Supporting information for

Label Free Quantification from Direct Infusion Shotgun Proteome Analysis (DISPA-LFQ) with CsoDIAq Software

*Yuming Jiang [1,2], Alexandre Hutton[1], Caleb W. Cranney [2], Jesse G. Meyer[1,2]**

[1] Department of Computational Biomedicine, Cedars Sinai Medical Center, Los Angeles, CA 90048, USA.
[2] Department of Biochemistry, Medical College of Wisconsin, Milwaukee, Wisconsin 53226, United States
*Corresponding author: *Jesse G. Meyer*; E-mail: jesse.meyer@cshs.org

Table of contents

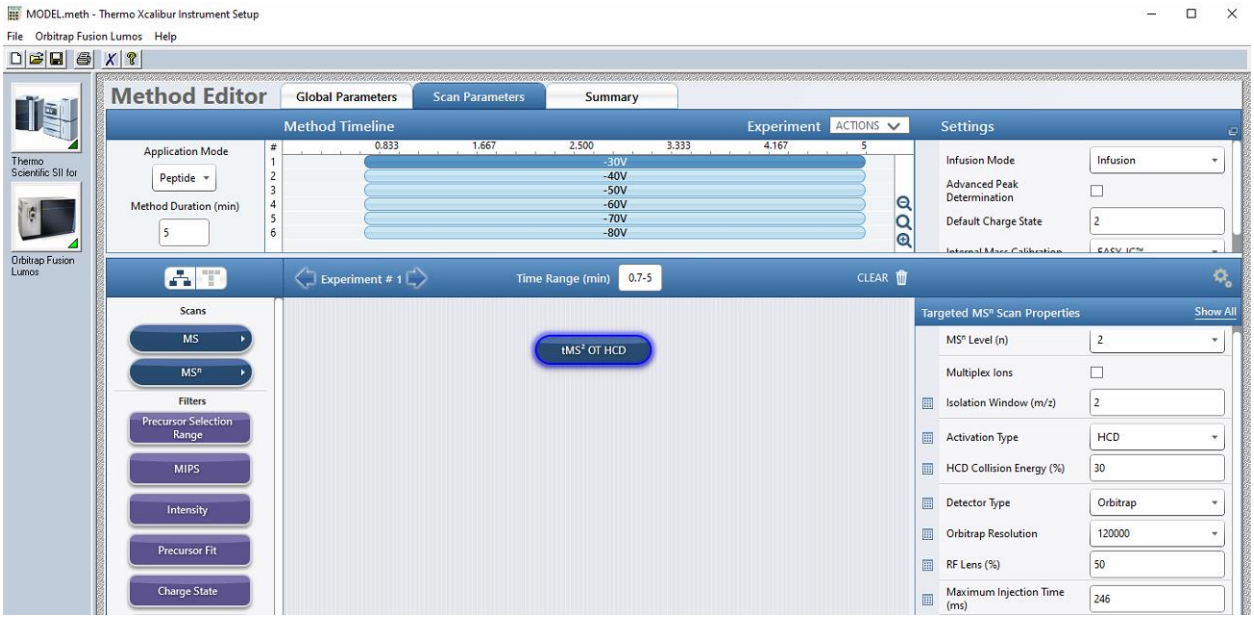
Supplemental Methods

Scouting experiment

A scouting experiment is used for finding potential targeted *m/z*. Peptides were analyzed by data independent acquisition (DIA) in targeted MS_2 (tMS_2) mode with 2 *m/z* quadrupole isolation window stepping across the *m/z* 400-1000 with 0.5 *m/z* overlap. The FAIMS compensation voltages stepped from -30V to -80V in a step of 10V in a single scouting experiment. The tune method was set at 300% normalized automatic gain control (AGC) target, 30% normal collisional energy of higher-energy collisional dissociation (HCD), standard maximum ion-accumulation time, and 120k orbitrap mass resolution. The spray voltage was maintained at 2000v for positive ion mode, ion transfer tube temperature was set at 310 °C. Sheath gas, AUC gas and sweep gas was set at 15, 3 and 12, respectively. The FAIMS Pro system was set to standard resolution mode with six FAIMS compensation voltages (-30v, -40v, -50v, -60v, -70v, -80v). The FAIMS experiment used a carrier N_2 gas flow at 5 L/min, asymmetric waveform dispersion voltage at −5000 V, and the inner and outer electrode temperatures set at 100 °C.

Targeted experiment

For targeted peptides analysis, peptides were analyzed with DIA in targeted MS_2 (tMS_2) mode with 1*m/z* quadrupole isolation window for all targeted peptide *m/z*. Targeted peptides' *m/z* are come from the peptides identified in scouting experiment. Here our CsoDIAq software can automatically identify all peptides and export .txt files of targeted m/z for each voltage, which can be further used for Targeted window settings. The source, FAIMS and tune settings of targeted experiment were the same as scouting experiment, and orbitrap mass resolution was varied according to experiment design. A screenshot of our method was shown below as an example.



Library build:

Procedures:

1. LC-MS/MS analysis with data dependent acquisition (DDA)

    - Perform LC-MS/MS analysis of experimental samples with DDA for same compensation voltages you plan to use for subsequent DISPA experiments. Here we used 11CVs from -30V to -80V in a step of 5V for 293T proteome analysis.

    - Tip: Extra compensation voltages or repeated injections can be collected if you want a better coverage in your library.

2. Get peptide identifications in pep.xml files from FragPipe

    - Download Fragpipe from https://fragpipe.nesvilab.org/ and install.

    - Using Fragpipe to produce pepxml files of peptides and proteins. Instructions of FragPipe can also be found on the above website.

    - Select appropriate initial reference fasta database for analysis. The initial reference fasta database we used here for 293T library build is 2022-08-02-decoys-reviewed-contam-UP000005640.fas, File contains 40840 entries (20420 decoys: 50.0%).

3. Produce spectra library files with SpectraST.

    - Here we use SpectraST, a spectral library building and searching tool, to build library for CsoDIAq software. The SpectraST library creation commands we used are provided as a file in supplementary .txt file

    - Details for commands of SpectraST can be found in website as below:

http://tools.proteomecenter.org/wiki/index.php?title=Software:SpectraST#Creating_Libraries_from_Sequence_Search_Results

Frequently Asked Questions (FAQ)

Q1. What is the role of FAIMS?

A1: Due to no LC separation in direct infusion analysis, all peptides are ionized at the same time, and thus the number of peptides for one DIA scan window is tremendous (see extended data figure 1 from Meyer et al., Nature Methods, 2020). As a result, the ion competition effect from high abundant peptides will be very strong and any detectable signals in the tandem mass spectrum will contain fragments from many peptides. To reduce this problem, an additional gas phase separation, such as FAIMS, was added here to separate peptides after ionization and decrease the strong ion suppression effect in mass analyzer. Thus, FAIMS reduces the complexity of ions entering the orbitrap to decrease the ion competition effect in the mass analyzer and increase the identification rate of peptides.

Q2. How much sample volume is required for each experiment?

A2: The total required volume per sample is determined by the total acquisition time of the experiment. The minimum requirement is to make sure sample volume is enough for the whole acquisition time. Here in targeted experiments, the flow rate we used is 0.25 µl/min and the acquisition time of targeted analysis is 3.6 min, thus 1µl is enough for the whole acquisition because 1 µl would last 4 minutes at that flow rate. In scouting experiments, the flow rate we used is still 0.25 µl/min and the total acquisition time is around 11 min at a resolution of 120K, thus 3µl is enough for the whole experiment. Specifically, we used a 5µl sample loop and inject needed sample volume with "User Defined Program" of ultimate 3000 sampler.

Q3. How many fragment matches are required for a csoDIAq match and what is the mass tolerance of matching a fragment?

A3: CsoDIAq requires at least 3 fragment ion matches for a match to receive a score, but CsoDIAq uses a score that includes the number of matched fragments, so more matched fragments give a better score (See original CsoDIAq paper, https://doi.org/10.1021/acs.analchem.1c02021). CsoDIAq also includes a mass correction module, where after the first search with user-defined ppm mass range for fragment matches (30 ppm), the true mass offset and range is discovered from the fragment matches. CsoDIAq outputs an image of a fragment mass error histogram with every search. A second search is then performed with the corrected tolerance. Usually for orbitrap data the range of the new mass tolerance is approximately 3ppm, and the offset depends on when the instrument was calibrated last and how much the mass accuracy has drifted."

Q4. How many MS/MS spectra are produced by DISPA and what is the identification rate of MS/MS spectra by csoDIAq?

A4: The number of MS/MS spectra is determined by the number of targeted m/z we decide to apply. Typically, in a non-targeted scouting experiment, we scanned every MS window between 400-1000 m/z in steps of 2 with 0.5 overlap giving 600 windows. We used 6 CVs*400 scans/CV = 2400 scans in each scouting experiment, which means 2400 tandem mass spectra. In a case of 1µg/µl and 120K mass resolution from HeLa peptides, we identified 5288 peptides and thus the identification rate of CsoDIAq software is 220%. However, when using scouting experiments, we found there are a lot of scans have very low signals

and no peptide identifications. Thus, we usually remove these “empty” scans and use targeted DISPA approach for real analysis. In typical targeted DISPA analysis of Hela sample under different experimental settings we showed in Figure 2, there are 768 scan segments were targeted and 768 MS/MS spectra were produced for one analysis. As shown in Figure S8, the number of identified peptide features varied with different settings like mass resolution and concentration. In a case of 1μg/μl and 120K mass resolution in HeLa, we identified 4640 peptides on average and the identification rate of CsoDIAq software is 604%.

Q5. What is a projected spectrum?

A5: A projected spectrum is a method of getting only the relevant fragment ions from an experimental spectrum before computing a match score to a library spectrum. The concept was originally introduced in 2010 by Wang et al. in Molecular and Cellular Proteomics (“Peptide Identification from Mixture Tandem Mass Spectra”, doi: 10.1074/mcp.M000136-MCP201) and subsequently applied to identify peptides from DIA data in 2015 by Wang et al. in Nature Methods (“MSPLIT-DIA: sensitive peptide identification for data-independent acquisition”, doi: 10.1038/nmeth.3655). The idea is that given an experimental spectrum that contains fragments from many peptides, we select only the subset of fragments in that spectrum that might match to a library spectrum generated from one pure peptide based on some mass tolerance. By doing this before computing the spectra-spectra match score, irrelevant fragment ions that would lower the score are ignored.

Supplemental Figures

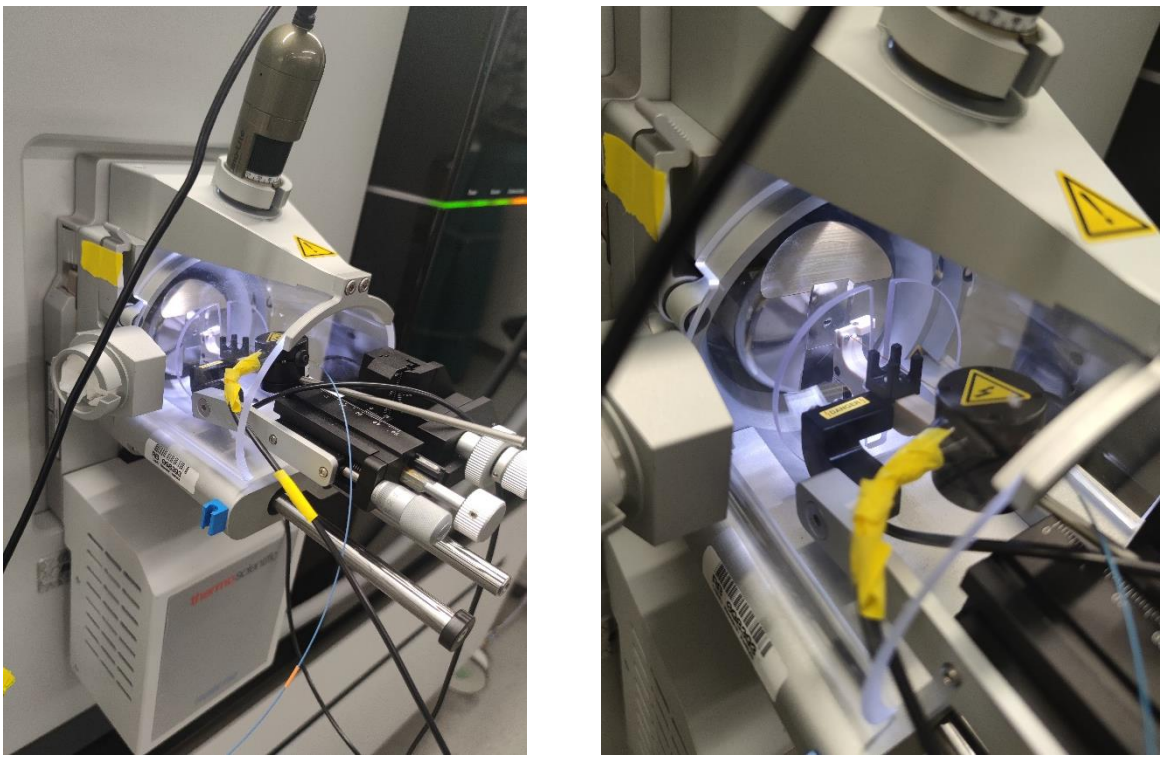


Figure S1. Picture of Nano-ESI and FAIMS for DISPA platform.

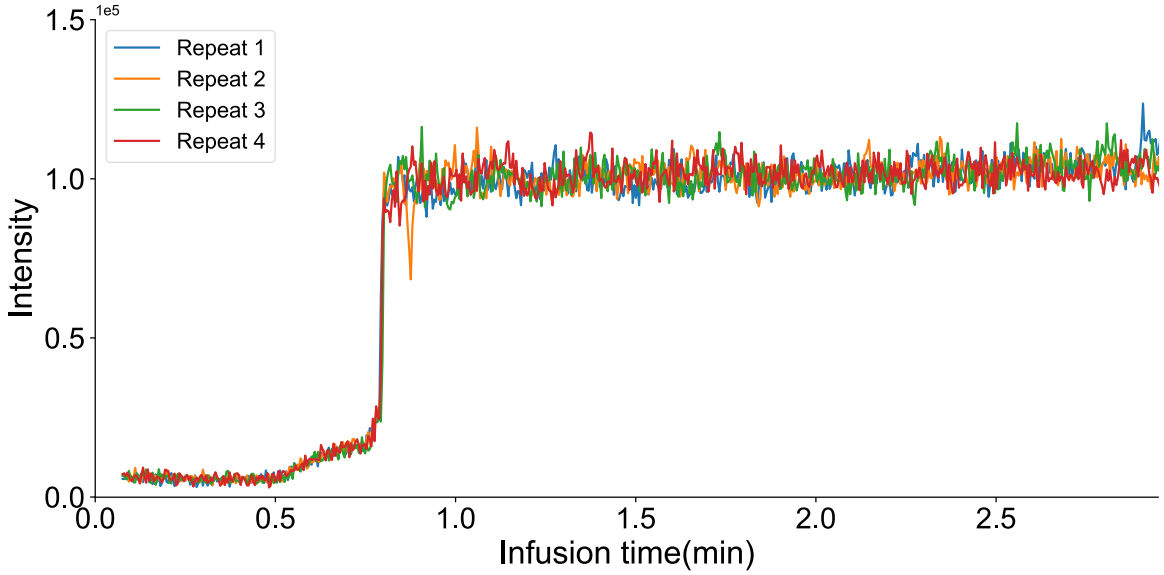


Figure S2. Total ion chromatogram (TIC) from infusion of 1 fmol/μL Angiotensin I of four repeats.

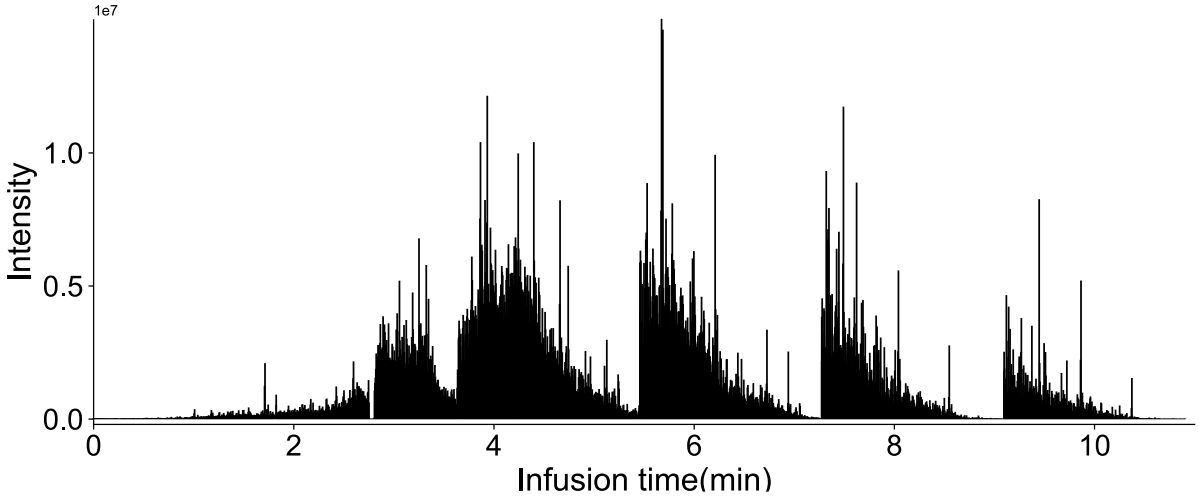


Figure S3. Typical total ion chromatogram (TIC) of scouting experiment of 0.15μg/μL Hela peptide standard.

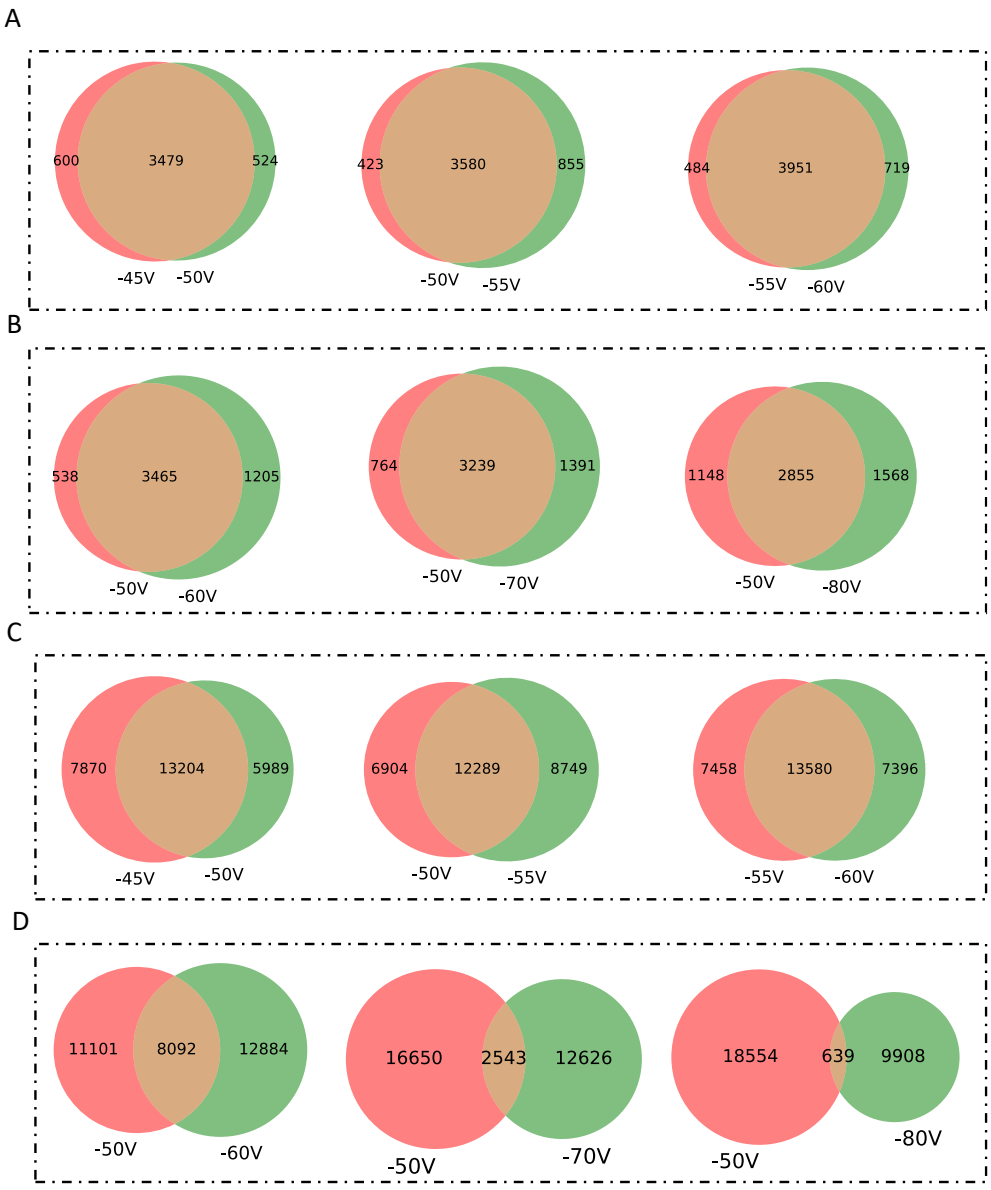


Figure S4. Comparative analysis of peptide and protein identifications under different compensation voltages (CVs) by LC-MS/MS in 293T proteome. (A) Venn diagram show duplicate proteins between adjacent CVs. (B) Venn diagram show duplicate proteins between distanced CVs. (C) Venn diagram show duplicate peptides between adjacent CVs. (D) Venn diagram show duplicate peptides between distanced CVs. Mass resolution: 120K. Sample Loading amount: 1μg.

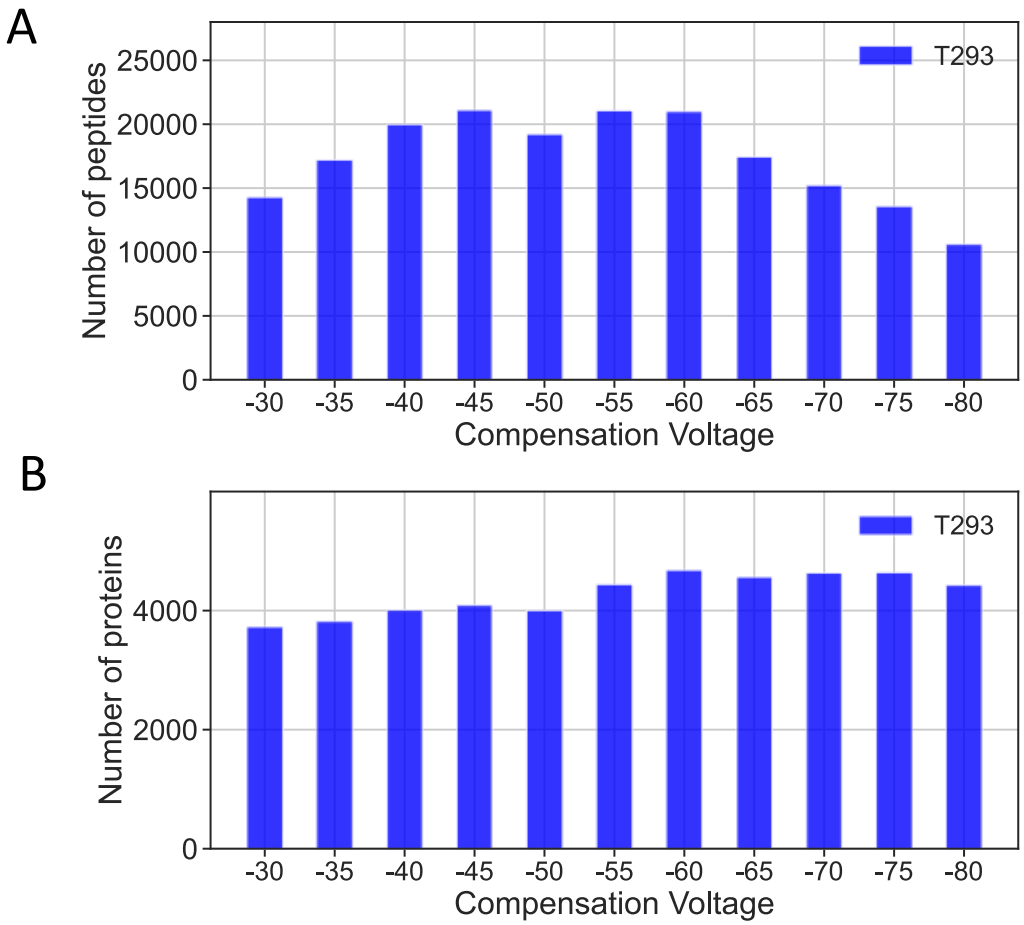
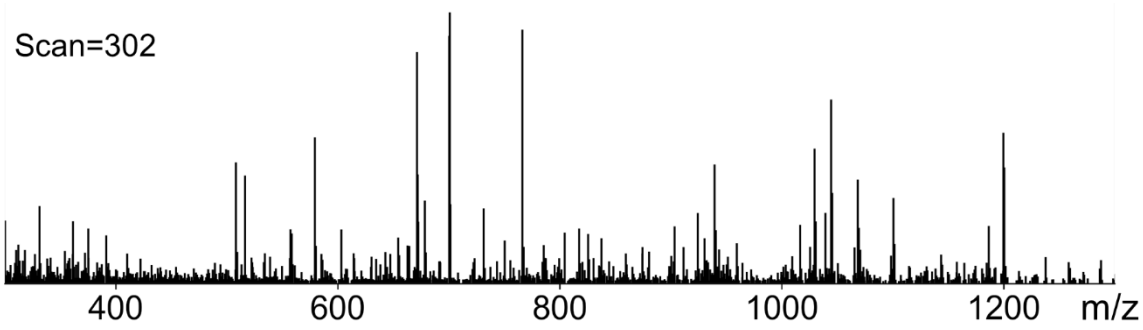
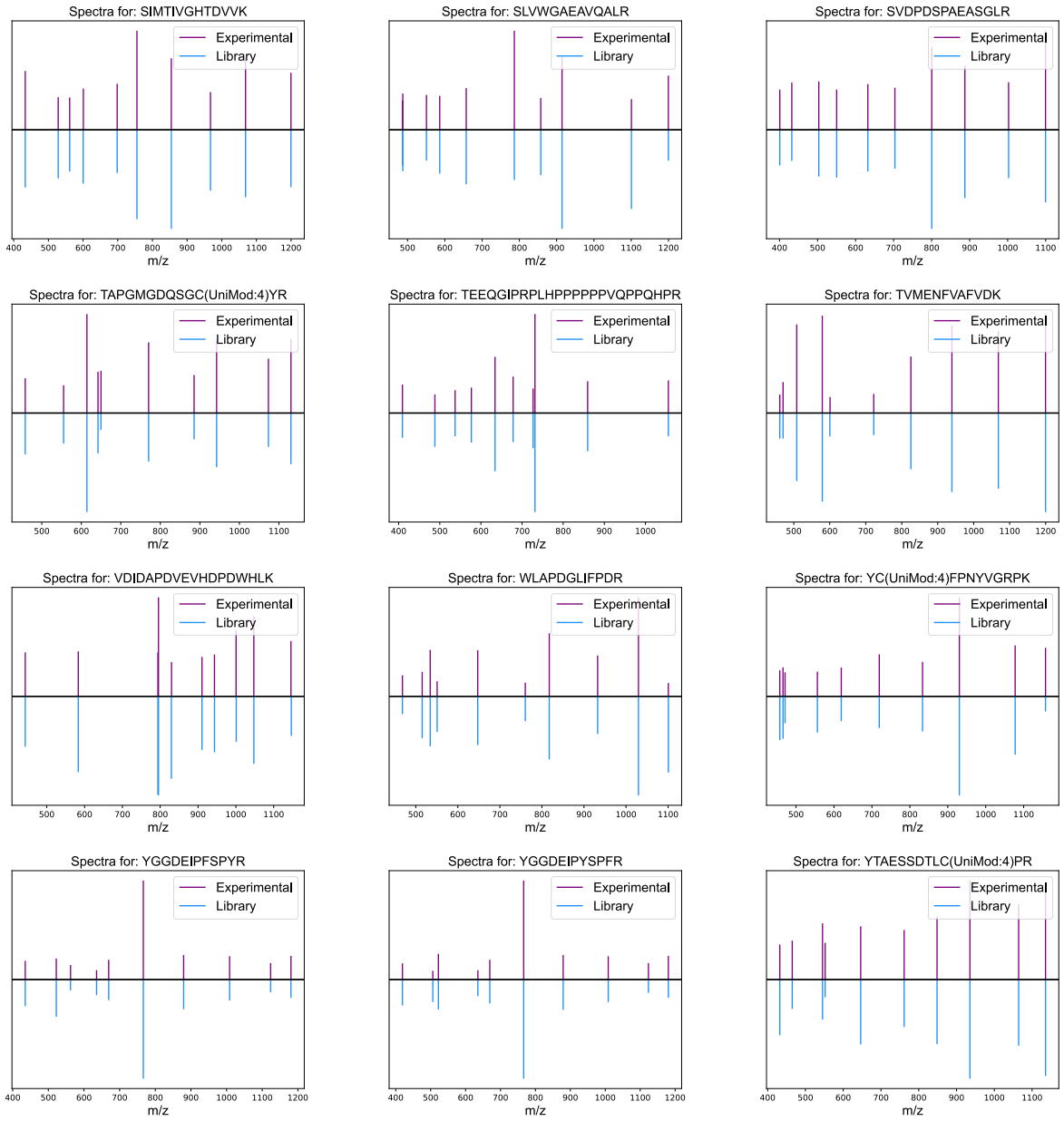


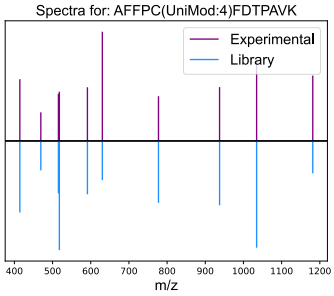
Figure S5. Peptide and protein identifications from stepped compensation voltages (CVs) with LC-MS/MS of the 293T proteome. (A) Peptide identifications of different CVs in 293T proteome. Mass resolution: 120K. Sample Loading amount: 1µg. (B) Protein identifications of different CVs in 293T proteome. Mass resolution: 120K. Sample Loading amount: 1µg.

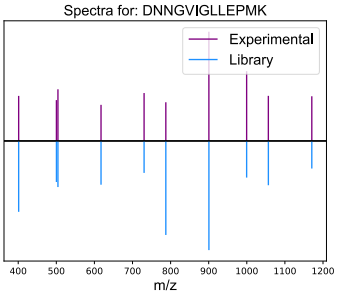
A

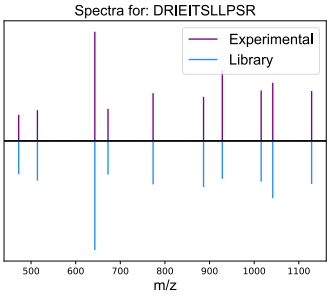

Scan=302
400
600
800
1000
1200
m/z

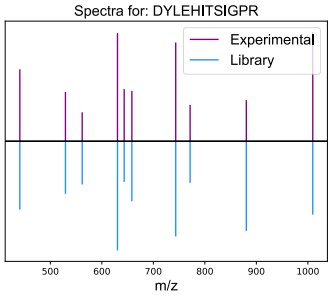
B

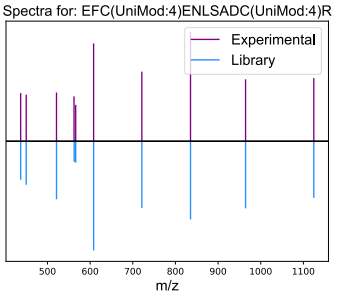

Spectra for: SIMTIVGHTDVVK
Spectra for: SLVWGAEAVQALR
Spectra for: SVDPDSPAEASGLR
Spectra for: TAPGMGDQSGC(UniMod:4)YR
Spectra for: TEEQGIPRPLHPPPPPPVQPPQHPR
Spectra for: TVMENFVAFVDK
Spectra for: VDIDAPDVEVHDPDWHLK
Spectra for: WLAPDGLIFPDR
Spectra for: YC(UniMod:4)FPNYVGRPK
Spectra for: YGGDEIPFSPYR
Spectra for: YGGDEIPYSPFR
Spectra for: YTAESSDTLC(UniMod:4)PR
Experimental
Library
m/z

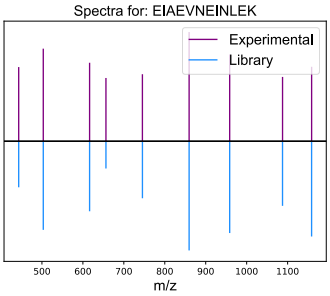

Spectra for: AFFPC(UniMod:4)FDTPAVK
Experimental
Library
400 500 600 700 800 900 1000 1100 1200
m/z

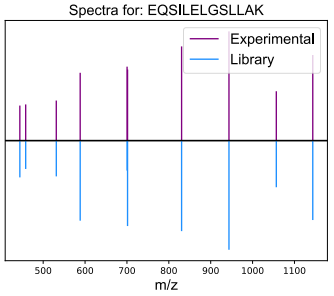

Spectra for: DNNGVIGLLEPMK
Experimental
Library
400 500 600 700 800 900 1000 1100 1200
m/z

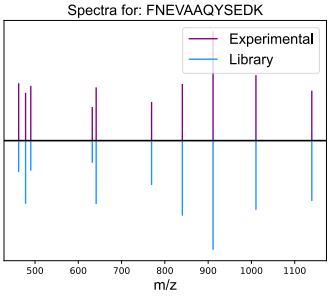

Spectra for: DRIEITSLLPSR
Experimental
Library
500 600 700 800 900 1000 1100
m/z

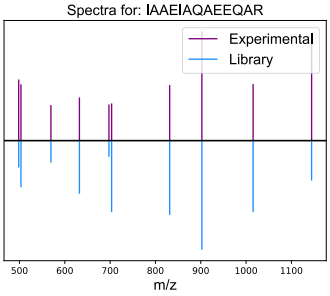

Spectra for: DYLEHITSIGPR
Experimental
Library
500 600 700 800 900 1000
m/z

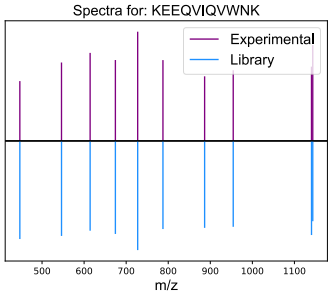

Spectra for: EFC(UniMod:4)ENLSADC(UniMod:4)R
Experimental
Library
500 600 700 800 900 1000 1100
m/z

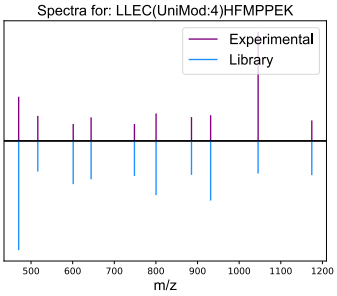

Spectra for: EIAEVNEINLEK
Experimental
Library
500 600 700 800 900 1000 1100
m/z

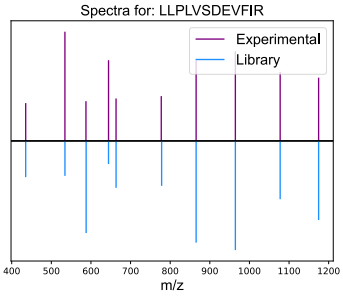

Spectra for: EQSILELGSLLAK
Experimental
Library
500 600 700 800 900 1000 1100
m/z


Spectra for: FNEVAAQYSEDK
Experimental
Library
500 600 700 800 900 1000 1100
m/z


Spectra for: IAAEIAQAEEQAR
Experimental
Library
500 600 700 800 900 1000 1100
m/z


Spectra for: KEEQVIQVWNK
Experimental
Library
500 600 700 800 900 1000 1100
m/z


Spectra for: LLEC(UniMod:4)HFMPPEK
Experimental
Library
500 600 700 800 900 1000 1100 1200
m/z


Spectra for: LLPLVSDEVFIR
Experimental
Library
400 500 600 700 800 900 1000 1100 1200
m/z

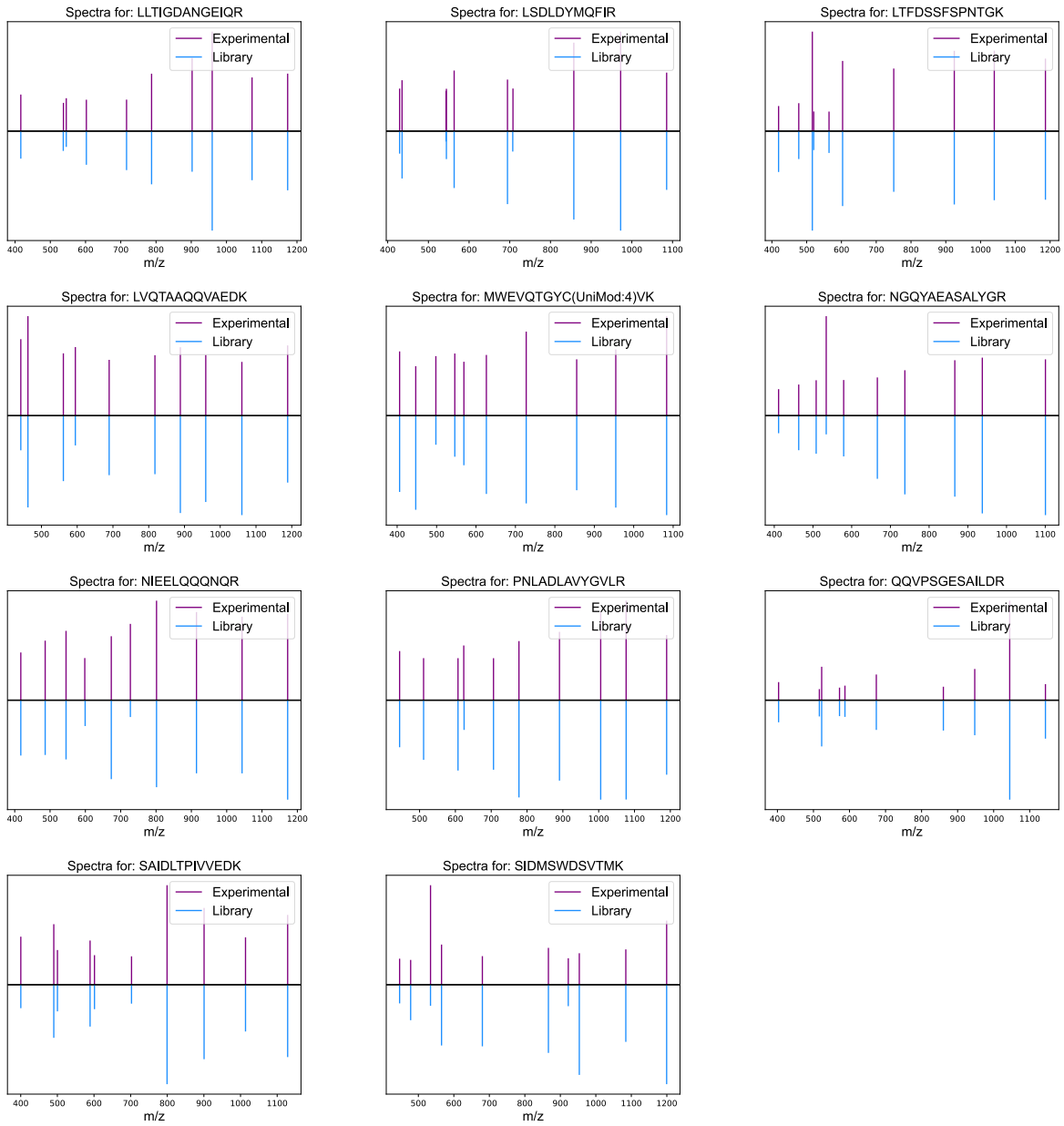


Figure S6. Comparison between library spectrum and projected spectrum of 35 peptides identified in a single scan. (A) raw scan #302 from HeLa sample in 1 μg/μl and 240K. (B) Comparison between library spectrum and projected spectrum of 35 peptides identified in this scan.

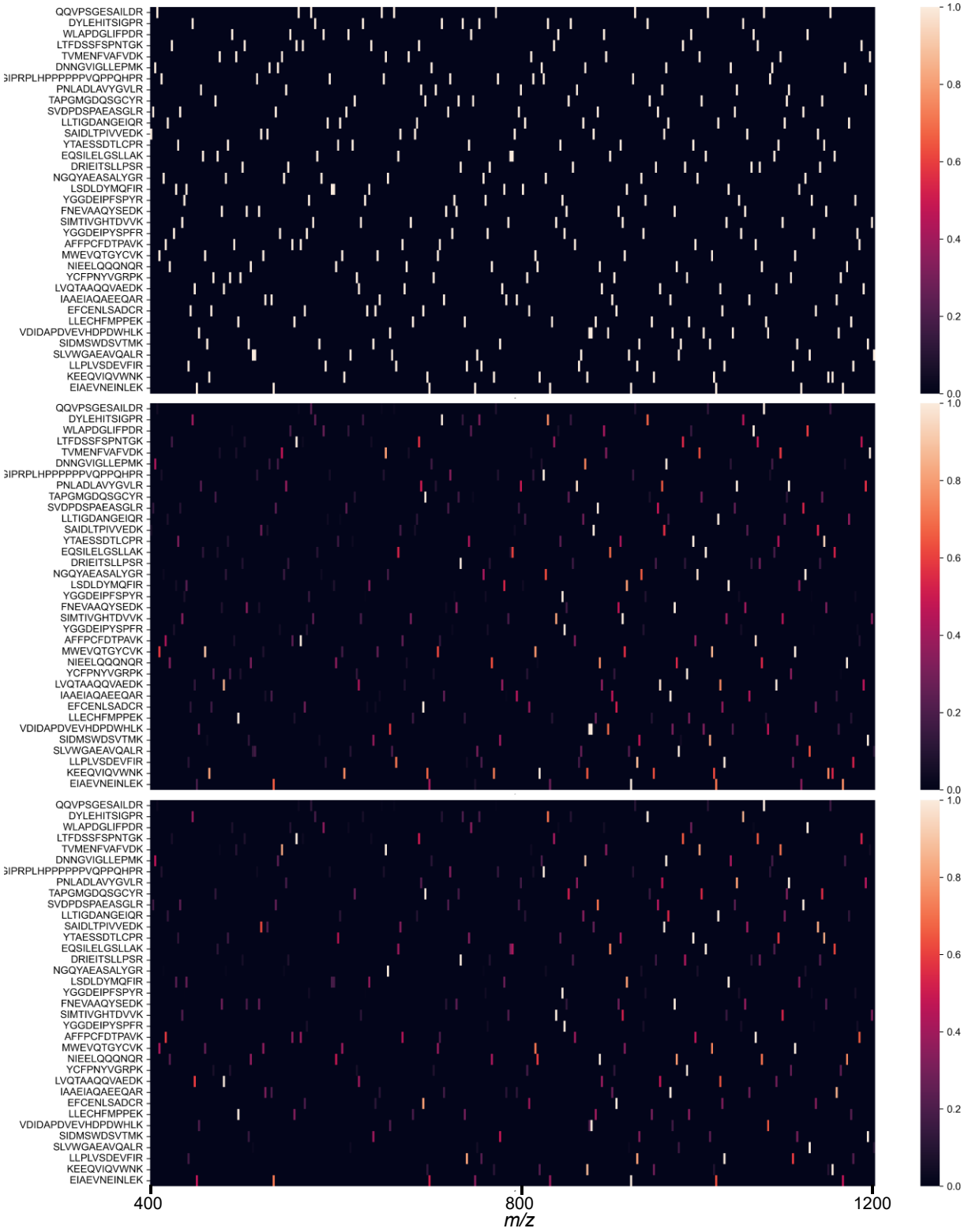

QQVPSGESAILDR
DYLEHITSIGPR
WLAPDGLIFPDR
LTFDSSFSPNTGK
TVMENFVAFVDK
DNNGVIGLLEPMK
GIPRPLHPPPPPPVQPPQHPR
PNLADLAVYGVLR
TAPGMGDQSGCYR
SVDPDSPAEASGLR
LLTIGDANGEIQR
SAIDLTPIVVEDK
YTAESSDTLCPR
EQSILELGSLLAK
DRIEITSLLPSR
NGQYAEASALYGR
LSDLDYMQFIR
YGGDEIPFSPYR
FNEVAAQYSEDK
SIMTIVGHTDVVK
YGGDEIPYSPFR
AFFPCFDTPAVK
MWEVQTGYCVK
NIEELQQQNQR
YCFPNYVGRPK
LVQTAAQQVAEDK
IAAEIAQAEEQAR
EFCENLSADCR
LLECHFMPPEK
VDIDAPDVEVHDPDWHLK
SIDMSWDSVTMK
SLVWGAEAVQALR
LLPLVSDEVFIR
KEEQVIQVWNK
EIAEVNEINLEK
1.0
0.8
0.6
0.4
0.2
0.0
400
800
1200
m/z

Figure S7. Comparison of fragment ions between all 35 peptide identifications from scan 302. (A) raw scan (302) from 1 μg/μl, 240K. (projected spectrum). The relative intensities of all fragments were set equal to 1.0. (B) Heatmap of all 35 peptides from projected spectra. The color gradient represents the relative intensities of fragments. (C) Heatmap of all 35 peptides from library. Color gradient represents the relative intensities of fragments.

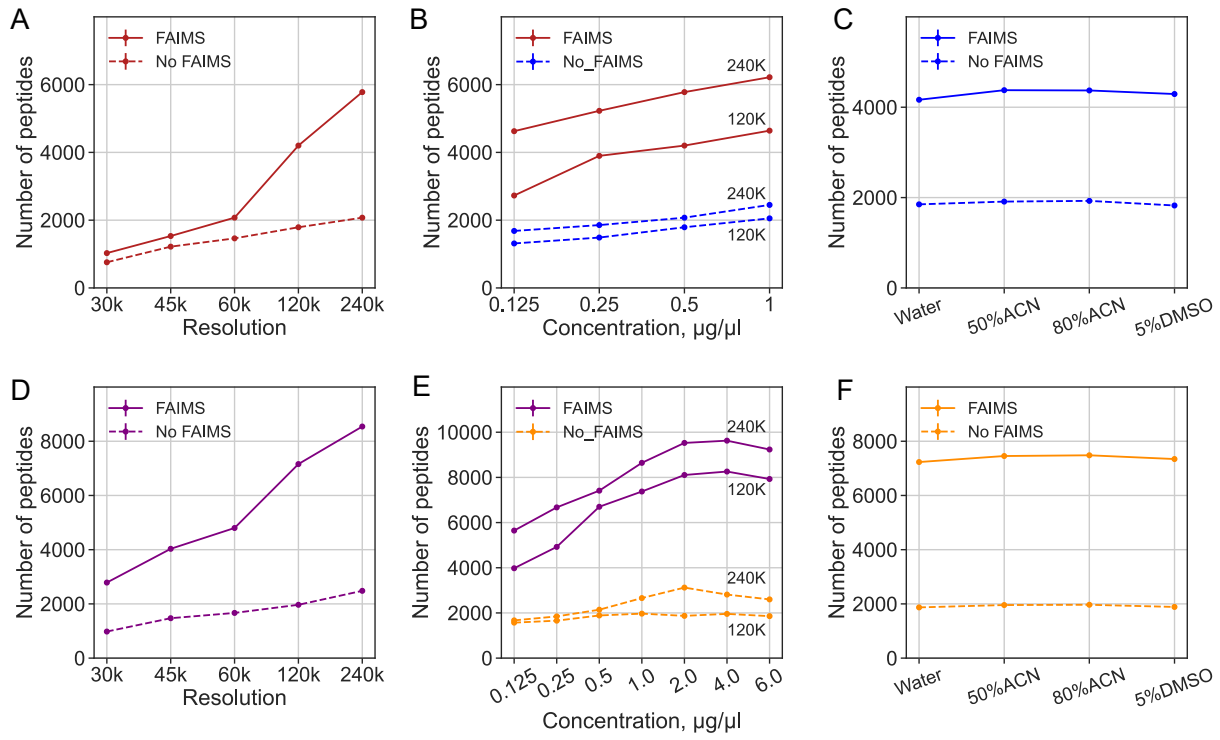


Figure S8. Peptide identifications of HeLa and 293T proteome by DISPA analysis with and without FAIMS. **A.B.C.** Number of peptide identifications in HeLa proteome as a function of different mass resolution (A), sample concentration (B) and resuspend solvent (C) with and without FAIMS. **D.E.F.** Number of peptide identifications in 293T proteome as a function of different mass resolution (D), sample concentration (E) and resuspend solvent (F) with and without FAIMS. For mass resolution analysis, Sample concentration: HeLa (0.5 μg/μl), 293T: (1μg/μl). For solvent analysis, Sample concentration: HeLa (0.5 μg/μl), 293T: (1μg/μl). Mass resolution: HeLa (120K), 293T: (120K).

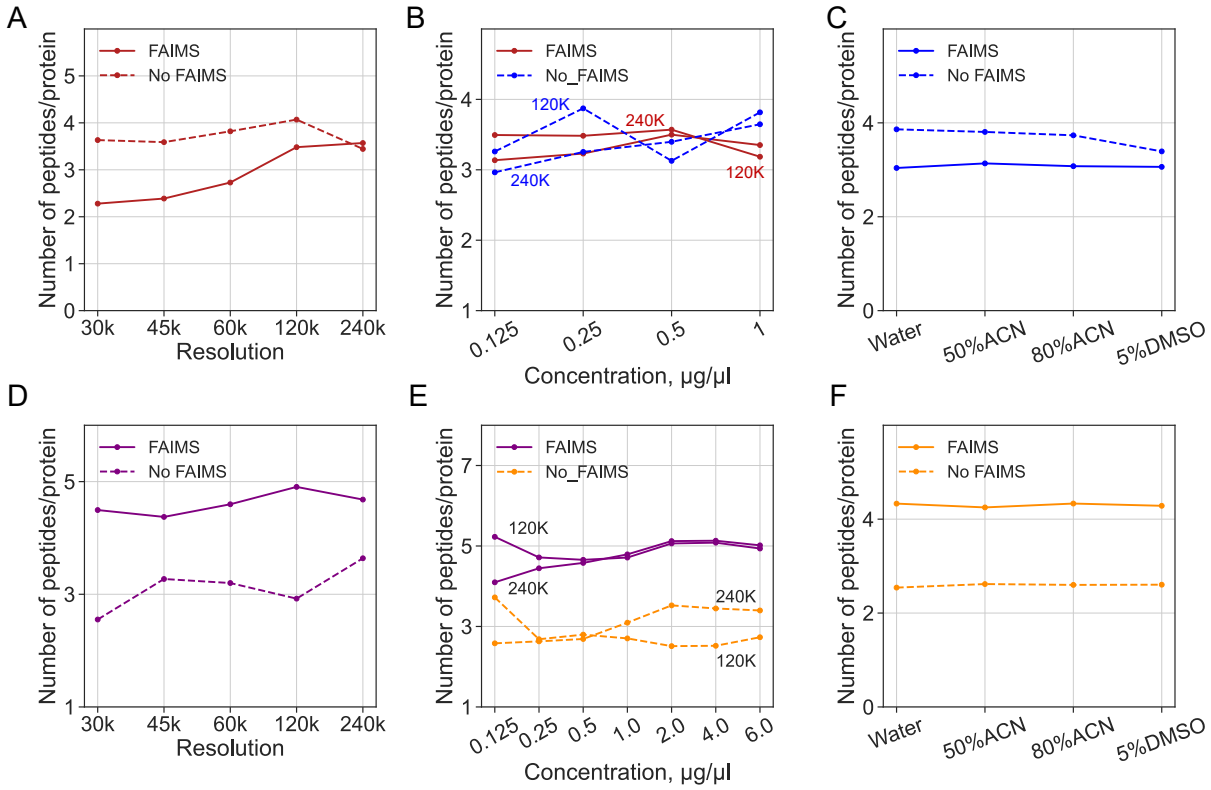


Figure S9. Peptide identifications per protein of HeLa and 293T proteome by DISPA analysis with and without FAIMS. A.B.C. Number of peptide identifications per protein in HeLa proteome as a function of different mass resolution (A), sample concentration (B) and resuspension solvent (C) with and without FAIMS. **D.E.F.** Number of peptide identifications per protein in 293T proteome as a function of different mass resolution (D), sample concentration (E) and resuspension solvent (F) with and without FAIMS. For mass resolution analysis, Sample concentration: HeLa (0.5 µg/µl), 293T: (1µg/µl). For solvent analysis, Sample concentration: HeLa (0.5 µg/µl), 293T: (1µg/µl). Mass resolution: HeLa (120K), 293T: (120K).

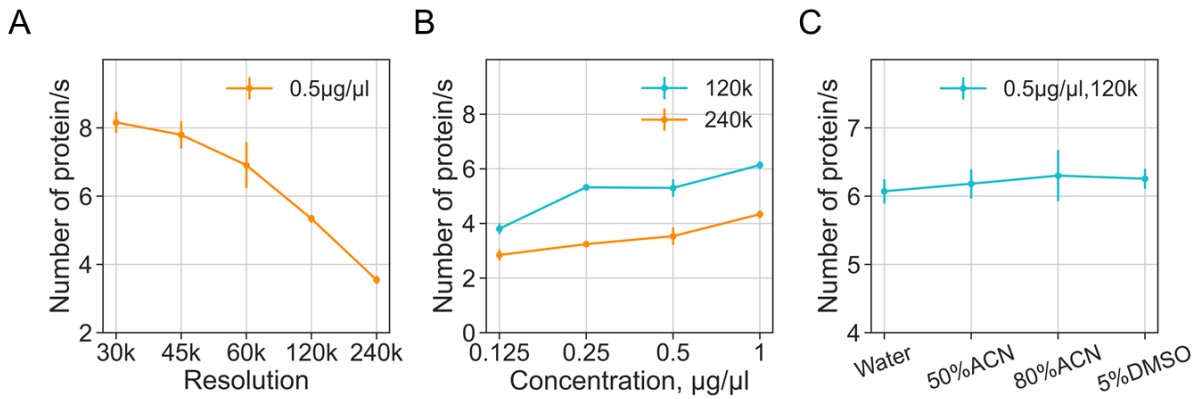


Figure S10. Number of unique proteins identified per second as a function of different mass resolution(A), sample concentration(B) and resuspend solvent(C). proteins per second was calculated from the total number of proteins identified divided by the total acquisition time. Sample: HeLa proteome.

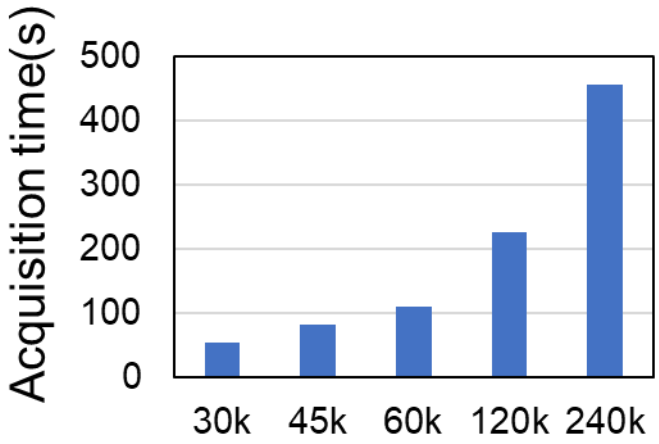


Figure S11. Acquisition time of different Mass resolution. Concentration: 0.5 µg/µl.

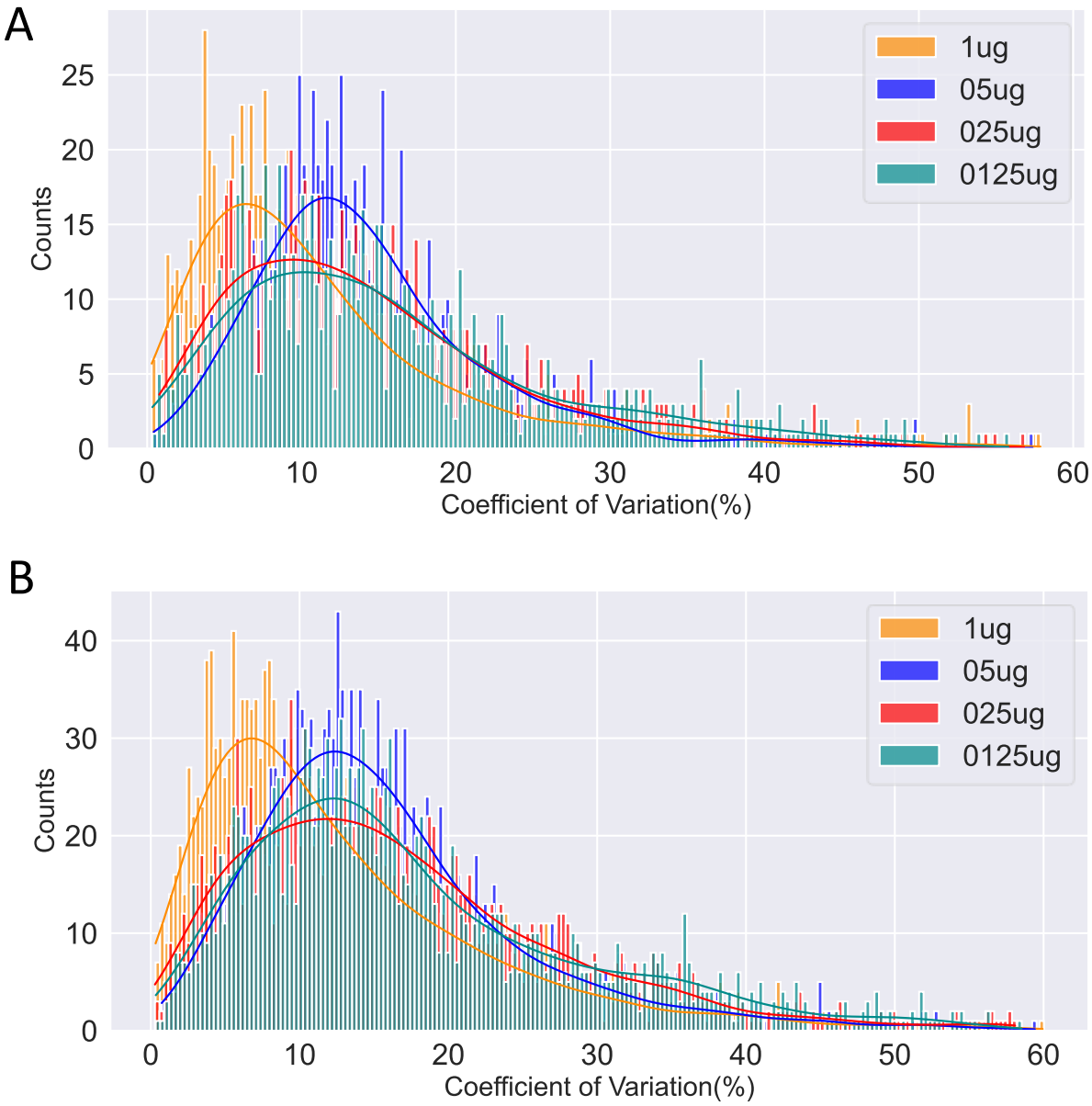


Figure S12. Analysis of Coefficient of Variations from all peptides and proteins across three replicates. (A) Histogram of Coefficient of Variations from all proteins across three replicates. (B) Histogram of Coefficient of Variations from all peptides across three replicates. Mass resolution: 240K. Sample: HeLa proteome.

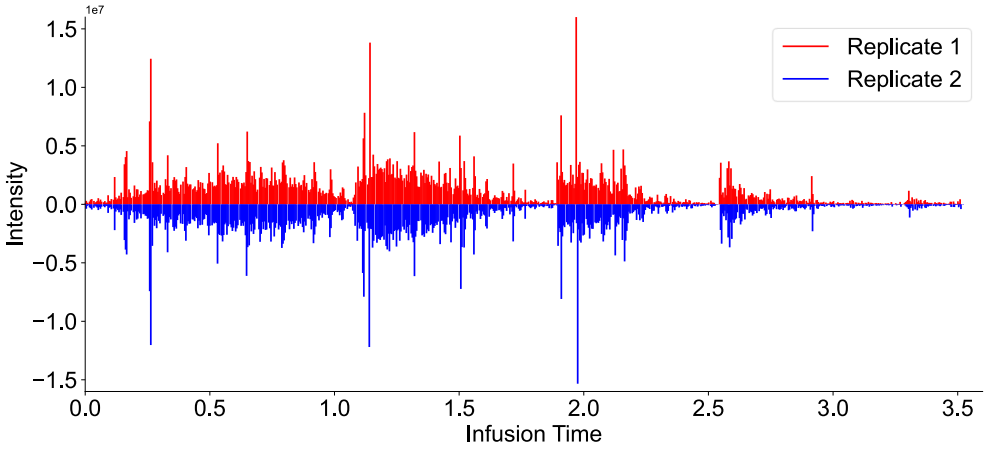


Figure S13. Comparison of total ion chromatogram (TIC) and mass spectra between two replicates of HeLa proteome. (A) Comparison of total ion chromatogram (TIC) between two replicates of HeLa proteome. Mass resolution: 120K. Concentration: 1μg/μl.

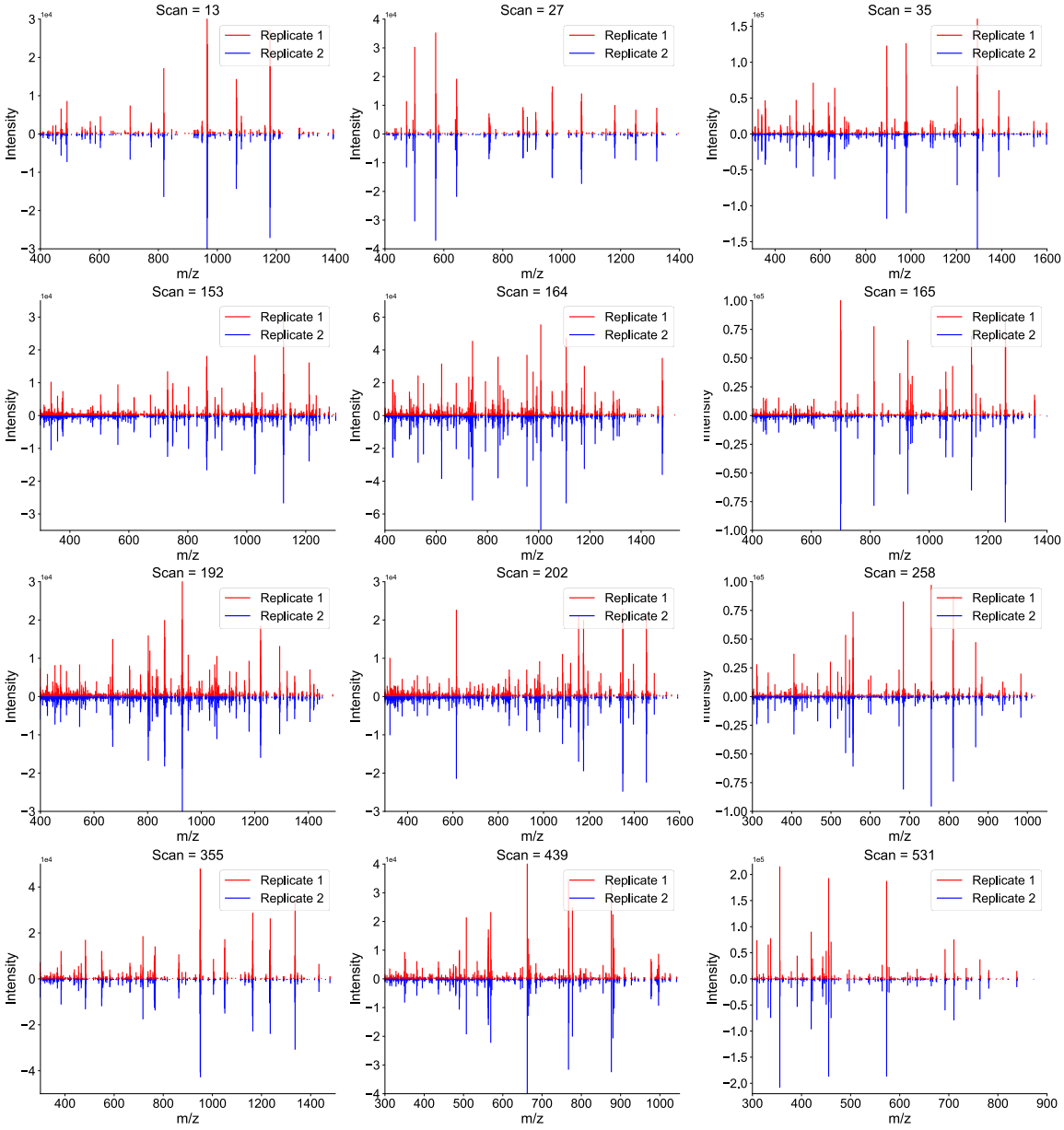


Figure S14. Comparison of mass spectrum of 12 randomly selected scans in two replicates of HeLa proteome. Mass resolution: 120K. Concentration: 1μg/μl.

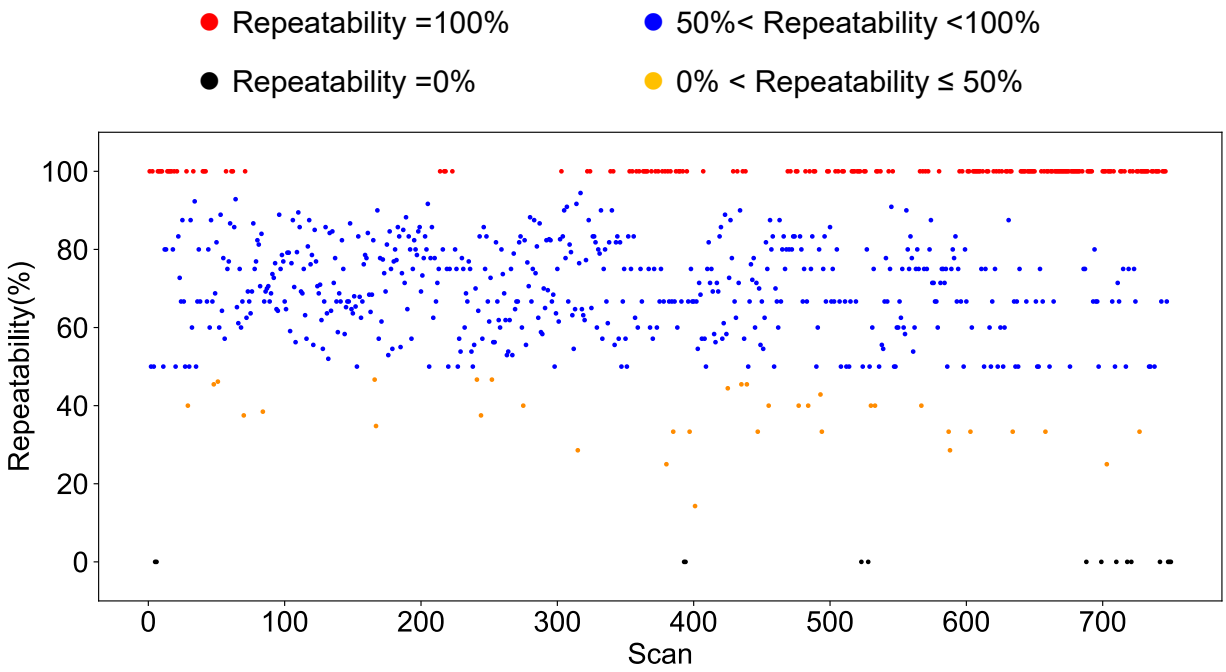


Figure S15. Repeatability assessment of all scans between two replicates of HeLa proteome. Repeatability is calculated by shared peptides/sum of peptides in every scan. Mass resolution: 240K. Concentration: 1μg/μl.

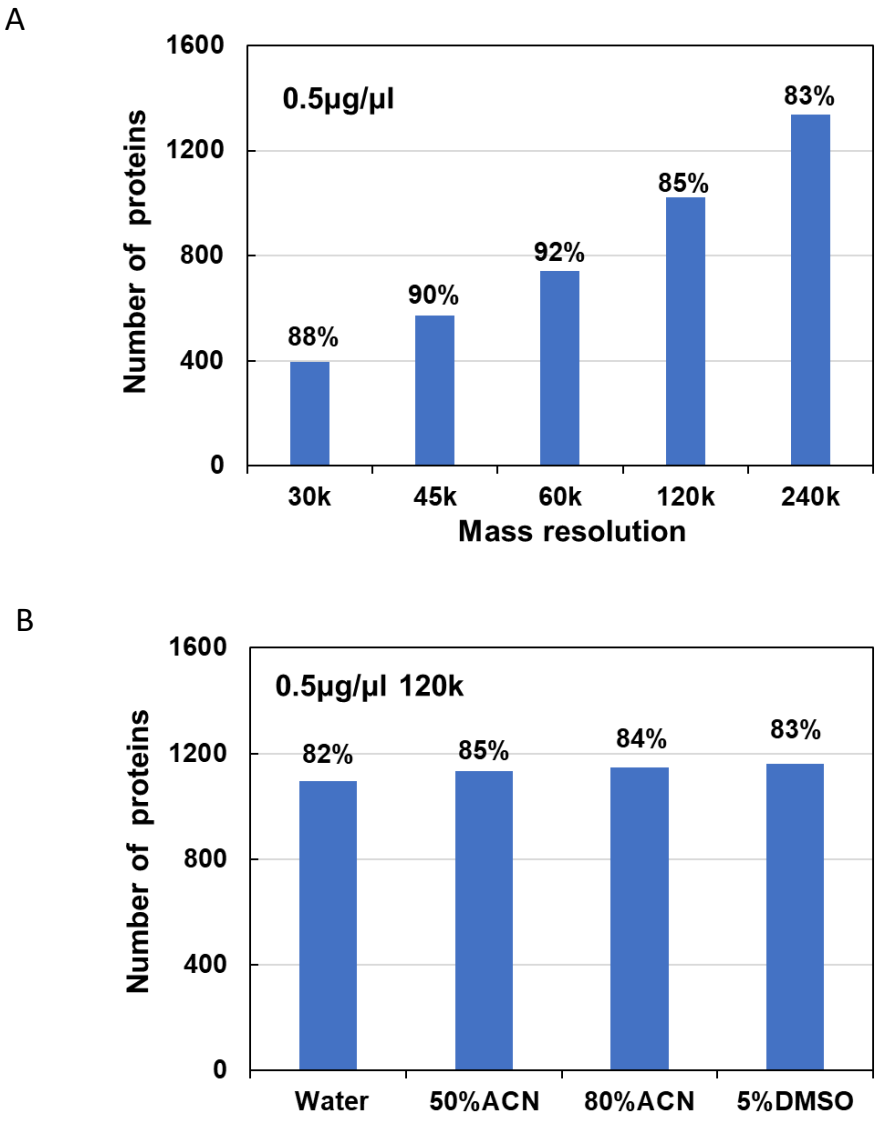


Figure S16. Repeatedly detected proteins in three replicates of different mass resolution (A) and solvents (B) in HeLa proteome. Percentages were calculated from repeated proteins detected in all three replicates divided by the average protein identifications of these three replicates.

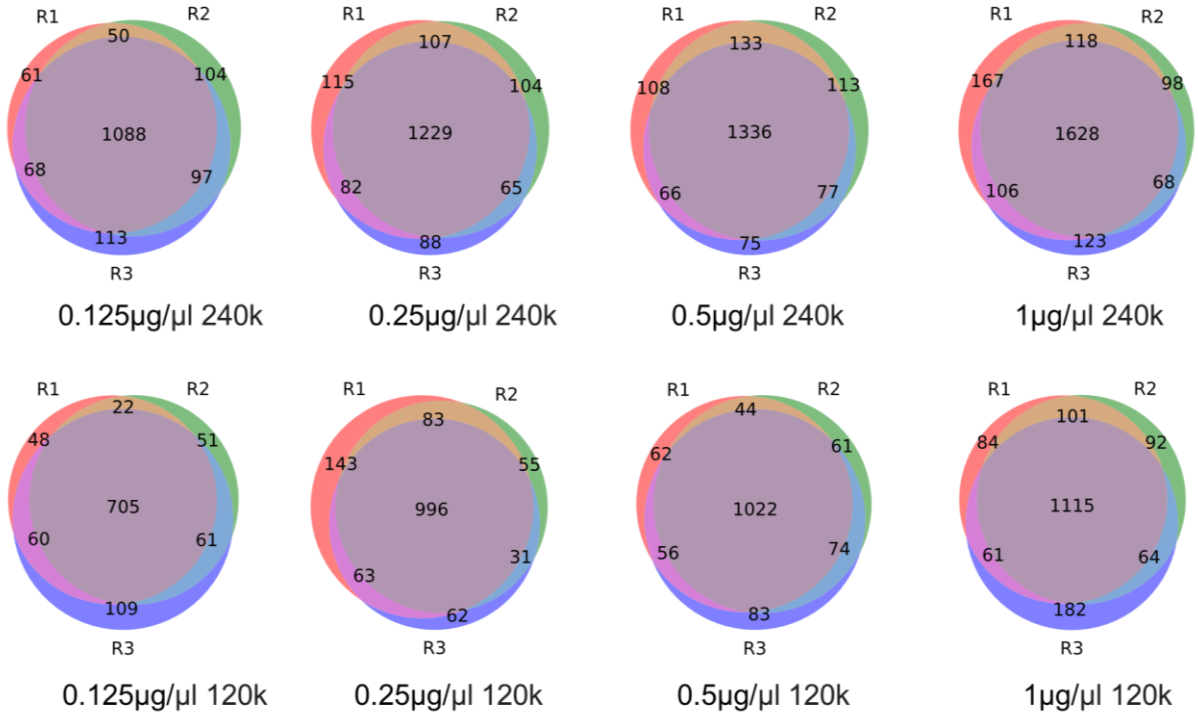


Figure S17. Venn diagram of protein identifications from three replicates for different experiment conditions. Sample: HeLa proteome.

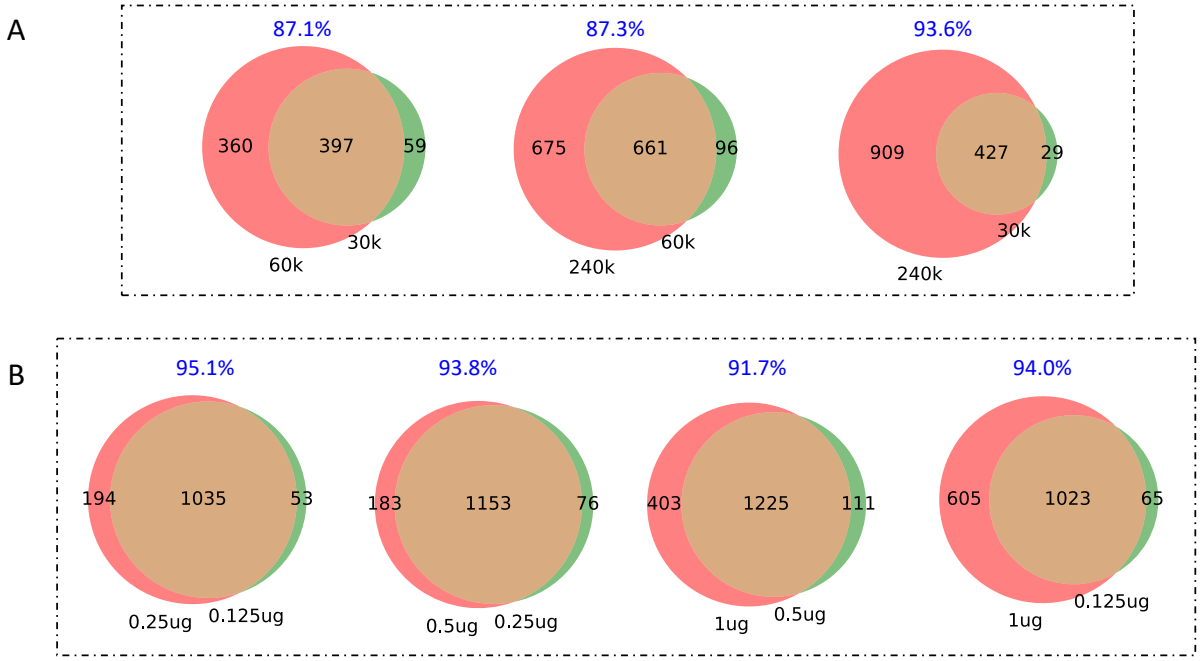


Figure S18. Comparative analysis of protein ID overlap across different mass resolution and concentration. (A) Venn diagram show duplicate proteins between different mass resolutions. Sample concentration: 0.5µg/µl. (B) Venn diagram show duplicate proteins between different concentrations. Mass resolution: 240K. Sample: HeLa proteome. Percentage was calculated by # Duplicate proteins / # dataset with less proteins.

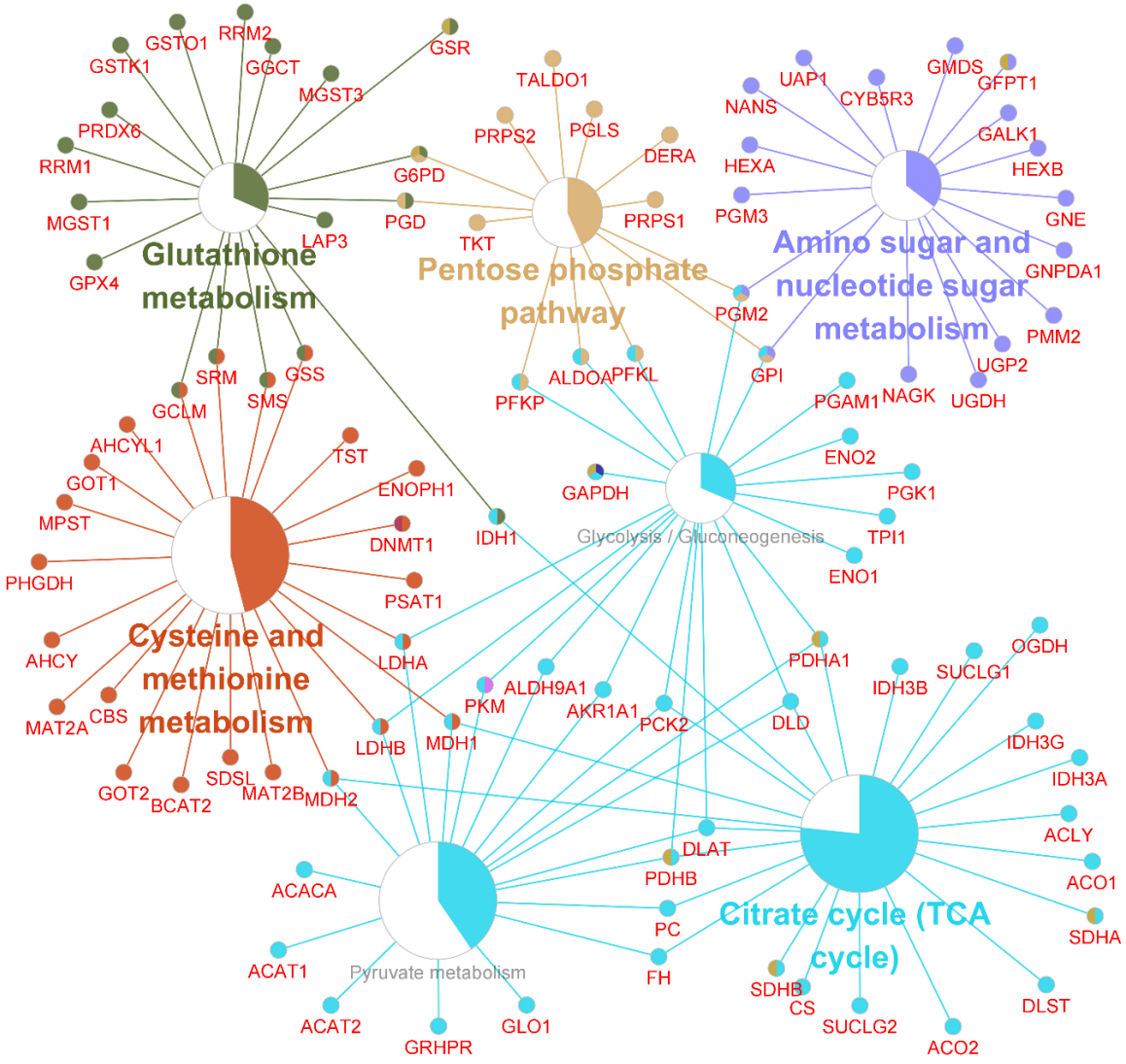


Figure S19. Pathway enrichment analysis showing proteins identified in pathways related to central carbon metabolism.

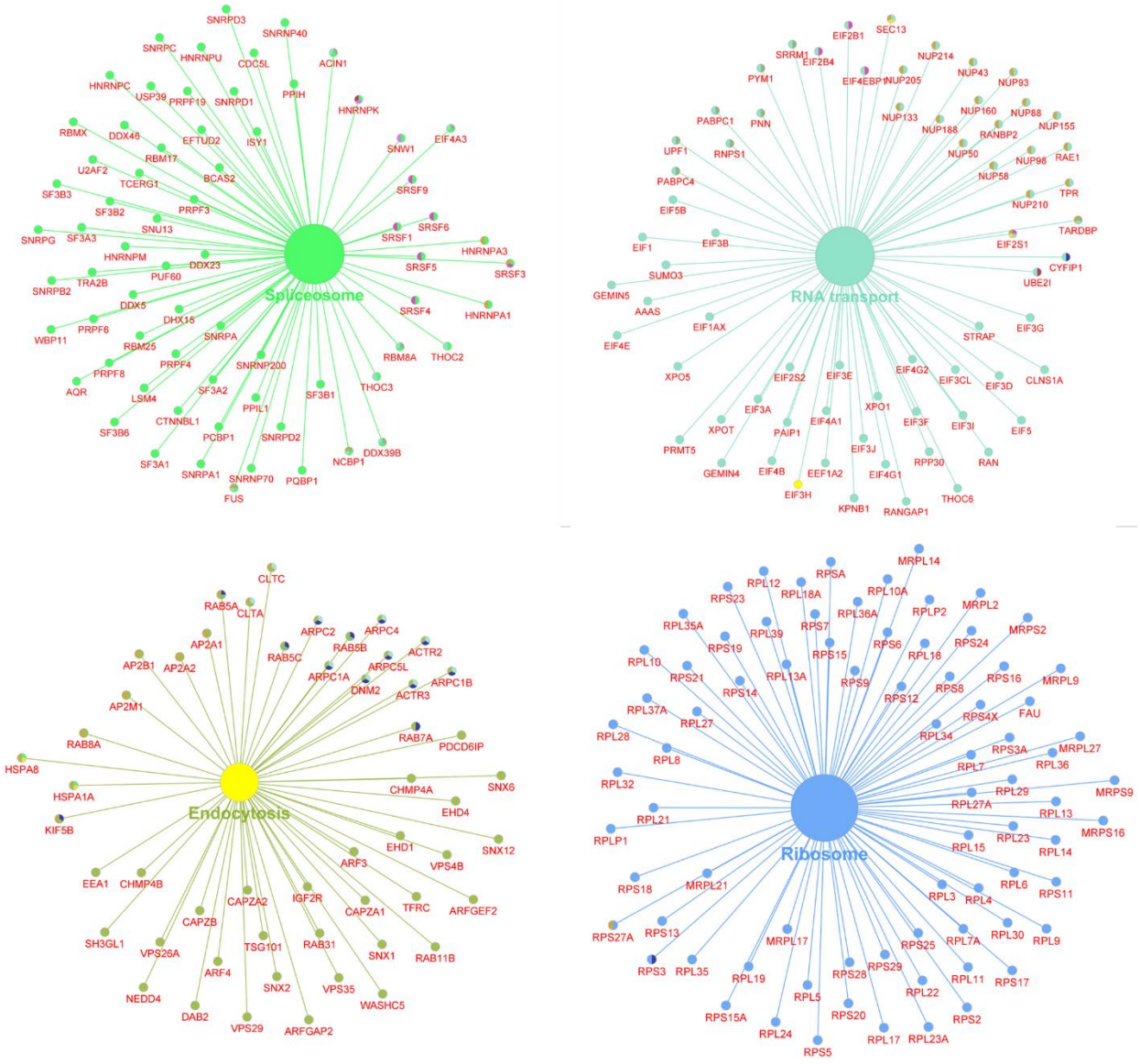


Figure S20. Pathway enrichment analysis showing proteins found in the following pathways: Spliceosome, RNA transport, Endocytois and Ribosome.

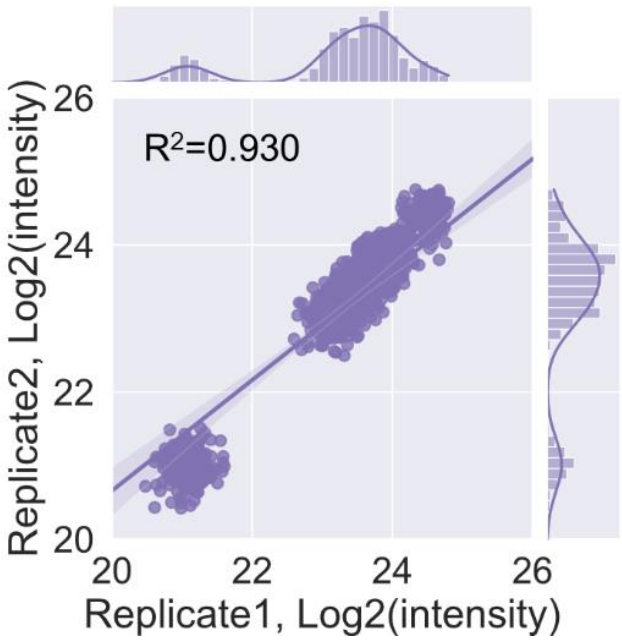


Figure S21. Scatter plot of two runs showing good reproducibility of peptide fragment intensity of angiotensin.

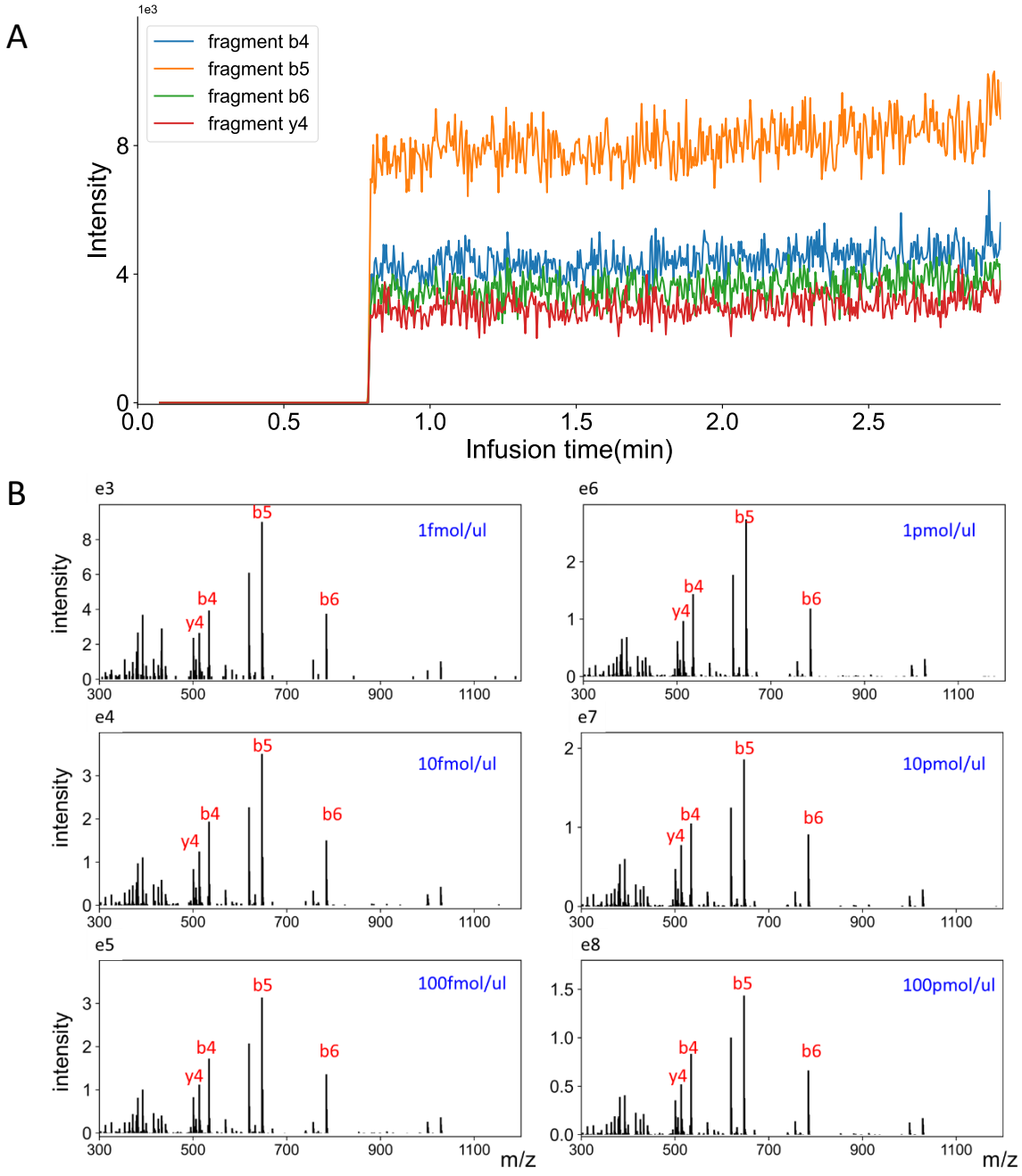


Figure S22. Label-free quantification of Angiotensin I from fragment ions. (A) illustration of TIC from four typical fragments from Angiotensin I. (B) MS/MS spectra of Angiotensin I with different concentrations.

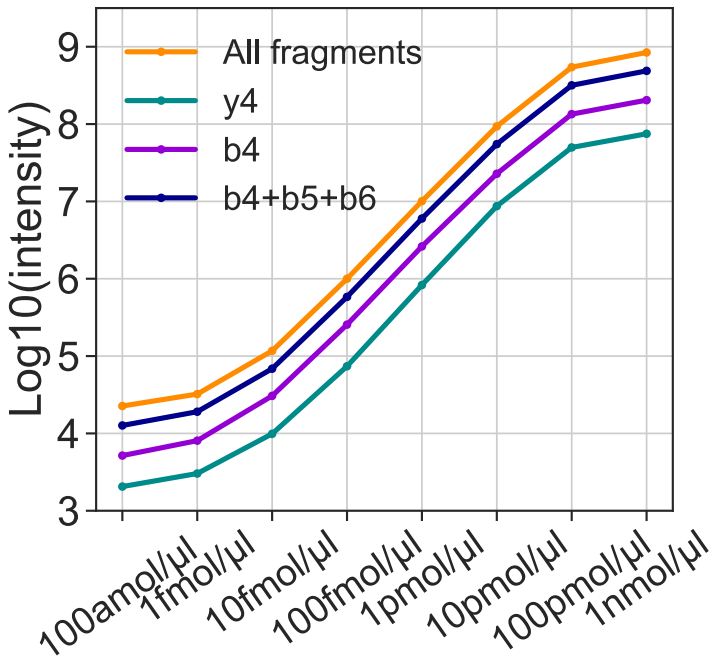


Figure S23. Quantitative curves of different fragment ions from Angiotensin I.

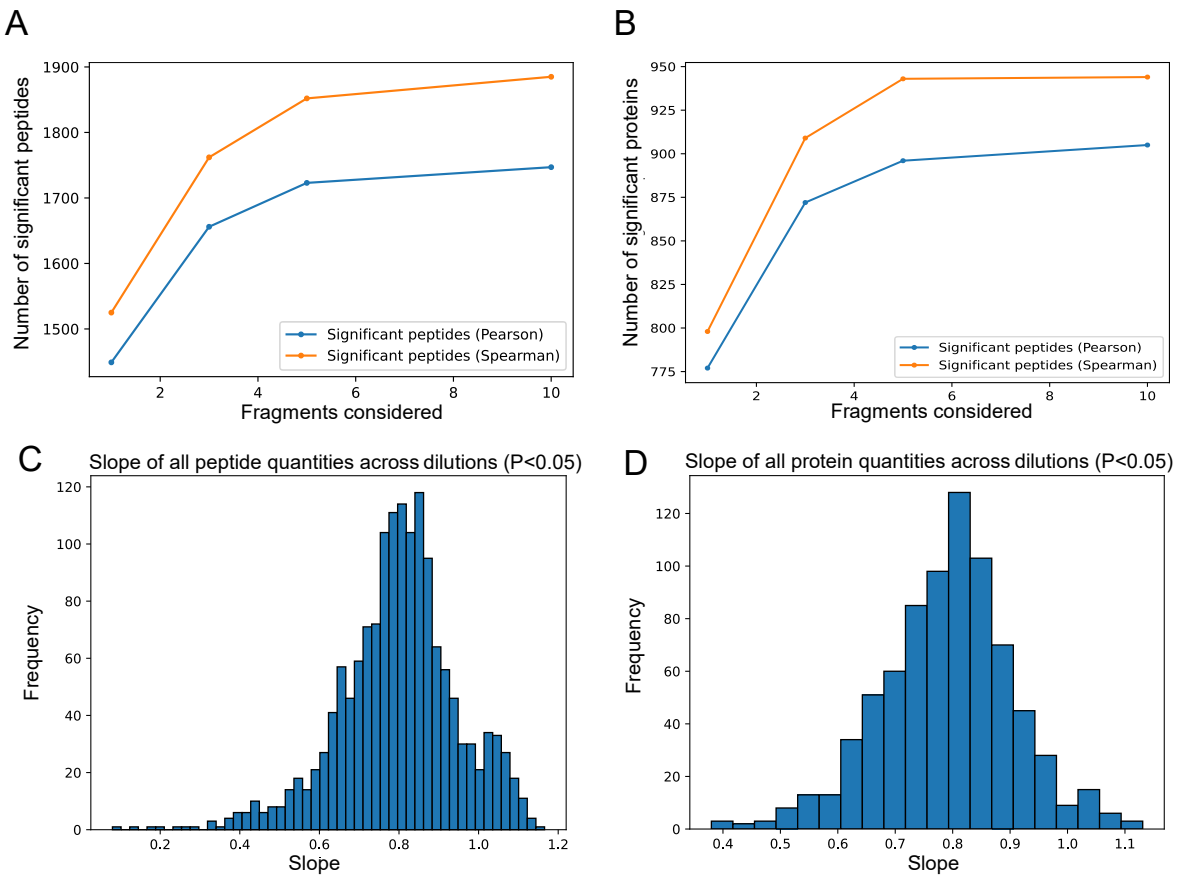


Figure S24. Quantification assessment of common peptides and proteins in HeLa proteome. (A) Number of significant peptides with different quantity of fragments used for quantification using a cutoff of p-value< 0.05. (B) Number of significant proteins with different quantity of fragments used for quantification using a cutoff of p-value< 0.05. (C) Diagram of slope of significant peptide quantities with positive slope across dilutions. This plot excludes three peptides with negative slopes. (D) Diagram of slope of all protein quantities across dilutions.

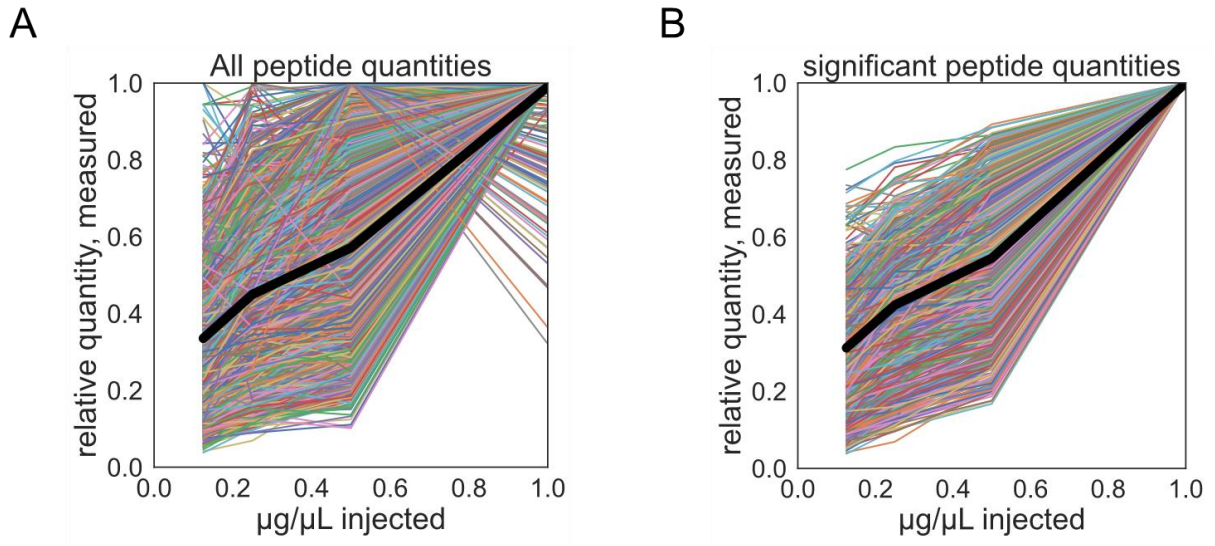


Figure S25. Common peptides from all runs in HeLa proteome quantification analysis. (A) Quantities of all peptides with positive slope that were identified in all 12 experiments of the HeLa dilution. (B) 1799 peptides that had a positive slope and significant correlation according to Pearson p value < 0.05. Sixty peptides with negative slope were excluded from A, and three peptides with negative slope were excluded from B.

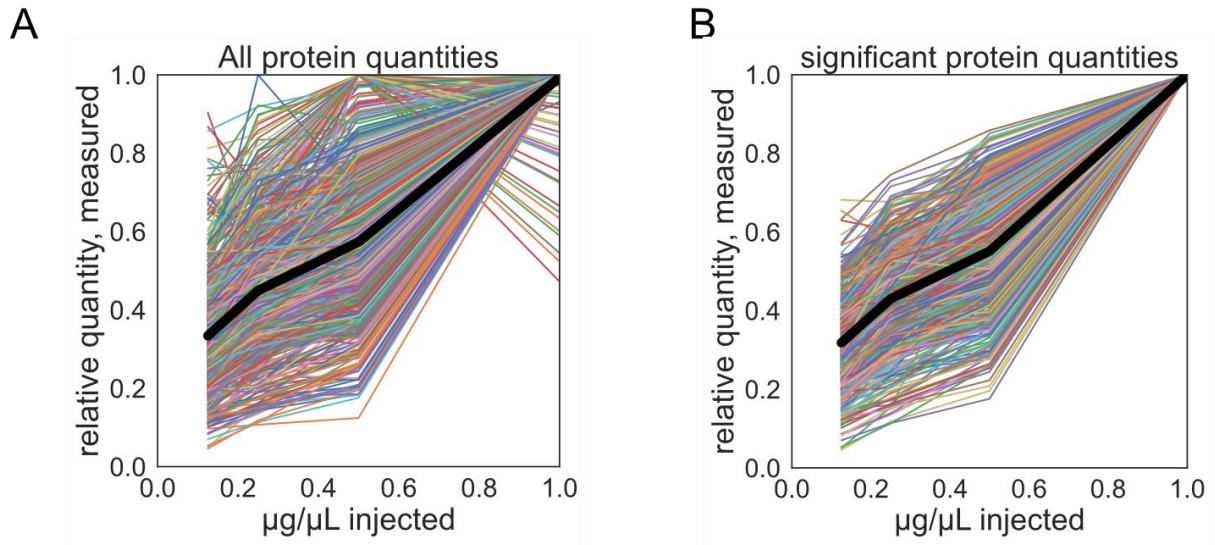


Figure S26. Common Proteins from all runs in HeLa proteome quantification analysis. (A) All protein quantities with positive slope that were identified in all 12 experiments of the HeLa dilution. (B) 905 proteins that had a positive slope and significant correlation according to Pearson p value < 0.05. Eighteen proteins with negative slope were excluded from A, and zero proteins with negative slope were excluded from B.

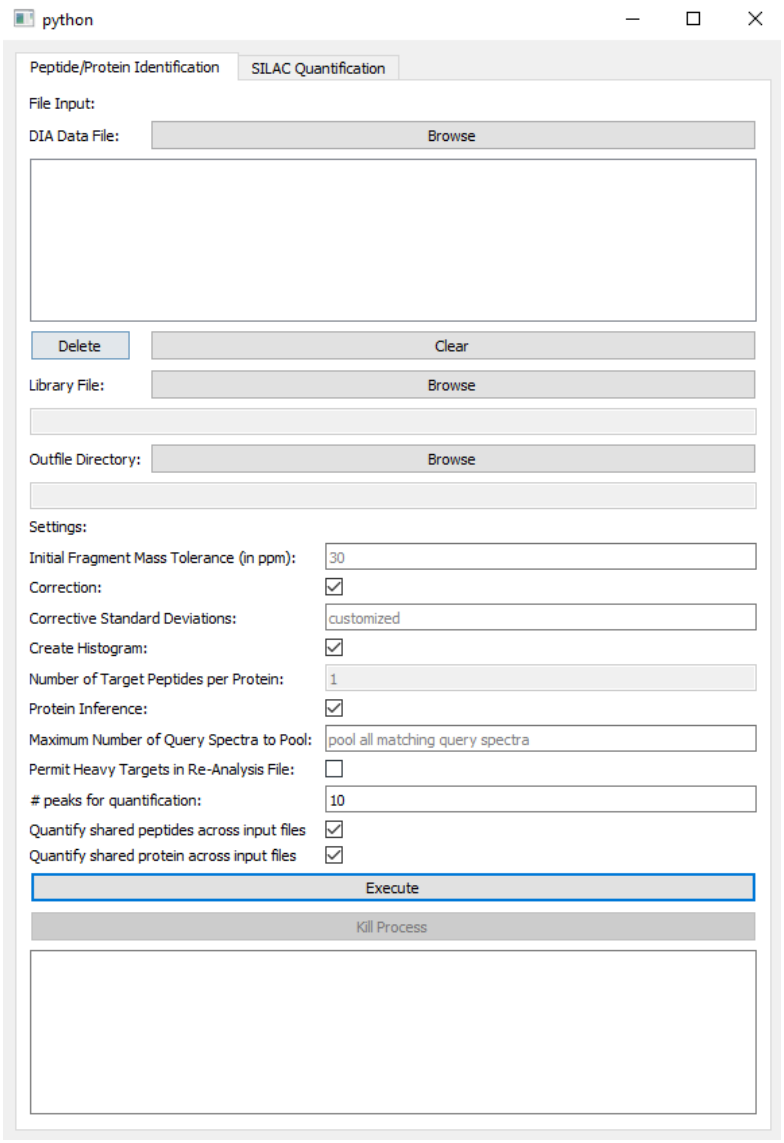


Figure S27. Screenshot of new GUI of CsoDIAq software with LFQ function implemented: "Quantify shared peptides across input files" and "Quantify shared protein across input files".