DIALib-QC an assessment tool for spectral libraries in data-independent acquisition proteomics

Mukul K. Midha, David S. Campbell, Charu Kapil, Ulrike Kusebauch, Michael R. Hoopmann, Samuel L. Bader & Robert L. Moritz

Data-independent acquisition (DIA) mass spectrometry, also known as Sequential Window Acquisition of all Theoretical Mass Spectra (SWATH), is a popular label-free proteomics strategy to comprehensively quantify peptides/proteins utilizing mass spectral libraries to decipher inherently multiplexed spectra collected linearly across a mass range. Although there are many spectral libraries produced worldwide, the quality control of these libraries is lacking. We present the DIALib-QC (DIA library quality control) software tool for the systematic evaluation of a library’s characteristics, completeness and correctness across 62 parameters of compliance, and further provide the option to improve its quality. We demonstrate its utility in assessing and repairing spectral libraries for correctness, accuracy and sensitivity.
Reliable identification and accurate quantification of peptides and proteins remains the central goal in mass spectrometry (MS) based proteomics\(^1\)\(^2\). Advancements in DIA methods have made them popular because of their increased coverage and unbiased MS/MS sampling of precursor ions detected or not, leading to more consistent peptide detection and quantitation across replicates\(^3\)\(^4\). The primary approach in DIA analysis requires prior knowledge of peptide fragmentation stored in spectral ion libraries\(^5\)\(^6\). As library quality directly influences DIA results, it is important to assess its characteristics and accuracy before drawing biological conclusions\(^8\)\(^9\)\(^10\). Several tools and pipelines are available to generate and merge spectral assay libraries, but few perform library quality assessment and are limited to a particular analysis pipeline, format or vendor, or are limited in scope, and none are able to correct a library. One such reported tool, SWATHXtend\(^3\), applies certain quality filters to generate clean hybrid libraries but does not evaluate spectral ion libraries for correctness. Here, our DIALib-QC software tool provides a path to ensure high-quality library usage for DIA analysis by evaluating and correcting spectral libraries for defects or biases (rather than blindly trusting and sharing for reuse). DIALib-QC performs an in-depth assessment, providing valuable characteristics of a library and also enables users to correct and generate an error-free library by eliminating problematic assays. These DIALib-QC features are extremely relevant to all DIA users regardless of the pipeline used (whether generating in-house libraries or using public repository libraries), library formats, and DIA analysis tools. DIALib-QC reports a tabular and graphical summary of common DIA library formats including OpenSWATH, Spectronaut, PeakView, and ProSight (generic text, Spectronaut compatible). We employ DIALib-QC and targeted data extraction with library-based DIA tools, PeakView and Spectronaut\(^1\)\(^1\)\(^1\) to demonstrate the importance of high-quality spectral libraries and their impact on DIA results. Further, we show DIALib-QC and its functionality to repair an ion library by correcting mass errors (recomputing to theoretical \(m/z\) values) and removing conflict assays to generate a clean and correct spectral library for DIA analysis. DIALib-QC is freely available to download and also integrated within the SWATHAtlas (http://www.swathatlases.org) interface, the free, extensively-used spectral ion library repository for DIA analysis, where libraries submitted by users are made accessible for reuse, thereby eliminating the effort of generating complex DIA spectral libraries a priori. We recommend assessing and correcting spectral libraries using DIALib-QC, prior to DIA analysis, to ensure correct identification and accurate quantitation of peptides and proteins from DIA data processing tools.

Results

Characteristics of the DIALib-QC tool. DIALib-QC inspects 62 criteria of a library and provides valuable insights of its composition in a tabulated form. These criteria are broadly organized into five categories (Fig. 1a, Supplementary Data 1) and are processed as depicted in Fig. 1b. These include (i) complexity of the library (number of peptides, peptide ions, and fragment ions); (ii) characteristics of the library that are physical attributes of precursor ions (e.g., charge distribution, retention time (RT) information, and properties of the enzymatic digestion) and fragment ions (e.g., number of fragments per precursor, ion series, and intensity distribution); (iii) modifications describing the number and types of posttranslational modifications in the library; (iv) completeness reflecting how comprehensively the library covers the target proteome, and percentage of proteotypic assays versus shared proteins; and (v) correctness highlighting diagnostic criteria that should be carefully evaluated before DIA analysis. DIALib-QC reports on these attributes as they are frequently encountered in poorly constructed libraries and as yet there has been no tool to interrogate them. Of the five, we focus hereafter on the correctness category that highly influences the DIA analysis.

1. Target/decoy percentage: DIALib-QC reports the percentage of target, mixed (map to both target and decoy), and decoy assays present in the library. Decoy assays are assays to peptides that are not present in the target proteome. These assays could be in silico derived, spiked into the library or inherited from DDA analysis during library generation. Some DIA pipelines require decoy assays for FDR estimation using a target-decoy approach, provided sufficient decoy assays are present for robust error modeling\(^1\)\(^2\).

2. Mass accuracy: DIALib-QC compares precursor and fragment ions \(m/z\) to theoretical values and provides the average delta mass, highlighting the accuracy of masses used in the library. The total number of these erroneous assays along with assays that cannot be verified with the input SWATH definition file (see category 4 below and Methods) are reported as the criterion problem assays. Such assays can be further explored in the DIALib-QC ion library repair feature to generate a clean library, free from mass errors, and conflict assays.

3. RT correlation fit: DIALib-QC estimates the RT fit by comparing the RT of [M + 2H]\(^+\) and [M + 3H]\(^3+\) charge states of the same peptide. This assesses the chromatography and RT normalization, based on reference peptides in the library. Poor RT correlation can lead to decreased identifications, due to reduced selectivity of DIA analysis, as chromatogram extraction occurs in a RT window centered around the expected elution time for the queried peptide\(^3\).

4. DIA verification: DIALib-QC reports the number of precursor ions that have at least one conflict fragment ion, and the total number of fragment ions that fall into the same DIA bin as the parent precursor ion. Such conflict assays may cause erroneous quantitation as un-fragmented precursor ion signals contribute to fragment ion intensities, thus overestimating peptides.

We sequentially explored these diagnostic criteria by explicitly generating defective ion libraries to demonstrate the DIALib-QC tool performance and effects on DIA analyses. We generated four spectral ion libraries from: (a) SWATHAtlas (pan-human library (PHL)) perturbed explicitly in q3 \(m/z\) value; viz. good (control), q3bad_0.01, q3bad_0.02, and q3bad_0.05, and seven sample-specific spectral ion libraries from (b) in-house K562 cell line, perturbed implicitly; viz. base, conflict, good (control), q3bad, raw, q3bad mz_corrected, and q3bad mz_corrected_clean (see Methods, Supplementary Fig. 1). All eleven libraries were assessed by DIALib-QC and their reports are provided in Supplementary Data 2.

Systematic quality evaluation of spectral ion libraries. Analysis of PHL and K562 good libraries by DIALib-QC shows >0.99 RT correlation (R-squared value), indicating high similarity between +2 and +3 charge states of a same peptide (Supplementary Fig. 2), and confirming high-quality libraries in the present DIA analysis. DIALib-QC identified the average \(m/z\) difference between the q3 \(m/z\) value in the library and the theoretical \(m/z\) value for each peptide sequence. Because these values may be computed differently from various software tools, or empirically derived, even small mass differences can have dramatic effects on DIA analysis. Specifically, DIALib-QC was used to
assess average fragment ion $m/z$ differences that increased from 0 $m/z$ to 0.05 $m/z$ across four PHL spectral libraries (see Methods). DIA analysis with PeakView resulted in a decrease of identified peptides from 38,944 to 878, and protein groups from 5801 to 653 (Fig. 2a). Increased fragment ion mass error decreased DIA sensitivity, even when those errors were small as reflected in their score distribution of target and decoy assays (Supplementary Fig. 3). Using the same libraries to analyze the data with Spectronaut, peptide and protein groups identification decreased from 48,947 to 37,932 and 5347 to 4276, respectively (Fig. 2a). Here, Spectronaut is less affected by fragment ion mass errors than PeakView because it implements a preprocessing MS2 calibration strategy onto the query library fragment ion $m/z$ values across the entire mass range (Supplementary Fig. 4). This calibration masks inherent issues with the library sufficiently enough to appear that the analysis is performing well. However,
DIALib-QC is able to identify the fragment ion mass error prior to DIA analysis, which produced a nearly 30% increase in the number of peptide identifications.

Further, seven sample-specific libraries were created (see Methods) and analyzed with DIALib-QC to highlight their differences juxtaposed with their effects on DIA analysis. The number of peptides and protein groups identified with PeakView or Spectronaut was consistent regardless of which library was used in the DIA analysis, except with the q3bad library (Fig. 2b), but the distinct differences in the libraries are evident from the DIALib-QC results (Supplementary Data 2). DIALib-QC identified the small fraction (0.1%) of decoy assays in the base library, that were left over from the DDA analysis during the library generation process. These assays were therefore removed in the conflict and good libraries and thus decoy numbers were confirmed as zero by DIALib-QC (Supplementary Data 2).

However, in silico decoy assays are often appended to the library tools such as OpenSWATH for false discovery rate (FDR) estimation. For DIA tools (such as Spectronaut and PeakView) that apply their own mechanism to estimate the FDR, in silico decoy assays can be added to the library to perform an entrapment analysis. During entrapment analysis, the false positive identifications of decoy peptides are estimated by matching to these decoy assays following DIA analysis, which can be used to calculate a realized FDR that is compared to the FDR estimate provided by the software. In either case, knowledge of the number of decoy assays is essential to determine if sufficient and not excessive assays are available to perform adequate DIA analysis.

Next, in the conflict library, DIALib-QC found 2057 peptide assays to have q3 fragment ions that fall in the same ion bin as the parent q1 precursor m/z (Fig. 3a). To demonstrate the effect of these conflict assays on the peptide quantitation, we compared the average peak quantities (PQ) between the two libraries (good and conflict) for the most intense 6 fragment ions of peptide ion GLQTSQDAR.2 in Spectronaut analysis (Fig. 3b). Both libraries showed similar contribution of five fragment ions in peptide quantitation whereas the sixth fragment ion, y4+ (489.24 m/z) in conflict contributes more than y3+ (361.18 m/z) in the good library (colored red to the total PQ) (Fig. 3c, d). The overestimating of peptide quantitation was mainly attributed by the residual precursor signal (denoted as p+) in MS2 spectrum (Fig. 3e, f). In the conflict library, the unfragmented precursor [M + 2H]2+ ion is overlapped with the y4+ ion (489.242 m/z) used for quantitation of the peptide, whereas in the good library, the y4+ ion was not used as it was removed during DIA verification (Fig. 3g, h). The correlation plots of relative intensities of fragment ions clearly indicates that y3+ ion in good library showed higher similarity in measured (black) and predicted (red) intensities whereas y4+ ion in conflict library was measured higher than expected intensity (Fig. 3i, j). This observation further supports that these conflict assays can affect peptide quantitation when overlapping residual precursor signal is measured with the q3...
**Fig. 3** Effect of conflict assay on the peptide quantitation in Spectronaut analysis. 

**a** DIALib-QC assessment reports the number of DIA conflict assays in both libraries.  
**b** Comparison of average peak quantities (PQ) of peptide GLQTSQDAR.2 between the two libraries (highlighted in orange and green) for the most intense six fragment ions observed in all five DIA/SWATH replicates. 
**c, d** Quantitative MS2 analysis based on extracted ion chromatograms (XICs) of peptide GLQTSQDAR.2 with both libraries eluted at the same RT. 
**e, f** MS2 spectrum of peptide GLQTSQDAR.2 in both libraries. p++ denotes un-fragmented precursor ion with [M + 2H]^{2+} charge. 
**g, h** Zoomed MS2 spectrum highlighting the overlap of un-fragmented precursor isotopic envelope (M, M + 1, M + 2) with y4+ ion in conflict library. 
**i, j** Correlation plots between measured (black) and predicted (red) relative fragment y4+ ion intensity showed higher similarity in the good than in the conflict library.
Methods

from DIA data processing tools. Identify libraries with DIALib-QC, prior to DIA analysis, to ensure correct to obtain a clean ion library without mass errors and con
libraries on the performance of DIA analysis along with an option. Two resultant fi
ed decoy spectra using PeptideProphet17 and iProphet18 within the
SpectraST20 work method using 100 variable acquisition windows14, two with a 1 Da overlap with the previous window. SWATH-MS2 spectra were collected in high-sensitivity mode from 100 to 1500 m/z with 25 milliseconds (ms) accumulation time. Before each SWATH-MS cycle, an MSI survey scan in high-resolution mode was recorded with a 250 ms accumulation time, resulting in a total duty cycle of ~2.8 s. Five analytical replicates of the K562 cell digest were measured for statistical confidence.

Spectral and assay library generation. All DIA wiff files were converted to profile mzML files using the ABSChic MS Data Converter version 1.3 beta. The Trans-Proteomic Pipeline (TPP)23 (5.2.0) was used for the analysis of the shotgun proteomics runs. The datasets were searched with Comrade24 against the redundant, canonical, and SWATH-MS like fragment sequences and iRT peptides (Biognosys). Carbamidomethyl (57.0214 Da) modification on cysteines was used as a fixed modification; oxidation (15.9949 Da) on methionine; and deamidation on asparagine and glutamine. Parent mass error was set to ±50 ppm, and fragment bin tolerance was set to 0.05 m/z. The search identifications were combined and statistically scored using PeptideProphet25 and iProphet26 within the

Modifications of assay libraries. We generated two different sets of spectral ion libraries (eleven libraries in total) to demonstrate the application of DIALib-QC. The first set is based on the PHL library by Rosenberger et al.21 (available at www.

K562 cell culture and protein digestion. K562 cells (ATCC CCL-243, human bone marrow myeloid leukemia cell line) were cultured at 37 °C with 5% CO2 in EMEM (ATCC 30-2003) supplemented with 10% fetal bovine serum. Cells were grown to 70% confluence, harvested, lysed, and proteins denatured in 8 M urea, 0.1% RapiGest, and 100 mM NH4HCO3. Protein content was determined by bicinchoninic acid assay (Thermo-Fisher). Proteins were reduced with 5 mM TCEP (30 min, 37 °C), alkylated with 10 mM iodoacetamide (30 min, room temperature, and darkness), digested using 1:50 trypsin (enzyme/protein, Promega), and samples desalted with tC18 SepPak cartridges (Waters).

Liquid chromatography-mass spectrometry. Peptides were separated with a NanoLC 400 interfacePlus HPLC system (Eksigent) configured in micro-flow mode and emitted into a TripleTOF 6600 mass spectrometer (SCIEX). Peptides were trapped on a 10 x 0.3 mm trap cartridge ChromXP C18EL, 5 µm, 120 Å (Eksigent) at 10 µL/min, and separated on a ReproSil-Pur C18-AQ, 2.4 µm, 150 x 0.2 mm (Dr. Maisch GmbH) at 5 µL/min using a gradient from 3 to 35% B in 98 min, 35 to 40% B in 5 min, 40 to 80% B in 2 min, isotropic flow at 80% B for 8 min, 80 to 3% B in 2 min, and isotropic at 3% B for 25 min. Data were acquired using data-dependent acquisition (DDA) and data-independent acquisition (DIA) mode. For both DDA and DIA acquisitions with mass spectrometry, the

DIA/SWATH mass spectrometry. DIA/SWATH data were collected with an MS/ MSN2 SWATHTM Acquisition method using 100 variable acquisition windows24, each with a 1 Da overlap with the previous window. SWATH-MS2 spectra were collected in high-sensitivity mode from 100 to 1500 m/z with 25 milliseconds (ms) accumulation time. Before each SWATH-MS cycle, an MIS survey scan in high-resolution mode was recorded with a 250 ms accumulation time, resulting in a total duty cycle of ~2.8 s. Five analytical replicates of the K562 cell digest were measured for statistical confidence.

Spectral and assay library generation. All DIA wiff files were converted to profile mzML files using the ABSChic MS Data Converter version 1.3 beta. The Trans-Proteomic Pipeline (TPP)23 (5.2.0) was used for the analysis of the shotgun proteomics runs. The datasets were searched with Comrade24 against the redundant, canonical, and SWATH-MS like fragment sequences and iRT peptides (Biognosys). Carbamidomethyl (57.0214 Da) modification on cysteines was used as a fixed modification; oxidation (15.9949 Da) on methionine; and deamidation on asparagine and glutamine. Parent mass error was set to ±50 ppm, and fragment bin tolerance was set to 0.05 m/z. The search identifications were combined and statistically scored using PeptideProphet25 and iProphet26 within the

Modifications of assay libraries. We generated two different sets of spectral ion libraries (eleven libraries in total) to demonstrate the application of DIALib-QC. The first set is based on the PHL library by Rosenberger et al.21 (available at www.

K562 cell culture and protein digestion. K562 cells (ATCC CCL-243, human bone marrow myeloid leukemia cell line) were cultured at 37 °C with 5% CO2 in EMEM (ATCC 30-2003) supplemented with 10% fetal bovine serum. Cells were grown to 70% confluence, harvested, lysed, and proteins denatured in 8 M urea, 0.1% RapiGest, and 100 mM NH4HCO3. Protein content was determined by bicinchoninic acid assay (Thermo-Fisher). Proteins were reduced with 5 mM TCEP (30 min, 37 °C), alkylated with 10 mM iodoacetamide (30 min, room temperature, and darkness), digested using 1:50 trypsin (enzyme/protein, Promega), and samples desalted with tC18 SepPak cartridges (Waters).

Liquid chromatography-mass spectrometry. Peptides were separated with a NanoLC 400 interfacePlus HPLC system (Eksigent) configured in micro-flow mode and emitted into a TripleTOF 6600 mass spectrometer (SCIEX). Peptides were trapped on a 10 x 0.3 mm trap cartridge ChromXP C18EL, 5 µm, 120 Å (Eksigent) at 10 µL/min, and separated on a ReproSil-Pur C18-AQ, 2.4 µm, 150 x 0.2 mm (Dr. Maisch GmbH) at 5 µL/min using a gradient from 3 to 35% B in 98 min, 35 to 40% B in 5 min, 40 to 80% B in 2 min, isotropic flow at 80% B for 8 min, 80 to 3% B in 2 min, and isotropic at 3% B for 25 min. Data were acquired using data-dependent acquisition (DDA) and data-independent acquisition (DIA) mode. For both DDA and DIA acquisitions with mass spectrometry, the
theoretical fragment ion m/z values and keeps them with the −e option enabled in the spectral library. The base library includes both target and decoy assays. Decoy assays contain peptides that do not exist in the human (target) proteome and are inherited from the DDA analysis during the library generation step. Peptide sequences identified from these decoy assays (referred to as decoy peptides in DIA analysis) are therefore known to be false. Further, the library was limited to the top 6 fragments per precursor without enabling the SWATH verification, which kept the number of conflicts low in the library.

**Library 2: K562 conflict (Consensus spectra + theoretical fragment ion m/z values + conflict assays):** This library was generated by removing the decoy assays from Library 1. Library 2 contains conflict assays (assays which are not verified with the SWATH definition file). The SWATH definition file contains precursor isolation windows with the information of mass range and width used in SWATH mode. The objective to construct this library was to understand how fragment ions that fall in the parent ion bin(s) impact the identification and quantification of a peptide using both Spectronaut and PeakView analysis. The DIALib-QC tool reports the total number of precursor and fragment ions m/z values that contribute to conflicts in the library.

**Library 3: K562 good (consensus spectra + theoretical fragment ion m/z values):** This library was generated from consensus spectra using theoretical fragment ion m/z values in SpectraST. K562 good was used as control library and was compared to other perturbed libraries. Library 3 was verified with the DIA/SWATH definition file used for sample acquisition and which ensures that fragments that fall in the parent ion swath bin(s) will not be selected.

**Library 4: K562 qbad (consensus spectra + observed fragment ion m/z values):** This library was generated from consensus spectra in the SpectraST pipeline without enabling the −e option, which resulted in storing the observed fragment ion m/z values in the spectral library.

**Library 5: K562 raw (redundant spectra + theoretical fragment ion m/z values):** The raw spectral library was generated directly from redundant spectra and contains non-unique entries resulting from multiple observations of the same peptide ion. All fragment ions of the same peptide ion found in all sample replicates above an iProphet probability threshold were used to generate the spectral library. The K562 raw spectral library uses theoretical fragment ion m/z values.

**Library 6: K562 qbad mz_corrected (consensus spectra + corrected theoretical fragment ion m/z values):** This library was generated from Library 4 (qbad) by correcting the observed to theoretical fragment ion m/z values.

**Library 7: K562 qbad mz_corrected_clean (consensus spectra + corrected theoretical fragment ion m/z values + removed conflict assays):** This library was generated from Library 6 by removing the conflict assays identified in reference to the SWATH definition file by using the filter conflict assays option of DIALib-QC. The resultant library is free from any mass errors and conflict assays. Further, for each spectral library (in both sets), two different library formats were used for analysis based on the DIA software, PeakView or Spectronaut. For each library, PKV.txt formatted libraries were generated using the SpectraST workflow. In the PKV format used in PeakView (SWATH 2.0) analysis, iRT values of all assays were shifted to positive values by adding −60.6 and 51.6 for PHL and K562 set of libraries respectively to prevent the failure in importing the complete information of the library. For Spectronaut analysis, negative iRT values are permissible in PKV inputs and did not require additional manipulation. DIALib-QC provides assessment reports for all libraries (DDA derived) in both library formats, PeakView and Spectronaut (Supplementary Data 2). In addition, DIALib-QC is fully capable to evaluate DIA derived libraries constructed using untargeted workflows in Spectronaut and DIA-Umpire. The tool supports Spectronaut format libraries regardless of, if it is built from DDA or DIA data. Similarly, DIA-Umpire analysis can be used to generate a library directly from DIA data using the SpectraST pipeline, which produces library formats (OpenSWATH and PeakView) compatible with DIALib-QC. In case of other formats, users can transform their library to formats compatible with DIALib-QC for its assessment and repair.

### DIA/SWATH data analysis

Two different DIA software tools were used in this study, Sciex PeakView and Biognoys Spectronaut, to perform the targeted data extraction of five analytical DIA replicates.

**PeakView (SWATH 2.0).** PeakView.tsv assay libraries generated as described above were appended with iRT peptide assays (protein labeled as [RT-Cal protein]). The RT calibration for all peptides in the ion library was performed based on the iRT peptide elution profiles for an adjusted RT window at a 75 ppm XIC extraction width with PeakView version 2.2 and the SWATH 2.0 plug-in MS/MS (ALL) using SWATH App module (v2.0) (SCIEX). To import all peptides from the library, the number of peptides per protein was set to the maximum value of 6 transitions per peptide. The FDR threshold was set at 1%, excluding both modified and shared peptides. The processed results were exported to Microsoft Excel files (Supplementary Data 4).

**Spectronaut.** For Spectronaut, the assay libraries were used directly as generated and described above. The HTRMS converter was used to convert the raw WIFF files into HTRMS files which were imported into Spectronaut version 13.3.190726.43655 (Laika, Biognoys, Switzerland). For the nonlinear iRT calibration strategy, a dynamic window was used for both mass tolerance and XIC RT extraction window. Preprocessing of MS1 and MS2 calibration strategy was enabled. The decoy assays were dynamically generated using the scrambled decoy method with a set size of 1 as a fraction of the input library size. The identification was performed using the normal distribution estimator with precursor and protein identification results filtered with a q value of <0.01. For quantification, MS2 ion peak areas of quantified peptides were summed to estimate the protein peak areas (Supplementary Data 5).

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### Data availability

Data that support the findings of this study are available via ProteomeXchange with identifier PXD020953 via the PRIDE23. Any other relevant data are available from the corresponding author on request.

### Code availability

The user-friendly web version of the DIALib-QC tool is available at https://db.systemsbiology.net/shemas/cgi/PeptideAtlas/AssessDIALibrary. DIALib-QC code, installation guide and instructions on how to run the full pipeline including repairing the spectral library can be found at http://www.swathlas.org/DIALibQC.php. DIALib-QC v1.0 is also freely available at https://github.com/CharuKMidha/DIALib-QC under a GNU licence.

Received: 2 March 2020; Accepted: 17 September 2020; Published online: 16 October 2020

### References

1. Aebersold, R. & Mann, M. Mass-spectrometric exploration of proteome structure and function. Nature 537, 347–355 (2016).
2. Kusebauch, U. et al. Human SRMAtlas: a resource of targeted assays to quantify the human proteome. Cell 166, 766–776 (2016).
3. Venable, J. D., Dong, M. Q., Wohlschlegel, J., Dillin, A. & Yates, J. R. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. Nat. Methods 1, 39–45 (2004).
4. Gillet, L. C. et al. Targeted data extraction of the MS/MS spectratra generated from Library 1. Library 2 contains consensus spectra using theoretical fragment ion m/z values in SpectraST. K562 good was used as control library and was compared to other perturbed libraries. Library 3 was verified with the DIA/SWATH definition file used for sample acquisition and which ensures that fragments that fall in the parent ion swath bin(s) will not be selected.
5. Kusebauch, U. et al. Human SRMAtlas: a resource of targeted assays to quantify the human proteome. Cell 166, 766–776 (2016).
6. Venable, J. D., Dong, M. Q., Wohlschlegel, J., Dillin, A. & Yates, J. R. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. Nat. Methods 1, 39–45 (2004).
7. Ting, Y. S. et al. Peptide-centric proteome analysis: an alternative strategy for quantification of proteomes. Mol. Cell. Proteom. 11, https://doi.org/10.1074/mcp.O111.016717 (2012).
8. Weisbrod, C. R., Eng, J. K., Hoopmann, M. R., Baker, T. & Bruce, J. E. Accurate peptide fragment mass analysis: multiplexed peptide identification and quantification. J. Proteome Res. 11, 1621–1632 (2012).
9. Ludwig, C. et al. Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. Mol. Syst. Biol. 14, e8126 (2018).
10. Reiter, L. et al. MProphet: automated data processing and statistical validation for dimensional liver microtissues. Mol. Cell. Proteomics 14, 1400–1410 (2015).
11. Bruderer, R. et al. Extending the limits of quantitative proteome profiling with data-independent acquisition and application to aetionaphthothen-treated three dimensional liver microtissues. Mol. Cell. Proteomics 14, 1400–1410 (2015).
12. Feng, X. Dong et al. Using the entrapment sequence method as a standard to quantify peptide modifications. Mol. Cell. Proteom. 15, 1400–1514 (2016).
13. Deutsch, E. W. et al. Expanding the use of spectral libraries in proteomics. J. Proteome Res. 17, 4031–4060 (2019).
14. Schilling, B., Gibson, B. W. & Hunter, C. L. Generation Of High-Quality SWATH* Acquisition Data For Label-free Quantitative Proteomics Studies Using TripleTOP Mass Spectrometers. Methods Mol Biol 1550, 233–233 (2017).
15. Deutsch, E. W. et al. Trans-proteomic pipeline, a standardized data processing pipeline for large-scale reproducible proteomics informatics. Proteom. Clin. Appl. 9, 745–754 (2015).
16. Eng, J. K., Jahan, T. A. & Hoopmann, M. R. Comet: an open-source MS/MS sequence database search tool. Proteomics 13, 22–24 (2013).
17. Keller, A., Nesvizhskii, A. I., Kolker, E. & Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* 74, 5383–5392 (2002).
18. Shteynberg, D. et al. iProphet: Multi-level Integrative Analysis of Shotgun Proteomic Data Improves Peptide and Protein Identification Rates and Error Estimates. *Mol. Cell. Proteom.* 10, M111.007690 (2011).
19. Reiter, L. et al. Protein identification false discovery rates for very large proteomics data sets generated by tandem mass spectrometry. *Mol. Cell. Proteom.* 8, 2405–2417 (2009).
20. Lam, H. et al. Development and validation of a spectral library searching method for peptide identification from MS/MS. *Proteomics* 7, 655–667 (2007).
21. Rosenberger, G. et al. A repository of assays to quantify 10,000 human proteins by SWATH-MS. *Sci. Data* 1, 1–15 (2014).
22. Tsou, C. C. et al. DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nat. Methods* 12, 258–264 (2015).
23. Vizcaíno, J. A. et al. 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* 44, D447–D456 (2016).

**Acknowledgements**

We like to thank Barbara Grimes for culturing K562 cells used in this study. This work was funded in part by the National Institutes of Health, National Institute of General Medical Sciences grant: R01GM087221, the National Institute of Allergy and Infectious Diseases grant: R21AI133335, and the National Institute on Aging grant: U19AG023122.

**Author contributions**

R.L.M., S.L.B., and D.C. conceived the study. M.K.M., D.C., U.K., and R.L.M. designed the study and R.L.M. supervised the study; D.C. and C.K. developed DIALib-QC tool; M.K.M. prepared the sample and acquired MS data, M.K.M. executed DIALib-QC, PeakView and Spectronaut analyses; M.K.M., D.C., U.K., M.R.H., and C.K. analyzed the data, and all authors contributed to the preparation of the paper.

**Competing interests**

The authors declare no competing interests.

**Additional information**

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-18901-y.

**Correspondence** and requests for materials should be addressed to R.L.M.

**Peer review information** *Nature Communications* thanks Ben Collins and the other, anonymous, reviewer for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at http://www.nature.com/reprints

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.