



# プロテオーム統合データベース jPOSTの構築

**代 表 : 石濱 泰 (京大院薬)**

**主な共同研究者 : 五斗 進 (京大化研)**

**荒木 令江 (熊本大学)**

**松本 雅記 (九州大学)**

**奥田 修二郎 (新潟大学)**

**河野 信 (DBCLS)**



Japan Proteome Standard  
Repository/Database

# 本日の発表

1. 全体の進捗： 石濱
2. リポジトリ開発： 奥田
3. 再解析プロトコル開発： 吉沢
4. 今後の計画： 石濱
5. ディスカッション： 全員

**随時ご質問ください**



# プロテオーム解析からのみ得られる情報

---

- 翻訳後修飾
- 発現プロファイル（臓器、組織、細胞内）
- 絶対発現量
- タンパク-タンパク相互作用

タンパク質は、発現調節（転写後調節）、局在、翻訳後修飾、分解、タンパク相互作用等によって、ダイナミックに姿を変え、場所を変え、消えたり現れたりを繰り返す。

**タンパク質は生物機能発現の立役者！  
生物現象や疾病のマーカーや薬の標的となる。**



**プロテオームはLS統合DBに欠けてはいけない必須情報**

# 研究開発の背景

## プロテオームDB: 世界の情勢

- HPP-HUPO  $\Rightarrow$  遅々として進まず
- ProteomicsDB (LCMSデータ)  $\Rightarrow$  Nature, 2014
- Human Protein Atlas (ヒト抗体ライブラリー)  $\Rightarrow$  Science, 2015

Journal of  
**proteome**  
research

[dx.doi.org/10.1021/pr500572z](https://doi.org/10.1021/pr500572z) | *J. Proteome Res.* 2014, 13, 3854–3855

Letter

[pubs.acs.org/jpr](https://pubs.acs.org/jpr)

## Analyzing the First Drafts of the Human Proteome

Iakes Ezkurdia,<sup>†</sup> Jesús Vázquez,<sup>§</sup> Alfonso Valencia,<sup>‡</sup> and Michael Tress<sup>\*,‡</sup>

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.

# Nature 2014, the Human Proteome

---

Nature. 2014 , DOI: [10.1038/nature13302](https://doi.org/10.1038/nature13302), PMID: [24870542](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/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

# 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 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

17,294 genes (84%)



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

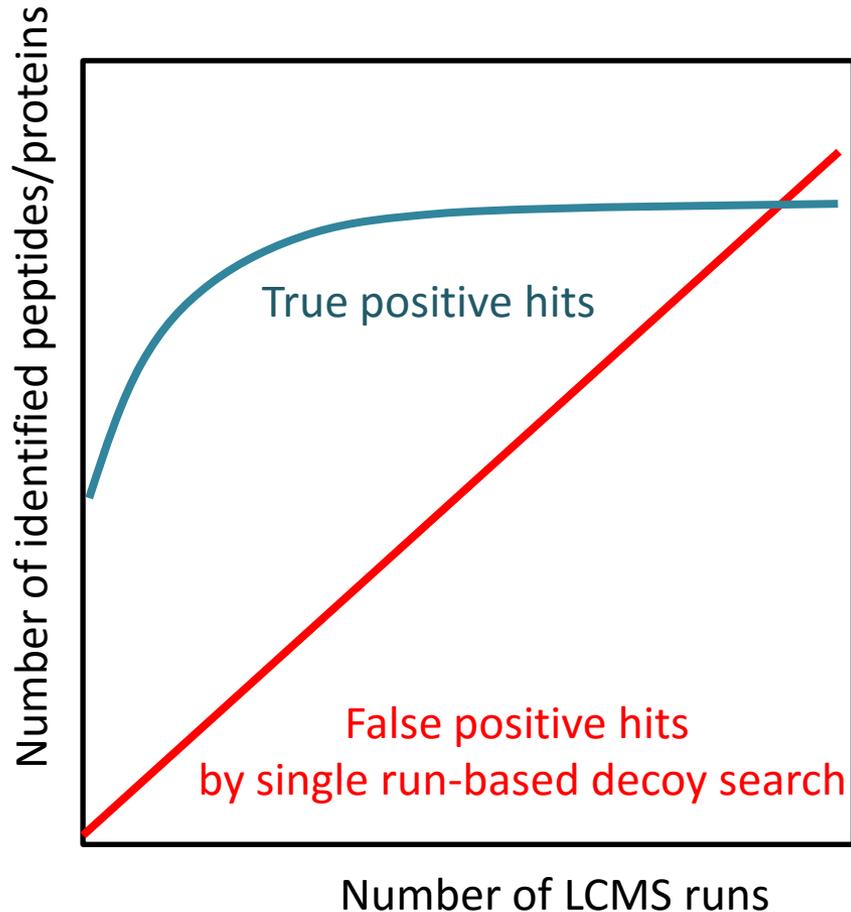
11,206 genes (57%)

1/3は偽ヒットだった。。。。



Juergen Cox (Max Planck)

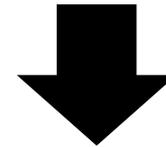
# ProteomicsDBの問題点とjPOSTの戦略



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解析データの標準化が必須

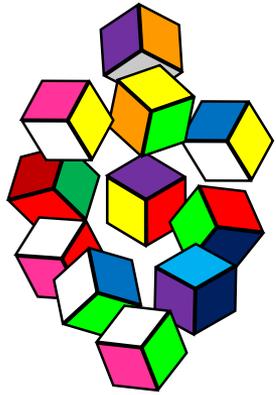
- 世界標準リポジトリ (PX)の採用
- 研究機関・プロジェクト毎に異なるフォーマットや解析法、信頼度評価法やアノテーションの標準化。
- 全データに対する統一したフィルタリング



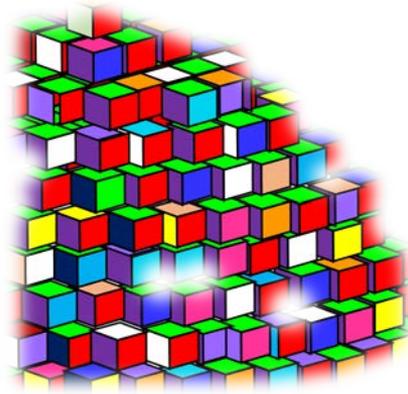
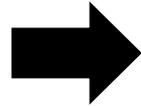
多彩な生物種・翻訳後修飾・絶対発現量も含めた世界初の横断的統合プロテオームDB

jPOST (Japan ProteOme **S**Tandardization Repository/Database)

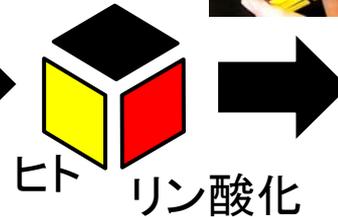
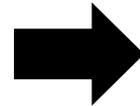
# jPOSTの構成 (ルービックキューブ型DB)



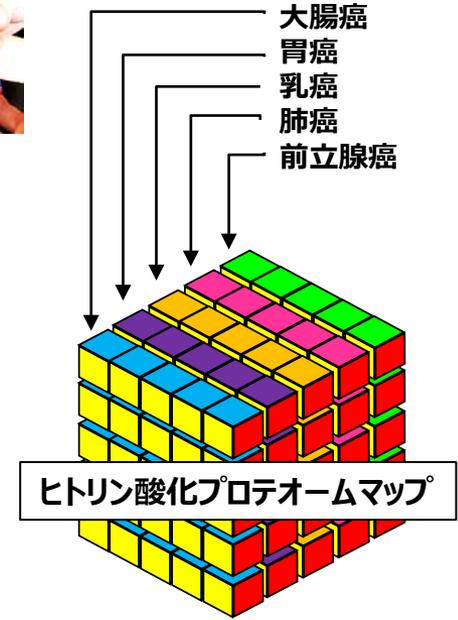
各プロジェクト毎のDB  
「Cube」



Cubeを集積した全DB  
「Globe」

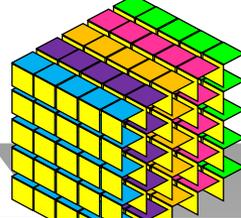


間引きとアライメント



Globeから抽出・調整した  
「Slice」

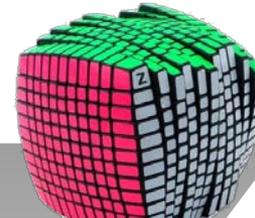
フォーカスドDB  
(Slice)



疾患別ヒトリン酸化  
プロテオームマップ

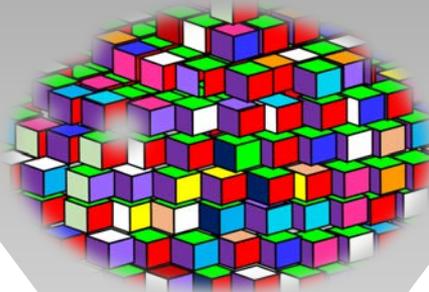


生物種別代謝酵素PTMs  
絶対発現量マップ



ユーザーカスタムマップ

集積DB  
(Globe)



例えば：ヒトとマウスの疾病別の転写因子群の発現量、リン酸化、アセチル化を見たい

プロジェクトDB  
(Cube)



標準化 (再解析)



リポジトリ



Japan Proteome Standard  
Repository/Database

測定生データ

抗体

質量分析

電気泳動

試料

臨床検体

植物

モデル生物

培養細胞

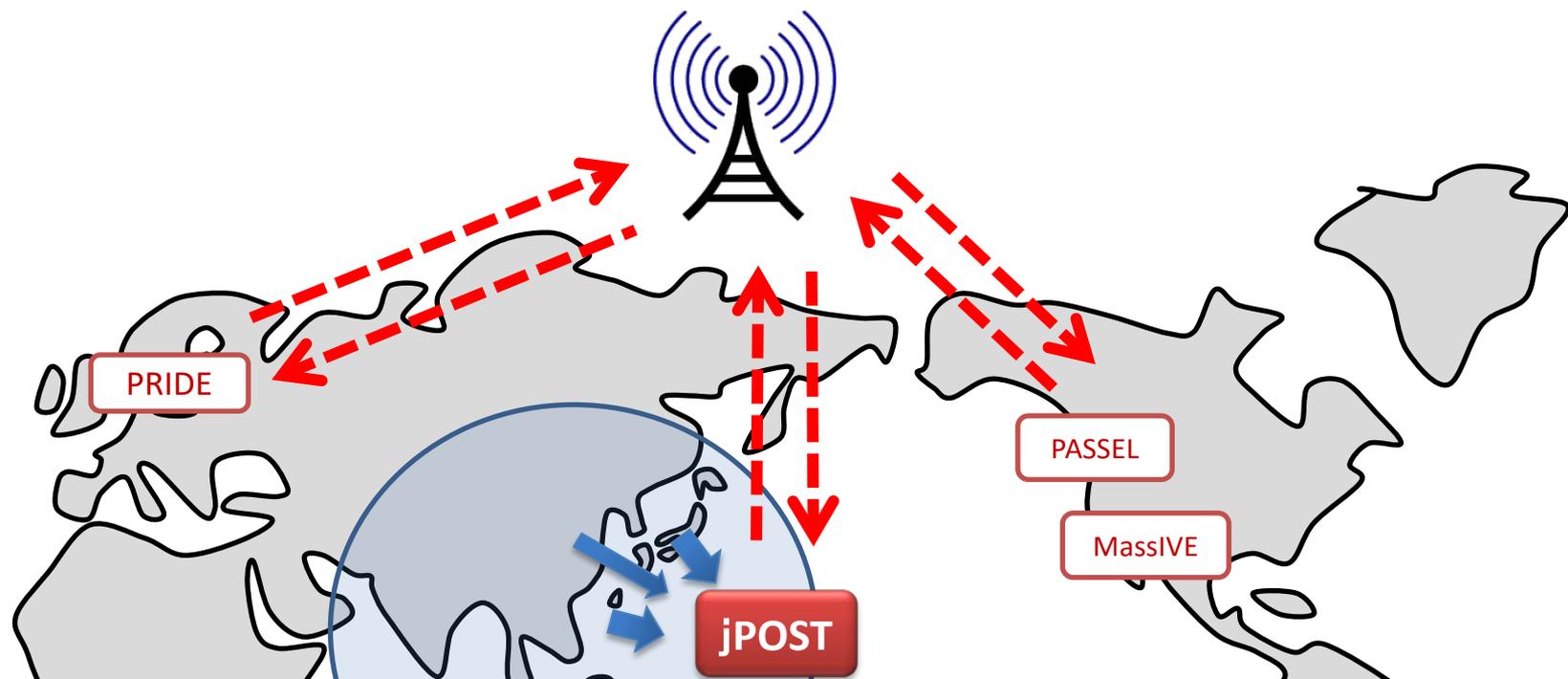
細菌

希少生物

# jPOSTの特徴

---

- データ標準化ステップを含む。
- ProteomeXchange コンソーシアムに参加。  
(開発費の抑制、データソースの確保、持続性)
- 段階ごとのDBシステム
- カスタマイズ可能なSlice DB
- 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, HUPPO)



- **ProteomeXchange** コンソーシアムに参加し、世界共通のIDを取得します
- アジア・オセアニア地域で唯一の国際リポジトリです

# ProteomeXchange: 3,802 datasets up until 1<sup>st</sup> April, 2016

## Origin:

885 USA  
465 Germany  
342 United Kingdom  
264 China  
194 France  
158 Netherland  
136 Canada  
126 Switzerland  
107 Denmark  
104 Spain  
99 Australia  
95 Japan  
72 Belgium  
68 Austria  
63 Sweden  
61 India  
51 Norway  
43 Taiwan  
30 Italy  
29 Brazil  
28 Singapore  
28 Finland  
27 Ireland  
27 Russia  
26 Israel ...

## Type:

2429 PRIDE partial  
1016 PRIDE complete  
250 MassIVE  
84 PeptideAtlas/PASSEL complete  
23 Reprocessed

## Datasets/year:

2012: 102  
2013: 527  
2014: 963  
**2015: 1758**  
2016: 452

## Publicly Accessible:

**1973 datasets, 52% of all**  
91% PRIDE  
885 5% MassIVE  
4% PASSEL

## Top Species studied by at least 20 datasets:

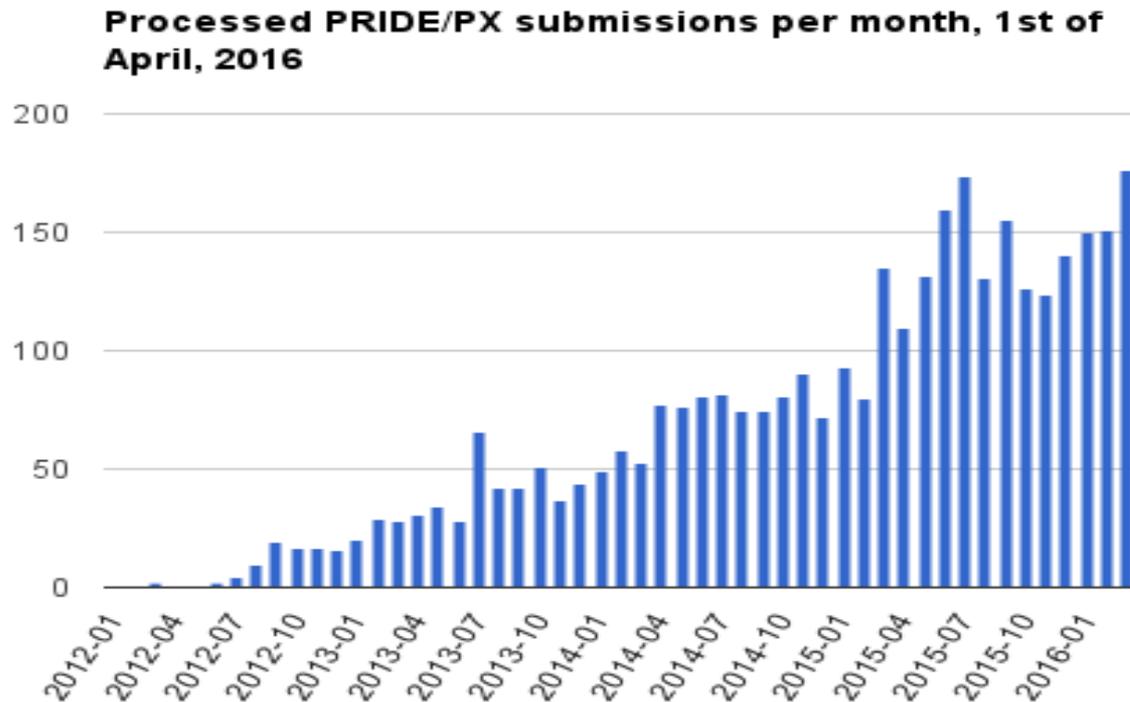
1526 *Homo sapiens*  
485 *Mus musculus*  
150 *Saccharomyces cerevisiae*  
121 *Arabidopsis thaliana*  
102 *Rattus norvegicus*  
86 *Escherichia coli*  
44 *Bos taurus*  
35 *Drosophila melanogaster*  
32 *Glycine max*

~ 700 species in total

## Data volume:

Total: ~220 TB  
Number of all files: ~560,000

# PRIDE Archive submitted datasets up until 1<sup>st</sup> April, 2016



- In the last year: **~150 submitted datasets per month**

# 研究開発体制

代表：石濱 泰（京大院薬）

## 主な共同研究者：

- 五斗 進（京大化研）
- 奥田 修二郎（新潟大学）
- 河野 信（DBCLS）
- 荒木 令江（熊本大学）
- 松本 雅記（九州大学）

## 連携研究者 & 保有データ：

| 連携研究者 | 所属                                | アドバイザー | 所属  |
|-------|-----------------------------------|--------|---|
| 朝長 毅  | 医薬基盤研、JHUPO 会長、<br>HUPO・AOHUPO 理事 | 平野 久   | 横浜市立大学  |
|       |                                   | 戸田 年総  | 横浜市立大学  |
| 中神 弘史 | 理化学研究所 環境資源科学研究センター               | 村井 純   | 慶應義塾大学  |
| 近藤 格  | 国立がん研究センター                        | 成松 久   | 産業総合技術研究所、<br>統合化プログラム糖鎖 DB<br>(JCGGDB) 研究代表者 |
| 尾野 雅哉 | 国立がん研究センター                        |        |   |
| 中林 潤  | 横浜市立大学                            |        |   |

# 研究開発計画

| 項目                           | 担当者                               | H27                            | H28                             | H29             |
|------------------------------|-----------------------------------|--------------------------------|---------------------------------|-----------------|
| サーバー管理・運用                    | <u>五斗</u> 、河野                     | 開発サーバー<br>立ち上げ<br>運用サーバー<br>導入 | 定期的なデータ移行（開発→運用サーバー、定期的なシステム増設） |                 |
| リポジトリ                        | <u>河野</u> 、<br>荒木、奥田              | レポジトリシステム（PXC）導入               | コアデータを用いたカスタマイズ                 | 大規模データを用いた検証    |
|                              |                                   | ストレージ導入                        | データ投稿システムの開発                    | 外部公開に向けた検討      |
| 標準化                          | <u>石濱</u> 、松本、<br>荒木、河野           | ワークフローの設計<br>（プラットフォーム別）       | プロセスシステム開発                      | プロセス自動化         |
|                              |                                   |                                |                                 | 統合化対応           |
| プロジェクト別DB<br>（Cube）構築        | <u>松本</u> 、石濱、<br>荒木、五斗           | Cubeの設計                        | グローバルプロテオミクスDB開発                | ターゲットプロテオミクス開発  |
|                              |                                   |                                | PTMプロテオミクスDB開発                  | 抗体・電気泳動DBの開発    |
| 集積キューブ型DB<br>（Globe）<br>構築   | <u>奥田</u> 、荒木、<br>石濱、松本           | RDFスキーマ、<br>オントロジー設<br>計       | Globeの設計                        | Globeの作製        |
|                              |                                   |                                |                                 | 統合化対応           |
| フォーカスドDB<br>（SliceMap）<br>構築 | <u>荒木</u> 、河野、<br>奥田、石濱、<br>松本、五斗 |                                | スライスシステムの開発                     | プリセットスライスマップの作製 |
|                              |                                   |                                | ビューワーの開発                        | カスタムマップ用ツール開発   |

# 昨年度の活動

---

- 全体ミーティング(計19回)
- インフォチームミーティング (随時)
- サイト相互訪問
- 拡大ミーティング (7月22日—熊本)
- JHUPO2015 – jPOSTシンポジウム (7月、熊本)
- HUPO2015 – PSI session (9月、バンクーバー)
- 統合の日
- ワークショップ (10月13日: JST本部)
- 分子生物学会ブース・JPrOSシンポジウム

# 昨年度の活動（続き）

---

- 国際学会および国内学会発表
  - 招待講演（国内 7件、国際 5件）
  - 口頭講演（国内 0件、国際 3件）
  - ポスター発表（国内 11件、国際 2件）



- 1 Shin Kawano, jPOST: Japan ProteOme STandard repository/database, HUPO-PSI meeting 2015, Seattle (USA), 13-15 April 2015
- 2 Shin Kawano, PRIDE to RDF project, HUPO-PSI meeting 2015, Seattle (USA), 13-15 April 2015
- 3 Yasushi Ishihama, Introduction of jPOST: Japan ProteOme STandard Repository/Database, HUPO-PSI-PXC session, HUPO 14th Annual World Congress, Vancouver (Canada), 27-30 September 2015

## 主催した活動（主催したワークショップ等）

| 年月日         | 名称                            | 場所         | 参加人数 | 概要                                  |
|-------------|-------------------------------|------------|------|-------------------------------------|
| 2015年7月24日  | JHUP02015シンポジウム               | くまもと森都心プラザ | 200人 | JHUP02015年大会内で開催した公開シンポジウム          |
| 2015年10月13日 | jPOSTワークショップ                  | JST東京本部別館  | 20人  | プロテオーム解析ソフトウェアのチュートリアルを中心としたワークショップ |
| 2015年10月29日 | 生命医薬情報学連合大会2015<br>年大会企画セッション | 京都大学       | 30人  | 質量分析インフォマティクスに関する企画セッションを<br>オーガナイズ |
| 2015年4月7日   | チーム内ミーティング（非公開）               | 京都大学       | 10人  | 研究進捗報告のためのミーティング                    |
| 2015年4月21日  | チーム内ミーティング（非公開）               | ウェブ会議      | 10人  | 研究進捗報告のためのミーティング                    |
| 2015年5月1日   | チーム内ミーティング（非公開）               | 東工大サテライト   | 5人   | 研究進捗報告のためのミーティング                    |
| 2015年5月8日   | チーム内ミーティング（非公開）               | ウェブ会議      | 11人  | 研究進捗報告のためのミーティング                    |
| 2015年5月11日  | チーム内ミーティング（非公開）               | ウェブ会議      | 5人   | 研究進捗報告のためのミーティング                    |
| 2015年5月29日  | チーム内ミーティング（非公開）               | 京都大学       | 9人   | 研究進捗報告のためのミーティング                    |
| 2015年6月18日  | チーム内ミーティング（非公開）               | ウェブ会議      | 11人  | 研究進捗報告のためのミーティング                    |
| 2015年7月14日  | チーム内ミーティング（非公開）               | ウェブ会議      | 5人   | 研究進捗報告のためのミーティング                    |
| 2015年7月22日  | チーム内ミーティング（非公開）               | 熊本大学       | 17人  | 研究進捗報告のためのミーティング                    |
| 2015年8月21日  | チーム内ミーティング（非公開）               | ウェブ会議      | 11人  | 研究進捗報告のためのミーティング                    |
| 2015年9月15日  | チーム内ミーティング（非公開）               | ウェブ会議      | 11人  | 研究進捗報告のためのミーティング                    |
| 2015年10月14日 | チーム内ミーティング（非公開）               | JST東京本部別館  | 9人   | 研究進捗報告のためのミーティング                    |
| 2015年11月10日 | チーム内ミーティング（非公開）               | ウェブ会議      | 11人  | 研究進捗報告のためのミーティング                    |
| 2015年12月1日  | チーム内ミーティング（非公開）               | 神戸国際会議場    | 10人  | 研究進捗報告のためのミーティング                    |
| 2015年12月10日 | チーム内ミーティング（非公開）               | DBCLS      | 5人   | 研究進捗報告のためのミーティング                    |
| 2016年1月15日  | チーム内ミーティング（非公開）               | ウェブ会議      | 11人  | 研究進捗報告のためのミーティング                    |
| 2016年2月5日   | チーム内ミーティング（非公開）               | 新潟大学       | 5人   | 研究進捗報告のためのミーティング                    |
| 2016年2月19日  | チーム内ミーティング（非公開）               | ウェブ会議      | 11人  | 研究進捗報告のためのミーティング                    |
| 2016年3月15日  | チーム内ミーティング（非公開）               | ウェブ会議      | 11人  | 研究進捗報告のためのミーティング                    |

# 昨年度の成果

---

- サーバーについては、開発サーバー（五斗グループ担当）および運用サーバー（河野グループ担当）を立ち上げ、定期的にデータを運用サーバーに移行することを開始した。  
<http://jpost.org/> を開設し、ポータルサイトを運用開始した。
- リポジトリについては、国際標準であるProteomeXchangeコンソーシアム加入にむけてシステム開発を行った。正式加入は4月に行われるHUPO-PSIミーティングにて可否投票で決定予定である。
- リポジトリから吸い上げた異なる様式のデータを再解析するためのキュレーションワークフローについては、複数のピークピッキング、複数のサーチエンジンおよび複数の偽陽性率に基づくクライテリア設定法を組み合わせ、用いるフローを確立した。
- プロジェクト別Cube開発においては、リポジトリへのデータ入力時に様々な異なる測定プラットフォームデータ（具体的には、グローバル定量プロテオミクス、翻訳後修飾定量プロテオミクスおよびターゲット定量プロテオミクス）に対応できるようにシステムの設計を行った。
- GlobeおよびSlice開発においては、リポジトリへのデータ入力時に必要な情報を吸い上げるシステム設計を行い、データ統合に向けた基本設計をおこなった。

以上より、1年目終了時の目安としていたリポジトリシステムの稼働およびキュレーションワークフロー決定を達成した。

# jPOST repository

## 2016年5月2日プレスリリース



配布先：京都大学記者クラブ、文部科学記者会、科学記者会、熊本県内報道機関、九州大学記者クラブ  
新潟県政記者クラブ  
解禁時間（テレビ、ラジオ、WEB）：平成28年5月2日（月）午前0時  
（新聞）：平成28年5月2日（月）付 朝刊

平成28年4月26日  
京都大学  
熊本大学  
九州大学  
新潟大学  
情報・システム研究機構  
科学技術振興機構

### プロテオーム統合データベース jPOST を開発 —アジア・オセアニア唯一の国際標準データリポジトリをスタート—

#### 概要

jPOST (Japan ProteOme Standard Repository/Database, <http://jpost.org/>)<sup>※1</sup>は、京都大学を中心としたオールジャパン体制で開発が進められているプロテオーム<sup>※2</sup>統合データベースです。国内外に散在している種々のプロテオームデータを標準化・統合化・一元管理したもので、多彩な生物種・翻訳後修飾<sup>※3</sup>・絶対発現量<sup>※4</sup>などの情報を含み、さまざまな解析が可能であるという特徴を有しています。

今回、jPOST データリポジトリ<sup>※5</sup>システムを新たに開発し、全世界に向けて公開しました。本システムは、アジア・オセアニア地域における初めての国際標準プロテオームデータリポジトリであり、2016年度国際ヒトプロテオーム機構・プロテオミクス標準化構想会議（HUPO-PSI, 2016年4月18-20日、ベルギー・ゲント市）において、国際標準のデータリポジトリシステムを提供する ProteomeXchange コンソーシアム<sup>※6</sup>への加盟が宣言されました。今後、アジアを中心に世界中のプロテオームデータを jPOST に収集することが可能となります。

### プロテオーム統合データベースjPOSTを開発—アジア・オセアニア唯一の国際標準データリポジトリをスタート—

2016年05月06日 |

石濱泰 薬学研究科教授、五斗進 化学研究所准教授、吉沢明康 化学研究所特定研究員らの協力を得て、生体中に存在する各種タンパク質のすべてを指すプロテオームのアジア・オセアニア地域唯一の国際標準データベースを開発・公開しました。

#### 研究者からのコメント



左から石濱教授、吉沢特定研究員

このデータベースは、タンパク質はもちろん、他の学系データベースとの統合も念頭に入れて設計したユーザーに入力してもらう属性情報を用いて、データベースや遺伝子発現データベースなど複数のデータベースと連携した検索が可能になる予定です。国内外の研究者と連携し、サイエンスの進歩に貢献する世界的なデータベースとなることを期待しております。



Japan Proteome Standard  
Repository/Database

### Recent posts

other

#### jPOST joined to ProteomeXchange

🕒 2016-07-6 👤 jpost

We are pleased to announce that the jPOST repository has joined to ProteomeXchange consortium on July 6, 2016.

other

#### Server maintenance

🕒 2016-05-27 👤 jpost

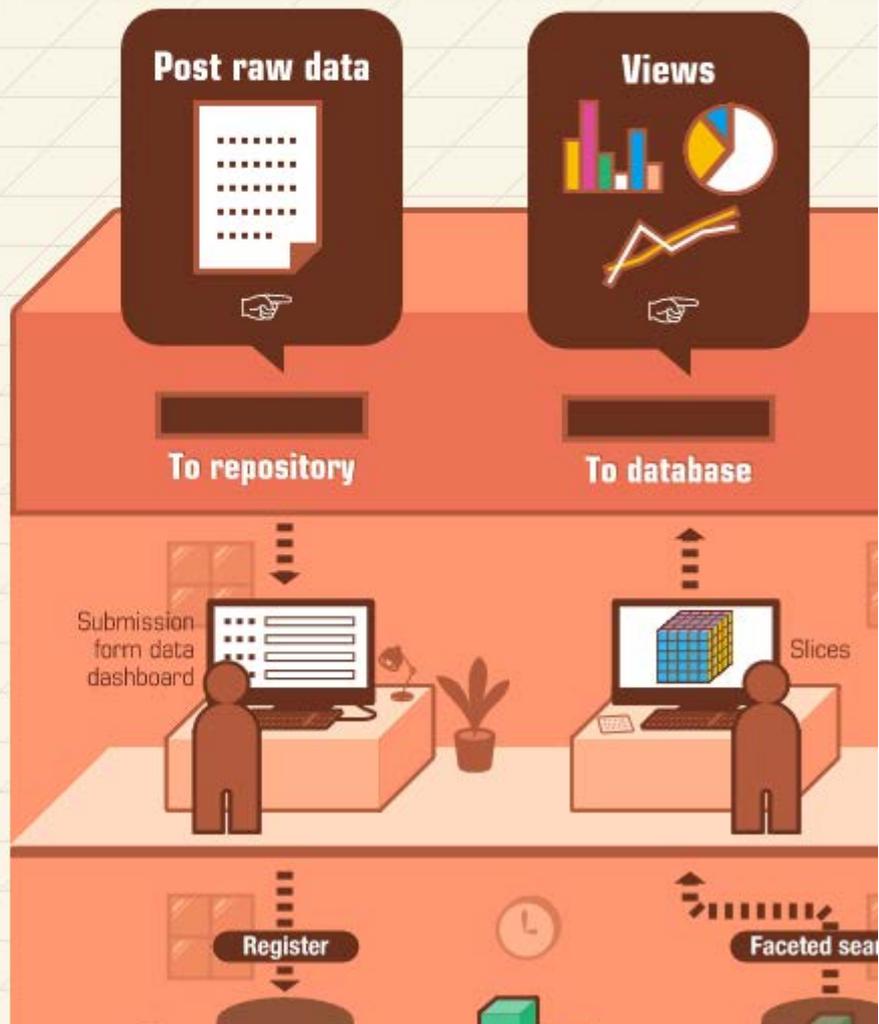
jPOST repository server will be temporarily unavailable. May 29, 9:30 - 19:00 (UTC+9)

other

#### Announcement of jPOST repository

🕒 2016-05-2 👤 jpost

We are pleased to announce that our jPOST





## About jPOST

The jPOST Repository is a new data repository of sharing MS raw/processed data. It consists of a newly-developed, high-speed file upload process, flexible file management system and easy-to-use interfaces. Users can release their "raw/processed" data via this site with a unique identifier number for the paper publication. Users also can suspend (or "embargo") their data until their paper is published. The file transfer from users' computer to the jPOST server is very fast (roughly ten times faster than usual file transfer) and uses only web browsers – it does not require installing any additional software.

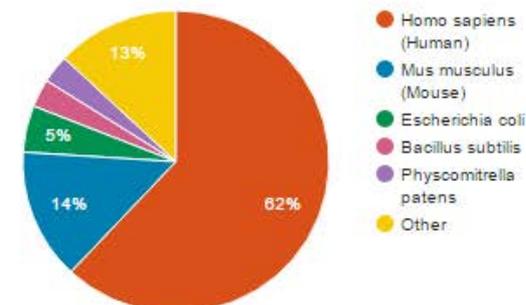
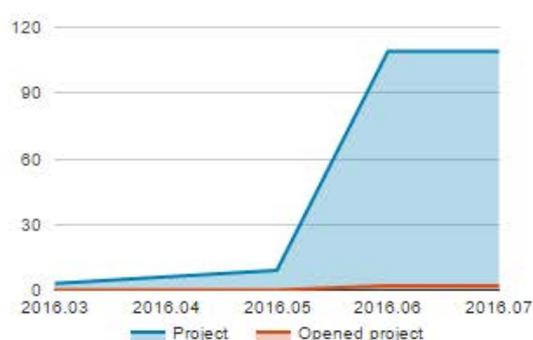


## Statistics

109 projects are registered. 2 are opened.

12607 files amount to 1.4 TB.

13 species.



# 商標登録 について

差出人

件名 Re: Pre-submission inquiries for the description of NEW databases

2016/07/01 22:4

宛先

Cc

Sorry, I just wanted to add that you better refer to your database as jPOSTdb. JPOST is a trademark of Jerusalem Post newspaper, and you wouldn't want to create any confusion.

MYG

On Fri, Jul 1, 2016 at 9:36 AM, Michael Galperin and Daniel Rigden <...> wrote:

Dear Dr. Okuda,

Thank you for your interest in the 2017 Nucleic Acids Research Database Issue. Your proposal sounds reasonable to me and I am glad to invite submission of your manuscript.

Please carefully read the latest instructions at the URL: [http://www.oxfordjournals.org/nar/for\\_authors/msprep\\_database.html](http://www.oxfordjournals.org/nar/for_authors/msprep_database.html) Then go ahead and submit your manuscript by the August 15th deadline.

With best wishes,  
Michael Galperin

On Fri, Jul 1, 2016 at 5:51 AM, Shujiro Okuda <...> wrote:

Dear Dr. Michael Galperin,

We would like to submit our paper about the jPOST repository. We are now promoting a proteomic database project named iPOST (Japan ProteOme Standard)

TMview

Home About News Contact Help FAQ Tutorials Enquiries

Log in |

Find term

jpost

Search

Clear

Advanced search

Filters

List of results

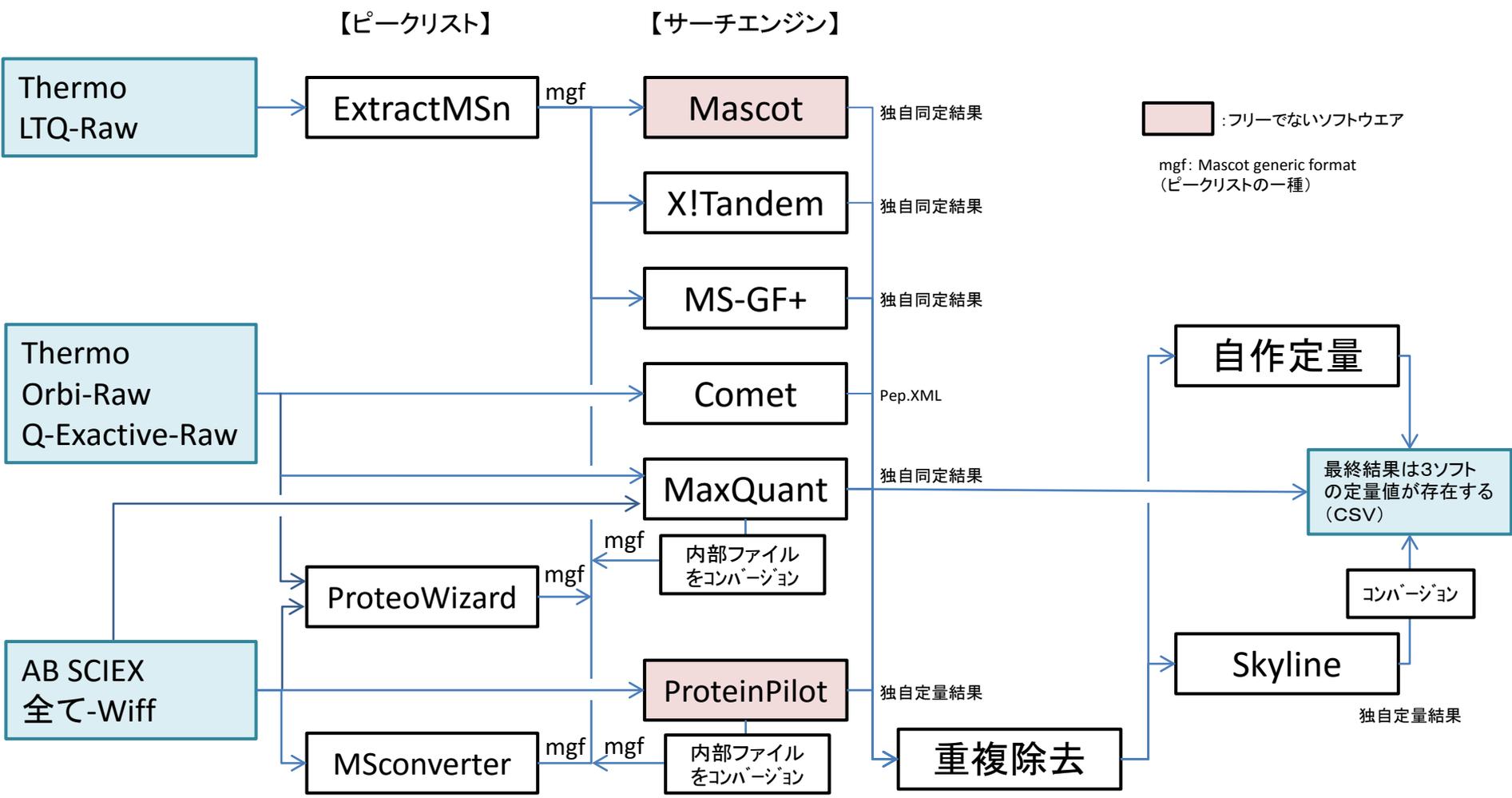
WIPO Global Brand DB

Page 1 of 1

10 20 30 50 75 100 | Displaying results

|                          | Graphic representation | Trade mark name | Trade mark of... | Designated te... | Application number / R... | Trade mark status | Nice class | Applicant name | Application date | Trade mark type | Registration date |
|--------------------------|------------------------|-----------------|------------------|------------------|---------------------------|-------------------|------------|----------------|------------------|-----------------|-------------------|
| <input type="checkbox"/> | -                      | JPOST           | JP               | JP               | 1987024965<br>0002221016  | Expired           |            | 株式会社日立製作所      | 11-03-1987       | Word            | 23-04-1990        |
| <input type="checkbox"/> | -                      | JPOST           | JP               | JP               | 1987024966<br>0002190575  | Expired           |            | 株式会社日立製作所      | 11-03-1987       | Word            | 28-11-1989        |
| <input type="checkbox"/> | -                      | J-POST          | JP               | JP               | 2001092539<br>0004614773  | Expired           | 39         | 日本郵政株式会社       | 16-10-2001       | Word            | 25-10-2002        |
| <input type="checkbox"/> | -                      | J-POST          | JP               | JP               | 2001092533<br>0004633907  | Expired           | 36,41      | 日本郵政株式会社       | 16-10-2001       | Word            | 10-01-2003        |

# 再解析フロー



# jPOST repository

奥田修二郎

新潟大学大学院医歯学総合研究科

2016年7月7日@京都大学薬学部本館

3階セミナー室

# Interface

## 主な特徴

オントロジーによる詳細メタ情報  
入力

メタ情報とファイルとの柔軟な対  
応付け

わかりやすいUI

データ公開までの期間の設定  
(embargo)

未公開状態でのReviewerのアク  
セス

超高速ファイルアップロード  
ウェブブラウザのみで完結

## 現状

109プロジェクト

12607ファイル

1400GB

13 species

The screenshot displays the jPOST Repository interface. At the top, there is a navigation bar with the jPOST logo, 'REPOSITORY', and buttons for 'Repository', 'Submit', 'Help', 'Sign in', and 'Sign up'. Below the navigation bar, the 'About jPOST' section provides a description of the repository's purpose and features. To the right of the text is an illustration of a globe with various data storage and management icons. The 'Statistics' section includes a line graph showing the number of projects registered and opened from March 2016 to July 2016, and a pie chart showing the distribution of projects across different species. The 'Data list' section features a search bar and a table of project entries.

**About jPOST**

The jPOST Repository is a new data repository of sharing MS raw/processed data. It consists of a newly-developed, high-speed file upload process, flexible file management system and easy-to-use interfaces. Users can release their "raw/processed" data via this site with a unique identifier number for the paper publication. Users also can suspend (or "embargo") their data until their paper is published. The file transfer from users' computer to the jPOST server is very fast (roughly ten times faster than usual file transfer) and uses only web browsers – it does not require installing any additional software.

**Statistics**

109 projects are registered. 2 are opened.

12607 files amount to 1.4 TB.

13 species.

**Data list**

| jPOST ID    | Project title   | Description  | Type     | Publication | Principal investigator   | Announcement date | Created    | Last modified |
|-------------|---|--|----------|-------------|--|-------------------|------------|---------------|
| JPOST000052 | Human iPS cell proteome   | Rapid and deep profiling of human induced pluripot ... | Complete | 23210603    | Yasushi Ishihama<br>Kyoto university                             | 2016-06-20        | 2016-06-20 | 2016-06-20    |
| JPOST000051 | Comparative Transcriptomics and Proteomics on Pi Depleted Arabidopsis Roots | Plant adaptation to limited phosphate (Pi) availab ... | Complete | 27121119    | Wolfgang Hoehenwarter<br>Leibniz Institute of Plant Biochemistry | 2016-06-20        | 2016-06-20 | 2016-06-20    |



ID      Name  
 Password      Affiliation  
                     ORCID

## Project

**Project**

Rev.1

Title, description, keywords,  
 PubMed, Principal investigator, etc...

FILE

**Project**

Rev.1

FILE

**Project**

Rev.0

FILE

## Preset

**Sample**

Species, Tissue, Cell type, Disease  
**Sample**  
 Species, Tissue, Cell type, Disease

**Fractionation**

Subcellular, Protein, Peptide  
**Fractionation**  
 Subcellular, Protein, Peptide

**Enzyme/Mod**

Enzyme, Modification, Taxonomy  
**Enzyme/Mod.**  
 Enzyme, Modification, Taxonomy

**MS mode**

Instrument, Purpose, Platform, etc...  
**MS mode**  
 Instrument, Purpose, Platform, etc...

# Preset

## New Sample profile

Title\*

Species\*

(select)

Tissue

(select)

Cell type

(select)

Disease

(select)

Note

**Title \***

MOCK-FLAG-IP

**Species \***

Arabidopsis thaliana (Mouse-ear cress)



Looking for more items

**Tissue**

(select)



Looking for more items

Search from EBI Ontology Lookup Service



or Enter free word

Enter



Close

(select)



Looking for more items

1  
TAR

4-low res.

ative quantification

ictra count  
duct ion (reporter quantification)

# Project

## New Project profile

Title \*

PC12 NF1, TCTP FLAG-IP

Keywords \*

rat itraq, PPI

Description \*

Must be more than 50 characters

Announcement \*

Now  Unfixed

Now: project will be open right after submission and confirmation finished.  
Unfixed: temporary set 1 year after and you can change the date later.

Principal investigator \*

Affiliation \*

Note

# Project

  Apply profile to files

  Apply profile  Add profile  Remove selected  Reset

| Select                              | File Name                     | Sample      | Fractionation     | Enzyme/Mod. | MS mode                   | Location |
|-------------------------------------|-------------------------------|-------------|-------------------|-------------|---------------------------|----------|
| <input checked="" type="checkbox"/> | 131229tk-RPmonolith-4h-3.txt  | S0000000002 | F0000000002 [111] | E0000000002 | M0000000002 [113: label1] | Local    |
| <input checked="" type="checkbox"/> | 131229tk-RPmonolith-4h-2.txt  | S0000000002 | F0000000002 [111] | E0000000002 | M0000000002 [113: label1] | Local    |
| <input checked="" type="checkbox"/> | 131229tk-RPmonolith-4h-1.txt  | S0000000002 | F0000000002 [111] | E0000000002 | M0000000002 [113: label1] | Local    |
| <input type="checkbox"/>            | 131229tk-RP-monolith-8h-5.txt | S0000000002 | F0000000002 [111] | E0000000002 | M0000000002 [113: label1] | Local    |
| <input type="checkbox"/>            | 131229tk-RP-monolith-8h-4.txt | S0000000002 | F0000000002 [111] | E0000000002 | M0000000002 [113: label1] | Local    |
| <input type="checkbox"/>            | 131229tk-RP-monolith-8h-3.txt | S0000000002 | F0000000002 [111] | E0000000002 | M0000000002 [113: label1] | Local    |
| <input type="checkbox"/>            | 131229tk-RP-monolith-8h-2.txt | S0000000002 | F0000000002 [111] | E0000000002 | M0000000002 [113: label1] | Local    |

# File upload

Uploading files... 1/16 finished. Estimated time remaining: 3 min.

Pause all Resume all Cancel all Minimize this window

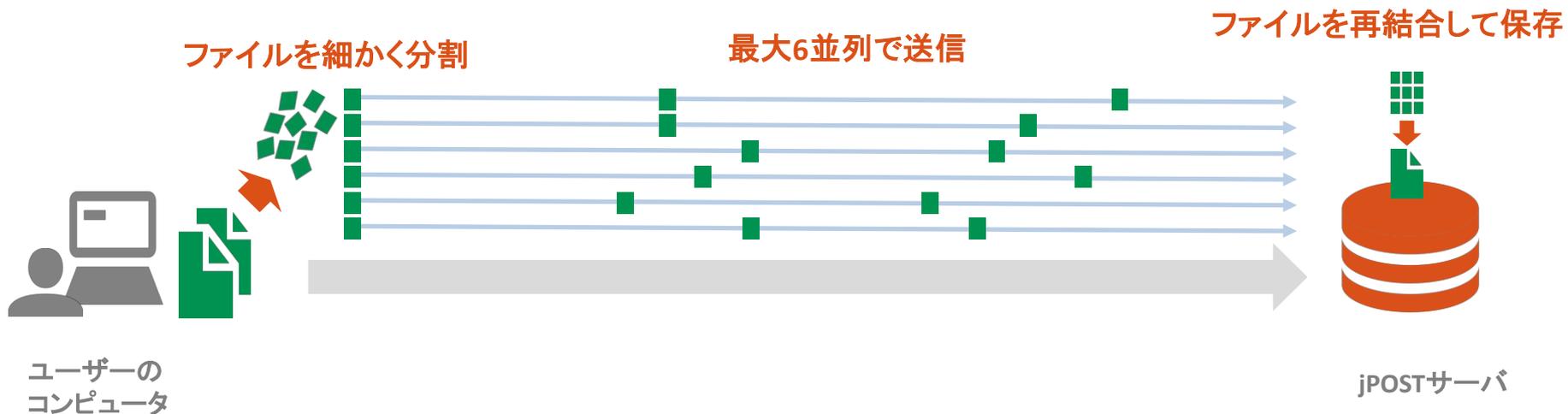
|                                      |     |  |
|--------------------------------------|-----|--|
| 131228tk-HILIC-monolith-4h-1ug-2.txt | 44% |  |
| 131228tk-HILIC-monolith-4h-1ug-3.txt | 0%  |  |
| 131228tk-HILIC-monolith-8h-1ug-1.txt | 0%  |  |
| 131228tk-HILIC-monolith-8h-1ug-2.txt | 0%  |  |
| 131228tk-HILIC-monolith-8h-1ug-3.txt | 0%  |  |
| 131228tk-HILIC-monolith-8h-1ug-4.txt | 0%  |  |
| 131228tk-HILIC-monolith-8h-1ug-5.txt | 0%  |  |
| 131229tk-RP-monolith-8h-1.txt        | 0%  |  |
| 131229tk-RP-monolith-8h-2.txt        | 0%  |  |
| 131229tk-RP-monolith-8h-3.txt        | 0%  |  |
| 131229tk-RP-monolith-8h-4.txt        | 0%  |  |
| 131229tk-RP-monolith-8h-5.txt        | 0%  |  |
| 131229tk-RPmonolith-4h-1.txt         | 0%  |  |

# ファイル転送のしくみ

## ■ 通常のファイルアップロード



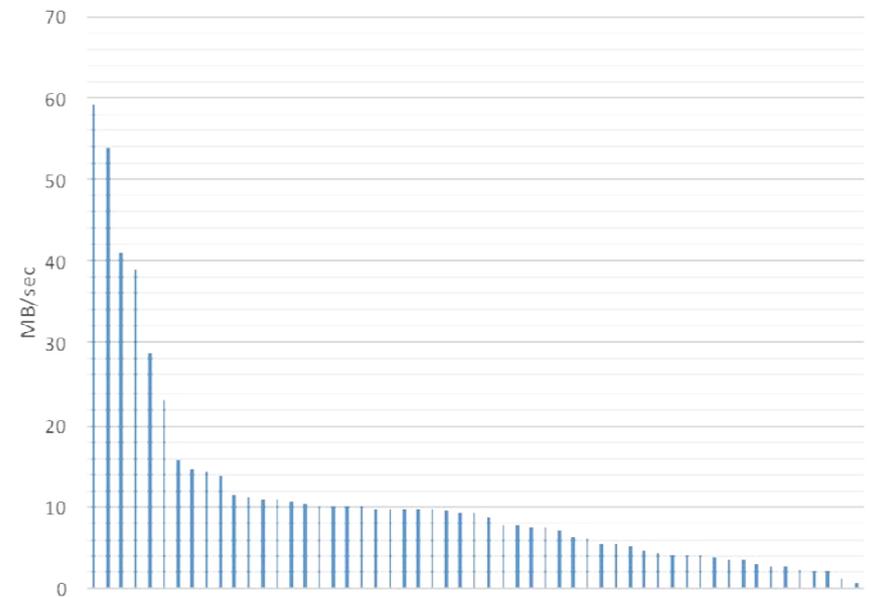
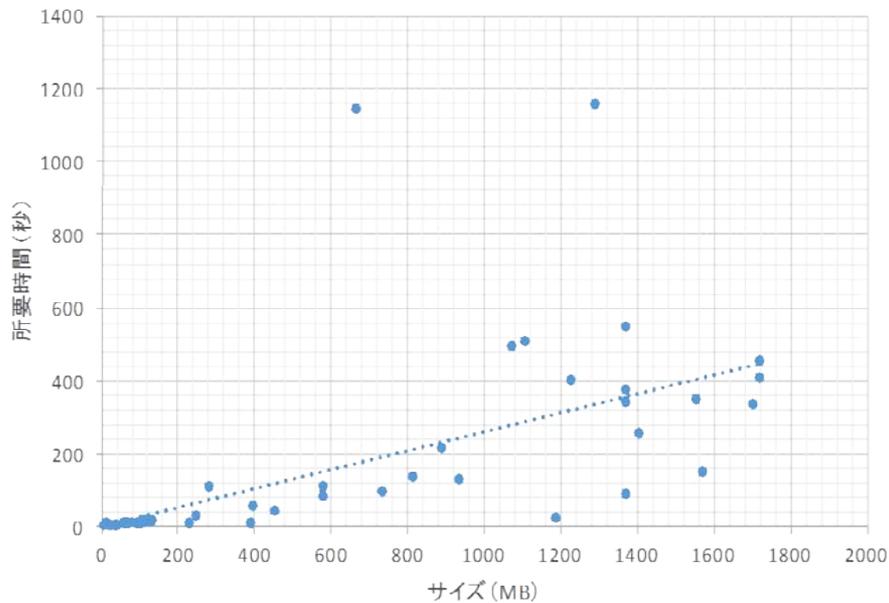
## ■ jPOSTのファイルアップロード



# ファイル転送速度

平均速度: **11.11MB/sec**    最高: 59.35MB/sec、最低: 0.58MB/sec

平均サイズ: **2.94GB/件**    最大: 126.1GB



# セキュリティ

---

## ① HTTPSを採用

すべての通信が暗号化されるため、通信途中で盗聴、改ざんを防ぐことができる。  
また、SSL証明書によりWebページの発行元を保証することができるので、なりすましの防止になる。

## ② CSP(Content Security Policy)を設定

CSPはインターネット上の各種技術の標準化を推進する標準化団体「W3C」によって整備されている仕様で、Webページのセキュリティレベルを設定することができる。

jPOST Repositoryでは、

- HTML内部でのJavaScriptの実行を禁止
  - 外部サーバからの各種ファイル(画像ファイル、JavaScriptファイル等)の読み込みを禁止
  - jPOST repository内の各ページが、他のWebサイトからiframeで呼び出されることを禁止
- 等の設定を行っている。

これにより、XSS(クロスサイトスクリプティング)、クリックジャッキング等の被害事例の多い攻撃手法を防ぐことができる。

## ③ CookieをSSL Onlyに設定

これにより、常に暗号化されたCookieが発行されるので、なりすましログインを防止できる。

# セキュリティ

---

## ④ ブラウザ(JavaScript)でのCookieの操作を禁止

ログイン中のユーザーが、他のユーザーになりすましてデータを取得するのを防ぐことができる。

## ⑤ ユーザーエージェントが一致しなければ再ログイン時にパスワードを要求

ログイン中のユーザーとしてリクエストがあった場合でも、接続元やブラウザが変わった可能性があれば、パスワードの入力を求める。

これにより、Cookieの盗聴や成りすましによるログインを防止する。

## ⑥ ウイルスチェック

ファイルのアップロードが完了するとすぐウイルスチェックを実施する。

ウイルス定義ファイルも自動的に更新されており、常に最新のウイルスチェックを実施できる。

いずれも、ユーザー情報(Cookie含む)の盗聴、なりすましや、Webページの改ざんを防止する目的の設定で、

複数のセキュリティ対策を組み合わせることで、よりセキュアなWebサービスになっている。

# Partial and complete submission

---

## STEP.0

### Complete submissionとしてsubmitできる条件：

#### 【Partial submissionと共通のチェック項目】

- Projectの必須項目 (Title, Keywords, Description(>50chars) Principal investigator, Affiliation) がすべて入力されている。
- Rawファイル、Resultファイルがそれぞれ1ファイル以上ある
- Species、Modification(FixedまたはVariable)、Instrumentがそれぞれ1種類以上関連付けられている。

※ここまでの条件をクリアしないと、「Submit project」ボタンを押すことができない

#### 【ファイルをグルーピングする際のチェック項目】

- 1つのグループには、Rawファイル、Resultファイル、Peakファイルがそれぞれ1ファイル以上ある



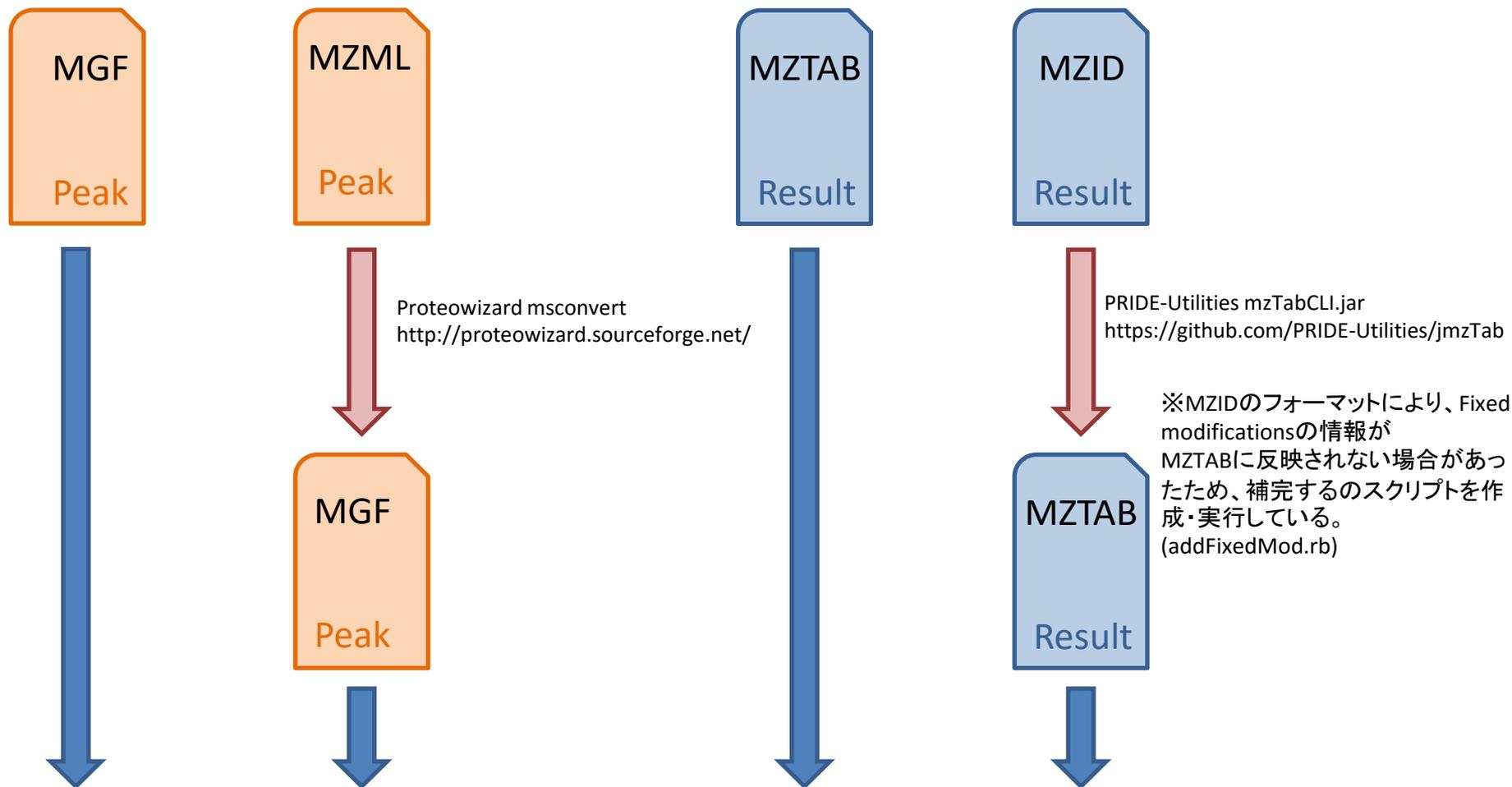
ユーザーが「Complete」を選択して「Submit project」ボタンを押す  
Projectのステータスが「validating」になり、一時的に編集不可になる

# STEP.1

## ファイルの変換

処理を共通化するために、ファイルフォーマットをMGFとMZTABに統一する。変換ツールは既製のものを利用。

※ファイルフォーマットは、ユーザーが入力した情報を信用するため、実際のファイルと食い違っていた場合は、変換できずにエラーとなり処理を中止する(validation failed)。



## STEP.2

### バリデーション

ユーザーが入力したファイルのグルーピングが正しいかどうかを1グループずつチェックする。



MZTABから、resultをランダムに200取り出し、  
対応するPeakが存在するかどうかをチェックする。  
対応するPeakが95%以上存在すれば「valid」とみなす。

※1グループに含まれるPeakファイル、Resultファイルの数は、1対1でも多対多でもOK。  
ただし、総当たりでチェックするため、少ない数ごとにまとまっていたほうが処理は速く済む。  
また、1つのResultファイルが複数のPeakファイル由来であるケースには対応できない。

※1つでもinvalidなグループがあれば、その時点で処理を終了する。

## STEP.3

### Peak view用データの作成

Protein/Peptideテーブル(ex. <https://repository.jpostdb.org/result/JPST000052.0>)に表示するデータをSQLiteファイルに格納する。(MZTABファイル1つにつき、1つのSQLiteファイルが生成される。)

## STEP.4

### Validation結果の反映

Validationにパスした場合：

- ユーザーにメールで通知
- Projectのステータスは「Submitted」に。(編集不可で、ReviseおよびPubMed、Announcement dateの入力ができる状態)
- Projectの公開設定が「Now」の場合には、即時Announceされる。

Validationに失敗した場合：

- ユーザーにメールで通知
- Adminにもメールで通知
- Projectのステータスは「Editing」に戻る。

# Complete submission

JPST000052.0 ▶ F024605.mzid ▾

Filter table:  Sequence  Protein  Variable modifications

keyword

**Filter** Reset

| Peak view | PSM ID | Sequence    | Protein | Charge | Spec e-value | Raw | Variable |
|-----------|--------|-------------|---------|--------|--------------|-----|----------|
|           | 1      | ETVAGALR    |         |        |              |     |          |
|           | 2      | VLDSRLGK    |         |        |              |     |          |
|           | 3      | VHELNEEIGK  |         |        |              |     |          |
|           | 3      | VHELNEEIGK  |         |        |              |     |          |
|           | 3      | VHELNEEIGK  |         |        |              |     |          |
|           | 3      | VHELNEEIGK  |         |        |              |     |          |
|           | 4      | LSVMESSK    |         |        |              |     |          |
|           | 5      | QAEILQESR   |         |        |              |     |          |
|           | 5      | QAEILQESR   |         |        |              |     |          |
|           | 6      | VEAEPAEVTER |         |        |              |     |          |
|           | 7      | VTWIQLK     |         |        |              |     |          |
|           | 8      | RQRLEEDK    |         |        |              |     |          |

**ETVAGALR, MH+ 816.4574, m/z 408.7323**

File: 120118ry\_20187-32\_1\_1.wiff, 120118ry\_20187-32\_1\_1.wiff.scan, DeltaMass: -0.000616000000363652, Scan: 1, Exp. m/z: 408.73202, Charge: 2

Click and drag in the plot to zoom X:  Y:  Zoom Out Print  Enable tooltip  Plot mass error

| b+       | # | Seq # | y+              |
|----------|---|-------|-----------------|
| 130.0499 | 1 | E 8   |                 |
| 231.0975 | 2 | T 7   | 687.4148        |
| 330.1660 | 3 | V 6   | 586.3671        |
| 401.2031 | 4 | A 5   | <b>487.2987</b> |
| 458.2745 | 5 | G 4   | 416.2616        |
| 529.2617 | 6 | A 3   | <b>359.2401</b> |
| 642.3457 | 7 | L 2   | 288.2030        |
|          | 8 | R 1   | 175.1190        |

[Click] to move table  
Static Modifications:  
C: 57.021464

# まとめ

---

- jPOST repository (<https://repository.jpostdb.org>)  
を公開した
- Partial submission, Complete submissionともに  
PXCに準拠
- PXC加盟が承認された
- 登録プロジェクト数は順調に増加中
- 日本プロテオーム学会2016年大会にて、参加者  
のデータのリポジトリへの登録サポート
- NAR DB Issueへの論文投稿



# 再解析プロトコルの開発

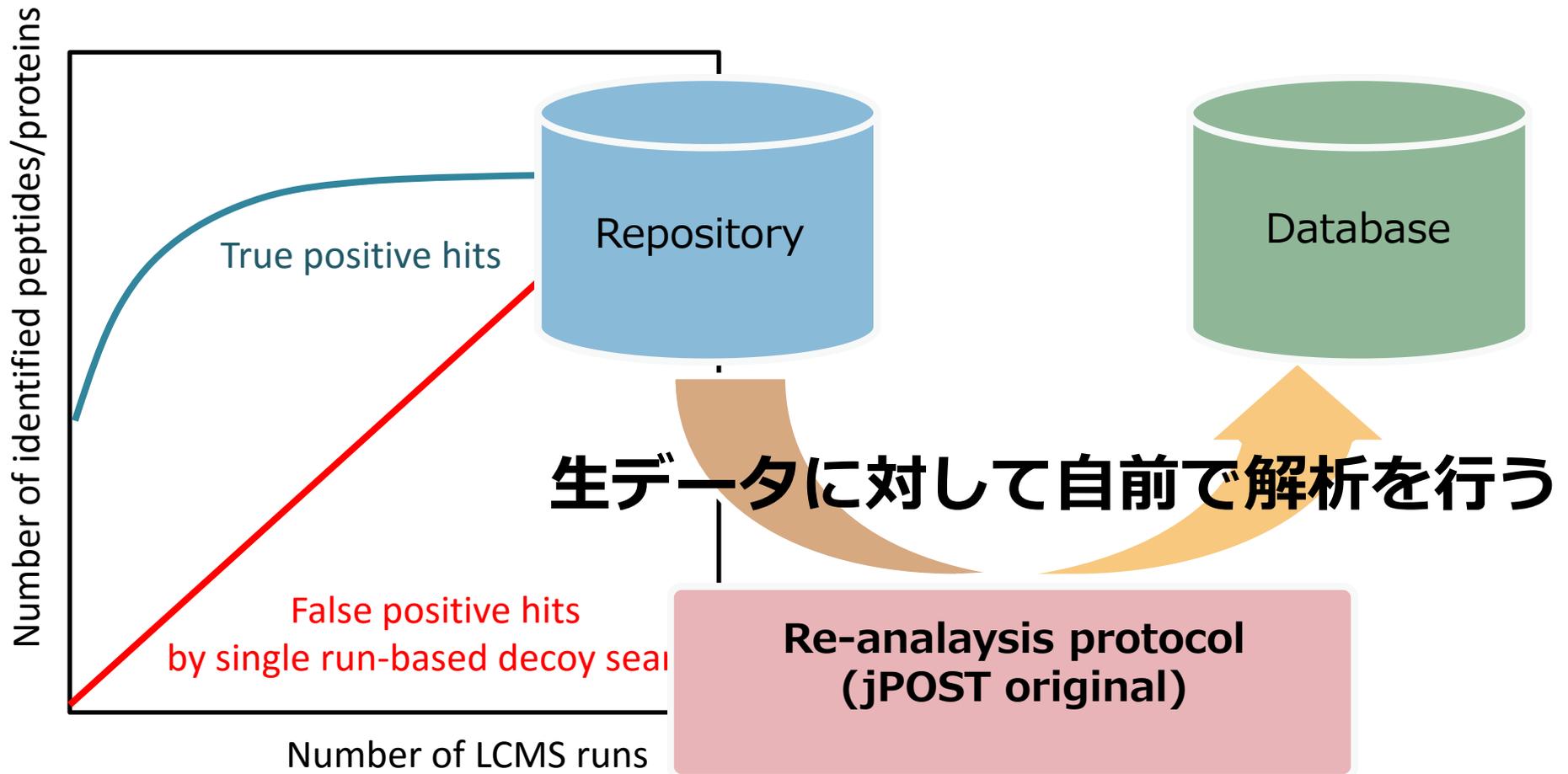
July 7, 2016

jPOST project

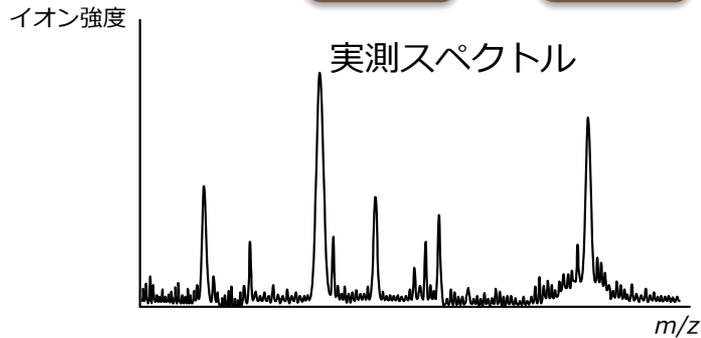
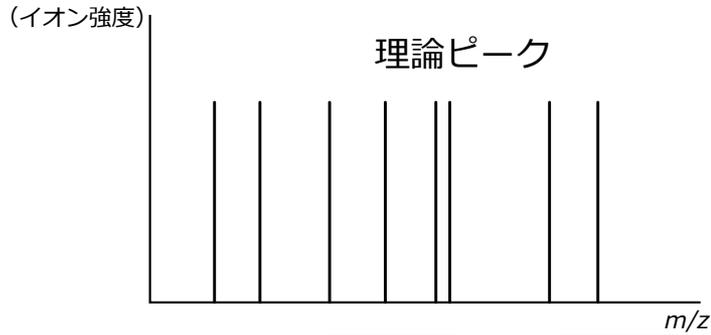
○吉沢明康, 田畑剛, 守屋勇樹, 河野信, 奥田修二郎, 渡辺由,  
山本格, 松本雅記, 高見知世, 小林大樹, 荒木令江, 杉山直幸,  
五斗進, 石濱泰



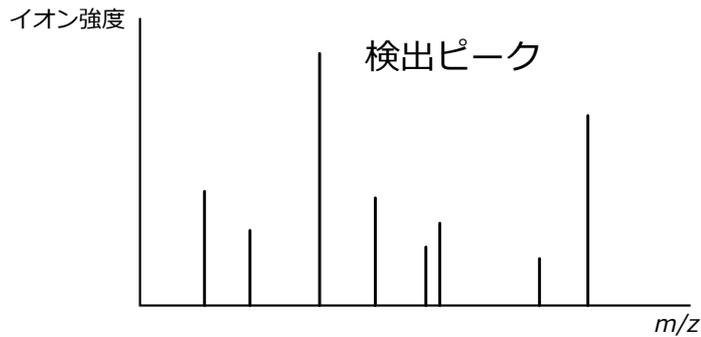
# 再解析プロトコル



# ピーク検出



↓ ピーク検出



# 配列同定の手法



- 大別して3種類

- スペクトル・ライブラリ検索

- *de novo* sequencing

- データベース検索

# 配列同定の手法

- スペクトル・ライブラリ検索
  - スペクトル（ピークリスト）のライブラリと、サンプルのスペクトル（ピークリスト）を比較
    - 他の分野（分析化学、メタボロミクスなど）での同定法と同じ
  - プロテオミクスではあまり用いられない
- *de novo* sequencing
- データベース検索

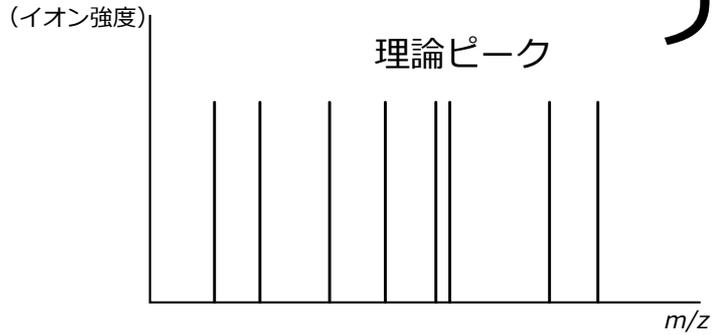
# 配列同定の手法

- スペクトル・ライブラリ検索
- *de novo sequencing*
  - 塩基配列決定のSanger法に類似
    - 長さ1個違いの配列断片を（仮想的に）ラダー状に並べることで、アミノ酸を1個ずつ確定
  - 必要なサンプル量が多いので、あまり用いられない
- データベース検索

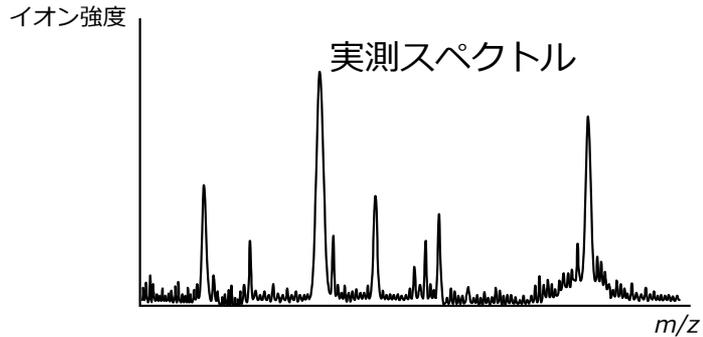
# 配列同定の手法

- スペクトル・ライブラリ検索
- *de novo* sequencing
- データベース検索
  - ゲノム情報を利用
    - 既知の配列（データベース）の中から、矛盾のないものを探す
    - 十分な量のイオンがない（1個違いの長さのイオンを網羅できない）場合の不利を補う
  - プロテオミクスでは主流

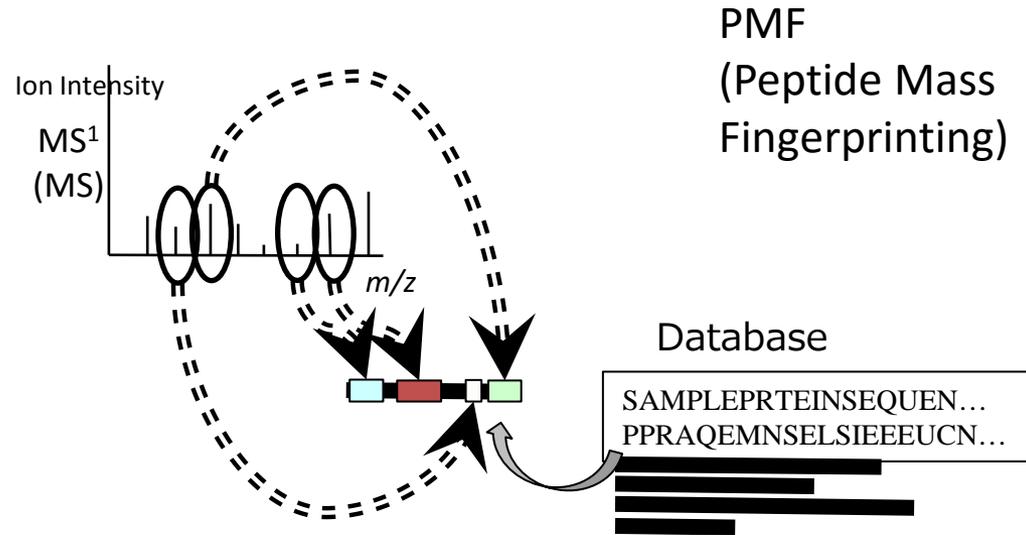
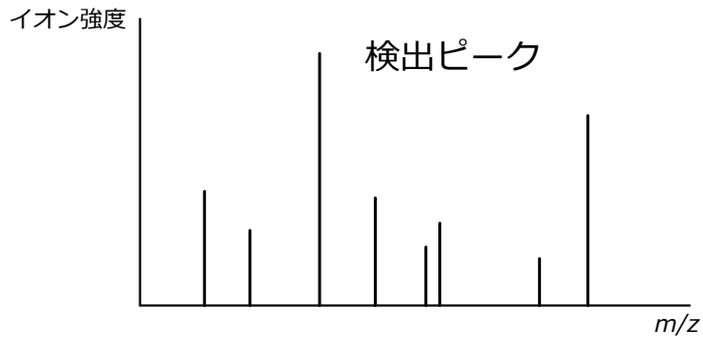
# データベース検索の原理



↓ イオン化効率,  
機器特性, ノイズ

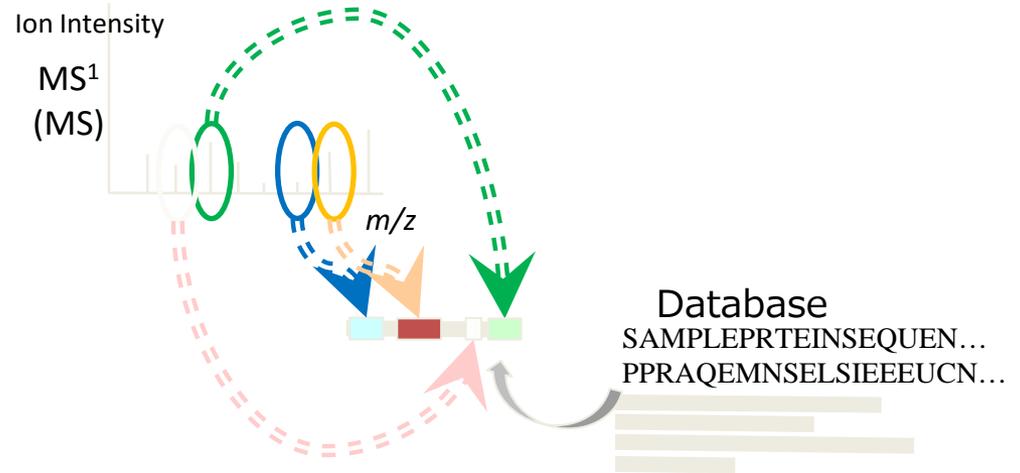


↓ ピーク検出

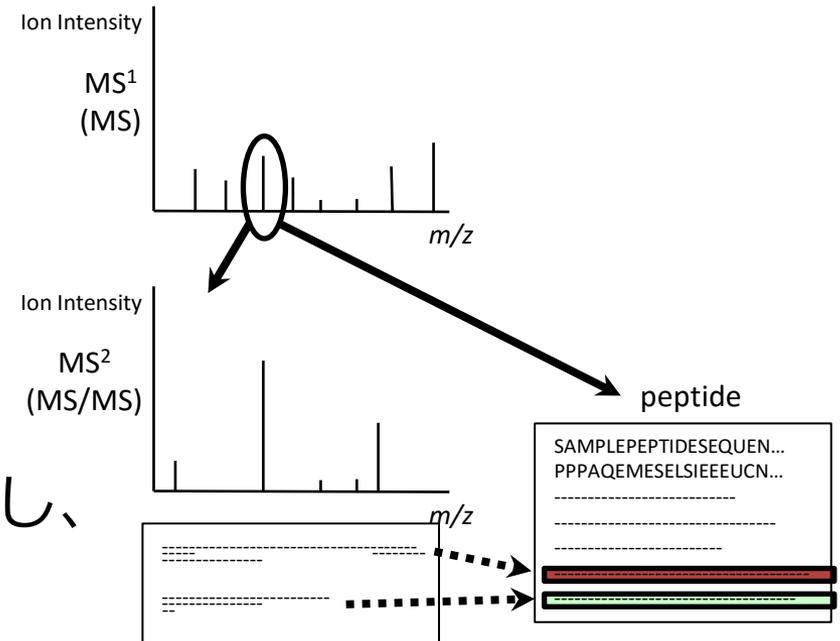


# データベース検索の原理

PMF  
(Peptide Mass  
Fingerprinting)

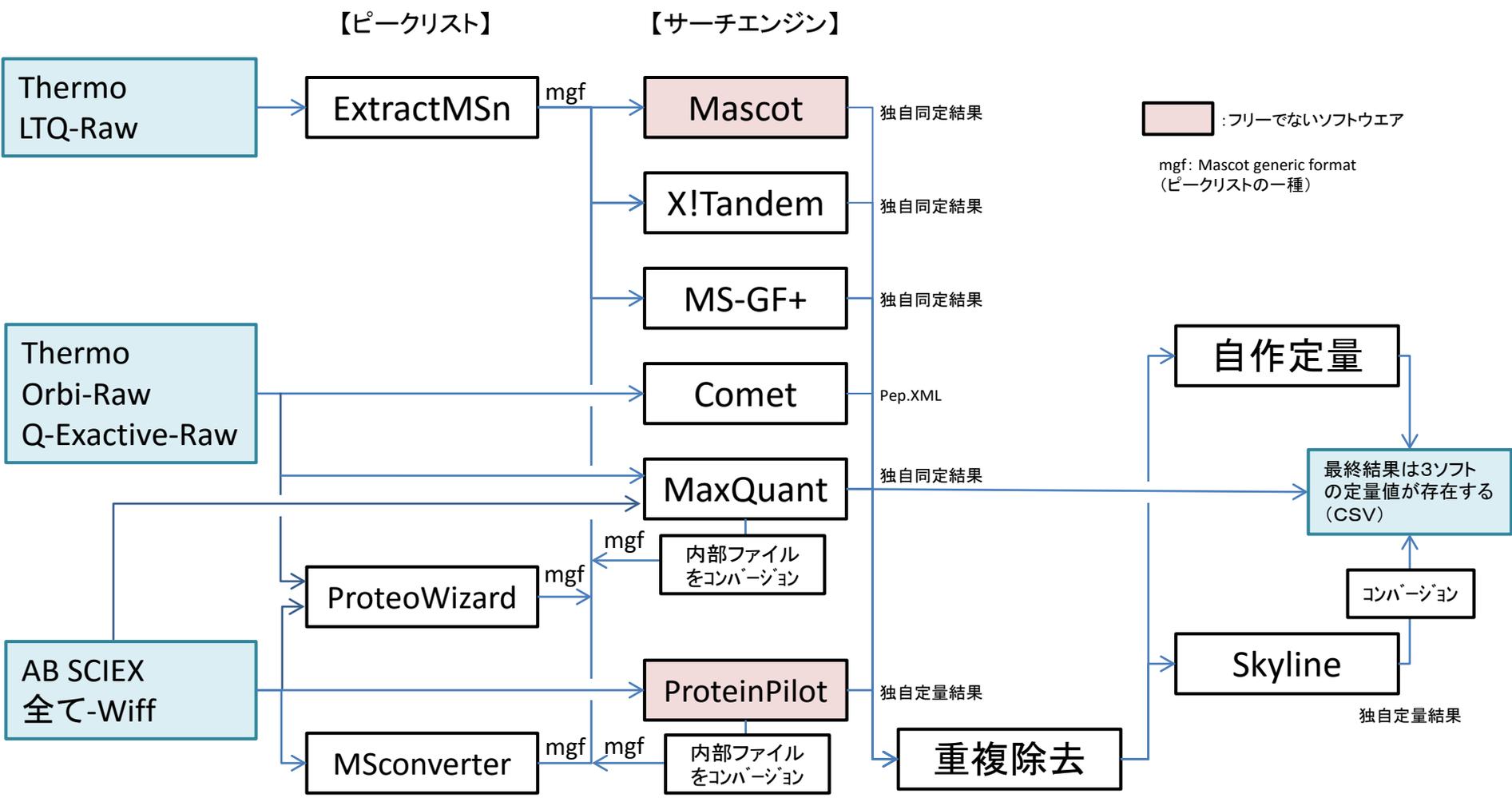


PFF  
(Peptide Fragmentation  
Fingerprinting)



- ペプチドを（1カ所だけ）切断し、断片の同定結果も用いる

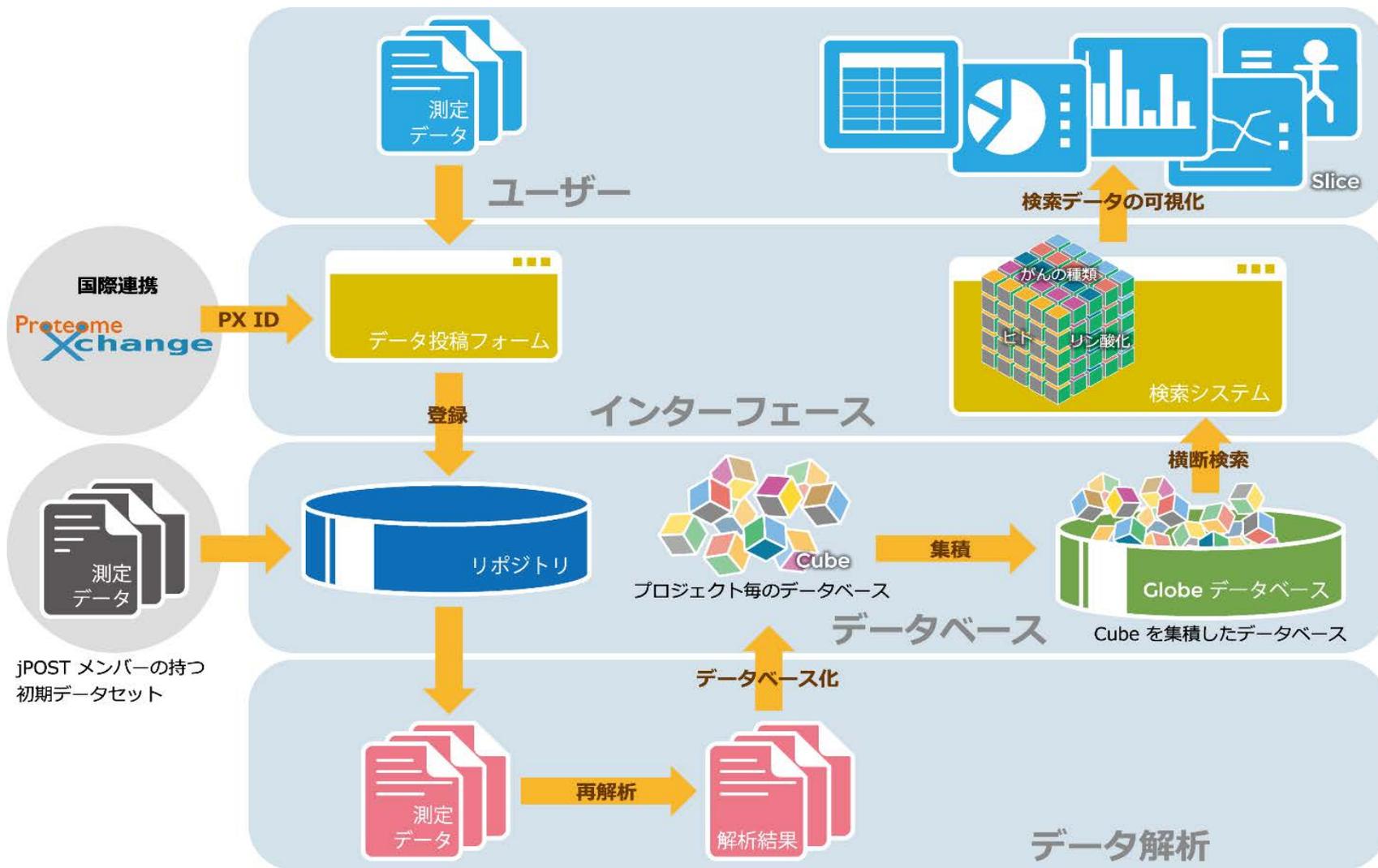
# 再解析フロー



# 再解析プロトコルの開発方針

- (データベース検索で) 探知可能な配列数を増加させる
  - (結果数の増加、取りこぼし (第2種エラー) の減少)
- 偽陽性の結果(false positive assignment)を減少させる
  - (結果品質の向上、誤った同定 (第1種エラー) の減少)

# 今後の計画



# 今年度の計画

---

- 1 サーバーについては、前年度から引き続き、開発サーバーは五斗グループが担当し、運用サーバーはNBCLS河野グループが管理する。ディスクを増設し、定期的にデータを運用サーバーに移行する。
- 2 リポジトリについては、4月に国際標準であるProteomeXchangeコンソーシアムに正式加入する。コアデータを格納してシステムをカスタマイズする。
- 3 前年度に設定したリポジトリデータ再解析ワークフローを実装する。
- 4 プロジェクト別Cube開発においては、前年度に行った設計に基づき、様々な異なる測定プラットフォームデータ（具体的には、グローバル定量プロテオミクス、翻訳後修飾定量プロテオミクスおよびターゲット定量プロテオミクス）を実装する。
- 5 GlobeおよびSlice開発においては、設計したRDFスキーマ、オントロジーに基づき設計を行い、またビューワー開発をおこなう。

なお、2年目終了時には、jPOST全体としてコアデータを格納したプロトタイプで稼働することを達成度の目安とする。

# 今後の計画

---

- 7/27 jPOST workshop (北里大)
- 7/28-29 JPrOS大会(北里大：口頭・ポスター・ブース)
- 9/18-21 HUPO (Taipei: 口頭・ポスター)
- 9/22-23 AOHUPO (Sun Moon Lake:口頭・ポスター)
- 10/5 統合の日 (東京: 口頭・ポスター)
- 11/30 分子生物学会 (横浜、DBシンポジウム、ブース)

# 将来展望

---

- 本プロジェクトでは、UniProtのような寄せ集めタイプのプロテオーム知識ベースではなく、実験データをただやみくもに集めたProteomicsDBの失敗の教訓を活かし、国際的にもユニークな日本発の高質・高機能かつ多視点の統合プロテオームデータベースjPOSTの構築を目指す。
- 将来的には、本DBとDBCLSとの連繋を強化して、連邦型統合ライフサイエンスDBに必須のプロテオーム情報の供給源として貢献する。
- プロテオミクスを専門とする情報科学者を本プロジェクトを介して育成し、今後永続的なプロテオームDB維持のみならず、これらの情報を介して、生命科学、医薬分野における基礎および応用学の進歩に大きく貢献できる多くの人材の輩出に貢献する。