Editorial Summary: MassIVE.quant is a data repository and data resource for reproducible quantitative mass spectrometry-based proteomics

**MassIVE.quant: a community resource of quantitative mass spectrometry-based proteomics datasets**

Meena Choi¹, Jeremy Carver², Cristina Chiva³,⁴, Manuel Tzouros⁵, Ting Huang¹, Tsung-Heng Tsai¹, Benjamin Pullman², Oliver M. Bernhardt⁶, Ruth Hüttenhain⁷, Guo Ci Teo⁸, Yasset Perez-Riverol⁹, Jan Muntel⁶, Maik Müller¹⁰, Sandra Goetze¹⁰,¹¹, Maria Pavlou¹⁰, Erik Verschueren⁷, Bernd Wollscheid¹⁰,¹¹, Alexey I. Nesvizhskii⁸, Lukas Reiter⁶, Tom Dunkley⁵, Eduard Sabidó³,⁴, Nuno Bandeira²,* Olga Vitek¹,*

¹ Khoury College of Computer Sciences, Northeastern University, Boston, MA, USA
² Department of Computer Science and Engineering, University of California, San Diego, CA, USA
³ Proteomics Unit, Center for Genomics Regulation, The Barcelona Institute of Science and Technology, Barcelona, Spain
⁴ Proteomics Unit, Universitat Pompeu Fabra, Barcelona, Spain
⁵ Roche Pharma Research and Early Development, Pharmaceutical Sciences, Roche Innovation Center Basel, Hoffmann-La Roche Ltd, Basel, Switzerland
⁶ Biognosys, Zurich, Switzerland
7 Department of Molecular and Cellular Pharmacology, University of California, San Francisco, CA, USA

8 Department of Pathology, University of Michigan, Ann Arbor, MI, USA

9 Proteomics Services, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK.

10 Department of Health Sciences and Technology, Institute of Translational Medicine, ETH, Zurich, Switzerland

11 Swiss Institute of Bioinformatics, Lausanne, Switzerland

Present address

360 Huntington Ave. Boston, Massachusetts, USA, 02115, Northeastern University

Correspondence (marked as *)

e-mail: O.V. (o.vitek@northeastern.edu) or N.B. (bandeira@ucsd.edu)
Abstract

MassIVE.quant is a repository infrastructure and data resource for reproducible quantitative mass spectrometry-based proteomics, which is compatible with all mass spectrometry data acquisition types and computational analysis tools. A branch structure enables MassIVE.quant to systematically store raw experimental data, metadata of the experimental design, scripts of the quantitative analysis workflow, intermediate input and output files, as well as alternative reanalyses of the same dataset.
Main

Quantitative mass spectrometry data analysis currently has multiple unmet reproducibility goals\(^1\). At minimum, the mass spectrometry-based workflows must provide enough information to enable its full independent replication\(^2\). Beyond that, conclusions of data analysis should not be dependent upon particular tuning parameters or software tools. Data analysis should demonstrate that alternative and equally appropriate parameter settings or software lead to qualitatively similar conclusions.

In mass spectrometry-based proteomics, data analysis is broadly categorized into peptide ion identification and quantification. Much progress in identification has been made in terms of open availability of tools and transparency of their algorithms. Archival resources MassIVE, PRIDE\(^3\), Panorama\(^4,5\), PASSEL\(^6\) component of Peptide Atlas, and jPOST in ProteomeXchange\(^7,8\) store raw data, peak lists, search engine output, identification results, and corresponding mass spectra.

Unfortunately, reproducibility and transparency of data analysis for relative protein quantification is less satisfactory. First, given the great diversity of biological objectives and experiments, quantitative analyses require richer metadata describing experimental design and biological samples. Second, quantitative experiments require many data processing steps, which are distinct from similar steps in quantitative transcriptomic investigations. These include detection and identification of chromatographic peaks and reporter ions, and propagating those identities across multiple runs. Finally, existing analysis tools (such as Skyline\(^9\), MaxQuant\(^10\), OpenMS\(^11\), OpenSWATH\(^12\), DIA-Umpire\(^13\), Proteome Discoverer, or Spectronaut\(^14\)) integrate, in their own unique ways, diverse functionalities for identification and quantification. These tools offer various
parameters and options, and output different details in various storage formats. Many tools also offer graphical user interfaces, for which analyses are difficult to document.

Benchmarking of individual analysis strategies and tools for quantitative proteomics workflows has become increasingly prevalent\textsuperscript{15}, but lack infrastructure to store, document, annotate and reanalyze the full diversity of analyses.

To meet these reproducibility needs, we implemented MassIVE.quant, an infrastructure that supports quantitative mass spectrometry-based proteomics experiments. MassIVE.quant is integrated with an existing repository, the Mass Spectrometry Interactive Virtual Environment (MassIVE).

MassIVE.quant systematically stores the intermediate output files of every tool and workflow in a way which allows the user to easily inspect, reproduce or modify any component of the workflow, beginning with well-defined intermediate files. To accomplish this, we first developed a series of steps which represent a quantitative proteomics experiment with any experimental design, data acquisition, and data analysis tools (Figure 1). These steps consist of (1) annotations of experimental design; in particular, descriptors of biological samples and conditions; (2) strategies of sample preparation and data acquisition; (3) peptide ions identification; (4) quantification; and (5) statistical analysis. At each step, MassIVE.quant provides the infrastructure to store all intermediate descriptions, annotations, analysis scripts, and results.

MassIVE.quant does not prescribe a standard format, but meets scientists where they are by
directly accommodating the diverse nature of existing workflows. Each dataset contains links to the original publications or to metadata, which can be used to gain deeper insight into the biological context of the experiment. While analyte identification is represented by existing standard formats (such as mzTab, thousands of examples available for MassIVE datasets) and some of these can represent some quantitative information (such as the quant section in mzTab files, e.g., the mzTab file for RMSV000000249.18), the output files produced by quantitative analyses tools can be of any nature. None of the formats mentioned above support reporting of the results of statistical analyses of quantitative data. We chose a tabular format (i.e., csv) as a common representation of the output of quantitative and statistical analyses for all tools. This format emphasizes biologically relevant aspects of the output, such as the identity of differentially abundant peptides or proteins, the magnitudes of fold changes, and the associated variation.

Next, at each step, a branch structure enables the user to view reanalyses of each experiment. The reanalyses can be performed by the user offline with any combination of software tools and settings. MassIVE.quant stores the intermediate files and allows the user to check for the presence of script files, accuracy of parameters and completeness of documentation.

In order to scale the submission procedure and to ensure the reproducibility of a quantitative workflow, MassIVE.quant maintains datasets with four levels of curation (Bronze, Silver, Gold, and Platinum), reflecting the documentation and the reproducibility of the quantitative workflow. During the submission, the infrastructure checks whether the submission of the dataset or reanalysis meets the minimal requirements for the entry-level of curation. The submitter can then request the advanced review to level up.
MassIVE.quant further automates the statistical analysis of quantified proteins with an online MSstats workflow accessible with a user-friendly interface. This workflow can be used to reproduce the statistical analysis steps in MassIVE.quant reanalyses, as well as to analyze new private or public datasets. The MSstats Comparison workflow in MassIVE.quant automatically compares MSstats outputs across alternative reanalyses, and produces figures such as Figure 2(o). This enables the user to evaluate the implications of alternative reanalyses on conclusions regarding differential protein abundance.

The online user documentation clearly describes the structure and the vocabulary used by MassIVE.quant, and provides detailed instructions for contributing data, reanalyses and comparisons (Supplementary Note 1). Supplementary Figure 1 and 2, Supplementary Table 1 describe the infrastructure of MassIVE.quant and gives an example of the reanalysis submission workflow.

The impact of the choice between various analysis options is best understood in the presence of some notion of ground truth. Therefore, we populated MassIVE.quant with a collection of ten datasets with controlled mixtures with known changes in protein abundance. These include three datasets collected using label-free DDA, four datasets collected using DIA, one dataset collected using SRM with heavy labeled isotope peptides, and two DDA dataset collected using chemical labeling (TMT) (Supplementary Table 2). These datasets vary in background proteomes, number of conditions and replicates, and type and number of differentially abundant proteins.
We also populated MassIVE.quant with 95 reanalyses of these ten controlled datasets using multiple software tools, performed by the developers of the tools or by expert users. All the DDA experiments in Figure 1 were processed with up to six tools for identification and four tools for quantification. For example, data from Choi et al.\textsuperscript{16} (DDA:Choi2017) were processed with eight different combinations of parameter settings in Skyline. The combinations of algorithms, tools, and settings generated 10 distinct quantification reports. Finally, up to five different types of downstream statistical analysis per dataset using MSstats\textsuperscript{17} generated 22 distinct tests for differential protein abundance for DDA:Choi2017.

To demonstrate the utility of documentation, reanalysis, and curation in basic biology investigations, we further populated MassIVE.quant with a collection of biological datasets, at the time of publication, including 8 DIA/SWATH, 7 SRM, 12 DDA, and 6 experiments collected DDA-TMT acquisition, analyzed with multiple tools; 25 datasets with platinum level of curation, 18 datasets with gold (Supplementary Table 3). For example, the DIA experiment by Selevsek et al.\textsuperscript{18} (DIA:Selevsek2015) was reanalyzed four times using different analysis strategies and different processing tools and parameter settings. Figure 2, Supplementary Table 4, Supplementary Figure 3 and 4 illustrate how changes in data processing propagated themselves into discrepancies in the number of quantified proteins, frequency of missing values, and lists of differentially abundant proteins. Figure 2(a-d) illustrate these discrepancies in the special case of one protein. The analysis strategies and processing tools affected protein-level summaries in terms of scale, variation, and patterns of missing values. This in turn impacted the estimates of fold changes (Figure 2(e-h)) and tests for differential abundance (Figure 2(i-l)). Analysis with filtering in Skyline, applied to limit the DIA features to those known to be informative \emph{a priori}
(Skyline:lowCV, RMSV00000251.1) detected a smoother, and therefore more biologically plausible, pattern of differential abundance in time (Figure 2(m)). While the true differential abundance is unknown, changes identified by the majority of tools are more likely to be real (Figure 2(n-o)). Such comparisons help curate the results of biological investigations.

To summarize, MassIVE.quant provides an opportunity for large-scale deposition of heterogeneous experimental datasets and facilitates a community-wide conversation about the benefits of its use. We hope that the community will find the resource useful and welcome user-driven submissions of both new datasets and documented reanalyses of the existing datasets.

Acknowledgements

This work was supported in part by NSF CAREER award DBI-1054826, NSF DBI-1759736 and the Chan-Zuckerberg foundation to O.V., NIH-NLM 1R01LM013115 to N.B. and O.V., NSF award ABI 1759980, NIH awards P41GM103484, and R24GM127667 to N.B., and the Personalized Health and Related Technologies (PHRT 0-21411-18) strategic focus area of ETH to B.W. The CRG/UPF Proteomics Unit is part of the Spanish Infrastructure for Omics Technologies (ICTS OmicsTech) and it is a member of the ProteoRed PRB3 consortium which is supported by grant PT17/0019 of the PE I+D+i 2013-2016 from the Instituto de Salud Carlos III (ISCIII) and ERDF. We acknowledge support from the Spanish Ministry of Science, Innovation and Universities, “Centro de Excelencia Severo Ochoa 2013-2017”, SEV-2012-0208, and “Secretaria d’Universitats i Recerca del Departament d’Economia i Coneixement de la Generalitat de Catalunya” (2017SGR595). This project has received funding from the European Union's Horizon
2020 research and innovation program under grant agreement No 823839 (EPIC-XS). Y.P.-R. acknowledges the Wellcome Trust (grant number 208391/Z/17/Z). We thank the MacCoss lab (Department of Genome Sciences, University of Washington) for the Skyline analyses and contributing the processed data, the Slavov lab (College of Engineering, Northeastern University) for providing the data, and the Guo lab (School of Life Sciences, Westlake University, China) for providing the data.

Author contributions
M.C., J.C., N.B., O.V. designed the research. M.C., T.H. collected datasets and performed statistical analysis. J.C., B.P. implemented MassIVE.quant. T.T. performed statistical analysis. C.C., E.S., M.T. experimented with new controlled mixtures. C.C., M.T., R.H., G.C.T., Y.P.-R., J. M., M. M., S.G., M.P., E.V., B.W., O.M.B., A.I.N., L.R., T.D., E.S. analyzed data up to quantification. M.C., N.B., O.V. wrote the manuscript, with input from all authors.

Competing financial interests
O.M.B., J.M. and L.R. are employees of Biognosys AG. Spectronaut is a trademark of Biognosys AG. M.T. and T.D. are employees of Hoffmann-La Roche Ltd. All other authors declare no competing financial interests.
References

1. Peng, R. D. Reproducible research in computational science. *Science* **334**, 1226-1227, doi:10.1126/science.1213847 (2011).

2. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018, doi:10.1038/sdata.2016.18 (2016).

3. Perez-Riverol, Y. *et al.* The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* **47**, D442-D450, doi:10.1093/nar/gky1106 (2019).

4. Sharma, V. *et al.* Panorama: a targeted proteomics knowledge base. *J Proteome Res* **13**, 4205-4210, doi:10.1021/pr5006636 (2014).

5. Sharma, V. *et al.* Panorama Public: A Public Repository for Quantitative Data Sets Processed in Skyline. *Mol Cell Proteomics* **17**, 1239-1244, doi:10.1074/mcp.RA117.000543 (2018).

6. Farrah, T. *et al.* PASSEL: the PeptideAtlas SRMexperiment library. *Proteomics* **12**, 1170-1175, doi:10.1002/prmic.201100515 (2012).

7. Vizcaino, J. A. *et al.* ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat Biotechnol* **32**, 223-226, doi:10.1038/nbt.2839 (2014).

8. Deutsch, E. W. *et al.* The ProteomeXchange consortium in 2017: supporting the cultural change in proteomics public data deposition. *Nucleic Acids Res* **45**, D1100-D1106, doi:10.1093/nar/gkw936 (2017).

9. MacLean, B. *et al.* Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966-968, doi:10.1093/bioinformatics/btq054 (2010).
Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-1372, doi:10.1038/nbt.1511 (2008).

Rost, H. L. *et al.* OpenMS: a flexible open-source software platform for mass spectrometry data analysis. *Nat Methods* **13**, 741-748, doi:10.1038/nmeth.3959 (2016).

Rost, H. L. *et al.* OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nat Biotechnol* **32**, 219-223, doi:10.1038/nbt.2841 (2014).

Tsou, C. C. *et al.* DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nat Methods* **12**, 258-264, 257 p following 264, doi:10.1038/nmeth.3255 (2015).

Bruderer, R. *et al.* Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Mol Cell Proteomics* **14**, 1400-1410, doi:10.1074/mcp.M114.044305 (2015).

Navarro, P. *et al.* A multicenter study benchmarks software tools for label-free proteome quantification. *Nat Biotechnol* **34**, 1130-1136, doi:10.1038/nbt.3685 (2016).

Choi, M. *et al.* ABRF Proteome Informatics Research Group (iPRG) 2015 Study: Detection of Differentially Abundant Proteins in Label-Free Quantitative LC-MS/MS Experiments. *J Proteome Res* **16**, 945-957, doi:10.1021/acs.jproteome.6b00881 (2017).

Choi, M. *et al.* MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics* **30**, 2524-2526, doi:10.1093/bioinformatics/btu305 (2014).
Selevsek, N. et al. Reproducible and consistent quantification of the Saccharomyces cerevisiae proteome by SWATH-mass spectrometry. *Mol Cell Proteomics* **14**, 739-749, doi:10.1074/mcp.M113.035550 (2015).

**Figure Legends**

**Figure 1**: Outline of MassIVE.quant repository structure, and reanalysis of three DDA-based experiments. Each step can be performed with multiple algorithms and software tools, generating tool-specific files in diverse formats. For the experiments in the figure, MassIVE.quant stores the intermediate outputs from combinations of algorithms and tools for peptide ion identification and quantification. For example, DDA:Choi2017 was processed with eight combinations of parameter settings in Skyline. Each reanalysis is saved with a unique reanalysis ID, prefixed by RMSV, under the experiment repository prefixed by MSV in MassIVE.quant.

**Figure 2**: Re-analyses of DIA:Selevsek2015, profiling changes in proteome abundance of *S. cerevisiae* over 6 time points: T0(0 min), T1(15 min), T2(30 min), T3(60 min), T4(90 min), T5(120 min), n=3 biologically independent samples per each time point, in response to osmotic stress *(RMSV000000251)* (a)-(d) Discrepancies of quantification of protein YKL096W across data processing tools. Gray lines: fragments reported by each tool. Red lines: protein quantification summarized by MSstats. (a) Skyline:lowCV used Skyline to quantify a subset of the fragments with low coefficient of variation; (b) Skyline:All used Skyline to quantify all detectable peptides,
with a maximum of 6 fragments each; (c) data processed by Spectronaut; (d) data processed by DIA-Umpire. (e)-(h) Discrepancies in detecting differential abundance for protein YKL096W across data processing tools, with statistical analysis by MSstats. Dark red dot, center for error bars: model-based estimates of log₂ fold change of protein abundance, as determined by MSstats. Error bars: 95% confidence intervals for the log₂ fold changes, as determined by MSstats. Stars(*) correspond to adjusted p-values < 0.05. (i)-(l) Volcano plots, summarizing differential abundance between T5 and T0. Dashed line: FDR=0.05; blue dots: significantly down-regulated proteins; red dots: significantly up-regulated proteins (counts are shown at the top left corner; other time points are shown in Supplementary Figure 3). (m) Number of differentially abundant proteins across all time points and all tools, FDR=0.05. (n) Venn diagram of differentially abundant proteins between two processing approaches by Skyline, comparing T5 versus T0. (o) Venn diagram of differentially abundant proteins across all tools, comparing T5 versus T0 (other time points are shown in Supplementary Figure 4).

**Online Methods**

**Manual and tutorials**

MassIVE.quant user manuals and video tutorials explain how to 1) create an account in MassIVE.quant, 2) upload files to MassIVE account via FTP, 3) submit quantification reanalysis to MassIVE.quant, 4) access reanalyses, 5) run MSstats workflow in MassIVE.quant, and 6) compare the results of statistical analysis by MSstats. User manuals are available at MassIVE.quant section in https://ccms-ucsd.github.io/MassIVEDocumentation/. The video tutorial is available in https://www.youtube.com/channel/UCPeNhYF1iabsoOITPZBc5Q
Data analysis for DIA:Selevsek2015

Skyline: All – All files for ProteomeXchange ID PXD001010 were downloaded, along with the forward and reversed FASTA and mzXML files (n = 46) used in the library peptide search (requested of the authors). Using Skyline (daily version 3.7.1.11571) first a spectral library was built using the iProphet score cut-off 0.0242 suggested in the paper as achieving 1% FDR at the PSM level, with 1,031 ambiguously matched peptides excluded, resulting in 82,439 unique peptides (104,993 entries). Because the Biognosys iRT standard peptides used were not included in the search, an iRT library was created from these files by adding all detected peptides with the iRT standards added as targets in Skyline and importing the mzXML files for MS1 extraction. The iRT values were then calculated using both the extracted peaks for the iRT standards and target peptide MS1 peaks where the peak contained a matching MS/MS ID, because the runs used fractionated samples and all peptides were not expected in all runs. For DIA, allowing unique peptides of length 7 to 45 resulting from semi-tryptic cleavage with up to 2 missed cleavages, with Carbamidomethyl (C) and optionally Oxidation (M), precursors of charge states 2, 3, or 4, from 400-1,200 m/z (the range covered by the DIA method), with 6 product transitions found in the library, of y or b type and 1 or 2 charge state (excluding y1, y2, b1, and b2). Chromatogram extraction was set to use TOF extraction at 18,000 resolving power with high-selectivity extraction applying to all MS/MS spectra within 10 minutes of predicted retention times using the iRT library. Importing the FASTA file and then removing duplicate peptides and empty proteins resulted in targets for 4,603 proteins, 68,910 peptides, 87,042 precursors, and 522,252 fragment ions, at 32% protein, 2.7% unique peptide FDR by reversed sequence decoy counting (decoys/targets). The protein FDR is likely overstated because the FASTA file contains only 6,717 protein sequences, which means as many as 2/3 of false peptides can be expected to occur in a true protein, while the
same is not true for detections of reversed peptides. Even at 10% protein FDR, however, this target set seemed to contain a higher error rate than we felt desirable. For these experiments, we decided to rebuild the library using the iProphet score cut-off 0.9, with 361 ambiguously matched peptides excluded, resulting in 64,501 unique peptides (84,245 entries). For our most inclusive method, we chose to include only fully-tryptic peptides and no variable modifications (dropping Oxidation (M)), which resulted in targets for 4,152 proteins, 36,889 peptides, 48,082 precursors, and 288,492 fragment ions, at 2.6% protein, 0.29% unique peptide FDR by reversed sequence decoy counting. An equal number of shuffled sequence decoys were generated for mProphet model generation. The 18 runs were then imported into the template an mProphet model trained and applied. The MSstats report was exported for further analysis.

Skyline:lowCV – Using Skyline (daily version 3.7.1.11571) and starting from the settings for the broadly inclusive test, we restored semi-tryptic cleavage and the variable modification Oxidation (M), expecting unstable peptides to be filtered by our first experiment. Importing the forward and reversed FASTA file resulted in targets for 4,246 proteins, 58,168 peptides, 74,314 precursors, and 445,884 fragment ions, at 4.5% protein, 0.34% unique peptide FDR by reversed sequence decoy counting. Reversed sequences, accounting for 193 proteins and 200 peptides, were left in the targets list to be carried through the subsequent “Reproducibility” experiment. An equal number of shuffled sequence decoys were generated for mProphet model generation. The runs (n = 8 – reported as n = 4 in the paper) acquired for the “Reproducibility” experiment were then imported into the template, an mProphet model trained and applied. Next, the targets were filtered for peptides detected at q value less than 0.01 in at least 4 (of 8) runs, and where the CV of the detected peak areas, median normalized, was less than 20%, resulting in targets for 2,212 proteins, 13,744 peptides, 18,910 precursors, and 113,460 fragment ions, at 1% protein, 0.18% unique peptide FDR.
by reversed sequence decoy counting. The remaining peptides comprised 249 of 2,367 –10.5% –
Oxidation (M), 679 of 18,479 – 3.7% – unmodified semi-tryptic, and 12,816 of 37,322 – 34.3% –
unmodified tryptic peptides. As expected, Oxidation (M) and semi-tryptic peptides made it through
this filter at much lower rates than unmodified fully tryptic peptides. An equal number of shuffled
sequence decoys of matching types were generated for mProphet model generation. The 18 runs
were then imported into the template an mProphet model trained and applied. The MSstats report
was exported for further analysis.

Spectronaut – A spectral library was generated using all available raw files from the original
publication using the Pulsar search engine integrated in Spectronaut 11(11.0.15038) with default
settings. The uniprot yeast reference proteome was used for the spectrum centric data analysis for
library generation. DIA data was then analyzed using default settings and exported using the built
in MSstats Report (v3.7.3) export schema.

DIA-Umpire – The raw files were converted to mzXML files with centroiding. The resulting
mzXML files were processed by the signal extraction (SE) module of DIA-Umpire to generate
pseudo-MSMS spectra. The generated pseudo-MS/MS spectra were searched using X! Tandem,
Comet and MSGF+ search engines. The output files from the search engines were further analyzed
by PeptideProphet and combined by iProphet. FDR filtering was done with PeptideProphet and
ProteinProphet. DIA-Umpire's Quant module was for the quantification analysis. The outputs for
all-level quantification (FragSummary, PeptideSummary, ProtSummary) were used in further
analysis.

**Statistical analysis for DIA:Selevsek2015**
R package MSstats v3.10.6 was used to pre-process the output from Skyline, Spectronaut, and DIA-Umpire before statistical analysis, to have protein quantification, and to perform differential abundance analysis. MSstats estimated log₂ fold change and the standard error by linear mixed effect model for each protein. To test two-sided null hypothesis of no changes in abundance, the model-based test statistics were compared to the Student T distribution with the degrees of freedom appropriate for each protein and each dataset. The resulting p-values were adjusted to control the False Discovery Rate with the method by Benjamini-Hochberg. Parameter settings as well as the R code used to analyze DIA:Selevsek2015 are available in reanalysis container, RMSV000000251 in MassIVE.quant.

**Data Availability**

All the datasets that support this study are publicly available in MassIVE.quant (https://massive.ucsd.edu/ProteoSAFe/static/massive-quant.jsp) with MassIVE and ProteomeXchange identifiers. Additionally, identifiers for all the datasets are listed in Supplementary Table 2, 3, and 5.
## Sample Preparation

- Data acquisition: MS Data Converter
  - Raw spectrum files: raw, wiff
  - Peak list files: mzXML, mzML, mgf

## Peptide Identification

- Identification: Comet, OMSSA, MS-GF+, X!Tandem, Mascot, Andromeda, ...
  - Sequence databases / Spectral library: .fasta or library file
  - Identification result files: mzIdentML, mzTab, pep.xml, msf, dat

## Quantification

- Quantification: MaxQuant, Skyline, OpenMS, OpenSWATH, DIA-Umpire, Spectronaut, SpectroMine, Progenesis QI, Proteome Discoverer, ...
  - Quantification file with peak intensities: txt, csv, tsv, mzTab
  - Tool-specific files: skd, sne, PDresult
  - Report for parameters setting

## Statistical Analysis

- R, Python, MatLab, Persus...
  - Annotation for experimental design
  - Statistical analysis result: txt, csv, pdf
  - Description or Codes: txt or script

## Data Submission

- MassIVE, PRIDE, Panorama, JPOST, ...

---

### Experiment

- DDA: Choi2017
  - MS-GF+
  - Comet
  - OMSSA
  - Andromeda

- DDA: Cox2014
  - MaxQuant
  - Skyline
  - Progenesis
  - Proteome Discoverer

- DDA: ControlledMix
  - Proteome Discoverer

---

### Data Submission

- RMSV000000261.4
- RMSV000000261.3
- RMSV000000261.2
- RMSV000000261.1
- RMSV000000255.3
- RMSV000000255.1
- RMSV000000255.2
- RMSV000000249.1
- RMSV000000249.9
- RMSV000000249.3
- RMSV000000249.2

### Data Submission Services

- MassIVE
- PRIDE
- Panorama
- JPOST

---

### R package, MSstats

RMSV000084181
RMSV000081831
RMSV000079843
MassIVE.quant: a community resource of quantitative mass spectrometry-based proteomics datasets

Supplementary Information

Table of Contents
Supplementary Note 1 : Links to the documentation and the video tutorials for MassIVE.quant .................................................2
Supplementary Figure 1 : Repository structure for submitting reanalyses in MassIVE.quant DDA:CHoi2017 was used as an example to show how reanalyses are organized in MassIVE.quant ..................................................3
Supplementary Figure 2 : Example of data analysis steps and corresponding input or output files .................................................4
Supplementary Table 1 : Categories for intermediate files for data processing tools .................................................................5
Supplementary Table 2 : List of benchmark experiments with controlled mixtures, and resulting datasets, available in MassIVE.quant ..................................................................................................................................6
Supplementary Table 3 : List of biological investigations in MassIVE.quant ....................................................................................7
Supplementary Table 3, continued : List of benchmark biological investigations in MassIVE.quant ..................................................8
Supplementary Table 4 : Summary for reanalyzed DIA:Selevsek2015 by three processing tools - Skyline, Spectronaut, DIA-Umpire. ......................................................................................................................................9
Supplementary Figure 3 : Volcano plots for pairwise comparisons between OMIN and others in DIA:Selevsek2015 .......10
Supplementary Figure 4 : Venn diagram of differentially abundant proteins in DIA:Selevsek2015 .................................11
Supplementary Table 5 : List of datasets for reference ................................................................................................................12
Supplementary references for datasets ........................................................................................................................................13
Supplementary Note 1: Links to the documentation and the video tutorials for MassIVE.quant

- MassIVE.quant website:
  [http://massive.ucsd.edu/ProteoSAFe/static/massive-quant.jsp](http://massive.ucsd.edu/ProteoSAFe/static/massive-quant.jsp)
  MassIVE.quant website contains a summary of the datasets in MassIVE.quant and links to the next steps of data exploration and reanalysis.

- MassIVE.quant user manuals: MassIVE.quant section in [https://ccms-ucsd.github.io/MassIVEDocumentation/](https://ccms-ucsd.github.io/MassIVEDocumentation/)

- MassIVE.quant video tutorials: [https://www.youtube.com/channel/UCPeNhYFlIabso0IiTPZBc5Q](https://www.youtube.com/channel/UCPeNhYFlIabso0IiTPZBc5Q)

User manuals and video tutorials explain how to:
1. Create an account in MassIVE.quant
2. Upload files to MassIVE account via FTP
3. Submit quantification reanalysis to MassIVE.quant: ‘Add analysis’ workflow
4. Access reanalyses
5. Run MSstats workflow in MassIVE.quant: ‘MSstats’ workflow
   a. For a new quantification, using offline data processing tools
   b. For a published quantification, using MassIVE.quant
6. Compare the results of statistical analysis by MSstats: ‘MSstats comparison’ workflow
Supplementary Figure 1: Repository structure for submitting reanalyses in MassIVE.quant

DDA:Choi2017 was used as an example to show how reanalyses are organized in MassIVE.quant. 'Dataset' is set up for each experiment or publication. For example, DDA:Choi2017, including all RAW files, metadata, and identifications, is stored as a 'Dataset' with an unique identifier, prefixed by MSV, which in this case is MSV00079843. MSV00079843 is the MassIVE dataset identifier for the DDA:Choi2017 experiment. Users can set up a 'Reanalyses Container' under each dataset (i.e., MassIVE dataset) to aggregate groups of related reanalyses, such as the MassIVE.quant sets of reanalyses for each dataset reported here. Each container has a unique reanalysis identifier (for example, RMSV000000249 and RMSV000000259) to capture container-specific metadata (e.g., related experiment) and to facilitate aggregating single-identifier references to related sets of reanalysis. Any number of reanalyses can be stored under each 'Reanalyses container'. Finally, each reanalysis also has a unique reanalysis identifier (such as RMSV000000249.1 and RMSV000000249.2) to enable direct references to individual reanalyses, regardless of the context of datasets to which each reanalysis is related.
Supplementary Figure 2: Example of data analysis steps and corresponding input or output files.
DDA:Cox2014, quantified by MaxQuant, was used for demonstration.
| Data processing tool | File type                | Which category in MassIVE.quant this file should be stored |
|----------------------|--------------------------|----------------------------------------------------------|
| All                  | Input.csv                | Quantification results                                    |
|                      | Annotation.csv           | Metadata                                                  |
|                      | testResult_byMSstats.csv | Statistical analysis of quantified analytes               |
|                      | msstats.log              | Methods and protocols                                     |
|                      | Rscript.R                | Methods and protocols                                     |
|                      | Description.pdf          | Methods and protocols                                     |
| MaxQuant             | evidence.txt             | Quantification results                                    |
|                      | proteinGroups.txt        | Quantification results                                    |
|                      | parameters.txt           | Quantification results                                    |
|                      | mqpar.xml                | Methods and protocols                                     |
|                      | fasta                    | Sequence database                                         |
| Spectronaut          | input.xls                | Quantification results                                    |
|                      | SpectronautAnalysis.sne  | Supplementary files                                       |
| DIAUmpire            | FragSummary.xls          | Quantification results                                    |
|                      | PeptideSummary.xls       | Quantification results                                    |
|                      | ProtSummary.xls          | Quantification results                                    |
| OpenSWATH            | input.tsv or input.txt   | Quantification results                                    |
| Skyline              | .sky.zip                 | Supplementary files                                       |

**Supplementary Table 1: Categories for intermediate files for data processing tools.**

All examples are quantification files and the output of MSstats for statistical analysis.

**Quantification results**: quantification result files from data processing tools/quantification tools. These should be used for downstream analysis (for example, PSM-level or transition-level or protein-level quantified peak intensities data).

**Metadata**: An annotation file should include the information about design of experiments and is submitted in the Metadata section. For example, it should contain information regarding biological replicate, technical replicate, fraction, condition or group for the corresponding MS runs, and corresponding channels.

Supplementary **files**: If the processing software provides software-specific files, it can be shared. For example, sky, skyd, sky.zip, blib, irtdb, protdb, optdb, and midas for Skyline; .pdresult for Proteome Discoverer; .sne for Spectronaut.

**Methods and protocols**: Any file types for any methods and protocols related to the corresponding reanalysis. For example, a text file to record parameter settings or extra steps, parameter files themselves, script file containing command lines for other statistical analysis tools or any downstream analysis. For example, R or Python script.

**Statistical analysis of quantified analytes**: Result files for statistical analysis. It could be a table or figure of results for statistical analysis. For example, table (.txt, .csv), r data format or pdf for figures.
Supplementary Table 2: List of benchmark experiments with controlled mixtures, and resulting datasets, available in MassIVE.quant

MassIVE.quant stores the analysis results of six previously published dataset (DDA:Choi2017, DDA:Cox2014, DIA:Navarro2016, DIA:Bruderer2015, DIA:Röst, DIA:Huang2020) and four new experiments (DDA:ControlledMix, SRM:ControlledMix, TMT:ControlledMix-MS3, TMT:ControlledMix-MS2) obtained by up to 95 combinations of processing methods, software tools, and parameter settings. In all cases, statistical analyses of the identified and quantified proteins were performed with MSstats. Four experiments (DDA:ControlledMix, DDA:Choi2017, DIA:Navarro2016, DIA:Bruderer2015) were reanalyzed with up to five different MSstats options.

| Dataset name          | Publication          | Type of acquisition | (Background) Organism | No. of tech replicates/condition | No. of conditions | Dataset ID in repository | Database search | Data processing tool | Reanalysis ID in MassIVE.quant |
|-----------------------|----------------------|---------------------|-----------------------|-------------------------------|------------------|---------------------------|----------------|----------------------|-----------------------------|
| DDA:ControlledMix     | Chiva et al., 2014   | DDA                 | E.coli                | 3                             | 5                | PXD005642 MSV000084181   | Andromeda       | Skyline              | RMSV000030261:1            |
|                       |                      |                     |                       |                               |                  |                           | MaxQuant        | RMSV000030261:2       |                             |
|                       |                      |                     |                       |                               |                  |                           | Proteome Discoverer, Mascot | RMSV000030261:3 |                             |
|                       |                      |                     |                       |                               |                  |                           | Proteome Discoverer | RMSV000030261:4     |                             |
|                       |                      |                     |                       |                               |                  |                           | + 16 reanalyses    | RMSV000000303:0      |                             |
| DDA:Choi2017          | Choi et al., 2017    | DDA                 | E.coli                | 3                             | 4                | PXD015300 MSV000079845   | Comet, OXSSA, MS-GH+ | Skyline              | RMSV000000249:0          |
|                       |                      |                     |                       |                               |                  |                           | Andromeda         | RMSV000000249:2       |                             |
|                       |                      |                     |                       |                               |                  |                           | MaxQuant          | RMSV000000249:3       |                             |
|                       |                      |                     |                       |                               |                  |                           | Comet             | RMSV000000249:4       |                             |
|                       |                      |                     |                       |                               |                  |                           | + 8 reanalyses     | RMSV000000029:7      |                             |
| DDA:Cox2014           | Cox et al., 2014     | DDA                 | E.coli                | 4                             | 2                | PXD0000279 MSV000081831  | Andromeda         | Skyline              | RMSV000000025:7          |
|                       |                      |                     |                       |                               |                  |                           | MaxQuant          | RMSV000000025:1       |                             |
|                       |                      |                     |                       |                               |                  |                           | Comet             | RMSV000000025:2       |                             |
|                       |                      |                     |                       |                               |                  |                           | + 12 reanalyses    | RMSV0000000025:9     |                             |
| DIA:Navarro2016       | Navarro et al., 2016 | DIA                 | Human / Yeast / E.coli| 3                             | 2                | PXD022952 MSV000081004   | Comet, Mascot     | Skyline              | RMSV000000250:1          |
|                       |                      |                     |                       |                               |                  |                           | Spectronaut       | RMSV000000250:2       |                             |
|                       |                      |                     |                       |                               |                  |                           | DIA/Limpire       | RMSV000000250:3       |                             |
|                       |                      |                     |                       |                               |                  |                           | + 16 reanalyses    | RMSV0000000025:7     |                             |
| DIA:Bruderer2015      | Bruderer et al., 2015| DIA                 | Human                 | 3                             | 8                | PASS00589 MSV000081838   | Pulsar           | Skyline              | RMSV000000025:2          |
|                       |                      |                     |                       |                               |                  |                           | Pulsar            | RMSV000000025:2       |                             |
|                       |                      |                     |                       |                               |                  |                           | Spectronaut       | RMSV000000025:2       |                             |
|                       |                      |                     |                       |                               |                  |                           | DIA/Limpire       | RMSV000000025:3       |                             |
|                       |                      |                     |                       |                               |                  |                           | + 12 reanalyses    | RMSV0000000025:8     |                             |
| DIA:Röst2014          | Röst et al., 2014    | DIA                 | Human                 | 3                             | 8                | PASS000258 MSV000081829  | -                | Skyline              | RMSV000000025:1          |
|                       |                      |                     |                       |                               |                  |                           | OpenSWATH         | RMSV000000025:2       |                             |
| DIA:Huang2020         | Huang et al., 2020   | DIA                 | Mouse                 | 5                             | 5                | PXD016647 MSV000084864   | Andromeda         | Skyline              | RMSV000000027:1          |
|                       |                      |                     | Human / C.elegans / S.cerevisiae | 3    | 10               |                           | Andromeda In MaxQuant | Spectronaut       | RMSV000000028:1         |
|                       |                      |                     |                       |                               |                  |                           | DIA/Limpire       | RMSV000000028:2       |                             |
|                       |                      |                     |                       |                               |                  |                           | + 8 reanalyses     | RMSV0000000028:3      |                             |
| SRM:ControlledMix     |                       | SRM                 | Yeast                 | 3                             | 5                | PXD015304 MSV000084220   | -                | MultiQuant           | -                          |
| TMT:ControlledMix-MS3 | New                  | TMT                 | Human                 | 4                             | 10               | PXD015328 MSV000084264   | Proteome Discoverer, Mascot | Proteome Discoverer       | RMSV00000000265:0       |
| TMT:ControlledMix-MS2 | New                  | TMT                 | Human                 | 4                             | 10               | PXD015261 MSV000084266   | Proteome Discoverer, Mascot | Proteome Discoverer       | RMSV00000000266:0       |
Supplementary Table 3: List of biological investigations in MassIVE.quant

The 33 published datasets with Gold or Platinum level of curation were currently available in MassIVE.quant. One experiment, DIA:Selevsek2015, was analyzed and quantified using four combinations of processing methods, software tools, and parameter settings. In all cases, differential abundance analysis of the quantified proteins were performed with MSstats or t-test in R. DIA:Selevsek2015 was analyzed using five different MSstats options. These information available for each dataset is sufficient to performing statistical analyses of differential abundance. Each dataset contains links to the publications or to metadata with additional descriptions provided by the original submitter.
| Dataset name | Publication | Type of acquisition | Organism | N of bio. replicates | N of conditions | Dataset ID in repository | Database search | Data processing tool | Reanalysis ID in MassIVE |
|-------------|-------------|---------------------|----------|----------------------|---------------|-------------------------|----------------|---------------------|-------------------------|
| SRM:Hüttenhain2019 | Hüttenhain et al., 2019 | SRM | Human | 234 | 2 | PXD001503 MSV000846048 | - | Skyline | - |
| SRM:Hüttenhain2019.2 | Hüttenhain et al., 2019 | DDA | Human | 4 | 5 | PXD0009012 MSV00084855 | Andromeda in MaxQuant | MaxQuant | - |
| | | SRM | Human/ | 4 | 5 | MSV00084962 | - | Skyline | - |
| | | SRM | Human | 4 | 4 | PXD017565 MSV00084963 | - | Skyline | - |
| | | SRM | E.coli | 4 | 4 | PXD017566 MSV00084964 | - | Skyline | - |
| | | SRM | E.coli | 4 or 5 | 4 | PXD017567 MSV00084965 | - | Skyline | - |
| SRM:Vertterli2017 | Vertterli et al., 2017 | DDA | Mouse | 3 | 12 | PXD005758 MSV00084857 | Andromeda in MaxQuant | MaxQuant | - |
| | | DDA | E.coli | 3 | 3 | PXD015048 MSV00084979 | - | Skyline | - |
| | | SRM | Human | 3 | 8 | PXD015049 MSV00084967 | - | Skyline | - |
| DDA:Luther2019 | Luther et al., 2019 | DDA | E.coli | 3 or 4 | 2 | PXD010174 MSV00084899 | Comet / TPP(Proteomic-Pipeline) | Progenesis | - |
| | | DDA | E.coli | 3 or 4 | 2 | PXD010174 MSV00084899 | Comet / TPP(Proteomic-Pipeline) | Progenesis | - |
| | | DDA | E.coli | 3 | 2 | PXD010174 MSV00084899 | Comet / TPP(Proteomic-Pipeline) | Progenesis | - |
| DDA:Jung2017 | Jung et al., 2017 | DDA | Mouse | 3 | 34 | PXD004263 MSV00084979 | Andromeda in MaxQuant | MaxQuant | - |
| DDA:Malmström2016 | Malmström et al., 2016 | DDA | Mouse | 1 | 13 | PXD002896 MSV00084979 | Andromeda in MaxQuant | MaxQuant | - |
| DDA:Meierhofer2016 | Meierhofer et al., 2016 | DDA | Mouse | 6 | 2 | PXD001555 MSV00084933 | Andromeda in MaxQuant | MaxQuant | - |
| DDA:Rakus2016 | Rakus et al., 2016 | DDA | Mouse | 5 | 3 | PXD0016125 MSV00084934 | Andromeda in MaxQuant | MaxQuant | - |
| DDA:Deshmukh2016 | Deshmukh et al., 2016 | DDA | Mouse | 3 | 2 | PXD001293 MSV00084938 | Andromeda in MaxQuant | MaxQuant | - |
| DDA:Guo2015 | Guo et al., 2015 | DDA | Mouse | 3 | 2 | PXD001293 MSV00084938 | Andromeda in MaxQuant | MaxQuant | - |
| DDA:Azimia2014 | Azimia et al., 2014 | DDA | Mouse | 3 | 21 | PXD000867 MSV00084937 | Andromeda in MaxQuant | MaxQuant | - |
| DDA:Hornburg2014 | Hornburg et al., 2014 | DDA | Mouse | 3 | 6 | PXD000866 MSV00084933 | Andromeda in MaxQuant | MaxQuant | - |
| DDA:Zhang2014 | Zhang et al., 2014 | DDA | Human | 5 | 4 | PXD015905 MSV00084238 | Peptide / MyriMatch / MS-GF+ | IDPicker | - |

Supplementary Table 3, continued: List of benchmark biological investigations in MassIVE.quant
Supplementary Table 4: Summary for reanalyzed DIA:Selesevk2015 by three processing tools - Skyline, Spectronaut, DIA-Umpire.

Two different processing steps were performed in Skyline. Detailed description of processing steps for four reanalyses are available in MassIVE.quant RMSV000000251. ‘No. of Proteins: in report or output from tool’ is the number of protein ids reported from data processing tools. ‘No. of Proteins after preprocessing by converter in MSstats’ is the number of proteins available the output of converter function in MSstats. ‘No. of Proteins possible to test’ is the number of proteins that can be fitted in a linear mixed model. It is less than ‘No. of Proteins’ because some proteins cannot be fitted and analyzed due to missingness issue. ‘Proportion of missing’ = the number of NA or normalized intensities less than or equal to 1 / total number of intensities. ‘Total No. Intensity’ = the number of unique features times the number of MS runs.

| Software or tool for quantification | Skyline | Spectronaut | DIA-Umpire |
|-----------------------------------|---------|-------------|------------|
| Different processing steps        | low CV  | all         |            |
| No. of Proteins:                  |         |             |            |
| - in report or output from tool   | 2,213   | 4,154       | 3,019      | 2,247      |
| - after preprocessing by converter in MSstats | 2,212 | 2,189 | 2,512 | 2,030 |
| No. of Proteins possible to test  | 2,205   | 1,944       | 2,499      | 1,870      |
| No. of Peptides per protein:      |         |             |            |
| - median                          | 4       | 4           | 7          | 6          |
| - range                           | 1-145   | 1-157       | 1-227      | 1-202      |
| No. of Fragments per peptide:     |         |             |            |
| - median                          | 6       | 6           | 3          | 6          |
| Proportion of missing:            |         |             |            |
| No. of NA + No. of Intensity ≤ 1  | 31,688/ | 500,627/    | 291,629/   | 542,632/   |
| / total No. Intensity             | 0.0155  | 0.2207      | 0.1342     | 0.2537     |
Supplementary Figure 3: Volcano plots for pairwise comparisons between 0min and others in DIA:Selevesek2015

The plots summarize and show discrepancies of differential abundance between T5 and T0 across data processing tools, with identical statistical analysis using MSstats. To test two-sided null hypothesis of no changes in abundance, the model-based test statistics were compared to a Student T distribution. The p-values were adjusted to control the False Discovery Rate with the method by Benjamini-Hochberg. Each row is the time point comparison from starting point (0 min). Each column shows reanalysis by a different processing tool or step. Four reanalyses were performed by three spectral processing tools: Skyline:lowCV (first column) and Skyline:All (second column), Spectronaut (third column), and DIA-Umpire (fourth column). Detailed descriptions of processing steps for the four reanalyses are available in MassIVE.quant RMSV000000251. Dashed horizontal line: FDR=0.05, blue dots: significantly down-regulated proteins, red dots: significantly up-regulated proteins. The numbers of significantly up- or down-regulated proteins are shown top left corner of each volcano plot.
Supplementary Figure 4: Venn diagram of differentially abundant proteins in DIA:Selevsek2015 (a, c, e, g) between two processing approaches by Skyline (Skyline:lowCV and Skyline:All), (b, d, f, h) across three processing tools (Skyline:lowCV, Spectronaut, DIA-Umpire). Each row is the time point comparison from starting point (0 min).
| Name                   | Controlled mixtures          | Type of acquisition | ID in repository       |
|------------------------|-----------------------------|---------------------|-------------------------|
| DDA:ControlledMix      | Chiva et al., 2014<sup>1</sup> | DDA                 | MSV000084181 (PXD005642) |
| DDA:Choi2017           | Choi et al., 2017<sup>2</sup> | DDA                 | MSV000079843 (PXD015300) |
| DDA:Cox2014            | Cox et al., 2014<sup>3</sup> | DDA                 | MSV000081831 (PXD000279) |
| DIA:Navarro2016        | Navarro et al., 2016<sup>4</sup> | DIA                 | MSV000081024 (PXD002952) |
| DIA:Bruderer2015       | Bruderer et al., 2015<sup>5</sup> | DIA                 | MSV000081828 (PASS00589) |
| DIA:Röst2014           | Röst et al., 2014<sup>6</sup> | DIA                 | MSV000081829 (PASS00289) |
| DIA:Choi2017           | Choi et al., 2017<sup>2</sup> | DDA                 | MSV000084266 (PXD015258) |
| DIA:Navarro2016        | Navarro et al., 2016<sup>4</sup> | DIA                 | MSV000084266 (PXD015261) |
| DIA:Selevsek2015       | Selevsek et al., 2015<sup>7</sup> | TMT                 | MSV000081677 (PXD001010) |
| DIA:Varadarajan2020    | Varadarajan et al., 2020<sup>8</sup> | TMT                 | MSV000083881 (PXD014891) |
| DIA:Huang2020          | Huang et al., 2020<sup>9</sup> | TMT                 | MSV000084864 (PXD016647) |
| DIA:Amon2019           | Amon et al., 2019<sup>10</sup> | TMT                 | MSV000084898 (PXD009255) |
| DIA:Muntel2019         | Muntel et al., 2019<sup>11</sup> | TMT                 | MSV000084878 (PXD013658) |
| DIATMT:Muntel2019      | Muntel et al., 2019<sup>12</sup> | TMT/TMT             | MSV000084859 (PXD011691) |
| DIA:Andolina2018       | Andolina et al., 2018<sup>13</sup> | DIA                 | MSV000084898 (PXD008577) |
| TMT:Shen2020           | Shen et al., 2020<sup>14</sup> | TMT                 | MSV000085507 (PXD019221) |
| TMT:Djomehri2020       | Djomehri et al., 2020<sup>15</sup> | TMT                 | MSV000085343 (PXD014414) |
| TMT:Speer2019          | Specht et al., 2019<sup>16</sup> | TMT                 | MSV000084660 (PXD015303) |
| TMT:Yang2019           | Yang et al., 2019<sup>17</sup> | TMT                 | MSV000084662 (PXD010175) |
| TMT:Plubell2017        | Plubell et al., 2017<sup>18</sup> | TMT                 | MSV000082569 (PXD005953) |
| TMT:Paulo2016          | Paulo et al., 2016<sup>19</sup> | TMT                 | MSV000084263 (PXD015257) |
| SRM:Hüttenhain2019     | Hüttenhain et al., 2019<sup>20</sup> | SRM                 | MSV000084048 (PXD015303) |
| SRM:Hüttenhain2019_2   | Hüttenhain et al., 2019<sup>21</sup> | SRM                 | MSV000084555 (PXD009012) |
| SRM:Lobingier2017      | Lobingier2017 et al., 2017<sup>22</sup> | SRM                 | MSV000084585 (PXD013758) |
| DDA:Luther2019         | Luther et al., 2019<sup>23</sup> | DDA                 | MSV000084999 (PXD010174) |
| DDA:Vetterli2018       | Vetterli et al., 2018<sup>24</sup> | DDA                 | MSV000084897 (PXD010988) |
| DDA:Jung2017           | Jung et al., 2017<sup>25</sup> | DDA                 | MSV000079789 (PXD004263) |
| DDA:Malmström2016      | Malmström et al., 2016<sup>26</sup> | DDA                 | MSV000081695 (PXD002896) |
| DDA:Meierhofer2016     | Meierhofer et al., 2016<sup>27</sup> | DDA                 | MSV000084933 (PXD001355) |
| DDA:Rakus2016          | Rakus et al., 2016<sup>28</sup> | DDA                 | MSV000084934 (PXD002152) |
| DDA:Deshmukh           | Deshmukh et al., 2015<sup>29</sup> | DDA                 | MSV000084940 (PXD000288) |
| DDA:Guo2015            | Guo et al., 2015<sup>30</sup> | DDA                 | MSV000084938 (PXD001293) |
| DDA:Azimifa2014        | Azimifa et al., 2014<sup>31</sup> | DDA                 | MSV000084937 (PXD000867) |
| DDA:Hornburg2014       | Hornburg et al., 2014<sup>32</sup> | DDA                 | MSV000084939 (PXD000666) |
| DDA:Zhang2014          | Zhang et al., 2014<sup>33</sup> | DDA                 | MSV000084238 (PXD015905) |

Supplementary Table 5: List of datasets for reference
Supplementary references for datasets

1. Chiva, C., Ortega, M. & Sabido, E. Influence of the digestion technique, protease, and missed cleavage peptides in protein quantitation. *J Proteome Res* **13**, 3979-3986, doi:10.1021/pr500294d (2014).
2. Choi, M. *et al.* ABRF Proteome Informatics Research Group (iPRG) 2015 Study: Detection of Differentially Abundant Proteins in Label-Free Quantitative LC-MS/MS Experiments. *J Proteome Res* **16**, 945-957, doi:10.1021/acs.jproteome.6b00881 (2017).
3. Cox, J. *et al.* Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* **13**, 2513-2526, doi:10.1074/mcp.M113.031591 (2014).
4. Navarro, P. *et al.* A multicenter study benchmarks software tools for label-free proteome quantification. *Nat Biotechnol* **34**, 1130-1136, doi:10.1038/nbt.3685 (2016).
5. Bruderer, R. *et al.* Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Mol Cell Proteomics* **14**, 1400-1410, doi:10.1074/mcp.M114.044305 (2015).
6. Rost, H. L. *et al.* OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nat Biotechnol* **32**, 219-223, doi:10.1038/nbt.2841 (2014).
7. Selevsek, N. *et al.* Reproducible and consistent quantification of the Saccharomyces cerevisiae proteome by SWATH-mass spectrometry. *Mol Cell Proteomics* **14**, 739-749, doi:10.1074/mcp.M113.035550 (2015).
8. Varadarajan, A. R. *et al.* A Proteogenomic Resource Enabling Integrated Analysis of Listeria Genotype-Proteotype-Phenotype Relationships. *J Proteome Res*, doi:10.1021/acs.jproteome.9b00842 (2020).
9. Huang, T. *et al.* Combining Precursor and Fragment Information for Improved Detection of Differential Abundance in Data Independent Acquisition. *Mol Cell Proteomics* **19**, 421-430, doi:10.1074/mcp.RA119.001705 (2020).
10. Amon, S. *et al.* Sensitive Quantitative Proteomics of Human Hematopoietic Stem and Progenitor Cells by Data-independent Acquisition Mass Spectrometry. *Mol Cell Proteomics* **18**, 1454-1467, doi:10.1074/mcp.TIR119.001431 (2019).
11. Muntel, J. *et al.* Surpassing 10,000 identified and quantified proteins in a single run by optimizing current LC-MS instrumentation and data analysis strategy. *Mol Omics* **15**, 348-360, doi:10.1039/c9mo00082h (2019).
12. Muntel, J. *et al.* Comparison of Protein Quantification in a Complex Background by DIA and TMT Workflows with Fixed Instrument Time. *J Proteome Res* **18**, 1340-1351, doi:10.1021/acs.jproteome.8b00898 (2019).
13. Andolina, G. *et al.* A Peptidomimetic Antibiotic Interacts with the Periplasmic Domain of LptD from Pseudomonas aeruginosa. *ACS Chem Biol* **13**, 666-675, doi:10.1021/acschembio.7b00822 (2018).
14. Shen, B. *et al.* Proteomic and Metabolomic Characterization of COVID-19 Patient Sera. *Cell*, doi:10.1016/j.cell.2020.05.032 (2020).
15. Djomehri, S. I. *et al.* Quantitative proteomic landscape of metaplastic breast carcinoma pathological subtypes and their relationship to triple-negative tumors. *Nat Commun* **11**, 1723, doi:10.1038/s41467-020-15283-z (2020).
Specht, H. et al. Single-cell mass-spectrometry quantifies the emergence of macrophage heterogeneity. *bioRxiv* 665307, doi: https://doi.org/10.1101/665307 (2019).

Yang, M. et al. Proteogenomics and Hi-C reveal transcriptional dysregulation in high hyperdiploid childhood acute lymphoblastic leukemia. *Nat Commun* 10, 1519, doi:10.1038/s41467-019-09469-3 (2019).

Plubell, D. L. et al. Extended Multiplexing of Tandem Mass Tags (TMT) Labeling Reveals Age and High Fat Diet Specific Proteome Changes in Mouse Epididymal Adipose Tissue. *Mol Cell Proteomics* 16, 873-890, doi:10.1074/mcp.M116.065524 (2017).

Paulo, J. A., O'Connell, J. D. & Gygi, S. P. A Triple Knockout (TKO) Proteomics Standard for Diagnosing Ion Interference in Isobaric Labeling Experiments. *J Am Soc Mass Spectrom* 27, 1620-1625, doi:10.1007/s13361-016-1434-9 (2016).

Huttenhain, R. et al. A targeted mass spectrometry strategy for developing proteomic biomarkers: a case study of epithelial ovarian cancer. *Mol Cell Proteomics*, doi:10.1074/mcp.RA118.001221 (2019).

Huttenhain, R. et al. ARIH2 Is a Vif-Dependent Regulator of CUL5-Mediated APOBEC3G Degradation in HIV Infection. *Cell Host Microbe* 26, 86-99 e87, doi:10.1016/j.chom.2019.05.008 (2019).

Lobingier, B. T. et al. An Approach to Spatiotemporally Resolve Protein Interaction Networks in Living Cells. *Cell* 169, 350-360 e312, doi:10.1016/j.cell.2017.03.022 (2017).

Luther, A. et al. Chimeric peptidomimetic antibiotics against Gram-negative bacteria. *Nature* 576, 452-458, doi:10.1038/s41586-019-1665-6 (2019).

Vetterli, S. U. et al. Thanatin targets the intermembrane protein complex required for lipopolysaccharide transport in Escherichia coli. *Sci Adv* 4, eaau2634, doi:10.1126/sciadv.aau2634 (2018).

Jung, S. Y. et al. An Anatomically Resolved Mouse Brain Proteome Reveals Parkinson Disease-relevant Pathways. *Mol Cell Proteomics* 16, 581-593, doi:10.1074/mcp.M116.061440 (2017).

Malmstrom, E. et al. Large-scale inference of protein tissue origin in gram-positive sepsis plasma using quantitative targeted proteomics. *Nat Commun* 7, 10261, doi:10.1038/ncomms10261 (2016).

Meierhofer, D., Halbach, M., Sen, N. E., Gispert, S. & Auburger, G. Ataxin-2 (Atxn2)-Knock-Out Mice Show Branched Chain Amino Acids and Fatty Acids Pathway Alterations. *Mol Cell Proteomics* 15, 1728-1739, doi:10.1074/mcp.M115.056770 (2016).

Rakus, D., Gizak, A. & Wisniewski, J. R. Proteomics Unveils Fibroblast-Cardiomyocyte Lactate Shuttle and Hexokinase Paradox in Mouse Muscles. *J Proteome Res* 15, 2479-2490, doi:10.1021/acs.jproteome.5b01149 (2016).

Deshmukh, A. S. et al. Deep proteomics of mouse skeletal muscle enables quantitation of protein isoforms, metabolic pathways, and transcription factors. *Mol Cell Proteomics* 14, 841-853, doi:10.1074/mcp.M114.044222 (2015).

Guo, M. et al. High-resolution quantitative proteome analysis reveals substantial differences between phagosomes of RAW 264.7 and bone marrow derived macrophages. *Proteomics* 15, 3169-3174, doi:10.1002/pmic.201400431 (2015).
Azimifar, S. B., Nagaraj, N., Cox, J. & Mann, M. Cell-type-resolved quantitative proteomics of murine liver. *Cell Metab* **20**, 1076-1087, doi:10.1016/j.cmet.2014.11.002 (2014).

Hornburg, D. *et al.* Deep proteomic evaluation of primary and cell line motoneuron disease models delineates major differences in neuronal characteristics. *Mol Cell Proteomics* **13**, 3410-3420, doi:10.1074/mcp.M113.037291 (2014).

Zhang, B. *et al.* Proteogenomic characterization of human colon and rectal cancer. *Nature* **513**, 382-387, doi:10.1038/nature13438 (2014).