quantification. Our open source software is available as standalone executable at http://www.openswath.org (**Supplementary Source Code File 1** and **Supplementary Data 4–6**). The OpenSWATH algorithms are provided as a C++ software library, allowing integration of our algorithms into a multitude of popular proteomics software, such as OpenMS[37] or Skyline[23]. The software is integrated and distributed together with OpenMS[37], which will make targeted DIA data analysis immediately accessible to a large research community. Owing to the nature of DIA data, which contain a complete record of all fragment ions of a biological sample, reanalysis of a data set is possible completely *in silico*, allowing researchers to re-query data with their specific hypothesis in mind. The availability of fast DIA-capable instruments, assay libraries (available in proteome-wide coverage owing to large-scale peptide synthesis efforts) and, now, an automated software for DIA targeted data analysis should facilitate the widespread use of this technology.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper (doi:10.1038/nbt.2841).

ACKNOWLEDGMENTS

H.L.R. was funded by ETH (ETH-30 11-2), G.R. and S.M.M. were funded by the Swiss Federal Commission for Technology and Innovation CTI (13539.1 PFFLI-LS), P.N., L.G. and R.A. were funded by the advanced European Research Council grant Proteomics v3.0 (233226), L.G. and R.A. were funded by PhosphonetX project of SystemsX.ch, J.M. was funded by the Swedish Research Council (project 2008-3356), the Crafoord Foundation (20100892) and the Swedish Foundation for Strategic Research (FFL4). Further funding was provided to R.A. by the Swiss National Science Foundation. We would like to thank the SyBIT project of SystemsX.ch for support and maintenance of the lab-internal computing infrastructure, the ITS HPC team (Brutus) and the OpenMS developers for including OpenSWATH in the OpenMS framework and fixing MS Windows compatibility bugs.

AUTHOR CONTRIBUTIONS

H.L.R. and G.R. designed, implemented and executed the C++ code and the analysis workflow. H.L.R. acquired and analyzed the *S. pyogenes* data. H.L.R., G.R., L.M. and R.A. wrote the manuscript. G.R. and L.G. provided the SGS sample. P.N., L.G. and B.C.C. provided critical input on the project. H.L.R., G.R. and P.N. analyzed the SGS data set manually. L.G., S.M.M. and O.T.S. performed all the measurements and provided important feedback. W.W. did code review and assisted in software design. B.C.C. performed testing of the software. J.M. performed the biological experiments. L.M. and R.A. designed and supervised the study.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper (doi:10.1038/nbt.2841).

Hannes L Röst[1,2,8], George Rosenberger[1,2,8], Pedro Navarro[1], Ludovic Gillet[1], Saša M Miladinović[1,3], Olga T Schubert[1,2], Witold Wolski[4], Ben C Collins[1], Johan Malmström[5], Lars Malmström[1] & Ruedi Aebersold[1,6,7]

[1]Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Zurich, Switzerland. [2]PhD Program in Systems Biology, University of Zurich and ETH Zurich, Zurich, Switzerland. [3]Biognosys AG, Schlieren, Switzerland. [4]SyBIT project of SystemsX.ch, ETH Zurich, Zurich, Switzerland. [5]Department of Immunotechnology, Lund University, Lund, Sweden. [6]Competence Center for Systems Physiology and Metabolic Diseases, Zurich, Switzerland. [7]Faculty of Science, University of Zurich, Zurich, Switzerland. [8]These authors contributed equally to this work.
e-mail: aebersold@imsb.biol.ethz.ch

1. Aebersold, R. & Mann, M. *Nature* **422**, 198–207 (2003).
2. Domon, B. & Aebersold, R. *Nat. Biotechnol.* **28**, 710–721 (2010).
3. Purvine, S., Eppel, J.-T.T., Yi, E.C. & Goodlett, D.R. *Proteomics* **3**, 847–850 (2003).
4. Venable, J.D., Dong, M.-Q., Wohlschlegel, J., Dillin, A. & Yates, J.R. *Nat. Methods* **1**, 39–45 (2004).
5. Plumb, R.S. *et al. Rapid Commun. Mass Spectrom.* **20**, 1989–1994 (2006).
6. Panchaud, A. *et al. Anal. Chem.* **81**, 6481–6488 (2009).
7. Panchaud, A., Jung, S., Shaffer, S.A., Aitchison, J.D. & Goodlett, D.R. *Anal. Chem.* **83**, 2250–2257 (2011).
8. Bern, M. *et al. Anal. Chem.* **82**, 833–841 (2010).
9. Wong, J., Schwahn, A. & Downard, K. *BMC Bioinformatics* **10**, 244 (2009).
10. Carvalho, P.C. *et al. Bioinformatics* **26**, 847–848 (2010).
11. Geromanos, S.J. *et al. Proteomics* **9**, 1683–1695 (2009).
12. Li, G.-Z. *et al. Proteomics* **9**, 1696–1719 (2009).
13. Blackburn, K., Mbeunkui, F., Mitra, S.K., Mentzel, T. & Goshe, M.B. *J. Proteome Res.* **9**, 3621–3637 (2010).
14. Huang, X. *et al. Anal. Chem.* **83**, 6971–6979 (2011).
15. Geiger, T., Cox, J. & Mann, M. *Mol. Cell. Proteomics* **9**, 2252–2261 (2010).
16. Gillet, L.C. *et al. Mol. Cell. Proteomics* **11**, O111.016717 (2012).
17. Lange, V., Picotti, P., Domon, B. & Aebersold, R. *Mol. Syst. Biol.* **4,** 222 (2008).
18. Domon, B. & Aebersold, R. *Science* **312**, 212–217 (2006).
19. Sherman, J., McKay, M.J., Ashman, K. & Molloy, M.P. *Mol. Cell. Proteomics* **8**, 2051–2062 (2009).
20. Röst, H., Malmström, L. & Aebersold, R. *Mol. Cell. Proteomics* **11**, 540–549 (2012).
21. Michalski, A., Cox, J. & Mann, M. *J. Proteome Res.* **10**, 1785–1793 (2011).
22. Picotti, P., Bodenmiller, B., Mueller, L.N., Domon, B. & Aebersold, R. *Cell* **138**, 795–806 (2009).
23. MacLean, B. *et al. Bioinformatics* **26**, 966–968 (2010).
24. Ince, D.C., Hatton, L. & Graham-Cumming, J. *Nature* **482**, 485–488 (2012).
25. Martens, L. *et al. Mol. Cell. Proteomics* **10**, R110.000133 (2010).
26. Deutsch, E.W. *Mol. Cell. Proteomics* **11**, 1612–1621 (2012).
27. Escher, C. *et al. Proteomics* **12**, 1111–1121 (2012).
28. Reiter, L. *et al. Nat. Methods* **8**, 430–435 (2011).
29. Malmström, L., Malmström, J., Selevsek, N., Rosenberger, G. & Aebersold, R. *J. Proteome Res.* **11**, 1644–1653 (2012).
30. Wenschuh, H. *et al. Biopolymers* **55**, 188–206 (2000).
31. Hilpert, K., Winkler, D.F. & Hancock, R.E. *Nat. Protoc.* **2**, 1333–1349 (2007).
32. Elias, J.E. & Gygi, S.P. *Nat. Methods* **4**, 207–214 (2007).
33. Malmström, J. *et al. J. Biol. Chem.* **287**, 1415–1425 (2012).
34. Deutsch, E.W., Lam, H. & Aebersold, R. *EMBO Rep.* **9**, 429–434 (2008).
35. Shea, P.R. *et al. Proc. Natl. Acad. Sci. USA* **108**, 5039–5044 (2011).
36. Malke, H., Steiner, K., McShan, W.M. & Ferretti, J.J. *Int. J. Med. Microbiol.* **296**, 259–275 (2006).
37. Sturm, M. *et al. BMC Bioinformatics* **9**, 163 (2008).

ProteomeXchange provides globally coordinated proteomics data submission and dissemination

To the Editor:

There is a growing trend toward public dissemination of proteomics data, which is facilitating the assessment, reuse, comparative analyses and extraction of new findings from published data[1,2]. This process has been mainly driven by journal publication guidelines and funding agencies. However, there is a need for better integration of public repositories and coordinated sharing of all the pieces of information needed to represent a full mass spectrometry (MS)–based proteomics experiment. An editorial in your journal in 2009, 'Credit where credit is overdue'[3], exposed the situation in the proteomics field, where full data disclosure is still not common practice. Olsen and Mann[4] identified different levels of information in the typical experiment: from raw data and going through peptide identification and quantification, protein identifications and protein ratios and the resulting biological conclusions. All of these levels should be captured and properly annotated in public databases, using the existing MS proteomics repositories for the MS data (raw data, identification and quantification results) and metadata, whereas the resulting biological information should be integrated in protein knowledge bases, such as UniProt[5]. A recent editorial[6] in *Nature Methods* again highlighted the need for a

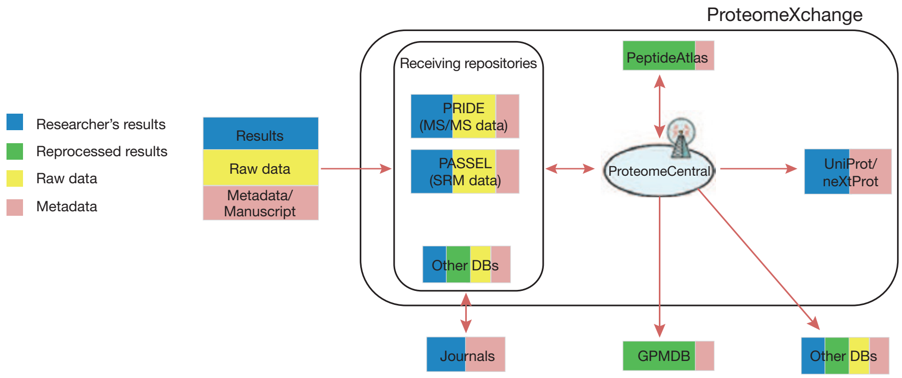

Figure 1 Representation of the ProteomeXchange workflow for MS/MS and SRM data. Raw data represents mass spectrometer output files. ProteomeCentral is the portal for all public ProteomeXchange datasets.

stable repository for raw MS proteomics data. In this Correspondence, we report the first implementation of the ProteomeXchange consortium, an integrated framework for submission and dissemination of MS-based proteomics data.

Among the existing MS proteomics repositories with a broad target audience, the PRIDE (PRoteomics IDEntifications) database[7] (European Bioinformatics Institute, EBI, Cambridge, UK; http://www.ebi.ac.uk/pride) and PeptideAtlas[8] (Institute for Systems Biology, ISB, Seattle, USA; http://www.peptideatlas.org) are two of the most prominent. Both are mainly focused on tandem MS (MS/MS) data storage. Whereas PRIDE represents the information as originally analyzed by the researcher (thus constituting a primary resource), data in PeptideAtlas are reprocessed through a common pipeline (the Trans-Proteomic Pipeline) to provide a uniformly analyzed view of the data with a focus on low protein false discovery rates (constituting a secondary resource). In addition, ISB has set up the first repository for selected reaction monitoring (SRM) data, PASSEL[9] (PeptideAtlas SRM Experiment Library, http://www.peptideatlas.org/passel/). There are other resources dedicated to storing MS proteomics data, each of them with different focuses and functionalities, for instance the Global Proteome Machine Database (GPMDB; where data are reprocessed using the search engine X!Tandem)[10]. At a higher abstraction level, resources such as UniProt and neXtProt are integrating proteomics results into a wider context of functional annotation from many different sources, including antibody-based methods.

Although most of the proteomics resources mentioned have existed for a long time, they have acted independently with limited coordination of their activities. As a result, data providers were unclear to which repository they should submit their data set, and in what form, with choices ranging from full raw data to highly processed identifications and quantifications. In addition, no repository could store both raw data and processed results. Similar issues arose for data consumers, who could not always find the data supporting a protein modification in UniProt, or know whether a particular data set from PRIDE had been integrated into PeptideAtlas.

The ProteomeXchange consortium (http://www.proteomexchange.org) was formed in 2006 (ref. 11) to overcome these challenges, developing from a loose collaboration into an international consortium of major stakeholders in the domain, comprising, among others, primary (PRIDE and PASSEL) and secondary resources (PeptideAtlas and UniProt), proteomics bioinformaticians, investigators (including some involved in the HUPO Human Proteome Project), and representatives from journals regularly publishing proteomics data (**Supplementary Note**, section 6). The aim of the ProteomeXchange consortium is to provide a common framework and infrastructure for the cooperation of proteomics resources by defining and implementing consistent, harmonized, user-friendly data deposition and exchange procedures among the major public proteomics repositories.

ProteomeXchange provides unified data submission for multiple MS data types and delivers different 'views' of the deposited data, such as the raw data suitable for reprocessing, the author-generated identifications and highly filtered composite results in resources like UniProt, all linked by a universal shared identifier. Authors are able to cite the resulting ProteomeXchange accession number for data sets reported in their publications. As such, a data set (with appropriate metadata) is becoming publishable *per se* and can be tracked if used by various consumers in different publications.

Individual resources can join ProteomeXchange by implementing the ProteomeXchange data submission and dissemination guidelines, and metadata requirements. In the current version (http://www.proteomexchange.org/concept), the mandatory information includes the following: first, mass spectrometer output files (raw data, either in a binary format, or in a standard open format such as mzML); second, processed identification results (two submission modes are available, see below); and third, sufficient metadata to provide a suitable biological and technological background, including method information such as transition lists in the case of SRM data. Other types of information, such as peak list files (processed versions of mass spectra most often used in the identification process) and quantification results can also be provided.

Two main MS proteomics workflows are now fully supported: tandem MS and SRM data (**Fig. 1** and **Supplementary Fig. 1**). PRIDE acts as the initial submission point for MS/MS data, whereas PASSEL is the initial submission point for SRM data. It is expected that, in most cases, one ProteomeXchange data set will correspond to data from one publication, and it will be clearly linked to it. However, this concept is flexible and a mechanism for grouping different ProteomeXchange data sets is also available, for example, for large-scale collaborative studies. At present, two different submission modes ('complete' and 'partial') are available for MS/MS data.

Complete submission requires peptide and protein identification results to be fully supported and integrated in the receiving repository (PRIDE at present). The search engine output files (plus the associated spectra) must therefore first be converted to PRIDE XML or mzIdentML format (a process supported by several popular and user-friendly tools; **Supplementary Note**, section 5). Complete submissions make the data fully available for querying, and thus maximize the potential for data re-use in MS. This in turn increases the visibility of the associated publication. A DOI (digital object identifier) is assigned to each data set, allowing formalized credit to be given to submitters and their principal investigators, through a citation index, as proposed in your editorial[3].

In a partial submission, peptide or protein identification results cannot be integrated in PRIDE because data converters and exporters to the supported formats are not yet available. In this case, search engine output files can be directly provided in their original format. Although partial submissions are searchable by their metadata, they are not fully searchable by results, such as protein identifiers, and will not receive a DOI. However, partial submissions are important as they allow data from newly developed experimental approaches to be deposited into the ProteomeXchange resources, rather than having to reject these until the workflows have been mapped into a representation in PRIDE or another ProteomeXchange partner.

For the submission of MS/MS data sets, a stand-alone, open-source Java tool has been made available, the 'ProteomeXchange submission tool' (http://www.proteomexchange.org/submission) (**Supplementary Figs. 2–10** and **Supplementary Note**, section 5). The tool allows interactive submission of small data sets as well as large-scale batch submissions.

For SRM data sets, a web form (http://www.peptideatlas.org/submit) can be used for submission to PASSEL. Similar to the guidelines stated above for MS/MS data sets, PASSEL submissions require mass spectrometer output files, study metadata, peptide reagents, analysis result files and the actual SRM transition lists, the information that drives the instrument data acquisition. Once data sets are submitted, they are checked by a curator and then loaded into the main PASSEL database, which facilitates interactive exploration of the data and results.

The submitted information and files can selectively be made available to journal editors and reviewers during manuscript peer review. Once the manuscript is accepted for publication or the submitter informs the receiving repository directly, the data will be publicly released (**Fig. 1**). At this point, the availability of the data set, as well as basic metadata, will be disseminated through a public RSS feed (http://groups.google.com/group/proteomexchange/feed/rss_v2_0_msgs.xml). The RSS feed includes a link to an XML message (ProteomeXchange XML), which is created by the receiving repository (**Supplementary Note**, section 3), and made available from ProteomeCentral, the portal for all public ProteomeXchange data sets (http://proteomecentral.proteomexchange.org) (**Supplementary Note**, section 2). Repositories, such as PeptideAtlas or GPMDB, as well as any interested end users can subscribe to this RSS feed and trigger actions, including incorporation of the data into local resources, re-processing or biological analysis. This reprocessing is already occurring in practice. For example, two ProteomeXchange data sets (PXD000134 and PXD000157) have been used in the latest build of the human proteome in PeptideAtlas, and PXD000013 (ref. 12) was reprocessed and nominated as technical data set of the year 2012 by GPMDB (http://www.thegpm.org/dsotw_2012.html#201210071).

ProteomeXchange started to accept regular submissions in June 2012. As of the beginning of February 2014, 685 ProteomeXchange data sets have been submitted (consisting of 656 tandem MS and 29 SRM data sets; **Fig. 2**), a total of ~32 Tb of data. The largest submission so far (data sets PXD000320–PXD000324) comprised 5 Tb of data. For a current list of the publicly available data sets, see http://proteomecentral.proteomexchange.org/.

In summary, ProteomeXchange provides an infrastructure for efficient and reliable public dissemination of proteomics data, supporting crucial validation, analysis and re-use. By providing and linking different interpretations of the data, we aim to maximize data set visibility as well as their potential benefit to different communities. Citability and traceability are addressed through the assignment of DOIs and a common identifier space. The consortium is open to the participation of additional resources (**Supplementary Note**, section 9). Although all repositories depend on continuous funding for continuous operation, the ProteomeXchange core repositories PRIDE and PeptideAtlas are well established, with first publications in 2005 (refs. 7,8), and have strong institutional backing (**Supplementary Note**, section 8), ensuring that the data will remain reliably available for the foreseeable future. We are confident that the ProteomeXchange infrastructure will support the growing trend toward public availability of proteomics data, maximizing its benefit to the scientific community through increased ease of access, greater ability to re-assess interpretations and extract further biological insight, and greater citation rates for the submitters.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper (doi:10.1038/nbt.2839).

ACKNOWLEDGMENTS
We thank all the members of the community who participated as stakeholders in the ProteomeXchange meetings. This work was supported by the EU FP7 grant ProteomeXchange (grant number 260558). J.A.V., A.C., F.R. and D.R. were also funded by the Wellcome Trust (grant number WT085949MA). E.W.D., Z.S. and T.F. are also funded in part by US National Institutes of Health/National Institute General Medical Sciences grant no. R01 GM087221, US National Science Foundation Major Research Instrumentation Program (grant number 0923536), and the Luxembourg Centre for Systems Biomedicine and the University of Luxembourg. M.E. is funded by Protein Unit for Research in Europe (PURE), a project of Nordrhein-Westfalen. L.G. was supported by the EU FP7 PRIME-XS project (grant number 262067). R.W. was supported by the UK Biotechnology and Biological Sciences Research Council 'PRIDE Converter' grant (reference BB/I024204/1). G.S.O. acknowledges support from US National Institute of Health grants RM-08-029, P30U54ES017885 and UL1RR24986.

AUTHOR CONTRIBUTIONS

J.A.V., H.H. and E.W.D. led the current implementation of the ProteomeXchange data workflow, guidelines and related software. R.W. developed the ProteomeXchange submission tool. The remaining authors (A.C., F.R., D.R., J.A.D., Z.S., T.F., N.B., P.-A.B., I.X., M.E., G.M., L.G.,

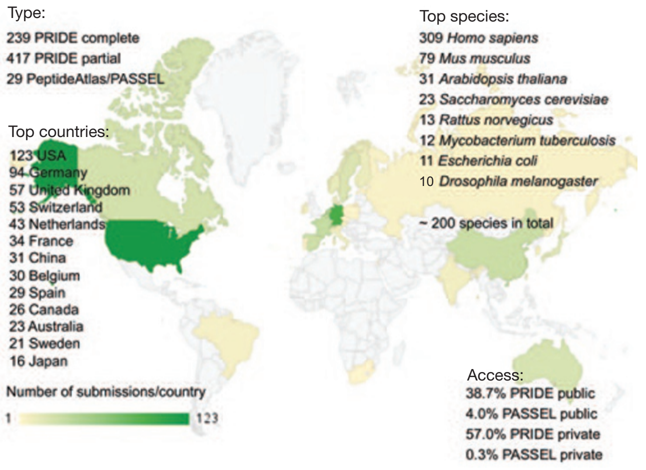

Figure 2 Summary of the main metrics of ProteomeXchange submissions (as of February 2014). The number of data sets is indicated for submission type, data access status and for the top species and countries represented.

A.C., R.J.C., H.-J.K., J.P.A., S.M.-B.,R.A., G.S.O., L.M. and A.R.J.) contributed to the development of the ProteomeXchange consortium in different ways, for example, by contributing to the initial ProteomeXchange prototypes in the past, developing software and data standards, or contributing in different aspects to the implementation of the guidelines and the data workflow. J.A.V., E.W.D. and H.H. wrote the manuscript. All authors have agreed to all the content in the manuscript, including the data as presented.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Juan A Vizcaíno[1,17], Eric W Deutsch[2,17], Rui Wang[1], Attila Csordas[1], Florian Reisinger[1], Daniel Ríos[1], José A Dianes[1], Zhi Sun[2], Terry Farrah[2], Nuno Bandeira[3], Pierre-Alain Binz[4], Ioannis Xenarios[4–6], Martin Eisenacher[7], Gerhard Mayer[7], Laurent Gatto[8], Alex Campos[9], Robert J Chalkley[10], Hans-Joachim Kraus[11], Juan Pablo Albar[12], Salvador Martinez-Bartolomé[12], Rolf Apweiler[1], Gilbert S Omenn[2,13], Lennart Martens[14,15], Andrew R Jones[16] & Henning Hermjakob[1]

*[1]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, UK. [2]Institute for Systems Biology, Seattle, Washington, USA. [3]Center for Computational Mass Spectrometry, University of California, San Diego, La Jolla, California, USA. [4]Swiss-Prot Group, SIB Swiss Institute of Bioinformatics, Geneva, Switzerland. [5]University of Lausanne, Lausanne, Switzerland, and Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland. [6]Vital-IT Group, SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland. [7]Medizinisches Proteom-Center, Ruhr-Universität Bochum, Bochum, Germany. [8]Computational Proteomics Unit and Cambridge Centre for Proteomics, Department of Biochemistry, University of Cambridge, Cambridge, United Kingdom. [9]Integromics SL, Santiago Grisolia, Madrid, Spain. [10]Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, California, USA. [11]Wiley-VCH Verlag, Weinheim, Germany. [12]ProteoRed-ISCIII, National Center for Biotechnology-CSIC, Madrid, Spain. [13]Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, USA. [14]Department of Medical Protein Research, VIB, Ghent, Belgium. [15]Department of Biochemistry, Ghent University, Ghent, Belgium. [16]Institute of Integrative Biology, University of Liverpool, UK. [17]These authors contributed equally to this work.
e-mail: juan@ebi.ac.uk.*

1. Hahne, H. & Kuster, B. *Mol. Cell. Proteomics* **11**, 1063–1069 (2012).
2. Matic, I., Ahel, I. & Hay, R.T. *Nat. Methods* **9**, 771–772 (2012).
3. Editors. *Nat. Biotechnol.* **27**, 579 (2009).
4. Olsen, J.V. & Mann, M. *Sci. Signal.* **4**, pe7 (2011).
5. The UniProt Consortium. *Nucleic Acids Res.* **40**, D71–D75 (2012).
6. Editors. *Nat. Methods* **9**, 419 (2012).
7. Martens, L. *et al. Proteomics* **5**, 3537–3545 (2005).
8. Deutsch, E.W. *et al. Proteomics* **5**, 3497–3500 (2005).
9. Farrah, T. *et al. Proteomics* **12**, 1170–1175 (2012).
10. Craig, R., Cortens, J.P. & Beavis, R.C. *J. Proteome Res.* **3**, 1234–1242 (2004).
11. Hermjakob, H. & Apweiler, R. *Expert Rev. Proteomics* **3**, 1–3 (2006).
12. Vaudel, M. *et al. J. Proteome Res.* **11**, 5072–5080 (2012).