BIOINFORMATICS MESSAGE FROM THE ISCB Vol. 30 no. 14 2014, pages 2089–2090
doi:10.1093/bioinformatics/btu116

Advance Access publication March 3, 2014

Beyond the proteome: Mass Spectrometry Special Interest Group (MS-SIG) at ISMB/ECCB 2013

MS-SIG 2013 Organizers: So Young Ryu[1], Samuel H. Payne[2], Christoph Schaab[3,4] and Wenzhong Xiao[1,5,*]

[1]Stanford Genome Technology Center, Stanford University, Palo Alto, CA 94304, [2]Pacific Northwest National Laboratory, Richland, WA 99352, USA [3]Max Planck Institute of Biochemistry and [4]Evotec AG 82152, Martinsried, Germany and [5]Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA

Contact: wenzhong.xiao@mgh.harvard.edu

Mass spectrometry special interest group (MS-SIG) aims to bring together experts from the global research community to discuss highlights and challenges in the field of mass spectrometry (MS)-based proteomics and computational biology. The rapid technological developments in MS-based proteomics have enabled the generation of a large amount of meaningful information on hundreds to thousands of proteins simultaneously from a biological sample; however, the complexity of the MS data require sophisticated computational algorithms and software for data analysis and interpretation. This year's MS-SIG meeting theme was 'Beyond the Proteome' with major focuses on improving protein identification/quantification and using proteomics data to solve interesting problems in systems biology and clinical research.

IMPROVING PROTEIN IDENTIFICATION/ QUANTIFICATION

We as the proteomics community need to effectively convey the confidence levels in the MS measurements of protein abundance and the subsequent presentation of protein quantitation. The lack of correlation often found between studies of the proteome and transcriptome can create a perception challenge that proteomics data might not be accurate, and thus, it is important to develop and apply rigorous computational methods for protein quantitation. Besides, identifying all proteins in a biological complex is an ongoing challenge. The first session was dedicated to the current developments in these areas.

At the session, Avinash Shanmugam described a strategy to use RNASeq abundance values and GPMdb protein observation frequencies to adjust the confidence level of protein identification and to improve protein identification. Christoph Schaab addressed the issues of controlling false-positive identification when combining multiple MS datasets. In terms of protein quantification, Mikhail Savitski introduced quantification quality scores based on the local slope of the aligned retention times for label-free approach. For Tandem Mass Tag (TMT) labeled data, he presented methods to identify signal-to-interfering peak ratios and suggested filtering the quantification results based on these quality scores to improve the quantification accuracy. Working with data-independent spectral acquisition, Olga Vitek proposed linear models with an optimal subset of fragments to improve quantification accuracy in this emerging data type. So Young Ryu developed stochastic censored regression considering informative missing peptide intensities and unbalanced total number of quantified peptides across experiments.

*To whom correspondence should be addressed.

PROTEOMICS IN SYSTEMS BIOLOGY AND MEDICINE

A critical facet of proteomics is the experimental hypotheses being explored. The second session of the workshop focused on translating computational improvements to wet-bench biological research. Many biological problems require measuring multiple types of molecules, and currently, a major challenge is how to effectively integrate proteomics data with other omics data such as transcriptomics and metabolomics data. Moreover, there is an unmet need in systems biology and medicine to expand proteomics beyond protein quantitation and begin to systematically assay post-translational modifications of protein (e.g. phosphorylation).

In combining protein and mRNA measurements to study circadian pattern of mouse liver, Jürgen Cox showed that only one-third of cycling proteins are not accompanied by significant cycling of mRNA, thus relying on post-translational regulation. David Gibbs explored methods for integrating transcriptome and proteome data for use with biological networks. Tieliu Shi expanded the human proteome by mining large collections of MS and sequencing data. Using chemical–cross-linking MS, Lars Malmström characterized proteins from the human pathogen *Streptococcus pyogenes*, which play a critical role in the infection process. With top-down proteomics, Julia Chamot-Rooke characterized novel post-translational modifications that regulate infection by *Neisseria meningitides*.

SOFTWARE AND DATA STANDARD/AVAILABILITY

The computational MS community is beginning to invest in community tools and standards. This will facilitate an explosion of ideas that are interoperable and comparable. Frameworks for creating workflows and format standards will make this an exciting place for computational exploration. Throughout the workshop were talks about software, data standards and data repositories to demonstrate how increased compatibility benefits the community. Oliver Kohlbacher presented the *OpenMS/TOPP*

infrastructure. This portable data management and analysis system allows users to create analysis workflows using a variety of open-source tools. There is also community-wide effort to standardize MS-data formats and make data available. Andrew Jones introduced the Proteomics Standards Initiative's new quantitative data format, mzQuantML. Henning Hermjakob showed how PRIDE and the ProteomeXchange consortium provide a data repository with globally coordinated infrastructure.

OPPORTUNITIES FOR BIOINFORMATICIANS IN PROTEOMICS RESEARCH

Our panel session discussed the current opportunities in computational proteomics. Many exciting technological developments have been made in the past decade in the field of MS-based proteomics, and many researchers are now interested in using proteomics to answer questions in systems biology and medicine. A consensus of the panelists is that this is an exciting time for bioinformaticians to study proteomics data and contribute in this area. To effectively integrate the community efforts, open-source tools and well-organized platforms for such tools are essential. The panel also emphasized the importance for bioinformaticians to communicate effectively with chemists and biologists to develop useful tools for proteomics research. Lastly, young computational biologists would benefit from experiences in MS laboratory and interactions with experimental scientists.

Additional Information

MS-SIG will be held at ISMB 2014 and aims to bring together experts from the global research community to discuss current progress and challenges in the field of MS-based proteomics and computational biology. The meeting will contain three portions: oral presentation, panel discussion and poster presentation; it invites abstract submissions on a wide range of topics.

Visit http://www.iscb.org/ismb2014-program/ismb2014-sigs-satellite-meetings#ms-sig for more information.