



Proceedings of the EuBIC Winter School 2019



Dominik Kopczynski^a, Wout Bittremieux^b, David Bouyssié^c, Viktoria Dorfer^d,
Marie Locard-Paulet^e, Bart Van Puyvelde^f, Veit Schwämmle^g, Alessio Soggiu^h, Sander Willems^f,
Julian Uszkoreit^{i,*}

^a Leibniz-Institut für Analytische Wissenschaften – ISAS – e.V., Bunsen-Kirchhoff-Str. 11, D-44139, Dortmund, Germany

^b University of Antwerp, Antwerp, Belgium

^c Institute of Pharmacology and Structural Biology, University of Toulouse, CNRS, UPS, Toulouse, France

^d Bioinformatics Research Group, University of Applied Sciences Upper Austria, Hagenberg, Austria

^e Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Denmark¹

^f Laboratory of Pharmaceutical Biotechnology, Ghent University, Ghent, Belgium

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

^h Department of Veterinary Medicine, University of Milan, Milan, Italy

ⁱ Ruhr University Bochum, Faculty of Medicine, Medizinisches Proteom-Center, Gesundheitscampus 4, D-44801, Bochum, Germany

ARTICLE INFO

Keywords:

EuBIC, Computational proteomics,
Bioinformatics, Protein quantification, Data
independent acquisition

ABSTRACT

The 2019 European Bioinformatics Community (EuBIC) Winter School was held from January 15th to January 18th 2019 in Zakopane, Poland. This year's meeting was the third of its kind and gathered international researchers in the field of (computational) proteomics to discuss (mainly) challenges in proteomics quantification and data independent acquisition (DIA). Here, we present an overview of the scientific program of the 2019 EuBIC Winter School. Furthermore, we can already give a small outlook to the upcoming EuBIC 2020 Developer's Meeting.

1. Introduction

The 2019 EuBIC Winter School on proteomics bioinformatics, held from January 15th to January 18th 2019 in Zakopane, Poland, was the third conference of its kind organised by the European Bioinformatics Community (EuBIC, [1,2]), an initiative of the European Proteomics Association (EuPA) for user-oriented bioinformatics. The Winter School brought together scientists from both academia and industry in the field of computational mass spectrometry (MS)-based proteomics to present and discuss their research in workshops, keynote lectures, flash talks and poster presentations. The 93 participants (see Fig. 1) came from 18 different European countries, together with guests from the USA, Canada and China. Besides the scientific programme, which included 12 workshops, 10 keynote talks, 10 flash talks and 35 poster presentations, the participants were also given time to network, socialize and meet representatives of our main sponsors BSI, ProFI, Thermo Fisher Scientific and SVA. To trigger interactions and alleviate the mood, the conference bags contained tongue-in-cheek advice, such as “cite the tools you use”, “do calculate FDRs” or “join EuBIC” as well as a personal

“Conference Bingo” sheet. For the Bingo, each participant had a number of words and events to cross out, which might occur in a talk (“open data”, “I would like to thank...”, “somebody starts clapping too early”, “presentation with font too small to read”). This Bingo culminated on the final day with a “BINGO” shout mid-session as a participant completed his version by crossing off “Speaker offers post-doc position”. Finally, as the main social event of the Winter School, a pub quiz with questions concerning proteomics and informatics with free drinks was hosted by the organizers at a local venue in Zakopane.

As the EuBIC has a special focus on teaching and education, the first day of the Winter School was dedicated to educational workshops on computational MS-based proteomics organised primarily by the community itself. On the second and third days the invited speakers and sponsors gave interactive keynotes in the morning and more advanced workshops in the afternoon. The fourth and final day consisted of another half day of keynotes given by the invited speakers and final discussions. The full conference program can be consulted online.²

* Corresponding author.

E-mail address: julian.uszkoreit@rub.de (J. Uszkoreit).

¹ The center is supported financially by the Novo Nordisk Foundation (Grant agreement NNF14CC0001).

² https://github.com/eubic/eubic_winterschool_programme/releases/tag/final2019.



Fig. 1. Participants of the EuBIC Winter School 2019. In the background, the mountains surrounding the conference venue in the city of Zakopane can be seen.

2. Workshops

The first day of the Winter School featured two full day workshops and two half-day workshops on four different educational topics in parallel: (1) introduction to computational mass spectrometry using OpenMS (H. Röst, L. Bichmann and O. Alka) (2), computational introduction into data independent acquisition (DIA) (M. Dhaenens and B. Searle), and (3a) Label-free quantification: concepts and algorithms and (3b) quantitative proteomics: statistics, clustering and complexes (D. Bouyssie and V. Schwämmle). These workshops consisted partly of presentations giving an overview of the specific topic as well as hands-on tutorials to encourage the participants to work with these concepts.

After the keynote lectures on the second day, the participants could attend one of four parallel workshops organised by the invited speakers and sponsors. M. Eisenacher and J. Uszkoreit talked about “The essentials before and after spectrum identification: Selecting the appropriate database and inference strategy” [3] and guided participants through practicals exploring the protein database exports of the UniProt KB [4], analyses in the workflow environment KNIME [5] and performing protein inference using PIA [6,7]. In the workshop entitled “Quality Control and Benchmarking of Label-Free Quantification Workflows with LFQBench” S. Tenzer and M. Lacki explained how to apply the R package LFQBench [8] to proteomics data and how to interpret the results. Furthermore there were two workshops of our main sponsors: “Discovering the open-source Proline software suite, a new efficient and user friendly solution for label-free quantification” by ProFI (<http://www.profi-proteomics.fr/>) and a workshop organised by ThermoFisher Scientific about the new features of Proteome Discoverer 2.3.

On the afternoon of the third day, four more workshops were offered. L. J. Jensen presented the “Network visualization with Cytoscape and stringApp” [9] with hands-on practicals, while M. Wilhelm presented a “Validation of peptide identifications” using ProteomicsDB [10]. In parallel,

E. Petsalaki guided the participants through an analysis using SELPHI in “SELPHI: using data-driven approaches for analysis of phosphoproteomics datasets” [11] in the first half of the workshop before F. Meier who showed the “Advanced data acquisition methods with MaxQuant.Live” [12] during the second half.

3. Keynotes

For this year’s Winter School we had in total ten keynote slots with 45 min talk and discussion each: four on both Wednesday and Thursday and two on Friday before the closing session. In keeping with the tradition of the first Winter Schools, the presenters were asked to talk

about problems and challenges in the field of bioinformatics for proteomics they encountered during their recent research. To emphasize this, this year’s special focus was “Challenges in Quantification and DIA”.

On Tuesday, after the welcome and announcement, B. Searle (Institute for Systems Biology, Seattle, WA, USA) kicked off the keynotes with a talk about a new database search engine to detect genetic variants from DIA data using the tool XCorDIA [13]. After this, H. Röst (The Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Canada) spoke about developing tools for the personalized medicine revolution and using mass spectrometry for longitudinal molecular profiling [14–16]. Here, he showed how new tools could be used to align the features of mass spectrometry analyses measured over a long time period using spanning trees. S. Tenzer (Institute for Immunology, University Medical Center of the Johannes-Gutenberg University Mainz, Mainz, Germany) highlighted how his lab was able to perform label-free quantification of complex proteomes using the still emerging ion-mobility-based DIA approaches [17]. Additionally, he noted the value and importance of reproducible chromatographic approaches and recent advances of microflow methods in this area. The fourth talk on this day was given by M. Eisenacher (Ruhr University Bochum, Bochum, Germany). He talked about historical researches which later were falsified, but seemed to be a relevant hypothesis at their time. With respect to this, he stated the question about the existence of a “reverse translase” and its hypothetical applications in proteomics.

On the third day, the first talk was given by L. J. Jensen (University of Copenhagen, Copenhagen, Denmark) about STRING [18] and recent updates to its database. Exemplary stated in this regard were the text mining procedures and performances used to find protein-protein-interactions. M. Wilhelm (Chair of Proteomics and Bioanalytics, Technical University of Munich, Freising, Germany) presented Prosit, a tool to predict MS/MS spectra from a given peptide sequence, which was trained on the data generated by the ProteomeTools project [19,20]. The following talks on Thursday were oriented towards MS-based data analysis and how to drive biological insights from the huge amount of data generated nowadays. E. Petsalaki (EMBL-EBI, Wellcome Genome Campus, Hinxton, UK) discussed re-analysing phosphoproteomics data (“Using phosphoproteomics data to study context-specific signalling”) and illustrated the associated challenges [21]. Then, K. Lilley (University of Cambridge, Cambridge, UK) ended the morning session with a talk entitled “Insights into the multi-functioning proteome” where she exchanged views on the importance and difficulties of accurately measuring the proteome system-wise in a space- and time-defined fashion [22].

On the last day, S. Degroevé (VIB-UGent Center for Medical Biotechnology, Ghent, Belgium) presented Ionbot, a fully data-driven search engine for open modification and mutation searches. He highlighted the importance of accurate MS/MS spectra prediction [23–25] and could directly connect his thoughts to the talk of M. Wilhelm the day before. The final keynote of the Winter School was about trapped ion mobility spectrometry, which adds a new dimension for mass spectrometry-based proteomics and was given by F. Meier (Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany) [26].

4. Flash talks and poster presentations

Each participant was encouraged to present their recent research in a poster presentation. Ten of the 35 submitted and peer reviewed posters were selected for a five-minute flash talk of the presented topic. The keynote speakers had the responsibility to vote for the best flash talk and the best poster. This was undoubtedly hard work during the lively poster session, which started directly after the workshops and lasted until well after midnight. During this time, all participants had the opportunity to talk about their research and discuss ideas about



Fig. 2. The winners of the flash talks and poster presentations. From left to right: Dominik Kopczynski of the organizing committee, the poster price winners Tim van den Bossche and Leon Bichmann, followed by the winners of the flash talks Ralf Gabriels and Henning Schiebenhöfer.

bioinformatic analyses of proteomics data. The awards for the best flash talk were given to Ralf Gabriels (VIB-UGent Center for Medical Biotechnology, Ghent, Belgium) and his talk on “MS2 peak intensity prediction (MS2PIP) for specific PTMs, fragmentation techniques and instruments” and Henning Schiebenhöfer (Robert Koch-Institute, Berlin, Germany) with “Prophane – Metaproteomic Data Analysis and Interpretation Made Simple”. The awards for the best poster went to Tim van den Bossche (VIB-UGent Center of Medical Biotechnology, Ghent, Belgium) and the poster “Prediction-based reduction of the search space in metaproteomics” and Leon Bichmann (University of Tübingen, Tübingen, Germany) for the poster “Immunopeptidomics of human tissues using DDA and DIA” (Fig. 2).

5. Outlook

In this third EuBIC Winter School, we were again able to attract many renowned keynote speakers presenting their interesting and exciting research topics in proteomics. Even if the number of 93 participants was slightly smaller than the previous edition, we are still very happy about the success of the conference. A post-conference survey confirmed the organizers’ impression of the participants’ general high satisfaction. The few negative experiences included the rather expensive venue, but most attendees will very likely join another EuBIC event and about 20% are even eager to have a longer Winter School with more workshops next time.

Next year’s EuBIC conference will again be a *Developer’s Meeting*, like the one in 2018 [2], with a more *hackathon*-like program comprising more hands-on coding and less keynote talks. The meeting will take place in Denmark in January 2020 and the planning of the program is already underway.

We, the organizing committee, emphasize that any member of the community who wants to become active in the EuBIC can join our channels³ and help us in ongoing and future activities with as much effort as each one has available. We would also like to remind everyone in the proteomics community that we have a special “Q&A” mailing list⁴ for all bioinformatics for proteomics or mass spectrometry related questions that often yields prompt answers by the community.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgements

The EuBIC likes to thank the the main sponsors EuPA and Elsevier, platinum sponsor BSI, the gold sponsors ProFI, ThermoFisher Scientific and SVA and poster price sponsor Molecular Omics for their support of the Winter School.

The organizing committee especially thanks the speakers and workshop organizers for bringing the Winter School to such a level of excellence, companies and exhibitors for their active participation in this community initiative, and all participants for the active and fruitful interactions.

References

- [1] S. Willems, D. Bouyssié, M. David, M. Locard-Paulet, K. Mechtler, V. Schwämmle, J. Uszkoreit, M. Vaudel, V. Dorfer, *Proceedings of the EuBIC Winter School 2017*, *J. Proteomics* 161 (2017) 78–80.
- [2] S. Willems, D. Bouyssié, D. Deforce, V. Dorfer, V. Gorshkov, D. Kopczynski, K. Laukens, M. Locard-Paulet, V. Schwämmle, J. Uszkoreit, D. Valkenburg, M. Vaudel, W. Bittremieux, *Proceedings of the EuBIC developer’s meeting 2018*, *J. Proteomics* 187 (2018) 25–27.
- [3] M. Turewicz, M. Kohl, M. Ahrens, G. Mayer, J. Uszkoreit, W. Naboulsi, T. Bracht, D.A. Megger, B. Sitek, K. Marcus, M. Eisenacher, *BiolInfra.Prot: A comprehensive proteomics workflow including data standardization, protein inference, expression analysis and data publication*, *J. Biotechnol.* 261 (2017) 116–125.
- [4] U. Consortium, UniProt: a worldwide hub of protein knowledge, *Nucleic Acids Res.* 47 (2019) D506–D515.
- [5] M.R. Berthold, N. Cebon, F. Dill, T.R. Gabriel, T. Kotter, T. Meinl, P. Ohl, C. Sieb, K. Thiel, B. Wiswedel, KNIME: the konstanz information miner, *Data Anal. Mach. Learning App.* (2008) 319–326.
- [6] J. Uszkoreit, A. Maerkens, Y. Perez-Riverol, H.E. Meyer, K. Marcus, C. Stephan, O. Kohlbacher, M. Eisenacher, PIA: an intuitive protein inference engine with a web-based user interface, *J. Proteome Res.* 14 (2015) 2988–2997.
- [7] J. Uszkoreit, Y. Perez-Riverol, B. Eggers, K. Marcus, M. Eisenacher, Protein inference using PIA workflows and PSI standard file formats, *J. Proteome Res.* 18 (2019) 741–747.
- [8] P. Navarro, J. Kuharev, L.C. Gillet, O.M. Bernhardt, B. MacLean, H.L. Röst, S.A. Tate, C.C. Tsou, L. Reiter, U. Distler, G. Rosenberger, Y. Perez-Riverol, A.I. Nesvizhskii, R. Aebersold, S. Tenzer, A multicenter study benchmarks software tools for label-free proteome quantification, *Nat. Biotechnol.* 34 (2016) 1130–1136.
- [9] N.T. Doncheva, J.H. Morris, J. Gorodkin, L.J. Jensen, Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data, *J. Proteome Res.* 18 (2019) 623–632.
- [10] T. Schmidt, P. Samaras, M. Frejino, S. Gessulat, M. Barnert, H. Kienegger, H. Krccmar, J. Schlegl, H.C. Ehrlich, S. Aiche, B. Kuster, M. Wilhelm, *ProteomicsDB*, *Nucleic Acids Res.* 46 (2018) D1271–D1281.
- [11] E. Petsalaki, A.O. Helbig, A. Gopal, A. Pasculescu, F.P. Roth, T. Pawson, SELPHI: correlation-based identification of kinase-associated networks from global phosphoproteomics data sets, *Nucleic Acids Res.* 43 (2015) W276–282.
- [12] C. Wichmann, F. Meier, S. Virreira Winter, A.D. Brunner, J. Cox, M. Mann, MaxQuant.Live Enables Global Targeting of More Than 25,000 Peptides, *Mol. Cell Proteomics* 18 (2019) 982–994.
- [13] B.C. Searle, L.K. Pino, J.D. Egerton, Y.S. Ting, R.T. Lawrence, B.X. MacLean, J. Villén, M.J. MacCoss, Chromatogram libraries improve peptide detection and quantification by data independent acquisition mass spectrometry, *Nat. Commun.* 9 (2018) 5128.
- [14] S. Gupta, S. Ahadi, W. Zhou, H. Röst, DIALignR Provides Precise Retention Time Alignment Across Distant Runs in DIA and Targeted Proteomics, *Mol. Cell Proteomics* 18 (2019) 806–817.
- [15] B.D. Piening, W. Zhou, K. Contrepois, H. Röst, G.J. Gu Urban, T. Mishra, B.M. Hanson, E.J. Bautista, S. Leopold, C.Y. Yeh, D. Spakowicz, I. Banerjee, C. Chen, K. Kukurba, D. Perelman, C. Craig, E. Colbert, D. Salins, S. Rego, S. Lee, C. Zhang, J. Wheeler, M.R. Sailani, L. Liang, C. Abbott, M. Gerstein, A. Mardinoglu, U. Smith, D.L. Rubin, S. Pitteri, E. Sodergren, T.L. McLaughlin, G.M. Weinstock, M.P. Snyder, Integrative personal omics profiles during periods of weight gain and loss, *Cell Syst.* 6 (2018) 157–170 e158.
- [16] H. Fröhlich, R. Balling, N. Beerenwinkel, O. Kohlbacher, S. Kumar, T. Lengauer, M.H. Maathuis, Y. Moreau, S.A. Murphy, T.M. Przytycka, M. Rebhan, H. Röst, A. Schuppert, M. Schwab, R. Spang, D. Stekhoven, J. Sun, A. Weber, D. Ziemek, B. Zupan, From hype to reality: data science enabling personalized medicine, *BMC Med.* 16 (2018) 150.
- [17] U. Distler, J. Kuharev, P. Navarro, Y. Levin, H. Schild, S. Tenzer, Drift time-specific collision energies enable deep-coverage data-independent acquisition proteomics, *Nat. Methods* 11 (2014) 167–170.
- [18] D. Szklarczyk, A.L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas,

³ <https://www.proteomics-academy.org/bioinformatics>

⁴ <https://qa.proteomics-academy.org/>

- M. Simonovic, N.T. Doncheva, J.H. Morris, P. Bork, L.J. Jensen, C.V. Mering, STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets, *Nucleic Acids Res.* 47 (2019) D607–D613.
- [19] D.P. Zolg, M. Wilhelm, K. Schnatbaum, J. Zerweck, T. Knaute, B. Delanghe, D.J. Bailey, S. Gessulat, H.C. Ehrlich, M. Weininger, P. Yu, J. Schlegl, K. Kramer, T. Schmidt, U. Kusebauch, E.W. Deutsch, R. Aebersold, R.L. Moritz, H. Wenschuh, T. Moehring, S. Aiche, A. Huhmer, U. Reimer, B. Kuster, Building ProteomeTools based on a complete synthetic human proteome, *Nat. Methods* 14 (2017) 259–262.
- [20] 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 (2019) 509–518.
- [21] G. Giudice, E. Petsalaki, Proteomics and phosphoproteomics in precision medicine: applications and challenges, *Brief Bioinform* (2017).
- [22] A. Geladaki, N. Kočevár Britovšek, L.M. Breckels, T.S. Smith, O.L. Vennard, C.M. Mulvey, O.M. Crook, L. Gatto, K.S. Lilley, Combining LOPIT with differential ultracentrifugation for high-resolution spatial proteomics, *Nat. Commun.* 10 (2019) 331.
- [23] S. Degroeve, L. Martens, MS2PIP: a tool for MS/MS peak intensity prediction, *Bioinformatics* 29 (2013) 3199–3203.
- [24] S. Degroeve, D. Maddelein, L. Martens, MS2PIP prediction server: compute and visualize MS2 peak intensity predictions for CID and HCD fragmentation, *Nucleic Acids Res.* 43 (2015) W326–330.
- [25] R. Gabriels, L. Martens, S. Degroeve, Updated MS²PIP web server delivers fast and accurate MS² peak intensity prediction for multiple fragmentation methods, instruments and labeling techniques, *Nucleic Acids Res.* (2019).
- [26] F. Meier, A.D. Brunner, S. Koch, H. Koch, M. Lubeck, M. Krause, N. Goedecke, J. Decker, T. Kosinski, M.A. Park, N. Bache, O. Hoerning, J. Cox, O. Räther, M. Mann, Online parallel accumulation-serial fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer, *Mol. Cell Proteomics* 17 (2018) 2534–2545.