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

Hannes L. Röst,[1,2,*] George Rosenberger,[1,2,*] 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] and Ruedi Aebersold[1,6,7,†]

[1]*Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, CH-8093 Zurich, Switzerland*
[2]*Ph.D. Program in Systems Biology, University of Zurich and ETH Zurich, CH-8057 Zurich, Switzerland*
[3]*Biognosys AG, CH-8952 Schlieren, Switzerland*
[4]*SyBIT project of SystemsX.ch, ETH Zurich, CH-8092 Zurich, Switzerland*
[5]*Department of Immunotechnology, Lund University, S-22100 Lund, Sweden*
[6]*Competence Center for Systems Physiology and Metabolic Diseases, CH-8093 Zurich, Switzerland*
[7]*Faculty of Science, University of Zurich, CH-8057 Zurich, Switzerland*

(Dated: October 19, 2015)

(This article has been published in Nat Biotechnol. 2014 Mar 10;32(3):219-23.)
(http://dx.doi.org/10.1038/nbt.2841)

To the Editor:

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. In the choice between the two, the researcher is faced with the trade-off between obtaining snapshots of extensive fragment ion data (full MS/MS spectra) of a population of peptides sampled from the pool of available peptides or obtaining time-resolved fragment ion intensities (ion chromatograms) for a selected number of predetermined peptides targeted in the measurement [1, 3–6]. While shotgun proteomics allows discovery-driven research and offers high throughput, its sensitivity is strongly sample dependent and it can suffer from inconsistent identification reproducibility across samples, sampling bias and ambiguity in spectra assignments to peptides. In contrast, SRM offers high reproducibility, a larger dynamic range, more sensitivity and good signal-to-noise ratio but comes at the cost of significantly lower throughput [2, 7, 8].

As an alternative to data-dependent shotgun proteomics and targeted SRM, some mass spectrometers can also be operated in data-independent acquisition (DIA) mode [9–21]. There, the instrument fragments all precursors generated from a sample that are within a predetermined m/z and retention time range. Usually, the instrument cycles through the precursor ion m/z range in segments of specified width. Multiple variations of the DIA theme have been described with different instrument types and setups, duty cycles and window widths. Methods like MS^E fragment all precursors [11] while others such as PAcIFIC use precursor selection windows as small as 2.5 Da [12] (see Gillet *et al.* [22] for a recent overview of different DIA approaches). The obvious advantage of these methods is that they create a complete record of the fragment ion spectra of all precursors generated from a sample, therefore combining the high throughput of shotgun proteomics with the high reproducibility of SRM. The resulting data is continuous in time *and* fragment ion intensity, thus increasing the dimensionality of shotgun proteomics data where fragment ion intensities are recorded only at selected time points or SRM data where continuous time profiles are acquired but only for selected fragment ions. However, to limit analysis time and sample amount, larger precursor isolation windows than in shotgun proteomics or SRM are typically used. 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. In most previous studies, researchers have either searched the multiplexed spectra from DIA data directly [10, 12] or after computation of pseudo-spectra containing fragments assigned to a precursor based on their coelution profiles [14, 15, 17–19]. However, 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 have proposed a fundamentally different approach for the analysis of DIA data and implemented it in a method called SWATH-MS [22]. In SWATH-MS, precursor ions from sequential segments of 25 mass units are concurrently fragmented and the resulting composite fragment ions are recorded at high mass accuracy in a time of flight (TOF) analyzer. The SWATH-MS data analysis strategy has

*These authors contributed equally to this work.
†Corresponding author: aebersold@imsb.biol.ethz.ch; Phone: +41 44 633 31 70; Fax: +41 44 633 10 51

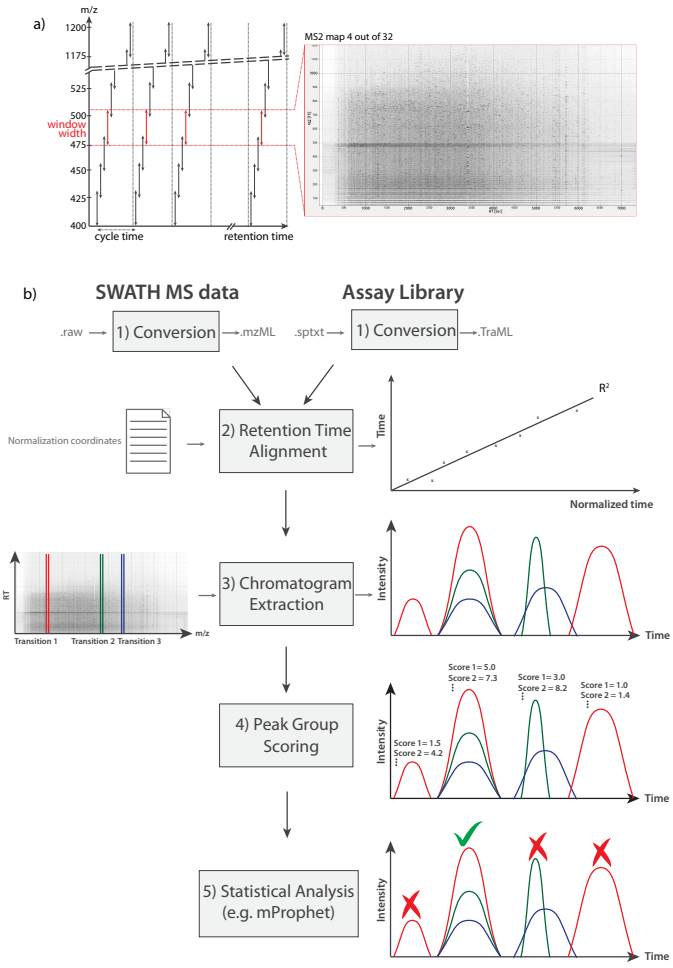

FIG. 1: **SWATH-MS data-independent acquisition and OpenSWATH analysis. a)** The DIA method used here consists of sequential acquisition of fragment ion spectra with overlapping precursor isolation windows. Here, a swath window width of 25 Da is depicted which allows stepping through a mass range of 400 - 1200 m/z in 32 individual steps. If all fragment ion spectra of the same isolation window are aligned, a MS2 map (so-called swath) is obtained (right side, swath 4 out of 32 is schematically shown). Figure adapted from Gillet *et al.* [22]. **b)** The individual steps performed by the OpenSWATH software are (illustrated by a peptide precursor with 3 transitions: red, green and blue): 1) Data conversion, 2) Retention time alignment, 3) Chromatogram extraction, 4) Peak group scoring, 5) Statistical analysis to estimate an FDR (false discovery rate). See main text (Results) for a more detailed explanation of the workflow.

its roots in the idea of targeted data analysis, whereby extracted ion chromatograms (XIC) of the most intense transitions of a targeted peptide are generated from all corresponding MS2 spectra, producing chromatographic data that is similar to SRM traces (Figure 1a). This approach reduces the complexity of the data significantly, facilitating data analysis while retaining the complete fragment ion information of all precursors. So far, the data analysis was performed semi-manually and to our knowledge, no automated workflow has been published.

Here we present OpenSWATH, the first open source (Modified BSD Licence) software that allows targeted analysis of DIA data in an automated fashion. The cross-platform C++ software relies only on open data formats, allowing it to analyze DIA data from multiple instrument vendors (see Supplementary Note 1) [23]. The algorithm can be summarized in the following 5 steps (see Figure 1 for a concise description of the algorithm as well as Supplementary Notes 2-4):

1) Data Conversion: The acquired SWATH-MS data together with an assay library comprise the input data which are converted to suitable open file formats (mzML and TraML [24, 25]). 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 using the OpenSwathDecoyGenerator for later classification and error rate estimation.

2) 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 (e.g. spiked-in peptides), as described by Escher *et al.* [26]. Outlier detection is applied afterwards to remove wrongly assigned reference peptides and to evaluate the quality of the alignment.

3) Chromatogram Extraction: Using the m/z and retention time information from the assay library, the workflow extracts an ion chromatogram from the corresponding MS2 map, producing integrated fragment ion count *vs.* retention time data. The extraction function (Top-hat or Bartlett) and m/z window-width can be specified to account for the instrument-specific MS2 resolution.

4) Peak group scoring: The core algorithm identifies so-called "peak groups" (positions in the chromatograms where individual fragment traces co-elute), 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 MS2 spectrum at the chromatographic peak apex.

5) 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) can then be estimated for example by the mProphet algorithm [27]. 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 DDA and SRM data [28].

To validate and benchmark our SWATH-MS data analysis algorithms, we created a "gold standard" dataset of known composition (termed SGS for SWATH-MS Gold Standard), consisting of 422 chemically synthesized, stable isotope-labeled standard (SIS) peptides [29, 30].

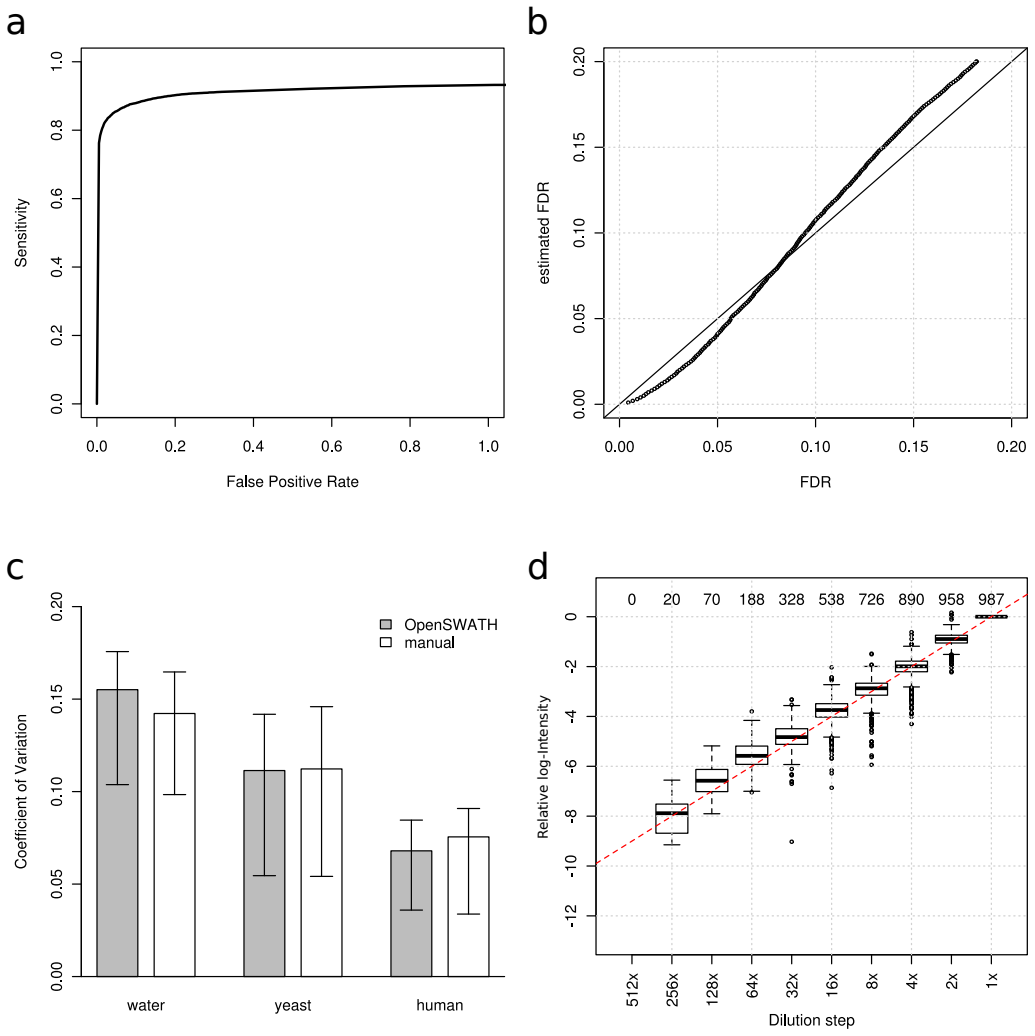

FIG. 2: **Identification and quantification accuracy of OpenSWATH on the SGS dataset.** 422 peptides were spiked into 3 different proteomic backgrounds in a 10-step dilution series to produce a "gold standard" dataset (see main text). **a)** Pseudo-ROC curve showing sensitivity (recall of true signals) *vs.* the false positive rate, achieving an AUC > 0.9 using OpenSWATH (since 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) *vs.* manually curated, true FDR on the SGS dataset. The continuous line at 45 degree shows the optimal values. **c)** Coefficients of variation (CVs) across the 3 technical replicates are below 20 % CV (no significant difference between OpenSWATH and manual quantification for yeast and human backgrounds using the Mann-Whitney test). **d)** Peptide intensities quantified by OpenSWATH for all 10 dilution steps, normalized to the most intense concentration shown for the yeast proteomic background. The red dotted line indicates the ideal values (two-fold 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 simulate differently abundant peptides in proteomic backgrounds of varying complexity, the peptides were added in 10 dilution steps at final concentrations ranging from 0.058 fmol/μL to 30.0 fmol/μL into three different backgrounds (trypsinized whole-cell protein extracts from *Homo sapiens*, *Saccharomyces cerevisiae* or water, normalized to 1 μg of total protein, see Supplementary Note 5). To explore the lower end of the dynamic range, we chose specifically 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 [22]. Using an assay library for 342 peptides, the 30,780 chromatograms were extracted in Skyline [31] and manually analyzed to determine the true peak group (if present). In parallel, the

same data were processed with OpenSWATH and results were compared to those generated by the manual analysis.

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 AUC > 0.9 (Figure 2a). At a fixed FDR of 5 % (as computed by mProphet [27]), the software could achieve a recall of 87.5 % and a precision of 94.3 %. Furthermore, we noticed that the misidentification rate (i.e. 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, see Figure 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; see Supplementary Note 1.5 [32]).

In analogy to SRM, 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 % (Figure 2c). By normalizing the intensities of each peptide signal to the intensity of the most concentrated run (1x dilution) we could evaluate the quantification accuracy achieved by the software over large fold changes (Figure 2d). Since studying quantification accuracy was our goal here, 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, yeast: 2.03±0.45, human: 2.11±0.53, mean fold change ± standard deviation) matched closely with the changes determined using OpenSWATH (water: 2.62±1.43, yeast: 2.02±0.44, human: 1.96±0.39). From this we computed a deviation from the theoretical value of 31.2 %, 1.0 % and 2.0 % and a coefficient of variation 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 were 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 *Streptococcus 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. First, we created a spectral library of *S. pyogenes* by combining the measurements of 10 fractions of the *S. pyogenes* proteome in DDA (shotgun) mode on the same instrument, providing an extensive coverage of the expressed *S. pyogenes* proteome, with 1322 proteins (out of 1905 ORFs) mapping to 20,027 proteotypic peptide precursors at 1 % peptide-spectrum match FDR (see Figure 3a).

Using OpenSWATH, we identified and quantified 927 proteins (out of 1322 targeted proteins) of *S. pyogenes* consistently in each of the four LC-MS/MS runs at 1 % FDR. Out of these, 767 proteins were quantified by more than one peptide per protein. Thus, we achieved over 70 % coverage of the expressed proteome spanning more than three orders of dynamic range in estimated protein ion count (see Figure 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) [28, 33]. The fraction of the assay library which 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 differences in abundance between the two conditions in two biological replicates (see Figure 3 c-d and Supplementary Note 7 for a complete list). 10 out of 13 proteins associated with fatty acid biosynthesis (FAB) are significantly downregulated, consistent with results of previous studies on *S. pyogenes* [33]. As expected, we also found several known virulence factors to be upregulated (e.g. HasA, HasB, Slo, SpeC and CovR) [35, 36]. Additionally, we observed downregulation of an ABC transporter complex for inorganic phosphate import (PstB1, PstB2 and PstS), as well as significant upregulation of six proteins involved in pyrimidine biosynthesis (PyrF, PyrD, PyrE, PyrB, PyrR and Upp). While 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 datasets analyzed with OpenSWATH are consistent with many previous suppositions about bacterial virulence but additionally are able to provide the foundation for new hypotheses (see Supplementary Note 7).

By combining the most advanced DIA technology with a software capable of analyzing the resulting complex datasets, we were able to significantly scale-up the targeted proteomic approach described earlier in Gillet *et al.* and show that targeted analysis of DIA data allows

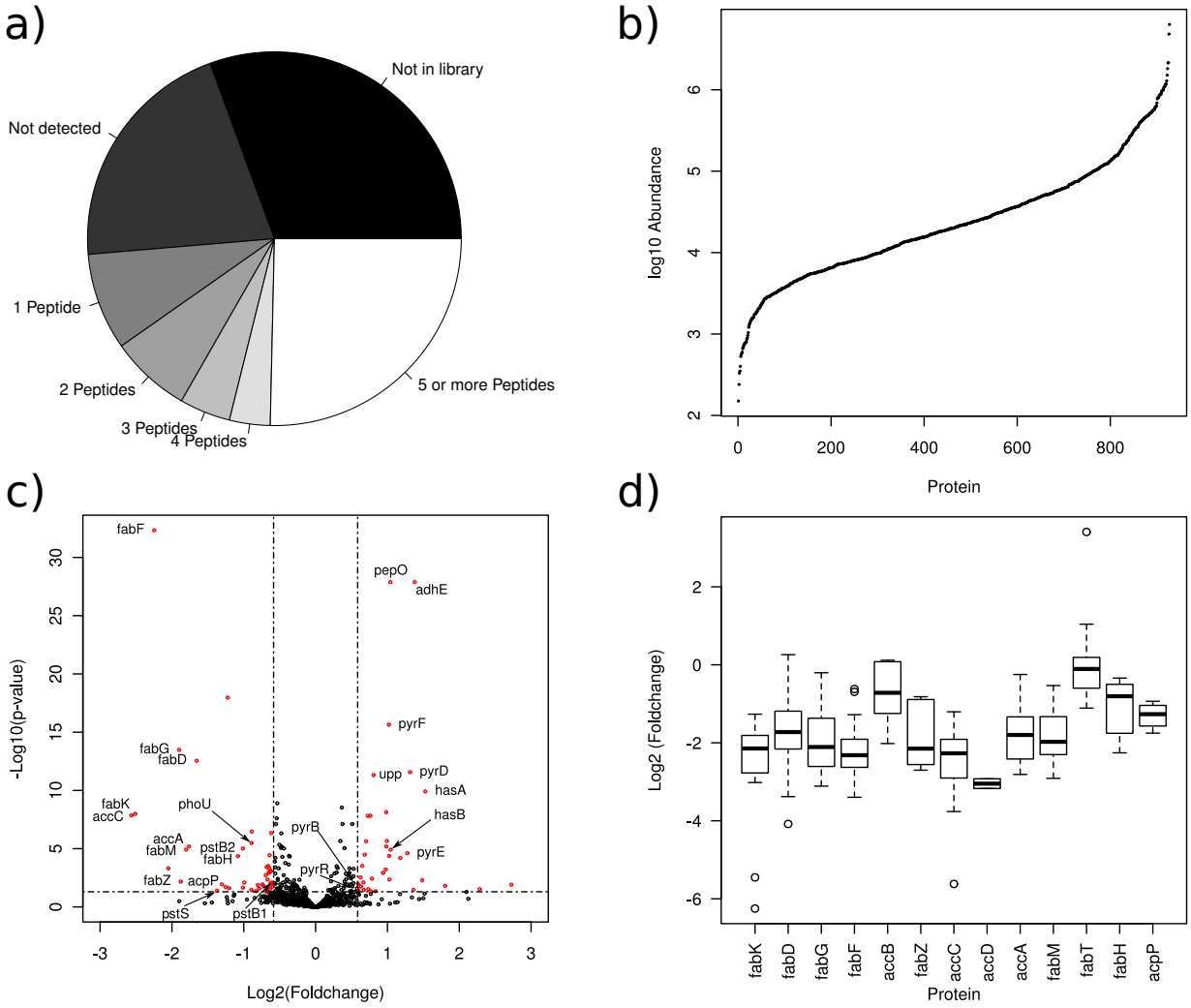

FIG. 3: ***Streptococcus pyogenes* exposed to human plasma (0 % plasma *vs.* 10 % plasma).** Analysis of two biological replicates with OpenSWATH at 1 % assay FDR yields over 900 proteins and 6000 peptides consistently quantified over four SWATH-MS runs. **a)** Proteome coverage of *S. pyogenes*: of 1905 annotated ORFs, 1322 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 *vs.* log-p-value) of protein expression determined by ANOVA analysis on two biological replicates. Red dots indicates 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 significantly downregulated except accD, accB and fabT (where fabT is a transcriptional repressor and not expected to be downregulated).

high-throughput analysis of microbial whole cell lysates, as demonstrated on the example of *S. pyogenes*. Using the SGS validation dataset 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` and is also packaged within OpenMS [37], which will make targeted DIA data analysis immediately accessible to a large research community. Due to the nature of DIA data, which contain a complete record of all fragment ions of a biological sample, reanalysis of a dataset is possible completely *in silico* which allows researchers to re-query data with their specific hypothesis in mind. With the availability of fast DIA-capable instruments, assay libraries, available in proteome-wide coverage due to large-scale peptide synthesis efforts and now, an automated software for DIA targeted data analysis, all pieces for a successful and more widespread use of this powerful technology are now available.

Contributions

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

Acknowledgements

H.R. was funded by ETH (ETH-30 11-2), G.R. & S.M. were funded by the Swiss Federal Commission for Technology and Innovation CTI (13539.1 PFFLI-LS), P.N. & L.G. & R.A. were funded by ERC Proteomics v3.0 (233226), L.G. & R.A. were funded by PhosphonetX project of SystemsX.ch, JM was funded by the Swedish Research Council (project 2008-3356), the Crafoord Foundation (20100892) and the Swedish Foundation for Swedish Research (FFL4). Further funding was provided to R.A. by the Swiss National Science Foundation (SNSF).

We would like to thank 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.

Competing financial interests

J.M. is co-founder and board member of Biognosys AG. J.M. and R.A. hold shares of Biognosys AG, which operates in the field covered by the article. The research group of R.A. is supported by AB SCIEX by providing access to prototype instrumentation.

Supplementary information and Data

Mass spectrometry data are available at the PeptideAtlas raw data repository (accession number PASS00289).

We further provide the following supplementary information and data:

- Supplementary Text
- Table 1 with fold changes of all quantified proteins of *S. pyogenes*
- Table 2 with all SIS peptide sequences in the SGS samples
- Data File 3 with all manual and OpenSWATH results for the SGS dataset
- Data File 4 Assay library for the SGS SIS peptides
- Data File 5 Assay library for the *S. pyogenes* peptides
- Data File 6 ini file for the OpenSWATH TOPP-Tools
- Data File 7 OpenSWATH TOPPAS workflow for OpenMS 1.10

[1] R. Aebersold and M. Mann, "Mass spectrometry-based proteomics," *Nature*, vol. 422, pp. 198–207, Mar. 2003.

[2] B. Domon and R. Aebersold, "Options and considerations when selecting a quantitative proteomics strategy.," *Nature Biotechnology*, vol. 28, pp. 710–721, July 2010.

[3] V. Lange, P. Picotti, B. Domon, and R. Aebersold, "Selected reaction monitoring for quantitative proteomics: a tutorial.," *Molecular Systems Biology*, vol. 4, pp. 222+, Oct. 2008.

[4] B. Domon and R. Aebersold, "Mass Spectrometry and Protein Analysis," *Science*, vol. 312, pp. 212–217, Apr. 2006.

[5] J. Sherman, M. J. McKay, K. Ashman, and M. P. Molloy, "Unique ion signature mass spectrometry, a deterministic method to assign peptide identity.," *Molecular & cellular proteomics : MCP*, vol. 8, pp. 2051–2062, Sept. 2009.

[6] H. Röst, L. Malmström, and R. Aebersold, "A computational tool to detect and avoid redundancy in selected reaction monitoring.," *Molecular & Cellular Proteomics*, vol. 11, pp. 540–549, Aug. 2012.

[7] A. Michalski, J. Cox, and M. Mann, "More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS.," *Journal of proteome research*, vol. 10, pp. 1785–1793, Apr. 2011.

[8] P. Picotti, B. Bodenmiller, L. N. Mueller, B. Domon, and R. Aebersold, "Full dynamic range proteome analysis of S. cerevisiae by targeted proteomics.," *Cell*, vol. 138,

pp. 795–806, Aug. 2009.

[9] S. Purvine, J.-T. T. Eppel, E. C. Yi, and D. R. Goodlett, "Shotgun collision-induced dissociation of peptides using a time of flight mass analyzer.," *Proteomics*, vol. 3, pp. 847–850, June 2003.

[10] J. D. Venable, M.-Q. Dong, J. Wohlschlegel, A. Dillin, and J. R. Yates, "Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra," *Nature Methods*, vol. 1, pp. 39–45, Sept. 2004.

[11] R. S. Plumb, K. A. Johnson, P. Rainville, B. W. Smith, I. D. Wilson, J. M. Castro-Perez, and J. K. Nicholson, "UPLC/MS(E); a new approach for generating molecular fragment information for biomarker structure elucidation.," *Rapid. Commun. Mass. Spectrom.*, vol. 20, no. 13, pp. 1989–1994, 2006.

[12] A. Panchaud, A. Scherl, S. A. Shaffer, P. D. von Haller, H. D. Kulasekara, S. I. Miller, and D. R. Goodlett, "Precursor acquisition independent from ion count: how to dive deeper into the proteomics ocean.," *Analytical Chemistry*, vol. 81, pp. 6481–6488, Aug. 2009.

[13] A. Panchaud, S. Jung, S. A. Shaffer, J. D. Aitchison, and D. R. Goodlett, "Faster, Quantitative, and Accurate Precursor Acquisition Independent From Ion Count," *Anal. Chem.*, vol. 83, pp. 2250–2257, Feb. 2011.

[14] M. Bern, G. Finney, M. R. Hoopmann, G. Merrihew, M. J. Toth, and M. J. MacCoss, "Deconvolution of Mixture Spectra from Ion-Trap Data-Independent-Acquisition Tandem Mass Spectrometry," *Anal. Chem.*, vol. 82, pp. 833–841, Dec. 2009.

[15] J. Wong, A. Schwahn, and K. Downard, "ETISEQ - an algorithm for automated elution time ion sequencing of concurrently fragmented peptides for mass spectrometry-based proteomics," *BMC Bioinformatics*, vol. 10, pp. 244+, Aug. 2009.

[16] P. C. Carvalho, X. Han, T. Xu, D. Cociorva, M. da Gloria Carvalho, V. C. Barbosa, and J. R. Yates, "XDIA: improving on the label-free data-independent analysis," *Bioinformatics*, vol. 26, pp. 847–848, Mar. 2010.

[17] S. J. Geromanos, J. P. C. Vissers, J. C. Silva, C. A. Dorschel, G.-Z. Li, M. V. Gorenstein, R. H. Bateman, and J. I. Langridge, "The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS," *Proteomics*, vol. 9, pp. 1683–1695, Mar. 2009.

[18] G.-Z. Li, J. P. C. Vissers, J. C. Silva, D. Golick, M. V. Gorenstein, and S. J. Geromanos, "Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures," *Proteomics*, vol. 9, pp. 1696–1719, Mar. 2009.

[19] K. Blackburn, F. Mbeunkui, S. K. Mitra, T. Mentzel, and M. B. Goshe, "Improving Protein and Proteome Coverage through Data-Independent Multiplexed Peptide Fragmentation," *J. Proteome Res.*, vol. 9, pp. 3621–3637, May 2010.

[20] X. Huang, M. Liu, M. J. Nold, C. Tian, K. Fu, J. Zheng, S. J. Geromanos, and S.-J. Ding, "Software for Quantitative Proteomic Analysis Using Stable Isotope Labeling and Data Independent Acquisition," *Anal. Chem.*, vol. 83, pp. 6971–6979, Aug. 2011.

[21] T. Geiger, J. Cox, and M. Mann, "Proteomics on an Orbitrap Benchtop Mass Spectrometer Using All-ion Fragmentation," *Molecular & Cellular Proteomics*, vol. 9, pp. 2252–2261, Oct. 2010.

[22] L. C. Gillet, P. Navarro, S. Tate, H. Röst, N. Selevsek, L. Reiter, R. Bonner, and R. Aebersold, "Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis.," *Molecular & Cellular Proteomics*, vol. 11, June 2012.

[23] D. C. Ince, L. Hatton, and J. Graham-Cumming, "The case for open computer programs," *Nature*, vol. 482, pp. 485–488, Feb. 2012.

[24] L. Martens, M. Chambers, M. Sturm, D. Kessner, F. Levander, J. Shofstahl, W. H. Tang, A. Rompp, S. Neumann, A. D. Pizarro, L. Montecchi-Palazzi, N. Tasman, M. Coleman, F. Reisinger, P. Souda, H. Hermjakob, P. A. Binz, and E. W. Deutsch, "mzML–a Community Standard for Mass Spectrometry Data," *Molecular & Cellular Proteomics*, vol. 10, pp. R110.000133–1, Dec. 2010.

[25] E. W. Deutsch, M. Chambers, S. Neumann, F. Levander, P.-A. Binz, J. Shofstahl, D. S. Campbell, L. Mendoza, D. Ovelleiro, K. Helsens, L. Martens, R. Aebersold, R. L. Moritz, and M.-Y. Brusniak, "TraML—A Standard Format for Exchange of Selected Reaction Monitoring Transition Lists," *Molecular & Cellular Proteomics*, vol. 11, Apr. 2012.

[26] C. Escher, L. Reiter, B. MacLean, R. Ossola, F. Herzog, J. Chilton, M. J. MacCoss, and O. Rinner, "Using iRT, a normalized retention time for more targeted measurement of peptides.," *Proteomics*, vol. 12, pp. 1111–1121, Apr. 2012.

[27] L. Reiter, O. Rinner, P. Picotti, R. Huttenhain, M. Beck, M.-Y. Brusniak, M. O. Hengartner, and R. Aebersold, "mProphet: automated data processing and statistical validation for large-scale SRM experiments," *Nature Methods*, vol. 8, pp. 430–435, May 2011.

[28] L. Malmström, J. Malmström, N. Selevsek, G. Rosenberger, and R. Aebersold, "Automated Workflow for Large-Scale Selected Reaction Monitoring Experiments," *J. Proteome Res.*, vol. 11, pp. 1644–1653, Jan. 2012.

[29] H. Wenschuh, R. Volkmer-Engert, M. Schmidt, M. Schulz, J. Schneider-Mergener, and U. Reineke, "Coherent membrane supports for parallel microsynthesis and screening of bioactive peptides.," *Biopolymers*, vol. 55, no. 3, pp. 188–206, 2000.

[30] K. Hilpert, D. F. Winkler, and R. E. Hancock, "Peptide arrays on cellulose support: SPOT synthesis, a time and cost efficient method for synthesis of large numbers of peptides in a parallel and addressable fashion.," *Nature Protocols*, vol. 2, pp. 1333–1349, June 2007.

[31] B. MacLean, D. M. Tomazela, N. Shulman, M. Chambers, G. L. Finney, B. Frewen, R. Kern, D. L. Tabb, D. C. Liebler, and M. J. MacCoss, "Skyline: an open source document editor for creating and analyzing targeted proteomics experiments," *Bioinformatics*, vol. 26, pp. 966–968, Apr. 2010.

[32] J. E. Elias and S. P. Gygi, "Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry.," *Nature methods*, vol. 4, pp. 207–214, Mar. 2007.

[33] J. Malmström, C. Karlsson, P. Nordenfelt, R. Ossola, H. Weisser, A. Quandt, K. Hansson, R. Aebersold, L. Malmström, and L. Björck, "Streptococcus pyogenes in Human Plasma," *Journal of Biological Chemistry*, vol. 287, pp. 1415–1425, Jan. 2012.

[34] E. W. Deutsch, H. Lam, and R. Aebersold, "PeptideAt-

las: a resource for target selection for emerging targeted proteomics workflows.," *EMBO reports*, vol. 9, pp. 429–434, May 2008.

[35] P. R. Shea, S. B. Beres, A. R. Flores, A. L. Ewbank, J. H. Gonzalez-Lugo, A. J. Martagon-Rosado, J. C. Martinez-Gutierrez, H. A. Rehman, M. Serrano-Gonzalez, N. Fittipaldi, S. D. Ayers, P. Webb, B. M. Willey, D. E. Low, and J. M. Musser, "Distinct signatures of diversifying selection revealed by genome analysis of respiratory tract and invasive bacterial populations.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, pp. 5039–5044, Mar. 2011.

[36] H. Malke, K. Steiner, W. M. McShan, and J. J. Ferretti, "Linking the nutritional status of Streptococcus pyogenes to alteration of transcriptional gene expression: The action of CodY and RelA," *International Journal of Medical Microbiology*, vol. 296, pp. 259–275, Aug. 2006.

[37] M. Sturm, A. Bertsch, C. Gröpl, A. Hildebrandt, R. Hussong, E. Lange, N. Pfeifer, O. Schulz-Trieglaff, A. Zerck, K. Reinert, and O. Kohlbacher, "OpenMS - an open-source software framework for mass spectrometry.," *BMC bioinformatics*, vol. 9, no. 1, 2008.