Expanded View Figures

Figure EV1. Technical feasibility assessment of ultra-high sensitivity mass spectrometry and liquid chromatography.

A   Ranked protein identifications for six-cell measurements with and without matching between runs. Zero-cell protein identifications are highlighted in orange and overlaid on the six-cell protein rank plot. The top 10 protein identifications of the zero-cell runs are depicted.

B   True nanoflow at 25, 50, and 100 nl/min flow rate on the EvoSep One liquid chromatography system.

C   Standardized 100 nl/min true nanoflow gradient on the EvoSep One liquid chromatography system. Pressure (Left) and flow profile (right) of the gradient of more than 1,000 consecutive runs (Day 1–Run #1 = gray; Day 20–Run #500 = orange; and Day 45–Run #1,000 = blue).

D   Data completeness (Blue) and coefficient of variation (Orange) evaluation of different diaPASEF consecutive scan repetitions merged for the analysis of 1 ng tryptic HeLa digest. Scans were varied from one, three, and five repetitions.
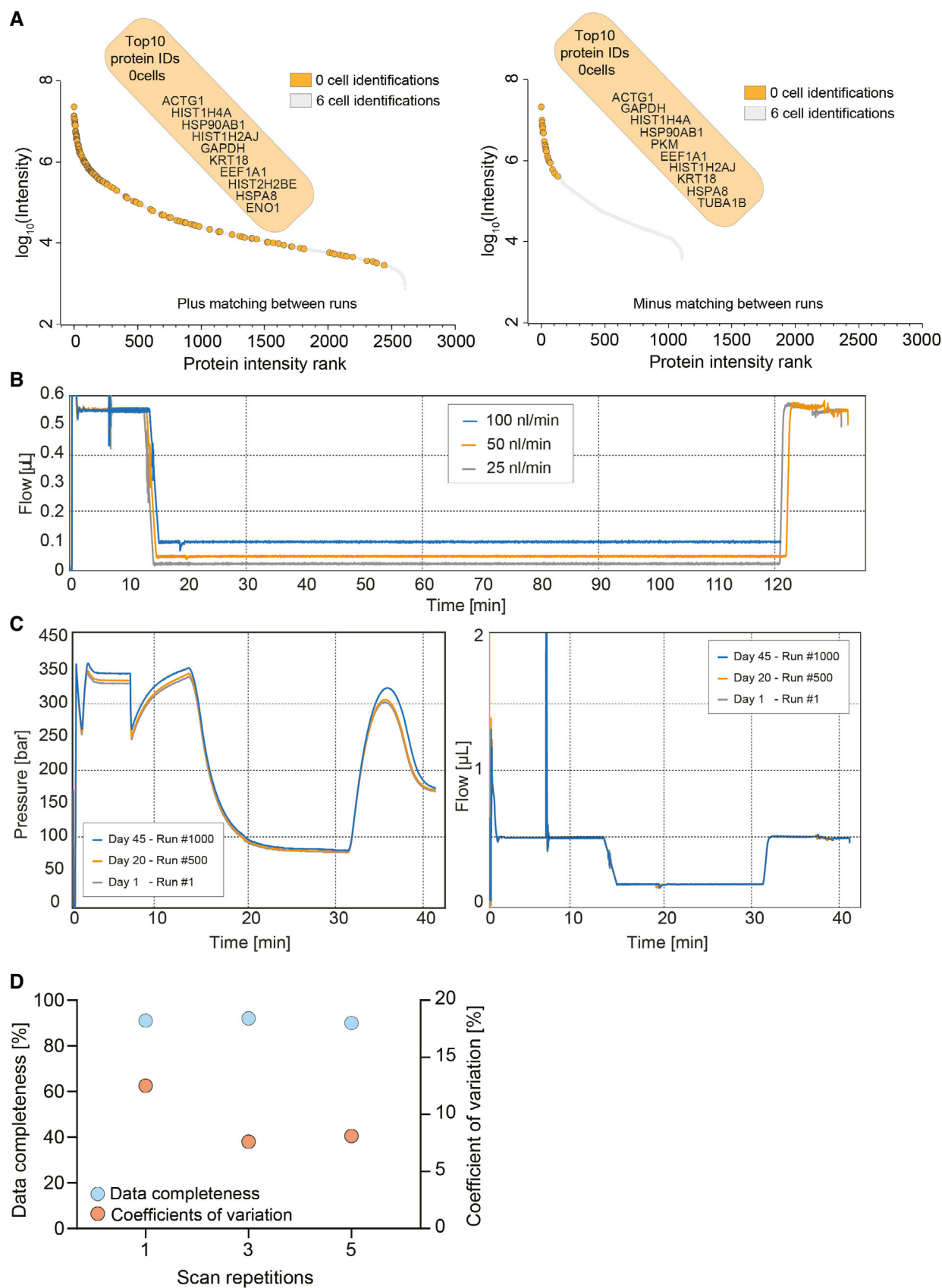
A

Top10
protein IDs
0cells

ACTG1
HIST1H4A
HSP90AB1
HIST1H2AJ
GAPDH
KRT18
EEF1A1
HIST2H2BE
HSPA8
ENO1

■ 0 cell identifications
□ 6 cell identifications

Plus matching between runs

Top10
protein IDs
0cells

ACTG1
GAPDH
HIST1H4A
HSP90AB1
PKM
EEF1A1
HIST1H2AJ
KRT18
HSPA8
TUBA1B

■ 0 cell identifications
□ 6 cell identifications

Minus matching between runs

B

— 100 nl/min
— 50 nl/min
— 25 nl/min

C

— Day 45 - Run #1000
— Day 20 - Run #500
— Day 1 - Run #1

D

○ Data completeness
○ Coefficients of variation

Figure EV1.

Figure EV2. True single-cell proteomics data set description.

A   Frequency plot for coefficient of variation occurrence within the 420 single-cell proteomics data set.

B   Protein $\log_{10}$ intensity versus coefficient of variation.

C   Raw $\log(x + 1)$-transformed intensity values of proteins per cell plotted against the number of identified proteins per cell (Left) and after normalization by local regression to cancel out those differences to enable downstream analysis (Right).

D   Principal component analysis of cell cycle stage enriched single-cell proteomics measurements and three cell culture batches projected on top.

E   Category count of gene ontology annotations for cellular compartment and biological process terms. Exemplary, category count terms are shown for the cellular compartment (Left) and biological process (Right) for more than 430 single-cell proteomics data set.

F   Cell cycle stage prediction for G2 versus G1 phase cells (Left) and G2/M versus G1 phase cells (Right) using the 60 topmost differentially expressed proteins reported by Geiger and coworkers (Aviner *et al*, 2015) as input.
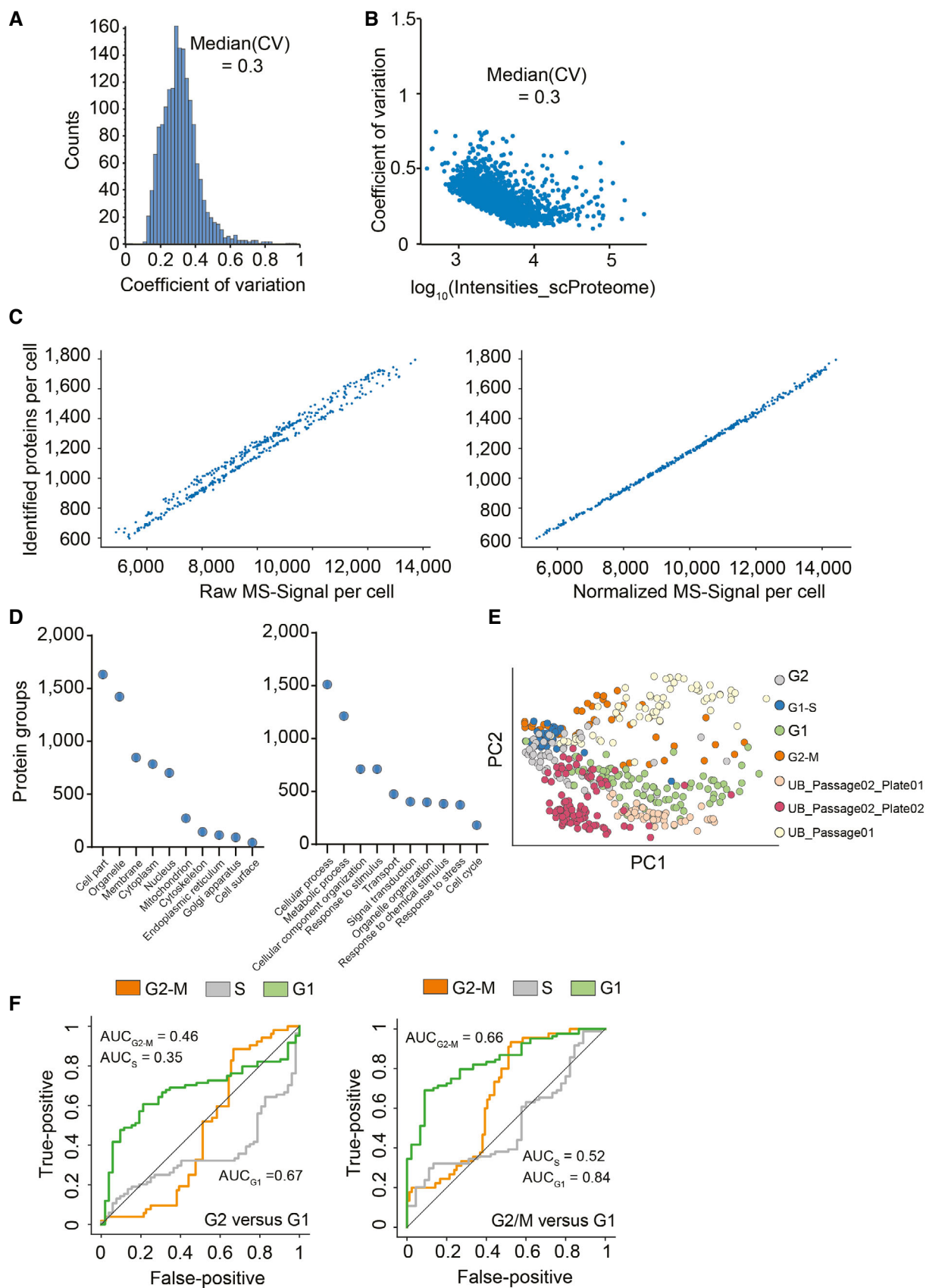
Figure EV2.

Figure EV3. Peptide fragment ion intensities of several proteins.

Fragment ion intensities of peptides for several differentially expressed proteins (HDAC2, FDR = 1.2E-3; UBE2C, FDR = 4.7E-6; UBE2S, FDR = 4.1E-15; TMSB10, FDR = 4.1E-15; BCCIP, FDR = 4.7E-1; HMGA1, FDR = 1.4E-2; NACA, and FDR = 4.1E-15) in the comparison of nocodazole- (G2-M transition) and thymidine (G1-S transition)-treated cells. Boxplots represent the intensity distribution of indicated peptide fragment ion intensities.
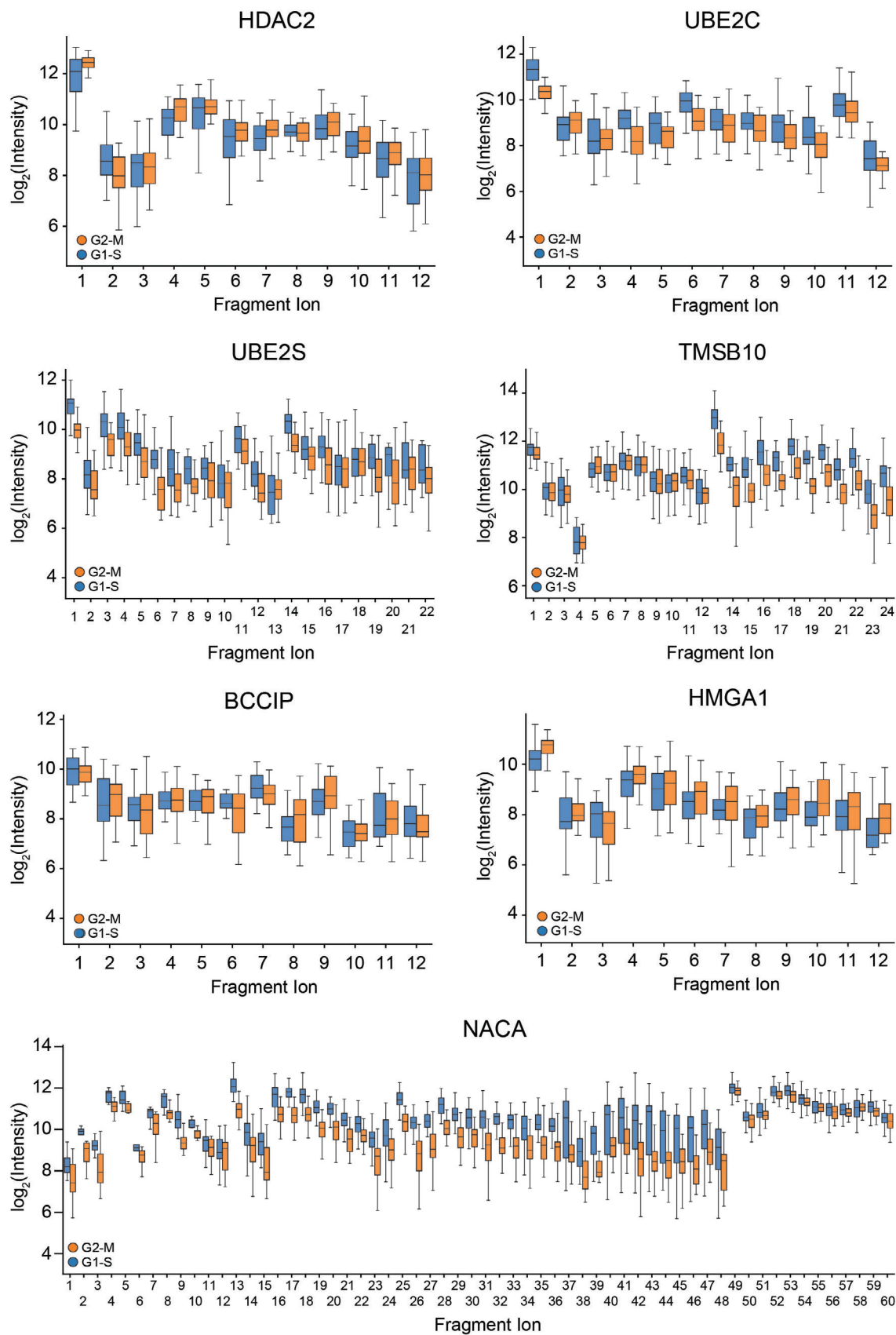
Figure EV3.

Figure EV4. Correlation and gene/protein completeness analysis of single-cell transcriptome sequencing and our LC–MS-based single-cell proteomics data set of the same cell line.

A Pearson correlation of observations for each cell within each of the technologies on all genes (MS-based proteomics, SMART-Seq2 RNA sequencing, and droplet-based RNA sequencing; Left) and for each cell within each of the technologies on shared genes between technologies (MS-based proteomics, SMART-Seq2 RNA sequencing, and droplet-based RNA sequencing; Right).

B Gene/Protein expression completeness per cell on all shared genes between the three technologies (scProteomics; SMART-seq2; and DROP-seq).

C Gene and protein expression completeness as a function of ranked genes/proteins for all three technologies (Proteomics, DROP-seq, and SMART-Seq2). Arrows indicate a bimodal distribution for single-cell RNAseq data in both technologies, which is absent in proteomics.

D Data completeness across single cells as a function of mean protein abundance for MS-based single-cell proteomics and both single-cell RNA sequencing (Drop-Seq, SMART-Seq2). Expected poison dropout distribution shown in red.

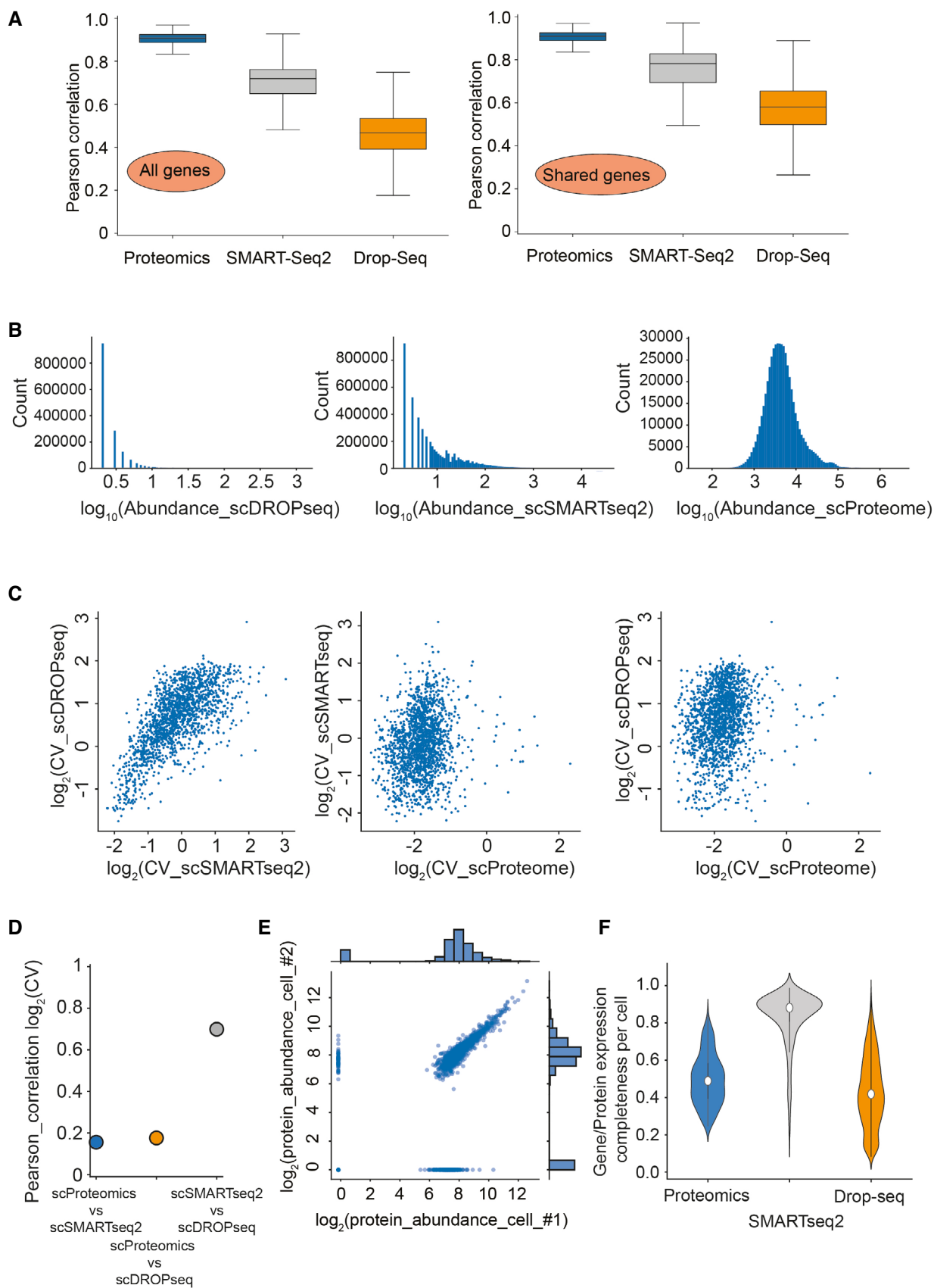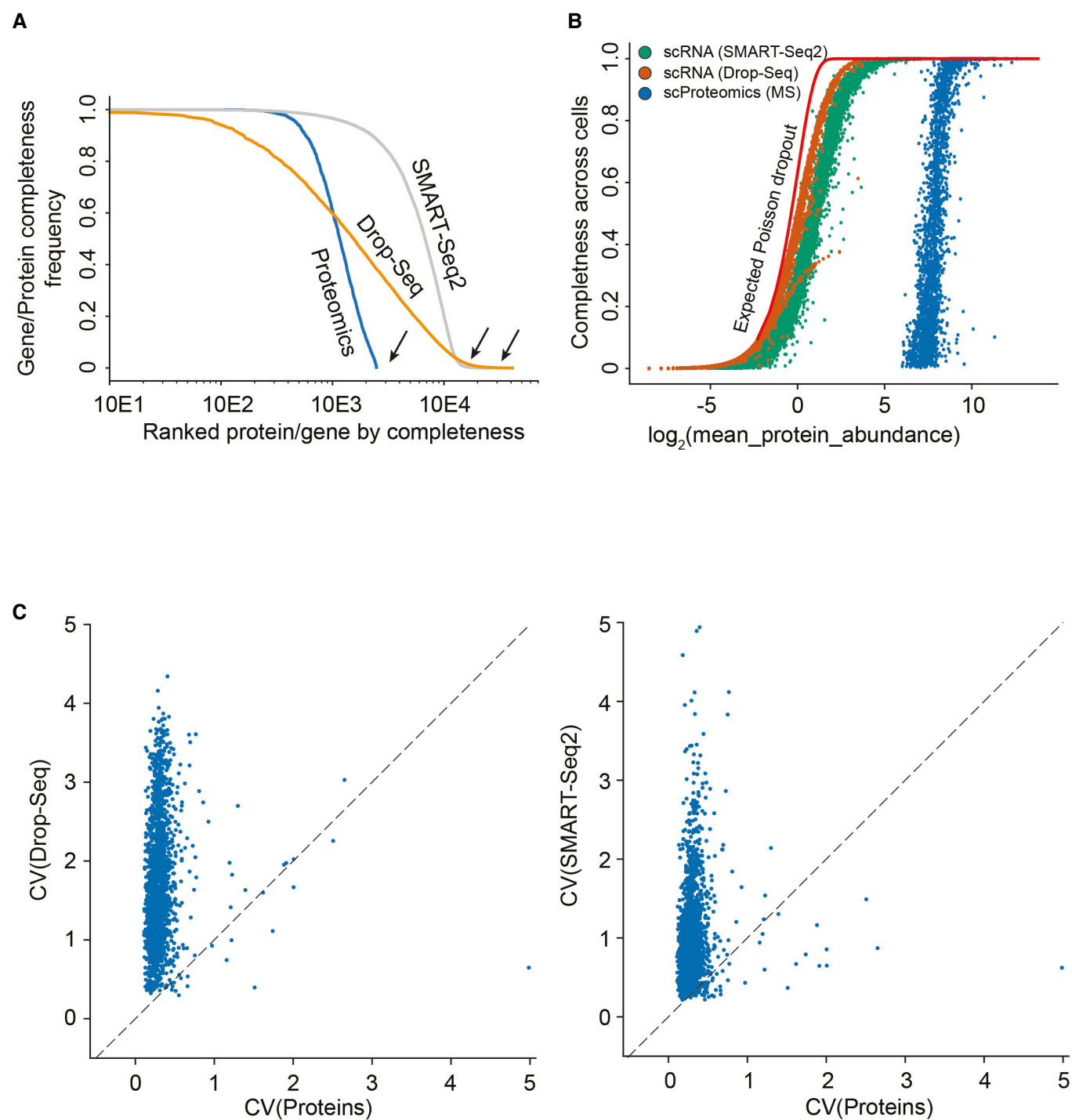E Scatter plot of two independently measured single-cell proteome expression values.
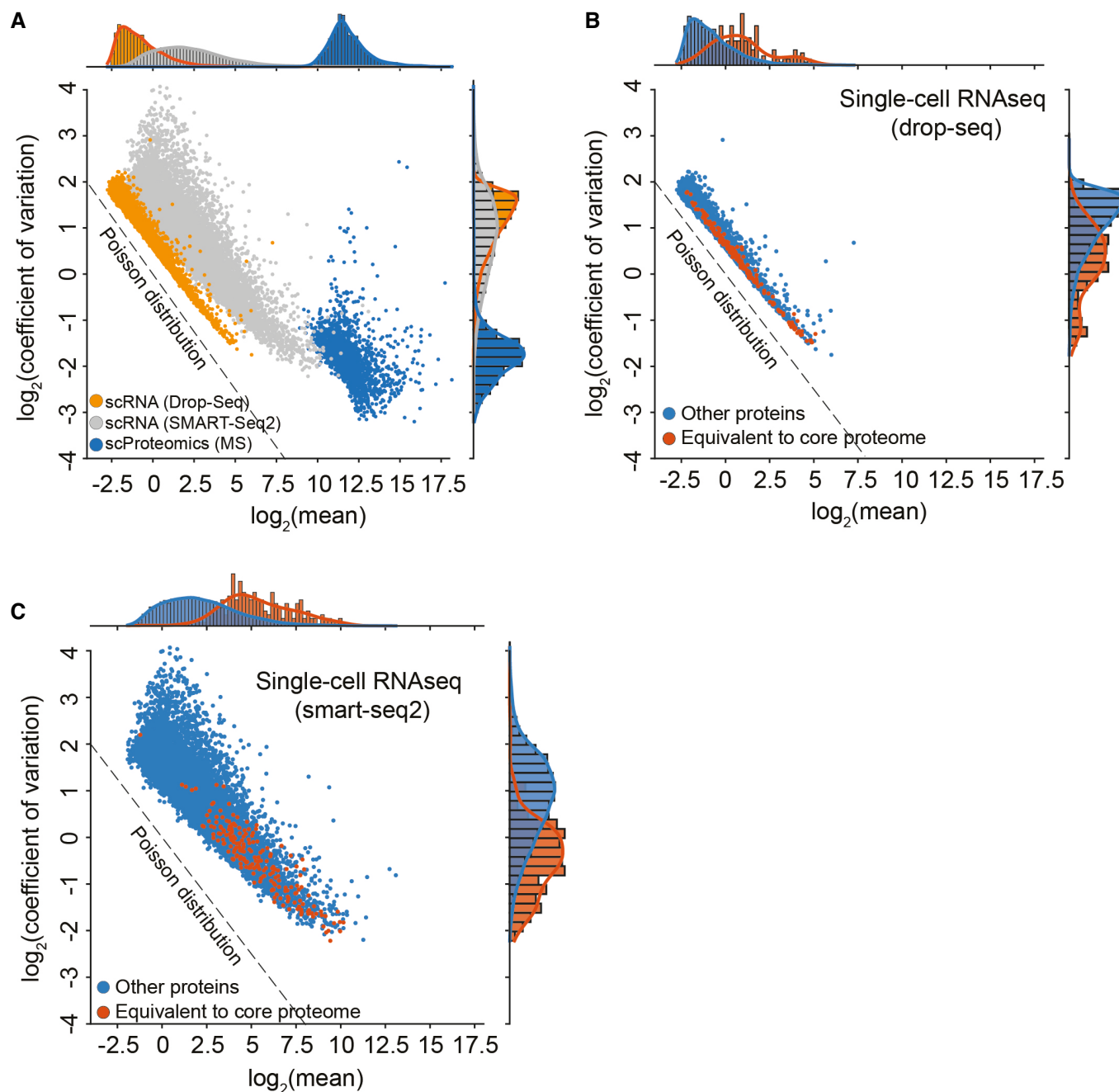
Figure EV4.

Figure EV5. Data distribution analysis of single-cell transcriptome sequencing and our LC–MS-based single-cell proteomics data set of the same cell line.

A   Histogram of $\log_{10}$ abundance of scDROPseq (left), scSMARTseq2 (middle), and scProteomics data (right).
B   The coefficient of variation of a gene measured by either Drop-Seq technology (Left) or SMART-Seq2 (Right) compared to the coefficient of variation of the corresponding protein measured by MS-based single-cell proteomics.
C   Pearson correlation of coefficients of variation for each gene shared within each comparison.

Figure EV6. Single-cell core proteome distribution compared to matched gene expression distribution in single-cell transcriptome data of the same cell line.

A Coefficient of variation distribution as a function of $\log_2$ mean gene or protein intensities for Drop-Seq (Orange), SMART-Seq2 (Gray), or MS-based single-cell proteomics (Blue). Expected Poisson distribution shown as dashed line.

B Coefficient of variation of single-cell RNA-sequencing (drop-seq) levels as a function of mean expression levels with the "core proteome" colored in orange and non-"core proteome" genes in blue. Expected Poisson distribution shown as dashed line.

C Coefficient of variation of single-cell RNA-sequencing (smart-seq2) levels as a function of mean expression levels with the "core proteome" colored in orange and non-"core proteome" genes in blue. Expected Poisson distribution shown as dashed line.
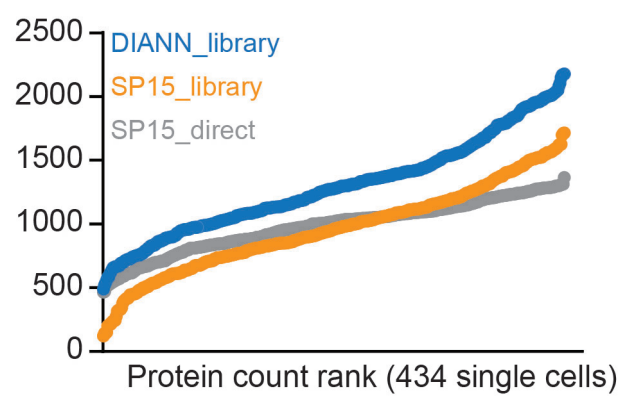
Figure EV7. Comparison of single-cell protein identifications across available DIA software solutions.

The 434 single-cell proteome data were processed with either DIA-NN or Spectronaut using a spectral library, or using Spectronaut in directDIA mode. Protein identifications were plotted as a function of the protein rank for each single cell.

Steps	Action	Total protein groups	Total cell count
1	Start	2501	434
2	At least 600 protein Ids	2501	421
3	At least 15% row completeness	1858	421
4	CV <= 0.75	1818	421

Figure EV8. Single-cell data filtering for processing.

The 434 single-cell proteome data were filtered first for at least 600 protein identifications, then for at least 15% data completeness across rows and finally for coefficients of variation of below 0.75 before the downstream processing and biological interpretation was performed.