ZENBU is freely available as a web service at http://fantom.gsc.riken.jp/zenbu/. ZENBU can also be installed locally from the open-source source code (**Supplementary Data**) or via preconfigured virtual machines which we provide. There is also a wiki-based documentation available on the website containing a detailed manual and set of case studies (**Supplementary Note 1** and **Supplementary Figs. 2–4,7–13**).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper (doi:10.1038/nbt.2840).

ACKNOWLEDGMENTS

We would like to acknowledge C. Plessy, P. Carninci and the FANTOM5 consortium members for critical feedback during development of the system. The work was funded by a research grant for RIKEN Omics Science Center from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) to Y.H. and a grant of the Innovative Cell Biology by Innovative Technology (Cell Innovation Program) from the MEXT, Japan to Y.H. This study is also supported by Research Grants from MEXT through RIKEN Preventive Medicine and Diagnosis Innovation Program to Y.H. and RIKEN Center for Life Science Technologies, Division of Genomic Technologies to Piero Carninci.

AUTHOR CONTRIBUTIONS

J.S. designed and developed the software; J.S., N.B., C.O.D. and A.R.R.F. contributed to the design of the interface and data views; C.O.D., A.R.R.F. and Y.H. supervised the project; J.S., M.L., J.H. and N.B. contributed to the loading and curation of the data; J.S., J.H. and N.B. contributed to the source code; J.S., N.B., H.K. and A.R.R.F. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Jessica Severin[1,2], Marina Lizio[1,2], Jayson Harshbarger[1,2], Hideya Kawaji[1-3], Carsten O Daub[1,2], Yoshihide Hayashizaki[2,3], The FANTOM Consortium, Nicolas Bertin[1,2,4] & Alistair R R Forrest[1,2]

[1]RIKEN Center for Life Science Technologies (Division of Genomic Technologies), Suehiro-cho, Tsurumi-ku, Yokohama, Japan. [2]RIKEN Omics Science Center (OSC), Yokohama, Japan. [3]RIKEN Preventive Medicine and Diagnosis Innovation Program, Wako, Japan. [4]Present address: Cancer Science Institute of Singapore, National University of Singapore, Singapore. e-mail: nicolas.bertin@gmail.com or alistair.forrest@gmail.com

1. Chadwick, L.H. *Epigenomics* **4**, 317–324 (2012).
2. The ENCODE Project Consortium. *Nature* **489**, 57–74 (2012).
3. Li, H. *et al. Bioinformatics* **25**, 2078–2079 (2009).
4. Nielsen, C.B., Cantor, M., Dubchak, I., Gordon, D. & Wang, T. *Nat. Methods* **7**, S5–S15 (2010).
5. Saito, T.L. *et al. Bioinformatics* **25**, 1856–1861 (2009).
6. Stein, L.D. *et al. Genome Res.* **12**, 1599–1610 (2002).
7. Kuhn, R.M., Haussler, D. & Kent, W.J. *Brief. Bioinform.* **14**, 144–161 (2013).
8. Hubbard, T. *et al. Nucleic Acids Res.* **30**, 38–41 (2002).
9. Robinson, J.T. *et al. Nat. Biotechnol.* **29**, 24–26 (2011).
10. Zhang, J. *et al. Database (Oxford)* **2011**, bar038 (2011).
11. Derrien, T. *et al. Genome Res.* **22**, 1775–1789 (2012).
12. Frith, M.C. *et al. Genome Res.* **18**, 1–12 (2008).
13. Wei, G. *et al. Immunity* **35**, 299–311 (2011).
14. Severin, J. *et al. BMC Bioinformatics* **11**, 240 (2010).
15. Carninci, P. *et al. Science* **309**, 1559–1563 (2005).
16. Suzuki, H. *et al. Nat. Genet.* **41**, 553–562 (2009).
17. Contrino, S. *et al. Nucleic Acids Res.* **40**, D1082–D1088 (2012).
18. Giardine, B. *et al. Genome Res.* **15**, 1451–1455 (2005).
19. The Cancer Genome Atlas Research Network *et al. Nat. Genet.* **45**, 1113–1120 (2013).

OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data

To the Editor:

Liquid chromatography tandem mass spectrometry (LC-MS/MS)-based proteomics is the method of choice for large-scale identification and quantification of proteins in a sample[1]. Several LC-MS/MS methods have been developed that differ in their objectives and performance profiles[2]. Among these, shotgun proteomics (also referred to as discovery proteomics) using data-dependent acquisition (DDA) and targeted proteomics using selected reaction monitoring (SRM, also referred to as multiple reaction monitoring, MRM) have been widely adopted. Alternatively, some mass spectrometers can also be operated in data-independent acquisition (DIA) mode[3–15]. In DIA mode, the instrument fragments all precursors generated from a sample that are within a predetermined mass-to-charge ratio (m/z) and retention-time range. Usually, the instrument cycles through the precursor-ion m/z range in segments of specified width, at each cycle producing a highly multiplexed fragment-ion spectrum. Multiple DIA methods have been described with different instrument types and setups, duty cycles and window widths. Methods such as MS^E (simultaneous acquisition of exact mass at high and low collision energy) fragment all precursors[5], whereas others, such as PAcIFIC (precursor acquisition independent from ion count), use precursor selection windows as small as 2.5 m/z[6] (see ref. 16 for a recent overview). In this Correspondence, we describe OpenSWATH, a software for automated targeted DIA analysis, benchmark it against manual analysis of >30,000 chromatograms from 342 synthesized peptides and use it to analyze the proteome of *Streptococcus pyogenes*.

DIA methods offer several potential advantages over shotgun proteomics and SRM. Specifically, data acquired in DIA mode is continuous in time and fragment-ion intensity, thus increasing the dimensionality of the data relative to shotgun proteomics, in which full fragment-ion intensity scans are recorded only at selected time points (MS/MS spectra), or SRM, in which continuous time profiles are acquired but only for selected fragment ions (ion chromatograms)[1,17–20]. Thus, DIA methods produce a complete two-dimensional record of the fragment-ion signal of all precursors generated from a sample (**Fig. 1a**). By acquiring time-resolved data of all fragment ions, DIA has the potential to overcome some of the limitations of the current proteomic methods and to combine the high-throughput of shotgun proteomics with the high reproducibility of SRM[21,22].

However, DIA data has historically been more difficult to analyze than shotgun or SRM data. To limit the time needed for data analysis and the amount of sample required, one typically uses larger precursor-isolation windows than in shotgun proteomics or SRM[16]. This leads to highly complex, composite fragment-ion spectra from multiple precursors and thus to a loss of the direct relationship between a precursor and its fragment ions, making subsequent data analysis nontrivial. To date, DIA data have been analyzed by one of two strategies. In the first, fragment-ion spectra[4,6] or pseudo fragment-ion spectra (which are computationally reconstructed from the complex data sets[8,9,11–13]) are searched by methods developed for DDA. In these approaches, a proteomics search engine compares experimental spectra to theoretical spectra generated by an *in silico* tryptic digest of a proteome, assuming that the fragment-ion spectrum is derived from a single precursor. These approaches suffer from the high complexity of the data and the fact that errors in the generation of pseudo-spectra will propagate through the analysis workflow.

Recently, we proposed an alternative, fundamentally different DIA data analysis

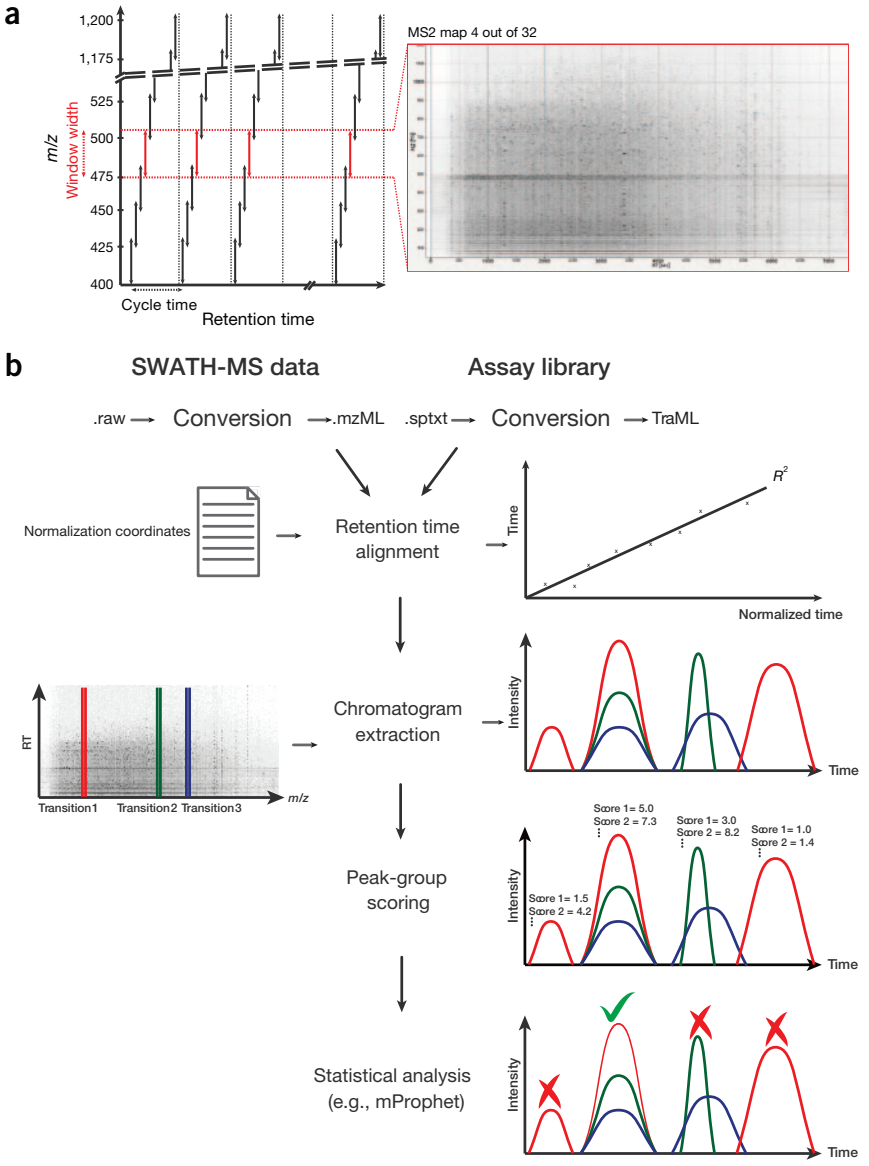

Figure 1 SWATH-MS data-independent acquisition and OpenSWATH analysis. (**a**) The DIA method used here consists of sequential acquisition of fragment-ion spectra with overlapping precursor isolation windows. Here, a swath window width of 25 *m/z* is depicted which allows stepping through a mass range of 400–1,200 *m/z* in 32 individual steps. If all fragment-ion spectra of the same isolation window are aligned, an MS2 map (so-called swath) is obtained (right side, swath 4 out of 32 is schematically shown). Figure adapted from ref. 16. (**b**) The individual steps performed by the OpenSWATH software are illustrated for a peptide precursor with three transitions: red, green and blue. The steps are data conversion, retention-time alignment, chromatogram extraction, peak-group scoring and statistical analysis to estimate an FDR (false-discovery rate). See main text for a more detailed explanation of the workflow.

approach, which is derived from methods for analyzing SRM-based targeted proteomics data. We implemented it in a method called SWATH-MS[16]. In SWATH-MS, precursor ions from sequential segments of 25 *m/z* units are concurrently fragmented and the resulting composite fragment ions are recorded at high mass accuracy in a time-of-flight (TOF) analyzer. In our targeted data analysis strategy, extracted ion chromatograms (XIC) of the most intense transitions of a targeted peptide are generated from all corresponding MS/MS spectra, producing chromatographic data that are similar to SRM traces (**Fig. 1a**). This approach reduces the complexity of the data substantially, facilitating data analysis while retaining the complete fragment-ion information of all precursors. So far, such data analysis has been performed semi-manually and to our knowledge, no automated workflow has been published. However, certain specialized software, such as Skyline[23] and PeakView (AB SCIEX), can visualize the XIC data, making manual analysis possible. Automation of this process is critical, however, because in a single typical SWATH-MS data set tens of thousands of peptides are identified.

Here we present OpenSWATH, an open-source (Modified BSD License) software that allows targeted analysis of DIA data in an automated, high-throughput fashion. OpenSWATH is cross-platform software, written in C++, that relies only on open data formats, allowing it to analyze DIA data from multiple instrument vendors (**Supplementary Note 1**)[24]. The algorithm can be summarized in the following five steps (**Fig. 1** and **Supplementary Notes 2–4**).

Data conversion. OpenSWATH takes as input the acquired SWATH-MS data and an assay library. These are first converted to suitable open file formats (mzML and TraML[25,26]). The assay library contains precursor- and fragment-ion *m/z* values (transitions) as well as relative fragment-ion intensities and normalized peptide retention times. Decoy assays are appended to the target assay library for later classification and error rate estimation.

Retention-time alignment. Each run is aligned against a previously determined normalized retention-time space using reference peptides whose mappings to the normalized space are known (for example, spiked-in peptides), as described previously[27]. Outlier detection is subsequently applied to remove wrongly assigned reference peptides and to evaluate the quality of the alignment.

Chromatogram extraction. Using the *m/z* and retention-time information from the assay library, the workflow extracts an ion chromatogram from the corresponding MS/MS map, producing integrated fragment-ion counts versus retention-time data. The extraction function (Top-hat or Bartlett) and *m/z* window-width can be specified to account for the instrument-specific MS/MS resolution.

Peak-group scoring. The core algorithm identifies 'peak groups' (that is, positions in the chromatograms where individual fragment traces coelute) and scores them using multiple, orthogonal scores (**Supplementary Note 4**). These scores are based on the elution profiles of the fragment ions, the correspondence of the peak group with the expected retention time and fragment-ion intensity from the assay library, as well as the properties of the full MS/MS spectrum at the chromatographic peak apex.

Statistical analysis. The separation between true and false signal is achieved

using a set of decoy assays that were scored exactly the same way as the target assays. The false-discovery rate (FDR = false positives/(true positives + false positives)) can then be estimated, for example by the mProphet algorithm[28]. If multiple runs are present, a peak-group alignment can be performed to annotate signals that could not be confidently assigned using data from a single run alone, as described previously for data-dependent acquisition and SRM data[29].

To validate and benchmark our SWATH-MS data analysis algorithms, we created a 'gold standard' data set of known composition (termed SGS for SWATH-MS Gold Standard), consisting of 422 chemically synthesized, stable isotope–labeled standard (SIS) peptides[30,31] (**Supplementary Table 2**). To simulate differently abundant peptides in proteomic backgrounds of varying complexity, we added the peptides in ten dilution steps at final concentrations ranging from 0.058 fmol/µL to 30.0 fmol/µL into three different backgrounds (water or trypsinized whole-cell protein extracts from *Homo sapiens* or *Saccharomyces cerevisiae*, normalized to 1 µg of total protein; **Supplementary Note 5**). We deliberately chose to explore the lower end of the dynamic range in this experiment, allowing us to study the influence of background complexity on ion suppression and signal-to-noise (see below and **Supplementary Note 5.4**). These samples were measured on the AB SCIEX TripleTOF 5600 System in DIA mode as described previously[16] (**Supplementary Note 6**). Using an assay library for 342 peptides (not all 422 peptides generated high-quality fragment-ion spectra, **Supplementary Data 1**), the 30,780 chromatograms were extracted in Skyline[23] and manually analyzed to determine the true peak group (if present). In parallel, the same data were processed with OpenSWATH and results were compared with those generated by the manual analysis (**Supplementary Data 2**).

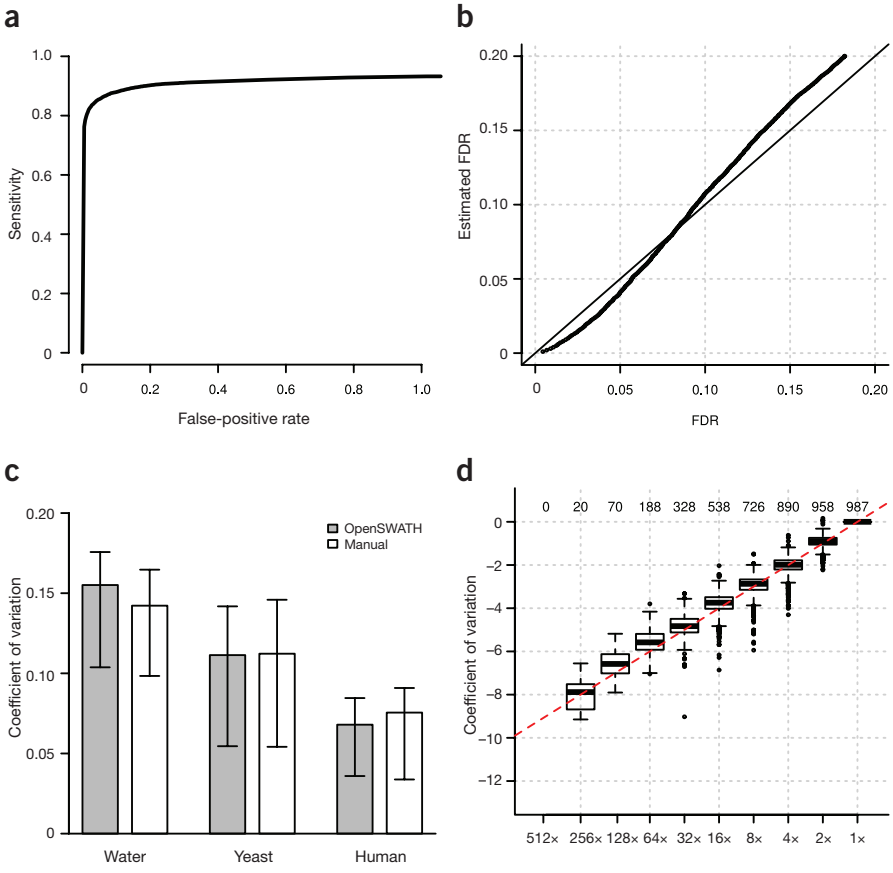

Figure 2 Identification and quantification accuracy of OpenSWATH on the SGS data set. We spiked 422 peptides into three different proteomic backgrounds in a ten-step dilution series to produce a 'gold standard' data set. (**a**) Pseudo-ROC curve showing sensitivity (recall of true signals) versus the false-positive rate, achieving an AUC >0.9 using OpenSWATH. Because misidentified peaks cannot be recovered, even at high score cutoff values, a sensitivity of 1.0 cannot be reached. (**b**) The estimated FDR (by mProphet[28]) versus manually curated, true FDR on the SGS data set. The continuous line at 45 degrees shows the optimal values. (**c**) Mean CVs across the three technical replicates are below 20% CV (no significant difference between OpenSWATH and manual quantification for yeast and human backgrounds using the Mann-Whitney test; whiskers indicate 25% and 75% quantiles). (**d**) Peptide intensities quantified by OpenSWATH for all ten dilution steps, normalized to the most intense concentration shown for the yeast proteomic background. The red dashed line indicates the ideal values (twofold difference to the next dilution). The number of peaks considered is given on the top. For panels **c** and **d**, only peptides that were detectable above a cutoff of 1% FDR were analyzed and only true positives were considered. For panel **c**, only peptides present in all triplicates were analyzed.

To assess the identification accuracy of OpenSWATH, we calculated the pseudo-receiver operator characteristics (ROC) using the best peak group per chromatogram and computed an area under the curve (AUC) >0.9 (**Fig. 2a**). At a fixed FDR of 5% (as computed by mProphet[28]), the software could achieve a recall of 87.5% and a precision of 94.3%. Furthermore, we noticed that the misidentification rate (that is, cases where the highest scoring peak group is not the correct peak group) is below 0.7%. Thus, most of the false identifications were caused by peak groups that were not confidently assigned by manual curation, rather than by misidentification by OpenSWATH. Furthermore, we found a good correspondence between the estimated FDR and the true, manually determined false-positive rate (with a slight underestimation of 0.9% at 1% FDR, **Fig. 2b**), indicating that OpenSWATH can identify peptides with high precision and that it supports the accurate selection of the desired false-positive rate. However, accurate error rate estimations critically depend on a suitable decoy strategy[32] (**Supplementary Note 1.5**). Similar to methods for SRM data analysis, OpenSWATH uses the sum of the integrated chromatographic fragment-ion peak areas of SWATH-MS data to quantify peptides. When analyzing the coefficients of variation (CV) of quantified signals reported in all technical replicates, we consistently found mean CVs below 20% (**Fig. 2c**). By normalizing the intensities of each peptide signal to the intensity of the most concentrated run (1 × dilution), we could evaluate the quantification accuracy achieved by the software over large fold changes (**Fig. 2d**). Because our goal was to study quantification accuracy, we did not include misidentified peptides in our analysis. We found that the manually determined changes between subsequent dilution steps (water, 2.35 ± 1.0 (mean fold change ± s.d.); yeast, 2.03 ± 0.45; and human, 2.11 ± 0.53) matched closely with the changes determined using OpenSWATH (water, 2.62 ± 1.43; yeast, 2.02 ± 0.44; and human, 1.96 ± 0.39). From this, we computed

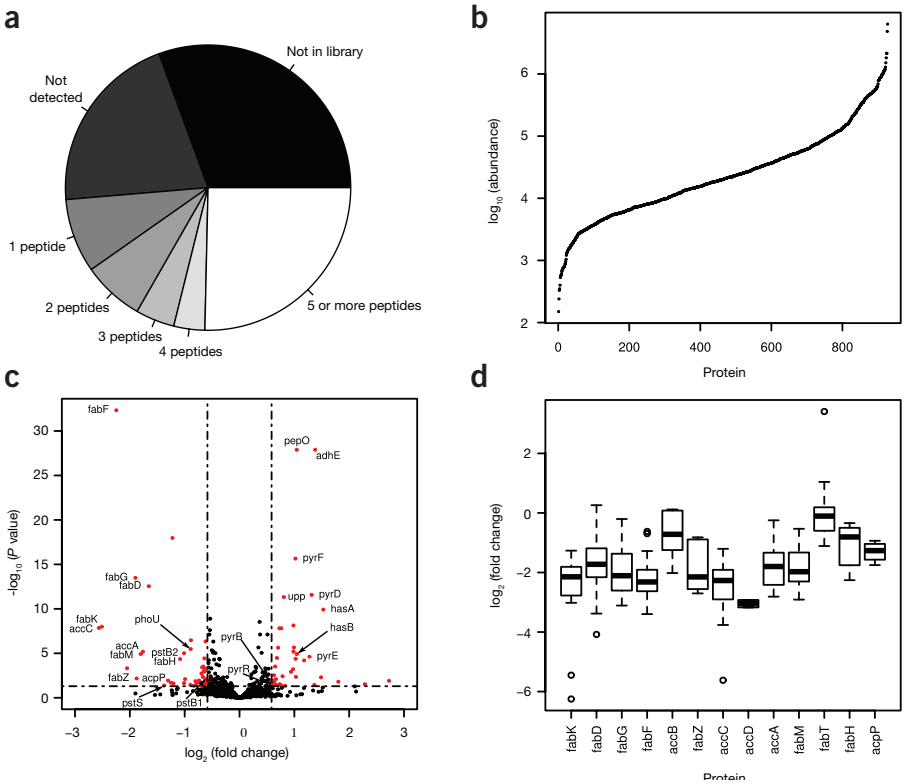

Figure 3 *Streptococcus pyogenes* exposed to human plasma (0% plasma versus 10% plasma). Analysis of two biological replicates with OpenSWATH at 1% assay FDR yields over 900 proteins and 6,000 peptides consistently quantified over four SWATH-MS runs. (**a**) Proteome coverage of *S. pyogenes*: of 1,905 annotated ORFs, 1,322 were detected using mass spectrometry after extensive fractionation (constituting the assay library) and 927 could be detected consistently in each of four unfractionated samples using SWATH-MS. (**b**) Protein abundances of *S. pyogenes* as detected by SWATH-MS estimated by the ion count of the most intense peptide. (**c**) Volcano plot (log fold change versus log *P* value) of protein expression determined by ANOVA analysis on two biological replicates. Red dots indicate fold changes above 1.5 fold and a Benjamini-Hochberg–corrected *P* value below 0.05. (**d**) The fold changes of all 13 proteins involved in fatty-acid biosynthesis (FAB) in *S. pyogenes*, in the same order as they appear on their respective operons. All proteins are substantially upregulated, except accD, accB and fabT (where fabT is a transcriptional repressor and not expected to be upregulated).

a deviation from the theoretical value of 31.2%, 1.0% and 2.0% and a CV of 54.6%, 21.9% and 20.2% for the OpenSWATH quantification (respectively for the three backgrounds, outliers removed), suggesting that OpenSWATH quantification is suitable for obtaining relative quantification values for differentially abundant peptides. The quantification in water is less accurate and precise than in the yeast and human backgrounds, because without a matrix, the spiked-in SIS peptides are prone to surface adsorption during sample preparation (**Supplementary Note 5.5**).

We next explored the performance of OpenSWATH in identifying and quantifying peptides from a full tryptic digest of a *S. pyogenes* microbial sample. To study proteomic changes that occur upon vascular invasion of the pathogen, we grew *S. pyogenes* (strain SF370) in 0% and 10% human plasma in biological duplicates and analyzed the samples in SWATH-MS mode on an AB SCIEX TripleTOF 5600 System (**Supplementary Note 6**). First, we created a spectral library of *S. pyogenes* by combining the measurements of ten fractions of the *S. pyogenes* proteome in data-dependent aquisition (shotgun) mode on the same instrument (**Supplementary Data 3**), providing an extensive coverage of the expressed *S. pyogenes* proteome, with 1,322 proteins (out of 1,905 open reading frames; ORFs) mapping to 20,027 proteotypic peptide precursors at 1% peptide-spectrum match FDR (**Fig. 3a**).

Using OpenSWATH, we identified and quantified 927 proteins (out of 1,322 targeted proteins) of *S. pyogenes* consistently in each of the four LC-MS/MS runs at 1% FDR (**Supplementary Table 1**). Out of these, 767 proteins were quantified by more than one peptide per protein. Thus, we achieved >70% coverage of the expressed proteome spanning more than three orders of dynamic range in estimated protein ion count (**Fig. 3b**) in a single injection. The results from these analyses surpassed previous shotgun proteomics and SRM approaches in terms of number of quantified proteins at 1% FDR (765 proteins were quantified in an extensive SRM study with multiple injections per sample and 523 proteins were identified in a shotgun proteomics study with 98.92% overlap with our data, see **Supplementary Note 1.4**)[29,33]. The fraction of the assay library that could not be detected may be partially explained by the fact that not all proteins were expressed under the conditions studied and that these proteins have also rarely been identified in earlier studies (nearly 80% were never identified in PeptideAtlas[34]).

OpenSWATH identified 82 proteins, which showed significant ($P < 0.05$ in an multiple testing-corrected ANOVA test) differences in abundance between the two conditions in two biological replicates (**Fig. 3c,d**; see **Supplementary Note 7**). Ten out of 13 proteins associated with fatty-acid biosynthesis are significantly ($P < 0.05$) downregulated, consistent with results of previous studies on *S. pyogenes*[33]. As expected, we also found several known virulence factors to be upregulated (for example, HasA, HasB, Slo, SpeC and CovR)[35,36]. Additionally, we observed significant ($P < 0.05$) downregulation of an ABC transporter complex for inorganic phosphate import (PstB1, PstB2 and PstS), as well as significant ($P < 0.05$) upregulation of six proteins involved in pyrimidine biosynthesis (PyrF, PyrD, PyrE, PyrB, PyrR and Upp). Although these results agree with previous observations on *S. pyogenes*, they also provide the first indications that the Pst system is involved in responding to human plasma in *S. pyogenes*. In conclusion, our results derived from SWATH-MS data sets analyzed with OpenSWATH are consistent with many previous suppositions about bacterial virulence but additionally are able to provide the foundation for new hypotheses (**Supplementary Note 7**).

By combining the most advanced DIA technology with a software capable of analyzing the resulting complex data sets, we were able to substantially scale-up the targeted proteomic approach described previously[16] and show that targeted analysis of DIA data facilitates high-throughput analysis of microbial whole-cell lysates, as demonstrated on the example of *S. pyogenes*. Using the SGS validation data set, we further demonstrate high sensitivity of the method and software for identification and

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