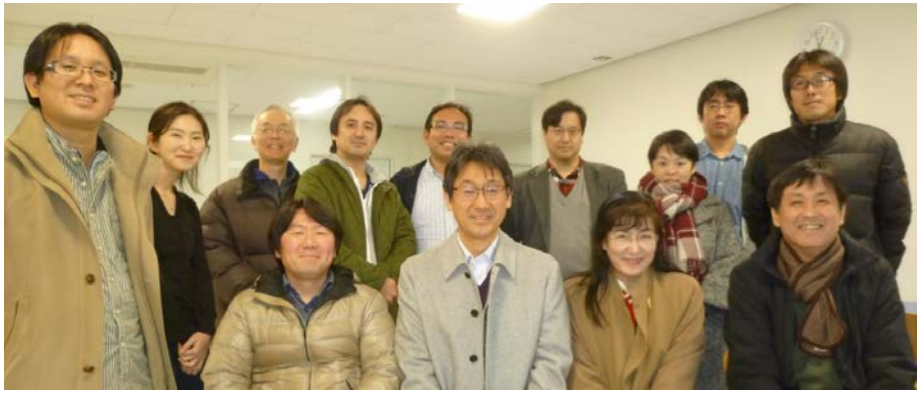
Since 2015 JPOST Repository/Database
Japan Proteome Standard Repository/Database

Proteomic Data Integration and Sharing by jPOST Repository/Database

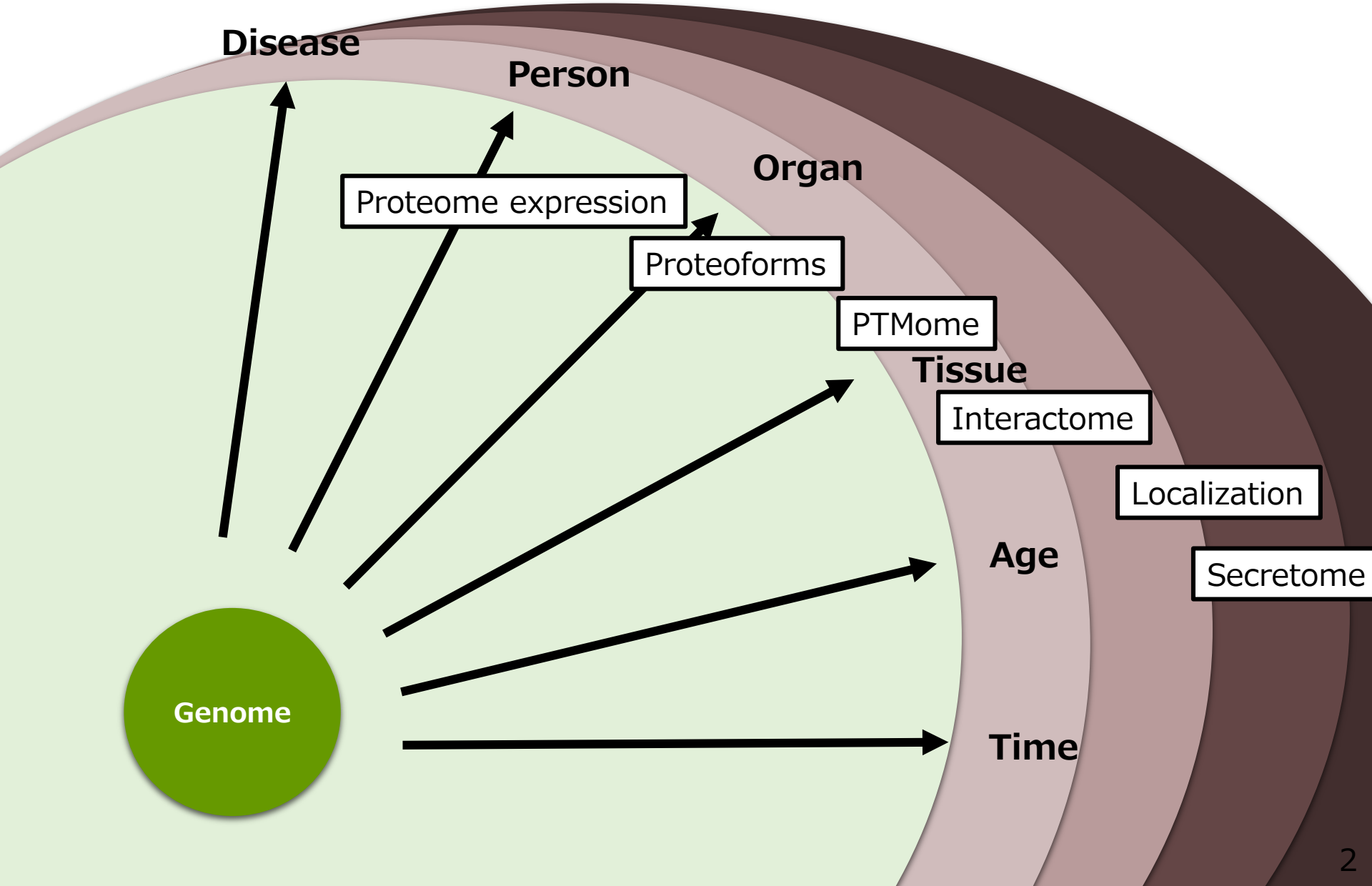
Yasushi Ishihama (Kyoto Univ) and jPOST project team

Y. Moriya[1], S. Kawano[1], S. Okuda[2], Y. Watanabe[2], M. Matsumoto[3], T. Takami[3], D. Kobayashi[4], N. Araki[4], AC. Yoshizawa[5], T. Tabata[5], M. Iwasaki[5], S. Goto[1]

1 DBCLS, 2 Niigata Univ, 3 Kyushu Univ, 4 Kumamoto Univ, 5 Kyoto Univ



Genome → Proteome → Function



Human Proteome Map/Atlas

Major Human Proteome DBs

UniProt/SwissProt, NCBInr, …

- HPP-HUPO
  (Human Proteome Project) → Start in 2010, international efforts
- **ProteomicsDB**
  (LC/MS/MS) → Nature, 2014
- Human Protein Atlas
  (Antibody-based) → Science, 2015

Nature 2014, the Human Proteome

Nature. 2014 , DOI: 10.1038/nature13302, PMID: 24870542

A draft map of the human proteome

Min-Sik Kim; Sneha M Pinto; Derese Getnet; Raja Nirujogi; Srikanth S Manda; Raghothama Chaerkady; Anil K Madugundu; Dhanashree S Kelkar; Ruth Isserlin; Shobhit Jain; Joji K Thomas; Babylakshmi Muthusamy; Pamela Leal-Rojas; Praveen Kumar; Nandini A Sahasrabuddhe; Lavanya Balakrishnan; Jayshree Advani; Bijesh George; Santosh Renuse; Lakshmi N Selvan; Arun H Patil; Vishalakshi Nanjappa; Aneesha Radhakrishnan; Samarjeet Prasad; Tejaswini Subbannayya; Rajesh Raju; Manish Kumar; Sreelakshmi K Sreenivasamurthy; Arivusudar Marimuthu; Gajanan J Sathe; Sandip Chavan; Keshava K Datta; Yashwanth Subbannayya; Apeksha Sahu; Soujanya D Yelamanchi; Savita Jayaram; Pavithra Rajagopalan; Jyoti Sharma; Krishna R Murthy; Nazia Syed; Renu Goel; Aafaque A Khan; Sartaj Ahmad; Gourav Dey; Keshav Mudgal; Aditi Chatterjee; Tai-Chung Huang; Jun Zhong; Xinyan Wu; Patrick G Shaw; ... (22 more)

The availability of human genome sequence has transformed biomedical research over the past decade. However, an equivalent map for the human proteome with direct measurements of proteins and peptides does not exist yet. Here we present a draft map of the human proteome using high-resolution Fourier-transform mass spectrometry. In-depth proteomic profiling of 30 histologically normal human samples, including 17 adult tissues, 7 fetal tissues and 6 purified primary haematopoietic cells, resulted in identification of proteins encoded by 17,294 genes accounting for approximately 84% of the total annotated protein-coding genes in humans. A unique and comprehensive strategy for proteogenomic analysis enabled us to discover a number of novel protein-coding regions, which includes translated pseudogenes, non-coding RNAs and upstream open reading frames. This large human proteome catalogue (available as an interactive web-based resource at http://www.humanproteomemap.org) will complement available human genome and transcriptome data to accelerate biomedical research in health and disease.

17,294 gene products

Nature. 2014 , DOI: 10.1038/nature13319

Mass-spectrometry-based draft of the human proteome

Mathias Wilhelm; Judith Schlegl; Hannes Hahne; Amin Moghaddas Gholami; Marcus Lieberenz; Mikhail M. Savitski; Emanuel Ziegler; Lars Butzmann; Siegfried Gessulat; Harald Marx; Toby Mathieson; Simone Lemeer; Karsten Schnatbaum; Ulf Reimer; Holger Wenschuh; Martin Mollenhauer; Julia Slotta-Huspenina; Joos-Hendrik Boese; Marcus Bantscheff; Anja Gerstmair; Franz Faerber; Bernhard Kuster

Proteomes are characterized by large protein-abundance differences, cell-type- and time-dependent expression patterns and post-translational modifications, all of which carry biological information that is not accessible by genomics or transcriptomics. Here we present a mass-spectrometry-based draft of the human proteome and a public, high-performance, in-memory database for real-time analysis of terabytes of big data, called ProteomicsDB. The information assembled from human tissues, cell lines and body fluids enabled estimation of the size of the protein-coding genome, and identified organ-specific proteins and a large number of translated lincRNAs (long intergenic non-coding RNAs). Analysis of messenger RNA and protein-expression profiles of human tissues revealed conserved control of protein abundance, and integration of drug-sensitivity data enabled the identification of proteins predicting resistance or sensitivity. The proteome profiles also hold considerable promise for analysing the composition and stoichiometry of protein complexes. ProteomicsDB thus enables navigation of proteomes, provides biological insight and fosters the development of proteomic technology.

18,097 gene products

Criticism for merging datasets

Journal of proteome research

dx.doi.org/10.1021/pr500572z | J. Proteome Res. 2014, 13, 3854−3855

Letter

pubs.acs.org/jpr

Analyzing the First Drafts of the Human Proteome

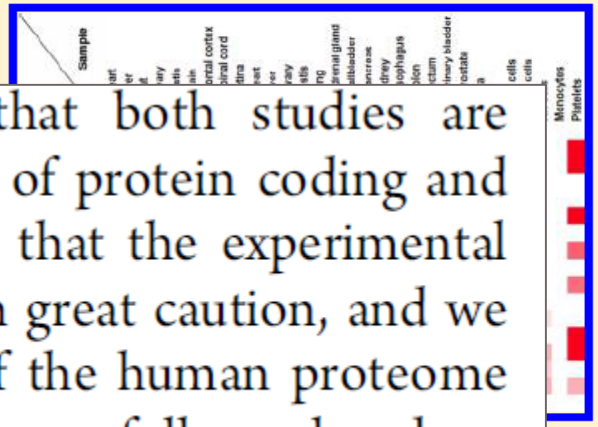
Iakes Ezkurdia,[†] Jesús Vázquez,[§] Alfonso Valencia,[‡] and Michael Tress*[,‡]

[†]Unidad de Proteómica, [§]Laboratorio de Proteómica Cardiovascular, Centro Nacional de Investigaciones Cardiovasculares, Melchor Fernández Almagro, 3, Madrid 28029, Spain

[‡]Spanish National Cancer Research Centre (CNIO), Melchor Fernandez Almagro, 3, Madrid 28029, Spain

Supporting Information

ABSTRACT: This letter analyzes two large-scale proteomics studies published in the same issue of *Nature*. At the time of the release, both studies were



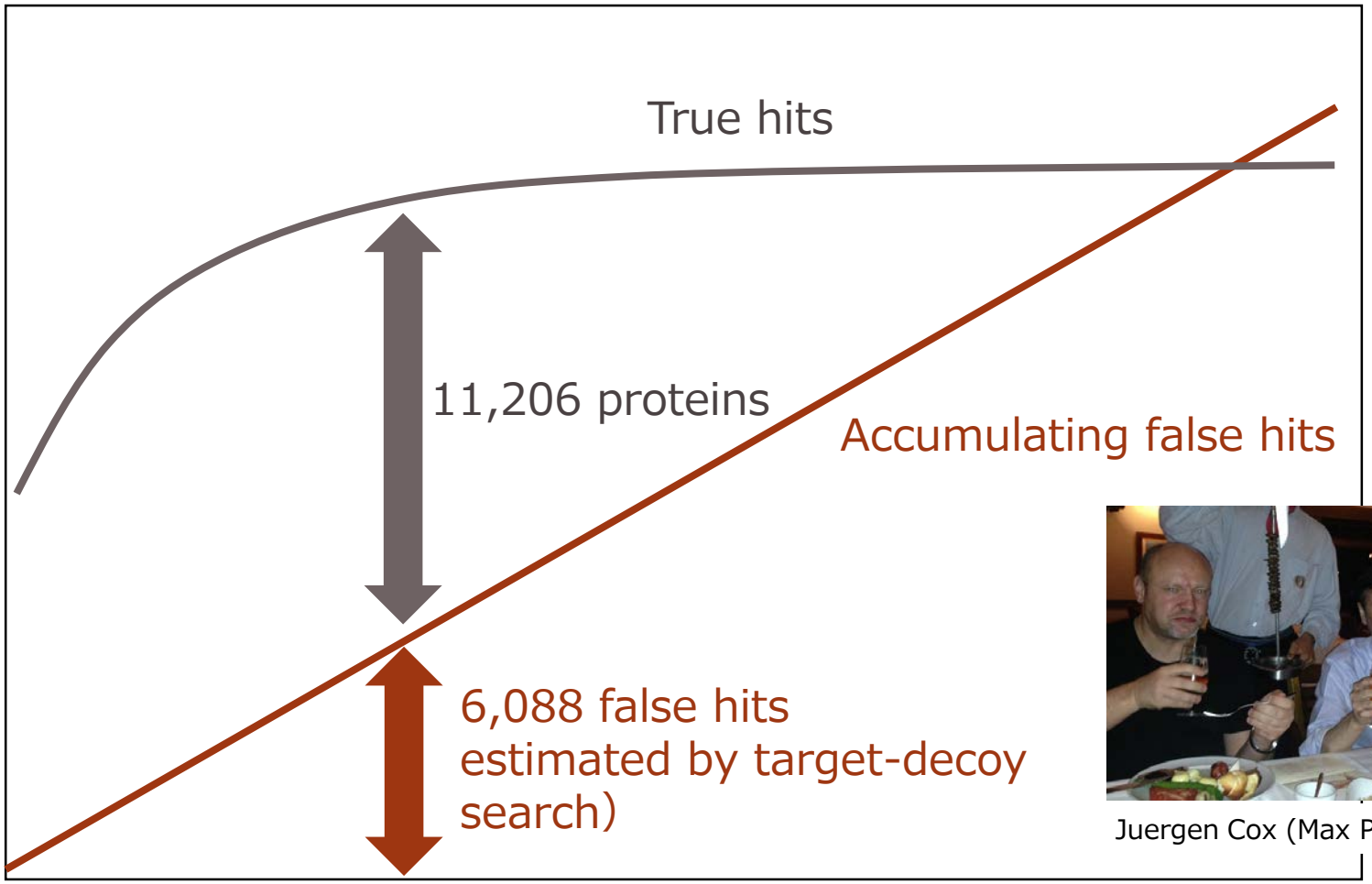
The results of our analysis show that both studies are substantially overestimating the number of protein coding and noncoding genes they find. We suggest that the experimental data from these two should be used with great caution, and we feel that these two unique draft maps of the human proteome should be put on hold until they can be carefully analyzed.

We conclude that

KEYWORDS

These draft maps should be withdrawn!

30% false positives!



Juergen Cox (Max Planck)

ProteomicsDB; self-corrected

Molecular & cellular proteomics : MCP. 2015 , DOI: 10.1074/mcp.M114.046995, PMID: 25987413

A scalable approach for protein false discovery rate estimation in large proteomic data sets.

Mikhail M Savitski; Mathias Wilhelm; Hannes Hahne; Bernhard Kuster; Marcus Bantscheff

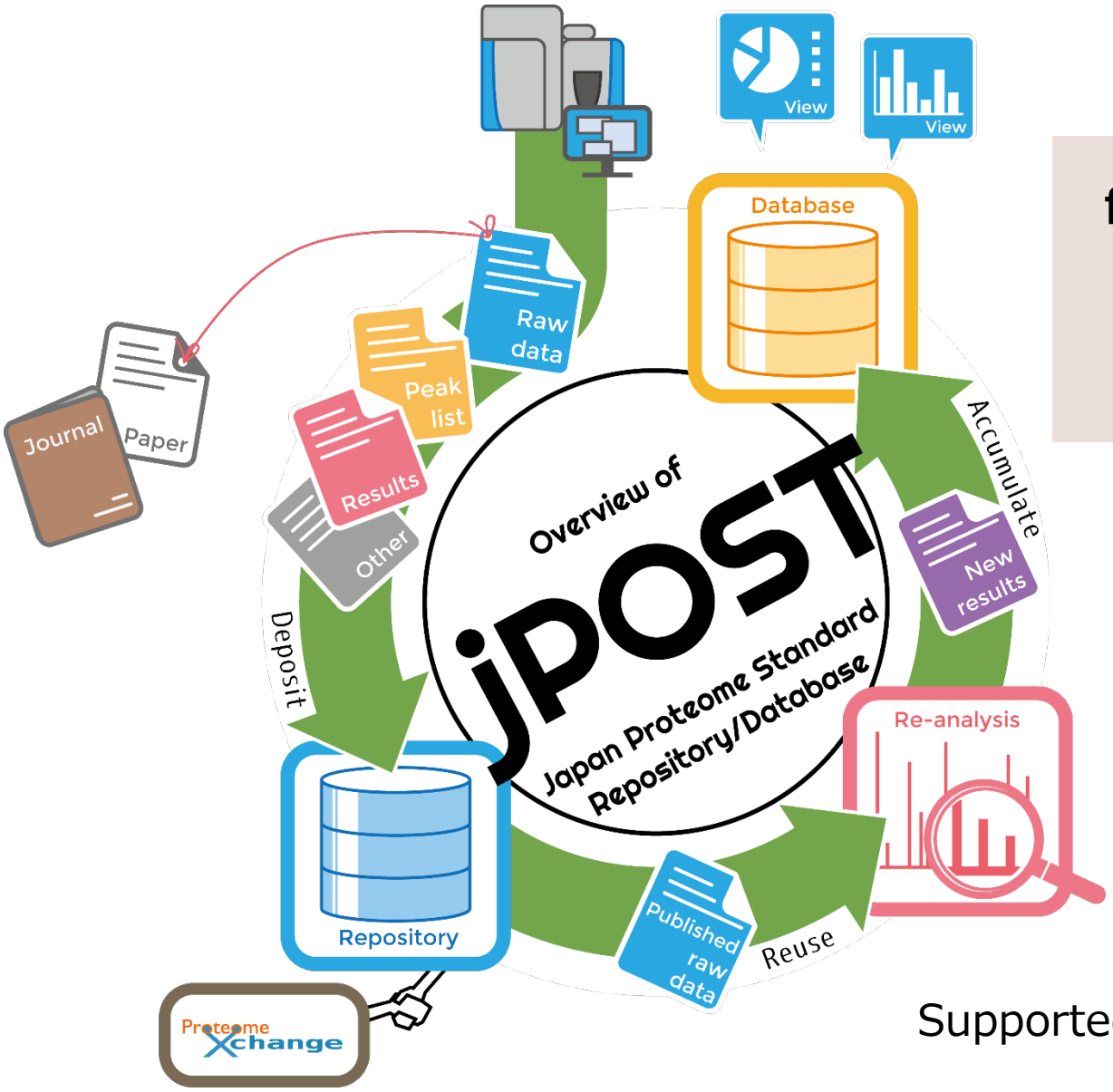
Calculating the number of confidently identified proteins and estimating false discovery rate (FDR) is a challenge when analyzing very large proteomic datasets such as entire human proteomes. Biological and technical heterogeneity in proteomic experiments further add to the challenge and there are strong differences in opinion regarding the conceptual validity of a protein FDR and no consensus regarding the methodology for protein FDR determination. There are also limitations inherent to the widely used classic target-decoy strategy (TDS) that particularly show when analyzing very large data sets and that lead to a strong over-representation of decoy identifications. In this study, we investigated the merits of the decoy-based protein FDR estimation approach taking advant collection comprised of ≈19,000 LC-MS/MS runs deposited i www.proteomicsdb.org). The "picked" protein FDR approach the same protein as a pair rather than as individual entities a decoy sequence depending on which receives the highest sc of this approach in combination with q-value based peptide instrument and search engine-specific differences. The "pick best when protein scoring was based on the best peptide q- stable number of true positive protein identifications over a demonstrate that this simple and unbiased strategy eliminat commonly used, "classic" protein FDR approach that causes protein identification in large data sets. The approach scales without losing performance, consistently increases the number of true positive protein identifications and is readily implemented in proteomics analysis software.

18,097 proteins (original)

for this difference (supplemental Figure 7). We next applied the described data analysis strategy to the subset of data stored in proteomicsDB corresponding to our earlier publication on a mass spectrometry based draft of the human proteome (9). Using the classic FDR strategy 14,035 proteins were observed at 1% protein FDR compared to 14,714 proteins using the picked strategy. Applying the picked strategy without any protein score threshold yielded 17,326 proteins of the target database at 11.3% protein FDR corresponding to 15,290 true positive protein identifications in the dataset. When analyzing the complete current content of proteomicsDB (including the data of the Pandey proteome (10) and a number of further datasets), the number of protein identifications at 1% FDR increased to 14,638 (classic) and 15,375 proteins (picked) respectively.

What is jPOST?

http://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.

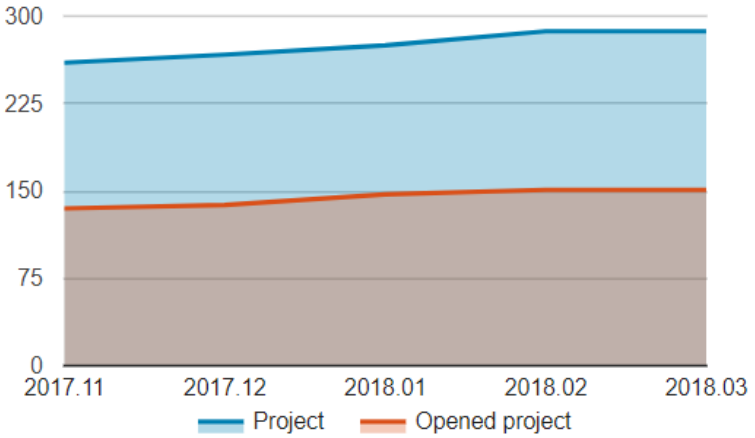
jPOST Repository

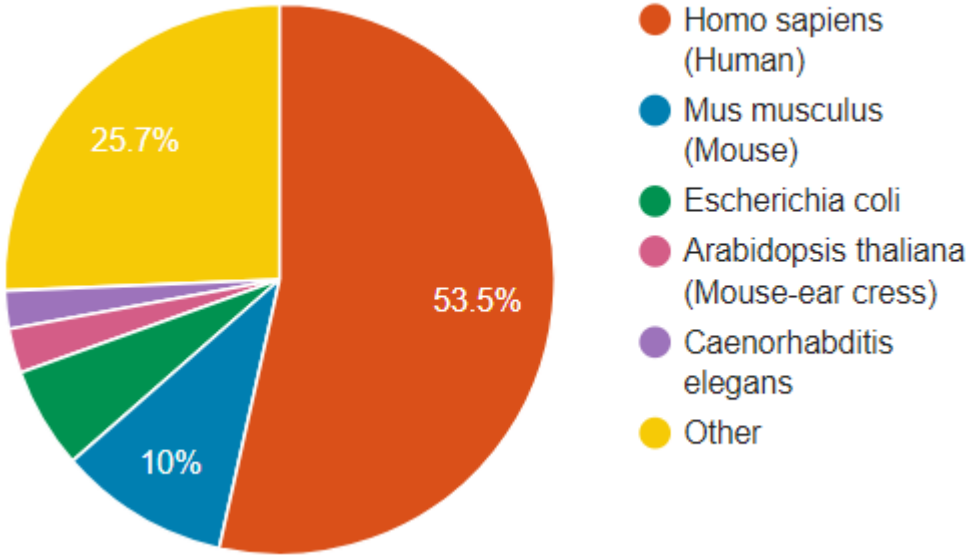
Statistics

287 projects are registered. **151** are opened.

30559 files amount to **8.0 TB**.

43 species.





Molecular & Cellular Proteomics
Providing Access to Annotated Spectra

Currently, the guidelines for MCP require that annotated spectra be provided in these two cases:

- proteins identified on the basis of single peptide
- post-translationally modified peptides

The purpose of this document is to summarize in one location existing tools that an author may choose to utilize to convert different proteomic results formats from a variety of software tools into files that satisfy the MCP requirement for access to annotated spectra.

It is possible to make annotated spectra available from most search engines, although the options for how to do this differ between software. MCP requires the files required for annotated spectra to be stored in a public repository that is beyond the control of the authors, so a lab website is not a compliant location. It may be possible to submit them as supplementary files with the manuscript submission. However, these files are often large (>100 MB). If this is the case, there are a handful of public repositories that can be used to store these files and the authors just need to provide a link to the location where they have been uploaded at the time of manuscript submission. MCP does not officially endorse any one repository. However, repositories that are part of the proteomeXchange consortium (proteomexchange.org) are suitable choices.

ACS Publications
Most Trusted. Most Cited. Most Read.

Journal of proteome research

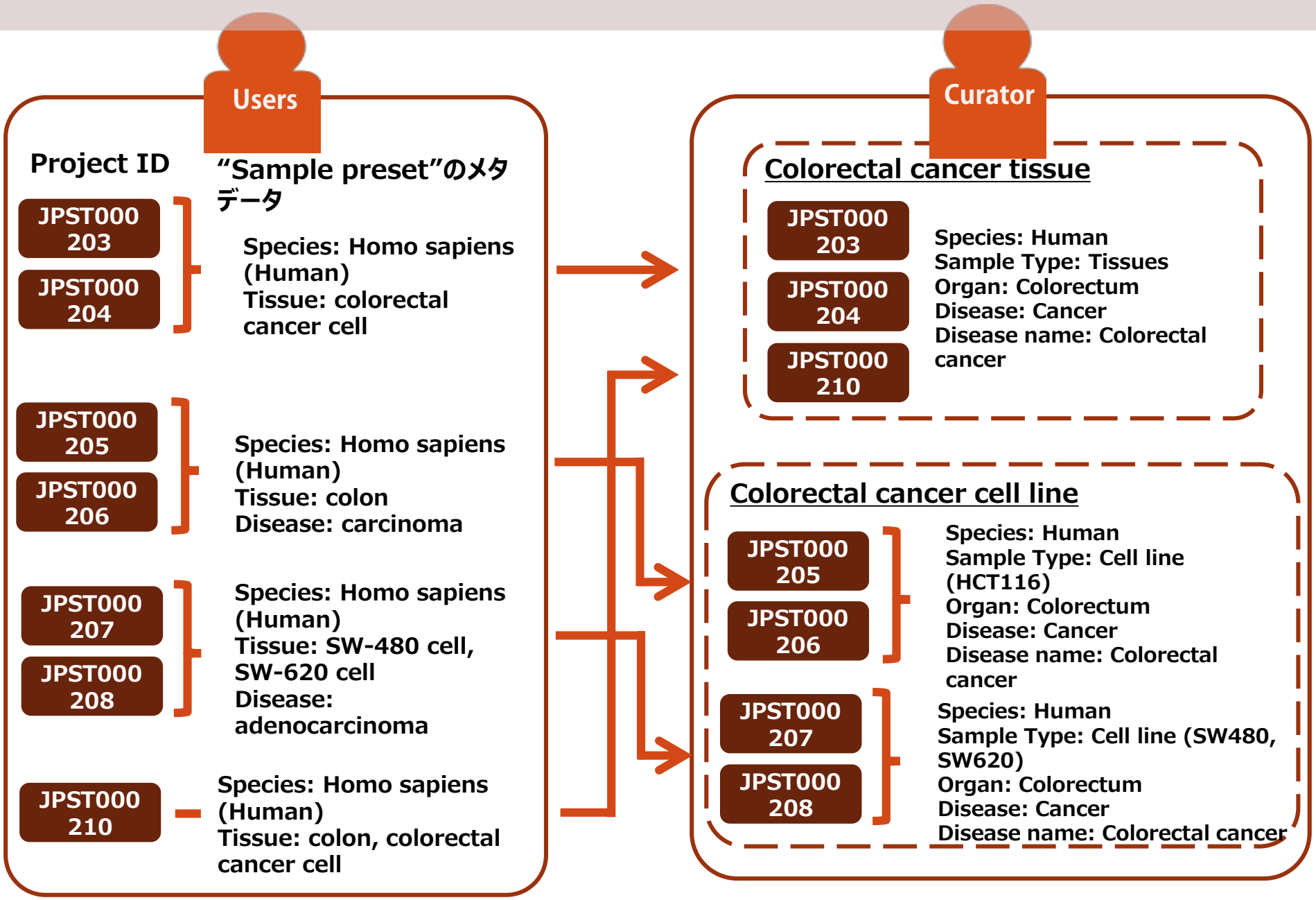
Instructions to Authors of *Journal of Proteome Research*

(Revised February 2018)

Important change about data deposition:

Author(s) are REQUIRED to deposit raw files and associated metadata in repositories such as ProteomeXchange (preferred) or other public repositories and to provide access to the information in the manuscript, including both the link as well as any necessary passwords (example shown below). Access to the information will be kept confidential while the manuscript is under review but will be open to the public upon publication. Please note: Providing this information on a link managed by the author(s) is not acceptable.

Manual curation of metadata

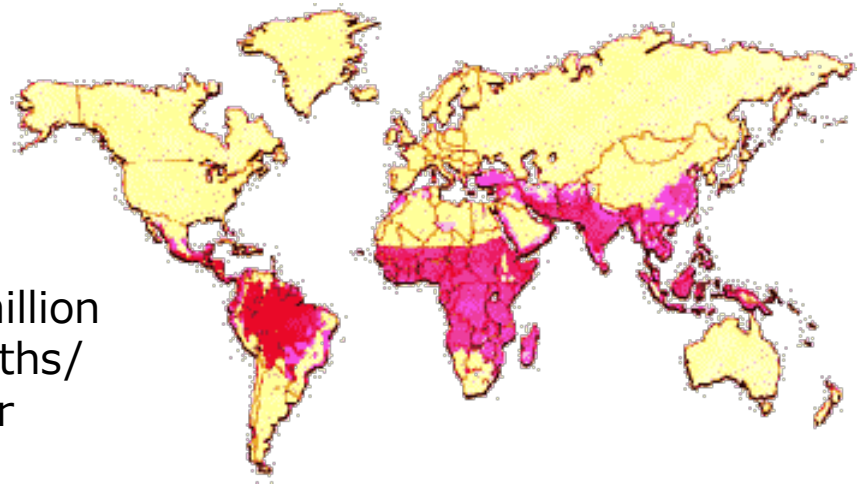


Proteomics community efforts against false hits

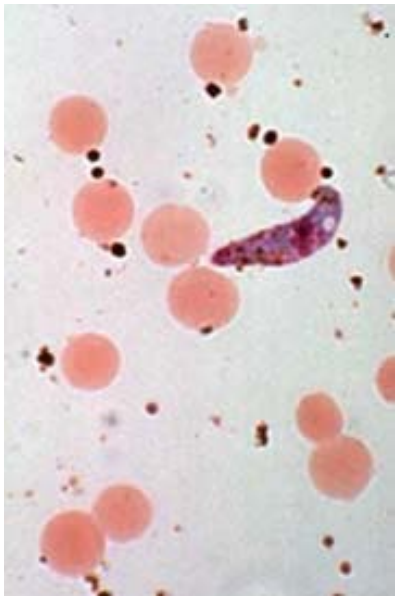
Malaria *Plasmodium falciparum* proteome by high-accuracy mass spectrometry in 2002

Hosts

- Anopheles mosquito
- human

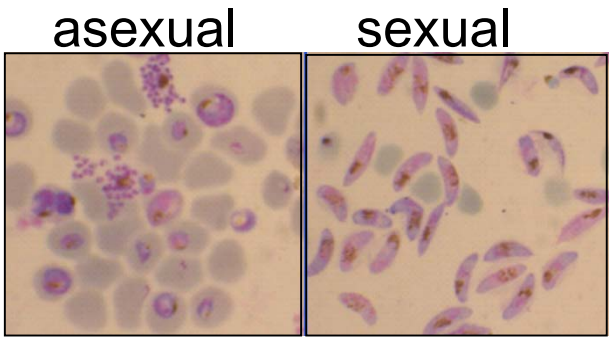


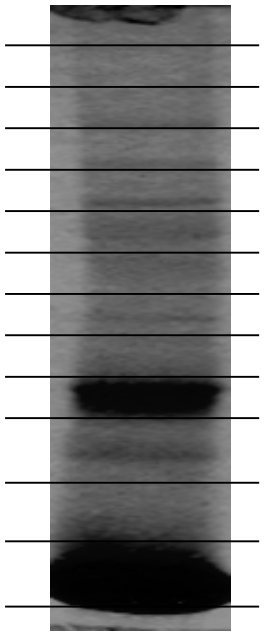
2 million deaths/ year

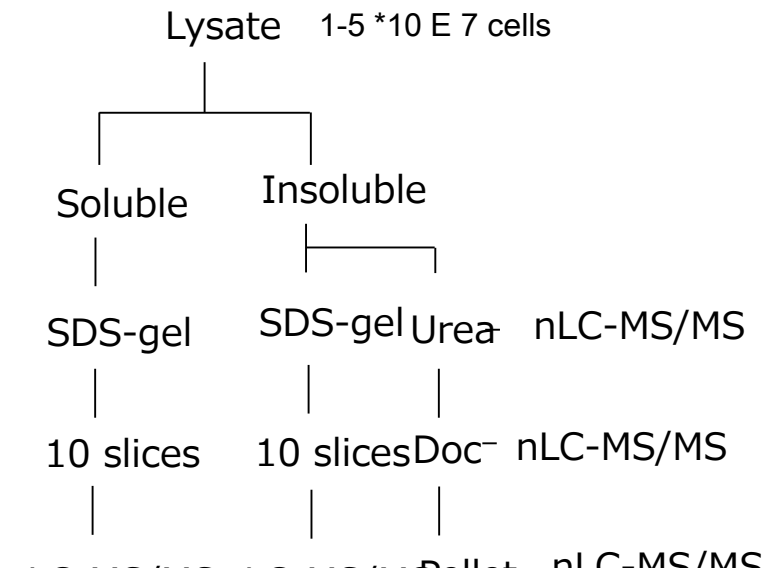


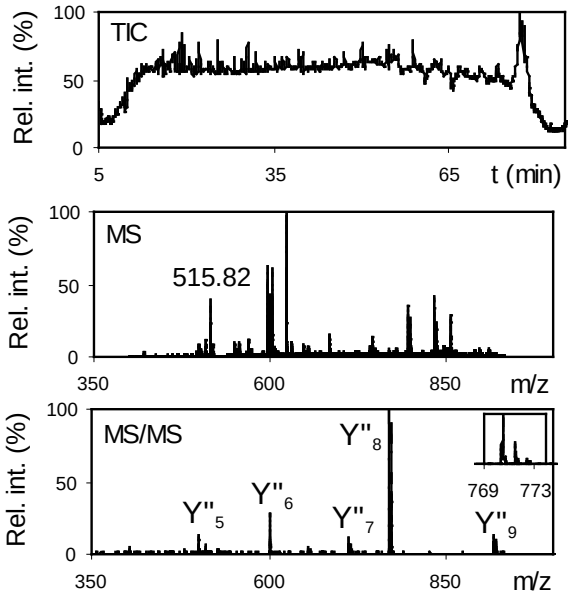
Proteomics of blood stages

infected RBCs



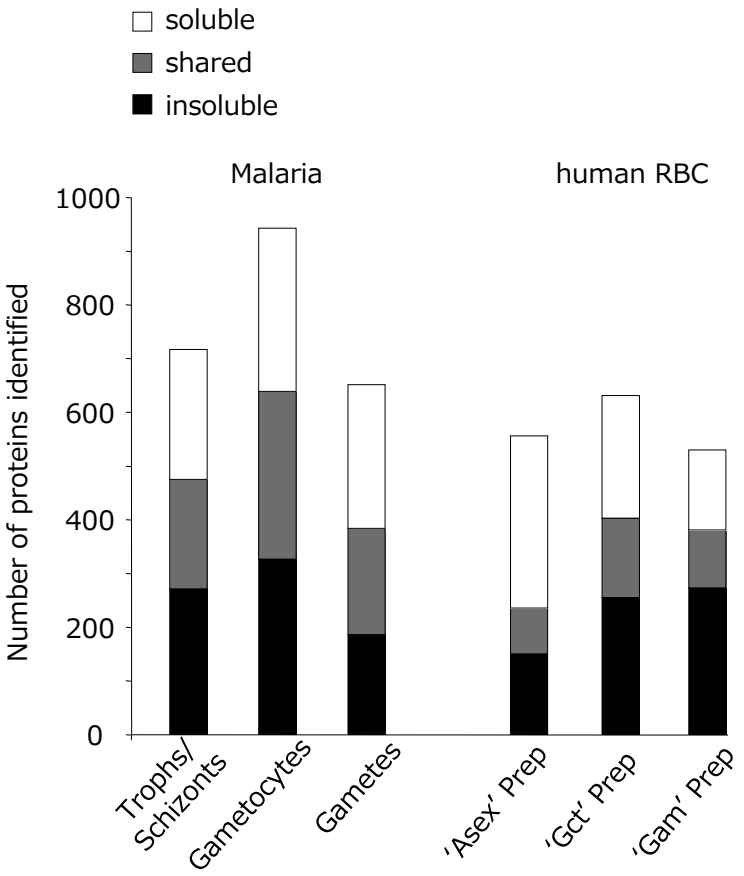




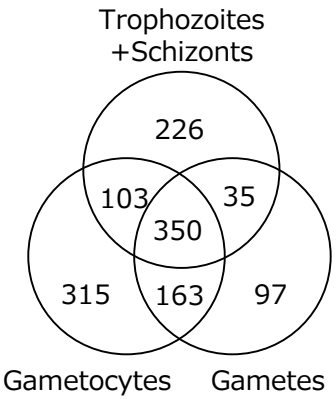


Nature, 2002, **419**, 537-541

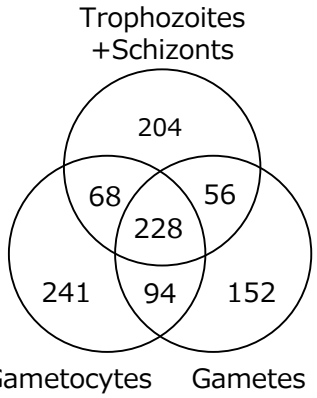
Identified proteins



Malaria proteins
Total unique: 1289



human proteins
Total unique: 1043



Most known merozoite surface proteins identified.

ca. 200 hypothetical proteins with transmembrane domains

New vaccine candiates

Nature, 2002, **419**, 537-542

Nature: Malaria special issue in 2002

articles

A proteomic view of the *Plasmodium falciparum* life cycle

Laurence Florens*, Michael P. Washburn†, J. Dale Raine‡, Robert M. Anthony§, Munira Grainger‖, J. David Haynes§¶, J. Kathleen Moch§, Nemone Muster*, John B. Sacci§#, David L. Tabb*☆, Adam A. Witney§#, Dirk Wolters†#, Yimin Wu**, Malcolm J. Gardner††, Anthony A. Holder‖, Robert E. Sinden‡, John R. Yates*† & Daniel J. Carucci§

* Department of Cell Biology, The Scripps Research Institute, SR-11, 10550 North Torrey Pines Road, La Jolla, California 92037, USA
† Department of Proteomics and Metabolomics, Torrey Mesa Research Institute, Syngenta Research & Technology, 3115 Merryfield Row, San Diego, California 92121-1125, USA
‡ Infection and Immunity Section, Department of Biological Sciences, Imperial College of Science, Technology & Medicine, Sir Alexander Fleming Building, South Kensington, London SW7 2AZ, UK
§ Naval Medical Research Center, Malaria Program (IDD), 503 Robert Grant Avenue, Room 3A40; and ¶ Department of Immunology, Walter Reed Army Institute of Research, Silver Spring, Maryland 20910-7500, USA
‖ The Division of Parasitology, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK
☆ Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA
** Malaria Research and Reference Reagent Resource Center, American Type Culture Collection, 10801 University Boulevard, Manassas, Virginia 20110-2209, USA
†† The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, Maryland 20850, USA

The completion of the *Plasmodium falciparum* clone 3D7 genome provides a basis on which to conduct comparative proteomics studies of this human pathogen. Here, we applied a high-throughput proteomics approach to identify new potential drug and vaccine targets and to better understand the biology of this complex protozoan parasite. We characterized four stages of the parasite life cycle (sporozoites, merozoites, trophozoites and gametocytes) by multidimensional protein identification technology. Functional profiling of over 2,400 proteins agreed with the physiology of each stage. Unexpectedly, the antigenically variant proteins of *var* and *rif* genes, defined as molecules on the surface of infected erythrocytes, were also largely expressed in sporozoites. The detection of chromosomal clusters encoding co-expressed proteins suggested a potential mechanism for controlling gene expression.

So sad.....😢😢

Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry

Edwin Lasonder*†, Yasushi Ishihama*, Jens S. Andersen*, Adriaan M. W. Vermunt†, Arnab Pain‡, Robert W. Sauerwein§, Wijnand M. C. Eling§, Neil Hall‡, Andrew P. Waters‖, Hendrik G. Stunnenberg† & Matthias Mann*

* Center for Experimental BioInformatics, Department of Biochemistry and Molecular Biology, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark
† Department of Molecular Biology, NCMLS, University of Nijmegen, Geert Grooteplein 26, 6525 GA Nijmegen, The Netherlands
‡ The Wellcome Trust Sanger Institute, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK
§ Department of Medical Microbiology, NCMLS, University Medical Centre, PO Box 9101, 6500 HB Nijmegen, The Netherlands
‖ Leiden Malaria Research Group, Department of Parasitology, Centre for Infectious Disease, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, The Netherlands

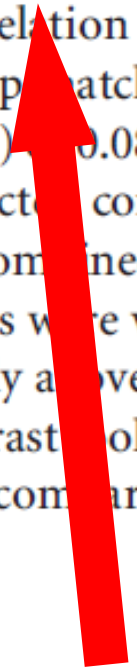
The annotated genomes of organisms define a 'blueprint' of their possible gene products. Post-genome analyses attempt to confirm and modify the annotation and impose a sense of the spatial, temporal and developmental usage of genetic information by the

letters to nature

organism. Here we describe a large-scale, high-accuracy (average deviation less than 0.02 Da at 1,000 Da) mass spectrometric proteome analysis[1–3] of selected stages of the human malaria parasite *Plasmodium falciparum*. The analysis revealed 1,289 proteins of which 714 proteins were identified in asexual blood stages, 931 in gametocytes and 645 in gametes. The last two groups provide insights into the biology of the sexual stages of the parasite, and include conserved, stage-specific, secreted and membrane-associated proteins. A subset of these proteins contain domains that indicate a role in cell–cell interactions, and therefore can be evaluated as potential components of a malaria vaccine formulation. We also report a set of peptides with significant matches in the parasite genome but not in the protein set predicted by computational methods.

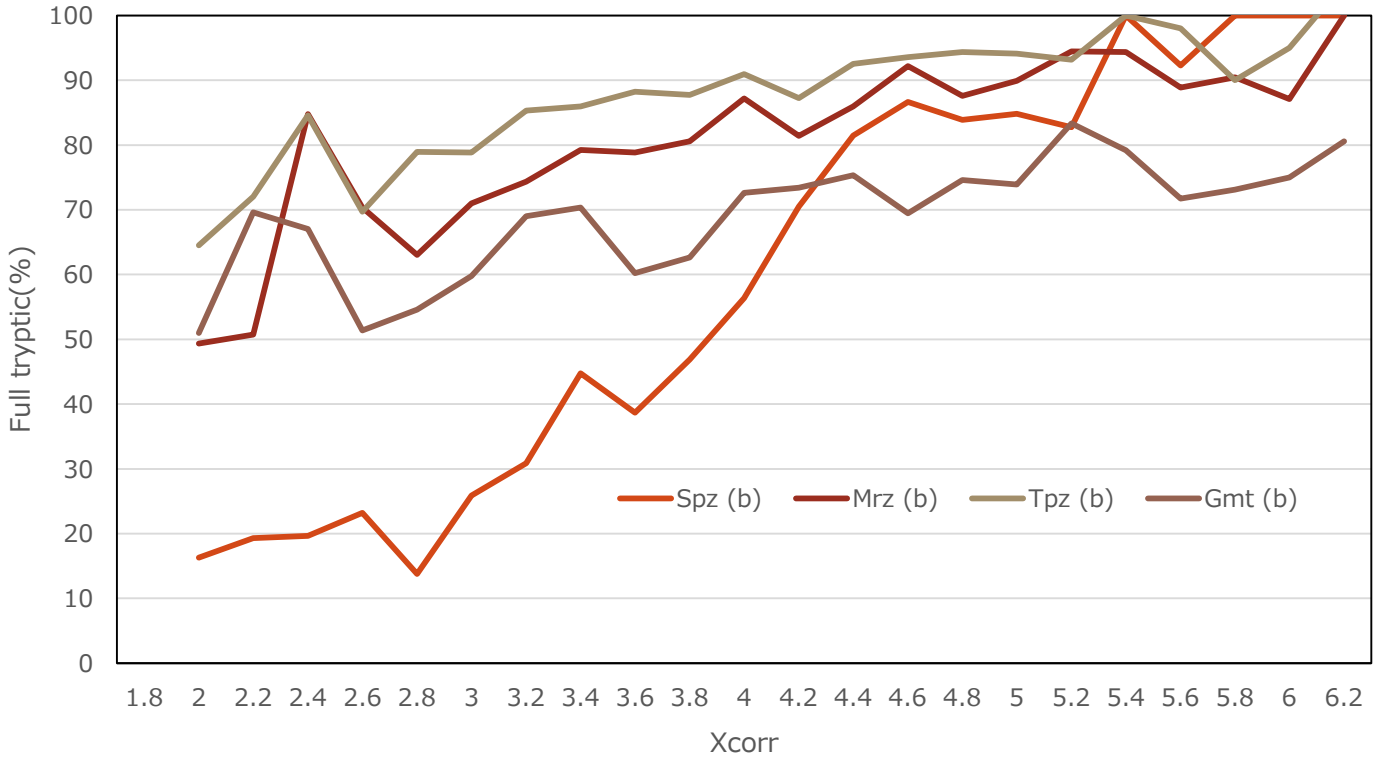
MS/MS data set analysis

The SEQUEST algorithm was used to match MS/MS spectra to peptides in the sequence databases[41]. To account for carboxyamidomethylation, MS/MS data sets were searched with a relative molecular mass of 57,000 (M_r, 57K) added to the average molecular mass of cysteines. Peptide hits were filtered and sorted with DTASelect[42]. Spectra/peptide matches were only retained if they were at least half-tryptic (Lys or Arg at either end of the identified peptide) and with minimum cross-correlation scores (XCorr) of 1.8 for +1, 2.5 for +2, and 3.5 for +3 spectra and DeltaCn (top match's XCorr minus the second-best match's XCorr divided by the top match's XCorr) of 0.08. Peptide hits were deemed unambiguous only if they were not found in non-infected controls and were uniquely assigned to parasite proteins by searching against combined parasite–host databases. Finally, for low coverage loci, peptide/spectrum matches were visually assessed on two main criteria: any given MS/MS spectrum had to be clearly above the baseline noise, and both *b* and *y* ion series had to show continuity. The Contrast tool[42] was used to compare and merge protein lists from replicate sample runs and to compare the proteomes established for the four stages.



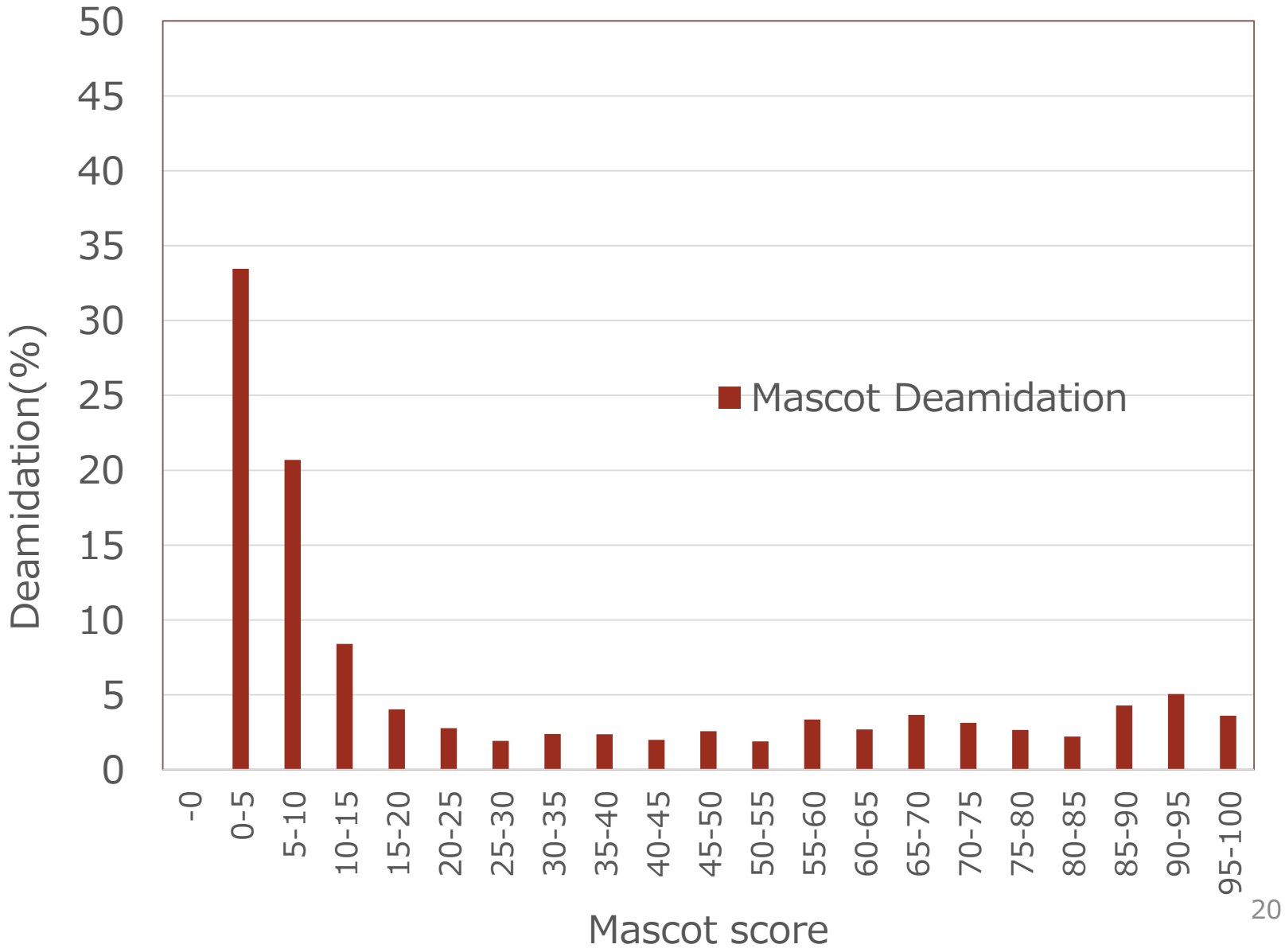
Umm, too wide search space??

Score distribution of fully tryptic peptides (%)

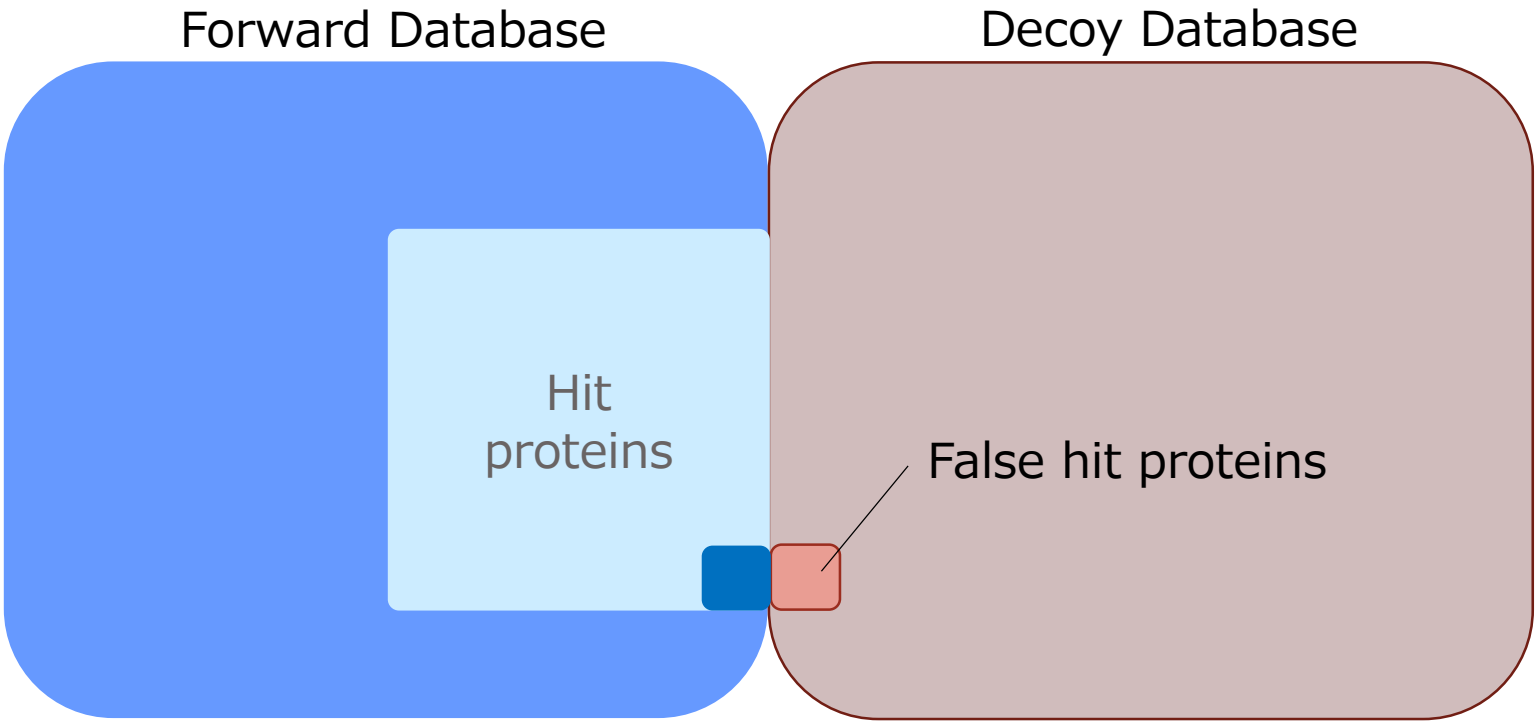


This should be prohibited!!

Variable mod: De-amidation (N, Q)



Target-Decoy Approach for estimating FDR(%)



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

Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry

Joshua E Elias[1] & Steven P Gygi[1,2]

NATURE METHODS | VOL.4 NO.3 | MARCH 2007 | 207

Rather than deciding exactly which peptide-spectral matches (PSMs) are correct or incorrect, the composite target-decoy database evaluates FP rates in large PSM populations. It permits estimation of the likelihood that a PSM is correct given that it came from a collection of PSMs with a measured FP rate. This is not to suggest that the search strategy removes all false identifications. Instead, the target-decoy approach allows the estimation of how many FP are associated with an entire data set.

Target-decoy search for all merged data

Juergen Cox (Max Planck)

Nature. 2014 , DOI: 10.1038/nature13302, PMID: 24870542

A draft map of the human proteome

Min-Sik Kim; Sneha M Pinto; Derese Getnet; Raja Nirujogi; Srikanth S Manda; Raghothama Chaerkady; Anil K Madugundu; Dhanashree S Kelkar; Ruth Isserlin; Shobhit Jain; Joji K Thomas; Babylakshmi Muthusamy; Pamela Leal-Rojas; Praveen Kumar; Nandini A Sahasrabuddhe; Lavanya Balakrishnan; Jayshree Advani; Bijesh George; Santosh Renuse; Lakshmi N Selvan; Arun H Patil; Vishalakshi Nanjappa; Aneesha Radhakrishnan; Samarjeet Prasad; Tejaswini Subbannayya; Rajesh Raju; Manish Kumar; Sreelakshmi K Sreenivasamurthy; Arivusudar Marimuthu; Gajanan J Sathe; Sandip Chavan; Keshava K Datta; Yashwanth Subbannayya; Apeksha Sahu; Soujanya D Yelamanchi; Savita Jayaram; Pavithra Rajagopalan; Jyoti Sharma; Krishna R Murthy; Nazia Syed; Renu Goel; Aafaque A Khan; Sartaj Ahmad; Gourav Dey; Keshav Mudgal; Aditi Chatterjee; Tai-Chung Huang; Jun Zhong; Xinyan Wu; Patrick G Shaw; ... (22 more)

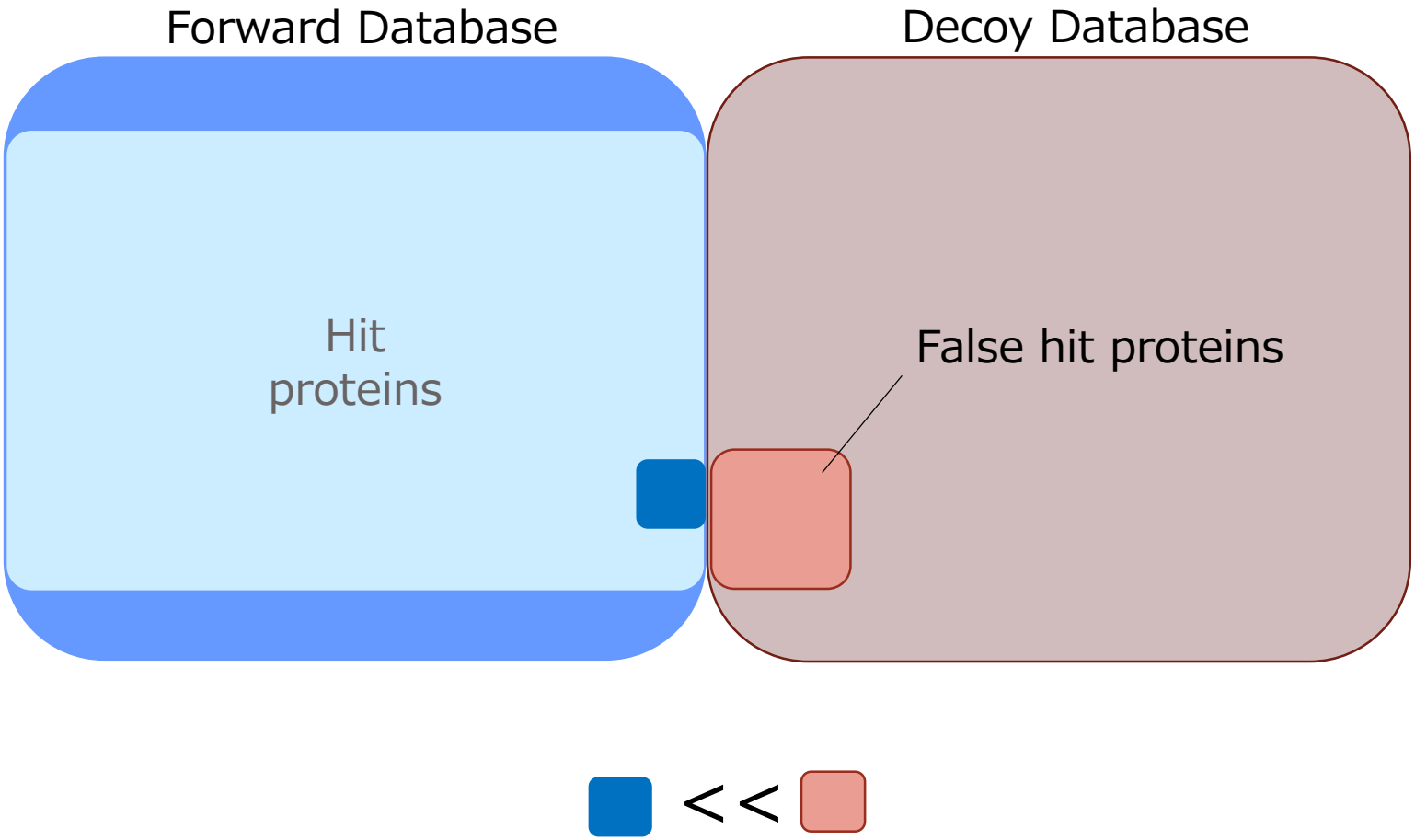
The availability of human genome sequence has transformed biomedical research over the past decade. However, an equivalent map for the human proteome with direct measurements of proteins and peptides does not exist yet. Here we present a draft map of the human proteome using high-resolution Fourier-transform mass spectrometry. In-depth proteomic profiling of 30 histologically normal human samples, including 17 adult tissues, 7 fetal tissues and 6 purified primary haematopoietic cells, resulted in identification of proteins encoded by 17,294 genes accounting for approximately 84% of the total annotated protein-coding genes in humans. A unique and comprehensive strategy for

17,294 genes (84%)

↓ Target-decoy search for all merged data
1% FDR at protein level

11,206 genes (57%)

Target-Decoy Approach for Ultra Large Datasets



ProteomicsDB; self-corrected

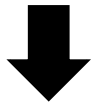
Molecular & cellular proteomics : MCP. 2015 , DOI: 10.1074/mcp.M114.046995, PMID: 25987413

A scalable approach for protein false discovery rate estimation in large proteomic data sets.

Mikhail M Savitski; Mathias Wilhelm; Hannes Hahne; Bernhard Kuster; Marcus Bantscheff

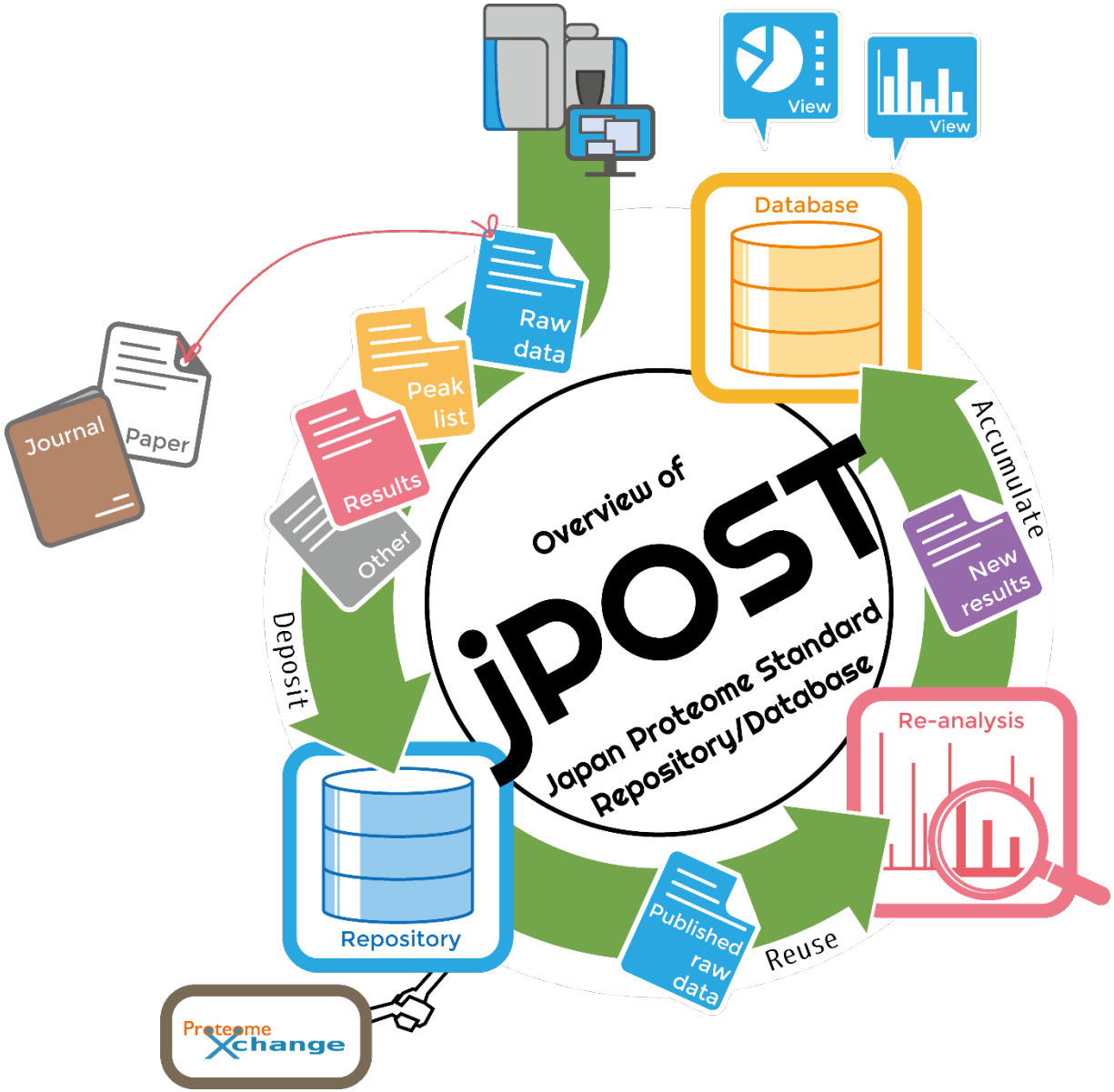
Calculating the number of confidently identified proteins and estimating false discovery rate (FDR) is a challenge when analyzing very large proteomic datasets such as entire human proteomes. Biological and technical heterogeneity in proteomic experiments further add to the challenge and there are strong differences in opinion regarding the conceptual validity of a protein FDR and no consensus regarding the methodology for protein FDR determination. There are also limitations inherent to the widely used classic target-decoy strategy (TDS) that particularly show when analyzing very large data sets and that lead to a strong over-representation of decoy identifications. In this study, we investigated the merits of the decoy-based protein FDR estimation approach taking advant collection comprised of ≈19,000 LC-MS/MS runs deposited i www.proteomicsdb.org). The "picked" protein FDR approach the same protein as a pair rather than as individual entities a decoy sequence depending on which receives the highest sc of this approach in combination with q-value based peptide instrument and search engine-specific differences. The "pick best when protein scoring was based on the best peptide q- stable number of true positive protein identifications over a demonstrate that this simple and unbiased strategy eliminat commonly used, "classic" protein FDR approach that causes protein identification in large data sets. The approach scales without losing performance, consistently increases the number of true positive protein identifications and is readily implemented in proteomics analysis software.

18,097 proteins (original)



for this difference (supplemental Figure 7). We next applied the described data analysis strategy to the subset of data stored in proteomicsDB corresponding to our earlier publication on a mass spectrometry based draft of the human proteome (9). Using the classic FDR strategy 14,035 proteins were observed at 1% protein FDR compared to 14,714 proteins using the picked strategy. Applying the picked strategy without any protein score threshold yielded 17,326 proteins of the target database at 11.3% protein FDR corresponding to 15,290 true positive protein identifications in the dataset. When analyzing the complete current content of proteomicsDB (including the data of the Pandey proteome (10) and a number of further datasets), the number of protein identifications at 1% FDR increased to 14,638 (classic) and 15,375 proteins (picked) respectively.

jPOST re-analysis



Universal index for annotated MSMS spectra

How can we merge the results from different sources?

jPOST score

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

Sequence Query

Introduction

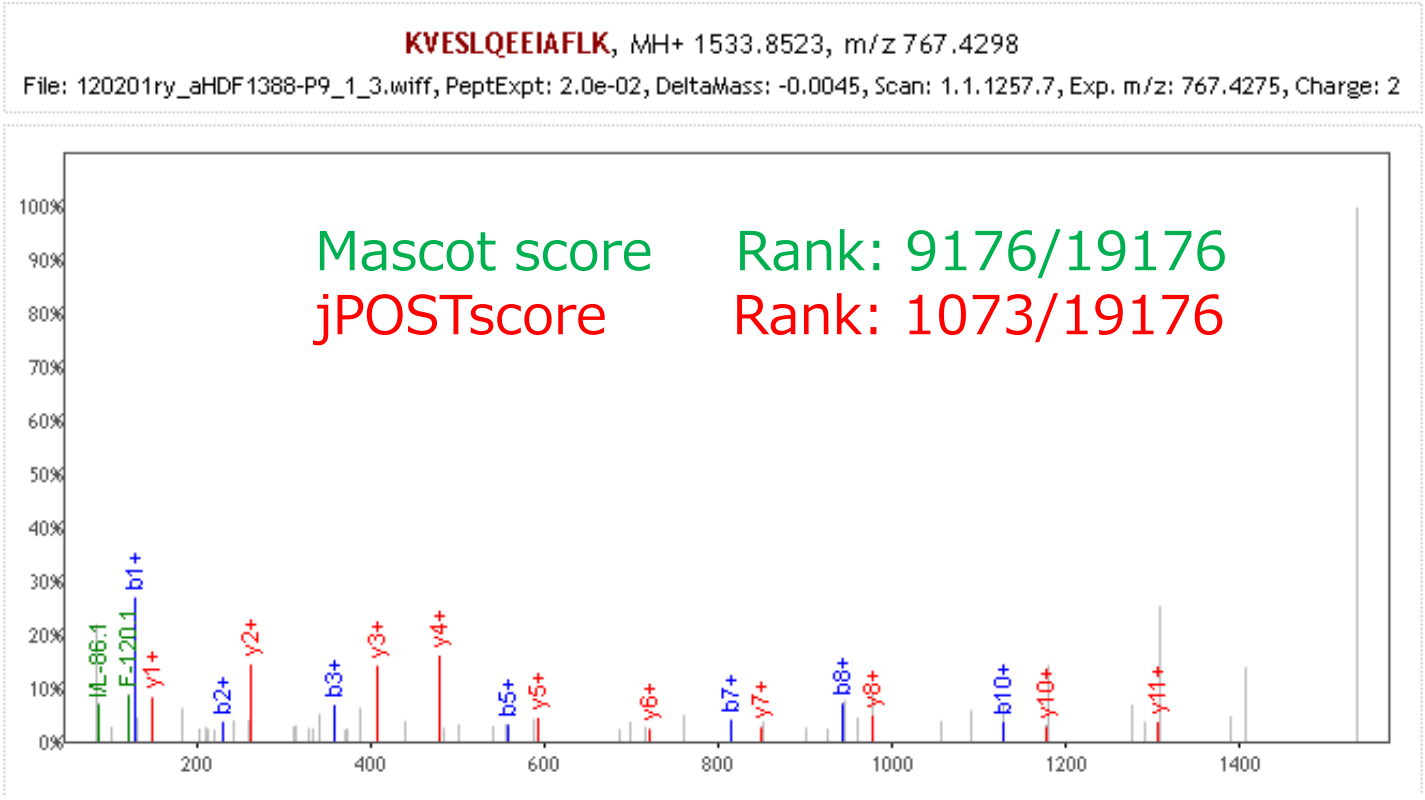
The sequence query, in which one or more peptide molecular masses are combined with sequence, composition and fragment ion data, is potentially the most powerful search of all. The usual source of the sequence information is interpretation of an MS/MS spectrum. While it is very difficult to determine a complete and unambiguous peptide sequence from an MS/MS spectrum, it is often possible to find a series of peaks providing 3 or 4 residues of reliable sequence data.

This general approach was pioneered by Mann and co-workers at EMBL, who used the term "sequence tag" for the combination of a few residues of sequence data combined with molecular weight information [Mann, 1994]. They defined a sequence tag derived from an MS/MS spectrum as the mass of the precursor peptide, the mass of the first peak of the identified sequence ladder, a stretch of interpreted sequence, and the mass of the final peak of the ladder.

PST-based jPOST SCORE

jPOST score

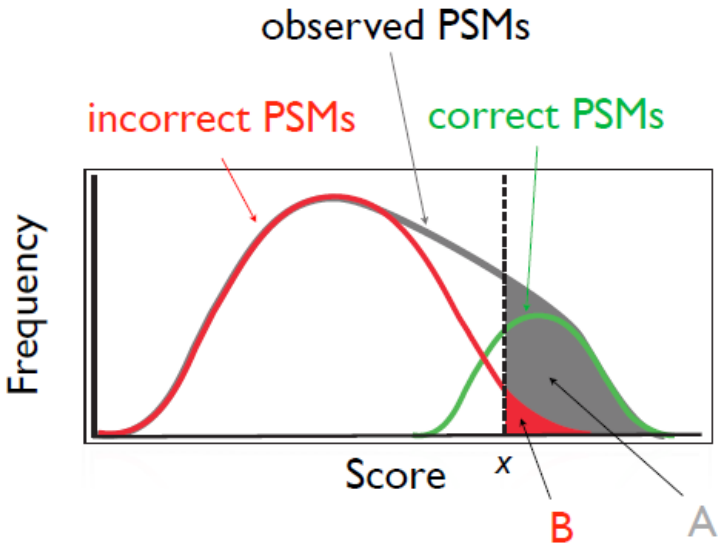
1. b, y-ion coverage
2. tag length
3. uncovered length



b+	#	Seq	#	y+
129.1022	1	K	13	
228.1707	2	V	12	1405.7573
357.2132	3	E	11	1306.6889
444.2453	4	S	10	1177.6463
557.3293	5	L	9	1090.6143
685.3879	6	Q	8	977.5302
814.4305	7	E	7	849.4716
943.4731	8	E	6	720.4291
1056.5572	9	I	5	591.3865
1127.5943	10	A	4	478.3024
1274.6627	11	F	3	407.2653
1387.7468	12	L	2	260.1969
	13	K	1	147.1128

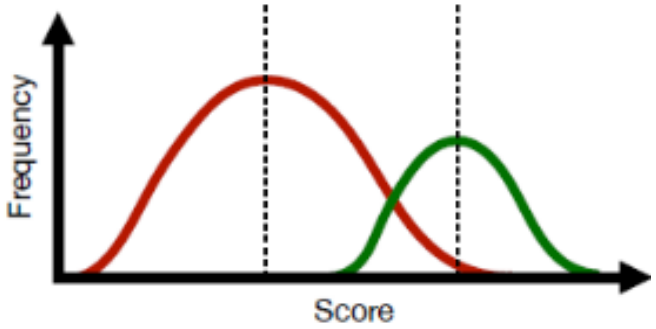
[Click] to move table

False Discovery Rate

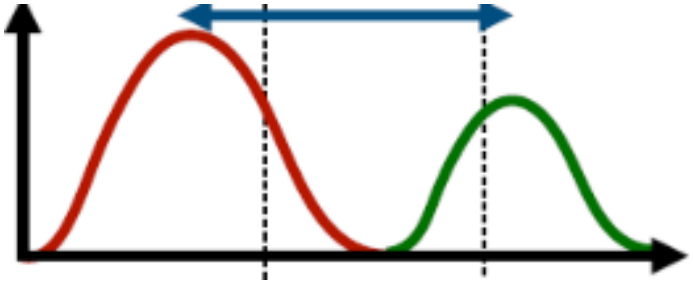


$$FDR(x) = \frac{B}{A}$$

FDR(x) is the expectation value of the fraction of detections above threshold *x* that are incorrect

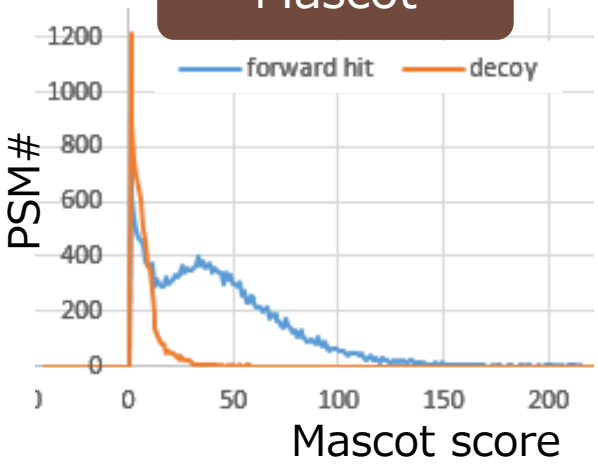


Improve the score function

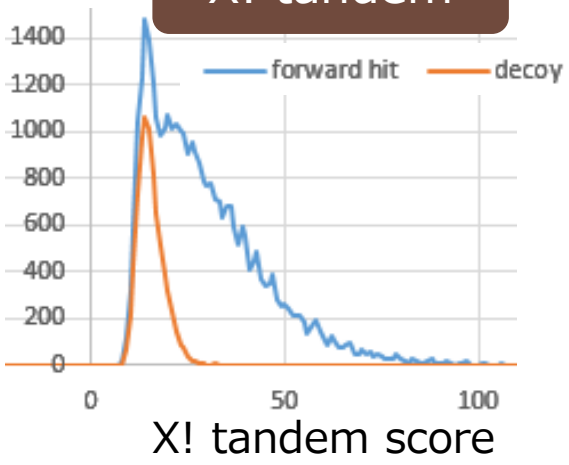


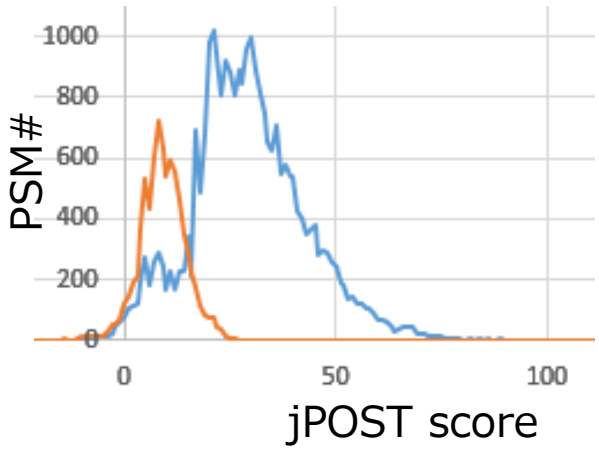
Lukas Kall, 2017

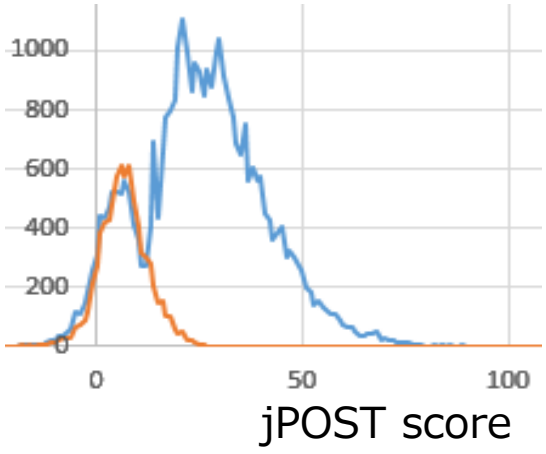
Mascot



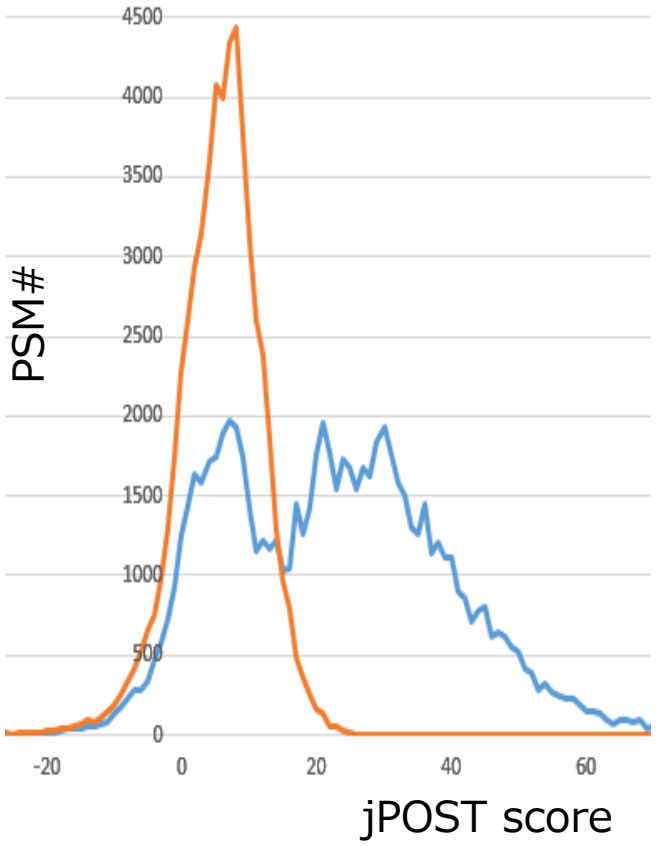
X! tandem







Mascot + X!tandem + Comet + MQ



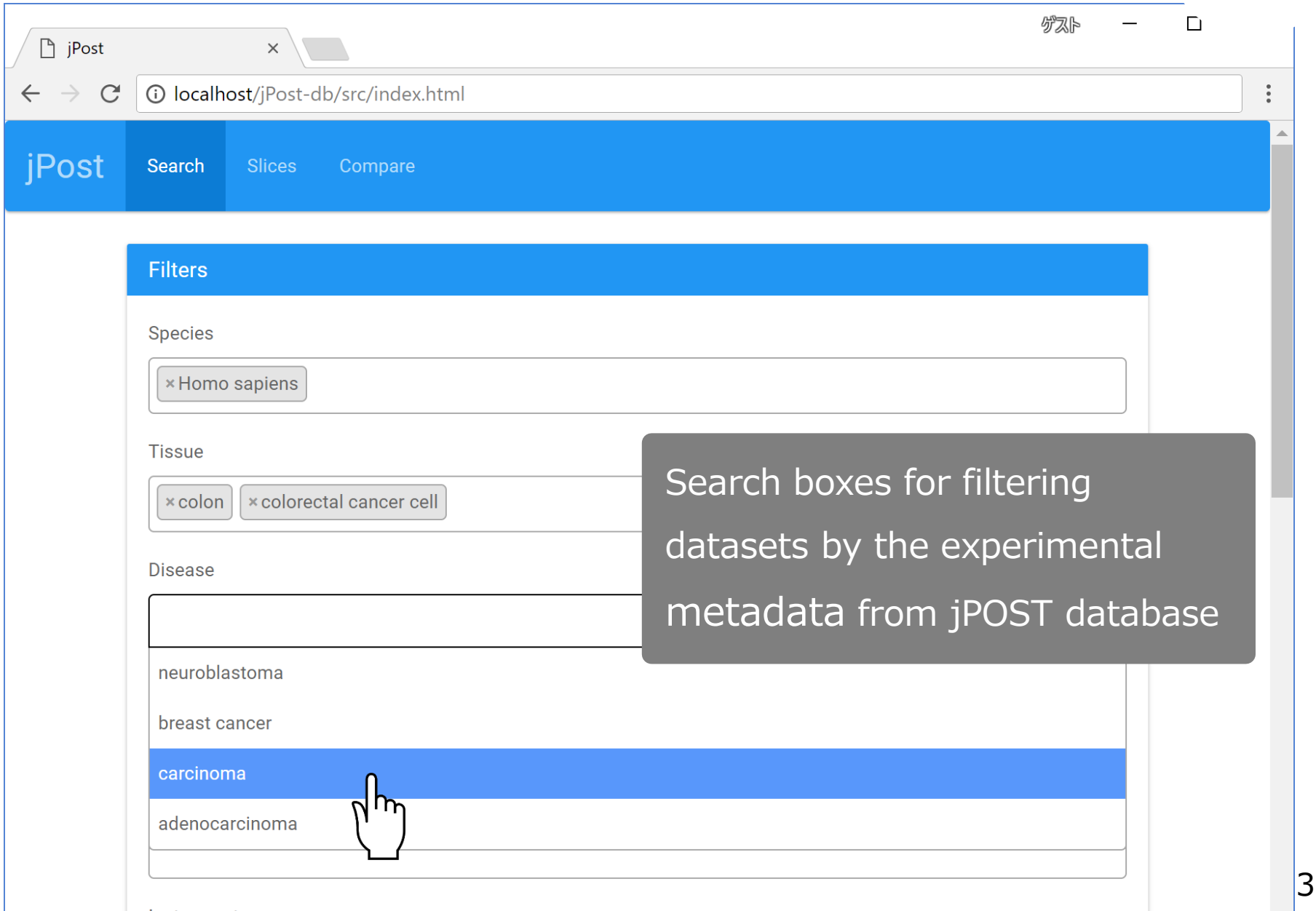
Dataset:PXD005159

Tryptic peptides from human HeLa cells

by Thermo Q-Exactive

jPOST customizable database 'Slice'

jPOST slice database



jPOST slice database

iPS colon +

Dataset Protein

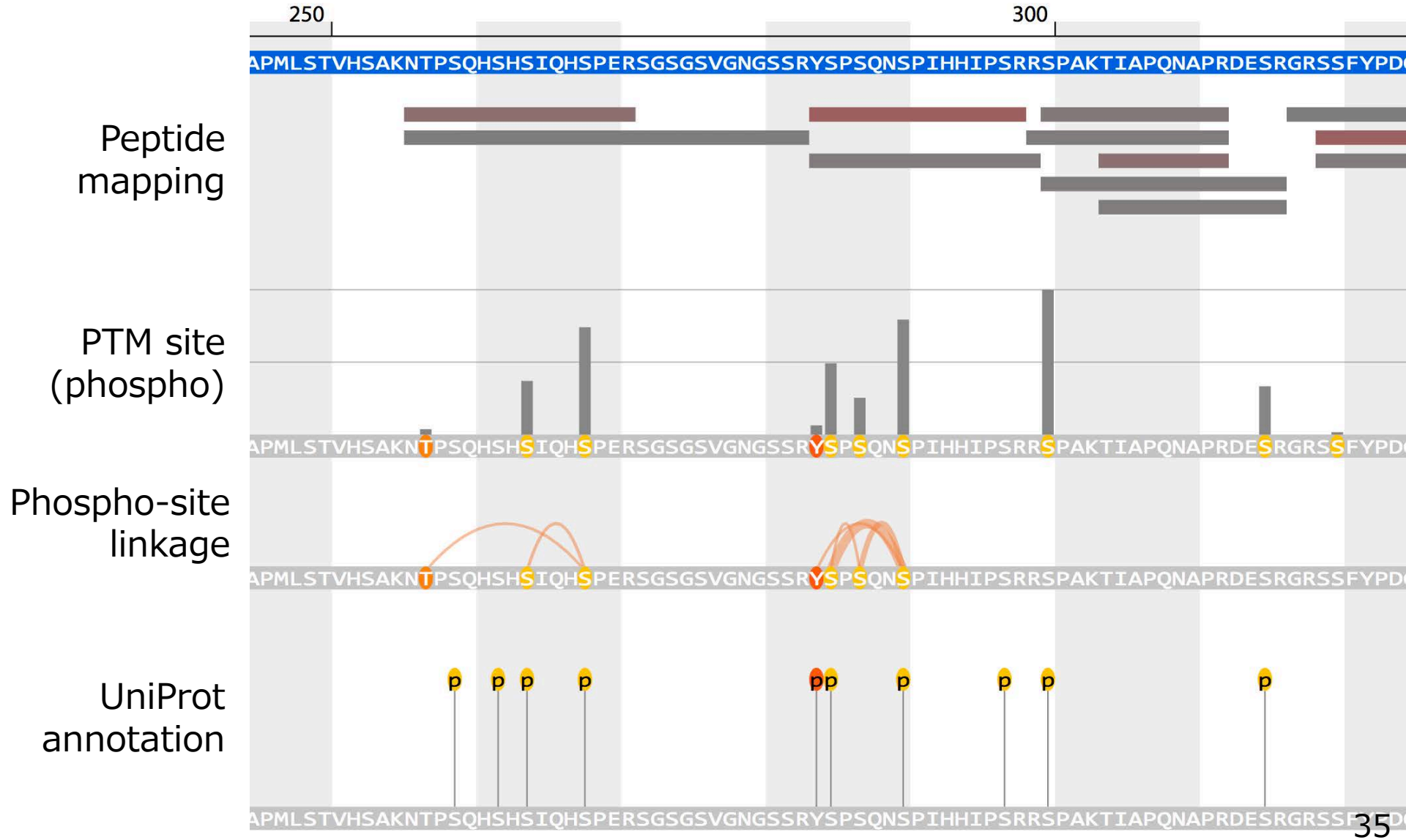
Selected dataset list called ' **slice** '

Showing 1 to 6 of 6 entries

ID	Project ID	Project Title	Project Date
DS203_1	JPST000203	Quantitative proteomics of colorectal cancer tissues	2016-10-18
DS204_1	JPST000204	Quantitative phosphoproteomics of colorectal cancer tissues	2016-10-18
DS205_1	JPST000205	Proteomic data of HCT116 cells	2016-10-18
DS206_1	JPST000206	Phosphoproteomic data of HCT116 cells	2016-10-18
DS210_1	JPST000210	Phosphoproteomics data of colon tissues (tumor and non-tumor)	2016-10-18
DS210_2	JPST000210	Phosphoproteomics data of colon tissues (tumor and non-tumor)	2016-10-18

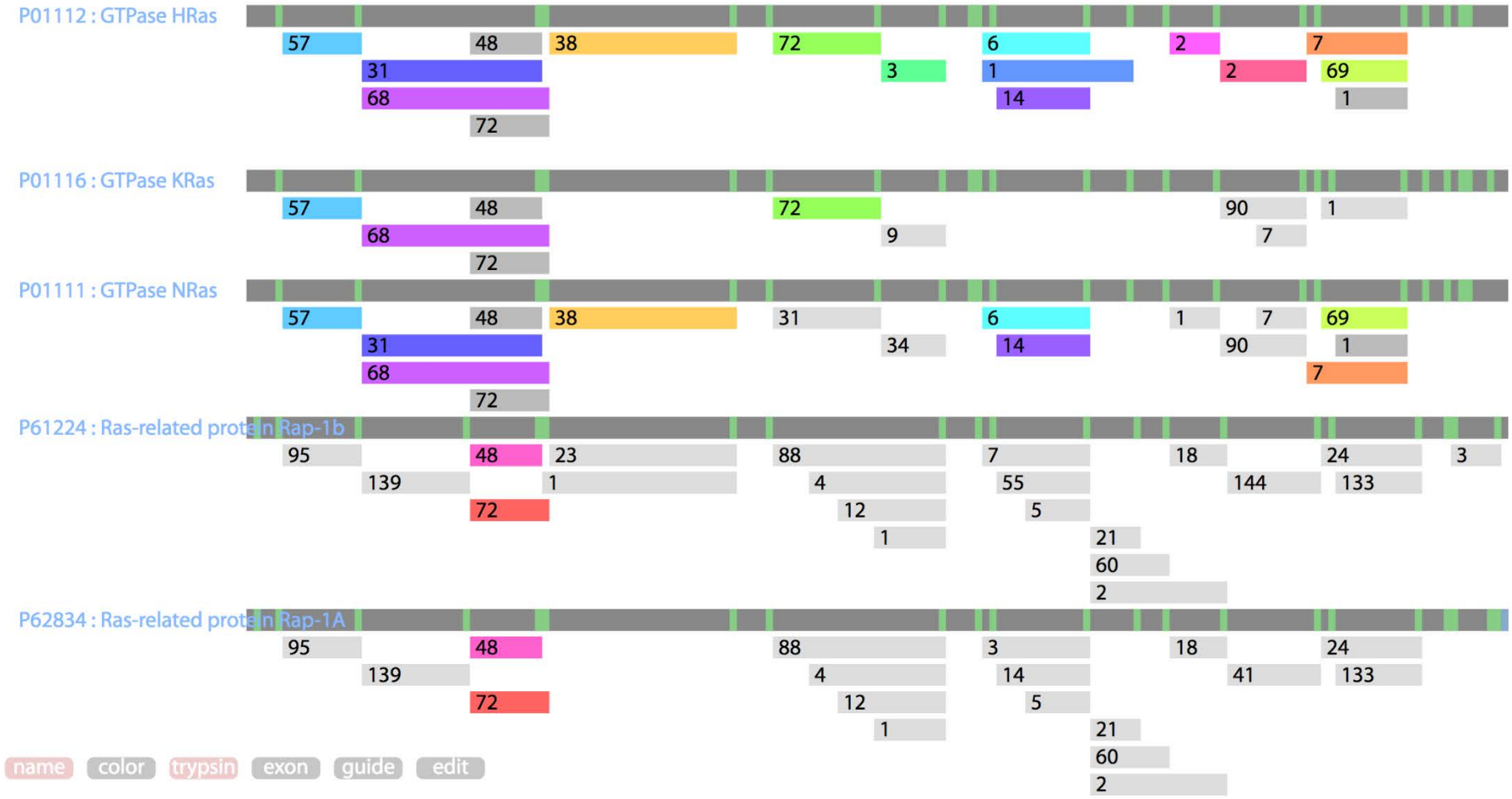
jPOST slice database

Protein browser



jPOST slice database

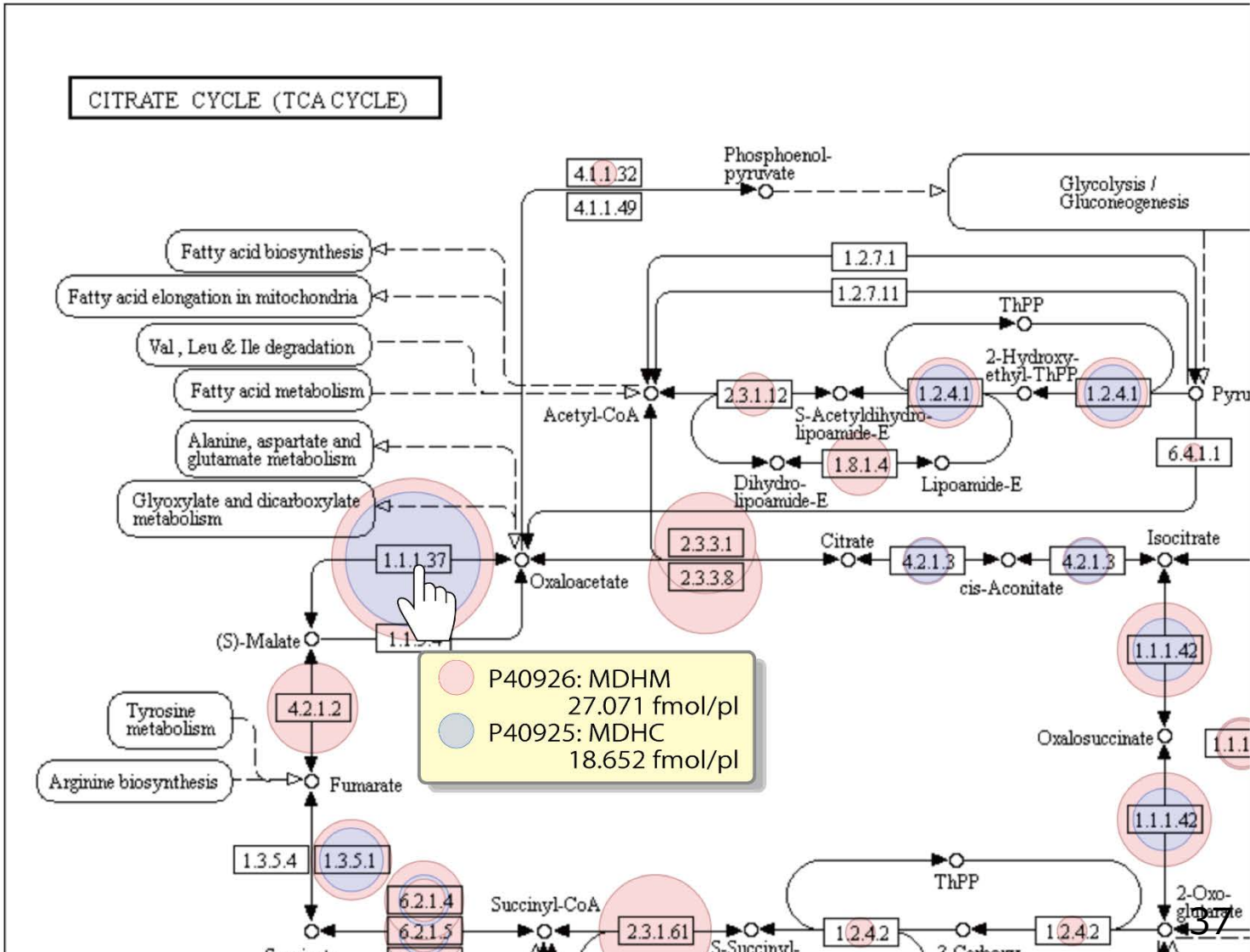
Proteoform browser shows peptide sharing



jPOST database

KEGG pathway mapping with absolute quantitative value

- Metabolism
  - Carbohydrate metabolism
    - Glycolysis / Gluconeogenesis
    - Amino sugar and nucleotide sugar metabolism
    - Pyruvate metabolism
    - Inositol phosphate metabolism
    - Citrate cycle (TCA cycle)
    - Propanoate metabolism
    - Fructose and mannose metabolism
    - Pentose phosphate pathway
    - Glyoxylate and dicarboxylate metabolism
    - Galactose metabolism
    - Starch and sucrose metabolism
    - Butanoate metabolism
    - Pentose and glucuronate interconversions
    - Ascorbate and aldarate metabolism
  - Amino acid metabolism
    - Valine, leucine and isoleucine degradation
    - Cysteine and methionine metabolism
    - Lysine degradation
    - Arginine and proline metabolism
    - Glycine, serine and threonine metabolism
    - Alanine, aspartate and glutamate metabolism
    - Tryptophan metabolism



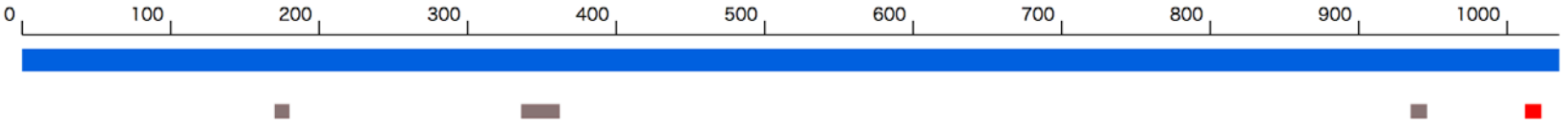
Missing protein search

using latest neXtprot & peptide uniqueness checker

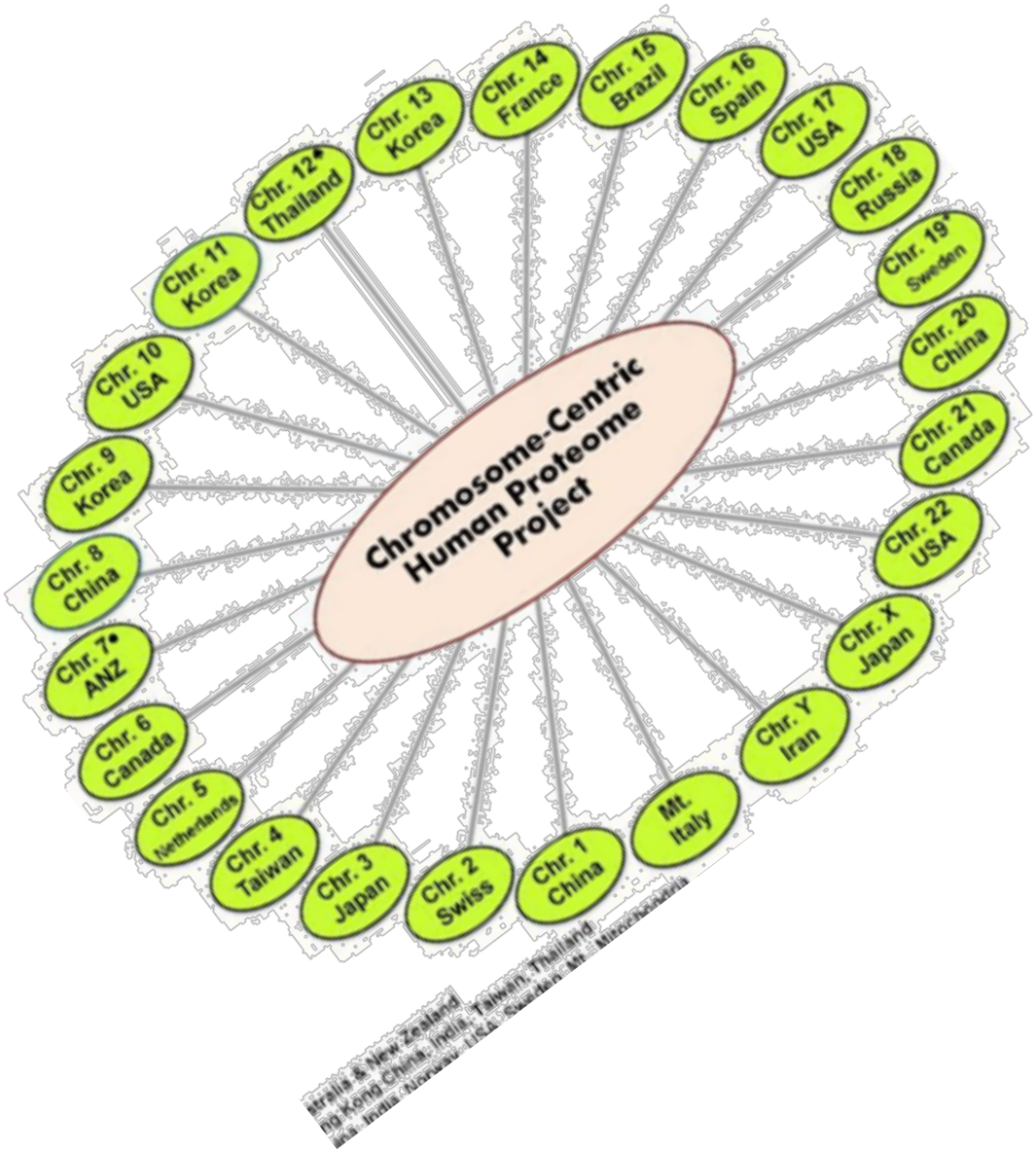
chromosome: X
protein evidence: 2-4
peptide length: >=9
number of peptide: >=2
unique peptide: has unique

PE	Chromosome	UniProt	Gene Symbol	Name	#Peptode	#Unique Peptide
3	X	A6NGH7	CC160_HUMAN	Coiled-coil domain-containing protein 160	2	2
2	X	Q5HY64	FA47C_HUMAN	Putative protein FAM47C	4	3
2	X	Q5HYW3	RGAG4_HUMAN	Retrotransposon gag domain-containing protein 4	3	3
2	X	Q6PI77	BHLH9_HUMAN	Protein BHLH	3	3
2	X	Q8IZF6	AGRG4_HUMAN	Adhesion G-protein coupled receptor G4	2	2
2	X	Q8N7E2	ZN645_HUMAN	E3 ubiquitin-protein ligase ZNF645	2	2

1 - 6 / 6 1



jPOST meets C-HPP



Chr No.	Leader
Chr. 1	Ping Xu
Chr. 2	Lydie Lane
Chr. 3	Takashi Kawamura
Chr. 4	Yu Ju Chen
Chr. 5	Peter Horvatovich
Chr. 6	Christoph Borchers
Chr. 7	Edouard Nice
Chr. 8	Pengyuan Yang
Chr. 9	Je-Yoel Cho
Chr.10	Joshua Labaer
Chr.11	Jong Shin Yoo
Chr.12	Ravi Sirdeshmukh
Chr.13	Young-Ki Paik
Chr.14	Charles Pineau
Chr.15	Gilberto B. Domont
Chr.16	Fernando Corrales
Chr.17	Gilbert S. Omenn
Chr.18	Alexander Archakov
Chr.19	György Marko-Varga
Chr.20	Siqi Liu
Chr.21	Albert Sickmann
Chr.22	Akhileshi Pandey
Chr. X	Yasushi Ishihama
Chr. Y	Ghasem Hosseini Salekdeh
Mitochondria	Andrea Urbani

Hunting Missing Proteins using jPOST

Journal of proteome research

Article

pubs.acs.org/jpr

Rapid and Deep Profiling of Human Induced Pluripotent Stem Cell Proteome by One-shot NanoLC–MS/MS Analysis with Meter-scale Monolithic Silica Columns

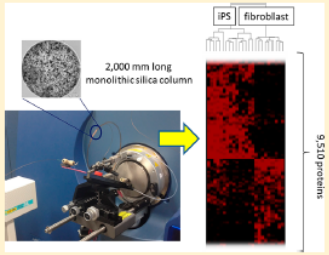
Ryota Yamana,[†,‡] Mio Iwasaki,[†,‡] Masaki Wakabayashi,[†] Masato Nakagawa,[§] Shinya Yamanaka,[§] and Yasushi Ishihama*[,†]

[†]Department of Molecular & Cellular BioAnalysis, Graduate School of Pharmaceutical Sciences, Kyoto University, Sakyo-ku, Kyoto 606-8501, Japan

[§]Center for iPS Cell Research and Application, Kyoto University, Sakyo-ku, Kyoto 606-8507, Japan

Supporting Information

ABSTRACT: Proteome analyses of human induced pluripotent stem cells (iPSC) were carried out on a liquid chromatography–tandem mass spectrometry system using meter-scale monolithic silica-C18 capillary columns without prefractionation. Tryptic peptides from five different iPSC lysates and three different fibroblast lysates (4 μg each) were directly injected onto a 200 cm long, 100 μm i.d. monolithic silica-C18 column and an 8-h gradient was applied at 500 nL/min at less than 20 MPa. We identified 98 977 nonredundant tryptic peptides from 9510 proteins (corresponding to 8712 genes), including low-abundance protein groups (such as 329 protein kinases) from triplicate measurements within 10 days. The obtained proteome profiles of the eight cell lysates were categorized into two groups, iPSC and fibroblast, by hierarchical cluster analysis. Further quantitative analysis based on an exponentially modified protein abundance index approach combined with UniProt keyword enrichment analysis revealed that the iPSC group contains more "transcription regulation"-related proteins, while the fibroblast group contained more "transport"-related proteins. Our results indicate that this simplified one-shot proteomics approach with long monolithic columns is advantageous for rapid, deep, sensitive, and reproducible proteome analysis.

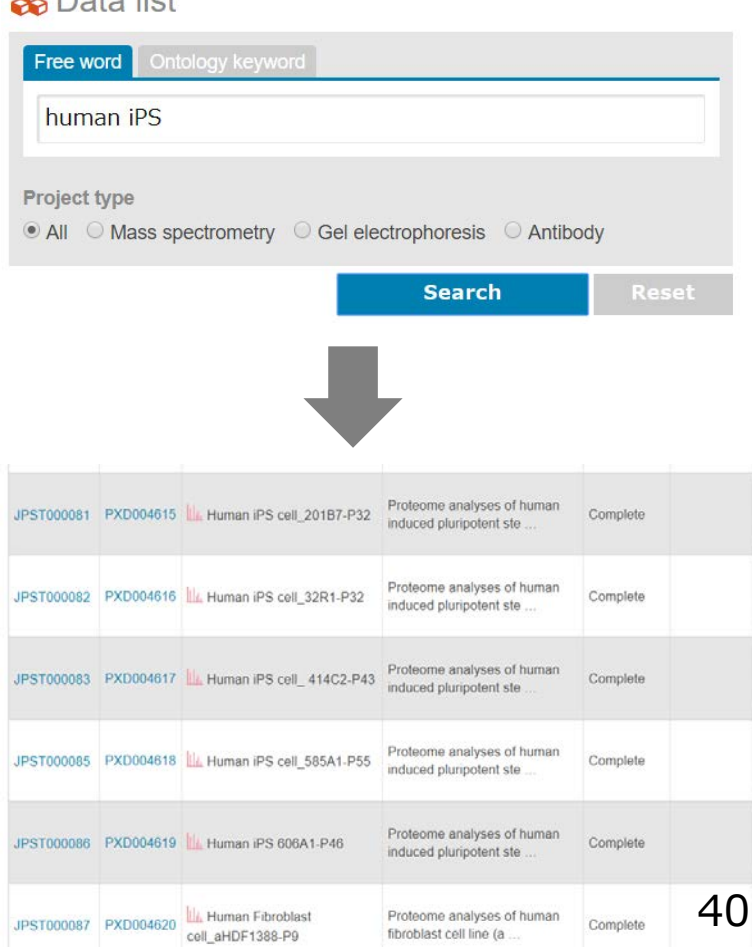


KEYWORDS: *shotgun proteomics, monolithic silica column, iPS cell, one-shot proteomics*

Special Issue: Chromosome-centric Human Proteome Project

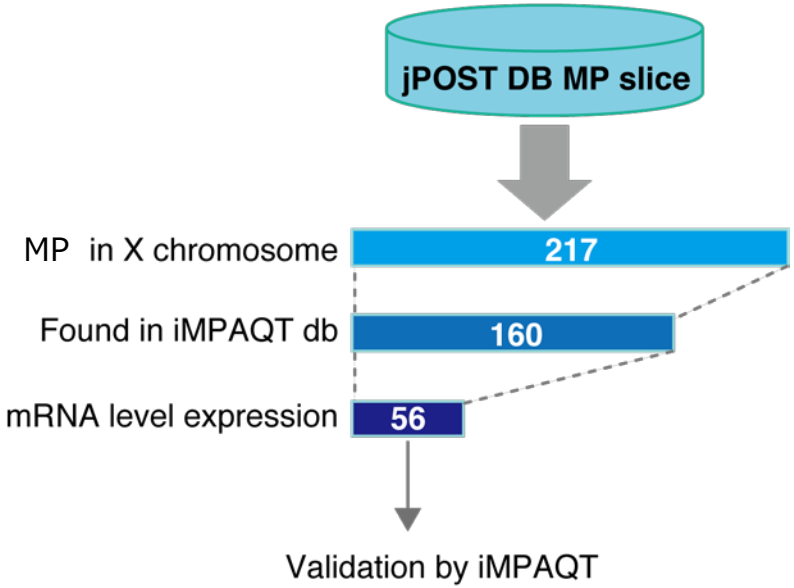
Received: September 2, 2012
Published: December 4, 2012

dx.doi.org/10.1021/pr300837u | *J. Proteome Res.* 2013, 12, 214–221

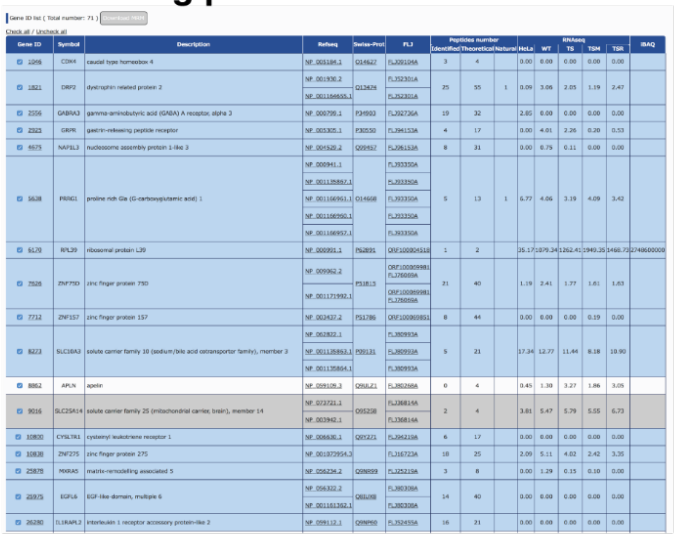


iMPAQT- Validation of Missing Proteins

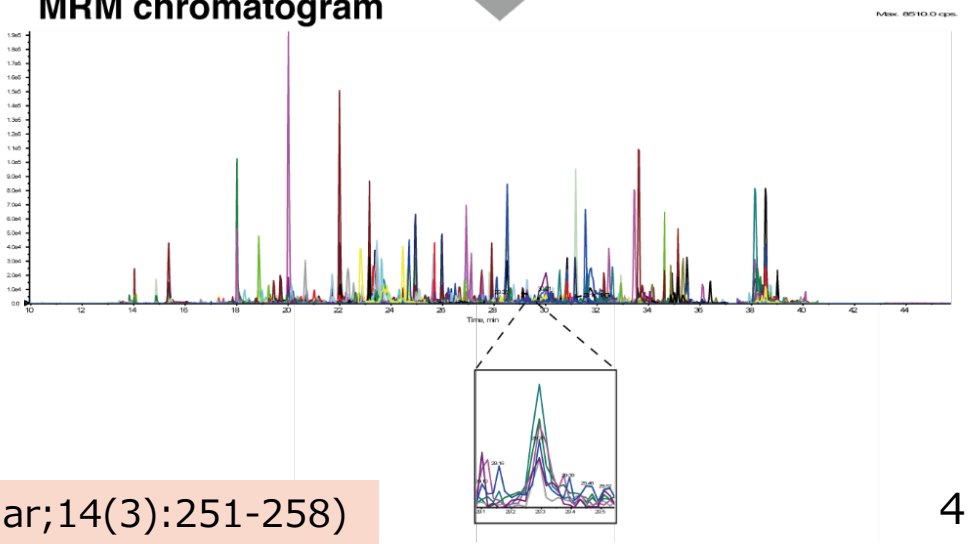
impaqt.jpost.org



Missing protein MRM transitions



MRM chromatogram



IMPAQT: Matsumoto et al., Nat Methods. 2017 Mar;14(3):251-258)

Conclusions

- The jPOST repository, re-analysis and database have been successfully developed.
  - The jPOST repository is a part of ProteomeXchange.
  - Re-analysis is based on jPOST score, independent of search engines.
  - The jPOST team is involved in HPP, missing protein “next 50 challenge” projects.
- The jPOST scoring could be extended to proteomic analysis with wider search space such as proteogenomics and metaproteomics.

JPrOS Japanese Proteomics Society

JPOST Repository / Database