



Technical Note

Proceedings of the EuBIC-MS developers meeting 2023



Pedro Beltrao^a, Tim Van Den Bossche^{b,c}, Ralf Gabriels^{b,c}, Tanja Holstein^{b,c}, Tobias Kockmann^d, Alireza Nameni^{b,c}, Christian Panse^{d,e,*}, Ralph Schlapbach^d, Ludwig Lautenbacher^f, Matthias Mattanovich^g, Alexey Nesvizhskii^h, Bart Van Puyveldeⁱ, Jonas Scheid^{j,k}, Veit Schwämmle^l, Maximilian Strauss^m, Anna Klimovskaia Susmelj^{n,o}, Matthew The^p, Henry Webel^q, Mathias Wilhelm^f, Dirk Winkelhardt^r, Witold E. Wolski^{d,e}, Muyao Xi^g

^a Department of Biology, ETH Zürich, Switzerland

^b VIB-UGent Center for Medical Biotechnology, 9052 Zwijnaarde, Belgium

^c Department of Biomolecular Medicine, Ghent University, 9000 Ghent, Belgium

^d Functional Genomics Center Zürich, ETH Zürich/University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

^e Swiss Institute of Bioinformatics, Quartier Sorge - Batiment Amphipole, 1015 Lausanne, Switzerland

^f Computational Mass Spectrometry, TUM School of Life Sciences, Technical University of Munich, Maximus-von-Imhof-Forum 3, D - 85354 Freising, Germany

^g Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Blegdamsvej 3B, DK-2200 Copenhagen, Denmark

^h Departments of Pathology and Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48105, USA

ⁱ ProGenTomics, Laboratory of Pharmaceutical Biotechnology, Ghent University, Ottergemsesteenweg 460, BE-9000 Ghent, Belgium

^j Department of Peptide-based Immunotherapy, Institute of Immunology, University and University hospital Tübingen, Auf der Morgenstelle 15, D-72076 Tübingen, Germany

^k Quantitative Biology Center (QBiC), University of Tübingen, Auf der Morgenstelle 10, D-72076 Tübingen, Germany

^l Department of Biochemistry and Molecular Biology, University of Southern Denmark, Campusvej 55, 5230 Odense, Denmark

^m Proteomics Program, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

ⁿ Biognosys, Schlieren, Zürich 8952, Switzerland

^o Now AI Center ETH Zürich, Switzerland

^p TUM School of Life Sciences Technische Universität München, D - 85354 Freising, Germany

^q Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

^r Ruhr-Universität Bochum, Germany

ARTICLE INFO

Keywords:

Artificial intelligence
Benchmark development
Bioinformatics
Computational mass spectrometry
Deep learning
Mass spectrometry
Metabolomics
Metaproteomics
Predicted spectra
Proteomics
Spectrum clustering
Visualization

ABSTRACT

The 2023 European Bioinformatics Community for Mass Spectrometry (EuBIC-MS) Developers Meeting was held from January 15th to January 20th, 2023, in Congressi Stefano Franscin at Monte Verità in Ticino, Switzerland. The participants were scientists and developers working in computational mass spectrometry (MS), metabolomics, and proteomics. The 5-day program was split between introductory keynote lectures and parallel hackathon sessions focusing on “Artificial Intelligence in proteomics” to stimulate future directions in the MS-driven omics areas. 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 actively contribute to highly relevant research projects. We successfully produced several new tools applicable to the proteomics community by improving data analysis and facilitating future research.

* Corresponding author at: Functional Genomics Center Zürich, ETH Zurich/University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland.

E-mail address: cp@fgcz.ethz.ch (C. Panse).

<https://doi.org/10.1016/j.jprot.2024.105246>

Received 26 April 2024; Received in revised form 19 June 2024; Accepted 27 June 2024

Available online 2 July 2024

1874-3919/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The EuBIC-MS Developers Meeting is organized every other year by the European Bioinformatics Community for Mass Spectrometry (EuBIC-MS, <https://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, early-career researchers, and long-standing experts from 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 <https://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]. 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, we alternate the Winter School every other year with a Developers Meeting, which targets software developers and computation-aware end-users. The first official Developers Meeting occurred in Ghent, Belgium, in 2018 [2], followed by the one in Nyborg, Denmark, in 2020 [3]. The 2023 European Bioinformatics Community for Mass Spectrometry (EuBIC-MS) Developers Meeting was held from 15 January to 20 January 2023, in Congressi Stefano Franscini at Monte Verità in Ticino, Switzerland. The participants were scientists and developers working in computational MS, metabolomics, and proteomics. The five-day program was divided between introductory keynote lectures and parallel hackathon sessions focusing on “Artificial Intelligence in proteomics” with the objective of stimulating future directions in the MS-driven omics areas. During the latter, 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 actively contribute to highly relevant research projects. The participants of the meeting jointly developed several novel tools that enhance the efficiency and scope of proteomics data analysis that will facilitate future research. It took off with five keynotes and a poster flash talk session where 26 participants gave a 60-s ‘elevator pitch’ on their poster. Detailed statistics on the participants and the meeting are presented in Table 1. In the next four and a half days, the participants split up into seven teams to each work on a project that was proposed and selected through an open call and selection process [4]. On the final day, we showcased two sponsor presentations and unveiled the output from the hackathons. Additionally, we

Table 1
Statistics of the meeting.

Location	Monte Verità, CH
Duration	5 days
# Participants	58
# Keynotes	5
# Hackathons	7
# Posters	26
% Female	>25%
# Institutions	25
# EU Countries	>10
# Sponsors	4

announced the winner of the best poster. This meeting, organized by and for the European computational MS community, provided a unique opportunity to learn, network, and participate in developing promising tools. The detailed program is available on <https://eubic-ms.org/events/2023-developers-meeting>. (See Fig. 1.)

2. Keynote presentations

2.1. Alexey Nesvizhskii - open search: open-minded exploration of proteomics data

Alexey Nesvizhskii has been fascinated with the subject of the “dark proteome” since his entry in the field proteomics in 2001. What are all those spectra that we cannot identify in a typical search, and how do we design efficient computational strategies to move beyond standard peptide identification searches. To this end, his group has designed a highly efficient fragment ion indexing algorithm and implemented it in MSFragger [5], which has become a widely used tool and the engine behind our FragPipe computational platform. MSFragger “open” and “mass offset” searches have empowered new strategies for faster and more sensitive identification of biologically or chemically modified peptides. In his presentation, he described recent improvements in their algorithms, including localization-aware open search [6] and new methods for the identification of N- and O-linked glycopeptides [7], labile PTMs [8], and chemically labeled peptides in chemoproteomics experiments [9]. He also provided an overview of the various quantification workflows (DIA, TMT, LFQ-MBR) available in FragPipe <https://fragpipe.nesvilab.org/>.

2.2. Maximilian Strauss - towards the next generation of proteomics analysis with the AlphaPept ecosystem

Like other omics fields, mass spectrometry (MS)-based proteomics experiences ever-increasing amounts of data with data, demanding swift and effective analysis tools. Tapping into recent computer science strides, Maximilian Strauss presented AlphaPept [10], a Python framework for rapid processing of high-resolution MS data. Utilizing Numba for on-the-fly compilation, AlphaPept offers a hundred-fold increase in processing speed over native Python and can leverage Python’s scientific ecosystem and machine learning innovations. Its Jupyter Notebooks facilitate community contributions, while robust engineering practices ensure reliability. AlphaPept can process proteomes within minutes, and supports automated pipelines via a user-friendly interface or use as a library, with its codebase openly available on GitHub. Building on the same design principles, we have since expanded to the AlphaPept ecosystem, which now offers a suite of libraries catering to various stages, from raw data processing to advanced deep learning predictions and downstream statistical and machine learning analyses for biomarker discovery.

2.3. Mathias Wilhelm - The DOME recommendations for machine learning exemplified on Prosit

In light of rising concerns over a reproducibility crisis in machine and deep learning [11], Mathias Wilhelm (Computational Mass Spectrometry, Technical University of Munich, Freising, Germany) presented DOME (data, optimization, model and evaluation), a set of community-wide recommendations for reporting supervised machine learning-based analyses [12], and its interpretation in proteomics research [13]. He showed this on the example of developing Prosit, a neural network developed by his group that predicts chromatographic retention time and fragment ion intensities. As machine and deep learning is seeing rapid adoption in proteomics, best practices and reporting standards like DOME are good starting points for scientists, as author and reviewer, to ensure high quality research. The presentation ended by a call to the community: let’s learn from each other instead of struggling



Fig. 1. Participants of the EuBIC-MS Developers Meeting 2023 on Monte Verità, Ticino, Switzerland. source: <https://twitter.com/kusterlab/status/1614974632569749504>

alone.

2.4. Anna Susmelj - predicting physicochemical properties of peptides using deep learning

Anna Susmelj's presentation delves into the crucial field of domain generalization in machine learning, with a focus on strategies to bolster model adaptability across diverse datasets while ensuring domain shift robustness. The talk addresses methods to mitigate performance degradation when models confront previously unseen domains, highlighting their practical significance. Additionally, the presentation discusses common pitfalls and testing strategies for AI models in various domains, aiming to contribute to the development of robust and versatile machine learning systems tailored for Mass Spectrometry applications.

2.5. Pedro Beltrao - towards a structurally resolved human protein interaction network

All cellular functions are governed by complex molecular machines that assemble through protein-protein interactions. Their tissue or cell type specificity and atomic details are critical to the study of their molecular and cellular mechanisms but fewer than 5% of hundreds of thousands of human interactions have been structurally characterized [14] and their tissue specificity is essentially unknown. Pedro Beltrao (ETH Zurich) discussed the potential and limitations of recent progress in deep-learning methods using AlphaFold2 [15] to predict structures for human interactions [16]. Higher confidence AlphaFold2 models were correctly enriched in interactions supported by affinity purification or structure based methods and can be orthogonally confirmed by spatial constraints defined by cross-link data. In addition, there were several examples of how the predicted binary complexes can be used to build larger assemblies. To study tissue specificity, it was shown how large-scale protein abundance measurements across samples can be used

to predict tissue-specific protein interactions [17,18]. Protein co-expression is a stronger predictor of protein interactions than mRNA co-expression and can be used to build tissue specific protein interaction networks. These tissue specific networks can be used to link trait/disease associated genes to specific tissues and can serve as a resource for studying tissue specific differences in cell biology and disease.

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 [4] 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. Maximilian Strauss - DeepScore: community curated scoring, supercharged with AI

State of the art for identification in mass-spectrometry-based proteomics is to generate "non-sense", decoy data and to determine a score cutoff based on a false discovery rate. Nowadays, machine (ML) and deep learning (DL) algorithms are used to learn how to optimally distinguish targets from decoys. While this drastically increases sensitivity, it comes at the cost of explainability, and human-chosen acceptance criteria are replaced with black-box models. This can hinder acceptance in clinical practice, e.g., in peptidomics, where upregulated proteins are investigated by investigating raw data peaks. While other DL-driven domains allow straightforward human validation of the models, e.g., imaging, speech, text, or inspecting a predicted structure from AlphaFold, for proteomics, this is much more challenging. Here, we aimed to explore the limitations of the current scoring approach and provide potential solutions. We revisited the idea of confidence by trying to artificially increase identifications with non-sense features, hard

decoys, or leaking data. Next, we will build an interactive tool to validate identifications manually and assign human confidence scores. With this, we created a training dataset and build an ML or DL- model to rescore identifications and assign predicted human-level confidence scores (see [19]).

3.2. Matthias Mattanovich and Muyao Xi - MS2 spectra matching for metabolite identification

One major open topic in untargeted metabolomics is identifying unknown compounds from mass spectra. As MS1 comparisons can be ambiguous (especially for small molecules), we need to look at MS2 spectra, and compare them to public MS2 databases, to differentiate compounds in the same mass range. Currently, the best performing methods for compound identification are GNPS and Sirius. They provide a user with a list of potential compounds, but in some cases the uncertainty is very high or multiple candidates are suggested, making the downstream analysis labor intensive. GNPS improves their predictions by using molecular networks and taking biological information into account[a]. Sirius improves their predictions by comparing structural similarity of the compounds. We would like to set up a novel system, with modular parts that can be tested separately. Each aspect of the pipeline can be improved/modified individually, and multiple methods can be combined as an ensemble. In doing so, this can also serve as a benchmark of existing scoring and matching functions and a testing playground for novel ideas. During this hackathon [20], different spectral similarity scoring functions were compared and their target and off-target scores were evaluated based on similarity of the compounds.

3.3. Ludwig Lautenbacher - shaping a European prediction service for biological data

We developed Koina (see <http://koina.proteomicsdb.org/>), an open-source and modular machine learning (ML) inference server for biological data based on NVIDIA Triton. This platform allows centralizing the hosting of models originating from various training/prediction pipelines such as PyTorch, Sklearn, XGBoost and TensorFlow. Most importantly, models can be accessed by a remote and language-independent interface. A similar system has been used for the last 5 years within Skyline (C#), the Prosit website (Python), and ProteomicsDB (JavaScript) to access Prosit models. We envision that such a platform will be centrally hosted by, for example, the EBI, and model submission will become a standard practice for publishing new ML models. Although accompanying codebases on platforms like GitHub describe the access of models, they require technical skill to set up. Standardizing the access of the models will dramatically improve reproducibility, and ensure easy access for as many people as possible. As part of this hackathon [21], we successfully extended our high-performance prediction service, named Koina, by adding multiple machine-learning models from the DeepLC [22], MS2PIP [23,24], AlphaPept [25] and Prosit [26–28] ecosystems. Since then, we have established a common interface between all these models, enabling the seamless generation of predictions with all models [29].

3.4. Matthew The - easy-to-use interactive HTML plots collection for data exploration and web tools

Interactive data exploration and publicly accessible web applications for bioinformatics tools are an essential part of bioinformatics. Ideally, one uses the same framework for both to reduce development time. Shiny apps are popular for this purpose, but are limited to the R language, while Python's interactive plotting solutions (Plotly, Bokeh) tend to have a steep learning curve and are tricky to deploy. For ProteomicsDB <https://www.proteomicsdb.org/>, we developed over a dozen interactive, general-purpose plots based on D3.js and Vue.js <https://github.com/wilhelm-lab/proteomicsdb-components>.

These range from simple scatter and bar plots to interaction graphs, heatmaps and pathway viewers. We managed to easily reuse and adapt these components in ongoing projects, allowing quick deployment upon publication. For the hackathon, we propose releasing these plots as Web Components <https://www.webcomponents.org/introduction>, the HTML standard that will shape the web for the coming decade. Web Components are easily includable in web pages as simple HTML tags without JavaScript knowledge. As this is purely a frontend framework, data processing and provisioning can be done in any programming language. We aim to provide examples and documentation to make integration for bioinformaticians easy and straightforward in any project. During the hackathon [30], five teams employed pair programming techniques to each develop a new component for a fundamental plot type: histogram, bar plot, swarm plot, violin plot, and line plot. Each team successfully produced a functional component, complete with customization options and unit tests. These components are now available from the node package manager (npm) as part of the @biowc organization.

3.5. Dirk Winkelhardt - Rusteomics - community driven toolbox for omic-research

The proteomics community created some exceptional toolboxes over the past years, like OpenMS, Pyteomics or mzR. Most of these toolboxes implement general computational tasks, like reading and preprocessing data. However, most of them do not rely on a mutual code base or the same internal data representation, which makes interoperability only possible by using (PSI) standard formats. Reading/writing them without a common implementation may introduce another layer of errors. The aim of Rusteomics is to build a collaborative community-driven toolbox, which provides read and write access to the most common file formats, as well as low-level and well established algorithms like (de)isotoping, deconvolution, MS/MS spectra annotation, etc.). While similar solutions exist in various programming languages, this project will be the opportunity to tailor these new components to be highly compatible with scientific (scripting) languages like Python / R. Moreover, the reimplementation in Rust should bring some major benefits: The modern compiler and building system makes Rust-based projects easier to maintain than C++ – based projects, while providing the same performance. During the hackathon [31] the participants build the basis for a community driven mass-spectrometry based proteomics framework written in Rust, starting with prototypes for a spectral library writer (MSP) and a FASTA reader, including easy to use Python bindings.

3.6. Veit Schwämmle and Arthur Grimaud - making sense of internal fragment ions

Peptide identification from fragment mass spectra uses only part of the contained information. Here, internal fragment ions, i.e. peptides with both termini cleaved, have a high potential to provide further evidence about peptide identity. Despite the option to include internal ions in several database search engines, their actual use has so far been explored only poorly. A big challenge lies in the large number of possible ions, and thus the difficulty in distinguishing them from background signals or other fragment ions. This hackathon project [32] aimed to shed more light into the applicability of internal ions by creating a framework to determine their characteristic patterns in MS data. We provided statistics and extensive visualizations for internal ions in a given data set. For that we employed both raw data files and identifications from a database search. This framework established the grounds for the detection and utilization of characteristic internal ions in a dataset, explore potential “fragment motifs”, and facilitate the distinction of actual internal ions from background noise. A clearer understanding and exploration of internal fragmentation will channel future efforts towards a more extensive use of them in MS data processing leading to higher peptide identification rates.

3.7. Tim van Den Bossche - exploring and solving functional analysis gaps in metaproteomics

Metaproteomics, the study of the complete protein complement of microbial communities, is becoming an increasingly popular approach for examining microbiomes as it reveals their actual functions [33]. These functions are described using functional annotations such as InterPro and Gene Ontology (GO) terms, which can often be mapped to functional pathways [34]. During the 2020 Developers Meeting [3], we developed a tool that enables researchers to better compare multiple sets of GO terms, improving upon traditional exact mapping methods [35]. Although this tool facilitates functional comparisons based on GO terms, significant challenges remain in optimizing the functional analyses of microbiomes, particularly in relation to metabolic pathways. To address these issues, we first explored various options and identified gaps in the field during the hackathon. This led to the creation of a tool called *PathwayPilot*. After developing the initial proof-of-concept at the hackathon [36], we further refined *PathwayPilot* into a user-friendly interface for exploring and visualizing metabolic pathways (<https://pathwaypilot.ugent.be/>, [37]).

4. Poster presentations

Twenty-six posters took center stage, accompanied by dynamic 60-s flash talk presentations. One event's highlight was the election of the poster price winner (see Fig. 2. The jury consists of the five keynote speakers.



Fig. 2. The winner of the poster presentations and the poster price - Tanja Holstein Federal Institute of Materials Research and Testing (BAM), Germany; now with the VIB. source: <https://twitter.com/HolsteinTanja/status/1616410121600483328>.

4.1. Tanja Holstein - probability based taxonomic profiling of viral and microbiome samples using PepGM and Unipept

Taxonomic inference in mass spectrometry-based metaproteomics is complex. The exact composition of metaproteomic samples is usually unknown, requiring large reference databases for peptide identification. The presence of proteins and corresponding taxa is then inferred from the identified peptides, which is complicated by protein homology: many proteins do not only share peptides within a single taxon but also between different taxa. Current taxonomic profiling approaches rely on strategies such as peptide-spectrum-match counting or the use of unique peptides. The metaproteomic analysis tool Unipept relies on a pre-computed database of tryptic peptides and performs taxonomic identifications based on a threshold of peptide-taxon matches. It does not provide confidence estimates for its taxonomic assignments and restricts its results to lowest common ancestor level. We present PepGM, a graphical model that uses belief propagation to compute the marginal distributions of peptides and taxa. Developed for the probability-based taxonomic identification of viral samples against a large reference database background, we extend it to the analysis of metaproteomic samples. Taxonomic information for peptides identified is now retrieved using the database provided by Unipept, which was extended with functionalities to restrict the taxa queried and provide all potential taxonomic origins of the peptides queried. As a multitude of taxa may be present, this streamlines the process of obtaining the peptide-taxon relationships required by PepGM. We evaluate the combination of PepGM and Unipept using viral and metaproteomic datasets. We show that our approach results in taxonomic profiles for metaproteomes with statistically computed confidence values. To address the complexity of the taxonomic composition of metaproteomes, results can be aggregated at different taxonomic levels. PepGM enables a robust taxonomic analysis of metaproteomic samples supporting statistically sound inference of taxa in mass spectrometric datasets. This eliminates the need for error-prone heuristics, providing the added benefit of confidence estimates for taxonomic assignments and improved taxonomic resolution (see also the preprint [38]).

5. Conclusion and outlook

The EuBIC-MS Developers Meeting 2023 follows the successful previous winter schools and developers meetings.

In a follow-up survey, 50% of the participants expressed their overall satisfaction with the meeting, with two-thirds of the survey respondents giving it a perfect score [5] and one-third a good score [4]. Overall, opinions varied, with a significant portion favoring either more leisure time or more organized activities, while a notable percentage found the existing balance to be satisfactory.

The EuBIC-MS Developers Meeting spanning a few days, witnessed substantial progress across all hackathon sessions, showcasing participants' impressive productivity and teamwork. The commitment of hackathon groups to ongoing collaboration and project completion holds promise for future scientific publications and improved computational mass spectrometric solutions to support proteomics and metabolomics.

Encouraged by the community's enthusiastic support, the next EuBIC-MS Developers Meeting is scheduled to take place in February 2025, in Brixen, Italy <https://eubic-ms.org/events/2025-developers-meeting/>.

CRedit authorship contribution statement

Pedro Beltrao: Writing – original draft. **Tim Van Den Bossch:** Writing – original draft. **Ralf Gabriels:** Writing – original draft. **Tanja Holstein:** Writing – original draft. **Tobias Kockmann:** Project administration, Conceptualization. **Alireza Nameni:** Writing – original draft. **Christian Panse:** Writing – review & editing, Writing – original draft,

Supervision, Project administration, Conceptualization. **Ralph Schlappbach**: Project administration, Funding acquisition. **Ludwig Lautenbacher**: Writing – original draft. **Matthias Mattanovich**: Writing – original draft. **Alexey Nesvizhskii**: Writing – original draft. **Bart Van Puyvelde**: Writing – review & editing, Project administration, Conceptualization. **Jonas Scheid**: Writing – original draft. **Veit Schwämmle**: Writing – review & editing, Writing – original draft, Project administration, Conceptualization. **Maximilian Strauss**: Writing – original draft. **Anna Klimovskaia Susmelj**: Writing – original draft. **Matthew The**: Writing – original draft. **Henry Webel**: Writing – review & editing, Conceptualization. **Mathias Wilhelm**: Writing – original draft. **Dirk Winkelhardt**: Writing – original draft. **Witold E. Wolski**: Writing – review & editing. **Muyao Xi**: Writing – original draft.

Declaration of competing interest

The authors declare no conflict of interest.

Data availability

No

Acknowledgements

Sponsoring for the EuBIC-MS 2023 Developers Meeting was provided by Matrix Science, MSAID and Biognosys AG. The poster price was sponsored by the Functional Genomics Center Zürich of ETH Zürich and University of Zürich. The organization was supported by Congressi Stefano Franscini on Monte Verità, Switzerland. We would like to thank EuPA for their continuing support. We would like to thank 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] S. Willems, D.e.M. David, M. Locard-Paulet, K. Mechtler, V. Mmle, J. Uszkoreit, M. Vaudel, V. Dorfer, Proceedings of the EuBIC Winter School 2017, J. Proteome 161 (2017) 78–80, <https://doi.org/10.1016/j.jprot.2017.04.001>.
- [2] S. Willems, D.e.D. Deforce, V. Dorfer, V. Gorchkov, D. Kopczynski, K. Laukens, M. Locard-Paulet, V. Mmle, J. Uszkoreit, D. Valkenborg, M. Vaudel, W. Bittremieux, Proceedings of the EuBIC developer's meeting 2018, J. Proteome 187 (2018) 25–27, <https://doi.org/10.1016/j.jprot.2018.05.015>.
- [3] Christopher Ashwood, Wout Bittremieux, Eric W. Deutsch, Nadezhda T. Doncheva, Viktoria Dorfer, Ralf Gabriels, Vladimir Gorchkov, Surya Gupta, Andrew R. Jones, Lukas Käll, Dominik Kopczynski, Lydie Lane, Ludwig Lautenbacher, Marc Legeay, Marie Locard-Paulet, Bart Mesuere, Yasset Perez-Riverol, Eugen Netz, Julianus Pfeuffer, Timo Sachsenberg, Renee Salz, Patroklos Samaras, Henning Schiebenhoefer, Tobias Schmidt, Veit Schwämmle, Alessio Soggiu, Julian Uszkoreit, Tim Van Den Bossche, Bart Van Puyvelde, Joeri Van Strien, Pieter Verschaffelt, Henry Webel, Sander Willems, Proceedings of the eubic-ms 2020 developers' meeting, in: EuPA Open Proteomics, 24: 1–6, December 2020, <https://doi.org/10.1016/j.euprot.2020.11.001>. ISSN 2212–9685.
- [4] Christian Panse, Tobias Kockmann, EuBIC 2023 Developers Meeting, URL, <https://github.com/EuBIC/EuBIC2023/issues>, 2022.
- [5] A.T. Kong, F.V. Leprevost, D.M. Avtonomov, D. Mellacheruvu, A.I. Nesvizhskii, MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics, Nat. Methods 14 (5) (2017) 513–520, <https://doi.org/10.1038/nmeth.4256>.
- [6] F. Yu, G.C. Teo, A.T. Kong, S.E. Haynes, D.M. Avtonomov, D.J. Geiszler, A. I. Nesvizhskii, Identification of modified peptides using localization-aware open search, Nat. Commun. 11 (1) (2020) 4065, <https://doi.org/10.1021/ac0341261>.
- [7] D.A. Polasky, F. Yu, G.C. Teo, A.I. Nesvizhskii, Fast and comprehensive N- and O-glycoproteomics analysis with MSFragger-Glyco, Nat. Methods 17 (11) (2020) 1125–1132, <https://doi.org/10.1093/nar/gkwy936>.
- [8] D.A. Polasky, D.J. Geiszler, F. Yu, K. Li, G.C. Teo, A.I. Nesvizhskii, MSFragger-labile: A flexible method to improve labile PTM analysis in proteomics, Mol. Cell. Proteomics 22 (5) (2023) 100538, <https://doi.org/10.1101/2022.09.12.507594>.
- [9] D.J. Geiszler, D.A. Polasky, F. Yu, A.I. Nesvizhskii, Detecting diagnostic features in MS/MS spectra of post-translationally modified peptides, Nat. Commun. 14 (1) (2023) 4132, <https://doi.org/10.1021/pr070542g>.
- [10] Maximilian T. Strauss, Isabell Bludau, Wen-Feng Zeng, Eugenia Voytik, Constantin Ammar, Julia P. Schessner, Rajesh Ilango, Michelle Gill, Florian Meier, Sander Willems, Matthias Mann, AlphaPept: a modern and open framework for ms-based proteomics, Nat. Commun. 15 (1) (2024), <https://doi.org/10.1038/s41467-024-46485-4>. ISSN 2041–1723.
- [11] E. Gibney, Could machine learning fuel a reproducibility crisis in science? Nature 608 (7922) (2022) 250–251, [https://doi.org/10.1016/S2589-7500\(19\)30123-2](https://doi.org/10.1016/S2589-7500(19)30123-2).
- [12] I. Walsh, D. Fishman, D. Garcia-Gasulla, T. Titma, G. Pollastri, J. Harrow, F. E. Psomopoulos, S.C.E. Tosatto, E. Capriotti, R. Casadio, S. Capella-Gutierrez, D. Cirillo, A. Del Conte, A.C. Dimopoulos, V.D. Del Angel, J. Dopazo, P. Fariselli, J. M. Ndez, F. Huber, A. Kreshuk, T. Lenaerts, P.L. Martelli, A. Navarro, Ó. Broin, J. Ero, D. Piovesan, M. Reczko, F. Ronzano, V. Satagopam, C. Savojardo, V. Spiwok, M.A. Tangaro, G. Tartari, D. Salgado, A. Valencia, F. Zambelli, DOME: recommendations for supervised machine learning validation in biology, Nat. Methods 18 (10) (2021) 1122–1127, <https://doi.org/10.1038/nbt926>.
- [13] M. Palmblad, S. Böcker, S. Degroove, O. Kohlbacher, L. Li, W.S. Noble, M. Wilhelm, Interpretation of the DOME recommendations for machine learning in proteomics and metabolomics, J. Proteome Res. 21 (4) (2022) 1204–1207, <https://doi.org/10.1038/s41592-021-01241-0>.
- [14] Roberto Mosca, Arnaud Céol, Patrick Aloy, Interactome3D: adding structural details to protein networks, Nat. Methods 10 (1) (January 2013) 47–53, <https://doi.org/10.1038/nmeth.2289>. ISSN 1548–7105.
- [15] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Židek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A.A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, Demis Hassabis, Highly accurate protein structure prediction with AlphaFold, Nature 596 (7873) (July 2021) 583–589, <https://doi.org/10.1038/s41586-021-03819-2>. ISSN 1476–4687.
- [16] David F. Burke, Patrick Bryant, Inigo Barrio-Hernandez, Danish Memon, Gabriele Pozzati, Aditi Shenoy, Wensi Zhu, Alistair S. Dunham, Pascal Albanese, Andrew Keller, Richard A. Scheltema, James E. Bruce, Alexander Leitner, Petras Kundrotas, Pedro Beltrao, Arne Elofsson, Towards a structurally resolved human protein interaction network, Nat. Struct. Mol. Biol. 30 (2) (January 2023) 216–225, <https://doi.org/10.1038/s41594-022-00910-8>. ISSN 1545–9985.
- [17] Emanuel Gonçalves, Athanasios Fragoulis, Luz Garcia-Alonso, Thorsten Cramer, Julio Saez-Rodriguez, Pedro Beltrao, Widespread post-transcriptional attenuation of genomic copy-number variation in cancer, Cell Syst. 5 (4) (October 2017), <https://doi.org/10.1016/j.cels.2017.08.013>, 386–398.e4. ISSN 2405–4712.
- [18] Theodoros I. Roumeliotis, Steven P. Williams, Emanuel Gonçalves, Clara Alsinet, Martin Del Castillo Velasco-Herrera, Nanne Aben, Fatemeh Zamanzad Ghavideh, Magali Michaut, Michael Schubert, Stacey Price, James C. Wright, Lu Yu, Mi Yang, Rodrigo Dienstmann, Justin Guinney, Pedro Beltrao, Alvis Brazma, Mercedes Pardo, Oliver Stegle, David J. Adams, Lodewyk Wessels, Julio Saez-Rodriguez, Ultan McDermott, Jyoti S. Choudhary, Genomic determinants of protein abundance variation in colorectal cancer cells, Cell Rep. 20 (9) (August 2017) 2201–2214, <https://doi.org/10.1016/j.celrep.2017.08.010>. ISSN 2211–1247.
- [19] Maximilian Strauss, DeepScore: Community Curated Scoring, Supercharged with AI, URL, <https://github.com/EuBIC/EuBIC2023/issues/14>, 2022.
- [20] Matthias Mattanovich, Muyao Xi, Ms2 Spectra Matching for Metabolite Identification, URL, <https://github.com/EuBIC/EuBIC2023/issues/13>, 2022.
- [21] Ludwig Lautenbacher, Shaping A European Prediction Service for Biological Data, URL, <https://github.com/EuBIC/EuBIC2023/issues/12>, 2022.
- [22] R. Bouwmeester, R. Gabriels, N. Hulstaert, L. Martens, S. Degroove, DeepLC can predict retention times for peptides that carry as-yet unseen modifications, Nat. Methods 18 (11) (2021) 1363–1369, <https://doi.org/10.1007/s13361-016-1460-7>.
- [23] S. Degroove, L. Martens, MS2PIP: a tool for MS/MS peak intensity prediction, Bioinformatics 29 (24) (2013) 3199–3203, <https://doi.org/10.1093/bioinformatics/btt544>.
- [24] S. Degroove, D. Madelein, L. Martens, MS2PIP prediction server: compute and visualize MS2 peak intensity predictions for CID and HCD fragmentation, Nucleic Acids Res. 43 (W1) (2015) W326–W330, <https://doi.org/10.1093/nar/gkv542>.
- [25] W.F. Zeng, X.X. Zhou, S. Willems, C. Ammar, M. Wahle, I. Bludau, E. Voytik, M. T. Strauss, M. Mann, AlphaPeptDeep: a modular deep learning framework to predict peptide properties for proteomics, Nat. Commun. 13 (1) (2022) 7238, <https://doi.org/10.1093/bioinformatics/bt921>.
- [26] S. Gessulat, T. Schmidt, D.P. Zolg, P. Samaras, K. Schnatbaum, J. Zerweck, T. Knaute, J. Rechenberger, B. Delanghe, A. Huhmer, U. Reimer, H.C. Ehrlich, S. Aiche, B. Kuster, M. Wilhelm, Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning, Nat. Methods 16 (6) (2019) 509–518, <https://doi.org/10.1021/acs.jproteome.6b00915>.
- [27] W. Gabriel, M. The, D.P. Zolg, F.P. Bayer, O. Shouman, L. Lautenbacher, K. Schnatbaum, J. Zerweck, T. Knaute, B. Delanghe, A. Huhmer, H. Wenschuh, U. Reimer, G. dard, B. Kuster, and M. Wilhelm., Prosit-TMT: deep learning boosts identification of TMT-labeled peptides, Anal. Chem. 94 (20) (2022) 7181–7190, <https://doi.org/10.1021/acs.analchem.1c05435>.
- [28] M. Wilhelm, D.P. Zolg, M. Graber, S. Gessulat, T. Schmidt, K. Schnatbaum, C. Schwencke-Westphal, P. Seifert, N. Tzig, J. Zerweck, T. Knaute, E. Unlein, P. Samaras, L. Lautenbacher, S. Klaeger, H. Wenschuh, R. Rad, B. Delanghe, A. Huhmer, S.A. Carr, K.R. Clauser, A.M. Krackhardt, U. Reimer, B. Kuster, Deep learning boosts sensitivity of mass spectrometry-based immunopeptidomics, Nat. Commun. 12 (1) (2021) 3346, <https://doi.org/10.1182/blood-2010-12-326231>.
- [29] Ludwig Lautenbacher, Kevin L. Yang, Tobias Kockmann, Christian Panse, Matthew Chambers, Elias Kahl, Yu Fengchao, Wassim Gabriel, Dulguun Bold,

- Tobias Schmidt, Kai Li, Brendan MacLean, Alexey I. Nesvizhskii, Mathias Wilhelm, Koina: Democratizing machine learning for proteomics research, *bioRxiv* (2024), <https://doi.org/10.1101/2024.06.01.596953>.
- [30] Matthew The, Easy-to-use Interactive HTML Plots Collection for Data Exploration and web TOOLS, URL, <https://github.com/EuBIC/EuBIC2023/issues/11>, 2022.
- [31] Dirk Winkelhardt, Rusteomics - Community Driven Toolbox for Omic-research, URL, <https://github.com/EuBIC/EuBIC2023/issues/10>, 2022.
- [32] Veit Schwämmle, Arthur Grimaud, Making Sense of Internal Fragment Ions, URL, <https://github.com/EuBIC/EuBIC2023/issues/9>, 2022.
- [33] Tim Van Den Bossche, Magnus Ø. Arntzen, Dörte Becher, Dirk Benndorf, Vincent G. H. Eijssink, Céline Henry, Pratik D. Jagtap, Nico Jehmlich, Catherine Juste, Benoit J. Kunath, Bart Mesuere, Thilo Muth, Phillip B. Pope, Jana Seifert, Alessandro Tanca, Sergio Uzzau, Paul Wilmes, Robert L. Hettich, Jean Armengaud, The metaproteomics initiative: a coordinated approach for propelling the functional characterization of microbiomes, *Microbiome* 9 (1) (December 2021), <https://doi.org/10.1186/s40168-021-01176-w>. ISSN 2049-2618.
- [34] Henning Schiebenhoefer, Tim Van Den Bossche, Stephan Fuchs, Bernhard Y. Renard, Thilo Muth, Lennart Martens, Challenges and promise at the interface of metaproteomics and genomics: an overview of recent progress in metaproteogenomic data analysis, *Expert Rev. Proteomics* 16 (5) (April 2019) 375–390, <https://doi.org/10.1080/14789450.2019.1609944>. ISSN 1744-8387.
- [35] Pieter Verschaffelt, Tim Van Den Bossche, Wassim Gabriel, Michał Burdukiewicz, Alessio Soggiu, Lennart Martens, Bernhard Y. Renard, Henning Schiebenhoefer, Bart Mesuere, Megago: a fast yet powerful approach to assess functional gene ontology similarity across meta-omics data sets, *J. Proteome Res.* 20 (4) (March 2021) 2083–2088, <https://doi.org/10.1021/acs.jproteome.0c00926>. ISSN 1535-3907.
- [36] Tim Van Den Bossche, Exploring and Solving Functional Analysis Gaps in Metaproteomics, URL, <https://github.com/EuBIC/EuBIC2023/issues/7>, 2022.
- [37] Tibo Vande Moortele, Pieter Verschaffelt, Nadezhda T. Doncheva, Tanja Holstein, Qingyao Huang, Caroline Jachmann, Peter Dawyndt, Lennart Martens, Bart Mesuere, Tim Van Den Bossche, PathwayPilot: A User-friendly Tool for Visualizing and Navigating Metabolic Pathways, June 2024 doi: 10.1101/2024.06.21.599989.
- [38] Tanja Holstein, Pieter Verschaffelt, Tim Van den Bossche, Lennart Martens, Thilo Muth, The peptonizer2000: Graphical Model Based Taxonomic Identifications of Metaproteomic Samples, May 2024, <https://doi.org/10.1101/2024.05.20.594958>.