A "Proteomic Ruler" for Protein Copy Number and Concentration Estimation without Spike-in Standards*⑤

Jacek R. Wiśniewski‡§¶, Marco Y. Hein‡§, Jürgen Cox‡, and Matthias Mann‡¶

Absolute protein quantification using mass spectrometry (MS)-based proteomics delivers protein concentrations or copy numbers per cell. Existing methodologies typically require a combination of isotope-labeled spike-in references, cell counting, and protein concentration measurements. Here we present a novel method that delivers similar quantitative results directly from deep eukaryotic proteome datasets without any additional experimental steps. We show that the MS signal of histones can be used as a "proteomic ruler" because it is proportional to the amount of DNA in the sample, which in turn depends on the number of cells. As a result, our proteomic ruler approach adds an absolute scale to the MS readout and allows estimation of the copy numbers of individual proteins per cell. We compare our protein quantifications with values derived via the use of stable isotope labeling by amino acids in cell culture and protein epitope signature tags in a method that combines spike-in protein fragment standards with precise isotope label quantification. The proteomic ruler approach yields quantitative readouts that are in remarkably good agreement with results from the precision method. We attribute this surprising result to the fact that the proteomic ruler approach omits error-prone steps such as cell counting or protein concentration measurements. The proteomic ruler approach is readily applicable to any deep eukaryotic proteome dataset— even in retrospective analysis—and we demonstrate its usefulness with a series of mouse organ proteomes. *Molecular & Cellular Proteomics 13: 10.1074/mcp.M113.037309, 3497–3506, 2014.*

Mass spectrometry (MS)[1] is now capable of analyzing the proteome to considerable depth, and more than 10,000 proteins have been reported in single mammalian cell types (1). In the past decade, MS-based proteomics has gone from sole identification to the quantification of proteins, which has typically meant relative quantification between samples (2–4). Apart from the presence of a protein and its relative fold changes between different conditions (5), it is often desirable to estimate absolute quantities such as molar concentrations or copy numbers per cell, which can be compared for different proteins (6). For instance, in systems biology, even a rough estimate of the copy number can help to establish initial parameters for simulation (7). Likewise, clinical protein measurements are typically done in absolute terms of titers, such as milligrams per deciliter. For this purpose various approaches have been utilized, including correlating total MS signals to visualized structures in the cell (8) and extrapolating from spiked-in reference protein mixtures (9) or from endogenous proteins quantified via accurately characterized, isotopically labeled peptide (10) or protein fragment standards (11). Absolute quantification is then achieved through quantification relative to a known reference. In all cases, results scale with the amount of input material or amount of spiked-in standard. Accurate protein concentration measurements are thus an essential and often limiting factor for overall accuracy. Commonly used dye-based protein determination methods rely on the reactivity of few amino acid residues—mainly tryptophan and tyrosine (12) in the case of the Lowry and BCA assays, or a hydrophilic/hydrophobic balance of the proteins in the case of Bradford reagent (13). Systematic errors of up to a factor of 2 may therefore arise from the selection of a non-optimal protein standard (14). An additional, often ignored source of errors is the cross-reactivity of the reagents with non-proteinaceous cell components such as thiols, nucleic acids, and phospholipids.

To convert protein quantities to copies per cell, all methods require knowledge of the number of cells used for the analysis. This can be obtained directly via cell counting or indirectly through knowledge of the total protein amount per cell, which in turn is a function of cell volume and total protein concentration. However, cells are not necessarily uniform; therefore scaling by cell numbers may be inaccurate, as a 25% variation

[1] The abbreviations used are: MS, mass spectrometry; SILAC, stable isotope labeling by amino acids in cell culture; PrEST, protein epitope signature tag; FASP, filter-aided sample preparation; PTM, posttranslational modification.

of the diameter of a sphere-shaped cell corresponds to a 2-fold change in cell volume. In tissues, not only are cell sizes variable, but visual counting of cells is also problematic. For instance, up to 5-fold differences in calculated cell volumes have been reported for enterocytes of the intestinal mucosa (15).

Any deviations in protein determination or cell counts will inevitably carry over to the final readout, even when very precise MS methods are used. This limits the overall accuracy, without showing up as a decrease in the precision of the quantification, as measured by standard deviations or coefficients of variation.

In the course of studying the colon cancer proteome, we recently devised a method for estimating absolute amounts of individual proteins or protein classes based on the proportion of their MS signals to the total MS signal (16). We termed the method the Total Protein Approach, because we relate this proportion to a total protein mass. To obtain copy numbers, we specifically used the total protein mass per cell, which needs to be determined or estimated separately.

In this study, we expanded the method by a concept we call the "proteomic ruler" to further allow correct absolute scaling of the readout without additional steps. We made use of the defined amount of genetic information in each cell, encoded in a known amount of DNA. We show that an accurate determination of the DNA content in a proteomic sample helps to directly determine the number of cells. We then demonstrate that the MS signal derived from histones, around which DNA is wrapped in a defined ratio, can be used as a natural standard in a whole proteome dataset. It serves as a proteomic ruler that allows the estimation of total protein amounts per cell. Thereby the quantitative readout can be absolutely scaled to copies per cell without the need for cell counting or protein concentration determination.

EXPERIMENTAL PROCEDURES

Plasma Lysate—The author's blood was capillary-collected via skin puncture of the middle finger. It was immediately supplemented with 0.05 M EDTA and centrifuged at 5000 × *g* for 1 min to separate blood cells from plasma. Plasma was diluted 10-fold with lysis buffer containing 0.1 M Tris-HCl, pH 8.0, 0.1 M DTT, and 2% SDS, and the mixture was incubated at 70 °C for 5 min.

Whole Cell and Tissue Lysates—U87-MG, A549, PC-3, and Hep-G2 cells were grown in DMEM supplemented with 10% FBS and 1% streptomycin. The cells were harvested at 70% confluence and dissolved in lysis buffer at 100 °C for 5 min. After being chilled to room temperature, the lysates were briefly sonicated to reduce the viscosity of the sample. Frozen mouse tissues (Pel-Freez, Rogers, AR) were homogenized with T10 basics Ultra-Turrax dispenser in the lysis buffer at a tissue-to-buffer ratio of 1:10. The homogenates were incubated at 100 °C for 5 min. Finally, the cell and tissue lysates were clarified by centrifugation at 16,000 × *g* for 10 min.

Protein Determination—Protein content was determined using a Cary Eclipse Fluorescence Spectrometer (Varian, Palo Alto, CA) as described previously (17). Briefly, aliquots of 1 to 3 μl of whole cell lysates were mixed with 2 ml of 8 M urea in 10 mM Tris-HCl, pH 8.5. The fluorescence was measured at 295 nm for excitation and 350 nm

for emission. The slits were set to 5 nm and 20 nm for excitation and emission, respectively. Tryptophan was used as a standard. The protein content was calculated from the following relationship: the fluorescence of 0.1 μg of tryptophan equals 9 μg of total protein, which reflects an average 1.1% weight content of tryptophan in whole lysates of human cells.

Cell Counting—Tissue cultures were trypsinized at 37 °C for 2 min, and the released cells were washed with PBS and collected at 1000 × *g* for 1 min. Then the pellets were suspended in PBS and the cells were stained with 0.2% Trypan Blue (Invitrogen). Cell counting was carried out on an automated cell counter (Countess, Invitrogen).

FASP-based Protein Processing—Aliquots of lysates containing 100 μg of total protein were processed according to the multi-enzyme digestion FASP protocol (18). Briefly, protein lysates were depleted from the detergent using 8 M urea in 0.1 M Tris/HCl, pH 8.5, thiols were alkylated with iodoacetamide, and proteins were consecutively digested with endoproteinase LysC and trypsin. Digests of plasma fractions were fractionated using a pipette tip strong anion exchange method into four and two fractions as described previously (19).

FASP-based Cleavage and Determination of RNA and DNA—After collection of the peptides released by trypsin, the material remaining in the filter was washed once with TE buffer (10 mM Tris-HCl, pH 8.0) and then was digested with 0.5 μl (0.5 U) of RiboShredder (Epicenter, Madison, WI) in 60 μl of TE buffer at 37 °C for 1 h to digest RNA. The released ribonucleotides were collected via centrifugation at 14,000 × *g*. Next the material on filters was washed twice with 80 μl of TE buffer, and then it was cleaved with 6 μg of DNAse (DN25, Sigma, St. Louis, MO) in 60 μl of 10 mM Tris-HCl, pH 7.8, containing 2.5 mM MgCl₂ and 0.5 mM CaCl₂ at 37 °C for 1 h. The obtained deoxynucleotides were collected via centrifugation. The RNA and DNA contents were determined by means of UV spectrometry using extinction coefficients of 0.025 and 0.030 (μg/ml)$^{-1}$cm^{-1} at 260 nm, respectively. The ratio of the spectral densities at 260 nm to 280 nm was ~2, indicating an absence of protein contamination that could contribute to A260 measurement.

LC-MS/MS and Data Analysis—Peptides were quantified by tryptophan fluorescence as described above, with the exception that the measurements were performed directly in 0.2 ml of 0.05 M Tris/HCl, pH 8.5, in 5 mm × 5 mm quartz cells. 4-μg aliquots of total peptide were loaded onto C₁₈ reverse phase columns (20 cm long, 75 μm inner diameter, in-house packed with ReproSil-Pur C₁₈-AQ 1.8-μm resin (Dr. Maisch GmbH, Ammerbuch-Entringen, Germany)) with buffer A (0.5% acetic acid). Peptides were eluted with a linear gradient of 5% to 30% buffer B (80% acetonitrile and 0.5% acetic acid) at a flow rate of 250 nl/min over 195 min. This was followed by 10 min from 30% to 60% buffer B, a washout of 95% buffer B, and re-equilibration with buffer A. Peptides were electrosprayed and analyzed on Q Exactive mass spectrometers using a data-dependent top-10 method with higher energy collisional dissociation fragmentation. Mouse organ samples were loaded onto a 15-cm reverse-phase column packed with 3-μm resin, separated over 320 min of gradient time, and analyzed on an LTQ Orbitrap mass spectrometer using collision-induced dissociation fragmentation. MS data were analyzed using the MaxQuant software environment (20), version 1.3.10.18, and its built-in Andromeda search engine (21). Proteins were identified by searching MS and MS/MS data against the human and mouse complete proteome sequences from UniProtKB (May 2013 version containing 88,820 and 50,807 sequences, respectively). Carbamidomethylation of cysteines was set as a fixed modification. N-terminal acetylation and oxidation of methionines were set as variable modifications. Up to two missed cleavages were allowed. The initial allowed mass deviation of the precursor ion was up to 6 ppm, and for the fragment masses it was up to 20 ppm (higher energy collisional dissociation, Orbitrap readout) and 0.5 Da (collision-induced dissoci-

ation, ion trap readout). The mass accuracy of the precursor ions was improved by time-dependent recalibration algorithms of MaxQuant. The "match between runs" option was enabled to match identifications across samples within a time window of 30 s of the aligned retention times. The maximum false peptide and protein discovery rates were set to 0.01. Protein matching to the reverse database and proteins identified only with modified peptides were filtered out. Protein abundances and copy numbers were calculated on the basis of summed peptide intensities of unique and "razor" peptides as reported by MaxQuant using the Perseus plugin described in this study. Finally, we removed all protein groups with fewer than two unique peptides (with the exception of two isoforms of creatine kinase in our plasma analysis), as they were less likely to yield highly accurate copy numbers.

Software Availability—The proteomic ruler Perseus plugin is available as a source code and as compiled binary from the Perseus website.

RESULTS

The Total Protein Approach Gives Accurate Estimates of Protein Concentrations—Using our Total Protein Approach, we previously demonstrated that a protein's abundance within the cell as a fraction of the total protein is reflected by the proportion of its MS signal to the total MS signal (16).

$$\frac{\text{Protein mass}}{\text{Total protein mass}} \approx \frac{\text{Protein MS signal}}{\text{Total MS signal}} \quad \text{(Eq. 1)}$$

This proportion can easily be extracted from any MS-based proteomics measurement, and its accuracy will improve with the depth of measurement. The value has to be scaled by a total protein mass, which can conceptually be the entire protein amount of a cell, the protein amount in a given volume of body fluid, or even a fixed unit such as 1 g. In that way we obtain the absolute amount of the protein or protein class per cell, per unit of volume, or per 1 g of total protein. To show that this principle is universally applicable, beyond the cell line and cancer tissue cases that we investigated before (16), we used it to estimate the concentrations of different diagnostically relevant proteins or protein classes in blood plasma after digesting plasma proteins using the FASP method (18). The total protein concentration in plasma varies around a typical value of 70 g/l within a narrow margin (22), so we scaled the MS readout by a total amount of 70 g to obtain grams per liter. We were able to quantify proteins within their expected physiological ranges over at least 5 orders of magnitude (Fig. 1, supplemental Table S1).

Nucleic Acid Quantification and Cell Counting via FASP-based Sample Preparation—In the case of a body fluid such as plasma, the total protein concentration is a readily accessible scaling parameter, and protein concentrations are meaningful and relevant. In the case of a cellular proteome, however, many applications require quantities of copies per cell, which necessitates cell counting. We wondered whether cell counting could be replaced by accurate DNA quantification when the genome size and ploidy were known. DNA concentration was shown to be proportional to the cell count
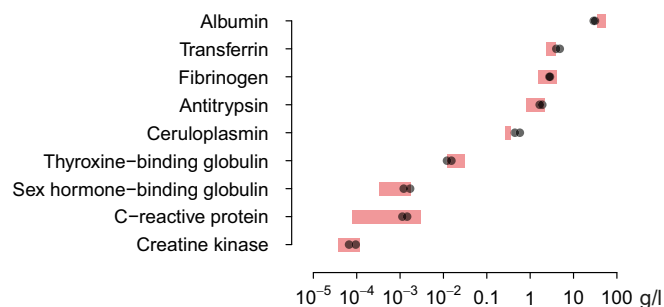

FIG. 1. **Analysis of protein abundances in human plasma using the Total Protein Approach.** Whole plasma was processed using the multi-enzyme digestion FASP approach with strong anion exchange peptide fractionation before LC-MS/MS analysis as described in "Experimental Procedures." Quantifications of selected target proteins are indicated as black dots; the reference values (red bars) are from Refs. 22 and 41. Two isoforms of creatine kinase were identified with one peptide each, for which we provide annotated MS/MS spectra in supplemental Fig. S1.

and was successfully used to normalize enzyme activities, transcript and protein amounts, and metabolome data (23–25). We hypothesized that DNA quantities could be measured directly from the proteomic sample, provided that the chromatin fraction was retained during sample preparation. In contrast to in-solution or in-gel approaches, the FASP method is reactor based (26) and allows sequential processing of the sample and separation of reaction products. Detergents are washed out at the beginning of the FASP procedure, and RNA and DNA, the major components remaining after protease digestion, can be cleanly released from the filter via RNase or DNase digestion (Fig. 2*A*). To test the feasibility of nucleic acid determination in the FASP format after digestion of proteins and elution of peptides, we consecutively digested the material retained on the filter with RNase and DNase. After each cleavage we collected the digestion products and determined their content based on UV absorbance at 260 nm. We observed a linear correlation between the amount of the eluted nucleotides and the amount of the sample. In parallel, we processed samples supplemented with defined amounts of purified calf thymus RNA and DNA. Yields were greater than 95% and were independent of the protein content (Fig. 2*B*), indicating that post-FASP digestion of a sample with DNase and RNase is a suitable method for determination of the RNA and DNA content in a proteomic sample that does not require additional preparative steps.

Next, we processed aliquots of total lysates prepared from counted numbers of four different human cell lines using two-step LysC/trypsin digestion of the proteins (multi-enzyme digestion FASP) (27). Both the starting protein amounts and the generated peptides were quantified. We then quantified the ribonucleotides and deoxyribonucleotides eluted after RNase and DNase treatment, respectively. The tryptic and LysC peptides obtained in the multi-enzyme digestion FASP-processed cell lysates (above) were analyzed in 4-h LC-
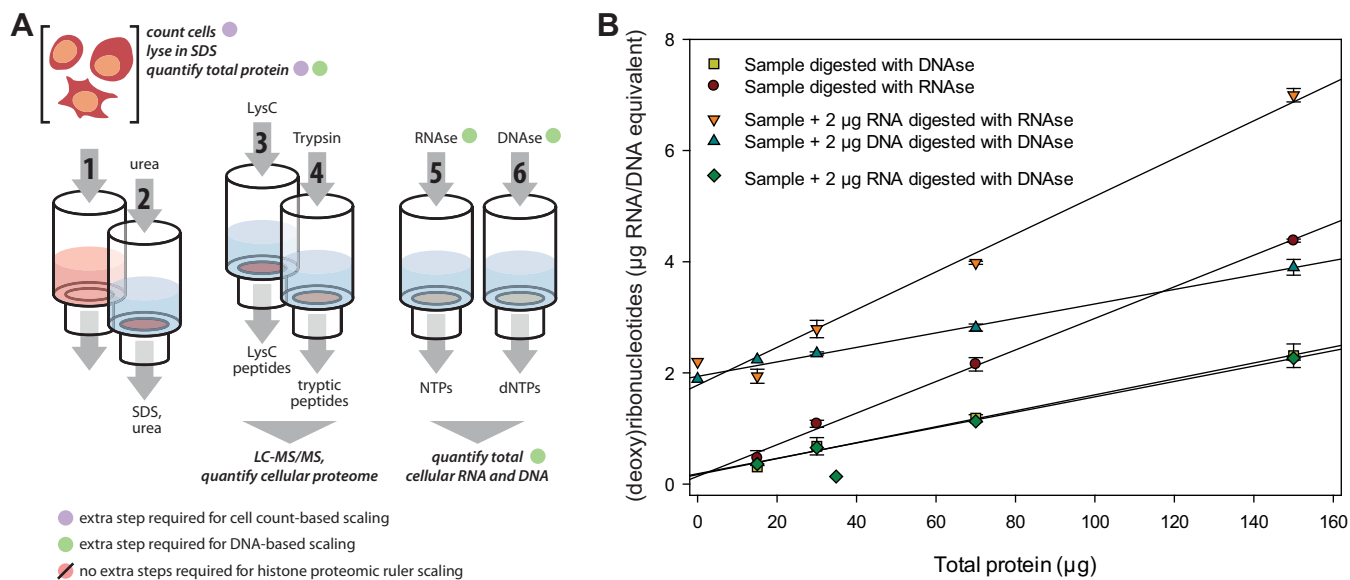
FIG. 2. *A*, the proteomic workflow. Cells were counted and lysed in a buffer containing SDS. Protein concentrations in the whole lysates were determined, and 100-μg aliquots of the whole lysates were successively processed in the proteomic reactor (FASP) format. After detergent removal, proteins were consecutively cleaved with endoproteinase LysC and trypsin. The released LysC and tryptic peptides were subjected to proteomic analysis. Next, RNA and DNA were digested, and the released ribo- and deoxyribonucleotides were spectrophotometrically quantified at 260 nm. Protein contents per single cell were calculated from the cell numbers and the protein concentrations. Alternatively, values of protein mass of single cells were obtained from DNA contents and the protein concentrations. *B*, determination of the efficiency and yield of RNase and DNase cleavages. Aliquots of mouse liver lysates were processed with the FASP method, and the residual high-molecular-weight material was sequentially cleaved with RNase and DNase (labeled "samples digested with DNase and RNase"). The released ribo- and deoxyribonucleotides were quantified spectrophotometrically at 260 nm. To demonstrate the completeness of digestion over the analyzed range, samples were supplemented with constant amounts of 2 μg of purified DNA or RNA prior to sample processing (labeled "samples + 2 μg RNA/DNA digested with DNase/RNase"). To demonstrate the specificity of the initial RNase digestion, samples were supplemented with DNA and digested with RNase (labeled "samples + 2 μg DNA digested with RNase").

MS/MS runs. In triplicate analyses, MaxQuant identified about 7000 proteins in each of the cell lines (supplemental Table S1). The human genome contains around 3.2×10^9 base pairs (28). Multiplying this number by the average mass of a base pair (615.9 Da) and by the ploidy of the respective cell type yields an expected amount of cellular DNA. We used a value of 6.5 pg for a diploid human cell to calculate cell numbers. Dividing the total amount of protein input by these cell numbers, we obtained a protein mass per cell that was very similar to that obtained by dividing the total protein input amount by the counted cell numbers (supplemental Table S2).

Histones Serve as a "Proteomic Ruler" for Absolute Scaling of Proteomic Data—In eukaryotic cells, DNA is packaged in chromatin by histones, and the mass of the DNA is about equal to the combined mass of histones (29). We therefore wondered whether the summed intensity of histones in a deep, eukaryotic proteome could serve as a proxy for the amount of DNA and therefore for the cell number. There are five major histone types, which are expressed in many isoforms and variants that are relevant for many aspects of chromatin biology. For our approach, however, we employed the summed MS signal of all histone-derived peptides, irrespective of which histone they mapped to or how they were assembled in protein groups. This value reflects the cumula-

tive histone mass. In this way, we used the MS signal of an entire class of proteins as a proteomic ruler and related it to a quantity that is not directly amenable to mass spectrometry. Our hypothesis of the histone proteomic ruler predicts the following relationship (Fig. 3*A*):

$$\frac{\text{Histone mass}}{\text{total protein mass}} \approx \frac{\text{Histone MS signal}}{\text{Total MS signal}}$$

$$\approx \frac{\text{Cellular DNA mass}}{\text{Cellular protein mass}} \quad \text{(Eq. 2)}$$

In our four-cell-line dataset, the histone MS signal amounted to 2.07% to 4.03% of the total MS signal. Equating this fraction with 6.5 pg as the DNA mass of diploid human cells, we obtained cellular protein masses within a factor of 1.24 ± 0.29 compared with the value obtained via cell counting (Fig. 3*B*; supplemental Table S2). This is close to the hypothesized value of 1 and implies that the ratio of histone MS signal to total MS signal allows the estimation of the total cellular protein mass without any additional measurements.

The error of the histone MS signal fraction depends on how accurately the histone MS signal and the total MS signal can be determined. For histones, a large number of various posttranslational modifications (PTMs) have been identified, lysine

FIG. 3. **Estimation of protein mass per cell using two biochemical approaches and the proteomic ruler method.** *A*, the histone proteomic ruler concept. The mass of cellular DNA is approximately equal to the protein mass of histones. Relating the histone MS signal to the total MS signal therefore allows one to estimate the protein mass per cells at a given cell ploidy and genome size. This method requires neither cell counting nor the determination of protein concentration. *B*, *C*, comparison of the values of total protein per cell obtained based on cell counting, DNA determination, and the histone proteomic ruler method. *D*, cell sizes obtained from retrospective analysis of published proteome datasets of CD4 or CD8a positive or double negative (DN) dendritic cell subtypes and plasmacytoid dendritic cells (pDCs) (36). All values represent the mean of two (cell counting) or three replicates (DNA and histone proteomic ruler quantifications) ± S.D.
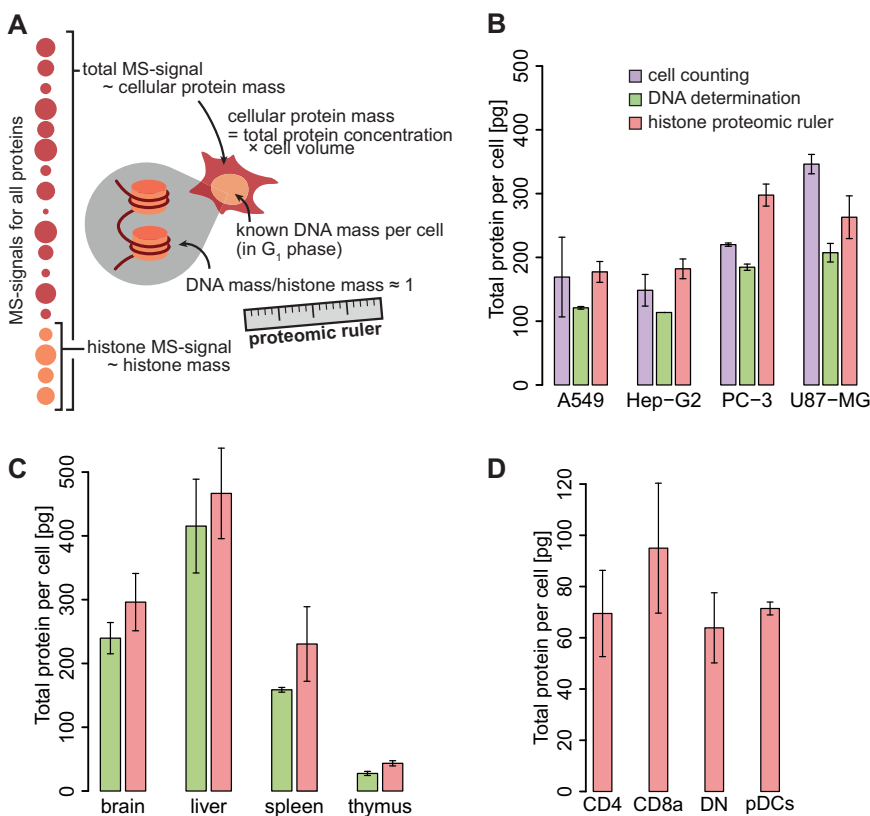
acetylation, serine and threonine phosphorylation, and lysine methylation being the most frequent. In most standard proteomics workflows, these modifications are not routinely included in the database search, and we were wondering whether this affects the ratio of histone MS signal to total MS signal, which is critical for our scaling approach. To address this question, we searched the data again with combinations of acetylation, phosphorylation, and methylation set as variable modifications. Although individual histones had changes in their relative abundances, in particular histone H3 (Figs. 4*A*–4*C*), the fraction of the cumulative histone to total MS signal changed only by 5% to 10% (Fig. 4*D*). This indicates that, with the exception of histone H3, the fraction of the MS signal derived from histone peptides that have PTMs is low and can be neglected in the overall data scaling process.

The accuracy of the total MS signal depends on the depth of the proteomic analysis. To estimate the required depth for a robust readout, we ranked all peptides by intensity and calculated the histone-MS fraction as a function of the number of identified peptides (Fig. 4*E*). Because peptide intensities span many orders of magnitude, the most intense peptides contribute a large part of the total intensity. Within the first few thousand peptides, the histone fraction is overestimated because histones contribute some of the most intense peptides. From a depth of around 12,000 or more peptides, however, the histone fraction stabilizes within tight margins. This depth of analysis is easily attainable with minimal sample

fractionation and also with single run analyses on latest-generation machines (30).

For each protein in the measured proteome, we can now estimate its mass per cell solely from its MS signal as the product of its MS signal fraction and the cellular protein mass. This value easily converts to copies per cell.

Protein copies per cell

$$= \frac{\text{Protein MS signal}}{\text{total MS signal}} \times \frac{N_A}{M} \times \text{cellular protein mass} \quad \text{(Eq. 3)}$$

$$= \text{protein MS signal} \times \frac{N_A}{M} \times \frac{\text{DNA mass}}{\text{Histone MS signal}}$$

where N_A is Avogadro's constant and M is the molar mass of the protein.

Ribosomal Proteins as a Proteomic Ruler for Cellular RNA—Next, we investigated whether the proteomic ruler concept is also applicable to cellular RNA. Ribosomal RNA typically represents about 80% of total RNA (31), and in eukaryotic ribosomes there is a ratio of about 1:1 between RNA and protein (32). The summed MS signal for all ribosomal proteins amounted to values between 3.61% and 5.27% of the total MS signal across the cell lines. We compared this result by the biochemical quantification of the total RNA content using the FASP method in relation to the total protein input (supplemental Table S2). Our results were within a factor of 1.01 ±
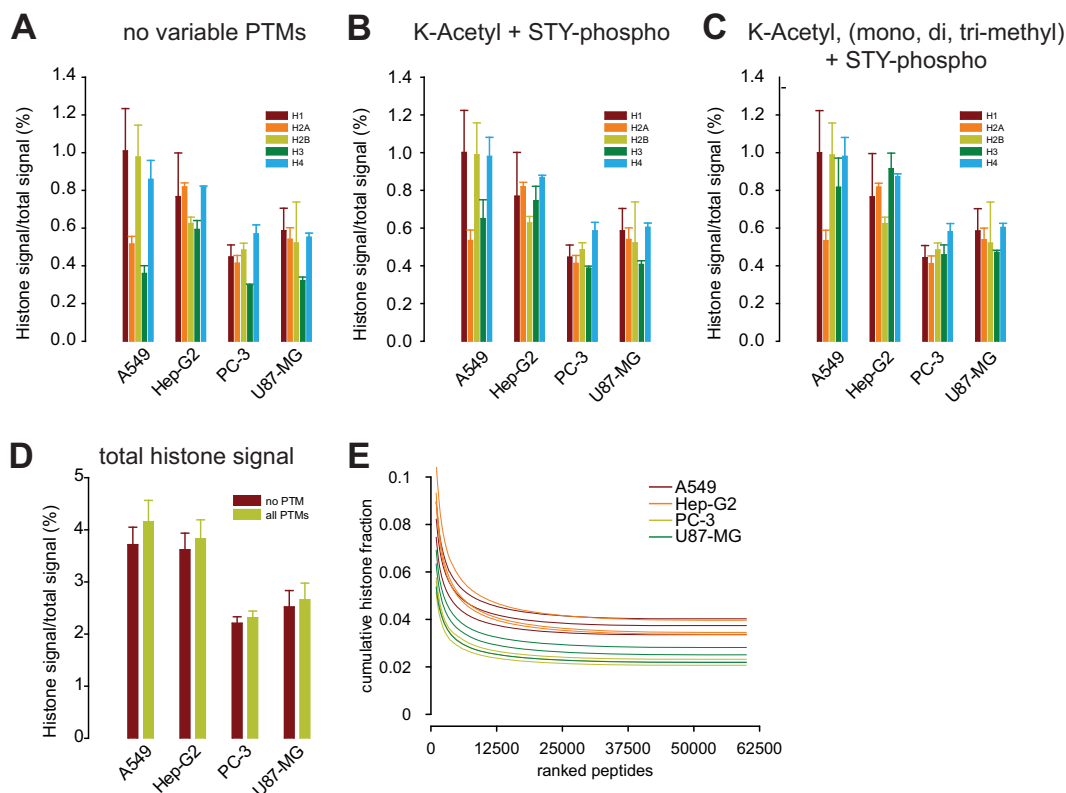
FIG. 4. **The contribution of PTMs to the estimated total protein content of histones.** Comparison of the fractions of the MS signals of individual histones, accumulated by histone type, derived by including different combinations of variable modifications in the database search. *A*, no variable PTMs (except for the default methionine oxidation and N-terminal acetylation). *B*, lysine acetylation and serine/threonine/tyrosine phosphorylation. *C*, lysine mono-, di-, and trimethylation in addition to the modifications searched in *B*. Comparison of the sum of all histone MS signals without PTMs (from *A*) and with all PTMs (from *C*). *D*, histone MS signal fraction as a function of the depth of analysis, simulated by intensity-based ranking of peptides.

0.13 of the biochemical measurements, indicating that the MS signal of ribosomal proteins can indeed be used as a proteomic ruler to estimate cellular RNA amounts.

Histone Proteomic Ruler Provides Estimates of Cell Sizes in Tissues—Counting cells in tissue samples is not trivial. However, determining the DNA and RNA content using our proteomic reactor format is equally straightforward as for cell lines. We prepared lysates from mouse brain, liver, and thymus; measured protein, RNA, and DNA contents; and performed proteomic analysis. There was excellent agreement between the total cellular protein mass values derived from the DNA-based method and our histone proteomic ruler approach (Fig. 3*C*; supplemental Table S3). This demonstrates that the histone proteomic ruler serves as a good proxy for estimating cellular protein masses in tissues.

The total cellular protein concentration typically lies within a range of 20% to 30% (w/v) (*i.e.* 200 to 300 g/l) in many cell types and organisms (33). This constraint can be used to convert between cellular protein mass and cell volume. Hepatocytes, the predominant cell type in liver, are roughly cubical cells with a 15-μm edge length (34). Assuming a total protein concentration of 200 g/l, this translates to 675 pg of protein

per cell. This compares to our estimate of 464 $\pm$ 35 pg total protein per average liver cell, which is reasonable given that non-hepatocytes contribute the same amount of DNA or histones but less overall protein mass. Thymocytes are at the other end of the size scale with an average volume of 250 μm^3 (35). This translates to 50 pg of protein, as compared with our estimate of 59 $\pm$ 31 pg.

To test the applicability of the histone proteomic ruler to the retrospect analysis of existing datasets, we reevaluated whole-proteome measurements of murine dendritic cell populations published by our group in 2010 (36). Samples had been prepared via one-dimensional SDS gel electrophoresis followed by in-gel digestion, an approach distinct from our FASP-based method and incompatible with direct DNA quantification from the proteomic sample. Mature dendritic cells have diameters between 10 and 15 μm (37). We compared these cell sizes to our proteomic ruler estimates that ranged between 64 $\pm$ 14 and 95 $\pm$ 25 pg total protein per cell for the different dendritic cell subtypes (Fig. 3*D*). These values translated to diameters of 8.5 to 9.7 μm for spherical cell shapes, which is expected to be slightly smaller than observed cell sizes, given the numerous dendrites projecting from the cell surfaces. Interestingly, our
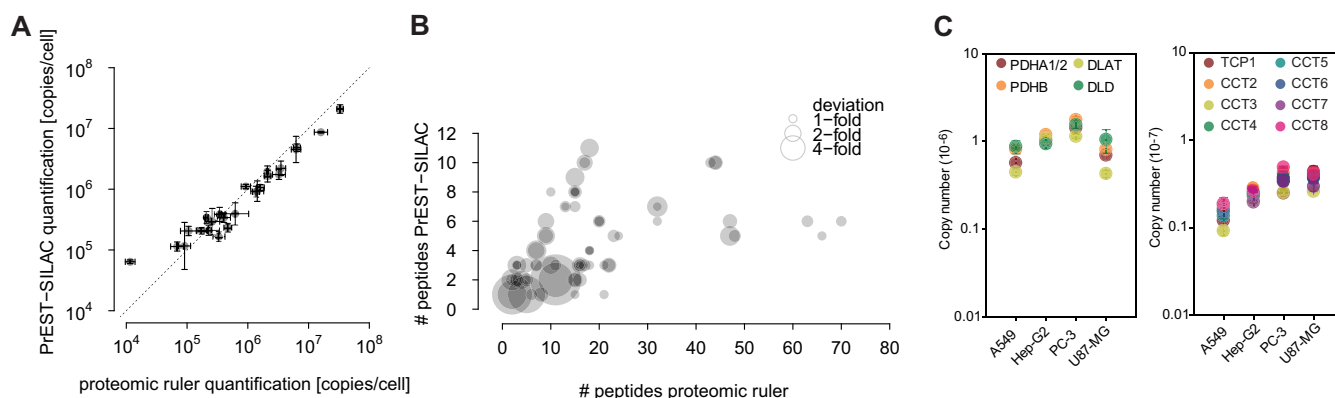
FIG. 5. **Comparison of absolute protein abundances calculated using the spike-in and proteomic ruler approaches.** *A*, comparison of protein copy numbers of selected proteins in HeLa cells obtained using spiked-in protein fragments (PrESTs) of known quantities and isotopic label quantification (11) to those calculated using the label-free histone proteomic ruler method. Values represent the mean of three replicates ± S.D. *B*, comparison of the numbers of peptides overlapping with the PrEST standard used for the SILAC quantification and the total number of peptides used for the proteomic ruler quantification. The deviations of the label-free values from the PrEST-SILAC values are represented as the sizes of the points. *C*, *D*, label-free protein copy number estimates correlate with the composition of protein complexes. *C*, pyruvate dehydrogenase complex. *D*, TRiC chaperonin.

observed similarities in cell sizes correlate with overall patterns of proteomic similarity on the level of individual proteins that were observed in the original study (36).

Label-free Copy Number Estimations Are Strikingly Close to Precise Spike-in Quantifications—We previously employed spiked-in protein epitope signature tags (PrESTs) of known quantities in combination with isotopic labeling, cell counting, and total protein concentration determination to obtain highly reliable copy number values of selected proteins (11). To assess the accuracy of our proteomic-ruler-derived protein copy numbers, we reanalyzed the same dataset used in the original PrEST-SILAC study and applied our calculations on the "heavy" labeled proteome without considering the ratio information from the "light" PrEST peptides. We recapitulated not only the correct scaling of the total protein mass, but also the copy numbers of the individual PrEST-quantified proteins within an average deviation of 1.5-fold (Fig. 5*A*; supplemental Table S4) and comparable precisions judged by the standard deviations from three replicates. We attribute the surprisingly good performance of the proteomic ruler quantifications to the fact that our label-free quantification on average made use of 19.4 peptides along the entire length of the proteins, whereas the PrEST-SILAC quantification used 4.7 peptides on average. This might compensate for some of the principal limitations of the label-free approach. Looking at the deviations of individual quantifications, we saw that the minority of larger deviations occurred exclusively with PrEST-SILAC quantifications based on two or fewer peptides or label-free quantifications based on 11 or fewer peptides (Fig. 5*B*). This observation underlines the benefits of approaches that rely on multiple independent quantifications instead of single peptide ratios, as commonly used, for example, with AQUA peptides. We conclude that for those proteins quantified with more than a few peptides, the proteomic ruler approach could offer a

surprisingly high level of accuracy, making it an attractive alternative to label-based methods.

In addition to the comparison with spike-in quantification data, macromolecular complexes offer another option for validating protein copy numbers. Many obligate protein complexes are well characterized in terms of their composition and stoichiometry with subunits expressed at equimolar levels. Fig. 5*C* shows that our histone proteomic-ruler-derived copy numbers of members of the pyruvate dehydrogenase complex and the TRiC chaperone closely match the expected 1:1 stoichiometry among subunits.

The Muscle Proteome Is Quantitatively Dominated by Large, Abundant Proteins—As a practical example of the usefulness of "easy" absolute protein quantification, we determined cell sizes and cellular copy numbers of proteins in a panel of other mouse organs (Fig. 6*A*). Ovaries consist predominantly of small follicular cells and showed the least protein per cell (42 pg). Leg muscle cells, in contrast, had around 675 pg of protein per nucleus. Considering that muscle fibers are syncytial, multi-nucleated cells, the histone proteomic ruler delivered protein amounts per nucleus and not per cell in this particular case. Despite the huge differences in cellular protein amounts, we observed much less variation in the dependence of the abundance of a protein and its molecular mass, irrespective of the tissue of origin. This is reflected in the average molecular mass of a protein, which is calculated as the ratio of the total protein mass per cell to the total number of protein molecules (Fig. 6*B*). This number is rather similar across tissues, with the notable exception of muscle tissues. The reason for this becomes apparent when we look at the distribution of protein sizes across the dynamic range of the individual proteins (Figs. 6*C* and 6*D*). Independent of the tissue of origin, low-abundant proteins had an average molecular mass of around 100 kDa, and this value decreased
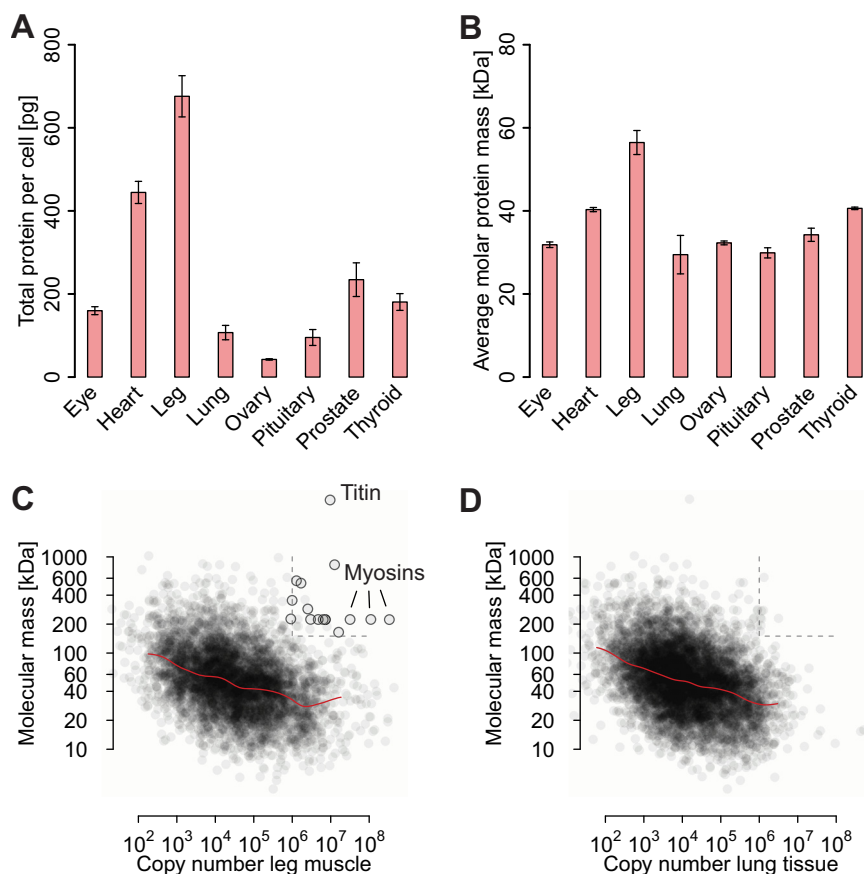
FIG. 6. **Application of the histone proteomic ruler to the global characterization of proteomes.** *A*, average total protein mass per cell. *B*, average molecular masses of proteins. Values represent the mean of three replicates ± S.D. *C*, *D*, abundant proteins tended to be smaller than low-abundance proteins. Motorproteins and filaments were notable exceptions in skeletal muscle.

with increasing cellular abundance of the proteins to around 40 kDa for the most abundant proteins. This dependence was observed in earlier studies and is thought to reflect the evolutionary advantage of decreasing the size of abundant proteins for reasons of biosynthetic cost (38). As a consequence of this trend, the average molecular mass of a protein in a cell is much smaller than the nominal average of the sizes of all proteins when their abundances are not taken into account. Notably, in skeletal muscle cells, filaments and motorproteins such as titin and myosins are notable exceptions to the trend of abundant proteins being smaller, as they are both large (>150 kDa) and very abundant (>1 million copies per cell) in this tissue, resulting in a profound increase in the average molecular protein mass in a muscle cell (Fig. 6*C*, circles).

Plugin for the Perseus Data Analysis Software for Calculation of Absolute Protein Abundances—The calculation of the protein abundances is a simple arithmetic task and can be performed using commonly available table calculation tools. To make the proteomic ruler approach easily usable for a wide community, we have implemented it as a plugin for the Perseus data analysis software. Perseus is part of the freely available MaxQuant suite (20). The proteomic ruler plugin supports all modes of label-free absolute quantification de-

scribed in this study and takes user-configurable variables such as the ploidy and the total protein concentration. Optionally, it can incorporate an additional level of protein-specific correction: our copy number calculation assumes a direct proportionality between a protein's cumulative mass in the proteomic sample and the MS signals summed over all peptides derived from it (see Eq. 3). Hence the protein's molar mass serves as a protein-specific normalization factor for copy number estimation. Because the combination of the sequence of a protein, the specificity of the protease used for digestion, and the characteristics of the mass spectrometric analysis can introduce protein-specific biases (39), our plugin allows the user to employ alternative normalization factors, such as the number of theoretically expected peptides that is used by some methods (9, 40).

In addition, we have implemented auxiliary functionalities. For instance, molecular weights and numbers of theoretical peptides can be calculated from protein I.D.s in combination with the FASTA database. Moreover, the plugin allows the categorization of proteins according to the expected accuracy of absolute quantification: proteins having a high fraction of theoretical peptides per sequence length and a high num-

ber of actually identified peptides, most of which are group-unique, are expected to yield better quantification.

DISCUSSION

In this paper, we propose that accurate absolute quantification is possible without the use of spike-in standards through the use of a concept we call the "proteomic ruler." Using the MS signal derived from histones and relating it to a known amount of DNA per cell provides accurate estimates of the total protein amount per cell that can be used as scaling factors for calculating cellular copy numbers of any protein of interest. We note that our approach makes a number of assumptions that allow us to omit any spike-in standards. At the same time, it eliminates several experimental steps such as cell counting and absolute protein concentration determination, which are themselves prone to errors, in particular stemming from issues with protein determination assays.

We found the quantitative results of our proteomic ruler approach to be typically within a factor of 2 of precision measurements or literature values. Importantly, this information comes for free, in that it incorporates absolute quantification into any kind of in-depth proteome dataset, even in retrospective analysis. The only prerequisite is a eukaryotic, whole-cell proteome dataset where the chromatin fraction is not over- or underrepresented as a result of sample handling. The latter is a specific requirement for an accurate estimation of the total protein mass per cell, but all whole proteome datasets should aim at an unbiased representation of all protein classes. A reasonable depth of proteomic analysis is needed to ensure a robust contribution of the histone MS signal, but the necessary depth should be readily attainable with many experimental setups. We expect that in the future, more and more proteomics projects will reach the required depth of proteome coverage and will be able to incorporate absolute quantification via the histone proteomic ruler. Additionally, individual protein copy numbers will become more accurate with increased peptide coverage in deep datasets.

Furthermore, we envision a generalization of the proteomic ruler concept beyond using the histone signal to estimate cellular protein amounts. For instance, using characteristic protein classes such as membrane or mitochondrial proteins, it should be possible to infer insights into subcellular architecture solely from proteomics datasets.

Ⓢ This article contains supplemental material.

¶ To whom correspondence should be addressed: E-mail: jwisniew@biochem.mpg.de; E-mail: mmann@biochem.mpg.de.

§ These authors contributed to this work equally.

REFERENCES

1. Beck, M., Claassen, M., and Aebersold, R. (2011) Comprehensive proteomics. *Curr. Opin. Biotechnol.* **22,** 3–8
2. Altelaar, A. F., Munoz, J., and Heck, A. J. (2013) Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat. Rev. Genet.* **14,** 35–48
3. Hein, M. Y., Sharma, K., Cox, J., and Mann, M. (2012) Proteomic analysis of cellular systems. *Handbook of Systems Biology*, pp. 3–25 A. J. Marian Walhout, Marc Vidal, Job Dekker (eds), Academic Press/Elsevier, London, UK
4. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* **422,** 198–207
5. Cox, J., Hein, M. Y., Luber, C. A., Paron, I., Nagaraj, N., and Mann, M. (2014) Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. *Mol. Cell. Proteomics* **13,** 2513–2526
6. Bantscheff, M., Lemeer, S., Savitski, M. M., and Kuster, B. (2012) Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal. Bioanal. Chem.* **404,** 939–965
7. Bork, P., and Serrano, L. (2005) Towards cellular systems in 4D. *Cell* **121,** 507–509
8. Malmstrom, J., Beck, M., Schmidt, A., Lange, V., Deutsch, E. W., and Aebersold, R. (2009) Proteome-wide cellular protein concentrations of the human pathogen Leptospira interrogans. *Nature* **460,** 762–765
9. Schwanhausser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011) Global quantification of mammalian gene expression control. *Nature* **473,** 337–342
10. Beck, M., Schmidt, A., Malmstroem, J., Claassen, M., Ori, A., Szymborska, A., Herzog, F., Rinner, O., Ellenberg, J., and Aebersold, R. (2011) The quantitative proteome of a human cell line. *Mol. Syst. Biol.* **7,** 549
11. Zeiler, M., Straube, W. L., Lundberg, E., Uhlen, M., and Mann, M. (2012) A protein epitope signature tag (PrEST) library allows SILAC-based absolute quantification and multiplexed determination of protein copy numbers in cell lines. *Mol. Cell. Proteomics* **11,** O111.009613
12. Wiechelman, K. J., Braun, R. D., and Fitzpatrick, J. D. (1988) Investigation of the bicinchoninic acid protein assay: identification of the groups responsible for color formation. *Anal. Biochem.* **175,** 231–237
13. Fountoulakis, M., Juranville, J. F., and Manneberg, M. (1992) Comparison of the Coomassie Brilliant Blue, bicinchoninic acid and Lowry quantitation assays, using non-glycosylated and glycosylated proteins. *J. Biochem. Biophys. Methods* **24,** 265–274
14. Noble, J. E., and Bailey, M. J. A. (2009) Quantitation of Protein. In Methods in Enzymology (Richard, R. B., and Murray, P. D., Eds.), pp. 73–95, Academic Press/Elsevier, London, UK
15. Crowe, P. T., and Marsh, M. N. (1993) Morphometric analysis of small intestinal mucosa. IV. Determining cell volumes. *Virchows Archive A Pathol. Anat. Histopathol.* **422,** 459–466
16. Wisniewski, J. R., Ostasiewicz, P., Dus, K., Zielinska, D. F., Gnad, F., and Mann, M. (2012) Extensive quantitative remodeling of the proteome between normal colon tissue and adenocarcinoma. *Mol. Syst. Biol.* **8,** 611
17. Wisniewski, J. R. (2013) Proteomic sample preparation from formalin fixed and paraffin embedded tissue. *J. Vis. Exp.* **79,** e50589
18. Wisniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009) Universal sample preparation method for proteome analysis. *Nat. Methods* **6,** 359–362
19. Wisniewski, J. R., Dus, K., and Mann, M. (2012) Proteomic workflow for analysis of archival formalin fixed and paraffin embedded clinical samples to a depth of 10,000 proteins. *Proteomics. Clin. Applicat*. **7,** 225–233
20. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26,** 1367–1372
21. Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., and Mann, M. (2011) Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10,** 1794–1805
22. Kratz, A., Ferraro, M., Sluss, P. M., and Lewandrowski, K. B. (2004) Case records of the Massachusetts General Hospital. Weekly clinicopatholog-

ical exercises. Laboratory reference values. *N. Engl. J. Med.* **351,** 1548–1563

23. Papadimitriou, E., and Lelkes, P. I. (1993) Measurement of cell numbers in microtiter culture plates using the fluorescent dye Hoechst 33258. *J. Immunol. Methods* **162,** 41–45

24. Shimada, H., Obayashi, T., Takahashi, N., Matsui, M., and Sakamoto, A. (2010) Normalization using ploidy and genomic DNA copy number allows absolute quantification of transcripts, proteins and metabolites in cells. *Plant Methods* **6,** 29

25. Silva, L. P., Lorenzi, P. L., Purwaha, P., Yong, V., Hawke, D. H., and Weinstein, J. N. (2013) Measurement of DNA concentration as a normalization strategy for metabolomic data from adherent cell lines. *Anal. Chem.* **85,** 9536–9542

26. Zhou, H., Ning, Z., Wang, F., Seebun, D., and Figeys, D. (2011) Proteomic reactors and their applications in biology. *FEBS J.* **278,** 3796–3806

27. Wisniewski, J. R., and Mann, M. (2012) Consecutive proteolytic digestion in an enzyme reactor increases depth of proteomic and phosphoproteomic analysis. *Anal. Chem.* **84,** 2631–2637

28. International Human Genome Sequencing C. (2004) Finishing the euchromatic sequence of the human genome. *Nature* **431,** 931–945

29. van Holde, K. E. (1989) *Chromatin*, Springer Verlag, New York

30. Nagaraj, N., Alexander Kulak, N., Cox, J., Neuhauser, N., Mayr, K., Hoerning, O., Vorm, O., and Mann, M. (2012) System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra HPLC runs on a bench top Orbitrap. *Mol. Cell. Proteomics* **11,** M111.013722

31. Warner, J. R. (1999) The economics of ribosome biosynthesis in yeast. *Trends Biochem. Sci.* **24,** 437–440

32. Melnikov, S., Ben-Shem, A., Garreau de Loubresse, N., Jenner, L., Yusupova, G., and Yusupov, M. (2012) One core, two shells: bacterial and eukaryotic ribosomes. *Nat. Struct. Mol. Biol.* **19,** 560–567

33. Brown, G. C. (1991) Total cell protein concentration as an evolutionary constraint on the metabolic control distribution in cells. *J. Theor. Biol.* **153,** 195–203

34. Lodish, H., Berk, A., Zipursky, S., Matsudaira, P., Baltimore, D., and Darnell, J., Eds. (2000), *Molecular Cell Biology* 4th ed., W.H. Freeman, New York

35. Salinas, F. A., Smith, L. H., and Goodman, J. W. (1972) Cell size distribution in the thymus as a function of age. *J. Cell. Physiol.* **80,** 339–345

36. Luber, C. A., Cox, J., Lauterbach, H., Fancke, B., Selbach, M., Tschopp, J., Akira, S., Wiegand, M., Hochrein, H., O'Keeffe, M., and Mann, M. (2010) Quantitative proteomics reveals subset-specific viral recognition in dendritic cells. *Immunity* **32,** 279–289

37. Dumortier, H., van Mierlo, G. J., Egan, D., van Ewijk, W., Toes, R. E., Offringa, R., and Melief, C. J. (2005) Antigen presentation by an immature myeloid dendritic cell line does not cause CTL deletion in vivo, but generates CD8+ central memory-like T cells that can be rescued for full effector function. *J. Immunol.* **175,** 855–863

38. Warringer, J., and Blomberg, A. (2006) Evolutionary constraints on yeast protein size. *BMC Evolutionary Biol.* **6,** 61

39. Peng, M., Taouatas, N., Cappadona, S., van Breukelen, B., Mohammed, S., Scholten, A., and Heck, A. J. (2012) Protease bias in absolute protein quantitation. *Nat. Methods* **9,** 524–525

40. Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., and Mann, M. (2005) Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol. Cell. Proteomics* **4,** 1265–1272

41. James, M. M., Verhofste, M., Franklin, C., Beilman, G., and Goldman, C. (2010) Dissection of the left main coronary artery after blunt thoracic trauma: case report and literature review. *World J. Emerg. Surg.* **5,** 21