

Supporting Information

A Comparison of Database Searching Programs for the Analysis of Single-Cell Proteomics Data

Jiaxi Peng^{1,3,4}, Calvin Chan¹, Fei Meng², Yechen Hu^{1,3,4}, Lingfan Chen⁵, Ge Lin^{2,6},
Shen Zhang^{2,*}, Aaron R. Wheeler^{1,3,4,*}

¹Department of Chemistry, University of Toronto, Toronto, Ontario, M5S 3H6, Canada

²Clinical Research Center for Reproduction and Genetics in Hunan Province, Reproductive and Genetic Hospital of CITIC-XIANGYA, Changsha, Hunan, 410000, China

³Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario, M5S 3E1, Canada

⁴Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, Ontario, M5S 3G9, Canada

⁵Fujian Province New Drug Safety Evaluation Centre, Fujian Medical University, Fuzhou Fujian, 350108, China

⁶Laboratory of Reproductive and Stem Cell Engineering, NHC Key Laboratory of Human Stem Cell and Reproductive Engineering, Central South University, Changsha, Hunan, 410075, China

*Contact details for the corresponding authors:

Shen Zhang

E-mail: szhang231@126.com

Address: 567 Tongzipo West Road, Clinical Research Center for Reproduction and Genetics in Hunan Province, Reproductive and Genetic Hospital of CITIC-XIANGYA, Changsha, Hunan, 410000, China

Aaron R. Wheeler

Email: aaron.wheeler@utoronto.ca; Tel: 416-946-3866; Fax: 416-946-3865

Address: 80 St. George Street, Department of Chemistry, University of Toronto, Toronto, ON, M5S 3H6, Canada

Table of Contents

1. Figure S1. Peptide and protein identifications integrated by iProphet.	S3
2. Figure S2. PSM numbers of proteins identified by Proteome Discoverer and MSGF+ from three datasets.....	S4
3. Figure S3. Plot of percentages of proteins identified in replicates of Hela digest dilutions.....	S5
4. Table S1. The percentage distributions of protein numbers in each abundance rank in the AutoPOTS dataset.....	S6
5. Table S2. The percentage distributions of protein numbers in each abundance rank in the nanoPOTS dataset.....	S7

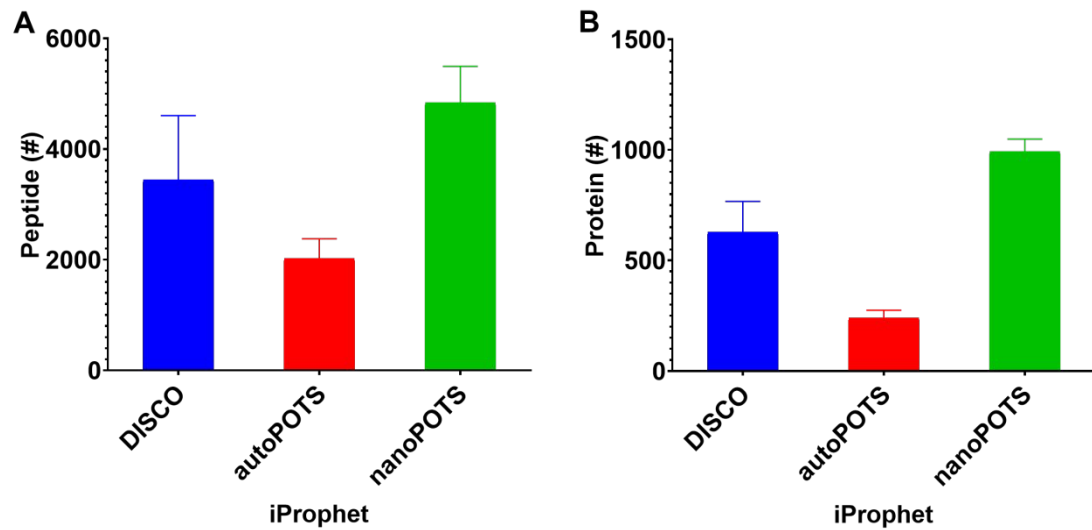


Figure S1. Plots of total, combined peptide (A) and protein (B) identifications from DISCO (blue), autoPOTS (red), and nanoPOTS (green) datasets using the iProphet (ProHits) integration function combining the outputs of Comet, Mascot, X!Tandem, MSGF+, and MSFragger. Error bars represent standard deviations from the analysis of three different cells in each study (n = 3).

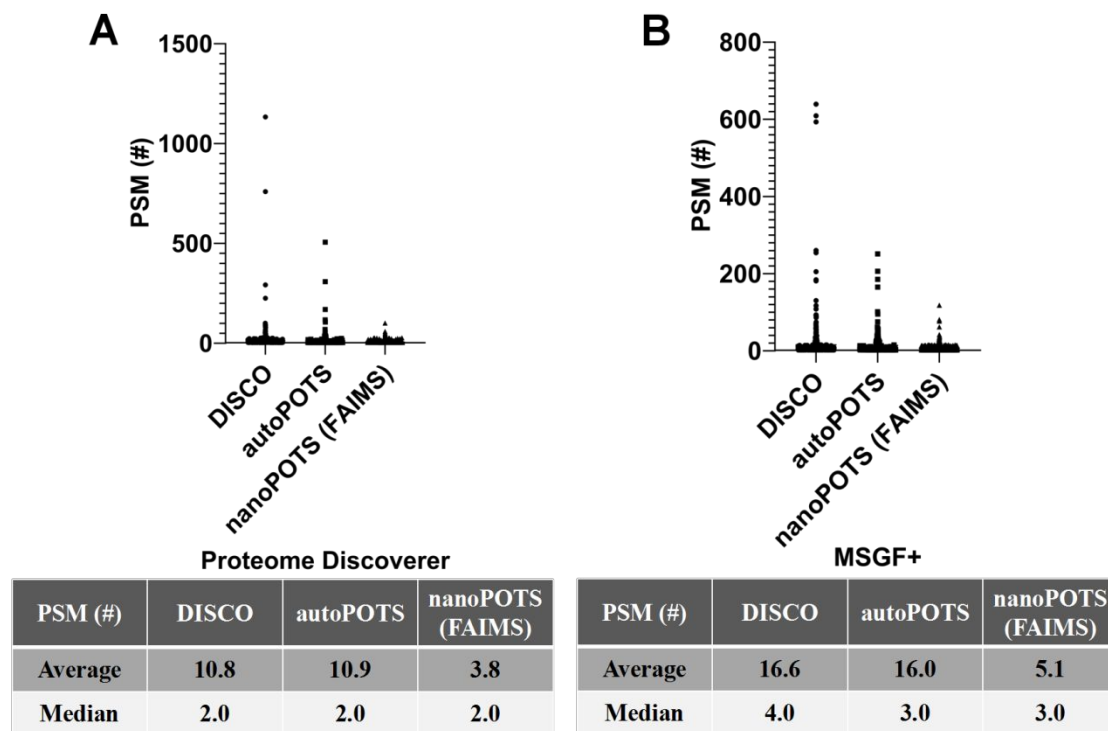


Figure S2. Plots of PSM numbers (black markers, above), and tables of average and median PSM numbers (below) of proteins identified by Proteome Discoverer (A) and MSGF+ (B) from single U87 cells in the DISCO dataset, and single HeLa cells in the autoPOTS and nanoPOTS (FAIMS) datasets.

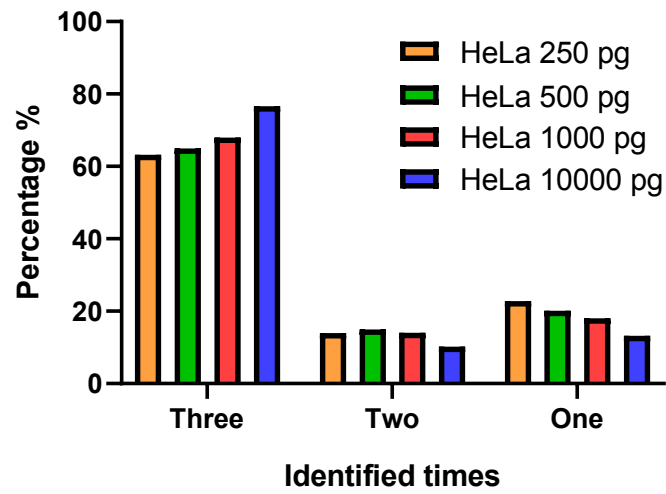


Figure S3. Plot of percentages of proteins identified in replicates (n=3) of HeLa digest dilutions containing 250 pg (orange), 500 pg (green), 1,000 pg (red) and 10,000 pg (blue) of total protein evaluated by HPLC-MS/MS (QE HFX) and searched by MaxQuant. Percentages are given for proteins identified in all three replicates (left), in two of the three replicates (middle), and only one of the three replicates (right).

Table S1. Percentage distributions of protein numbers in each abundance rank in the AutoPOTS dataset (as shown in Figure 3A and 3C) by Comet, Mascot, X!Tandem, MSGF+, MSFragger, Proteome Discoverer, and MaxQuant.

Copy number range	Comet	Mascot	X!Tandem	MSGF+	MSFragger	Proteome Discoverer	MaxQuant
$1 \times 10^2 - 1 \times 10^3$	6%	6%	5%	5%	4%	4%	10%
$1 \times 10^3 - 1 \times 10^4$	10%	10%	8%	9%	8%	12%	16%
$1 \times 10^4 - 1 \times 10^5$	8%	7%	7%	8%	7%	10%	12%
$1 \times 10^5 - 1 \times 10^6$	12%	13%	13%	14%	14%	15%	13%
$1 \times 10^6 - 1 \times 10^7$	44%	40%	49%	49%	50%	45%	36%
$1 \times 10^7 - 1 \times 10^8$	21%	24%	19%	15%	17%	14%	14%
Matched protein number	160	157	199	247	229	264	210

Table S2. Percentage distributions of protein numbers in each abundance rank in the nanoPOTS dataset (as shown in Figure 3B and 3D) by Comet, Mascot, X!Tandem, MSGF+, MSFragger, Proteome Discoverer, and MaxQuant.

Copy number range	Comet	Mascot	X!Tandem	MSGF+	MSFragger	Proteome Discoverer	MaxQuant
$1 \times 10^2 - 1 \times 10^3$	2%	1%	2%	2%	1%	2%	5%
$1 \times 10^3 - 1 \times 10^4$	4%	2%	4%	5%	3%	5%	8%
$1 \times 10^4 - 1 \times 10^5$	5%	2%	4%	5%	5%	6%	9%
$1 \times 10^5 - 1 \times 10^6$	31%	9%	24%	35%	26%	36%	23%
$1 \times 10^6 - 1 \times 10^7$	53%	68%	60%	49%	59%	47%	50%
$1 \times 10^7 - 1 \times 10^8$	5%	17%	6%	4%	6%	4%	5%
Matched protein number	1041	207	833	1215	887	1264	777