

2014. 7. 7

【日本プロテオーム学会通信】は、日本プロテオーム学会会員の皆様に配信しています。【アドレス変更/配信中止】【ご質問・お問合せ】【掲載希望】は、日本プロテオーム学会事務局 (info@jhupo.org) をお願いいたします。

会員の皆様

昨年に引き続き、ライフサイエンス統合データベースセンターの河野先生より HUPO-PSI/ProteomeXchange meeting 2014 のレポートをいただきました。本学会でもデータベースセンターを設立し、今月 19 日にはワークショップも開催予定です。ご興味のある方は是非ご一読ください。

=====

HUPO-PSI/ProteomeXchange meeting 2014 レポート
情報・システム研究機構 ライフサイエンス統合データベースセンター ([DBCLS](#))
河野 信

2014 年 4 月 13 日から 16 日にかけて [HUPO-PSI](#) (Proteome Standards Initiative) のミーティングが、16 日 17 日と [ProteomeXchange](#) (PX) ミーティングがドイツフランクフルトの郊外エルバツハで開催された。昨年に引き続き筆者は両ミーティングに参加したので、その様子について報告する。

前半の PSI ミーティングには約 50 名が参加して行われた。まずは基調講演が行われ、EMBL の Martin Beck から質量分析計を利用した新技術として、リンカー化合物で化学的に相互作用しているタンパク質を架橋し、タンパク質間相互作用を調べる方法が紹介された。昨年に引き続き基調講演を行った EMBL-EBI の Reza Salek からは、メタボロミクス分野での標準化の取り組み ([COSMOS](#)=Coordination Of Standards in MetabOlomicS) について、HUPO-PSI での標準化の取り組みをほぼ踏襲する形で標準化を進め、質量分析のデータについて HUPO-PSI で開発したマススペクトルの生データを保存するオープンフォーマットである mzML をそのまま活用すること、NMR のデータについては独自でファイル標準や統制語彙などを開発していく方針が示され、引き続いて発表した Leibniz Institute of Plant Biochemistry の Daniel Schober から NMR 標準 (nmrML) についての説明がなされた。また、データ共有の仕組みとして ProteomeXchange を模した MetabolomeXchange を構築する構想が発表された。ミシガン大学の Gilbert Omenn からは現在

HUPO で進めている Human Proteome Project ([HPP](#)) について、その中で HUPO-PSI や PX が果たした役割の紹介があり、今後の課題にとしていて missing protein をいかに見つけていくかについて議論がなされた。

引き続き、Molecular Interaction (MI)、Mass Spectrometry (MS)、Proteome Informatics (PI) の各グループから昨年度の進捗状況と本ミーティングで話し合われる内容についての紹介があった。ここで示された各課題について 2 日目以降に詳細な議論が行われた。

筆者の参加した MS/PI のグループでは、メタボローム分野で mzML が正式採用になったことに伴う拡張、各質量分析計メーカーが提供している生データよりもファイルサイズが大きくなってしまふ mzML について効率的な圧縮方法の検討、SWATH-MS などの新規測定技術への対応、ペプチド同定データを記述するための mzIdentML での翻訳後修飾部位が確定できない場合の記述方法、定量データを記述するための mzQuantML の絶対定量法への対応などについて活発な議論が交わされた。特に mzML のファイルサイズの問題に対しては、スペクトルデータを記述する部分 (m/z 値や intensity) を効率的に圧縮するアルゴリズム NumPep ([Teleman, J. et al., Mol. Cell. Proteomics 2014, 13, 1537-1542](#)) と、ファイルをいくつかのブロックに分割してそれぞれのブロックごとに圧縮する blocked gzip を組み合わせることで、効率的な圧縮とランダムなデータアクセスを実現することに合意した(これまでではファイル全体をまとめて圧縮していたため、使用したいデータが一部分でも一度すべてのファイルを解凍展開してからファイルにアクセスする必要があり非効率的だった)。また、複数のタンパク質-タンパク質相互作用データベースを統一的なインターフェイスで取り扱うことのできる [PSICQUIC](#) に倣って、複数のペプチドマスレポジトリ ([PRIDE](#)、[GMPDB](#)、[PeptideAtlas](#)、[nextProt](#) など) を統一的なインターフェイスからアクセス可能にする PROXI (PROteomeXchange common query Interface) の開発が提案があり、これが了承されて開発が開始されることとなった。

PSI ミーティングに引き続いて開催された PX ミーティングでは、これまでの進捗状況として、ほぼデータ登録の仕組みが出来上がり 2014 年 4 月時点で 821 データセットの登録があったこと(2013 年時点では 204 データセット)、FTP より高速にファイルをアップロード可能な Aspera がファイル転送ソフトとして正式採用されたことが報告された(どこからアップロードするかによるが 10 倍程度の高速化)。また現在 PX の登録窓口となっている [PRIDE](#) (PROteomics IDentifications database) からは、新しいインターフェイスとして PRIDE Archive の提供を開始したこと、データベースに登録されているデータをウェブ上で再解析することが可能な PeptideShaker を開発中であることが紹介された。定量データの PX 登録窓口となっている [PASSEL](#) (PeptideAtlas SRM Experiment Library) も登録システムが完成し、既に 50 のデータセットが登録されたこと、今後定量データや SWATH データの検索インターフェイスとして SRMQuantAtlas、SWATHAtlas の開発が進んでいることが紹介された。引き続いてこれから PX コンソーシアムに参加を目指す [MassIVE](#) (Mass Spectrometry Interactive Virtual Environment)、[ProteomicsDB](#)、[iProX](#) (integrated Proteome resources) からそれぞれのデータベ

ースについての紹介があった(その後、6月16日に MassIVE が ProteomeXchange consortium に加盟することが正式に発表された)。

本会議での発表資料は HUPO-PSI (<http://www.psidev.info/psi2014>) および ProteomeXchange (<http://www.proteomexchange.org/storys/meeting-frankfurt-2014>) のウェブページから公開されているので、興味を持たれた方は参考にさせていただきたい。

世界的にデータの共有に対する機運が高まっており、Nature 誌は今年になって研究データ(データの解釈ではなくデータそのもの)を公開するためのジャーナル「Scientific Data」を創刊した。PX コンソーシアムは Scientific Data に投稿する際のプロテオームデータの推奨リポジトリとして指定されており、今後日本のプロテオミクスコミュニティがどのようにこの流れに貢献していくか考えていかねばならないと感じた。今回、JST の協力により European Bioinformatics Institute (EBI) にて PRIDE データベースの開発責任者、ProteomeXchange Consortium のコーディネーターをされている、Juan Antonio Vizcaino 博士を本年の JHUPO 大会に招待することができた。7月18日のシンポジウムにてプロテオームのデータ標準について、翌19日のサテライトシンポジウムにてプロテオームのデータベースについて発表していただくので、お時間のある方は是非ご参加いただき、今後の日本のプロテオミクスコミュニティが進むべき道について活発なご議論をいただきたい。