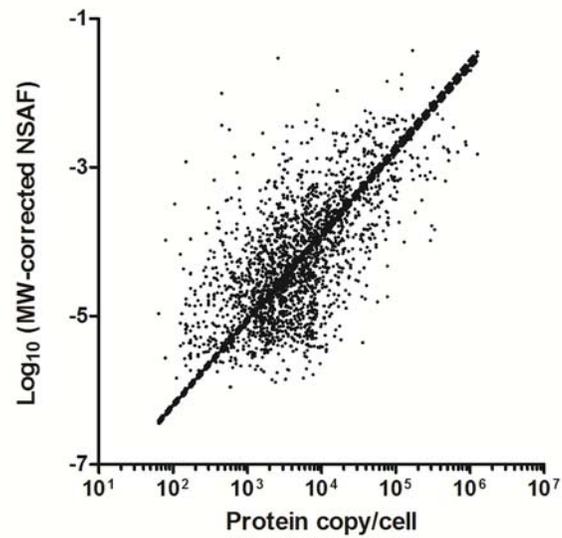


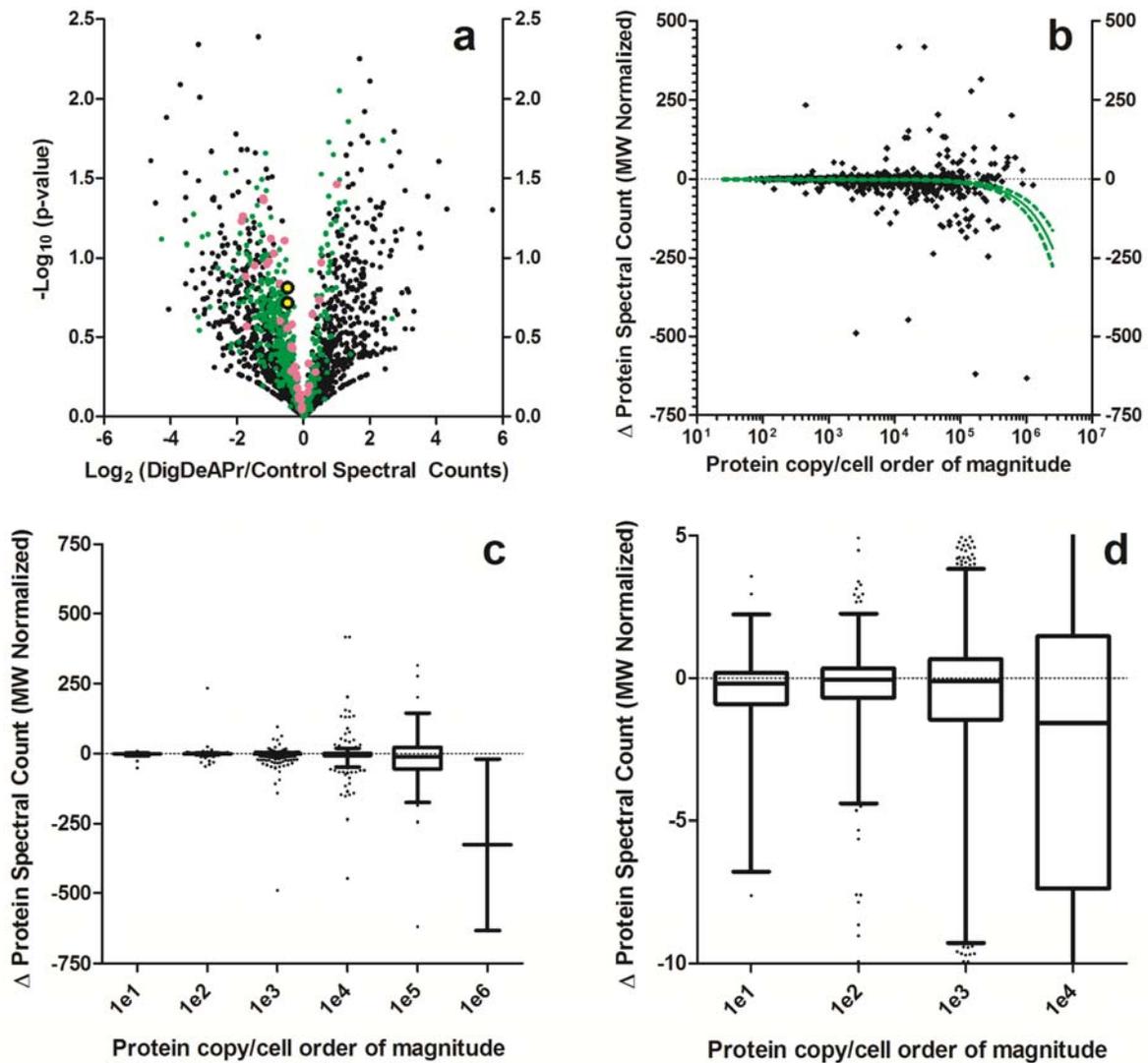
Supplementary Figures and Notes for *Digestion and depletion of abundant proteins improves proteomic coverage*

Bryan R. Fonslow, Benjamin D. Stein, Kristofor J. Webb, Tao Xu, Jeong Choi, Sung Kyu Park, and John R. Yates III

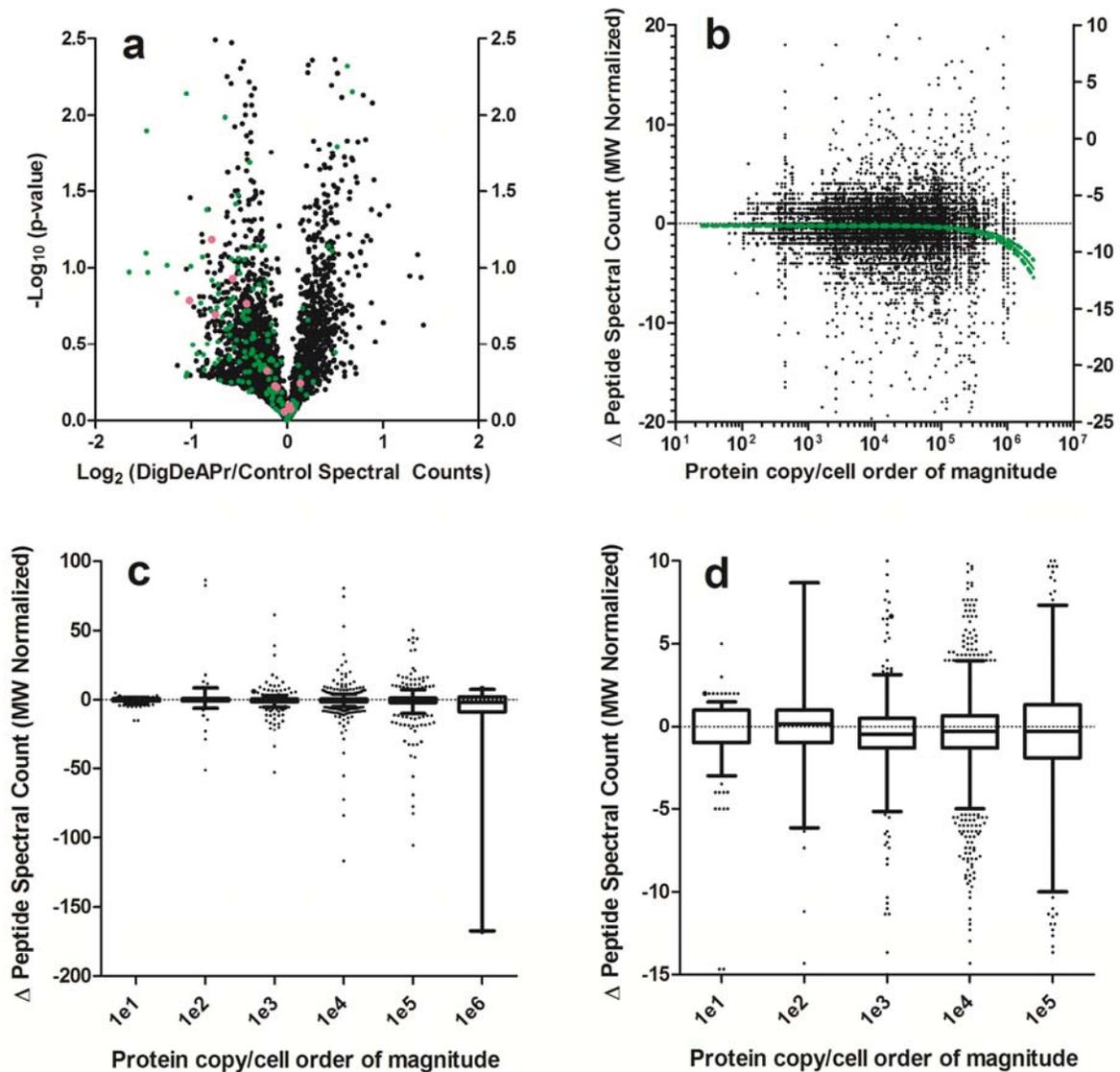
Supplementary Figure 1	Log-log correlation plot of MW-corrected normalized spectral abundance factor (NSAF) for yeast proteins to their absolute cell copy number
Supplementary Figure 2	Analysis of digestion and depletion of proteins from a tryptic yeast lysate
Supplementary Figure 3	Analysis of digestion and depletion of tryptic peptides from a yeast lysate
Supplementary Figure 4	Comparison of proteins and peptides from triplicate Control and DigDeAPr runs of tryptic HEK lysates
Supplementary Figure 5	Statistical comparison of peptide quality scores from HEK cell lysates
Supplementary Figure 6	Peptide precursor intensity histogram comparison from HEK cell lysates
Supplementary Figure 7	Surface plots of HEK protein physicochemical properties based on relative spectral count changes
Supplementary Figure 8	Statistical comparison of proteins between two separate triplicate Control analyses of tryptic HEK lysates
Supplementary Note 1	Derivation of Michaelis-Menten equation to describe protein digestion in a complex mixture
Supplementary Note 2	Estimation and evaluation of initial DigDeAPr conditions
Supplementary Note 3	Validation of DigDeAPr with known yeast protein copy numbers
Supplementary Note 4	Improvements to peptide quantitation metrics with DigDeAPr
Supplementary Note 5	Comparison of dynamic range compression mechanisms between DigDeAPr and hexapeptide-based ProteoMiner
Supplementary Note 6	Analysis of potential protein losses from the MWCO spin-filter



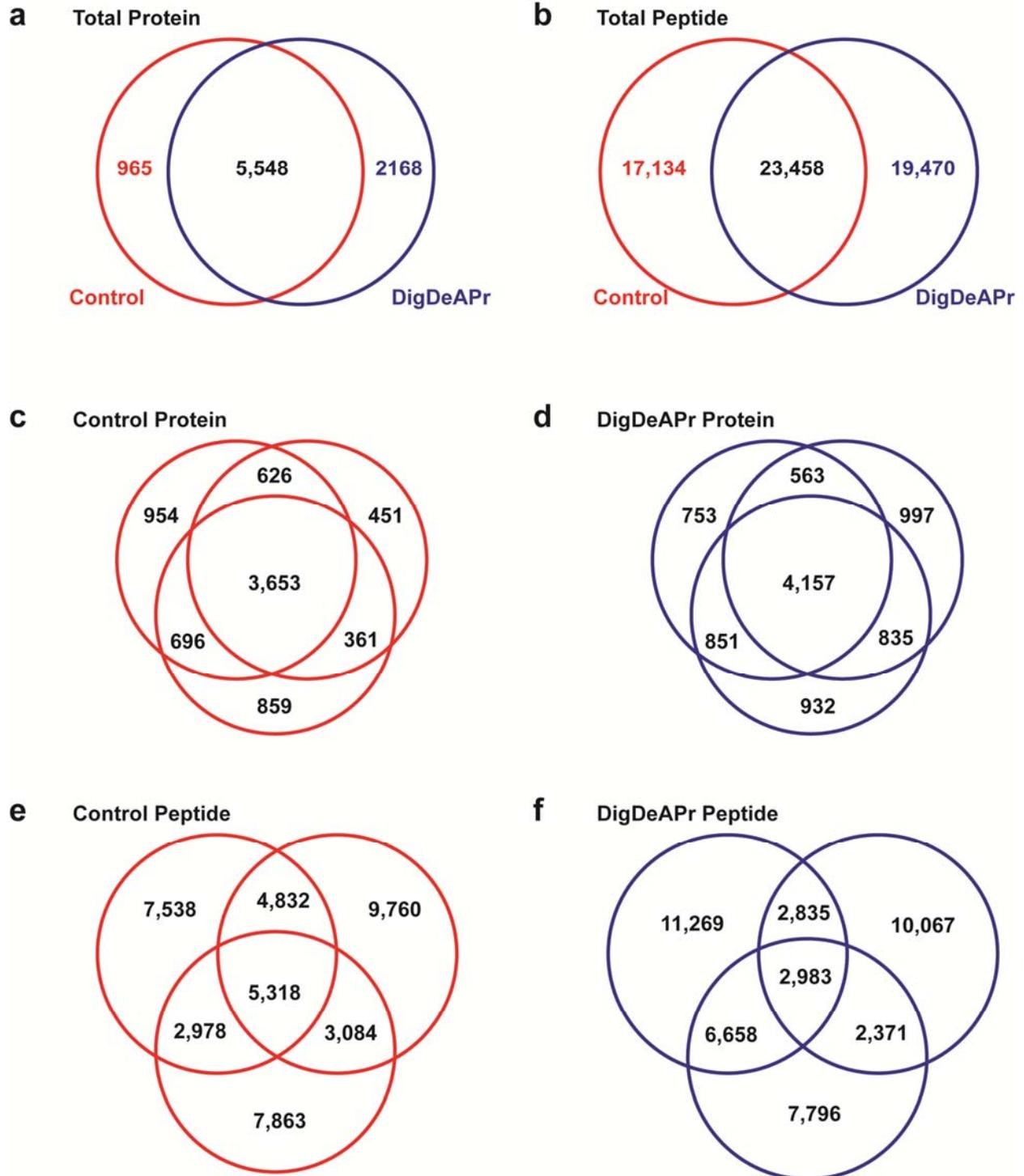
Supplementary Figure 1: Log-log correlation plot of molecular weight-corrected normalized spectral abundance factor (NSAF) for yeast proteins to their absolute cell copy number. The data was fit using a nonlinear regression. The best fit line and 95% confidence interval are plotted.



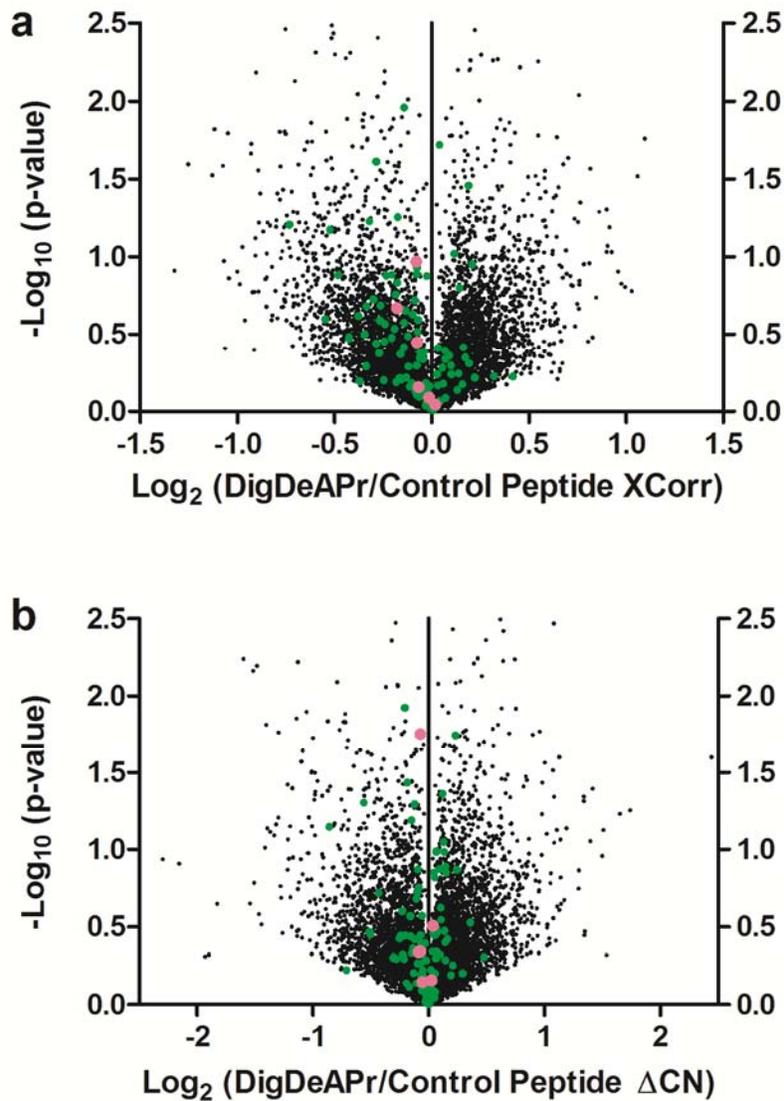
Supplementary Figure 2: Analysis of digestion and depletion of proteins from a tryptic Yeast lysate. (a) Volcano plot of the log₂ average protein spectral count ratio between triplicate DigDeAPr and Control runs plotted against the p-value. Data points are plotted based on average spectral counts from triplicate Control runs: 1-9 spectral counts (black), 10-99 (green), 100-999 (magenta), and more than 1000 spectral counts (yellow with black outline). (b) Correlation of spectral count depletion to absolute protein abundance in Yeast. The data was fit using a linear regression. The fit line and 95% confidence interval are plotted. Box whisker plots of spectral count depletion for (c) all identified Yeast proteins and (d) low abundance Yeast proteins.



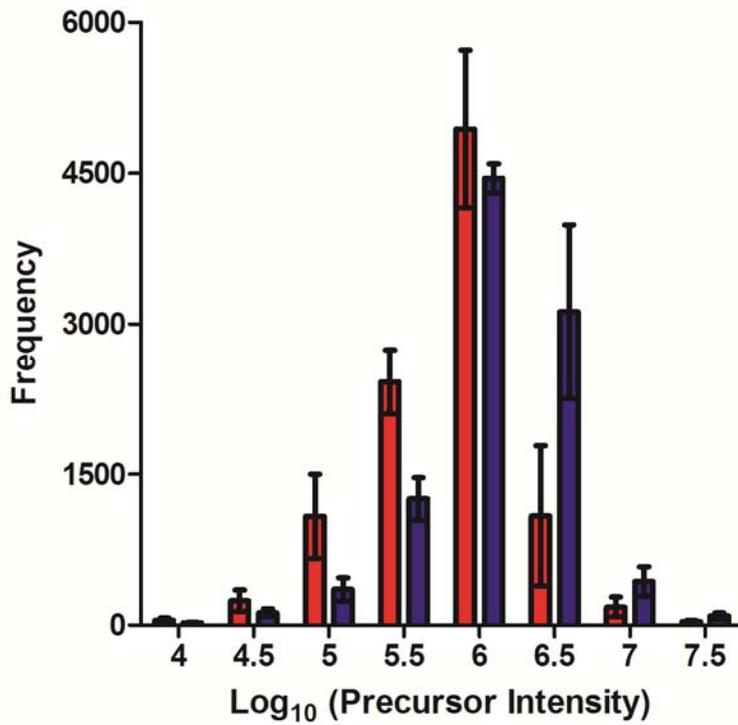
Supplementary Figure 3: Analysis of digestion and depletion of peptides from a yeast lysate. (a) Volcano plot of the \log_2 average peptide spectral count ratio between triplicate DigDeAPr and Control runs plotted against the p-value. Data points are plotted based on average spectral counts from triplicate Control runs: 1-9 spectral counts (black), 10-99 (green), and 100-999 (magenta) spectral counts. (b) Correlation of peptide spectral count changes to absolute protein abundance in yeast. The data was fit using a linear regression. The best fit line and 95% confidence interval are plotted. Box whisker plots of peptide spectral count changes for (c) all identified yeast peptides and (d) low abundance yeast peptides plotted binned according to their protein abundance.



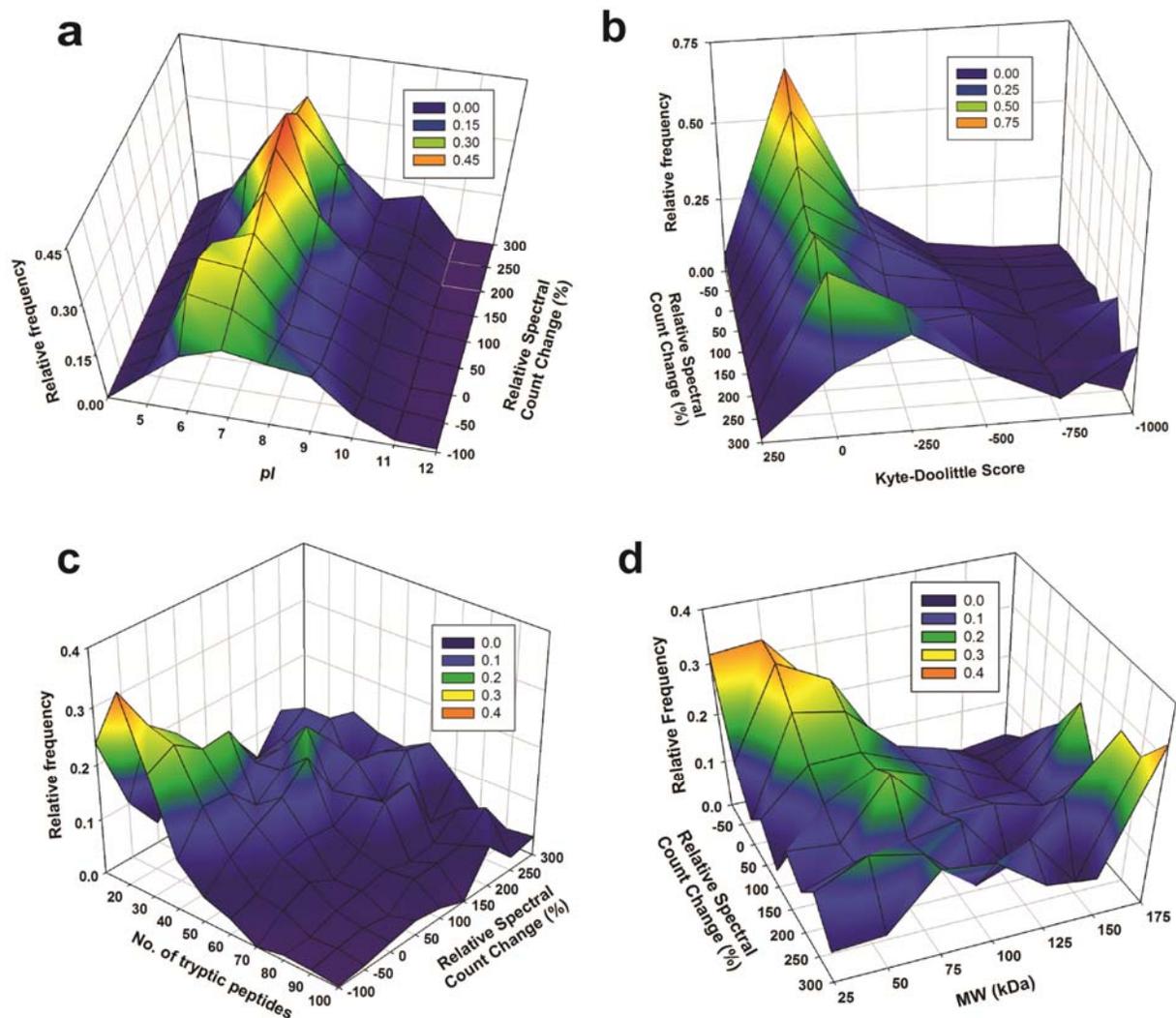
Supplementary Figure 4: Comparison of proteins and peptides from triplicate Control and DigDeAPr runs of HEK lysates. Venn diagrams for total (a) proteins and (b) peptides among triplicate Control and DigDeAPr runs. Venn diagrams for proteins among triplicate (c) Control and (d) DigDeAPr runs. Venn diagrams for peptides among triplicate (e) Control and (f) DigDeAPr runs.



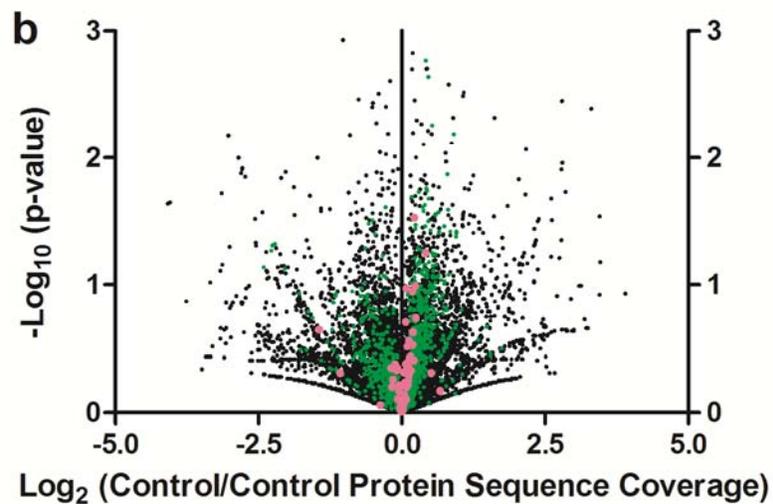
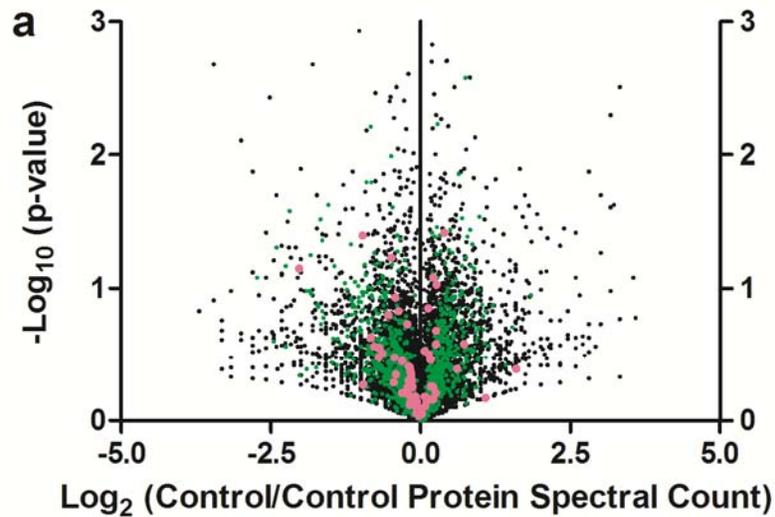
Supplementary Figure 5: Statistical comparison of peptide quality scores from HEK cell lysates. Volcano plots of the \log_2 ratio of the average (a) peptide XCorr and (b) peptide ΔCN between triplicate DigDeAPr and Control runs plotted against the p-value. Data points are plotted based on average spectral counts from triplicate Control runs: 1-9 spectral counts (black), 1-99 spectral counts (green), and more than 100 spectral counts (magenta). XCorr is a measure of the spectral matching quality between theoretical and experimental spectra. ΔCN is the difference between XCorr values of the 1st and 2nd candidate peptide sequences and is an indicator of peptide spectrum match correctness.



Supplementary Figure 6: HEK cell lysate peptide precursor intensity histogram comparison for triplicate Control (red) and DigDeAPr (blue) runs with error bars representing standard deviation. A systematic increase in peptide precursor intensity was found for all peptides identified in DigDeAPr runs relative to Control runs.



Supplementary Figure 7: Surface plots of HEK protein physicochemical properties based on relative spectral count abundance changes. Histograms for the relative frequency of protein (a) isoelectric point, (b) Kyte-Doolittle Score, (c) number of tryptic peptides, and (d) molecular weight were determined for each relative spectral count change bin. Only proteins with greater than 5 spectral counts in either Control or DigDeAPr runs were considered.



Supplementary Figure 8. Statistical comparison of proteins between two separate triplicate Control analyses of tryptic HEK lysates. Volcano plots of the log_2 ratio of the average from replicate triplicate Control runs for (a) protein spectral counts and (b) protein sequence coverage plotted against the p-value. Data points are plotted based on average spectral counts from triplicate Control runs: 1-9 spectral counts (black), 10-99 (green), and more than 100 (magenta).

Supplementary Note 1. Derivation of Michaelis-Menten equation to describe abundance-based proteome digestion.

The selective digestion of high abundance proteins and not low abundance proteins from the proteome requires specific conditions. Qualitatively, we knew that these conditions would be both diffusion-limited and trypsin-limited to ensure that trypsin would digest proteins at a controllable rate, based on the rate at which it forms a complex with abundant proteins. Since the complex formation rate is in part defined by K_M , we also knew the protein concentration relative to K_M would be an important parameter to consider. To quantitatively estimate the appropriate conditions for selective digestion of high abundance proteins we first derived an equation from the Michaelis-Menten equation to describe the digestion of low abundance proteins in a proteome from a complex cell lysate.

For any enzyme reaction dependent on formation of a complex between the enzyme (E) and substrate (S) to form products (P)



the velocity (v) or rate of product formation ($d[P]/dt$) can be defined by the rate constant k_2 and concentration of the Michaelis enzyme:substrate complex ($[ES]$), assuming the complex has reached steady-state equilibrium ($d[ES]/dt = 0$):

$$v = \frac{d[P]}{dt} = k_2[ES] \quad (2)$$

The enzyme:substrate complex is commonly defined by the total enzyme concentration ($[E_T]$), substrate concentration ($[S]$), and Michaelis constant (K_M) as:

$$[ES] = \frac{[E_T][S]}{K_M} \quad (3)$$

Inserting equation 3 into equation 2 and defining the maximum reaction velocity (v_{max}) as $[E_T]k_2$ yields the Michaelis-Mention equation:

$$v = \frac{v_{max}[S]}{K_M + [S]} \quad (4)$$

These derivations are well known. This form of the equation can be easily used to define the rate of digestion of one protein substrate at different concentrations of substrates relative to K_M . In this equation, the relative relationship between the substrate concentration and the Michaelis constant are important. Most simply, when K_M and the substrate concentration ($[S]$) are of the same order, equation 4 defines the velocity of the reaction (v). When a reaction is performed where $[S]$ is much greater than K_M , such as with highly abundant proteins, the K_M drops out of the equation and the $[S]$'s cancel to yield a velocity that is the maximum velocity:

$$v_{[S] \gg K_M} = V_{max} \quad (5)$$

Conversely, when the substrate concentration is less than K_M , such as when a protein is of low abundance in a lysate, the $[S]$ in the denominator drops out and the velocity is equal to maximum velocity times the ratio of the substrate concentration and the Michaelis constant:

$$v_{[S] \ll K_M} = V_{max} \frac{[S]}{K_M} \quad (6)$$

As stated, during a protease digestion highly abundant proteins will be digested at a rate defined by equation 5 or 6, depending on whether they are above or below the K_M . Note that under typical digestion condition for shotgun proteomics, where the protein:protease ratio is ~100:1, a majority of the proteins will be digested at V_{max} . Unlike high abundance proteins, low abundance proteins will not be digested at a rate defined by equation 5 or 6. That is, low abundance protein digestion velocities should be defined by both their concentrations and the total protein concentration of the lysate. Thus, we derived an equation to describe the rate of digestion of individual low abundance proteins within a lysate when their concentrations are below the proteases K_M . The K_M of sequence grade trypsin is ~3 μM ,¹ well above the concentration of low abundance proteins (< 1 nM) in a common protein digestion concentration of 1 mg/mL (~10 μM). We started with the equation generated from inserting equation 3 into equation 2:

$$v = \frac{k_2[E_T][S]}{K_M + [S]} \quad (5)$$

Under the conditions described, the total enzyme concentration ($[E_T]$) is no longer available for formation of a complex with low abundance proteins since it is in a complex with high abundance proteins. Thus we can substitute $[E_T]$ for the free protease concentration ($[E]$) and the individual low abundance protein concentration ($[P_i]$) for the substrate concentration ($[S]$). Further, since the protein concentration is much lower than K_M , the $[P_i]$ in the denominator drops out of the equation, yielding:

$$v_{[P_i] \ll K_M} = \frac{k_2[E][P_i]}{K_M} \quad (6)$$

The k_2 and K_M can be separated out and the total enzyme concentration ($[E_T]$) can be substituted by the difference of the total enzyme concentration and the protease:protein complex concentration ($[E_T] - [EP_T]$). Note the total protein (P_T) denotation to illustrate the complex concentration is defined by the total protein concentration and not the individual low abundance protein concentration (P_i).

$$v_{[P_i] \ll K_M} = \frac{k_2}{K_M} ([E_T] - [EP_T])[P_i] \quad (7)$$

The protease:protein complex concentration can be replaced by equation 3 yielding:

$$v_{[P_i] \ll K_M} = \frac{k_2}{K_M} \left([E_T] - \frac{[P_T][E_T]}{K_M + [P_T]} \right) [P_i] \quad (8)$$

A common denominator is created for $[E_T]$ by multiplying the numerator and denominator by $K_M + [P_T]$:

$$v_{[P_i] \ll K_M} = \frac{k_2}{K_M} \left(\frac{[E_T](K_M + [P_T])}{K_M + [P_T]} - \frac{[P_T][E_T]}{K_M + [P_T]} \right) [P_i] \quad (9)$$

The numerator is further expanded:

$$v_{[P_i] \ll K_M} = \frac{k_2}{K_M} \left(\frac{[E_T]K_M + [E_T][P_T]}{K_M + [P_T]} - \frac{[P_T][E_T]}{K_M + [P_T]} \right) [P_i] \quad (10)$$

and subtracted using the common denominator:

$$v_{[P_i] \ll K_M} = \frac{k_2}{K_M} \left(\frac{[E_T]K_M + [E_T][P_T] - [P_T][E_T]}{K_M + [P_T]} \right) [P_i] \quad (11)$$

The two $[E_T][P_T]$ terms subtract out leaving :

$$v_{[P_i] \ll K_M} = \frac{k_2}{K_M} \left(\frac{[E_T]K_M}{K_M + [P_T]} \right) [P_i] \quad (12)$$

The K_M 's cancel leaving:

$$v_{[P_i] \ll K_M} = \frac{k_2[E_T]}{K_M + [P_T]} [P_i] \quad (13)$$

The numerator is the definition of V_{max} , simplifying the equation to:

$$v_{[P_i] \ll K_M \approx [P_T]} = \frac{V_{max}[P_i]}{K_M + [P_T]} \quad (14)$$

This equation takes the same form of the general Michaelis-Menten equation. However, the substrate concentration for an individual protein ($[P_i]$) and the total protein substrate concentration ($[P_T]$) now together define the velocity of digestion for the individual protein (P_i). Further, the equation clearly defines that if the total protein concentration ($[P_T]$) is on the same order as K_M , that the individual protein digestion velocity is a function of the sum of K_M and the total protein concentration ($[P_T]$). Under conditions where the total protein concentration ($[P_T]$) is greater than K_M and the individual protein concentrations $[P_i]$ are still less than K_M , then the velocity for digestion of individual low abundance proteins is:

$$v_{[P_i] \ll K_M \ll [P_T]} = V_{max} \frac{[P_i]}{[P_T]} \quad (15)$$

Through inspection of equations 14 and 15 it becomes apparent that in order to achieve selective increases to the digestion velocity of abundant proteins over low abundant proteins, the total protein concentration ($[P_T]$) should not be less than K_M . That is, the digestion velocity of abundant proteins will approach V_{max} and be proportional to the mole percentage of the protein within the proteome. From equation 14, the digestion velocity for low abundance proteins will be proportional to their concentration ($[P_i]$) and

inversely proportional to the sum of the K_M and the total protein concentration ($[P_T]$). Further, the extent to which $[P_T]$ is greater than K_M will define the number of proteins within a proteome that have a concentration also equal to or greater than K_M . Equations 14 and 15 illustrate that the adequate conditions for digestion depletion are dependent on five parameters. Two of the parameters are biologically defined constants – the K_M of the protease and the protein abundance dynamic range of a proteome ($[P_T] - [P_i]$), assumed to be greater than 10^6 for the HEK lysate.^{2,3} The other three parameters are biochemical variables - the protein and enzyme concentrations and the time of digestion. The next section (**Supplementary Note 2**) describes the estimation and evaluation of the three biochemical variables.

Supplementary Note 2. Estimation and evaluation of initial DigDeAPr conditions.

As with most analytical work, we estimated that an order-of-magnitude adjustment in protein dynamic range was necessary to see a noticeable improvement in proteomic metrics. Thus we wanted to digest ~90% of the proteome away. This defined a reasonable starting protein mass of 1 mg since 100 μg of protein lysate is regularly used for protein analysis with MudPIT. Our protein lysates were around 10 mg/mL, thus our digestions could be performed at a concentration of 10 mg/mL or lower. Somewhat arbitrarily assuming an average molecular weight of proteins in a lysate to be 67 kDa, this would yield protein concentrations less than 150 μM . Assuming we want high abundance proteins to be above the K_M of trypsin and low abundance proteins to be below it described in the previous section, we chose 1 mg/mL, corresponding to 15 μM (five times greater than the 3 μM K_M of trypsin).¹ Next we estimated a trypsin concentration based on a reasonable digestion time (8 hrs). In order to digest 90% of proteins at a concentration of 15 μM in 8 hours, a V_{max} of 500 pM s^{-1} would be required. Thus, an adequate trypsin concentration can be calculated by dividing the V_{max} by the k_{cat} of 0.44 s^{-1} ,¹ yielding a trypsin concentration of 1 nM or 25 ng/mL. These were the initial conditions we tested which also ultimately gave some of the best results. However, we also tested longer digestion times (12, 18, and 24 hr), different trypsin concentrations (2.5 ng/mL, 10 ng/mL, and 100 ng/mL), and a 10-fold lower 1 mg HEK protein lysate concentration (100 $\mu\text{g}/\text{mL}$).

Since the mass spectrometry-based proteomic analysis of all the initial digestion depletion optimization conditions would have been excessively time consuming, we

focused first on the mass balance of proteins during the course of digestion depletion (Online Methods). We would suggest using a similar mass balance strategy for implementing DigDeAPr on other sample types where the protein abundance dynamic range may be different. Time point analysis of the digestion depletion progress at fixed trypsin and lysate concentrations was the most informative optimization experiment we performed. It provided a confirmation of the kinetics of the digestion depletion and an adequate timeframe for which other concentration modifications could be performed. With varied digestion depletion time the progress of the depletion was established as either incomplete (< 75%), complete (75 – 95%), or over-depleted (> 95%), based on protein mass depletion measured by BCA mass balance. Appropriate concentration and/or digestion time changes were made. For instance, if the sample was over-depleted the trypsin concentration or digestion depletion time were reduced 50%. Incomplete depletion could be increased by doubling either of these parameters.

The less straightforward, and potentially most important, optimization condition was selection of protein lysate concentration. An optimum lysate concentration is likely to be different even for human cell lines. For instance, from our previous analysis of cancer-derived HeLa cells with hexapeptide bead protein abundance adjustment we found 7 proteins with greater than 1000 spectral counts and another 8 proteins with greater than 500 spectral counts.⁴ While with neuron-like HEK cells in this study, we only found one protein with close to 1000 spectral counts and 5 proteins around 500 spectral counts (which were the same as the HeLa proteins with greater than 1000 spectral counts). The protein lysate concentration defines which portion of the proteome is above (depleted), equal to (partially depleted), or below (non-depleted) the

K_M of trypsin. Thus, in order to optimize the protein lysate concentration, proteomic analysis was performed and would be suggested for adopters of this method since it provides protein-specific abundance changes not evident with the mass balance approach. After proteomic analysis if all proteins appear over-depleted (including low abundance ones), decrease the lysate concentration. This shifts more individual protein concentrations below the K_M of trypsin, focusing the depletion on the highest abundance proteins. If the sample is under-depleted (i.e. too few high abundance proteins are depleted), the lysate concentration can be increased. Conversely, this concentration change will shift more individual protein concentrations above the K_M of trypsin, expanding the depletion range of abundant proteins. These concentration changes may also affect the rate of digestion depletion, so the appropriate changes to either protein mass, trypsin concentration, or digestion depletion time may be necessary. Fortunately, these adjustments can be monitored by mass balance without more time consuming mass spectrometry experiments.

Supplementary Note 3. Validation of digestion and depletion of abundant proteins with known yeast protein copy numbers

To directly correlate protein spectral count changes to protein abundance changes we also performed triplicate Control and DigDeAPr runs on a yeast lysate. The analysis of log-phase yeast allows for direct comparison of spectral counts to absolute protein copy numbers per cell (**Supplementary Fig. 1**) determined by global western blot analysis.⁵ In this case, we use protein copy numbers to correlate spectral count changes to protein abundance changes. We found similar changes in yeast protein abundance (**Supplementary Fig. 2a**) as HEK proteins (**Fig. 2a**) based on spectral counting using the same digestion depletion conditions for the yeast lysate as the HEK cell lysate. Further, similar trends were also observed at the peptide level for yeast proteins (**Supplementary Fig. 3a**) as with the HEK peptides (**Fig. 2e**). This highlights the robustness of DigDeAPr since mammalian and yeast cell lysates have significantly different numbers of proteins (~6,750 versus >10,000) and protein abundance profiles (10^6 versus 10^7).^{2, 3, 5} Nevertheless, improvements in protein identifications and sequence coverages from DigDeAPr were more dramatic due to the higher complexity and dynamic range of the HEK proteome. For the yeast lysate data, when spectral count changes of yeast proteins are plotted versus their absolute abundance it is evident that the higher abundance proteins and peptides are selectively depleted as expected (**Supplementary Figs. 2b and 3b**). These results are further highlighted when the data is represented in box-whisker plots for all (**Supplementary Figs. 2c and 3c**) and lower abundance (**Supplementary Figs. 2d and 3d**) proteins and peptides.

Supplementary Note 4. Improvements to peptide quantitation metrics with DigDeAPr

A comparison of the protein spectral count relative standard deviations (RSDs) between Control and DigDeAPr runs (**Fig. 2f**) can be used to assess the capabilities of DigDeAPr for spectral counting-based quantitation. High abundance protein spectral counting quantitation can be easily performed without DigDeAPr, thus the comparison of lower abundance protein spectral count RSDs are more relevant for applying DigDeAPr for quantitation. For 10-99 spectral counts, median RSDs were 0.30 (Control) and 0.33 (DigDeAPr) and for 1-9 spectral counts, 0.43 and 0.46, respectively. The similarity of RSDs within these spectral count ranges indicates DigDeAPr should be just as precise as routinely used shotgun proteomic spectral counting methods. The higher RSDs for low spectral count proteins in both Control and DigDeAPr runs highlight the potential quantitation gains of DigDeAPr, as the methodology shifts low abundance proteins into more reproducible, higher spectral count ranges (i.e. 10-100 spectral counts). Additionally, these results indicate that our digestion depletions with MWCO spin-filters were within the spectral counting error of data-dependent shotgun proteomics. Despite the obvious expected changes to spectral counts of proteins with DigDeAPr, it should be a viable method for spectral counting-based protein quantitation.

For peptides identified by both Control and DigDeAPr runs, we also investigated their changes in precursor intensity, an important factor for fragmentation spectra quality and quantitation. We found dramatic, statistically significant increases in precursor intensity from DigDeAPr for all peptide abundances in both Control and DigDeAPr runs (**Fig. 3a**) and for all peptides identified (**Supplementary Fig. 6**). Increased precursor intensities also directly led to greater chromatographic peak areas (**Fig. 3b**), as they are

theoretically proportional and their trends are quite similar. Precursor intensity and peak area are both relevant to label-free and metabolically-labeled protein quantitation methods. However, the potential quantitation gains may be best illustrated by improved peptide precursor S/N (**Fig. 3c**). Slightly more than half of the peptides (13,358 out of 23,932) with calculated ratio changes had low S/N (< 20) in Control runs. Notably, 79% of these low S/N peptides had higher S/N from employing DigDeAPr. Further, of the 6,442 peptides with S/N less than the limit of quantitation (LOQ = S/N > 10) in Control runs, 89% of these were raised above the LOQ threshold with DigDeAPr. Presumably these S/N gains would increase the number of total peptides that could be used for quantitation by ~25%, corresponding mostly to low abundance proteins. Conversely, peptides with high S/N in Control runs had minimal, acceptable decreases in S/N from digestion depletion. That is, of the 10,574 peptides with S/N greater than 20, only 5% (562) fell below the LOQ with an average S/N decrease of 18%. Thus, these precursor intensity, peak area, and S/N gains should improve accuracy, precision, and comprehensiveness of MS-based quantitation.

Multiple existing and developing quantitative methods also employ MS/MS-based quantitation. Data-independent acquisition (DIA) methods for fragmenting, identifying, and quantifying peptides are more reproducible and sensitive than traditional data-dependent acquisition (DDA) of peptide precursors⁶ and with other similar methods emerging,^{7, 8} is expected to become the norm. With DIA, peptide fragment ion intensities that contribute to the peptide identification can be summed within an MS/MS spectra and used to generate a reconstructed chromatogram from successive identifications with better S/N than from MS precursor ion spectra. Thus, we also

investigated the changes in summed MS/MS ion intensities (**Fig. 3d**) from the use of DigDeAPr. As with MS precursor intensities, all summed MS/MS intensities had similar gains to precursor intensity across all peptide abundances. Since we used DDA in this work, we could not reconstruct chromatograms from successive MS/MS spectra of the same peptide as in DIA. However, the MS/MS signal gains should translate very similarly to those in the MS scan where high precursor intensity yields both greater peak areas and higher S/N for quantitation. Additionally, although we haven't directly investigated how reporter-based MS/MS quantitation would be affected by DigDeAPr, the higher precursor intensities and S/N we found in the MS scans are also an indicator of potential gains in reporter ion intensities and S/N. Thus, DigDeAPr should ultimately also improve MS2-based reporter ion and data-independent quantitation methods.

Supplementary Note 5. Comparison of dynamic range compression mechanisms between DigDeAPr and hexapeptide-based ProteoMiner

We previously showed that freeing of chromatographic and ion trap space using hexapeptide equalization of proteins leads to more new peptide identifications through sampling of higher precursor intensities for low abundance peptides.⁴ This also appears to be the case for DigDeAPr as precursor intensities increased over all peptide abundances (**Fig. 3a**). These results were not surprising as the mechanisms for protein abundance adjustment using hexapeptide ligand libraries and DigDeAPr are similar, relying on either formation of protein:hexapeptide complexes with an affinity constant (K_A) or the formation of protein:protease complexes with a Michaelis constant (K_M), respectively. However, DigDeAPr appears to be less sensitive to protein isoelectric point (**Supplementary Fig. 7a**), with a nearly uniform relative distribution of isoelectric points over all spectral count changes, and is similarly insensitive to protein hydrophobicity (**Supplementary Fig. 7b**). Thus, digestion depletion may be more versatile than hexapeptide depletion and easier to implement than other common targeted depletion methods such as antibody depletion arrays. The advantages of relying on a single protease K_M instead of many different protein-ligand K_A 's are (1) the selectivity differences are almost entirely based on protein abundance, not affinity and (2) the K_M is less biased for a defined set of proteins as with a hexapeptide ligand library. Protein size may provide a slight bias with DigDeAPr since the number of tryptic sites generally scale with protein molecular weight (**Supplementary Fig. 7c-d**), but protein abundance remains the dominant depletion factor.

Supplementary Note 6. Analysis of potential protein loss from the MWCO spin-filter

In addition to the expected abundance changes, some low spectral count proteins (< 10 spectral counts) also appeared depleted from DigDeAPr. Of these proteins, ~ 50 were consistently depleted (\log_2 ratio < -1 and $p < 0.05$) in comparison to Control runs. However, when we examined this small number of proteins' physicochemical properties, their isoelectric point, molecular weight, and hydrophobicity appeared equally distributed in comparison to all the proteins identified in Control runs. Overall, the low abundance protein spectral count decreases could be attributed to aspects of the DigDeAPr methodology, such as losses from the MWCO spin-filter, but without an obvious physicochemical trend it appears unlikely. Most of the low abundance protein depletion trend can likely be attributed to the pseudo-random sampling of low abundance peptide precursor ions in shotgun proteomics⁹ and the inability to measure losses of low abundance proteins with data-dependent acquisition. This is illustrated by the comparison of two separate sets of triplicate Control analyses (Control versus Control). The comparison shows uniform ratio distributions for spectral counts and sequence coverages for all abundances (**Supplementary Fig. 8**), unlike DigDeAPr versus Control comparisons (**Fig. 2c-d**) where high abundance proteins are skewed towards decreasing ratios and low abundance proteins towards increasing ratios. This comparison not only further validates the spectral count changes observed with DigDeAPr, but also illustrates the expected variance in spectral counts based on abundance. That is, the low abundance proteins which appear depleted by DigDeAPr actually fall within the variance of spectral counting comparison. The spread of spectral count ratios is also represented as spectral count relative standard deviations (RSDs)

for Control and DigDeAPr runs (**Fig. 2f**). Comparison of these RSDs allows for the evaluation of the reproducibility of our digestion depletions. We found independently performed digestion depletions of HEK lysates, estimated by BCA mass balance analysis to be $85 \pm 10\%$ depleted, resulted in similar spectral count RSDs as Control runs for proteins with less than 100 spectral counts. Proteins with greater than 100 spectral counts had slightly higher median RSDs for DigDeAPr (0.32) versus Control (0.24). These RSDs directly illustrate the variation of the digestion depletion since high abundance proteins should be most affected by these steps. Thus, the 8% higher RSD than Control runs is consistent with our ability to measure the digestion depletion with BCA mass balance to $\pm 10\%$.

References

1. Finehout, E.J., Cantor, J.R. & Lee, K.H. *Proteomics* **5**, 2319-2321 (2005).
2. Mense, S.M. et al. *Physiological genomics* **25**, 435-449 (2006).
3. Schwanhausser, B. et al. *Nature* **473**, 337-342 (2011).
4. Fonslow, B.R. et al. *Journal of proteome research* **10**, 3690-3700 (2011).
5. Ghaemmaghami, S. et al. *Nature* **425**, 737-741 (2003).
6. Venable, J.D., Dong, M.Q., Wohlschlegel, J., Dillin, A. & Yates, J.R. *Nature methods* **1**, 39-45 (2004).
7. Gillet, L.C. et al. *Mol Cell Proteomics* **11**, O111 016717 (2012).
8. Panchaud, A., Jung, S., Shaffer, S.A., Aitchison, J.D. & Goodlett, D.R. *Anal Chem* **83**, 2250-2257 (2011).
9. Liu, H., Sadygov, R.G. & Yates, J.R., 3rd *Anal Chem* **76**, 4193-4201 (2004).