Supporting Information

Proteome-wide evaluation of two common protein quantification methods

Jeremy D. O'Connell[1], *Joao A. Paulo*[1], *Jonathon J. O'Brien*[1], *and Steven P. Gygi*[1*]

[1]Department of Cell Biology, Harvard Medical School, Boston, MA

*Correspondence should be addressed to Steven P. Gygi (steven_gygi@hms.harvard.edu)

Table of Contents

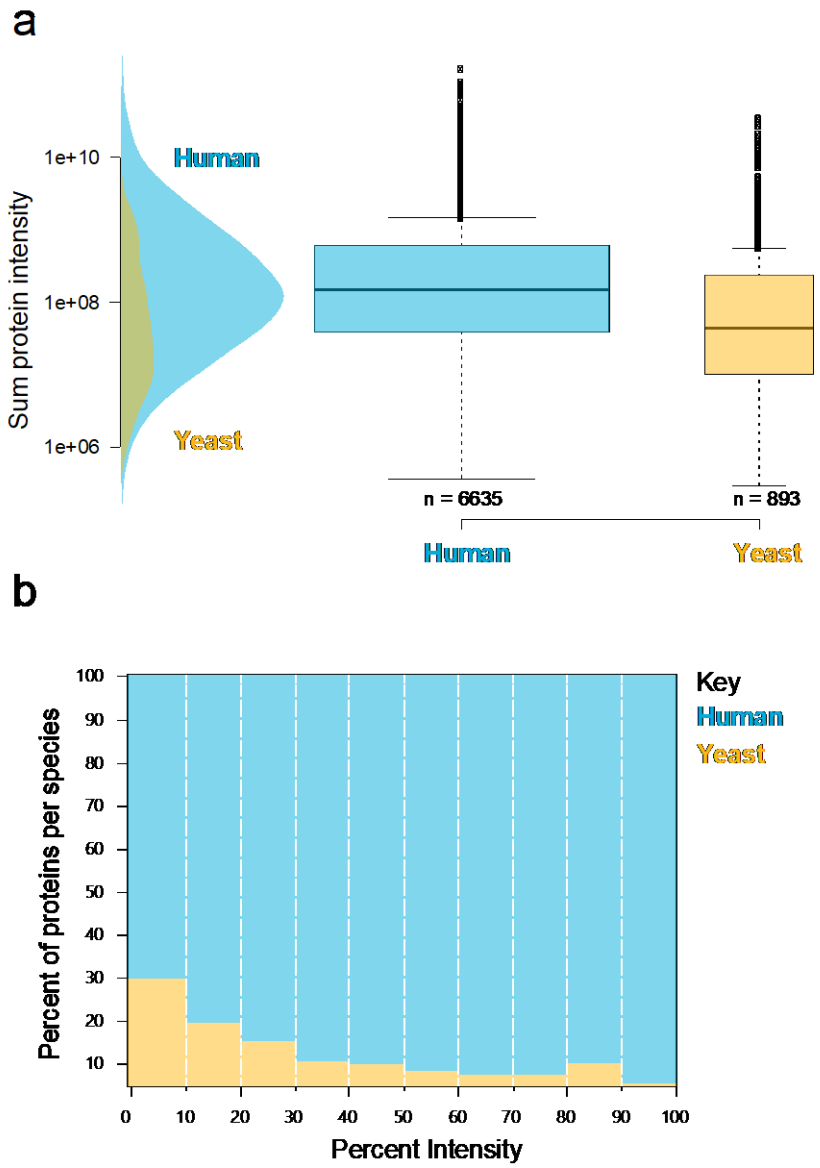
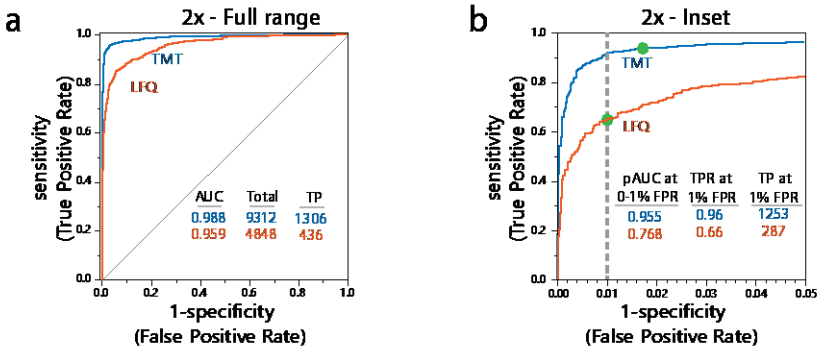
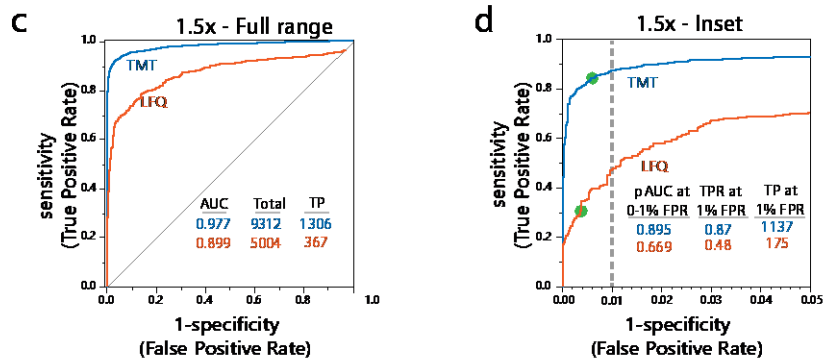


Figure S1. Distribution of protein intensities for the two proteomes mixed in the samples shows yeast proteins present across the entire range of protein abundances, but with a distinct enrichment among the less abundant proteins. a) Density and box plots show the summed intensities for each protein quantified in the human and yeast proteomes across all 11 label-free samples as reported by MaxQuantLFQ. **b)** Stacked histograms showing the distribution of proteins contributed by each species across the deciles of protein intensity (summed across the 11 samples for each protein). There is a general trend for yeast proteins to be present at lower intensities, reflecting lower relative abundance within the samples.





e

Perseus t-test results

	total proteins	total yeast	total human	3x at q<0.05 yeast proteins # changing	3x at q<0.05 yeast proteins %	2x at q<0.05 yeast proteins # changing	2x at q<0.05 yeast proteins %	1.5x at q<0.05 yeast proteins # changing	1.5x at q<0.05 yeast proteins %
TMT	9312	1306	8006	1270	97	1255	96	1065	80
LFQ	7731	945	6786	296	80	287	66	107	19
LFQ imputed	7731	945	6786	304	32	282	30	101	11

Figure S2. The TMT method had superior sensitivity likely due to a combined effect of reduced measurement error and fewer missing values. a,c) ROC plots of the adjusted p-values from a 2-sided t-test for the 2-fold and 1.5-fold change comparisons for proteins with >2 quantitation values highlighting the sensitivity and specificity for the TMT (blue) and LFQ (red) methods, which at the whole proteome level were both quite good at distinguishing proteins that changed abundance between groups (yeast) from proteins that did not (human). b,d) In the FPR range from 0 to 0.05, which contains the proteins typically chosen for follow-up experiments, the TMT method notably outperformed the LFQ method. Green filled circles represent the actual point where the adjusted p-value was 0.05. Note that Figure 4 shows the ROC plots for the 3-fold differences. e) Summary statistics for each method's performance at a common significance threshold ($p < 0.05$ after multiple hypothesis-testing correction) for both the protein that should be changing (yeast) and those that should not (human) demonstrates the trade-off between sensitivity and specificity, particularly for imputed LFQ data where imputed values reduce the false positive rate at the cost of also reducing the true positive rate. TP = true positives, TPR = true positive rate, FP = false positives, FPR = false positive rate, AUC = area under the curve, pAUC = partial area under the curve.

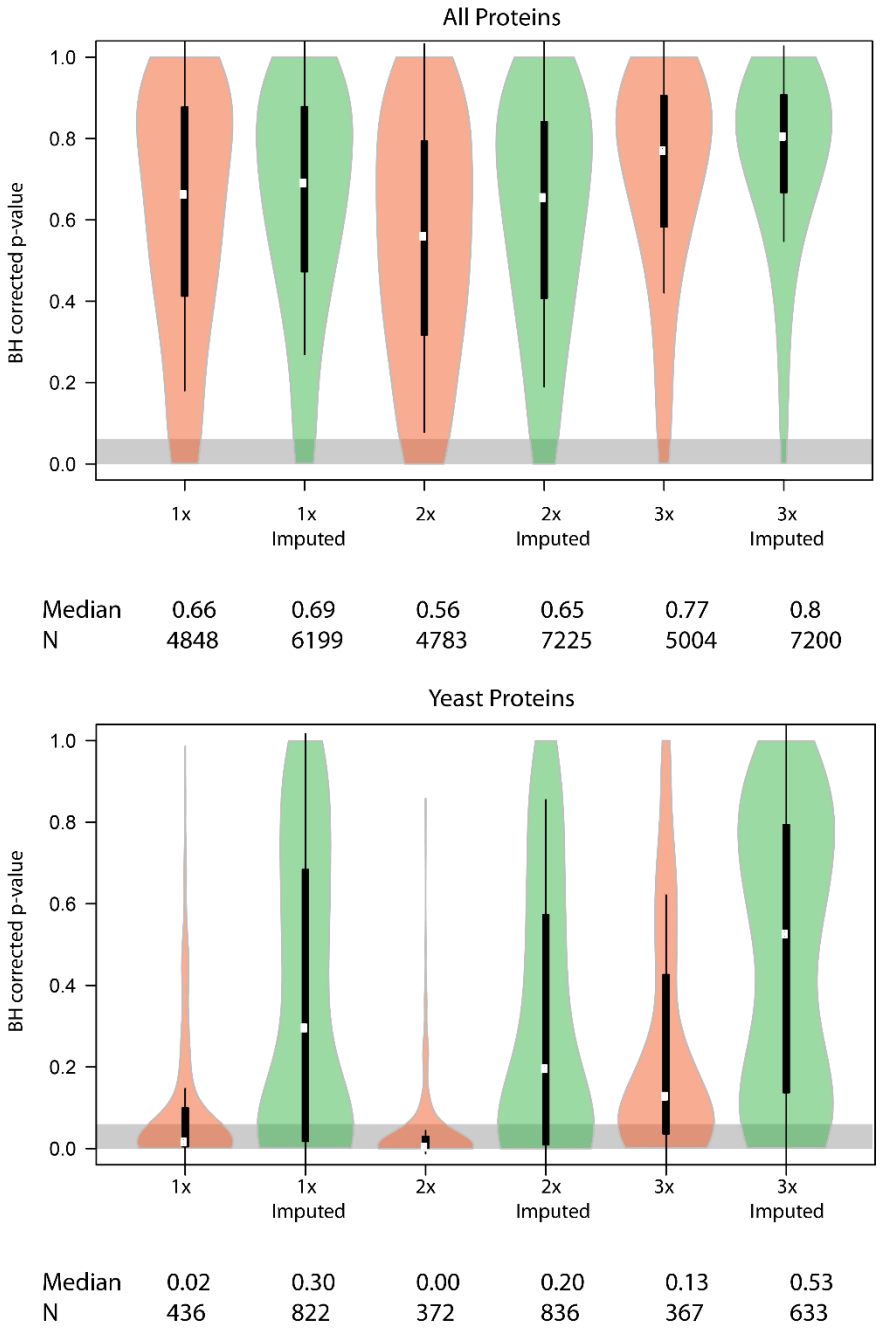


Figure S3. Imputation (green) markedly increased the number of proteins available for statistical inference, nearly doubling the number of yeast proteins (bottom) compared to the non-imputed set (red), but at the cost is an increased false negative rate among the class of proteins that change abundance between samples (the fraction above the grey bar in the lower plot). Violin plots show the distribution of Benjamini-Hochberg multiple hypothesis testing corrected p-values for all proteins (top) or yeast protein specifically (bottom), with the grey bar highlighting the fraction of those distributions that would pass a conventional significance test of $p<0.05$.

	Fractionated	Number Samples	MS Time	Total Proteins	Total * Peptides	Yeast Proteins	Yeast * Peptides	MS2 Scans	MS3 Scans
Sequest TMT	Yes	11	33h	9,312	112,663	1,306	8,999	437,663	437,645
MQ TMT	Yes	11	33h	8,912	105,442	1,247	8,788	437,663	437,645
Sequest LFQ	No	11	33h	7,903	489,499	1,071	31,248	805,513	0
MQ LFQ	No	11	33h	7,731	454,545	945	29,391	805,513	0

Supplemental Table 3. MaxQuant and a Sequest based pipeline produce nearly similar dataset level statistics when run with matching parameters on the same set of raw files, suggesting experimental design rather than choice of search engine is the major determinant of the resulting data structures. In the top two rows (blues), the TMT -11plex sample .raw files were searched with either the Sequest pipeline or MaxQuant. In the bottom two rows (reds), the label-free sample .raw files were searched with either the Sequest pipeline or MaxQuant. *Match-between-runs increases the Total Peptides for LFQ by MaxQuant 627,579 and to 43,605 for the Yeast Peptides. Match-between-runs was not enabled on the TMT data.