The HUPO-PSI Quality Control Working Group : making quality control more accessible for biological mass spectrometry

The HUPO-PSI Quality Control Working Group: Making quality control more accessible for biological mass spectrometry

Wout Bittremieux,[†,‡] Mathias Walzer,[¶,§] Stefan Tenzer,[‖] Weimin Zhu,[⊥] Reza M. Salek,[#] Martin Eisenacher,[@] and David L. Tabb[*,△]

†*Department of Mathematics and Computer Science, University of Antwerp, Middelheimlaan 1, 2020 Antwerp, Belgium*

‡*Biomedical Informatics Research Center Antwerp (biomina), University of Antwerp/Antwerp University Hospital, Wilrijkstraat 10, 2650 Edegem, Belgium*

¶*Department of Computer Science, University of Tübingen, Tübingen, Germany*

§*Center for Bioinformatics, University of Tübingen, Tübingen, Germany*

‖*Institute for Immunology, University Medical Center of the Johannes-Gutenberg University Mainz, Germany*

⊥*National Center for Protein Science - Beijing, No. 38, Science Park Road, Changping District, Beijing, 102206, China*

#*European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, United Kingdom*

@*Medical Bioinformatics, Medizinisches Proteom-Center, Ruhr-University Bochum, Germany*

△*Division of Molecular Biology and Human Genetics, Stellenbosch University Faculty of Medicine and Health Sciences, Tygerberg Hospital, Francie Van Zijl Drive, Cape Town, 7505, South Africa*

E-mail: dtabb@sun.ac.za

Phone: +27 21 938-9403. Fax: +27 21 938-9863

Abstract

In order to be confident of the results acquired during biological mass spectrometry experiments, a systematic approach to quality control is of vital importance. Nonetheless, until now only scattered initiatives have been undertaken to this end, and these individual efforts have often not been complementary. To address this issue, the Human Proteome Organization (HUPO) – Proteomics Standards Initiative (PSI) has established a new working group on quality control at its meeting in the spring of 2016. The goal of this working group is to provide a unifying framework for quality control data. The initial focus will be on providing a community-driven standardized file format for quality control. For this purpose the previously proposed qcML format will be adapted to support a variety of use cases for both proteomics and metabolomics applications, and it will be established as an official PSI format. An important consideration is to avoid enforcing restrictive requirements on quality control, but instead provide the basic technical necessities required to support extensive quality control for any type of mass spectrometry-based workflow. We want to emphasize that this is an open community effort, and we seek participation from all scientists with an interest in this field.

As mass spectrometry proteomics and metabolomics have matured over the past few years, a growing emphasis has been placed on quality assurance (QA) and quality control (QC), which are of crucial importance to endorse the generated experimental results. Mass spectrometry is a highly complex technique, and because its results can be subject to significant variability[1], suitable quality control is necessary to model the influence of this variability on experimental results. Potential sources of variability can include the introduction of unexpected modifications during sample preparation[2], limited stability of proteolytic digestion[3], the presence of contaminants[4], variability in the chromatography[5] and mass measurements[6], etc. A systematic approach to quality control makes it possible to quantify the technical variability within experimental results, which can inform subsequent data analysis steps and can be fed back to optimize the mass spectrometry set-up. For example, Sle-

bos et al.[7] have combined a meticulous experimental design with advanced computational quality control procedures to detect deviating measurements in a high-profile proteogenomic cancer study. This allowed them to trace the variability in the experimental results to both batch effects from instrument drift and biological variability, which would not have been possible by only examining high-level identification performance. Quality control plays an increasingly important role in such large-scale multi-site projects[7–10] to enable an intra- and inter-laboratory comparison of experimental results. Additionally, although suitable quality control procedures are beneficial for any biological mass spectrometry application, a formal approach to quality control is of particular importance in a clinical setting[11,12].

As a result, computationally derived QC metrics have been defined to objectively assess the quality of mass spectrometry experiments[6]. These QC metrics capture quantitative information that may be related to the performance of the various processes in a mass spectrometry experiment. The importance of quality control is exemplified by the recent proliferation of tools that can compute such metrics[13,14]. However, most of these tools require specialized and non-standard experimental and software environments, which significantly hinders their evaluation and universal applicability. This problem is further exacerbated by the fact that each tool extracts different types of metrics from mass spectral data, and uses different frameworks to store, visualize, interpret, and communicate these metrics. In other words, the interoperability and comparability of these tools is essentially non-existent. These problems significantly hinder the systematic adoption of existing QC tools in well-established informatic pipelines. Consequently these tools are not yet adopted as a standard in the field, hindering biological mass spectrometry from reaching its full potential.

To address these issues a unifying framework for QC data is required, which would make it possible to bring these scattered initiatives together in a concerted approach, enabling a long-term strategy for quality control in proteomics. Moreover, we envision that in the future QC data will accompany mass spectrometry data and associated results in repositories such as those coordinated by the ProteomeXchange Consortium[15] and MetaboLights[16]. This will,

for instance, enable scientists interested in the reuse of these data to easily assess the quality of heterogeneous datasets in the increasingly extensive catalog of publicly available mass spectrometry data. Therefore, the Human Proteome Organization (HUPO) – Proteomics Standards Initiative (PSI)[17] and members of the Metabolomics Standards Initiative (MSI) Data Standards task group[18] have established a new working group on quality control at the HUPO-PSI meeting during April of 2016 in Ghent, Belgium. This working group consists of a wide range of stakeholders from the proteomics and metabolomics communities and is composed of academic, government, and industry researchers, software developers, journal representatives, and instrument manufacturers. Its main goal is to define a community standard format for QC data and associated controlled vocabulary (CV) terms, in order to facilitate the use of QC metrics more broadly in the proteomics and metabolomics communities and to enable the data exchange and archiving of mass spectrometry-derived QC metrics. It is important to emphasize that the working group does not seek to impose restrictive requirements on how quality control should be performed and how QC metrics should be interpreted. Instead, its aim is to provide the basic technical necessities to support extensive quality control practices over the whole course of a mass spectrometry experiment and to bolster a strong community-driven ecosystem of quality control tools and methodologies. We believe that involvement from experts in all mass spectrometry-related aspects, from analytical chemists to bioinformaticians, from both the proteomics and metabolomics communities, is essential to develop robust QC procedures and ensure their comprehensive adoption, and to this end we welcome contributions from all biological mass spectrometry researchers.

This aim fits into the overall objective of the HUPO-PSI to define community standards for proteomics data to facilitate data comparison, exchange, and verification[17]. At the same time a strong emphasis is placed on full interoperability with metabolomics approaches. The new working group will exclusively address applications related to quality control, with previously established HUPO-PSI working groups focusing on mass spectrometry, proteomics

informatics, molecular interactions, and protein separations. These working groups have previously established several standard data formats[19–23], controlled vocabularies[24,25], and minimum information guidelines[26,27], which have significantly contributed to the maturation and unification of proteomics research.

Methods

Mass spectrometry quality control

Myriad applications of mass spectrometry have been used in biology, each with specific properties and considerations. Therefore, no single strategy for performing quality control will be appropriate in all scenarios. Both external factors, such as sample preparation and environmental conditions, and instrumental factors, from autosampler to LC pump to column to mass spectrometry method, contribute to the overall performance and should be measured in an appropriate quality control regime. For example, in shotgun proteomics, the total number of identified tandem mass spectra is a high-level QC metric that is often used to assess the performance of an experiment. However, this is mostly useful for discovery experiments, where the aim is to identify as many proteins as possible. It would not be reasonable, however, to count identified spectra in a selected reaction monitoring (SRM) experiment, since tandem mass spectra are only sometimes collected in this process, and they are generally not used as an input to database search. This implies that different types of QC metrics are needed to suit a wide variety of experiment types.

QC metrics can vary in the time scale of assessment, spanning the retention time of a single mass spectrometry LC gradient or comparing among multiple experiments over the lifetime of an instrument. QC metrics may reveal information about different aspects of the experimental apparatus[6]. The metrics can be identification-free metrics that are computed from raw spectral data[28], which can be applied to some extent to both proteomics and metabolomics use cases. Alternatively, the metrics may depend upon application-specific

results, such as identification performance, to draw inferences (for example, the extent of oxidation or carbamylation observed in a sample). Additionally, metrics that are closely tied to data from a particular class of instrument may retrieve information directly from the control software, such as column temperature or back pressure[29].

Different types of QC samples of varying complexity can be used. In proteomics the QC samples can range from a simple peptide mixture or a single protein digest, such as BSA, to a complex whole-cell lysate, such as a yeast or HeLa cell lysate. Complementary information can be provided by spiking synthetic mixtures into the experimental samples, which enables monitoring specific peptides of interest to measure the dynamic range (as one example). In metabolomics the QC samples can similarly exhibit different levels of complexity: they can be composed of either pooled samples (combining a small aliquot of each biological sample), or of mixtures of different compounds or chemical standards[30]. Pooled QC samples characterize the entire collection of samples included in the study qualitatively and quantitatively by providing an average metabolome representation. As an alternative to pooled QC samples, predefined mixtures of certain biological fluids such as serum, plasma, or urine are commercially available (e.g. via National Institute of Standards and Technology (NIST)). Additionally, synthetic QC samples prepared under identical conditions and consisting of a mixture of compounds representing the different classes of metabolites expected to be present in the study samples can be used.

These different types of QC samples are not mutually exclusive; instead how they are used is closely linked to the experimental design[31], as they are each able to measure specific performance characteristics, and they should be used in combination. One consideration is how many QC samples of each type should be used; another is how to interleave them with experimental samples[13]. Further, besides these dedicated QC samples, other commonly used sample types to assess the quality of a run or an instrument in proteomics and/or metabolomics include: blanks, used to monitor or control instrument contamination; calibration curve samples, spanning a wide dynamic range and consisting of pure reference

compounds; and internal standards, usually consisting of synthetic peptides or of a single metabolic compound or a mixture of compounds that is added to all samples to monitor the reproducibility of the analytical methods. Another source of qualitative information comes from replicate measurements; based on the experimental design these replicates provide identical inputs for each injection and are typically used to compensate for variance across the analytical study[32].

A community-driven standard file format for QC data

As there is such a wide variety in the composition of mass spectrometry workflows and corresponding quality control methodologies, it is impossible to define a single fixed QC directive. Instead, the HUPO-PSI Quality Control working group wants to facilitate quality control for a wide variety of configurations by providing the basic technical foundation. To successfully communicate and interpret advanced quality control information a unified frame of reference is required. To this end a community-driven standardized format for the archival, transmission, analysis, and visualization of QC metrics derived from mass spectrometry will function as the focal point supporting various advanced tasks, as represented in figure 1. The definition of an unambiguous and expressive standard file format supported by powerful application program interfaces (APIs) and robust tools will allow bioinformaticians to focus on uncovering novel biological knowledge instead of being encumbered by low-level implementation details. This requisite technical infrastructure will be developed in the context of the HUPO-PSI Quality Control working group. Crucially, the file format constituting the centerpiece of the QC ecosystem will support metrics of an arbitrary type to accommodate QC information relevant for all kinds of experimental configurations. For this we will adopt the previously proposed qcML file format[33] as the starting point. Although this format is not an official PSI standard yet, it has been developed according to the same philosophy, and it has received considerable feedback at the 2016 HUPO-PSI meeting. Based on this feedback, an updated version of the qcML format will be developed, which will subsequently be sub-

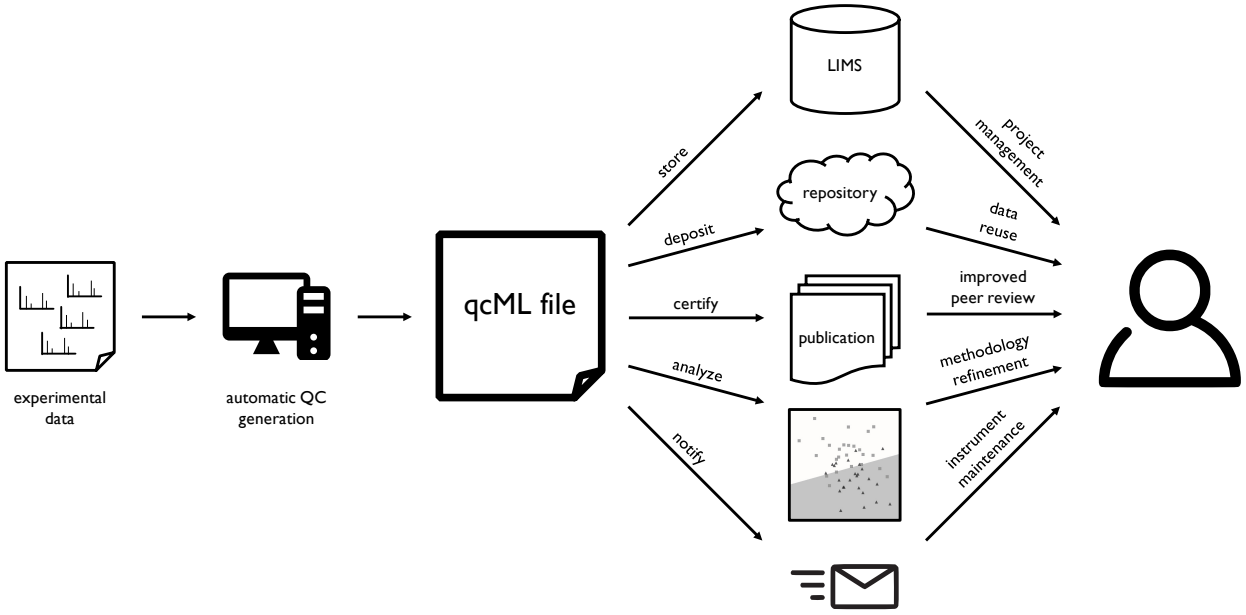

Figure 1: The qcML format is intended as the focal point for all QC applications. QC metrics are automatically generated as experiments are performed, whereupon the qcML data can be stored locally in a laboratory information management system (LIMS) where it can be used to support project management and data provenance. Additionally, the qcML data can be submitted to a public data repository, empowering data reuse, and it can facilitate peer review by certifying the data quality upon publication. Advanced analyses can aid informed decision-making to drive methodology refinements and automatically notify the instrument operator when a malfunction is detected.

mitted to the PSI formal document process[34] to establish it as a PSI standard format. A key part of this work will connect the qcML format and a CV in accordance with the previously established PSI CV[24] to enable direct interpretability of the collected metrics and addition of further metrics without the need to update the standard format to a new version. When metrics are defined in CV terms, they are more comprehensible, even as QC pipelines gain complexity and cover thousands of mass spectrometry acquisitions. This will hopefully also elevate the ease of informed decision-making during analysis processes[28,35,36] and contribute to the prevalence of applied QC in biological mass spectrometry. An important aspect that ties in with this semantic interpretability is the development of a Minimum Information About a Proteomics Experiment (MIAPE)-like document for quality control, as suggested earlier[37]. The information specified in the MIAPE-QC document will present opportunities for extensive linking of the QC data to the experimental results. For metabolomics, some earlier work on capturing the minimum information for reporting on quality control already exists[38,39], but this will need to be revisited in coordination with the Metabolomics Society Data Quality and Data Standards task groups[18,40].

To make the qcML format more attractive to the authors of quality metric generators, we will create a software library that is able to import and export information in the qcML format, which will enable developers to easily create new qcML files and extract information from existing qcML files. This will mainly be mediated in the form of the jqcML Java API[41]. Although at present the jqcML API only structurally validates qcML files against the schema definition, additional functionality to semantically validate the qcML files based on the terms defined in the CVs and the MIAPE-QC specification will be added[27]. Similar to APIs for other standard formats[19–23], the availability of the jqcML API will assist developers to support the qcML format, fostering the interoperability of QC tools. For example, this will enable the construction of a custom tool workflow where a first tool generates various QC metrics, a second tool applies advanced algorithms to draw inferences from the data, and a third tool provides long-term storage and visualization. Instead of a rigid, monolithic

framework, the qcML format and the jqcML API will support the construction of modular, highly customizable QC pipelines. Furthermore, besides updates to existing support for the qcML format in OpenMS[42] and SimpatiQCo[43], native support will be added to other tools developed by members of the working group as well, such as QuaMeter[44] and iMonDB[45]. We will develop a user-friendly graphical user interface (GUI) tool for visualization of quality control data in the qcML format, including established outlier detection techniques to automatically identify low-quality experiments[28,36]. This tool will enable visual data exploration and easy downstream processing of QC data. These steps will foster broader interest in and adoption for the qcML format, both from developers and end users.

Broadening the applicability of quality control

Although shotgun LC-MS/MS proteomics and metabolomics can assuredly benefit from wider adoption and automation of quality control, other classes of data in proteomics and metabolomics can also benefit from these tools. The working group particularly emphasizes the use of quality control in quantitative proteomics methods, from iTRAQ to data-independent acquisition (DIA) datasets, where details of the experimental procedures can be employed to compute further relevant and focused QC metrics. For example, for iTRAQ experiments the isobaric tagging reagents can be a source of variability, influenced by the protein abundances, and fold changes are biased towards $1:1$ ratios[46]. The labeling efficiency can broadly be determined by verifying the fraction of MS/MS spectra for which reporter ions are observed. Additionally, as iTRAQ reagents bond to primary amines both at the N-terminus and on lysine residues the labeling efficiency can be determined in full detail by evaluating the extent to which the labels are present on one or both of these sites. Furthermore, the labeling stability can be evaluated based on the evolution of the reporter ion intensity over the course of an experiment and their signal to noise ratios.

In contrast to during a data-dependent acquisition (DDA) experiment, during a DIA experiment MS/MS scans are measured with wide isolation windows that do not target any

particular peptide precursor[47]. This way all analytes within the desired precursor mass range can be measured in an unbiased fashion, potentially leading to an increase in reproducibility. To evaluate the performance of a DIA experiment general QC metrics from the DDA setting can likewise be applied. In both cases the consistency of the LC and MS performance can be evaluated using well-characterized standard QC samples[48], which can for example be visualized using the powerful Skyline software tool[49]. Furthermore, specialized QC metrics can be defined based on the characteristics of a DIA experiment. For example, the isolation window size can be evaluated based on the rate of ion interference[50]. This is sample-dependent, as complex samples will lead to a lower precursor selectivity resulting in highly complex chimeric MS/MS spectra. Because all analytes are reproducibly measured during a DIA experiment consecutive MS/MS scans can be compared to each other to further evaluate the LC and MS performance. Measurements of the same analyte over repeat scans can be used to assess the mass accuracy, while successive scans covering the same isolation window (separated by the duty cycle) can provide information on the chromatographic sample rate. An important step during DIA spectrum identification is the correlation of a precursor ion with its corresponding product ions. Although a high rate of cofragmentation complicates an accurate precursor–product correlation this is essential for obtaining correct peptide identifications. Some QC metrics that can be used to evaluate whether precursor and product ions are accurately correlated are for example the fraction of MS isotopic packets that match product ions[51], the retention time (RT) variability of associated ions[52], and whether or not the elution profiles of corresponding precursor and product ions match a similar exponentially modified Gaussian peak shape[53].

MALDI imaging mass spectrometry is an increasingly popular technique for molecular imaging, yet a standardized approach to quality control has not been established thus far. Currently, quality assessments are typically done manually through visual inspection of the spectra. Additionally, simple plots can be employed to evaluate the variation in peak intensities among different measurements[54]. This can, for example, be complemented by an ion

intensity histogram to highlight specific regions of poor signal. However, these QC methods for MALDI imaging usually disregard the spatial information present in the data. Conversely, important qualitative measures can be defined by comparing nearby measurements as, for example, spatial proximity has an influence on the intensity similarity. A recent result has shown how image analysis measures based on sliding windows of increasing sizes, which consider the spatial information implicitly, can be used to accurately replicate manual expert quality assessments[55].

These are but a few examples of how quality control can be expanded to further mass spectrometry technologies. To adequately cover all these different workflows the novel qcML standard will have to be flexible enough to support mass spectrometry data ranging from shotgun proteomics to SRM to MALDI imaging data. This effort will require a broader perspective than has dominated QC software to date. The HUPO-PSI QC working group has members from both the proteomics and the metabolomics communities and welcomes any contributions, ensuring wide applicability of the qcML open data standard in a variety of mass spectrometry-based settings.

Conclusions

Quality control will indubitably play a growing role in aiding the maturation of biological mass spectrometry as a field. A systematic approach to quality control will aid researchers to assess their workflows over time or to compare data among different laboratories. Furthermore, it can stimulate public data reuse to harvest new knowledge[56]. We envision that the inclusion of QC metrics will become an integral component when submitting datasets to public data repositories or for peer-reviewed publication, similar to the situation for three-dimensional structures in the Worldwide Protein Data Bank (wwPDB)[57].

The HUPO-PSI seeks to broaden the conversation surrounding quality control in our community. This effort will provide a format definition along with examples and software

infrastructure that will enable new research in the interpretation of QC metrics. We emphasize that this is very much a community effort, and any and all contributions are welcome. Our group charter is available on the HUPO-PSI website (`http://psidev.info/groups/quality-control`), including a summary of the milestones the working group wants to achieve. You can connect with us through our GitHub repository (`https://github.com/HUPO-PSI/qcML-development`) and through our mailing list (`Psidev-qc-dev@lists.sourceforge.net`). These online sources contain further detailed instructions on how to get started and how to contribute. We cordially invite anyone to join the discussion and participate in this important community effort.

Acknowledgement

WB acknowledges the support of the VLAIO grant "InSPECtor" (IWT project 120025). ST acknowledges support by the BMBF (Express2Present, 0316179C). DLT is funded by a South African MRC grant to Gerhard Walzl for the South African Tuberculosis Bioinformatics Initiative.

References

(1) Tabb, D. L.; Vega-Montoto, L.; Rudnick, P. A.; Variyath, A. M.; Ham, A.-J. L.; Bunk, D. M.; Kilpatrick, L. E.; Billheimer, D. D.; Blackman, R. K.; Cardasis, H. L.; Carr, S. A.; Clauser, K. R.; Jaffe, J. D.; Kowalski, K. A.; Neubert, T. A.; Regnier, F. E.; Schilling, B.; Tegeler, T. J.; Wang, M.; Wang, P.; Whiteaker, J. R.; Zimmerman, L. J.; Fisher, S. J.; Gibson, B. W.; Kinsinger, C. R.; Mesri, M.; Rodriguez, H.; Stein, S. E.; Tempst, P.; Paulovich, A. G.; Liebler, D. C.; Spiegelman, C. *J. Proteome Res.* **2010**, *9*, 761–776.

(2) Karty, J. A.; Ireland, M. M. E.; Brun, Y. V.; Reilly, J. P. *J. Chromatogr. B* **2002**, *782*, 363–383.

(3) Vandermarliere, E.; Mueller, M.; Martens, L. *Mass Spectrom. Rev.* **2013**, *32*, 453–465.

(4) Keller, B. O.; Sui, J.; Young, A. B.; Whittal, R. M. *Anal. Chim. Acta* **2008**, *627*, 71–81.

(5) Barwick, V. J. *J. Chromatogr. A* **1999**, *849*, 13–33.

(6) Rudnick, P. A.; Clauser, K. R.; Kilpatrick, L. E.; Tchekhovskoi, D. V.; Neta, P.; Blonder, N.; Billheimer, D. D.; Blackman, R. K.; Bunk, D. M.; Cardasis, H. L.; Ham, A.-J. L.; Jaffe, J. D.; Kinsinger, C. R.; Mesri, M.; Neubert, T. A.; Schilling, B.; Tabb, D. L.; Tegeler, T. J.; Vega-Montoto, L.; Variyath, A. M.; Wang, M.; Wang, P.; Whiteaker, J. R.; Zimmerman, L. J.; Carr, S. A.; Fisher, S. J.; Gibson, B. W.; Paulovich, A. G.; Regnier, F. E.; Rodriguez, H.; Spiegelman, C.; Tempst, P.; Liebler, D. C.; Stein, S. E. *Mol. Cell. Proteomics* **2010**, *9*, 225–241.

(7) Slebos, R. J. C.; Wang, X.; Wang, X.; Zhang, B.; Tabb, D. L.; Liebler, D. C. *Sci. Data* **2015**, *2*, 150022.

(8) Zhang, B.; Wang, J.; Wang, X.; Zhu, J.; Liu, Q.; Shi, Z.; Chambers, M. C.; Zimmerman, L. J.; Shaddox, K. F.; Kim, S.; Davies, S. R.; Wang, S.; Wang, P.; Kinsinger, C. R.; Rivers, R. C.; Rodriguez, H.; Townsend, R. R.; Ellis, M. J. C.; Carr, S. A.; Tabb, D. L.; Coffey, R. J.; Slebos, R. J. C.; Liebler, D. C.; Carr, S. A.; Gillette, M. A.; Klauser, K. R.; Kuhn, E.; Mani, D. R.; Mertins, P.; Ketchum, K. A.; Paulovich, A. G.; Whiteaker, J. R.; Edwards, N. J.; McGarvey, P. B.; Madhavan, S.; Wang, P.; Chan, D.; Pandey, A.; Shih, I.-M.; Zhang, H.; Zhang, Z.; Zhu, H.; Whiteley, G. A.; Skates, S. J.; White, F. M.; Levine, D. A.; Boja, E. S.; Kinsinger, C. R.; Hiltke, T.; Mesri, M.; Rivers, R. C.; Rodriguez, H.; Shaw, K. M.; Stein, S. E.; Fenyo, D.; Liu, T.; McDermott, J. E.; Payne, S. H.; Rodland, K. D.; Smith, R. D.; Rudnick, P.; Snyder, M.; Zhao, Y.; Chen, X.; Ransohoff, D. F.; Hoofnagle, A. N.; Liebler, D. C.; Sanders, M. E.; Shi, Z.;

Slebos, R. J. C.; Tabb, D. L.; Zhang, B.; Zimmerman, L. J.; Wang, Y.; Davies, S. R.; Ding, L.; Ellis, M. J. C.; Reid Townsend, R. *Nature* **2014**, *513*, 382–387.

(9) Campos, A.; Díaz, R.; Martínez-Bartolomé, S.; Sierra, J.; Gallardo, O.; Sabidó, E.; López-Lucendo, M.; Ignacio Casal, J.; Pasquarello, C.; Scherl, A.; Chiva, C.; Borras, E.; Odena, A.; Elortza, F.; Azkargorta, M.; Ibarrola, N.; Canals, F.; Albar, J. P.; Oliveira, E. *J. Proteomics* **2015**, *127, Part B*, 264–274.

(10) Tabb, D. L.; Wang, X.; Carr, S. A.; Clauser, K. R.; Mertins, P.; Chambers, M. C.; Holman, J. D.; Wang, J.; Zhang, B.; Zimmerman, L. J.; Chen, X.; Gunawardena, H. P.; Davies, S. R.; Ellis, M. J. C.; Li, S.; Townsend, R. R.; Boja, E. S.; Ketchum, K. A.; Kinsinger, C. R.; Mesri, M.; Rodriguez, H.; Liu, T.; Kim, S.; McDermott, J. E.; Payne, S. H.; Petyuk, V. A.; Rodland, K. D.; Smith, R. D.; Yang, F.; Chan, D. W.; Zhang, B.; Zhang, H.; Zhang, Z.; Zhou, J.-Y.; Liebler, D. C. *J. Proteome Res.* **2016**, *15*, 691–706.

(11) Tabb, D. L. *Clin. Biochem.* **2013**, *46*, 411–420.

(12) Martens, L. *PROTEOMICS - Clin. Appl.* **2013**, *7*, 388–391.

(13) Bereman, M. S. *PROTEOMICS* **2015**, *15*, 891–902.

(14) Bittremieux, W.; Valkenborg, D.; Martens, L.; Laukens, K. *PROTEOMICS*

(15) Vizcaíno, J. A.; Deutsch, E. W.; Wang, R.; Csordas, A.; Reisinger, F.; Ríos, D.; Dianes, J. A.; Sun, Z.; Farrah, T.; Bandeira, N.; Binz, P.-A.; Xenarios, I.; Eisenacher, M.; Mayer, G.; Gatto, L.; Campos, A.; Chalkley, R. J.; Kraus, H.-J.; Albar, J. P.; Martinez-Bartolomé, S.; Apweiler, R.; Omenn, G. S.; Martens, L.; Jones, A. R.; Hermjakob, H. *Nat. Biotechnol.* **2014**, *32*, 223–226.

(16) Haug, K.; Salek, R. M.; Conesa, P.; Hastings, J.; de Matos, P.; Rijnbeek, M.; Mahendraker, T.; Williams, M.; Neumann, S.; Rocca-Serra, P.; Maguire, E.; González-

Beltrán, A.; Sansone, S.-A.; Griffin, J. L.; Steinbeck, C. *Nucleic Acids Res.* **2013**, *41*, D781–D786.

(17) Deutsch, E. W.; Albar, J. P.; Binz, P.-A.; Eisenacher, M.; Jones, A. R.; Mayer, G.; Omenn, G. S.; Orchard, S.; Vizcaíno, J. A.; Hermjakob, H. *J. Am. Med. Inform. Assoc.* **2015**, *22*, 495–506.

(18) Salek, R. M.; Arita, M.; Dayalan, S.; Ebbels, T.; Jones, A. R.; Neumann, S.; Rocca-Serra, P.; Viant, M. R.; Vizcaíno, J.-A. *Metabolomics* **2015**, *11*, 782–783.

(19) Côté, R. G.; Reisinger, F.; Martens, L. *PROTEOMICS* **2010**, *10*, 1332–1335.

(20) Helsens, K.; Brusniak, M.-Y.; Deutsch, E.; Moritz, R. L.; Martens, L. *J. Proteome Res.* **2011**, *10*, 5260–5263.

(21) Reisinger, F.; Krishna, R.; Ghali, F.; Ríos, D.; Hermjakob, H.; Vizcaíno, J. A.; Jones, A. R. *PROTEOMICS* **2012**, *12*, 790–794.

(22) Qi, D.; Krishna, R.; Jones, A. R. *PROTEOMICS* **2014**, *14*, 685–688.

(23) Xu, Q.-W.; Griss, J.; Wang, R.; Jones, A. R.; Hermjakob, H.; Vizcaíno, J. A. *PROTEOMICS* **2014**, *14*, 1328–1332.

(24) Mayer, G.; Montecchi-Palazzi, L.; Ovelleiro, D.; Jones, A. R.; Binz, P.-A.; Deutsch, E. W.; Chambers, M.; Kallhardt, M.; Levander, F.; Shofstahl, J.; Orchard, S.; Vizcaíno, J. A.; Hermjakob, H.; Stephan, C.; Meyer, H. E.; Eisenacher, M. *Database* **2013**, *2013*, bat009–bat009.

(25) Mayer, G.; Jones, A. R.; Binz, P.-A.; Deutsch, E. W.; Orchard, S.; Montecchi-Palazzi, L.; Vizcaíno, J. A.; Hermjakob, H.; Oveillero, D.; Julian, R.; Stephan, C.; Meyer, H. E.; Eisenacher, M. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2014**, *1844*, 98–107.

(26) Taylor, C. F.; Paton, N. W.; Lilley, K. S.; Binz, P.-A.; Julian, R. K.; Jones, A. R.; Zhu, W.; Apweiler, R.; Aebersold, R.; Deutsch, E. W.; Dunn, M. J.; Heck, A. J. R.; Leitner, A.; Macht, M.; Mann, M.; Martens, L.; Neubert, T. A.; Patterson, S. D.; Ping, P.; Seymour, S. L.; Souda, P.; Tsugita, A.; Vandekerckhove, J.; Vondriska, T. M.; Whitelegge, J. P.; Wilkins, M. R.; Xenarios, I.; Yates, J. R.; Hermjakob, H. *Nat. Biotechnol.* **2007**, *25*, 887–893.

(27) Montecchi-Palazzi, L.; Kerrien, S.; Reisinger, F.; Aranda, B.; Jones, A. R.; Martens, L.; Hermjakob, H. *PROTEOMICS* **2009**, *9*, 5112–5119.

(28) Wang, X.; Chambers, M. C.; Vega-Montoto, L. J.; Bunk, D. M.; Stein, S. E.; Tabb, D. L. *Anal. Chem.* **2014**, *86*, 2497–2509.

(29) Scheltema, R. A.; Mann, M. *J. Proteome Res.* **2012**, *11*, 3458–3466.

(30) Godzien, J.; Alonso-Herranz, V.; Barbas, C.; Armitage, E. G. *Metabolomics* **2015**, *11*, 518–528.

(31) Maes, E.; Kelchtermans, P.; Bittremieux, W.; De Grave, K.; Degroeve, S.; Hooyberghs, J.; Mertens, I.; Baggerman, G.; Ramon, J.; Laukens, K.; Martens, L.; Valkenborg, D. *Expert Rev. Proteomics* **2016**, *13*, 495–511.

(32) Dunn, W. B.; Wilson, I. D.; Nicholls, A. W.; Broadhurst, D. *Bioanalysis* **2012**, *4*, 2249–2264.

(33) Walzer, M.; Pernas, L. E.; Nasso, S.; Bittremieux, W.; Nahnsen, S.; Kelchtermans, P.; Pichler, P.; van den Toorn, H. W. P.; Staes, A.; Vandenbussche, J.; Mazanek, M.; Taus, T.; Scheltema, R. A.; Kelstrup, C. D.; Gatto, L.; van Breukelen, B.; Aiche, S.; Valkenborg, D.; Laukens, K.; Lilley, K. S.; Olsen, J. V.; Heck, A. J. R.; Mechtler, K.; Aebersold, R.; Gevaert, K.; Vizcaino, J. A.; Hermjakob, H.; Kohlbacher, O.; Martens, L. *Mol. Cell. Proteomics* **2014**, *13*, 1905–1913.

(34) Vizcaíno, J. A.; Martens, L.; Hermjakob, H.; Julian, R. K.; Paton, N. W. *PROTEOMICS* **2007**, *7*, 2355–2357.

(35) Amidan, B. G.; Orton, D. J.; LaMarche, B. L.; Monroe, M. E.; Moore, R. J.; Venzin, A. M.; Smith, R. D.; Sego, L. H.; Tardiff, M. F.; Payne, S. H. *J. Proteome Res.* **2014**, *13*, 2215–2222.

(36) Bittremieux, W.; Meysman, P.; Martens, L.; Valkenborg, D.; Laukens, K. *J. Proteome Res.* **2016**, *15*, 1300–1307.

(37) Eisenacher, M.; Schnabel, A.; Stephan, C. *PROTEOMICS* **2011**, *11*, 1031–1036.

(38) Fiehn, O.; Wohlgemuth, G.; Scholz, M.; Kind, T.; Lee, D. Y.; Lu, Y.; Moon, S.; Nikolau, B. *Plant J.* **2008**, *53*, 691–704.

(39) Sansone, S.-A.; Schober, D.; Atherton, H. J.; Fiehn, O.; Jenkins, H.; Rocca-Serra, P.; Rubtsov, D. V.; Spasic, I.; Soldatova, L.; Taylor, C.; Tseng, A.; Viant, M. R.; Ontology Working Group Members, *Metabolomics* **2007**, *3*, 249–256.

(40) Bearden, D. W.; Beger, R. D.; Broadhurst, D.; Dunn, W.; Edison, A.; Guillou, C.; Trengove, R.; Viant, M.; Wilson, I. *Metabolomics* **2014**, *10*, 0.

(41) Bittremieux, W.; Kelchtermans, P.; Valkenborg, D.; Martens, L.; Laukens, K. *J. Proteome Res.* **2014**, *13*, 3484–3487.

(42) Röst, H. L.; Sachsenberg, T.; Aiche, S.; Bielow, C.; Weisser, H.; Aicheler, F.; Andreotti, S.; Ehrlich, H.-C.; Gutenbrunner, P.; Kenar, E.; Liang, X.; Nahnsen, S.; Nilse, L.; Pfeuffer, J.; Rosenberger, G.; Rurik, M.; Schmitt, U.; Veit, J.; Walzer, M.; Wojnar, D.; Wolski, W. E.; Schilling, O.; Choudhary, J. S.; Malmström, L.; Aebersold, R.; Reinert, K.; Kohlbacher, O. *Nat. Methods* **2016**, *13*, 741–748.

(43) Pichler, P.; Mazanek, M.; Dusberger, F.; Weilnböck, L.; Huber, C. G.; Stingl, C.;

Luider, T. M.; Straube, W. L.; Köcher, T.; Mechtler, K. *J. Proteome Res.* **2012**, *11*, 5540–5547.

(44) Ma, Z.-Q.; Polzin, K. O.; Dasari, S.; Chambers, M. C.; Schilling, B.; Gibson, B. W.; Tran, B. Q.; Vega-Montoto, L.; Liebler, D. C.; Tabb, D. L. *Anal. Chem.* **2012**, *84*, 5845–5850.

(45) Bittremieux, W.; Willems, H.; Kelchtermans, P.; Martens, L.; Laukens, K.; Valkenborg, D. *J. Proteome Res.* **2015**, *14*, 2360–2366.

(46) Mahoney, D. W.; Therneau, T. M.; Heppelmann, C. J.; Higgins, L.; Benson, L. M.; Zenka, R. M.; Jagtap, P.; Nelsestuen, G. L.; Bergen, H. R.; Oberg, A. L. *J. Proteome Res.* **2011**, *10*, 4325–4333.

(47) Gillet, L. C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. *Mol. Cell. Proteomics* **2012**, *11*, O111.016717–O111.016717.

(48) Escher, C.; Reiter, L.; MacLean, B.; Ossola, R.; Herzog, F.; Chilton, J.; MacCoss, M. J.; Rinner, O. *PROTEOMICS* **2012**, *12*, 1111–1121.

(49) Egertson, J. D.; MacLean, B.; Johnson, R.; Xuan, Y.; MacCoss, M. J. *Nat. Protoc.* **2015**, *10*, 887–903.

(50) Zhang, Y.; Bilbao, A.; Bruderer, T.; Luban, J.; Strambio-De-Castillia, C.; Lisacek, F.; Hopfgartner, G.; Varesio, E. *J. Proteome Res.* **2015**, *14*, 4359–4371.

(51) Geromanos, S. J.; Vissers, J. P. C.; Silva, J. C.; Dorschel, C. A.; Li, G.-Z.; Gorenstein, M. V.; Bateman, R. H.; Langridge, J. I. *PROTEOMICS* **2009**, *9*, 1683–1695.

(52) Broeckling, C. D.; Heuberger, A. L.; Prince, J. A.; Ingelsson, E.; Prenni, J. E. *Metabolomics* **2013**, *9*, 33–43.

(53) Kalambet, Y.; Kozmin, Y.; Mikhailova, K.; Nagaev, I.; Tikhonov, P. *J. Chemom.* **2011**, *25*, 352–356.

(54) Karlsson, O.; Bergquist, J.; Andersson, M. *Mol. Cell. Proteomics* **2014**, *13*, 93–104.

(55) Palmer, A.; Ovchinnikova, E.; Thuné, M.; Lavigne, R.; Guével, B.; Dyatlov, A.; Vitek, O.; Pineau, C.; Borén, M.; Alexandrov, T. *Bioinformatics* **2015**, *31*, i375–i384.

(56) Martens, L. *EuPA Open Proteomics* **2016**, *11*, 42–44.

(57) Berman, H.; Henrick, K.; Nakamura, H. *Nat. Struct. Biol.* **2003**, *10*, 980–980.

Graphical TOC Entry