

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)

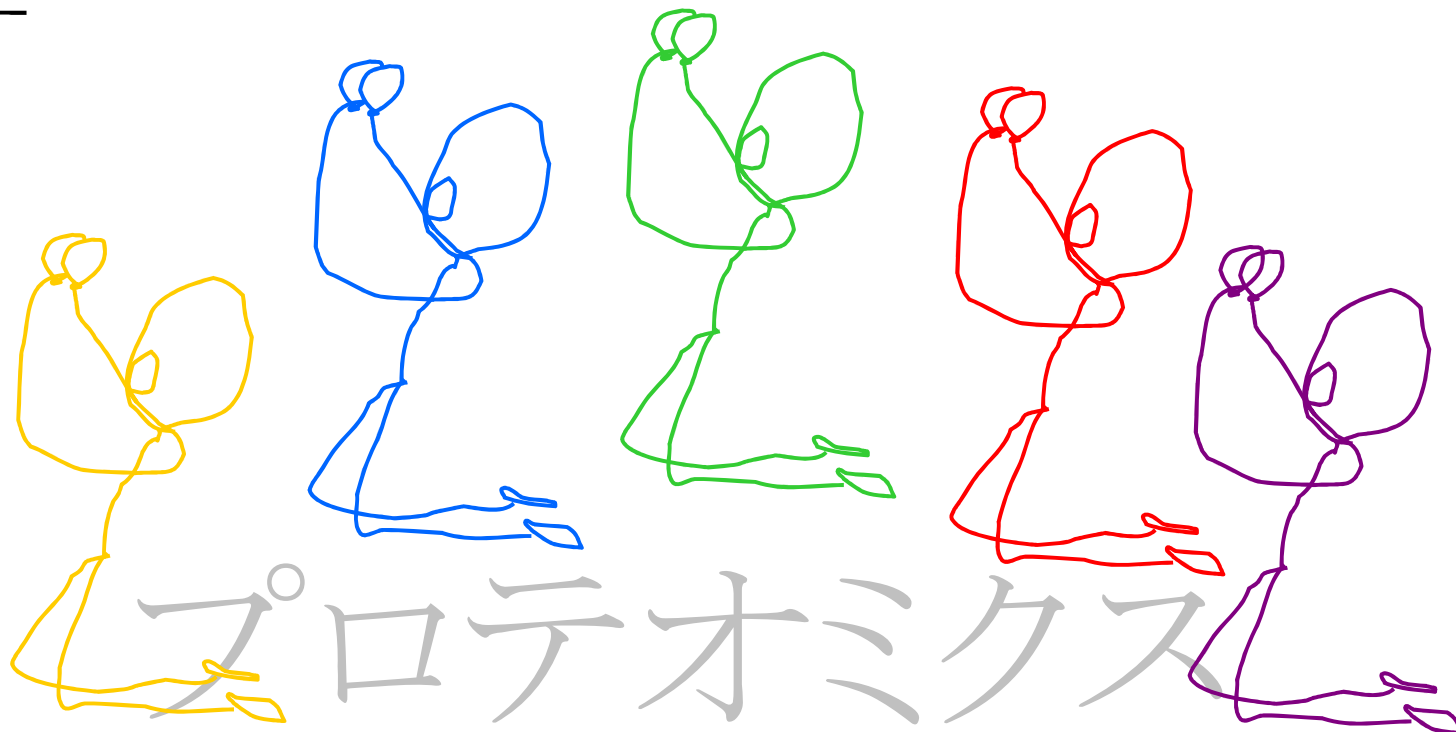
荒木 令江 (熊本大学)

松本 雅記 (九州大学)

奥田 修二郎 (新潟大学)

河野 信 (DBCLS)

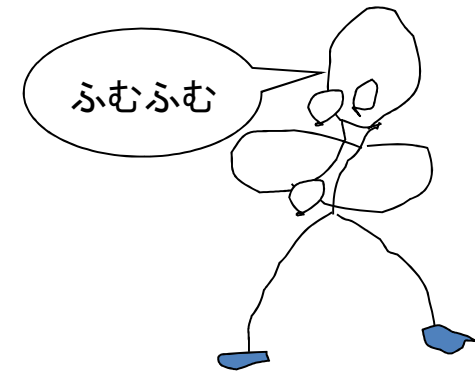
プロテ
オーム
の神



ああ、タンパク質、最高！

Protein, the best!!

なんで最高か？



- だって細胞を直接動かしているのは**タンパク**だし
- だから病気になるのは**タンパク**がおかしくなった時だし
- だから、売ってる薬のほとんどのものの標的は**タンパク**だし。

植物だっておなじだよん



プロテオーム解析からのみ得られる情報

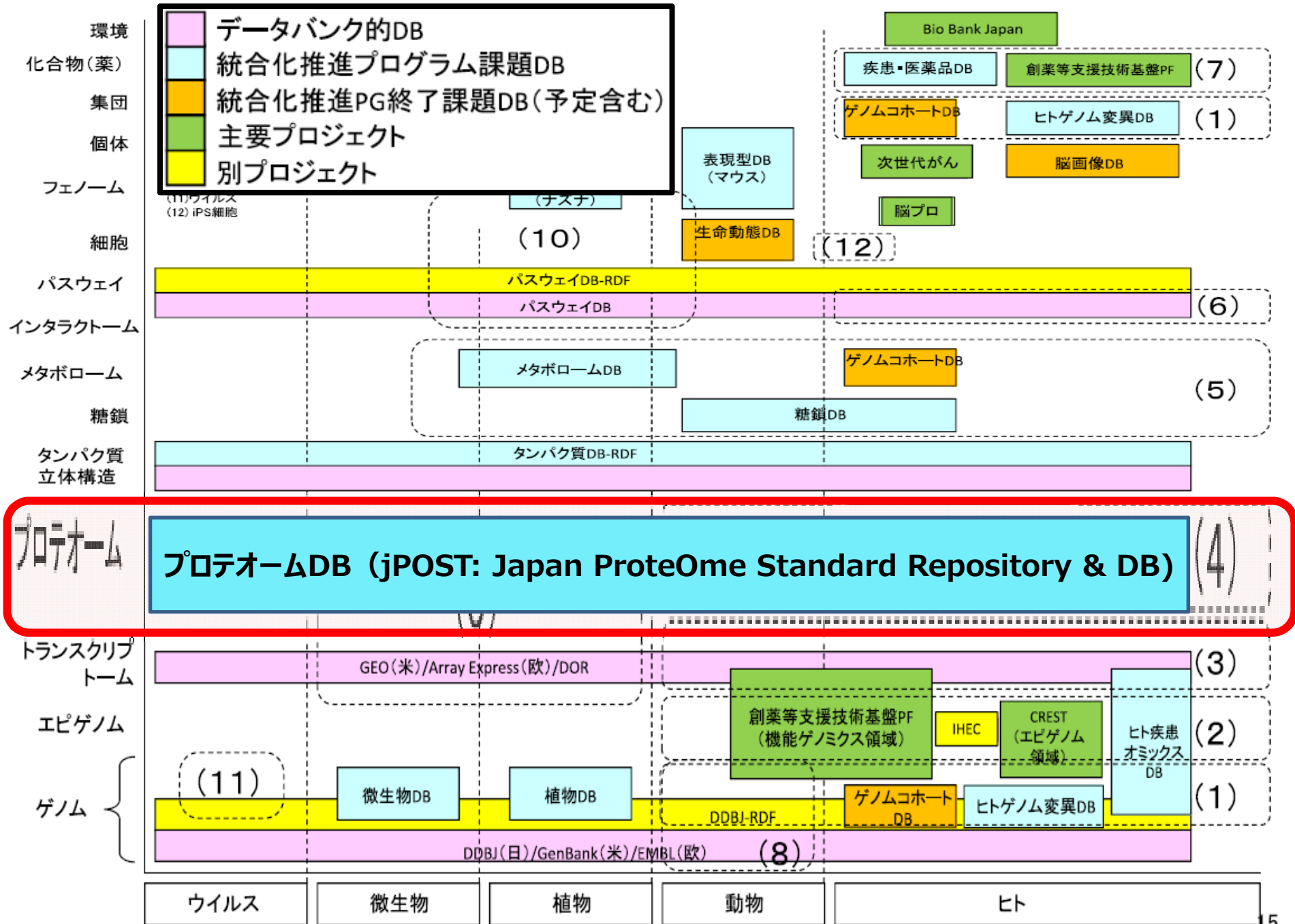
- 翻訳後修飾
- 発現プロファイル（臓器、組織、細胞内）
- 絶対発現量
- タンパク-タンパク相互作用

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

Analyzing the First Drafts of the Human Proteome

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

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, PMID: 24870542

A draft map of the human proteome

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

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

17,294 gene products

Nature. 2014 , DOI: 10.1038/nature13319

Mass-spectrometry-based draft of the human proteome

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

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

18,097 gene products

Target-decoy search for all merged data



Juergen Cox (Max Planck)

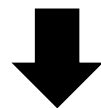
Nature. 2014 , DOI: 10.1038/nature13302, PMID: 24870542

A draft map of the human proteome

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

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

17,294 genes (84%)

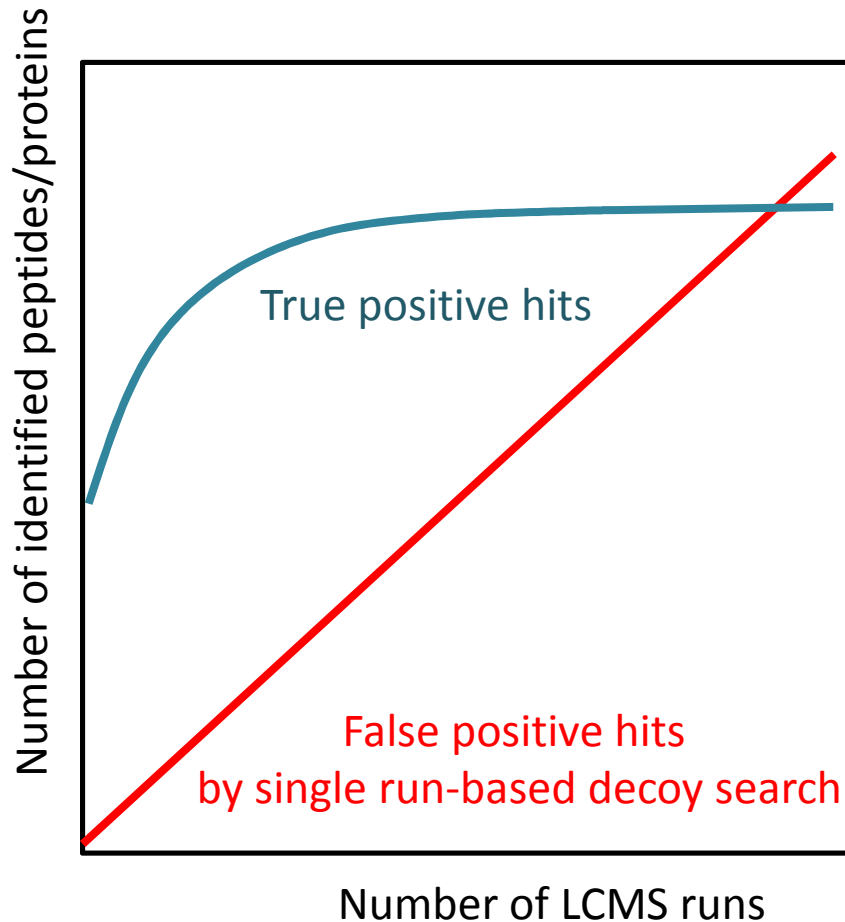


Target-decoy search for all merged data
1% FDR (protein level)

11,206 genes (57%)

1/3は偽ヒットだった。。。。

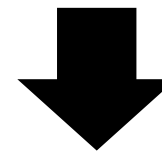
ProteomicsDBの問題点とjPOSTの戦略



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

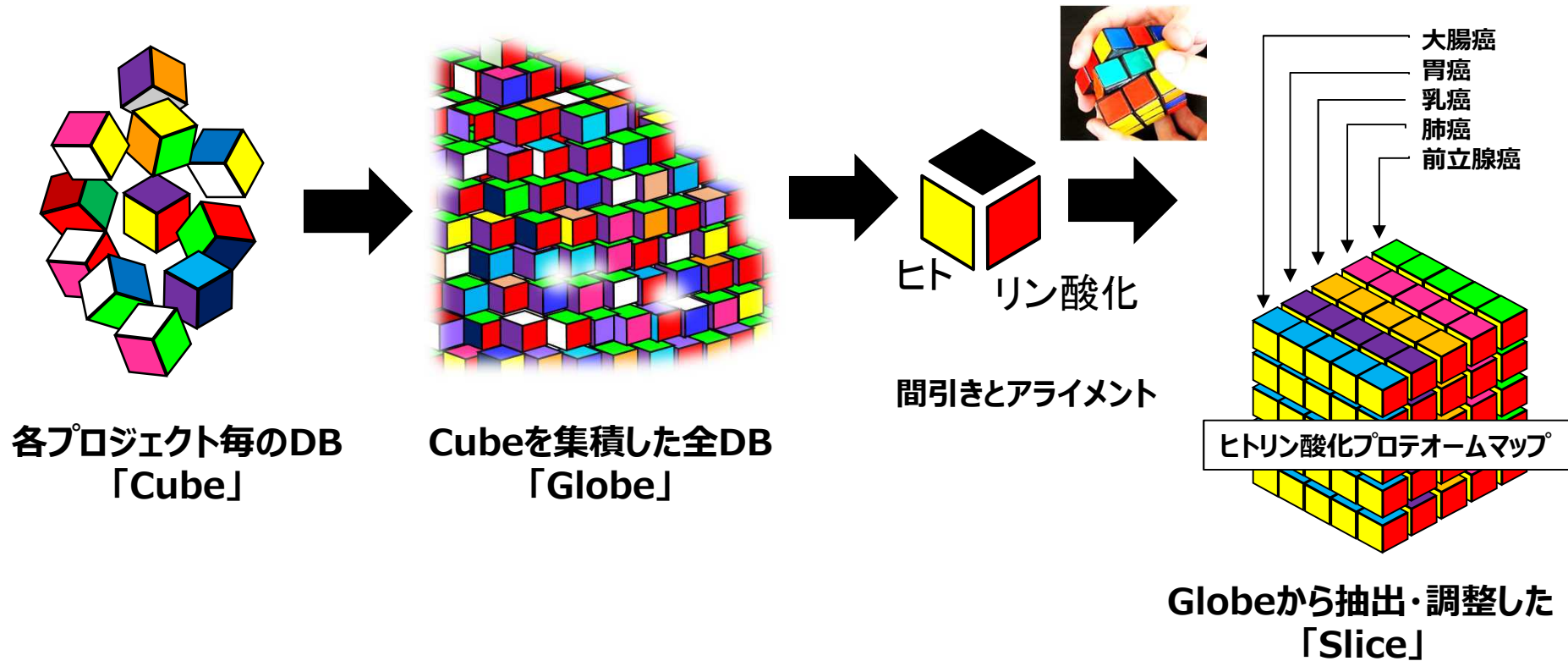
- 世界標準リポジトリ (PX)の採用
- 研究機関・プロジェクト毎に異なるフォーマットや解析法、信頼度評価法やアノテーションの標準化。
- 全データに対する統一したフィルタリング



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 **ST**andardization Repository & Database)

jPOSTの構成（ルービックキューブ型DB）



フォーカスドDB
(Slice)

集積DB
(Globe)

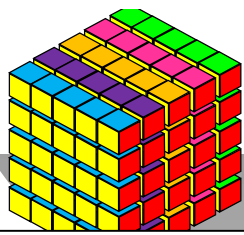
プロジェクトDB
(Cube)

標準化

リポジトリ

測定生データ

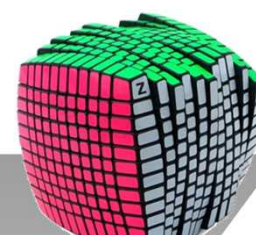
試料



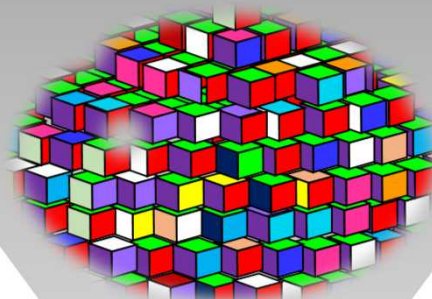
疾患別ヒトリン酸化
プロテオームマップ



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



抗体

質量分析

電気泳動

臨床検体

植物

モデル生物

培養細胞

細菌

希少生物

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

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)



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、 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, 統合化プログラム糖鎖 DB (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 | | |

4月以降の活動

- 全体ミーティング(計9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)
- jPOSTセッション in JHUPO2015 (7月24日)
- HUPO2015 (9月末バンクーバー)
- 統合の日
- ワークショップ (10月13日@JST本部)
- EBI・ISB訪問 (PXC, PRIDE, PASSEL)

jPOST: Next steps



- ✓ Fix the re-processing workflow
 - Peak-picking/DB search/Post-Dbsearch (FDR)
- ✓ Set the main servers
- ✓ Use the PSI terminology for data registration

jPOST Repository Ready (early 2016)

Vote for **PXC partnership** (at PSI meeting, April 2016)



jPOST Repository Start

jPOST Database Start (2017)

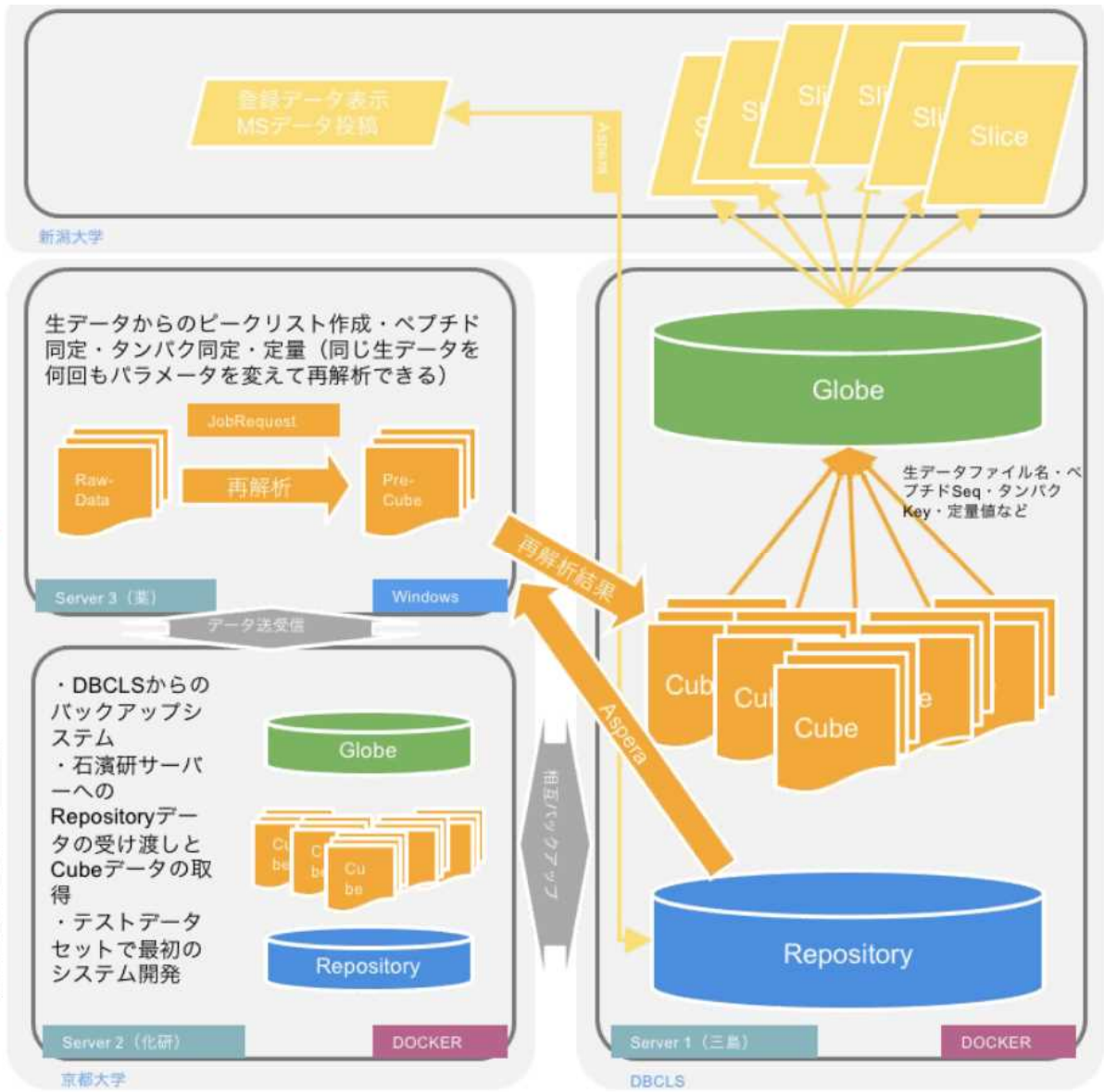


jPOST DBオーバービュー改訂版



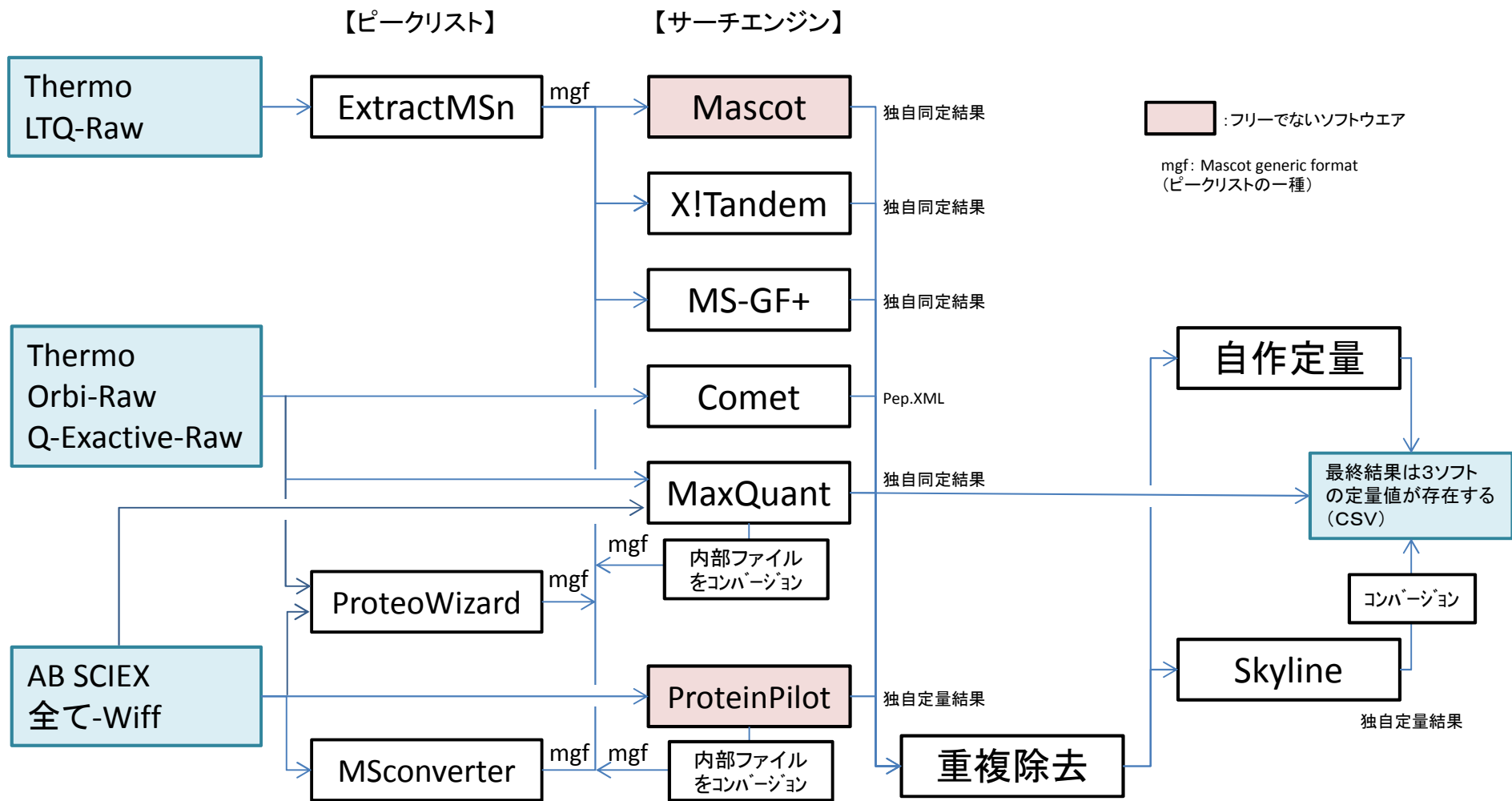
| 項目 | 担当 | R17 | R18 | R19 |
|---------------------|-------------|--|---------------------------------|------|
| サーバ管理・運用 | 高木、河野 | 開発サーバの立ち上げ 運用サーバの導入 | 定期的なデータ移行（開発→運用サーバへ、定期的なシステム更新） | |
| ソフトウ | 河野、高木、河野 | ソフトウェアのインストール（PXID導入）、ソフトウェアのインストールと更新 | 大規模データ移行の検証 | |
| セキュリティ | 高木、河野、高木、河野 | 脆弱性診断、ファイアウォール設定、データバックアップシステムの構築 | 外部公開 | |
| スケーラビリティ | 高木、河野、高木、河野 | クラウド移行の検討（AWS導入） | クラウド移行の実施 | 外部公開 |
| スケーラビリティ (Data) 構築 | 高木、河野、高木、河野 | Cubeの設計 | データ移行の実施 | 外部公開 |
| 監視システム (Data) 構築 | 高木、河野、高木、河野 | 監視システムの構築 | 監視システムの運用 | 外部公開 |
| パフォーマンス向上 (Data) 構築 | 高木、河野、高木、河野 | パフォーマンス向上の検討 | パフォーマンス向上の実施 | 外部公開 |
| パフォーマンス向上 (Data) 構築 | 高木、河野、高木、河野 | パフォーマンス向上の検討 | パフォーマンス向上の実施 | 外部公開 |

(参考) 申請書上の役割分担とスケジュール



- ユーザー**
ウェブインターフェイス
データ可視化
Mass++拡張
- データ統合**
SQLデータ管理
RDF化・SPARQL
オントロジー対応
- キュレーション**
ピークピッキング
ペプチド同定
タンパク同定・定量
- データ登録**
MS登録データベース/
ツール
PX IDの取得

解析フロー(例)



jPOST current status (www.jpostdb.org)

The screenshot shows the jPOST website interface. At the top right, there is a navigation menu with links: ABOUT, REPOSITORY, DATABASE, WORKFLOW, and CONTACT. Below the menu is a large diagram illustrating the data workflow. The diagram shows 'Raw data' being stored into a 'Repository' (represented by a blue cylinder). From the 'Repository', data is processed into a 'Cube' (represented by a cluster of colorful cubes), which is noted as 'is DB for each project'. The 'Cube' is then stored into a 'Database' (represented by a green cylinder). The 'Database' is used for 'cross search' to generate 'Slices' (represented by a 3D grid of cubes). The 'Slices' are further processed into 'Ice' (represented by a stack of yellow rectangles). The 'Database' is also connected to a 'Globe' (represented by a green cylinder with a globe icon). The diagram is set against a background of a 'User Interface'.

RECENT POSTS

- jPOST Workshop: Computational Proteomics Togo Symposium 2015
- HUPO 2015
- JHUPO 2015
- Site Visit

RECENT COMMENTS

ARCHIVES

- September 2015
- July 2015
- May 2015

CATEGORIES

EVENT

JPOST WORKSHOP: COMPUTATIONAL PROTEOMICS

© 2015-09-07


[English : [日本語](#)]

Date: October 13, 2015 9:30 – 17:00

Venue:

JST Tokyo Headquarters K's Gobancho 2F meeting room A1,2
7-2-1, Gopancho, Chiyoda-ku, Tokyo 100-0071, Japan




REPOSITORY

[Database](#)
[Submit](#)
[Help](#)
[Sign in](#)

| 29990519755f26f7a511ba | |
|---------------------------------|--|
| Project title | Sample Project |
| Keywords | keyword |
| Project description | description |
| Parent Project | Human Proteome Project |
| PubMed ID(s) | 000000 |
| Links to other 'Omics' datasets | Another DB |
| Workflow | Protein complex → DDA-high res. → Relative quantification → Precursor ion stable isotope labeling |
| When to publicize | 2015-4-1 |
| Contact | New name Somewhere else. |
| XML file | 29990519755f26f7a511ba.xml |
| File location | ftp://... |

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。