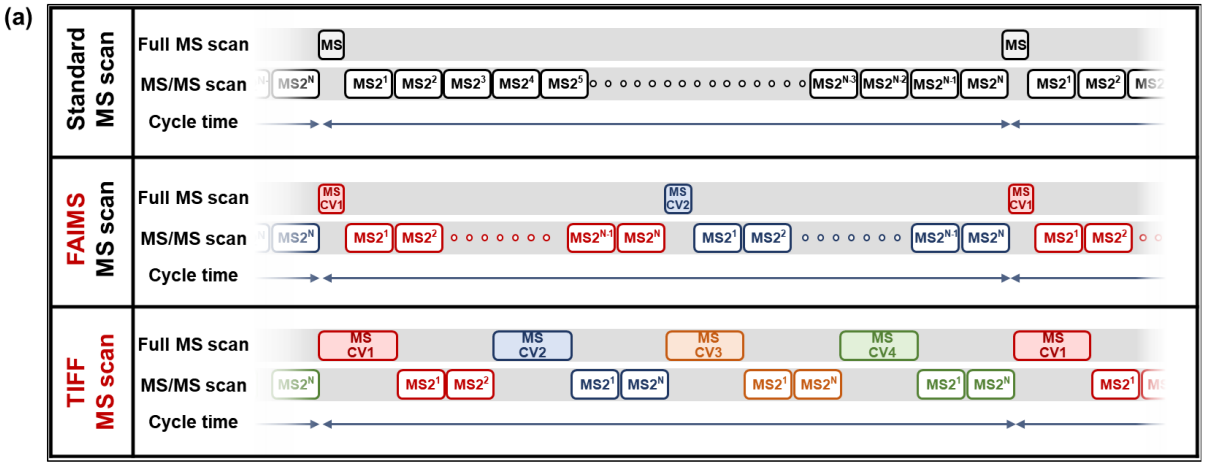
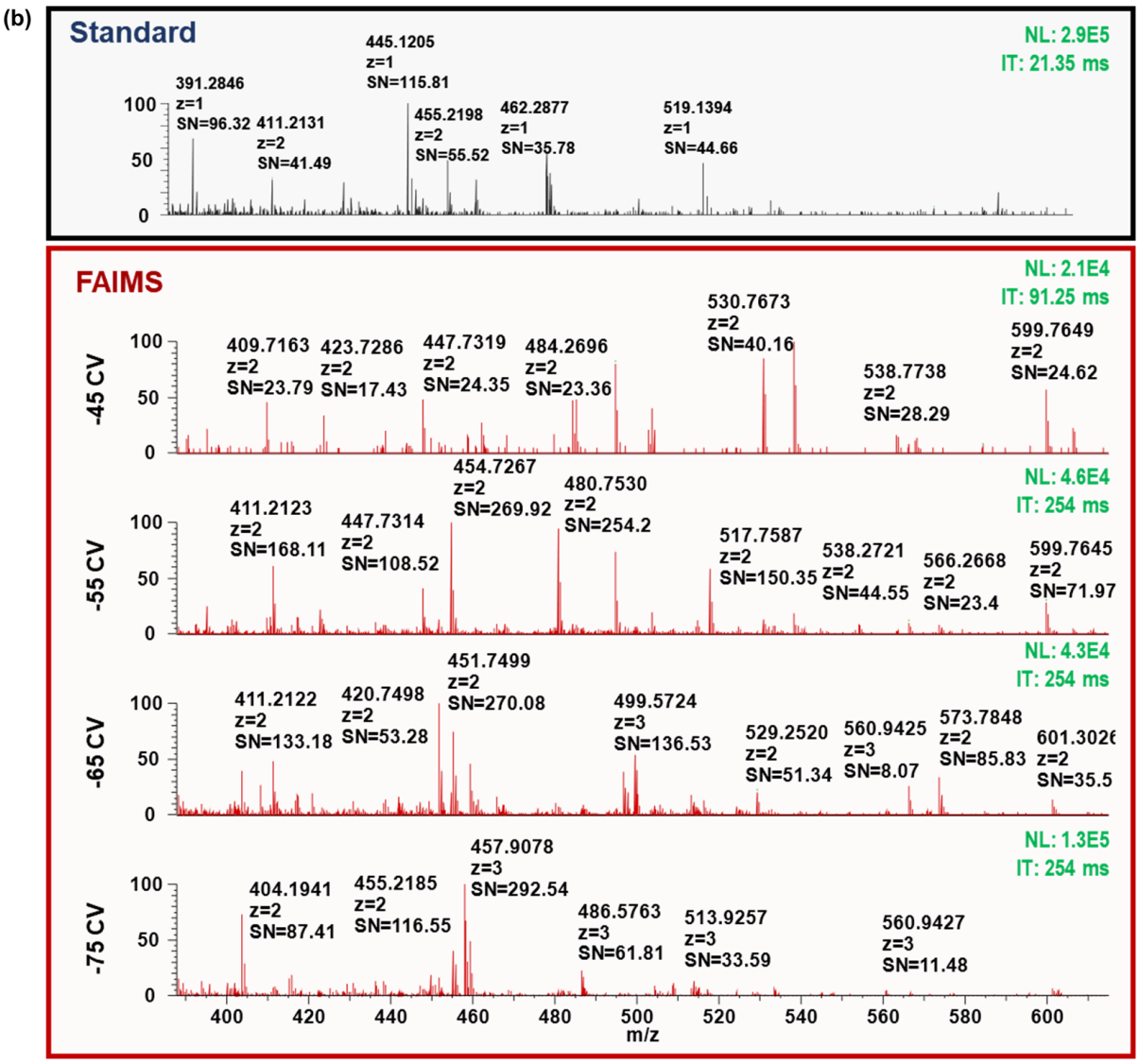
Supplementary Information for

Robust, sensitive, and quantitative single-cell proteomics based on ion mobility filtering

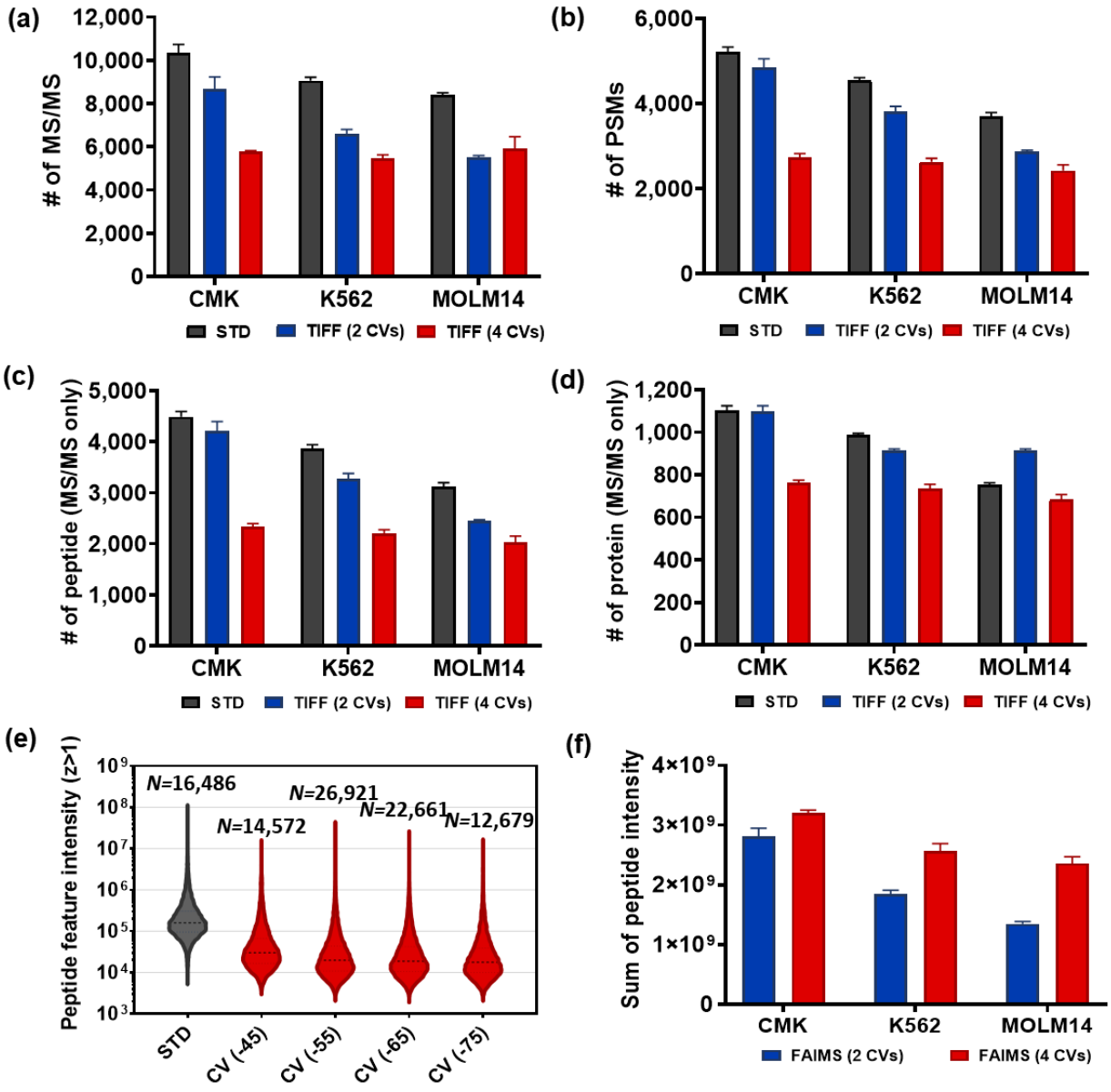
Woo et al.

Supplementary Materials



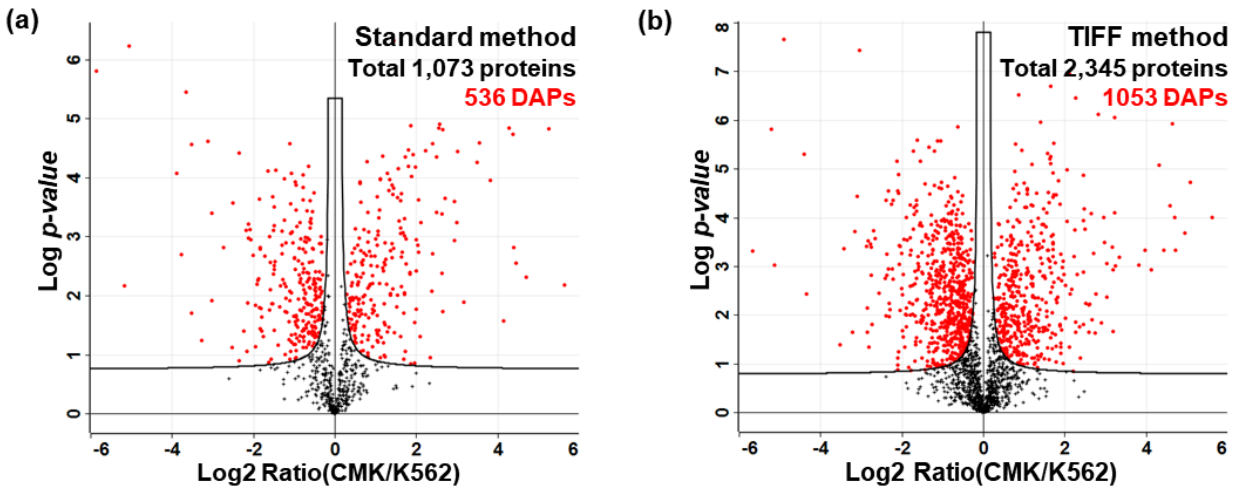


Supplementary Figure 1. (a) Schematic illustration of three different MS acquisition methods. (Upper) standard MS method without FAIMS; (Middle) standard FAIMS-MS method; (Bottom) The transferring identification based on FAIMS filtering (TIFF) method. In the TIFF method, the elongated ion accumulations for MS1 scan will increase the sensitivity of MS1-level peptide detection. The peptide features are identified by matching to a spectral library based on 3D tags (LC retention time, accurate m/z, and FAIMS CV). Small number of MS/MS scans are used for non-linear alignment during MaxQuant search. **(b)** Representative spectra are chosen from the RAW files of a standard method (in the blue box) and a FAIMS method with 4 CVs (in the red box). The spectra are extracted from a similar retention time. Spectra are labeled with m/z, ion charge state, and signal to noise (SN) value.

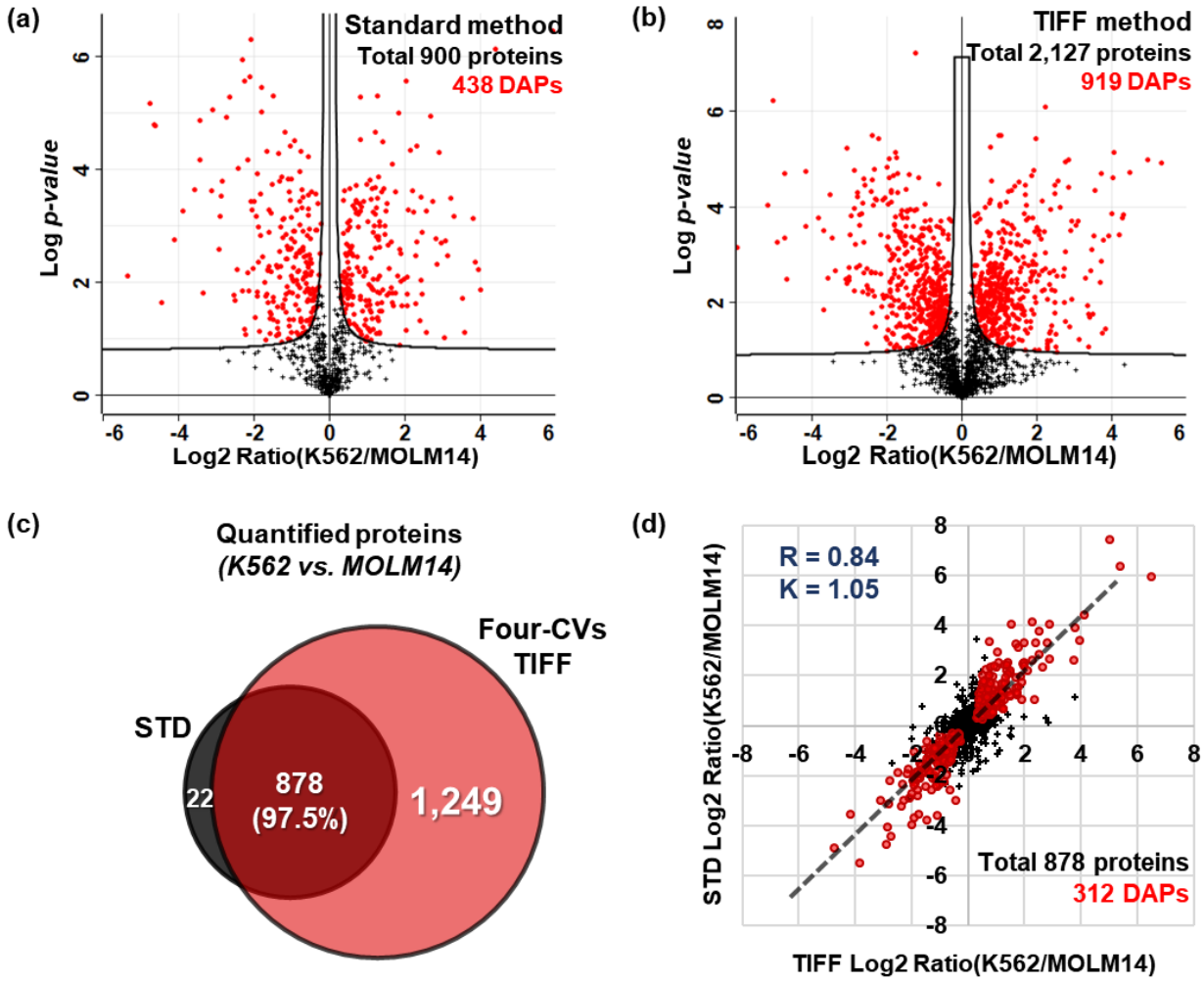


Supplementary Figure 2. (a-d) Benchmarking of the standard, 2-CV-TIFF, and 4-CV TIFF methods using single-cell level peptides (0.2 ng) from three cell lines (CMK, K562, and MOLN14). **(a)** The numbers of MS/MS events, **(b)** peptide spectrum matches (PSM), **(c)** unique peptides and **(d)** proteins identified by MS/MS. **(e)** Intensity distributions of peptide features (z > +1) obtained by the standard and 4-CV-TIFF methods using 0.2-ng CMK peptides. Labeled numbers indicate the numbers of detected peptide features. An in-house MASIC tool was used to select the peptide features from MSGF+ results. **(f)** The

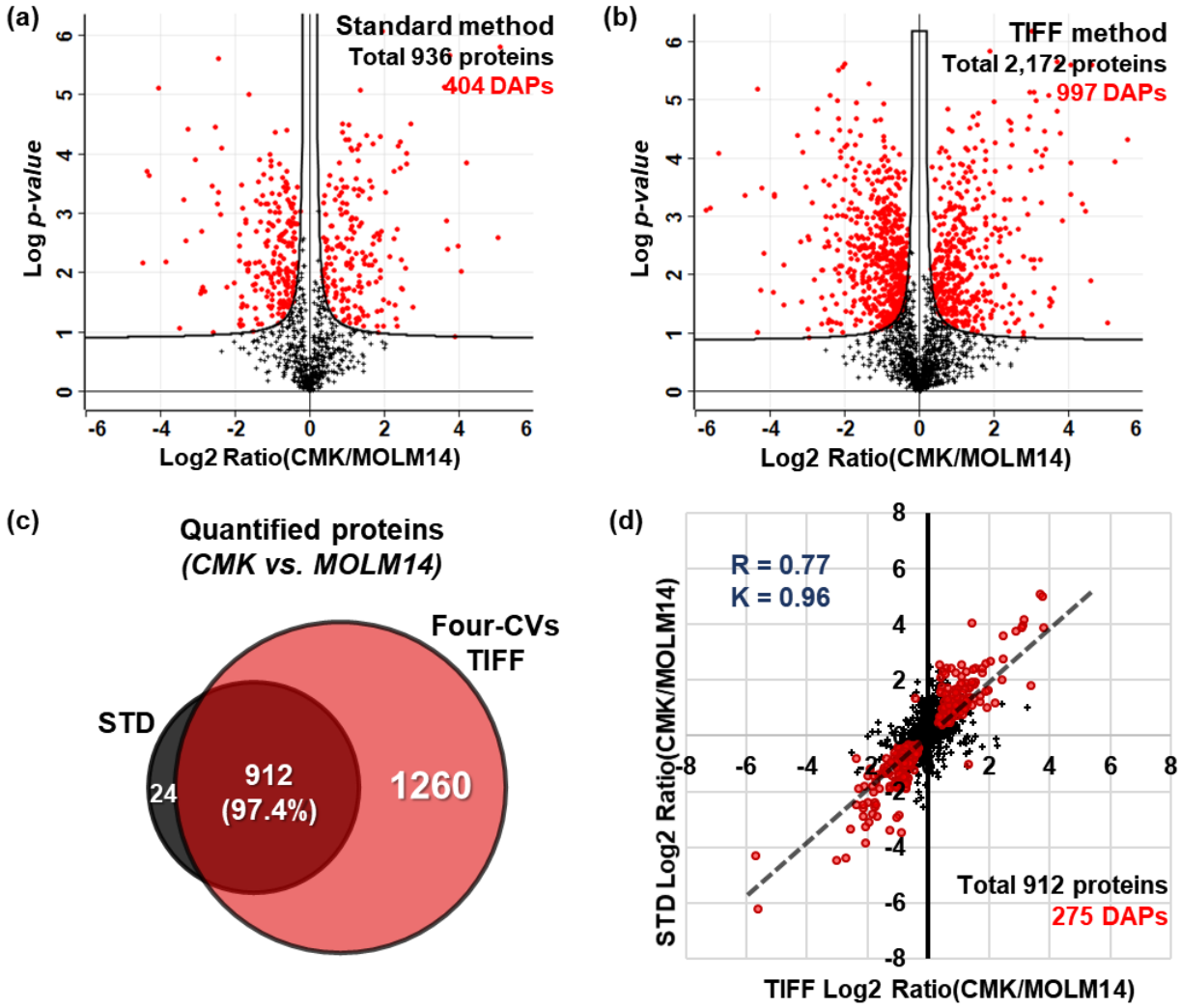
summed peptide intensities from the 2-CV and 4-CV TIFF methods. All the error bar on the graph include triplicate of the sample.



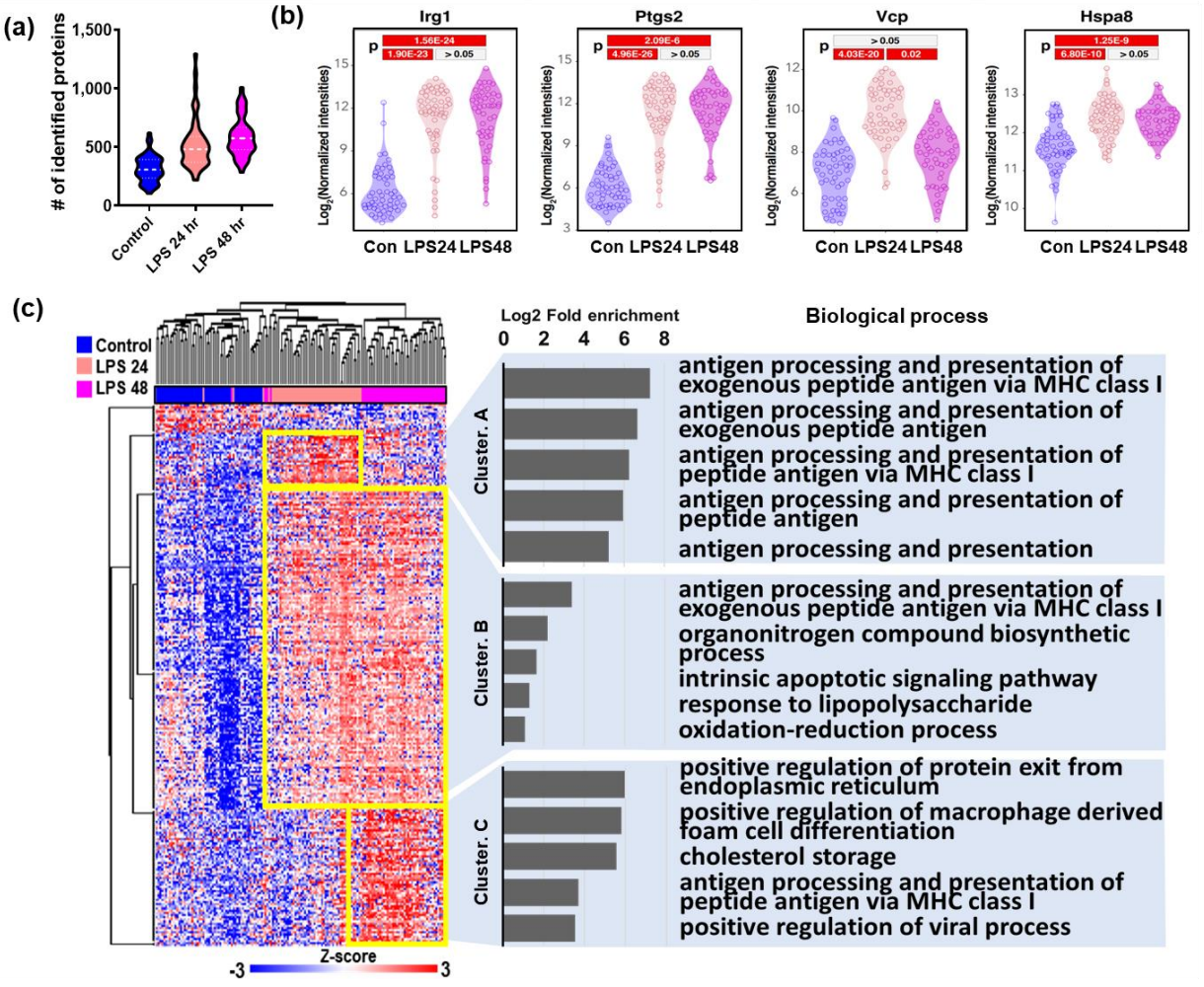
Supplementary Figure 3. (a-b) Statistics analysis to identify differentially abundant proteins (DAPs) between CMK and K562 cells using iBAQ intensities (t-test FDR < 0.05 and S_0 = 0.1). Volcano plots for **(a)** standard method and **(b)** the 4-CV TIFF method. Total quantified proteins and DAPs were labeled with red color.



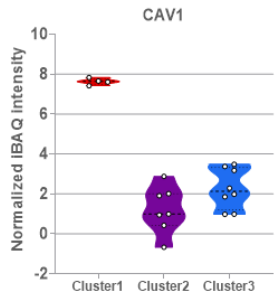
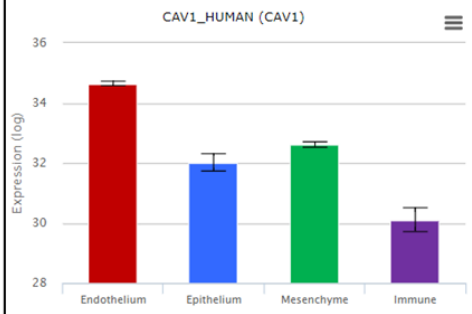
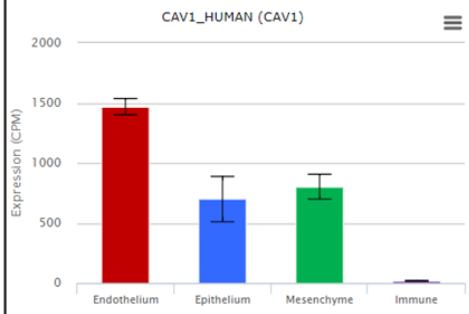
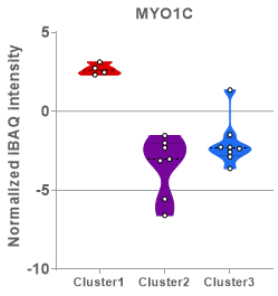
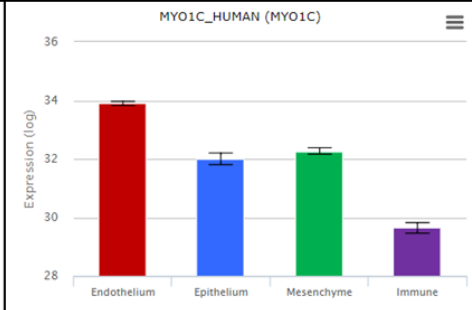
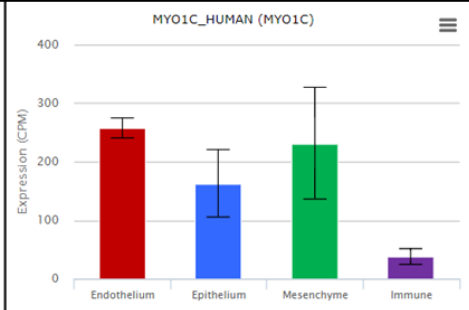
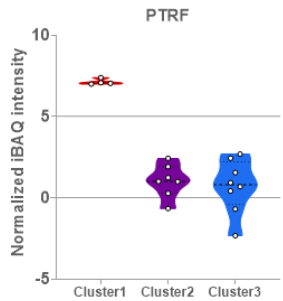
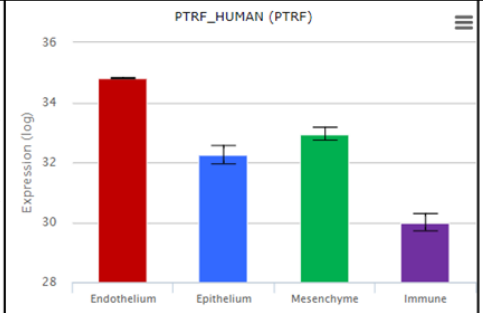
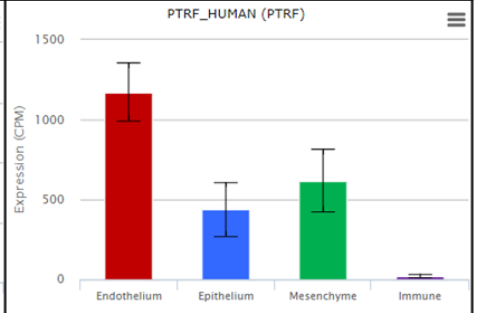
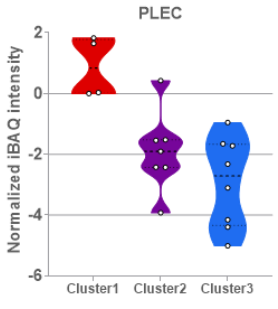
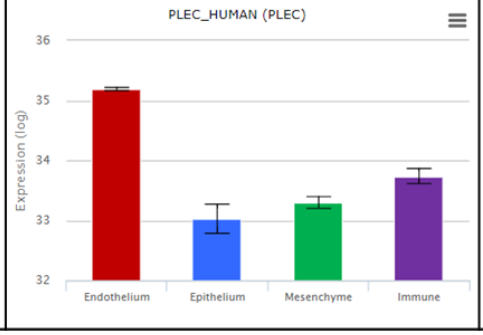
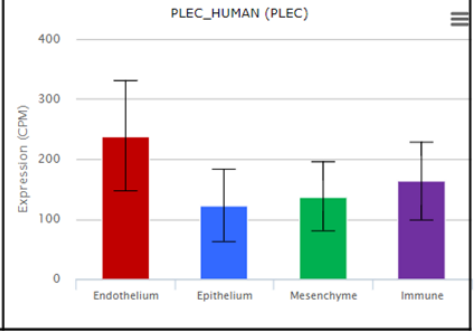
Supplementary Figure 4. (**a-b)** Statistics analysis to identify differentially abundant proteins (DAPs) between K562 and MOLM14 cells (t-test FDR < 0.05 and S_0 = 0.1). Volcano plots for **(a)** the standard and **(b)** 4-CV TIFF methods. **(c)** Overlap of quantifiable proteins between K562 and MOLM14 cells measured by standard and TIFF methods (4 CVs). **(d)** The linear correlation and slope of log2 transformed fold changes of K562 and MOLM14 proteins between the 4-CV TIFF and STD methods. Red dots indicate DAPs in both methods calculated by t-test (FDR<0.05, S_0=0.1).

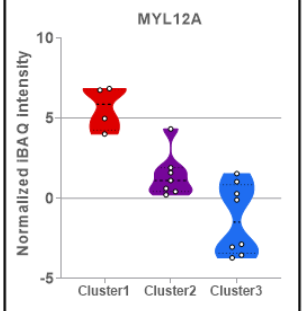
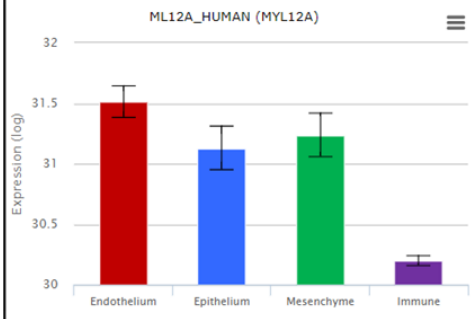
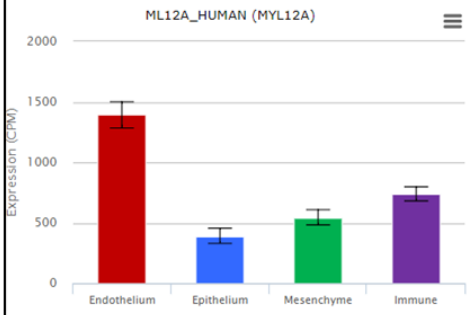
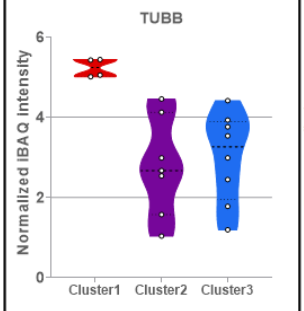
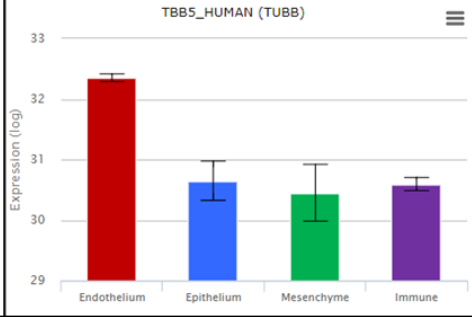
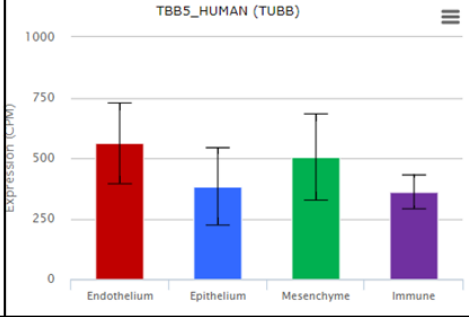
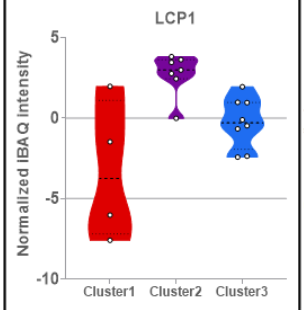
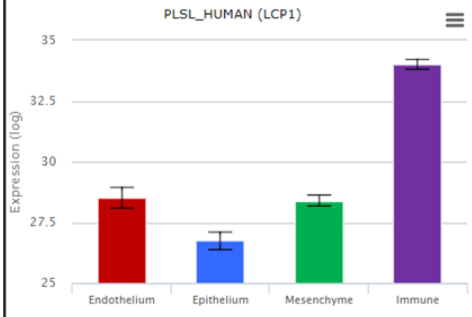
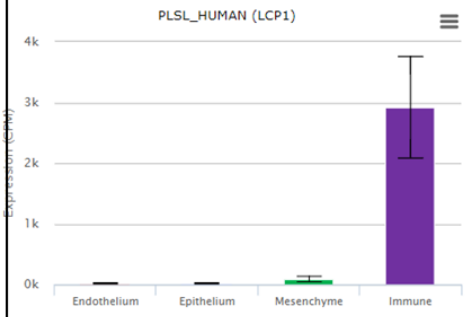
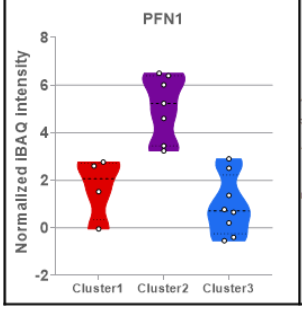
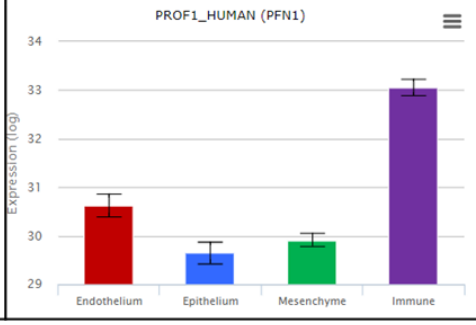
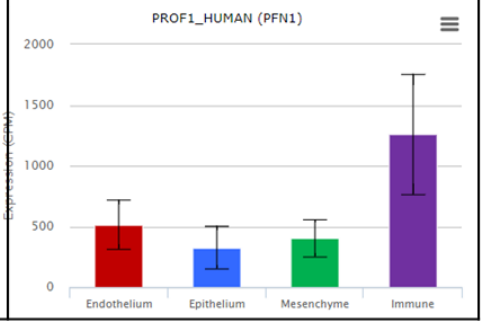


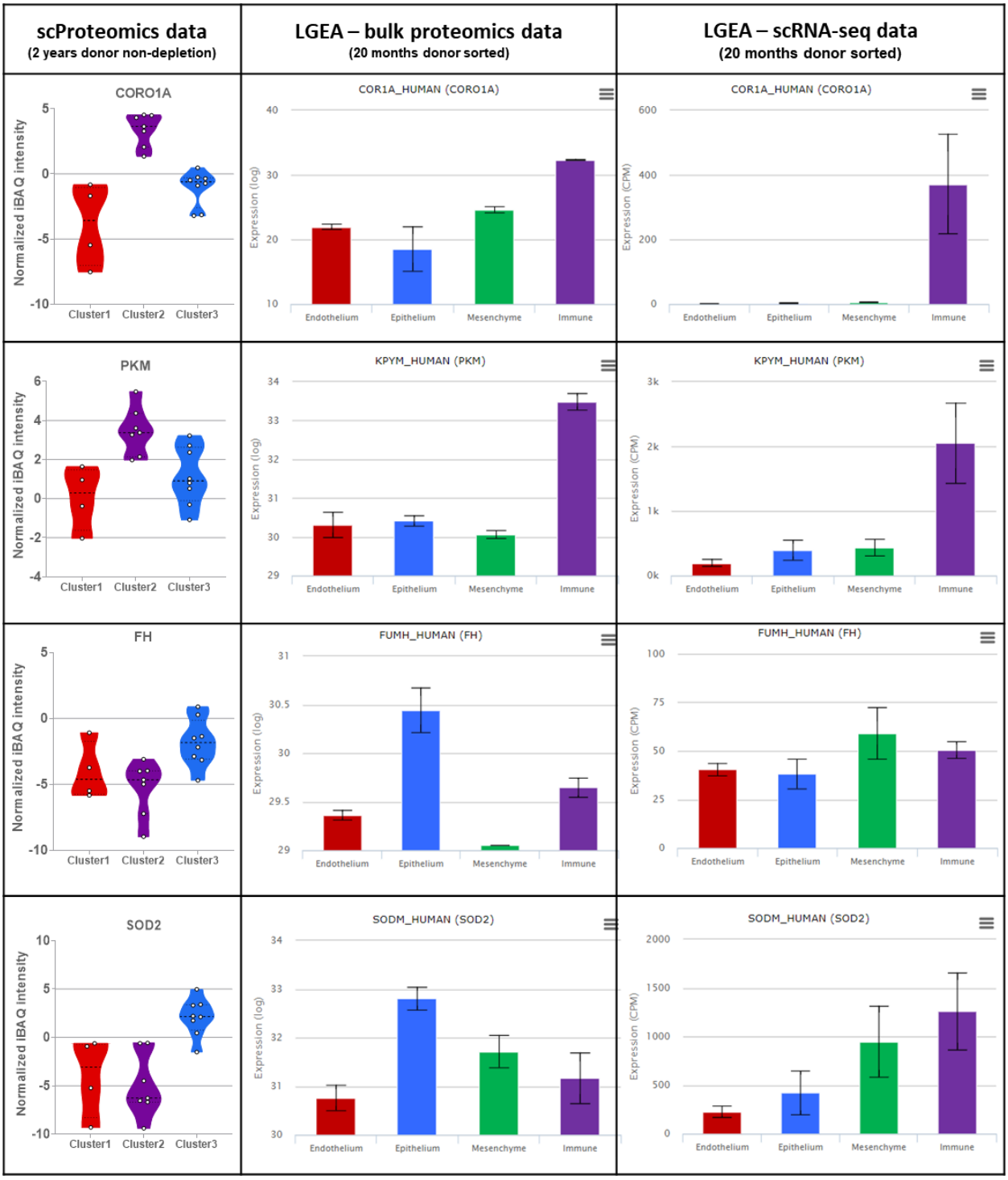
Supplementary Figure 5. (a-b) Statistics analysis to identify differentially abundant proteins (DAPs) between CMK and MOLM14 cells (t-test FDR < 0.05 and S_0 = 0.1). Volcano plots for **(a)** the standard and **(b)** 4-CV TIFF methods. **(c)** Overlap of quantifiable proteins between CMK and MOLM14 cells measured by the standard and 4-CV TIFF methods. **(d)** The linear correlation of log2-transformed fold changes of CMK and MOLM14 proteins between the 4-CV TIFF and STD methods. Red dots indicate (DAPs) calculated by t-test (FDR<0.05, S_0=0.1).

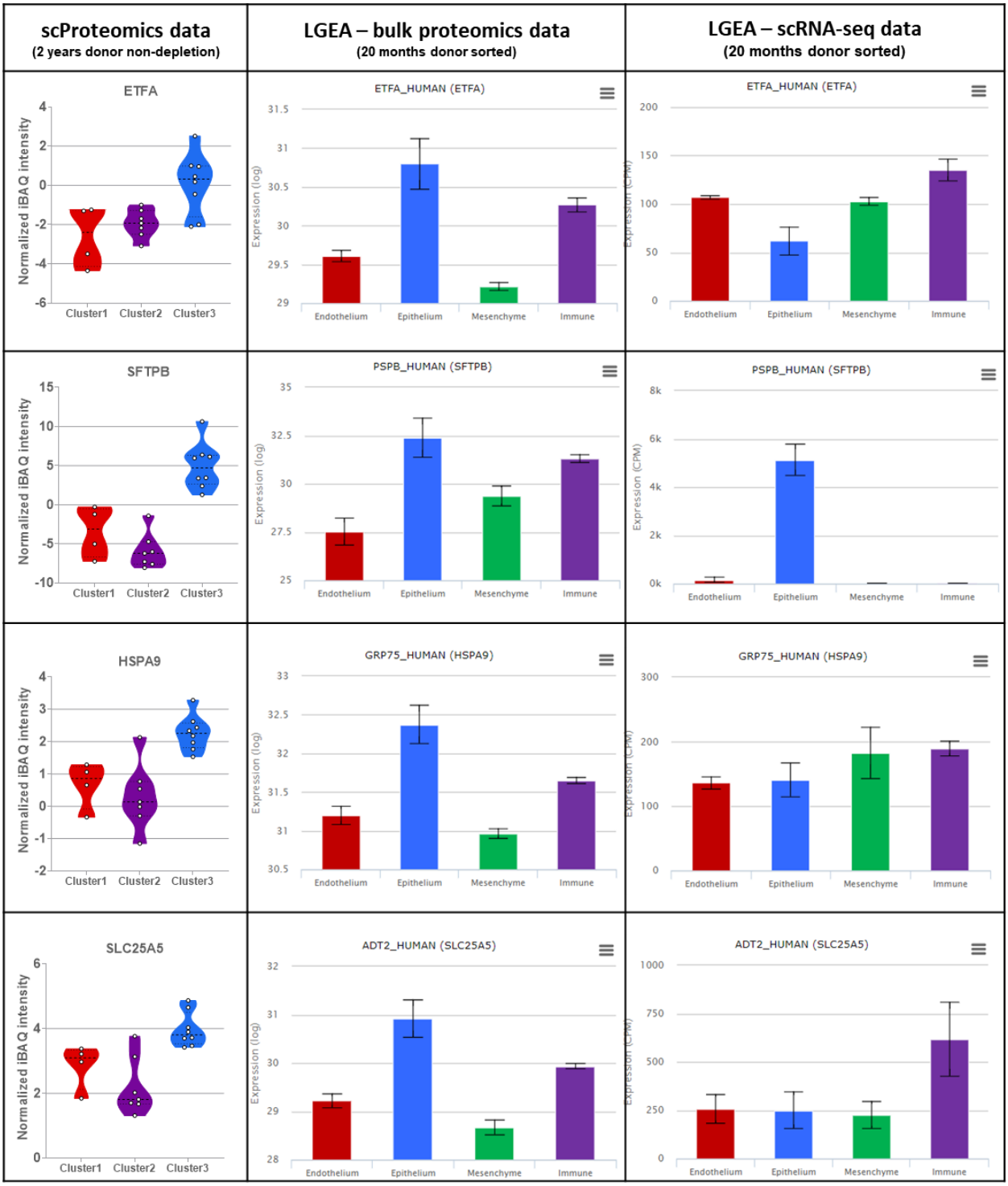


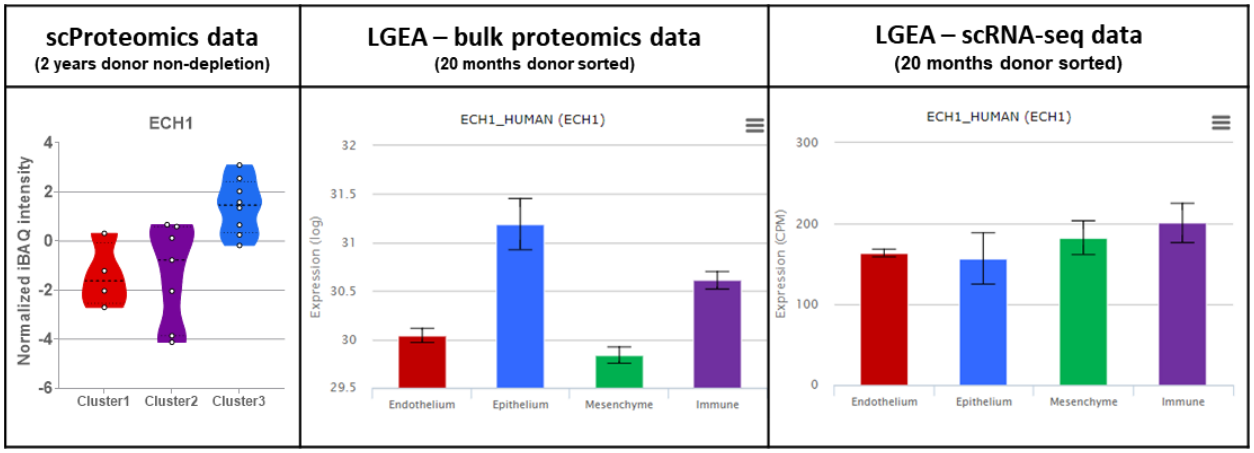
Supplementary Figure 6. (a) The numbers of identified proteins in single macrophage cells at different conditions. The median values of each condition were marked with a white dotted line in violin plots. **(b)** Abundance distributions of representative regulated proteins from different treatment conditions. **(c)** Heatmap showing the protein abundance differences across the 155 macrophage cells after statistical test using ANOVA (FDR <0.001, S_0 = 5). The hierarchical clustering was performed using the Euclidean method with 6 number of clusters for 250 DAPs by ANOVA test. Proteins in cluster A to C were applied to enrichment analysis using DAVID bioinformatics tools.

scProteomics data (2 years donor non-depletion)	LGEA – bulk proteomics data (20 months donor sorted)	LGEA – scRNA-seq data (20 months donor sorted)
		
		
		
		

scProteomics data (2 years donor non-depletion)	LGEA – bulk proteomics data (20 months donor sorted)	LGEA – scRNA-seq data (20 months donor sorted)
		
		
		
		


scProteomics data
(2 years donor non-depletion)
LGEA – bulk proteomics data
(20 months donor sorted)
LGEA – scRNA-seq data
(20 months donor sorted)
CORO1A
Normalized iBAQ intensity
Cluster1
Cluster2
Cluster3
COR1A_HUMAN (CORO1A)
Expression (log)
Expression (CPM)
Endothelium
Epithelium
Mesenchyme
Immune
PKM
KPYM_HUMAN (PKM)
FH
FUMH_HUMAN (FH)
SOD2
SODM_HUMAN (SOD2)


scProteomics data
(2 years donor non-depletion)
LGEA – bulk proteomics data
(20 months donor sorted)
LGEA – scRNA-seq data
(20 months donor sorted)
ETFA
Normalized iBAQ intensity
Cluster1
Cluster2
Cluster3
ETFA_HUMAN (ETFA)
Expression (log)
Expression (CPM)
Endothelium
Epithelium
Mesenchyme
Immune
SFTPB
PSPB_HUMAN (SFTPB)
HSPA9
GRP75_HUMAN (HSPA9)
SLC25A5
ADT2_HUMAN (SLC25A5)



Supplementary Figure 7. The abundance distributions of representative proteins markers in the scProteomics data and lung gene expression analysis (LGEA) database (https://research.cchmc.org/pbge/lunggens/mainportal.html) containing sorted human lung endothelial, epithelial, immune and mesenchymal cells measured by bulk proteomics and single-cell RNA sequencing [1].

Single-cell approach	LC system (column I.D. in µm, flow rate in nL/min)	MS instrument	Cell Type	Protein Groups (by MS/MS)	Proteins Groups (MBR)	Reference
nanoPOTS	30 / 50	Lumos	HeLa	211	669	Zhu et. al, *Angew chem*, 2018 [2]
nanoPOTS with autosampler	50 / 150	Lumos	MCF10	250	773	Sarah et. al. *Analytical Chemistry*, 2020 [3]
nanoPOTS with narrow-bore LC	20 / 20	Eclipse	HeLa	362	874	Cong et. al. *Analytical Chemistry*, 2020 [4]
nanoPOTS with FAIMS	20 / 20	Eclipse	HeLa	683 (By MQ)/ 1056 (By PD)	1475	Cong et. Al. *bioRxiv*, 2020 [5]
nanoPOTS with TIFF	50 / 100	Lumos	HeLa	209	1212	*This study*

Supplementary Table 1. Numbers of identified proteins in single mammalian cells from previously published papers.

Proteins were identified by MS/MS or by matching between runs (MBR) algorithm in Maxquant software [6] (MQ: MaxQuant, PD: Proteome Discoverer).

Supplementary references

1. Du, Y. et al. Integration of transcriptomic and proteomic data identifies biological functions in cell populations from human infant lung. *Am J Physiol Lung Cell Mol Physiol* **317**, L347-L360 (2019).
2. Zhu, Y. et al. Proteomic Analysis of Single Mammalian Cells Enabled by Microfluidic Nanodroplet Sample Preparation and Ultrasensitive NanoLC-MS. *Angew Chem Int Ed Engl* **57**, 12370-12374 (2018).
3. Williams, S.M. et al. Automated Coupling of Nanodroplet Sample Preparation with Liquid Chromatography-Mass Spectrometry for High-Throughput Single-Cell Proteomics. *Anal Chem* (2020).
4. Cong, Y. et al. Improved Single-Cell Proteome Coverage Using Narrow-Bore Packed NanoLC Columns and Ultrasensitive Mass Spectrometry. *Anal Chem* **92**, 2665-2671 (2020).
5. Cong, Y. et al. Ultrasensitive single-cell proteomics workflow identifies >1000 protein groups per mammalian cell. *bioRxiv*, 132449 (2020).
6. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc* **11**, 2301-2319 (2016).