Supporting Information

Evaluation of spectral counting for relative quantitation of proteoforms in top-down proteomics

Lucía Geis-Asteggiante[1]; Suzanne Ostrand-Rosenberg[2]; Catherine Fenselau[1]; Nathan J. Edwards[3,*]

[1] University of Maryland, College Park, MD. 20742; [2] University of Maryland Baltimore County, Baltimore, MD. 21250; [3] Georgetown University Medical Center, Washington, DC. 20007.

Corresponding Author: *Email: nje5@georgetown.edu

TABLE OF CONTENTS

Methods and Materials

Results and Discussion

Table S1. Eight aliquots of an exosome lysate were spiked as described in the table below in order to reach a constant protein amount of 680 ng per injection. Note that even though 5 standard proteins were spiked, apolipoprotein D was not identified in the mixture and apomyoglobin co-eluted with S100A8 proteoforms, which are the most intense peaks in the spectra, hindering apomyoglobin identification

Lysate (ng)	**TNF-a (ng)**	**CAH (ng)**	**RNase (ng)**	**Apomyoglobin (ng)**	**Apolipoprotein D (ng)**	**Total amount of protein (ng)**
500	40	10	50	60	20	680
500	20	50	10	40	60	680
500	60	-	60	20	40	680
500	50	20	40	60	10	680
500	50	40	20	10	60	680
500	10	40	20	50	60	680
500	5	100	5	70	-	680
500	1	100	1	78	-	680

Table S2. Complete list of proteoforms identified and comparison of differential abundances calculated by spectral counting, normalized intensity and area. See Excel file "Supporting Information Table S2."

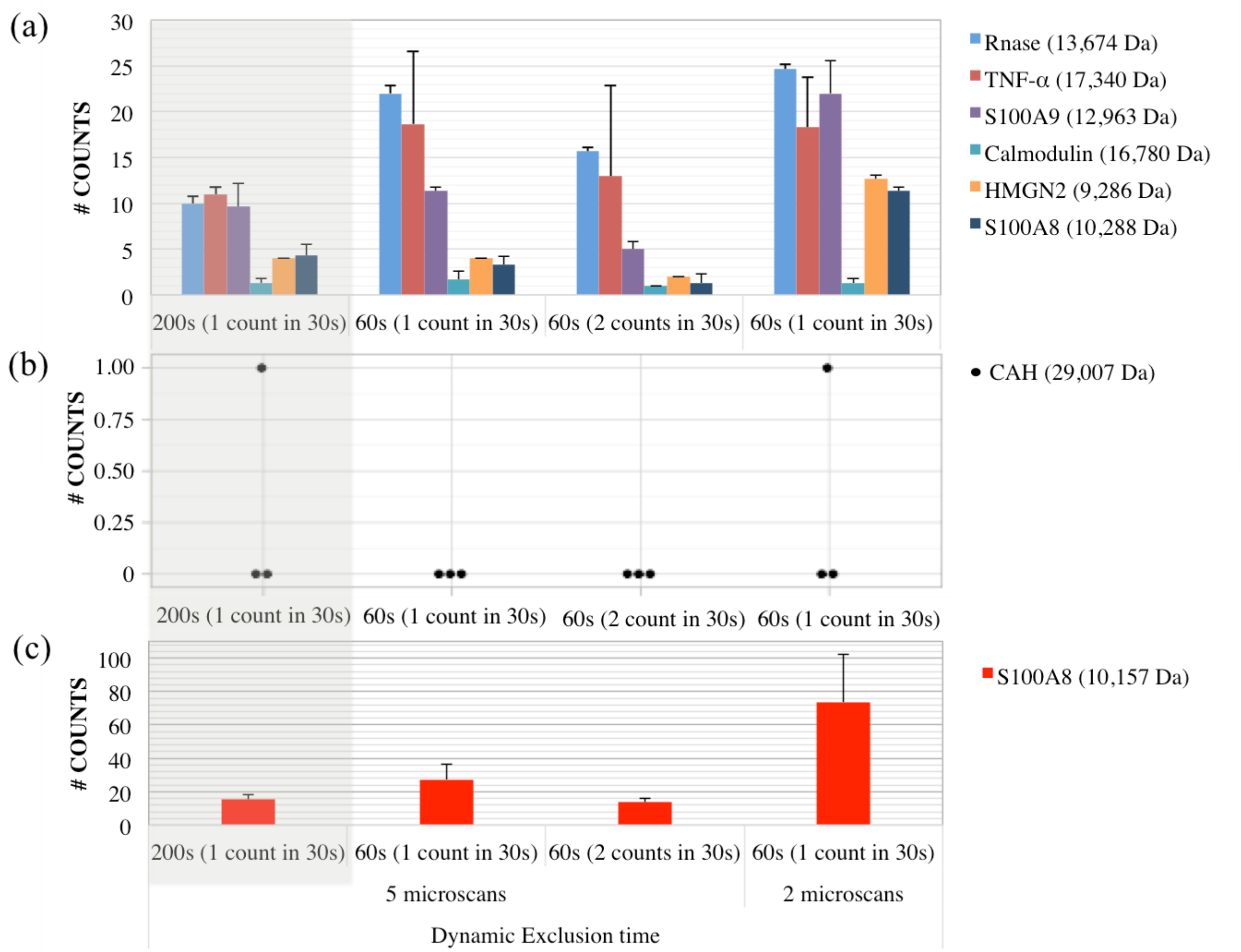
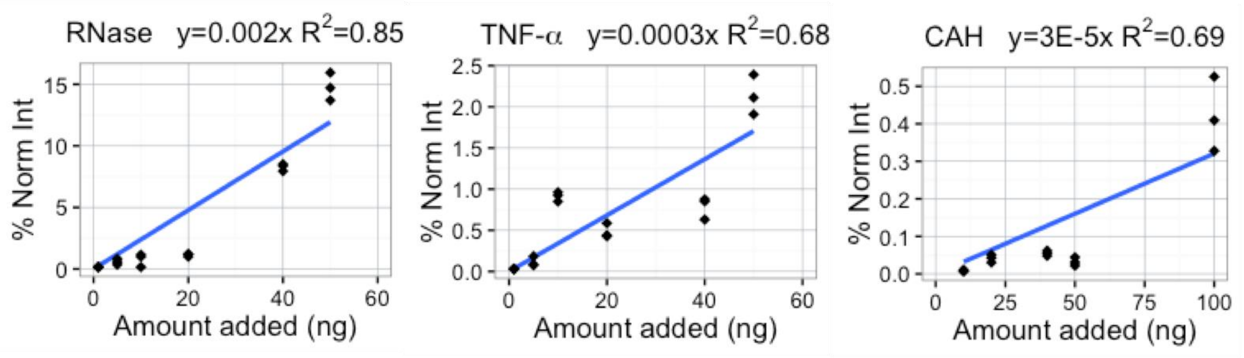
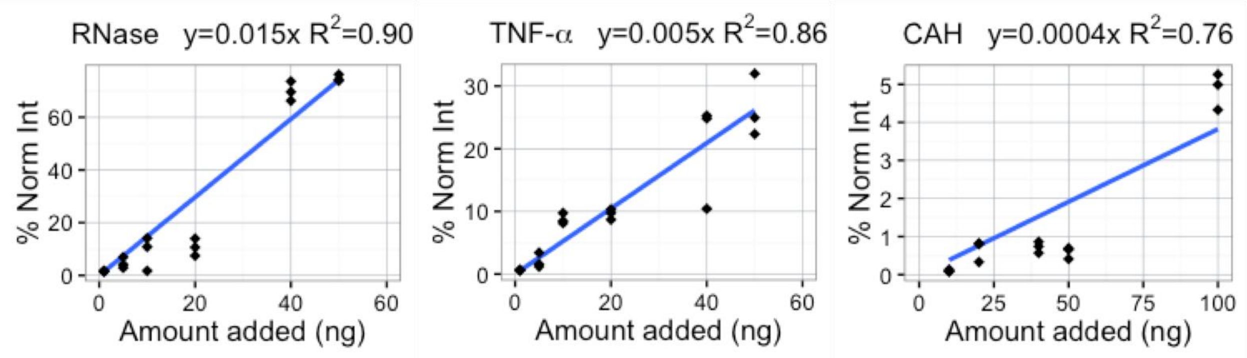


Figure S1. Effect of dynamic exclusion and microscans averaging on PrSMs counts of (a) 2 protein standards spiked (50 ng) and 4 proteoforms found in the exosome lysate (representing in average 0.2 – 4.2% of the TIC intensity), (b) CAH standard spiked, and (c) S100A8 (10,157 Da) proteoform, which represents in average 37% of the TIC intensity. Error bars represent the standard deviation of the mean. Typically used dynamic exclusion settings for top-down proteomics are highlighted in light gray.

(a) Normalized areas



(b) Normalized intensities



(c) Spectral counts

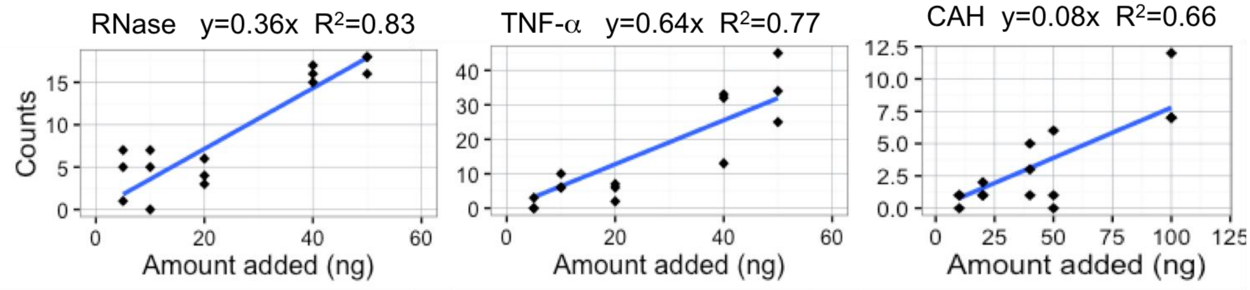


Figure S2. Calibration curves of (a) normalized area, (b) normalized intensity, (c) spectral counts vs. amount spiked into the exosome sample for each protein standard used.

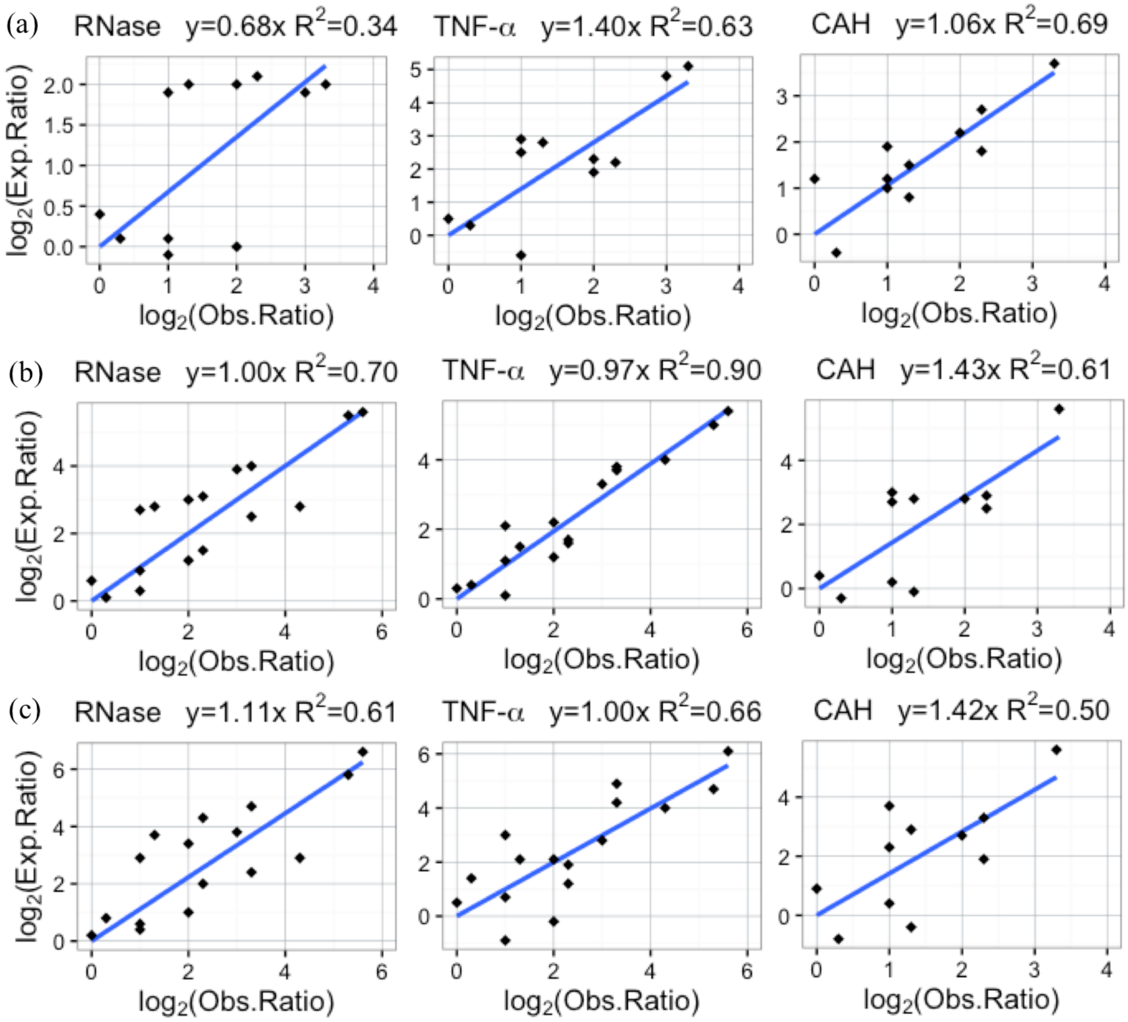
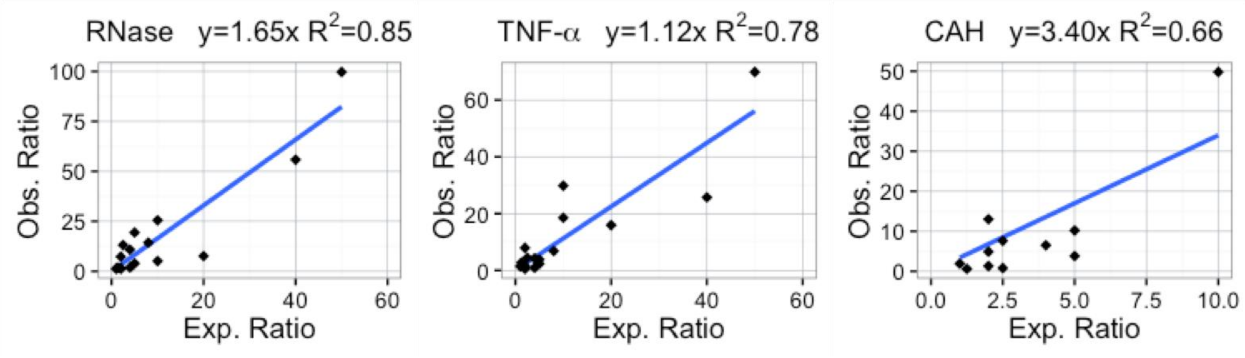
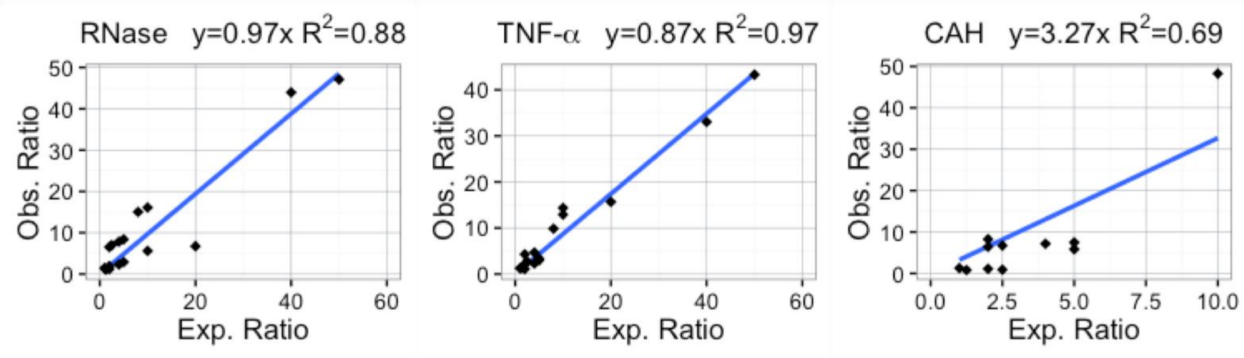


Figure S3. Comparison of expected and observed protein ratios estimated by (a) spectral counting, (b) normalized intensities and (c) normalized areas from deconvoluted ion extracted chromatograms.

(a) Normalized areas



(b) Normalized intensities



(c) Spectral counts

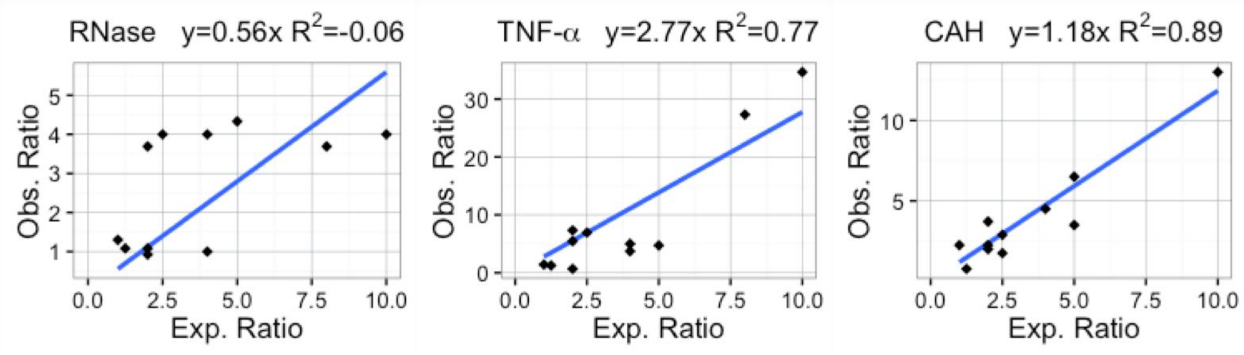


Figure S4. Linear regression plots of observed and expected protein ratios without $\log_2$ transformation obtained for (a) normalized area, (b) normalized intensity, (c) spectral counts.

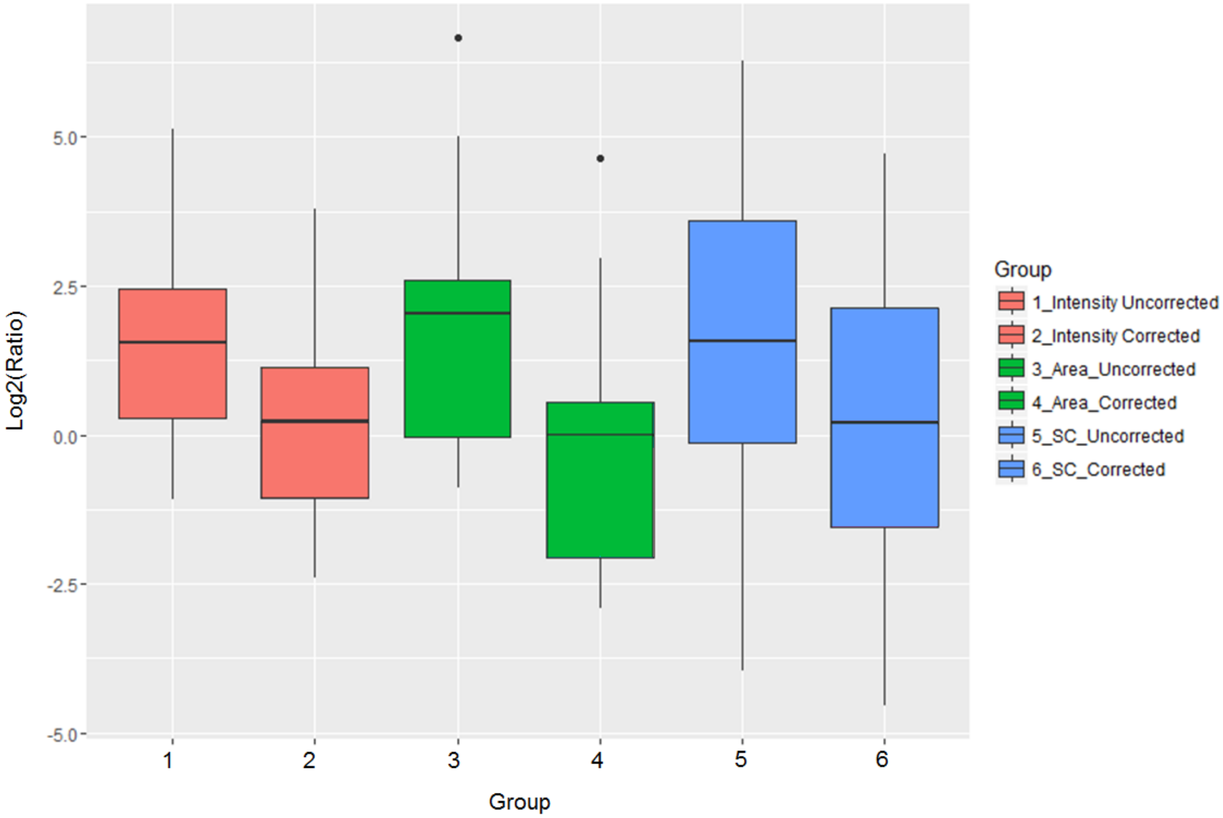


Figure S5. Distribution of $\log_2$(ratios) estimated using intensities, areas and spectral counts, before and after correction by median normalization.

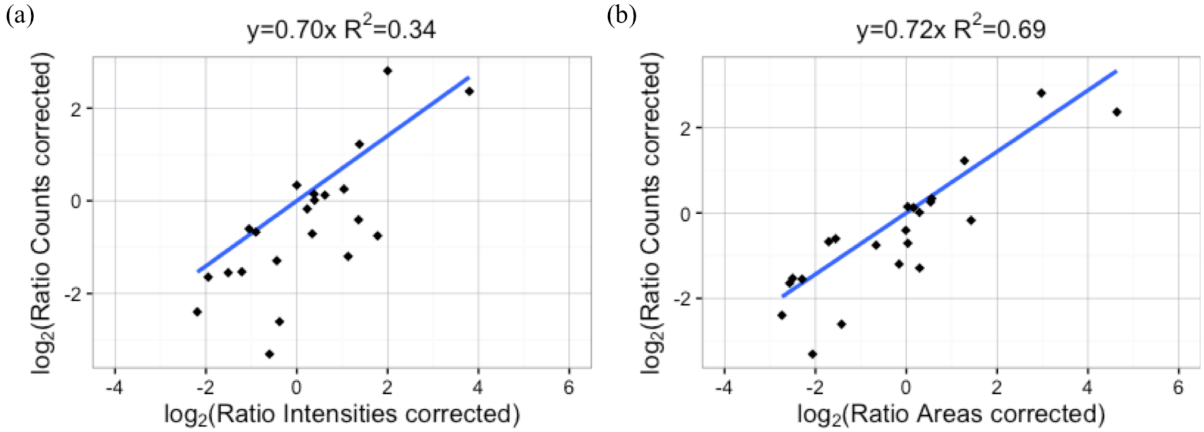


Figure S6. Comparison of $\log_2$(ratios) found for the 22 quantified proteoforms in the exosome samples estimated using peak (a) intensities and (b) areas vs. those obtained by spectral counting.