



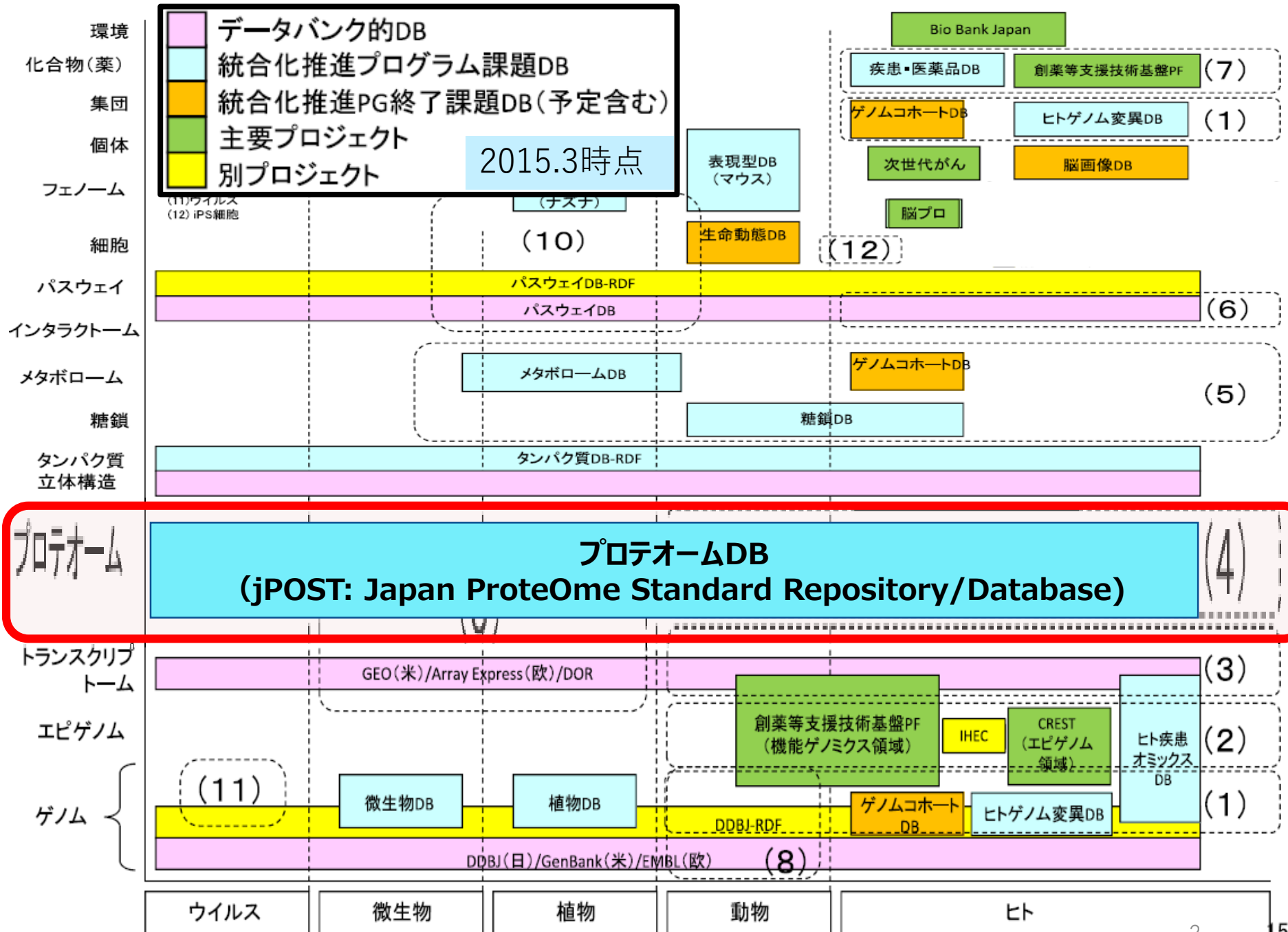
プロテオームデータベースの 機能深化と連携基盤強化

京大院薬 石濱 泰
& jPOST TEAM

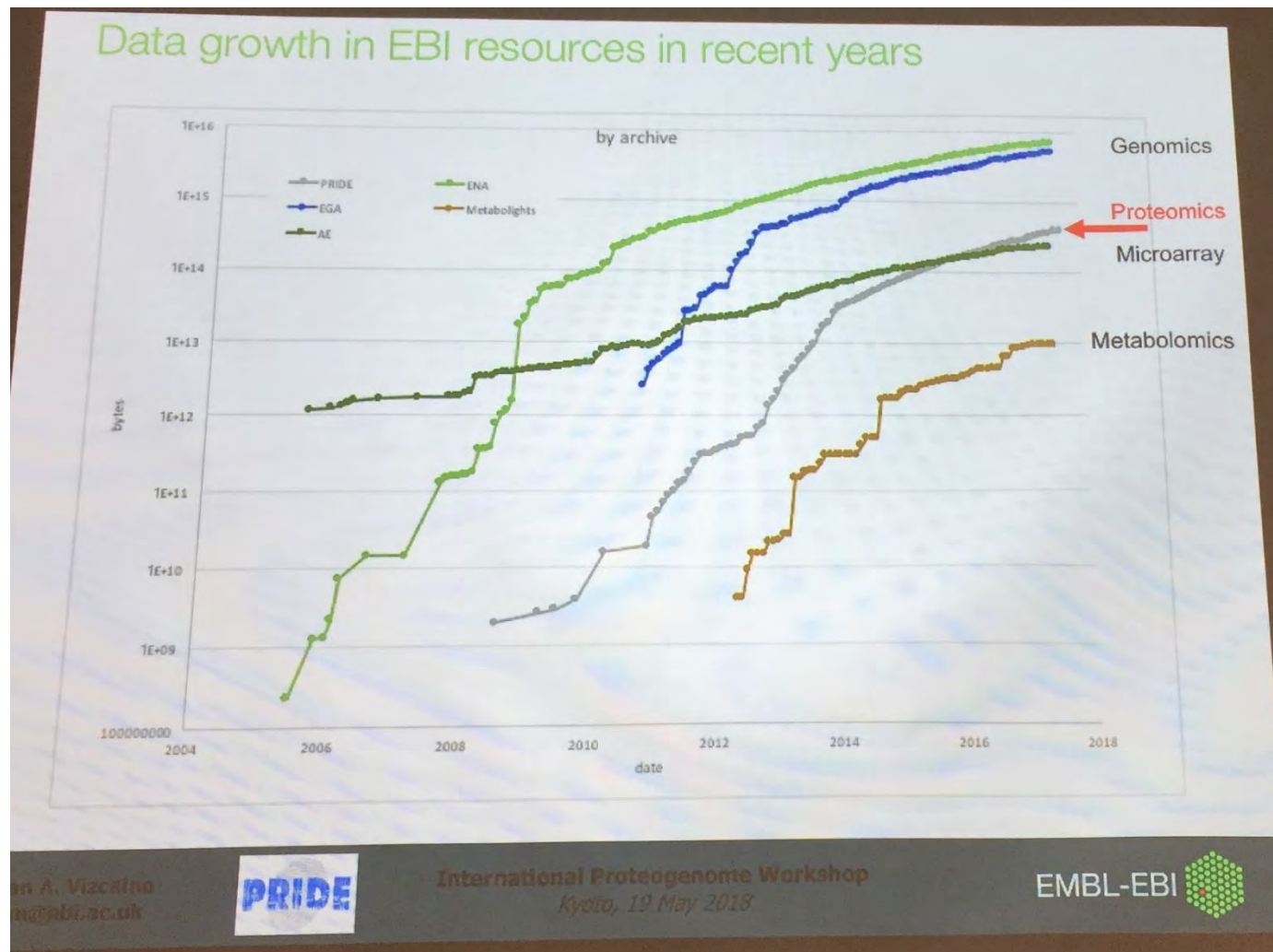


©2019 石濱 泰(京都大学) licensed under CC 表示 4.0 国際 ライセンス





EMBL-EBIでは？

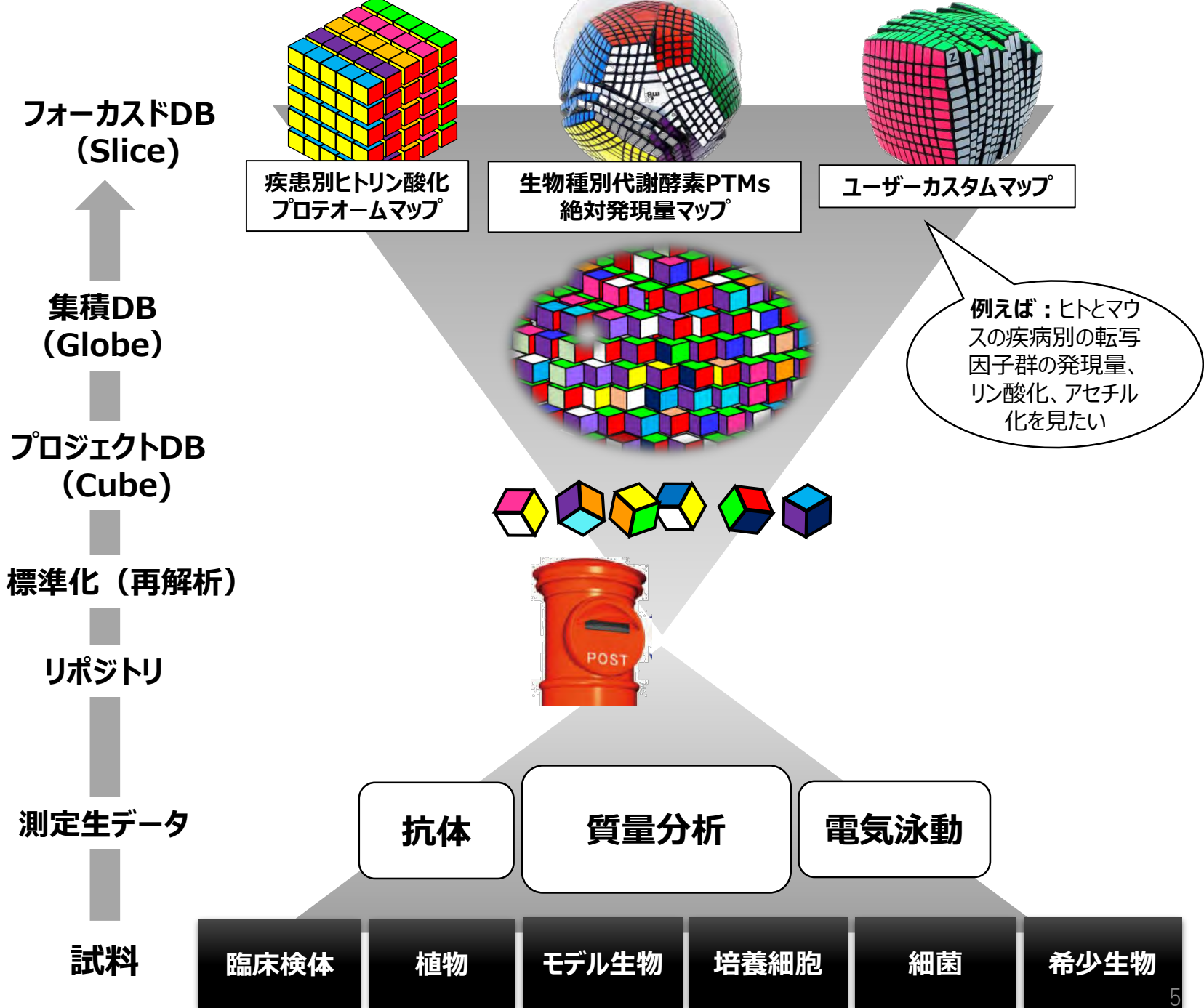


By JA Biscaino (EBI), 2018.5.17 MSP2018での講演より

研究機関や解析法の異なる、冗長性の高いMSデータをやみくもにかき集めた結果、多くの偽ヒットがProteomicsDBに登録。

MSデータ解析の標準化が必須

- 世界標準リポジトリ ProteomeXchange に参加。
- データ標準化ステップおよび統一したフィルタリングの導入(jPOSTスコアの開発)。
- データベースパートはカスタマイズ可能
- RDFベース（将来の統合へ向けて）
- 翻訳後修飾および絶対量情報を含むプロテオームDB
- ヒトだけではなく、多彩な生物種も網羅。
- 国内・国際連携（JHUPO, AOHUPO, HUPO）



機能深化と連携基盤強化

- より多様なLCMSデータ, 電気泳動データに対する対応
- 巨大国際プロジェクト対応
- 世界共通スペクトルID対応
- プロテオゲノミクス・グライコプロテオミクス対応
- メタプロテオミクス対応
- マルチオミクスデータ解析対応
- ネットワーク解析機能
- メタデータ半自動化取得

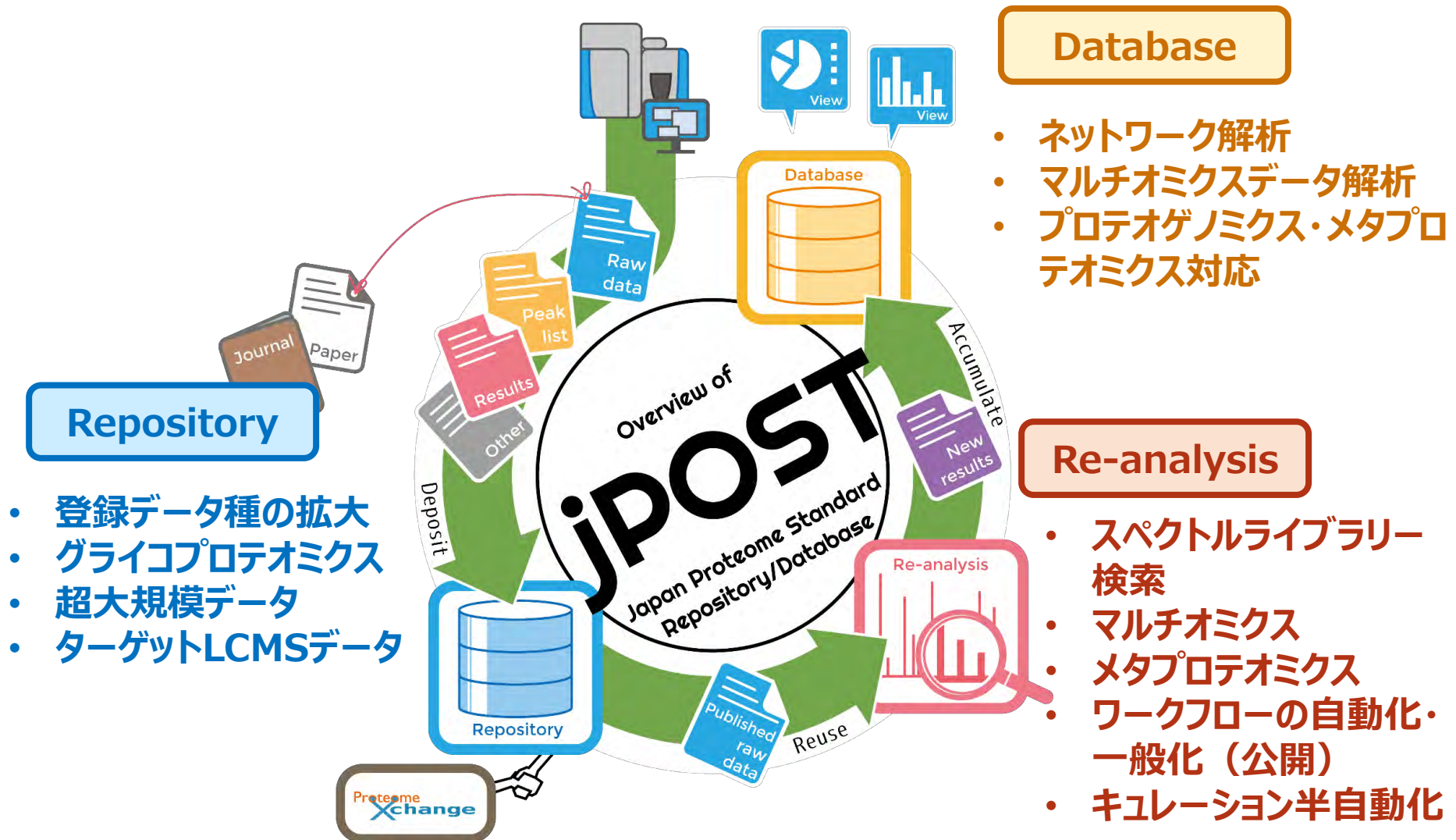
代 表：石濱 泰（京大院薬）

主な共同研究者：

- G1: 五斗 進（DBCLS）
- G2: 奥田 修二郎（新潟大学）
- G3: 松本 雅記（九州大学）
- G4: 荒木 令江（熊本大学）

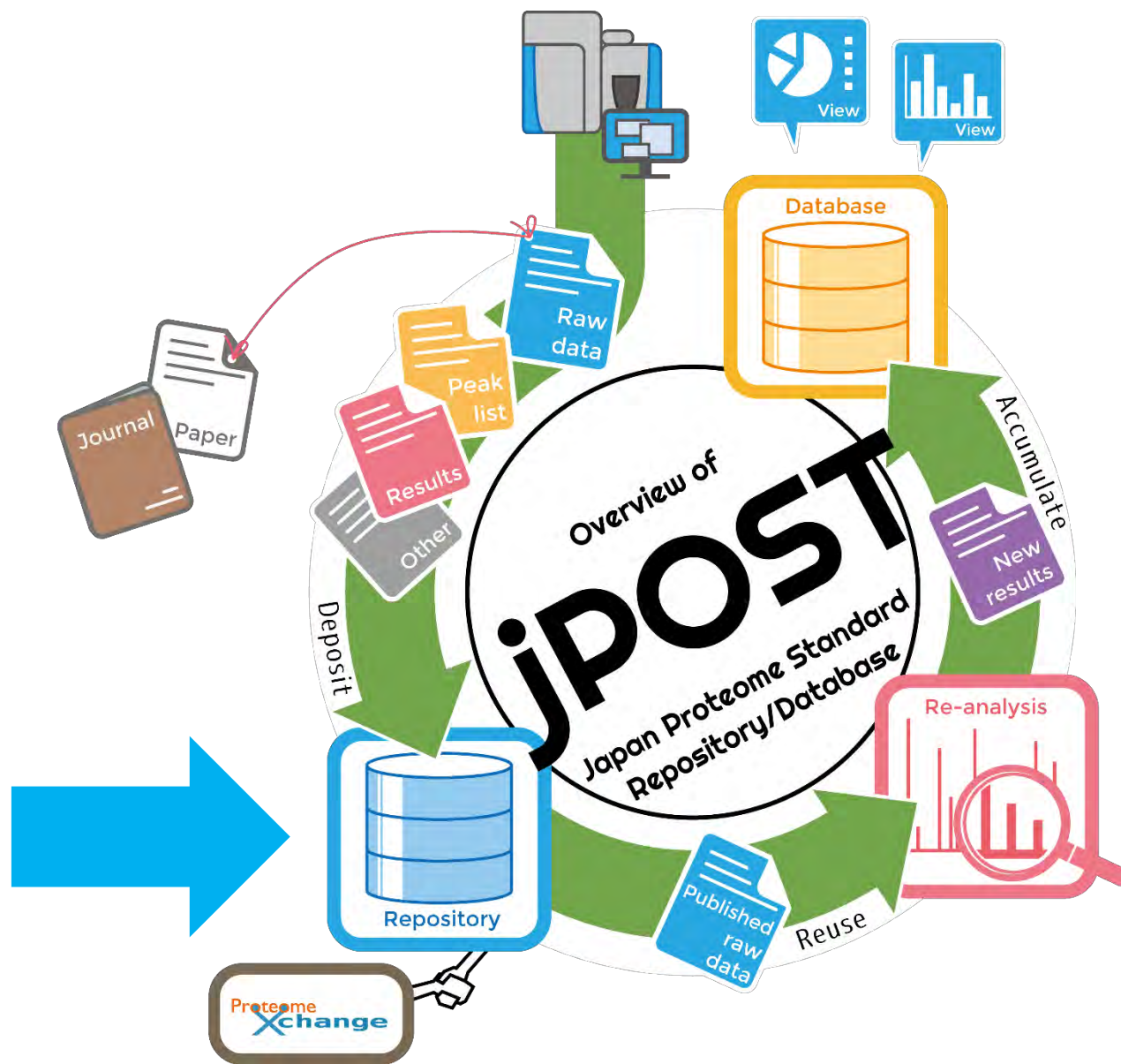
外部有識者
会議メンバー：

氏名	所属	氏名	所属
足立淳	医薬基盤研	小松節子	筑波大
小田吉哉	エーザイ	小迫英尊	徳島大
川村猛	東京大	植田幸嗣	がん研究会
榊原陽一	宮崎大	川島 祐介	理研
梶裕之	産総研	曾川一幸	麻布大
紀藤圭治	明治大	堂前直	理研
久保田一石	第一三共	肥後大輔	サームサイエンティフィック
近藤格	国立がん研究センター	本田一文	国立がん研究センター
高尾敏文	大阪大	木村弥生	横浜市立大



研究開発計画

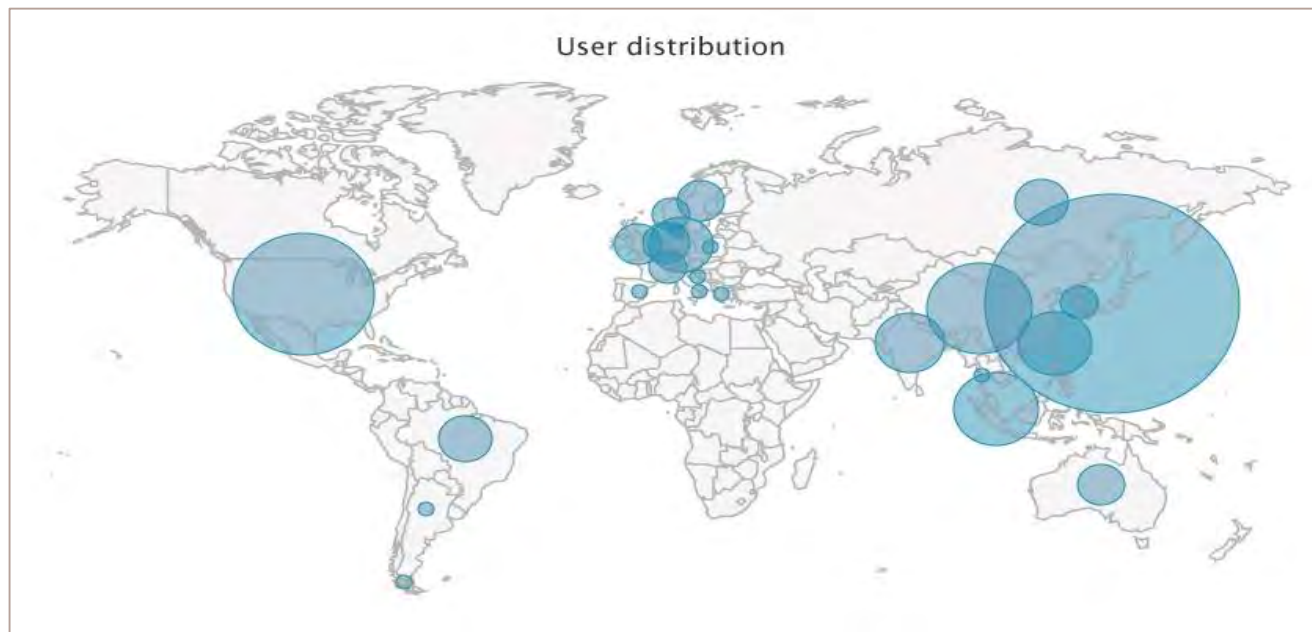
研究開発テーマ	タスク	担当グループ	2018	2019
リポジトリの機能強化と新機能開発	Help機能の充実	奥田G	■	■
	PXCのXML対応	奥田G	■	■
	データの修正に対する対応	奥田G	■	■
	大規模メタデータ一括インポートシステムの開発	奥田G、荒木G	■	■
	SRM, PRM, DIAデータ対応	奥田G、松本G	■	■
	グライコプロテオミクスデータ対応	奥田G、	■	■
再解析ワークフローの機能強化と新機能開発	jPOSTスコアの確立	石濱G	■	■
	ワークフローの全自動化	石濱G	■	■
	プロテオゲノミクスデータの再解析	石濱G	■	■
	メタゲノム配列データベースの整備	奥田G、石濱G	■	■
データベースツールの深化とプロテオゲノミクスデータの可視化およびネットワーク解析ツールの開発	Proteoform解析ツールの開発	五斗G、松本G	■	■
	絶対量やバリエーション選択的スライスデータベースの深化	松本G、五斗G	■	■
	プロテオゲノミクス用スライスデータベース	五斗G、荒木G	■	■
	既知/新規パスウェイ解析ツールの開発	奥田G、石濱G	■	■
オントロジーの深化と強化	オントロジーの充実	五斗G	■	■
	プロテオゲノミクスデータに対する対応	五斗G	■	■
キュレーションの深化と強化	メタボロームデータ対応	荒木G、松本G	■	■
	プロテオゲノミクスデータ対応	荒木G、松本G	■	■
	専門論文誌の発行	荒木G、石濱G、五斗G	■	■



jPOST repository – 利用状況

データ登録数は目標としていた年間100プロジェクトを大きく上回り、
2018.4～2019.4の1年間で153プロジェクトが登録された。

	2019年7月 現在	2018年度 登録数
プロジェクト	476	153
ファイル	46,175 (14.1TB)	12,163 (5.5TB)
ユーザー	265	114



jPOST repository – 利用状況

Help機能の大幅な見直しを行い、動画やチュートリアルをウェブ上で公開した。

ファイルは以下の4つのカテゴリごとに追加します。

- Raw data
- Peak list
- Search results
- Others

Raw data

1. ファイルカテゴリのタブで「Raw data」を選択します。（初期状態では選択されています）
2. Raw dataのファイルにメタデータのプリセットを関連づけるので、追加したいファイルに適したプリセットを選択します。（選択されたプリセットが下部のパネルに表示されます）
3. 4つのカテゴリのプリセットをそれぞれ選択していきます。

results were filtered at 1% false discovery rate (FDR) (500000) by searching against a randomized decoy database using Mascot at peptide and protein levels.

Revision 0

Raw data Peak list Search results Others

Sample TFractionation Enzyme/Mod. cMS mode

Search profile... sort (Modified date newer)

Sample SELECTED

Sample Create

Title: 50000000040
Species: Homo sapiens (Human)
Tissue: cell culture
Cell type:
Disease:
Note:

Fractionation (Not selected)

Enzyme/Mod. (Not selected)

MS mode (Not selected)

4. プリセットの選択が終わったら、コンピュータからファイルを選択・追加します。点線で囲まれたエリアにドラッグ&ドロップ、または、エリアをクリックしてダイアログから選択してください。（下部のリストにファイルが追加され、関連づけたプリセットのタイトルも表示されます）



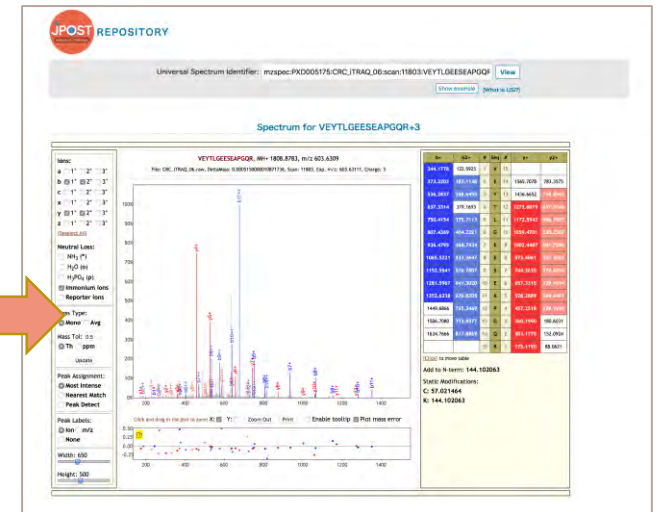
jPOST repository – ProteomeXchange連携

加盟する国際コンソーシアムProteomeXchangeでの仕様変更・追加への対応を行った。

- メタデータ記述に使用するControlled vocabularyの変更に対応
- Universal Spectrum Identifier(USI)に対応
 - 共通のルールに従って記述されたURLにアクセスすると、特定のスペクトラムの情報が表示される。

The figure is a world map illustrating the global distribution of proteomics data sources. Callouts point to specific regions: North America (PrideAtlas, MassIVE, Panorama), Europe (Pride), Asia (iProX), and Australia (JPOST). A central box labeled 'ProteomeXchange' lists 'Raw data', 'ID & Quant', and 'Meta-data'.

https://repository.jpostdb.org/spectrum/?USI=mzspec:PXD005175:CRC_iTRAQ_06:scan:11803:VEYTLGEESEAPGQR/3



- PROXI APIの提供を開始
 - 登録データの情報を検索できるAPI。
 - 検索クエリの指定方法やデータフォーマットの仕様がコンソーシアムで定義されている。

<https://repository.jpostdb.org/proxi/datasets?resultType=full&pageSize=5>

```
[{"accession":{"name":"JPOST dataset identifier","value":"JPST000004","accession":"MS:100263"}, {"publications":{"name":"PubMed identifier","accession":"MS:1000879","value":"2826"}, {"submitter":{"name":"Fukui K, Horimoto K, Nagashima T, Funayama R, Nakayama K, Nakayama KI A large-scale targeted proteome","accession":"MS:1002037","cvLabel":"MS"},"contact name":{"accession":"MS:1000590","value":"Div. Proteomics, Med. Inst. Bioreg., Kyushu Univ."}, {"affiliation":{"accession":"MS:1000590","value":"N/A","cvLabel":"MS"},"species":{"name":"tax","value":"9606","accession":"MS:1001467","cvLabel":"MS"},"instruments":{"name":"Q","accession number":"value":"PXD004528","accession":"MS:1001919","cvLabel":"MS"},"title":"Hirai Saki M, Wakabayashi M, Nakagawa M, Yamanaka S, Ishihama Y Rapid and deep profiling of human i","name":"dataset submitter","accession":"MS:1002037","cvLabel":"MS"},"contact name":{"accession":"MS:1000590","value":"Kyoto university","cvLabel":"MS"},"contact affiliation":{"accession":"MS:1000590","value":"N/A","cvLabel":"MS"},"species":{"name":"tax","value":"9606","accession":"MS:1001467","cvLabel":"MS"},"instruments":{"name":"Tr","value":"ProteomeXchange accession number","value":"PXD004845","accession":"MS:1001919"}
```

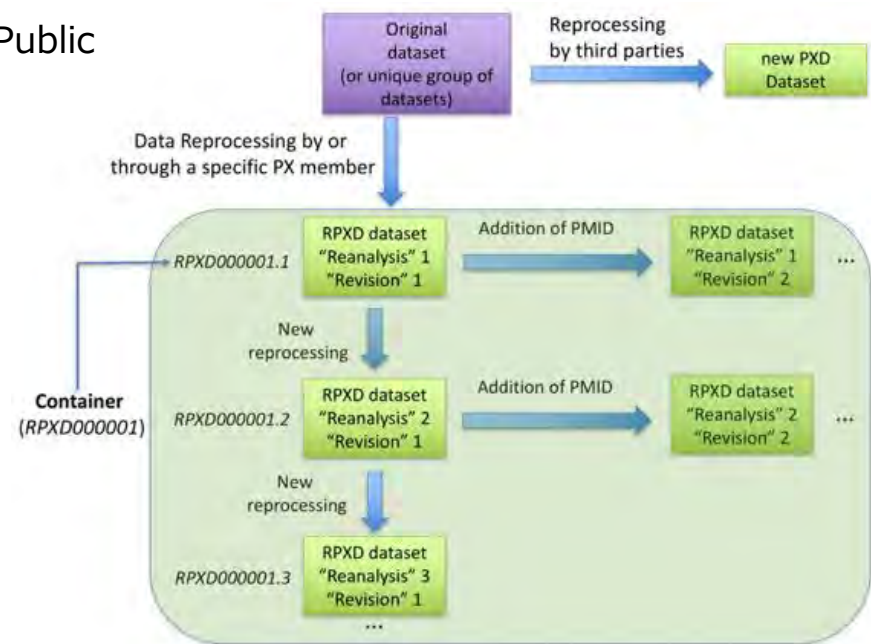
国際連携

再解析データ登録ガイドライン作成（進行中）

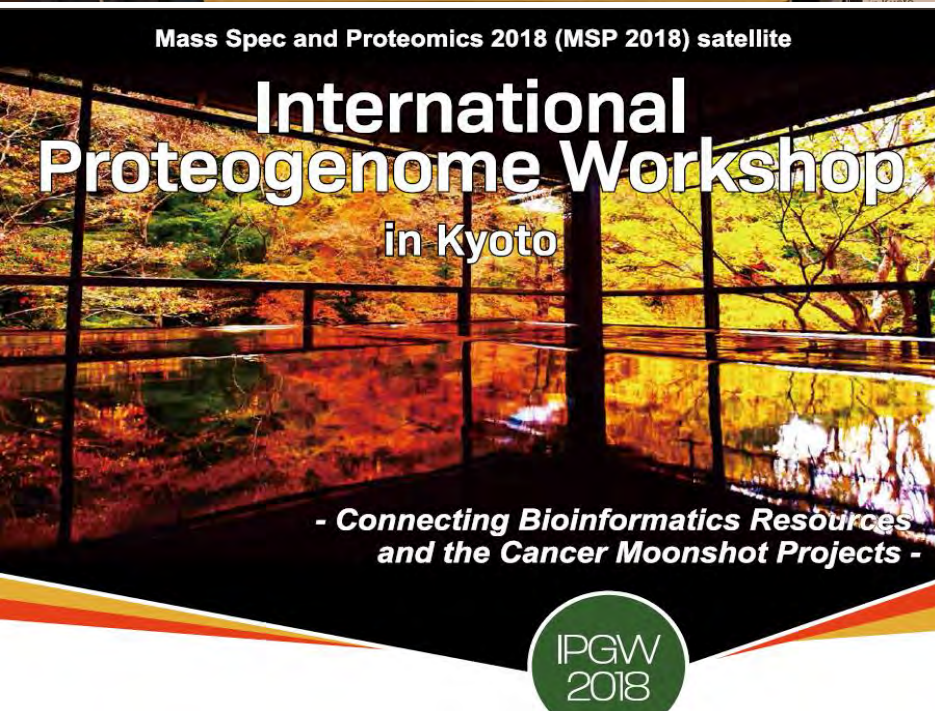
PX: jPOST/PRIDE/MassIVE/PASSEL/iProX/PanoramaPublic

- リポジトリに登録されている生データを再解析したデータを PX に登録する方法について共通化

- RPXD ID を発行し、PXD エントリ（オリジナルの結果）と区別
- jPOST の再解析結果もガイドラインにそって登録予定



International Proteogenome Workshop in Kyoto, May 19, 2018



- EBIとの連携：グラント申請（メタプロテオミクス）
- ICPC－Japan: NCCとの連携（データ解析担当）
- ICPC – Taiwan Cancer Moonshot との連携

International Proteogenome Workshop in Kyoto, May 19, 2018

Time	Title	Speaker
9:30	Registration open	
9:50-10:00	Opening	Yasushi Ishihama (Kyoto Univ, Japan)
10:00-12:30	Current status of Cancer Proteogenomics in AO regions: focusing on bioinformatics infrastructure and data analysis	• Yu-Ju Chen (Academia Sinica, Taiwan) • Jau-Song Yu (Chang Gung Univ, Taiwan) • Jun Qin (BPRC, China) • Min-Sik Kim (Kyung Hee Univ, Korea) • Tadashi Kondo (NCC, Japan)
12:30-14:00	Lunch	
14:00-16:00	Current status of bioinformatics for proteogenomics	• Juan Antonio Vizcaíno and Andrew Jarnuczak (EBI): PRIDE • Henning Hermjakob (EBI, BPRC): Reactome • Jyoti Choudhary (Institute for Cancer Research, London): Genomic Determinants of Protein Abundance Variation in Colorectal Cancer Cells • Yasushi Ishihama (Kyoto Univ): jPOST
16:00-17:00	Discussion	All participants

メタオミクスワークショップ

Meta-Omics Workshop in Kyoto 2019

- 主催 京都大学融合チーム研究プログラム SPIRITS メタプロテオミクス・プロジェクト
- 後援 一般社団法人 日本マイクロバイオームコンソーシアム (JMBC)
日本人腸内環境の全容解明と産業応用プラットフォーム (JCHM)
京都生体質量分析研究会 (KBMS)
- 協賛 JST ライフサイエンスデータベース統合推進事業 jPOST プロジェクト

● 参加費 無料

● お申し込み

<http://seizai.pharm.kyoto-u.ac.jp/mow2019/>



● プログラム

12:30 開場

12:55 挨拶 石濱 泰 (京都大学・院薬)

13:00 特別講演 1 黒川 顕 (国立遺伝学研究所)

微生物統合データベース「MicrobeDB.jp」

13:40 依頼講演 1 奥田 修二郎 (新潟大学・院医歯)

「がんゲノムインフォマティクスからプロテオゲノミクスへ」

14:10 依頼講演 2 須田 互 (理化学研究所・生命医科学研究センター)

「日本人腸内細菌叢のメタゲノム解析」

14:40 依頼講演 3 遠藤 寿 (京都大学・化研)

「大規模観測とメタゲノム解析による海洋微生物の多様性研究」

(休憩 30分)

15:40 依頼講演 4 杉山 直幸 (京都大学・院薬)

「細菌叢の高深度メタプロテオーム解析」

16:10 依頼講演 5 早川 英介 (沖縄科学技術大学院大学)

「質量スペクトルデータの多層ネットワークで見る未知代謝物の世界」

16:40 特別講演 2 水口 賢司 (医薬基盤研究所)

「マイクロバイオームと表現形データの統合解析プラットフォーム」

17:20 総合討論

18:00 終了

2019年3月9日 土

京都大学薬学部キャンパス内

藤多記念ホール (医薬系総合研究棟)

国内 ワークショップ 開催



Tokyo Cancer Moonshot Workshop

Proteogenomics to solve a mystery of the cancer genome

国際がんプロテオゲノムコンソーシアム

(ICPC, International Cancer Proteogenome Consortium) 代表

Henry Rodriguez博士 (NCI, USA) を迎えて



Tokyo Cancer
Moonshot Workshop,
June 24, 2019

開催概要



National Cancer Center Japan

日時：6月24日（月） 受付：14:00～ セミナー：15:00～18:20

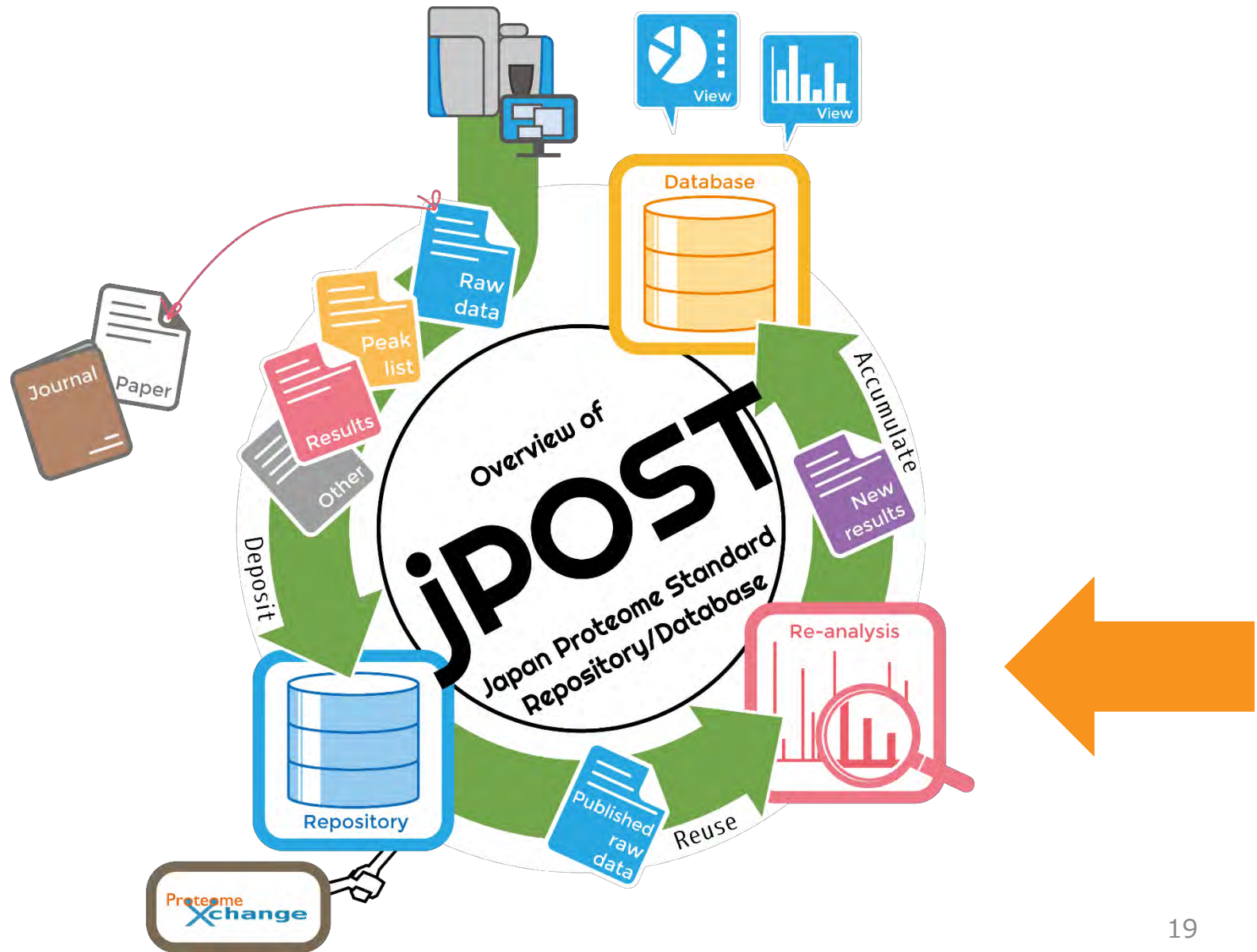
開催場所：国立がん研究センター研究棟 1階 大会議室

(104-0045東京都中央区築地5-1-1)

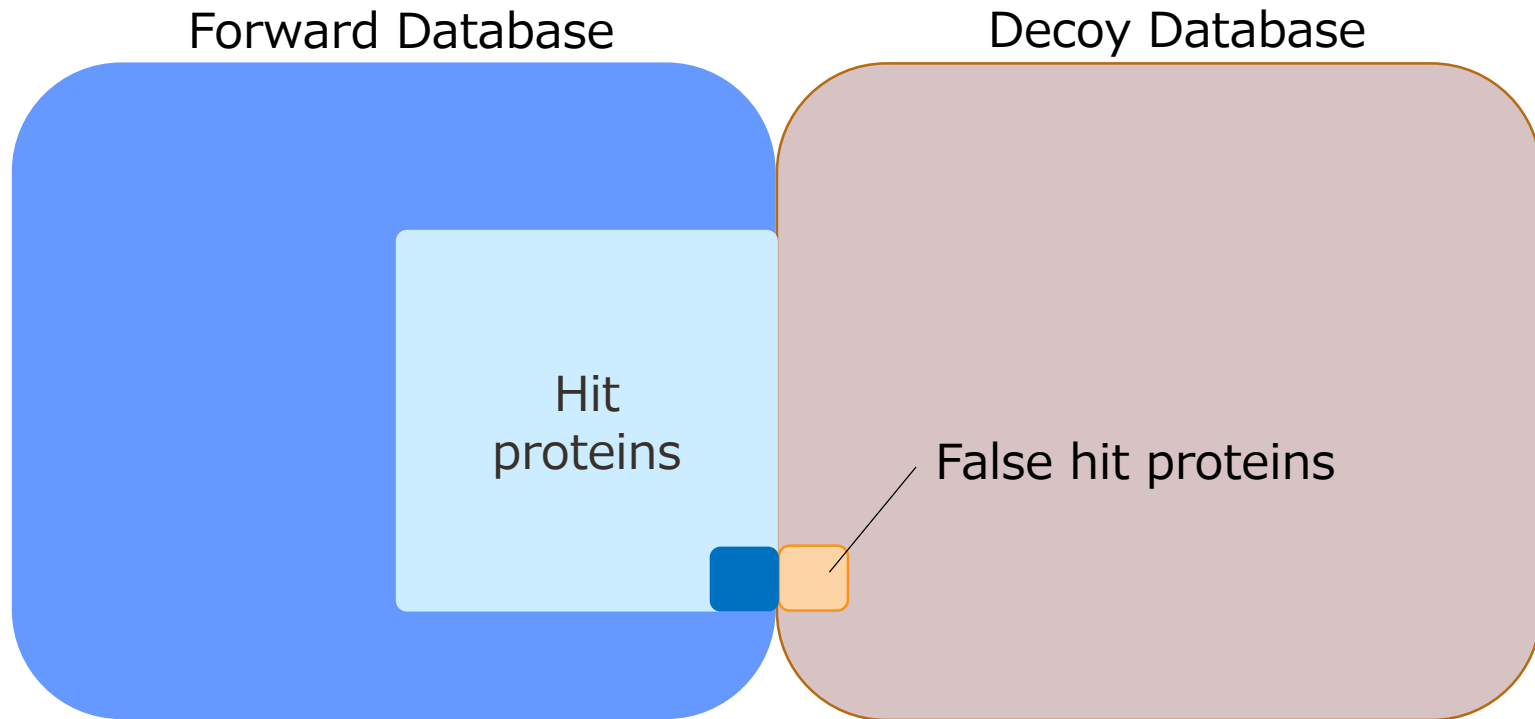
プログラム

15:00～15:05	Introduction Tesshi Yamada (Human Proteome Organization) 山田哲司（ヒトプロテオーム機構アジア・オセアニア地区理事）
15:05～15:50	Proteogenomics - Proteomics joins forces with Genomics to Advance Basic and Translational Cancer Research Henry Rodriguez (Deputy Director (Acting), Center for Strategic Scientific Initiatives, Director, Office of Cancer Clinical Proteomics Research, National Cancer Institute, National Institutes of Health)
15:50～16:15	Proteogenomic landscapes of cancer-associated intra- and extracellular signaling プロテオゲノミクスによるがん細胞内外シグナル経路の解明 Koji Ueda (Cancer Precision Medicine Center, Japanese Foundation for Cancer Research) 植田幸嗣（がん研究会 がんプレシジョン医療研究センター）
16:15～16:40	Utility of reverse-phase protein array for proteogenomic analysis プロテオゲノム解析における逆相タンパクアレイ法の有用性 Mari Masuda (National Cancer Center Research Institute) 増田万里（国立がん研究センター研究所）
16:40～17:00	休憩
17:00～17:25	Proteogenomic analysis of 259 liver cancers 肝臓がん259症例のプロテオゲノム解析 Masashi Fujita (RIKEN Center for Integrative Medical Sciences) 藤田征志（理化学研究所 生命医科学研究センター）
17:25～17:55	Advancing Mass Spectrometry-based Large-Cohort Proteomics for Precision Medicine – An International Cancer Moonshot Multiple Site Study Xuan Yue (Thermo Fisher Scientific)
17:55～18:20	Challenges in the jPOST framework towards proteogenome analysis プロテオゲノム解析に向けたjPOSTの挑戦 Yasushi Ishihama (Graduate School of Pharmaceutical Sciences, Kyoto University) 石濱泰（京都大学 薬学研究科）

jPOST re-analysis



Target-Decoy Approach for estimating FDR(%)



$$\text{False Discovery Rate} = \text{False Positive} / (\text{False Positive} + \text{True Positive})$$

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

Joshua E Elias¹ & 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.

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 Sahasrabudde](#); [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

Analyzing the First Drafts of the Human Proteome

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

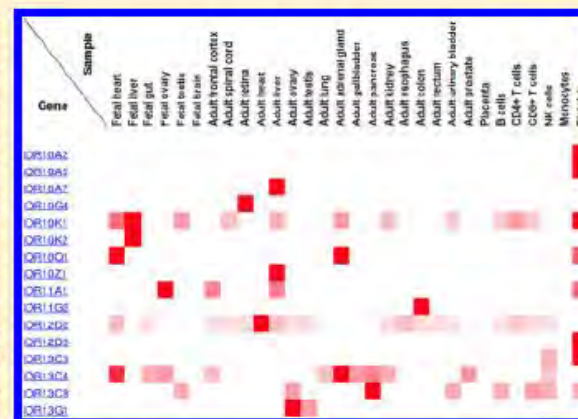
[†]Unidad de Proteómica, [§]Laboratorio de Proteómica Cardiovascular, Centro Nacional de Investigaciones Cardiovasculares, Melchor Fernández Almagro, 3, Madrid 28029, Spain

[‡]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 portrayed as draft maps of the human proteome and great advances in the field. As with the initial publication of the human genome, these papers have broad appeal and will no doubt lead to a great deal of further analysis by the scientific community. However, we were intrigued by the number of protein-coding genes detected by the two studies, numbers that far exceeded what has been reported for the multinational Human Proteome Project effort. We carried out a simple quality test on the data using the olfactory receptor family. A high-quality proteomics experiment that does not specifically analyze nasal tissues should not expect to detect many peptides for olfactory receptors. Neither of the studies carried out experiments on nasal tissues, yet we found peptide evidence for more than 100 olfactory receptors in the two studies. These results suggest that the two studies are substantially overestimating the number of protein coding genes they identify. We conclude that the experimental data from these two studies should be used with caution.

KEYWORDS: proteomics, *Nature*, human proteome, protein coding genes, olfactory receptors



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

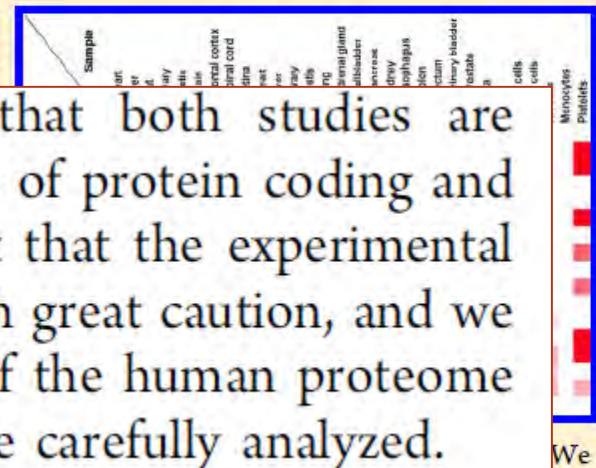
Supporting Information

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

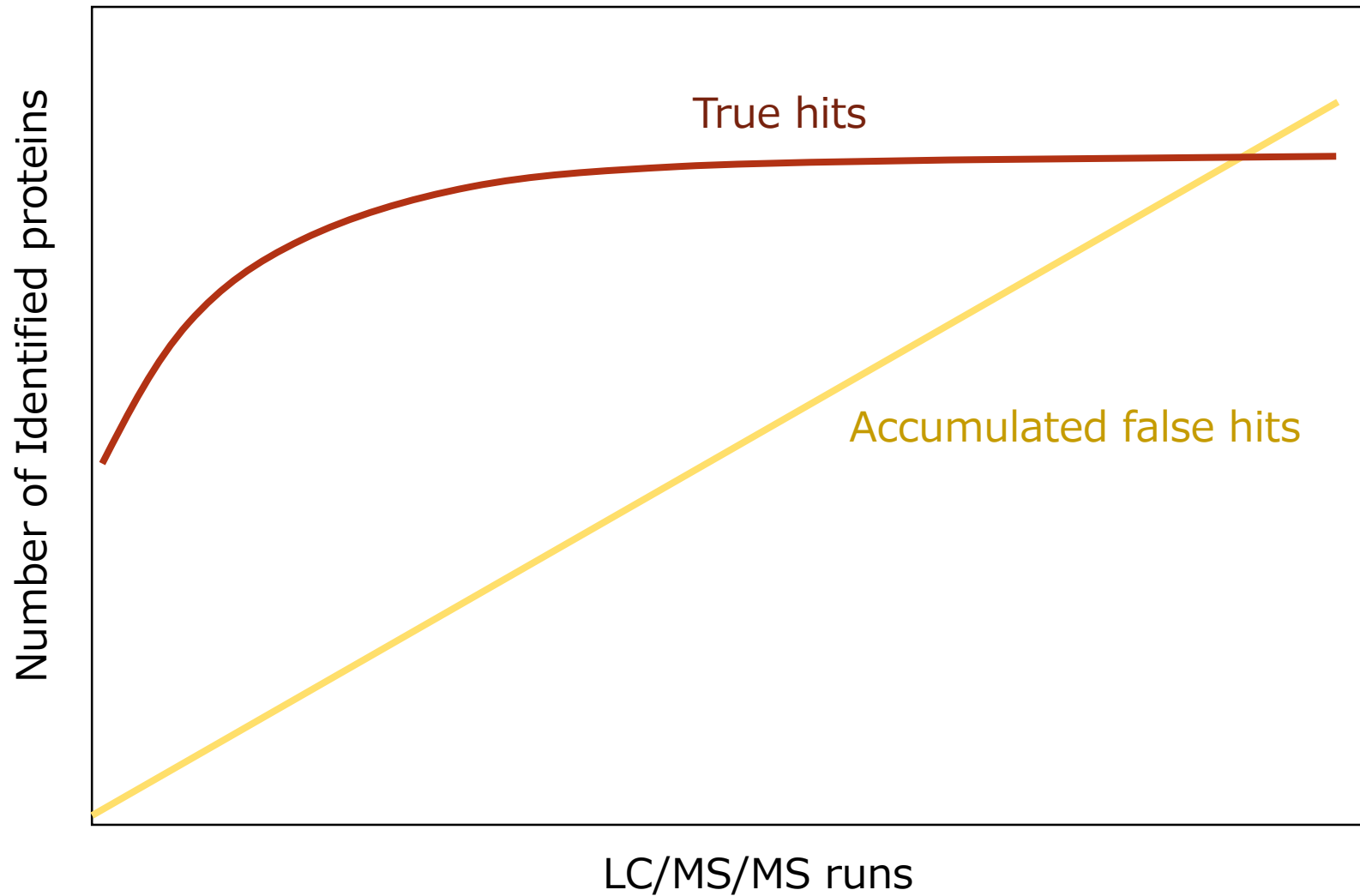
The results of our analysis show that both studies are substantially overestimating the number of protein coding and noncoding genes they find. We suggest that the experimental data from these two should be used with great caution, and we feel that these two unique draft maps of the human proteome should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

We conclude that the experimental data from these two studies should be used with caution.

KEYWORDS: proteomics, *Nature*, human proteome, protein coding genes, olfactory receptors



False hit content is $> 30\%$



Target-decoy search for all merged data

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%)



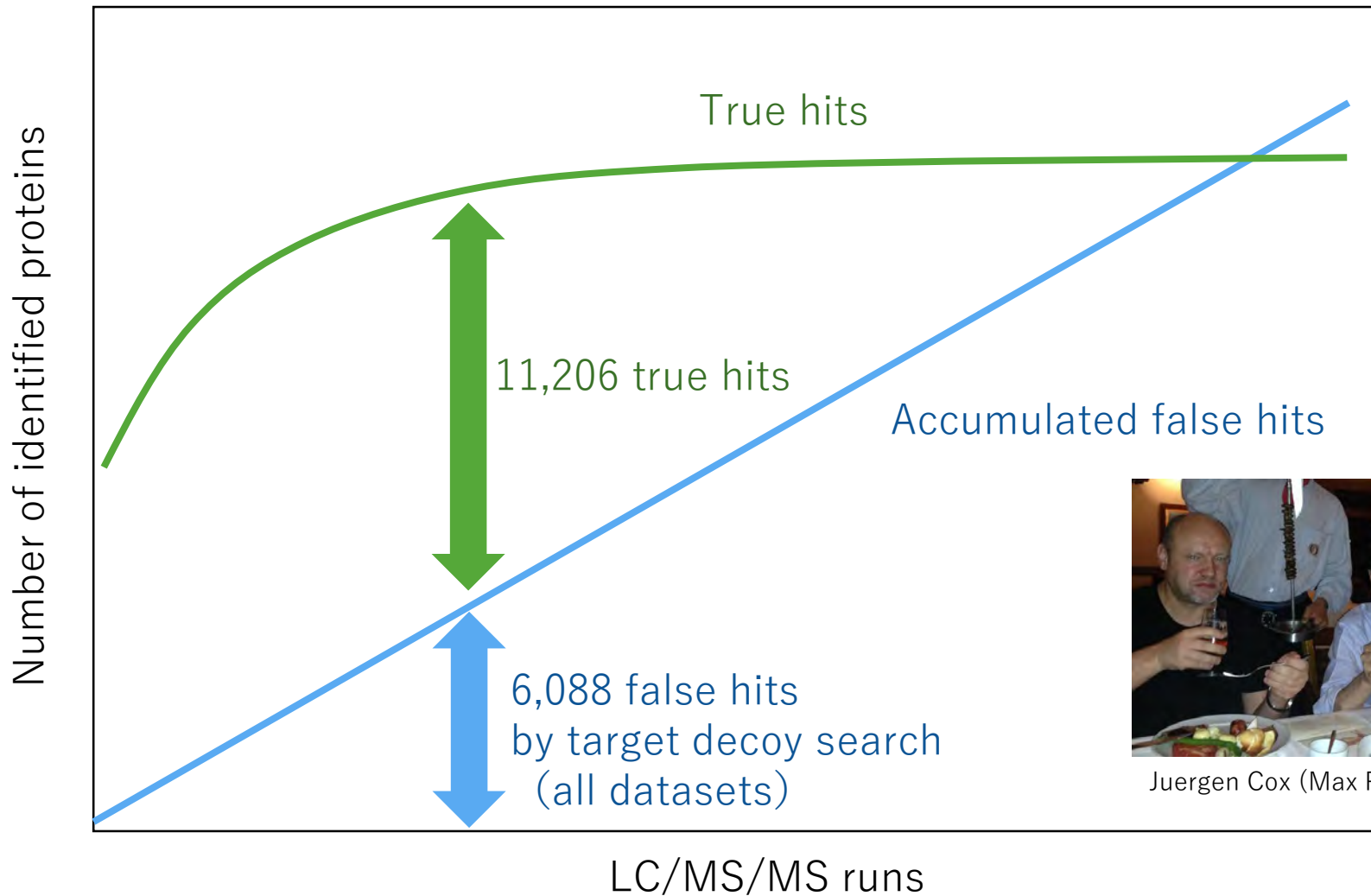
Target-decoy search for all merged data
1% FDR at protein level

11,206 genes (57%)

Juergen Cox (Max Planck)

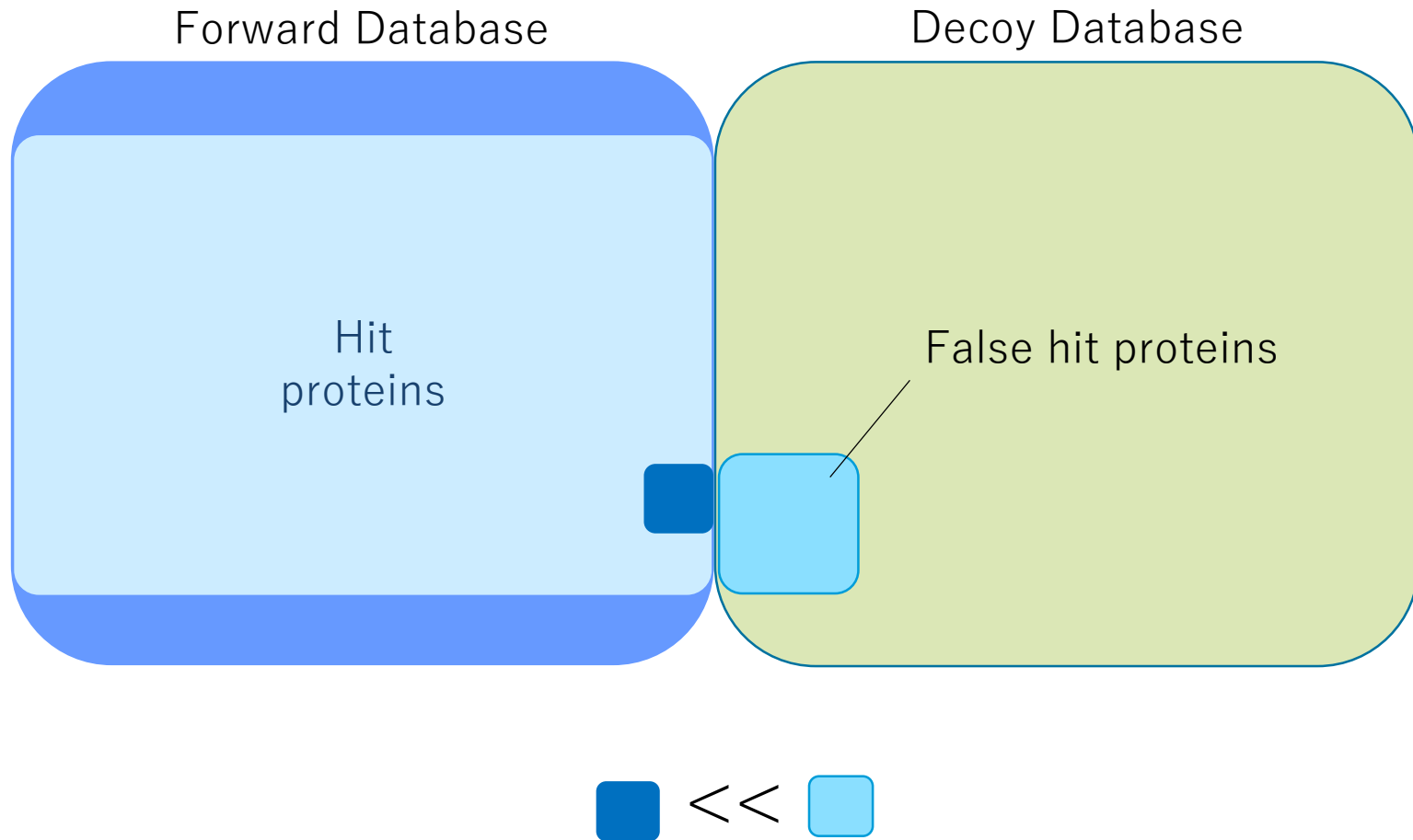


False hit content is $> 30\%$



Juergen Cox (Max Planck)

Target-Decoy Approach for Ultra-large Datasets



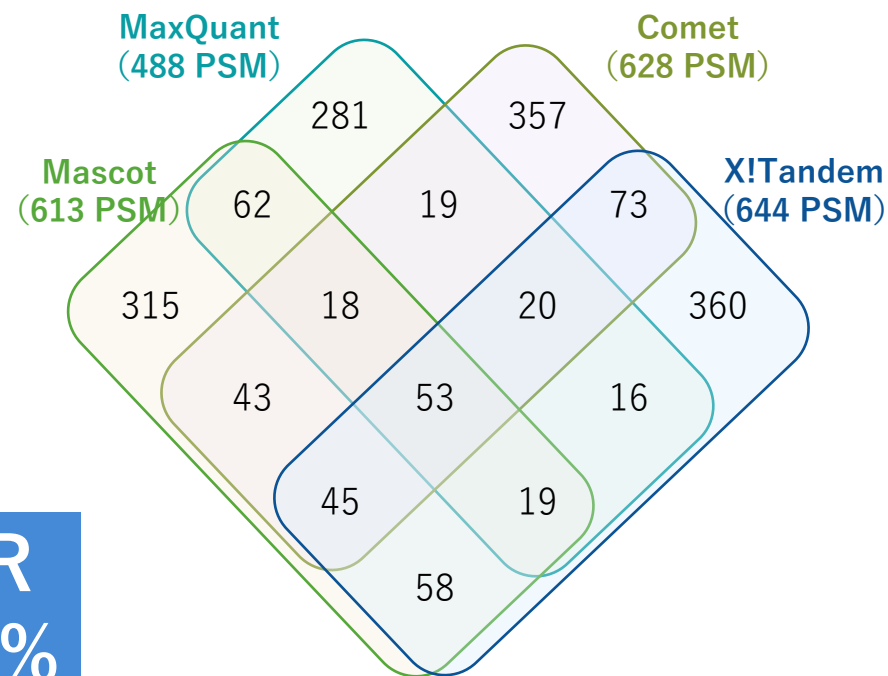
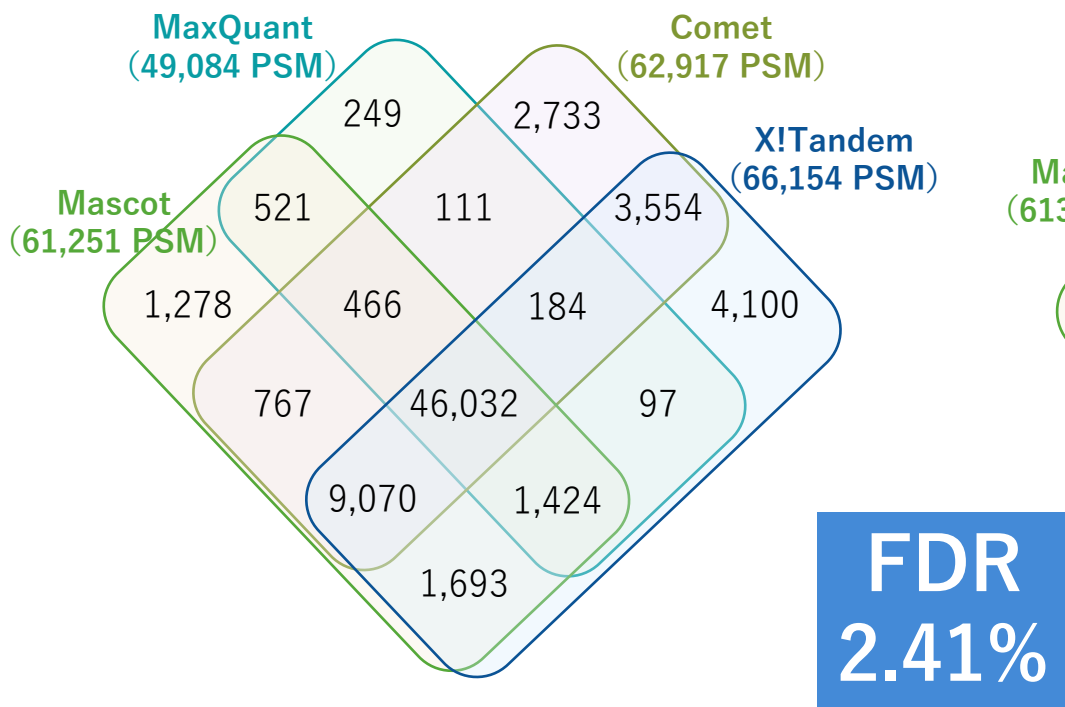
Multiple search engines with FDR control

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

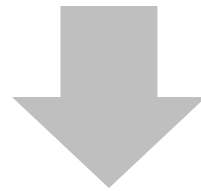
▣ HeLa global proteome dataset (JPST000200)

Target
72,279 PSM in total

Decoy
1,739 PSM in total



How can we merge the results
from different sources?



jPOST score

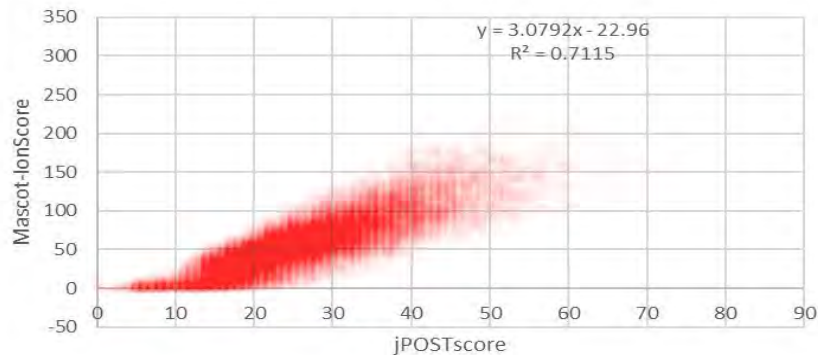
- based on peak annotation in MSMS
- search engine independent
- MS instrument independent
- search DB independent
- can be used as universal threshold for peptide identification

Relationship between different scores

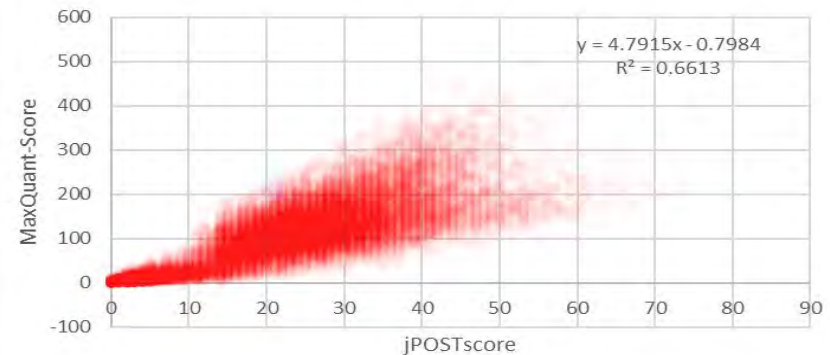
Dataset:JPST000200(5 files)

Tryptic peptides from HeLa cells, by Thermo Q-Exactive with 8h gradient

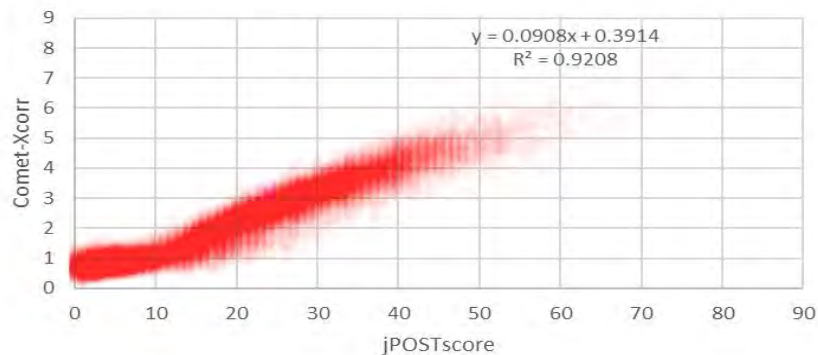
Mascot



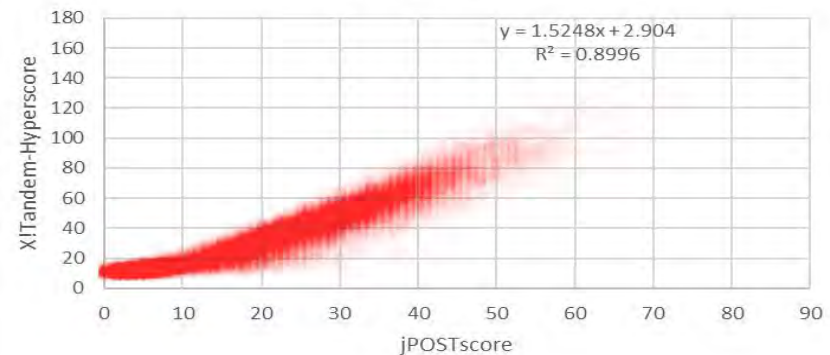
MaxQuant



Comet



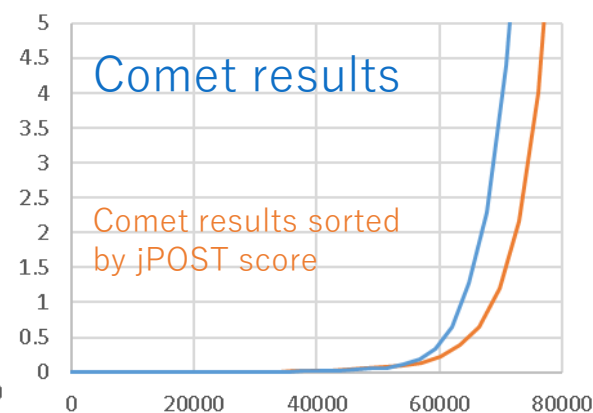
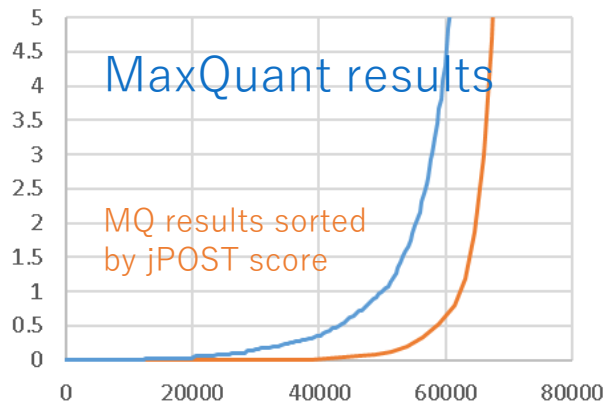
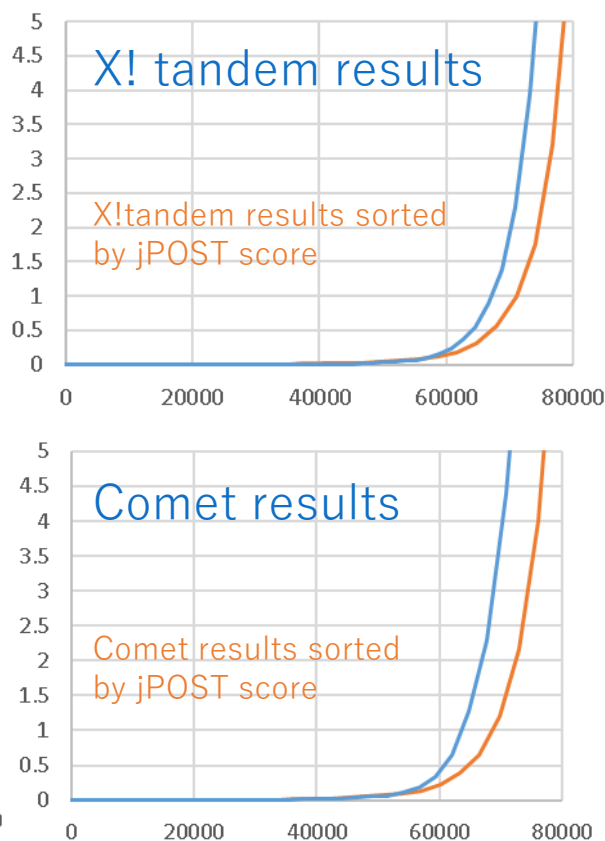
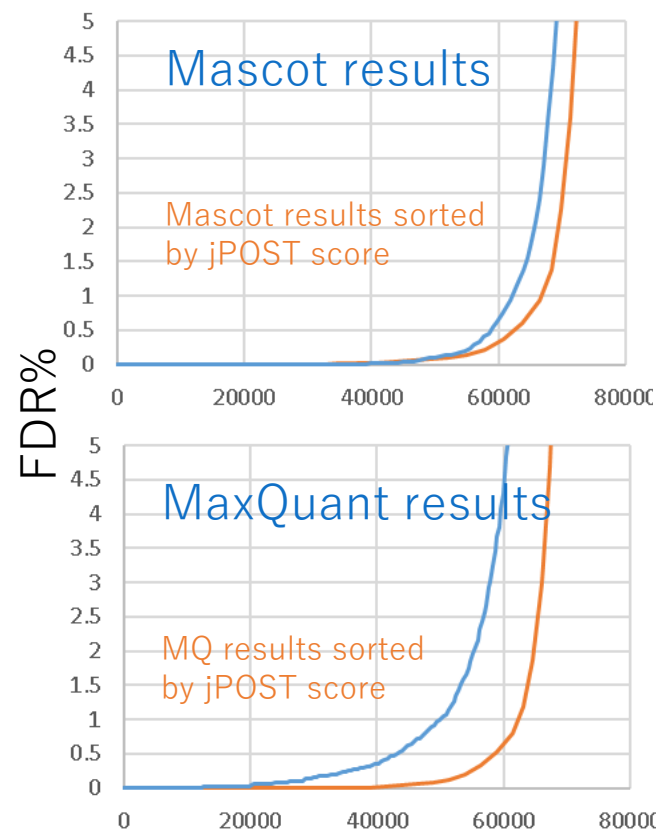
X!Tandem



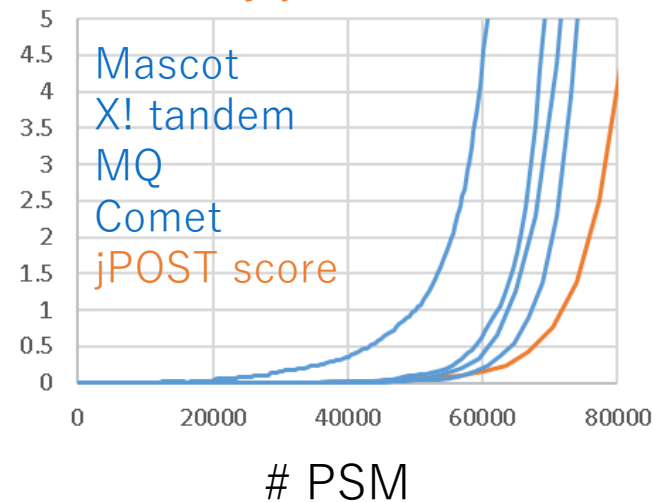
Optimized jPOST score to rank down the decoy hits

Dataset:JPST000200(5 files)

Tryptic peptides from HeLa cells, by Thermo Q-Exactive with 8h gradient



4 engines merged results
Sorted by jPOST score

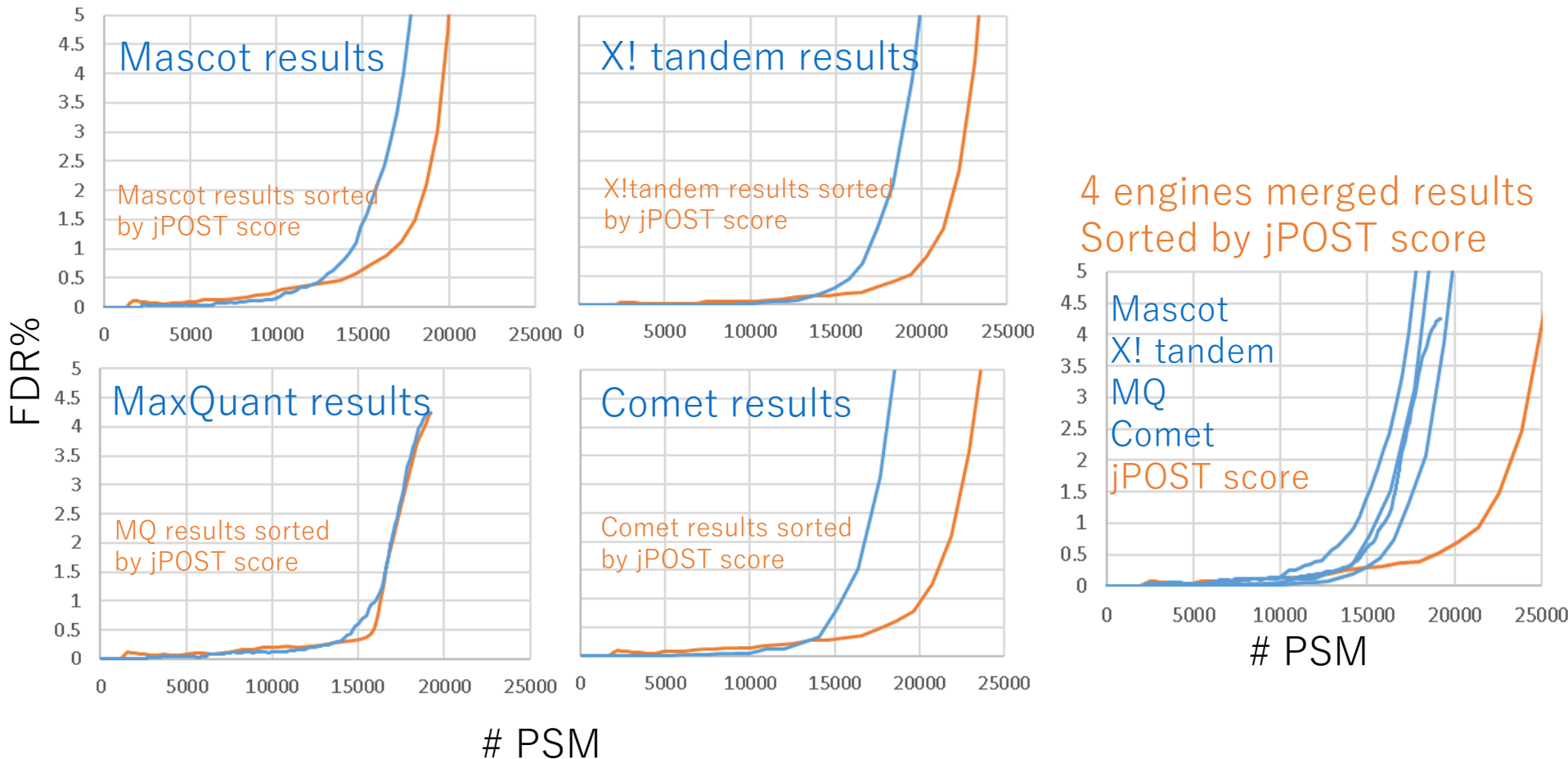


PSM

Optimized jPOST score to rank down the decoy hits

Dataset: PXD001792

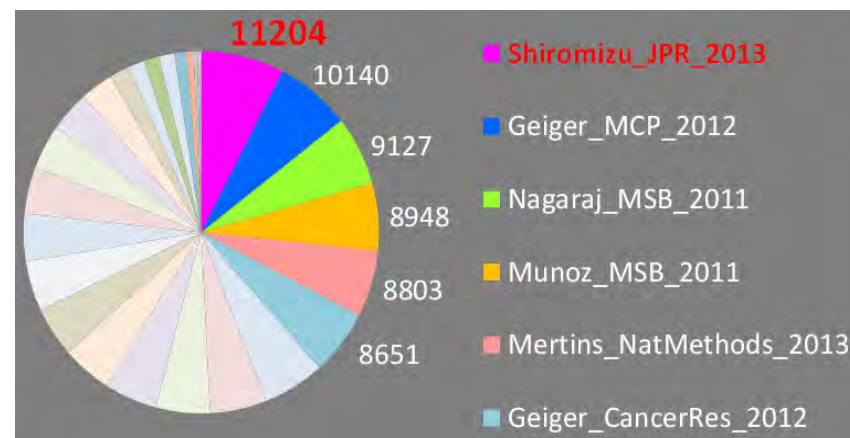
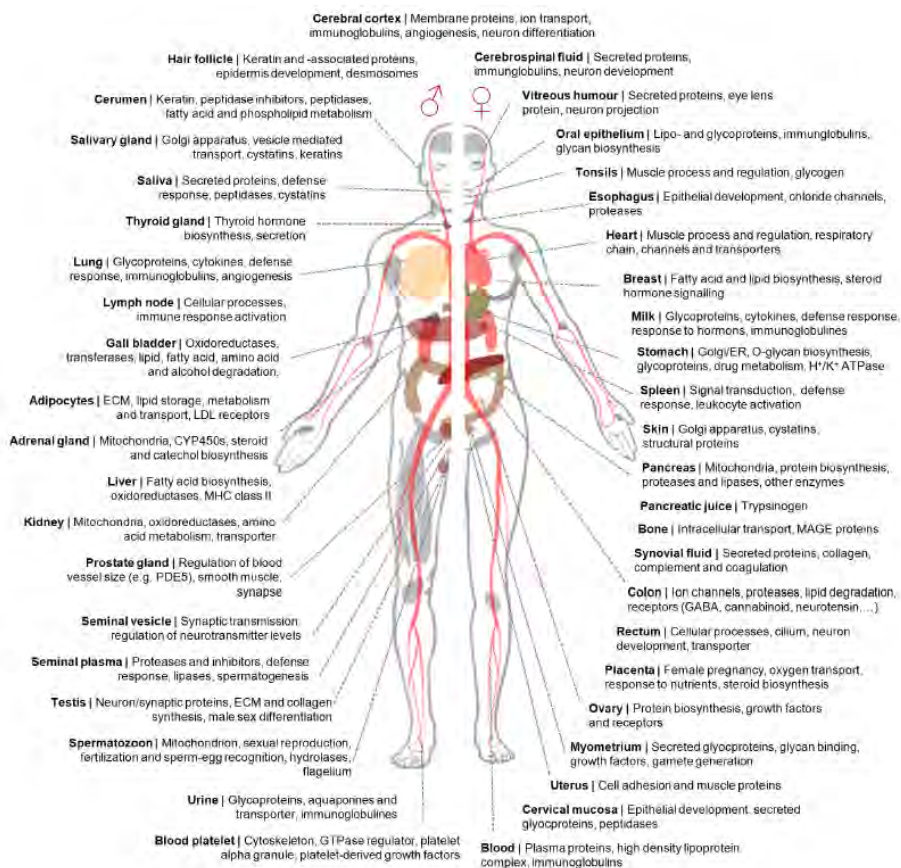
Phosphopeptides from mouse Hepa cells by Thermo Q-Exactive



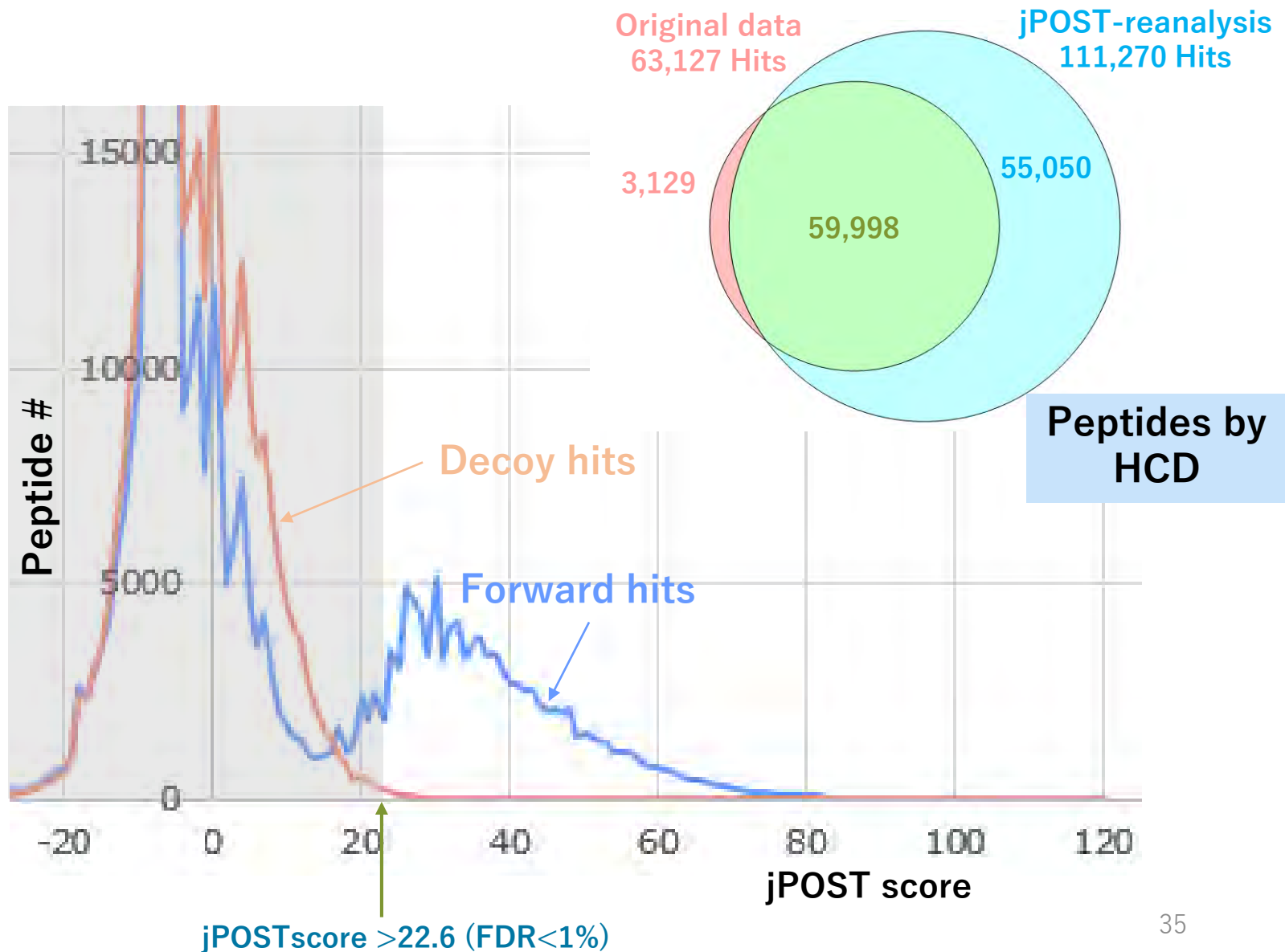
ARTICLE

doi:10.1038/nature13319

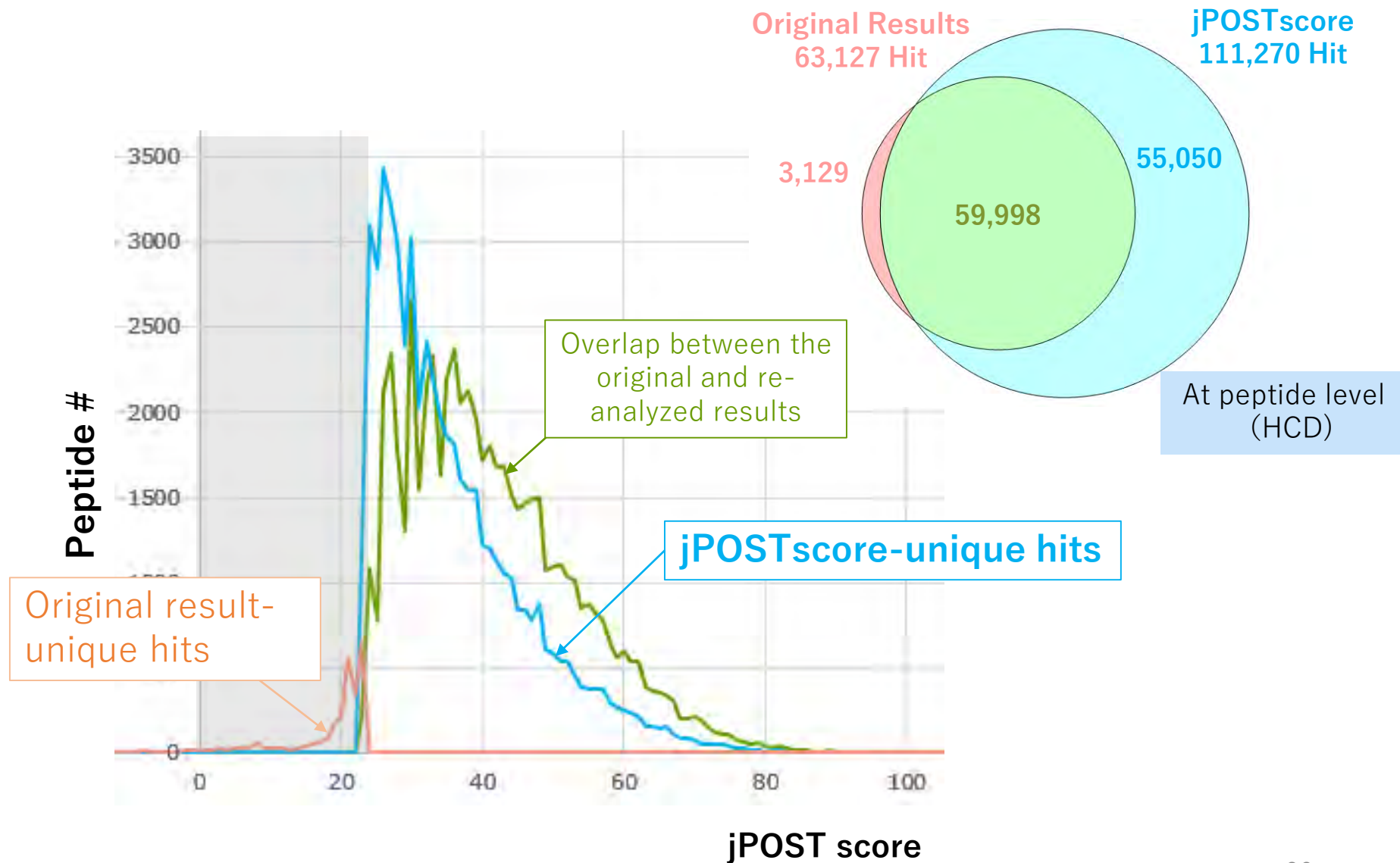
Mass-spectrometry-based draft of the human proteome



jPOST score-based re-analysis



HPP datasets reanalyzed by jPOST score



Phospho-site localization (リン酸化プロテオミクス)

現状の課題

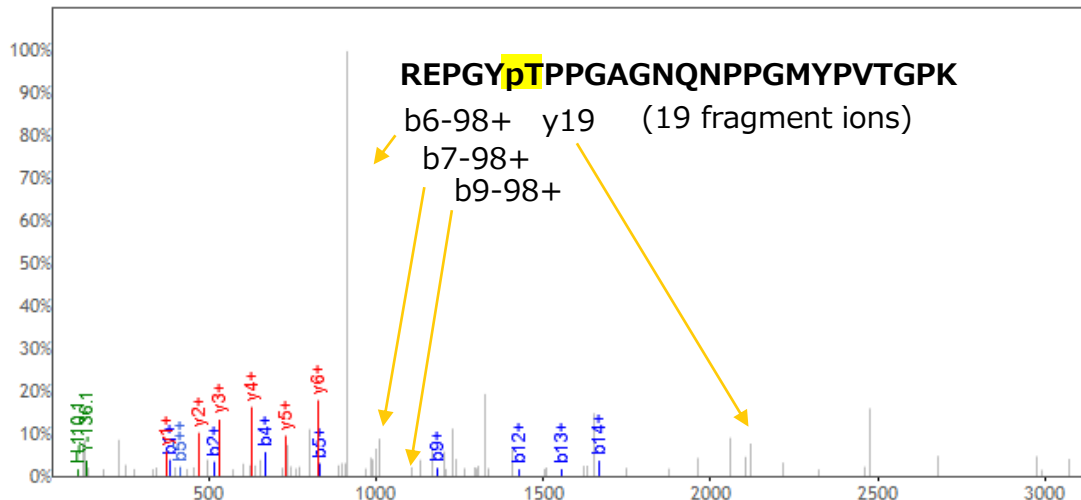
- ◆ 使用する解析ツールによって異なるリン酸化サイトが同定される
- ◆ キメラMS/MSスペクトルの場合、サイトの特定がより困難になる

既存の手法は改良の余地があり、新しいリン酸化サイトの評価手法の確立が必要

キメラMS/MSスペクトルの例

REPGYTPPGAGNQNPPGMpYPVTGPK, MH+ 3136.5326, m/z 784.8886

File: 20170821_LC_TMTB13_phospho_F01_02.maxq.txt, PeptExpt: 1.6e-04, DeltaMass: -0.983310, Mod: TMT6plex (N-term);Oxidation (M);Phospho (Y);TMT6plex (K), Scan: 20318, Exp. m/z: 784.6428157021699, Charge: 4



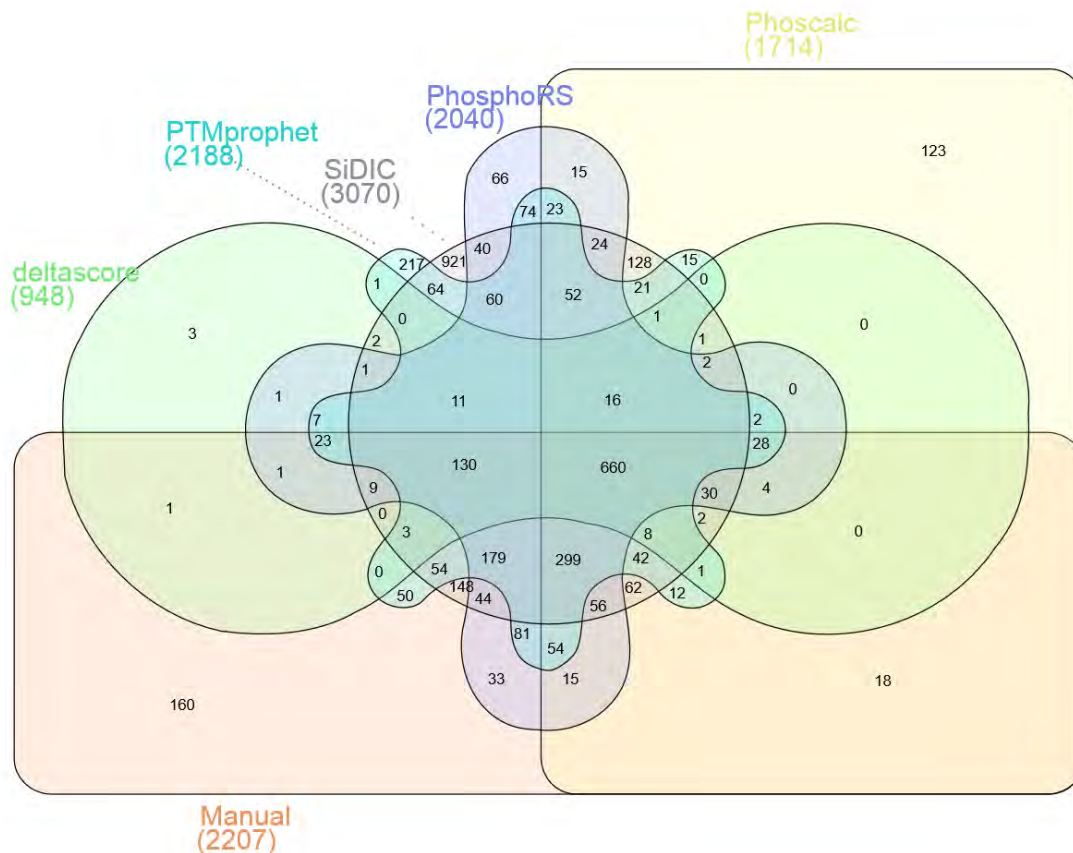
- Manual inspection: Phospho@Y:19
- PTM score: Phospho@T:6
- SiDIC* (site-determining ion combination):
@Y:19, Phospho@T:6 (chimeric)

*Nakagami, H. et al. Plant Physiol 153, 1161-1174 (2010).

Phospho-site localization (リン酸化プロテオミクス)

Taiwan Cancer Moonshot program (Academia Sinica)との共同研究

- ◆ 大規模リン酸化プロテオームデータを用いた解析ツール間の同定結果の比較
- ◆ リン酸化サイトの評価手法の改良、新規な手法の確立

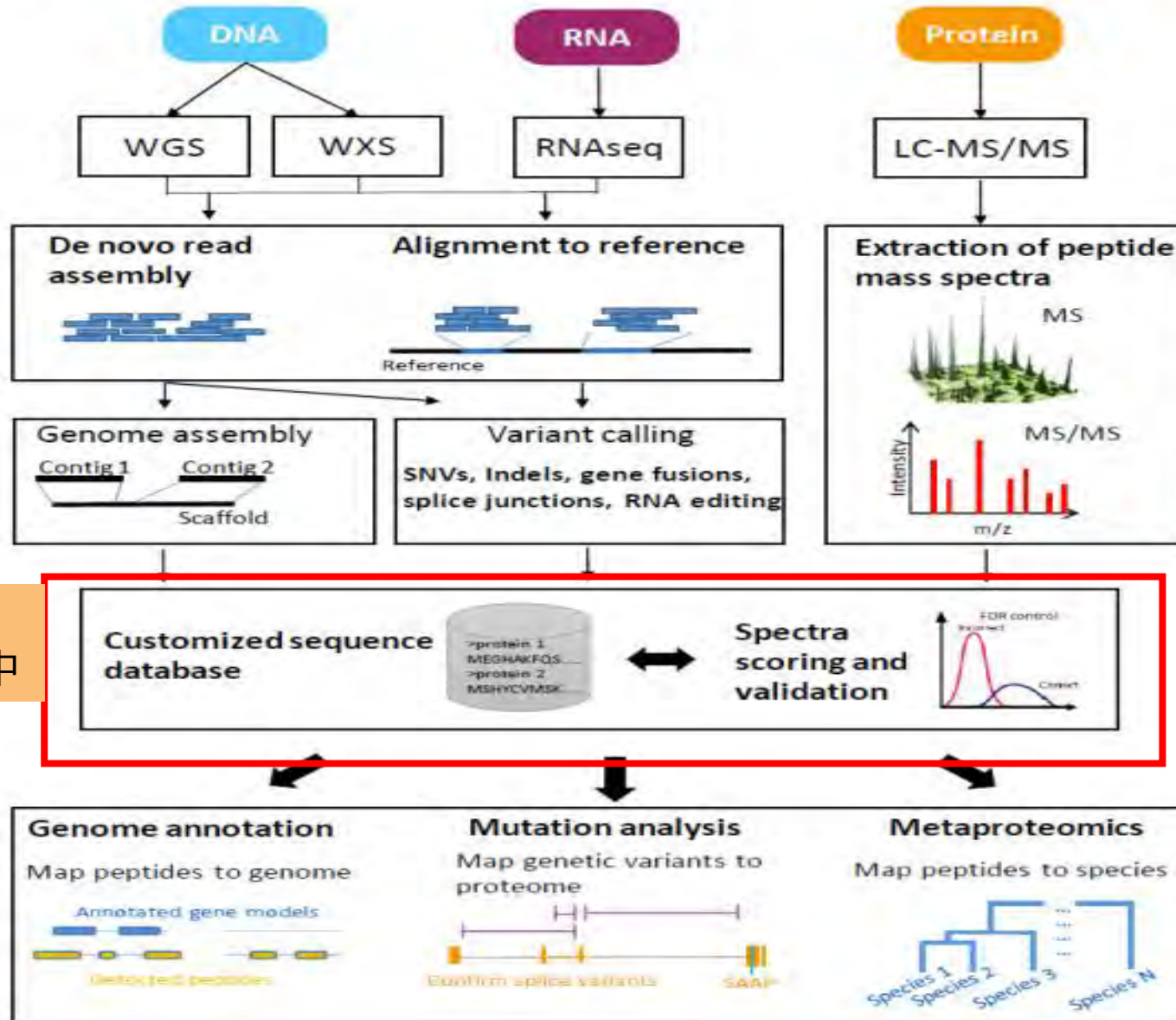


6種類の解析ツールにより
同定されたリン酸化サイトの比較

- Manual inspection
- Phoscalc
- SiDIC
- MD-score
- PhosphoRS
- PTMprophet

SiDICが最も多くの
リン酸化サイトを同定

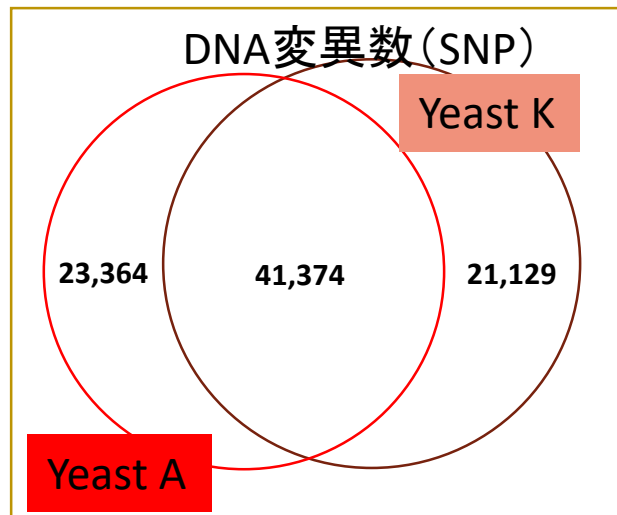
プロテオゲノム解析のためのDB



現在
作成中

プロテオゲノム解析応用例（アルコール酵母株）

S288C株 (S. Cerevisiae) **リファレンス配列**
5939遺伝子(ゲノムサイズ：12 Mb)



蛋白質アミノ酸配列情報(ORF)
missense変異をデータベース化



タンパク質サンプル
Yeast K (condition1,2)
Yeast A(condition3,4)

質量分析

スペクトル
800000~/sample

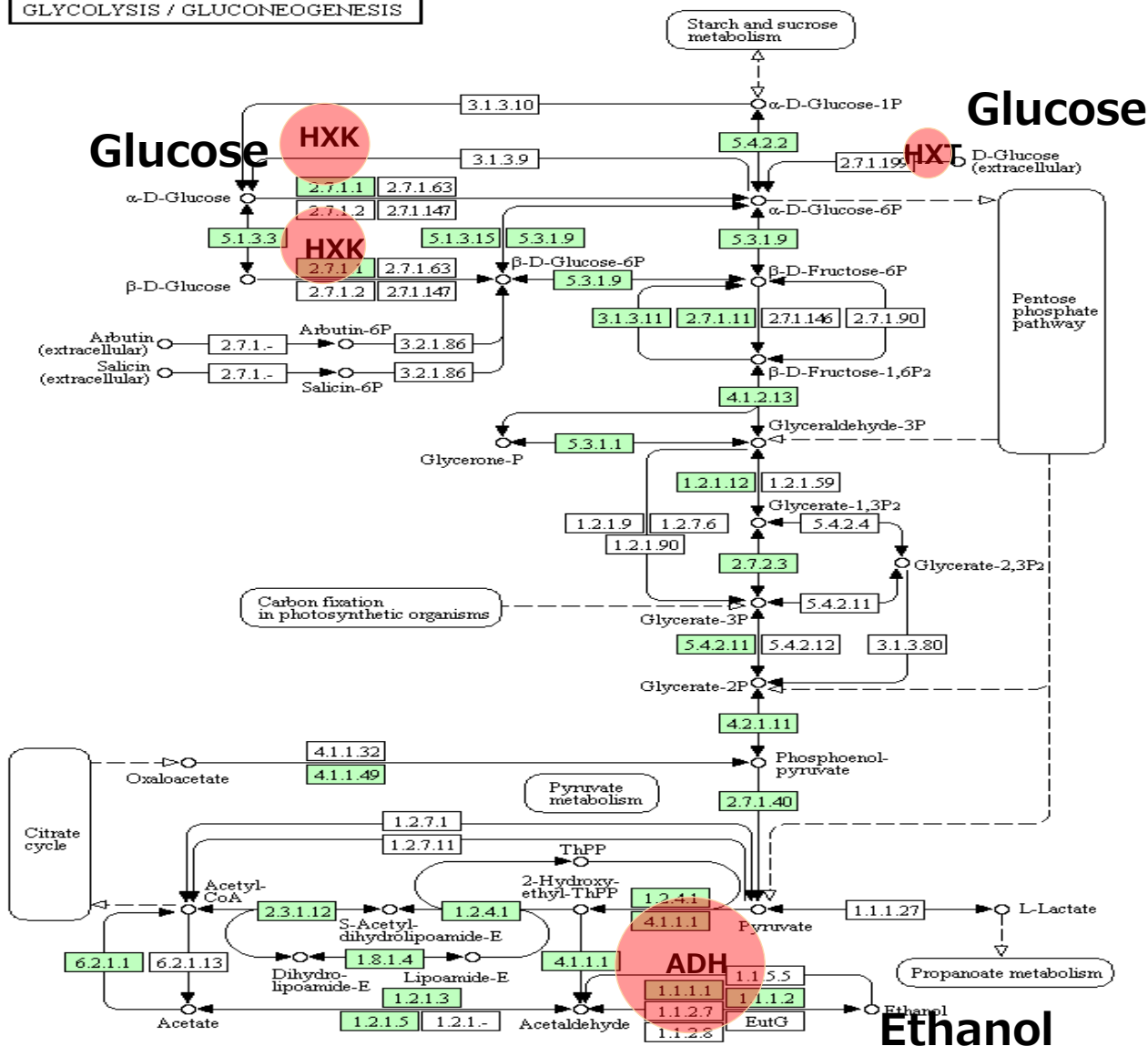
検索

変異配列を含むDBによって同定/定量された変異タンパク質, ペプチド, 変異部位

Sample	変異タンパク質数	変異ペプチド数	変異部位(箇所)	全同定タンパク質数	Up /down K2 vs K1 A2 vs A1
Yeast-K1	559	781	762	3602	82/136
Yeast-K2	469	631	617	3377	
Yeast-A3	557	771	753	3558	79/83
Yeast-A4	513	697	682	3331	

プロテオゲノム解析応用例 (Yeast, *S. Cerevisiae*)

GLYCOLYSIS / GLUCONEOGENESIS



Yeastプロテオゲノムによって同定されたアルコール高発酵株の特異的パスウェイ

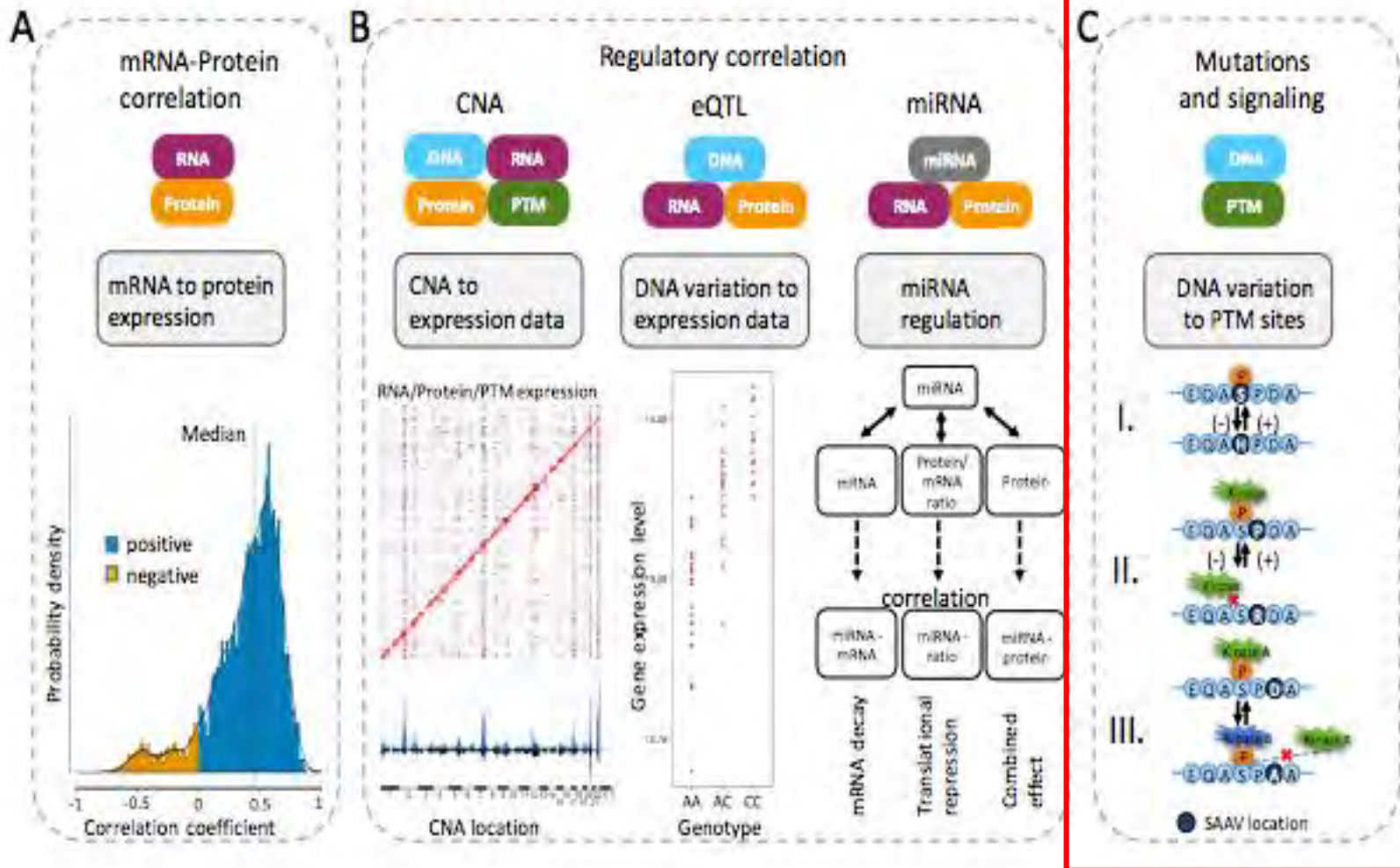
アルコール発酵状態
VS
通常培養状態

Green proteins were specifically up-regulated in glycolysis Pathway

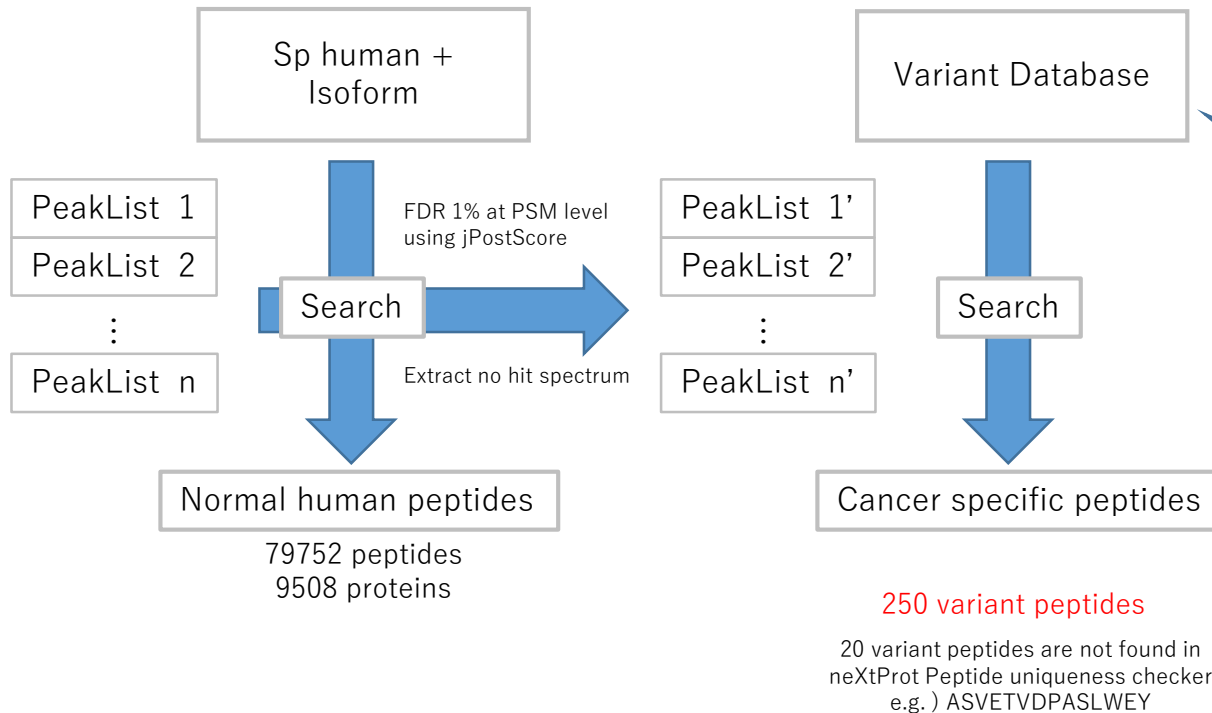
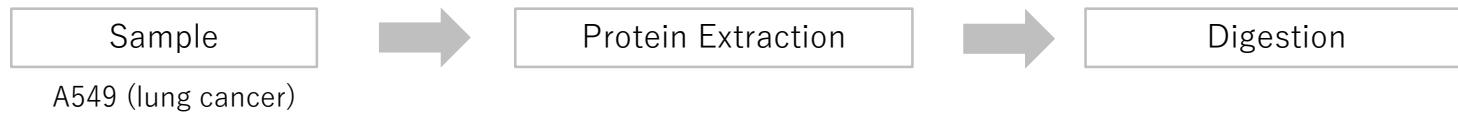
Pink proteins were identified via the mutant Database

プロテオゲノム解析の応用(参考まで)

Integrative analysis of genetic variants and PTM sites



Another example: Proteogenomics with DBKERO



DBKERO
Database of Gene-Expression of Regulating Peptides

Tools

- Genome Browser (Lathal)
- Search from Keyword or Genomic Position (human and mouse)
- Search from SNV-enriched Gene in Cancers
- TF Binding Site Search
- Transcription Factor Binding Site Search (Red)
- Pathway Map
- Human Pathway Map
- RDF for NGS Analysis Results (Trial) (Lung adenocarcinoma 26 cell lines: RNA-seq, ChIP-seq, SNV, BS-seq, TSS-seq)

News

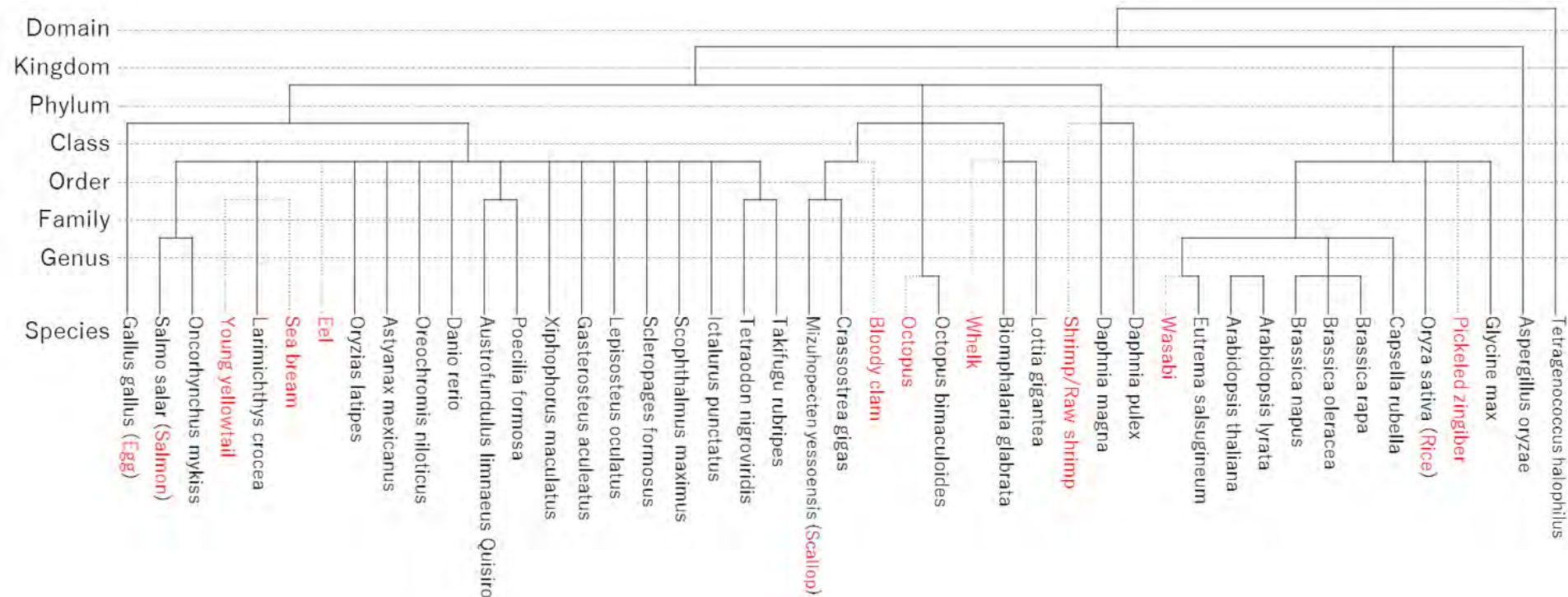
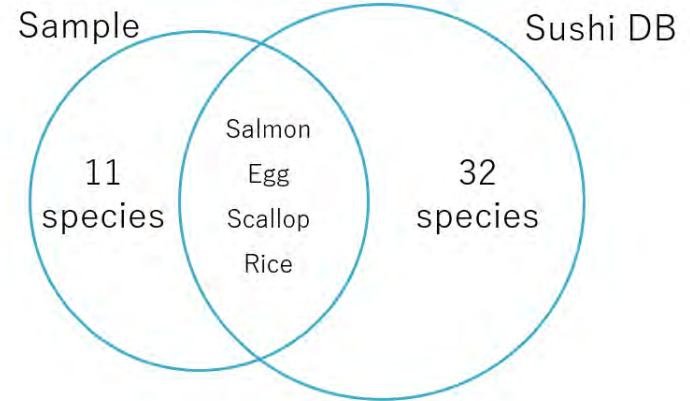
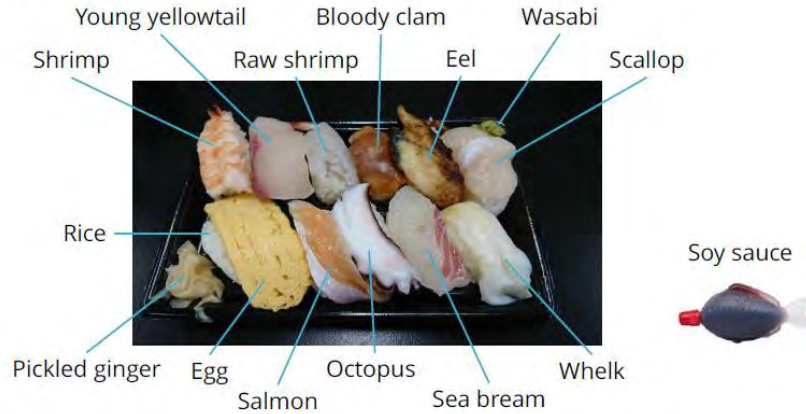
- 23 May 2019: New six OMAS studies were applied to genome browser: OMAS viewer was released (a2a30c).
- 27 Jun 2018: The function of "TF binding site search" was released (d6).
- 21 Dec 2017: New KERO (Release 1.1) opened.
- 1 Sep 2017: New C1 data of human lung adenocarcinoma cell line (LC-2/ad (triplicate)) are now available (See browser: Single cell -> C1 -> LC-2/ad (triplicate)).

From gene variant information of A549 (DBKERO)

<http://kero.hgc.jp/>

The Sushi Proteomics without Genome DB

The used sample



The Sushi Proteomics without Genome DB

Sequence around substitution

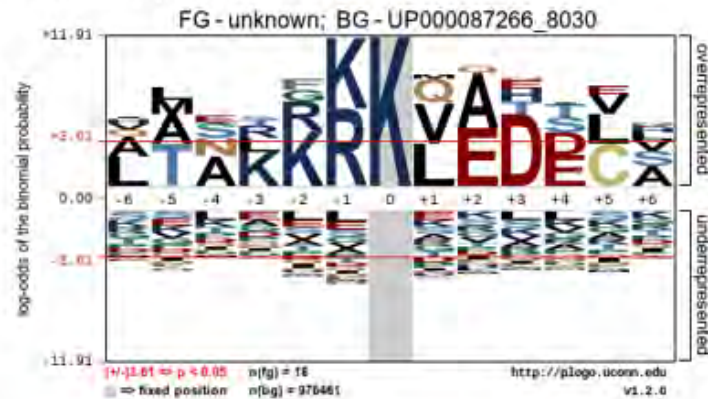
Salmon sample (Sushi DB)

Foreground

18 peptides substituted lysine to glycine

Background

Salmon proteomes

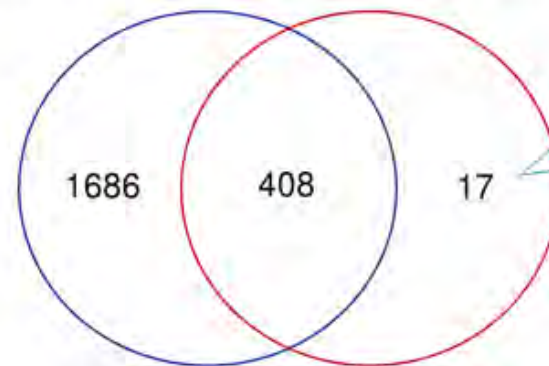


Sushi DB vs. modified Sushi DB(KK/RK $\rightarrow$ KG/RG)

Salmon sample (sushi DB)

sushi DB

Modified sushi DB

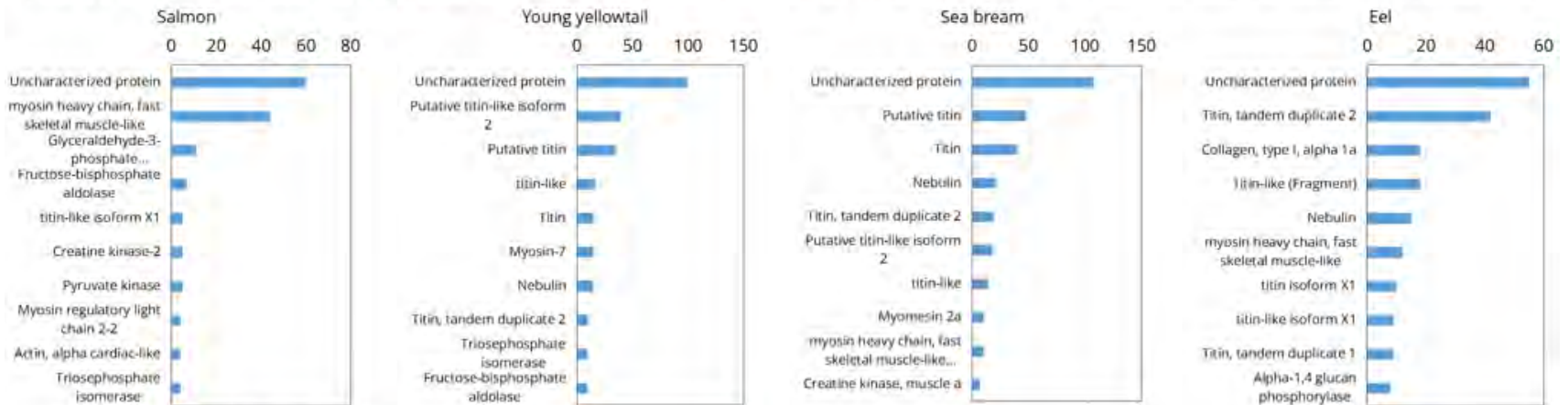


GVNMAHMCIVGSHAVNGVAR
GMEGDLNEMEIQLSHSNR
KLETDLVQVQGEVDDIVQEAR
GMEGDLNEMEIQLSQANR
GQADSVAEIGEIDNLQR
GGGSMQTVSSQFR
KMEGDLNEMEIQLSHSNR
KRGLEDECSSELK
KMEGDLNEM(Oxi)EIQLSHSNR
GVQHELEEAER
GDLYANNVLSGGTTMYPGIADR
GGSM(Oxi)QTVSSQFR
GVNMAHLCIVGSHAVNGVAR
GAITAMDVVYALK
GYGYTHLSSGDLR
GVAETELVDASER
GEFLQELLTTQCDR

The Sushi Proteomics without Genome DB

Frequently substituted protein

Top 10 proteins whose peptides were substituted from error tolerant search (y axis is the number of peptides) .



➤ Titin is likely to be substituted.

Sushi DB vs. Sushi DB + Titin all taxonomy DB (overlap of each unique peptides)



jPOST customizable database 'Slice'



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



データの絞り込み検索 (slicing)

- オントロジーで整理されたメタデータをもとに、ユーザが自由にデータを切り取ることが可能

Filters ▲

Species

* Homo sapiens

Tissue

Disease

cancer

Cell Type

cancer

neuroblastoma

Modification

breast cancer

Instrument

carcinoma

adenocarcinoma

pancreas cancer

Instrument Mode

Dataset Keyword

Protein Keyword

species

90.7%

disease

78%

データの絞り込み検索 (slicing)

- 絞り込まれたデータセットのリストから、一部、または全部を切り出し(Slice)

Dataset Protein

Showing 1 to 7 of 7 entries (filtered from 43 total entries)

☐	ID	Project ID	Project Title
☒	DS201_1	JPST000201	In-depth membrane proteomic study of breast cancer tissues
☒	DS205_1	JPST000205	Proteomic data of HCT116 cells
☒	DS206_1	JPST000206	Phosphoproteomic data of HCT116 cells
☐	DS207_1	JPST000207	Phosphoproteomics comparison between SW480 and SW620 cells
☐	DS207_2		between
☐	DS208_1		SW480 and
☐	DS208_2		SW480 and SW620 cells

必要なデータセットにチェックを入れ、“Add to Slice”ボタンをクリックすることで、Sliceをブラウザに記憶

Show 10 entries

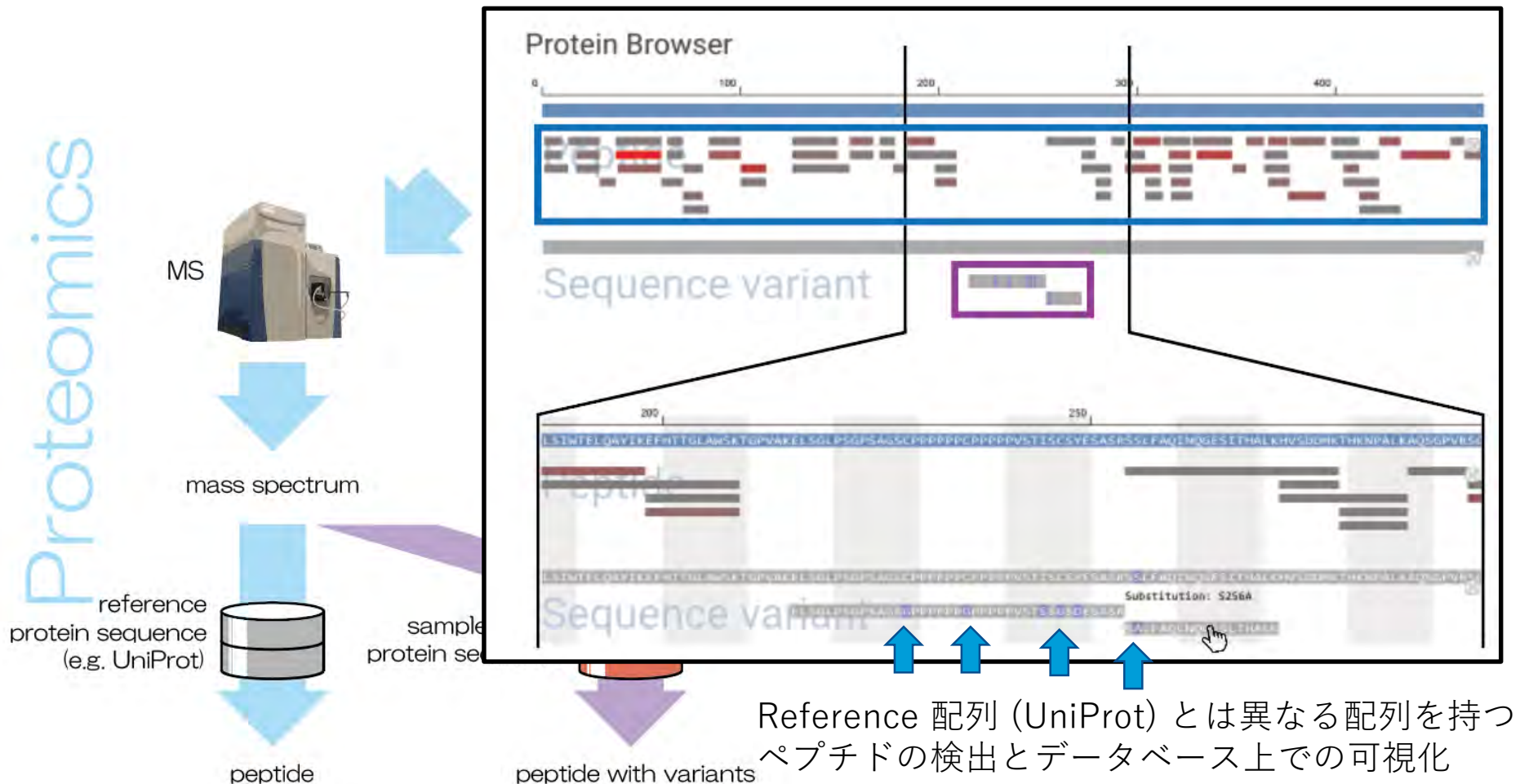
Exclude Add to Slice

Previous 1 Next

“ヒト”の”癌”で検索した例⁵⁰

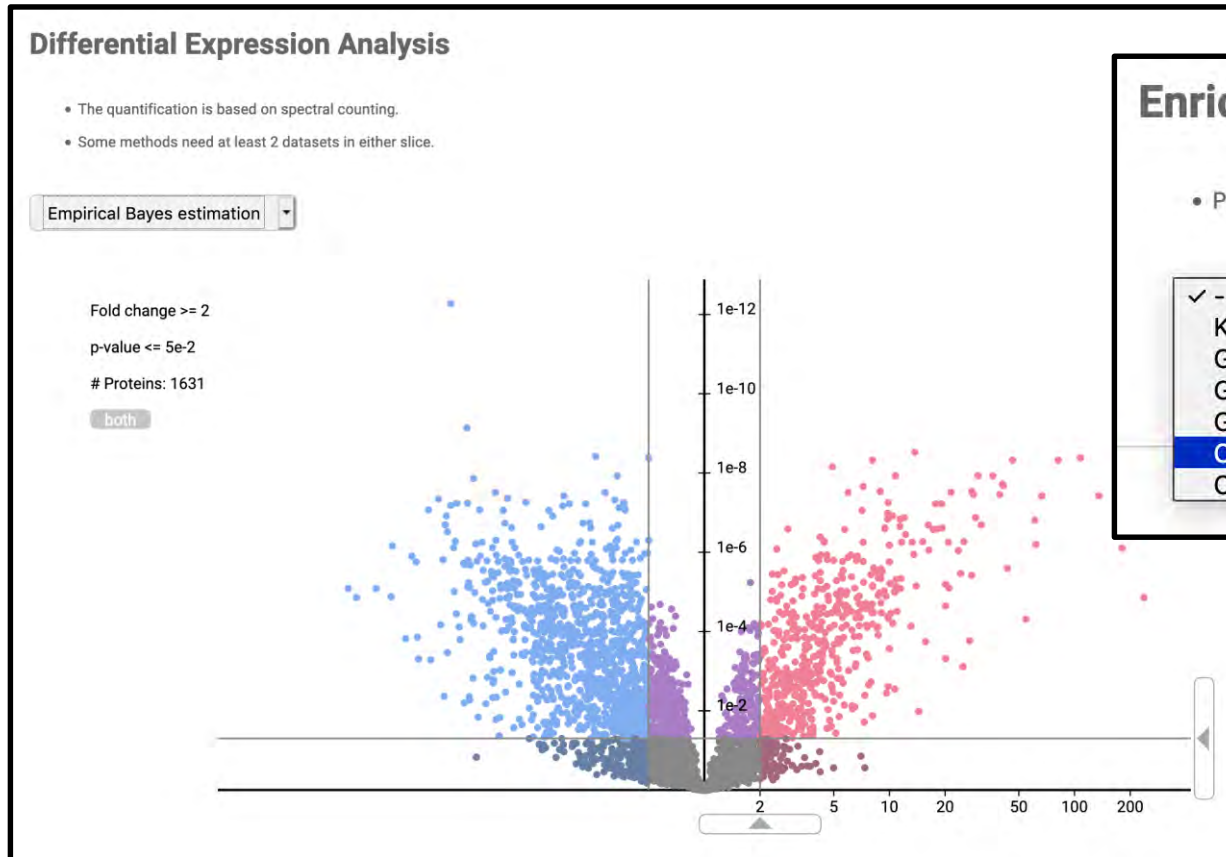
プロテオゲノミクス対応

解析プロトコル、データベースのプロテオゲノミクス対応を進めている



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



Enrichment Analysis

- Protein set enrichment analysis for selected proteins

✓ -- Select target --
KEGG Pathway
GO: biological process
GO: molecular function
GO: cellular component
ChIP-seq in ChIP-Atlas
ChIP-seq up/down comparison

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 - Enrichment Analysis

Analyze your data with public ChIP-seq data.

H. sapiens M. musculus R. norvegicus D. melanogaster C. elegans S. cerevisiae

1. Antigen Class

All antigens (43086)
DNase-seq (1632)
Histone (10578)
RNA polymerase (1532)
TFs and others (9868)
Input control (5080)
Unclassified (10053)
No description (4343)

4. Select your data

Genomic regions (BED) or sequence motif
☒ Gene list (Gene symbols) ①

COL6A3
FN1
PRPH
MVP
AHNAK2
ANPEP
HSPG2
COL6A2

選択... ファイルが選択されています
Choose local file Try with example

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

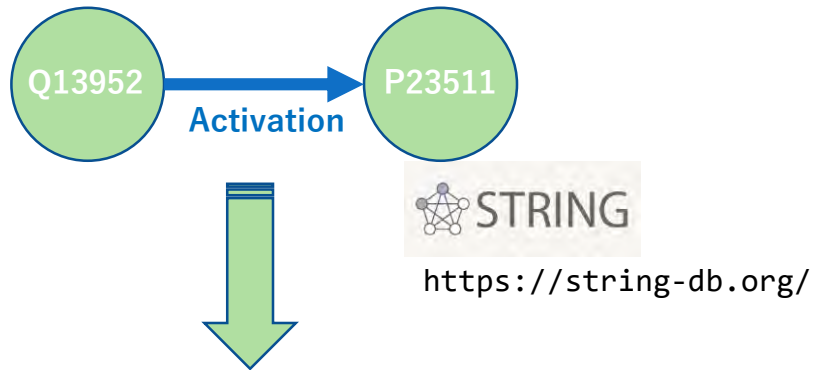
変動遺伝子リスト

ChIP-Atlas サイト

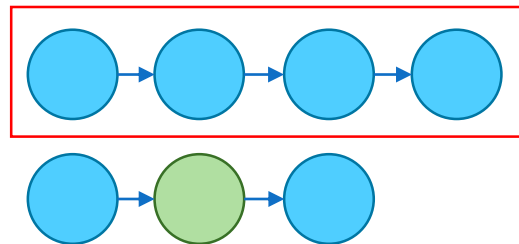
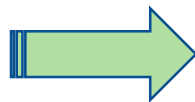
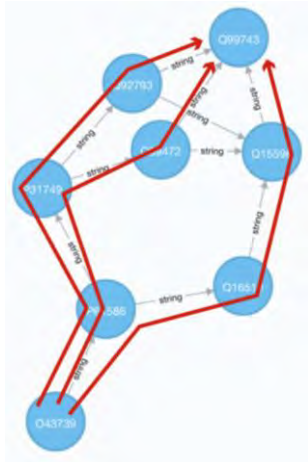
変動遺伝子の発現を制御している転写因子などのリスト

ネットワーク解析ツールの開発

1. STRING databaseからタンパク質間相互作用データを取得



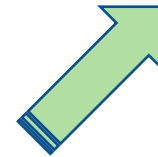
2. 可能なパスの探索（グラフデータベース）



3. 取得したパスにデータをマッピングし、タンパク質発現が連続して変化（Up/Down）しているパスの抽出



4. ドナーごとに変異タンパク質からのタンパク質発現ネットワークを可視化



データ利用者と連携：UniProt

Cross-references between UniProt and jPOST

UniProt UniProtKB

BLAST Align Retrieve/ID mapping Peptide search

UniProtKB - P40925 (MDHC_HUMAN)

Display BLAST Align Format Add to basket History Other tutorials and

Entry **Protein** Malate dehydrogenase, cytoplasmic

Publications **Gene** MDH1

Feature viewer **Organism** Homo sapiens (Human)

Feature table **Status** Reviewed - Annotation score: ●●●●● - Experimental evidence at protein level¹

None

Display

Entry

Publications

Feature viewer

Feature table

None

Modified residue¹ 332 Phosphoserine By similarity

Modified residue¹ 333 Phosphoserine Combined sources

Post-translational modification¹

ISGylated. 1 Publication

Acetylation at Lys-118 dramatically enhances enzymatic activity and promotes adipogenic differentiation. 2 Publications

Keywords - PTM¹

Acetylation, Methylation, Phosphoprotein, Ubl conjugation

Proteomic databases

EPD ¹	P40925
jPOST ¹	P40925
PaxDb ¹	P40925
PeptideAtlas ¹	P40925
PRIDE ¹	P40925
ProteomicsDB ¹	55384
	55385 [P40925-2]
TopDownProteomics ¹	P40925-1 [P40925-1]

jPOSTdb Search Slides Contacts Help

Protein: P40925

Protein Name Malate dehydrogenase, cytoplasmic

Protein ID MDHC_HUMAN

Gene Name MDH1

Accession P40925

Length 334 aa

Sequence

MSSEIRVLTGAAGQIAYSLLYSIGNSVFGKDDPITLVLDITPMGVLUGVLMELDICALPLLKDVIAIDKERVAFKDLDAVLV
GSRPRREGNEKDKLKAANKLFPKQQAALDNYAKFSVNVLYGVDPAMTICLTASKSAFSPKPEPSELFLDNNWAKAGIALKLGVI
ANDVKXNVIWGNHSTQYDPVNNHAKVLOGKEGVGYEALKDDSWLKGEFVTTVQHRGAIVKARKLSSAMSAKAICDVRDIDWF
PEGEFVSMGVISOGNSYGVDPDLLYSFPVVIKNTWKFVEGLPINDFSREKMDTAKELTEEKESAFEFLLSA

Location Chromosome 2

Statistics

# peptides:	42
# spectra:	1,834
# unique peptides (UniProt entry level):	31
# unique peptides (gene name level):	42

for Absolute GeneID: 4190

Quantification search in iMPAQT

Protein Browser

1 100 200

Peptide

-- Add view --

Peptide Sharing

shared in isoforms

P40925

1 145 19 5 1 142 10 136 30 112 73 35 13

82 155 16 29 22 32 118 41 116 72 33 142

2 20 10 210 52 118 50 162 42

749 2 21 20 115 21

16 16

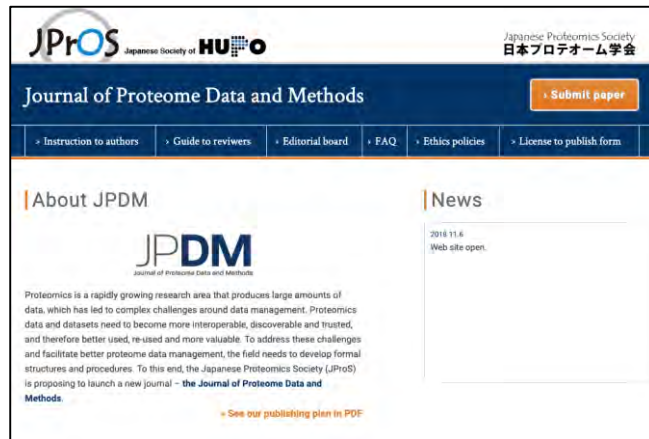
15 16

オントロジーの深化と強化 ～HUPO Proteomics Standard Initiative (PSI) を通じた標準化～

毎週木曜 or 金曜深夜1時から（夏期は0時から）ウェブミーティング

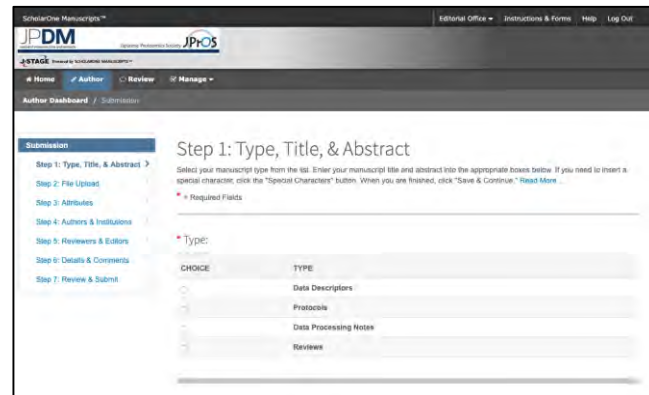
- Universal Spectrum Identifier (USI)
 - すべてのスペクトルに ID を付与するための仕様
 - mzspec:<collection>:<msRun>:<indexType>:<indexNumber>
- ProteomeXchange guideline
 - 再解析データの取り扱い
 - データのライセンス
 - ヒトデータの取り扱い
- PROteomics eXpression Interface (PROXI)
 - 各プロテオームリポジトリ間でデータを交換するためのAPI仕様
- Spectral Library format
 - Spectral Libraryを登録するためのファイルフォーマットの策定
 - Spectral Libraryを登録するためのメタデータ項目の選定
 - メタデータを記述するための用語の策定

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



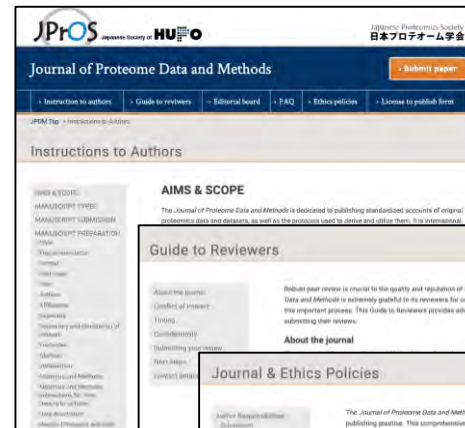
ウェブサイト

<https://www.jhupo.org/jpdm/>



投稿システム (ScholarOne)

<https://mc.manuscriptcentral.com/jpdm>



Instructions to Authors

Guide to Reviewers

Ethics Policies

FAQ

License to Publish form

License to Publish	
Manuscript Number	Journal Name (the "Journal")
Article Title (the "Title")	
Authors (the "Author(s)") (Please list all authors on a new sheet if necessary)	
The Japanese Proteomics Society (the "Society") will consider publishing the Article in the Journal pursuant to the terms below. The Authors grant to the Society the right and 4 worldwide, irrevocable license to:	
1) Publish, disseminate, copy, display, store, commercially exploit and otherwise use the Article, including any supplementary material, in all forms and all media (now or at any future time);	
2) Create any translations, extracts or derivative works based on the Article and exercise all of the rights set forth in (1) in any such translations, extracts or derivatives; and	
3) Sub-license others to do all or any of (1) and (2).	

- 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

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

Editorial Board members

Juan Antonio Vizcaino	EMBL-EBI, UK	PRIDE Archive* leader
Eric Deutsch	Institute for Systems Biology, USA	PeptideAtlas/PASSEL* leader
Nuno Bandeira	UCSD, USA	MassIVE* leader
Yunping Zhu	Beijing Proteomics Center, China	iProX* leader
Brendan MacLean	U. Washington, USA	Panorama Public* leader
Henning Hermjacob	EBI/Beijing Proteomics Center	PRIDE* and iProX* adviser
Jau-Song Yu	Chang Gung U., Taiwan	President of TPS
Je-Yoel Cho	Seoul Nat. U., Korea	President of KHUPO
Chung Ching Ming	Nat. U. Singapore	President of AOHUPO
Yu-Ju Chen	Academia Sinica, Taiwan	HUPO council
Lukas Kall	Royal Institute of Technology, Sweden	Percolator developer

2019年度予定

- お手本論文準備中
- 2019 年 9 月創刊予定

上記に加えてJPrOS理事20名

* PX member

メタデータキュレーション自動化に向けて

メタデータのマニュアルキュレーション

リポジトリ-jPOSTrepo
Xmlファイル



メタデータを抽出
生データを再解析セット毎に分類
(マニュアル)



サンプル情報
質量分析測定条件



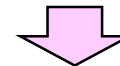
メタデータの確認・修正
参考文献からの情報付加
(マニュアル)



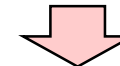
メタデータ

メタデータキュレーションの半自動化

リポジトリ-jPOSTrepo
Xmlファイル、関連論文



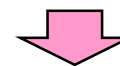
メタデータを抽出
生データを再解析セット毎に分類
(自動化)



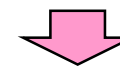
サンプル情報
質量分析測定条件



自動言語処理による
参考文献からの情報付加



メタデータの確認・修正



メタデータ

メタデータ自動抽出

jPOSTrepoに登録されているメタデータ

Experimental presets

Sample	Fractionation	Enzyme/Mod.	MS mode
Title: Hela-S3 Species: Homo sapiens (Human) Tissue: Hela-S3 cell Cell type: cervical epitheloid carcinoma Disease: cervix carcinoma Note:	(No profile)	Title: ATPome_trypsin Enzyme: Trypsin/P Fixed modification: Carbamidomethyl (C) Variable modification: Biotin:Thermo-88310 (K), Oxidation (M) Taxonomy: Homo sapiens (Human) Note:	Title: QE_SILAC K8R10 Instrument: Q Exactive Instrument mode: DDA-high res. Purpose: Relative quantification Quantification platform: Precursor ion stable isotope labeling Plex: 2

Xmlファイル

This XML file does not appear to have any data within the element(s) listed below.

```
<?xml version="1.0" encoding="UTF-8"?>
<JPOST-XML>
  <Sample>
    <Title>Hela-S3</Title>
    <Species>Homo sapiens (Human)</Species>
    <Tissue>Hela-S3 cell</Tissue>
    <Cell type>cervical epitheloid carcinoma</Cell type>
    <Disease>cervix carcinoma</Disease>
    <Note></Note>
  </Sample>
  <Fractionation>
    <Title>(No profile)</Title>
  </Fractionation>
  <Enzyme/Mod.>
    <Title>ATPome_trypsin</Title>
    <Enzyme>Trypsin/P</Enzyme>
    <Fixed modification>Carbamidomethyl (C)</Fixed modification>
    <Variable modification>Biotin:Thermo-88310 (K), Oxidation (M)</Variable modification>
    <Taxonomy>Homo sapiens (Human)</Taxonomy>
    <Note></Note>
  </Enzyme/Mod.>
  <MS mode>
    <Title>QE_SILAC K8R10</Title>
    <Instrument>Q Exactive</Instrument>
    <Instrument mode>DDA-high res.</Instrument mode>
    <Purpose>Relative quantification</Purpose>
    <Quantification platform>Precursor ion stable isotope labeling</Quantification platform>
    <Plex>2</Plex>
  </MS mode>
</JPOST-XML>
```

プロテオミクス関連論文 (JPDM)

2. Materials and methods

2.1. Cell cultures

Hela-S3 cells were grown in DMEM with 10% fetal bovine serum plus antibiotics in 10% CO₂ at 37 °C. For **SILAC labeling**, Hela-S3 cells were cultured in DMEM supplemented with 10% dialyzed fetal bovine serum and either 28.0 mg/L normal isotopic abundance arginine and 48.7 mg/L normal isotopic abundance lysine (Light) or 28.0 mg/L arginine with six ¹³C and four ¹⁵N atoms and 48.7 mg/L lysine with six ¹³C and two ¹⁵N atoms (Heavy) [2]. Labeling efficiency was confirmed after five passages.

2.2. Sample preparation

Approximately 5 × 10⁸ **Hela-S3 cells** after each of the **light** and **heavy SILAC labeling** conditions were harvested with 5 ml of ice-cold lysis buffer (25 mM Tris pH 7.5, 150 mM NaCl, 1% CHAPS, 1% Nonidet P-40, protease inhibitor cocktail and phosphatase inhibitor cocktail). Cell lysate was incubated on ice for

2.3. LC-MS/MS analysis

All nanoflow LC-MS/MS experiments were performed on a **Q Exactive mass spectrometer** (Thermo Scientific). Data were acquired in data-dependent mode using Xcalibur software. The precursor ion scan MS spectra (m/z 350–1800) were acquired in the Orbitrap with **70,000 resolution** after accumulation of ions to a 3 × 10⁶ target value. The twelve most intensive ions were isolated and fragmented in the octopole collision cell by higher-energy collisional dissociation (HCD) with a maximum injection time of 120 ms and 35,000 resolution. In data-dependent LC/MS² experiments dynamic exclusion was used with 20-s exclusion duration. Typical mass spectrometric conditions were as follows: spray voltage, 2 kV; no sheath and auxiliary gas flow; heated capillary temperature, 250 °C; normalized HCD collision energy 25%. The MS/MS ion selection threshold was set to 5 × 10⁴ counts. A 3.0 Da isolation width was chosen.

2.4. Data analysis

All raw files were processed by MaxQuant software suite (version 1.3.0.5) supported by the Andromeda search engine for peptide identification [4]. MaxQuant was used to score peptides for identification based on a search with an initial allowed mass deviation of the precursor ion of up to 7 ppm. The allowed fragment mass deviation was 20 ppm. Peak lists were searched against a concatenated forward and reversed version of the **UniProt human database** (release 2011_11) combined with 262 common contaminants was performed using the Andromeda search engine. Enzyme specificity was set as C-terminal to Arg and Lys with allowed cleavage at proline bonds and a maximum of three missed cleavages. **Carbamidomethylation of cysteine was set as a fixed modification** and **methionine oxidation and desthiobiotinylated lysine (+196.1212 Da) as variable modifications**. Peptides and proteins were accepted with a false discovery rate of < 1%. Feature matching between raw files was enabled for ATP competition datasets using a retention time window of 2 min. We removed peptides that were inhibited by < 80% at 10 μM ATP in order to extract ATP-competitive peptides/sites from label-free quantitation data. Peptides with quantitation values that were not reproducible were also rejected, based on an average of < 0.5 for three correlation coefficients of intensity values in triplicate experiments (correlation coefficients of experiments 1 and 2, 2 and 3, 3 and 1); 89.6% of peptides were rejected by this filter. MS spectra for all ATP-competitive peptides were checked for missed assignments

Xmlファイルからのメタデータ抽出

メタデータ

Sample

Species	Human	9606
Sample Type	Cell Line	C16403
Cell line	Hela-S3	CLO_0003696
Organ		
Disease class		
Disease		
Note	Enrichment of kinases in Hela	

Fractionation

peptide fractionation	Enrichment of ATP binding proteins
peptide fractions	1
peptide replicates	1

Enzyme_Mod

Fixed Modification	Carbamidomethyl (UNIMOD_4)	Chemical derivative	C
Variable Modification	Oxidation (M)	UNIMOD_35	Artefact
Variable Modification	Biotin:Thermo-8831	UNIMOD_1031	Chemical derivative
Variable Modification	Label:13C(6)15N(4)	UNIMOD_267	Isotopic label
Variable Modification	Label:13C(6)15N(2)	UNIMOD_259	Isotopic label
Enzyme	Trypsin/P	MS:1001313	
Taxonomy	Homo sapiens	9606	

MS_Mode

MS instrument	Q Exactive	MS:1001911
Instrument Mode	DDA-high res	JPO_005
Purpose	Relative quan	JPO_011
Quantification Platform	Precursor ion	JPO_014
Quantification Method	SILAC quantit	MS:1001835
Plex number		2

自然言語処理による

1. サンプル
2. サンプル処理
3. データベース設定
4. 質量分析測定条件

等の情報取得

2019年度の達成目標・タスク

達成目標	タスク
リポジトリの機能強化と新機能開発	・ Help機能の充実 ・ PXCのXML対応 ・ データの修正に対する対応 ・ 大規模メタデータ一括インポートシステムの開発 ・ SRM, PRM, DIAデータ対応 ・ グライコプロテオミクスデータ対応
再解析ワークフローの機能強化と新機能開発	・ jPOSTスコアの確立 ・ ワークフローの全自動化 ・ プロテオゲノミクスデータの再解析 ・ メタゲノム配列データベースの整備
データベースツールの深化とプロテオゲノミクスデータの可視化およびネットワーク解析ツールの開発	・ Proteoform解析ツールの開発 ・ 絶対量やバリエーション選択的なスライスデータベースの深化 ・ プロテオゲノミクス用スライスデータベース ・ 既知/新規パスウェイ解析ツールの開発
オントロジーの深化と強化	・ オントロジーの充実 ・ プロテオゲノミクスデータに対する対応
キュレーションの深化と強化	・ メタボロームデータ対応 ・ プロテオゲノミクスデータ対応 ・ 専門論文誌の発行

現在までの進捗

研究開発テーマ	タスク	担当グループ	2018					2019		
リポトリの機能強化と新機能開発	Help機能の充実	奥田G								
	PXCのXML対応	奥田G								
	データの修正に対する対応	奥田G								
	大規模メタデータ一括インポートシステムの開発	奥田G、荒木G								
	SRM, PRM, DIAデータ対応	奥田G、松本G								
	グライコプロテオミクスデータ対応	奥田G、								
再解析ワークフローの機能強化と新機能開発	jPOSTスコアの確立	石濱G								
	ワークフローの全自動化	石濱G								
	プロテオゲノミクスデータの再解析	石濱G								
	メタゲノム配列データベースの整備	奥田G、石濱G								
データベースツールの深化とプロテオゲノミクスデータの可視化およびネットワーク解析ツールの開発	Proteoform解析ツールの開発	五斗G、松本G								
	絶対量やバリエーション選択的スライスデータベースの深化	松本G、五斗G								
	プロテオゲノミクス用スライスデータベース	五斗G、荒木G								
	既知/新規パスウェイ解析ツールの開発	奥田G、石濱G								
オントロジーの深化と強化	オントロジーの充実	五斗G								
	プロテオゲノミクスデータに対する対応	五斗G								
キュレーションの深化と強化	メタボロームデータ対応	荒木G、松本G								
	プロテオゲノミクスデータ対応	荒木G、松本G								
	専門論文誌の発行	荒木G、石濱G、五斗G								

現在までの進捗まとめ

- Journal of Proteome Data and Methods: 本年度もJ-Stage コンサルティング対象に採択→科研費再チャレンジへ準備中
- 各国の状況把握：第1回International Proteogenome Workshopの開催
→TPS, KHUPO、Tokyo Workshopへの参加（終了）
→ AOMSC(2020.1), AOHUPO(2020.3)
- PSI/PXCとの連携：USI/PROXI/再解析/Spectral library/CC-BY
- EBIとの連携：グラント申請再チャレンジへ準備中（メタプロテオミクス）
- ICPC-Taiwanとの連携（月例ミーティングの開催）
- スウェーデンチームとの連携（スペクトルライブラリー、jPOSTスコア）
- UniProtとの連携（クロスリファレンス化）
- ICPC-Japan: NCCとの連携（データ解析担当）
- 糖鎖DB, ChIP Atlas, DBKEROとの連携開始
- 農学プロテオームDB(イネ、ダイズ) との連携模索中