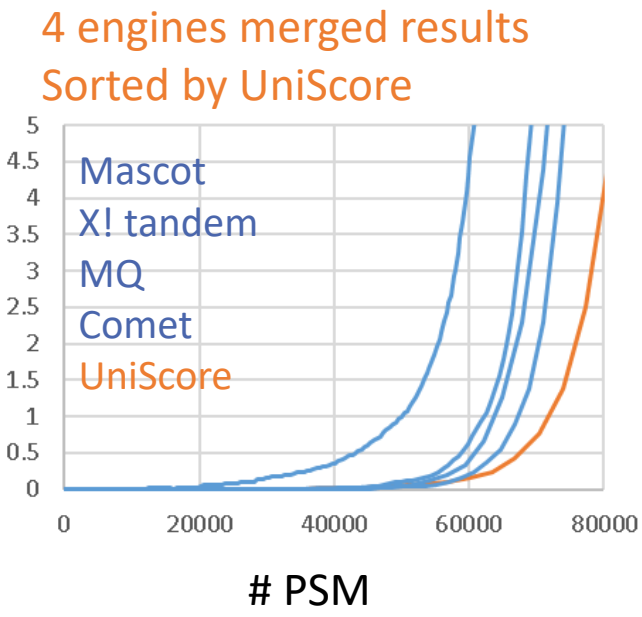




Optimized UniScore to rank down the decoy hits

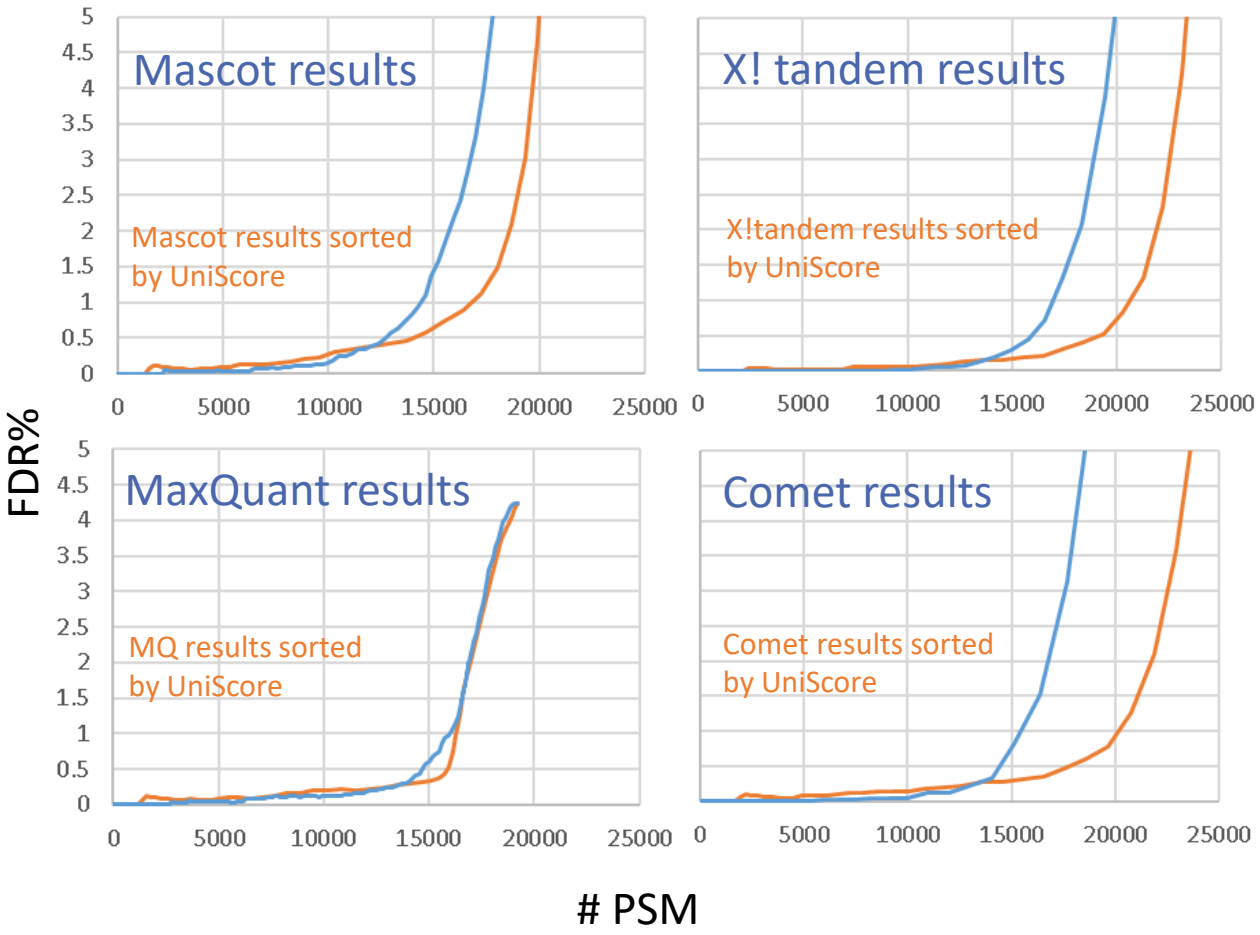
Dataset:JPST000200(5 files)
Tryptic peptides from HeLa cells, by Thermo Q-Exactive with 8h gradient

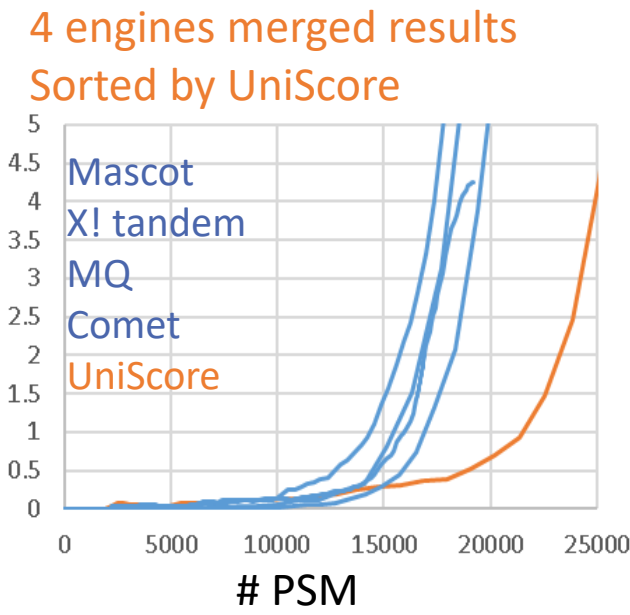




Optimized UniScore to rank down the decoy hits

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





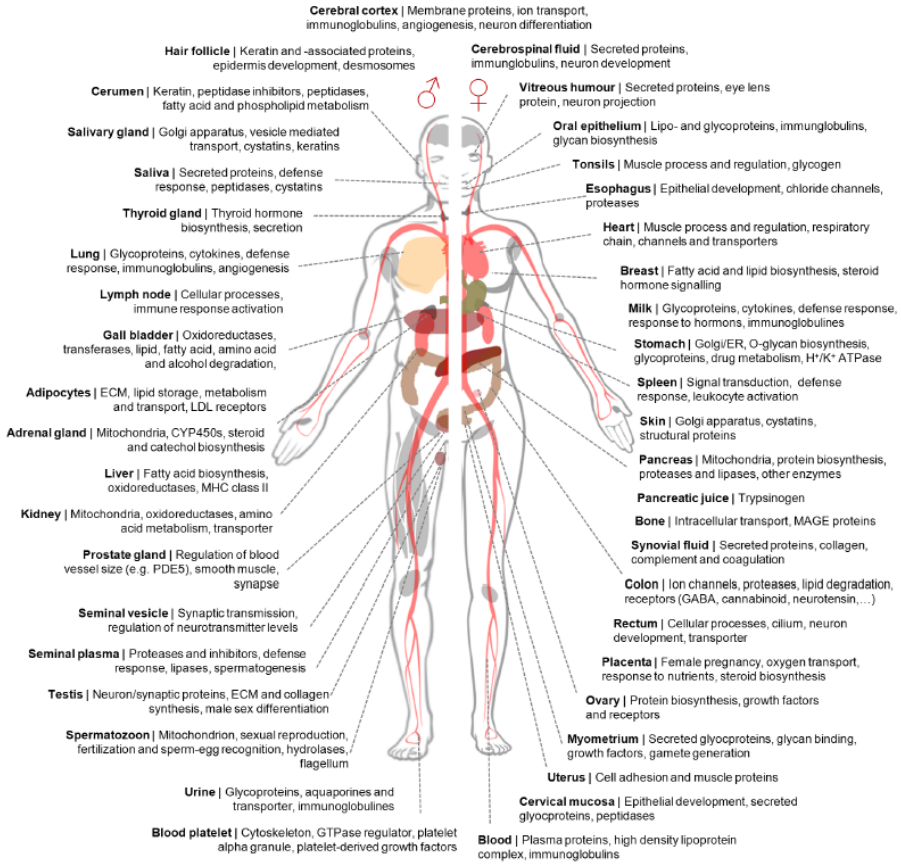
Re-analysis of large-scale HPP datasets

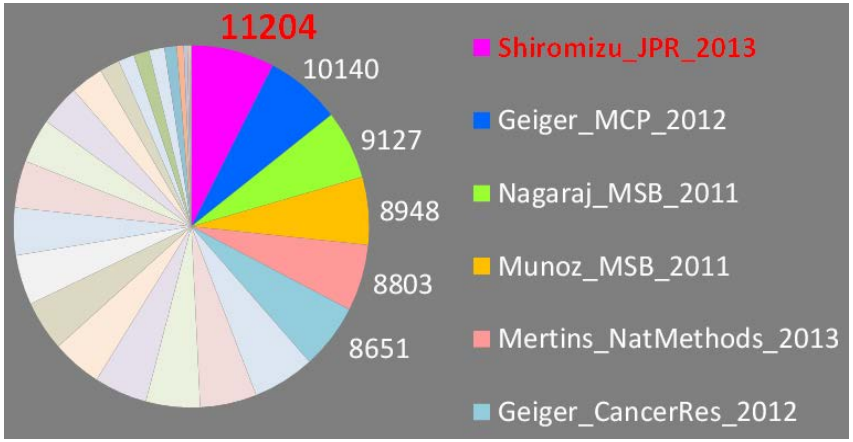
582 | NATURE | VOL 509 | 29 MAY 2014

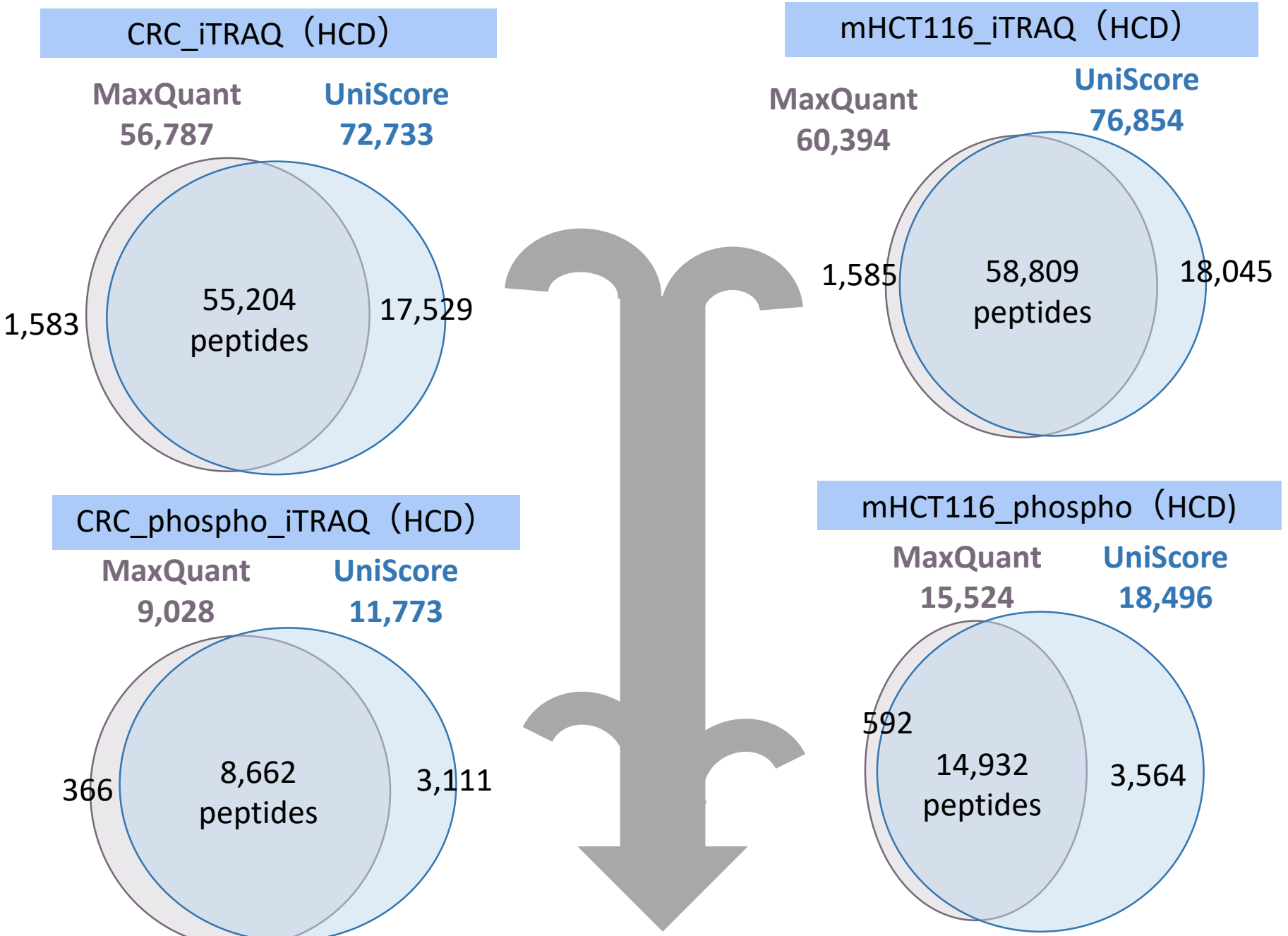
ARTICLE

doi:10.1038/nature13319

Mass-spectrometry-based draft of the human proteome



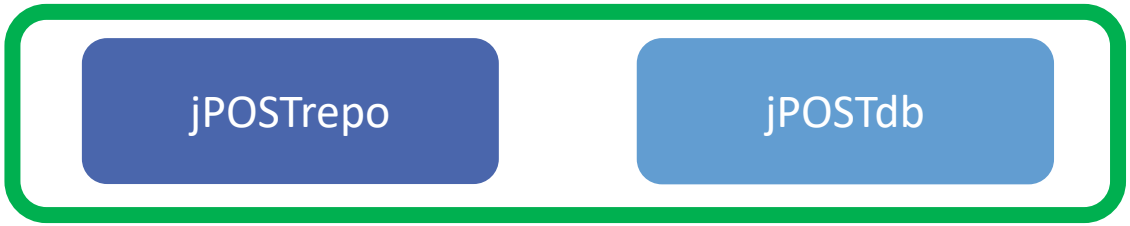


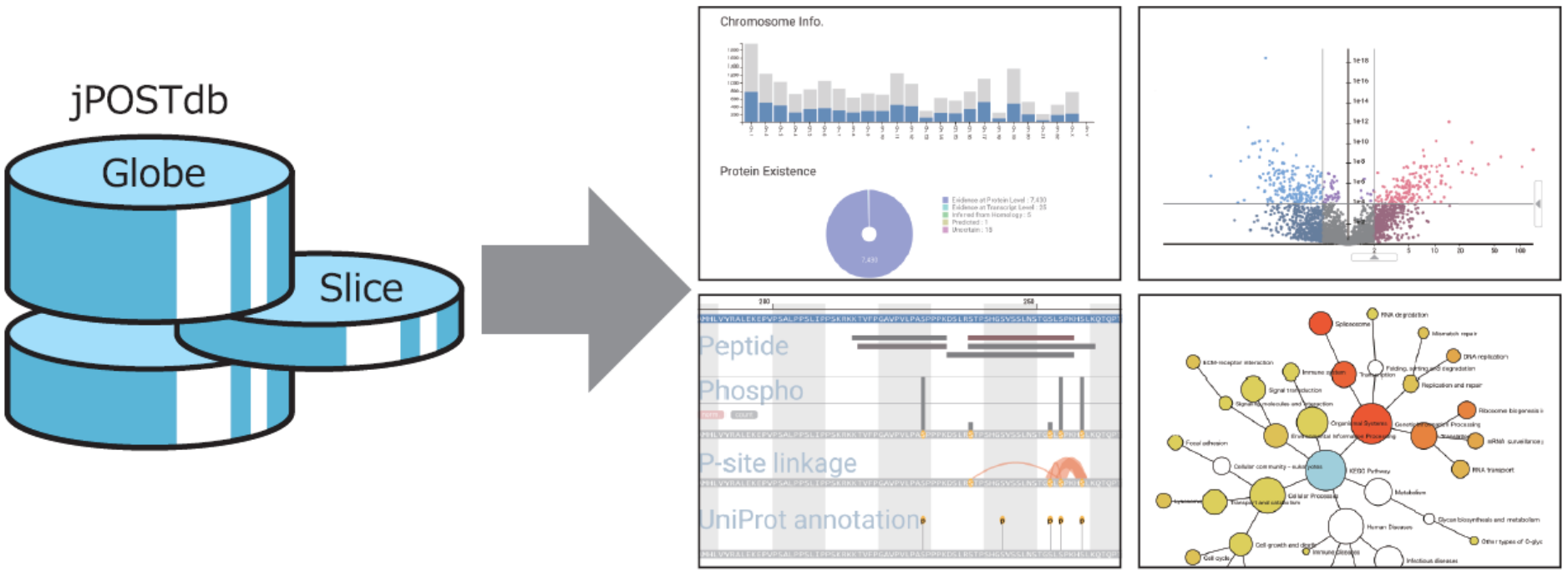

CRC_iTRAQ（HCD）
MaxQuant
56,787
UniScore
72,733
1,583
55,204
peptides
17,529
mHCT116_iTRAQ（HCD）
MaxQuant
60,394
UniScore
76,854
1,585
58,809
peptides
18,045
CRC_phospho_iTRAQ（HCD）
MaxQuant
9,028
UniScore
11,773
366
8,662
peptides
3,111
mHCT116_phospho（HCD）
MaxQuant
15,524
UniScore
18,496
592
14,932
peptides
3,564

The jPOST Environment

The jPOST environment: an integrated proteomics data repository and database

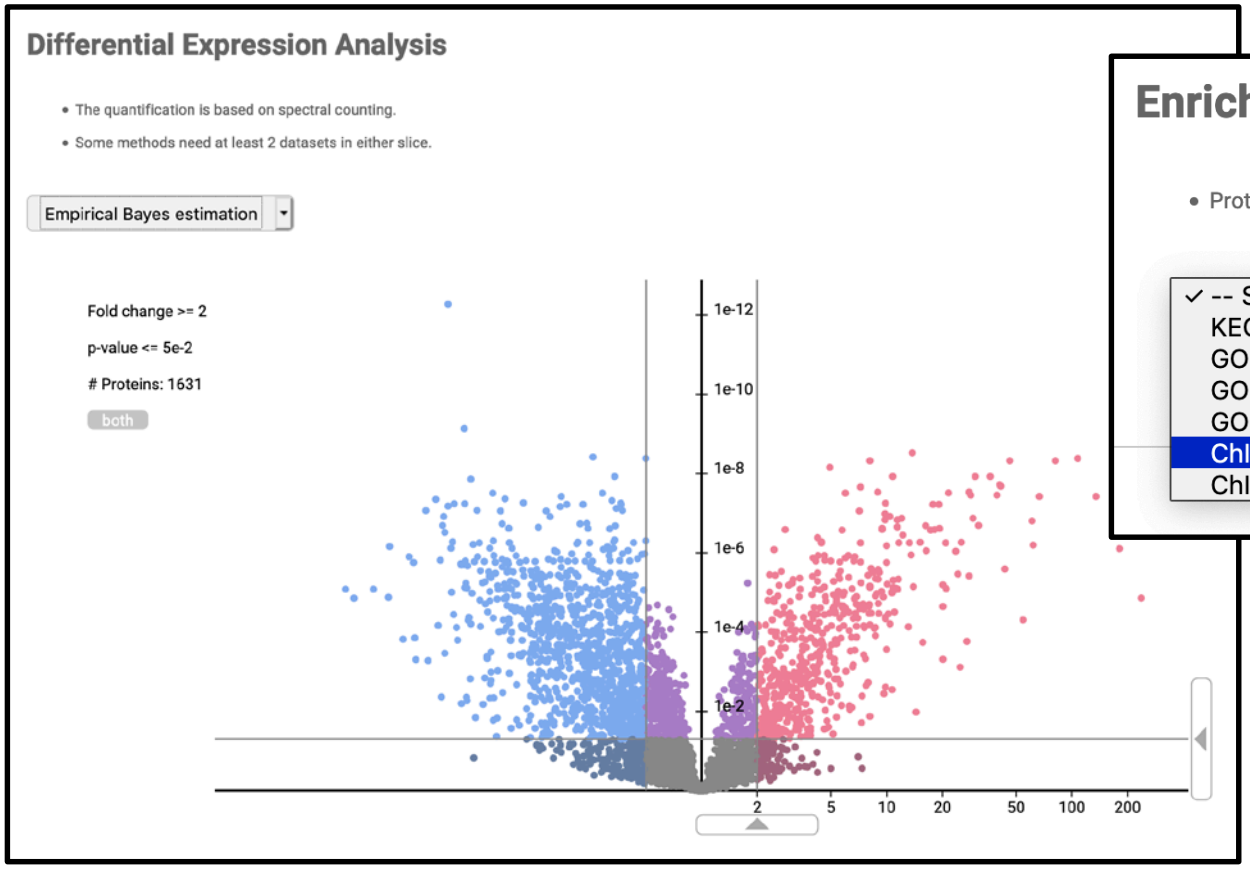
Moriya et al., *Nucleic Acids Res,* 2019 Jan 8;47(D1):D1218-D1224

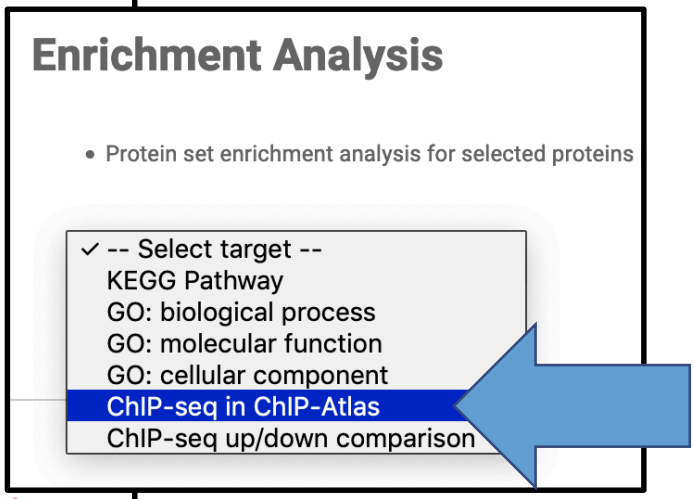




ChIP-Atlas 連携

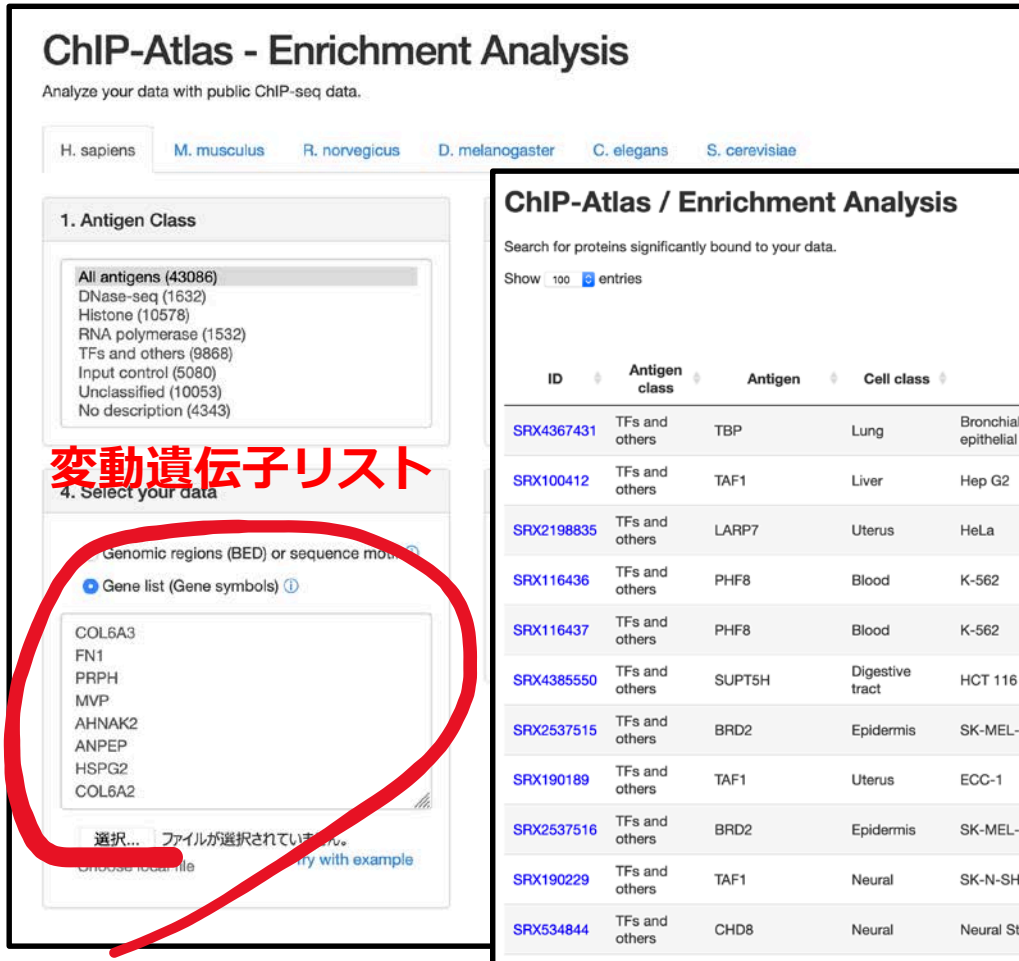
jPOST では Slice (データセットグループ） 間で変動解析するための簡易的なツールを提供しており、有意な変動のあったタンパク質（遺伝子）リストを用いてシームレスに ChIP-Atlas でエンリッチメント解析を行うことが可能



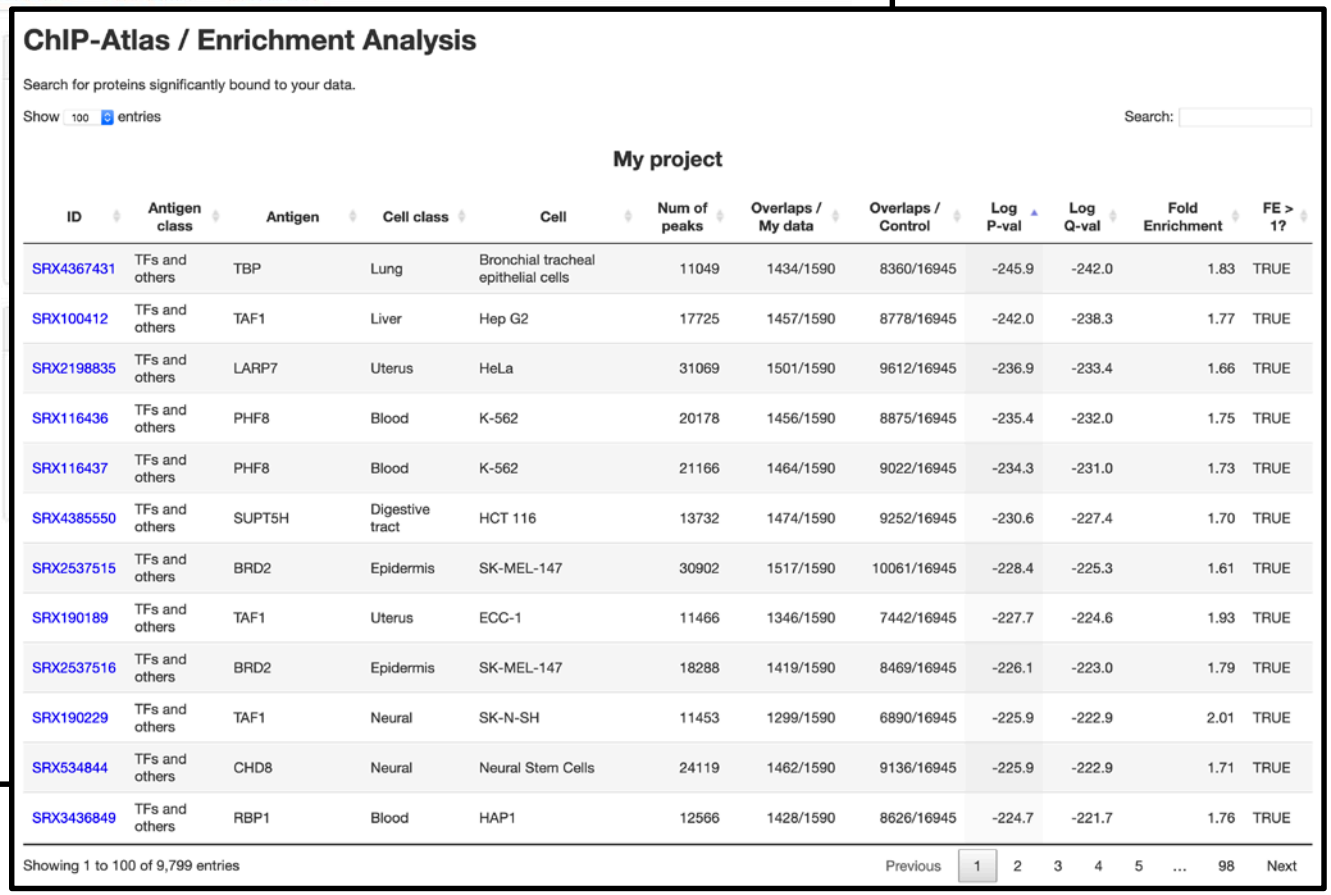


これまで提供していた
KEGG, GOの他に
ChIP-Atlasが追加

ChIP-Atlas 連携



ChIP-Atlas サイト



ChIP-Atlas / Enrichment Analysis

Search for proteins significantly bound to your data.

Show 100 entries

Search:

My project

ID	Antigen class	Antigen	Cell class	Cell	Num of peaks	Overlaps / My data	Overlaps / Control	Log P-val	Log Q-val	Fold Enrichment	FE > 1?
SRX4367431	TFs and others	TBP	Lung	Bronchial tracheal epithelial cells	11049	1434/1590	8360/16945	-245.9	-242.0	1.83	TRUE
SRX100412	TFs and others	TAF1	Liver	Hep G2	17725	1457/1590	8778/16945	-242.0	-238.3	1.77	TRUE
SRX2198835	TFs and others	LARP7	Uterus	HeLa	31069	1501/1590	9612/16945	-236.9	-233.4	1.66	TRUE
SRX116436	TFs and others	PHF8	Blood	K-562	20178	1456/1590	8875/16945	-235.4	-232.0	1.75	TRUE
SRX116437	TFs and others	PHF8	Blood	K-562	21166	1464/1590	9022/16945	-234.3	-231.0	1.73	TRUE
SRX4385550	TFs and others	SUPT5H	Digestive tract	HCT 116	13732	1474/1590	9252/16945	-230.6	-227.4	1.70	TRUE
SRX2537515	TFs and others	BRD2	Epidermis	SK-MEL-147	30902	1517/1590	10061/16945	-228.4	-225.3	1.61	TRUE
SRX190189	TFs and others	TAF1	Uterus	ECC-1	11466	1346/1590	7442/16945	-227.7	-224.6	1.93	TRUE
SRX2537516	TFs and others	BRD2	Epidermis	SK-MEL-147	18288	1419/1590	8469/16945	-226.1	-223.0	1.79	TRUE
SRX190229	TFs and others	TAF1	Neural	SK-N-SH	11453	1299/1590	6890/16945	-225.9	-222.9	2.01	TRUE
SRX534844	TFs and others	CHD8	Neural	Neural Stem Cells	24119	1462/1590	9136/16945	-225.9	-222.9	1.71	TRUE
SRX3436849	TFs and others	RBP1	Blood	HAP1	12566	1428/1590	8626/16945	-224.7	-221.7	1.76	TRUE

Showing 1 to 100 of 9,799 entries

Previous 1 2 3 4 5 ... 98 Next

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

コメ

Free word | Ontology keyword

rice

Project type

All | Mass spectrometry | Gel electrophoresis | Antibody

Search | Reset

Search result rice

jPOST ID	PXID	Project title	Description
JPST000358		Shotgun Phosphoproteomic in Plant	
JPST000156	PXD004603	Rice (Oryza sativa) proteome	

ダイズ

Search result soybean

jPOST ID	PXID	Project title	Description	Complete / Partial	Publication	Principal investigator
JPST000585	PXD013575	SK07_CatharanthusRoseus_Phosphoproteomics	Quantitative phosphoproteomic analysis unravels me …	Partial	31356092	Setsuko Komatsu Faculty of Life and Environmental and Information Sciences, Fukui University of Technology.
JPST000374	PXD008603	TK03_Proteome_Soybean	Proteomic analysis of the effect of plant-derived …	Partial	29704570	Setsuko Komatsu Faculty of Life and Environmental Sciences, University of Tsukuba

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

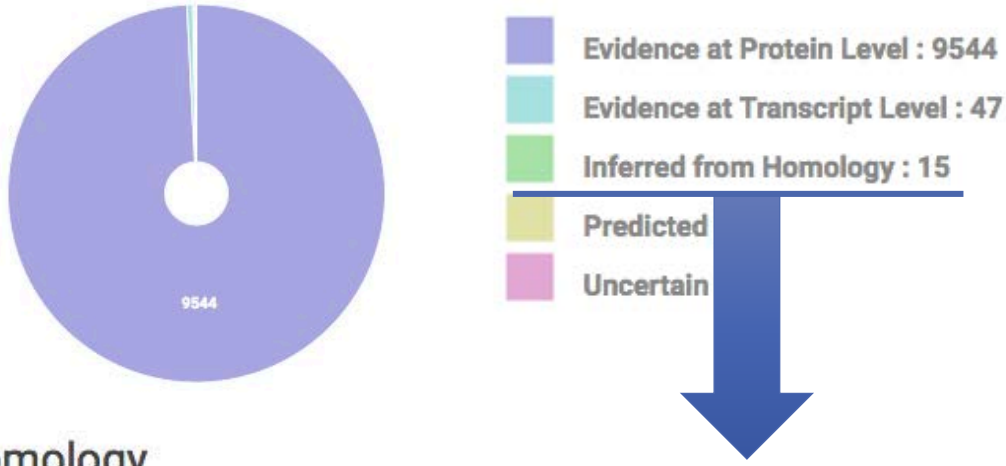
- プロテオーム発現データベース化と比較定量
- 各種ツールによるデータ可視化、ネットワーク解析
- Missing protein の同定

On-going

統計情報

- UniProt(ヒトはneXtProt)におけるタンパク質の同定レベル

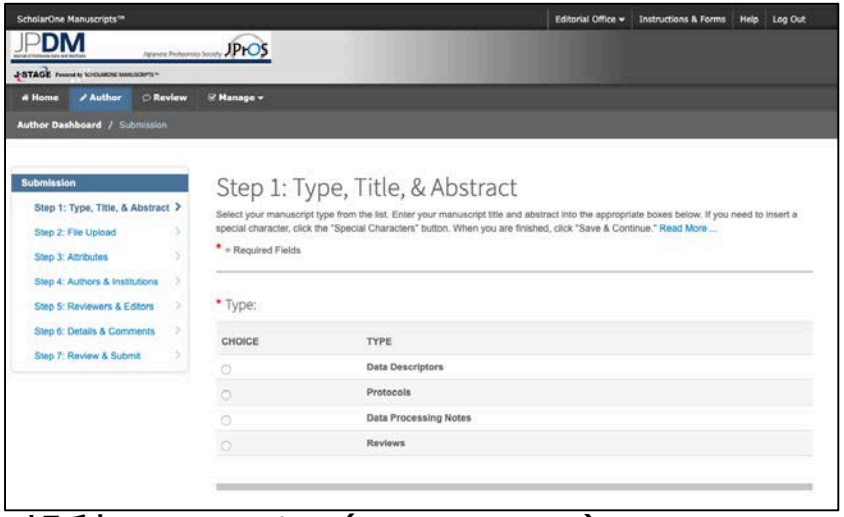
Protein Evidence



Inferred from Homology

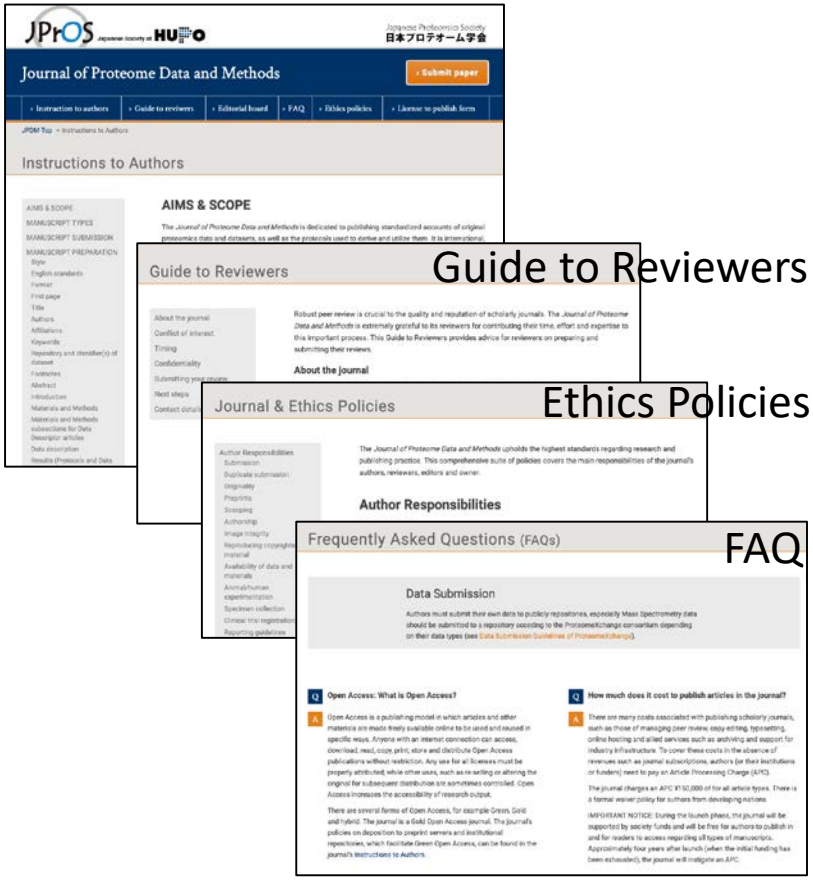
Protein Name	ID	Accession	#Peptide	#PSM	Unique Peptide
Arylacetamide deacetylase-like 4	ADCL4_HUMAN	Q5VUY2	1	2	yes
COBW domain-containing protein 6	CBWD6_HUMAN	Q4V339	1	2	
Cytochrome b-c1 complex subunit 6-like, mitochondrial	QCR6L_HUMAN	A0A096LP55	1	1	
Eukaryotic translation initiation factor 3 subunit C-like protein	EIFCL_HUMAN	B5ME19	6	141	
Golgin subfamily A member 6-like protein 6	GG6L6_HUMAN	A8MZA4	1	1	
HIG1 domain family member 1C	HIG1C_HUMAN	A8MV81	2	2	
Homeobox protein unc-4 homolog	UNC4_HUMAN	A6NJT0	2	2	yes
Immunoglobulin kappa variable 1-12	KV112_HUMAN	A0A0C4DH73	1	1	
Immunoglobulin kappa variable 3D-15	KVD15_HUMAN	A0A087WSY6	2	2	
Olfactory receptor 5AC2	O5AC2_HUMAN	Q9NZP5	1	1	

キュレーションの深化と強化
～Journal of Proteome Data and Methods (JPDM) 創刊～



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

Instructions to Authors



Guide to Reviewers

Ethics Policies

FAQ

- 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

2019.9.30
創刊！

Editorial

JPDM
Journal of Proteome Data and Methods

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

JPOST
Repository / Database