Extensive and accurate benchmarking of DIA acquisition methods and software tools using a complex proteomic standard

Clarisse Gotti[1,2]*, Florence Roux-Dalvai[1,2]*, Charles Joly-Beauparlant[2], Mickaël Leclercq[2], Loïc Mangnier[2,3] and Arnaud Droit[1,2#]

[1] Proteomics platform, CHU de Québec - Université Laval Research Center, Québec City, Québec, Canada.

[2] Computational Biology Laboratory, CHU de Québec - Université Laval Research Center, Québec City, Québec, Canada.

[3] Cervo, Brain research Center, Quebec City, Quebec, Canada

* these authors contributed equally to the work

#To whom correspondence should be addressed: arnaud.droit@crchudequebec.ulaval.ca

Running title: Benchmark of proteomic DIA workflows

Keywords: Proteomic standard, software benchmark, Data Independent Acquisition, label-free quantification

ABSTRACT

In the past few years, LC-MS/MS in DIA mode has become a strategy of choice for deep coverage of complex proteomes and accurate quantification of low abundant species. However, there is still no consensus in the literature on the best acquisition parameters and processing tools to use. We present here the largest benchmark of DIA proteomic workflows on Orbitrap instruments ever published. Using a complex proteomic standard, we tested 36 workflows including 4 different acquisition schemes and 6 different software tools (DIA-NN, DIA-Umpire, OpenSWATH, Scaffold-DIA, Skyline and Spectronaut-Pulsar-X) with or without the use of a spectral library. By reporting the protein identifications, linearity, reproducibility, as well as sensitivity and specificity in 28 pairwise comparisons, we propose here a guide book for the users to choose the best DIA strategy for their data. Moreover, our data has been deposited on ProteomeXchange for testing or developing new DIA processing tools.

INTRODUCTION

Since the last decades, proteomics, i.e. the analysis of the protein content of a biological sample, has become an essential strategy for the comprehension of systems biology and molecular events involved in health and disease. *Bottom-up proteomics* especially, which rely on the Liquid Chromatography-tandem Mass Spectrometry (LC-MS/MS) analysis of peptides resulting from protein proteolysis, allows the identification and quantification of up to thousands of proteins in a whole cell extract[1,2].

Among the acquisition strategies used for bottom-up proteomics, the most widely used, so-called *shotgun proteomics* or *discovery proteomics*, use Data-Dependent Acquisition (DDA) analysis which is based on the successive acquisition of MS1 survey scans, containing the accurate mass measurement of co-eluting peptides, followed by the acquisition of fragmentation spectra (MS2 spectra) of automatically selected precursor ions, containing the peptide sequence information[3]. MS1 and MS2 spectra used to search against publicly available protein databases in order to obtain peptide identification and then infer the corresponding proteins. Moreover, quantification information can be obtained, using, for instance in *Label Free Quantification* (LFQ), the area under the peptide chromatographic curve reconstructed from the MS1 precursor ion intensities[4]. Therefore, using DDA analysis, thousands of proteins can be identified and relatively quantified from fractionated, purified or whole protein extracts, bringing essential information on biological mechanisms in a large-scale manner.

Conversely, in *targeted proteomics*, a previously chosen list of proteins of interest can be monitored in protein extracts, using Selected Reaction Monitoring (SRM)[5,6] or Parallel Reaction Monitoring (PRM)[7], by the successive measurements of specific precursor/fragment transitions of a given peptide along the chromatographic elution. In that case, the detection is validated by the overlap of the chromatographic elution profiles from different transitions of the same peptide and peptide quantification is obtained by integration of the area under these chromatographic peaks. As compared to *discovery proteomics*, targeted strategies display a better sensitivity, quantification accuracy and reproducibility across a large number of samples. For these reasons, they are well-suited for monitoring protein candidates in large sample cohorts in biomarker validation studies.

Although these two approaches have been widely adopted by the community, they have several drawbacks. DDA analysis is based on a stochastic event of precursor selection prior to the acquisition of fragmentation spectra which leads to a non-reproducibility in protein identification from run to run in the analysis of complex samples. Although, the post-acquisition bioinformatical tools are now better able to deal with this undersampling effect by recovering the precursor MS1 signal in all the runs of a same experiment, the dynamic range of intensity within MS1 spectra often limits the detection and quantification of low abundances species. Conversely, by the narrow quadrupole filtering of the precursor to monitor, SRM or PRM analyses avoid, in most of the cases, the concomitant selection of high-abundant species which makes this strategy more sensitive and more reproducible than DDA. However, even when the signal acquisition of transitions is finely scheduled during the total chromatographic time, this strategy is limited to the analysis of several hundreds of peptides due the number of transitions to monitor in parallel without exceeding an appropriate cycle time for the reconstruction of chromatographic peaks.

For those reasons, a new strategy has emerged in the past few years, named Data-Independent Acquisition (DIA), which combines the strengths of both DDA and targeted approaches in order to obtain reproducible and accurate peptide/protein quantification at proteome level (i.e. several thousands of proteins). In this method, initially described by Venable *et al.*[8] and later popularized by Gillet *et al.*[9,10], precursors are not individually selected, but systematic windows of precursors are used for fragmentation all along the mass range. These results in the concomitant fragmentation of all peptide ions contained in the same window. The main advantages of this strategy is to cover all detectable ions present in a sample and reduce the dynamic range of spectra by using a small window rather than a full mass range acquisition. This makes DIA analyses more reproducible than DDA in quantification, especially for low abundance species, while able to identify thousands of species[10,11]. However, the acquisition of high complexity fragmentation spectra due the concomitant selection of multiple precursors makes the database search engines usually used for DDA analysis, unable to match experimental data to theoretical masses obtained from the public protein databases. Therefore, two main strategies, named spectrum-centric and peptide-centric, have been developed to obtain peptide identification from the chimeric DIA-MS2 spectra. In the spectrum-centric strategy,

identifications are obtained by the deconvolution of the complex spectra, mostly using fragment co-elution profiles, followed by protein database search, while in peptide-centric strategy, m/z and retention time information from spectral libraries, previously obtained in DDA mode and searched against public protein databases, are used to query the DIA spectra. In the recent years, several open-source or proprietary software tools have been developed for DIA analysis[12–16]. However, each of them has its own specificity, using one or both of the spectrum- and peptide-centric strategies, including or not normalization and missing values imputation algorithms, having or not a graphical user interface. Moreover, the scheme of windows for DIA acquisition appears to be a critical point since wide windows allow the coverage of the whole mass range in a reasonable cycle time while narrow windows produce lower complexity spectra for the subsequent analysis. Overlapped windows have also been shown to increase the identification while keeping a good cycle time for each analyte[17]. Therefore, among the plethora of methods for DIA acquisition and data treatment found in the literature, it appears difficult for the user to choose the best strategy for the analysis of complex proteomes and especially for the detection and accurate quantification of low abundant species. In 2016, Navarro *et al.* in collaboration with the developers of several DIA tools, made a first attempt of software comparison using a proteomic dataset coming from the mixture of 3 protein extracts from human, yeast and *E.coli* acquired in SWATH-MS DIA mode on TripleTOF (Sciex) instruments[18]. If this experimental design allowed the measurement of protein ratios across the different proteomes, no information on the abundance of each protein in the mixture can be obtained which makes the interpretation of the software efficiency more difficult especially for low abundance species or very large protein ratios. Moreover, since the publication of this work, other tools have been developed and some others extend their capabilities to library-free (spectrum-centric) strategies. For these reasons, we present here an extensive benchmark of proteomic strategies for DIA analysis with different acquisition schemes and analysis software tools on a proteomic standard composed of whole *E.coli* protein extract background spiked in with the UPS1 commercial mixture (Sigma) composed of 48 human proteins in known concentration to mimic a real complex sample. A similar standard has already been used in the past for acquisition methods or processing tools testing for DDA analysis[19–21]. In our study, eight

concentrations of UPS1 in *E.coli* extract, have been measured on an Orbitrap Fusion (Thermo) instrument with four different window sizes (narrow, wide, variable and overlapped) and analyzed with six different DIA processing tools (DIA-NN, DIA-Umpire, OpenSWATH, Scaffold-DIA, Skyline and Spectronaut-Pulsar-X) with or without the use of a large spectral library containing more than 2800 *E.coli* and UPS1 proteins. This work allowed us to compile information on protein identification, quantification linearity, reproducibility and accuracy, sensitivity and specificity in pairwise comparisons, for 36 different DIA workflows, in order to identify the best strategies for the protein quantification regarding their abundance in complex proteomes. We also make our data (96 DIA raw files, software outputs and quantification tables) accessible to the community via ProteomeXchange for other statistical analyses or software testing and development.

METHODS

Sample preparation

Bacterial culture and protein extraction

E.coli culture (#CCRI-12923) was obtained from the Culture Collection of Centre de Recherche en Infectiologie of Laval University (CCRI, Quebec, Canada) registered as WDCM861 at the World Data Centre for Microorganisms. Six aliquots of 1mL of the broth culture in Brain Heart Infusion (BHI) medium at a concentration of $8x10^8$ CFU/mL were pelleted by centrifugation at 10,000 x *g* during 15 minutes and were washed three times with 1mL of 50mM Tris. Dried pellets were frozen and stored at -20°C.

For protein extraction, each aliquot was resuspended in 135µL of extraction buffer (50mM ammonium bicarbonate (ABC), 1% sodium deoxycholate (SDC) , 20mM dithiothreitol (DTT)). Bacteria were inactivated by heating 10 minutes at 95°C and the lysis was achieved by sonication with Bioruptor® (Diagenode), for 15 minutes with 30s/30s ON/OFF cycles at high intensity level.

Samples were centrifuged at 13,000 x *g* for 10 minutes to remove insoluble cell debris, supernatants of the six aliquots were then pooled together and protein concentration was measured by Bradford Assay. Protein extracts were stored at -80°C.

Proteomic standard

E.coli whole protein extract was diluted at 0.1µg/µL and used to resuspend a Universal Proteomic Standard-1 (UPS1, Sigma) vial containing 48 human proteins at 5pmol of each. Dilution series were performed in order to obtain 8 different UPS1 concentrations: 50, 25, 10, 5, 2.5, 1, 0.25 and 0.1fmol of UPS1 per µg of *E.coli*. Dissulfide bridges were reduced by heating 30 minutes at 37°C and alkylated with 50mM iodoacetamide during 30 minutes in obscurity. Prior to trypsin digestion, pH was adjusted to 8.0 with sodium hydroxide. Proteolytic digestion was performed with trypsin (Promega) in an enzyme:protein ratio of 1:50 at 37°C overnight. The reaction was stopped by acidification of the samples at pH 2.0 with formic acid (FA). After centrifugation at 16,000 x *g* for 5 minutes to remove SDC precipitate, the supernatants (peptides) were cleaned up using an Oasis HLB cartridge 10mg (Waters), aliquoted and vacuum-dried.

Spectral library

250µg of *E.coli* protein extract were used to generate a spectral library. Reduction, alkylation and trypsin digestion was performed under the same conditions as described above.

Peptides were fractionated on an Agilent 1200 Series HPLC system equipped with an Agilent Extend C_{18} (1.0mm x 150mm, 3.5µm) column. Peptides were loaded at 1mL/min of solvent A (10mM ABC pH 10) and eluted by addition of solvent B (90% acetonitrile (ACN) / 10% ABC pH 10) with a gradient of 5 to 35% B during 60 minutes and 35 to 70% B during 24 minutes. Fractions were collected in a 96-well plate at every 1 minute and finally pooled into 48 fractions and vacuum-dried.

A sample of 200fmol UPS1 per µg of *E.coli* was also prepared in the same conditions to complete the spectral library with UPS1 human proteins.

Data acquisition

NanoLC-MS/MS analysis

For both proteomic standard or spectral library acquisition, samples were resuspended at a 0.2µg/µL concentration with loading buffer (2% ACN / 0.05% trifluoroacetic acid (TFA) containing 0.5X iRT peptides (Biognosys)), and 5µL were injected on a nanoLC-MS/MS system composed of a U3000 NanoRSLC liquid chromatography device (ThermoScientific, Dionex Softron GmbH, Germering, Germany) online with an Orbitrap Fusion Tribrid – ETD mass spectrometer (ThermoScientific, San Jose, CA, USA) driven by the Orbitrap Fusion Tune Application 3.0.2041 and equipped with a Nanospray Flex™ ion source. Peptides were trapped at 20µL/min in loading solvent (2% ACN / 0.05% TFA) on a 300 mm i.d x 5mm, C_{18} PepMap100, 5 mm, 100Å precolumn cartridge (Thermo Fisher Scientific) during 5 minutes. Then, the precolumn was switched online with a PepMap100 RSLC, C_{18} 3mm, 100Å, 75mm i.d. x 50cm length column (Thermo Fisher Scientific) and the peptides were eluted with a linear gradient from 5-40% solvent B (A: 0.1% FA, B: 80% ACN / 0.1% FA) in 90 minutes, at 300nL/min flow rate. Spectra were acquired using Thermo XCalibur software version 4.1.50. Internal calibration using lock mass on the m/z 445.12003 siloxane ion was used.

DDA acquisition parameters – Spectral library

Full scan mass spectra (350 to 1800m/z) were acquired in the Orbitrap using an AGC target of 4e5, a maximum injection time (IT) of 50ms and a resolution of 120,000. Each MS scan was followed by acquisition of fragmentation MS/MS spectra of the most intense ions for a total cycle time of 3 seconds (top speed mode). The selected ions were isolated using the quadrupole analyser in a window of 1.6 m/z and fragmented by Higher energy Collision-induced Dissociation (HCD) with 35% of collision energy (CE). The resulting fragments were detected in the Orbitrap at a 15,000 resolution with an AGC target of 5e4 and a maximum IT of 22ms. Dynamic exclusion of previously fragmented peptides was set for a period of 30sec and a tolerance of 10ppm.

DIA acquisition parameters – Proteomic standard

Full scan mass spectra were acquired in the Orbitrap using an AGC target of 4e5, a maximum IT of 50ms and a resolution of 60,000. Each MS scan is followed by 75 DIA scans with different isolation window width depending on each scheme: 8m/z windows covering from 350m/z to 950m/z precursor mass range for *Narrow* scheme, two cycles of 8m/z isolation windows shifted by 4m/z for *Overlap* scheme, covering from 350m/z to 954m/z, combination of 8m/z (from 455 to 711m/z) and 15m/z (from 350 to 455m/z and from 711 to 1251m/z) windows for *Variable* scheme and 15m/z windows covering from 350m/z to 1475m/z for *Wide* scheme (Fig.1 and SuppTable S1). Precursor ions were selected in the quadrupole, fragmented by HCD with 35% CE and the detection of the fragments was performed in the Orbitrap at a 15,000 resolution and with an AGC target of 4e5 and a maximum IT of 22ms.

Data processing and statistics

Spectral library generation

Raw data from *E.coli* fractionation and 200fmol UPS1 DDA acquisition were search against an *E.coli* database (UniProt Reference Proteome – Taxonomy 83333 – Proteome ID UP000000625 – 4314 entries – 2016.03.15) and the UPS1 database (downloaded from Sigma – 48 entries), using Mascot search engine version 2.5.1 (Matrix Science) as a node in Proteome Discoverer 2.2. Trypsin enzyme parameter setting was selected with two possible missed cleavages. Carbamidomethylation of cysteines (C) was set as fixed modification, methionine (M) oxidation, asparagine (N) and glutamine (Q) deamidation were set as variable modifications. Mass search tolerances were 10ppm and 25mmu for MS and MS/MS respectively. Peptides and proteins were filtered at 1% False Discovery Rate (FDR). Mascot *.dat* files were then imported into Skyline software[22] (version 19.1.0.193) to generate a spectral library in a *.blib* format, with a cut-off score of 0.95.

DIA files treatment

Six different tools (DIA-NN[15] v02/04/2020, DIA-Umpire v2.0[14], OpenSWATH v2.4.0[12], Scaffold-DIA v1.2.1 (Proteome Software, Inc.), Skyline v19.1.0.193[22], Spectronaut-Pulsar-X (Biognosys AG)) were used for extracting peptide signal from raw files using the previously generated spectral library (DIA-NN, OpenSWATH, Scaffold-DIA, Skyline, Spectronaut-Pulsar-X) or for searching them against an *E.coli* and UPS1 fasta file (UniProt Reference Proteome – Taxonomy 83333 – Proteome ID UP000000625 – 4314 entries – 2016.03.15 and the 48 sequences of UPS1) (DIA-NN, DIA-Umpire, Scaffold-DIA, Spectronaut-Pulsar-X)

Raw files were converted in .mzML for the analysis by DIA-NN and OpenSWATH and in .mzXML files for DIA-Umpire analysis, according to software author recommendations. Raw files were directly used for the analysis of the other tools (Scaffold-DIA, Skyline and Spectronaut-Pulsar-X). For OpenSWATH, the spectral library was converted into a .tsv file.

The following processing parameters were common to all software: maximum of 2 trypsin missed cleavages, cysteine carbamidomethylation as fixed modification, methionine oxidation and deamidation of glutamine and asparagine as variable modifications, a minimum of 3 fragments and a maximum of 6 were used to consider a peptide for quantification and only peptides with charges 2+ to 4+ were kept. Specific parameters to each software tool are detailed in Supplementary Table S1. For all tools, quantification tables at precursor level were used to perform DIA label-free quantification.

Data processing and evaluation of DIA label-free quantification

All data processing and visualization was performed using the R software[23] (SuppFig.S1). After removing contaminants and decoy features, data were filtered out on specific validation criteria for each tool: 1% FDR for Scaffold-DIA and DIA-NN; *q-value* < 0.01 with filtering on intensity (intensity < 100) to remove outliers for Spectronaut-Pulsar-X; mProphet *q-value* < 0.01, library dotP > 0.70 and mass tolerance < 10ppm for Skyline; PyProphet score (m_score) < 0.01 for OpenSWATH and PeptideProphet score (MaxProb) > 0.99 for DIA-Umpire.

While DIA-NN, Scaffold-DIA and Spectronaut-Pulsar-X performed their own data normalization, data from the other tools (DIA-Umpire, OpenSWATH and Skyline) were normalized by applying a normalization factor calculated on the median of the intensities for each sample injection. For protein and peptide identification counting, precursors were considered as identified for each of the 8 UPS1 concentrations, if 3 quantification values were observed for the three replicates before imputation of missing values, and the number corresponding to the corresponding unique stripped sequences and accession numbers were reported.

Missing values were then replaced by a noise value corresponding to the 1st percentile of all precursor intensities for each sample injection. Precursors intensities were then summed respectively by stripped sequences and accession number in order to obtain quantification at the peptide and protein level. Coefficients of variation were computed as CV % = 100 * standard deviation / mean.

For the evaluation of the linearity, only proteins corresponding to precursors identified in all 3 injection replicates of at least one of the 6 highest UPS1 concentrations (i.e. 1fmol/µg to 50fmol/µg) were considered. The quantification values from the 3 replicates were then averaged and log2 transformed. Linear regressions between log2(intensity) and log10(UPS1 concentration) were then calculated and their corresponding coefficient of determination (r^2) reported.

All the 28 possible pairwise comparisons between two UPS1 concentrations have been performed to assess the accuracy of the quantification, as well as the sensitivity and sensitivity of each workflow in a differential expression analysis. For each comparison, only peptides having 3 quantification values before imputing missing values in one of the 2 groups to compare were summed to obtain a quantification at protein level. Protein ratios and the corresponding Fold Change values (FC = log2(ratio)) were calculated for each of the 28 comparisons and the ratio accuracy was assessed by computing the Mean Absolute Percentage Error (MAPE % = 1/n * (|Expected FC - Experimental FC| / expected FC) * 100). Welch tests *p*-values were also calculated and adjusted by the Benjamini-Hochberg method for multiple testing, and the resulting *q*-values were used to plot ROC curves using the pROC R package[24]. Finally, we considered a protein as variant (i.e. differentially expressed) between two groups of samples if |*z*-score| > 1.96 and if *q*-value

< 0.05, and classified them as true positives (TP, variant UPS1 proteins), true negatives (TN, non-variant *E.coli* proteins), false positives (FP, variants *E.coli* proteins) and false negatives (FN, non-variants UPS1 proteins). Sensitivity (%) = TP / (TP+FN) * 100 and False Discovery Proportion FDP (%) = FP / (FP+TP)) * 100 were reported.

RESULTS AND DISCUSSION

DIA analysis of the E.coli-UPS1 proteomic standard

In order to obtain a complex sample reflecting the difficulties encountered when analyzing whole proteomes by bottom-up proteomics, especially the detection and quantification of low abundant species, we used an *E.coli* whole protein extract background spiked-in with 8 different concentrations of 48 human proteins standard (UPS1, Sigma), ranking from 0.1fmol to 50fmol per µg of *E.coli* (0.1, 0.25, 1, 2.5, 5, 10, 25 and 50fmol/µg) (Fig.1A) and trypsin digested according to the protocol described in experimental section. Prior to DIA experiments, a spectral library was created to be further used for peptide-centric analysis strategy. This library was generated by the analysis in DDA mode of 48 fractions of an *E.coli* protein extract, trypsin digested and fractionated on a high pH reversed-phase chromatography and a single injection of 200fmol UPS1 spiked in one microgram of *E.coli*. The final spectral library contains a total of 2892 proteins and 30656 peptides allowing a deep proteome coverage for further DIA analysis.

Moreover, the DDA single injection of 200fmol UPS1 in *E.coli* was used to optimize the mass range and window sizes to be used in DIA experiments by plotting the distribution of m/z precursors across the LC gradient (Fig.1B). This non-homogeneous distribution highlights the importance of adjusting DIA windows to obtain a maximum coverage of the eluted peptides. However, DIA acquisition parameters have to be a compromise between chromatographic resolution (cycle time), sensitivity (window size) and sequencing coverage (mass range). In order to obtain the minimum of 10 measurements per chromatographic peak required for an accurate quantification, while acquiring MS2 spectra at 15k resolution in the Orbitrap (transient time of 32ms), a maximum number of 75 DIA can be used in our conditions. Thus, we obtain a cycle time around 2.5s for 20-30s chromatographic peak elution.

Originally, the SWATH-MS analysis described by Gillet *et al.*[9] used 32 scans of 25m/z, allowing to sequence peptides from 400m/z to 1200m/z. In that case, using wide DIA windows allowed a large coverage of the mass range but with the effect to select a large number of co-eluting peptides for fragmentation, resulting in highly complex chimeric MS2 spectra. By setting narrower DIA windows, more ions from the same species are accumulated prior to fragmentation, increasing the sensitivity of the analysis and reducing the complexity of the resulting MS2 spectra. However, the covered mass range has to be reduced to keep an appropriate cycle time. To equally distribute the number of precursors in the different isolation windows, Zhang *et al.*[25] proposed to set different window sizes across the mass range, as a compromise between sensitivity (window size) and coverage of the sequenced mass range. Also, Amodei *et al*.[17] have demonstrated an increase of sensitivity without sacrificing the mass range selected for fragmentation, while using overlapping windows including two different cycles of 20m/z DIA windows shifted by 10m/z.

Taken into account these methodologies described in the literature and based on the observed precursors repartition during the chromatographic run of our DDA experiment on UPS1 spiked in *E.coli* extract (Fig.1B), 4 different DIA acquisition modes have been selected : 8m/z isolation windows for the *Narrow* scheme, 15m/z isolation windows for the *Wide* scheme, combination of 8m/z and 15m/z windows for *Variable* scheme and 2 cycles of 8m/z isolation windows shifted by 4m/z for the *Overlap* scheme. It is to note that the use of *Narrow* and *Overlap* DIA windows allow us to cover 85.6% of the precursors identified in DDA with the same gradient, while *Variable* and *Wide* schema respectively cover 98.7% and 99.9% of them. These different acquisition schemes allowed us to assess how the mass range selected for fragmentation and the complexity of the MS/MS spectra may influence DIA signal extraction and quantification.

Finally, six different proprietary or open-source software tools (DIA-NN[15], DIA-Umpire[14], OpenSWATH[12], Scaffold-DIA (Proteome Software, Inc.), Skyline[22] and Spectronaut-Pulsar-X (Biognosys AG)) were used to perform protein identification and quantification. A total of 36 proteomic workflows have been benchmarked, corresponding to *Narrow*, *Wide*, *Variable* and *Overlap* schemes treated either in a spectrum-centric manner, further called *Fasta* (DIA-NN, DIA-Umpire, Scaffold-DIA, Spectronaut-Pulsar-X) and or in a

peptide-centric analysis manner, further called *Library* (DIA-NN, OpenSWATH, Scaffold-DIA, Skyline, Spectronaut-Pulsar-X). To accurately benchmark all the workflows, data treatment from precursors to proteins (data normalization, imputation of missing values, filtering and protein aggregation) were done in the same way for all the DIA tools using R, except for tools having their own normalization process (see experimental procedures, Supp.Figure.S1 and S2). It is to note that an unexplained issue in the DIA signal extraction occurred with OpenSWATH in *Variable* acquisition, that was not observed for other tools or acquisition schemes and that had an impact on the final results (Supp.Fig.S2B).

Protein and peptide identifications

In order to assess the ability of our workflows to identify a large number of protein species in a complex biological sample and their efficiency to detect low abundant species, the number of *E.coli* and UPS1 detected in each of the 8 concentrations was reported in figure 2.

Background E.coli *identifications*

Since the *E.coli* background is expected to be the same for all analyzed samples, we reported in the average of *E.coli* proteins and peptides across the 8 tested UPS1 concentrations for each workflow (Fig.2A and B, left panels). We can observe that all the software tools better perform with narrow DIA windows than with the other acquisition schemes. Indeed, an average of 20.4% more proteins have been identified with the *Narrow* scheme as compared to the three others (Supp.Table.S2A). This can be explained by the fact that, with *Narrow* scheme, despite the reduced precursor mass range coverage, the lower complexity of MS/MS spectra seems to help the software algorithms to extract DIA MS2 signals from the library or to deconvolute the MS2 spectra in library-free approaches, thus improving the detection of peptides. With this acquisition mode, an average of 1900 *E.coli* proteins were identified with Spectronaut-Pulsar-X-*Library*, respectively 1894 and 1824 with DIA-NN-*Library* and *Fasta*, 1730 with Scaffold-DIA-*Library*, 1564 with Spectronaut-Pulsar-X-*Fasta*, 1487 with Skyline, 1459 with Scaffold-DIA-*Fasta*, 1135 with OpenSWATH and 1108 with DIA-Umpire. Moreover, with this *Narrow* acquisition scheme, a better protein sequence coverage is reached with

Spectronaut-Pulsar-X-*Fasta* and Scaffold-DIA-*Fasta* than with the other tools, with in average respectively 7.2 and 7.1 peptides per protein, while 6.9, 6.8, 6.3, 6.1, 5.8, 4.6 and 2.5 peptides per protein were identified with Scaffold-DIA-*Library*, Spectronaut-Pulsar-X-*Library*, DIA-NN-*Library*, DIA-NN-*Fasta*, DIA-umpire, Skyline and OpenSWATH respectively. For the three tools working in both spectrum- and peptide-centric modes, the use of the spectral library to extract DIA data allowed us to improve the number of protein identifications by average of 5.9%, 18% and 89% with DIA-NN, Spectronaut-Pulsar-X and Scaffold-DIA respectively. It is to note that Scaffold-DIA encountered difficulties to extract signals with wide isolation windows in spectrum-centric mode. It is particularly visible for *Wide* and *Variable* acquisitions schemes (the last one being a combination of both narrow and wide windows), with respectively 202% and 122 % more identified proteins when using the spectral library.

UPS1 proteins identifications

Overall, the best numbers of protein and peptide identification are again obtained with the *Narrow* acquisition scheme. With this one, at the highest UPS1 concentration (50fmol/µg), all the 48 human proteins could be identified with DIA-NN, Scaffold-DIA-*Fasta,* Skyline and Spectronaut-Pulsar-X-*Library*. A maximum of 47 UPS1 proteins were identified for DIA-Umpire and Spectronaut-Pulsar-X-*Fasta*, 46 for Scaffold-DIA-*Library* and 39 for OpenSWATH (Fig.2A and 2B, right panels; Supp.Table.S2A). Even if almost all UPS1 proteins are identified at high concentrations, the number of UPS1 identifications rapidly decreases with the amount of spiked-in UPS1, for Skyline, DIA-Umpire and OpenSWATH. This reveals the difficulties for these tools to extract signals and thus to identify peptides and proteins of low abundance species. Conversely, at very low concentrations of UPS1 (i.e. 0.1fmol and 0.25fmol/µg), Scaffold-DIA still reported an important number of identifications with up to 28 UPS1 proteins. This is also the case for Spectronaut Pulsar X-*Fasta*, which allowed the identification of a maximum of 11 and 17 proteins, respectively for 0.1fmol and 0.25fmol/µg UPS1. These high number of identifications at very low concentrations might be explained by the LC-MS runs realignment and by the low rate of missing values obtained by these tools. Indeed, the missing values percentages in these datasets are between 1.2 and 7.7% whereas in the

other ones, this percentage varies from 9.6% to 33.1% (Supp.Fig.S2A). Finally, we observed that the use of the *Overlap* and *Variable* acquisition schemes does not significantly improve the detection of low abundance species, as we could have expected.

Reproducibility between technical replicates

Since technical replicates (i.e. three injections of the same sample) have been acquired, the reproducibility of the signal extraction of the different workflows has been assessed by the calculation of coefficients of variation (CV) of the intensity values across the three replicates (Fig.3 and Supp.Table.S2B). For *E.coli*, the average of CV values for all datasets, were 13% and 10% for peptides and proteins respectively, highlighting the high reproducibility of DIA analysis for label-free quantification. The lowest CV values were found for Skyline and Spectronaut-Pulsar-X-*Library* with 4.74% and 5.39% respectively at peptide and protein levels. CVs were also computed for each concentration of UPS1 (right panels). As expected, an increase of CV values can be observed at low concentrations of UPS1, with averages of 26% and 24% at 0.1fmol/µg for peptides and proteins respectively, whereas 8.8% and 4.6% were obtained for 50fmol/µg samples. Overall, at low concentrations, the best results were obtained by the same tools, Skyline and Spectronaut-Pulsar-X-*Library* (for instance, average CV values were 9.0% and 6.5% respectively for the 2.5fmol/µg UPS1 concentration) demonstrating a better ability of these software tools to properly extract signals from low abundance species. However, no CV averages could be computed at very low concentrations (≤ 0.25fmol/µg) for Skyline, while Spectronaut-Pulsar-X-*Library* reported a CV average of 10.6% for the lowest tested concentration (0.1fmol/µg).

Linearity across the UPS1 protein concentrations

Another important parameter to assess the reliability of a quantification method is the linearity of the instrument response as a function of peptide concentration in the sample. Indeed, to obtain accurate ratio measurements in pairwise comparisons, this response has to be linear across a wide range of concentrations. To assess the linearities obtained with the tested acquisition modes and software processing, we plotted the intensities values obtained for the different UPS1 concentrations. Because the limit of detection of

these proteins is reached at 1fmol/µg UPS1 in *E.coli* with most of the tools (Fig.2), only the concentrations above this point were retained. Thus, linear regressions were calculated for each UPS1 protein across the 6 highest concentrations (1fmol/µg to 50fmol/µg) and the distribution of the corresponding coefficients of determination (r^2) was plotted (Fig.4 and Supp.Table.S2C). Surprisingly a large disparity can be observed between the tested tools, with less dependency on the acquisition scheme. Indeed, DIA-NN, Scaffold-DIA and Spectronaut-Pulsar-X display high r^2 average values of 0.964, 0.980 and 0.959 respectively, suggesting a very strong correlation between the amount of spiked-in UPS1 proteins and the intensities extracted by the software, while DIA-Umpire, OpenSWATH and Skyline reported lower values (0.858, 0.844 and 0.886 on average respectively). For most of the tools, better linearities for UPS1 proteins were obtained with data acquired with *Narrow* DIA windows, suggesting again that the lower complexity of MS2 spectra acquired with narrower windows might help to properly detect and quantify low abundance signals and therefore improve the overall linearity response. For software tools performing both extraction modes, DIA-NN reported better linearities with the spectral library (0.955 vs 0.973 on average, *Fasta* vs *Library*) and no significant difference was noticed for Scaffold-DIA (0.981 vs 0.980). Conversely, Spectronaut-Pulsar-X performed slightly better without the spectral library (0.961 vs 0.956).

Ratio accuracy in pairwise comparisons

While the aim of most of the proteomics studies is to highlight significant quantitative differences in complex proteomes by performing differential protein expression analysis, the accuracy of the ratio measurement in pairwise comparison is an important factor to consider. To assess this parameter for each of the 36 tested workflows, protein ratios have been calculated for each possible pair of the 8 UPS1 concentrations analyzed, resulting in a total of 28 pairwise comparisons.

Because UPS1 proteins are spiked-in different and known concentrations, the accuracy of the protein ratio measurement can be assessed by comparison of the observed ratio to the expected one. Thus, we calculated the Fold-Change (FC = log2(ratio)) for each protein in a pairwise comparison and reported the corresponding Mean Absolute Percentage Error to the expected FC (Fig.5 and Supp.Table.S2D). The overall tendency

is to get higher errors for pairwise comparisons involving one, and even more two, low UPS1 concentrations. On average for all workflows, 12.6% and 42.6% errors were respectively observed for 50fmol/10fmol (expected ratio = 50) and 50fmol/0.1fmol (expected ratio = 500) comparisons. When the expected ratio is kept at 50 but using lower UPS1 concentrations (5fmol/1fmol), an average of 27.6% error is reached. Such a distortion of the ratio values might be due to the difficulty for the tools to distinguish the signal from the noise for low abundance species. Indeed, when very low concentrations of UPS1 are used (*i.e.* both below 2.5fmol/µg) the average MAPE reaches 41.5% to 87.9% for all tested workflows. However, this difficulty to recover low abundant signals does not seem related to the size of the DIA windows, since no significant differences can be observed between MAPE computed for *Narrow* and for *Wide* acquisition schemes. Across the 28 ratios, Scaffold-DIA-*Library* reports the better ratio accuracy with a MAPE average of 23.6%, followed by DIA-NN-*Library* (25.2%), DIA-NN-*Fasta* (27.1%), Scaffold-DIA-*Fasta* (27.7%), Spectronaut-Pulsar-X-*Library* and -*Fasta* (28.7% and 34.4% respectively), Skyline (35.1%), DIA-Umpire (36.6%) and OpenSWATH (38.8%). Since the best accuracies are obtained for tools having their own normalization process, one can assume that the way of normalizing the data is an important parameter to consider. Although the median normalization we used for the other tools seems to be efficient (SuppFig.S2B), other methods of normalization might be applied. For certain tools (DIA-NN, Skyline and OpenSWATH), we can also observe the highest errors for comparisons involving close UPS1 concentrations (*i.e.* ratios of 2 or 2.5) even when medium concentrations (5 to 10fmol/µg) are used. This observation reveals an overall lower quantification accuracy for those tools which especially affects the computation of low protein ratios and might eventually have an impact on the detection of low variations in differential protein expression studies.

Sensitivity and specificity in differential expression analysis

The use of high-performance and fast-scanning LC-MSMS instruments now enables the identification and quantification of several hundreds or thousands proteins in a single dataset as it is the case in this study. However, the final objective of a differential proteome analysis is to detect, in this large amount of data, as many truly variant proteins

as possible but with a low number of false positives by controlling the False Discovery Rate (FDR). Our study, in which variant proteins (UPS1) are spiked-in at various concentrations in a fixed background (*E.coli*), is a good opportunity to assess this FDR for each of the different acquisition schemes and software tools used. To do so, we used Receiver Operating Characteristic (ROC) curves of the *q*-values and reported their corresponding Areas Under the Curve (AUC) (Fig.6A and Supp.Table.S2E). An AUC of 1 represents a perfect distinction between the two groups compared (i.e. UPS1 variant and *E.coli* non-variant proteins), while an AUC of 0.5 represents a model which is not able to distinguish the two groups. In contrast, AUC = 0, the model is reversing the classes. As expected and already noticed for the assessment of ratio accuracy, all the software tools seem to be less able to differentiate UPS1 proteins (variant) from the *E.coli* background proteins (non-variant) when comparing low or very low concentrations of UPS1 (2.5fmol/µg and below). Indeed, on average for all the experiments, AUC dropped from 0.954 on average (all workflows and comparisons) for high concentrations (ie. involving at least one UPS1 concentration ≥ 5 fmol/µg) to less than 0.761 for low concentrations (both concentrations ≤ 2.5fmol/µg). It can be explained by the fact that at these concentrations, only few UPS1 proteins are identified and the identified ones are quantified with a lower accuracy than for higher concentrations. We can also notice some particularities for certain tools. For example, OpenSWATH-*Variable* reports low AUC values due to the issue in signal extraction previously described; DIA-NN-*Fasta* reported a poor AUC for comparisons involving the 50fmol/µg UPS1 concentration when acquired with *Wide* scheme. But, surprisingly, this issue is only observed when using the *Fasta* mode and not in *Library* mode despite the use of the same raw files. All the other tools show approximately the same capacity to distinguish UPS1 proteins from *E.coli* ones, better AUC can be obtained with *Narrow* and *Variable* acquisition schemes for both *Fasta* and *Library* modes, highlighting once again the importance to reduce the complexity of MS2 spectra. In the end, it is also important to mention that we used three technical replicates for each sample but that better AUC values might have been obtained with more replicates since the *q*-values are dependent on this parameter.

Since the only use of a *q*-value cut-off (often set at $q = 0.05$) to define variants proteins in a proteomic analysis may lead to a high rate of false positives, in most of studies, another

filtering criterion based on Fold Change or *z*-score is usually applied. *z*-score calculation (*z* = (FC - FC average / FC standard deviation) allows to transform the distribution of fold changes into a normal distribution and avoiding biases related to not centered distribution while setting an arbitrary FC cut-off. To determine the efficiency of our workflows to distinguish UPS1 variant proteins from the background, we chose a double cut on *q-value* < 0.05 and |*z*-score| > 1.96 (corresponding to the external boundaries of 95% of the FC distribution) and plotted the corresponding specificity and False Discovery Proportion (FDP) obtained by counting the number of true and false positives and negatives (see calculation in the experimental section) (Fig.6B and Supp.Table. S2F). For the *narrow* acquisition scheme, and on average for all the ratios, Spectronaut-Pulsar-X reported the best combination 'Sensitivity - FDP' (respectively 68.6% - 5.2% for *Fasta*, and 69.4% - 10.3% for *Library*), as well as DIA-NN-*Library* (60.1% - 7.7%). High sensitivity was also reached by DIA-Umpire and Skyline but with a higher FDP (respectively 71.8% - 29.3% and 65.9% - 38%). Conversely, Scaffold-DIA reported less sensitivity but with extremely low FDP (44.6% - 1.8% for *Fasta*, 34.5% - 2.8% for *Library*). Depending on the acquisition scheme and software used, more stringent or relaxed criteria could be chosen to maximize the detection of variant proteins while keeping a low false discovery proportion.

CONCLUSIONS

In this study, using a mix of *E.coli* background and UPS1 proteins at eight different concentrations, we were able to challenge DIA proteomic workflows for protein identification and quantification by recovering thousands of measurements to assess their performance and reliability. Although certain software tools generally performed better than others in our conditions, each of them showed their own strengths and weaknesses in terms of identification, linearity, reproducibility and accuracy.

The use of four different acquisition schemes showed that the size of DIA windows is an important parameter to consider when performing DIA experiments. Indeed, overall, narrow windows performed better than wide ones, despite a lower coverage of the mass range. This observation suggests that the improvement of Orbitrap scanning speed on the most recent instruments should participate in the improvement of DIA analysis

performance by the use of a larger mass range or the reduction of window width. Moreover, our results also show that, in our conditions, some published acquisition schemes (overlapped or variable windows) do not perform significantly better than traditional ones.

Our study also demonstrates that the use of a very large spectral library acquired from 48 fractions of *E.coli*, does not strongly improve the number of identified and quantified proteins. Indeed, it seems that the signal of low abundance peptides detected by this extensive fractionation is not easily recovered by DIA processing tools on unfractionated DIA acquisition. This finding is confirmed by recently published articles and interrogates on the usefulness to generate time-consuming libraries prior to DIA analysis[26,27]. However, other types of libraries such as repetitive measurement of non-fractionated sample library[27] or chromatogram library[28] might improve the results shown here. Furthermore, publicly-available libraries[29], gas-phase fractionation libraries[30] or *in silico* generated libraries[31] could also be used to avoid performing additional injections.

Finally, this work shows that the most important parameter which can affect the result of a DIA study is the choice of the processing software, followed by the acquisition scheme (size and number of DIA windows) and finally the use or not of a spectral library. Nevertheless, in addition to their performance, the choice of using a DIA software, requires to consider their cost (Spectronaut-Pulsar-X and Scaffold DIA are proprietary, when the others are open source), their ease of use (OpenSWATH and DIA-Umpire do not have a graphical user interface and require several steps that might be difficult to install and execute without bioinformatical knowledges) and the availability of an integrated intensity normalization (it is the case for Spectronaut-Pulsar-X, Scaffold-DIA and DIA-NN).

In the end, by analyzing 96 DIA raw files with 36 workflows and performing 28 pairwise comparisons for each of them, our study is the largest benchmark of DIA methods ever published on Orbitrap mass spectrometers, that are the most widely used instruments in the proteomic community. Therefore, this work could be a very helpful guide book for the scientists to efficiently choose the best acquisition parameters and software processing for their DIA study, but also through the availability of all our data on ProteomeXchange repository, it will make them able to reapply their own statistics or test other software tools

and it could help developers to improve their algorithms, with the final objective for all of them to maximize the capabilities of DIA strategy in proteomics studies.

DATA AVAILABILITY

All raw mass spectrometry data, software outputs and quantification results are publicly available on ProteomeXchange repository (www.proteomexchange.org) with the identifier PXD022206.

ACKNOWLEDGMENTS

The authors are grateful to David Bouyssié for critical reading of the manuscript.

AUTHORS CONTRIBUTIONS

C.G and F.R.D. designed experiments; C.G., C.J.B., M.L. and L.M. performed experiments; C.G. and F.R.D. performed data analysis; C.G and F.R.D. wrote the manuscript; A.D. supervised research.

COMPETING INTERESTS

The authors declare no competing interests.

REFERENCES

1. Müller, J. B. *et al.* The proteome landscape of the kingdoms of life. *Nature* **582**, 592–596 (2020).
2. Gillet, L. C., Leitner, A. & Aebersold, R. Mass Spectrometry Applied to Bottom-Up Proteomics: Entering the High-Throughput Era for Hypothesis Testing. *Annu. Rev. Anal. Chem.* **9**, 449–472 (2016).
3. Zhang, Y., Fonslow, B. R., Shan, B., Baek, M.-C. & Yates, J. R., 3rd. Protein analysis by shotgun/bottom-up proteomics. *Chem. Rev.* **113**, 2343–2394 (2013).
4. 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 (2014).

5. Lange, V., Picotti, P., Domon, B. & Aebersold, R. Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **4**, 222 (2008).
6. Picotti, P. & Aebersold, R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat. Methods* **9**, 555–566 (2012).
7. Peterson, A. C., Russell, J. D., Bailey, D. J., Westphall, M. S. & Coon, J. J. Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Mol. Cell. Proteomics* **11**, 1475–1488 (2012).
8. 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).
9. Gillet, L. C. *et al.* Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* **11**, O111.016717 (2012).
10. Ludwig, C. *et al.* Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **14**, e8126 (2018).
11. Bruderer, R. *et al.* Optimization of Experimental Parameters in Data-Independent Mass Spectrometry Significantly Increases Depth and Reproducibility of Results. *Mol. Cell. Proteomics* **16**, 2296–2309 (2017).
12. Röst, H. L. *et al.* OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nat. Biotechnol.* **32**, 219–223 (2014).
13. Egertson, J. D., MacLean, B., Johnson, R., Xuan, Y. & MacCoss, M. J. Multiplexed peptide analysis using data-independent acquisition and Skyline. *Nat. Protoc.* **10**, 887–903 (2015).
14. Tsou, C.-C. *et al.* DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nat. Methods* **12**, 258–64, 7 p following 264 (2015).
15. Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser, M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat. Methods* **17**, 41–44 (2020).

16. 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 (2015).
17. Amodei, D. *et al.* Improving Precursor Selectivity in Data-Independent Acquisition Using Overlapping Windows. *J. Am. Soc. Mass Spectrom.* **30**, 669–684 (2019).
18. Navarro, P. *et al.* A multicenter study benchmarks software tools for label-free proteome quantification. *Nat. Biotechnol.* **34**, 1130–1136 (2016).
19. Tu, C., Li, J., Sheng, Q., Zhang, M. & Qu, J. Systematic assessment of survey scan and MS2-based abundance strategies for label-free quantitative proteomics using high-resolution MS data. *J. Proteome Res.* **13**, 2069–2079 (2014).
20. Ramus, C. *et al.* Benchmarking quantitative label-free LC-MS data processing workflows using a complex spiked proteomic standard dataset. *J. Proteomics* **132**, 51–62 (2016).
21. Välikangas, T., Suomi, T. & Elo, L. L. A comprehensive evaluation of popular proteomics software workflows for label-free proteome quantification and imputation. *Brief. Bioinform.* **19**, 1344–1355 (2018).
22. MacLean, B. *et al.* Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966–968 (2010).
23. Team, R. C. RA language and environment for statistical computing. Versión 3.4. 3, Vienna, Austria, R Foundation for Statistical Computing. (2017).
24. Robin, X. *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**, 77 (2011).
25. Zhang, Y. *et al.* The Use of Variable Q1 Isolation Windows Improves Selectivity in LC-SWATH-MS Acquisition. *J. Proteome Res.* **14**, 4359–4371 (2015).
26. Govaert, E. *et al.* Comparison of fractionation proteomics for local SWATH library building. *Proteomics* **17**, (2017).
27. Barkovits, K. *et al.* Reproducibility, Specificity and Accuracy of Relative Quantification Using Spectral Library-based Data-independent Acquisition. *Mol. Cell. Proteomics* **19**, 181–197 (2020).

28. Searle, B. C. *et al.* Chromatogram libraries improve peptide detection and quantification by data independent acquisition mass spectrometry. *Nat. Commun.* **9**, 5128 (2018).
29. Rosenberger, G. *et al.* A repository of assays to quantify 10,000 human proteins by SWATH-MS. *Sci Data* **1**, 140031 (2014).
30. Pino, L. K., Just, S. C., MacCoss, M. J. & Searle, B. C. Acquiring and Analyzing Data Independent Acquisition Proteomics Experiments without Spectrum Libraries. *Mol. Cell. Proteomics* **19**, 1088–1103 (2020).
31. Yang, Y. *et al.* In silico spectral libraries by deep learning facilitate data-independent acquisition proteomics. *Nat. Commun.* **11**, 146 (2020).

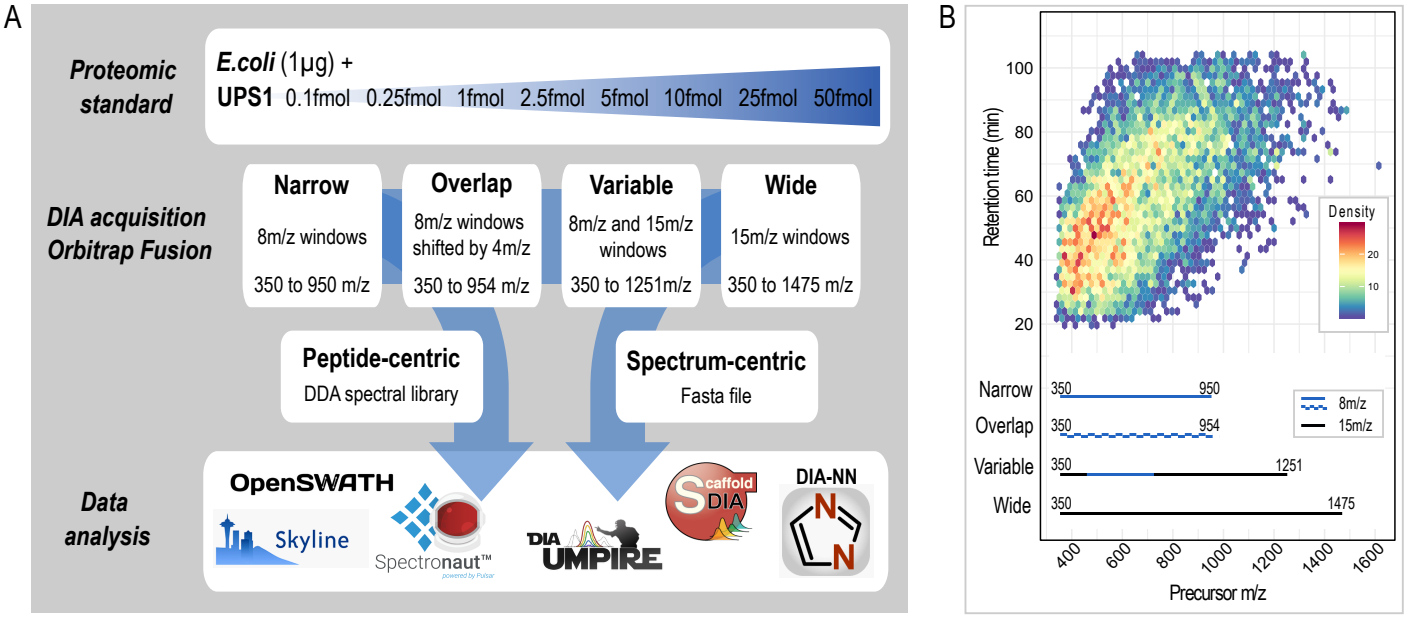


Figure 1 - Proteomic standards and processing workflows - (A) The proteomic standards, consisting in 8 concentrations UPS1(Sigma) spiked in *E.coli* background were injected in LC-MS/MS with 4 different DIA acquisition modes. Files were then processed with 6 different softwares, working in peptide-centric or spectrum-centric mode. (B) Distribution, along the mass range and the analysis time, of ion precursors from an *E.coli* whole cell extract spiked-in with 200fmol of UPS1 and acquired in DDA mode. Selection of windows schemes for DIA analysis is shown below the graph.

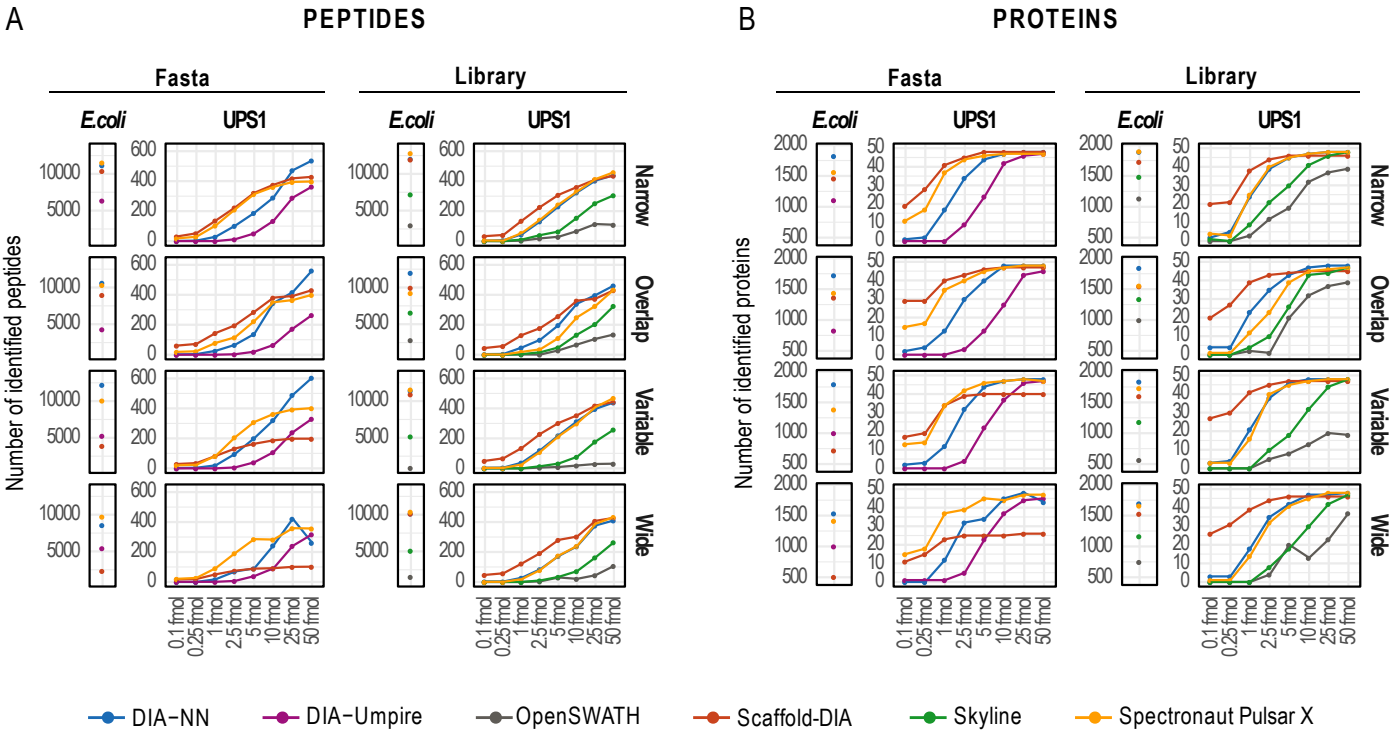


Figure 2 - Peptides and proteins identifications - For each workflow (acquisition scheme associated to one software in 'fasta' or 'library' mode), the number of UPS1 (right panels) or *E.coli* (left panels) peptides (A) or proteins (B) found in the three replicates of each sample (i.e. without missing values) are reported. For *E.coli* proteins and peptides, the average of identifications from the 8 samples of a same dataset (corresponding to one workflow) is given.

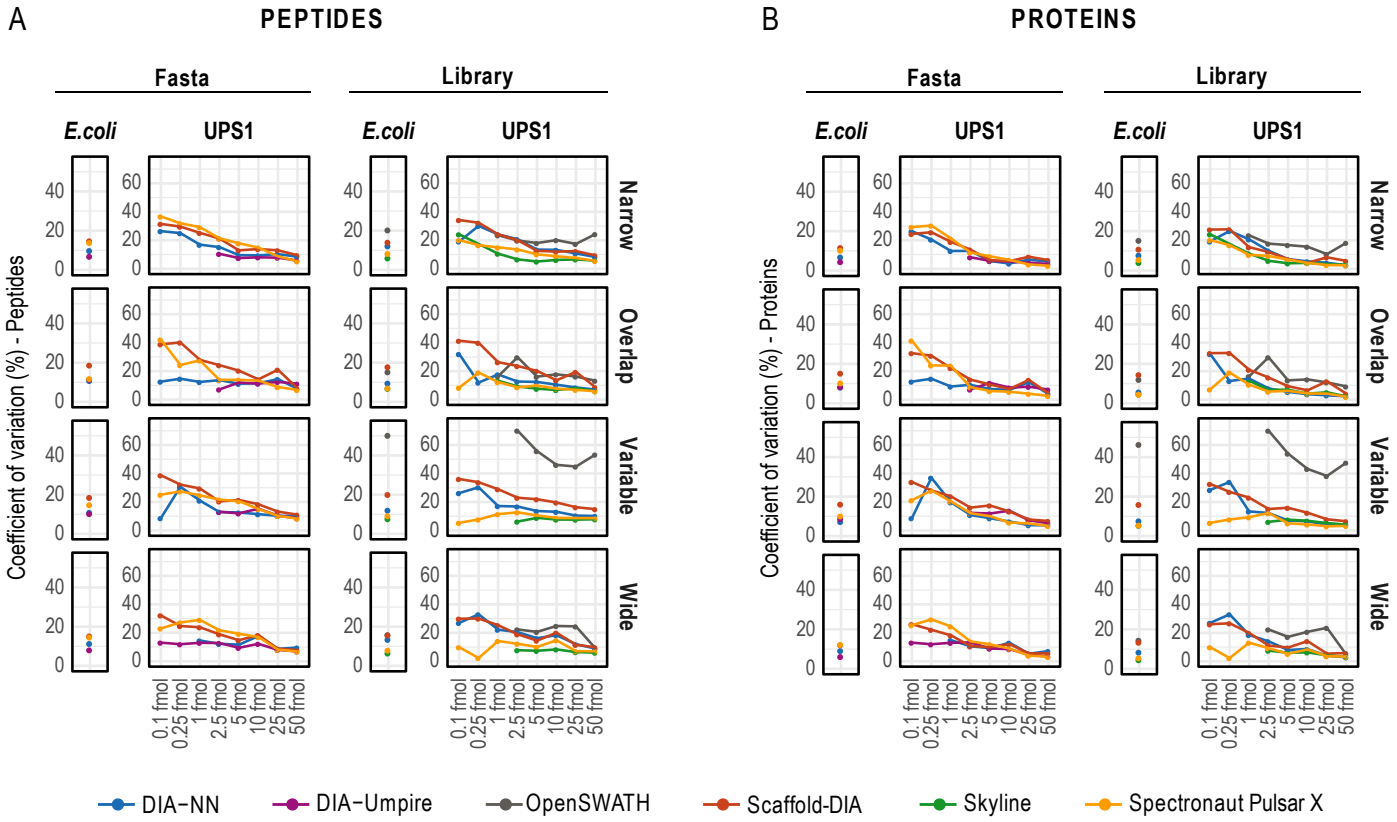


Figure 3 - Reproducibility of signal extraction - For each workflow (acquisition scheme associated to one software in fasta or library mode), the coefficients of variation were calculated over the three injection replicates of the same sample for UPS1 (right panels) or *E.coli* (left panels) peptides (A) or proteins (B). Only precursor intensities found in the three replicates before missing values imputation were used for this calculation and summed by stripped sequence and accession number in order to obtain respectively quantification at the peptide and protein levels. For *E.coli* proteins and peptides, the average of coefficient of variation from the 8 samples of a same dataset (corresponding to one workflow) is given.

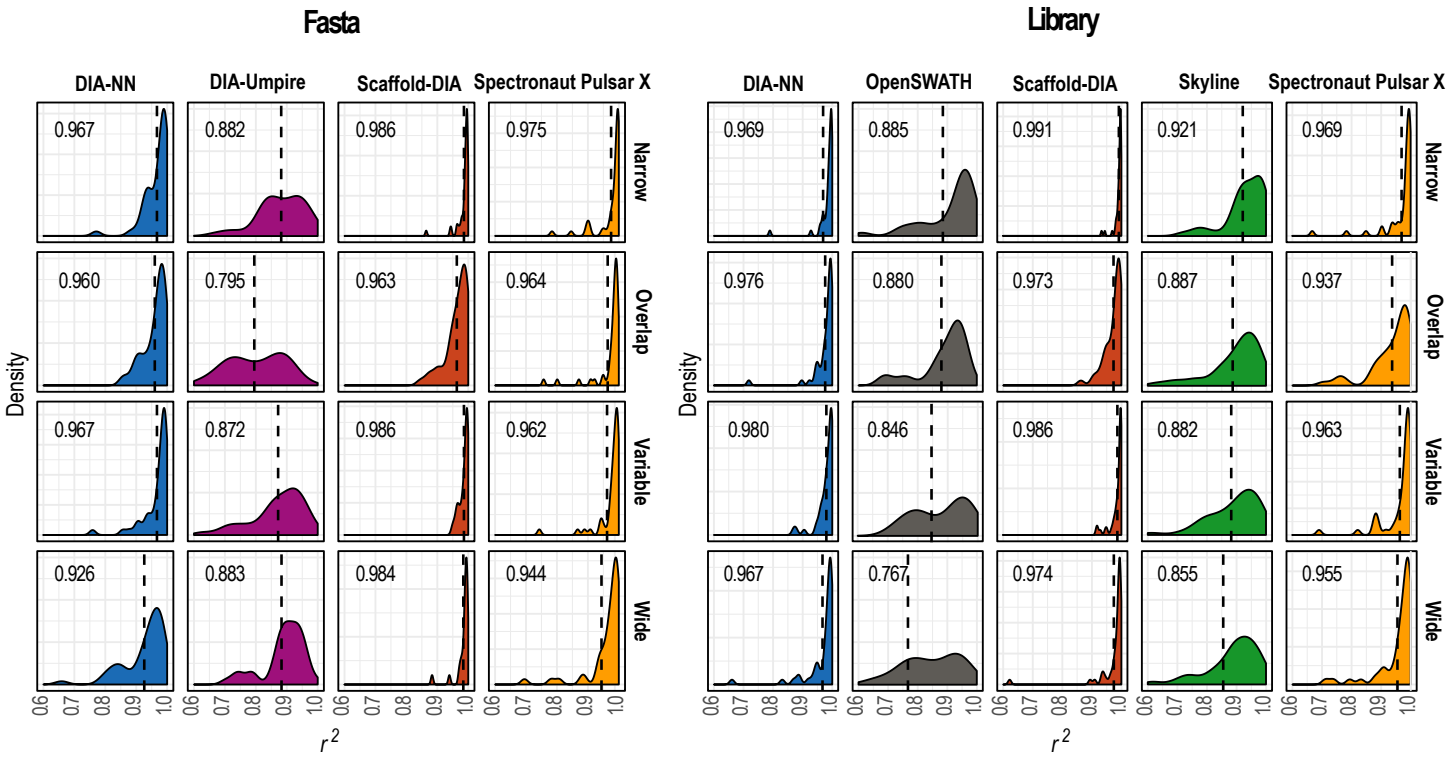


Figure 4 - Linearity across UPS1 concentrations - For each workflow (acquisition scheme associated to one software in fasta or library mode), linear regressions have been plotted using quantification values of each UPS1 protein across their 6 highest concentrations in *Ecoli* background and the distribution of their coefficients of determination (r^2) were reported. Only proteins detected in the three replicates of at least one of the 6 concentration samples were kept. The dotted lines and values reported for each distribution represent the average of the r^2 values of this distribution.

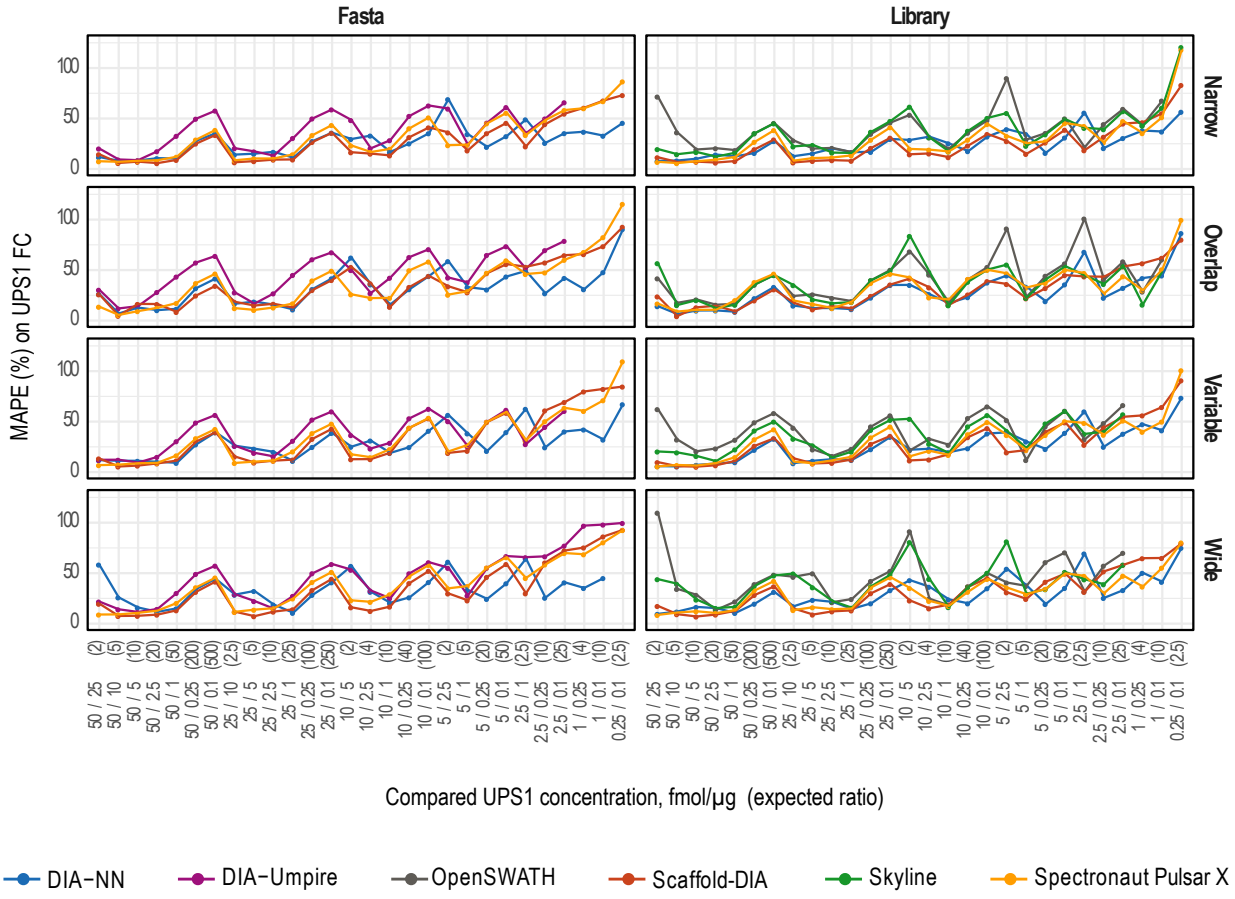


Figure 5 - Accuracy of protein ratios - For each workflow (acquisition scheme associated to one software in fasta or library mode), Mean Absolute Percentage Error (MAPE) on fold changes (FC) calculated for UPS1 proteins in 28 pairwise comparisons. MAPE (%) = 1/n * (|Expected FC - Experimental FC| / expected FC) * 100.

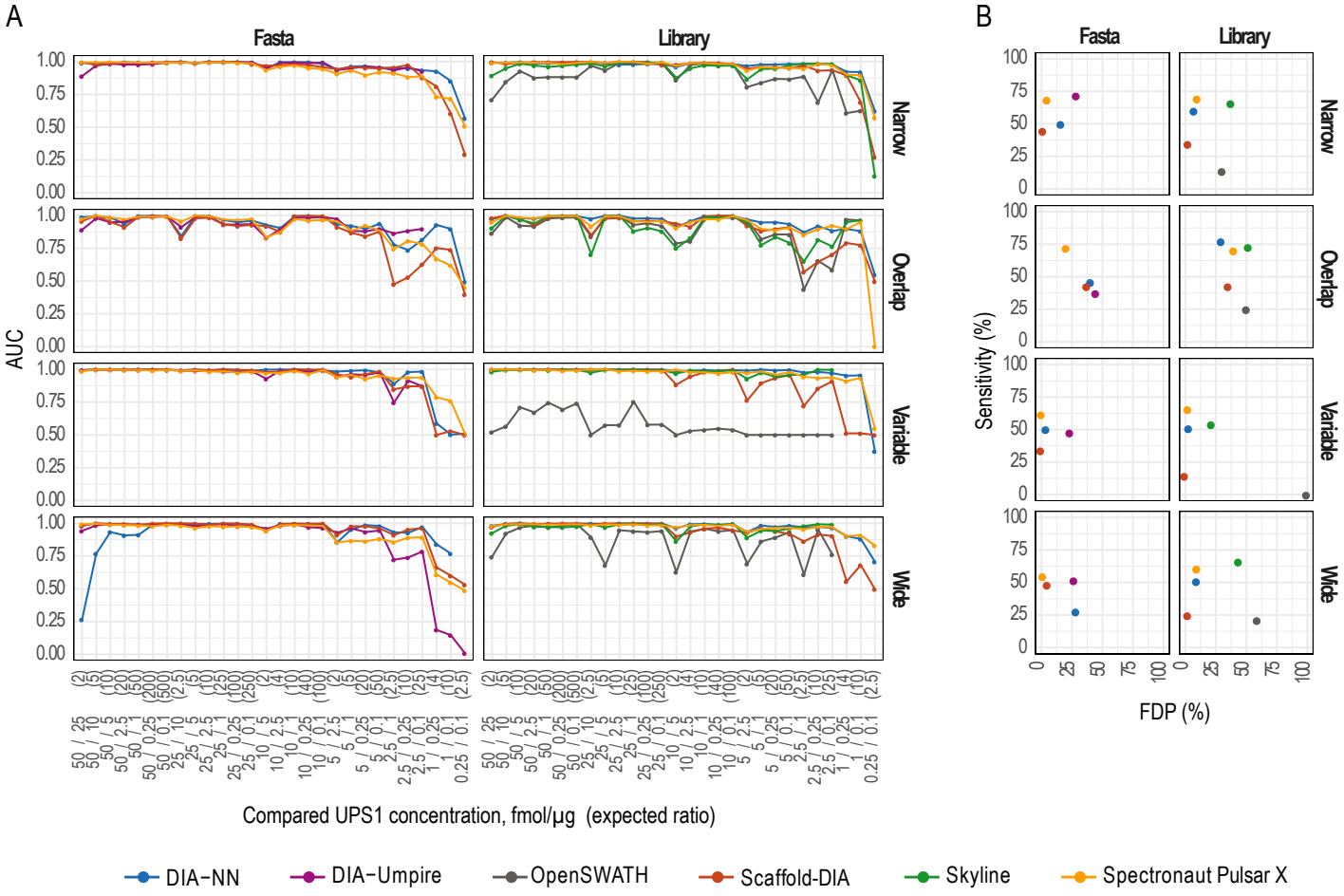
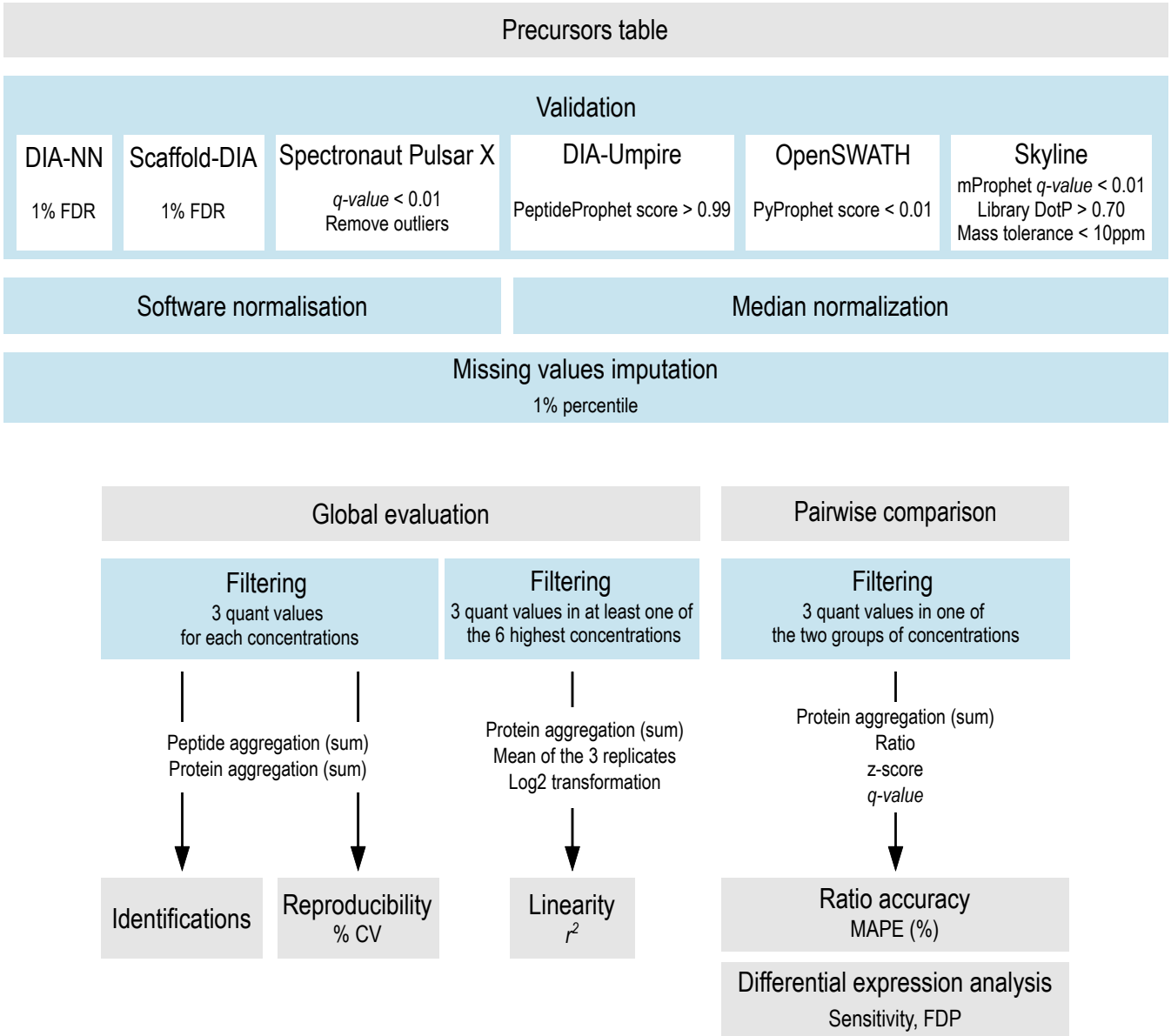
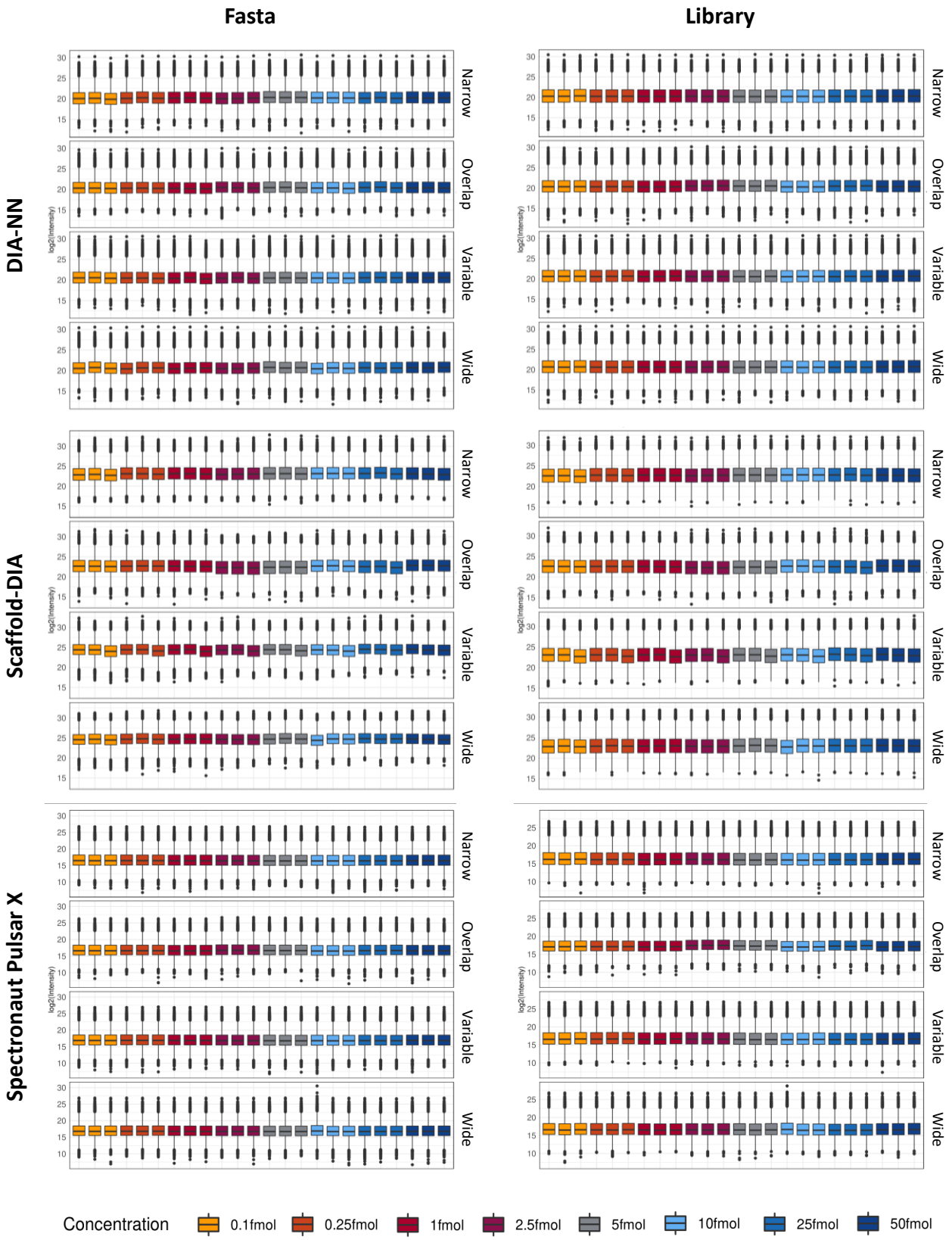


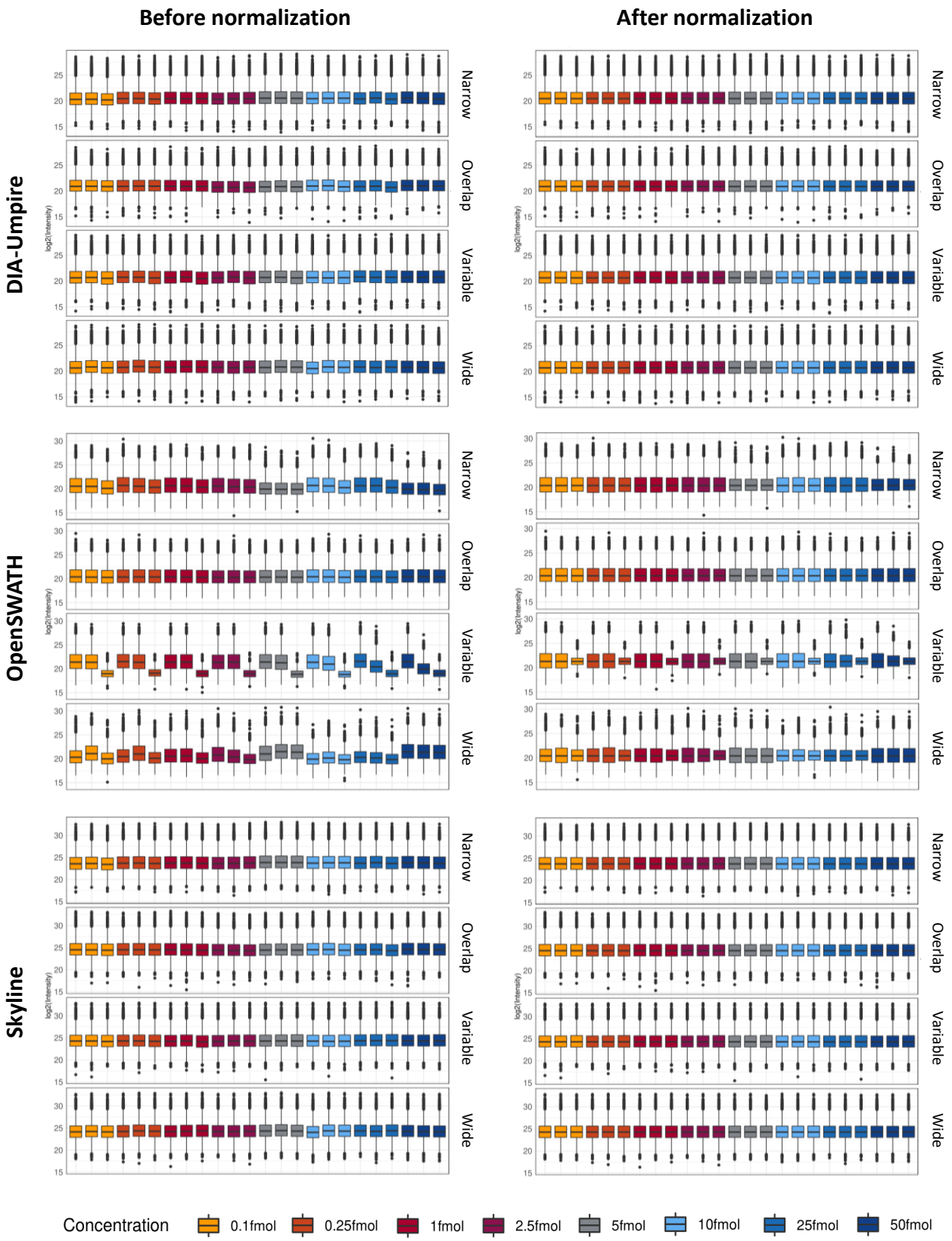
Figure 6 - Sensitivity and specificity of the differential expression analysis - For each workflow (acquisition scheme associated to one software in fasta or library mode), (A) Areas Under the Curve corresponding to Receiver Operating Characteristic (ROC) curves on *q-values* plotted for 28 pairwise comparisons. (B) Average True Positive Rate (Sensitivity) and False Discovery Proportion (FDP) obtained from 28 pairwise comparisons with cut-offs on |z-score| > 1.96 and on *q-value* < 0.05. Sensitivity (%) = TP / (TP+FN) * 100 and FDP (%) = TP / (TP+FP) * 100 where TP (True Positives) = UPS1 proteins found as variant, TN (True Negatives) = *E.coli* proteins found as non-variant, FP (False Positives) = *E.coli* proteins found as variant and FN (False Negatives) = UPS1 proteins found as non-variant.



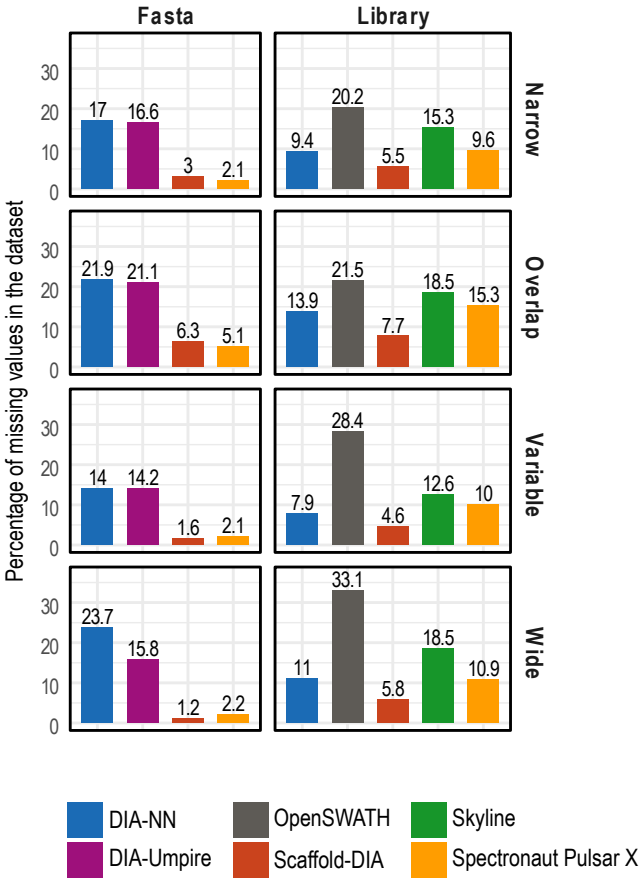
Supplementary Figure S1 - Data Treatment - Post-processing (validation, normalization and missing value imputation) applied on the outputs of DIA softwares (top panel); Filtering and calculation performed on the data to obtain identification, reproducibility, linearity, ratio accuracy, sensitivity and fale discovery proportion values.



Supplementary Figure S2A – Automatic normalization - Distribution of log2 intensities for each sample of the dataset. Normalization was performed by the software, in fasta and library mode.



Supplementary Figure S2B – Manual normalization - Distribution of log2 intensities for each sample of the dataset, before and after performing a median normalization.



Supplementary Figure S2C - Missing values in each dataset - Percentage of missing values in outputs from each workflow (acquisition scheme associated to one software in fasta or library mode).