Zou *et al. Clinical Proteomics* (2026) 23:2
https://doi.org/10.1186/s12014-025-09572-2

Clinical Proteomics

RESEARCH Open Access

In-depth analysis of data characteristics and comparative evaluation of dda and dia accuracy in label-free quantitative proteomics of biological samples

Xun Zou[1†], Lulu Wang[2,3†], Yulu Chen[1,3], Hang Fu[3], Yuan Gao[2,3], Bin Liu[1], Minjia Tan[1,2,4] and Linhui Zhai[4*]

Abstract

Data-Dependent Acquisition (DDA) and Data-Independent Acquisition (DIA) are widely used in MS-based proteomics. However, a comprehensive evaluation of their data characteristics—including protein and peptide identification, differential expression analysis, and the performance in revealing biological insights—remains lacking. In this study, we conducted a systematic comparison of DDA and DIA across three model sample types: one disease model, two drug-treated models, and their respective controls. Our analysis extended beyond conventional metrics such as total protein and peptide counts, precision, and accuracy, to include data completeness, detection of positive control markers, reproducibility, functional annotation reliability, and sources of methodological variation. The results demonstrated that DIA outperformed DDA in terms of protein identification (disease group: 7,735 vs. 5,067; drug-treated group 1: 7,987 vs. 4,605), quantitative coverage (average quantifiable protein ratio: DIA 98–99% vs. DDA 95–96%), and reproducibility (intragroup correlation coefficients: DIA > 0.98 vs. DDA 0.93–0.98). We also found DIA exhibited lower variability (intragroup CV < 10% vs. > 15% for DDA) and improved accuracy for low-abundance and housekeeping proteins. Additionally, the functional enrichment analyses further revealed DIA's superior capability in detecting pathway activation. Finally, discrepancies between DIA and DDA were primarily attributed to proteins identified with ≤ 5 peptides, the exclusion of single-peptide proteins enhanced overall data quality. Overall, this study systematically assess the overall capabilities of DDA and DIA approaches in uncovering biologically relevant findings and driving mechanistic insights within authentic pharmacological and disease models, thereby offering practical guidance for methodological choices in future research.

Keywords DDA, DIA, Mass spectrometry quantification, Proteomics, Reproducibility, Interferon pathway

†Xun Zou and Lulu Wang these authors contribute equally.

*Correspondence:
Linhui Zhai
zhailinhui@tongji.edu.cn
[1]College of Pharmacy, Jiangsu Ocean University, Lianyungang 222000, Jiangsu, China
[2]School of Chinese Materia Medica, School of Pharmacy, Nanjing University of Chinese Medicine, Nanjing, Jiangsu, China
[3]State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China
[4]Translational Research Institute of Brain and Brain-Like Intelligence, School of Medicine, Shanghai Fourth Peoples Hospital, Tongji University, Shanghai 200434, China

Introduction

In the post-genomic era, proteomics research has become essential for unraveling the complexities of life processes. Mass spectrometry (MS), due to its high-throughput capability and high accuracy in molecular mass detection, has emerged as the core technology for systematically addressing the complexity of the proteome [1]. A typical MS-based shotgun proteomics workflow includes protein extraction, enzymatic digestion into peptides, separation of peptides via liquid chromatography (LC), detection by MS, and subsequent bioinformatics analysis [2, 3]. Data-Dependent Acquisition (DDA) and Data-Independent Acquisition (DIA) are two widely used strategies in MS-based proteomics analysis. In the DDA approach, the MS detection begins with a primary mass spectrometry scan (MS1) to acquire m/z and abundance information. Then the specific precursor ions are selected for fragmentation (e.g., via Collision-Induced Dissociation, CID; Higher-energy collisional dissociation, HCD; Electron transfer dissociation, ETD), generating the MS/MS mass spectra (MS2), which facilitate peptide identification, and quantification [4, 5]. DDA has been widely adopted in discovery proteomics, protein identification, protein interactomics studies due to its high throughput capabilities. However, the stochastic nature of precursor ion selection in DDA leads to insufficient detection of low-abundance ions and reduced the reproducibility [6, 7].

To overcome the limitations of DDA, DIA has been developed and also widely adopted in large-scale cohort studies, clinical proteomics, biomarker discovery and validation, as well as in studies requiring highly reproducible quantification [8, 9]. DIA eliminates the need for precursor ion selection by utilizing fixed or dynamically cycling isolation windows [10], thereby systematically fragmenting all precursor ions within each window. This unbiased acquisition strategy improves data completeness, and reproducibility [11, 12].

With the continuous advancement of the DIA data analysis algorithms, DIA has exhibited superior efficiency in mass spectrometry identification compared to DDA. Consequently, its application has become increasingly widespread across various research fields. In recent years, numerous comparative studies for evaluating DDA versus DIA have been reported, particularly in fields demanding deep coverage and high-precision quantification, such as clinical sample analysis [13, 14]. For example, in studies using FFPE clinical samples, comparisons between the two methods in terms of quantitative completeness and depth, workflow applicability and robustness, and ability to quantify low-abundance proteins strongly favored DIA over DDA in this specific application scenario. In host–pathogen interaction research [15], results indicated that the DIA-PASEF approach significantly outperformed DDA-PASEF in nearly all aspects, particularly in identification depth, detection capability for low-abundance proteins, and quantitative accuracy, providing substantial advantages in complex host–pathogen systems. Similarly, in drug mechanism exploration [16], comparisons covered aspects such as identification depth, quantitative completeness (missing values), detection of antimicrobial peptides, and quantitative reproducibility. In cancer research [17], evaluations included identification and quantitative coverage, data completeness, quantitative precision, biological discovery capability, and methodological flexibility. For complex tissue profiling [16, 18], studies compared identification capability, sensitivity, detection and quantification limits, and data integrity. In the model microbiome, analyses were compared in terms of identification depth (number of proteins/peptides), data reproducibility, quantitative accuracy, and false discovery rate [19, 20]. Additionally, in the field of phosphoproteomics, a key area within functional proteomics, the comparative evaluation of DDA and DIA methodologies has attracted considerable attention [21]. These studies aim to assess the suitability of each approach for specific scientific questions or sample types, typically focusing on key performance indicators such as identification depth, quantitative precision, and accuracy [22, 23]. The majority of these reports emphasize the advantages of DIA over DDA in terms of the number of identified proteins and their contributions to uncovering critical biological insights. However, comprehensive comparative analyses based on label-free DDA and DIA remain limited. A systematic and in-depth evaluation of the data characteristics of both strategies-including features of identified proteins and peptides, differential protein expression analysis, and the capacity to reveal meaningful biological information-is still lacking. This gap, to some extent, hinders a more thorough understanding of the distinct data profiles associated with these two acquisition strategies.

In this study, we aim to systematically assess the overall capabilities of DDA and DIA approaches in uncovering biologically relevant findings and driving mechanistic insights within authentic pharmacological and disease models, thereby offering practical guidance for methodological choices in future research. We conducted systematic DDA and DIA comparative analyses using three distinct model sample types. Specifically, one disease model and two drug-treated model-followed by a comprehensive multi-dimensional evaluation. Our analytical framework extends beyond conventional metrics such as total protein and peptide counts, precision, and accuracy, incorporating assessments of data completeness, accuracy in detecting positive control markers, reproducibility, reliability of functional annotations, and the identification of sources contributing to methodological discrepancies. This holistic profiling approach

aims to provide deeper and more actionable insights for selecting optimal data acquisition strategies, particularly in complex studies requiring deep proteomic coverage, highly reproducible quantification, and robust biological discoveries.

Methods

Instruments and reagents

The Vanquish Neo UHPLC system coupled to an Orbitrap Exploris 480 mass spectrometer, as well as a SpeedVac Concentrator and a centrifuge, were purchased from Thermo Fisher Scientific (MA, USA). An electrothermal incubator was purchased from Shanzhi Equipment (Shanghai, China).

Hep G2 (Human Hepatocarcinoma Cells) and HeLa cells were purchased from ATCC. DMEM medium was from Invitrogen (USA). Urea, ammonium bicarbonate (NH_4HCO_3, HPLC grade), cysteine (97%), trifluoroacetic acid (TFA, HPLC grade), and formic acid (FA, MS grade) were purchased from Sigma-Aldrich (Germany). Protease inhibitors were from Roche (Switzerland). Dithiothreitol (DTT, 99%) and iodoacetamide (IAA, 98%) were purchased from Acros Organics (Belgium). The Enhanced BCA Protein Concentration Assay Kit was purchased from Beyotime Biotechnology (Jiangsu, China). MS-grade acetonitrile and water were purchased from Thermo Fisher Scientific (USA). MS-grade trypsin was purchased from Hualishi Scientific (Beijing, China). SepPak tC_{18} cartridges were purchased from Waters (Milford, MA, USA).

Cell culture and model construction

Cell culture

Hep G2 and HeLa cells were cultured in Dulbecco's modified Eagle's medium (DMEM, Invitrogen) supplemented with 10% fetal bovine serum (FBS, Invitrogen) and 100 units/mL penicillin–streptomycin (Hyclone) at 37 °C under 5% CO_2.

Drug-treated model

Upon reaching 80–90% confluency, HepG2 cell were treated with either 100 ng/ml Interferon α–2a (Model Group 1) or 100 ng/ml Interferon λ1 (Model Group 2) in fresh medium, while control cells received medium lacking interferons. All groups were cultured under standard conditions (37 °C, 5% CO_2) for 12 h to establish treatment effects. Following incubation, the medium was discarded, and cells were washed twice with ice-cold phosphate-buffered saline (PBS) to prepare the interferon-treated and control proteomes. Three independent biological replicates were prepared for each group (M1, M2, and control) to ensure statistical reliability, with each replicate originating from separate cell culture passages.

Disease model

After reaching 80–90% confluency, HeLa cells were exposed to ionizing radiation (IR) with a total dose of 10 Gy. Culture medium was replaced with fresh complete medium 1 h prior to irradiation. Post-irradiation, cells were returned to the incubator (37 °C, 5% CO_2) for a recovery period of 1 h to establish the injury phenotype. After discarding the medium, cells were washed twice with ice-cold phosphate-buffered saline (PBS) to prepare IR-treated and control proteomes. Two biological replicates were generated per condition, with each replicate originating from independent cell passages.

Protein extraction and digestion

Lysis buffer (8 M urea, 100 mM NH_4HCO_3, pH 8.0, 2 × phosSTOP phosphatase inhibitor (Roche), 1 mM PMSF, 10 mM NaF, 1:100 Phosphatase Inhibitor Cocktail 2, 1:100 Phosphatase Inhibitor Cocktail 3) was added to washed cells. Cells were scraped off, and the suspension transferred to tubes for lysis. Samples were sonicated on ice (4 min, 2 s on/5 s off, 30% power), centrifuged (21,300 g, 15 min, 4 °C). Supernatants (protein mixtures) were collected; concentration was determined using a BCA kit (Beyotime).

Protein samples were reduced with 5 mM DTT (56 °C, 30 min), alkylated with 15 mM IAA (room temperature, dark, 30 min), and the reaction quenched with 30 mM cysteine (room temperature, 30 min).

Urea concentration was diluted five-fold using 100 mM NH_4HCO_3 prior to digestion. Samples were digested with MS-grade trypsin (1:50 w/w enzyme-to-protein) at 37 °C, pH 8.0 for 16 h. Digested peptides underwent C_{18}-based solid-phase extraction (SepPak cartridges) and were dried by SpeedVac.

Liquid chromatography and mass spectrometry

The mass spectrometry analysis was performed on an Orbitrap Exploris 480 mass spectrometer (Thermo Fisher Scientific) equipped with Nanospray Flex ion source and coupled to Vanquish Neo UHPLC system (Thermo Fisher Scientific). All desalted peptide samples were loaded onto a homemade C18 analytical column (25 cm,75 μm Inner Diameter, 1.9 μm Dr. Maisch C_{18} particles) and then separated on it. The analytical column was maintained at 60 °C using a column oven, with inter-run equilibration performed using solvent A (0.1% formic acid in water).

Regarding the Data dependent acquisition (DDA) detection mode, the 60-min gradient method was used for the drug-treated groups and the 120-min gradient method was used for the disease model group, the liquid chromatographic separation gradient and scan parameters were set as followed:

1) 90 min: 0 to 2% solvent B (0.1% FA in 90% ACN) for 1 min, 2 to 13% solvent B for 35 min, 13 to 26% solvent B for 40 min, 26 to 45% solvent B for 5.1 min, 45 to 80% solvent B for 0.5 min, and maintained at 80% solvent B for 8.4 min. The flow rate was always set to 300 nL/min.The full-scan MS spectra were acquired in the Orbitrap mass analyzer with a resolution of 120,000 (at m/z 200), covering a mass range of m/z 350–1300. The RF lens was set to 50% with a normalized AGC target of 300%, and the maximum ion injection time was limited to 25 ms. The dynamic exclusion time was set to 35 s. The ten most abundant precursor ions were selected for fragmentation. MS2 spectra were acquired in the Orbitrap with HCD (30%) in centroid mode with an isolation window = 0.8 m/z, a resolving power of 15,000, and normalized AGC target of 50%.
2) 120 min: 5 to 13% solvent B (0.1% FA in 90% ACN) for 32 min, 13 to 26% solvent B for 68 min, 26 to 45% solvent B for 8 min, 45 to 80% solvent B for 2 min, and maintained at 80% solvent B for 10 min. The flow rate was always set to 300 nL/min. The full-scan MS spectra were acquired in the Orbitrap mass analyzer with a resolution of 120,000 (at m/z 200), covering a mass range of m/z 350–1800. The RF lens was set to 50% with a normalized AGC target of 300%, and the maximum ion injection time was limited to 50 ms. The dynamic exclusion time was set to 45 s. The time between master scans was set to 2 s. MS2 spectra were acquired in the Orbitrap with HCD (32%) in centroid mode with an isolation window = 2 m/z, a resolving power of 15,000.

Regarding the Data independent acquisition (DIA) detection mode, the 90-min gradient method was used for the drug-treated groups, and the 120-min gradient method was used for the disease model group, the liquid chromatographic separation gradient and scan parameters were set as followed:

1) 90 min: 0 to 2% solvent B (0.1% FA in 90% ACN) for 1 min, 2 to 13% solvent B for 35 min, 13 to 26% solvent B for 40 min, 26 to 45% solvent B for 5.1 min, 45 to 80% solvent B for 0.5 min, and maintained at 80% solvent B for 8.4 min. All master scan parameters remained identical to those used in DDA mode, except for the RF lens (30%), and maximum ion injection time (50 ms). From each MS1 scan, all possible precursor ions were selected for fragmentation with a 30,000 m/z resolution and a maximum injection time of 54 ms using Higher Collisional Dissociation (HCD) in the orbitrap with 28% normalized collision energy (NCE).
2) 120 min: 5 to 13% solvent B (0.1% FA in 90% ACN) for 32 min, 13 to 26% solvent B for 68 min, 26 to 45% solvent B for 8 min, 45 to 80% solvent B for 2 min, and maintained at 80% solvent B for 10 min. The flow rate was always set to 300 nL/min. All master scan parameters remained identical to those used in DDA mode, except for the RF lens (30%), maximum ion injection time (50 ms), and adjusted scan range (m/z 350–1300). From each MS1 scan, all possible precursor ions were selected for fragmentation with a 30,000 m/z resolution and a maximum injection time of 54 ms using Higher Collisional Dissociation (HCD) in the orbitrap with 28% normalized collision energy (NCE).

Mass spectrometry data processing

For proteomic analysis, all raw files generated in DDA mode were searched against UniProt Human protein database (version: 2025–10-09, 573,661 entries, http://www.uniprot.org) using MaxQuant (version: 2.4.14.0) with the following searching parameters: Trypsin/P was specified as cleavage enzyme allowing a maximum of two missing cleavages. Carbamidomethyl (C) was set as a fixed modification, while Oxidation (M) and Acetylation (protein N-term) were included as variable modifications. The identified peptides were required to be ≥ 7 amino acids, only those within the 7–30 residue range were included in quantitative analyses. Peptide and protein identifications were filtered at a 1% false discovery rate (FDR). Match between runs (MBR) was enabled to transfer identifications across samples. The alignment time window and the match time window were set to 20 min and 0.4 min, respectively.

Additionally, all raw files generated in DIA mode were analyzed in Spectronaut (version: 18.5.231110.55695) using a library-free workflow with default settings. The fixed modification was set as carbamidomethyl (C). The variable modification was set as Oxidation (M) and Acetylation (protein N-term). The max variable modification number was set to 5. Both peptide-spectrum matches (PSMs) and protein-level identifications were controlled at a global FDR of 1%. The strategy for fragment ion selection was based on intensity. Quantification was performed using MS2 extracted ion chromatograms (XICs) with the following criteria: tryptic peptide length 7–30 residues, ≤ 2 missed cleavages, and a post-search filtering threshold of q-value < 0.01 (1% FDR). Cross-run normalization was enabled with the normalization strategy and row selection both set to automatic. The quantification window was configured to synchronized mode to ensure consistent peak integration across all runs.

Data preprocessing and standardization

Initial median normalization was applied to the raw quantitative proteomics data to eliminate systematic biases. Proteins with missing values (NAs) exceeding 50% in any experimental group were removed based on stringent quality control criteria. For the retained protein dataset, missing values were imputed by random sampling from a normal distribution parameterized using the mean and standard deviation derived from the bottom 5% of observed values: Statistical characteristics (arithmetic mean, μ, and standard deviation, σ) were calculated from the lowest 5th percentile of complete observations within the dataset. Missing values in the original data were then filled using random numbers generated from a normal distribution defined by these μ and σ parameters. This approach preserves the biological variability inherent in low-abundance proteins while mitigating potential statistical biases introduced by over-imputation. For both DDA and DIA data, group means were used to calculate differences, two-sample equal variance t-tests were applied to compute p-values, and the FDR method was used for multiple testing correction.

All data visualization and downstream analyses were performed using R (version 4.2.3), primarily leveraging the following bioinformatics packages: Data Visualization: ggplot2 (v3.5.2) for generating high-quality statistical graphics; enrichplot (v1.18.4), clusterProfiler (v4.6.2), for visualizing enrichment analysis results. Data Manipulation: dplyr (v1.1.4) for data transformation and manipulation; tibble (v3.2.1) for constructing standardized data frames; tidyr (v1.3.1) for data tidying and reshaping.

A multi-tiered functional annotation strategy was employed for biological interpretation: Pathway Enrichment Analysis: Gene Set Enrichment Analysis (GSEA) was performed using the MSigDB Hallmark gene sets (h.all.v2023.1.Hs.symbols.gmt) to characterize the regulatory pathway features of differentially abundant proteins. Gene Ontology (GO) Analysis: Hierarchical annotation for Molecular Function, Biological Process, and Cellular Component was conducted using the DAVID bioinformatics resource (https://david.ncifcrf.gov/). Kyoto Encyclopedia of Genes and Genomes (KEGG) Analysis: Topological analysis of metabolic pathways and signal transduction networks was performed concurrently on the DAVID platform. The background gene sets for all GO and KEGG pathway enrichment analyses consisted of the union of all proteins identified by each specific analytical method in the respective experiments. This approach prevents potential bias in enrichment p-values that could arise from differences in protein identification numbers across methods.

Results

Sample collection and data acquisition strategy for DIA and DDA

To systematically evaluate the characteristics and performance of label-free quantification in proteomics results based on DDA and DIA strategies, complex models including a disease cell model and drug-treated cell models were established for proteome sample preparation. In the disease model, disease cells were exposed to IR to induce DNA damage repair responses, while the control group remained untreated. For the drug-treated models, cells were treated with either type I or type III interferon, whereas the control cells received DMSO treatment. In total, 16 samples were prepared, comprising one disease model (2 treatment replicates and 2 controls) and two drug-treated models (3 treatment replicates and 3 controls each). After cell harvesting and collecting, the cell pellets were pre-washed with cold PBS buffer. Subsequently, the cell pellets were then subjected to lysis, protein extraction, protein reduction, alkylation, and sequential digestion with trypsin. Equal amounts of the resulting peptide samples were analyzed using an Orbitrap Exploris 480 mass spectrometer operated in both DDA and DIA modes. DDA data were processed using MaxQuant, while DIA data were analyzed with Spectronaut. The identified proteins were filtered with 1% FDR at protein level and peptide level, with the comprehensive experimental workflow illustrated in Fig. 1 and detailed sample processing parameters and instrument configurations summarized in Table 1.

Evaluation of protein-level data from the DDA and DIA acquisition methods

To systematically evaluate the differences and characteristics in comprehensive performance of identified proteins between DIA and DDA in proteomic analysis, we primarily focused on key evaluation metrics including identification breadth, quantification depth, reproducibility, data completeness, and precision.

We first compared the depth of protein and peptide identification and quantification across each sample group (Fig. 2A). In the disease model (D), the DDA method identified only 5,067 proteins, with an average of approximately 4,800 proteins quantified per group. In contrast, the DIA method identified a total of 7,735 unique proteins—nearly 50% more than with DDA. Moreover, with an average of approximately 7,550 proteins quantified per group, the average number of identified proteins per sample was nearly 50% higher in DIA than in DDA. In drug-treated model 1 (M1), DDA identified only 4,605 proteins (average quantification ~ 4,450), whereas DIA identified 7,987 proteins (average quantification ~ 7,600), demonstrating a significantly higher protein identification coverage than DDA. Across all

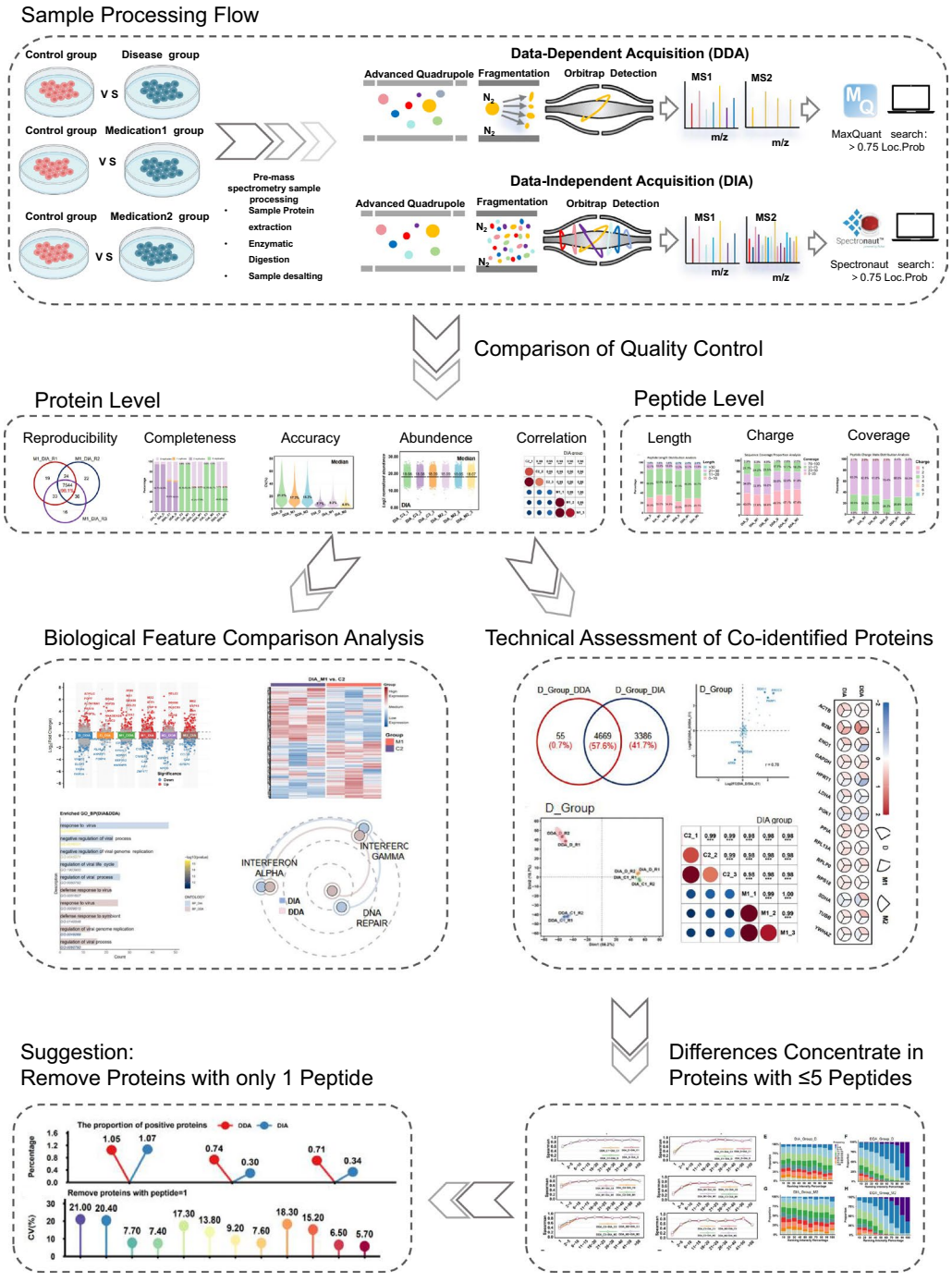

Fig. 1 Sample Collection and Data Acquisition Strategy for DIA and DDA. The disease model control and treatment groups were analyzed with two biological replicates each, while the drug-treated model control and treatment groups were analyzed with three biological replicates each. All data were subsequently analyzed by MaxQuant (DDA) and Spectronaut (DIA) for database searching and detection

experimental models, the number of proteins identified by DIA consistently exceeded that identified by DDA. On average, the identification depth achieved by DIA was 60% higher than that of DDA. These results clearly highlight the superior performance of DIA in terms of both protein and peptide identification.

To evaluate the reproducibility of these two technologies, we conducted an intersection analysis to identify proteins consistently detected across all biological replicates within each experimental group (Fig. 2B, C; S1A-D). The results showed that, in proteomic analysis using DDA mode, only 95.6%, 95.4%, and 96.2% of the proteins were consistently identified in all three replicates of

Table 1 Instrument and sample information summary

MS2	Sample	Sample process	Cell line	Biological repetition	MS instrument	loading amount (ug)	Gradient
DDA	Disease model	IR illumination	HeLa	2	Orbitrap exploris 480	4	120 min
	Drug-treated model	IFNα-2a: 100 ng/mL	HepG2	3		4	90 min
		IFNλ1: 100 ng/mL	HepG2	3		4	90 min
DIA	Disease model	IR illumination	HeLa	2		4	120 min
	Drug-treated model	IFNα-2a: 100 ng/mL	HepG2	3		4	90 min
		IFNλ1: 100 ng/mL	HepG2	3		4	90 min

the D, M1, and M2 model groups, respectively. In contrast, when using DIA mode, this proportion increased to 99.0%, 98.2%, and 98.2%, respectively. These findings indicate that DIA mode offers superior consistency in cross-replicate protein identification. We further analyzed the distribution of missing values and fluctuations in quantitative abundance across replicate samples obtained using different acquisition modes. Bar charts illustrate the distribution of missing quantitative values for proteins within each group (Fig. 2D). The distribution of missing quantitative values for the three DIA data sets was generally slightly lower than the corresponding DDA data. The data completeness rate, defined as the proportion of proteins quantified across all replicates, was consistently above 97% for DIA, compared with a range of 92–96% for DDA.

Further comparative analysis was conducted on the quantitative results of proteins obtained by two acquisition modes. Analysis of the protein abundance distribution in the experimental group (Fig. 2E–F, S1E–F) revealed that the average Log2 protein abundance calculated by DIA was approximately 18, whereas that obtained by DDA was approximately 29. This significant difference in abundance levels is primarily attributed to the inherent algorithmic differences between the two search engines.

Correlation analysis assessing intra-group reproducibility (Fig. 2H-I**,** S1G-J) revealed that the pearson correlation coefficients among the three biological replicates in the DIA mode consistently exceeded 98%, demonstrating a high level of intra-group consistency. In contrast, the intra-group reproducibility coefficients for the DDA mode ranged from 93 to 98%. This difference is likely attributable to the presence of missing values in the DDA and DIA datasets or variations in quantitative measurements. The coefficient of variation (CV) analysis yielded the following results: 20.9% for DDA-D compared to 7.7% for DIA-D; 21.7% for DDA-M1 compared to 9.4% for DIA-M1; and 17.4% for DDA-M2 compared to 6.6% for DIA-M2. These findings indicate that the precision of the DIA method was markedly higher than that of the DDA method across all three experimental groups (Fig. 2G), suggesting that DIA technology generates lower quantitative noise. In summary, through evaluation across multiple critical dimensions —including identification depth, quantification robustness, data completeness, reproducibility, and precision—the results demonstrate that the DIA mode offers significant comprehensive advantages over the conventional DDA mode.

Evaluation of peptide-level data from the DDA and DIA acquisition methods

To systematically evaluate and compare peptide identification performance between DDA and DIA, we analyzed various metrics including peptide counts, length distribution, sequence coverage, and charge state. In D, DIA identified 107,024 peptides, with an average quantification depth of 104,500 peptides per replicate, significantly outperforming DDA, which identified 52,976 peptides on average (42,000 per replicate) (Fig. 3A). In M1, DIA detected 123,598 peptides (average: 118,000), compared to DDA's 49,276 peptides (average: 39,500). Similarly, in M2, DIA identified 120,474 peptides (average: 117,000), whereas DDA yielded 47,747 peptides on average (average: 40,000). We found both methods predominantly identified peptides of approximately 7–18 amino acids in length (Fig. 3B). The slight difference lies in the fact that DIA can identify a higher proportion of short peptide segments with fewer than 12 amino acids compared to DDA. Conversely, for longer peptide segments, DDA demonstrates superior identification capability. For example, in the identification of peptide segments longer than 12 amino acids, DIA identifies such segments at an average proportion of 47.7% of its total identified peptides, whereas DDA achieves a higher proportion of 51.8%. These findings suggest that DDA and DIA exhibit distinct preferences in spectral interpretation. A plausible explanation is that DIA involves the interpretation of mixed spectra, whereas DDA focuses on individual peptide spectra, making spectral interpretation in DIA inherently more complex than in DDA.

Our data also showed that DIA achieved higher sequence coverage (Fig. 3C), with approximately 2.1% of all identified peptides covering 76–100% of the protein

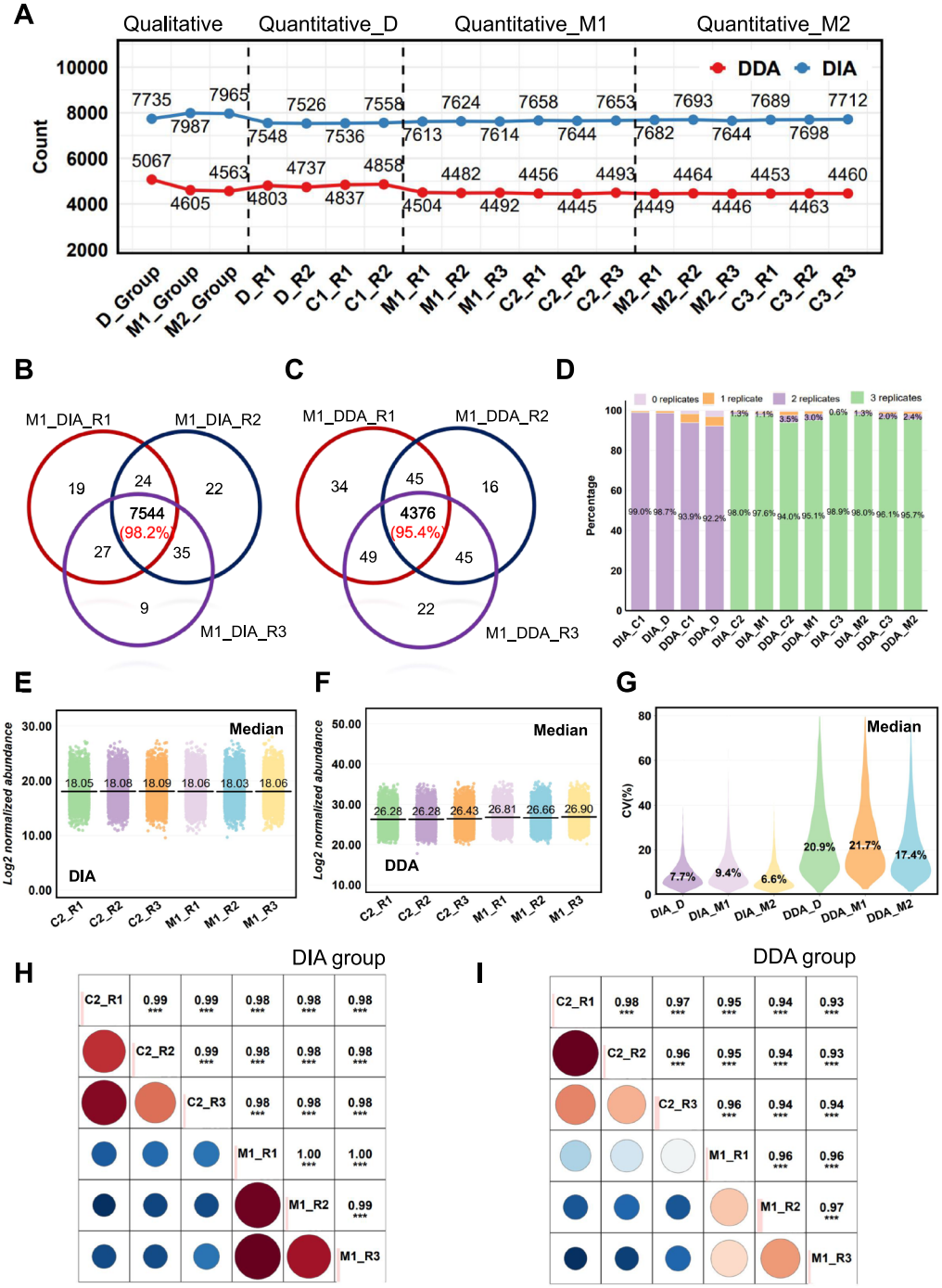

Fig. 2 Evaluation of Protein-Level Data from the Two Acquisition Methods. (**A**) Qualitative (identified) and quantitative (quantified) protein data per replicate group across the three models. (**B-C**) Reproducibility visualization for DIA and DDA data in M1 group. DIA achieved a reproducible identification rate of 98.2%, compared to 95.4% for DDA. (**D**) Distribution of missing values for DIA and DDA data across the three models. Color key: Green = No missing values (present in all 3 replicates), Purple = Missing in 1 replicate, Yellow = Missing in 2 replicates, Light purple = Missing in all 3 replicates. (**E–F**) Abundance distribution of identified proteins for DIA and DDA data in M1 group. (**G**) Precision (Coefficient of Variation, CV) for DIA and DDA data across the three models (y-axis: CV%). (**H-I**) Intra-group correlation analysis (Spearman) for M1 group

sequences (compared to ~ 1.1% in DDA), and approximately 21.1% covering 50–75% (compared to 11.8% in DDA). These results indicate that DIA achieved higher protein sequence coverage compared to DDA. Charge state analysis revealed that DIA was enriched in 3 + charged peptides, while DDA showed a preference for 2 + charged peptides (Fig. 3D). Interestingly, in the DDA data, almost no + 1 charged ions were detected, whereas

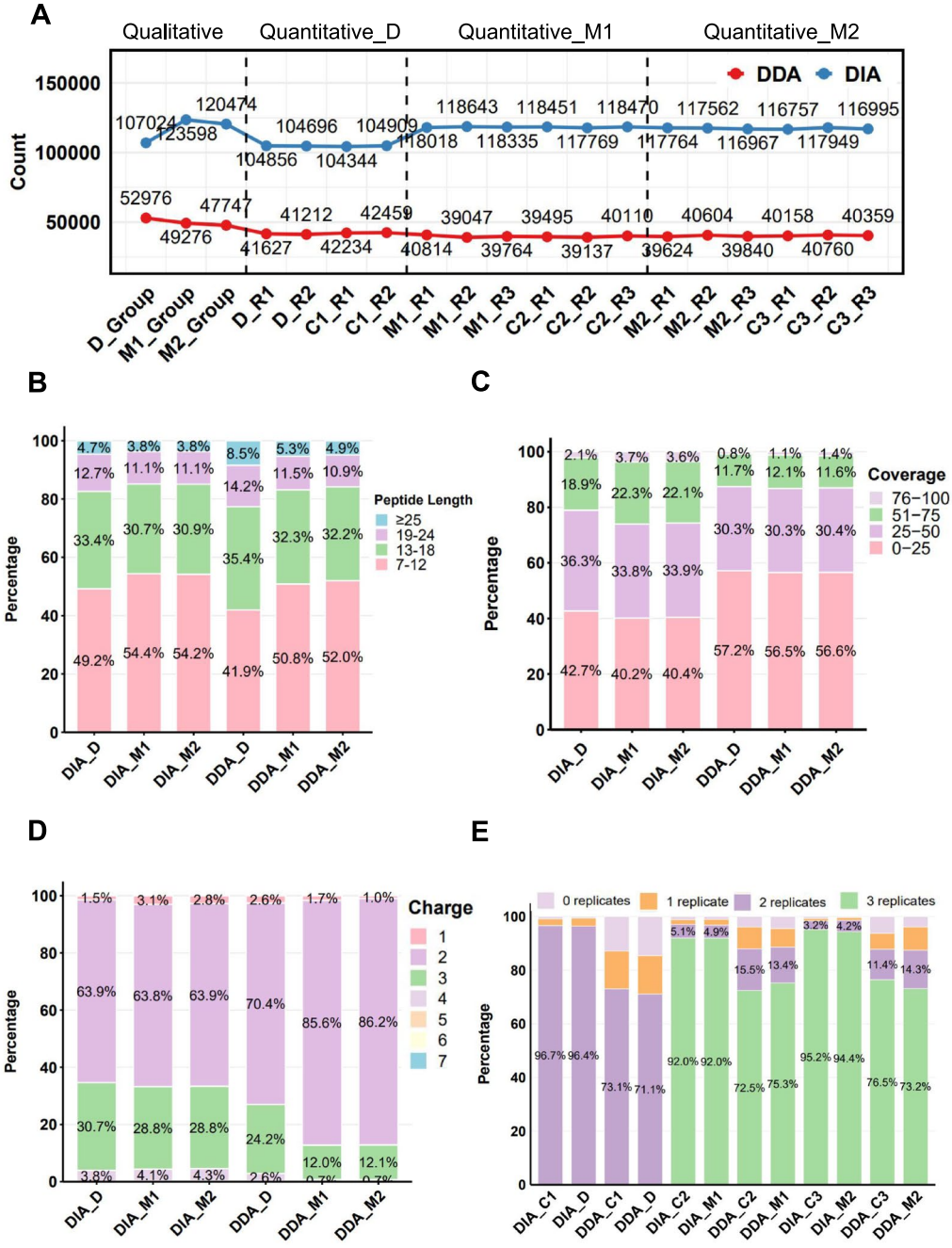

Fig. 3 Evaluation of Peptide-Level Data from the Two Acquisition Methods. (**A**) Qualitative (identified) and quantitative (quantified) peptide data per replicate group across the three models. (**B**) Distribution analysis of peptide length for DIA and DDA data across the three models. (**C**) Distribution analysis of sequence coverage proportion for DIA and DDA data across the three models. (**D**) Distribution analysis of peptide charge state for DIA and DDA data across the three models. (**E**) Distribution of missing values for DIA and DDA data across the three models. Color key: Green = No missing values (present in all 3 replicates), Purple = Missing in 1 replicate, Yellow = Missing in 2 replicates, Light purple = Missing in all 3 replicates

a small percentage of + 1 ions were observed in the DIA data. This difference can be attributed to the distinct mass spectrometry detection parameters employed in DDA and DIA. Specifically, DDA typically excludes + 1 charged ions from secondary fragmentation, while DIA fragments all ions within a defined isolation window, regardless of their charge state. Furthermore, we analyzed the reproducibility of peptide identification at the peptide level. The results indicated that, on average, over 92% of peptides could be consistently identified in DIA data, whereas less than 77% were consistently identified in DDA data (Fig. 3E). These findings align with those observed in protein identification and further demonstrate that DIA technology offers higher reproducibility

and reduced randomness in identifying proteins or peptides.

Comparative analysis of biological features across proteomics data using DDA and DIA

To systematically evaluate the differences between DDA and DIA modes in terms of quantitative accuracy and the effectiveness in identifying positively differentially expressed proteins, as well as to uncover underlying pathways and biological characteristics associated with these differences, we first screened for differentially expressed proteins (DEPs) between the disease model vs. control or drug-treated vs. control cells using stringent criteria: Benjamini–Hochberg adjusted p-value < 0.05 and fold-change (FC) > 1.5 (upregulated) or < 0.667 (downregulated). Volcano plots illustrating the DEPs depict the differential expression patterns across the three groups (Fig. 4A). In the D group, DIA detected 52 significantly upregulated and 57 significantly downregulated proteins, while DDA identified significantly more DEPs (160 upregulated and 100 downregulated). In M1, the DIA method identified 275 significantly upregulated and 319 significantly downregulated proteins, whereas the DDA mode identified 283 upregulated and 190 downregulated proteins. In M2, DIA identified 162 upregulated and 156 downregulated proteins, compared to 186 upregulated and 194 downregulated proteins identified by DDA.

Additionally, a heatmap clustering analysis of three model datasets was performed to visualize the global expression patterns and identify coherent protein clusters with similar regulation profiles across experimental conditions (Fig. 4B**,** S2A). This analysis revealed distinct protein expression signatures specific to each treatment group and highlighted the consistency of biological replicates within groups.

Further functional annotation was conducted and compared for the differentially expressed proteins (DEPs) identified using DDA and DIA technologies, covering Gene Ontology (GO) functional enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses (Fig. 4C–D, S2C–D). Notably, both methods successfully identified several key positive control pathways that were significantly activated following drug intervention, such as the "response to virus" pathway in the GO Biological Process (BP) category for the IFN α–2a and IFN λ1 treated groups (M1 and M2) compared to the control groups. However, when comparing both the number of enriched genes (Count) and the significance level of enrichment (P-value), the results indicated that DIA data demonstrated superior performance compared to DDA. This finding was further supported by Gene Set Enrichment Analysis (GSEA) (Fig. 4E-G). To quantitatively evaluate the extent of pathway activation, we constructed radar plots based on the Normalized Enrichment Score (NES), allowing for a comparative assessment of key positive control pathways across experimental groups (with a larger radar plot radius reflecting more pronounced pathway activation). The GSEA results demonstrated significant enrichment of the "DNA_REPAIR" pathway in the D group, the "INTERFERON_ALPHA_RESPONSE" pathway in the M1 group, and the "INTERFERON_GAMMA_RESPONSE" pathway in the M2 group. Supplementary Figure S2D–E presents detailed GSEA plots illustrating the enrichment of the "INTERFERON_ALPHA_RESPONSE" pathway in the M1 group for both the DIA and DDA datasets. Collectively, based on p-values and count values from GOBP and KEGG analyses, as well as NES values from GSEA, these findings consistently indicate that the pathway activation effects detected using the DIA mode were substantially more pronounced than those observed with the DDA mode.

Accuracy assessment of the DDA and DIA

To gain deeper insight into the factors contributing to the differences in quantitative accuracy between DDA and DIA modes, we analyzed the overlap of quantified proteins using Venn diagrams. In the D group, only 4.0% of the proteins were quantified exclusively by DDA, whereas 37.7% were quantified exclusively by DIA (Fig. 5A). A total of 4,586 proteins were quantified by both acquisition methods, representing 58.2% of all quantified proteins. In the M1 group, 4,287 proteins (53.5%) were quantified commonly by both methods, with DDA-specific quantifications accounting for 3.6% and DIA-specific quantifications for 42.8% (Fig. 5B). Similarly, in the M2 group, 4,252 proteins (53.1%) were quantified by both approaches, while DDA-specific and DIA-specific quantifications accounted for 3.6% and 43.3%, respectively (Fig. S3A). These findings align with the protein quantification depth comparison presented in Fig. 2, which indicated that DIA achieves greater quantification depth compared to DDA. To further investigate the discrepancies between the DDA and DIA datasets, we conducted correlation analyses on overlapping proteins within each experimental model. The results revealed that inter-method correlation coefficients for these intersecting proteins across the three models ranged from 0.82 to 0.87, which were significantly lower than the intra-group correlations observed within each method (Fig. S3B–D). A similar trend was observed at the peptide level (Fig. S4A–F). Principal Component Analysis (PCA) showed that DDA data exhibited greater variability compared to DIA data (Fig. 5C).

To validate the consistency in detecting positive control markers, we selected proteins that were commonly quantified by both acquisition modes within each group and cross-referenced them with UniProt-curated sets of DNA

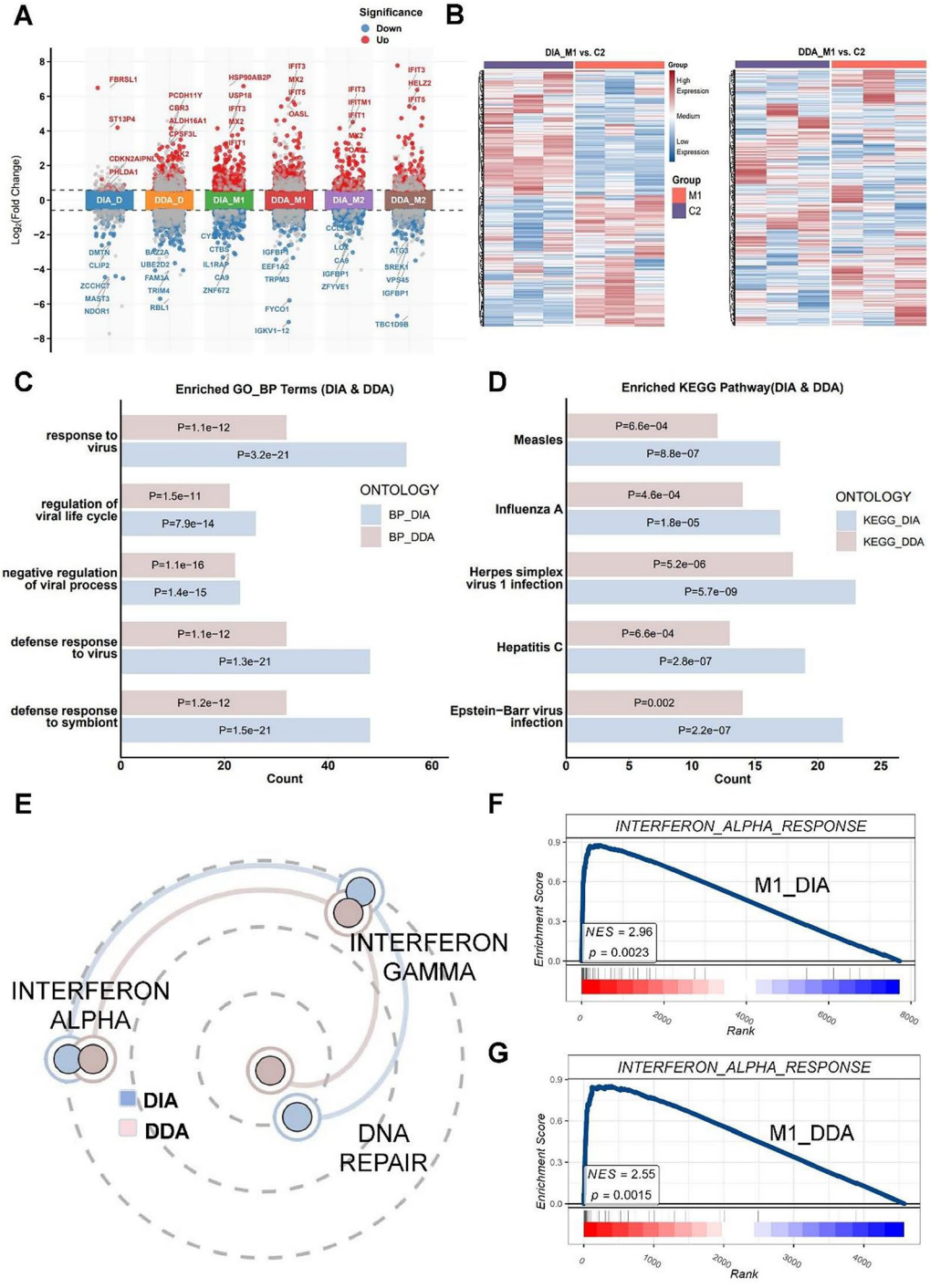

Fig. 4 Evaluation of Enrichment Capabilities for the Two Data Acquisition Methods. (**A**) Summary of differential protein analysis results for protein data across the three models (Significance: $p < 0.05$). (**B**) Heatmap visualization of DIA data for M1 group. (**C-D**) Comparison of GO Biological Process (BP) and KEGG pathway enrichment results between DIA and DDA data for M1 group. (**E**) Radar plot showing the activation levels of key positive con-trol pathways (INTERFERON_ALPHA_RESPONSE for M1 group, INTERFERON_GAMMA_RESPONSE for M2 group, DNA_REPAIR for D group) in DIA versus DDA data across the three models. (**F-G**) Detailed GSEA enrichment plots specifically show the regulation of the INTERFER-ON_ALPHA_RESPONSE pathway by DIA and DDA data in M1

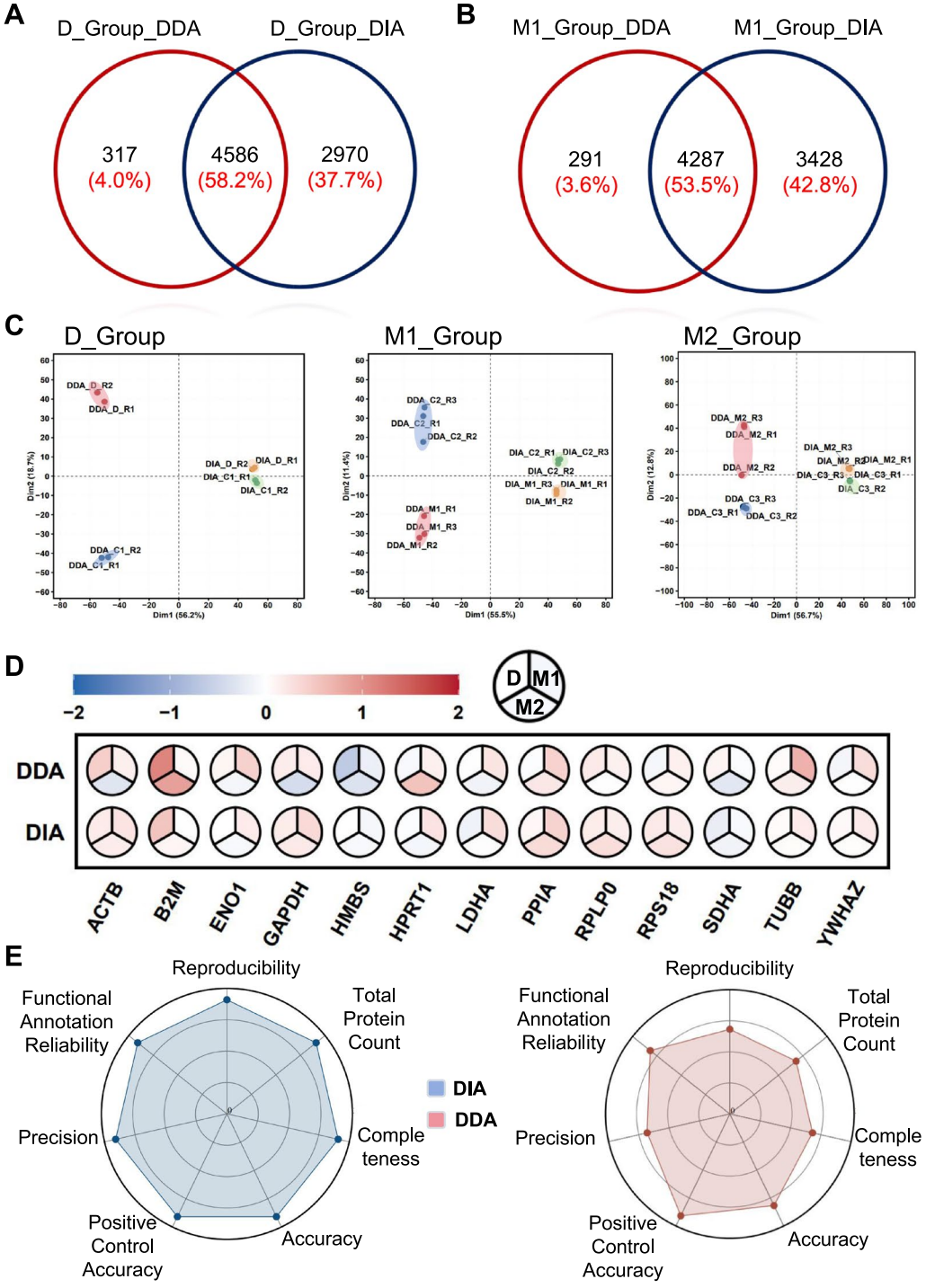

Fig. 5 Accuracy Assessment of the Two Data Acquisition Methods. (**A-B**) Venn diagram analysis of overlapping proteins quantified by DIA and DDA in D and M1. (**C**) PCA analysis of all samples across the three experimental model groups. (**D**) Quantitative analysis of fold changes in housekeeping proteins detected by both Data-Dependent Acquisition (DDA) and Data-Independent Acquisition (DIA) methods across the three experimental models. The diagram is divided into three sectors corresponding to each model: Group D (top-left sector), Model M1 (top-right sector), and Model M2 (bottom sector). Consistent expression patterns of housekeeping proteins observed across methodologies underscore their potential role as internal controls. (**E**) Radar chart comparing the overall performance of DDA and DIA

damage response proteins and interferon-stimulated genes (ISGs), which were used as positive controls. In the D model, DDA quantified 291 DNA damage response proteins, compared to 173 quantified by DIA. For ISG proteins in the M1 and M2 models, DDA quantified 40 (M1) and 38 (M2), respectively, whereas DIA quantified 57 in both models. A Spearman rank correlation analysis was then performed on these overlapping proteins. The results revealed a high degree of quantitative consistency between the two acquisition modes at the level of these positive markers (Fig. S3E), indicating their comparable capacity to capture relevant biological signals.

In addition, housekeeping proteins are a class of cellular proteins that perform essential cellular functions and typically exhibit stable expression levels across various experimental conditions. Therefore, we used housekeeping proteins as a reference to assess the consistency of quantification across different methods in each model. To further validate the quantitative accuracy of DIA and DDA, we analyzed the fold changes of 13 housekeeping proteins across all experimental groups (Fig. 5D). Notably, 12 of these proteins showed significantly reduced variability in the DIA dataset compared to DDA (excluding ACTB), aligning more closely with their expected stable expression profiles. These findings support the superior quantitative precision of DIA over DDA.

Based on the multi-dimensional evaluation framework established above—which encompasses identification breadth (measured by the number of identified proteins), data completeness (evaluated by the percentage of missing values), precision (assessed using the coefficient of variation, CV), quantitative accuracy (determined by fold changes in housekeeping proteins), positive marker identification accuracy (calculated based on the number of proteins overlapping with UniProt positive control sets), reproducibility (assessed through intra-group correlation), and functional annotation reliability (evaluated via GO, KEGG, and GSEA analysis results)—a radar chart was constructed to visually compare the overall performance of DIA and DDA approaches (Fig. 5E). The results indicate that, across most evaluation metrics, DIA demonstrates superior performance compared to DDA.

DDA and DIA data differences concentrate in proteins with ≤ 5 peptides

Whether DDA or DIA is used in label-free quantification proteomics, the final identification and quantification of protein abundance fundamentally rely on the detection and quantification of the corresponding original peptide segments. To better understand the underlying reasons for the weaker inter-method correlation, we initially classified the overlapping proteins according to the number of peptides identified for each protein in both the DDA and DIA datasets. Proteins were categorized based on their peptide counts into the following groups: 1, 2–5, 6–10, 11–15, 16–20, 21–25, 26–30, 31–40, 41–50, and > 50 peptides (Fig. 6A–B, S5A–D). Analysis of peptide segment distribution revealed distinct differences between DIA and DDA. In DIA data, peptide coverage across proteins was more uniform, with most proteins identified by two or more peptide segments. Notably, 276 proteins in the M1 group were identified by more than 50 peptide segments. In contrast, DDA data exhibited sparse coverage: 7.0% of proteins were identified by only a single peptide segment, and 65.7% had fewer than 10 segments in the M1 group, a pattern that remained consistent across other models (Fig. S5A–D, S6A–F). These findings further demonstrate that DIA offers greater identification depth and improved accuracy in protein quantification due to its higher peptide coverage.

Subsequently, correlation analyses further validated these findings across protein groups defined by peptide counts in both DDA and DIA (Fig. 6C). The results demonstrated a clear pattern: the correlation between DIA and DDA data improved notably as the number of peptides per protein increased. Specifically, when the number of identified peptide segments for a protein in DDA was fewer than 5, the correlation between the corresponding protein quantification in DDA and DIA was significantly weaker compared to proteins with a higher number of identified peptides. The reason for this result may be attributed to the relatively small number of peptide segments used for protein quantification, which results in limited quantitative information and consequently reduces the accuracy of quantification for these proteins.

To further investigate the impact of low peptide counts on protein quantification intensity, we ranked the common proteins shared across all replicates based on their signal intensities within each replicate group. Specifically, the intensities in each replicate were sorted in descending order, and the top 10% of proteins based on intensity were selected. Proteins that consistently appeared in this top 10% across all replicates were classified into the "Top 10% Intensity" group. This process was repeated for subsequent percentile intervals. Subsequently, we calculated the proportion of proteins from each peptide count category within these intensity percentiles and visualized the results using stacked bar plots (Fig. 6D–E, S5E–H). In the D group, the results showed that proteins identified by only one peptide were predominantly found in the lowest abundance bins. Similarly, proteins identified by 2–5 peptides were primarily enriched in the low-abundance categories. In conclusion, we observed that proteins with fewer identified peptides in DDA datasets exhibited limited peptide-to-protein inference, resulting in lower protein abundance estimates. The relatively small number of peptide segments used for protein quantification, which

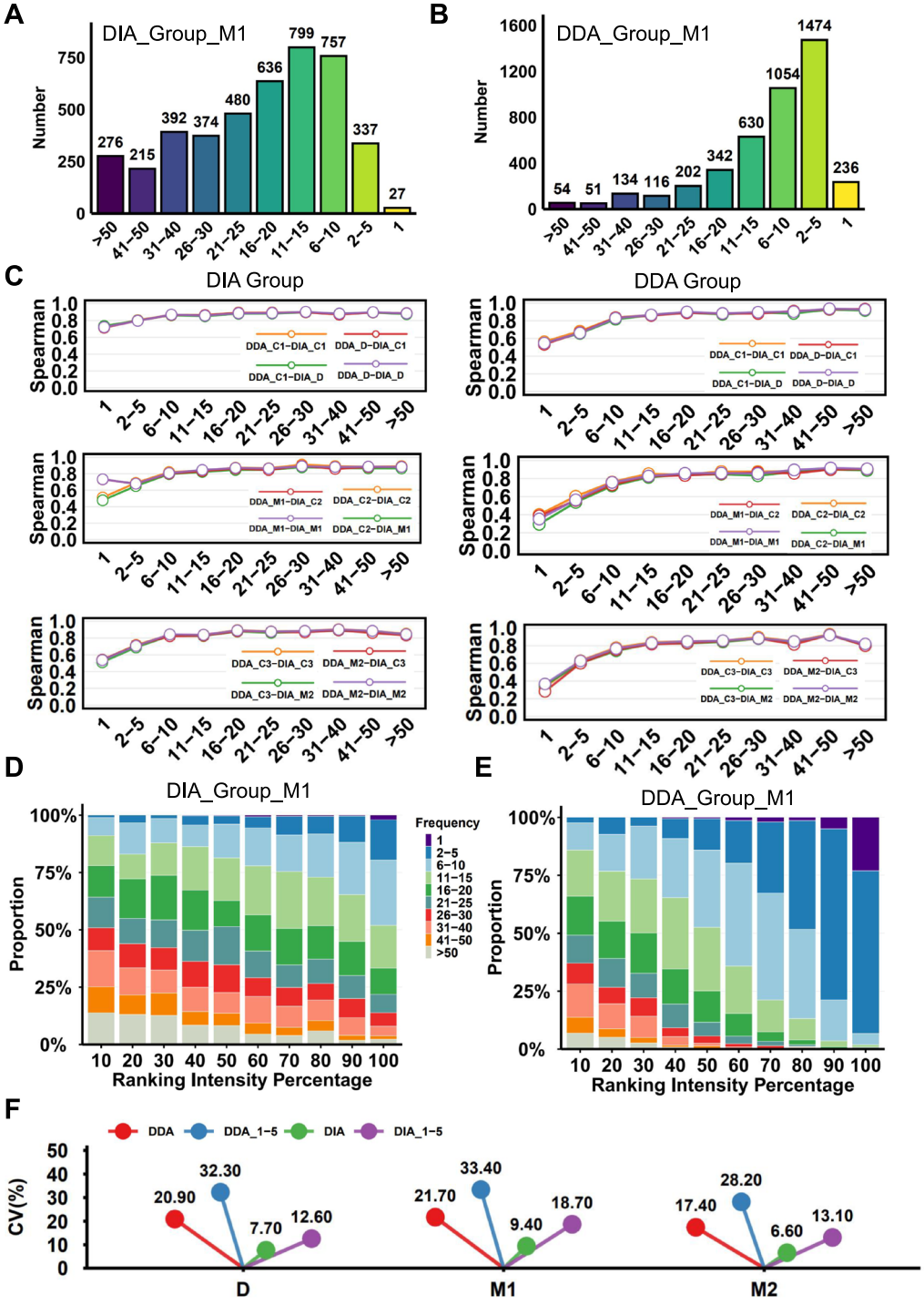

Fig. 6 DDA and DIA Data Differences Concentrate in Proteins with ≤ 5 Peptides. (**A-B**) Distribution of peptide counts per protein for common proteins identified by all models in M1, visualized for DIA (**A**) and DDA (**B**) data. (**C**) Correlation between DIA and DDA measurements for proteins grouped by peptide count (based on DIA data(left),based on DDA data(right) across three experimental groups. (**D-E**) Stacked bar represent the percentage of proteins within different peptide count groups (≤ 1, 2–5, 6–10, > 10 peptides) across intensity percentiles (Top 10%, 10–20%, etc.) for the M1 model, using DDA (**D**) and DIA (**E**) data. (**F**) Precision analysis (CV) showing significantly lower precision for proteins identified by 1–5 peptides in both DIA and DDA data compared to the full intersection set

results in limited quantitative information and consequently reduces the accuracy of quantification for these proteins.

Precision analysis, evaluated using the coefficient of variation (CV), was conducted on the complete set of overlapping proteins and further categorized based on the number of peptides used for identification, with a specific focus on proteins identified by ≤ 5 peptides. The results demonstrated significant differences in variability between the DDA and DIA datasets. For the DDA dataset: in the D group, the CV increased by 54.5% (from 20.9 to 32.3); in the M1 group, the increase was 53.9%; and in the M2 group, it was 62.7%. For the DIA dataset: in the D group, the CV increased by 63.6%, while in both the M1 and M2 groups, the increase reached 98.9% and 98.5%. these results indicate that the observed differences between DDA and DIA data are primarily driven by proteins identified using ≤ 5 peptides (Fig. 6F). This result further demonstrates that the Precision of protein quantification is influenced by the number of peptide segments identified through spectral matching. A lower number of identified peptide segments correlates with reduced accuracy in protein quantification.

The number of peptide segments matched by the quantified protein serves as a key indicator for assessing data quantifiability

To further elucidate the relationship between peptide count and the reliability of protein quantification, we subdivided the intersection proteins identified by ≤ 5 peptides into groups based on their exact peptide counts: 1, 2, 3, 4, 5, and > 5 (Fig. S7A–F), followed by correlation analysis within each group. When proteins were classified according to peptide counts derived from DDA data, a strong positive correlation between peptide count and inter-platform correlation strength was observed. Specifically, the DDA–DIA correlation coefficients increased progressively with increasing peptide counts. This trend was consistently observed across all three experimental models.

Protein grouping based on DIA-derived peptide counts revealed model-specific correlation patterns. While the D group preserved the peptide count-dependent correlation trend observed in DDA-based grouping, the M1 and M2 groups exhibited distinct correlation behaviors. Notably, proteins identified by only a single peptide consistently showed the lowest DDA-DIA correlation across all models (Fig. 7A). Precision analysis, assessed through the coefficient of variation (CV), indicated that lower peptide counts were associated with reduced measurement reproducibility (Fig. 7B), particularly for single-peptide identifications, which significantly compromised data quality across all experimental systems. Given the substantial compromise in data quality associated with single-peptide identifications, we sought to evaluate the potential impact of removing these less reliable identifications. To this end, we performed an intersection analysis with established positive control protein sets, which demonstrated that single-peptide proteins accounted for only 1% of validated positive control proteins (Fig. 7C).

Following an assessment of the impact of these proteins on overall quantitative reliability, considering both improvements in accuracy and the potential trade-off of data loss, we determined that excluding proteins identified by a single peptide was beneficial. The subsequent removal of these proteins from both DIA and DDA datasets, followed by reassessment of precision, NES from enrichment analysis, and differential protein expression, consistently supported the effectiveness of this filtering strategy (Fig. 7D–E, S7G). In contrast, proteins identified by only two peptides made up a substantially larger proportion of the positive control set. To maximize the retention of biologically relevant data while upholding quality standards, we recommend retaining proteins identified by at least two peptides (Fig. S7H).

Conclusions and discussions

In proteomics research, the choice of a data acquisition strategy directly affects the quantitative accuracy, depth of coverage, and throughput of experimental results. DDA and DIA are two core strategies in mass spectrometry and are widely used in label-free protein quantification. However, a comprehensive comparison of the characteristics of label-free quantitative data generated by these two approaches—such as their performance in identifying proteins and peptides, accuracy of quantification, and ability to support downstream biological interpretation—remains lacking.

In this study, we conducted a systematic evaluation of DDA and DIA across three experimental models, including disease and drug-treated cell models, focusing on the quantity and quality of protein and peptide identifications, the distribution of peptide charge states and lengths, and the effectiveness in extracting meaningful biological information. Additionally, we proposed strategies to improve the reliability of data analysis. We found that DIA, which employs an unbiased acquisition strategy by systematically fragmenting all precursor ions, demonstrates significantly enhanced detection sensitivity for low-abundance proteins (e.g., a 66.54% increase in the number of identified proteins under DIA compared to DDA) and significantly reduces the occurrence of missing quantitative values (with less than 5% missing values in DIA). The high technical reproducibility of DIA (with intra-group correlation coefficients exceeding 98%) is attributed to its design, which involves cyclically repeated scanning windows that effectively reduce random sampling variability. Moreover, the stable

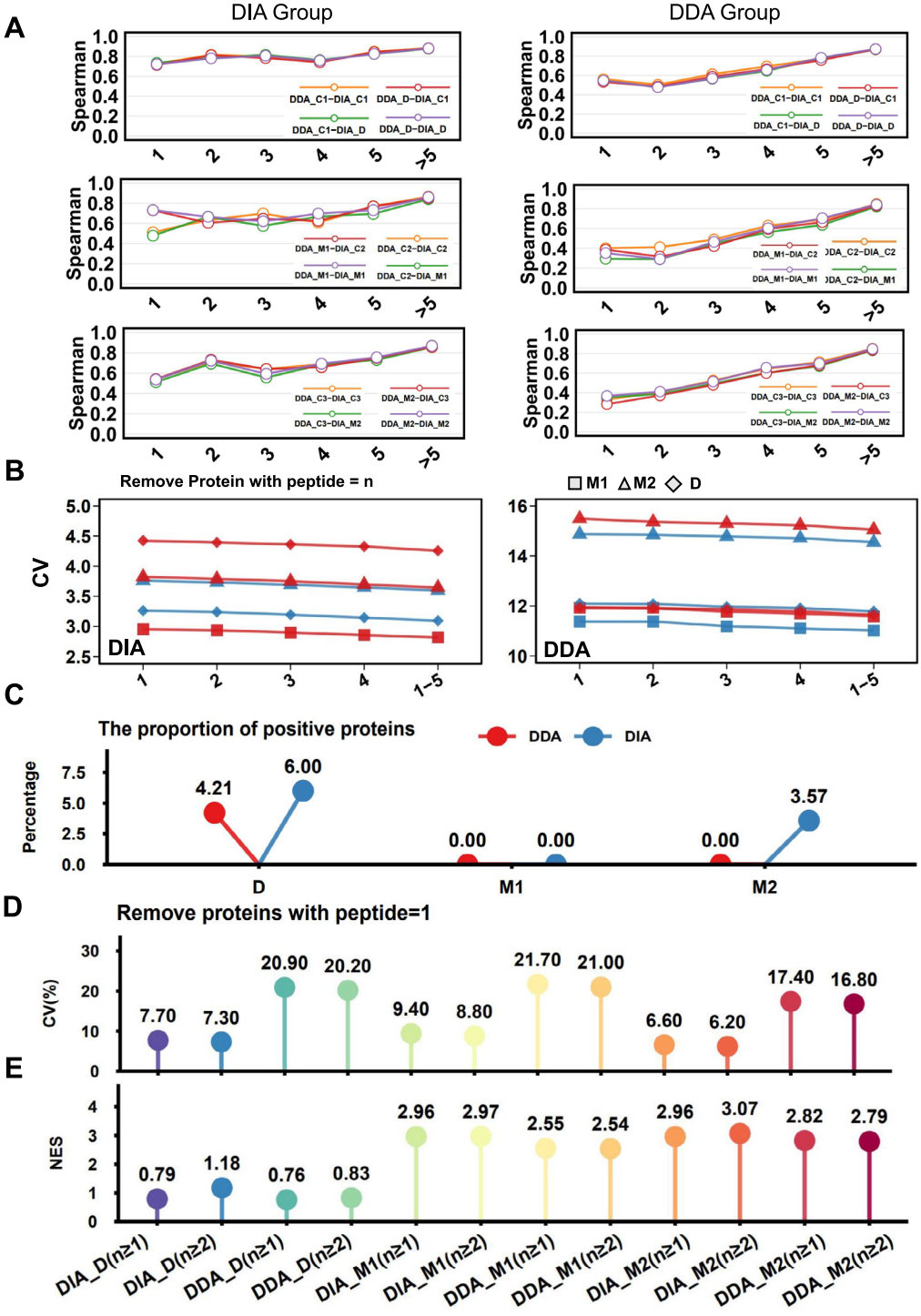

Fig. 7 Peptide Count as a Criterion for Assessing Data Quantifiability. (**A**) Correlation analysis (DIA vs. DDA) for commonly identified proteins grouped by exact peptide count categories (1, 2, 3, 4, 5, > 5 peptides). Left: Grouping based on peptide counts based on DIA-derived peptide counts; Right: Grouping based on DDA-derived peptide counts. (**B**) Precision (CV) values after sequentially removing proteins identified by 1 to 5 peptides in the three experimental models. (**C**) Proportion of single-peptide-hit positive control proteins within the total positive control protein set. (**D-E**) Lollipop chart quantitatively demonstrates the impact of single-peptide-hit proteins removal on enrichment results (NES) and precision (CV)

quantification of housekeeping proteins achieved with DIA (e.g., lower coefficients of variation for 12 out of 13 housekeeping proteins compared to DDA) further confirms its superior quantitative precision. Additionally, our findings indicate that the precision of protein quantification is influenced by the number of identified peptide segments. Specifically, proteins with fewer identified peptide segments tend to exhibit lower quantification accuracy. Therefore, proteins matched with only a single peptide segment or a limited number of peptide segments should be interpreted with caution regarding their quantitative reliability. To improve the overall accuracy of protein quantification, it is advisable to exclude such proteins from further analysis.

The discrepancy in absolute protein intensities between DDA and DIA primarily stems from fundamental differences in their algorithmic and quantitative principles. However, this does not impact the downstream differential expression or GSEA analyses, as our interpretations rely on relative quantification values computed within each respective technological platform. Specifically, the data from each platform were independently normalized, and subsequent differential analysis was based on fold-changes between conditions, while GSEA utilized ranked gene sets. Therefore, all comparisons were made internally within each platform, and the conclusions are robust against the differences in absolute intensities across platforms.

The advantages of DIA can be attributed to several key mechanisms. First, the unbiased acquisition nature of DIA inherently provides a wider dynamic range, capturing both high and low-abundance ions. To further mitigate challenges like co-elution and maximize this advantage, experimental optimizations such as extending chromatographic gradients or increasing sample loading can be employed, thereby improving the signal-to-noise ratio for low-abundance peptides. Second, DIA ensures data completeness. Unlike methods that selectively fragment precursor ions, DIA captures information from all ions, enabling comprehensive post-acquisition analysis. In contrast, DDA is prone to inherent precursor selection bias, which often results in the missed detection of low-abundance peptides. Third, DIA benefits from computational algorithm optimization. Advanced software tools such as Spectronaut and DIA-NN employ spectral library matching and machine learning techniques—including deep learning models—to markedly enhance the accuracy and efficiency of DIA data interpretation.

Although DDA did not exhibit the same level of robustness as DIA in this comparative analysis, it demonstrates superior compatibility with multiplexed labeling strategies, such as Tandem Mass Tag (TMT)-based quantification methods. TMT technology enables the simultaneous analysis of up to 32 samples, depending on the specific reagent used, within a single LC-MS2 run, thereby significantly improving throughput and experimental efficiency. This feature is particularly valuable for studies requiring accurate relative quantification across multiple experimental conditions—such as comparisons among various treatment groups, time points, doses, or cell lines versus controls—while effectively minimizing potential variations in LC retention times across samples. Moreover, due to its longer development history, DDA benefits from a more established software and database ecosystem, supported by widely used analytical platforms such as Sequest, Mascot, and MaxQuant. Additionally, in DDA, selected precursor ions (typically the most intense within a given cycle) are subjected to targeted fragmentation, generating corresponding MS2 spectra, which enhances the accuracy of spectral interpretation. This characteristic is especially beneficial for the precise identification and analysis of specific proteins, peptides, and post-translational modifications, including their types and modification sites.

Ultimately, the choice between DIA and DDA should not be viewed merely as a binary comparison of superiority, but rather as a decision that requires careful and comprehensive evaluation by the researcher. When selecting between the two approaches, researchers must consider multiple factors, such as the specific scientific objectives, sample complexity, availability of instrumentation, computational resources, and budgetary constraints. A solid understanding of the fundamental principles, strengths, limitations, and optimization strategies of both methods is crucial for designing proteomics experiments that produce high-quality and biologically meaningful results. As instrument performance continues to advance—exemplified by the ultra-high speed and resolution of platforms such as the Astral analyzer—and computational algorithms evolve—such as deep learning-based approaches for DIA data analysis—the functional capabilities and applicability of both DIA and DDA will continue to expand and evolve.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12014-025-09572-2.

Additional file 1.

Additional file 2.

Additional file 3.

Additional file 4.

Additional file 5.

Additional file 6.

Additional file 7.

Author contributions
Xun Zou and Lulu Wang conducted the experiments, including cell culture, proteomics experiments, and LC–MS/MS detection. They also analyzed the proteomics data and drafted the manuscript. Yulu Chen and Yuan Gao assisted with the proteomics experiments. Hang Fu contributed to the analysis of proteomics data. Bin Liu provided support in project supervision. Linhui Zhai supervised the project, reviewed and edited the manuscript. All authors reviewed and contributed to the final version of the manuscript.

Funding
This work was supported by grants from the National Natural Science Foundation of China, grant number: 32171434, 22577093, 22225702, the Program of Shanghai Academic Research Leader, grant number: No. 22XD1420900, the State Key Laboratory of Drug Research, grant number: SIMM2105KF-13, the Postgraduate Research & Practice Innovation Program of jiangsu Ocean University, grant number: KYCX2024-70.

Data availability
All the mass spectrometry data and the database search result have been deposited to the ProteomeXchange Consortium via the iProX partner with the project ID IPX0013929001.

Declarations

Competing interests
The authors declare no competing interests.

Received: 22 August 2025 / Accepted: 10 November 2025
Published online: 11 December 2025

References

1. Noreldeen HAA. Recent advances in lung cancer lipidomics: analytical techniques and their applications. Clin Chim Acta. 2025;577:120470. https://doi.org/10.1016/j.cca.2025.120470.
2. Li Y, Zhang S, Wang X, Li X, Guo L. LC-MS/MS quantification of nusinersen in rat cerebrospinal fluid and preclinical pharmacokinetics study application. Bioanalysis. 2025. https://doi.org/10.1080/17576180.2025.2535949.
3. David PA, Norazhar AI, Che Zain MS. Integrating LC-MS/MS and molecular networking for advance analysis and comprehensive flavonoids annotation. J Sep Sci. 2025;48(7):e70230. https://doi.org/10.1002/jssc.70230.
4. Zheng R, Matzinger M, Mayer RL, Valenta A, Sun X, Mechtler K. A high-sensitivity low-nanoflow LC-MS configuration for high-throughput sample-limited proteomics. Anal Chem. 2023;95(51):18673–8. https://doi.org/10.1021/acs.analchem.3c03058.
5. Yu F, Teo GC, Kong AT, et al. Analysis of DIA proteomics data using MSFragger-DIA and FragPipe computational platform. Nat Commun. 2023;14(1):4154. https://doi.org/10.1038/s41467-023-39869-5.
6. Xue Z, Zhu T, Zhang F, et al. DPHL v.2: an updated and comprehensive DIA pan-human assay library for quantifying more than 14,000 proteins. Patterns. 2023;4(7):100792. https://doi.org/10.1016/j.patter.2023.100792.
7. Woessmann J, Petrosius V, Uresin N, et al. Assessing the role of trypsin in quantitative plasma and single-cell proteomics toward clinical application. Anal Chem. 2023;95(36):13649–58. https://doi.org/10.1021/acs.analchem.3c02543.
8. Wen C, Wu X, Lin G, et al. Evaluation of DDA library-free strategies for phosphoproteomics and ubiquitinomics data-independent acquisition data. J Proteome Res. 2023;22(7):2232–45. https://doi.org/10.1021/acs.jproteome.2c00735.
9. Wang H, Lim KP, Kong W, et al. Multiprop: DDA-PASEF and diaPASEF acquired cell line proteomic datasets with deliberate batch effects. Sci Data. 2023;10(1):858. https://doi.org/10.1038/s41597-023-02779-8.
10. van Leeuwen SJM, Proctor GB, Staes A, et al. The salivary proteome in relation to oral mucositis in autologous hematopoietic stem cell transplantation recipients: a labelled and label-free proteomics approach. BMC Oral Health. 2023;23(1):460. https://doi.org/10.1186/s12903-023-03190-w.
11. Suran J, Radic B, Trevisan-Silva D, Cindric M, Hozic A. First proteome analysis of poplar-type propolis. Plant Foods Hum Nutr. 2024;79(1):83–9. https://doi.org/10.1007/s11130-023-01127-w.
12. Rudt E, Feldhaus M, Margraf CG, et al. Comparison of data-dependent acquisition, data-independent acquisition, and parallel reaction monitoring in trapped ion mobility spectrometry-time-of-flight tandem mass spectrometry-based lipidomics. Anal Chem. 2023;95(25):9488–96. https://doi.org/10.1021/acs.analchem.3c00440.
13. Xin L, Qiao R, Chen X, et al. A streamlined platform for analyzing tera-scale DDA and DIA mass spectrometry data enables highly sensitive immunopeptidomics. Nat Commun. 2022;13(1):3108. https://doi.org/10.1038/s41467-022-30867-7.
14. Coscia F, Doll S, Bech JM, et al. A streamlined mass spectrometry-based proteomics workflow for large-scale FFPE tissue analysis. J Pathol. 2020;251(1):100–12. https://doi.org/10.1002/path.5420.
15. Ball B, Sukumaran A, Krieger JR, Geddes-McAlister J. Comparative cross-kingdom DDA- and DIA-PASEF proteomic profiling reveals novel determinants of fungal virulence and a putative druggable target. J Proteome Res. 2024;23(9):3917–32. https://doi.org/10.1021/acs.jproteome.4c00255.
16. Wang A, Fekete EEF, Creskey M, Cheng K, Ning Z, Pfeifle A, et al. Assessing fecal metaproteomics workflow and small protein recovery using DDA and DIA PASEF mass spectrometry. Microbiome Res Rep. 2024;3(3):39. https://doi.org/10.20517/mrr.2024.21.
17. Hoffman MA, Fang B, Haura EB, Rix U, Koomen JM. Comparison of quantitative mass spectrometry platforms for monitoring kinase ATP probe uptake in lung cancer. J Proteome Res. 2018;17(1):63–75. https://doi.org/10.1021/acs.jproteome.7b00329.
18. Shah SMZ, Ali A, Khan MN, et al. Sensitive detection of pharmaceutical drugs and metabolites in serum using data-independent acquisition mass spectrometry and open-access data acquisition tools. Pharmaceuticals (Basel). 2022. https://doi.org/10.3390/ph15070901.
19. Rajczewski AT, Blakeley-Ruiz JA, Meyer A, Vintila S, McIlvin MR, Van Den Bossche T, et al. Data-independent acquisition mass spectrometry as a tool for metaproteomics: interlaboratory comparison using a model microbiome. Proteomics. 2025;25:e202400187. https://doi.org/10.1002/pmic.202400187.
20. Wandy J, McBride R, Rogers S, et al. Simulated-to-real benchmarking of acquisition methods in untargeted metabolomics. Front Mol Biosci. 2023;10:1130781. https://doi.org/10.3389/fmolb.2023.1130781.
21. Bekker-Jensen DB, Bernhardt OM, Hogrebe A, et al. Rapid and site-specific deep phosphoproteome profiling by data-independent acquisition without the need for spectral libraries. Nat Commun. 2020;11(1):787. https://doi.org/10.1038/s41467-020-14609-1.
22. Reilly L, Lara E, Ramos D, et al. A fully automated FAIMS-DIA mass spectrometry-based proteomic pipeline. Cell Rep Methods. 2023;3(10):100593. https://doi.org/10.1016/j.crmeth.2023.100593.
23. Frankenfield AM, Ni J, Ahmed M, Hao L. Protein contaminants matter: building universal protein contaminant libraries for DDA and DIA proteomics. J Proteome Res. 2022;21(9):2104–13. https://doi.org/10.1021/acs.jproteome.2c00145.

Publisher's Note