Proceedings of the EuBIC-MS 2020 Developers' Meeting

Ashwood, Christopher; Bittremieux, Wout; Deutsch, Eric W.; Doncheva, Nadezhda T.; Dorfer, Viktoria; Gabriels, Ralf; Gorshkov, Vladimir; Gupta, Surya; Jones, Andrew R.; Käll, Lukas; Kopczynski, Dominik; Lane, Lydie; Lautenbacher, Ludwig; Legeay, Marc; Locard-Paulet, Marie; Mesuere, Bart; Perez-Riverol, Yasset; Netz, Eugen; Pfeuffer, Julianus; Sachsenberg, Timo; Salz, Renee; Samaras, Patroklos; Schiebenhoefer, Henning; Schmidt, Tobias; Schwämmle, Veit; Soggiu, Alessio; Uszkoreit, Julian; Van Den Bossche, Tim; Van Puyvelde, Bart; Van Strien, Joeri; Verschaffelt, Pieter; Webel, Henry; Willems, Sander

Published in:
EuPA Open Proteomics

DOI:
10.1016/j.euprot.2020.11.001

Publication date:
2020

Document version:
Final published version

EuPA Open Proteomics 24 (2020) 1–6

Contents lists available at ScienceDirect

EuPA Open Proteomics

journal homepage: www.elsevier.com/locate/euprot

Proceedings of the EuBIC-MS 2020 Developers' Meeting

ARTICLE INFO

Keywords
Computational mass spectrometry
Mass spectrometry
Proteomics
Bioinformatics
Spectrum clustering
Phosphoproteomics
XIC extraction
Proteomics graph networks
Predicted spectra
Metaproteomics
Functional annotations
Benchmark development

ABSTRACT

The 2020 European Bioinformatics Community for Mass Spectrometry (EuBIC-MS) Developers' meeting was held from January 13th to January 17th 2020 in Nyborg, Denmark. Among the participants were scientists as well as developers working in the field of computational mass spectrometry (MS) and proteomics. The 4-day program was split between introductory keynote lectures and parallel hackathon sessions. During the latter, the participants developed bioinformatics tools and resources addressing outstanding needs in the community. The hackathons allowed less experienced participants to learn from more advanced computational MS experts, and to actively contribute to highly relevant research projects. We successfully produced several new tools that will be useful to the proteomics community by improving data analysis as well as facilitating future research. All keynote recordings are available on https://doi.org/10.5281/zenodo.3890181.

1. Introduction

The EuBIC-MS Developers' Meeting is organized every other year by the European Bioinformatics Community for Mass Spectrometry (EuBIC-MS, eubic-ms.org), an initiative of the European Proteomics Association (EuPA) for user-oriented bioinformatics. EuBIC-MS promotes the use of bioinformatics for mass spectrometry (MS). Our goal is to bring together the European MS bioinformatics community, including students and early-career researchers as well as long-standing experts from both academia and industry. Through the setup of community-driven initiatives, EuBIC-MS mainly focuses on improving education in computational methods, highlighting job and funding opportunities, promoting international collaborations, publications of specialized studies, and training in specialized software tools. To this end, EuBIC-MS maintains, in collaboration with EuPA, several web resources that include educational videos, grant overviews, a job fair, and tutorials, all available on www.proteomics-academy.org. Besides these online resources, EuBIC-MS regularly organizes workshops and hubs at major international conferences on MS and proteomics. Additionally, an annual conference on computational MS is organized by EuBIC-MS itself, forming an important community outreach effort to bring together bioinformatics researchers from all over Europe.

Since 2017, EuBIC-MS's Winter School takes place every two years [1,2]. These events are highly attended and are a unique opportunity to learn, discover new tools and methods, and discuss current challenges in the field. However, not all computational expertise is utilized to its full potential in a typical conference setup. Therefore, every other year we alternate the Winter School with a Developers' Meeting, which targets software developers and computation-aware end-users. The first official Developers' Meeting took place in Ghent, Belgium in 2018 [3]. This year's meeting was organized in Nyborg, Denmark, from January 13th to January 17th 2020. A total of 54 participants, including trainees and keynote speakers participated in the meeting. It took off with an educational introduction to MS and proteomics by the OpenMS team (www.openms.de) [4] and was followed by 6 keynotes. The next three days, the participants split up into six teams to each work on a project that was proposed and selected through an open call and selection process (github.com/eubic/EuBIC2020). This meeting, organized by and for the European computational MS community, provided a unique opportunity to learn, network, and participate in the development of promising tools. The full program is available on eubic-ms.org/events/2020-developers-meeting (Fig. 1).

2. Keynote presentations

2.1. Eric W. Deutsch - application of the Universal Spectrum Identifier

Eric W. Deutsch (Institute for Systems Biology, Seattle, WA, USA) kicked off the keynote sessions with a presentation on the Universal Spectrum Identifier (USI), a community standard under development in the Mass Spectrometry Standards Working Group of the Human Proteome Organization - Proteomics Standards Initiative (HUPO-PSI, www.psidev.info). The main goal of the USI is to provide a straight-forward and easy-to-use identifier for publicly available MS spectra, optionally including their interpretation (peptidoform and charge). This will be useful to refer to individual key spectrum identifications in research articles, spectral libraries, and spectrum visualization applications. More information on the USI and its development can be found at www.psidev.info/usi.

https://doi.org/10.1016/j.euprot.2020.11.001
Received 30 September 2020
Available online 24 November 2020

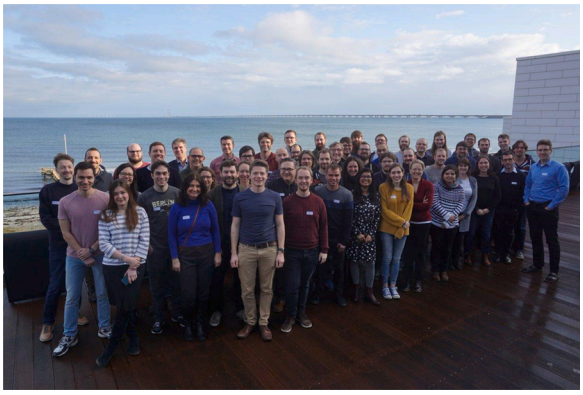

Fig. 1. Participants of the EuBIC-MS Developers' Meeting 2020.

2.2. Andy Jones - Statistical considerations for determining PTM site localization confidence, including discovery of rare or unusual modification types

Joining the Developers' Meeting remotely, Andy Jones (Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Liverpool, GB) addressed some of the open questions and issues in the statistical analysis of post-translational modification (PTM) site localization. Although PTMs can be analyzed in high-throughput by mass spectrometry-based proteomics, pinpointing the exact location of the PTM on the peptide sequence remains challenging. Several tools have been developed to score a PTM's localization, and the concept of false localization rate (FLR) –analogous to the false discovery rate (FDR)– has been established. The score threshold that is expected to result into a given FLR is, however, not generally determined for each data set individually, as is the case for the FDR, but is based on a small amount of benchmark data sets. In his presentation, Andy Jones highlighted the fact that the relation of the localization score and the FLR heavily depends on the data set characteristics, such as instrument settings and the search engine that was used [5]. He therefore proposed to use decoy modifications, for instance phosphorylated alanine in the case of phosphoproteomics, to assess the FLR on a case-by-case basis.

2.3. Olga Vitek - Components of reproducible quantitative mass spectrometry-based research: a statistician's perspective

Covering the statistical side of computational proteomics, Olga Vitek (Northeastern University, Boston, MA, US) described aspects of reproducibility that vary across quantitative analyses, with the end goal of being able to obtain the same results with new experiments and new subjects for a given hypothesis. Towards that goal, she provided a case-study of MSstats [6], a software package developed by her lab to statistically model differentially abundant proteins in proteomics studies. The future of reproducible research was proposed through public data re-analysis with MassIVE.Quant (massive.ucsd.edu/ProteoSAFe/static/massive-quant.jsp) and education/training (May Institute courses on computation and statistics for mass spectrometry and proteomics).

2.4. Alexander Peltzer - Scalable, reproducible bioinformatics workflows using Nextflow & nf-core

The development of workflows is time consuming and complex for beginners in the field. Building on the Nextflow workflow framework, Alexander Peltzer (Quantitative Biology Center, University of Tübingen, Tübingen, DE) presented a community effort to standardize and collect robust workflows. nf-core (https://nf-co.re) offers a collection of tested and curated workflows with documentation and error management, which is portable to any Linux-based platform due to the capability of Nextflow to provide a reproducible and scalable execution framework. This allows the user to run the data analysis workflow –composed of multiple tools for the often computationally demanding data operations– on a local computer, a research high-performance cluster (HPC) or cloud providers. The project supports existing workflows as well as new implementations and will contribute to the reproducibility and comparability of proteomics research.

2.5. Ole N. Jensen - Top-down and middle-down proteomics: experimental methods and computational challenges

Ole N. Jensen (University of Southern Denmark, Odense, DK) presented his team's work on modified proteoforms using top-down and middle-down approaches. Digesting with the endopeptidase Glu-C, which generates long peptides, allows the study of variant-specific N-terminal modifications on histones. With this approach, Tvardovskiy et al. analyzed the relative quantities of modified histone H3 and H3.3 in several mouse tissues and their evolution over time [7]. Computational tools were developed, such as the PTM interplay score that describes the functional links occurring between PTMs: positive as well as negative crosstalk. These can be visualised on a two-dimensional plot with CrossTalkMapper [8] (https://github.com/veitveit/CrossTalkMapper). Nowadays, an increasing number of laboratories utilize intact protein MS to study proteoforms and PTM crosstalk. Although this technology is on the rise, comprehensive proteoform characterization using top-down proteomics remains a challenge. A major bottleneck is that isolation and fragmentation parameters for getting interpretable MS2 spectra are highly protein-dependent, requiring individualized instrument parameters for each protein. Ole N. Jensen's team developed the *topdownr* tool that allows automatic batch-testing of multiple instrument parameters and reports the quality of MS2 spectra for each parameter set [9]. Using this strategy on an increasing number of proteins should provide the basis to predict protein-specific optimal top-down data acquisition parameters in the future.

2.6. Lydie Lane - Mining the dark human proteome using neXtProt's SPARQLing tools

Lydie Lane (Swiss Institute of Bioinformatics, CMU, Geneva, CH) presented the neXtProt database [10] (www.nextprot.org), a protein knowledge base that combines selected publicly available genomic, transcriptomic and proteomic datasets with human UniprotKB/Swiss-Prot sequences and annotations. She described in detail the data structure (RDF format), which has the particularity to be isoform centric. Its web interface provides many visualization options that allow data mining such as mapping of MS-identified peptide sequences to protein sequences, PTM localization, and mapping of variants. Lydie Lane went through a guided tour on how to mine neXtProt data using SPARQL for protein-level data, or SNORQL for any type of data (https://snorql.nextprot.org). These are facilitated by a set of available pre-filled queries, and detailed guide/help pages. neXtProt took on the challenge of mapping functional annotation to the dark proteome and published the list of nearly 2000 proteins without functional annotation from free-text (UniProt summaries), pathway annotations, gene ontology (GO) terms, amongst others, and tackled this challenge using SPARQL queries. They propose hypotheses on the function of 26 of them [11]. The presentation ended by a call to the community: neXtProt provides tools dedicated to assessing peptide unicity, either based on the canonical sequences or on all variant sequences available [12]; or for *in silico* peptide digestion, which are freely available for their integration in independent tools. Furthermore, neXtProt integrates independent tools developed by the community. This paves the way for a federated system of interoperable public databases and tools that would improve functional analysis.

3. Hackathons

During the subsequent days, the participants split up into small groups to actively develop bioinformatics applications. Project proposals for the hackathon sessions were crowd-sourced in a transparent and open process: prior to the meeting, community members could submit project proposals for hackathon sessions, which were subsequently evaluated on scientific merit and community interest.

3.1. Batch XIC and spectra extraction in ThermoRawFileParser (Vladimir Gorshkov, Niels Hulstaert, Yasset Perez-Riverol)

ThermoRawFileParser [13] is an open-source cross-platform software tool that converts raw files from ThermoFisher Scientific MS instruments to open data formats, namely mzML and MGF. Each open format has its specific use case, for example, mzML is best for data retention, *i.e.* preserving the MS data to the fullest extent. The downside is, however, the complex ("heavy") structure of the data. On the other hand, the MGF format –built for supplying spectra to database search engines– is very "light" and thus lacks important metadata. During the hackathon, ThermoRawFileParser was extended with two modules. The first one creates extracted ion chromatograms (XICs) directly from raw files and exports them to a JSON format. The parameters used for XIC retrieval are supplied in the structured JSON: m/z value and tolerance, m/z range, peptide sequence and retention time cutoffs. The second module retrieves individual spectra from a raw file and returns them in a PROXI format that was recently drafted by the Human Proteome Organization Proteomics Standard Initiative (HUPO-PSI, https://github.com/HUPO-PSI/proxi-schemas). Both modules facilitate programmatic access to partial data from a raw file avoiding overhead for converting a complete raw file to mzML or other open formats. They serve as a middle layer between data in a raw file and other services accessing these data. Use cases would be accessing individual scans from raw files stored in repositories using universal spectrum identifier (USI), or batch retrieval of XICs to be displayed on a website. Apart from adding specific functionality to ThermoRawFileParser, the hackathon contributed to creating a community that collectively supports further developments of the software. As a result, one major and three minor releases of ThermoRawFileParser (fixing bugs and adding new features) were published after the hackathon.

3.2. How can we best use Cytoscape for proteomics data analysis? (Nadezhda T. Doncheva, Marc Legeay)

Cytoscape [14] is an open-source software to visualize and analyze biological networks. Developers can add new features to Cytoscape by developing their own Java-based apps. The Cytoscape community also maintains a REST API [15] that enables Cytoscape and most of its apps to be used from R and Python in a more automated way. In this hackathon, we assembled several such automation workflows to facilitate more reproducible proteomics data analysis in Cytoscape. We made use of the core Cytoscape functionality as well as two apps, stringApp [16] and Omics Visualizer [17], which were developed with a focus on the analysis and visualization of proteomics and phosphoproteomics data. We also performed an exploratory study of tissue expression visualization for aging-related proteins in a STRING [18] network by combining gene expression from the TISSUES database [19] and peptide based protein expression evidence for each mapped tissue from the neXtProt database [10]. The resulting workflows, implemented in both R and python, are freely available on GitHub (https://github.com/scaramoncheEuBIC2020_Cytoscape).

3.3. Mapping proteins to functions: method and benchmark development (Bart Mesuere, Pieter Verschaffelt, Henning Schiebenhoefer)

Understanding how the microbiome works requires knowledge about the functions of the expressed proteins [20]. Retrieving these functions is currently one of the major challenges in metaproteomics [21]. This results in an increasing number of proteomics tools that either provide functional information themselves (*e.g.* MetaProteomeAnalyzer [22]), or connect to functional annotation tools such as Unipept [23,24]. Gene Ontology (GO) terms are commonly used to describe protein functionality. Comparing GO terms between tools or datasets is a complex task, hindered, for example, by the different annotation levels of proteins. Moreover, due to the high similarity between some GO terms (*e.g.* one list provides the parent term, while another list provides the child term), exact matching fails in providing the user with an optimal result. Therefore, we developed MegaGO [39], which compares two lists of GO-terms while taking this inherent similarity into account. The tool calculates the relevance semantic similarity metric [25] and returns a single number for similarity. The tool is freely available on Github (github.com/MEGA-GO/MegaGo).

3.4. Online spectrum identification validation by comparison to predicted/experimental spectra (Tobias Schmidt, Patroklos Samaras)

Confirmation of a single MS2 spectrum identification by comparison to other spectra is still not simple (enough). Although many different resources such as PeptideAtlas [26], MassIVE [27], and ProteomicsDB [28,29] have a deep back catalog of high-confidence mass spectra, they are not easily obtainable. During this hackathon, ProteomicsDB extended its API following the FAIR principles [30]. It now offers a direct and query-able interface to $>$ 108,000,000 experimental spectra covering $>$ 70 % of the human and arabidopsis proteome. Every experimental spectrum comes with detailed metadata of the underlying experimental and acquisition schema. Additionally, virtually any unmodified peptide can be predicted in high quality by the deep-learning algorithm Prosit [31] and can also be accessed *via* the above-mentioned API. In the hackathon, we combined this REST interface with the core visualizations of the Interactive Peptide Spectral Annotator (IPSA) [32] to build a minimal website for researchers to visualize their acquired spectra and get an instantaneous comparison to an external source, which can either be a predicted spectrum or one included in ProteomicsDB. In addition, by enabling the use of the Universal Spectrum Identifier (http://psidev.info/usi) every user can request, compare, and validate spectra from any resource implementing this standard (https://www.proteomicsdb.org/use/). The API and the online spectrum identification validator will be available soon.

3.5. Simulating a quantified phosphoproteome for software benchmarking and algorithm development (Marie Locard-Paulet, Veit Schwämmle, Vasileios Tsiamis, Ludwig Lautenbacher)

Many cellular processes are controlled by phosphorylation events, often forming cascades for tight and fast control/adaptation of cellular function. Bottom-up mass spectrometry provides a highly sensitive and high throughput platform to measure these events, and most studies base their results on peptide quantities. Such strategy still suffers from several data analysis challenges, coming from the difficulty to assemble the quantitative behavior of phosphoproteins –real players of cell signaling events– from short peptides and incomplete fragmentation patterns [33,34]. The development of new and better algorithms to determine the quantitative phosphoproteome is hindered by the absence of golden standard datasets with known abundance changes of selected phosphoproteins. Since this is not possible yet (or extremely expensive) to produce such a sample, we undertook the development of a computational pipeline that creates peptide-level MS-like data from a known theoretical proteoform-level gold standard (from a FASTA file). This hackathon resulted in the new PhosFake tool that simulates the full experimental data acquisition of a typical phosphoproteomics experiment. Proteomes (protein sequences with PTMs) with arbitrary *a priori* defined changes at the proteoform level are transformed into

quantitative peptide profiles over any number of experimental conditions and replicates. Currently, the tool implements a total of 38 parameters to simulate the many factors that contribute to experimental design, biological composition, protein digestion, PTM enrichment, MS data acquisition and subsequent computational analysis of the resulting raw spectra. PhosFake will be used to study if and how quantitative changes on the proteoform level can be determined from the noisy and incomplete data obtained in standard phosphoproteomics experiments. The tool, still in development, is available on https://github.com/veitveit/PhosFake.

3.6. Formation of spectral libraries by representative spectra (by Lukas Käll, Yasset Perez-Riverol)

Spectral clustering algorithms aim to accurately and efficiently group large numbers of spectra based on their similarity, such that all spectra in a given cluster stem from the same analyte (peptides in this case) [35]. A frequent output from clustering of spectra are consensus spectra, *i.e.* a common representation of the spectra in each cluster. A consensus spectrum can be used either as a means to annotate the spectra, instead of annotating all the cluster members one can search a single spectrum, or as an optimal representation of the analyte they stem from. Consensus spectra have successfully been utilized for the purpose of quality control of existing peptide identifications in proteomics archives [36], improvement of peptide identification and performing or refining label-free quantification based on the consensus spectra [37, 38]. In such applications, the performance could be assumed to rely on which algorithm constructs the consensus spectra. During the hackathon we implemented a workflow to benchmark different algorithms for assembling consensus spectra (github.com/statisticalbiotechnology/representative-spectra-benchmark). A synthetic peptide library dataset was used to benchmark the different methods for which best-spectrum, binned-spectrum, most-similar-spectrum, and clustering-spectrum were determined. Preliminary results showed no major differences between binned-spectrum and best-spectrum algorithms, which are the two algorithms that perform best in peptide identifications. The group is still working on the project to benchmark more data sets including phosphoproteomics data sets to explore how clustering algorithms and consensus spectra generation can affect the phospho-localization algorithms.

4. Poster presentations

The poster presentations took place the evening of the first day in a relaxed atmosphere. Two prices of 250€ were funded by EuPA for the best posters, which were selected by the keynote speakers (Fig. 2).

5. Conclusion and outlook

The keynote speakers provided a wide overview of the current challenges faced by the community of computational MS. These were tackled by some of the hackathons, such as the development of an *in silico* pipeline for generation of a proteoform centric quantitative gold standard that could be utilized in some of the topics developed by Ole N. Jensen and Andy Jones. The invited speakers that stayed for the entire meeting were seen in several hackathons, providing guidance, and helping out with the integration of their tools in the different projects. For example, the USI presented by Eric W. Deutsch was utilized in several hackathons, and Lydie Lane contributed to the integration of the neXtProt database in one of the pipelines developed for Cytoscape. In our opinion, this Developers' Meeting was a success in providing training, favorizing networking, nesting collaborations, and resulted in several tools that are either already available or under active development.

We believe that working as a community for the community is the most efficient way to go forward, and we hope that our next events will be as productive as this one. If you are interested in joining us, please contact us directly (info@eubic-ms.org). Our next events will be advertised on our website (eubic-ms.org) and posted on our mailing list.

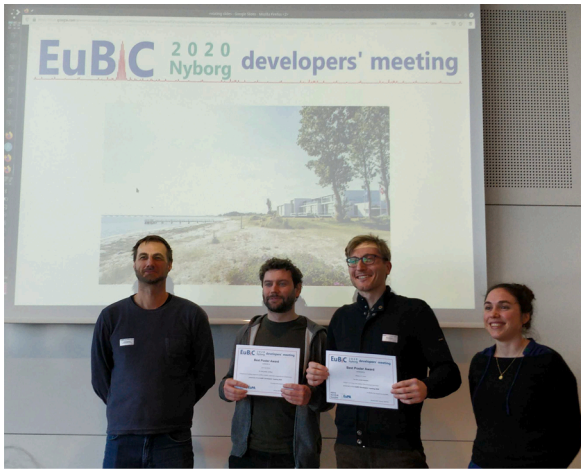

Fig. 2. The winners of the poster presentations: Joeri Van Strien for "Complexome Profiling Alignment (COPAL) enables systematic comparison of complexomes" and Mateusz K. Łacki for "IsoSpec 2.0: Crazy Fast Isotopic Fine Structure Calculator" surrounded by the two main organisers of the event.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgements

Funding for the EuBIC-MS 2020 Developer's Meeting was provided by the Novo Nordisk Foundation. We would like to thank EuPA and the local MS societies of Austria (APMA) and France (SFEAP) for funding travel grants.

We would like to thank Lene B Hørning for an incredibly smooth organization, all EuBIC-MS members who volunteered to help on all fronts, as well as all keynote speakers, hackathon organizers, and participants who contributed to the success of the Developers' Meeting.

References

[1] D. Kopczynski, et al., Proceedings of the EuBIC Winter School 2019, EuPA Open Proteom. 22-23 (2019) 4–7.
[2] S. Willems, et al., Proceedings of the EuBIC Winter School 2017, J. Proteomics 161 (2017) 78–80.
[3] S. Willems, et al., Proceedings of the EuBIC developer's meeting 2018, J. Proteomics 187 (2018) 25–27.
[4] J. Pfeuffer, et al., OpenMS – a platform for reproducible analysis of mass spectrometry data, J. Biotechnol. 261 (2017) 142–148.
[5] S. Ferries, et al., Evaluation of parameters for confident phosphorylation site localization using an orbitrap fusion tribrid mass spectrometer, J. Proteome Res. 16 (9) (2017) 3448–3459.
[6] E. Dogu, et al., MSstatsQC 2.0: R/Bioconductor package for statistical quality control of mass spectrometry-based proteomics experiments, J. Proteome Res. 18 (2) (2019) 678–686.
[7] A. Tvardovskiy, et al., Accumulation of histone variant H3.3 with age is associated with profound changes in the histone methylation landscape, Nucleic Acids Res. 45 (16) (2017) 9272–9289.
[8] R. Kirsch, O.N. Jensen, V. Schwammle, Visualization of the dynamics of histone modifications and their crosstalk using PTM-CrossTalkMapper, Methods (2020), https://doi.org/10.1016/j.ymeth.2020.01.012.
[9] P.V. Shliaha, et al., Maximizing sequence coverage in top-down proteomics by automated multimodal gas-phase protein fragmentation, Anal. Chem. 90 (21) (2018) 12519–12526.
[10] M. Zahn-Zabal, et al., The neXtProt knowledgebase in 2020: data, tools and usability improvements, Nucleic Acids Res. 48 (D1) (2020) D328–D334.
[11] P. Duek, et al., Exploring the uncharacterized human proteome using neXtProt, J. Proteome Res. 17 (12) (2018) 4211–4226.
[12] M. Schaeffer, et al., The neXtProt peptide uniqueness checker: a tool for the proteomics community, Bioinformatics 33 (21) (2017) 3471–3472.

[13] N. Hulstaert, et al., ThermoRawFileParser: modular, scalable, and cross-platform RAW file conversion, J. Proteome Res. 19 (1) (2020) 537–542.
[14] P. Shannon, et al., Cytoscape: a software environment for integrated models of biomolecular interaction networks, Genome Res. 13 (11) (2003) 2498–2504.
[15] D. Otasek, et al., Cytoscape Automation: empowering workflow-based network analysis, Genome Biol. 20 (1) (2019) 185.
[16] N.T. Doncheva, et al., Cytoscape StringApp: network analysis and visualization of proteomics data, J. Proteome Res. 18 (2) (2019) 623–632.
[17] M. Legeay, et al., Visualize omics data on networks with Omics Visualizer, a Cytoscape App, F1000Res 9 (2020) 157.
[18] D. Szklarczyk, et al., STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets, Nucleic Acids Res. 47 (D1) (2019) D607–D613.
[19] O. Palasca, et al., TISSUES 2.0: an integrative web resource on mammalian tissue expression, Database (Oxford) 2018 (2018).
[20] R. Sajulga, et al., Survey of metaproteomics software tools for functional microbiome analysis, bioRxiv (2020).
[21] H. Schiebenhoefer, et al., Challenges and promise at the interface of metaproteomics and genomics: an overview of recent progress in metaproteogenomic data analysis, Expert Rev. Proteomics 16 (5) (2019) 375–390.
[22] T. Muth, et al., The MetaProteomeAnalyzer: a powerful open-source software suite for metaproteomics data analysis and interpretation, J. Proteome Res. 14 (3) (2015) 1557–1565.
[23] R. Gurdeep Singh, et al., Unipept 4.0: functional analysis of metaproteome data, J. Proteome Res. 18 (2) (2019) 606–615.
[24] T. Van Den Bossche, et al., Connecting MetaProteomeAnalyzer and PeptideShaker to unipept for seamless end-to-end metaproteomics data analysis, J. Proteome Res. 19 (8) (2020) 3562–3566.
[25] A. Schlicker, et al., A new measure for functional similarity of gene products based on Gene Ontology, BMC Bioinformatics 7 (2006) 302.
[26] J.M. Schwenk, et al., The human plasma proteome draft of 2017: building on the human plasma PeptideAtlas from mass spectrometry and complementary assays, J. Proteome Res. 16 (12) (2017) 4299–4310.
[27] M. Wang, et al., Assembling the community-scale discoverable human proteome, Cell Syst. 7 (4) (2018) 412–421, e5.
[28] P. Samaras, et al., ProteomicsDB: a multi-omics and multi-organism resource for life science research, Nucleic Acids Res. 48 (D1) (2020) D1153–D1163.
[29] T. Schmidt, et al., ProteomicsDB, Nucleic Acids Res. 46 (D1) (2018) D1271–D1281.
[30] M.D. Wilkinson, et al., The FAIR Guiding Principles for scientific data management and stewardship, Sci. Data 3 (2016) 160018.
[31] S. Gessulat, et al., Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning, Nat. Methods 16 (6) (2019) 509–518.
[32] D.R. Brademan, et al., Interactive peptide spectral annotator: a versatile web-based tool for proteomic applications, Mol. Cell Proteomics 18 (8 suppl 1) (2019) S193–S201.
[33] V. Schwammle, T. Verano-Braga, P. Roepstorff, Computational and statistical methods for high-throughput analysis of post-translational modifications of proteins, J. Proteomics 129 (2015) 3–15.
[34] C. Jorgensen, M. Locard-Paulet, Analysing signalling networks by mass spectrometry, Amino Acids 43 (3) (2012) 1061–1074.
[35] Y. Perez-Riverol, J.A. Vizcaino, J. Griss, Future Prospects of Spectral Clustering Approaches in Proteomics, Proteomics 18 (14) (2018) e1700454.
[36] J. Griss, et al., Recognizing millions of consistently unidentified spectra across hundreds of shotgun proteomics datasets, Nat. Methods 13 (8) (2016) 651–656.
[37] J. Griss, et al., Spectral clustering improves label-free quantification of low-abundant proteins, J. Proteome Res. 18 (4) (2019) 1477–1485.
[38] M. The, L. Käll, Focus on the spectra that matter by clustering of quantification data in shotgun proteomics, Nat. Commun. 11 (1) (2020) 3234.
[39] P. Verschaffelt, T. Van Den Bossche, W. Gabriel, M. Burdukiewicz, A. Soggiu, L. Martens, B. Renard, H. Schiebenhoefer, B. Mesuere, MegaGO: a fast yet powerful approach to assess functional similarity across meta-omics data sets, bioRxiv (2020), https://doi.org/10.1101/2020.11.16.384834, 2020.11.16.384834.

Christopher Ashwood[a,b]
[a] European Bioinformatics Community for Mass Spectrometry (info@eubic-ms.org), Denmark
[b] Department of Cellular and Integrative Physiology, University of Nebraska Medical Center, Omaha, USA

Wout Bittremieux[a,b,c]
[a] European Bioinformatics Community for Mass Spectrometry (info@eubic-ms.org), Denmark
[b] Department of Computer Science, University of Antwerp, Antwerp, Belgium
[c] Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA 92093, USA

Eric W. Deutsch[a,b]
[a] European Bioinformatics Community for Mass Spectrometry (info@eubic-ms.org), Denmark
[b] Institute for Systems Biology, Seattle, WA 98109, USA

Nadezhda T. Doncheva[a,b,c]
[a] European Bioinformatics Community for Mass Spectrometry (info@eubic-ms.org), Denmark
[b] Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Denmark
[c] Center for non-coding RNA in Technology and Health, Department of Veterinary and Animal Sciences, University of Copenhagen, Denmark

Viktoria Dorfer[a,b]
[a] European Bioinformatics Community for Mass Spectrometry (info@eubic-ms.org), Denmark
[b] Bioinformatics Research Group, University of Applied Sciences Upper Austria, Hagenberg, Austria

Ralf Gabriels[a,b,c]
[a] European Bioinformatics Community for Mass Spectrometry (info@eubic-ms.org), Denmark
[b] VIB-UGent Center for Medical Biotechnology, VIB, Ghent, Belgium
[c] Department of Biomolecular Medicine, Ghent University, Ghent, Belgium

Vladimir Gorshkov[a,b]
[a] European Bioinformatics Community for Mass Spectrometry (info@eubic-ms.org), Denmark
[b] Department of Biochemistry and Molecular Biology, University of Southern Denmark, Denmark

Surya Gupta[a,b,c]
[a] European Bioinformatics Community for Mass Spectrometry (info@eubic-ms.org), Denmark
[b] VIB-UGent Center for Medical Biotechnology, VIB, Ghent, Belgium
[c] Department of Biomolecular Medicine, Ghent University, Ghent, Belgium

Andrew R. Jones[a,b]
[a] European Bioinformatics Community for Mass Spectrometry (info@eubic-ms.org), Denmark
[b] Institute of Systems, Molecular and Integrative Biology, University of Liverpool, UK

Lukas Käll[a,b]
[a] European Bioinformatics Community for Mass Spectrometry (info@eubic-ms.org), Denmark
[b] Science for Life Laboratory, KTH - Royal Institute of Technology, Solna, Stockholm, Sweden

Dominik Kopczynski[a,b]
[a] European Bioinformatics Community for Mass Spectrometry (info@eubic-ms.org), Denmark
[b] Leibniz-Institut für Analytische Wissenschaften - ISAS - e.V., Dortmund, Germany

Lydie Lane[a,b]
[a] European Bioinformatics Community for Mass Spectrometry (info@eubic-ms.org), Denmark
[b] SIB-Swiss Institute of Bioinformatics and Department of Microbiology and Molecular Medicine, Faculty of Medicine, Geneva University, Geneva, Switzerland

Ludwig Lautenbacher[a,b]
[a] European Bioinformatics Community for Mass Spectrometry (info@eubic-ms.org), Denmark
[b] Technical University of Munich, Chair for Proteomics and Bioanalytics, Freising, Germany

Marc Legeay[a,b]
[a] European Bioinformatics Community for Mass Spectrometry (info@eubic-ms.org), Denmark
[b] Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Denmark

Marie Locard-Paulet[a,b,*]

[a] European Bioinformatics Community for Mass Spectrometry (info@eubic-ms.org), Denmark
[b] Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Denmark

Bart Mesuere[a,b,c]
[a] European Bioinformatics Community for Mass Spectrometry (info@eubic-ms.org), Denmark
[b] VIB-UGent Center for Medical Biotechnology, VIB, Ghent, Belgium
[c] Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Ghent, Belgium

Yasset Perez-Riverol[a,b]
[a] European Bioinformatics Community for Mass Spectrometry (info@eubic-ms.org), Denmark
[b] European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK

Eugen Netz[a,b,c]
[a] European Bioinformatics Community for Mass Spectrometry (info@eubic-ms.org), Denmark
[b] Applied Bioinformatics, Dept. of Computer Science, University of Tübingen, Tübingen, Germany
[c] Biomolecular Interactions, Max Planck Institute for Developmental Biology, Tübingen, Germany

Julianus Pfeuffer[a,b]
[a] European Bioinformatics Community for Mass Spectrometry (info@eubic-ms.org), Denmark
[b] Applied Bioinformatics, Dept. of Computer Science, University of Tübingen, Tübingen, Germany

Timo Sachsenberg[a,b]
[a] European Bioinformatics Community for Mass Spectrometry (info@eubic-ms.org), Denmark
[b] Applied Bioinformatics, Dept. of Computer Science, University of Tübingen, Tübingen, Germany

Renee Salz[a,b]
[a] European Bioinformatics Community for Mass Spectrometry (info@eubic-ms.org), Denmark
[b] Center for Molecular and Biomolecular Informatics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen, the Netherlands

Patroklos Samaras[a,b]
[a] European Bioinformatics Community for Mass Spectrometry (info@eubic-ms.org), Denmark
[b] Technical University of Munich, Chair for Proteomics and Bioanalytics, Freising, Germany

Henning Schiebenhoefer[a,b,c]
[a] European Bioinformatics Community for Mass Spectrometry (info@eubic-ms.org), Denmark
[b] Data Analytics and Computational Statistics, Hasso-Plattner-Institute, Faculty of Digital Engineering, University of Potsdam, Potsdam, Germany
[c] Bioinformatics Unit (MF1), Department for Methods Development and Research Infrastructure, Robert Koch Institute, Berlin, Germany

Tobias Schmidt[a,b]
[a] European Bioinformatics Community for Mass Spectrometry (info@eubic-ms.org), Denmark
[b] Technical University of Munich, Chair for Proteomics and Bioanalytics, Freising, Germany

Veit Schwämmle[a,b,**]
[a] European Bioinformatics Community for Mass Spectrometry (info@eubic-ms.org), Denmark
[b] Department of Biochemistry and Molecular Biology, University of Southern Denmark, Denmark

Alessio Soggiu[a,b]
[a] European Bioinformatics Community for Mass Spectrometry (info@eubic-ms.org), Denmark
[b] Department of Biomedical, Surgical and Dental Sciences, One Health Unit, University of Milan, Milan, Italy

Julian Uszkoreit[a,b,c]
[a] European Bioinformatics Community for Mass Spectrometry (info@eubic-ms.org), Denmark
[b] Ruhr University Bochum, Center for Protein Diagnostics (PRODI), Medical Proteome Analysis, Bochum, Germany
[c] Ruhr University Bochum, Medical Faculty, Medizinisches Proteom-Center, Bochum, Germany

Tim Van Den Bossche[a,b,c]
[a] European Bioinformatics Community for Mass Spectrometry (info@eubic-ms.org), Denmark
[b] VIB-UGent Center for Medical Biotechnology, VIB, Ghent, Belgium
[c] Department of Biomolecular Medicine, Ghent University, Ghent, Belgium

Bart Van Puyvelde[a,b]
[a] European Bioinformatics Community for Mass Spectrometry (info@eubic-ms.org), Denmark
[b] ProGenTomics, Laboratory of Pharmaceutical Biotechnology, Ghent University, Ghent, Belgium

Joeri Van Strien[a,b]
[a] European Bioinformatics Community for Mass Spectrometry (info@eubic-ms.org), Denmark
[b] Center for Molecular and Biomolecular Informatics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen, the Netherlands

Pieter Verschaffelt[a,b,c]
[a] European Bioinformatics Community for Mass Spectrometry (info@eubic-ms.org), Denmark
[b] VIB-UGent Center for Medical Biotechnology, VIB, Ghent, Belgium
[c] Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Ghent, Belgium

Henry Webel[a,b]
[a] European Bioinformatics Community for Mass Spectrometry (info@eubic-ms.org), Denmark
[b] Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Denmark

Sander Willems[a,b]
[a] European Bioinformatics Community for Mass Spectrometry (info@eubic-ms.org), Denmark
[b] ProGenTomics, Laboratory of Pharmaceutical Biotechnology, Ghent University, Ghent, Belgium

[*] Corresponding author at: Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Denmark.

[**] Corresponding author at: Department of Biochemistry and Molecular Biology, University of Southern Denmark, Denmark.
E-mail address: marie.locard-paulet@cpr.ku.dk (M. Locard-Paulet).
E-mail address: veits@bmb.sdu.dk (V. Schwämmle).